Statistical Properties of Matrix Decomposition Factor Analysis

Yoshikazu Terada*
Graduate School of Engineering Science, Osaka University
Center for Advanced Integrated Intelligence Research, RIKEN

1 Introduction

Exploratory factor analysis, often referred to as factor analysis, is an important technique of multivariate analysis (Anderson 2003). Factor analysis is a method for exploring the underlying structure of a set of variables and is applied in various fields. In factor analysis, we consider the following model for a p-dimensional observation x:

$$x = \mu + \Lambda f + \epsilon, \tag{1}$$

where $\mu \in \mathbb{R}^p$ is a mean vector, m is the number of factors ($m < p$), $\Lambda \in \mathbb{R}^{p\times m}$ is a factor loading matrix, f be a m-dimensional centered random vector with the identity covariance, ϵ be a p-dimensional uncorrelated centered random vector, which is independent from f, with diagonal covariance matrix $\mathrm{Var}(\epsilon) = \Psi^2 = \mathrm{diag}(\sigma_1^2, \dots, \sigma_p^2)$. Each component of f and ϵ are called the common and unique factors, respectively.

For a constant $c_\Lambda > 0$, let $\Theta_\Lambda := \{\Lambda \in \mathbb{R}^{p\times m} \mid |\lambda_{jk}| \le c_\Lambda \ (j = 1, \dots, p;\ k = 1, \dots, m)\}$ be the parameter space for the factor loading matrix Λ. For positive constants $c_L, c_U > 0$, define the parameter space for Ψ as $\Theta_\Psi := \{\mathrm{diag}(\sigma_1, \dots, \sigma_p) \mid c_L \le |\sigma_j| \le c_U \ (j = 1, \dots, p)\}$. Let $\Phi = [\Lambda, \Psi] \in \mathbb{R}^{p\times(m+p)}$, and define $\Theta_\Phi = \{\Phi = [\Lambda, \Psi] \mid \Lambda \in \Theta_\Lambda \text{ and } \Psi \in \Theta_\Psi\}$. For the factor model (1) with $\Phi = [\Lambda, \Psi]$, the covariance matrix of x is represented as $\Phi\Phi^\top = \Lambda\Lambda^\top + \Psi^2$.

We assume that the factor model (1) is true with some unknown parameter $\Phi_* = [\Lambda_*, \Psi_*] \in \Theta_\Phi$. Let $\Sigma_* = \Phi_*\Phi_*^\top = \Lambda_*\Lambda_*^\top + \Psi_*^2$ denote the true covariance matrix. It should be noted that the statistical properties described later still hold as a minimum contrast estimator even when the factor model (1) is not true. Let $(x_1, f_1, \epsilon_1), \dots, (x_n, f_n, \epsilon_n)$ be i.i.d. copies of (x, f, ϵ), where $(f_1, \epsilon_1), \dots, (f_n, \epsilon_n)$ are not observable in practice. Throughout the paper, it is assumed that $n > m+p$. In factor analysis, we aim to estimate (Λ_*, Ψ_*) from the observations $X_n = (x_1, \dots, x_n)^\top$. Here, we note that the factor model (1) has an indeterminacy. For example, for any $m \times m$ orthogonal matrix R, a rotated loading matrix $\Lambda_* R$ can also serve as a true loading matrix. Thus, let $\Theta_\Phi^* = \{\Phi \in \Theta_\Phi \mid \Sigma_* = \Phi\Phi^\top\}$ be the set of all possible true parameters.

*This research was supported by JSPS KAKENHI Grant (JP20K19756, JP20H00601, JP23H03355, and JP24K14855).

There are several estimation approaches to estimate the parameter $\Phi = [\Lambda, \Psi]$, e.g., maximum likelihood estimation, least-squares estimation, generalized least-squares estimation (Jöreskog & Goldberger 1972), minimum rank factor analysis (ten Berge & Kiers 1991), and non-iterative estimation (Ihara & Kano 1986, Kano 1990). The theoretical properties of these estimation approaches have been extensively studied. Moreover, most of these estimators can be formulated as minimum discrepancy function estimators. Thus, we can apply the general theory of minimum discrepancy function estimators to derive the theoretical properties of the estimators (Shapiro 1983, 1985).

One might think that the maximum likelihood estimator is the best choice from the viewpoint of efficiency. However, it is known that the maximum likelihood estimator is sensitive to the model error on (1) while it is robust to the distributional assumptions. For more details, see MacCallum & Tucker (1991) and Briggs & MacCallum (2003).Thus, other estimators could be better choices than the maximum likelihood estimator in practice.

