Some inequalities contrasting principal component and factor analyses solutions

Journal Item

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.1007/s42081-018-0024-4

SOME INEQUALITIES CONTRASTING PRINCIPAL COMPONENT AND FACTOR ANALYSES SOLUTIONS

February 14, 2018

KOHEI ADACHI
OSAKA UNIVERSITY

NICKOLAY T. TRENDAFILOV
OPEN UNIVERSITY

Abstract

Principal component analysis (PCA) and factor analysis (FA) are two time-honored dimension reduction methods. In this paper, some inequalities are presented to contrast PCA and FA solutions for the same data set. For this reason, we take advantage of the recently established matrix decomposition (MD) formulation of FA. In summary, the resulting inequalities show that [1] FA gives a better fit to the data than PCA, [2] PCA extracts a larger amount of common "information" than FA, and [3] For each variable, its unique variance in FA is larger than its residual variance in PCA minus the one in FA. The resulting inequalities can be useful to suggest whether PCA or FA should be used for a particular data set. The answers can also be valid for the classic FA formulation not relying on the MD-FA definition, as both "types" FA provide almost equal solutions. Additionally, the inequalities give theoretical explanation of some empirically observed tendencies in PCA and FA solutions, e.g., that the absolute values of PCA loadings tend to be larger than those for FA loadings, and that the unique variances in FA tend to be larger than the residual variances of PCA.

Key words: Matrix decomposition; Dimesion reduction; Common parts; Unique parts; Loadings; Residuals.

1. Introduction

Principal component analysis (PCA) was conceived by Pearson (1901) and formulated by Hotelling (1933) who named the procedure PCA. On the other hand, factor analysis (FA) was proposed by Spearman (1904) and further developed to its modern form as known today by Thurstone (1935). Both procedures are time-honored dimension reduction methods for an n-observations × q-variables column-centered data matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_q]$. Thus, PCA and FA are often performed for an identical data set (e.g., Adachi, 2016; Jolliffe, 2002). Their solutions are compared mathematically and numerically in this paper. Throughout the paper, $n \geq \text{rank}(\mathbf{X}) = q$ is supposed with rank($\mathbf{X}$) denoting the rank of $\mathbf{X}$.

PCA can be formulated in a number of different ways (Okamoto, 1969; ten Berge & Kiers, 1996). One of them is to define PCA as "composing scores by variables", i.e., summing weighted observed variables to provide composite scores (Hotelling, 1933). This formulation of PCA is rather opposite to the FA assumption of "composing variables by scores", i.e., summing the weighted unobserved (factor) scores to provide observed variables. An approach, which is comparable with FA, is to formulate PCA as

$$\mathbf{X} = \mathbf{PC}' + \mathbf{E}_{\text{PC}}, \tag{1}$$

where $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_m]$ is an n-observation × m-components PC score matrix, $\mathbf{C} = (c_{jk})$ is a $q \times m$ component loading matrix, and $\mathbf{E}_{\text{PC}}$ ($n \times q$) contains errors, with $m \leq q$. (e.g., Adachi, 2016). The implication of (1) can be illustrated as in Figure 1(A): in (1) the variables $\mathbf{x}_1, \ldots, \mathbf{x}_q$ are commonly explained by the PC scores $\mathbf{p}_1, \ldots, \mathbf{p}_m$ which are weighted by their loadings c_{jk}, and the errors in $\mathbf{E}_{\text{PC}}$ remain unexplained. In this point, we call $\mathbf{PC}'$ a common part.. The

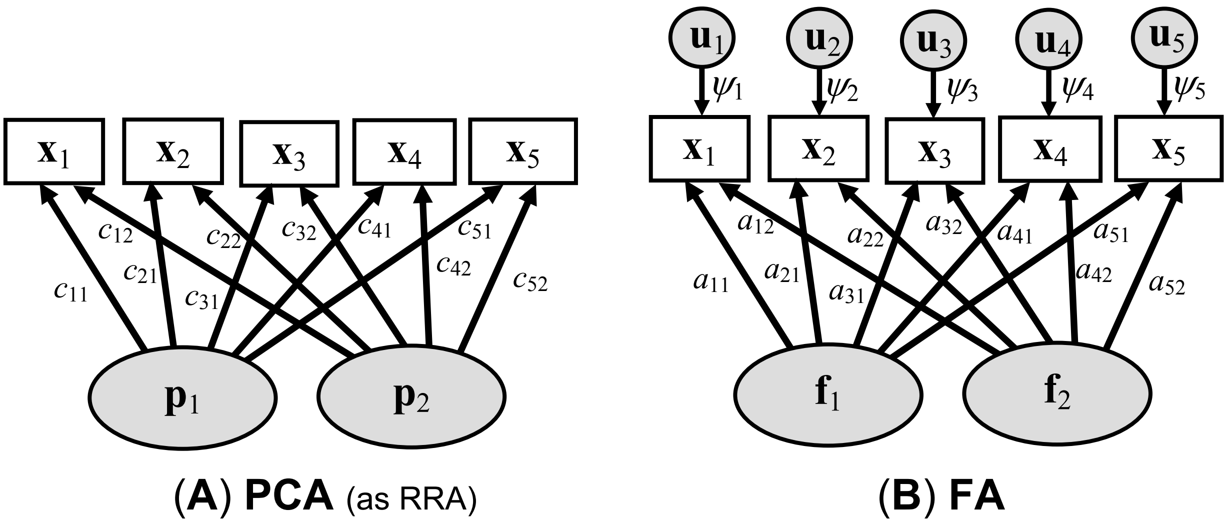

Figure 1. Graphical representation of PCA as reduced rank approximation (RRA) and FA with $q = 5$ and $m = 2$

matrices **P** and **C** that minimize a least square error $||\mathbf{E}_{PC}||^2 = \text{tr}\mathbf{E}_{PC}'\mathbf{E}_{PC} = ||\mathbf{X} - \mathbf{PC}'||^2$ is given through the singular value decomposition (SVD) of **X** (Eckart & Young, 1936). Thus, PCA can be regarded as a matrix decomposition problem for approximating **X** by a lower rank matrix **PC**′ with rank(**PC**′) = $m \leq q$, also known as the truncated SVD.

In a similar manner as PCA, FA can be formulated as a matrix decomposition problem. It was firstly proposed by Henk A. L. Kiers as described in Sočan (2003, pp. 19-20) and recently established (Unkel & Trendafilov, 2010; Stegeman, 2016; Adachi & Trendafilov, 2017). In this formulation, FA is modeled as

$$\mathbf{X} = \mathbf{FA}' + \mathbf{U\Psi} + \mathbf{E}_{FA} . \tag{2}$$

Here, $\mathbf{F} = [\mathbf{f}_1, \ldots , \mathbf{f}_m]$ is the $n \times m$ matrix containing common factor scores, $\mathbf{U} = [\mathbf{u}_1, \ldots , \mathbf{u}_q]$ is the $n \times q$ matrix of unique factor scores, $\mathbf{A} = (a_{jk})$ is a $q \times m$ factor loading matrix, and $\mathbf{\Psi}$ is the $q \times q$ diagonal matrix, the squares of whose diagonal elements $\psi_1, \ldots , \psi_q$ are called unique variances, with $\mathbf{E}_{FA}$ ($n \times q$) containing unsystematic errors. The factor score matrices are constrained as

$$\frac{1}{n}\mathbf{F}'\mathbf{F} = \mathbf{I}_m, \quad \frac{1}{n}\mathbf{U}'\mathbf{U} = \mathbf{I}_q, \quad \text{and} \quad \mathbf{F}'\mathbf{U} = \mathbf{O} \tag{3}$$

with **O** being a matrix of zeros. The implications of (2) can be illustrated in Figure 1(B): **FA**′ is the common part as **PC**′ in (1), while a unique part **UΨ** is added in FA with the *j*th unique factor $\mathbf{u}_j$ being weighted by ψ_j, which affects only (uniquely) the corresponding variable $\mathbf{x}_j$.

