Exploratory factor and principal component analyses: some new aspects

Nickolay T. Trendafilov*, Steffen Unkel

Department of Mathematics and Statistics,

The Open University, Milton Keynes, UK

and

Wojtek Krzanowski

School of Engineering, Mathematics and Physical Sciences,
University of Exeter, UK

March 24, 2010

*Corresponding author (N.Trendafilov@open.ac.uk).

Abstract

Exploratory Factor Analysis (EFA) and Principal Component Analysis (PCA) are popular techniques for simplifying presentation of, and investigating structure of, an $(n \times p)$ data matrix. However, these fundamentally different techniques are frequently confused, and the differences between them are obscured, because they give similar results in some practical cases. We therefore investigate conditions under which factor scores and principal component scores are expected to be close to each other, by considering EFA as a matrix decomposition so that it can be directly compared with the data matrix decomposition underlying PCA. We propose a QR decomposition as an alternative to the usual singular value decomposition for PCA, provide iterative algorithms for the computation, and derive a matrix condition that has to be satisfied for the two techniques to give similar results. Throughout, we consider separately the cases $n > p$ and $p \geq n$. All derived algorithms and matrix conditions are illustrated on two data sets, one for each of these two cases.

Key words: Data matrix decomposition, SVD and QR factorization, Projected gradients, Optimality conditions, Procrustes problems.

1 Introduction

Multivariate data, in which n independent observations have been made on p variables and collected in an $n \times p$ matrix $\mathbf{X}$, arise frequently in many quantitative fields of study. Such data are traditionally considered to be $p-$dimensional, since the values of each individual can be represented by a point whose coordinates on p orthogonal axes are given by its respective variable values. However, since the same variables have been measured on each individual, the ensuing correlations between the variables implies that the swarm of n points will generally lie in fewer than p dimensions. The objective of descriptive multivariate techniques is therefore to simplify the data presentation as far as possible, and to uncover its true dimensionality. Two of the most popular techniques for doing this are Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA).

PCA (Jolliffe 2002) replaces the p original variables by p linear combinations of them, chosen such that they are ordered by decreasing size of variance. These combinations are the "principal components"; data simplification is achieved by discarding those components that have small variance and retaining only the r largest components. This procedure defines a rotation of the original $p-$dimensional space, followed by projection into the $r-$dimensional subspace that causes the least distortion of data points. The components are obtained as eigenvectors of the sample covariance matrix, and the coordinates of the projected points (the component "scores") by postmultiplying $\mathbf{X}$ by these components. More usually nowadays, both components and scores are obtained directly from the Singular Value Decomposition (SVD) of $\mathbf{X}$.

EFA (Mulaik 2010) takes a superficially opposite viewpoint, and seeks to explain the inter-variable correlations by modelling each variable as a linear combination of $r < p$ common "factors" plus a specific "residual". By assuming that the observed data (i.e. rows of $\mathbf{X}$) are independently drawn replicates from this model and by making appropriate distributional assumptions about the common factors

and specific residuals, the coefficients of the linear combinations can be estimated using standard statistical procedures such as maximum likelihood or least squares. However, unlike in PCA, factor scores can no longer be calculated directly but need further estimation based on additional assumptions.

While PCA and EFA thus have fundamentally different bases and standpoints, it is unfortunate that the differences between them have often been obscured because both techniques produce very similar solutions in a number of practical cases. The standard warning that PCA is a descriptive technique, while EFA is a model based one, seems not quite convincing and transparent. Instead, it might be more acceptable if PCA and EFA solutions were to be directly compared according to their performance on any particular data set. This idea is not new and follows the suggestion made by Rao (1996): "Some conditions under which the factor scores and principal components are close to each other have been given by Schneeweiss and Mathes (1995). It would be of interest to pursue such theoretical investigations and also examine in individual data sets the actual differences between principal components and factor scores." One of our main aims in this paper is to investigate how this can be done.

Our principal mode of attack is to consider EFA as a data matrix decomposition, so that it can be directly compared with the data matrix decomposition underlying PCA. In Section 2 we build on the work of De Leeuw (2004) for EFA as a matrix decomposition, and then in Section 3 we consider how the SVD of PCA can viewed in this light. Of course, the SVD is not the only possible matrix decomposition that could be used, and in Section 3 we also introduce the idea of PCA based on the QR decomposition of the data matrix. Section 4 is then the central section in which comparison of EFA and PCA is discussed. To enable an exact comparison to be made, the structures of the two decompositions need to be first aligned. For this we introduce what we call an EFA-like decomposition of the PCA error term, and provide an iterative algorithm for its computation. By examining the optimality conditions of both EFA and EFA-like PCA, we are then

able to deduce the exact conditions under which the EFA and PCA solutions will be similar on any given data set. Section 5 provides some numerical illustrations of the methodology.

Note, however, that most classical theory for EFA and PCA was developed when data sets were almost without exception "vertical", i.e. such that $n > p$. In recent years, areas of study have proliferated in which samples are relatively difficult to obtain but measurement of many variables on them is relatively easy, as for example in chemometrics or genomic studies. These studies give rise to "horizontal" data sets having $p \geq n$ (frequently $p >> n$), and such data sets are becoming very common. These situations cause some complications, both for standard EFA and for the comparison with PCA, so the two distinct cases are treated separately in each section below. For simplicity of presentation, we also assume throughout that all variables have been standardised prior to analysis.

2 EFA as a data matrix decomposition

2.1 The classical case $n > p$

Let $\mathbf{Z}$ be the $n \times p$ data matrix collecting n independent centered observations on p ($n > p$) variables with unit variance. The classic k-factor model and related assumptions (Mulaik 2010, pp.135-137) imply that EFA represents the data matrix $\mathbf{Z}$ as:

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

where the EFA (unknown) parameters $\mathbf{F}, \mathbf{U}, \mathbf{\Lambda}$ and $\mathbf{\Psi}$ satisfy the following constraints:

$$\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k, \ \mathbf{U}^\top\mathbf{U} = \mathbf{I}_p, \ \mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k} \ , \tag{2}$$

with $\mathbf{\Lambda}$ of full column rank k (to secure the k-factor model) and $\mathbf{\Psi}$ diagonal. Thus, the observations $\mathbf{Z}$ are in the space spanned by the common and unique factors.

The EFA data representation (1) – (2) implies the following representation of the sample correlation matrix:

$$\mathbf{Z}^\top \mathbf{Z} \approx \mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Psi}^2 \ . \tag{3}$$

The equation (3) is fundamental for parameter estimation in standard EFA (with random common factors), whereby a pair $\{\mathbf{\Lambda}, \mathbf{\Psi}\}$ is sought which gives the best fit to $\mathbf{Z}^\top \mathbf{Z}$ with respect to some discrepancy measure. The most frequently used goodness-of-fit measures are maximum likelihood (ML) and least squares (LS). The former requires the solution of the following optimization problem, known as ML factor analysis:

$$\min_{\mathbf{\Lambda},\mathbf{\Psi}} [\log(\det(\mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Psi}^2)) + \mathrm{trace}((\mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Psi}^2)^{-1}(\mathbf{Z}^\top \mathbf{Z}))] \ , \tag{4}$$

while the latter requires solution of LS factor analysis:

$$\min_{\mathbf{\Lambda},\mathbf{\Psi}} ||\mathbf{Z}^\top \mathbf{Z} - \mathbf{\Lambda}\mathbf{\Lambda}^\top - \mathbf{\Psi}^2||^2 \ , \tag{5}$$

where $||\mathbf{X}|| = \sqrt{\mathrm{trace}(\mathbf{X}^\top \mathbf{X})}$ denotes the Frobenius norm of a matrix $\mathbf{X}$.

In general, the process of finding $\{\mathbf{\Lambda}, \mathbf{\Psi}\}$, for some specified value of k, is called factor extraction. Various factor extraction methods have been proposed, of which (4) and (5) are the most popular (Mulaik 2010). Once the pair $\{\mathbf{\Lambda}, \mathbf{\Psi}\}$ is obtained, then common factor scores $\mathbf{F}$ can be found as a function of $\mathbf{Z}$, $\mathbf{\Lambda}$ and possibly $\mathbf{\Psi}$ in a number of ways (Harman 1976; Mulaik 2010). However, the estimation of $\mathbf{F}$ raises a number of conceptual and computational problems related to the factor indeterminacy (Mulaik 2005).

