PSYCHOMETRIKA—VOL. 67, NO. 1, 79–94
MARCH 2002

STATISTICAL INFERENCE OF MINIMUM RANK FACTOR ANALYSIS

ALEXANDER SHAPIRO

SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING
GEORGIA INSTITUTE OF TECHNOLOGY

JOS M.F. TEN BERGE

HEYMANS INSTITUTE OF PSYCHOLOGICAL RESEARCH
UNIVERSITY OF GRONINGEN

For any given number of factors, Minimum Rank Factor Analysis yields optimal communalities for an observed covariance matrix in the sense that the unexplained common variance with that number of factors is minimized, subject to the constraint that both the diagonal matrix of unique variances and the observed covariance matrix minus that diagonal matrix are positive semidefinite. As a result, it becomes possible to distinguish the explained common variance from the total common variance. The percentage of explained common variance is similar in meaning to the percentage of explained observed variance in Principal Component Analysis, but typically the former is much closer to 100 than the latter. So far, no statistical theory of MRFA has been developed. The present paper is a first start. It yields closed-form expressions for the asymptotic bias of the explained common variance, or, more precisely, of the unexplained common variance, under the assumption of multivariate normality. Also, the asymptotic variance of this bias is derived, and also the asymptotic covariance matrix of the unique variances that define a MRFA solution. The presented asymptotic statistical inference is based on a recently developed perturbation theory of semidefinite programming. A numerical example is also offered to demonstrate the accuracy of the expressions.

Key words: factor analysis, communalities, proper solutions, explained common variance, semidefinite programming, large samples asymptotics, asymptotic normality, asymptotic bias.

1. Introduction

Factor analysis is based on the notion that given a set of variables $z_1, \ldots, z_p$, each variable z_j can be decomposed into a common part c_j and a unique part u_j, assumed to be uncorrelated with any variable except z_j, $j = 1, \ldots, p$. Upon writing

$$z_j = c_j + u_j, \tag{1}$$

$j = 1, \ldots, p$, and using the assumption on u_j, we have

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c + \boldsymbol{\Psi}, \tag{2}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the variables, $\boldsymbol{\Psi}$ is the diagonal matrix of unique variances, and $\boldsymbol{\Sigma}_c$ is the variance-covariance matrix of the common parts c_j, $j = 1, \ldots, p$, of the variables. The variances of these common parts are in the diagonal of $\boldsymbol{\Sigma}_c$. They are the so-called communalities of the variables.

The ideal of factor analysis is to find a decomposition (2) with $\boldsymbol{\Sigma}_c$ of low rank r, which can be factored as $\boldsymbol{\Sigma}_c = \mathbf{F}\mathbf{F}'$, with $\mathbf{F}$ a $p \times r$ matrix. To accomplish this, communalities are required that reduce the rank of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}$ to some small value. Although the early days of factor analysis were characterized by great optimism in this respect (Ledermann, 1937), the ideal of low reduced rank will never be attained in practice, see Guttman (1958) and Shapiro (1982). A

This work was supported, in part, by grant DMS-0073770 from the National Science Foundation.

Requests for reprints should be sent to Alexander Shapiro, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205. E-Mail: ashapiro@isye.gatech.edu

0033-3123/2002-1/2000-0824-A $00.75/0

historical overview of how the ideal of low reduced rank was shattered can be found in ten Berge (1998).

For practical purposes, there is no choice other than trying to approximate the ideal low rank situation by using the eigendecomposition $\mathbf{\Sigma}_c = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2$, of the covariance matrix of the common parts, where both matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are positive semidefinite, matrix $\mathbf{\Sigma}_1$ has rank r (for some small value of r), and r nonzero eigenvalues of $\mathbf{\Sigma}_1$ coincide with r largest eigenvalues of $\mathbf{\Sigma}_c$. This leads to the following decomposition of $\mathbf{\Sigma}$,

$$\mathbf{\Sigma} = \mathbf{F}\mathbf{F}' + (\mathbf{\Sigma}_c - \mathbf{F}\mathbf{F}') + \mathbf{\Psi}. \tag{3}$$

This means that the variances of the variables (diagonal elements of $\mathbf{\Sigma}$) are decomposed into explained common variances (diagonal elements of $\mathbf{F}\mathbf{F}'$), unexplained common variances (diagonal elements of $\mathbf{\Sigma}_c - \mathbf{F}\mathbf{F}'$), and unique variances (diagonal elements of $\mathbf{\Psi}$).