In (2), all **F**, **A**, **U**, and **Ψ** are treated as fixed unknown matrices. In contrast, the classic formulation of FA treats the elements of **F** and **U** as the random variables following distributional assumptions associated with (3). Then, the covariance matrix among the columns of **FA**′ + **UΨ** can be expressed as $\mathbf{AA}'+\mathbf{\Psi}^2$, which is supposed to approximate the sample counterpart $\mathbf{S}_{XX} = n^{-1}\mathbf{X}'\mathbf{X}$:

$$\mathbf{S}_{XX} \cong \mathbf{AA}'+\mathbf{\Psi}^2 . \tag{4}$$

In the classic FA, the discrepancy between $\mathbf{S}_{XX}$ and $\mathbf{AA}'+\mathbf{\Psi}^2$ is minimized over **A** and $\mathbf{\Psi}^2$ (e.g., Harman, 1976; Mulaik, 2010). It is known that this approach and the one in the last paragraph provide almost equivalent solutions (Adachi, 2012, 2015; Stegeman, 2016). However, it is difficult to compare (4) with (1), as the former concerns moments (i.e., covariances), while the latter directly fits the data. Nevertheless, several studies exist comparing the classic FA (4) with PCA (e.g., Bentler & Kano, 1990; Ogasawara, 2000; Sato, 1990). In contrast, comparing (2) with (1) is straightforward, as they both fit the data. The only difference is that (2) also involves an unique part. In this paper, we compare the properties of the PCA solutions with those obtained by the FA procedure based on (2), referred to simply as FA and the classic one based on (4) - as random FA (RFA).

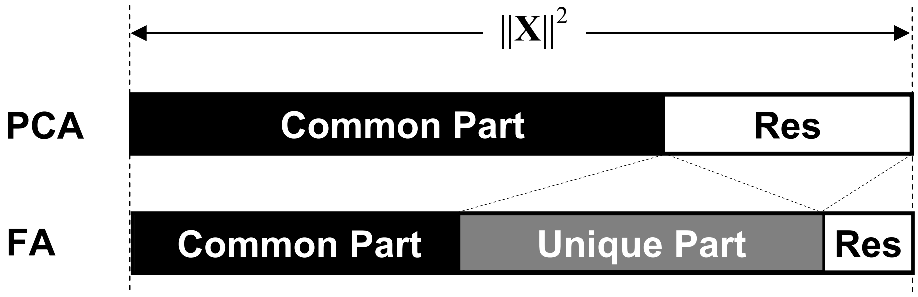

Figure 2. Relative largeness of common parts, unique part, and residuals in PCA and FA solutions,

The main goal of this paper is to quantify the illustration in Figure 2. It depicts how $||\mathbf{X}||^2$, i.e., the total sum of squares (SS) of a data set is decomposed into some SS's. There, the areas of the common part and residuals for PCA stand for $||\hat{\mathbf{P}}\hat{\mathbf{C}}'||^2$ and $||\hat{\mathbf{E}}_{\text{PC}}||^2$, respectively. On the other hand, the areas of the common part, unique part, and residuals for FA correspond to $||\hat{\mathbf{F}}\hat{\mathbf{A}}'||^2$, $||\hat{\mathbf{U}}\hat{\mathbf{\Psi}}||^2$, and $||\hat{\mathbf{E}}_{\text{FA}}||^2$, respectively. Here, $\hat{\mathbf{P}}\hat{\mathbf{C}}'$ denotes the PCA solution for $\mathbf{PC}'$ with $\hat{\mathbf{E}}_{\text{PC}}$ containing the resulting residuals, while $\hat{\mathbf{F}}\hat{\mathbf{A}}'$ and $\hat{\mathbf{U}}\hat{\mathbf{\Psi}}$ are the FA solutions for $\mathbf{FA}'$ and $\mathbf{U\Psi}$, respectively, with $\hat{\mathbf{E}}_{\text{FA}}$ the resulting residual matrix. The relative largeness of the areas in Figure 2 show that:

[1] The common part for PCA is larger than that for FA;
[2] The residual part for FA is smaller than that for PCA;
[3] The unique part for FA is larger than the residual one for PCA.

Here, [1] and [2] always hold, while it is suggested that [3] is often observed. Those assertions are proved in Section 3 in the form of several inequalities, after preliminary results are presented in the Section 2. The theoretical results obtained in Section 3 are illustrated in Sections 4 and 5.

2. Preliminary Notes

In this section, the solutions of PCA and FA are described, which are followed by their rotational indeterminacy. It serves as the preparations for the next section.

2.1. PCA Solution

As described in the last section, PCA can be formulated as minimizing

$$f_{\text{PC}}(\mathbf{P}, \mathbf{C}) = ||\mathbf{E}_{\text{PC}}||^2 = ||\mathbf{X} - \mathbf{PC}'||^2 \quad (5)$$

over $\mathbf{P}$ and $\mathbf{C}$. It is attained through the SVD of $\mathbf{X}$ defined as $\mathbf{X} = \mathbf{V\Theta W}'$ with $\mathbf{V}'\mathbf{V} = \mathbf{W}'\mathbf{W} = \mathbf{I}_q$ and $\mathbf{\Theta}$ the $q \times q$ diagonal matrix whose diagonal elements are arranged in decreasing order.

The solution satisfies $\hat{\mathbf{P}}\hat{\mathbf{C}}' = \mathbf{V}_m\Theta_m\mathbf{W}_m'$. Here, $\mathbf{V}_m$ ($n \times m$) and $\mathbf{W}_m$ ($q \times m$) contain the first m columns of $\mathbf{V}$ and $\mathbf{W}$, respectively, and Θ_m is the first $m \times m$ diagonal block of Θ. For an identification purpose, the condition

$$\frac{1}{n}\mathbf{P}'\mathbf{P} = \mathbf{I}_m \tag{6}$$

is introduced. Then, we can choose $\hat{\mathbf{P}}$ and $\hat{\mathbf{C}}$ as

$$\hat{\mathbf{P}} = n^{1/2}\mathbf{V}_m = n^{1/2}\mathbf{X}\mathbf{W}_m\Theta_m^{-1} \text{ and } \hat{\mathbf{C}} = n^{-1/2}\mathbf{W}_m\Theta_m \tag{7}$$

(e.g., Adachi, 2016), though rotational indeterminacy remains as explained in Section 2.3.

2.2. FA Solution

The FA model (2) leads to the minimization of the following least squares function

$$f_{\mathrm{FA}}(\mathbf{F}, \mathbf{A}, \mathbf{U}, \Psi) = \|\mathbf{E}_{\mathrm{FA}}\|^2 = \|\mathbf{X} - (\mathbf{F}\mathbf{A}' + \mathbf{U}\Psi)\|^2 \tag{8}$$

over $\mathbf{F}$, $\mathbf{A}$, $\mathbf{U}$, and Ψ subject to (3). Though the solution cannot be given explicitly and must be obtained through iterative algorithms, the optimal $\mathbf{A}$ and Ψ are known to satisfy

$$\hat{\mathbf{A}} = \mathbf{S}_{\mathrm{X}\hat{\mathrm{F}}} \text{ and } \hat{\Psi} = \mathrm{diag}(\mathbf{S}_{\mathrm{X}\hat{\mathrm{U}}}), \tag{9}$$