These difficulties can be avoided if EFA is considered as a specific data matrix decomposition, without embedding the data in a replication framework and assuming the observations to be realizations of random variables. Such modification of EFA was proposed by De Leeuw (2004). Then, the EFA problem is to minimize the following least squares goodness-of-fit criterion:

$$f(\mathbf{F}, \mathbf{\Lambda}, \mathbf{U}, \mathbf{\Psi}) \ = \ ||\mathbf{Z} - \mathbf{F}\mathbf{\Lambda}^\top - \mathbf{U}\mathbf{\Psi}||^2 \ , \tag{6}$$

$$\text{subject to} \qquad \mathbf{F}^\top \mathbf{F} = \mathbf{I}_k, \ \mathbf{U}^\top \mathbf{U} = \mathbf{I}_p, \ \mathbf{U}^\top \mathbf{F} = \mathbf{O}_{p \times k} \ , \tag{7}$$

and $\mathbf{\Psi}$ diagonal. De Leeuw (2004) suggested two alternating least squares (ALS) procedures for solving (6) – (7). The idea is that the common and unique factor scores $\mathbf{F}$ and $\mathbf{U}$ can be found simultaneously by solving a Procrustes problem (Gower and Dijksterhuis 2004) for given or estimated $\mathbf{\Lambda}$ and $\mathbf{\Psi}$, and the latter are then updated by $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ and $\mathbf{\Psi} = \text{diag}(\mathbf{U}^\top\mathbf{Z})$. Recently, the proposed ALS procedures were further developed and elaborated for EFA of data matrices containing more variables than observations (Trendafilov and Unkel 2009), where it is proved that the updating formula $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ secures $\text{rank}(\mathbf{\Lambda}) = k$.

It is well known that for any orthogonal $k \times k$ matrix $\mathbf{Q}$ the transformed matrices $\widetilde{\mathbf{F}} = \mathbf{F}\mathbf{Q}$, $\widetilde{\mathbf{\Lambda}} = \mathbf{\Lambda}\mathbf{Q}$ and $\widetilde{\mathbf{U}} = \mathbf{U}\mathbf{Q}$ are also solutions of the EFA problem (6) – (7) (Mulaik 2010). This rotational freedom can be reduced considerably by restricting the loadings matrix $\mathbf{\Lambda}$ to be lower triangular $\mathbf{L}$. Then the EFA problem (6) – (7) can be formulated alternatively as:

$$f(\mathbf{F}, \mathbf{L}, \mathbf{U}, \mathbf{\Psi}) \;=\; ||\mathbf{Z} - \mathbf{F}\mathbf{L}^\top - \mathbf{U}\mathbf{\Psi}||^2 \;, \tag{8}$$

$$\text{subject to} \qquad \mathbf{F}^\top\mathbf{F} = \mathbf{I}_k,\; \mathbf{U}^\top\mathbf{U} = \mathbf{I}_p,\; \mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k} \;, \tag{9}$$

where $\mathbf{L}$ is lower triangular and $\mathbf{\Psi}$ diagonal. Trendafilov and Unkel (2009) show that the ALS algorithms for solving the original EFA problem (6) – (7) remain valid with a modified updating formula $\mathbf{L} = \texttt{tril}(\mathbf{Z}^\top\mathbf{F})$, where $\texttt{tril}(\mathbf{X})$ is defined as:

$$\{\texttt{tril}(\mathbf{X})\}_{ij} = \begin{cases} x_{ij} & i \geq j \\ 0 & \text{otherwise} \end{cases} \;.$$

2.2 New applications with $p \geq n$

If $p \geq n$, the sample covariance/correlation matrix is singular. Moreover, the constraint $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$ cannot be fulfilled. In order to make progress, Trendafilov and Unkel (2009) show that the classical EFA problem (6) – (7) can be generalized

to the following minimization problem:

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

$$\text{subject to} \qquad \text{rank}(\mathbf{F}) = k, \mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n, \mathbf{U}^\top\mathbf{U}\mathbf{\Psi} = \mathbf{\Psi} \tag{11}$$

and $\mathbf{\Psi}$ assumed diagonal. Note, that under the new constraints $\mathbf{\Psi}$ has at most $n - k$ nonzero entries (Trendafilov and Unkel 2009). The rotational freedom can be reduced as in the classical case, by replacing $\mathbf{\Lambda}$ with $\mathbf{L}$ in (10).

3 PCA as a data matrix decomposition

3.1 PCA based on the SVD decomposition

As a matrix decomposition, PCA is based on the SVD (Golub and Van Loan 1996) of the $n \times p$ data matrix $\mathbf{Z}$, which is quite different from the EFA matrix decomposition (1) of $\mathbf{Z}$. To appreciate the difference between EFA and PCA, consider the SVD of $\mathbf{Z}$ which has the form:

$$\mathbf{Z} = \mathbf{P}\mathbf{\Sigma}\mathbf{Q}^\top \;, \tag{12}$$

where $\mathbf{P} \in \mathbb{R}^{n\times p}$ is orthonormal, $\mathbf{Q} \in \mathbb{R}^{p\times p}$ is orthogonal and $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ 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 (12) of $\mathbf{Z}$ is exact, while the EFA decomposition (1) can be achieved only approximately.

For some k, the SVD (12) of $\mathbf{Z}$ can be partitioned and rewritten as:

$$\mathbf{Z} = \mathbf{P}_1\mathbf{\Sigma}_1\mathbf{Q}_1^\top + \mathbf{P}_2\mathbf{\Sigma}_2\mathbf{Q}_2^\top \;, \tag{13}$$

where $\mathbf{\Sigma}_1 = \text{diag}(\sigma_1, ..., \sigma_k)$, $\mathbf{\Sigma}_2 = \text{diag}(\sigma_{k+1}, ..., \sigma_p)$ 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 $\mathbf{E}_{SVD} = \mathbf{P}_2\mathbf{\Sigma}_2\mathbf{Q}_2^\top$ is

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

By defining $\mathbf{F} := \mathbf{P}_1$ and $\mathbf{\Lambda} := \mathbf{Q}_1\mathbf{\Sigma}_1$, the PCA decomposition (13) of $\mathbf{Z}$ turns into

$$\mathbf{Z} = \mathbf{F}\mathbf{\Lambda}^\top + \mathbf{E}_{SVD} \ . \tag{14}$$

Note also, that $\mathbf{F}$ ($= \mathbf{P}_1$) in both EFA and PCA is orthogonal to the second ('error') term, i.e. $\mathbf{F}^\top\mathbf{U}\mathbf{\Psi} = \mathbf{O}_{k\times p}$ and $\mathbf{F}^\top\mathbf{E}_{SVD} = \mathbf{O}_{k\times p}$.

3.2 PCA based on the QR decomposition

Traditionally, PCA accomplished by the SVD reduced-rank approximation is considered to be the optimal method for reducing the dimensionality of the data. This is due to the optimal LS property of the SVD (Golub and Van Loan 1996, Theorem 2.5.3). However, as a rank-reducing method, the SVD can be expensive to compute for a large data matrix $\mathbf{Z}$.

Recently, it has been shown that "if any reduced-rank approximation is accurate then it contains good approximations to the singular vectors corresponding to large singular values" (Berry *et al.* 2005, Theorem 6.1). This is true, in particular, for the QR decomposition (Golub and Van Loan 1996) which possesses a number of attractive numerical properties that the SVD lacks (Stewart 1998). For computational and interpretational reasons, one can perform a PCA-like analysis based on the QR factorization of $\mathbf{Z}$ which is given by

$$\mathbf{Z} = \mathbf{Q}\mathbf{R} \ , \tag{15}$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is orthogonal and $\mathbf{R} \in \mathbb{R}^{n\times p}$ is upper triangular. After partitioning, (15) can be written as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix} = \mathbf{Q}_1\mathbf{R}_1 + \mathbf{Q}_2\mathbf{R}_2 \ , \tag{16}$$

where $\mathbf{Q}_1 \in \mathbb{R}^{n\times k}$ and $\mathbf{Q}_2 \in \mathbb{R}^{n\times(n-k)}$ are orthonormal and $\mathbf{R}_1 \in \mathbb{R}^{k\times p}$ and $\mathbf{R}_2 \in \mathbb{R}^{(n-k)\times p}$ are upper triangular. From (16) it also follows that $\mathbf{Q}_1^\top\mathbf{Q}_1 =$

$\mathbf{I}_k, \mathbf{Q}_1^\top \mathbf{Q}_2 = \mathbf{O}_{k\times(n-k)}, \mathbf{Q}_2^\top \mathbf{Q}_2 = \mathbf{I}_{n-k}$ and $\mathbf{Q}_1\mathbf{Q}_1^\top + \mathbf{Q}_2\mathbf{Q}_2^\top = \mathbf{I}_n$. By defining $\mathbf{F} := \mathbf{Q}_1$ and $\mathbf{L} := \mathbf{R}_1^\top$, (16) turns into

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^\top + \mathbf{E}_{QR} \ , \tag{17}$$

where $\mathbf{L} \in \mathbb{R}^{p\times k}$ is a lower triangular matrix and $\mathbf{E}_{QR} = \mathbf{Q}_2\mathbf{R}_2$ is the error term. Note that $\mathbf{F}^\top \mathbf{E}_{QR} = \mathbf{O}_{k\times p}$. The norm of $\mathbf{E}_{QR}$ is $||\mathbf{E}_{QR}|| = ||\mathbf{R}_2||$.