In the early 2000s, a novel estimator based on matrix factorization was developed for factor analysis (Socan 2003, de Leeuw 2004). According to Adachi & Trendafilov (2018), this method was originally developed by Professor Henk A. L. Kiers and first appeared in Socan's dissertation (Socan 2003). This method is called matrix decomposition factor analysis (MDFA for short). The MDFA algorithm always provides proper solutions (i.e., no Heywood cases in MDFA); thus, it is computationally more stable than the maximum likelihood estimator. From the aspect of computational statistics, matrix decomposition factor analysis has been well-studied, and several extensions have been developed (see, e.g., Trendafilov & Unkel 2011, Trendafilov et al. 2013, Adachi 2022, Cho & Hwang 2023, Yamashita 2024). An important extension for high-dimensional data is the sparse estimation of matrix decomposition factor analysis with the ℓ_0-constraint. Although the sparse estimation with the ℓ_0-constraint has no bias, unlike other sparse regularizations, the optimization process with the ℓ_0-constraint is generally challenging. There is no sparse version of classical factor analysis with the ℓ_0-constraint. Surprisingly, the sparse MDFA estimator with the ℓ_0-constraint can be easily obtained, as described in Section 2.

In matrix decomposition factor analysis, the estimator is obtained by minimizing the following principal component analysis-like loss function:

$$\mathcal{L}_n(\mu, \Lambda, \Psi, F, E) = \frac{1}{n} \sum_{i=1}^{n} \|x_i - (\mu + \Lambda f_i + \Psi e_i)\|^2,$$

where $e_i = (e_{i1}, \ldots, e_{ip})^\top$, $E = (e_1, \ldots, e_n)^\top \in \mathbb{R}^{n \times p}$, and $F = (f_1, \ldots, f_n)^\top \in \mathbb{R}^{n \times m}$. As described in Section 2, certain constraints are imposed on the common factor matrix F and the normalized unique factor matrix E. It is known that we cannot consider the maximum likelihood estimation to the problem of simultaneous estimation of (Λ, Ψ) and latent factor scores $f_1, \ldots, f_n$ (see Section 7.7 and Section 9 of Anderson & Rubin (1956)). However, interestingly, we consider the simultaneous estimation in matrix decomposition factor analysis, and the MDFA estimator can be interpreted as the maximum likelihood estimator of the semiparametric model.

It is important to note that the loss function of principal component analysis can be written as

$$\mathcal{L}_{\text{PCA}}(\mu, \Lambda, F) = \frac{1}{n} \sum_{i=1}^{n} \|x_i - (\mu + \Lambda f_i)\|^2,$$

with the same constraints on F. The equivalence between this formulation and other standard formulations of principal component analysis can be found in Adachi (2016).

This formulation clearly shows that the term Ψe_i is the only difference between principal component analysis and matrix decomposition factor analysis. Although the loss function of matrix decomposition factor analysis is very similar to that of principal component analysis, the MDFA estimator empirically behaves like other consistent estimators used in factor analysis rather than principal component analysis. For high-dimensional data, it is well-known that principal component analysis and factor analysis are approximately the same (e.g., Bentler & Kano (1990) and Section 2.1 of Fan et al. (2013)). Thus, we can expect the MDFA estimator to exhibit similar behavior in high-dimensional settings. On the other hand, even in low-dimensional cases, matrix decomposition factor analysis yields results close to those of other consistent estimators for factor analysis.

Unlike classical factor analysis, matrix decomposition factor analysis treats the common factors F and normalized unique factors E as parameters that are estimated simultaneously with $\Phi = [\Lambda, \Psi]$. The number of parameters linearly depends on the sample size n, and the standard asymptotic theory of classical M-estimators cannot be directly applied to analyze its theoretical properties. As a result, the statistical properties of the MDFA estimator have yet to be discussed, leading to the open problem: Can matrix decomposition factor analysis truly be regarded as "factor analysis"?

In this paper, we establish the statistical properties of matrix decomposition factor analysis to answer this question. We show that as the sample size n goes to infinity, the MDFA estimator converges to the true parameter $\Phi_* \in \Theta_\Phi^*$. First, we formulate the MDFA estimator as the semiparametric profile likelihood estimator and derive the explicit form of the profile likelihood. Next, we reveal the population-level loss function of matrix decomposition factor analysis and its fundamental properties. Then, we show the statistical properties of matrix decomposition factor analysis.

Throughout the paper, let us denote by $\lambda_j(A)$ the jth largest eigenvalue of a symmetric matrix A. Let $\|\cdot\|_2$ and $\|\cdot\|_F$ represent the operator norm and the Frobenius norm for a matrix, respectively. Let I_p denote the identity matrix of size p, and let $O_{p\times q}$ denote the $p \times q$ matrix of zeros. The p-dimensional vectors of ones will be denoted by 1_p, and the vectors of zeros will be denoted by 0_p. For matrix A, let A^+ denote the Moore-Penrose inverse of A. We will denote by $\mathcal{O}(p \times q)$ the set of all $p \times q$ column-orthogonal matrices and will denote by $\mathcal{O}(p)$ the set of all $p \times p$ orthogonal matrices.