with $\mathbf{S}_{\mathrm{X}\hat{\mathrm{F}}} = n^{-1}\mathbf{X}'\hat{\mathbf{F}}$, $\mathbf{S}_{\mathrm{X}\hat{\mathrm{U}}} = n^{-1}\mathbf{X}'\hat{\mathbf{U}}$, and $\mathrm{diag}(\mathbf{S}_{\mathrm{X}\hat{\mathrm{U}}})$ denotes the diagonal matrix containing the main diagonal of $\mathbf{S}_{\mathrm{X}\hat{\mathrm{U}}}$ (e.g., Adachi & Trendafilov, 2017; Stegeman, 2016). The optimal factor score matrices $\hat{\mathbf{F}}$ and $\hat{\mathbf{U}}$ are undetermined, but $\mathbf{S}_{\mathrm{X}\hat{\mathrm{F}}}$ and $\mathbf{S}_{\mathrm{X}\hat{\mathrm{U}}}$ can be uniquely determined (Adachi & Trendafilov, 2017). Using (4), the function (8) is rewritten as $f_{\mathrm{FA}}(\mathbf{F}, \mathbf{A}, \mathbf{U}, \Psi) = \mathrm{tr}(\mathbf{X}'\mathbf{X} + n\mathbf{A}\mathbf{A}' + n\mathbf{U}\mathbf{U}' + \mathbf{X}'\mathbf{F}\mathbf{A}' + \mathbf{X}'\mathbf{U})$. Here, we can substitute (9) to have

$$f_{\mathrm{FA}}(\hat{\mathbf{F}}, \hat{\mathbf{A}}, \hat{\mathbf{U}}, \hat{\Psi}) = n\,\mathrm{tr}(\mathbf{S}_{\mathrm{XX}} - \hat{\mathbf{A}}\hat{\mathbf{A}}' - \hat{\Psi}^2). \tag{10}$$

It is irrelevant to this paper how $\hat{\mathbf{F}}$, $\hat{\mathbf{U}}$, $\mathbf{S}_{\mathrm{X}\hat{\mathrm{F}}}$, and $\mathbf{S}_{\mathrm{X}\hat{\mathrm{U}}}$ are expressed. Also, in spite of (8) being a data-fitting problem, the solution $\{\hat{\mathbf{A}}, \hat{\Psi}\}$ can be obtained without explicitly given $\mathbf{X}$, if only $\mathbf{S}_{\mathrm{XX}}$ is available (Adachi, 2012; Adachi & Trendafilov, 2017).

2.3. Rotational Indeterminacy

The rotational indeterminacy of the loading matrices in PCA and FA affect some of the results to be presented in the next section.

Let $\mathbf{T}_{\mathrm{P}}$ and $\mathbf{T}_{\mathrm{F}}$ be $m \times m$ arbitrary orthogonal matrices, with

$$\mathbf{T}_P'\mathbf{T}_P = \mathbf{T}_P\mathbf{T}_P' = \mathbf{T}_F'\mathbf{T}_F = \mathbf{T}_F\mathbf{T}_F' = \mathbf{I}_m . \quad (11)$$

Since $\mathbf{PT}_P$ and $\mathbf{FT}_F$ can be substituted for $\mathbf{P}$ in (6) and for $\mathbf{F}$ in (3), respectively, with $\mathbf{PC}' = \mathbf{PT}_P\mathbf{T}_P'\mathbf{C}'$ and $\mathbf{FA}' = \mathbf{FT}_F\mathbf{T}_F'\mathbf{A}'$, PCA and FA solutions have the rotational indeterminacy: score and loading matrices can be rotated without affecting the PCA and FA fit. This property is exploited to obtain the unique orthogonal $\mathbf{T}_P$ or $\mathbf{T}_F$ which make the rotated loading matrices $\hat{\mathbf{P}}\mathbf{T}_P$ or $\hat{\mathbf{A}}\mathbf{T}_F$ most interpretable in some sense. This procedure is called orthogonal rotation (e.g., Adachi, 2016; Mulaik, 2010).

The constraint (6) and $n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$ in (3) are sometimes relaxed to $n^{-1}\mathrm{diag}(\mathbf{F}'\mathbf{F})$ and $n^{-1}\mathrm{diag}(\mathbf{P}'\mathbf{P})$, respectively. Then, PCA and FA have the following rotational indeterminacy: if the $m \times m$ nonsingular matrices $\mathbf{N}_P$ and $\mathbf{N}_F$ satisfy

$$\frac{1}{n}\mathrm{diag}(\mathbf{N}_P'\mathbf{P}'\mathbf{P}\mathbf{N}_P) = \mathbf{I}_m \quad \text{and} \quad \frac{1}{n}\mathrm{diag}(\mathbf{N}_F'\mathbf{F}'\mathbf{F}\mathbf{N}_P) = \mathbf{I}_m , \quad (12)$$

then $\mathbf{PN}_P$ and $\mathbf{FN}_F$ can be substituted for $\mathbf{P}$ and $\mathbf{F}$ in the above relaxed constraints, with $\mathbf{PC}' = \mathbf{PN}_P\mathbf{N}_P^{-1}\mathbf{C}'$ and $\mathbf{FA}' = \mathbf{FN}_F\mathbf{N}_F^{-1}\mathbf{A}'$. These are called oblique rotations and are also used to obtain $\mathbf{N}_P$ or $\mathbf{N}_F$ satisfying (11), such that $\hat{\mathbf{C}}\mathbf{N}_P'^{-1}$ and $\hat{\mathbf{A}}\mathbf{N}_F$ are most interpretable (e.g., Adachi, 2016; Mulaik, 2010).

3. Results

In this section, we present four theorems, which help to contrast the PCA and FA solutions minimizing (5) subject to (6) and minimizing (8) under (3), respectively.

We start with introducing Trendafilov, Unkel, and Krzanowski's (2013) FA-like PCA which is utilized in the proof for the following theorems. In the FA-like PCA, the PCA solution $\hat{\mathbf{P}}\hat{\mathbf{C}}'$ minimizing (5) is substituted for $\mathbf{FA}'$ in the FA loss function (8) as

$$g(\mathbf{U}^*,\mathbf{\Psi}^*) = \|\mathbf{X} - (\hat{\mathbf{P}}\hat{\mathbf{C}}' + \mathbf{U}^*\mathbf{\Psi}^*)\|^2 . \quad (13)$$

Then, it is minimized over $\mathbf{U}^*$ and $\mathbf{\Psi}^*$ subject to $\hat{\mathbf{P}}'\mathbf{U}^* = {}_p\mathbf{O}_m$ and $n^{-1}\mathbf{U}^{*\prime}\mathbf{U}^* = \mathbf{I}_p$, with $\mathbf{\Psi}^*$ being diagonal.

Theorem 1. FA model (2) always fits better for a certain data matrix $\mathbf{X}$ than PCA (1):

$$\|\hat{\mathbf{E}}_{PC}\|^2 \geq \|\hat{\mathbf{E}}_{FA}\|^2 \quad (14)$$

Proof. Obviously, $\|\mathbf{X} - \hat{\mathbf{P}}\hat{\mathbf{C}}'\|^2 \geq \|\mathbf{X} - \hat{\mathbf{P}}\hat{\mathbf{C}}'\|^2 - \|\mathbf{U}^*\mathbf{\Psi}^*\|^2$ holds true. Further, the FA-like PCA optimality condition $\mathbf{\Psi} = \mathrm{diag}(\mathbf{U}^{*\prime}\mathbf{X})$ (Trendafilov, et al's, 2013, 5.2.1) shows $\|\mathbf{X} - \hat{\mathbf{P}}\hat{\mathbf{C}}'\|^2 - \|\mathbf{U}^*\mathbf{\Psi}^*\|^2 = \|\mathbf{X} - (\hat{\mathbf{P}}\hat{\mathbf{C}}' + \mathbf{U}^*\mathbf{\Psi}^*)\|^2$. Thus, the value of PCA loss function (5) cannot be less than the (13) value:

$$||\mathbf{X} - \hat{\mathbf{P}}\hat{\mathbf{C}}'||^2 \geq ||\mathbf{X} - (\hat{\mathbf{P}}\hat{\mathbf{C}}' + \mathbf{U}^*\mathbf{\Psi}^*)||^2. \tag{15}$$