It is essential to note that a proper solution for (3) requires that both matrices $\mathbf{\Sigma}_c$ and $\mathbf{\Psi}$ should be positive semidefinite (denoted $\mathbf{\Sigma}_c \succeq \mathbf{0}$ and $\mathbf{\Psi} \succeq \mathbf{0}$, respectively). Negative elements in $\mathbf{\Psi}$, known as Heywood cases, have drawn a lot of attention, and are usually not tolerated. However, when $\mathbf{\Sigma}_c$, the covariance matrix for the common parts of the variables, would appear to be indefinite, that would be no less embarrassing than having a negative unique variance in $\mathbf{\Psi}$. Nevertheless, popular methods of common factor analysis generally ignore the constraint that $\mathbf{\Sigma}_c - \mathbf{F}\mathbf{F}'$ must be positive semidefinite. The only exception seems to be Minimum Rank Factor Analysis (MRFA). This method was originally proposed by ten Berge and Kiers (1991) as AMRFA, but the "A" of *approximate* has worn off in the meantime. MRFA offers a decomposition of $\mathbf{\Sigma}$ that satisfies (3), with both $\mathbf{\Psi}$ and $\mathbf{\Sigma}_c - \mathbf{F}\mathbf{F}'$ positive semidefinite. Subject to these two constraints, MRFA constructs the solution that *minimizes the unexplained common variance for any fixed number of factors r*. In other words, MRFA approximates the ideal of low reduced rank by minimizing the amount of common variance that is left unexplained when as few as r factors are maintained.

For a $p \times p$ symmetric matrix $\mathbf{S}$ we denote by $\lambda_1(\mathbf{S}) \geq \cdots \geq \lambda_p(\mathbf{S})$ its eigenvalues arranged in decreasing order. Formally, MRFA minimizes, for fixed r, the function

$$f_{\text{mrfa}}(\mathbf{\Psi}) := \sum_{i=r+1}^{p} \lambda_i(\mathbf{\Sigma} - \mathbf{\Psi}) \tag{4}$$

subject to $\mathbf{\Sigma} - \mathbf{\Psi} \succeq \mathbf{0}$ and $\mathbf{\Psi} \succeq \mathbf{0}$. This is very similar to MINRES/IPFA/ULS, where the function

$$f_{\text{minres}}(\mathbf{\Psi}) := \sum_{i=r+1}^{p} \lambda_i^2(\mathbf{\Sigma} - \mathbf{\Psi}) \tag{5}$$

is minimized, without any constraint on the sign of these eigenvalues (Harman & Jones, 1966, Jöreskog, 1967). A similar eigenvalue interpretation of Maximum Likelihood Factor Analysis has been given by Jöreskog (1967, p. 449), also see ten Berge (1998) for a discussion.

A key feature of (3) is the distinction between communalities as variances "to be explained" on the one hand, and the explained variances of the variables, the diagonal elements of $\mathbf{F}\mathbf{F}'$ on the other. The difference rests in the unexplained parts of the communalities, in the diagonal of $\mathbf{\Sigma}_c - \mathbf{F}\mathbf{F}'$. When this distinction is preserved, it is possible to evaluate to what extent the common variance is accounted for by the common factors. This can be expressed as "the percentage of explained common variance", analogous to the percentage of explained observed variance in Principal Component Analysis. So far, MRFA is the only method that solves (3) subject to its constraints. Hence it is the only method so far that yields a percentage of explained common variance, to guide decisions about the number of factors to retain. Consider, for instance, the situation where only one common factor is hypothesized to account for the correlations between variables. MRFA will give the smallest possible percentage of common variance that is left unex-

plained under the one-factor hypothesis. This means that MRFA evaluates the extent to which a one factor hypothesis is untenable, under the most favorable conditions for the that very hypothesis, namely, the MRFA communalitites with $r = 1$. The same logic applies to r hypothesized factors in general.

