Noname manuscript No.
(will be inserted by the editor)

Zig-zag exploratory factor analysis with more variables than observations

Steffen Unkel · Nickolay T. Trendafilov

Received: date / Accepted: date

Abstract In this paper, the problem of fitting the exploratory factor analysis (EFA) model to data matrices with more variables than observations is reconsidered. A new algorithm named 'zig-zag EFA' is introduced for the simultaneous least squares estimation of all EFA model unknowns. As in principal component analysis, zig-zag EFA is based on the singular value decomposition of data matrices. Another advantage of the proposed computational routine is that it facilitates the estimation of both common and unique factor scores. Applications to both real and artificial data illustrate the algorithm and the EFA solutions.

Keywords Factor analysis · Horizontal data matrices · Procrustes problems · Singular value decomposition · Gene expression data · Thurstone's box data

1 Setting the scene

This paper reconsiders the problem of fitting the exploratory factor analysis (EFA) model (Harman 1976; Mulaik 2010) to a data matrix. Let $\mathbf{Z} \in \mathbb{R}^{n \times p}$ be a given data matrix of p manifest variables on n cases ($n > p$). To facilitate the notation, assume that the columns of $\mathbf{Z}$ are mean-centered and scaled to

Dr. Steffen Unkel
Department of Mathematics and Statistics, Faculty of Mathematics, Computing and Technology, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
Tel.: +44-1908-652703
Fax: +44-1908-655515
E-mail: S.Unkel@open.ac.uk

Dr. Nickolay T. Trendafilov
Department of Mathematics and Statistics, Faculty of Mathematics, Computing and Technology, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
Tel.: +44-1908-652030
Fax: +44-1908-655515
E-mail: N.Trendafilov@open.ac.uk

have unit length, that is, the columns have variance equal to $1/(n-1)$.
Suppose that the EFA model holds which states that $\mathbf{Z}$ can be written in the form (Mulaik 2010, p. 136):

$$\mathbf{Z} = \mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{U}\boldsymbol{\Psi} \ , \tag{1}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times k}$ is a matrix of factor loadings with $\text{rank}(\boldsymbol{\Lambda}) = k$ ($k \ll p$), $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$ is a diagonal matrix of uniquenesses; the columns of $\mathbf{F} \in \mathbb{R}^{n\times k}$ are called common factors and those of the matrix $\mathbf{U} \in \mathbb{R}^{n\times p}$ unique factors. The choice of k in EFA is subject to some limitations (Mulaik 2010, p. 174), which will not be discussed here. Assume that the columns of $\mathbf{F}$ and $\mathbf{U}$ have a mean of zero and are scaled to have unit norm. Furthermore, let $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$ and $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$, where $\mathbf{I}_p$ is an identity matrix of order p. Hence, all common factors and unique factors are mutually uncorrelated (orthogonal). Finally, suppose that the unique factors are orthogonal to the common factors, that is, $\mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k}$, where $\mathbf{O}_{p\times k}$ is a $p \times k$ matrix of zeros.
The idea behind the model (1) is that the common factors account for the correlation structure among the set of manifest variables, while each unique factor corresponds to that portion of a particular observed variable that cannot be accounted for by the common factors. As such, a unique factor contains the specificity of that variable as well as errors in measurement.
To fit the EFA model (1) the classical approach is to find the pair $\{\boldsymbol{\Lambda}, \boldsymbol{\Psi}\}$ which gives the best fit, for some specified value of k, of the well-known EFA correlation structure:

$$\mathbf{R} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2 \tag{2}$$

to the sample correlation matrix $\mathbf{Z}^\top\mathbf{Z}$ with respect to some discrepancy measure. The process of finding this pair is called *factor extraction*. Various factor extraction methods have been proposed (Harman 1976; Mulaik 2010). The most frequently used goodness-of-fit measures are the maximum likelihood (ML) and least squares (LS) (weighted and unweighted) for which a number of iterative algorithms are available (e.g., Bartholomew and Knott 1999; Jöreskog 1977).
Suppose that a pair $\{\boldsymbol{\Lambda}, \boldsymbol{\Psi}\}$ is obtained by solving the factor extraction problem. Then, common factor scores can be computed as a function of $\mathbf{Z}$, $\boldsymbol{\Lambda}$ and possibly $\boldsymbol{\Psi}$ in a number of ways (see Harman 1976; Mulaik 2010), for example by means of the formula proposed by Anderson and Rubin (1956) which satisfies the correlation preserving constraint $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$.
In formulating EFA models, the standard approach is to embed the data in a replication framework by assuming the observations are realizations of random variables. Instead, De Leeuw (2004, 2008) formulated the EFA model (1) as a specific data matrix decomposition with fixed unknown matrix parameters. Let $||\mathbf{X}||_F = \sqrt{\text{trace}(\mathbf{X}^\top\mathbf{X})}$ denote the Frobenius norm of $\mathbf{X}$. De Leeuw (2004, 2008) solved the EFA fitting problem by minimizing the following LS goodness-of-fit criterion:

$$\mathcal{F}(\mathbf{F}, \boldsymbol{\Lambda}, \mathbf{U}, \boldsymbol{\Psi}) = ||\mathbf{Z} - \mathbf{F}\boldsymbol{\Lambda}^\top - \mathbf{U}\boldsymbol{\Psi}||_F^2 \ , \tag{3}$$

subject to

$$\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k,\ \mathbf{U}^\top\mathbf{U} = \mathbf{I}_p,\ \mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k} \text{ and } \boldsymbol{\Psi} \text{ diagonal } . \tag{4}$$

De Leeuw (2004, 2008) proposed an alternating least squares (ALS) algorithm to minimize (3) subject to (4). Note that the approach by De Leeuw (2004, 2008) is equivalent to a method developed by Henk A. L. Kiers in some unpublished notes (H. A. L. Kiers, personal communication, 2009). Sočan (2003) called this approach 'Direct-simple factor analysis' and gives a description in some detail.

The approach of De Leeuw (2004, 2008) is designed for the classical case of 'vertical' data with $n > p$, though. In many modern applications, the number of available observations is less than the number of variables (see Hastie et al. 2009, Chapter 18). Consider for example data arising in atmospheric science, where a meteorological variable is measured at p spatial locations at n different points in time. Typically, these data are high-dimensional with $p >> n$.

Trendafilov and Unkel (2010) extended the approach of De Leeuw (2004, 2008) to EFA for horizontal data matrices with $p \geq n$. New assumptions are imposed on the EFA model parameters which necessarily require the acceptance of unique factors having zero variance. In the proposed numerical procedure, and update for the common and unique factor scores $\mathbf{F}$ and $\mathbf{U}$ is found simultaneously by solving an augmented Procrustes problem (for given or estimated $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$).

In this paper, a new iterative routine for the *simultaneous* estimation of *all* EFA model unknowns is presented, which updates $\mathbf{F}$ and $\mathbf{U}$ *successively*. The new computational technique is named 'zig-zag EFA'. As in principal component analysis (PCA) (Jolliffe 2002), zig-zag EFA is based on the well-known and computationally efficient numerical procedure for singular value decomposition (SVD) (Golub and Van Loan 1996) of data matrices.

The present paper is organized as follows. The next Section briefly reviews the approach of Trendafilov and Unkel (2010) for fitting the EFA model to horizontal data matrices. The new zig-zag routine is presented in Section 3. Section 4 discusses convergence issues of zig-zag EFA. In Section 5, the new method is applied to a real high-dimensional data set from genome research (Subsection 5.1) and artificial data from psychometrics (Subsection 5.2). Concluding comments are given in Section 6.