Now, $\hat{\mathbf{P}}\hat{\mathbf{C}}'$, $\mathbf{U}^*$, and $\mathbf{\Psi}^*$ in the right-hand side of (15) can be replaced by the corresponding FA solutions $\hat{\mathbf{F}}\hat{\mathbf{A}}', \hat{\mathbf{U}}$, and $\hat{\mathbf{\Psi}}$. The function value after this substitution cannot exceed the right-hand side of (15):

$$||\mathbf{X} - (\hat{\mathbf{P}}\hat{\mathbf{C}}' + \mathbf{U}^*\mathbf{\Psi}^*)||^2 \geq ||\mathbf{X} - (\hat{\mathbf{F}}\hat{\mathbf{A}}' + \hat{\mathbf{U}}\hat{\mathbf{\Psi}})||^2, \tag{16}$$

because the FA loss function $||\mathbf{X} - (\mathbf{FA}' + \mathbf{U\Psi})||^2$ is minimized for $\mathbf{FA}' + \mathbf{U\Psi} = \hat{\mathbf{F}}\hat{\mathbf{A}}' + \hat{\mathbf{U}}\hat{\mathbf{\Psi}}$ but not for $\mathbf{FA}' + \mathbf{U\Psi} = \hat{\mathbf{P}}\hat{\mathbf{C}}' + \mathbf{U}^*\mathbf{\Psi}^*$. Thus, the inequalities (15) and (16) lead to (14). □

The theorem suggests that for better fit to the data, FA should be preferred over PCA.

The next theorem concerns the largeness of squared loadings and common parts:

Theorem 2. For certain **X**, the sum of squared PCA loadings is always equal to or larger than the sum of squared FA ones (under constraints (3) and (6)),

$$||\hat{\mathbf{C}}||^2 \geq ||\hat{\mathbf{A}}||^2, \tag{17}$$

$$||\hat{\mathbf{C}}\mathbf{T}_{\mathrm{P}}||^2 \geq ||\hat{\mathbf{A}}\mathbf{T}_{\mathrm{F}}||^2, \tag{18}$$

with $\mathbf{T}_{\mathrm{P}}$ and $\mathbf{T}_{\mathrm{F}}$ satisfying (11). This implies that the common part in PCA is always equal to or larger than that one for FA:

$$||\hat{\mathbf{P}}\hat{\mathbf{C}}'||^2 \geq ||\hat{\mathbf{F}}\hat{\mathbf{A}}'||^2, \tag{19}$$

$$||\hat{\mathbf{P}}\mathbf{N}_{\mathrm{P}}\mathbf{N}_{\mathrm{P}}^{-1}\hat{\mathbf{C}}'||^2 \geq ||\hat{\mathbf{F}}\mathbf{N}_{\mathrm{F}}\mathbf{N}_{\mathrm{F}}^{-1}\hat{\mathbf{A}}'||^2. \tag{20}$$

with $\mathbf{N}_{\mathrm{P}}$ and $\mathbf{N}_{\mathrm{F}}$ arbitrary nonsingular $m \times m$ matrices.

Proof. The PCA loss function (5) is expanded as $||\mathbf{X}||^2 - 2\mathrm{tr}\mathbf{X}'\hat{\mathbf{P}}\hat{\mathbf{C}}' + ||\hat{\mathbf{P}}\hat{\mathbf{C}}'||^2$, which is equal to $n(\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\hat{\mathbf{C}}\hat{\mathbf{C}}')$ since of (6) and (7). That is, we have

$$f_{\mathrm{PC}}(\hat{\mathbf{P}}, \hat{\mathbf{C}}) = n(\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\hat{\mathbf{C}}\hat{\mathbf{C}}'). \tag{21}$$

Now, let us consider $f_{\mathrm{PC}}(\hat{\mathbf{F}}, \hat{\mathbf{A}}) = ||\mathbf{X} - \hat{\mathbf{F}}\hat{\mathbf{A}}'||^2$, i.e., the PCA function (5) with the FA solution $\hat{\mathbf{F}}\hat{\mathbf{A}}'$ substituted for $\mathbf{PC}'$. Using (3) and (9), $f_{\mathrm{PC}}(\hat{\mathbf{F}}, \hat{\mathbf{A}})$ can be rewritten as $n(\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - 2\mathrm{tr}\mathbf{S}_{\mathrm{X}\hat{\mathrm{F}}}\hat{\mathbf{A}}' - \mathrm{tr}\hat{\mathbf{A}}\hat{\mathbf{A}}') = n(\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\hat{\mathbf{A}}\hat{\mathbf{A}}')$. Clearly, $f_{\mathrm{PC}}(\hat{\mathbf{F}}, \hat{\mathbf{A}}) = n(\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\hat{\mathbf{A}}\hat{\mathbf{A}}')$ cannot be lower than (21), since the PCA solution is known as the best low rank approximation (Eckart & Young, 1936). Thus, we finally have

$$\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\hat{\mathbf{C}}\hat{\mathbf{C}}' \leq \mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\hat{\mathbf{A}}\hat{\mathbf{A}}', \tag{22}$$

which gives (17). It also implies (18), because of the orthogonality propety (11).

Inequality (17) leads to (19), since $\|\hat{\mathbf{F}}\hat{\mathbf{A}}'\|^2 = n\text{tr}\,\hat{\mathbf{A}}\hat{\mathbf{A}}' = n\|\hat{\mathbf{A}}\|^2$ and $\|\hat{\mathbf{P}}\hat{\mathbf{C}}'\|^2 = n\text{tr}\,\hat{\mathbf{C}}\hat{\mathbf{C}}' = n\|\hat{\mathbf{C}}\|^2$ follow from (3) and (6), respectively. Obviously, (19) leads to (20). □

This theorem shows that the common part in PCA is larger than in FA. Inequality (20) shows that the common part is larger in PCA solutions even after oblique rotation. On the other hand, $\mathbf{T}_\text{P}$ and $\mathbf{T}_\text{F}$ in (18) cannot be replaced by $\mathbf{N}_\text{P}$ and $\mathbf{N}_\text{F}$. That is, after the oblique rotation, $\|\hat{\mathbf{C}}\mathbf{N}_\text{P}'^{-1}\|^2 \geq \|\hat{\mathbf{A}}\mathbf{N}_\text{F}'^{-1}\|^2$ does not necessarily hold.

Though Theorem 2 discuss the magnitudes of the squared loadings and common part in PCA solutions, the next one shows their upper limits.