In practical applications, MRFA decomposes the sample analog of $\mathbf{\Sigma}$. A key question, therefore, is the extent to which the sample statistics in general, and the explained common variance in particular, are reliable. The present paper deals with the bias of the explained common variance (ECV) with MRFA, or, more precisely, with the bias of its counterpart, the "unexplained common variance" (UCV).

The organization of this paper is as follows. First, an example of MRFA is given, to demonstrate the concept of explained common variance. Next, the algorithm of MRFA is explained in some detail, to set the stage for asymptotic theory. Then the asymptotic theory for the bias of UCV is developed. It is shown, in particular, that the UCV is asymptotically unbiased when the r-factor hypothesis is correct. In addition, however, the asymptotic bias is also established under misspecification of the number of factors. Also, the asymptotic covariance matrix of the unique variances is derived. Finally, some numerical results are given which demonstrate the accuracy of the asymptotics. We start with an example.

Example 1: The Schutz data. The covariance matrix $\mathbf{\Sigma}$ for nine tests (Schutz, 1958; also see Carroll, 1993, p. 94) is given in Table 1. The 9 eigenvalues of the reduced correlation matrix, from MRFA, are given in Table 2, when the number r of common factors is 2, 3, and 4, respectively. Note that in this example we assume that the corresponding variables are standardized to have variance one, that is, the matrix $\mathbf{\Sigma}$ is in fact a correlation matrix. Note, however, that the sample covariance matrix $\mathbf{S}$, used in the asymptotic theory of section 3, is the usual unbiased estimate of $\mathbf{\Sigma}$ and its diagonal entries may be different from one.

Let us first look at the solution for $r = 2$. MRFA minimizes the UCV, defined as $\lambda_3 + \cdots + \lambda_9$, and the minimum value is 1.1741. This means that, among all solutions for $\mathbf{\Psi}$ such that $\mathbf{\Psi}$ and $\mathbf{\Sigma} - \mathbf{\Psi}$ are positive semidefinite, there is no solution with lower UCV than 1.1741. Because the total common variance (the sum of communalitites, and also the sum of the nine reduced eigenvalues) is 6.2730, the percentage of UCV is $100 \times 1.1741/6.2730 = 18.72\%$. Hence the explained common variance is 81.28%.

Next, consider the solution with $r = 3$. Now the UCV is $\lambda_4 + \cdots + \lambda_9$ and its minimum is .4296. This amounts to a percentage ECV of 93.16%. The solution with 4 factors reaches an ECV as high as 99.1%. Carroll (1993) went to considerable length explaining why, for these data, a four factor solution was needed to fully account for the data. The ECV of 99.1%, compared to the 93.16 per cent when $r = 3$, fully supports his decision. To demonstrate the difference between MRFA, see (4), and MINRES, see (5), we also report Varimax rotated factor loadings and "communalities" for both methods for the two factor solutions. Table 3 gives these values for the MRFA solution.

TABLE 1.
The correlation matrix for the Schutz data

1.00	0.80	0.28	0.29	0.41	0.38	0.44	0.40	0.41
0.80	1.00	0.31	0.33	0.49	0.44	0.50	0.44	0.46
0.28	0.31	1.00	0.71	0.32	0.34	0.41	0.41	0.30
0.29	0.33	0.71	1.00	0.32	0.36	0.42	0.41	0.31
0.41	0.49	0.32	0.32	1.00	0.77	0.50	0.39	0.37
0.38	0.44	0.34	0.36	0.77	1.00	0.48	0.35	0.37
0.44	0.50	0.41	0.42	0.50	0.48	1.00	0.56	0.48
0.40	0.44	0.41	0.41	0.39	0.35	0.56	1.00	0.46
0.41	0.46	0.30	0.31	0.37	0.37	0.48	0.46	1.00

