

Computational Identification of Confirmatory Factor Analysis Model with Simplimax Procedures

Jingyu Cai^{1*}, Henk A. L. Kiers², Kohei Adachi¹

¹Graduate School of Human Sciences, Osaka University, Osaka, Japan

²Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

Email: *caijingyu10@yahoo.co.jp

How to cite this paper: Cai, J.Y., Kiers, H.A.L. and Adachi, K. (2021) Computational Identification of Confirmatory Factor Analysis Model with Simplimax Procedures. *Open Journal of Statistics*, 11, 1044-1061.

<https://doi.org/10.4236/ojs.2021.116062>

Received: November 15, 2021

Accepted: December 25, 2021

Published: December 28, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Confirmatory factor analysis (CFA) refers to the FA procedure with some loadings constrained to be zeros. A difficulty in CFA is that the constraint must be specified by users in a subjective manner. For dealing with this difficulty, we propose a computational method, in which the best CFA solution is obtained optimally without relying on users' judgements. The method consists of the procedures at lower (L) and higher (H) levels: at the L level, for a fixed number of zero loadings, it is determined both which loadings are to be zeros and what values are to be given to the remaining nonzero parameters; at the H level, the procedure at the L level is performed over the different numbers of zero loadings, to provide the best solution. In the L level procedure, Kiers' (1994) simplimax rotation fulfills a key role: the CFA solution under the constraint computationally specified by that rotation is used for initializing the parameters of a new FA procedure called simplimax FA. The task at the H level can be easily performed using information criteria. The usefulness of the proposed method is demonstrated numerically.

Keywords

Confirmatory Factor Analysis, Cardinality of Loadings, Simplimax Rotation

1. Introduction

In factor analysis (FA), the variation of p observed variables is assumed to be explained by m common factors and p unique factors, with $m < p$ and the two types of factors mutually uncorrelated. The m common factors serve to explain the variations of all p variables. On the other hand, each of the p unique factors

has a one-to-one correspondence to each variable: a unique factor explains specifically the variation of the corresponding variable that remains unaccounted for by the common factors [1] [2] [3].

The parameters to be estimated in FA are a factor loading matrix $\Lambda = (\lambda_{ij})$ ($p \times m$), a unique variance matrix $\Psi = (\psi_{ii})$ ($p \times p$), and a factor correlation matrix $\Phi = (\phi_{jk})$ ($m \times m$). Here, the loadings in Λ stand for how the variables load on the common factors, Ψ is the diagonal matrix whose diagonal element ψ_{ii} expresses the variance of the i th unique factor, and Φ contains the correlation coefficients among the m common factors. Upon making certain distributional assumptions for the factors, the covariance matrix Σ among p observed variables is modeled as

$$\Sigma = \Lambda\Phi\Lambda' + \Psi. \tag{1}$$

(e.g., Adachi, 2019 [1]; Mulaik, 2010 [2]). Thus, a loss function $f(\Lambda, \Psi, \Phi | S)$ can be defined, which stands for the discrepancy between (1) and its sample counterpart S ($p \times p$): FA can be formulated as minimizing $f(\Lambda, \Psi, \Phi | S)$ over Λ, Ψ , and Φ , for a given S .

FA can be classified into two types; confirmatory (CFA) and exploratory (EFA) [2]. In CFA, some loadings in Λ are constrained to take zero values, while no constraints are imposed on the loadings of EFA. In this paper, we focus on CFA. An example of the CFA model with a particular constraint is illustrated in Figure 1. There, a constrained loading matrix Λ is shown on the left, and the corresponding CFA model is depicted on the right as a diagram, in which only pairs of variables and factors with unconstrained loadings are linked by paths. In Figure 1, the binary matrix B is also presented which specifies the constraints in Λ and the links in the right diagram. The elements in $B = (b_{ij})$ ($p \times m$) are defined generally as

$$b_{ij} = \begin{cases} 0, & \text{iff } \lambda_{ij} = 0 \\ 1, & \text{otherwise} \end{cases}, \tag{2}$$

in other words, $b_{ij} = 1$ if variable i is linked to factor j ; otherwise, $b_{ij} = 0$. In this sense, we call B a link matrix. Any CFA constraint can be expressed as $\Lambda = B \bullet \Lambda$. Here, \bullet denotes the element-wise Hadamard product with $B \bullet \Lambda = (b_{ij}\lambda_{ij})$. It should be kept in mind that specifying a CFA model amounts to selecting a particular link matrix B . Thus, CFA can be formally expressed as

$$\min_{\Lambda, \Psi, \Phi} f(\Lambda, \Psi, \Phi | S) \quad \text{s.t.} \quad \Lambda = B \bullet \Lambda \quad \text{for a specified } B \in \mathbb{S}_B \tag{3}$$

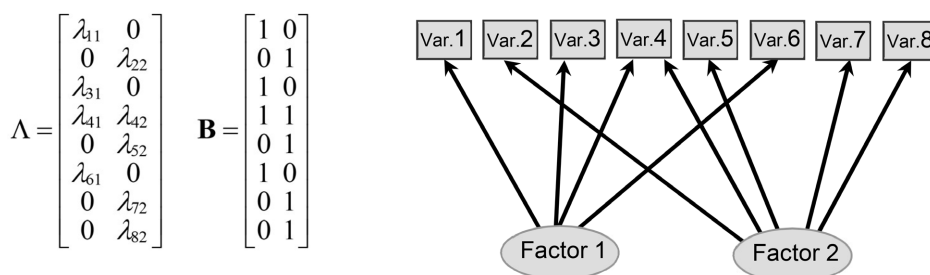


Figure 1. Illustration of constrained loadings and the corresponding CFA model.

with “s.t.” the abbreviation for “subject to” and $\mathbb{S}_{\mathbf{B}}$ denoting a set of considered matrices \mathbf{B} .

A problem in CFA (3) is that the link matrix $\mathbf{B} = (b_{ij})$ defined as (2) must be selected by users. That is, it must be decided in a subjective manner, which elements in \mathbf{B} are to be zeros/ones, in other words, which pairs of variables and factors are linked as in **Figure 1** whether the CFA model with a particular constraint is accepted or not is checked afterwards by referring to the goodness-of-fit of the solution [4] [5] [6]. However, even if the model is found acceptable, better CFA models with other link matrices \mathbf{B} may exist. It implies that all possible \mathbf{B} in $\mathbb{S}_{\mathbf{B}}$ must be considered for finding the best \mathbf{B} . However, this is unfeasible, unless pm is very small, as the number of all possible \mathbf{B} is enormous. That number can be calculated as 2^{pm} , since each of the pm elements in \mathbf{B} takes zero or one as in (2): for example, for $p = 12$ and $m = 3$ one finds $2^{pm} \cong 6.87 \times 10^{10}$. In short, the problems in CFA can be summarized next:

[P1] The link matrix \mathbf{B} (specifying a CFA model) must be selected subjectively by users.

[P2] An enormous number of possible matrices \mathbf{B} in $\mathbb{S}_{\mathbf{B}}$ must be considered for finding the best \mathbf{B} .

To the best of our knowledge, the problems [P1] and [P2] in CFA have not been considered in the existing papers. In order to deal with those problems, we propose an FA procedure for computationally and optimally identifying a suitable CFA model. Here, the model identification includes estimating the model parameter values. The outline of our proposed procedure is described in the next section. Then, we detail the procedure in Section 3, report its assessment in a simulation study in Section 4, give numerical examples in Section 5, and conclude this paper in Section 6.

2. Outline of the Proposed Procedure

First, we outline our approach to the CFA model identification under the condition that the number of zero loadings is fixed to a particular integer in Section 2.1. Then, in Section 2.2, the approach is extended to the cases with the number of zero loadings not being fixed. We then summarize the prospects for the following sections.

