

FACTOR ANALYSIS WITH EM ALGORITHM NEVER GIVES IMPROPER SOLUTIONS  
 WHEN SAMPLE COVARIANCE AND INITIAL PARAMETER MATRICES ARE PROPER

KOHEI ADACHI

OSAKA UNIVERSITY

Rubin and Thayer (Psychometrika, 47:69–76, 1982) proposed the EM algorithm for exploratory and confirmatory maximum likelihood factor analysis. In this paper, we prove the following fact: the EM algorithm always gives a proper solution with positive unique variances and factor correlations with absolute values that do not exceed one, when the covariance matrix to be analyzed and the initial matrices including unique variances and inter-factor correlations are positive definite. We further numerically demonstrate that the EM algorithm yields proper solutions for the data which lead the prevailing gradient algorithms for factor analysis to produce improper solutions. The numerical studies also show that, in real computations with limited numerical precision, Rubin and Thayer’s (Psychometrika, 47:69–76, 1982) original formulas for confirmatory factor analysis can make factor correlation matrices asymmetric, so that the EM algorithm fails to converge. However, this problem can be overcome by using an EM algorithm in which the original formulas are replaced by those guaranteeing the symmetry of factor correlation matrices, or by formulas used to prove the above fact.

Key words: factor analysis, EM algorithm, improper solutions, maximum likelihood method.

1. Introduction

Using  $\mathbf{y}$  for a  $p \times 1$  observation vector whose expectation  $E(\mathbf{y})$  equals the  $p \times 1$  zero vector  $\mathbf{0}_p$ , the factor analysis model is expressed as  $\mathbf{y} = \mathbf{Bz} + \mathbf{e}$ , with  $\mathbf{B} = (b_{jk})$  a  $p$ -variables  $\times m$ -factors loading matrix,  $\mathbf{z}$  the  $m \times 1$  latent random vector with unit variance,  $\mathbf{e}$  a  $p \times 1$  random error vector, and  $p > m$ . It is assumed that  $E(\mathbf{z}) = \mathbf{0}_m$ ,  $E(\mathbf{e}) = \mathbf{0}_p$ ,  $E(\mathbf{e}\mathbf{e}') = \Psi$ , and  $\mathbf{z}$  is uncorrelated with  $\mathbf{e}$ , where  $\Psi = \text{diag}\{\psi_1, \dots, \psi_p\}$  is the  $p \times p$  diagonal matrix whose diagonal elements  $\psi_1, \dots, \psi_p$  are called unique variances. Then, the covariance matrix of  $\mathbf{y}$  is expressed as

$$\Sigma = \mathbf{BRB}' + \Psi, \tag{1}$$

with  $\mathbf{R} = (r_{kl}) = E(\mathbf{z}\mathbf{z}')$  being an  $m \times m$  factor correlation matrix. The normality assumptions for  $\mathbf{z}$  and  $\mathbf{e}$  lead to the log-likelihood

$$\begin{aligned} l(\mathbf{B}, \Psi, \mathbf{R} | \mathbf{C}_{yy}) &= -\frac{n}{2} \log |\Sigma| - \frac{n}{2} \text{tr} \mathbf{C}_{yy} \Sigma^{-1} \\ &= -\frac{n}{2} \log |\mathbf{BRB}' + \Psi| - \frac{n}{2} \text{tr} \mathbf{C}_{yy} (\mathbf{BRB}' + \Psi)^{-1}, \end{aligned} \tag{2}$$

given a sample covariance matrix  $\mathbf{C}_{yy} = N^{-1} \mathbf{Y}'(\mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n') \mathbf{Y}$ , where  $\mathbf{Y}$  is the  $n \times p$  matrix with its rows the realizations of  $\mathbf{y}'$ ,  $N$  is  $n$  or  $n - 1$ ,  $\mathbf{I}_n$  denotes the  $n \times n$  identity matrix, and  $\mathbf{1}_n$  expresses the  $n \times 1$  vector of ones. Parameter matrices  $\mathbf{B}$ ,  $\Psi$ , and  $\mathbf{R}$  can be estimated by maximizing (2). Factor analysis is further classified into either exploratory factor analysis (EFA) or confirmatory factor analysis (CFA); in the latter, some elements of  $\mathbf{B}$  are constrained to be zero, while such constraints are not imposed in the former (e.g., Mulaik, 2010).

The author would like to thank the editor and anonymous associate editor and reviewers. This research was supported by grant (C)-23500347 from the Japan Society for the Promotion of Science.

Requests for reprints should be sent to Kohei Adachi, Graduate School of Human Sciences, Osaka University, 1-2 Yamadaoka, Suita, Osaka 565-0871, Japan. E-mail: [adachi@hus.osaka-u.ac.jp](mailto:adachi@hus.osaka-u.ac.jp)

In the prevailing procedures for maximum likelihood EFA and CFA, (2) is maximized by the gradient algorithms, such as the Newton–Raphson and quasi-Newton methods (Jennrich & Robinson, 1969; Jöreskog, 1967, 1969; Mulaik, 2010; Yanai & Ichikawa, 2007). A problem with those procedures is that they can give improper solutions that have at least one of two properties: (1) unique variances (i.e., the diagonal elements of  $\Psi$ ) include zero or a negative value; (2) factor correlation matrix  $\mathbf{R}$  includes an element whose absolute value exceeds one (e.g., Anderson & Gerbing, 1984; Gerbing & Anderson, 1987, Kano, 1998; Sato, 1987; Van Driel, 1978).

One purpose of this paper is to prove that such improper solutions are not given by Rubin and Thayer’s (1982) maximum likelihood EFA and CFA procedures using the EM algorithm (Dempster, Laird, & Rubin, 1977) if the sample covariance matrix  $\mathbf{C}_{yy}$  and the initial matrices of  $\Psi$  and  $\mathbf{R}$  are positive definite. We refer to the EFA and CFA with the EM algorithm as EM-EFA and EM-CFA, respectively, and generally name them EM factor analysis. After a brief introduction of its algorithms in Section 2, we provide some theorems to prove the above fact in Section 3. Furthermore, in Section 4, we present a modified EM algorithm in which the equations used for the proofs are substituted for Rubin and Thayer’s original formulas.

Another purpose is to numerically assess the behavior of EM factor analysis. In Section 5 we report real data examples, where EM factor analysis is shown to avoid the improper solutions produced by the factor analysis with the prevailing gradient algorithms, although the following observations are also reported: The EM algorithm with Rubin and Thayer’s original formulas can update factor correlation matrices into asymmetric ones so that the algorithm fails to converge in real computations where numbers are treated with finite precision by computers, but this problem is overcome by using the modified EM algorithm or an algorithm in which the original formulas are replaced by those guaranteeing the symmetry of factor correlation matrices. In Section 6, we further report simulation studies to answer the questions remaining to be elucidated with simulated data.

## 2. Rubin and Thayer’s (1982) EM Factor Analysis

In EM factor analysis, the maximization of log-likelihood (2) is attained by initializing  $\mathbf{B}$ ,  $\Psi$ , and  $\mathbf{R}$  to iterate the so-called *E* and *M* steps until convergence is reached. In the *E* step, it is required to find the expectation of the complete-data log-likelihood

$$l(\mathbf{B}, \Psi, \mathbf{R} | \mathbf{Y}, \mathbf{Z}) = -\frac{n}{2} \sum_{j=1}^p \log \psi_j - \frac{1}{2} \text{tr}(\mathbf{Y} - \mathbf{ZB}')\Psi^{-1}(\mathbf{Y} - \mathbf{ZB}')' - \frac{n}{2} \log |\mathbf{R}| - \frac{1}{2} \text{tr} \mathbf{ZR}^{-1}\mathbf{Z}', \quad (3)$$

supposing that an  $n \times m$  factor score matrix  $\mathbf{Z}$  were to be observed with its rows being realizations of  $\mathbf{z}'$ , and the expected log-likelihood is maximized over  $\mathbf{B}$ ,  $\Psi$ , and  $\mathbf{R}$  in the *M* step.

The computations actually required in the *E* step are simple: obtain

$$\delta = \Sigma^{-1}\mathbf{BR}, \quad (4)$$

$$\Delta = \mathbf{R} - \mathbf{RB}'\Sigma^{-1}\mathbf{BR}, \quad (5)$$

$$\mathbf{Q} = \delta'\mathbf{C}_{yy}\delta + \Delta \quad (6)$$

from the current  $\mathbf{B}$ ,  $\Psi$ , and  $\mathbf{R}$  (Rubin and Thayer, 1982, Equations (5), (7)). This *E* step is common between EFA and CFA, except that the factor correlation matrix  $\mathbf{R}$  is fixed at  $\mathbf{I}_m$  in EFA. The *M* step, however, is different for each of the two methods.