TABLE 2.
Reduced MRFA eigenvalues for the 2, 3, and 4 factors solution

Eigenvalues of $\mathbf{\Sigma} - \mathbf{\Psi}$	$r = 2$	$r = 3$	$r = 4$
λ_1	4.1550	4.1563	4.1572
λ_2	0.9439	0.9425	0.9481
λ_3	0.7417	0.7500	0.7472
λ_4	0.3726	0.3709	0.3751
λ_5	0.0457	0.0458	0.0442
λ_6	0.0142	0.0129	0.0125
λ_7	0.0000	0.0000	0.0000
λ_8	0.0000	−0.0000	−0.0000
λ_9	−0.0000	−0.0000	−0.0000
Common Variance	6.2731	6.2785	6.2843
Unexplained	1.1741	.4296	.0567
% Explained	81.28	93.16	99.10

TABLE 3.
The two factor solution for MRFA

MRFA loadings (rotated)		ECV	communalities
.7446	.1389	.5737	.6769
.8918	.1283	.8117	.9752
.1728	.7680	.6197	.6472
.1571	.8460	.7404	.7946
.6677	.3112	.5426	.7749
.6268	.3552	.5190	.8250
.5543	.4536	.5130	.5844
.4619	.4675	.4319	.5736
.4988	.3129	.3467	.4213
		5.0987	6.2731

It is clear that, for each variable, the communality exceeds the explained part ECV, and the percentage ECV of 81.28 shows up in this table as the sum of the nine ECV values 5.0987 divided by the sum of the nine communalities 6.2731. For Minres, the loadings, sums of squared loadings per variable, and reduced eigenvalues are given in Table 4.

Table 4 reveals that the MINRES loadings are very similar to those of MRFA. The real difference rests in the communalities. In common factor analysis, it is customary to call the sum of squared loadings per variable "the communality" of the variable. This is correct in cases of perfect fit, but leads to paradoxical results in cases of less than perfect fit. Specifically, the reduced eigenvalues are partly negative, when sums of squared loadings are inserted in the diagonal of the correlation matrix. In particular, when r factors are maintained, the last $p - r$ eigenvalues sum to zero (Harman, 1967, p. 195) also see the last column of Table 4. This implies that some of these are negative, unless they are all zero (perfect fit). It follows that the sum of squared loadings is not the communality of a variable, consistent with (3). But neither is it the explained variance of that variable: If we sums the sums of squared loadings over variables, we get the same as the sum of the sum of squared loadings per factor, implying that, for any number of factors, 100% of common variance would be explained, also see ten Berge (2000). It can be concluded that the

TABLE 4.
The two factor solution for MINRES

Minres loadings (rotated)		ssq	reduced eigenvalues
0.7334	0.1443	0.5587	3.9812
0.8129	0.1706	0.6899	0.8372
0.1702	0.8067	0.6798	0.4553
0.1904	0.7982	0.6733	0.2368
0.6321	0.2931	0.4854	−0.0327
0.5832	0.3245	0.4454	−0.0568
0.5829	0.4221	0.5180	−0.1125
0.4834	0.4228	0.4124	−0.1794
0.5219	0.2882	0.3554	−0.3106
ssq = 2.8494	ssq = 1.9690	sum = 4.8184	$\lambda_1 + \lambda_2 = 3.9812 + .8372$

main difference between MRFA and MINRES is the property of the former to yield meaningful communalities, allowing an evaluation of the (percentage of) ECV.

1.1. The Computational Solution of MRFA

A first description of the computations that produce the MRFA solution can be found in ten Berge and Kiers (1991). Here, some more detail is added, to set the stage for asymptotic theory. Minimizing the sum of the last $p - r$ eigenvalues of $\mathbf{\Sigma} - \mathbf{\Psi}$ can be written equivalently as minimizing the trace function

$$g(\mathbf{\Psi}, \mathbf{E}) := \operatorname{tr}\left[\mathbf{E}'(\mathbf{\Sigma} - \mathbf{\Psi})\mathbf{E}\right]$$