2 EFA of horizontal data matrices

It would be helpful to recall briefly some notation from Trendafilov and Unkel (2010). Recall the loss function (3) (De Leeuw 2004, 2008):

$$\mathcal{F}(\mathbf{F}, \boldsymbol{\Lambda}, \mathbf{U}, \boldsymbol{\Psi}) = ||\mathbf{Z} - \mathbf{F}\boldsymbol{\Lambda}^\top - \mathbf{U}\boldsymbol{\Psi}||_F^2 \ .$$

In the sequel, it is assumed that $p \geq n$. If the number of variables exceeds the number of observations, the classical EFA constraint $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$ cannot

be fulfilled because $\mathbf{U}$ has at most n linearly independent columns ($< p$). In fact, since the constraints $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$ and $\mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k}$ remain valid, $\text{rank}(\mathbf{U}) \leq n - k$. Trendafilov and Unkel (2010) introduce the new constraint $\mathbf{U}^\top\mathbf{U}\boldsymbol{\Psi} = \boldsymbol{\Psi}$ in order to keep the model correlation structure (2) unchanged. This new constraint in turn implies $\text{rank}(\boldsymbol{\Psi}) \leq n - k$, that is, unique factors having zero variances and hence $\boldsymbol{\Psi}^2$ being positive semi-definite should be accepted in the EFA model (1).
Trendafilov and Unkel (2010) proceed with a proof that if $p \geq n$, the EFA model constraints $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$ and $\mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k}$ are equivalent to the constraints $\mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n$ and $\text{rank}(\mathbf{F}) = k$. Consequently, the approach of Trendafilov and Unkel (2010) requires minimization of (3) subject to

$$\text{rank}(\mathbf{F}) = k,\ \mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n,\ \text{and}\ \mathbf{U}^\top\mathbf{U}\boldsymbol{\Psi} = \boldsymbol{\Psi}\ \text{diagonal}. \tag{5}$$

Define $\mathbf{B} := [\mathbf{F}\ \ \mathbf{U}]$ and $\mathbf{A} := [\boldsymbol{\Lambda}\ \ \boldsymbol{\Psi}]$ as block matrices with dimensions $n \times (k + p)$ and $p \times (k + p)$, respectively. Trendafilov and Unkel (2010) show that for given or estimated $\mathbf{A}$, the following optimization problem:

$$\min_{\mathbf{B}} \left|\left|\mathbf{Z} - \mathbf{B}\mathbf{A}^\top\right|\right|_F^2 , \tag{6}$$

subject to

$$\mathbf{B}\mathbf{B}^\top = \begin{bmatrix}\mathbf{F} & \mathbf{U}\end{bmatrix}\begin{bmatrix}\mathbf{F}^\top \\ \mathbf{U}^\top\end{bmatrix} = \mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n , \tag{7}$$

is a Procrustes problem (Gower and Dijksterhuis 2004), that is, its solution is found by maximizing $\text{trace}(\mathbf{B}^\top\mathbf{Z}\mathbf{A})$ and can be computed via the compact SVD (e.g., ten Berge 1993, pp. 4-5) of $\mathbf{A}^\top\mathbf{Z}^\top$. Unfortunately, since $\mathbf{A}^\top\mathbf{Z}^\top$ is rank deficient, the maximum is attained for a set of matrices and one can choose any of its elements to find $\mathbf{B}$ (see the appendix in Trendafilov and Unkel 2010). The fact that the maximizing solution of the Procrustes problem above is not unique is closely related to the factor score indeterminacy problem associated with the EFA model (cf. Section 3).
After solving (6) for $\mathbf{B} = [\mathbf{F}\ \ \mathbf{U}]$, $\mathbf{A} = [\boldsymbol{\Lambda}\ \ \boldsymbol{\Psi}]$ is updated making use of the following identities:

$$\mathbf{F}^\top\mathbf{Z} = \mathbf{F}^\top\mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{F}^\top\mathbf{U}\boldsymbol{\Psi} = \boldsymbol{\Lambda}^\top, \tag{8}$$

$$\mathbf{U}^\top\mathbf{Z} = \mathbf{U}^\top\mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{U}^\top\mathbf{U}\boldsymbol{\Psi} = \boldsymbol{\Psi}\ (\text{and thus diagonal}) , \tag{9}$$

which follow from the EFA model (1) and the imposed constraints. The ALS procedure of finding $\{\mathbf{F}, \mathbf{U}\}$ and $\{\boldsymbol{\Lambda}, \boldsymbol{\Psi}\}$ continues until a pre-specified convergence criterion is met.

3 Zig-zag routine

Another way to find $\mathbf{F}$ and $\mathbf{U}$ such that $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$ and $\mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k}$ is by updating $\mathbf{F}$ and $\mathbf{U}$ successively. The first step of zig-zag EFA is as follows:

(i) for given $\boldsymbol{\Lambda}, \boldsymbol{\Psi}$ and $\mathbf{U}$, find $\mathbf{F}$ that minimizes $\left|\left|(\mathbf{Z}-\mathbf{U}\boldsymbol{\Psi})-\mathbf{F}\boldsymbol{\Lambda}^\top\right|\right|_F^2$,
subject to $\mathbf{F}^\top\mathbf{F}=\mathbf{I}_k$.

The optimization problem (i) is again a standard Procrustes problem, i.e. its solution is found by maximizing $\text{trace}(\mathbf{F}^\top(\mathbf{Z}-\mathbf{U}\boldsymbol{\Psi})\boldsymbol{\Lambda})$.
Next, for available $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$ and $\mathbf{F}$, an update of $\mathbf{U}$ is found by solving

$$\min ||(\mathbf{Z}-\mathbf{F}\boldsymbol{\Lambda}^\top)-\mathbf{U}\boldsymbol{\Psi}||_F^2 \ , \tag{10}$$

subject to

$$\mathbf{U}^\top\mathbf{F}=\mathbf{O}_{p\times k} \ . \tag{11}$$

To find an update of $\mathbf{U}$, note that the constraint (11) requires that the columns of $\mathbf{U}$ should be in the nullspace of $\mathbf{F}$ in $\mathbb{R}^n$. Let $\mathbf{F}_\perp$ be the $n\times(n-k)$ matrix containing an orthonormal basis of the nullspace of $\mathbf{F}$ in $\mathbb{R}^n$, where $\mathbf{F}_\perp$ can be found by QR factorization (Golub and Van Loan 1996) of $\mathbf{F}$:

$$\mathbf{F}=\mathbf{Q}\mathbf{R}=\underbrace{\begin{bmatrix}\mathbf{F} & \mathbf{F}_\perp\end{bmatrix}}_{\mathbf{Q}}\underbrace{\begin{bmatrix}\mathbf{I}_k\\ \mathbf{O}_{(n-k)\times k}\end{bmatrix}}_{\mathbf{R}} , \tag{12}$$

where $\mathbf{Q}\in\mathbb{R}^{n\times n}$ is orthogonal and $\mathbf{R}\in\mathbb{R}^{n\times k}$ is upper triangular.
Then, there exists a full row-rank $(n-k)\times p$ matrix $\widetilde{\mathbf{U}}$ such that $\mathbf{U}=\mathbf{F}_\perp\widetilde{\mathbf{U}}$. As $\widetilde{\mathbf{U}}$ can be chosen row-wise orthonormal, i.e. $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^\top=\mathbf{I}_{n-k}$, it follows that $\mathbf{U}\mathbf{U}^\top=\mathbf{F}_\perp\mathbf{F}_\perp^\top$, and $\mathbf{F}\mathbf{F}^\top+\mathbf{F}_\perp\mathbf{F}_\perp^\top=\mathbf{I}_n$ implies the new constraint $\mathbf{F}\mathbf{F}^\top+\mathbf{U}\mathbf{U}^\top=\mathbf{I}_n$. By making use of (12), the objective function in (10) can be rewritten as:

$$\begin{aligned}
||(\mathbf{Z}-\mathbf{F}\boldsymbol{\Lambda}^\top)-\mathbf{U}\boldsymbol{\Psi}||_F^2 &= ||\mathbf{Q}^\top(\mathbf{Z}-\mathbf{F}\boldsymbol{\Lambda}^\top)-\mathbf{Q}^\top\mathbf{U}\boldsymbol{\Psi}||^2||_F\\
&= \left|\left|\begin{bmatrix}\mathbf{F}^\top\mathbf{Z}-\mathbf{F}^\top\mathbf{F}\boldsymbol{\Lambda}^\top-\mathbf{F}^\top\mathbf{U}\boldsymbol{\Psi}\\ \mathbf{F}_\perp^\top\mathbf{Z}-\mathbf{F}_\perp^\top\mathbf{F}\boldsymbol{\Lambda}^\top-\mathbf{F}_\perp^\top\mathbf{U}\boldsymbol{\Psi}\end{bmatrix}\right|\right|_F^2\\
&= \left|\left|\begin{bmatrix}\mathbf{F}^\top\mathbf{Z}-\boldsymbol{\Lambda}^\top\\ \mathbf{F}_\perp\mathbf{Z}-\widetilde{\mathbf{U}}\boldsymbol{\Psi}\end{bmatrix}\right|\right|_F^2\\
&= ||\mathbf{F}^\top\mathbf{Z}||_F^2+||\mathbf{F}_\perp^\top\mathbf{Z}||_F^2+||\boldsymbol{\Lambda}||_F^2-2\,\text{trace}(\mathbf{F}^\top\mathbf{Z}\boldsymbol{\Lambda})\\
&\quad +||\boldsymbol{\Psi}\widetilde{\mathbf{U}}^\top||_F^2-2\,\text{trace}(\mathbf{F}_\perp^\top\mathbf{Z}\boldsymbol{\Psi}\widetilde{\mathbf{U}}^\top) \ ,
\end{aligned} \tag{13}$$

assuming $\mathbf{F}^\top\mathbf{F}=\mathbf{I}_k$ and $\mathbf{U}^\top\mathbf{F}=\mathbf{O}_{p\times k}$. Note that the term $||\boldsymbol{\Psi}\widetilde{\mathbf{U}}^\top||_F^2$ is constant if the new constraint $\mathbf{U}^\top\mathbf{U}\boldsymbol{\Psi}=\boldsymbol{\Psi}$ holds. Indeed:

$$\begin{aligned}
\text{trace}(\boldsymbol{\Psi}\tilde{\mathbf{U}}^\top\tilde{\mathbf{U}}\boldsymbol{\Psi}) &= \text{trace}(\boldsymbol{\Psi}\mathbf{U}^\top\mathbf{F}_\perp\mathbf{F}_\perp^\top\mathbf{U}\boldsymbol{\Psi})\\
&= \text{trace}(\boldsymbol{\Psi}\mathbf{U}^\top\mathbf{U}\mathbf{U}^\top\mathbf{U}\boldsymbol{\Psi})=\text{trace}(\boldsymbol{\Psi}^2) \ ,
\end{aligned}$$

which shows that $\text{trace}(\boldsymbol{\Psi}\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{U}}\boldsymbol{\Psi})$ does not depend on $\widetilde{\mathbf{U}}$. Thus, the problem (10) is equivalent to the following Procrustes problem which constitutes the second step of the zig-zag algorithm:

(ii) for given $\boldsymbol{\Lambda}, \boldsymbol{\Psi}$ and $\mathbf{F}$, find $\widetilde{\mathbf{U}}$ that minimizes $\left|\left|\mathbf{F}_{\perp}^{\top}\mathbf{Z} - \widetilde{\mathbf{U}}\boldsymbol{\Psi}\right|\right|_F^2$,
subject to $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^{\top} = \mathbf{I}_{n-k}$.

The problem (ii) requires the maximization of $\mathrm{trace}(\boldsymbol{\Psi}\mathbf{Z}^{\top}\mathbf{F}_{\perp}\widetilde{\mathbf{U}})$ which can be solved analytically via the SVD of $\boldsymbol{\Psi}\mathbf{Z}^{\top}\mathbf{F}_{\perp}$. After solving this Procrustes problem, the original $\mathbf{U}$ is computed as $\mathbf{U} = \mathbf{F}_{\perp}\widetilde{\mathbf{U}}$. The final step of zig-zag EFA is then:

(iii) for given $\mathbf{F}$ and $\mathbf{U}$, find $\boldsymbol{\Lambda} = \mathbf{Z}^{\top}\mathbf{F}$ and $\boldsymbol{\Psi} = \mathrm{diag}(\mathbf{U}^{\top}\mathbf{Z})$.

The alternating procedure (i) – (iii) is continued until a pre-specified convergence criterion is met. Summarizing, the zig-zag EFA method described by means of user-friendly MATLAB code (The MathWorks 2008) is as follows:

```
% Initializing common factor scores F and unique factor scores U
[B,RB] = qr(rand(p+k,n)-.5,0); idx=find(diag(RB)<=0) ;
B(:,idx) = -B(:,idx) ; B=B' ;
F = B(:,1:k); U = B(:,k+1:p+k) ;

% Initializing loadings Lambda and uniquenesses Psi
Lambda = Z'*F ;
Psi = diag(U'*Z) ;

f = norm(Z-F*Lambda'-U*diag(Psi),'fro')^2 ; % Objective function
f0 = norm(Z,'fro')^2 ;

% Main loop for ALS algorithm
while abs(f0-f) > 1e-6  % Convergence criterion

% Solve Procrustes problem to update F
[U0,S0,V0] = svd((Z-U*diag(Psi))*Lambda,0) ;
F = U0*V0' ;

% Solve reduced Procrustes-like problem for Utilde, such that
% Utilde*Utilde' = eye(n-k) and hence F*F'+U*U' = eye(n)
[Q,upper_t] = qr(F) ;
Fbot = Q(:,k+1:n) ;

[y1,dy,y2] = svd(diag(Psi)*Z'*Fbot,0) ;
Utilde = y2*y1';

% Update U
U = Fbot*Utilde ;

% Update Lambda and Psi
Lambda = Z'*F ;
Psi = diag(U'*Z) ;
```

```
% Evaluating the objective function
f0 = f;
f = norm(Z-F*Lambda'-U*diag(Psi),'fro')^2 ;
end % End of main loop
```

The EFA model suffers from two indeterminacies: *rotational indeterminacy* and *factor score indeterminacy* (e.g., Mulaik 2010). To start with the former, the factor loadings $\boldsymbol{\Lambda}$ can be any $p \times k$ matrix of full column rank. For example, any matrix $\boldsymbol{\Lambda}\mathbf{T}$, where $\mathbf{T}$ is an arbitrary $k \times k$ orthogonal matrix, gives the same model fit if one compensates for this rotation in the common factor scores $\mathbf{F}$. The assumptions about the variables that make up the original model are not violated by this transformation. This means that $\boldsymbol{\Lambda}$ can only be determined up to an orthogonal rotation and there is an infinite number of factor loadings satisfying the original assumptions of the model. In other words, the parameters of the EFA model cannot be identified uniquely.
Adding suitable constraints, such as requiring $\boldsymbol{\Lambda}^\top\boldsymbol{\Lambda}$ or $\boldsymbol{\Lambda}^\top\boldsymbol{\Psi}^{-2}\boldsymbol{\Lambda}$ to be diagonal (Jöreskog 1977), eliminates the rotational indeterminacy of the EFA model. This indeterminacy-elimination feature is not always helpful, because such solutions are usually difficult to interpret. Instead, the parameter estimation is usually followed by some kind of rotation such that a certain simplicity criterion (e.g. Varimax) is optimized (Browne 2001).
The present paper promotes a different solution to circumvent this roatational indeterminacy problem and to enhance interpretation of the loadings. The property $\text{rank}(\boldsymbol{\Lambda}) = k$ can be accomplished by having $\boldsymbol{\Lambda}$ in the form of a $p \times k$ lower triangular matrix $\mathbf{L}$, with a triangle of $k(k-1)/2$ zeros (Anderson and Rubin 1956; Trendafilov 2005). Such a reparameterization removes the rotational ambiguity of the EFA model. The only indeterminacy left are simultaneous changes of the signs of $\mathbf{L}$ and $\mathbf{F}$, which do not alter their interpretation and the model fit. Consider the following slightly modified EFA loss function:

$$\mathcal{F}(\mathbf{L}, \boldsymbol{\Psi}, \mathbf{F}, \mathbf{U}) = \left|\left|\mathbf{Z} - [\mathbf{F} \quad \mathbf{U}] \begin{bmatrix} \mathbf{L}^\top \\ \boldsymbol{\Psi} \end{bmatrix}\right|\right|_F^2, \tag{14}$$

which should be minimized subject to the constraints (5). Then, the updating formula $\boldsymbol{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ should simply be replaced by $\mathbf{L} = \texttt{tril}(\mathbf{Z}^\top\mathbf{F})$, where `tril()` is the operator taking the lower triangular part of its argument, that is, $\texttt{tril}(\mathbf{Z}^\top\mathbf{F})$ is composed of the elements of $\mathbf{Z}^\top\mathbf{F}$ with the upper triangle of $k(k-1)/2$ elements replaced by zeros. By doing so, one obtains an alternative solution of zig-zag EFA.
The EFA model postulates the existence of k common and p unique factors such that the p observed variables can be represented as their linear combinations. Thus, the scores of the n observations on the common and unique factors are not uniquely identifiable. This form of indeterminacy is known as factor score indeterminacy (Mulaik 2010). Due to the factor score indeterminacy, the

common and unique factors will not be uniquely identified, even if $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ are. In other words, there exists an infinite set of distinct variables that one can construct as having the same pattern of correlations between the observed variables as do the common and unique factors for a given EFA solution.
Despite the factor score indeterminacy, the non-uniqueness of the common and unique factor scores is not a problem for a numerical procedure that finds estimates for all EFA model unknowns simultaneously. In this respect, zig-zag EFA avoids the conceptual problem with the factor score indeterminacy and facilitates the estimation of both common and unique factor scores.