In the  $M$  step for EFA,  $\mathbf{B}$  and  $\Psi$  are updated into  $\mathbf{B}_{\text{new}}$  and  $\Psi_{\text{new}}$  with

$$\mathbf{B}_{\text{new}} = \mathbf{C}_{\text{yy}} \delta \mathbf{Q}^{-1}, \quad (7)$$

$$\Psi_{\text{new}} = \text{diag}\{\mathbf{C}_{\text{yy}} - \mathbf{C}_{\text{yy}} \delta \mathbf{Q}^{-1} \delta' \mathbf{C}_{\text{yy}}\} \quad (8)$$

(Rubin & Thayer, 1982, Equation (10)), where  $\text{diag}\{\cdot\}$  denotes the diagonal matrix whose diagonal elements are those of a parenthesized matrix.

In the  $M$  step for CFA,  $\mathbf{B} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p]'$  and  $\Psi = \text{diag}\{\psi_1, \dots, \psi_p\}$  are updated per variable  $j = 1, \dots, p$  (Rubin & Thayer, 1982, Equation (11)). Consider the  $j$ th variable with loading vector  $\boldsymbol{\beta}_j$  on  $m$  factors. Permute its elements so that  $\mathbf{P}_j \boldsymbol{\beta}_j = [\boldsymbol{\beta}'_{1j}, \boldsymbol{\beta}'_{0j}]'$  with  $\mathbf{P}_j$  a permutation matrix, where  $m_j \times 1$  vector  $\boldsymbol{\beta}_{1j}$  contains  $m_j$  unconstrained loadings to be estimated, and  $(m - m_j) \times 1 \boldsymbol{\beta}_{0j}$  consists of the loadings constrained to be zero. Similarly, partition  $\mathbf{Q} (m \times m)$  and  $\delta' \mathbf{C}_{\text{yy}} (m \times p)$  so that their  $m_j \times m_j$  block  $(\mathbf{Q})_{1j}$  and  $m_j \times 1$  block  $(\delta' \mathbf{C}_{\text{yy}})_{1j}$  correspond to the factors with unconstrained loadings for variable  $j$ . The vector  $\boldsymbol{\beta}_{1j}$  and  $\psi_j$  are updated into  $\boldsymbol{\beta}_{1j}^{(\text{new})}$  and  $\psi_j^{(\text{new})}$  with

$$\boldsymbol{\beta}_{1j}^{(\text{new})} = (\mathbf{Q})_{1j}^{-1} (\delta' \mathbf{C}_{\text{yy}})_{1j}, \quad (9)$$

$$\psi_j^{(\text{new})} = (\mathbf{C}_{\text{yy}})_j - (\delta' \mathbf{C}_{\text{yy}})_{1j}' (\mathbf{Q})_{1j}^{-1} (\delta' \mathbf{C}_{\text{yy}})_{1j}, \quad (10)$$

where  $(\mathbf{C}_{\text{yy}})_j$  is the  $j$ th diagonal element of  $\mathbf{C}_{\text{yy}}$ . Further, factor correlation matrix  $\mathbf{R}$  is updated into

$$\mathbf{R}_{\text{new}} = \text{diag}\{\mathbf{Q}\}^{-1/2} \mathbf{Q} \text{diag}\{\mathbf{Q}\}^{-1/2}, \quad (11)$$

with  $\text{diag}\{\mathbf{Q}\}^{-1/2} \text{diag}\{\mathbf{Q}\}^{-1/2} = \text{diag}\{\mathbf{Q}\}^{-1}$ .

In the log-likelihood (3) for complete data, the positive definiteness of  $\Psi$  and  $\mathbf{R}$  is supposed. However, this supposition does not guarantee that they are estimated as positive definite matrices. We thus prove that  $\Psi$  and  $\mathbf{R}$  can always be updated into positive definite  $\Psi_{\text{new}}$  and  $\mathbf{R}_{\text{new}}$ , which implies that no improper solution occurs when  $\mathbf{C}_{\text{yy}}$  and the initial  $\Psi$  and  $\mathbf{R}$  are positive definite.

### 3. Theorems

The positive definiteness of sample covariance matrix  $\mathbf{C}_{\text{yy}}$  is not required in the formulas of (4) to (11) and the log-likelihood (2). In the following lemma and theorems, we thus consider not only positive definite  $\mathbf{C}_{\text{yy}}$  but also the cases where  $\mathbf{C}_{\text{yy}}$  is not positive but rather nonnegative definite.

**Lemma 1.** *If  $\Psi$  and  $\mathbf{R}$  are positive definite and  $\mathbf{C}_{\text{yy}}$  is nonnegative definite, then  $\Delta$  and  $\mathbf{Q}$  are positive definite.*

*Proof:* The positive definiteness of  $\Psi$  and  $\mathbf{R}$  guarantees that the inverse of (1) exists. Using this in (5) we have  $\Delta = \mathbf{R} - (\mathbf{R}\mathbf{B}')(\mathbf{B}\mathbf{R}\mathbf{B}' + \Psi)^{-1}(\mathbf{B}\mathbf{R})$ , which can be rewritten as

$$\begin{aligned} \Delta &= \mathbf{R}^{1/2'} \{ \mathbf{I}_m - (\mathbf{R}^{1/2} \mathbf{B}') [ (\mathbf{B}\mathbf{R}^{1/2'}) (\mathbf{R}^{1/2} \mathbf{B}') + \Psi ]^{-1} (\mathbf{B}\mathbf{R}^{1/2'}) \} \mathbf{R}^{1/2} \\ &= \mathbf{R}^{1/2'} [ \mathbf{I}_m + (\mathbf{R}^{1/2} \mathbf{B}') \Psi^{-1} (\mathbf{B}\mathbf{R}^{1/2'}) ]^{-1} \mathbf{R}^{1/2}, \end{aligned} \quad (12)$$

where  $\mathbf{R}^{1/2}$  is the  $m \times m$  positive definite matrix satisfying  $\mathbf{R}^{1/2'} \mathbf{R}^{1/2} = \mathbf{R}$ ; and we have used the SMW formula described in Appendix A. The nonnegative definiteness of  $(\mathbf{R}^{1/2} \mathbf{B}') \Psi^{-1} (\mathbf{B}\mathbf{R}^{1/2'})$  thus leads to the positive definiteness of  $\Delta$ . This fact and the nonnegative definiteness of  $\delta' \mathbf{C}_{\text{yy}} \delta$  imply the positive definiteness of (6).  $\square$

**Theorem 1.** *If  $\Psi$  is positive definite and  $\mathbf{C}_{yy}$  is nonnegative definite, then  $\Psi_{\text{new}}$  is nonnegative definite in EM-EFA. Moreover, if both  $\Psi$  and  $\mathbf{C}_{yy}$  are positive definite, then  $\Psi_{\text{new}}$  is positive definite in EM-EFA.*

*Proof:* Lemma 1 implies that the inverse of (6) exists. Substituting this inverse into (8), it can be rewritten as

$$\begin{aligned}\Psi_{\text{new}} &= \text{diag}\{\mathbf{C}_{yy} - \mathbf{C}_{yy}\delta(\delta'\mathbf{C}_{yy}\delta + \Delta)^{-1}\delta'\mathbf{C}_{yy}\} \\ &= \text{diag}\{\mathbf{C}_{yy}^{1/2'}[\mathbf{I}_p - \mathbf{C}_{yy}^{1/2}\delta(\delta'\mathbf{C}_{yy}^{1/2'}\mathbf{C}_{yy}^{1/2}\delta + \Delta)^{-1}\delta'\mathbf{C}_{yy}^{1/2'}]\mathbf{C}_{yy}^{1/2}\},\end{aligned}\quad (13)$$

with  $\mathbf{C}_{yy}^{1/2}$  the  $p \times p$  nonnegative definite matrix satisfying  $\mathbf{C}_{yy}^{1/2'}\mathbf{C}_{yy}^{1/2} = \mathbf{C}_{yy}$ . We can further use the SMW formula (Appendix A) to rewrite (13) as

$$\Psi_{\text{new}} = \text{diag}\{\mathbf{C}_{yy}^{1/2'}(\mathbf{I}_p + \mathbf{C}_{yy}^{1/2}\delta\Delta^{-1}\delta'\mathbf{C}_{yy}^{1/2'})^{-1}\mathbf{C}_{yy}^{1/2}\},\quad (14)$$

which is nonnegative definite. Moreover, if  $\mathbf{C}_{yy}$  is positive definite, then (14) is positive definite.  $\square$

Theorem 1 shows the following fact for EM-EFA: If unique variances are initialized at positive values, the resulting solution is always proper for a positive definite sample covariance matrix; but unique variances can be estimated at zero, though never at negative values, for a sample covariance matrix which is not positive but rather nonnegative definite.