subject to $\mathbf{\Sigma} - \mathbf{\Psi} \succeq \mathbf{0}$, $\mathbf{\Psi} \succeq \mathbf{0}$, and $\mathbf{E}'\mathbf{E} = \mathbf{I}_{p-r}$. For any given $\mathbf{\Psi}$, the best choice for $\mathbf{E}$ is a columnwise orthonormal matrix of eigenvectors of $\mathbf{\Sigma} - \mathbf{\Psi}$, which correspond to the $p - r$ smallest eigenvalues. Conversely, for any given $\mathbf{E}$, the best choice for $\mathbf{\Psi}$ is to minimize $-\operatorname{tr}(\mathbf{E}'\mathbf{\Psi}\mathbf{E}) = -\operatorname{tr}(\mathbf{D}\mathbf{\Psi})$, where $\mathbf{D} := \operatorname{diag}(\mathbf{E}\mathbf{E}')$. Equivalently, $\mathbf{\Psi}$ has to yield the minimum of $\operatorname{tr}\left(\mathbf{D}^{1/2}\mathbf{\Sigma}\mathbf{D}^{1/2} - \mathbf{D}^{1/2}\mathbf{\Psi}\mathbf{D}^{1/2}\right)$ subject to $\mathbf{\Sigma} - \mathbf{\Psi} \succeq \mathbf{0}$ and $\mathbf{\Psi} \succeq \mathbf{0}$. The latter problem is a weighted CMTFA problem, which can be solved by an iterative algorithm (Bentler & Woodward, 1980; ten Berge, Snijders and Zegers, 1981). The MRFA algorithm consists of alternatively updating $\mathbf{\Psi}$ by solving this weighted CMTFA problem, and $\mathbf{E}$ by taking the last eigenvectors of $\mathbf{\Sigma} - \mathbf{\Psi}$.

Upon convergence of the MRFA algorithm, and assuming that the constraint $\mathbf{\Psi} \succeq \mathbf{0}$ is inactive, we have a solution $\mathbf{W} = \mathbf{D}^{1/2}\mathbf{\Psi}\mathbf{D}^{1/2}$ for the weighted CMTFA problem, and a $p \times p$ matrix $\mathbf{T}$ with rows sums of squares equal to 1, spanning the null-space of $(\mathbf{D}^{1/2}\mathbf{\Sigma}\mathbf{D}^{1/2} - \mathbf{W})$. It follows that $(\mathbf{\Sigma} - \mathbf{\Psi})\mathbf{D}^{1/2}\mathbf{T} = \mathbf{0}$. Let $\mathbf{\Omega}$ be defined as $\mathbf{\Omega} := \mathbf{D}^{1/2}\mathbf{T}\mathbf{T}'\mathbf{D}^{1/2}$, with the same diagonal as $\mathbf{D}$.

Upon convergence, we also have an eigenvector matrix $\mathbf{E}$ such that $\operatorname{diag}(\mathbf{E}\mathbf{E}') = \mathbf{D}$. For future reference, it is important to notice that, when $\mathbf{\Sigma} - \mathbf{\Psi}$ has precisely rank $p - r$, $\mathbf{E}\mathbf{E}'$ and $\mathbf{\Omega}$ are equal. Otherwise, these matrices only share their diagonal elements.

2. Semidefinite Programming

In this section we discuss some general results from the theory of so-called semidefinite programming problems. For a thorough discussion of that topic and rigorous derivations of the following results the interested reader is referred to the recently published books: Saigal, Vandenberghe and Wolkowicz (2000) and Bonnans and Shapiro (2000).

We use the following notation throughout the paper. By $\mathbf{A}^\dagger$ we denote the Moore-Penrose pseudo-inverse of matrix $\mathbf{A}$, $\mathbf{A} * \mathbf{B}$ denotes the Hadamard (i.e., term by term) product of matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{I}_p$ denotes the $p \times p$ identity matrix, vec($\mathbf{A}$) denotes vector obtained by stacking columns of matrix $\mathbf{A}$, diag($\mathbf{H}$) denotes the vector formed by diagonal elements of matrix $\mathbf{H}$, tr($\mathbf{H}$) denotes the trace of matrix $\mathbf{H}$.