Theorem 3. For certain $\mathbf{X}$, the sum of the squared PCA loadings cannot exceed the sum of the squared loadings and unique variances in the FA solution:

$$\|\hat{\mathbf{C}}\|^2 \leq \|\hat{\mathbf{A}}\|^2 + \|\hat{\mathbf{\Psi}}\|^2, \tag{23}$$

$$\|\hat{\mathbf{C}}\mathbf{T}_\text{P}\|^2 \leq \|\hat{\mathbf{A}}\mathbf{T}_\text{F}\|^2 + \|\hat{\mathbf{\Psi}}\|^2, \tag{24}$$

with $\mathbf{T}_\text{P}$ and $\mathbf{T}_\text{F}$ satisfying (11). It implies that the squared norm of the PCA common part cannot exceed the FA model part one:

$$\|\hat{\mathbf{P}}\hat{\mathbf{C}}'\|^2 \leq \|\hat{\mathbf{F}}\hat{\mathbf{A}}' + \hat{\mathbf{U}}\hat{\mathbf{\Psi}}\|^2 \quad , \tag{25}$$

$$\|\hat{\mathbf{P}}\mathbf{N}_\text{P}\mathbf{N}_\text{P}^{-1}\hat{\mathbf{C}}'\|^2 \leq \|\hat{\mathbf{F}}\mathbf{N}_\text{F}\mathbf{N}_\text{F}^{-1}\hat{\mathbf{A}}' + \hat{\mathbf{U}}\hat{\mathbf{\Psi}}\|^2 \quad . \tag{26}$$

with $\mathbf{N}_\text{P}$ and $\mathbf{N}_\text{F}$ arbitrary nonsingular $m \times m$ matrices.
Proof. (10) and (21) are rewritten as $\|\hat{\mathbf{E}}_\text{FA}\|^2 = n(\text{tr}\mathbf{S}_{XX} - \text{tr}\,\hat{\mathbf{A}}\hat{\mathbf{A}}' - \text{tr}\,\hat{\mathbf{\Psi}}^2)$ and $\|\hat{\mathbf{E}}_\text{PC}\|^2 = n(\text{tr}\mathbf{S}_{XX} - \text{tr}\,\hat{\mathbf{C}}\hat{\mathbf{C}}')$, respectively. Using them in (14), we have (23) and it leads to (24), since of (11). Inequality (23) leads to (25) and thus (26), since $\|\hat{\mathbf{F}}\hat{\mathbf{A}}' + \hat{\mathbf{U}}\hat{\mathbf{\Psi}}\|^2 = n\|\hat{\mathbf{A}}\|^2 + n\|\hat{\mathbf{\Psi}}\|^2$ and $\|\hat{\mathbf{P}}\hat{\mathbf{C}}'\|^2 = n\|\hat{\mathbf{C}}\|^2$ follow from (3) and (6), respectively. □

Inequality (26) shows that the model part is larger in FA even after oblique rotation. However, after the rotation, the sum of squared loadings in PCA is not necessarily less than the sum of squared loadings and unique variances in FA, since $\mathbf{T}_\text{P}$ and $\mathbf{T}_\text{F}$ in (24) cannot be replaced by $\mathbf{N}_\text{P}$ and $\mathbf{N}_\text{F}$.

The following theorem concerns the magnitudes of the unique variances in FA:

Theorem 4. For a certain $\mathbf{X}$, the sum of the unique variances in FA is larger than the sum of squared residuals for PCA minus the sum for FA:

$$\| \hat{\Psi} \|^2 \geq \frac{1}{n} \| \hat{\mathbf{E}}_{\mathrm{PC}} \|^2 - \frac{1}{n} \| \hat{\mathbf{E}}_{\mathrm{FA}} \|^2 \ . \tag{27}$$

Proof. We can rewrite (10) as $n^{-1}\| \hat{\mathbf{E}}_{\mathrm{FA}} \|^2 = \mathrm{tr}(\mathbf{S}_{\mathrm{XX}} - \hat{\mathbf{A}}\hat{\mathbf{A}}' - \hat{\Psi}^2)$, which leads to $\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\,\hat{\mathbf{A}}\hat{\mathbf{A}}' = \hat{\Psi}^2 + n^{-1}\| \hat{\mathbf{E}}_{\mathrm{FA}} \|^2$. We can also rewrite (21) as $\mathrm{tr}\mathbf{S}_{\mathrm{XX}} - \mathrm{tr}\,\hat{\mathbf{C}}\hat{\mathbf{C}}' = n^{-1}\| \hat{\mathbf{E}}_{\mathrm{PC}} \|^2$. Their use in (22) we have to $n^{-1}\| \hat{\mathbf{E}}_{\mathrm{PC}} \|^2 \leq \hat{\Psi}^2 + n^{-1}\| \hat{\mathbf{E}}_{\mathrm{FA}} \|^2$, which can be rewritten as (27). □

The mathematical results presented so far always hold. However, Theorems 2 and 4 also make the following suggests, which are likely but not necessarily holding in every occasion:

[S1] The absolute value of each PCA loading before/after orthogonal rotation tends to be greater than the absolute one of the corresponding FA loading (though exceptions can also exist), which is suggested by (17) and (18).

[S2] If $\| \hat{\mathbf{E}}_{\mathrm{FA}} \|^2$ is small enough, $\| \hat{\Psi} \|^2$ tends to be larger than $n^{-1}\| \hat{\mathbf{E}}_{\mathrm{PC}} \|^2$. The unique variance $\hat{\psi}_j^2$ for variable j tends to be greater than the corresponding PCA residual variance $n^{-1}\| \hat{\mathbf{e}}_j^{\mathrm{PC}} \|^2$, where $\hat{\mathbf{e}}_j^{\mathrm{PC}}$ and $\hat{\psi}_j^2$ are the j-th column and diagonal element of $\hat{\mathbf{E}}_{\mathrm{PC}}$ and $\hat{\Psi}$, respectively. Note, $n^{-1}\| \hat{\mathbf{e}}_j^{\mathrm{PC}} \|^2$ is a variance, because $\hat{\mathbf{E}}_{\mathrm{PC}} = \mathbf{X} - \hat{\mathbf{P}}\hat{\mathbf{C}}'$ and $\mathbf{X}$ is column-centered and implies (7) for $\mathbf{P}$.

These features are numerically assessed in the following sections.

4. Illustration

In this section, two real data examples are used in order to illustrate the theorems in the last section as well as [S1] and [S2]. For the every data set, we carry out PCA and FA, together with two classic RFA procedures. One of the two RFA procedures is the least squares RFA (LS-RFA) with loss function $\|\mathbf{S}_{\mathrm{XX}} - (\mathbf{AA}' + \Psi^2)\|^2$. The other one is the maximum likelihood RFA (ML-RFA), whose loss function is $\mathrm{tr}\mathbf{S}_{\mathrm{XX}}(\mathbf{AA}' + \Psi^2)^{-1} - \log|\mathbf{AA}' + \Psi^2|$ following from certain normality assumptions, with |•| denoting the determinant of its argument. As the theorems in the last section are derived from the formulation of FA with (2), they are not guaranteed to hold in RFA with (4). Thus, it is of interest to see to what extent the RFA solutions follow the inequalities in the theorems. Of course, Theorem 1 is not considered because the error matrix $\hat{\mathbf{E}}_{\mathrm{FA}}$ is not relevant to RFA with (4). The resulting loadings in LS- and ML-RFA are expressed as $\hat{\mathbf{A}}_{\mathrm{L}}$ and $\hat{\mathbf{A}}_{\mathrm{M}}$, respectively, with the corresponding unique variances matrices denoted as $\hat{\Psi}_{\mathrm{L}}^2$ and $\hat{\Psi}_{\mathrm{M}}^2$. The loading matrices in all procedures are

Table 1. The solution of PCA, FA, LS-RFA, and ML-RFA for a part of Tanaka and Tarumi's (1995) test score data (Adachi & Trendafiov, 2017), with Res standing for residual variances.