**Theorem 2.** *If  $\mathbf{R}$  and  $\Psi$  are positive definite and  $\mathbf{C}_{yy}$  is nonnegative definite, then  $\Psi_{\text{new}}$  is nonnegative definite in EM-CFA. Moreover, if  $\mathbf{R}$ ,  $\Psi$ , and  $\mathbf{C}_{yy}$  are positive definite, then  $\Psi_{\text{new}}$  is positive definite in EM-CFA.*

*Proof:* The scalar  $(\mathbf{C}_{yy})_j$ ,  $m_j \times 1$  vector  $(\delta'\mathbf{C}_{yy})_{1j}$ , and  $m_j \times m_j$  matrix  $(\mathbf{Q})_{1j}$  in (10) can be rewritten as  $\mathbf{e}_j'\mathbf{C}_{yy}\mathbf{e}_j$ ,  $\mathbf{H}_j'\delta'\mathbf{C}_{yy}\mathbf{e}_j$ , and  $\mathbf{H}_j'\mathbf{Q}\mathbf{H}_j$ , respectively. Here,  $\mathbf{e}_j$  denotes the  $j$ th column of  $\mathbf{I}_p$ , and  $\mathbf{H}_j$  is the  $m \times m_j$  binary matrix satisfying  $\beta_j = \mathbf{H}_j\beta_{1j}$ ; for example, if  $\beta_j = \begin{bmatrix} b_{j1} \\ 0 \\ b_{j3} \end{bmatrix}$  and  $\beta_{1j} = \begin{bmatrix} b_{j1} \\ b_{j3} \end{bmatrix}$ , then  $\mathbf{H}_j = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$ ; and if  $\beta_j = \begin{bmatrix} 0 \\ b_{j2} \\ 0 \end{bmatrix}$  and  $\beta_{1j} = [b_{j2}]$  (a scalar), then  $\mathbf{H}_j = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ , with  $b_{jk}$  and 0 standing for unconstrained and constrained loadings, respectively. Therefore, (10) is rewritten as

$$\psi_j^{(\text{new})} = \mathbf{e}_j'\mathbf{C}_{yy}\mathbf{e}_j - \mathbf{e}_j'\mathbf{C}_{yy}\delta\mathbf{H}_j(\mathbf{H}_j'\mathbf{Q}\mathbf{H}_j)^{-1}\mathbf{H}_j'\delta'\mathbf{C}_{yy}\mathbf{e}_j.\quad (15)$$

Here, it should be noted that the positive definiteness of  $\mathbf{R}$  and  $\Psi$  guarantees that the inverse of  $\mathbf{H}_j'\mathbf{Q}\mathbf{H}_j$  exists, which followed from the fact that the positive definiteness of  $\mathbf{Q}$  (Lemma 1) and  $\mathbf{H}_j$  being of full column rank by definition imply the positive definiteness of  $\mathbf{H}_j'\mathbf{Q}\mathbf{H}_j$  (e.g., Lütkepohl, 1996, p. 152).

Substituting (6) into (15), we have

$$\begin{aligned}\psi_j^{(\text{new})} &= \mathbf{e}_j'\mathbf{C}_{yy}\mathbf{e}_j - \mathbf{e}_j'\mathbf{C}_{yy}\delta\mathbf{H}_j\{\mathbf{H}_j'(\delta'\mathbf{C}_{yy}\delta + \Delta)\mathbf{H}_j\}^{-1}\mathbf{H}_j'\delta'\mathbf{C}_{yy}\mathbf{e}_j \\ &= \mathbf{e}_j'\mathbf{C}_{yy}^{1/2'}\{\mathbf{I}_p - \mathbf{C}_{yy}^{1/2}\delta\mathbf{H}_j(\mathbf{H}_j'\delta'\mathbf{C}_{yy}^{1/2'}\mathbf{C}_{yy}^{1/2}\delta\mathbf{H}_j + \mathbf{H}_j'\Delta\mathbf{H}_j)^{-1}\mathbf{H}_j'\delta'\mathbf{C}_{yy}^{1/2'}\}\mathbf{C}_{yy}^{1/2}\mathbf{e}_j.\end{aligned}\quad (16)$$

The positive definiteness of  $\Delta$  (Lemma 1) and  $\mathbf{H}_j$  being of full column rank imply that  $\mathbf{H}_j'\Delta\mathbf{H}_j$  is positive definite. Using this property and the SMW formula (Appendix A), (16) is further rewritten as

$$\psi_j^{(\text{new})} = \mathbf{e}_j'\mathbf{C}_{yy}^{1/2'}\{\mathbf{I}_p + \mathbf{C}_{yy}^{1/2}\delta\mathbf{H}_j(\mathbf{H}_j'\Delta\mathbf{H}_j)^{-1}\mathbf{H}_j'\delta'\mathbf{C}_{yy}^{1/2'}\}^{-1}\mathbf{C}_{yy}^{1/2}\mathbf{e}_j.\quad (17)$$

The nonnegative definiteness of  $\mathbf{C}_{yy}^{1/2} \delta \mathbf{H}_j (\mathbf{H}'_j \Delta \mathbf{H}_j)^{-1} \mathbf{H}'_j \delta' \mathbf{C}_{yy}^{1/2}$  implies that (17) is nonnegative. Moreover, if  $\mathbf{C}_{yy}$  is positive definite, then (17) is positive.  $\square$

**Theorem 3.** *If  $\mathbf{R}$  and  $\Psi$  are positive definite and  $\mathbf{C}_{yy}$  is nonnegative definite,  $\mathbf{R}_{\text{new}}$  is a positive definite correlation matrix.*

*Proof:* Lemma 1 shows the positive definiteness of  $\mathbf{Q}$  and, thus, that of  $\text{diag}\{\mathbf{Q}\}$ . Further,  $\mathbf{Q}$  is symmetric, as found in (6). These facts imply that  $\mathbf{R}_{\text{new}}$  in (11) is a symmetric positive definite matrix, whose diagonal elements are unity and whose off-diagonal elements take the values within the range  $[-1, 1]$ .  $\square$

Theorems 2 and 3 show the following facts for EM-CFA: If unique variances and a factor correlation matrix are initialized at positive values and a positive definite matrix, respectively, then the resulting solution is always proper for a positive definite sample covariance matrix; but unique variances can be estimated at zero, though never at negative values, for a sample covariance matrix which is not positive but rather nonnegative definite. Further, if unique variances and a factor correlation matrix are initialized as above, factor correlations are properly estimated, even for a sample covariance matrix that is not positive but rather nonnegative definite.

Theorems 1, 2, and 3 imply that EM factor analysis is a constrained estimation procedure in which its solutions are restricted to proper ones for the sample covariance and initial parameter matrices satisfying a certain condition. On the other hand, factor analysis with gradient algorithms is an unconstrained estimation procedure, although a constrained variant exists in which the unique variance  $\psi_j$  is restricted to nonnegative values by treating the square root of  $\psi_j$  as the parameter to be estimated (e.g., SPSS Inc., 1997). Here, the constraint of this version is noted to be weaker than the constraint in EM factor analysis which restricts  $\psi_j$  to be positive, not only nonnegative, for the  $\mathbf{C}_{yy}$  and initial  $\mathbf{R}$  and  $\Psi$  which are positive definite.

Besides the properness of unique variances and factor correlations, we can add the following theorem on the loading matrices in EM-EFA:

**Theorem 4.** *If  $\Psi$  and  $\mathbf{C}_{yy}$  are positive definite and  $\text{rank}(\mathbf{B}) = m$ , then  $\text{rank}(\mathbf{B}_{\text{new}}) = m$  in EM-EFA, where  $\text{rank}(\mathbf{B})$  denotes the rank of  $\mathbf{B}$ .*

*Proof:* Substituting (4) into (7) while setting  $\mathbf{R} = \mathbf{I}_p$ , we have  $\mathbf{B}_{\text{new}} = \mathbf{C}_{yy} \Sigma^{-1} \mathbf{B} \mathbf{Q}^{-1}$ . The positive definiteness of  $\Psi$  and Lemma 1 imply  $\text{rank}(\Sigma^{-1}) = p$  and  $\text{rank}(\mathbf{Q}^{-1}) = m$ . Furthermore, the positive definiteness of  $\mathbf{C}_{yy}$  implies  $\text{rank}(\mathbf{C}_{yy}) = p$ . Here, we can use the fact that  $\text{rank}(\mathbf{A}\mathbf{W}) = \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{W}\mathbf{A}')$  holds for any matrix  $\mathbf{A}$  if  $\mathbf{W}$  is square of full rank (e.g., Magnus & Neudecker, 1991, p. 8). That is,  $\mathbf{C}_{yy} \Sigma^{-1}$  and  $\mathbf{Q}^{-1}$  are square of full rank, which leads to  $\text{rank}(\mathbf{C}_{yy} \Sigma^{-1} \mathbf{B}) = \text{rank}(\mathbf{B}) = \text{rank}(\mathbf{C}_{yy} \Sigma^{-1} \mathbf{B} \mathbf{Q}^{-1}) = \text{rank}(\mathbf{B}_{\text{new}})$ .  $\square$

The above theorem shows that if the initial unique variances and loading matrix are positive and of full column rank, respectively, EM-EFA always gives a loading matrix of full column rank for a positive definite sample covariance matrix.