4 Convergence issues and non-uniqueness of estimates

Let the EFA matrix parameters $\{\mathbf{F}, \boldsymbol{\Lambda}, \mathbf{U}, \boldsymbol{\Psi}\}$ indexed with the subscript $_0$ and $_+$ denote their current and updated values, respectively. The update $\mathbf{F}_+$ found by means of the first step (i) of zig-zag EFA (cf. Section 3) reduces the EFA objective function (3) as follows:

$$\begin{aligned}\mathcal{F}(\mathbf{F}_0, \boldsymbol{\Lambda}_0, \mathbf{U}_0, \boldsymbol{\Psi}_0) &= ||\mathbf{Z} - \mathbf{F}_0\boldsymbol{\Lambda}_0^\top - \mathbf{U}_0\boldsymbol{\Psi}_0||_F^2 = ||(\mathbf{Z} - \mathbf{U}_0\boldsymbol{\Psi}_0) - \mathbf{F}_0\boldsymbol{\Lambda}_0^\top||_F^2 \\ &\geq ||(\mathbf{Z} - \mathbf{U}_0\boldsymbol{\Psi}_0) - \mathbf{F}_+\boldsymbol{\Lambda}^\top||_F^2 \\ &= \mathcal{F}(\mathbf{F}_+, \boldsymbol{\Lambda}_0, \mathbf{U}_0, \boldsymbol{\Psi}_0) \ , \end{aligned} \tag{15}$$

where $\mathbf{F}_+^\top\mathbf{F}_+ = \mathbf{I}_k$. The solution (update) $\mathbf{F}_+$ is global and is given by the SVD of the $n \times k$ matrix $(\mathbf{Z} - \mathbf{U}_0\boldsymbol{\Psi}_0)\boldsymbol{\Lambda}_0$. If this matrix has full column rank k, the global solution $\mathbf{F}_+$ is unique (Gower and Dijksterhuis 2004).
The solution (update) $\mathbf{U}_+$ found by means of the second step (ii) of zig-zag EFA (cf. Section 3) reduces the EFA objective function (15) as follows:

$$\begin{aligned}\mathcal{F}(\mathbf{F}_+, \boldsymbol{\Lambda}_0, \mathbf{U}_0, \boldsymbol{\Psi}_0) &= ||(\mathbf{Z} - \mathbf{F}_+\boldsymbol{\Lambda}^\top) - \mathbf{U}_0\boldsymbol{\Psi}_0||_F^2 \\ &\geq ||(\mathbf{Z} - \mathbf{F}_+\boldsymbol{\Lambda}^\top) - \mathbf{U}_+\boldsymbol{\Psi}_0||_F^2 \\ &= \mathcal{F}(\mathbf{F}_+, \boldsymbol{\Lambda}_0, \mathbf{U}_+, \boldsymbol{\Psi}_0) \ , \end{aligned} \tag{16}$$

where $\mathbf{U}_+^\top\mathbf{F} = \mathbf{O}_{p\times k}$. The solution (update) $\mathbf{U}_+$ is global. Since $\boldsymbol{\Psi}\mathbf{Z}^\top\mathbf{F}_\perp$ has rank (at most) $n - k < p$, the update $\mathbf{U}_+$ is obtained by means of the SVD of a rank deficient matrix. Hence, the global solution $\mathbf{U}+$ is not unique.
Further reduction of (16) is achieved by choosing $\boldsymbol{\Lambda}_+$ that minimizes $\mathbf{Z}-\mathbf{F}_+\boldsymbol{\Lambda}^\top$ in (16), that is, $\boldsymbol{\Lambda}_+ = \mathbf{Z}^\top\mathbf{F}$ and thus:

$$\begin{aligned}\mathcal{F}(\mathbf{F}_+, \boldsymbol{\Lambda}_0, \mathbf{U}_+, \boldsymbol{\Psi}_0) &= ||(\mathbf{Z} - \mathbf{F}_+\boldsymbol{\Lambda}_0^\top) - \mathbf{U}_+\boldsymbol{\Psi}_0||_F^2 \\ &\geq ||(\mathbf{Z} - \mathbf{F}_+\mathbf{F}_+^\top\mathbf{Z}) - \mathbf{U}_+\boldsymbol{\Psi}_0||_F^2 \\ &= \mathcal{F}(\mathbf{F}_+, \boldsymbol{\Lambda}_+, \mathbf{U}_+, \boldsymbol{\Psi}_0) \ . \end{aligned} \tag{17}$$

Since $\boldsymbol{\Lambda}_+ = \mathbf{Z}_+^\top\mathbf{F}$ has full column rank k, it is a global minimizer at each step. Because of the rotational indeterminacy of the EFA model, this global minimizer is not unique. If one uses $\boldsymbol{\Lambda}_+ = \texttt{tril}(\mathbf{Z}^\top\mathbf{F})$ instead, then the update $\boldsymbol{\Lambda}_+$ is a unique minimizer as the subspace of all lower triangular $p \times k$ matrices

with rank k is convex.
Finally, an update of $\boldsymbol{\Psi}$ is needed which further reduces (17). Trendafilov and Unkel (2010) show that the update $\boldsymbol{\Psi}_+ = \text{diag}(\mathbf{Z}^\top \mathbf{U})$ is a global and unique minimizer at each step and leads to the following reduction of (17):

$$\begin{aligned} \mathcal{F}(\mathbf{F}_+, \boldsymbol{\Lambda}_+, \mathbf{U}_+, \boldsymbol{\Psi}_0) &\geq ||(\mathbf{Z} - \mathbf{F}_+\mathbf{F}_+^\top\mathbf{Z}) - \mathbf{U}_+\text{diag}(\mathbf{U}_+^\top\mathbf{Z})||_F^2 \\ &= ||\mathbf{Z} - \mathbf{F}_+\boldsymbol{\Lambda}_+^\top - \mathbf{U}_+\boldsymbol{\Psi}_+||_F^2 \\ &= \mathcal{F}(\mathbf{F}_+, \boldsymbol{\Lambda}_+, \mathbf{U}_+, \boldsymbol{\Psi}_+) \ , \end{aligned} \tag{18}$$

imposing the new constraint $\mathbf{U}_+^\top\mathbf{U}_+\boldsymbol{\Psi} = \boldsymbol{\Psi}$ being a diagonal matrix.
Thus, zig-zag EFA decreases the EFA objective function (3) monotonically and globally, i.e. convergence to a (local) minimizer is reached independently of the initial states. However, the estimated EFA parameters are *not* unique. Finally, it can be shown that the zig-zag routine is a specific gradient descent method. Thus, the ALS process has linear rate of convergence, as any other gradient method.

5 Applications

In the sequel, zig-zag EFA is applied to high-dimensional microarray gene expression data in Subsection 5.1 and a famous artificial data set from psychometrics in Subsection 5.2. Computations are carried out using the software package MATLAB R2009a (The MathWorks 2008) on a PC under the Windows XP operating system with an Intel Pentium 4 CPU having 2.4 GHz clock frequency and 1 GB of RAM. All computer code used in the examples is available upon request.