2.1. Model Identification for a Specified Number of Nonzero Loadings (Cardinality)

For dealing with the difficulties [P1] and [P2] in CFA, we can consider the two procedures introduced in the next paragraphs, on the condition that

$\text{Card}(\mathbf{A}) = \text{Card}(\mathbf{B})$, that is, the number of nonzero values in \mathbf{A} and \mathbf{B} the matrix between parentheses equals a specified integer c hence

$$\text{Card}(\mathbf{A}) = c, \text{ or equivalently, } \text{Card}(\mathbf{B}) = c. \quad (4)$$

Clearly, $pm - c$ equals the number of zeros in \mathbf{B} or \mathbf{A} .

The first procedure considered can be formulated as

$$\min_{\Lambda, \Psi, \Phi} f(\Lambda, \Psi, \Phi | \mathbf{S}) \quad \text{s.t. } \Lambda = \mathbf{B} \bullet \Lambda, \text{ after } \mathbf{B} \text{ is estimated optimally s.t. (4).} \quad (5)$$

This minimization is rewritten as performing CFA under the constraint indicated by link matrix \mathbf{B} , with \mathbf{B} estimated by another method in advance. Thus, the problem [P1] is overcome. Further, we can also consider that [P2] is dealt with, supposed that the value c in (4) and \mathbf{B} are suitable.

The above procedure consists of two stages: first \mathbf{B} is estimated, then CFA is performed, see (5). In contrast, the second procedure considered is formulated with a single stage as follows:

$$\min_{\mathbf{B}, \Lambda, \Psi, \Phi} f(\mathbf{B}, \Lambda, \Psi, \Phi | \mathbf{S}) \quad \text{s.t. } \Lambda = \mathbf{B} \bullet \Lambda \text{ and (4).} \quad (6)$$

Here, \mathbf{B} has been added to the subscripts of “min”: the link matrix \mathbf{B} , which indicates the pairs of variables and factors to be linked, is estimated jointly with the other parameters Λ , Ψ , and Φ .

Between (5) and (6) or (6'), we can find the following difference: in minimizing loss function $f(\Lambda, \Psi, \Phi | \mathbf{S})$, the \mathbf{B} value is kept fixed in (5), but allowed to change in (6). The difference implies that the resulting loss function value in (6) cannot exceed that value in (5):

$$\min_{\mathbf{B}, \Lambda, \Psi, \Phi} f(\mathbf{B}, \Lambda, \Psi, \Phi | \mathbf{S}) \leq \min_{\Lambda, \Psi, \Phi} f(\Lambda, \Psi, \Phi | \mathbf{S}). \quad (7)$$

This inequality shows that (6) can provide a better solution than (5). We will propose an iterative algorithm for (6), which will be started by the optimum for (5). As empirically shown later, in almost all cases, the solutions of (5) and (6) are equivalent: we can obtain the final solutions only by (5), without performing the iterative algorithm for (6). However, it is worth to perform the latter step, as (6) can provide a better solution in a few cases.

In this paper, we use the maximum likelihood (ML) method for estimating the parameters in (5) and (6). This implies that the loss function to be minimized is given as the negative of the log likelihood. It is explicitly expressed as

$$f(\Lambda, \Psi, \Phi | \mathbf{S}) = \log |\Sigma| + \text{tr} \mathbf{S} \Sigma^{-1} = \log |\Lambda \Phi \Lambda' + \Psi| + \text{tr} \mathbf{S} (\Lambda \Phi \Lambda' + \Psi)^{-1}, \quad (8)$$

following from the normality assumptions for factors [7]. For minimizing (8), we use Rubin & Thayer's [8] EM algorithm for FA, whose properties are discussed in Adachi [9]. This algorithm is detailed in **Appendix A1**.

2.2. Selection of the Best Cardinality

We should notice that the above approach is conditional upon c in (4). Thus, it remains to select the best value for c . This can be attained by the following procedure:

$$\text{Select the value } c \text{ with the lowest } IC(c) \text{ among } c = c_{\min}, \dots, c_{\max}. \quad (9)$$

Here, c_{\min}/c_{\max} expresses the reasonable minimum/maximum of c , and $IC(c)$ denotes the value of an information criterion statistic [10], with the statistic being a function of c . The information criteria consist of some statistics which can be defined in ML-based procedures, and they show smaller values for better solutions. Explicit expressions of $IC(c)$ are introduced later. We can rewrite (9) as

follows: the procedure of (6) following (5) (in the last subsection) is performed for $c = c_{\min}, \dots, c_{\max}$, so that we have multiple solutions, among which the solution with the least $IC(c)$ is selected as the best one. The largest interval of $[c_{\min}, c_{\max}]$ is obviously $[1, pm]$. It can be reasonably reduced as

$$c_{\min} = p \quad \text{and} \quad c_{\max} = pm - \frac{m(m-1)}{2} \tag{10}$$

This is because, solutions with c smaller than p would surely have one or more zero rows of loadings. On the other hand, it is considered in the c_{\max} value that $m(m-1)/2$ elements in Λ can be set to zeros without a change in the values of loss function (8).

Now, let us discuss how our proposed procedure is related to the set $\mathbb{S}_{\mathbf{B}}$ containing all possible \mathbf{B} , using $\mathbb{S}_{\mathbf{B}}(c)$ for the subset of $\mathbb{S}_{\mathbf{B}}$ that contains the matrices \mathbf{B} with c nonzeros. The number of \mathbf{B} in $\mathbb{S}_{\mathbf{B}}(c)$ is given by

$$N_{\mathbf{B}}(c) = {}_{pm}C_c, \tag{11}$$

i.e., the number of the combinations of the c elements being ones among all pm elements in \mathbf{B} . Thus, the number of all \mathbf{B} contained in $\mathbb{S}_{\mathbf{B}}$ is given by

$$N_{\mathbf{B}} = \sum_{c=c_{\min}}^{c_{\max}} N_{\mathbf{B}}(c) \tag{12}$$

with (10) and (11). For example, when $p = 12$ and $m = 3$, the value in (12) is approximately 1.25×10^9 , which is enormous (though less than $2^{pm} \cong 6.87 \times 10^{10}$ presented in the last section where (10) was not yet considered). However, it is not required to assess all $N_{\mathbf{B}}$ matrices \mathbf{B} in $\mathbb{S}_{\mathbf{B}}$ in our proposed procedure, as the performance of (6) following (5) allows us to find the optimal \mathbf{B} in $\mathbb{S}_{\mathbf{B}}(c)$ for a particular c , with this performance made for $c = c_{\min}, \dots, c_{\max}$ as in (9). That is, our proposed procedure can find the optimal \mathbf{B} among all $N_{\mathbf{B}}$ matrices \mathbf{B} in $\mathbb{S}_{\mathbf{B}}$ by the $c_{\max} - c_{\min} + 1$ runs of (6) following (5), but not by assessing all \mathbf{B} . The example of $c_{\max} - c_{\min} + 1 = 22$ for $p = 12$ and $m = 3$ demonstrates how easily we can arrive at the optimal \mathbf{B} .

The remaining parts of this paper are organized as follows: in Sect 3.1 and Sect 3.2, we detail (5) and (6) in turn. There, it is described that Kiers' [11] simplimax rotation is used for the optimal estimation of \mathbf{B} subject to (4) in (5) and the idea of this rotation also fulfills a key role in (6). Thus, the procedures for (5) and (6) are particularly called simplimax-based CFA (SbCFA) and simplimax FA (SimpFA), respectively. In the section after treating SbCFA (5) and SimpFA (6), we detail the whole process of the proposed procedure, *i.e.*, SimpFA (6) following SbCFA (5) with the cardinality selection by (9). The proposed procedure is studied in a simulation study and illustrated with real data examples, as reported in the sections before concluding this paper.