#### 4. Modified Algorithm

In order to prove the lemma and theorems in the last section, we rewrote Rubin and Thayer's (1982) original formulas (5), (8) and (10) into Equations (12), (14), and (17), respectively, which explicitly guarantee EM factor analysis to give proper solutions when  $\mathbf{C}_{yy}$  and the initial  $\mathbf{R}$  and  $\Psi$  are proper. We can thus consider the EM algorithm in which (12), (14), and (17) are substituted for (5), (8) and (10), respectively. This modified EM algorithm follows the steps below:

TABLE 1.  
Unique variances obtained by EFA with gradient and EM algorithms for Maxwell's (1961) data.

Variable	1	2	3	4	5	6	7	8	9	10
Gradient	0.40	0.63	0.30	0.64	0.32	0.80	0.28	-73.60	0.72	0.61
EM	0.38	0.62	0.30	0.64	0.35	0.78	0.29	0.03	0.69	0.60

- Step 1. Initialize  $\mathbf{B}$ ,  $\mathbf{R}$ , and  $\Psi$  so that  $\mathbf{R}$  and  $\Psi$  are positive definite and  $\mathbf{B}$  is of full column rank.
- Step 2. Obtain  $\delta$  by (4) and  $\Delta$  by (12) and  $\mathbf{Q}$  by (6).
- Step 3. Update  $\Psi$  with (14) for EFA and (17) for CFA.
- Step 4. Update  $\mathbf{B}$  with (7) for EFA and (9) for CFA.
- Step 5. Update  $\mathbf{R}$  with (11) for CFA
- Step 6. Finish if convergence is reached; otherwise, go back to Step 2.

Here,  $\mathbf{R}$  is fixed at  $\mathbf{I}_m$  for EFA.

The modified EM algorithm and the original one in Section 2 are mathematically equivalent, thus the former algorithm also decreases log-likelihood (2) monotonically with the parameter updates according to the principle of the EM algorithm (Dempster et al., 1977). However, the two algorithms are not necessarily guaranteed to behave equivalently in real computations, where numbers are treated with finite precision by computers. In such conditions, the modified EM algorithm, whose formulas were used for proving theorems in Section 3, is thought to follow those theorems more faithfully than the original one whose formulas do not explicitly incorporate the facts given by the theorems. Their different behaviors are reported in the second example in the next section.

### 5. Real Data Examples

We compare the original EM, the modified EM, and gradient algorithms for EFA and CFA, using Maxwell's (1961, p. 55)  $10 \times 10$  correlation matrix. All computations in this paper are performed with the computer programs detailed in Appendix B; gradient algorithms are run using the popular program package SAS (SAS Institute Inc., 2009), while EM algorithms are performed by our own FORTRAN programs, starting with the initial  $\Psi$  and  $\mathbf{R}$  that are positive definite and the initial  $\mathbf{B}$  of full column rank (though  $\mathbf{R} = \mathbf{I}_m$  in EFA).

First, we compared the three algorithms in EFA with  $m = 4$ . As found in the first row of results presented in Table 1, the unique variance of Variable 8 estimated by the gradient algorithm was far smaller than zero. On the other hand, the original and modified EM algorithms gave identical four-factor solutions with the proper unique variances, shown in the second row of results in Table 1, and a loading matrix of full column rank.

Next, we performed CFA with  $m = 3$  and loadings  $b_{41}, \dots, b_{10,1}, b_{12}, b_{22}, b_{62}, \dots, b_{10,2}$ , and  $b_{13}, \dots, b_{53}$  constrained to be zero. As a result, the gradient algorithm produced an improper solution with a factor correlation  $r_{21} > 1$ , as shown in Table 2. On the other hand, the modified EM algorithm gave a proper one with the correlations shown in Table 2, positive unique variances, and a loading matrix of full column rank. However, the original EM algorithm for CFA failed to converge with log-likelihood (2) rather decreasing at the 74th iteration. What happened in this algorithm is detailed in the next paragraph.

We found that the failure of the original EM algorithm originated when  $\mathbf{RB}'\Sigma^{-1}\mathbf{BR}$  in (5) became asymmetric, which led to the asymmetry of factor correlation matrix  $\mathbf{R}$  through updates (5), (6), and (11). Although  $\mathbf{RB}'\Sigma^{-1}\mathbf{BR} = (\mathbf{BR})'\Sigma^{-1}\mathbf{BR}$  being asymmetric seems to be strange, it did occur in our computer program, where  $(\mathbf{BR})'$  was post-multiplied by  $\Sigma^{-1}\mathbf{BR}$ ; the

TABLE 2.

Factor correlation matrices obtained with CFA with gradient and modified EM algorithms for Maxwell's (1961) data.

Procedure	Gradient			Modified EM		
	1	2	3	1	2	3
1	1			1		
2	1.08	1		0.98	1	
3	0.65	0.48	1	0.62	0.46	1

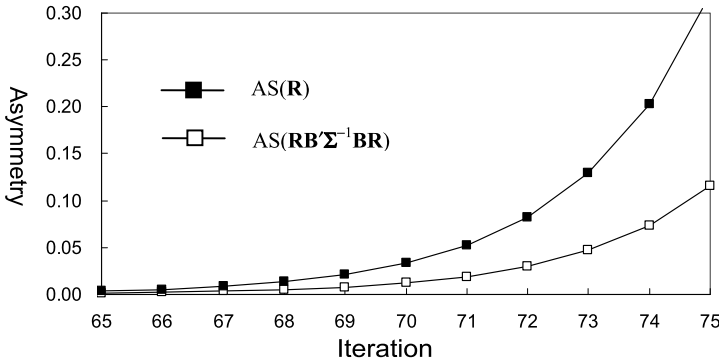


FIGURE 1.

Averaged asymmetry of  $\mathbf{RB}'\Sigma^{-1}\mathbf{BR}$  and  $\mathbf{R}$  as a function of the number of iterations in the original EM algorithm.

resulting  $(\mathbf{BR})'\Sigma^{-1}\mathbf{BR}$  was not guaranteed to be symmetric in real computation, since  $\mathbf{BR}$  and  $\Sigma^{-1}\mathbf{BR}$  are different matrices. Indeed, averaged asymmetry  $AS(\mathbf{R}) = 2\sum_{k<l}|r_{kl} - r_{lk}|/\{m(m - 1)\}$  and  $AS(\mathbf{RB}'\Sigma^{-1}\mathbf{BR})$  were 0.203 and 0.074, respectively, at the 74th iteration, which is due to limited numerical precision as described in Appendix C. Figure 1 shows the changes in averaged asymmetry. It is considered that the increase of asymmetry with the iteration found in Figure 1 was caused by  $\mathbf{RB}'\Sigma^{-1}\mathbf{BR}$  being a function of  $\mathbf{R}$ ; the latter asymmetry made the former more asymmetric, which further increased the latter asymmetry through (5), (6) and (11). We can thus consider the following scenario leading to the decrease in (2): at the 74th iteration,  $\mathbf{R}$  became substantially asymmetric to deviate from a correlation matrix, which led to the deviation of  $l(\mathbf{B}, \Psi, \mathbf{R}|\mathbf{C}_{yy})$  from the proper log-likelihood that guarantees its monotone increase with updates in the EM algorithm.

The above considerations are supported by the following fact: the original EM algorithm, in which (6) was replaced by

$$\mathbf{Q} = \frac{1}{2} \{ (\delta' \mathbf{C}_{yy} \delta + \Delta) + (\delta' \mathbf{C}_{yy} \delta + \Delta)' \} \tag{6'}$$

to guarantee the symmetry of  $\mathbf{Q}$  and  $\mathbf{R}$ , yielded the same proper solution that the modified EM algorithm gave for the data set. We refer to the above algorithm with (6') as a symmetric original EM algorithm, while continuing to simply call the one with (6) the original EM algorithm.

### 6. Simulation Studies

In this section, after describing the purposes and outline of our simulation studies, we report on the studies for EFA and CFA, which are followed by a summary of the major results.

TABLE 3.  
Setting of true parameters or matrix sizes specifying conditions in simulation studies.