2 Matrix decomposition factor analysis (MDFA)

We will briefly describe the matrix decomposition factor analysis. Without loss of generality, the data matrix X_n is centered by the sample mean, and we ignore the estimation of the mean vector μ. For simplicity of notation, we use the same symbol X_n for the centered data matrix. Let $F_n = (f_1, \ldots, f_n)^\top$ and $\mathcal{E}_n = (\epsilon_1, \ldots, \epsilon_n)^\top$. The factor model (1) can be expressed in the matrix form as $X_n = F_n\Lambda^\top + \mathcal{E}_n$. From this representation, we can naturally consider the following matrix factorization problem:

$$X_n \approx F\Lambda^\top + E\Psi,$$

where $F \in \mathbb{R}^{n\times m}$ and $E \in \mathbb{R}^{n\times p}$ are constrained by

$$1_n^\top F = 0_m^\top,\ 1_n^\top E = 0_p^\top,\ \frac{1}{n}F^\top F = I_m,\ \frac{1}{n}E^\top E = I_p,\ \text{and}\ F^\top E = O_{m\times p}. \tag{2}$$

In the constraint (2), the first two conditions are empirical counterparts of the assumptions $\mathbb{E}[f] = 0_m$ and $\mathbb{E}[\epsilon] = 0_p$. Moreover, the remaining conditions correspond to the covariance constraints for f and ϵ. This matrix factorization approach is called the matrix decomposition factor analysis (MDFA). Here, we note that the factor $1/n$ can be replaced by the factor $1/(n-1)$ in the constraint (2). This modification is essential in practice, as will be described later.

Let $\Theta_Z := \{Z = [F, E] \in \mathbb{R}^{n\times(m+p)} \mid Z$ satisfies the constraint (2)$\}$. In matrix decomposition factor analysis, the estimator $(\widehat{\Lambda}_n, \widehat{\Psi}_n, \widehat{F}_n, \widehat{E}_n)$ is obtained by minimizing the following loss function over $\Phi = [\Lambda, \Psi] \in \Theta_\Phi$ and $Z = [F, E] \in \Theta_Z$:

$$\mathcal{L}_n(\Phi, Z) = \frac{1}{n}\left\|X_n - Z\Phi^\top\right\|_F^2 = \frac{1}{n}\sum_{i=1}^n \|x_i - (\Lambda f_i + \Psi e_i)\|^2. \tag{3}$$

Thus, we can formulate the matrix decomposition factor analysis as the maximum likelihood estimation of the following semiparametric model:

$$x_i = \Lambda f_i + \Psi e_i + \xi_i \quad (i = 1, \ldots, n), \tag{4}$$

where both common and unique factor score vectors $f_1, \ldots, f_n, e_1, \ldots, e_n$ are fixed vectors satisfying the constraint (2), and $\xi_1, \ldots, \xi_n$ are independently distributed according to a p-dimensional centered normal distribution with a known variance $\tau_0^2 I_p$.

Now, we recall that the principal component analysis can be formulated as the minimization problem of the following loss function with constraints $1_n^\top F = 0_m^\top$ and $F^\top F/n = I_m$:

$$\frac{1}{n}\left\|X_n - F\Lambda^\top\right\|_F^2.$$

The only difference between matrix decomposition factor analysis and principal component analysis is the term $E\Psi$ in the loss function (3). When we impose the constraint that $\Lambda^\top\Lambda$ is a diagonal matrix whose diagonal elements are arranged in decreasing order to identify the parameter Λ uniquely, the principal component estimator can be represented as

$$\widehat{\Lambda}_{\text{PCA}} = L_m\Delta_m/\sqrt{n} \ \text{ and }\ \widehat{F}_{\text{PCA}} = \sqrt{n}K_m,$$

where Δ_m is a diagonal matrix with the first m singular values of X_n, and $K_m \in \mathcal{O}(n\times m)$ and $L_m \in \mathcal{O}(p\times m)$ are the matrices of first m left singular vectors and right singular vectors of X_n, respectively.