	PCA			FA				LS-RFA			ML-RFA		
	$\hat{\mathbf{C}}\mathbf{T}_{\mathrm{P}}$		Res	$\hat{\mathbf{A}}\mathbf{T}_{\mathrm{F}}$		$\hat{\Psi}^2$	Res	$\hat{\mathbf{A}}_{\mathrm{L}}\mathbf{T}_{\mathrm{L}}$		$\hat{\Psi}_{\mathrm{L}}^2$	$\hat{\mathbf{A}}\mathbf{T}_{\mathrm{M}}$		$\hat{\Psi}_{\mathrm{M}}^2$
Japanese	**0.51**	**0.62**	0.13	0.38	0.60	0.50	0.001	0.38	0.60	0.50	0.37	0.61	0.50
English	**0.25**	**0.81**	0.08	0.21	0.76	0.37	0.002	0.19	0.77	0.37	0.21	0.76	0.38
Social*	-0.02	**0.86**	0.07	0.03	0.65	0.58	0.002	0.03	0.64	0.59	0.02	0.65	0.58
Mathematics	**0.80**	0.26	0.08	0.59	0.34	0.53	0.003	0.58	0.35	0.54	0.58	0.34	0.55
Sciences	**0.90**	0.02	0.03	0.89	0.10	0.19	0.001	0.91	0.10	0.17	0.90	0.11	0.17
Sum of Squares	$\Vert\hat{\mathbf{C}}\Vert^2$ 3.62		$n^{-1}\Vert\hat{\mathbf{E}}_{\mathrm{PC}}\Vert^2$ 1.38	$\Vert\hat{\mathbf{A}}\Vert^2$ 2.81		$\Vert\hat{\Psi}\Vert^2$ 2.18	$n^{-1}\Vert\hat{\mathbf{E}}_{\mathrm{FA}}\Vert^2$ 0.008	$\Vert\hat{\mathbf{A}}_{\mathrm{L}}\Vert^2$ 2.83		$\Vert\hat{\Psi}_{\mathrm{L}}\Vert^2$ 2.17	$\Vert\hat{\mathbf{A}}_{\mathrm{M}}\Vert^2$ 2.82		$\Vert\hat{\Psi}_{\mathrm{M}}\Vert^2$ 2.18

*Social Studies

rotated by the orthogonal varimax rotation (Kaiser, 1958). We denote the rotated PCA, FA, LS-RFA, and ML-RFA loading matrices as $\hat{\mathbf{C}}\mathbf{T}_{\mathrm{P}}$, $\hat{\mathbf{A}}\mathbf{T}_{\mathrm{F}}$, $\hat{\mathbf{A}}_{\mathrm{L}}\mathbf{T}_{\mathrm{L}}$, and $\hat{\mathbf{A}}_{\mathrm{M}}\mathbf{T}_{\mathrm{M}}$, respectively.

The first example is the standardized version of the test score data with $q = 5$ courses for $n = 20$ examinees (Adachi & Trendafilov, 2017, Table 1), which is a part of Tanaka and Tarumi's (1995) data. Table 1 shows the solutions with $m = 2$. First, let us consider the bottom parts in the left two panels of PCA and FA. Those parts illustrate Theorems 1 to 4, as listed below.

[Theorem 1] $n^{-1}\Vert \hat{\mathbf{E}}_{\mathrm{PC}} \Vert^2 = 1.38 > n^{-1}\Vert \hat{\mathbf{E}}_{\mathrm{FA}} \Vert^2 = 0.008$;
[Theorem 2] $n^{-1}\Vert \hat{\mathbf{P}}\hat{\mathbf{C}}' \Vert^2 = \Vert \hat{\mathbf{C}} \Vert^2 = 3.62 \geq n^{-1}\Vert \hat{\mathbf{F}}\hat{\mathbf{A}}' \Vert^2 = \Vert \hat{\mathbf{A}} \Vert^2 = 2.81$;
[Theorem 3] $\Vert \hat{\mathbf{C}} \Vert^2 = 3.62 \leq \Vert \hat{\mathbf{A}} \Vert^2 + \Vert \hat{\Psi} \Vert^2 = 2.81 + 2.18 = 4.99$;
[Theorem 4] $\Vert \hat{\Psi} \Vert^2 = 2.18 \geq n^{-1}\Vert \hat{\mathbf{E}}_{\mathrm{PC}} \Vert^2 - n^{-1}\Vert \hat{\mathbf{E}}_{\mathrm{FA}} \Vert^2 = 1.38 - 0.008 = 1.37$.

Next, we consider the loadings, residuals, and unique variance in the left two panels in Table 1. Seven PCA loadings among all ten are bold-faced. Their absolute values are larger than their FA counterparts, which supports the suggestion by [S1]. We also find that $\Vert \hat{\mathbf{E}}_{\mathrm{FA}} \Vert^2$ is close to zero and $\Vert \hat{\Psi} \Vert^2 = 2.18 > n^{-1}\Vert \hat{\mathbf{E}}_{\mathrm{PC}} \Vert^2 = 1.38$ with all unique variances in FA larger than the corresponding "Res" (residual variances) in PCA, i.e. as suggested by [S2].

Now, we consider the right three panels. The panels for RFA do not have column "Res", since $\hat{\Psi}_{\mathrm{L}}^2 = \mathrm{diag}(\mathbf{S}_{\mathrm{XX}} - \hat{\mathbf{A}}_{\mathrm{L}}\hat{\mathbf{A}}_{\mathrm{L}}')$ and $\hat{\Psi}_{\mathrm{M}}^2 = \mathrm{diag}(\mathbf{S}_{\mathrm{XX}} - \hat{\mathbf{A}}_{\mathrm{M}}\hat{\mathbf{A}}_{\mathrm{M}}')$: the residual variances for variables are always estimated as zero. Besides "Res", all three FA solutions (loadings and unique variances) are almost identical. Thus, the RFA solutions show the same relationships to PCA ones as the FA solutions.

The second example is the Mullen's (1939) data set with $q = 8$ physical variables for $n = 305$ girls. The inter-variable correlation matrix is also available from Harman (1976, p. 22). Table 2 presents the $m = 2$ solutions. Again, we can empirically confirmed the theoretical results established in the last section. The findings are pretty similar to those observed in the first example, and are summarized as follows:

Table 2. The solution of PCA, FA, LS-RFA, and ML-RFA for a part of Mullen's (1939) physical variables data, with Res standing for residual variances.

	PCA			FA				LS-RFA			ML-RFA		
	$\hat{\mathbf{C}}\mathbf{T}_P$		Res	$\hat{\mathbf{A}}\mathbf{T}_F$		$\hat{\Psi}^2$	Res	$\hat{\mathbf{A}}_L\mathbf{T}_L$		$\hat{\Psi}_L^2$	$\hat{\mathbf{A}}\mathbf{T}_M$		$\hat{\Psi}_M^2$
Height	0.24	**0.91**	0.12	0.26	0.88	0.16	0.005	0.25	0.88	0.16	0.27	0.87	0.17
Arm span	**0.18**	**0.93**	0.10	0.17	0.93	0.10	0.006	0.18	0.93	0.11	0.16	0.93	0.11
Forearm[1]	0.14	**0.92**	0.13	0.16	0.89	0.17	0.002	0.16	0.89	0.18	0.16	0.90	0.17
Lower leg[1]	0.21	**0.90**	0.14	0.23	0.87	0.19	0.005	0.22	0.87	0.19	0.23	0.86	0.20
Weight	0.88	**0.27**	0.15	0.91	0.26	0.10	0.002	0.91	0.26	0.11	0.92	0.25	0.09
Bitrochanteric[2]	**0.84**	0.20	0.26	0.77	0.21	0.36	0.002	0.77	0.21	0.36	0.77	0.21	0.36
Chest girth	**0.84**	0.12	0.28	0.75	0.15	0.41	0.002	0.75	0.15	0.42	0.75	0.15	0.42
Chest width	**0.74**	0.27	0.38	0.64	0.28	0.52	0.002	0.64	0.28	0.51	0.62	0.29	0.54
Sum of Squares	$\|\hat{\mathbf{C}}\|^2$ 6.44		$n^{-1}\|\hat{\mathbf{E}}_{PC}\|^2$ 1.56	$\|\hat{\mathbf{A}}\|^2$ 5.97		$\|\hat{\Psi}\|^2$ 2.01	$n^{-1}\|\hat{\mathbf{E}}_{FA}\|^2$ 0.025	$\|\hat{\mathbf{A}}_L\|^2$ 5.96		$\|\hat{\Psi}_L\|^2$ 2.04	$\|\hat{\mathbf{A}}_M\|^2$ 5.95		$\|\hat{\Psi}_M\|^2$ 5.95

[1] length [2] diameters

[O1] The absolute values of the PCA loadings tend to be greater than those of FA.