5.1 Analysis of gene expression data

The DNA (deoxyribonucleic acid) microarray technology facilitates the quantitative study of thousands of genes simultaneously and is being applied in cancer research. A DNA is the basic material that makes up human chromosomes. The DNA microarrays measure the expression of a gene in a cell by measuring the amount of mRNA (messenger ribonucleic acid) present for that gene. The nucleotide sequences for a few thousand genes are printed on a glass slide. A target sample and a reference sample are labeled with red and green dyes, and each are hybridized with the DNA on the slide. Through fluoroscopy, the log (red/green) intensities of RNA hybridizing at each site is measured. The result is a few thousand numbers, measuring the expression level of each gene in the target relative to the reference sample. Positive values indicate higher expression in the target versus the reference, and vice versa for negative values. For a more detailed description of experiments in microarray genomic analysis, the reader is referred to McLachlan et al. (2004) and the

references therein.

The data arising from such experiments in genome research are usually in the form of large horizontal matrices of expression levels of p genes (variables) under n experimental conditions (observations) such as different times, cells, or tissues. Due to the large number of genes and to the complex relations between them, a reduction of the data dimensionality is needed in order to allow for a biological interpretation of the results and for subsequent information processing. Various methods have been proposed for the reduction of the data dimensionality and classification in genome research (Chipman et al. 2003).

Here, zig-zag EFA is applied to the lymphoma data set of Alizadeh et al. (2000), which contains gene expression levels for $p = 4026$ genes in $n = 62$ cells. The observations are composed of 42 cases of diffuse large B-cell lymphoma (DLCL), 9 cases of follicular lymphoma (FL) and 11 cases of B-cell chronic lymphocytic leukemia (CLL). The data set can be downloaded via `http://llmpp.nih.gov/lymphoma`. The gene expression data are summarized by a 62×4026 matrix $\mathbf{Z}$. Each data point in this matrix represents an expression measurement for a sample (row) and gene (column). Missing data were imputed by a k-nearest-neighbors algorithm (Hastie et al. 1999). Prior to the analysis the columns of the data matrix were standardized, so that the variables have commensurate means and scales.

When fitting an EFA model, it is important to decide over how many common factors, k, should be extracted. Various (often heuristic) criteria for choosing k are available (Mulaik 2010). However, note that EFA and similar techniques are typically examined to see whether they give an insightful low-dimensional representation of the data. In particular for gene expression data applied in this paper, there is less interest in the quality of the fit than whether the factors found can be usefully interpreted. In any case the model fit can be improved by increasing k. On the other hand, a k and $k + 1$ solution often produce quite different loadings and factor scores. We therefore examined solutions corresponding to different values of k and report here the the fitting solution for $k = 5$, which from our subjective perspective gives the most convincing interpretation.

For $k = 5$ and twenty random starts, the procedure required on average 39 iterations, taking about 85 seconds to converge. The algorithm was stopped when successive function values differed by less than $\epsilon = 10^{-3}$. Using a higher accuracy criterion such as $\epsilon = 10^{-5}$ needed considerably more CPU time but did not change the quality of the solution. Numerical experiments revealed that the algorithm converges to the same minimum of the loss function, up to the second decimal place. For comparison, factorizing the 4026×4026 sample correlation matrix and finding a numerical solution for minimizing the LS loss function $||\mathbf{Z}^\top\mathbf{Z} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top - \boldsymbol{\Psi}^2||_F^2$ based on principal axis factoring (with iterative estimation of communalities) takes about six times as much CPU time.

We also compared the performance of the zig-zag EFA algorithm with the alternative one for factorizing a horizontal data matrix proposed by Trendafilov and Unkel (2010). Whereas both methods perform similar with respect to computational speed and goodness-of-fit, Table 1 gives evidence that zig-zag

EFA is numerically more accurate with respect to the preservation of the EFA orthogonality constraints $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$ and $\mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k}$. The values in Table 1

Table 1 Evaluation of the numerical accuracy of zig-zag EFA and the algorithm by Trendafilov and Unkel (2010) for the lymphoma data of Alizadeh et al. (2000).

Method	$\|\|\mathbf{F}^\top\mathbf{F} - \mathbf{I}_k\|\|_F^2$	$\|\|\mathbf{U}^\top\mathbf{F}\|\|_F^2$
Trendafilov and Unkel (2010)	1.3249×10^{-9}	6.7859×10^{-5}
Zig-zag EFA	4.9059×10^{-31}	1.4003×10^{-31}

are averages taken over the results obtained by the twenty random starts.
In the context of gene expression analysis, $\mathbf{F}$ is the 62×5 matrix consisting of the scores of the 62 cells on the 5 latent factors. In gene expression analysis the underlying factors of the genes are also known as *eigengenes* if they were obtained by means of PCA (Alter et al. 2000). Figure 1 depicts the first three common factors for investigating the cell subclasses "DLCL", "FL" and "CLL". These three factors explain together 36.83% of the total sample variation. The subclass labels indicate some class grouping for all of the first three latent factors.
By means of the 4026×5 loading matrix $\boldsymbol{\Lambda}$, EFA provides a method of describing which genes play a leading role in the construction of the corresponding factor; see Figure 2. If they were obtained by means of PCA, the columns of the loading matrix are also sometimes called *eigenarrays* (Alter et al. 2000). The genes shown in Figure 2 are in the order produced by a agglomerative hierarchical clustering using the Cluster 3.0 software package of de Hoon et al. (2004). The (initial) loadings $\boldsymbol{\Lambda}$ could then be rotated towards a simple structure (Browne 2001), which would help to explain the patterns in the loadings and enhance their interpretation.
One might argue that results as displayed in Figure 1 and Figure 2 could also be obtained by means of PCA, which is probably the most popular technique for the reduction of the dimensionality. However, if there is a considerable amount of variation in each observed variable that is not accounted for by the k components, then one can expect to obtain a much better fit to the data using the EFA model than PCA. To clarify the difference between EFA and PCA, consider the SVD of

$$\mathbf{Z} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{Q}^\top \ , \tag{19}$$

where $\mathbf{P} \in \mathbb{R}^{n\times p}$ is orthonormal, $\mathbf{Q} \in \mathbb{R}^{p\times p}$ is orthogonal and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ is a diagonal matrix containing the singular values of $\mathbf{Z}$ sorted in decreasing order, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$, on its main diagonal. Note, that the decomposition (19) of $\mathbf{Z}$ is exact, while the EFA decomposition (1) can be achieved only approximately. As a matrix decomposition, PCA is based on (19) which is quite different from the EFA model decomposition (1) of $\mathbf{Z}$. For some k, the

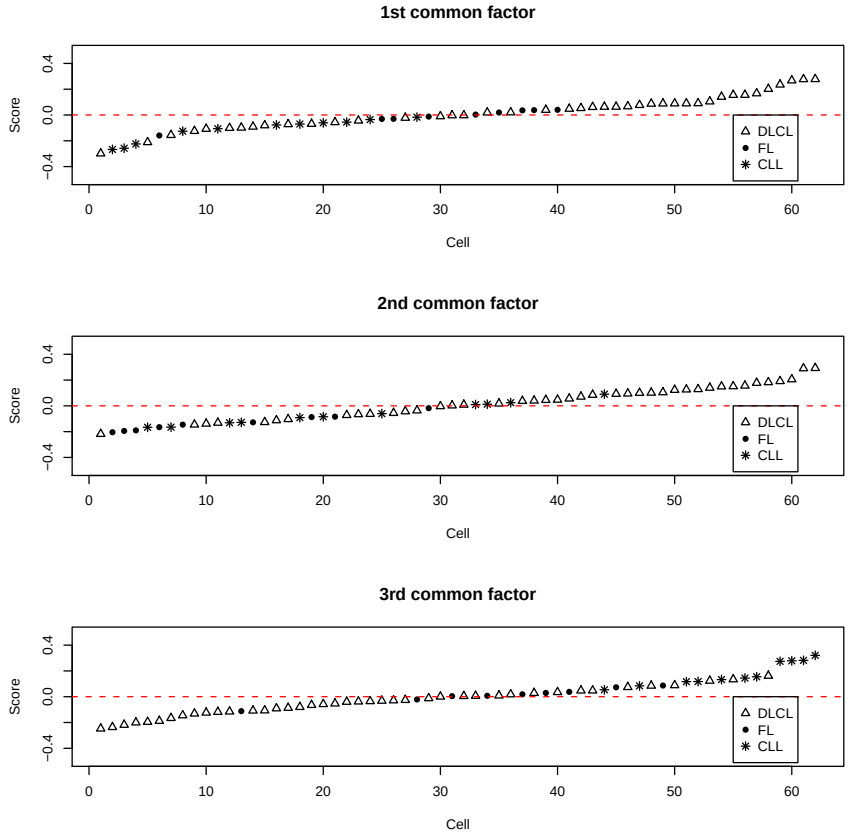

Fig. 1 Scores of the 62 cells (sorted in increasing order) on the first three common factors obtained from EFA of the lymphoma data of Alizadeh et al. (2000) ($k = 5$).