3.3 The case $p \geq n$

In this case, PCA is performed on the transposed data $\mathbf{Z}^\top$.

4 Comparing PCA and EFA solutions

In order to conduct the comparison, we first need to rewrite the PCA data matrix decompositions (14) and (17) in a form resembling the EFA decomposition of $\mathbf{Z}$, and we will refer to these latter decompositions as EFA-like PCA. Then, a comparison between PCA and EFA can be made by comparing the EFA and the EFA-like PCA optimality conditions.

4.1 EFA-like decomposition of the PCA error term

4.1.1 $n > p$

The error term $\mathbf{E}_{SVD}$ in the PCA decomposition (14) has a very different structure from $\mathbf{U}\boldsymbol{\Psi}$ in the EFA decomposition (1). It will be demonstrated that the form of $\mathbf{E}_{SVD}$ gives only superficial difference between EFA and PCA. What really matters is the underlying model of the EFA matrix decomposition (1). The EFA algorithms look for pairs of unknowns $\{\boldsymbol{\Lambda}, \boldsymbol{\Psi}\}$ and $\{\mathbf{F}, \mathbf{U}\}$, whereas the (EFA-like) PCA looks for $\{\boldsymbol{\Lambda}, \mathbf{F}\}$ and $\{\boldsymbol{\Psi}, \mathbf{U}\}$.

Formally speaking, to get an EFA-like decomposition from (14), $\mathbf{E}_{SVD}$ should be further decomposed as a product of orthonormal and diagonal matrices $\mathbf{U}$ and

$\mathbf{\Psi}$ of sizes $n \times p$ and $p \times p$, respectively. This decomposition can be formulated as the following LS optimization problem:

$$\min_{\mathbf{U},\mathbf{\Psi}} ||\mathbf{E}_{SVD} - \mathbf{U}\mathbf{\Psi}||^2 \quad , \tag{18}$$

subject to the constraints $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$ and $\mathbf{\Psi}$ being a diagonal matrix. In addition, $\mathbf{U}$ should be orthogonal to $\mathbf{F}$ already found in (13).

Alternatively, if the PCA decomposition (17) is used, for an EFA-like decomposition the following optimization problem needs to be solved:

$$\min_{\mathbf{U},\mathbf{\Psi}} ||\mathbf{E}_{QR} - \mathbf{U}\mathbf{\Psi}||^2 \quad , \tag{19}$$

subject to $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$, $\mathbf{U}^\top\mathbf{F} = \mathbf{O}_{p\times k}$ and $\mathbf{\Psi}$ being a diagonal matrix. Note, that the principal components $\mathbf{F}$ have already been found.

Thus, once an initial PCA solution is available from either the SVD or the QR factorization, i.e. either $\mathbf{Z} \approx \mathbf{F}\mathbf{\Lambda}^\top + \mathbf{E}$ or $\mathbf{Z} \approx \mathbf{F}\mathbf{L}^\top + \mathbf{E}$ are available, then, in both cases, the following Procrustes-like problem:

$$\min_{\mathbf{U},\mathbf{\Psi}} ||\mathbf{E} - \mathbf{U}\mathbf{\Psi}||^2 \quad , \tag{20}$$

should be solved, subject to $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$ and $\mathbf{\Psi}$ being a diagonal matrix. In addition, $\mathbf{U}$ should be orthogonal to $\mathbf{F}$, i.e. $\mathbf{F}^\top\mathbf{U} = \mathbf{O}_{k\times p}$.

4.1.2 $p \geq n$

In this case, the EFA error term in (10) has the same form $\mathbf{U}\mathbf{\Psi}$, but instead of the classical $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$, a new constraint $\mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n$ is imposed. Also, $\mathbf{\Psi}$ can have at most $n - k$ nonzero entries.

Once an initial PCA solution is available, i.e. either $\mathbf{Z}^\top = \mathbf{F}\mathbf{L}^\top + \mathbf{E}$ or $\mathbf{Z} = \mathbf{F}\mathbf{L}^\top + \mathbf{E}$ is available, then the following Procrustes-like problem:

$$\min_{\mathbf{U},\mathbf{\Psi}} \left|\left|\mathbf{E}^\top - \mathbf{\Psi}\mathbf{U}^\top\right|\right|^2 \quad , \tag{21}$$

should be solved, subject to $\mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n$ and $\mathbf{\Psi}$ being diagonal. In addition, $\mathbf{U}$ should be orthogonal to $\mathbf{F}$, i.e. $\mathbf{F}^\top\mathbf{U} = \mathbf{O}_{k\times p}$.

4.2 Iterative algorithm

The application of an EFA-like PCA to $\mathbf{Z}$ requires an efficient method for solving the Procrustes-like problem (20). The major problem with transforming (20) into a standard Procrustes problem is that $\mathbf{E}$ is always, by construction, rank deficient as $\mathbf{F}^\top\mathbf{E} = \mathbf{O}_{k\times p}$.

Let the QR decomposition of $\mathbf{F}$ be written as follows:

$$\mathbf{F} = \mathbf{QR} = \underbrace{\begin{bmatrix} \mathbf{F} & \mathbf{F}_\perp \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \mathbf{I}_k \\ \mathbf{O}_{(n-k)\times k} \end{bmatrix}}_{\mathbf{R}} \text{ with } \mathbf{Q}^\top\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_n. \tag{22}$$

Then the objective functions (20) and (21) can be transformed into

$$||\mathbf{E} - \mathbf{U}\mathbf{\Psi}||^2 = ||\mathbf{Q}^\top(\mathbf{E} - \mathbf{U}\mathbf{\Psi})||^2 = \left|\left| \begin{array}{c} \mathbf{F}^\top(\mathbf{E} - \mathbf{U}\mathbf{\Psi}) \\ \mathbf{F}_\perp^\top(\mathbf{E} - \mathbf{U}\mathbf{\Psi}) \end{array} \right|\right|^2 = ||\mathbf{F}_\perp^\top(\mathbf{E} - \mathbf{U}\mathbf{\Psi}||^2 \ . \tag{23}$$

4.2.1 $n > p$

In this case, the unknown $\mathbf{U}$ in (23) is sought in the form $\mathbf{U} = \mathbf{F}_\perp\widetilde{\mathbf{U}}$, where $\widetilde{\mathbf{U}}$ is a $(n-k)\times p$ orthonormal matrix with $\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{U}} = \mathbf{I}_p$. Clearly, the constraint $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$ is fulfilled, as $\mathbf{U}^\top\mathbf{U} = \widetilde{\mathbf{U}}^\top\mathbf{F}_\perp^\top\mathbf{F}_\perp\widetilde{\mathbf{U}} = \widetilde{\mathbf{U}}^\top\widetilde{\mathbf{U}} = \mathbf{I}_p$. Making use of the transformed objective function (23) and denoting $\mathbf{E} = \mathbf{F}_\perp\widetilde{\mathbf{E}}$, the Procrustes-like problem (20) reduces to a standard Procrustes problem:

$$\min_{\widetilde{\mathbf{U}},\mathbf{\Psi}} ||\widetilde{\mathbf{E}} - \widetilde{\mathbf{U}}\mathbf{\Psi}||^2 \ , \tag{24}$$

subject to $\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{U}} = \mathbf{I}_p$ and $\mathbf{\Psi}$ being a diagonal matrix. Hence (20) can be solved by an alternating procedure: solve for $\tilde{\mathbf{U}}$ the standard Procrustes problem (24) for fixed $\mathbf{\Psi}$, and then update $\mathbf{\Psi}$ using $\mathbf{\Psi} = \text{diag}(\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{E}}) = \text{diag}(\mathbf{U}^\top\mathbf{E})$; continue until convergence.