3. Simplimax-Based Confirmatory Factor Analysis

In this section, we detail the procedure formulated as (5) with $f(\Lambda, \Psi, \Phi | \mathbf{S})$ defined as (8). A key point is that we will use exploratory FA (EFA) followed by

Kiers' [11] simplimax rotation for estimating \mathbf{B} optimally subject to (4) in (5). That is, in the procedure, firstly EFA is performed for a data set, secondly simplimax is applied to the EFA solution for estimating the link matrix \mathbf{B} , and, finally, the resulting \mathbf{B} is used for minimizing (8) subject to $\mathbf{\Lambda} = \mathbf{B} \bullet \mathbf{\Lambda}$. These three stages are described respectively in the following subsections. The last one can be regarded as *confirmatory FA (CFA)* on the basis of the \mathbf{B} given by simplimax. We thus call the procedure in this section simplimax-based CFA (SbCFA).

3.1. Exploratory Factor Analysis

Let \mathbf{T} denote any $m \times m$ nonsingular matrix satisfying

$$\text{diag}(\mathbf{T}\mathbf{T}') = \mathbf{I}_m, \quad (13)$$

where $\text{diag}(\mathbf{T}\mathbf{T}')$ is the diagonal matrix whose diagonal elements are those of $\mathbf{T}\mathbf{T}'$. EFA has rotational indeterminacy as shown next:

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{T}^{-1}\mathbf{T}\mathbf{T}'\mathbf{T}^{-1}\mathbf{\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda}_T\mathbf{\Phi}\mathbf{\Lambda}_T' + \mathbf{\Psi} \quad (14)$$

with $\mathbf{\Phi} = \mathbf{T}\mathbf{T}'$ and $\mathbf{\Lambda}_T = \mathbf{\Lambda}\mathbf{T}^{-1}$. Here, the latter matrix can also be regarded as the loading matrix. Thus, even if $\mathbf{\Phi} = \mathbf{T}\mathbf{T}'$ is fixed to \mathbf{I}_m by choosing \mathbf{T} that meets $\mathbf{T}\mathbf{T}' = \mathbf{I}_m$, the (8) value remains unchanged when $\mathbf{\Lambda}$ is replaced by $\mathbf{\Lambda}\mathbf{T}^{-1}$. By taking account of this property, $f(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} = \mathbf{I}_m | \mathbf{S})$, i.e., (8) with $\mathbf{\Phi} = \mathbf{I}_m$ is minimized over $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ in EFA. This is attained with the EM algorithm described in **Appendix A1**. We use $\mathbf{\Lambda}_E$ for $\mathbf{\Lambda}$ resulting from EFA.

The indeterminacy (14) implies that $\mathbf{\Lambda}_T = \mathbf{\Lambda}_E\mathbf{T}^{-1}$ and $\mathbf{\Phi} = \mathbf{T}\mathbf{T}'$ can also be the EFA solutions of factor loading and correlation matrices, respectively, with $f(\mathbf{\Lambda}_E, \mathbf{\Psi}, \mathbf{\Phi} = \mathbf{I}_m | \mathbf{S}) = f(\mathbf{\Lambda}_T, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S})$. This fact is often exploited by obtaining \mathbf{T} that allows $\mathbf{\Lambda}_T = \mathbf{\Lambda}_E\mathbf{T}$ to be interpretable. The procedures for obtaining \mathbf{T} are generally referred to as factor rotation.

3.2. Simplimax Rotation

This subsection concerns the factor rotation intermedating between the first EFA and the final CFA stages. Here, the loading matrix is unconstrained in EFA, but constrained to meet $\mathbf{\Lambda} = \mathbf{B} \bullet \mathbf{\Lambda}$ in CFA as in (3). The latter constrainedness leads to

$$\min_{\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi}} f(\mathbf{B} \bullet \mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S}) \geq f(\mathbf{\Lambda}_E, \mathbf{\Psi}, \mathbf{\Phi} = \mathbf{I} | \mathbf{S}) = f(\mathbf{\Lambda}_T, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S}), \quad (15)$$

where $f(\mathbf{B} \bullet \mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S})$ stands for the loss function in (3) in which the CFA constraint $\mathbf{\Lambda} = \mathbf{B} \bullet \mathbf{\Lambda}$ substituted. Inequality (15) implies that the value of the EFA loss function $f(\mathbf{\Lambda}_E, \mathbf{\Psi}, \mathbf{\Phi} = \mathbf{I} | \mathbf{S}) = f(\mathbf{\Lambda}_T, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S})$ gives the lower limit of $f(\mathbf{B} \bullet \mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S})$ to be minimized in CFA (3). This suggests that the best attainable CFA solution would be one which provides the $\min_{\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi}} f(\mathbf{B} \bullet \mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S})$ value close to the EFA counterpart the $f(\mathbf{\Lambda}_E, \mathbf{\Psi}, \mathbf{\Phi} = \mathbf{I} | \mathbf{S}) = f(\mathbf{\Lambda}_T, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S})$ value.

Now if \mathbf{T} can be chosen in the factor rotation such that $\mathbf{\Lambda}_T = \mathbf{\Lambda}_E\mathbf{T}^{-1}$ is similar

to a matrix that can be written as $\mathbf{B} \bullet \mathbf{\Lambda}$ for particular matrices \mathbf{B} and $\mathbf{\Lambda}$, then we can expect that such \mathbf{B} and $\mathbf{\Lambda}$ will be good candidates for giving a CFA solution with a fit close to that of EFA. This can be attained by the simplimax rotation [11], which is formulated as minimizing the simplimax function

$$spX(\mathbf{B}, \mathbf{\Lambda}, \mathbf{T}) = \|\mathbf{B} \bullet \mathbf{\Lambda} - \mathbf{\Lambda}_E \mathbf{T}^{-1}\|^2 \tag{16}$$

over \mathbf{B} , $\mathbf{\Lambda}$, and \mathbf{T} subject to (4) and (13). Thus, this rotation serves as a suitable bridge between the first and final stages.

For the constrained minimization of (16), two steps are iterated alternately. In one of them, the simplimax function (16) is minimized over \mathbf{T} under (13) for a given $\mathbf{B} \bullet \mathbf{\Lambda}$, using Browne’s [12] algorithm. In another step, (16) is minimized over $\mathbf{B} = (b_{ij})$ and $\mathbf{\Lambda}$ under (8) for a given $\mathbf{\Lambda}_E \mathbf{T}^{-1}$. As this problem is also related to the procedure in the next section, we express the minimization in a generalized form as

$$\min_{\mathbf{B}, \mathbf{\Lambda}} spX(\mathbf{B}, \mathbf{\Lambda} | \mathbf{H}) = \|\mathbf{B} \bullet \mathbf{\Lambda} - \mathbf{H}\|^2 \text{ for a given } p \times m \text{ matrix } \mathbf{H} = (h_{jk}), \tag{17}$$

with $\mathbf{H} = \mathbf{\Lambda}_E \mathbf{T}^{-1}$ in the present context. As explained in **Appendix A3**, (17) can be attained for

$$b_{ij} = \begin{cases} 1, & \text{if } h_{ij}^2 \geq h_{[c]}^2 \\ 0, & \text{otherwise} \end{cases} \text{ and } \mathbf{\Lambda} = \mathbf{H}, \tag{18}$$

with $h_{[c]}^2$ denoting the c th largest value of all elements in $\mathbf{H} \bullet \mathbf{H} = (h_{ij}^2)$. The steps in the simplimax rotation algorithm is listed in **Appendix A2**.

3.3. Confirmatory Factor Analysis Following Simplimax Rotation

The final stage is simply to perform CFA using \mathbf{B} obtained by the simplimax rotation. Thus, SbcFA is formulated by making (5) concrete as

$$\min_{\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi}} f(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S}) \text{ s.t. } \mathbf{\Lambda} = \mathbf{B} \bullet \mathbf{\Lambda} \tag{5'}$$

after \mathbf{B} is estimated s.t. $\text{Card}(\mathbf{B}) = c$ by simplimax

with $f(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S})$ defined as (8). The minimization in (5) or (5') is attained with the EM algorithm, which is described in **Appendix A1**.