Condition	EFA	CFA
P ( <i>proper</i> )	$\psi_j = 0.5$	$\psi_j = 0.5, r_{jk} = 0.4$ for $j \neq k$
L ( <i>likely improper</i> )	[L <sub>E1</sub> ] $\psi_1 = 0.02$ [L <sub>E2</sub> ] $b_{j1} = 0$ for $j \geq 3$ [L <sub>E3</sub> ] $\hat{m} = m + 1$	[L <sub>C1</sub> ] $\psi_1 = 0.02$ [L <sub>C2</sub> ] $r_{21} = 0.98$ [L <sub>C3</sub> ] $\hat{m} = m + 1$
T ( <i>true improper</i> )	$\psi_1 = -0.05$	[T] $\psi_1 = -0.05$ ; [T'] $r_{21} = 1.05$
S ( <i>singular</i> )	$n < p$	$n < p$

### 6.1. Purposes and Outline

The observation in the last section raises the following questions: [Q1] How often does the original EM algorithm for CFA fail to converge, and [Q2] can the modified EM algorithm behave differently than the symmetric original EM algorithm? Also of interest is [Q3] how EM factor analysis behaves when sample covariance matrices are not positive but nonnegative definite, since such matrices were considered in Section 3. Furthermore, we should study [Q4] how well the true parameters are recovered, and [Q5] how many iterations are needed until convergence is reached, since the properness of solutions does not necessarily imply goodness of recovery and it is known that convergence is generally slow in EM algorithms (e.g., Minami, 2004).

In order to answer the above questions and demonstrate the results presented in Section 3, we compare the original EM, symmetric original EM, modified EM, and gradient algorithms in simulation studies, where sample correlations are synthesized and analyzed under the *proper* (P), *likely improper* (L), *true improper* (T), and *singular correlation* (S) conditions outlined in Table 3. In Condition P, model parameters are set at the values unlikely to lead to improper solutions. A change of one procedure in Condition P defines another condition. The procedures in which gradient algorithms are likely to produce improper solutions are used in Condition L, which are divided into three sub-conditions. In Condition T, a true parameter is set at an improper value; EM factor analysis should *incorrectly* produce proper solutions. Positive definite sample correlation matrices are synthesized in P, L, and T, but in Condition S, data matrices with  $n < p$  (more variables than observations) give singular correlation matrices which are not positive but rather nonnegative definite. In Condition S, only EM factor analysis is performed, as singular matrices cannot be analyzed by the gradient algorithms that we use in this paper. The studies for EFA and CFA are reported in the next subsections.

### 6.2. Exploratory Factor Analysis

In Condition P, we set the true  $\mathbf{B}(12 \times 3)$  and  $\Psi$  as in Panel (A) of Table 4 with  $\mathbf{R} = \mathbf{I}_3$  and synthesize  $180 \times 12$  data matrices  $\mathbf{Y}$ , whose rows were sampled from the multivariate normal distribution with its mean vector  $\mathbf{0}_{12}$  and covariance matrix (1). The EFA algorithms were applied to the correlation matrix  $\text{diag}(\mathbf{C}_{yy})^{-1/2} \mathbf{C}_{yy} \text{diag}(\mathbf{C}_{yy})^{-1/2}$  with the estimated number of factors  $\hat{m}$  set at  $m = 3$ . The procedures in other conditions were the same as those in Condition P, except for the one described next. The true parameters were set as in Panel (B) of Table 4, with  $\psi_1 = -0.05$ , in Condition T, and the number of rows in  $\mathbf{Y}$  was set to 11 ( $< p = 12$ ) in S. Condition L is divided into L<sub>E1</sub>, L<sub>E2</sub>, and L<sub>E3</sub>. The true parameters were set as in Panel (B) of Table 4 with  $\psi_1 = 0.02$  in Condition L<sub>E1</sub>, they were set as in Panel (C) of Table 4 in L<sub>E2</sub>, and EFA was performed with  $\hat{m} = 4$  ( $> \text{true } m = 3$ ) in L<sub>E3</sub>. It is known that the EFA model is not uniquely identified in Condition L<sub>E2</sub> (Anderson & Rubin, 1956) and gradient algorithms tend to produce improper solutions in L<sub>E1</sub>, L<sub>E2</sub>, and L<sub>E3</sub> (Kano, 1998; Sato, 1987; Van Driel, 1978).



TABLE 4.  
True loadings and unique variances used in simulation studies.

(A) Conditions P, L <sub>E3</sub> , L <sub>C2</sub> , L <sub>C3</sub> , S				(B) Conditions L <sub>E1</sub> , L <sub>C1</sub> , T				(C) Condition L <sub>E2</sub>			
<b>B</b>		$\Psi \mathbf{1}_p$		<b>B</b>		$\Psi \mathbf{1}_p$		<b>B</b>		$\Psi \mathbf{1}_p$	
0.5 <sup>1/2</sup>	0	0	0.5	(1 - $\psi_1$ ) <sup>1/2</sup>	0	0	$\psi_1$	0.5 <sup>1/2</sup>	0	0	0.5
0.5 <sup>1/2</sup>	0	0	0.5	0.5 <sup>1/2</sup>	0	0	0.5	0.5 <sup>1/2</sup>	0	0	0.5
0.5 <sup>1/2</sup>	0	0	0.5	0.5 <sup>1/2</sup>	0	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5
0.5 <sup>1/2</sup>	0	0	0.5	0.5 <sup>1/2</sup>	0	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5
0	0.5 <sup>1/2</sup>	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5
0	0.5 <sup>1/2</sup>	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5
0	0.5 <sup>1/2</sup>	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5
0	0.5 <sup>1/2</sup>	0	0.5	0	0.5 <sup>1/2</sup>	0	0.5	0	0	0.5 <sup>1/2</sup>	0.5
0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5
0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5
0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5
0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5	0	0	0.5 <sup>1/2</sup>	0.5

TABLE 5.

Percentages of improper solutions, averaged MAE (mean absolute error), averaged CN (condition number), and averaged number of iterations for EFA with gradient and original EM (O-EM) algorithms.

Condition			P	L <sub>E1</sub>	L <sub>E2</sub>	L <sub>E3</sub>	T	S
(A)	Improper Solutions	Gradient	0	<b>45</b>	<b>65</b>	<b>40</b>	<b>80</b>	
		O-EM	0	0	0	0	0	0
(B)	MAE( $\hat{\mathbf{B}}$ )	Gradient	0.045	0.042	<b>0.171</b>	0.033	0.040	
		O-EM	0.045	0.042	0.054	<b>0.034</b>	0.040	0.198
(C)	MAE( $\hat{\Psi}$ )	Gradient	0.058	<b>0.049</b>	<b>4.457</b>	<b>2.869</b>	0.048	
		O-EM	<b>0.059</b>	0.047	0.091	0.085	<b>0.052</b>	0.258
(D)	CN( $\hat{\mathbf{B}}$ )	Gradient	1.160	1.226	<b>3.787</b>	<b>4.661</b>	1.209	
		O-EM	1.160	1.226	1.530	2.712	<b>1.213</b>	1.512
(D)	Iteration	Gradient	6.0	6.3	<b>13510.8</b>	<b>12010.8</b>	6.4	
		O-EM	<b>20.7</b>	<b>416.9</b>	1727.7	1444.9	<b>409.2</b>	479.9

In each of Conditions P, L<sub>E1</sub>, L<sub>E2</sub>, L<sub>E3</sub>, S, and T, 20 data sets (correlation matrices) were synthesized and analyzed. Each estimated loading matrix was rotated so that it was optimally matched to the corresponding true **B**, as described in Appendix D. Let  $\hat{\mathbf{B}} = (\hat{b}_{jk})$  and  $\hat{\psi}_j$  denote the resulting loading matrix and unique variance, respectively. In Appendix D, it is also described how **B** was redefined so that its size equals that of  $\hat{\mathbf{B}}$  in Condition L<sub>E3</sub>.

Every algorithm produced  $\hat{\mathbf{B}}$  of full column rank for all data sets. The other results are summarized in Table 5, while those for the symmetric original and modified EM algorithms are not shown, since these two always gave the same solution as the original EM algorithm. In the table, the bold font in gray cells indicates the larger values between the two algorithms. Panel (A) in Table 5 shows the percentages of the data sets for which the improper solutions including  $\hat{\psi}_j \leq 0$  arose. There, we find that the EM algorithm always gave a positive  $\hat{\psi}_j$ , while the gradient algorithm produced improper solutions in Conditions L and T.

The recovery of **B** and that of  $\Psi$  were assessed by mean absolute errors MAE( $\hat{\mathbf{B}}$ ) =  $(p\hat{m})^{-1} \sum_{j=1}^p \sum_{k=1}^{\hat{m}} |\hat{b}_{jk} - b_{jk}|$  and MAE( $\hat{\Psi}$ ) =  $p^{-1} \sum_{j=1}^p |\hat{\psi}_j - \psi_j|$ , whose averages over data sets are presented in Table 5, Panels (B) and (C). We further obtained the condition number of  $\hat{\mathbf{B}}$ , CN( $\hat{\mathbf{B}}$ ) (which is the ratio of the largest singular value to the smallest one of  $\hat{\mathbf{B}}$ ), to assess the nearness of  $\hat{\mathbf{B}}$  to rank deficiency. The average of CN( $\hat{\mathbf{B}}$ ) is shown in Panel (D) of Table 5,

TABLE 6.

True parameters, the solutions of EFA with gradient and EM algorithms, and their MAE and CN values for a data set in Condition  $L_{E3}$ .