4.2.2 $p \geq n$

In this case, the unknown $\mathbf{U}$ in (23) is sought again in the form $\mathbf{U} = \mathbf{F}_\perp \widetilde{\mathbf{U}}$, but here $\widetilde{\mathbf{U}}$ is a $(n-k) \times p$ orthonormal matrix with $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^\top = \mathbf{I}_{n-k}$. One can check that $\mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{F}\mathbf{F}^\top + \mathbf{F}_\perp \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^\top \mathbf{F}_\perp^\top = \mathbf{F}\mathbf{F}^\top + \mathbf{F}_\perp \mathbf{F}_\perp^\top = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_n$, i.e. the new EFA constraint is fulfilled. Note also that $\mathbf{\Psi} = \mathbf{U}^\top \mathbf{U} \mathbf{\Psi} = \widetilde{\mathbf{U}}^\top \mathbf{F}_\perp^\top \mathbf{F}_\perp \widetilde{\mathbf{U}} \mathbf{\Psi} = \widetilde{\mathbf{U}}^\top \widetilde{\mathbf{U}} \mathbf{\Psi}$.

Then, making use of (23) and $\mathbf{E} = \mathbf{F}_\perp \widetilde{\mathbf{E}}$, the Procrustes-like problem (21) is reduced to the following standard Procrustes problem:

$$\min_{\widetilde{\mathbf{U}}, \mathbf{\Psi}} ||\widetilde{\mathbf{E}} - \widetilde{\mathbf{U}} \mathbf{\Psi}||^2 \ , \tag{25}$$

subject to $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^\top = \mathbf{I}_{n-k}$ and $\mathbf{\Psi}$ being diagonal. Hence (21) can be solved by an alternating procedure: solve for $\tilde{\mathbf{U}}$ the standard Procrustes problem (25) for fixed $\mathbf{\Psi}$, and then update $\mathbf{\Psi}$ using $\mathbf{\Psi} = \text{diag}(\widetilde{\mathbf{U}}^\top \widetilde{\mathbf{E}}) = \text{diag}(\mathbf{U}^\top \mathbf{E})$; continue until convergence.

4.3 EFA optimality conditions

4.3.1 $n > p$

Let the EFA problem (6) – (7) be rewritten as:

$$f(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{F}, \mathbf{U}) = \left|\left| \mathbf{Z} - [\mathbf{F} \ \ \mathbf{U}] \begin{bmatrix} \mathbf{\Lambda}^\top \\ \mathbf{\Psi} \end{bmatrix} \right|\right|^2 , \tag{26}$$

subject to $\mathbf{F}^\top \mathbf{F} = \mathbf{I}_k$, $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_p$, $\mathbf{U}^\top \mathbf{F} = \mathbf{O}_{p \times k}$, and $\mathbf{\Psi}$ being a $p \times p$ diagonal matrix. This reformulation helps the understanding of the ALS approach to EFA. Indeed, by defining the block matrices $\mathbf{B} = [\mathbf{F} \ \ \mathbf{U}]$ and $\mathbf{A} = [\mathbf{\Lambda} \ \ \mathbf{\Psi}]$ with dimensions $n \times (k+p)$ and $p \times (k+p)$, respectively, (26) is rewritten as

$$f = \left|\left| \mathbf{Z} - \mathbf{B}\mathbf{A}^\top \right|\right|^2 = \text{trace}(\mathbf{Z}^\top \mathbf{Z}) + \text{trace}(\mathbf{A}\mathbf{A}^\top) - 2 \ \text{trace}(\mathbf{B}^\top \mathbf{Z}\mathbf{A}), \tag{27}$$

where $\mathbf{B}$ is orthonormal, as

$$\mathbf{B}^\top \mathbf{B} = \begin{bmatrix} \mathbf{F}^\top \\ \mathbf{U}^\top \end{bmatrix} \begin{bmatrix} \mathbf{F} \ \ \mathbf{U} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_k & \mathbf{O}_{k \times p} \\ \mathbf{O}_{p \times k} & \mathbf{I}_p \end{bmatrix} = \mathbf{I}_{k+p} \ .$$

Thus, for available $\mathbf{A} = [\mathbf{\Lambda} \;\; \mathbf{\Psi}]$, the minimization of f in (27) requires solution of a standard Procrustes problem for $\mathbf{B}$. Then, after $\mathbf{B} = [\mathbf{F} \;\; \mathbf{U}]$ is found, update $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ and $\mathbf{\Psi} = \mathrm{diag}(\mathbf{Z}^\top\mathbf{U})$, and continue until convergence.

First-order optimality conditions for the EFA unknowns $\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{F}$ and $\mathbf{U}$ can be found by setting to zero the projected gradients (Trendafilov 2006) of f in (27) with respect to $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{B}^\top\mathbf{Z}\mathbf{A} = \mathbf{A}^\top\mathbf{Z}^\top\mathbf{B} \;, \tag{28}$$

$$(\mathbf{I}_n - \mathbf{B}\mathbf{B}^\top)\mathbf{Z}\mathbf{A} = \mathbf{O}_{n\times(k+p)} \;, \tag{29}$$

$$\mathbf{A} = \mathbf{Z}^\top\mathbf{B} \;. \tag{30}$$

The last condition (30) simply reduces to $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ and $\mathbf{\Psi} = \mathbf{Z}^\top\mathbf{U}$. The condition (28) is equivalent to the following three, which hold simultaneously:

$$\mathbf{F}^\top\mathbf{Z}\mathbf{\Lambda} = \mathbf{\Lambda}^\top\mathbf{Z}^\top\mathbf{F} \;,\; \mathbf{U}^\top\mathbf{Z}\mathbf{\Lambda} = \mathbf{\Psi}\mathbf{Z}^\top\mathbf{F} \;,\; \mathbf{U}^\top\mathbf{Z}\mathbf{\Psi} = \mathbf{\Psi}\mathbf{Z}^\top\mathbf{U} \;. \tag{31}$$

Substitution of $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ in the first condition of (31) shows that $\mathbf{F}^\top\mathbf{Z}\mathbf{\Lambda} = \mathbf{\Lambda}^\top\mathbf{Z}^\top\mathbf{F}$ is trivially fulfilled. Substitution of $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ and $\mathbf{\Psi} = \mathbf{Z}^\top\mathbf{U}$ in the second condition of (31) gives $\mathbf{U}^\top\mathbf{Z}\mathbf{\Lambda} = \mathbf{\Psi}\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{U}\mathbf{\Lambda}$. As $\mathbf{\Lambda}$ is a full column-rank matrix, the equation $(\mathbf{U}^\top\mathbf{Z} - \mathbf{Z}^\top\mathbf{U})\mathbf{\Lambda} = \mathbf{O}_{p\times k}$ does not have unique (zero) solution for $\mathbf{U}^\top\mathbf{Z} - \mathbf{Z}^\top\mathbf{U}$, i.e. $\mathbf{U}^\top\mathbf{Z}$ is not necessarily symmetric. Indeed, $\mathbf{U}^\top\mathbf{Z} - \mathbf{Z}^\top\mathbf{U} = \mathbf{Q}(\mathbf{I}_p - (\mathbf{\Lambda}^\top\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^\top)$, where $\mathbf{Q}$ is an arbitrary $p \times p$ matrix. As $\mathbf{\Psi}$ is assumed (required) diagonal, the third condition in (31) $\mathbf{U}^\top\mathbf{Z}\mathbf{\Psi} = \mathbf{\Psi}\mathbf{Z}^\top\mathbf{U}$ implies that $\mathbf{U}^\top\mathbf{Z}$ should be diagonal, provided $\mathbf{\Psi}$ is nonsingular diagonal with different entries, i.e. $\mathbf{\Psi} \neq \alpha\mathbf{I}_n$ for some $\alpha \in \mathbb{R}$.