3.4. Simplimax Factor Analysis

We will now describe simplimax FA (SimpFA). This is the procedure minimizing (6), using the solution from SbcFA as a starting configuration. Specifically, we minimize the negative of the log likelihood (see (8)), with $\mathbf{B} \bullet \mathbf{\Lambda}$ is substituted into $\mathbf{\Lambda}$, over \mathbf{B} , $\mathbf{\Lambda}$, $\mathbf{\Psi}$, and $\mathbf{\Phi}$ subject to $\text{card}(\mathbf{\hat{a}}) = c$ (see (4)). We call this procedure simplimax FA (SimpFA), as this is a new FA procedure, and the minimization of the simplimax function (17) fulfills a key role as will be seen in the next paragraph.

The EM algorithm described in **Appendix A1** can also be used for SimpFA. The algorithm for SimpFA differs from that for the other FA procedures, in that the former includes the step for minimizing (8) over both \mathbf{B} and $\mathbf{\Lambda}$ with $\mathbf{\Psi}$ and $\mathbf{\Phi}$ fixed. An innovative feature in the algorithm is to use Kiers’ [13] [14] majorization

method for dealing with difficulties in treating (8) as a function of \mathbf{B} and $\mathbf{\Lambda}$. In that method, an auxiliary function that majorizes (8) is minimized. For this minimization,

$$spx(\mathbf{B}, \mathbf{\Lambda} | \mathbf{W}) = \|\mathbf{B} \bullet \mathbf{\Lambda} - \mathbf{W}\|^2, \tag{19}$$

i.e., the simplimax function in (17) with $\mathbf{H} = \mathbf{W}$, fulfills a key role. Here, $\mathbf{W} = \mathbf{\Lambda}_C + \mathbf{Q}'\mathbf{\Lambda}'_C\mathbf{\Psi}^{-1} - \text{tr}\mathbf{C}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}_C$ with \mathbf{Q} and \mathbf{C} defined in A.1 and $\mathbf{\Lambda}_C$ is the current $\mathbf{\Lambda}$ value (before update), see **Appendix A4** for a derivation.

3.5. Multiple Runs of SimpFA Following SbcFA

Here, we describe the procedures in SimpFA following SbcFA, in which SbcFA provides the initial values of the parameters in SimpFA providing the final solution. We take a multiple run approach for SimpFA following SbcFA, in order to reduce the possibility of selecting a local minimizer as the optimal solution: the algorithm of SbcFA followed by SimpFA is run multiple times by starting SbcFA with mutually different initial values, and the best solution is selected among the multiple SimpFA solutions. The procedure is listed as follows:

Stage 1. For \mathbf{S} , perform EFA to provide $\mathbf{\Lambda}_E$ and set $l = 1$.

Stage 2. For each of $l = 1, \dots, 100$, perform the following sub-stages:

Stage 2.1. Initialize \mathbf{T} to \mathbf{T}_l and perform simplimax. Express the resulting \mathbf{B} as \mathbf{B}_l .

Stage 2.2. For \mathbf{S} , perform CFA using \mathbf{B}_l as \mathbf{B} . Express the resulting $\mathbf{\Lambda}$, $\mathbf{\Psi}$, and $\mathbf{\Phi}$ as $\tilde{\mathbf{\Lambda}}_l$, $\tilde{\mathbf{\Psi}}_l$, and $\tilde{\mathbf{\Phi}}_l$, respectively, with their set $\tilde{\mathbf{\Theta}}_l = \{\tilde{\mathbf{\Lambda}}_l, \tilde{\mathbf{\Psi}}_l, \tilde{\mathbf{\Phi}}_l\}$.

Stage 2.3. For \mathbf{S} , perform SimpFA with $\mathbf{\Lambda}$, $\mathbf{\Psi}$, and $\mathbf{\Phi}$ initialized at $\tilde{\mathbf{\Lambda}}_l$, $\tilde{\mathbf{\Psi}}_l$, and $\tilde{\mathbf{\Phi}}_l$, respectively. Express the resulting $\mathbf{\Lambda}$, $\mathbf{\Psi}$, and $\mathbf{\Phi}$ as $\mathbf{\Lambda}_b$, $\mathbf{\Psi}_b$, and $\mathbf{\Phi}_b$ respectively, with their set $\mathbf{\Theta}_l = \{\mathbf{\Lambda}_l, \mathbf{\Psi}_l, \mathbf{\Phi}_l\}$.

Stage 3. Select $\{\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Phi}}\}$ with $f(\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Phi}} | \mathbf{S}) = \min_l f(\mathbf{\Lambda}_l, \mathbf{\Psi}_l, \mathbf{\Phi}_l | \mathbf{S})$ as the optimal solution.

Here, the initial value \mathbf{T}_l for \mathbf{T} in Stage 2.1 is chosen as follows: \mathbf{T}_l is obtained by the varimax rotation for $\mathbf{\Lambda}_E$ if $l = 1$; otherwise, \mathbf{T}_l is set to $\text{diag}(\mathbf{T}_0\mathbf{T}'_0)^{-1/2}\mathbf{T}_0$ with the elements in \mathbf{T}_0 chosen randomly: the resulting \mathbf{T}_l can be substituted into \mathbf{T} . In **Figure 2**, the above stages are graphically illustrated, so that they can be captured visually.

3.6. Selection of Cardinality by Information Criteria

The final solution $\{\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Phi}}\}$ resulting in Stage 3 (the last subsection) depends on the cardinality c value in (8). We thus use $\{\hat{\mathbf{\Lambda}}_c, \hat{\mathbf{\Psi}}_c, \hat{\mathbf{\Phi}}_c\}$ for the solution $\{\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Phi}}\}$ for a particular value of c . The best value for c can be selected with the procedure (9). This can be rewritten using $\{\hat{\mathbf{\Lambda}}_c, \hat{\mathbf{\Psi}}_c, \hat{\mathbf{\Phi}}_c\}$ as

$$\text{Choose the } \{\hat{\mathbf{\Lambda}}_{c^*}, \hat{\mathbf{\Psi}}_{c^*}, \hat{\mathbf{\Phi}}_{c^*}\} \text{ for the value } c^* = \arg \min_{c_{\min} \leq c \leq c_{\max}} IC(c) \tag{20}$$

with c_{\min} and c_{\max} defined as (10).

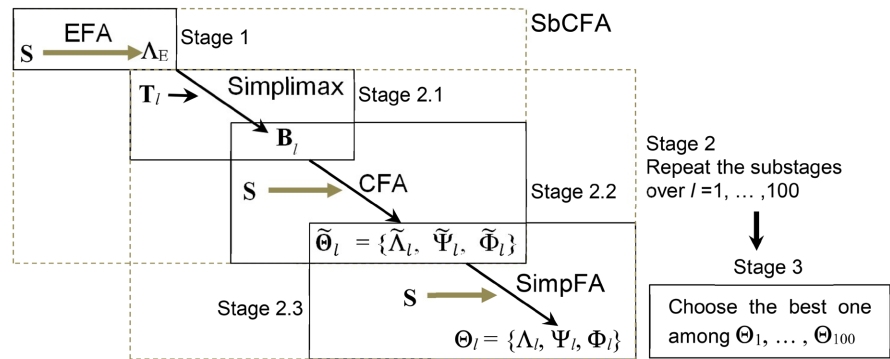


Figure 2. Graphical illustration of Stages 1 - 3.

As $IC(c)$ in (20), we consider using either of Akaike’s [15] information criterion (AIC) or Schwarz’s [16] Bayesian information criterion (BIC), which are particularly popular in the information criteria. AIC and BIC can be calculated as

$$AIC(c) = 2f(\hat{\Lambda}_c, \hat{\Psi}_c, \hat{\Phi}_c | \mathbf{S}) + 2\kappa(c), \tag{21}$$

$$BIC(c) = 2f(\hat{\Lambda}_c, \hat{\Psi}_c, \hat{\Phi}_c | \mathbf{S}) + \log n\kappa(c), \tag{22}$$

respectively, for a particular value of c . Here, n is the number of observations, and $\kappa(c) = c + a$ with a the number of unique variances plus the number of inter-factor correlations. Thus, (21) or (22) is substituted into $IC(c)$ in (20). Whether we should use (21) or (22) is assessed in the next section.