SVD (19) of $\mathbf{Z}$ can be partitioned and rewritten as

$$\mathbf{Z} = \mathbf{P}_1 \boldsymbol{\Sigma}_1 \mathbf{Q}_1^\top + \mathbf{P}_2 \boldsymbol{\Sigma}_2 \mathbf{Q}_2^\top \ , \tag{20}$$

where $\boldsymbol{\Sigma}_1 = \text{diag}(\sigma_1, ..., \sigma_k)$, $\boldsymbol{\Sigma}_2 = \text{diag}(\sigma_{k+1}, ..., \sigma_p)$ and $\mathbf{P}_1, \mathbf{P}_2, \mathbf{Q}_1$, and $\mathbf{Q}_2$ are the corresponding orthonormal matrices of left and right singular vectors with sizes $n \times k$, $n \times (p-k)$, $p \times k$, and $p \times (p-k)$, respectively. The norm of the error term $\mathbf{E}_{SVD} = \mathbf{P}_2 \boldsymbol{\Sigma}_2 \mathbf{Q}_2^\top$ is

$$||\mathbf{E}_{SVD}|| = ||\boldsymbol{\Sigma}_2|| = \sqrt{\sum_{i=k+1}^{p} \sigma_i^2} \ .$$

By defining $\mathbf{F} := \mathbf{P}_1$ and $\boldsymbol{\Lambda} := \mathbf{Q}_1 \boldsymbol{\Sigma}_1$, the PCA decomposition (20) of $\mathbf{Z}$ turns into

$$\mathbf{Z} = \mathbf{F} \boldsymbol{\Lambda}^\top + \mathbf{E}_{SVD} \ , \tag{21}$$

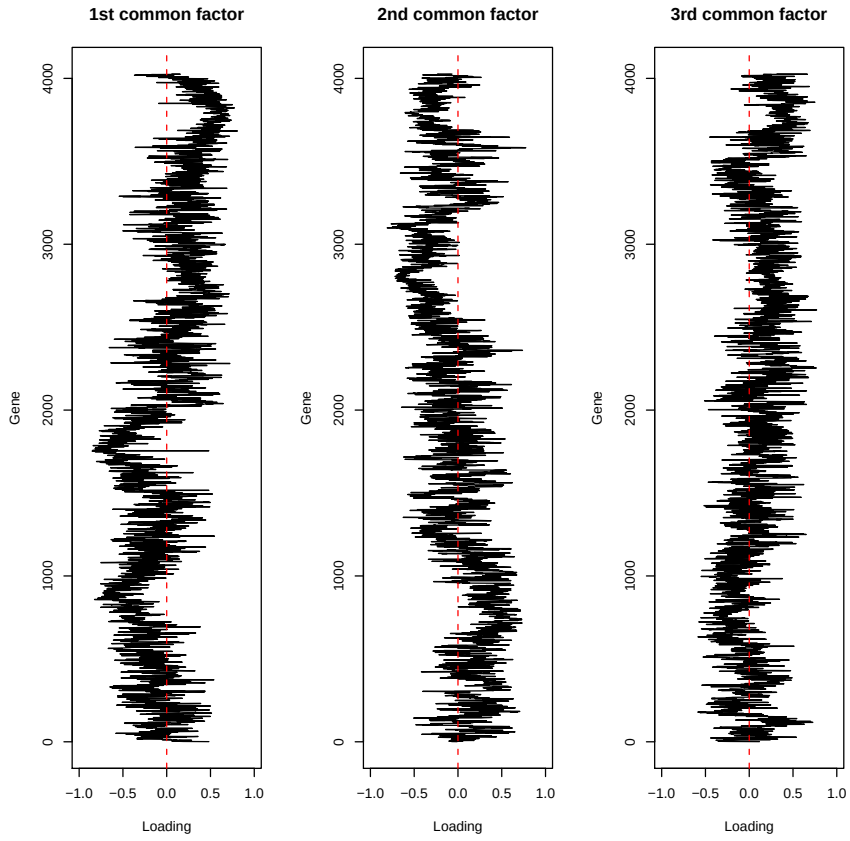

Fig. 2 Bar plots of the loadings of the 4026 genes on the first three common factors obtained from EFA of the lymphoma data of Alizadeh et al. (2000) ($k = 5$).

where $\mathbf{F}$ is the matrix of component scores of the n observations on the first k components and $\boldsymbol{\Lambda}$ is the corresponding matrix of coefficients or loadings of the p variables on the k components. In other words, PCA can be defined as the minimization of

$$||\mathbf{Z} - \mathbf{F}\boldsymbol{\Lambda}^{\top}||_F^2 \quad . \tag{22}$$

The two loss functions (22) and (3) are different and the error term $\mathbf{E}_{SVD}$ in the PCA decomposition (21) has a very different structure from $\mathbf{U}\boldsymbol{\Psi}$ in the EFA decomposition (1). To make the fit obtained by EFA and PCA directly comparable, a technique called 'EFA-like PCA', as advocated in Unkel (2009), is used. EFA-like PCA (for $p \geq n$) decomposes $\mathbf{E}_{SVD}$ further by solving the following LS optimization problem:

$$\min_{\mathbf{U},\boldsymbol{\Psi}} ||\mathbf{E}_{SVD} - \mathbf{U}\boldsymbol{\Psi}||^2 \quad , \tag{23}$$

subject to $\mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n$ and $\boldsymbol{\Psi}$ being a diagonal matrix. In addition, $\mathbf{U}$ should be orthogonal to the principal component scores $\mathbf{F}$ already found in (20). For an efficient alternating procedure to solve (23), the reader is referred to Unkel (2009).
Table 2 shows that for the lymphoma data, the fit obtained by EFA-like PCA compared to zig-zag EFA is considerably worse. The results indicate that the

Table 2 Quality of fit for zig-zag EFA and EFA-like PCA for the lymphoma data of Alizadeh et al. (2000).

Method	loss function	error of fit
EFA-like PCA	$\|\|\mathbf{E}_{SVD} - \mathbf{U}\boldsymbol{\Psi}\|\|_F^2$	1684.91
Zig-zag EFA	$\|\|\mathbf{Z} - \mathbf{F}\boldsymbol{\Lambda} - \mathbf{U}\boldsymbol{\Psi}\|\|_F^2$	1053.31

EFA model is preferable to PCA for reducing the dimensionality of this particular multivariate data set.