Finally, the optimality condition (29) expressed as

$$(\mathbf{I}_n - \mathbf{F}\mathbf{F}^\top - \mathbf{U}\mathbf{U}^\top)\mathbf{Z}\mathbf{A} = \mathbf{O}_{n\times(k+p)} \;,$$

results in two optimality conditions to hold simultaneously:

$$(\mathbf{I}_n - \mathbf{F}\mathbf{F}^\top - \mathbf{U}\mathbf{U}^\top)\mathbf{Z}\mathbf{\Lambda} = \mathbf{O}_{n\times k} \tag{32}$$

and

$$(\mathbf{I}_n - \mathbf{F}\mathbf{F}^\top - \mathbf{U}\mathbf{U}^\top)\mathbf{Z}\boldsymbol{\Psi} = \mathbf{O}_{n\times p} \ . \tag{33}$$

By making use of $\boldsymbol{\Lambda}^\top = \mathbf{F}^\top\mathbf{Z}$ and $\boldsymbol{\Psi} = \mathbf{U}^\top\mathbf{Z}$, and assuming $\boldsymbol{\Psi}$ nonsingular, the conditions (32) and (33) turn into

$$(\mathbf{Z} - \mathbf{F}\boldsymbol{\Lambda}^\top - \mathbf{U}\boldsymbol{\Psi})\boldsymbol{\Lambda} = \mathbf{O}_{n\times k} \tag{34}$$

and

$$\mathbf{Z} = \mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{U}\boldsymbol{\Psi} \ . \tag{35}$$

4.3.2 $p \geq n$

In this case the EFA objective function (26) should be minimized subject to the new constraints $\text{rank}(\mathbf{F}) = k, \mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n$ and $\mathbf{U}^\top\mathbf{U}\boldsymbol{\Psi} = \boldsymbol{\Psi}$. Denote again $\mathbf{B} = [\mathbf{F} \ \ \mathbf{U}]$ and $\mathbf{A} = [\boldsymbol{\Lambda} \ \ \boldsymbol{\Psi}]$. Then, for given $\mathbf{A}$, the minimization of (26) requires solution of the following problem:

$$\min_{\mathbf{B}} \left|\left|\mathbf{Z} - \mathbf{B}\mathbf{A}^\top\right|\right|_F^2 \ , \text{ subject to } \mathbf{B}\mathbf{B}^\top = \mathbf{F}\mathbf{F}^\top + \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n \ , \tag{36}$$

which is again a standard Procrustes problem, i.e. its solution is found by maximizing $\text{trace}(\mathbf{B}^\top\mathbf{Z}\mathbf{A})$. Indeed, the loss function $\left|\left|\mathbf{Z} - \mathbf{B}\mathbf{A}^\top\right|\right|_F^2 = ||\mathbf{Z}||_F^2 + \text{trace}(\mathbf{B}^\top\mathbf{B}\mathbf{A}^\top\mathbf{A}) - 2\ \text{trace}(\mathbf{B}^\top\mathbf{Z}\mathbf{A})$ contains two terms depending on $\mathbf{B}$, because $\mathbf{B}^\top\mathbf{B}$ is not an identity matrix as in (27). Nevertheless, making use of the constraint $\mathbf{U}^\top\mathbf{U}\boldsymbol{\Psi} = \boldsymbol{\Psi}$, Trendafilov and Unkel (2009) show that $\text{trace}(\mathbf{B}^\top\mathbf{B}\mathbf{A}^\top\mathbf{A}) = \text{trace}(\boldsymbol{\Lambda}^\top\boldsymbol{\Lambda}) + \text{trace}(\boldsymbol{\Psi}^2) = \text{trace}(\mathbf{A}^\top\mathbf{A})$. Thus, the solution of (36) is equivalent to the maximization of $\text{trace}(\mathbf{B}^\top\mathbf{Z}\mathbf{A})$ and simply requires the SVD of $\mathbf{A}^\top\mathbf{Z}^\top$. After solving (36) for $\mathbf{B} = [\mathbf{F} \ \ \mathbf{U}]$, the values of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ are updated making use of $\boldsymbol{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ and $\boldsymbol{\Psi} = \text{diag}(\mathbf{U}^\top\mathbf{Z})$.

As above, first-order optimality conditions for the EFA unknowns $\boldsymbol{\Lambda}, \boldsymbol{\Psi}, \mathbf{F}$ and $\mathbf{U}$ can be found by setting to zero the projected gradients (Trendafilov 2006) of f

in (36) with respect to $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{ZAB}^\top = \mathbf{BA}^\top\mathbf{Z}^\top \ , \tag{37}$$

$$\mathbf{ZA}(\mathbf{I}_{k+p} - \mathbf{B}^\top\mathbf{B}) = \mathbf{O}_{n\times(k+p)} \ , \tag{38}$$

$$\mathbf{A} = \mathbf{Z}^\top\mathbf{B} \ . \tag{39}$$

Most of the optimality conditions are the same as for the classical case $n > p$. Indeed, the last condition (39) gives $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ and $\mathbf{\Psi} = \mathbf{Z}^\top\mathbf{U}$. The conditions following from (37) are:

$$\mathbf{F}^\top\mathbf{Z\Lambda} = \mathbf{\Lambda}^\top\mathbf{Z}^\top\mathbf{F} \ , \mathbf{U}^\top\mathbf{Z\Lambda} = \mathbf{\Psi}\mathbf{Z}^\top\mathbf{F} \ , \ \mathbf{U}^\top\mathbf{Z\Psi} = \mathbf{\Psi}\mathbf{Z}^\top\mathbf{U} \ . \tag{40}$$

As with the $n > p$ case, the first optimality condition in (40) is trivial and the second one shows that $\mathbf{U}^\top\mathbf{Z}$ is not necessarily symmetric. The last optimality condition in (40) is more interesting, because $\mathbf{\Psi}$ can have at most $n - k$ nonzero entries if $p \geq n$ (Trendafilov and Unkel 2009). Then, in contrast to the classical case, $\mathbf{U}^\top\mathbf{Z\Psi} = \mathbf{\Psi}\mathbf{Z}^\top\mathbf{U}$ *cannot* imply $\mathbf{U}^\top\mathbf{Z}$ to be diagonal. Instead, it implies that $\mathbf{U}^\top\mathbf{Z}$ has identically zero rows corresponding to the zero entries of $\mathbf{\Psi}$.

The optimality condition (38) is equivalent to:

$$\mathbf{Z}\ [\mathbf{\Lambda}\ \mathbf{\Psi}] \begin{bmatrix} \mathbf{I}_k - \mathbf{F}^\top\mathbf{F} & \mathbf{F}^\top\mathbf{U} \\ \mathbf{UF}^\top & \mathbf{I}_p - \mathbf{U}^\top\mathbf{U} \end{bmatrix} = \mathbf{O}_{n\times(k+p)} \ ,$$

which, making use of $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$ and $\mathbf{F}^\top\mathbf{U} = \mathbf{O}_{k\times p}$, results in the following optimality condition:

$$\mathbf{Z\Psi}(\mathbf{I}_p - \mathbf{U}^\top\mathbf{U}) = \mathbf{Z}(\mathbf{\Psi} - \mathbf{\Psi}\mathbf{U}^\top\mathbf{U}) = \mathbf{O}_{n\times p} \ . \tag{41}$$

The equation (41) shows that $\mathbf{\Psi} - \mathbf{\Psi}\mathbf{U}^\top\mathbf{U} = (\mathbf{I}_p - \mathbf{Z}^\top(\mathbf{ZZ}^\top)^{-1}\mathbf{Z})\mathbf{X}$ for any $p \times p$ matrix $\mathbf{X}$. Thus, indeed, the new EFA constraint $\mathbf{\Psi} = \mathbf{\Psi}\mathbf{U}^\top\mathbf{U}$ needs to be imposed in advance.

If the loadings matrix $\mathbf{\Lambda}$ in the EFA decomposition (26) is required to be a lower triangular $\mathbf{L}$, then one simply needs to replace the updating formula by $\mathbf{L} = \texttt{tril}(\mathbf{Z}^\top\mathbf{F})$, and $\mathbf{\Lambda}$ by $\mathbf{L}$ in all expressions.

In the next Section it is shown that the conditions (34) and $\mathbf{U}^\top\mathbf{Z}\boldsymbol{\Lambda} = \boldsymbol{\Psi}\mathbf{Z}^\top\mathbf{F}$ from (40) are crucial for distinguishing between EFA and PCA solutions when $n > p$ and $p \geq n$ respectively.