4. Simulation Study

We performed a simulation study to assess the proposed procedure in the last section, with respect to [1] how often the c value in (8) is selected correctly by AIC and BIC ; [2] how similar the SbcCFA and SimpFA solutions are; [3] how well parameter values are recovered. The procedures for synthesizing and analyzing data in the study are described in the first subsection, then the results for [1] [2] and [3] are reported in the following ones.

4.1. Data Synthesis and Analysis

With $p = 12$, and $m = 3$, we set the true $\{\Lambda, \Psi, \Phi\}$ as in Table 1 and sampled $n = 300$ rows of an $n \times p$ data matrix \mathbf{X} from the p -variate normal distribution whose average is $\mathbf{0}_p$ and covariance matrix is defined as $\Sigma = \Lambda\Phi\Lambda' + \Psi$, see (1). The resulting \mathbf{X} provided the sample covariance matrix $\mathbf{S} = n^{-1}\mathbf{X}'\mathbf{X}$. We replicated this procedure 200 times to have 200 matrices \mathbf{S} .

For each \mathbf{S} , we carried out the procedure of SimpFA following SbcCFA with the selection of the best c by (20), using both AIC and BIC . It gives the SimpFA solutions $\{\hat{\Lambda}_c, \hat{\Psi}_c, \hat{\Phi}_c\}$, which are classified into two types according to whether c was chosen by AIC or BIC .

4.2. Correctness of Selected Cardinality

Let c_{true} denote the true $\text{Card}(\Lambda)$. As found in Table 1, $c_{\text{true}} = 15$. On the other

Table 1. True parameters in Λ , Ψ_{1_p} , and Φ with blank cells indicating zero elements.

Λ		Ψ_{1_p}	Φ		
-0.9		0.2	1	0.2	-0.3
0.8		0.3	0.2	1	0.1
0.7		0.5	-0.3	0.1	1
0.6	-0.6	0.4			
	0.9	0.2			
	-0.8	0.4			
	0.7	0.5			
	0.6	-0.6	0.3		
		0.9	0.2		
		0.8	0.3		
		-0.7	0.5		
-0.6		0.6	0.4		

hand, c^* expresses $\text{Card}(\Lambda)$ selected by the proposed procedure (20). We obtained the deviation $c^* - c_{\text{true}}$ for each of the 200 data set. The averages (standard deviation) of these deviations over all data sets are 2.43 (1.40) when using *AIC* and 0.34 (0.62) when using *BIC*. This result shows that *AIC* tends to overestimate $\text{Card}(\Lambda)$. Moreover, the mean absolute bias $|c^* - c_{\text{true}}|$ for *BIC* was smaller than that for *IC* for 184 data sets among the 200 ones. This result shows that *BIC* is substantially better than *AIC*. Accordingly, we take only the *BIC*-based solution into consideration from here.

4.3. Equivalence of SbcFA and SimpFA Solutions

It was found that 98 percent of the 200 solutions of SimpFA were equivalent to those of the SbcFA ones used for initializing the SimpFA parameters: no iteration in the SimpFA algorithm was required. Also in the remaining two percent of the runs, SbcFA and SimpFA solutions were almost equivalent, with the average of the differences between the two solutions over the 200 data sets being 0.001. Here, the difference is defined as

$$\left(\|\hat{\Lambda}_{c^*} - \tilde{\Lambda}_{c^*}\|_1 + \|\hat{\Psi}_{c^*} - \tilde{\Psi}_{c^*}\|_1 + \|\hat{\Phi}_{c^*} - \tilde{\Phi}_{c^*}\|_1 \right) / \{pm + p - m(m-1)\} \text{ with } \|\Gamma\|_1 = \sum_i \sum_j |\gamma_{ij}| \text{ denoting the } L_1 \text{ norm of a matrix } \Gamma = (\gamma_{ij}) \text{ and } \{\tilde{\Lambda}_{c^*}, \tilde{\Psi}_{c^*}, \tilde{\Phi}_{c^*}\} \text{ being the SbcFA counterpart of } \{\hat{\Lambda}_{c^*}, \hat{\Psi}_{c^*}, \hat{\Phi}_{c^*}\}.$$

The above results do not show that the SimpFA (following SbcFA) is useless, as SimpFA serves for showing the equivalence of its solution to the SbcFA one, which allows us to find the optimality of the SbcFA solution.

4.4. Recovery of Parameters

We assess how well the true parameter values are recovered by the SimpFA solution $\hat{\Theta}_{c^*} = \{\hat{\Lambda}_{c^*}, \hat{\Psi}_{c^*}, \hat{\Phi}_{c^*}\}$, which is provided by (20) with $IC(c) = BIC(c)$. As

the indices standing for the badness in the recovery for Λ , Ψ , and Φ , we obtained $\|\hat{\Lambda}_{c^*} - \Lambda_{\text{true}}\|_1 / (pm)$, $\|\hat{\Psi}_{c^*} - \Psi_{\text{true}}\|_1 / p$, and $\|\hat{\Phi}_{c^*} - \Phi_{\text{true}}\|_1 / M$, respectively. Further, for assessing the incorrectness in identifying the true zero and nonzero loadings, we recorded MIR_0 = the number of the true zero loadings estimated as nonzero / $(pm - c)$, and $MIR_{\#}$ = the number of the true nonzero loadings estimated as zero / c , for each data set, with $M = m(m - 1)$. Here MIR stands for misidentification rate.

The statistics of the resulting five index values over the 200 data sets are presented in **Table 2**. It shows that the proposed procedure recovered $\{\hat{\Lambda}_{c^*}, \hat{\Psi}_{c^*}, \hat{\Phi}_{c^*}\}$, *i.e.*, the links of variables to factors and parameter values, fairly well, except that the 95 percentile for factor correlations can be considered large. These results allow us to conclude that the true CFA model can be recovered satisfactorily, though the estimates of factor correlations might deviate from the true counterparts in a few cases.

5. Real Data Demonstration

In order to demonstrate how useful our proposed procedure is, we apply it to two data sets that have already been analyzed by CFA with zero constraints selected by users. The solutions of the latter user-based CFA are compared with those for our proposed procedure.

The first data set is Carlson and Mulaik’s [17] personality trait data matrix of $n = 280$ (participants) $\times p = 15$ (personality traits). Its correlation matrix (Mulaik, 2010, p. 198) has been analyzed by Mulaik [2] using CFA with his selected constraints, and its solution is shown in Mulaik (2010, p. 437) and also in **Table 3** together with the solution of the proposed procedure. Here, the latter solution is the same as the one of SbCFA before SimpFA: iteration was not required in the SimpFA algorithm. Comparing the BIC values in **Table 3**, we can find that the solution of the proposed procedure was better than Mulaik’s [2] one.

Another data set is Kojima’s housing preference one with $n = 1120$ (participants) and $p = 13$ (features). It describes to what degree the participants wish to

Table 2. Percentiles, average (Avg) and standard deviation (SD) of each index over 200 data sets.

Index	Percentiles					Avg.	SD
	5	25	50	75	95		
$\ \hat{\Lambda}_{c^*} - \Lambda_{\text{true}}\ _1 / (pm)$	0.011	0.013	0.015	0.017	0.088	0.023	0.035
$\ \hat{\Psi}_{c^*} - \Psi_{\text{true}}\ _1 / p$	0.029	0.035	0.039	0.045	0.053	0.040	0.008
$\ \hat{\Phi}_{c^*} - \Phi_{\text{true}}\ _1 / M$	0.020	0.035	0.049	0.069	0.227	0.071	0.089
MIR ₀	0.000	0.000	0.000	0.000	0.095	0.006	0.026
MIR _#	0.000	0.000	0.000	0.067	0.200	0.032	0.067

Table 3. Solution for personality trait data, with blank cells indicating zero elements.