5.2 An illustration of the lower triangular reparameterization

Developing analytical methods for factor rotation has a long history (Browne 2001). It is motivated by both solving the rotational indeterminacy problem in EFA and facilitating the factors' interpretation. The aim for analytic rotation is to find loadings with 'simple structure' in an objective manner. Thurstone (1947) set forth a number of general principles which, vaguely stated, say that a loading matrix with many small values and a small number of larger values is simpler than one with mostly intermediate values (Thurstone 1947, p. 335). Thurstone's 26-variable box problem (Thurstone 1947) was notorious for being difficult to solve by any analytic rotation method. Thurstone (1947) collected a random sample of 20 boxes and measured their three dimensions x (length), y (width) and z (height). In this data set, the boxes constitute the n observational units. The p variables of the example are twenty-six functions of these dimensions : x, y, z, xy, xz, yz, x^2y, xy^2, x^2z, xz^2, y^2z, yz^2, x/y, y/x, x/z, z/x, y/z, z/y, $2x+2y$, $2x+2z$, $2y+2z$, $\sqrt{x^2+y^2}$, $\sqrt{x^2+z^2}$, $\sqrt{y^2+z^2}$, xyz and $\sqrt{x^2+y^2+z^2}$. The observed variables are mean-centered and scaled to unit norm and the result is expressed in a 20×26 data matrix $\mathbf{Z}$. These are the data to which methods were applied. The problem is to find simple loadings which identify the dimensions of the boxes. Hence, three common factors are extracted ($k = 3$).
To illustrate the lower triangular (LT) reparameterization of the EFA model proposed in Section 3, Table 3 shows the classical EFA least squares solution $\{\boldsymbol{\Lambda}, \boldsymbol{\Psi}^2\}$ obtained by factorizing the correlation matrix and minimizing

$||\mathbf{Z}^\top\mathbf{Z} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top - \boldsymbol{\Psi}^2||_F^2$, the results $\{\boldsymbol{\Lambda}, \boldsymbol{\Psi}^2\}$ obtained by zig-zag EFA and minimizing (3), as well as the ideal pattern (simplest structure) of the loadings with respect to Thurstone's box problem. The stars in the ideal pattern indicate which of the dimensions x, y and z were used to generate the functions in the first column. The results reported are the 'best' obtained after twenty

Table 3 Ideal pattern of the loading matrix and initial solutions obtained by classical LS EFA and zig-zag EFA for loadings and unique factor variances for the 20×26 box data.

	Ideal pattern			Classical EFA				Zig-zag EFA			
Function	of loadings			$\boldsymbol{\Lambda}$			$\boldsymbol{\Psi}^2$	$\mathbf{L} = \text{tril}(\mathbf{Z}^\top\mathbf{F})$			$\boldsymbol{\Psi}^2$
x	★	0	0	.50	.53	.68	.0011	1.00	0	0	.0000
y	0	★	0	.47	.70	-.53	.0015	.24	.97	0	.0000
z	0	0	★	-.63	.78	-.00	.0022	.10	.23	.96	.0000
xy	★	★	0	.61	.79	-.07	.0079	.68	.73	-.00	.0000
xz	★	0	★	-.35	.89	.27	.0103	.49	.20	.84	.0000
yz	0	★	★	-.29	.92	-.22	.0130	.20	.59	.77	.0000
x^2y	★	★	0	.61	.76	.17	.0293	.82	.54	-.00	.0191
y^2x	★	★	0	.58	.76	-.25	.0256	.52	.84	-.03	.0000
x^2z	★	0	★	-.14	.87	.42	.0455	.68	.15	.68	.0198
xz^2	★	0	★	-.44	.86	.14	.0449	.33	.24	.90	.0000
y^2z	0	★	★	-.08	.92	-.30	.0570	.23	.73	.60	.0298
yz^2	0	★	★	-.42	.87	-.14	.0541	.16	.45	.85	.0000
x/y	★	★	0	-.06	-.30	.93	.0423	.44	-.87	-.05	.0279
y/x	★	★	0	.07	.27	-.94	.0322	-.46	.87	.02	.0290
x/z	★	0	★	.80	-.47	.23	.0929	.31	-.15	-.89	.0811
z/x	★	0	★	-.80	.46	-.30	.0666	-.36	.20	.88	.0476
y/z	0	★	★	.86	-.28	-.34	.0728	.04	.40	-.87	.0566
z/y	0	★	★	-.85	.30	.33	.0791	-.03	-.38	.88	.0651
$2x + 2y$	★	★	0	.61	.78	.09	.0032	.79	.61	.00	.0000
$2x + 2z$	★	0	★	-.09	.88	.46	.0042	.74	.15	.65	.0000
$2y + 2z$	0	★	★	-.09	.93	-.34	.0033	.22	.76	.61	.0000
$(x^2 + y^2)^{1/2}$	★	★	0	.61	.75	.23	.0094	.87	.49	-.01	.0000
$(x^2 + z^2)^{1/2}$	★	0	★	.18	.79	.58	.0163	.91	.10	.39	.0001
$(y^2 + z^2)^{1/2}$	0	★	★	.09	.90	-.42	.0132	.23	.86	.44	.0000
xyz	★	★	★	-.11	.98	-.00	.0290	.47	.54	.68	.0017
$(x^2 + y^2 + z^2)^{1/2}$	★	★	★	.37	.90	.20	.0142	.80	.52	.28	.0001

random starts. By 'best' a solution is meant that resembles the ideal pattern in Table 3 most. Unlike the initial loadings $\boldsymbol{\Lambda}$, the loadings $\mathbf{L}$ exhibit an interpretable and contextually meaningful relation between the observed variables and the common factors. If one ignores all loadings with magnitude .24 or less in $\mathbf{L}$, the remaining loadings perfectly identify which of the box dimensions x,

y and z were used to generate each of the manifest variables. Furthermore, the numerical experiments revealed that zig-zag EFA employing a lower triangular matrix $\mathbf{L} = \texttt{tril}(\mathbf{Z}^\top\mathbf{F})$ yields loadings that are pretty stable. In contrast, zig-zag EFA with the parameterization $\boldsymbol{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ and classical EFA converge to quite different $\boldsymbol{\Lambda}$.

Table 3 also shows the unique factor variances, each of which accounts for the unique variance in its associated observed variable. Since the LS algorithm factorizing the sample correlation matrix results in $\boldsymbol{\Psi}^2$ staying on the cone of positive definite diagonal matrices, the classical LS approach in EFA leads to uniquenesses being strictly positive. In contrast, zig-zag EFA allows the unique factors to have zero variance, which is in line with the EFA model assumptions for $p \geq n$ (cf. Section 2).

To compare the loadings $\mathbf{L}$ with what one might have obtained on the data using one of the well-known rotation-to-simplicity methods, the solution $\mathbf{L}$ in Table 3 shall be compared with the (orthogonal) Minimum entropy rotation (Browne 2001) of the initial loading matrix $\boldsymbol{\Lambda}$ in Table 3. Since one must look fairly hard at aligned loading matrices to form an opinion about the simple structure achieved, the sorted absolute loadings (SAL) plot as advocated by Jennrich (2004) is used for comparison instead. The SAL plot does not need to produce the column permutations and sign changes often required to align loadings matrices for comparison. Let $m = pk$ and let $|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_m|$ denote the absolute values of the loadings sorted in increasing order. The SAL plot is a plot of $|\lambda_j|$ against j for $j = 1, \ldots, m$. Generally, the greater the number of small loadings, the simpler the loading matrix.

Figure 3 is an SAL plot of the rotated loadings obtained by Minimum entropy applied to $\boldsymbol{\Lambda}$ in Table 3 and the loadings of the LT solution $\mathbf{L}$. According to the ideal pattern shown in Table 3, the optimal loading matrix should have 27 zero loadings. Whereas the LT solution has the smallest 12 loadings very close to zero, Minimum entropy encourages small loadings more than the LT parameterization and has 20 loadings very close to zero.

It is not surprising that the structure in the rotated loadings obtained by Minimum entropy is simpler than the one found by the LT solution, as Minimum entropy and other rotation-to-simplicity criteria were explicitly designed to tackle the rotational indeterminacy problem in EFA and to enhance the factors' interpretation. However, the LT solution is quite remarkable, as it identifies the dimensions of the boxes and hence facilitates the interpretation of the solution without recourse to rotation.