Adachi & Trendafilov (2018) shows several essential properties of matrix decomposition factor analysis. Here, we will introduce some of these properties. The loss function $\mathcal{L}_n$ can be written as follows:

$$\mathcal{L}_n(\Phi, Z) = \frac{1}{n}\|X_n\|_F^2 + \|\Phi\|_F^2 - \frac{2}{n}\text{tr}\{(X_n\Phi)^\top Z\} \tag{5}$$

$$= \frac{1}{n}\|X_n - ZZ^\top X_n/n\|_F^2 + \|X_n^\top Z/n - \Phi\|_F^2. \tag{6}$$

Algorithm 1 An algorithm of matrix decomposition factor analysis.

1: Initialize $t = 0$ and $\Phi_{(0)} = \left[\Lambda_{(0)}, \Psi_{(0)}\right] \in \Theta_\Phi$.

2: Update $t = t + 1$. By the singular value decomposition of $X_n\Phi_{(t-1)}/\sqrt{n}$, update the parameter Z as follows:

$$\widehat{Z}_{(t)} = \left[\widehat{F}_{(t)}, \widehat{E}_{(t)}\right] = \sqrt{n}\widehat{K}(\Phi_{(t-1)})\widehat{L}(\Phi_{(t-1)})^\top + \sqrt{n}\widehat{K}_\perp(\Phi_{(t-1)})\widehat{L}_\perp(\Phi_{(t-1)})^\top.$$

3: Update the parameter Φ as follows:

$$\widehat{\Phi}_{(t)} = \left[\widehat{\Lambda}_{(t)}, \widehat{\Psi}_{(t)}\right] = \left[X_n^\top \widehat{F}_{(t)}/n, \mathrm{diag}\big(X_n^\top \widehat{E}_{(t)}/n\big)\right].$$

4: Repeat Steps 2 and 3 until convergence.

Now, we consider minimizing the loss function $\mathcal{L}_n$. Let $\widehat{K}(\Phi)\widehat{\Delta}(\Phi)\widehat{L}(\Phi)^\top$ be the singular value decomposition of $X_n\Phi/\sqrt{n}$, where $\widehat{\Delta}(\Phi)$ is the diagonal matrix with the singular values, and $\widehat{K}(\Phi) \in \mathcal{O}(n \times p)$ and $\widehat{L}(\Phi) \in \mathcal{O}((m+p) \times p)$ are the matrix of the left singular vectors and the matrix of the right singular vectors, respectively. Let $\widehat{S}_n = X_n^\top X_n/n$ be the sample covariance matrix, and then the spectral decomposition of $\Phi^\top \widehat{S}_n \Phi$ can be written as

$$\Phi^\top \widehat{S}_n \Phi = \big(X_n^\top \Phi/\sqrt{n}\big)^\top \big(X_n^\top \Phi/\sqrt{n}\big) = \widehat{L}(\Phi)\widehat{\Delta}(\Phi)^2\widehat{L}(\Phi)^\top.$$

From (5), it follows that, for given Φ, the following $\widehat{Z}(\Phi)$ attains the minimum of $\mathcal{L}_n(\Phi, Z)$:

$$\widehat{Z}(\Phi) = \sqrt{n}\widehat{K}(\Phi)\widehat{L}(\Phi)^\top + \sqrt{n}\widehat{K}_\perp(\Phi)\widehat{L}_\perp(\Phi)^\top, \tag{7}$$

where $\widehat{K}_\perp(\Phi) \in \mathcal{O}(n \times m)$ and $\widehat{L}_\perp(\Phi) \in \mathcal{O}((m+p) \times m)$ are column-orthonormal matrices such that $1_n^\top \widehat{K}_\perp(\Phi) = 0_m^\top$ and

$$\widehat{K}(\Phi)^\top \widehat{K}_\perp(\Phi) = \widehat{L}(\Phi)^\top \widehat{L}_\perp(\Phi) = O_{p\times m}.$$

It is important to note that $\widehat{K}_\perp(\Phi)$ and $\widehat{L}_\perp(\Phi)$ are not uniquely determined.

For given $Z = [F, E] \in \Theta_Z$, the minimization of $\mathcal{L}_n$ with Φ is obvious. From (6), we conclude that, for given $Z \in \Theta_Z$, the optimal $\widehat{\Phi}(Z)$ is given by

$$\widehat{\Phi}(Z) = \left[\widehat{\Lambda}(Z), \widehat{\Psi}(Z)\right] = \left[X_n^\top F/n, \mathrm{diag}(X_n^\top E/n)\right], \tag{8}$$

where $\mathrm{diag}(X_n^\top E/n)$ is the diagonal matrix with diagonal elements of $X_n^\top E/n$. That is, $\widehat{\Lambda}(Z) = X_n^\top F/n$ and $\widehat{\Psi}(Z) = \mathrm{diag}(X_n^\top E/n)$.