Let us consider the following optimization problem:

$$\operatorname*{Min}_{\mathbf{x} \in I\!R^m_+} f(\mathbf{x}) \text{ subject to } \mathbf{G}(\mathbf{x}) \succeq \mathbf{0}. \tag{6}$$

Here $f(\mathbf{x})$ is a real valued function of $\mathbf{x} \in I\!R^m$, $\mathbf{G}(\mathbf{x})$ is a mapping from $I\!R^m$ into the space $\mathcal{S}^p$ of $p \times p$ symmetric matrices, and

$$I\!R^m_+ := \left\{ \mathbf{x} \in I\!R^m : x_i \geq 0, i = 1, \ldots, m \right\}.$$

Problems of the form (6) are called (nonlinear) semidefinite programming problems. For the sake of simplicity we consider the case where the constraint mapping $\mathbf{G}(\mathbf{x})$ is affine, i.e., $\mathbf{G}(\mathbf{x}) := \mathbf{A}_0 + \sum_{i=1}^m x_i \mathbf{A}_i$ with $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_m$ being given $p \times p$ symmetric matrices. MRFA can be considered in that framework if we use the objective function $f_{\text{mrfa}}(\cdot)$ and the constraint mapping $\mathbf{G}(\mathbf{x}) := \mathbf{\Sigma} - \mathbf{X}$, where $\mathbf{X}$ is a diagonal matrix and $\mathbf{x} := \text{diag}(\mathbf{X})$. We refer to such mapping $\mathbf{G}(\mathbf{x})$ as FA-*mapping*.

Consider the set $\mathcal{W}_s$ of $p \times p$ symmetric matrices of rank s. The set $\mathcal{W}_s$ forms a smooth manifold in the linear space $\mathcal{S}^p$. We denote by $T_{\mathcal{W}_s}(\mathbf{A})$ the tangent space to $\mathcal{W}_s$ at $\mathbf{A} \in \mathcal{W}_s$. A point $\bar{\mathbf{x}} \in I\!R^m$ is said to be a *feasible* point of problem (6) if it satisfies the corresponding constraints, i.e., $\mathbf{G}(\bar{\mathbf{x}}) \succeq \mathbf{0}$ and $\bar{\mathbf{x}} \in I\!R^m_+$. It is said that a feasible point $\bar{\mathbf{x}}$ is *nondegenerate* if

$$\mathcal{L}(\bar{\mathbf{x}}) + T_{\mathcal{W}_s}(\bar{\mathbf{A}}) = \mathcal{S}^p, \tag{7}$$

where $s := \text{rank}\, \mathbf{G}(\bar{\mathbf{x}})$, $\bar{\mathbf{A}} := \mathbf{G}(\bar{\mathbf{x}})$, $I(\bar{\mathbf{x}}) := \{i : \bar{x}_i = 0, i = 1, \ldots, m\}$ and

$$\mathcal{L}(\bar{\mathbf{x}}) := \left\{ \mathbf{Z} \in \mathcal{S}^p : \mathbf{Z} = \sum_{i=1}^m x_i \mathbf{A}_i, x_i = 0, i \in I(\bar{\mathbf{x}}) \right\}.$$

Note that both $\mathcal{L}(\bar{\mathbf{x}})$ and $T_{\mathcal{W}_s}(\bar{\mathbf{A}})$ are linear subspaces of $\mathcal{S}^p$, and that if $I(\bar{\mathbf{x}})$ is empty, i.e., all components of vector $\bar{\mathbf{x}}$ are positive, then $\mathcal{L}(\bar{\mathbf{x}}) = D\mathbf{G}(\bar{\mathbf{x}}) I\!R^m$. Here $D\mathbf{G}(\bar{\mathbf{x}})\mathbf{h} = \sum_{i=1}^m h_i \mathbf{A}_i$ is the differential of the mapping $\mathbf{G}$ and $D\mathbf{G}(\bar{\mathbf{x}}) I\!R^m$ is the image of the mapping $D\mathbf{G}(\bar{\mathbf{x}}) : I\!R^m \to \mathcal{S}^p$.