6 Conclusions

Over the past decade, new methods emerged in EFA that operate directly on the data matrix rather than on a sample covariance or correlation matrix. These methods produce factor loadings and factor scores simultaneously. We presented the new algorithm zig-zag EFA that is able to fit the EFA model to horizontal data matrices with $p \geq n$ and estimates all EFA matrix parameters

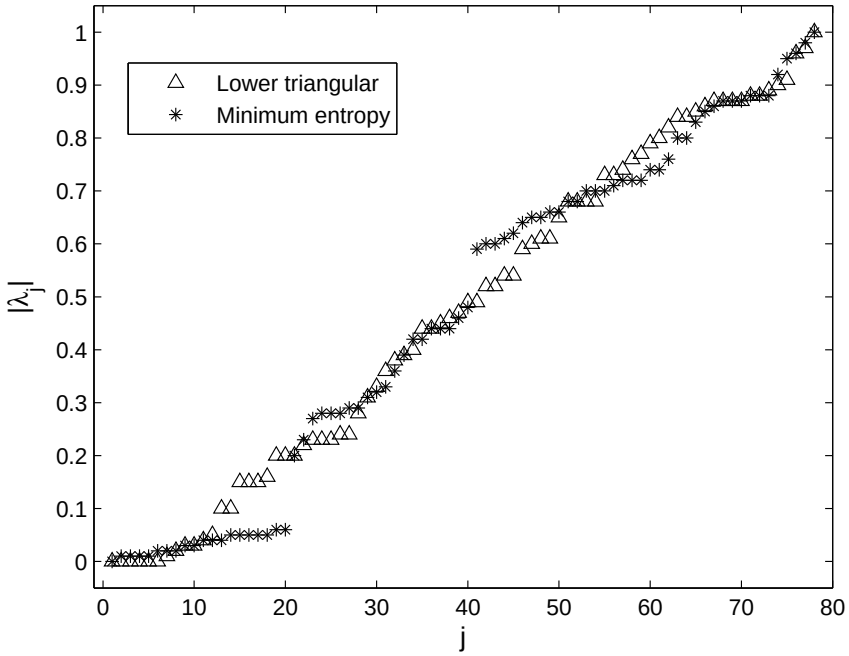

Fig. 3 SAL plot of the Minimum entropy criterion and the lower triangular parameterization applied to the 20×26 box problem.

simultaneously.

For the high-dimensional data set analyzed in Subsection 5.1, zig-zag EFA was shown to be considerably faster than classical least squares EFA factorizing the sample correlation matrix. Moreover, zig-zag EFA outperformed PCA with respect to the fit obtained. Despite the differences between PCA and EFA (e.g., Jolliffe 2002, pp. 158-161), both methods aim to reduce the dimensionality of a data set and, so far, PCA is probably the most popular technique for doing so, no matter whether the data arise in social and behavioural sciences, in genome research, atmospheric science or elsewhere. The methods developed by De Leeuw (2004, 2008), Trendafilov and Unkel (2010) and the one in this paper may be able to make EFA a more attractive competitor to PCA in the future, because they are based on the computationally efficient and transparent SVD of matrices, as PCA is.

A reparameterization of the EFA model was proposed to produce solutions with a lower triangular matrix $\mathbf{L}$ of loadings. Whereas the parameter matrix $\boldsymbol{\Lambda}$ in the classical EFA formulation (1) admits an infinite number of orthogonal rotations, the lower triangular reparameterization in (14) removes the rotational indeterminacy of the EFA model and leads to solutions already having an interpretable pattern. Moreover, the new parameters are subject to

the constraint $\text{rank}(\mathbf{L}) = k$ expressing the nature of the EFA model, rather than facilitating the numerical method for their estimation (as is the case with the condition $\boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}$ or $\boldsymbol{\Lambda}^\top \boldsymbol{\Psi}^{-2} \boldsymbol{\Lambda}$ being diagonal for the standard EFA estimation procedures).

Methods operating on the data matrix may also form the basis for constructing new approaches in robust EFA that can resist the effect of outliers. Since outliers can heavily influence the estimate of the model covariance/correlation matrix and hence also the parameter estimates, classical EFA techniques taking input data in the form of product moments are very vulnerable to the presence of outliers in the data. One may either use some robust modification of the sample correlation matrix to overcome the outlier problem (e.g., Pison et al. 2003) or look for alternative techniques working with the data matrix (Croux et al. 2003; Unkel and Trendafilov 2010).

References

Alizadeh, A. A., Eisen, M. B., Davis, R. E., and et al. (2000), "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, 403, 503–511.

Alter, O., Brown, P. O., and Botstein, D. (2000), "Singular value decomposition for genome-wide expression data processing and modeling," *Proceedings of the National Academy of Sciences*, 97, 10101–10106.

Anderson, T. W. and Rubin, H. (1956), "Statistical inference in factor analysis," in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, Vol. V*, ed. Neyman, J., University of California Press: Berkeley, pp. 111–150.

Bartholomew, D. J. and Knott, M. (1999), *Latent Variable Models and Factor Analysis*, Edward Arnold: London, 2nd ed.

Browne, M. W. (2001), "An overview of analytic rotation in exploratory factor analysis," *Multivariate Behavioral Research*, 36, 111–150.

Chipman, H., Hastie, T. J., and Tibshirani, R. (2003), "Clustering microarray data," in *Statistical Analysis of Gene Expression Microarray Data*, ed. Speed, T., Chapman & Hall/CRC: Boca Raton, Florida, pp. 159–200.

Croux, C., Filzmoser, P., Pison, G., and Rousseeuw, P. J. (2003), "Fitting multiplicative models by robust alternating regressions," *Statistics and Computing*, 13, 23–36.

de Hoon, M. J. L., Imoto, S., Nolan, J., , and Miyano, S. (2004), "Open Source Clustering Software," *Bioinformatics*, 20, 1453–1454.

De Leeuw, J. (2004), "Least squares optimal scaling of partially observed linear systems," in *Recent Developments on Structural Equation Models: Theory and Applications*, eds. van Montfort, K., Oud, J., and Satorra, A., Kluwer Academic Publishers: Dordrecht, pp. 121–134.

— (2008), "Factor analysis as matrix decomposition," Preprint series: Department of Statistics, University of California, Los Angeles.

Golub, G. H. and Van Loan, C. F. (1996), *Matrix Computations*, The John Hopkins University Press: Baltimore, Maryland, 3rd ed.

Gower, J. C. and Dijksterhuis, G. B. (2004), *Procrustes Problems*, Oxford University Press: Oxford.

Harman, H. H. (1976), *Modern Factor Analysis*, University of Chicago Press: Chicago, 3rd ed.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer: New York, 2nd ed.

Hastie, T., Tibshirani, R., Sherlock, G., Eisen, M., Brown, P., and Botstein, D. (1999), "Imputing missing data fro gene expression arrays," Technical Report, Division of Biostatistics, Stanford University.

Jennrich, R. I. (2004), "Rotation to simple loadings using component loss functions: The orthogonal case," *Psychometrika*, 69, 257–273.

Jolliffe, I. T. (2002), *Principal Component Analysis*, Springer: New York, 2nd ed.

Jöreskog, K. G. (1977), "Factor analysis by least-squares and maximum likelihood methods," in *Mathematical methods for digital computers*, eds. Enslein, K., Ralston, A., and Wilf, H. S., John Wiley & Sons: New York, pp. 125–153.

McLachlan, G. J., Do, K.-A., and Ambroise, C. (2004), *Analyzing Microarray Gene Expression Data*, John Wiley & Sons: Hoboken, New Jersey.

Mulaik, S. A. (2010), *Foundations of Factor Analysis*, Chapman & Hall/CRC: Boca Raton, Florida, 2nd ed.

Pison, G., Rousseeuw, P. J., Filzmoser, P., and Croux, C. (2003), "Robust factor analysis," *Journal of Multivariate Analysis*, 84, 145–172.

Sočan, G. (2003), "The Incremental Value of Minimum Rank Factor Analysis," PhD Thesis, University of Groningen: Groningen, The Netherlands.

ten Berge, J. M. F. (1993), *Least squares optimization in multivariate analysis*, DSWO Press: Leiden.

The MathWorks (2008), *Using MATLAB, Version 7.7 (R2008b)*, The MathWorks Inc.

Thurstone, L. L. (1947), *Multiple Factor Analysis*, University of Chicago Press: Chicago.

Trendafilov, N. T. (2005), "Fitting the factor analysis model in ℓ_1 norm," *British Journal of Mathematical and Statistical Psychology*, 58, 19–31.

Trendafilov, N. T. and Unkel, S. (2010), "Exploratory factor analysis of data matrices with more variables than observations," *Journal of Computational and Graphical Statistics*, under revision (also Open University Statistics Group Technical Report 09/13, available from http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm).

Unkel, S. (2009), "Factor Analysis of Data Matrices: New Theoretical and Computational Aspects with Applications," PhD Thesis, Open University: Milton Keynes.

Unkel, S. and Trendafilov, N. T. (2010), "A majorization algorithm for simultaneous parameter estimation in robust exploratory factor analysis," *Computational Statistics & Data Analysis* 54, 3348-3358.