Therefore, the minimization problem for $\mathcal{L}_n$ can be solved by a simple alternating minimization algorithm. An algorithm of matrix decomposition factor analysis is summarized in Algorithm 1.

Remark 2.1. *For estimating both Φ and Z, the original data matrix X_n is necessary. However, when only the estimator for Φ is needed (i.e., the estimator of Z is unnecessary), Adachi (2012) shows that the algorithm can be performed using only the sample covariance matrix $\widehat{S}_n$. Thus, for estimating only Φ, the computational cost of matrix decomposition factor analysis only depends on the dimension p. For more details of this algorithm, see Adachi & Trendafilov (2018).*

3 Asymptotic properties of MDFA

3.1 Main idea

First, we will describe the main idea for proving asymptotic properties of matrix decomposition factor analysis. Remark 2.1 indicates the possibility that the loss function $\mathcal{L}_n$ concentrated on Φ can be rewritten using the sample covariance matrix $\widehat{S}_n$ instead of the data matrix X_n. The following lemma shows that the concentrated loss function $\mathcal{L}_n(\Phi) = \min_{Z \in \Theta_Z} \mathcal{L}_n(\Phi, Z)$ has the explicit form with the sample covariance matrix. By considering $Z = [F, E]$ as the nuisance parameter in the semiparametric model (4), this loss $\mathcal{L}_n(\Phi)$ is related to the concentrating-out (or profile likelihood) approach (Newey 1994, Murphy & Van der Vaart 2000). That is, this loss $\mathcal{L}_n(\Phi)$ is the negative profile likelihood for the semiparametric model (4).

Lemma 3.1. *For any* $\Phi \in \Theta_\Phi$,

$$\begin{aligned}\mathcal{L}_n(\Phi) &= \min_{Z \in \Theta_Z} \mathcal{L}_n(\Phi, Z) \\ &= \mathrm{tr}\left[\{I_p - \widehat{A}(\Phi)\}^\top \widehat{S}_n \{I_p - \widehat{A}(\Phi)\}\right] + \left\| (\Phi^\top)^+ \left(\Phi^\top \widehat{S}_n \Phi\right)^{1/2} - \Phi \right\|_F^2,\end{aligned}$$

where $\widehat{A}(\Phi) = \Phi\Phi^\top \widehat{S}_n \Phi (\Phi^\top \widehat{S}_n \Phi)^+ = \Phi \widehat{L}(\Phi) \widehat{L}(\Phi)^\top \Phi^+$.

Remark 3.1. *When we replace the factor* $1/n$ *with* $1/(n-1)$ *in the loss fucntion (3) and the constraint (2), the sample covariance matrix* $\widehat{S}_n$ *is replaced by the unbiased covariance matrix* $\widehat{U}_n = X^\top X/(n-1)$ *in the explicit form of* $\mathcal{L}_n(\Phi)$. *This modification does not affect the asymptotic properties of the MDFA estimator but improves the finite-sample performance.*

3.2 Population-level loss function and its properties

For given $\Phi \in \mathbb{R}^{p \times (p+m)}$, the spectral decomposition of $\Phi^\top \Sigma_* \Phi \in \mathbb{R}^{(m+p) \times (m+p)}$ is denoted by

$$\Phi^\top \Sigma_* \Phi = L(\Phi) \Delta(\Phi)^2 L(\Phi)^\top,$$

where $\Delta(\Phi)^2$ is the diagonal matrix with ordered positive eigenvalues, and $L(\Phi) \in \mathcal{O}((m+p) \times p)$ is the matrix of eigenvectors. From Lemma 3.1, we can naturally consider the following population-level loss function:

$$\mathcal{L}(\Phi) = \mathrm{tr}\left[\{I_p - A(\Phi)\}^\top \Sigma_* \{I_p - A(\Phi)\}\right] + \left\| (\Phi^\top)^+ \left(\Phi^\top \Sigma_* \Phi\right)^{1/2} - \Phi \right\|_F^2,$$

where $A(\Phi) = \Phi(\Phi^\top \Sigma_* \Phi)(\Phi^\top \Sigma_* \Phi)^+ \Phi^+ = \Phi L(\Phi) L(\Phi)^\top \Phi^+$.

Since the loss $\mathcal{L}(\Phi)$ has a complex form, it is not immediately clear whether this loss $\mathcal{L}(\Phi)$ is reasonable for factor analysis. The following proposition shows that the loss function $\mathcal{L}(\Phi)$ is appropriate for factor analysis. That is, the population-level matrix decomposition factor analysis is identifiable up to the indeterminacy of the factor model.