(A) True parameters				(B) Gradient algorithm				(C) EM algorithm					
$\mathbf{B}$		$\boldsymbol{\psi}_{1_p}$		$\hat{\mathbf{B}}$		$\hat{\boldsymbol{\psi}}_{1_p}$		$\hat{\mathbf{B}}$		$\hat{\boldsymbol{\psi}}_{1_p}$			
0.71	0	0	0.5	0.61	-0.17	0.08	0.00	0.59	0.60	-0.16	0.06	-0.15	0.59
0.71	0	0	0.5	0.83	-0.04	-0.05	0.00	0.31	0.77	-0.03	-0.03	-0.10	0.40
0.71	0	0	0.5	0.70	-0.09	-0.01	0.01	0.50	0.72	-0.04	-0.04	-0.30	0.38
0.71	0	0	0.5	0.61	0.04	-0.01	0.00	0.63	0.77	-0.04	-0.02	0.60	0.04
0	0.71	0	0.5	-0.10	0.76	-0.01	0.01	0.41	-0.10	0.76	-0.05	0.10	0.39
0	0.71	0	0.5	0.05	0.80	-0.07	0.00	0.35	0.04	0.80	-0.04	0.07	0.36
0	0.71	0	0.5	-0.10	0.65	0.11	0.00	0.56	-0.10	0.64	0.11	0.10	0.56
0	0.71	0	0.5	-0.10	0.62	0.01	0.00	0.60	-0.09	0.60	0.01	0.18	0.59
0	0	0.71	0.5	-0.12	-0.01	0.81	0.00	0.34	-0.11	-0.02	0.73	0.04	0.45
0	0	0.71	0.5	0.06	0.03	0.68	-0.01	0.53	0.06	0.04	0.73	-0.04	0.46
0	0	0.71	0.5	0.00	0.00	0.71	-17.68	-312.25	-0.07	-0.02	0.80	-0.02	0.35
0	0	0.71	0.5	0.07	0.02	0.68	-0.01	0.53	0.08	0.03	0.70	-0.04	0.50
(Note $0.71 \cong 0.5^{1/2}$ )				MAE( $\hat{\mathbf{B}}$ ) = 0.044; MAE( $\hat{\boldsymbol{\psi}}$ ) = 26.15				MAE( $\hat{\mathbf{B}}$ ) = 0.043; MAE( $\hat{\boldsymbol{\psi}}$ ) = 0.116					
CN( $\mathbf{B}$ ) = 1				CN( $\hat{\mathbf{B}}$ ) = 13.995				CN( $\hat{\mathbf{B}}$ ) = 2.272					

with  $CN(\mathbf{B})$  (the condition number of the true  $\mathbf{B}$ ) being 1.114 in Condition  $L_{C2}$ , 1.581 in  $L_{C3}$ , 1.129 in T, and 1.0 in the others (P,  $L_{E3}$ , T, and S). In Table 5, we find that the gradient algorithm recovered parameters more poorly and gave  $\hat{\mathbf{B}}$  nearer to rank deficiency, than the EM algorithm in  $L_{E2}$  and  $L_{E3}$ , while the EM algorithm recovered parameters more poorly in Condition T, where that algorithm incorrectly gave proper solutions. The averages of  $MAE(\hat{\boldsymbol{\psi}})$  in  $L_{E2}$  and  $L_{E3}$  for the gradient algorithm were found to be very large, which was due to the fact that  $\psi_j$  was estimated to be far smaller than zero for some data sets. Such a solution of the gradient algorithm (with  $\hat{\psi}_{11} = -312.25$ ) in  $L_{E3}$  is presented in Table 6, where the  $CN(\hat{\mathbf{B}})$  shows that  $\hat{\mathbf{B}}$  for the gradient algorithm was nearer to rank deficiency than  $\hat{\mathbf{B}}$  for the EM algorithm.

Table 5, Panel (D), shows that the gradient algorithm needed more iterations for convergence in  $L_{E2}$  and  $L_{E3}$ , though the EM algorithm converged more slowly in the other conditions. Finally, let us note the performances of the EM algorithm in Condition S (Table 5). Though Theorem 1 shows that an improper solution with  $\hat{\psi}_j = 0$  could occur in S, such a solution was not found. However, averaged MAE values were large, which demonstrates the difficulty of estimating parameters from nonsingular sample correlation matrices.

### 6.3. Confirmatory Factor Analysis

The true parameters used in Conditions P, S, T, and  $L_{C1}$  are the same as those in P, S, T, and  $L_{E1}$  for EFA (Section 6.2), respectively, except that all true inter-factor correlations  $r_{kl}$  ( $k \neq l$ ) were set to 0.4 in P, S, T, and,  $L_{C1}$  for CFA. In those conditions, CFA was performed subject to the loadings whose true values were zero being constrained to zero. A change of one procedure in Condition P defines each of the remaining conditions. In  $T'$ , which is another *true improper* condition beside T, the true  $r_{21}$  and  $r_{12}$  were set to 1.05 (not 0.4), while they were set to 0.98 in  $L_{C2}$ . In Condition  $L_{C3}$ , CFA was performed with  $\hat{m} = 4$  ( $> m = 3$ ) subject to the constraint found in Table 8(B) and (C). Twenty data sets were synthesized in each condition. We use  $\hat{\mathbf{B}} = (\hat{b}_{jk})$ ,  $\hat{\boldsymbol{\psi}}_j$ , and  $\hat{\mathbf{R}} = (\hat{r}_{kl})$  for the solutions of  $\mathbf{B}$ ,  $\boldsymbol{\psi}_j$ , and  $\mathbf{R} = (r_{jk})$ .

The original EM algorithm always failed to converge with a decrease in (2) in all conditions except P (though such a failure did not occur in Condition P). The failures are considered to be due to  $\mathbf{R}$  becoming asymmetric, as observed in Section 5, which can be ascertained from the fact

TABLE 7.

Percentages of improper solutions, the averaged MAE (mean absolute error), averaged CN (condition number), and averaged number of iterations for CFA with gradient and modified EM (M-EM) algorithms.

Condition			P	L <sub>C1</sub>	L <sub>C2</sub>	L <sub>C3</sub>	T	T'	S
(A)	Improper $\hat{\psi}_j$	Gradient	0	<b>25</b>	0	0	<b>80</b>	0	
		M-EM	0	0	0	0	0	0	0
(B)	Improper $\hat{r}_{kl}$	Gradient	0	0	<b>35</b>	<b>70</b>	0	<b>100</b>	
		M-EM	0	0	0	0	0	0	0
(C)	MAE( $\hat{\mathbf{B}}$ )	Gradient	0.042	<b>0.036</b>	0.037	0.045	0.032	<b>0.034</b>	
		M-EM	0.042	0.034	0.037	<b>0.046</b>	<b>0.033</b>	0.033	0.196
(D)	MAE( $\hat{\Psi}$ )	Gradient	0.059	<b>0.050</b>	0.053	<b>0.125</b>	0.046	0.047	
		M-EM	0.059	0.048	0.053	0.107	<b>0.049</b>	<b>0.048</b>	0.242
(E)	MAE( $\hat{\mathbf{R}}$ )	Gradient	0.059	0.070	<b>0.057</b>	0.063	0.051	0.053	
		M-EM	0.059	0.070	0.053	0.063	0.051	<b>0.065</b>	0.333
(F)	CN( $\hat{\mathbf{B}}$ )	Gradient	1.051	1.138	<b>1.065</b>	<b>2.078</b>	1.149	<b>1.053</b>	
		M-EM	1.051	1.138	1.064	1.431	<b>1.154</b>	1.048	1.255
(G)	CN( $\hat{\mathbf{R}}$ )	Gradient	3.414	3.290	259.959	49.978	3.091	55.084	
		M-EM	3.414	<b>3.303</b>	<b>262.145</b>	<b>814.376</b>	<b>3.143</b>	<b>1043.060</b>	9019.451
(H)	Iteration	Gradient	7.9	8.8	7.5	43.1	8.3	7.9	
		M-EM	<b>27.2</b>	<b>3655.5</b>	<b>288.6</b>	<b>644.3</b>	<b>5220.4</b>	<b>292.0</b>	4676.9

that the symmetric original EM algorithm was successfully applied to all data sets. Thus, we do not summarize the solutions of the original EM algorithm. The other three algorithms always gave positive definite  $\hat{\mathbf{R}}$  and  $\hat{\mathbf{B}}$  of full column rank. The other results are summarized in Table 7. There, the results for the symmetric original EM algorithm have not been presented, since it and the modified EM algorithm showed equivalent behavior, except for the slightly different performances in Condition S, which were ignorable, as reported later. In the table, the bold font in gray cells indicates the larger values between the two algorithms.