Variable	User-based CFA: BIC= 924.1			Proposed: BIC = 12.3		
	Λ		Ψ_{1_p}	Λ		Ψ_{1_p}
<i>Friendly</i>	0.85		0.53	0.85		0.29
<i>Sympathetic</i>	0.92		0.39	1.03	-0.19	0.15
<i>Kind</i>	0.93		0.38	1.10	-0.28	0.11
<i>Affectionate</i>	0.90		0.43	0.90		0.19
<i>Intelligent</i>		0.88	0.48		0.88	0.23
<i>Capable</i>		0.93	0.38		0.93	0.15
<i>Competent</i>		0.93	0.38		0.93	0.15
<i>Smart</i>		0.90	0.43		0.90	0.19
<i>Talkative</i>			0.80	0.60		0.80
<i>Outgoing</i>			0.95	0.32		0.95
<i>Gregarious</i>			0.89	0.45		0.89
<i>Extrovert</i>			0.90	0.44		0.90
<i>Helpful</i>	0.73	0.22		0.59	0.74	0.20
<i>Cooperative</i>	0.72	0.23		0.60	0.73	0.21
<i>Sociable</i>	0.17		0.83	0.34	0.18	0.81
Factor	Φ			Φ		
1	1.00	0.22	0.56	1.00	0.25	0.64
2	0.22	1.00	0.30	0.25	1.00	0.30
3	0.56	0.30	1.00	0.64	0.30	1.00

live in the houses featured by the variables. The correlation matrix for this data set is presented in **Table 4**, as it is described in Japanese and not easily available. This matrix has been analyzed by Kojima using CFA with his selected constraints, and its solution was shown in Kojima and also in **Table 5** with the solution of the proposed procedure. Here, this solution differs from the SbCFA counterpart: iteration was required in the SimpFA algorithm. In **Table 5**, BIC shows that the solution of the proposed procedure was better.

In both of the above examples, the proposed procedure outperformed in the BIC values, but its solutions are similar to the counterparts of CFA with users' selected constraints (**Table 3** and **Table 5**). This similarity shows that the selected constraints were rational. However, the proposed procedure is fully computational and does not require any users' effort for considering constraints. This property would be helpful as soon as there is some uncertainty as to which loadings to constrain to 0.

6. Conclusions

A problem in the confirmatory factor analysis (CFA) is that users must select

Table 4. Correlation matrix for Kojima’s (2013) housing preference data.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Food services	1.000	0.423	0.404	0.127	0.179	0.132	0.148	0.214	0.225	0.252	0.125	0.122	0.192
2 Tea services	0.423	1.000	0.751	0.187	0.267	0.115	0.188	0.223	0.265	0.300	0.144	0.097	0.193
3 Home for the old	0.404	0.751	1.000	0.238	0.276	0.178	0.239	0.249	0.281	0.325	0.174	0.147	0.220
4 Lush greenery	0.127	0.187	0.238	1.000	0.411	0.399	0.350	0.286	0.274	0.273	0.150	0.192	0.176
5 Walking and joking	0.179	0.267	0.276	0.411	1.000	0.545	0.330	0.302	0.296	0.333	0.235	0.248	0.224
6 Large park	0.132	0.115	0.178	0.399	0.545	1.000	0.298	0.327	0.229	0.362	0.173	0.242	0.143
7 River and lake	0.148	0.188	0.239	0.350	0.330	0.298	1.000	0.266	0.205	0.240	0.140	0.153	0.161
8 Communal events	0.214	0.223	0.249	0.286	0.302	0.327	0.266	1.000	0.417	0.554	0.338	0.293	0.324
9 Multigenerational living	0.225	0.265	0.281	0.274	0.296	0.229	0.205	0.417	1.000	0.371	0.223	0.171	0.222
10 Active community	0.252	0.300	0.325	0.273	0.333	0.362	0.240	0.554	0.371	1.000	0.286	0.291	0.283
11 Interactions for hobbies	0.125	0.144	0.174	0.150	0.235	0.173	0.140	0.338	0.223	0.286	1.000	0.335	0.452
12 House party	0.122	0.097	0.147	0.192	0.248	0.242	0.153	0.293	0.171	0.291	0.335	1.000	0.333
13 Utilizing own’s careers	0.192	0.193	0.220	0.176	0.224	0.143	0.161	0.324	0.222	0.283	0.452	0.333	1.000

Table 5. Solution for housing preference data, with blank cells indicating zero elements.

Variable	(A) User-based CFA: BIC = 10915.5				(B) Proposed: BIC = 10864.2			
	Λ		Ψ_{1p}		Λ		Ψ_{1p}	
Food services	0.48		0.77	0.39	0.16		0.76	
Tea services	0.86		0.27	0.89			0.21	
Home for the old	0.88		0.23	0.81	0.09		0.28	
Lush greenery		0.59	0.65		0.58		0.67	
Walking and joking		0.74	0.46		0.82	-0.11	0.45	
Large park		0.69	0.52	-0.18	0.78		0.48	
River and lake		0.49	0.77		0.48		0.77	
Communal events			0.45	-0.17		0.84	0.40	
Multigenerational living			0.70			0.55	0.70	
Active community			0.47			0.72	0.49	
Interactions for hobbies			0.66	0.57			0.67	0.55
House party			0.53	0.72		0.15	0.43	0.73
Utilizing own’s careers			0.66	0.57			0.67	0.55
Factor	Φ				Φ			
1	1.00	0.38	0.46	0.32	1.00	0.41	0.50	0.28
2	0.38	1.00	0.65	0.47	0.41	1.00	0.69	0.44
3	0.46	0.65	1.00	0.65	0.50	0.69	1.00	0.62
4	0.32	0.47	0.65	1.00	0.28	0.44	0.62	1.00

what pairs of variables and factors are linked, in other words, what loadings are to be zero or nonzero, in a subjective manner. To deal with this problem, we proposed the procedure of SimpFA following SbCFA for computing an optimally suitable CFA model and its solution without relying on any user's judgment. The simulation study showed that the true CFA model and parameter values can be recovered fairly well by the proposed procedure. Real data examples demonstrated that it can outperform CFA with users' selected constraints in terms of the BIC statistic.

In Section 4.3, we found the SbCFA solutions to be equivalent to SimpFA in almost all cases. In particular, the good performance of SbCFA was somewhat surprising. To theoretically study, reasons for such a SbCFA performance are considered as a subject for a future study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Adachi, K. (2019) Factor Analysis: Latent Variable, Matrix Decomposition, and Constrained Uniqueness Formulations. *WIREs Computational Statistics*, **11**, e1458. <https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1458>
<https://doi.org/10.1002/wics.1458>
- [2] Mulaik, S.A. (2010) Foundations of Factor Analysis. 2nd Edition, CRC Press, Boca Raton.
- [3] Yanai, H. and Ichikawa, M. (2007) Factor Analysis. In: Rao, C.R. and Sinharay, S., Eds., *Handbook of Statistics*, Vol. 26, Elsevier, Amsterdam, 257-296. [https://doi.org/10.1016/S0169-7161\(06\)26009-7](https://doi.org/10.1016/S0169-7161(06)26009-7)
- [4] Arbuckle, J.L. (2008) AMOS 17 User's Guide. SPSS Inc., Chicago.
- [5] Bollen, K.A. (1989) Structural Equations with Latent Variables. Wiley, New York. <https://doi.org/10.1002/9781118619179>
- [6] Jöreskog, K.G. (1969) A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*, **34**, 183-202. <https://doi.org/10.1002/9781118619179>
- [7] Lawley, D.N. (1941) Further Investigation of Factor Estimation. *Proceedings of the Royal Society of Edinburgh A*, **61**, 176-185. <https://doi.org/10.1017/S0080454100006178>
- [8] Rubin, D.B. and Thayer, D.T. (1982) EM Algorithms for ML Factor Analysis. *Psychometrika*, **47**, 69-76. <https://doi.org/10.1007/BF02293851>
- [9] Adachi, K. (2013) Factor Analysis with EM Algorithm Never Gives Improper Solutions When Sample and Initial Parameter Matrices Are Proper. *Psychometrika*, **78**, 380-394. <https://doi.org/10.1007/s11336-012-9299-8>
- [10] Konishi, S. and Kitagawa, G. (2008) Information Criteria and Statistical Modeling. Springer, New York. <https://doi.org/10.1007/978-0-387-71887-3>
- [11] Kiers, H.A.L. (1994) Simplimax: Oblique Rotation to an Optimal Target with Simple Structure. *Psychometrika*, **59**, 567-579. <https://doi.org/10.1007/BF02294392>
- [12] Browne, M.W. (1972) Oblique Rotation to a Partially Specified Target. *British Journal*

of Mathematical and Statistical Psychology, **25**, 207-212.