Proposition 3.2 (Identifiability of the population-level MDFA)**.** *For any* $\Phi = [\Lambda, \Psi] \in \Theta_\Phi$, *the following two conditions are equivalent:*

$$\text{(i) } \Sigma_* = \Phi\Phi^\top = \Lambda\Lambda^\top + \Psi^2, \quad \textit{and} \quad \text{(ii) } \mathcal{L}(\Phi) = 0.$$

Interestingly, as shown in the proof, only the second term of the loss $\mathcal{L}$ is essential for this identifiability. Thus, the second term of the empirical loss $\mathcal{L}_n$ could be an appropriate loss function for factor analysis, but the optimization step will be more complicated.

Since $\mathcal{L}_n(\Phi)$ is not represented as the empirical mean over samples, even the smoothness of the loss $\mathcal{L}$ is non-trivial, unlike the classical theory of M-estimators. For example, the second term of the loss $\mathcal{L}(\Phi)$ involves the square root of the degenerate matrix $\Phi^\top \Sigma \Phi$. Generally, the differentiability of the square root function cannot be guaranteed for degenerate cases. Fortunately, from Theorem 2 of Freidlin (1968), we can ensure the smoothness of $(\Phi^\top \Sigma \Phi)^{1/2}$. The following proposition ensures the smoothness of the population-level loss $\mathcal{L}(\Phi)$. This smoothness can be directly obtained from the smoothness of $(\Phi^\top \Sigma \Phi)^{1/2}$ and the smoothness of $A(\Phi)$. For $\Lambda \in \mathbb{R}^{p\times m}$ and $\Psi \in \Theta_\Psi$, let $\mathrm{vec}(\Lambda)$ and $\mathrm{diag}(\Psi)$ be the vectorization of Λ and the diagonal vector of Ψ, respectively. Write $\Theta_\phi = \left\{(\mathrm{vec}(\Lambda)^\top, \mathrm{diag}(\Psi)^\top)^\top \in \mathbb{R}^{p(m+1)} \mid [\Lambda, \Psi] \in \Theta_\Phi\right\}$.

Proposition 3.3 (Smoothness of the population-level loss)**.** *The population-level loss function $\mathcal{L} : \Theta_\phi \to \mathbb{R}$ is smooth on the interior of the compact parameter space Θ_ϕ.*

3.3 Consistency

Now, we will show the strong consistency of matrix decomposition factor analysis. The following proposition ensures the uniform strong law of large numbers.

Proposition 3.4 (Uniform law of large numbers)**.** *For any $\Phi \in \Theta_\Phi$,*

$$|\mathcal{L}_n(\Phi) - \mathcal{L}(\Phi)| \le Const. \times \left(p\|\widehat{S}_n - \Sigma_*\|_F^{1/2} \vee p^4\|\widehat{S}_n - \Sigma_*\|_F\right),$$

where $a \vee b = \max(a, b)$, and $Const.$ is a global constant depending only on c_L, c_Λ, and c_U. Therefore, the following uniform strong law of large numbers holds:

$$\lim_{n\to\infty} \sup_{\Phi\in\Theta_\Phi} |\mathcal{L}_n(\Phi) - \mathcal{L}(\Phi)| = 0 \quad a.s. \tag{9}$$

By the above uniform law of large numbers for $\mathcal{L}_n$ and the continuity of the population-level loss $\mathcal{L}$, we can obtain the strong consistency of matrix decomposition factor analysis.

Theorem 3.5 (Consistency of the MDFA estimator)**.** *Assume the observation $X_n = (x_1, \ldots, x_n)^\top$ is an i.i.d. sample from the factor model (1). Let $\widehat{\Phi}_n = \left[\widehat{\Lambda}_n, \widehat{\Psi}_n\right]$ be the estimator of the matrix decomposition factor analysis for $\Phi = [\Lambda, \Psi]$. That is, $\widehat{\Phi}_n \in \arg\min_{\Phi\in\Theta_\Phi} \mathcal{L}_n(\Phi)$. Then,*

$$\lim_{n\to\infty} \mathcal{L}\big(\widehat{\Phi}_n\big) = 0 \ \ a.s., \quad and \quad \lim_{n\to\infty} \min_{\Phi_*\in\Theta_\Phi^*} \left\|\widehat{\Phi}_n - \Phi_*\right\|_F = 0 \ \ a.s.$$

From this theorem, the MDFA estimator converges to the true parameter, similar to other consistent estimators in factor analysis; thus, the MDFA estimator is appropriate for factor analysis. This theorem explains why matrix decomposition factor analysis provides results similar to those of other consistent estimators in factor analysis.

Under Anderson and Rubin's sufficient condition, a stronger consistency result can be achieved when combined with Theorem 1 of Kano (1983).

Moreover, even if the factor model (1) is incorrect, the MDFA estimator remains consistent as a minimum contrast estimator. That is, the MDFA estimator converges to a solution that minimizes the population-level loss $\mathcal{L}$.