4.4 EFA-like PCA optimality conditions

4.4.1 $n > p$

The first order optimality conditions for $\widetilde{\mathbf{U}}$ and $\boldsymbol{\Psi}$ that minimize (24) are:

$$\begin{aligned}
&\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{E}}\boldsymbol{\Psi} = \boldsymbol{\Psi}\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{E}} \ , \\
&(\mathbf{I}_{n-k} - \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^\top)\widetilde{\mathbf{E}}\boldsymbol{\Psi} = \mathbf{O}_{(n-k)\times p} \ , \\
&\boldsymbol{\Psi} = \mathrm{diag}(\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{E}}) \ .
\end{aligned}$$

Then, the optimality conditions for $\mathbf{U}$ and $\boldsymbol{\Psi}$ that minimize (20) are obtained by making use of the transformations $\widetilde{\mathbf{U}} = \mathbf{F}_\perp^\top\mathbf{U}$ and $\widetilde{\mathbf{E}} = \mathbf{F}_\perp^\top\mathbf{E}$:

$$\begin{aligned}
&\mathbf{U}^\top\mathbf{E}\boldsymbol{\Psi} = \boldsymbol{\Psi}\mathbf{E}^\top\mathbf{U} \ , \\
&\mathbf{F}_\perp^\top(\mathbf{I}_n - \mathbf{U}\mathbf{U}^\top)\mathbf{E}\boldsymbol{\Psi} = \mathbf{O}_{(n-k)\times p} \ , \\
&\boldsymbol{\Psi} = \mathrm{diag}(\mathbf{U}^\top\mathbf{E}) \ ,
\end{aligned}$$

which, after substitution of $\mathbf{E}$ from (14) or (17), are equivalent to:

$$\mathbf{U}^\top\mathbf{Z}\boldsymbol{\Psi} = \boldsymbol{\Psi}\mathbf{Z}^\top\mathbf{U} \ , \tag{42}$$

$$\mathbf{F}_\perp^\top(\mathbf{I}_n - \mathbf{U}\mathbf{U}^\top)\mathbf{Z}\boldsymbol{\Psi} = \mathbf{O}_{(n-k)\times p} \ , \tag{43}$$

$$\boldsymbol{\Psi} = \mathrm{diag}(\mathbf{U}^\top\mathbf{Z}) \ . \tag{44}$$

Condition (42) implies that $\mathbf{U}^\top\mathbf{Z}\boldsymbol{\Psi}$ is not necessarily symmetric. If $\boldsymbol{\Psi}$ is non-singular, then $\mathbf{U}^\top\mathbf{Z}$ is necessarily diagonal, provided $\boldsymbol{\Psi} \neq \alpha\mathbf{I}_n$ for some $\alpha \in \mathbb{R}$.

The conditions (42) and (44) together with the condition $\boldsymbol{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ or $\mathbf{L} = \mathbf{Z}^\top\mathbf{F}$, depending on the type of PCA adopted, i.e. either (13) or (15), are identical to the optimality conditions for simultaneous EFA.

If $\mathbf{\Psi}$ is nonsingular, then the condition (43) is equivalent to:

$$(\mathbf{I}_n - \mathbf{U}\mathbf{U}^\top)\mathbf{Z} = \mathbf{F}\mathbf{X}^\top \ , \tag{45}$$

where $\mathbf{X}$ is an arbitrary full column rank $p \times k$ matrix. This is a very loose optimality condition compared to the corresponding EFA optimality condition (35) rewritten as

$$(\mathbf{I}_n - \mathbf{U}\mathbf{U}^\top)\mathbf{Z} = \mathbf{F}\mathbf{\Lambda}^\top \ .$$

This marked difference reflects the fact that in PCA, and in the EFA-like PCA, the pair $\{\mathbf{F}, \mathbf{\Lambda}\}$ is found first with no consideration for $\{\mathbf{U}, \mathbf{\Psi}\}$, and similarly, the pair $\{\mathbf{U}, \mathbf{\Psi}\}$ is found independently from $\{\mathbf{F}, \mathbf{\Lambda}\}$ and their only relation is $\mathbf{F}^\top\mathbf{U} = \mathbf{O}_{k\times p}$.

The EFA optimality condition (34) does not have an equivalent in EFA-like PCA. Thus, the EFA and the PCA solutions will be similar if the EFA-like PCA solution meets the EFA optimality condition (34). This can be readily evaluated by checking the deviation of $(\mathbf{Z} - \mathbf{F}\mathbf{\Lambda}^\top - \mathbf{U}\mathbf{\Psi})\mathbf{\Lambda}$ from $\mathbf{O}_{n\times k}$. In particular, this accomplishes Rao's suggestion to examine "the actual differences between principal components and factor scores" for individual data sets with $n > p$.

4.4.2 $p \geq n$

The first order optimality conditions for $\widetilde{\mathbf{U}}$ and $\mathbf{\Psi}$ that minimize (25) are:

$$\begin{aligned}
&\widetilde{\mathbf{U}}\mathbf{\Psi}\widetilde{\mathbf{E}}^\top = \widetilde{\mathbf{E}}\mathbf{\Psi}\widetilde{\mathbf{U}}^\top \ , \\
&\widetilde{\mathbf{E}}\mathbf{\Psi}(\mathbf{I}_p - \widetilde{\mathbf{U}}^\top\widetilde{\mathbf{U}}) = \mathbf{O}_{(n-k)\times p} \ , \\
&\mathbf{\Psi} = \mathrm{diag}(\widetilde{\mathbf{U}}^\top\widetilde{\mathbf{E}}) \ .
\end{aligned}$$

Note, that the first optimality condition transforms into $\mathbf{\Psi}\widetilde{\mathbf{E}}^\top\widetilde{\mathbf{U}} = \mathbf{\Psi}\widetilde{\mathbf{E}}^\top\widetilde{\mathbf{U}}$ after left multiplication by $\widetilde{\mathbf{U}}^\top$, right multiplication by $\widetilde{\mathbf{U}}$, and making use of $\mathbf{\Psi} = \widetilde{\mathbf{U}}^\top\widetilde{\mathbf{U}}\mathbf{\Psi}$. Then, the optimality conditions for $\mathbf{U}$ and $\mathbf{\Psi}$ that minimize (21)

are obtained by making use of $\widetilde{\mathbf{U}} = \mathbf{F}_{\perp}^{\top}\mathbf{U}$ and $\widetilde{\mathbf{E}} = \mathbf{F}_{\perp}^{\top}\mathbf{E}$:

$$\begin{aligned}
&\boldsymbol{\Psi}\mathbf{E}^{\top}\mathbf{U} = \boldsymbol{\Psi}\mathbf{E}^{\top}\mathbf{U} \ , \\
&\mathbf{F}_{\perp}^{\top}\mathbf{E}(\boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{U}^{\top}\mathbf{U}) = \mathbf{O}_{(n-k)\times p} \ , \\
&\boldsymbol{\Psi} = \mathrm{diag}(\mathbf{U}^{\top}\mathbf{E}) \ ,
\end{aligned}$$

which, after substitution of $\mathbf{E}$ from (14) or (17), are equivalent to:

$$\mathbf{U}^{\top}\mathbf{Z}\boldsymbol{\Psi} = \boldsymbol{\Psi}\mathbf{Z}^{\top}\mathbf{U} \ , \tag{46}$$

$$\mathbf{F}_{\perp}^{\top}\mathbf{Z}(\boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{U}^{\top}\mathbf{U}) = \mathbf{O}_{(n-k)\times p} \ , \tag{47}$$

$$\boldsymbol{\Psi} = \mathrm{diag}(\mathbf{U}^{\top}\mathbf{Z}) \ . \tag{48}$$

The conditions (46) and (48) are the same as (42) and (44) for the classic $n > p$ case. The condition (47) is new and reflects the fact that $p \geq n$ and that the condition $\boldsymbol{\Psi} = \boldsymbol{\Psi}\mathbf{U}^{\top}\mathbf{U}$ is required.

The conditions (46) and (48) together with the condition $\boldsymbol{\Lambda} = \mathbf{Z}^{\top}\mathbf{F}$ or $\mathbf{L} = \mathbf{Z}^{\top}\mathbf{F}$, depending on the type PCA adopted, i.e. either (13) or (15), are identical to the optimality conditions for simultaneous EFA.

The condition (47) is equivalent to:

$$\mathbf{Z}(\boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{U}^{\top}\mathbf{U}) = \mathbf{F}\mathbf{X}^{\top} \ , \tag{49}$$

where $\mathbf{X}$ is an arbitrary full column rank $p \times k$ matrix. This is a very loose version of the corresponding EFA optimality condition (41), reflecting the fact that in PCA $\mathbf{F}$ is found before the specific error term decomposition is sought.

The EFA optimality condition $\mathbf{U}^{\top}\mathbf{Z}\boldsymbol{\Lambda} = \boldsymbol{\Psi}\mathbf{Z}^{\top}\mathbf{F}$ from (40) does not have an equivalent in EFA-like PCA. Thus, the EFA and the PCA solutions will be similar if the corresponding EFA-like PCA solution meets the EFA optimality condition $\mathbf{U}^{\top}\mathbf{Z}\boldsymbol{\Lambda} = \boldsymbol{\Psi}\mathbf{Z}^{\top}\mathbf{F}$. This can be readily evaluated by checking the deviation of $\mathbf{U}^{\top}\mathbf{Z}\boldsymbol{\Lambda} - \boldsymbol{\Psi}\mathbf{Z}^{\top}\mathbf{F}$ from $\mathbf{O}_{n\times k}$. As in Section 4.4.1 for $n > p$, this accomplishes Rao's suggestion to examine "the actual differences between principal components and factor scores" for individual data sets with $p \geq n$.