<https://doi.org/10.1007/BF02294392>

- [13] Kiers, H.A.L. (1990) Majorization as a Tool for Optimizing a Class of Matrix Functions. *Psychometrika*, **55**, 417-428. <https://doi.org/10.1007/BF02294758>
- [14] Kiers, H.A.L. (2002) Setting Up Alternating Least Squares and Iterative Majorization Algorithms for Solving Various Matrix Optimization Problems. *Computational Statistics and Data Analysis*, **41**, 157-170. [https://doi.org/10.1016/S0167-9473\(02\)00142-1](https://doi.org/10.1016/S0167-9473(02)00142-1)
- [15] Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716-723. <https://doi.org/10.1109/TAC.1974.1100705>
- [16] Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-464. <https://doi.org/10.1214/aos/1176344136>
- [17] Carlson, M. and Mulaik, S.A. (1993) Trait Ratings from Descriptions of Behavior as Mediated by Components of Meaning. *Multivariate Behavioral Research*, **28**, 111-159. https://doi.org/10.1207/s15327906mbr2801_7

Appendix

A1. EM Algorithm for Factor Analysis

In Rubin and Thayer's (1982) EM algorithm for FA, an auxiliary function of (8) is defined as

$$\phi(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi} | \mathbf{S}) = \log |\mathbf{\Psi}| + \text{tr} \mathbf{\Psi}^{-1} (\mathbf{S} - 2\mathbf{C}\mathbf{\Lambda}' + \mathbf{\Lambda}\mathbf{Q}\mathbf{\Lambda}') + \log |\mathbf{\Phi}| + \text{tr} \mathbf{Q}\mathbf{\Phi}^{-1}. \quad (\text{A1})$$

Here, $\mathbf{C} = \mathbf{S}\mathbf{A}$ and $\mathbf{Q} = \mathbf{A}'\mathbf{S}\mathbf{A} + \mathbf{U}$, with \mathbf{A} and \mathbf{U} being computed as

$$\mathbf{A} = (\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi})^{-1} \mathbf{\Lambda}\mathbf{\Phi} \quad \text{and} \quad \mathbf{U} = \mathbf{\Phi}^{1/2} \left[\mathbf{I}_m + (\mathbf{\Phi}^{1/2} \mathbf{\Lambda}') \mathbf{\Psi}^{-1} (\mathbf{\Lambda}\mathbf{\Phi}^{1/2}) \right]^{-1} \mathbf{\Phi}^{1/2} \quad (\text{A2})$$

using the current $\mathbf{\Lambda}$, $\mathbf{\Psi}$, and $\mathbf{\Phi}$ values. In the EM algorithm, (A1) rather than (8) is minimized iteratively. This minimization is known to allow (8) to be minimized (Rubin & Thayer, 1982). This fact also holds true, when $\mathbf{\Lambda}$ is constrained as $\mathbf{\Lambda} = \mathbf{B} \bullet \mathbf{\Lambda}$. Thus, the minimization of (8) in EFA, SbcFA (5) or (5'), and SimpFA (6) or (6') is attained with the EM algorithm.

The EM algorithm for EFA, SbcFA, and SimpFA can be summarized as follows:

- Step 1. Initialize $\mathbf{\Lambda}$, $\mathbf{\Psi}$, and $\mathbf{\Phi}$, with \mathbf{B} also initialized in SimpFA.
- Step 2. Obtain \mathbf{C} and \mathbf{Q} in (A.1) through (A.2).
- Step 3. Minimize (A.1) over $\mathbf{\Psi}$ with $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ fixed.
- Step 4. Minimize (A.1) over $\mathbf{\Lambda}$ with $\mathbf{\Psi}$ and $\mathbf{\Phi}$ fixed, with "over $\mathbf{\Lambda}$ " replaced by "over \mathbf{B} and $\mathbf{\Lambda}$ under (4)" only in SimpFA.
- Step 5. Minimize (A.1) over $\mathbf{\Phi}$ with $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ fixed.
- Step 6. Transform $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ so that $\mathbf{\Phi}$ is a correlation matrix and finish if convergence is reached; otherwise. Go back to Step 2.

Here, the minimization in Step 3 is attained when the i th diagonal elements in $\mathbf{\Psi}$ are updated as $\psi_{ii} = s_{ii} - 2\lambda_i' c_i + \lambda_i' \mathbf{Q} \lambda_i$ ($i = 1, \dots, p$), with λ_i' the i th row of $\mathbf{\Lambda}$, c_i' that of \mathbf{C} , and s_{ii} the (i, i) element of \mathbf{S} . The task of Step 5 is attained simply by updating $\mathbf{\Phi}$ as $\mathbf{\Phi} = \mathbf{Q}$. This is because $\mathbf{\Phi}$ may be regarded simply as a covariance matrix during the iteration and finally be transformed into a correlation matrix. Thus, if convergence is reached in Step 6, we must transform $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ as $\mathbf{\Phi}_R = \text{diag}(\mathbf{\Phi})^{-1/2} \mathbf{\Phi} \text{diag}(\mathbf{\Phi})^{-1/2}$ and $\mathbf{\Lambda}_R = \mathbf{\Lambda} \text{diag}(\mathbf{\Phi})^{1/2}$, then regard the resulting $\mathbf{\Phi}_R$ and $\mathbf{\Lambda}_R$ as the solutions of $\mathbf{\Phi}$ and $\mathbf{\Lambda}$, respectively. Here, we should notice that the transformation does not change the value of (8), which is shown by $\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' = \mathbf{\Lambda}\mathbf{D}(\mathbf{D}^{-1}\mathbf{\Phi}\mathbf{D}^{-1})(\mathbf{\Lambda}\mathbf{D})'$ with \mathbf{D} an $m \times m$ diagonal matrix. Convergence in Step 6 is defined here as the decrease in (6) from the previous round being less than 10^{-6} . The details in Steps 1 and 4 differ among EFA, SbcFA, and SimpFA, as described in the next paragraphs.

The initialization in Step 1 is detailed here. In EFA, the eigenvalue decomposition of \mathbf{S} defined as $\mathbf{S} = \mathbf{L}\mathbf{\Delta}^2\mathbf{L}'$ is used, with $\mathbf{\Delta}^2$ the $p \times p$ diagonal matrix whose diagonal elements are arranged in descending order, and $\mathbf{L}\mathbf{L}' = \mathbf{I}_p$. That is, the initial $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ are set to $\mathbf{L}_m\mathbf{\Delta}_m$ and $\text{diag}(\mathbf{S} - \mathbf{L}_m\mathbf{\Delta}_m^2\mathbf{L}_m')$, respectively, in EFA, with $\mathbf{\Delta}_m^2$ the first $m \times m$ diagonal block of $\mathbf{\Delta}^2$ and \mathbf{L}_m the $p \times m$ matrix containing the first m columns of \mathbf{L} . In SbcFA, the initial $\mathbf{\Psi}$ is set to the one obtained in the preceding EFA, while

the initial Λ and Φ are set to the matrices $\mathbf{B} \bullet \Lambda$ and $\mathbf{T}\mathbf{T}'$, respectively, that are obtained by the preceding simplimax rotation. In SimpFA, Λ , Ψ , and Φ are initialized at their SbCFA solution and \mathbf{B} at its simplimax rotation solution.