3.4 Asymptotic normality

Based on the unified approach of Shapiro (1983), we can show the asymptotic normality of the MDFA estimator. To eliminate the rotational indeterminacy, we consider the following identifiability condition introduced by Anderson & Rubin (1956):

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & 0 & \cdots & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & \lambda_{m2} & \lambda_{m3} & \cdots & \lambda_{mm} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \lambda_{p3} & \cdots & \lambda_{pm} \end{bmatrix}. \tag{10}$$

When we impose this condition on the true loading matrix with $\lambda_{jj} > 0$ $(j = 1, \ldots, m)$, we can avoid both rotational and sign indeterminacy. This condition can be easily handled in matrix decomposition factor analysis. In fact, from the representation (6), we simply set the upper triangle part of $\widehat{\Lambda}_{(t)}$ to zero in Step 3 of Algorithm 1. This adjustment allows us to obtain the MDFA estimator that minimizes the loss function $\mathcal{L}_n$ under the condition (10).

Moreover, for the identifiability of the factor decomposition, we also assume Anderson and Rubin's sufficient condition for the true loading matrix Λ_*. That is, we assume that if any row of Λ_* is deleted, there remain two disjoint submatrices of rank m.

Let vech$(\cdot)$ be the vech operator. For $\Phi = [\Lambda, \Psi]$ with the condition (10), let $\theta = (\lambda_{11}, \ldots, \lambda_{p1}, \lambda_{22}, \ldots, \lambda_{p2}, \ldots, \lambda_{mm}, \ldots, \lambda_{pm}, \sigma_1^2, \ldots, \sigma_p^2)^\top$ be the parameter vector. Let $\Theta = \{\theta_\Lambda \in \mathbb{R}^{pm-m(m-1)/2} \mid |\theta_{\Lambda,j}| < c_\Lambda\} \times [c_L^2, c_U^2]^p$ be the compact parameter space for the parameter θ. To derive the asymptotic normality of the MDFA estimator, we redefine the loss functions $\mathcal{L}_n$ and $\mathcal{L}$ as a unified function of the parameter θ and the covariance matrix Σ. For any $\Phi \in \Theta_\Phi$ and any $p \times p$ positive definite matrix $\Sigma > 0$, we rewrite the spectral decomposition of $\Phi^\top \Sigma \Phi \in \mathbb{R}^{(m+p)\times(m+p)}$ as

$$\Phi^\top \Sigma \Phi = L(\Phi, \Sigma)\Delta(\Phi, \Sigma)^2 L(\Phi, \Sigma)^\top,$$

where $\Delta(\Phi, \Sigma)^2$ is the diagonal matrix with ordered positive eigenvalues, and $L(\Phi, \Sigma) \in \mathcal{O}((m+p) \times p)$ is the matrix of eigenvectors. For any $\theta \in \Theta$, define

$$\mathcal{L}(\theta, \Sigma) = \mathcal{L}(\Phi, \Sigma) = \mathrm{tr}\left[\{I_p - A(\Phi, \Sigma)\}^\top \Sigma \{I_p - A(\Phi, \Sigma)\}\right] + \left\| (\Phi^\top)^+ \left(\Phi^\top \Sigma \Phi\right)^{1/2} - \Phi \right\|_F^2,$$

where Φ is the matrix representation of θ, and $A(\Phi, \Sigma) = \Phi L(\Phi, \Sigma) L(\Phi, \Sigma)^\top \Phi^+$. Using this notation, we have $\mathcal{L}_n(\Phi) = \mathcal{L}(\Phi, \widehat{S}_n)$ and $\mathcal{L}(\Phi) = \mathcal{L}(\Phi, \Sigma_*)$.

Theorem 3.6 (Asymptotic normality of the MDFA estimator)**.** *Assume that the observation $X_n = (x_1, \ldots, x_n)^\top$ is an i.i.d. sample from the factor model (1) with the true parameter $\Phi_* = [\Lambda_*, \Psi_*]$. It is assumed that the fourth moment of x_1 is bounded. For the identifiability, suppose that the true loading matrix Λ_* satisfies Anderson and Rubin's sufficient condition and the condition (10) with $\lambda_{jj}^* > 0$ $(j = 1, \ldots, m)$. Let θ_* be the true* *parameter vector corresponding to Φ_*. Moreover, suppose that the true parameter lies in the interior of Θ, and the Hessian matrix*

$$H_{\theta\theta}(\theta_*, \Sigma_*) = \frac{\partial^2 \mathcal{L}(\theta_*, \Sigma_*)}{\partial\theta\partial\theta^\top}$$

is nonsingular at the point (θ_*, Σ_*).