5 Numerical illustrations

We illustrate the developed algorithms on two data sets, the first for the case $n > p$ and the second for $p \geq n$.

5.1 Harman's five socio-economic variables data

This is the well-known data set (Harman 1976, Table 2.1, p. 14) containing $n = 12$ observations and $p = 5$ variables. The twelve observations are census tracts - small areal subdivisions of the city of Los Angeles. The five socio-economic variables are 'total population' (POPULATION), 'median school years' (SCHOOL), 'total employment' (EMPLOYMENT), 'miscellaneous professional services' (SERVICES) and 'median house value' (HOUSE).

The raw data were first preprocessed to produce variables having zero mean and unit length in a 12×5 matrix $\mathbf{Z}$, and this matrix was subjected to analysis using an EFA model with two common factors ($k = 2$) in terms of different LS loss functions.

Then, LS solutions for estimating $\{\mathbf{F}, \mathbf{\Lambda}, \mathbf{U}, \mathbf{\Psi}\}$ simultaneously were obtained by making use of the new ALS algorithms proposed in (Trendafilov and Unkel 2009). To reduce the chance of mistaking a locally optimal solution for a globally optimal one, the algorithm was run twenty times using different randomly chosen column-wise orthonormal matrices $\mathbf{F}$ and $\mathbf{U}$ as starting values. The algorithm was stopped when successive function values differed by less than $\epsilon = 10^{-6}$. Estimates of $\{\mathbf{\Lambda}, \mathbf{\Psi}^2\}$ applying two parameterizations for the loadings are provided in Table 1.

The results reported are the 'best' obtained after the twenty random starts. By 'best' is meant the solution employing the full column rank (FCR) preserving formula $\mathbf{\Lambda} = \mathbf{Z}^\top \mathbf{F}$ that most resembles the lower triangular (LT) one. The iterative algorithm gives the same goodness-of-fit and quite similar $\mathbf{\Psi}^2$ for both types of loadings .

For both algorithms the twenty runs led to the same minimum of the EFA

	FCR ($\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$) error of fit = .002835			LT ($\mathbf{L} = \text{tril}(\mathbf{Z}^\top\mathbf{F})$) error of fit = .002836		
Variable	$\mathbf{\Lambda}$		$\mathbf{\Psi}^2$	$\mathbf{L}$		$\mathbf{\Psi}^2$
POPULATION	.99	.05	.0150	1.00	0	.0173
SCHOOL	-.01	.88	.2292	.03	.88	.2307
EMPLOYMENT	.97	.16	.0182	.98	.11	.0158
SERVICES	.40	.80	.2001	.44	.78	.2009
HOUSE	-.03	.98	.0318	.02	.98	.0292

Table 1: Simultaneous EFA solutions for Harman's five socio-economic variables data.

objective function up to the fourth decimal place. Numerical experiments revealed that the algorithm employing a lower triangular matrix $\mathbf{L}$ is slower but yields pretty stable loadings. In contrast, the algorithm employing $\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ is faster, but converges to quite different $\mathbf{\Lambda}$.

It is of interest to compare the simultaneous EFA solutions in Table 1 with the ones obtained by means of EFA-like PCA based on the SVD and the QR decomposition of $\mathbf{Z}$. The algorithms described in Section 4.2 were run twenty times each and were stopped when successive function values differed by less than $\epsilon = 10^{-6}$. The initial value for $\mathbf{\Psi}$ was simply taken to be a diagonal matrix with diagonal entries randomly drawn from a uniform distribution on the unit interval. The EFA-like solutions obtained from the two types of PCA are shown in Table 2.

The most striking feature of both EFA-like PCA solutions compared to the simultaneous EFA solutions given in Table 1 is that the fits attained by the former are considerably worse.

One can further assess the difference between the simultaneous EFA solutions and the EFA-like PCA solutions by substituting them into the optimality condition

	SVD error of fit = .059281			QR decomposition error of fit = .029820		
Variable	$\mathbf{\Lambda}$		$\mathbf{\Psi}^2$	$\mathbf{L}$		$\mathbf{\Psi}^2$
POPULATION	.58	.81	.0000	1.00	0	.0000
SCHOOL	.77	-.54	.0945	.01	1.00	.0000
EMPLOYMENT	.67	.73	.0095	.97	.14	.0314
SERVICES	.93	-.10	.1019	.44	.69	.3114
HOUSE	.79	-.56	.0055	.02	.86	.2211

Table 2: EFA-like solutions for Harman's five socio-economic variables data.

(34) which is not satisfied for EFA-like PCA, that is, by calculating:

$$E_{n>p} = ||(\mathbf{Z} - \mathbf{F}\mathbf{\Lambda}^\top - \mathbf{U}\mathbf{\Psi})\mathbf{\Lambda}|| \ . \tag{50}$$

The value of the error $E_{n>p}$ in (50) is 0.4334 for EFA-like PCA based on the SVD and 0.1884 for EFA-like PCA based on the QR factorization. For comparison, the values of $E_{n>p}$ for the corresponding simultaneous EFA solutions (FCR and LT in Table 1) are 6.7578×10^{-4} and 6.9207×10^{-4}.

The conclusion is that for Harman's data PCA differs considerably from EFA, with the QR based EFA-like PCA being closer to the EFA solution. The numerical experiments show that both EFA-like PCA procedures are faster than the iterative algorithms for simultaneous EFA parameter estimation.

Estimated common factor scores are shown in Table 3. For both parameterizations of the loadings, the next two pairs of columns are the factor scores $\mathbf{F}_{\mathbf{\Lambda}}$ and $\mathbf{F}_{\mathbf{L}}$ found by the iterative algorithm for simultaneous parameter estimation. The last two pairs of columns $\mathbf{F}_{SVD}$ and $\mathbf{F}_{QR}$ show component scores obtained by PCA based on the SVD and the QR decomposition, respectively. Note that $\mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$ for all sets of scores. It can be seen from Table 3 that the factor scores $\mathbf{F}_{\mathbf{L}}$ as well as the component scores $\mathbf{F}_{QR}$, all obtained from a lower triangular

parameterization of the loadings matrix, are quite similar.

	Simultaneous EFA				PCA			
#	$\mathbf{F}_{\mathbf{\Lambda}}$		$\mathbf{F}_{\mathbf{L}}$		$\mathbf{F}_{SVD}$		$\mathbf{F}_{QR}$	
1	.04	.38	.02	.38	.29	.21	-.05	.37
2	-.41	-.32	-.41	-.32	-.40	.23	-.46	-.30
3	-.28	-.37	-.27	-.37	-.44	-.03	-.25	-.44
4	-.21	.39	-.20	.38	.14	.39	-.21	.37
5	-.22	.38	-.21	.38	.10	.35	-.20	.37
6	.13	-.25	.12	-.25	-.21	-.35	.17	-.27
7	-.47	-.03	-.48	-.04	-.31	.35	-.44	-.00
8	.25	-.15	.24	-.15	.02	-.26	.25	-.15
9	.25	.04	.25	.04	.24	-.17	.32	.05
10	.26	.37	.27	.37	.57	.02	.29	.38
11	.27	-.25	.27	-.25	-.07	-.40	.29	-.25
12	.39	-.19	.41	-.19	.08	-.35	.28	-.18

Table 3: Common factor scores for Harman's five socio-economic variables data.

5.2 Thurstone's 26-variable box data

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 observational units. The variables of the example are twenty-six functions of these dimensions.

After mean-centering and scaling to have unit norm, the variables form columns of the 20×26 data matrix $\mathbf{Z}$. The first three eigenvalues of the sample correlation matrix $\mathbf{Z}^\top\mathbf{Z}$, in decreasing order, are 12.4217, 7.1807, 5.5386. The fourth one is

much smaller than 1, and thus, according to Kaiser's rule, a solution having three common factors should be sought.

The data matrix $\mathbf{Z}$ was fitted making use of the new iterative algorithms for simultaneous EFA and EFA-like PCA. In contrast to the $n > p$ case, the simultaneous EFA as well as the EFA-like PCA procedures both allow the unique factors to have zero variance. The corresponding solutions are given in Table 4 and Table 5, respectively. Table 4 and Table 5 also show that the fit attained by the EFA-like solutions is worse than by the simultaneous EFA solutions. However the results for the 26-variable box data are much better than for the Harman's data from Section 5.1.