The minimization in Step 4 is attained by the update of Λ with $\Lambda = \mathbf{C}\mathbf{Q}^{-1}$ in EFA and the row-wise $\lambda_i = \mathbf{B}'_i(\mathbf{B}_i\mathbf{Q}\mathbf{B}'_i)^{-1}\mathbf{B}_i\mathbf{c}_i$ ($i = 1, \dots, p$) in SbCFA. Here, \mathbf{B}_i is the $m_i \times m$ binary matrix satisfying $\mathbf{1}'_{m_i}\mathbf{B}_i = \mathbf{b}'_i$ with \mathbf{b}'_i the i th row of the link matrix $\mathbf{B} = (b_{ij})$ defined as (2) and $m_i = \mathbf{b}'_i\mathbf{1}_m$: for example, if $\mathbf{b}'_i = [1, 0, 1]$, then $\mathbf{B}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. The Step 4 in SimpFA is detailed in **Appendix A4**.

A2. Algorithm for Simplimax Rotation

The algorithm for the simplimax rotation can be summarized as follows:

- Step 1. Initialize \mathbf{T}
- Step 2. Update \mathbf{B} and Λ as (18) with $\mathbf{H} = \Lambda\mathbf{E}\mathbf{T}^{-1}$
- Step 3. Update \mathbf{T} with Browne's (1972) algorithm
- Step 4. Finish if convergence is reached; otherwise, go back to Step 2.

How \mathbf{T} is initialized is described in the section for the whole process of the proposed procedure. The convergence is defined that the decrease in (16) from the previous round is less than 10^{-6} in this paper.

A3. Derivation of (18)

The simplimax function $spx(\mathbf{B}, \Lambda | \mathbf{H}) = \|\mathbf{B} \bullet \Lambda - \mathbf{H}\|^2$ in (17) can be rewritten as

$$spx(\mathbf{B}, \Lambda | \mathbf{H}) = \sum_{(i,j) \in \mathbb{N}} (b_{ij}\lambda_{ij} - h_{ij})^2 + \sum_{(i,j) \in \mathbb{N}^\perp} h_{ij}^2 \geq \sum_{(i,j) \in \mathbb{N}^\perp} h_{ij}^2. \quad (\text{A3})$$

with $\mathbf{H} = (h_{ij})$. Here, \mathbb{N} denotes the set of the index pairs (i, j) for $b_{ij} = 1$, while \mathbb{N}^\perp is the set of (i, j) for $b_{ij} = 0$, and we have used

$\sum_{(i,j) \in \mathbb{N}^\perp} (b_{ij}\lambda_{ij} - h_{ij})^2 = \sum_{(i,j) \in \mathbb{N}^\perp} h_{ij}^2$. The inequality in (A.3) shows that the lower limit of $spx(\mathbf{B}, \Lambda | \mathbf{H})$ is $\sum_{(i,j) \in \mathbb{N}^\perp} h_{ij}^2$ which is attained if $b_{ij}\lambda_{ij} = h_{ij}$. Furthermore, the limit $\sum_{(i,j) \in \mathbb{N}^\perp} h_{ij}^2$ is minimal when \mathbb{N}^\perp contains the (i, j) for $h_{ij}^2 \leq [h_{ij}^2]_q$, with $q = pm - c$ and $[h_{ij}^2]_q$ the q th smallest h_{ij}^2 value among h_{ij}^2 , $i = 1, \dots, p$; $j = 1, \dots, m$. This can be rewritten as (18). That is, (17) is attained for (18).

A4. Majorization Algorithm for SimpFA Loadings

Let us rewrite (A.1) as $\phi^*(\Lambda) + const$, where $const$ is a part independent of $\Lambda = \mathbf{B} \bullet \Lambda$ and

$$\phi^*(\Lambda) = -2\text{tr}\Psi^{-1}\mathbf{C}\Lambda' + \text{tr}\Psi^{-1}\Lambda\mathbf{Q}\Lambda' = \text{tr}\Psi^{-1}\Lambda\mathbf{Q}\Lambda' - 2\text{tr}\mathbf{C}'\Psi^{-1}\Lambda. \quad (\text{A4})$$

This minimization over $\mathbf{B} \bullet \Lambda$ is found to be the task of Step 4 in **Appendix A1** for SimpFA. Using $\Delta = \Lambda - \Lambda_C$ or $\Lambda = \Lambda_C + \Delta$ with Λ_C the current Λ value, (A4) can be rewritten as

$$\begin{aligned} \phi^*(\Lambda) &= \text{tr}\Psi^{-1}(\Lambda_C + \Delta)\mathbf{Q}(\Lambda_C + \Delta)' - 2\text{tr}\mathbf{C}'\Psi^{-1}(\Lambda_C + \Delta) \\ &= \phi^*(\Lambda_C) + \text{tr}\Psi^{-1}\Delta\mathbf{Q}\Delta' + \text{tr}2\mathbf{V}\Delta \end{aligned} \quad (\text{A4}')$$

with $\mathbf{V} = \mathbf{Q}'\mathbf{\Lambda}_c'\mathbf{\Psi}^{-1} - \mathbf{C}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}_c$. Kiers (1990, Theorem 1) shows that $\text{tr}\mathbf{\Psi}^{-1}\mathbf{\Delta}\mathbf{Q}\mathbf{\Delta}$ in (A4') satisfies the inequality $\text{tr}\mathbf{\Psi}^{-1}\mathbf{\Delta}\mathbf{Q}\mathbf{\Delta}'/\|\mathbf{\Delta}\|^2 \leq \alpha$, i.e., $\text{tr}\mathbf{\Psi}^{-1}\mathbf{\Delta}\mathbf{Q}\mathbf{\Delta}' \leq \alpha\|\mathbf{\Delta}\|^2$, where α is the greatest eigenvalue of $\mathbf{\Psi}^{-1} \otimes \mathbf{Q}$, with \otimes denoting the Kronecker product. Using that inequality, a function majorizing (A4') can be defined as

$$\eta(\mathbf{\Lambda}) = \phi^*(\mathbf{\Lambda}_c) + \beta \text{tr}\|\mathbf{\Delta}\|^2 + 2\text{tr}\mathbf{V}\mathbf{\Delta} = g(\mathbf{\Lambda}_c) + \beta\|\mathbf{\Delta} - \mathbf{V}\|^2 - \beta\|\mathbf{V}\|^2, \quad (\text{A5})$$

with $\beta \geq \alpha$ and $\beta \geq 0$: (14) satisfies $\phi^*(\mathbf{\Lambda}_c) = \eta(\mathbf{\Lambda}_c) \geq \eta(\mathbf{\Lambda}) \geq \phi^*(\mathbf{\Lambda})$ if $\mathbf{\Lambda}$ is the minimizer of $\phi^*(\mathbf{\Lambda}_c)$. Since only $\beta\|\mathbf{\Delta} - \mathbf{V}\|^2$ is a function of $\mathbf{\Lambda}$ with $\beta \geq 0$ on the right side of (A5), the task of Step 4 is attained by minimizing

$\eta^*(\mathbf{\Lambda}) = \|\mathbf{\Delta} - \mathbf{V}\|^2$. From $\mathbf{\Delta} = \mathbf{\Lambda} - \mathbf{\Lambda}_c$, we can rewrite $\eta^*(\mathbf{\Lambda})$ as $\eta^*(\mathbf{\Lambda}) = \|\mathbf{\Lambda} - \mathbf{\Lambda}_c - \mathbf{V}\|^2 = \|\mathbf{\Lambda} - \mathbf{W}\|^2$, which is the simplimax function (19), with $\mathbf{W} = \mathbf{\Lambda}_c + \mathbf{V}$. Thus, \mathbf{B} and $\mathbf{\Lambda}$ to be obtained in Step 4 are given by (18) with $\mathbf{H} = (h_{ij})$ and $h_{[c]}^2$ replaced by $\mathbf{W} = (w_{ij})$ and $w_{[c]}^2$, respectively. Here, $w_{[c]}^2$ is the c th largest value of all elements in $\mathbf{W} \bullet \mathbf{W} = (w_{ij}^2)$.