Let $\widehat{\theta}_n$ be the MDFA estimator with the condition (10) for the true parameter θ_*. Here, the elements $\widehat{\lambda}_{jj}$ ($j = 1, \ldots, m$) of the estimator $\widehat{\theta}_n$ are set to be nonnegative. Then, we obtain the asymptotic normality of the MDFA estimator $\widehat{\theta}_n$: $\sqrt{n}\big(\widehat{\theta}_n - \theta_*\big) \to N(0, V)$ in distribution as $n \to \infty$, where the matrix Γ is the asymptotic variance of the sample covariance, $V = J\Gamma J^\top$, and

$$J = -H_{\theta\theta}^{-1}(\theta_*, \Sigma_*)\frac{\partial^2 \mathcal{L}(\theta_*, \Sigma_*)}{\partial\theta\partial\mathrm{vech}(\Sigma)^\top}.$$

References

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

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

Adachi, K. (2022), 'Factor analysis procedures revisited from the comprehensive model with unique factors decomposed into specific factors and errors', *Psychometrika* **87**(3), 967–991.

Adachi, K. & Trendafilov, N. T. (2018), 'Some mathematical properties of the matrix decomposition solution in factor analysis', *Psychometrika* **83**, 407–424.

Anderson, T. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley.

Anderson, T. & Rubin, H. (1956), Statistical inference in factor analysis, *in* 'Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability', University of California Press, p. 111.

Bentler, P. M. & Kano, Y. (1990), 'On the equivalence of factors and components', *Multivariate Behavioral Research* **25**(1), 67–74.

Briggs, N. E. & MacCallum, R. C. (2003), 'Recovery of weak common factors by maximum likelihood and ordinary least squares estimation', *Multivariate Behavioral Research* **38**(1), 25–56.

Cho, G. & Hwang, H. (2023), 'Structured factor analysis: A data matrix-based alternative approach to structural equation modeling', *Structural Equation Modeling: A Multidisciplinary Journal* **30**(3), 364–377.

de Leeuw, J. (2004), Least squares optimal scaling of partially observed linear systems, *in* 'Recent developments on structural equation models: Theory and applications', Springer, pp. 121–134.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **75**(4), 603–680.

Freidlin, M. I. (1968), 'On the factorization of non-negative definite matrices', *Theory of Probability & Its Applications* **13**(2), 354–356.

Ihara, M. & Kano, Y. (1986), 'A new estimator of the uniqueness in factor analysis', *Psychometrika* **51**, 563–566.

Jöreskog, K. & Goldberger, A. (1972), 'Factor analysis by generalized least squares', *Psychometrika* **37**, 243–260.

Kano, Y. (1983), 'Consistency of estimators in factor analysis', *Journal of the Japan Statistical Society* **13**(2), 137–144.

Kano, Y. (1990), 'Noniterative estimation and the choice of the number of factors in exploratory factor analysis', *Psychometrika* **55**, 277–291.

MacCallum, R. C. & Tucker, L. R. (1991), 'Representing sources of error in the common-factor model: Implications for theory and practice.', *Psychological Bulletin* **109**(3), 502–511.

Murphy, S. A. & Van der Vaart, A. W. (2000), 'On profile likelihood', *Journal of the American Statistical Association* **95**(450), 449–465.

Newey, W. K. (1994), 'The asymptotic variance of semiparametric estimators', *Econometrica* **62**(6), 1349–1382.

Shapiro, A. (1983), 'Asymptotic distribution theory in the analysis of covariance structures', *South African Statistical Journal* **17**(1), 33–81.

Shapiro, A. (1985), 'Asymptotic equivalence of minimum discrepancy function estimators to G.L.S. estimators', *South African Statistical Journal* **19**(1), 73–81.

Socan, G. (2003), The incremental value of minimum rank factor analysis., PhD dissertation, University of Groningen.

ten Berge, J. M. & Kiers, H. A. (1991), 'A numerical approach to the approximate and the exact minimum rank of a covariance matrix', *Psychometrika* **56**, 309–315.

Trendafilov, N. T. & Unkel, S. (2011), 'Exploratory factor analysis of data matrices with more variables than observations', *Journal of Computational and Graphical Statistics* **20**(4), 874–891.

Trendafilov, N. T., Unkel, S. & Krzanowski, W. (2013), 'Exploratory factor and principal component analyses: some new aspects', *Statistics and Computing* **23**, 209–220.

Yamashita, N. (2024), 'Matrix decomposition approach for structural equation modeling as an alternative to covariance structure analysis and its theoretical properties', *Structural Equation Modeling: A Multidisciplinary Journal* **31**(5), 817–834.