The conclusion is that for the 26-variable box data PCA does not differ much from EFA, with the SVD based EFA-like PCA being quite close to the EFA solution. As with the Harman's data, the EFA-like PCA procedures are faster (at least 5 times) than the simultaneous EFA parameter estimation.

To assess further the difference between the simultaneous EFA solutions and the EFA-like PCA solutions the latter ones can be substituted into the optimality condition $\mathbf{U}^\top \mathbf{Z}\boldsymbol{\Lambda} = \boldsymbol{\Psi}\mathbf{Z}^\top \mathbf{F}$ which is not satisfied for EFA-like PCA, that is, by calculating:

$$E_{p \geq n} = \|\mathbf{U}^\top \mathbf{Z}\boldsymbol{\Lambda} - \boldsymbol{\Psi}\mathbf{Z}^\top \mathbf{F}\| \ . \tag{51}$$

The value of the error $E_{p \geq n}$ in (51) is 0.5130 for EFA-like PCA based on the SVD and 0.9219 for EFA-like PCA based on the QR factorization. For comparison, the values of $E_{p \geq n}$ for the corresponding simultaneous EFA solutions (FCR and LT in Table 4) are 6.7119×10^{-4} and 8.4540×10^{-5}. The uniquenesses of the EFA-like PCA and the ones obtained by simultaneous EFA are very similar.

All algorithms employing a LT parameterization give virtually identical loadings. Moreover, the loadings exhibit an interpretable and contextually meaningful relation between the observed variables and the common factors. If one ignores all loadings with magnitude .25 or less in the LT loading matrices in Table 4 and

Table 5, the remaining loadings perfectly identify which of the box dimensions x, y and z were used to generate each of the variables. Using the results for EFA-like PCA based on the QR decomposition in Table 5, this can be done by ignoring all loadings with magnitudes $\leq .26$.

	$\mathbf{\Lambda} = \mathbf{Z}^\top\mathbf{F}$ error of fit = .175174				$\mathbf{L} = \text{tril}(\mathbf{Z}^\top\mathbf{F})$ error of fit = .175184			
Formula	$\mathbf{\Lambda}$			$\mathbf{\Psi}^2$	$\mathbf{L}$			$\mathbf{\Psi}^2$
x	.98	.12	.17	.0000	1.00	0	0	.0000
y	.15	.98	-.09	.0000	.25	.97	0	.0000
z	-.11	.36	.92	.0000	.10	.23	.96	.0000
xy	.59	.80	.01	.0000	.68	.73	-.00	.0000
xz	.30	.36	.88	.0000	.49	.20	.84	.0000
yz	-.01	.71	.70	.0000	.20	.59	.77	.0000
x^2y	.75	.63	.06	.0191	.82	.54	-.00	.0191
xy^2	.44	.89	-.06	.0001	.52	.84	-.03	.0000
x^2z	.52	.32	.77	.0198	.68	.15	.68	.0198
xz^2	.13	.39	.90	.0000	.33	.24	.90	.0000
y^2z	.05	.82	.52	.0298	.25	.73	.60	.0298
yz^2	-.05	.58	.79	.0000	.16	.45	.85	.0000
x/y	.53	-.81	.16	.0279	.44	-.87	-.05	.0279
y/x	-.54	.80	-.19	.0290	-.46	.87	.02	.0290
x/z	.48	-.22	-.79	.0811	.31	-.15	-.89	.0811
z/x	-.54	.26	.76	.0476	-.36	.20	.88	.0476
y/z	.17	.29	-.90	.0566	.04	.40	-.87	.0566
z/y	-.17	-.27	.90	.0651	-.03	-.38	.88	.0651
$2x + 2y$	.71	.70	.05	.0000	.79	.61	.00	.0000
$2x + 2z$	.59	.32	.74	.0000	.74	.15	.65	.0000
$2y + 2z$	.03	.85	.52	.0000	.23	.76	.61	.0000
$(x^2 + y^2)^{1/2}$	.80	.59	.08	.0000	.87	.49	-.01	.0000
$(x^2 + z^2)^{1/2}$	.80	.26	.53	.0001	.91	.10	.39	.0001
$(y^2 + z^2)^{1/2}$	.07	.93	.34	.0000	.25	.86	.44	.0000
xyz	.27	.67	.67	.0017	.47	.54	.68	.0017
$(x^2 + y^2 + z^2)^{1/2}$	.68	.64	.34	.0001	.80	.52	.28	.0001

Table 4: Simultaneous EFA solutions for Thurstone's 26-variable box data.

	SVD				QR decomposition			
	error of fit = .198038				error of fit = .222473			
Formula	$\mathbf{\Lambda}$			$\mathbf{\Psi}^2$	$\mathbf{L}$			$\mathbf{\Psi}^2$
x	.50	.53	.68	.0000	1.00	0	0	.0000
y	.47	.70	-.53	.0000	.25	.97	0	.0000
z	-.62	.78	-.00	.0000	.10	.23	.97	.0000
xy	.61	.79	-.06	.0000	.68	.73	-.01	.0000
xz	-.34	.89	.27	.0000	.49	.20	.83	.0000
yz	-.29	.92	-.22	.0000	.20	.59	.76	.0000
x^2y	.61	.76	.17	.0157	.82	.54	-.00	.0285
xy^2	.59	.76	-.25	.0014	.52	.84	-.03	.0000
x^2z	-.14	.87	.42	.0177	.68	.15	.67	.0453
xz^2	-.44	.86	.14	.0000	.33	.24	.88	.0000
y^2z	-.08	.92	-.30	.0290	.24	.73	.58	.0571
yz^2	-.42	.87	-.14	.0000	.16	.45	.84	.0000
x/y	-.06	-.30	.94	.0149	.45	-.87	-.04	.0210
y/x	.07	.27	-.95	.0169	-.46	.87	.02	.0196
x/z	.81	-.47	.24	.0613	.31	-.16	-.88	.0768
z/x	-.80	.46	-.31	.0343	-.36	.20	.89	.0124
y/z	.86	-.28	-.34	.0394	.04	.40	-.87	.0560
z/y	-.86	.30	.34	.0459	-.04	-.37	.90	.0349
$2x + 2y$	.61	.78	.09	.0000	.79	.61	.00	0
$2x + 2z$	-.09	.88	.46	.0000	.74	.16	.65	0
$2y + 2z$	-.09	.93	-.34	.0000	.22	.76	.61	0
$(x^2 + y^2)^{1/2}$	.61	.75	.23	.0000	.87	.49	-.01	.0000
$(x^2 + z^2)^{1/2}$	.18	.79	.58	.0001	.91	.10	.40	.0001
$(y^2 + z^2)^{1/2}$	.09	.90	-.42	.0000	.24	.86	.44	.0000
xyz	-.10	.98	-.00	.0016	.46	.53	.67	.0057
$(x^2 + y^2 + z^2)^{1/2}$	.37	.90	.20	.0001	.80	.52	.28	.0001

Table 5: EFA-like solutions for Thurstone's 26-variable box data.

References

Berry, M. W., Pulatova, S. A., and Stewart, G. W. (2005). Algorithm 844: Computing sparse reduced-rank approximations to sparse matrices. *ACM Transactions on Mathematical Software (TOMS)*, **31**, 252–269.

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

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

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, IL, 3rd edition.

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

Mulaik, S. A. (2005). Looking back on the indeterminacy controversies in factor analysis. In A. Maydeu-Olivares and J. J. McArdle, editors, *Contemporary Psychometrics: A Festschift for Roderick P. McDonald*, pages 173–206,. Lawrence Erlbaum: Mahwah.

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

Rao, C. R. (1996). Principal component and factor analyses. In G. S. Maddala and C. R. Rao, editors, *Handbook of Statistics, Vol. 14*, pages 489–505. Elsevier: Amsterdam.

Schneeweiss, H. and Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis*, **55**, 105–124.

Stewart, G. W. (1998). *Matrix Algorithms, Vol I: Basic Decompositions*. John Wiley & Sons: New York, NY.

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

Trendafilov, N. T. (2006). Dynamical system approach to multivariate data analysis. *Journal of Computational and Graphical Statistics*, **15**(3), 628–650.

Trendafilov, N. T. and Unkel, S. (2009). Exploratory factor analysis of data matrices with more variables than observations. *Journal of Computational and Graphical Statistics*, page submitted.