[O2] $\|\hat{\Psi}\|^2$ tends to be larger than $n^{-1}\|\hat{\mathbf{E}}_{PC}\|^2$.

[O3] The unique variance $\hat{\psi}_j^2$ for variable j tends to be greater than the variance of PCA residuals $n^{-1}\|\hat{\mathbf{e}}_j^{PC}\|^2$ for j

[O4] FA and RFA solutions are broadly equivalent.

[O5] The inequalities in Theorems 2 to 4 also hold in RFA solutions.

Here [O1] corresponds to [S1] from Section 3, while [S2] is divided into [O2] and [O3].

5. Supplementary Simulation Studies

In this section, we explore whether [O1] to [O5] from the last section are fulfilled for most of the data sets in practice. It is not efficient to make such assessments with real data sets. We thus resort to use simulated data. Indeed, the correctness of [O4] was demonstrated in the past simulation studies in Adachi (2012, 2015) (unfortunately Adachi (2015) is in Japanese) and Stegeman (2016). Adachi (2012) and Stegeman (2016) have indirectly shown [O4]. This has been assessed without direct comparison of FA and RFA solutions. Instead, it has been shown that the true parameters are recovered well both by FA and RFA. Here, we assess [O4] with direct comparisons.

We simulate two types of data sets: one is the PCA-modeled data set synthesized with (1) and (6), while the other type is the FA-modeled data set following (2) and (3). For each of m = 1, … , 5, we synthesize a data set with the following procedure:

[1] Choose q from $DU(4m, 8m)$ and n from $DU(8q, 12q)$, with $DU(4m, 8m)$ the discrete uniform distribution defined for the integers within the range $[4m, 8m]$.

[2] Draw each element of **P**, **F**, **U**, and **E** ($n \times p$) from the standard normal distribution.

[3] Draw each element of $q \times m$ matrix $\mathbf{A}_0$ from $U(-1, 1)$ and each diagonal element of $q \times q$ diagonal matrix $\mathbf{\Psi}_0$ from $U(0.1, 1)$, with $U(-1, 1)$ the uniform distribution defined for $[-1, 1]$.

[4] Set $\mathbf{C} = \alpha\mathbf{A}_0$ and $\mathbf{E}_{\mathrm{PC}} = \mathbf{E}$ so that $||\mathbf{PC}'||^2/(||\mathbf{PC}'||^2 + ||\mathbf{E}_{\mathrm{PC}}||^2) = 0.75$ with $\alpha > 0$.

[5] Set $\mathbf{A} = \beta\mathbf{A}_0$, $\mathbf{\Psi} = \gamma\mathbf{\Psi}_0$, and $\mathbf{E}_{\mathrm{FA}} = \mathbf{E}$ so that $||\mathbf{FA}'||^2\ /SST = 0.55$ and $||\mathbf{U\Psi}||^2/SST = 0.42$ with $SST = ||\mathbf{FA}'||^2 + ||\mathbf{U\Psi}||^2 + ||\mathbf{E}_{\mathrm{FA}}||^2$, $\beta > 0$, and $\gamma > 0$.

[6] Generate **X** with (1) and another one with (2), followed by the column standardization.

The procedures were repeated 250 times to provide 250 (replications) × 5 (m) × 2 (PCA-FA) = 2500 data sets. They are analyzed as in the last section.

We first assess how [O4] (the equivalences of solutions between FA procedures) is fulfilled making use of the averaged absolute difference (AAD) of the elements between the resulting matrices. It is defined as AAD($\hat{\mathbf{A}}, \hat{\mathbf{A}}_{\mathrm{L}}$) = $(qm)^{-1}||\hat{\mathbf{A}} - \hat{\mathbf{A}}_{\mathrm{L}}\mathbf{R}||_{l1}$ for $\hat{\mathbf{A}}$ and $\hat{\mathbf{A}}_{\mathrm{L}}$, where $||\bullet||_{l1}$ denotes the l_1 matrix norm, and **R** is the $m \times m$ orthonormal matrix minimizing $||\hat{\mathbf{A}} - \hat{\mathbf{A}}_{\mathrm{L}}\mathbf{R}||^2 = ||\hat{\mathbf{A}}\mathbf{R}' - \hat{\mathbf{A}}_{\mathrm{L}}||^2$, i.e., perfoming Procrustes rotation (e.g., Gower & Dijksterhuis, 2004). This is required, since the loading matrices have rotational indeterminacy. The averages and 95 percentiles of the AAD values among $\hat{\mathbf{A}}, \hat{\mathbf{A}}_{\mathrm{L}}$, $\hat{\mathbf{A}}_{\mathrm{M}}$, and $\hat{\mathbf{P}}$ are presented for each of the two types of data in Table 3, where the cells concerning PCA solutions are colored gray. In the other cells, we can find that the FA loadings are broadly equivalent to RFA ones, with the averages less than 0.010 and even the highest 95 percentile 0.012. These statistics are rather higher within the RFA procedures than between them and FA. In contrast to the equivalence between FA and RFA loadings, it is found that the PCA ones differ from FA and RFA loadings, as the averaged AAD between FA and PCA are about seven to ten

Table 3. Averages and 95 percentiles of AAD values for loading matrices.

		PCA-modeled Data			FA-modeled Data		
		FA	LS-RFA	ML-RFA	FA	LS-RFA	ML-RFA
PCA	Ave	0.021	0.021	0.021	0.033	0.033	0.035
	[95%]	[0.048]	[0.048]	[0.048]	[0.069]	[0.069]	[0.074]
FA	Ave		0.002	0.002		0.004	0.005
	[95%]		[0.006]	[0.006]		[0.009]	[0.012]
LS-RFA	Ave			0.004			0.009
	[95%]			[0.011]			[0.020]

Table 4. Averages and 95 percentiles of AAD values for unique variances.

		PC-modeled Data		FA-modeled Data	
		LS-RFA	ML-RFA	LS-RFA	ML-RFA
FA	Ave	0.006	0.006	0.010	0.010
	[95%]	[0.011]	[0.010]	[0.016]	[0.017]
LS-RFA	Ave		0.007		0.013
	[95%]		[0.017]		[0.026]

times of the values within FA solutions. Table 4 shows the statistics of the AAD values for unique variances, with AAD($\hat{\boldsymbol{\Psi}}$, $\hat{\boldsymbol{\Psi}}_L$) = $(qm)^{-1}\|\hat{\boldsymbol{\Psi}} - \hat{\boldsymbol{\Psi}}_L\|_{l1}$. Clearly, the FA and RFA solutions are pretty close.

The RFA solutions were found to satisfy the inequalities in Theorems 2 to 4 for every simulated data set, i.e. [O5] is also verified:

Next, we consider [O2], i.e. that $\|\hat{\boldsymbol{\Psi}}\|^2$ tends to be larger than $n^{-1}\|\hat{\mathbf{E}}_{PC}\|^2$. It turns out, that this fulfilled for every data set, in the FA solutions, and also in the LS- and ML-RFA solutions.

To assess [O3], i.e. that $\hat{\psi}_j^2$ tends to be greater than $n^{-1}\|\hat{\mathbf{e}}_j^{PC}\|^2$, for each data set we count U/q, where U is the number of variables for which $n^{-1}\|\hat{\mathbf{e}}_j^{PC}\|^2 > \hat{\psi}_j^2$, where a deviation from [O3] occurs. The resulting statistics are presenting in Tables 5. The average and 95 percentiles are found to be substantially smaller than 0.5. It allows us to conclude that $n^{-1}\|\hat{\mathbf{e}}_j^{PC}\|^2$ tends to be smaller than $\hat{\psi}_j^2$.