Panels (A) and (B) in Table 7 show the percentages of improper solutions with  $\hat{\psi}_j \leq 0$  and  $|\hat{r}_{kl}| > 1$ , respectively. The gradient algorithm produced improper  $\hat{\psi}_j$  in Conditions L<sub>C1</sub> and T, and improper  $\hat{r}_{kl}$  in L<sub>C2</sub>, L<sub>C3</sub>, and T', while the EM algorithm always gave proper solutions. The recovery of parameters was assessed by MAE( $\hat{\mathbf{B}}$ ) =  $h^{-1} \sum_{j=1}^p \sum_{k=1}^m h_{jk} |\hat{b}_{jk} - b_{jk}|$ , MAE( $\hat{\Psi}$ ) =  $p^{-1} \sum_{j=1}^p |\hat{\psi}_j - \psi_j|$ , and MAE( $\hat{\mathbf{R}}$ ) =  $3^{-1} (|\hat{r}_{21} - r_{21}| + |\hat{r}_{31} - r_{31}| + |\hat{r}_{32} - r_{32}|)$ , whose averages over the data sets are shown in Panels (C), (D), and (E), respectively, where  $h_{jk} = 0$  if  $\hat{b}_{jk} = 0$  and  $h_{jk} = 1$  if otherwise, and  $h = \sum_{j,k} h_{jk}$ . In the panels, we found that the goodness of recovery was not very different between the two algorithms. Panels (F) and (G) show the averages of condition numbers CN( $\hat{\mathbf{B}}$ ) and CN( $\hat{\mathbf{R}}$ ), with the true counterpart CN( $\mathbf{B}$ ) being 1.114 in Condition L<sub>C2</sub>, 1.129 in T, and 1.0 in the others, and with CN( $\mathbf{R}$ ) being 111.92 in L<sub>C2</sub>, 45.935 in T', and 3.0 in the other conditions. We can find that CN( $\hat{\mathbf{R}}$ ) was very large in all conditions except P, L<sub>C1</sub>, and T. In particular, it is noted that CN( $\hat{\mathbf{R}}$ ) for the EM algorithms was far larger than CN( $\hat{\mathbf{R}}$ ) for the gradient one in L<sub>C3</sub> and T'. That is, in those conditions, the EM algorithm gave nearly rank deficient  $\hat{\mathbf{R}}$ , though all  $\hat{\mathbf{R}}$  were proper and positive definite. Such a solution in L<sub>C3</sub> is presented in Table 8 with the corresponding gradient algorithm's solution which includes improper  $\hat{r}_{42}$  and  $\hat{r}_{43}$ . Here, we should note the above definition of MAE( $\hat{\mathbf{R}}$ ) in which  $\hat{r}_{41}$ ,  $\hat{r}_{42}$ , and  $\hat{r}_{43}$  have not been considered, since their true counterparts do not exist (in L<sub>C3</sub> with  $\hat{m} = 4 > m = 3$ ). Thus, MAE( $\hat{\mathbf{R}}$ ) did not take large values.

Let us return back to Table 7. Panel (H) shows that the EM algorithm needed far more iterations until convergence than the gradient algorithm. The numbers were different between the modified and symmetric original EM algorithms for three data sets in Condition S. But the averaged number of iterations (4678.2) for the latter algorithm was very close to the number

TABLE 8.

True parameters, the solutions of CFA with gradient and modified EM algorithms, and their MAE and CN values for a data set in Condition  $L_{C3}$ , with blank cells standing for zero loadings.

(A) True parameters			(B) Gradient algorithm				(C) EM algorithm				
$\mathbf{B}$			$\hat{\mathbf{B}}$				$\hat{\mathbf{B}}$				
$\boldsymbol{\Psi} \mathbf{1}_p$			$\hat{\boldsymbol{\Psi}} \mathbf{1}_p$				$\hat{\boldsymbol{\Psi}} \mathbf{1}_p$				
0.71		0.5				0.23	0.95			0.29	0.91
0.71		0.5	0.84				0.29	0.85			0.28
0.71		0.5	0.65				0.58	0.65			0.58
0.71		0.5	0.61				0.62	0.61			0.63
	0.71	0.5				0.40	0.84			0.37	0.86
	0.71	0.5	0.74				0.46	0.73			0.47
	0.71	0.5	0.68				0.53	0.69			0.52
	0.71	0.5	0.64				0.59	0.64			0.59
		0.71	0.5			0.44	0.81			0.68	0.54
		0.71	0.5			0.75	0.44		0.75		0.44
		0.71	0.5			0.75	0.43		0.77		0.41
		0.71	0.5			0.72	0.48		0.72		0.48
$\mathbf{R}$			$\hat{\mathbf{R}}$				$\hat{\mathbf{R}}$				
1			1				1				
0.4	1		0.27	1			0.26	1			
0.4	0.4	1	0.35	0.43	1		0.34	0.43	1		
			0.55	1.10	1.17	1	0.34	0.63	0.97	1	
(Note $0.71 \cong 0.5^{1/2}$ )			MAE( $\hat{\mathbf{B}}$ ) = 0.057; MAE( $\hat{\boldsymbol{\Psi}}$ ) = 0.152;				MAE( $\hat{\mathbf{B}}$ ) = 0.057; MAE( $\hat{\boldsymbol{\Psi}}$ ) = 0.130;				
CN( $\mathbf{B}$ ) = 1;			MAE( $\hat{\mathbf{R}}$ ) = 0.069; CN( $\hat{\mathbf{B}}$ ) = 2.28;				MAE( $\hat{\mathbf{R}}$ ) = 0.075; CN( $\hat{\mathbf{B}}$ ) = 1.57;				
CN( $\mathbf{R}$ ) = 3			CN( $\hat{\mathbf{R}}$ ) = 7.29				CN( $\hat{\mathbf{R}}$ ) = 986.8				

(4676.9) for the former in Panel (H), and the other statistics for the symmetric original algorithm were equivalent to those in Table 7. We can find that the EM algorithm did not give improper solutions even in Condition S, but the recovery of parameters was not good with  $\hat{\mathbf{R}}$  nearly rank deficient in the condition.

#### 6.4. Summary of Major Results

The three EM algorithms for EFA gave equivalent solutions. Also, the symmetric original and modified EM algorithms for CFA yielded equivalent solutions, except for their slightly different behavior for singular sample correlation matrices. However, the original EM algorithm for CFA often failed to converge. The results of that algorithm are thus excluded from the following summary.

The results in Section 3 were demonstrated: the EM algorithm always gave a proper solution matrix of full rank. The recovery by the EM algorithm was better than, or almost the same as, that for the gradient algorithm, except for those cases with the true parameters being improper. However, the EM algorithm for CFA yielded factor correlation matrices whose ranks were nearly deficient for the cases where models were over-factorized with  $\hat{m} > m$ , and where true factor correlations were improper. The EM algorithm needed more iterations for convergence than the gradient algorithm, except for two conditions ( $L_{E2}$  and  $L_{E3}$ ) in EFA.

The gradient algorithms often produced improper solutions for the cases where the true parameters are improper and where models are over-factorized, but also gave proper solutions for some data sets in such cases (Table 5, Panel (A); Table 7, Panel (A)), the implications of which are mentioned in the next section.

## 7. Conclusions

In this paper, we proved the following mathematical fact: Rubin and Thayer's (1982) EM factor analysis always produces a proper solution when the covariance matrix to be analyzed and the initial matrices including unique variances and inter-factor correlations are positive definite. We further presented the symmetric original and modified EM algorithms in which Rubin and Thayer's (1982) original formulas are replaced by those with more numerically desirable properties. The real data examples and simulation studies showed that the original, symmetric original, and modified EM algorithms behave equivalently well for EFA, but that the original EM algorithm should not be used for CFA. It is inconclusive which is superior, the symmetric original or the modified EM algorithm, since they showed almost equivalent performances for CFA.

Obviously, EM factor analysis is welcomed by users who wish to avoid improper solutions. However, one drawback of EM factor analysis is that it needs more iterations for convergence (except for special conditions for EFA), although this difficulty may be dealt with by incorporating an acceleration technique into the factor analysis (Kuroda & Sakakihara, 2006). The following two points may also be drawbacks: First, EM factor analysis is a constrained estimation procedure that restricts solutions to be proper and, thus, involves difficulties with constrained maximum likelihood estimation procedures (e.g., Savalei & Kolenikov, 2008). Second, EM factor analysis yields proper solutions when models are over-factorized and even when true parameter values are improper, which implies that the impropriety of user model specifications and the impropriety of true parameters are not detected. On the other hand, such impropriety can be suggested by improper solutions in the unconstrained gradient algorithms. However, such a suggestion is not always given, since the gradient algorithms can also give proper solutions for true improper and over-factorization cases.

## Appendix A. SMW Formula

If a  $q \times q$  matrix  $\mathbf{W}$  is nonsingular, then  $\mathbf{I}_q - \mathbf{M}(\mathbf{M}'\mathbf{M} + \mathbf{W})^{-1}\mathbf{M}' = (\mathbf{I}_q + \mathbf{M}\mathbf{W}^{-1}\mathbf{M}')^{-1}$ . We refer to this equality as an SMW formula, as the equality is a special case of the so-called Sherman–Morrison–Woodbury formula (e.g., Seber, 2008, p. 309).

## Appendix B. Computer Programs

The algorithms for EM factor analysis were run using computer programs which we wrote in FORTRAN. In EM-EFA,  $\mathbf{A}_m\boldsymbol{\Omega}_m^{1/2}$  and  $\text{diag}\{\mathbf{C}_{yy} - \mathbf{A}_m\boldsymbol{\Omega}_m\mathbf{A}_m'\}$  were used for the initial matrices of  $\mathbf{B}$  and  $\boldsymbol{\Psi}$ , respectively. Here,  $\mathbf{A}_m(p \times m)$  and  $\boldsymbol{\Omega}_m(m \times m)$  were obtained using eigenvalue decomposition (EVD)  $\mathbf{C}_{yy} = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}'$ , with  $\boldsymbol{\Omega}(p \times p)$  being the diagonal matrix whose diagonal elements were arranged in descending order, and  $\mathbf{A}(p \times p)$  satisfying  $\mathbf{A}'\mathbf{A} = \mathbf{I}_p$ . That is,  $\mathbf{A}_m$  contained the first  $m$  columns of  $\mathbf{A}$ , and  $\boldsymbol{\Omega}_m$  was the first  $m \times m$  diagonal block of  $\boldsymbol{\Omega}$ . In EM-CFA, initial  $\mathbf{A}$ ,  $\boldsymbol{\Psi}$ , and  $\mathbf{R}$  were obtained by McDonald and Hartmann's (1992) method. In the programs, every square root of a symmetric positive definite matrix was obtained using EVD; for example,  $\mathbf{C}_{yy}^{1/2} = \mathbf{A}\boldsymbol{\Omega}^{1/2}\mathbf{A}'$ . Though  $\mathbf{C}_{yy}^{1/2}$  may also be set at  $\boldsymbol{\Omega}^{1/2}\mathbf{A}'$ , we chose symmetric  $\mathbf{C}_{yy}^{1/2} = \mathbf{A}\boldsymbol{\Omega}^{1/2}\mathbf{A}'$  for the sake of the convenience that  $\mathbf{C}_{yy}^{1/2}$  and  $\mathbf{C}_{yy}^{1/2'}$  can be treated equally.

The gradient algorithms were run using SAS software; EFA and CFA were performed with the FACTOR and CALIS procedures, respectively (SAS Institute Inc., 2009), in which improper solutions were not suppressed.

In all (FORTRAN and SAS) programs for EFA, iterations were stopped when the maximum of the changes in unique variances with their updates became less than  $0.1^5$ , or when

the number of iterations reached 30,000. In all programs for CFA, iterations were stopped when  $\max_{1 \leq i \leq I} \{|\theta_i^{(\text{new})} - \theta_i^{(\text{old})}| / \max(|\theta_i^{(\text{new})}|, |\theta_i^{(\text{old})}|), 0.01)\}$  became less than  $0.1^5$  with  $\theta_1^{(\text{new})}, \dots, \theta_I^{(\text{new})}$  expressing the current estimates of all unconstrained parameters and  $\theta_i^{(\text{old})}$  denoting the counterpart of  $\theta_i^{(\text{new})}$  before the current update.

### Appendix C. Results Regarding Numerical Precision

Real numbers were processed with *double* precision in the computer programs for EM factor analysis. However, we tried once to perform EM-CFA for the second example in Section 5, using the program of the original EM algorithm in which numerical precision was improved from *double* to *quadruple* precision. As a result, the algorithm did not decrease likelihood at the 74th iteration, with  $AS(\mathbf{R}) = 0.003$  and  $AS(\mathbf{R}\mathbf{B}'\Sigma^{-1}\mathbf{B}\mathbf{R}) = 0.001$  smaller than those values for *double* precision (Figure 1), and a decrease of likelihood was delayed until the 87th iteration. This shows that the failure of the original EM algorithm in the example is due to limited numerical precision.

### Appendix D. Procrustes Rotation

Let  $\hat{\mathbf{B}}_0$  denote an unrotated loading matrix obtained with EFA. We obtained the orthonormal matrix  $\mathbf{T}$  minimizing  $\|\hat{\mathbf{B}}_0\mathbf{T} - \mathbf{B}\|^2 = \text{tr}(\hat{\mathbf{B}}_0\mathbf{T} - \mathbf{B})'(\hat{\mathbf{B}}_0\mathbf{T} - \mathbf{B})$  and set  $\hat{\mathbf{B}} = \hat{\mathbf{B}}_0\mathbf{T}$  in all conditions except  $L_{E3}$ . In Condition  $L_{E3}$  with  $\hat{m} = m + 1$  (Table 3), we obtained the  $\hat{m} \times \hat{m}$  orthonormal  $\mathbf{T}$  and  $p \times 1$  unconstrained vector  $\mathbf{c}$  minimizing  $g(\mathbf{T}, \mathbf{c}) = \|\hat{\mathbf{B}}_0\mathbf{T} - [\mathbf{B}, \mathbf{c}]\|^2$ , set  $\hat{\mathbf{B}} = \hat{\mathbf{B}}_0\mathbf{T}$ , and redefined  $m$  as  $\hat{m}$  and  $\mathbf{B}$  as  $[\mathbf{B}, \mathbf{c}]$ , with  $[\mathbf{B}, \mathbf{c}]$  denoting the  $p \times (m + 1)$  matrix whose left and right blocks are  $\mathbf{B}$  and  $\mathbf{c}$ . The minimization of  $g(\mathbf{T}, \mathbf{c})$  is attained by the alternate iteration of [1] minimizing  $g(\mathbf{T}, \mathbf{c})$  over  $\mathbf{T}$  with  $\mathbf{c}$  kept fixed, and [2] minimizing  $g(\mathbf{T}, \mathbf{c})$  over  $\mathbf{c}$  with  $\mathbf{T}$  fixed, where [1] is the orthogonal Procrustes problem and [2] is achieved by setting  $\mathbf{c}$  at the last column of  $\hat{\mathbf{B}}_0\mathbf{T}$ .

### References

- Anderson, J.C., & Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155–173.
- Anderson, T.W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (pp. 111–150). Berkeley: University of California Press.
- Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, *39*, 1–38.
- Gerbing, D.W., & Anderson, J.C. (1987). Improper solutions in the analysis of covariance structures: their interpretability and a comparison of alternative respecifications. *Psychometrika*, *52*, 99–111.
- Jennrich, R.I., & Robinson, S.M. (1969). A Newton–Raphson algorithm for maximum likelihood factor analysis. *Psychometrika*, *34*, 111–123.
- Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*, 443–482.
- Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183–202.
- Kano, Y. (1998). Causes and treatment of improper solutions: exploratory factor analysis. *Bulletin of the Department of Human Sciences, Osaka University*, *24*, 303–327 (in Japanese).
- Kuroda, M., & Sakakihara, M. (2006). Accelerating the convergence of the EM algorithm using the vector epsilon algorithm. *Computational Statistics & Data Analysis*, *51*, 1549–1561.
- Lütkepohl, H. (1996). *Handbook of matrices*. Chichester: Wiley.
- Magnus, J.R., & Neudecker, H. (1991). *Matrix differential calculus with applications in statistics and econometrics* (2nd ed.). Chichester: Wiley.
- Maxwell, A.E. (1961). Recent trends in factor analysis. *Journal of the Royal Statistical Society. Series A*, *124*, 49–59.
- McDonald, R.P., & Hartmann, W.M. (1992). A procedure for obtaining initial values of parameters in the RAM model. *Multivariate Behavioral Research*, *27*, 57–76.

- Minami, M. (2004). Convergence speed and acceleration of the EM algorithm. In M. Watanabe & K. Yamaguchi (Eds.), *The EM algorithm and related statistical models* (pp. 85–94). New York: Marcel Dekker
- Mulaik, S.A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton: CRC Press.
- Rubin, D.B., & Thayer, D.T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, 47, 69–76.
- SAS Institute Inc. (2009). *SAS/STAT<sup>®</sup> 9.2 users guide*, version (2nd ed.). Cary: SAS Institute Inc.
- Sato, M. (1987). Pragmatic treatment of improper solutions in factor analysis. *Annals of the Institute of Statistical Mathematics*, 39, 443–455.
- Savalei, V., & Kolenikov, S. (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods*, 13, 150–170.
- Seber, G.A.F. (2008). *A matrix handbook for statisticians*. Hoboken: Wiley.
- SPSS Inc. (1997). *SPSS 7.5 statistical algorithm*. Chicago: SPSS Inc.
- Van Driel, O.P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, 43, 225–243.
- Yanai, H., & Ichikawa, M. (2007). Factor analysis. In C.R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 257–296). Amsterdam: Elsevier.

*Manuscript Received: 29 JUN 2011*

*Final Version Received: 19 MAY 2012*

*Published Online Date: 28 NOV 2012*