Now, let us consider [O1]. The proportion $L/(qm)$ measuring the deviation from [O1] is recorded for each data set. Here, L is the number of the PCA loadings whose absolute values are less than their FA counterparts. The resulting statistics are presenting in Tables 6(A). The averages are found to be substantially less than 0.5, which shows that [O1] is observed in around 30% of the data sets or less. But, the 95 percentiles in Table 5 are close to 0.5, suggesting that the solutions without feature [O1] is likely to be observed.

We further assess whether the relationships (18) and (24) often occur, even if the orthonormal matrices $\mathbf{T}_P$ and $\mathbf{T}_F$ are replaced by nonsingular matrices subject to (12), i.e.,

Table 5. Averages and 95 percentiles of the proportions of the squared sum of PCA residuals less than the FA unique variances for variables.

	PCA-modeled Data			FA-modeled Data		
	FA	LS-RFA	ML-RFA	FA	LS-RFA	ML-RFA
Ave	0.10	0.07	0.07	0.21	0.18	0.19
[95%]	[0.29]	[0.25]	[0.25]	[0.35]	[0.30]	[0.33]

Table 6. Averages and 95 percentiles of the proportions of the PCA loadings whose absolute values are less than the FA counterparts.

		(A) After Orthogonal Rotation			(B) After Oblique Rotation		
		FA	LS-RFA	ML-RFA	FA	LS-RFA	ML-RFA
PCA-modeled Data	Ave	0.27	0.27	0.28	0.26	0.25	0.27
	[95%]	[0.44]	[0.44]	[0.44]	[0.45]	[0.45]	[0.46]
FA-modeled Data	Ave	0.30	0.30	0.31	0.30	0.29	0.31
	[95%]	[0.46]	[0.47]	[0.46]	[0.46]	[0.46]	[0.47]

even after oblique rotation. For this assessment, we perform Jennrich's (2006) oblique rotation, in which $\|\hat{\mathbf{C}}\mathbf{N}_P\|_{l1}$ is minimized over $\mathbf{N}_P$ and $\|\hat{\mathbf{A}}\mathbf{N}_F\|_{l1}$ is minimized over $\mathbf{N}_F$ under (12) for PCA and FA solutions, respectively. As a result, it was found for every data set that the sum of the squares of obliquely rotated PCA loadings was greater than the sum for FA/RFA, but less than that sum plus the sum of FA/RFA unique variances.

6. Discussion

In this paper, we derive several theorems contrasting PCA and FA solutions, with both PCA and FA formulated as matrix decomposition problems. Next, the conclusions from the theorems are assessed numerically.

Theorems 1 and 2 show that FA fits better than PCA, but PCA extracts a larger common part than FA, for a certain data set. To the best of our knowledge, no research exists suggesting which technique, PCA or FA, should be used for a particular data set **X**. This might be due to the fact that PCA (1) had been originally considered the transformation of the observed variables (Hotelling, 1933), while the classic FA (4) looks for new latent variables. However, the whole comparative story makes perfect sense, when the FA formulation (2) is introduced. Then, PCA and FA can be considered purely as data matrix decompositions, and thus, comparable. In this respect, these theorems suggest:

[P] Choose PCA when a large common part is wished to be found in **X**.

[F] Choose FA when **X** is wished to be better explained.

The conclusions from the theorems are numerical assessessed in Sections 4 and 5. The experimental findings are summarized as follows:

[1] The absolute values of PCA loadings tend to be greater than the corresponding FA ones, though solutions can also occur in which this is not clearly found.

[2] It is a common result that the sum of unique variances in FA is larger than the sum of error variances of PCA. Further, the unique variance for each variable in FA tends to be greater than the corresponding residual variance in PCA.

Finding [1] can be restated as that the relationships of variables to components tend to be estimated as stronger than those to common factors. This suggests that in a number of cases interpreting component loadings may provide more reliable information than interpreting factor loadings. On the other hand, [2] impresses how important the role of the unique factors is in FA.

As the inequalities in Section 4 are derived from the matrix decomposition formulation of FA with (2), they are not guaranteed to hold in the classic random FA (RFA) formulated as (4). However, as found in Sections 4 and 5, the matrix decomposition FA solutions are broadly equivalent to the RFA ones. Thus, the inequalities in the theorems are likely to hold for RFA, except Theorem 1 which does not make sense in RFA.

The above statement "FA fits better than PCA" is to be carefully reconsidered. As found in (1) and (2), the addition of the unique part $\mathbf{U\Psi}$ to the PCA model leads to the FA model. Thus, PCA has fewer parameters than FA, and can be viewed as more parsimonious. This suggests that a model selection strategy taking into account the model's parsimony remains to be studied for prescribing whether PCA or FA is suitable for a particular data set.

References

Adachi, K. (2012). Some contributions to data-fitting factor analysis with empirical comparisons to covariance-fitting factor analysis. *Journal of the Japanese Society of Computational Statistics*, *25*, 25-38.

Adachi, K. (2015). A matrix-intensive approach to factor analysis. *Journal of the Japan Statistical Society*, *Japanese Issue*, *44*, 363-382 (in Japanese).

Adachi, K. (2016). *Matrix-based introduction to multivariate data analysis*. Singapore: Springer.

Adachi & Trendafilov (2017). Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika*, doi.org/10.1007/s11336-017-9600-y.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, **25**, 67-74.

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211-218.

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

Harman, H. H. (1976). *Modern factor analysis* (*3rd Edition*). Chicago: The University of Chicago Press.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Statistics*, *24*, 417-441.

Jennrich, R. I. (2006). Rotation to simple loadings using component loss function: The oblique case. *Psychometrika*, *71*, 173-191.

Jolliffe, I. T. (2002), *Principal component analysis*, *Second Edition*, New York: Springer.

Kaiser H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187-200.

Mullen, F. (1939). *Factors in the growth of girls seven to seventeen years of age*. Ph.D Dissertation. University of Chicago, Department of Education.

Mulaik, S. A. (2010). *Foundations of factor analysis*, 2nd Edition. Boca Raton: CRC Press. Mullen

Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika*, *65*, 167-185.

Okamoto, K. (1969). Optimality of principal components. In P. R. Krishinaiah (Ed.), *Multivariate* *analysis*, vol.II, pp. 673-687, New York: Academic Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazines*, *2*, 559-572.

Sato, M. (1990). Some remarks on principal component analysis as a substitute for factor analysis in monofactor cases. *Journal of the Japan Statistical Society*, **20**, 23-31.

Sočan, G. (2003). *The incremental value of minimum rank factor analysis*. PhD Thesis, University of Groningen: Groningen.

Spearman, C. (1904). "General Intelligence", objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.

Stegeman, A. (2016). A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts. *Computational Statistics and Data Analysis*, *99*, 189-203.

Tanaka, Y., & Tarumi, T. (1995). *Handbook for statistical analysis*: *Multivariate analysis* (*windows version*). Tokyo: Kyoritsu-Shuppan (in Japanese). ten Berge & Kiers

ten Berge, J. M. F., & Kiers, H. A. L. (1996). Optimality criteria for principal component analysis and generalizations. *British Journal of Mathematical and Statistical Psychology*, *49*, 335-345.

Thurstone, L. L. (1935). *The vectors of mind*. University if Chicago Press: Chicago.

Trendafilov, N. T. Unkel, S., & Krzanowski, W. (2013). Exploratory factor and principal component analyses: some new aspects. *Statistics and Computing*, *23*, 209-220.

Unkel, S. and Trendafilov, N. T. (2010a). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, *78*, 363-382.