It is known that

$$T_{\mathcal{W}_s}(\bar{\mathbf{A}}) = \left\{ \mathbf{Z} \in \mathcal{S}^p : \mathbf{\Xi}' \mathbf{Z} \mathbf{\Xi} = \mathbf{0} \right\}, \tag{8}$$

where $\mathbf{\Xi} = [\boldsymbol{\xi}_{s+1}, \ldots, \boldsymbol{\xi}_p]$ is a $p \times (p - s)$ matrix whose column vectors $\boldsymbol{\xi}_{s+1}, \ldots, \boldsymbol{\xi}_p$ form a basis of the null space of the matrix $\bar{\mathbf{A}} = \mathbf{G}(\bar{\mathbf{x}})$. We refer to $\mathbf{\Xi}$ as a *complement* of the matrix $\bar{\mathbf{A}}$. Note that although the complement matrix $\mathbf{\Xi}$ is not unique, the space given in the right hand side of the equation (8) is defined uniquely. In particular, one can take $\boldsymbol{\xi}_{s+1}, \ldots, \boldsymbol{\xi}_p$ to be a set of orthonormal eigenvectors of $\bar{\mathbf{A}}$ corresponding to its zero eigenvalue.

By using this description of the tangent space it can be shown that condition (7) is equivalent to the following condition. For $s + 1 \leq k \leq \ell \leq p$ consider the $(m - |I(\bar{\mathbf{x}})|)$-dimensional vectors with components $\boldsymbol{\xi}_k' \mathbf{A}_i \boldsymbol{\xi}_\ell$, $i \in \{1, \ldots, m\} \setminus I(\bar{\mathbf{x}})$. Then $\bar{\mathbf{x}}$ is nondegenerate iff these vectors are linearly independent. Note that the total number of such vectors is $(p-s)(p-s+1)/2$. Therefore a necessary condition for $\bar{\mathbf{x}}$ to be nondegenerate is that

$$\frac{(p-s)(p-s+1)}{2} \leq m - |I(\bar{\mathbf{x}})|, \tag{9}$$

where $|I(\bar{\mathbf{x}})|$ denotes the number of elements in the set $I(\bar{\mathbf{x}})$.

In particular, let $\mathbf{G}(\mathbf{x})$ be the FA-mapping, and hence $\mathbf{G}(\overline{\mathbf{x}}) = \boldsymbol{\Sigma} - \overline{\mathbf{X}}$. Suppose that $\operatorname{rank}(\boldsymbol{\Sigma} - \overline{\mathbf{X}}) = s$. Then $\boldsymbol{\xi}_{s+1}, \ldots, \boldsymbol{\xi}_p$ can be any set of linearly independent vectors such that $(\boldsymbol{\Sigma} - \overline{\mathbf{X}})\boldsymbol{\xi}_i = \mathbf{0}$, $i = s+1, \ldots, p$. We have then that $\overline{\mathbf{x}}$ is nondegenerate iff vectors $\boldsymbol{\xi}_k * \boldsymbol{\xi}_\ell$, $s+1 \leq k \leq \ell \leq p$, with components corresponding to $I(\overline{\mathbf{x}})$ deleted, are linearly independent. Therefore a necessary condition for the nondegeneracy is that

$$\frac{(p-s)(p-s+1)}{2} \leq p - |I(\overline{\mathbf{x}})|. \tag{10}$$

If the set $I(\overline{\mathbf{x}})$ is empty, and hence $|I(\overline{\mathbf{x}})| = 0$, then (10) is equivalent to

$$\frac{2p+1-(8p+1)^{1/2}}{2} \leq s. \tag{11}$$

The above lower bound for the rank s is called the Ledermann bound.

Consider the Lagrangian function

$$L(\mathbf{x}, \boldsymbol{\Omega}) := f(\mathbf{x}) - \operatorname{tr}\left[\boldsymbol{\Omega}\mathbf{G}(\mathbf{x})\right] \tag{12}$$

associated with the problem (6). We have that if $\overline{\mathbf{x}}$ is a locally optimal solution of (6), and a constraint qualification holds, then the following first-order necessary conditions are satisfied: there exists a matrix $\overline{\boldsymbol{\Omega}} \in \mathcal{S}^p$ such that

$$\frac{\partial L(\overline{\mathbf{x}}, \overline{\boldsymbol{\Omega}})}{\partial x_i} = 0, i \in \{1, \ldots, m\} \setminus I(\overline{\mathbf{x}}), \tag{13}$$

$$\frac{\partial L(\overline{\mathbf{x}}, \overline{\boldsymbol{\Omega}})}{\partial x_i} \geq 0, i \in I(\overline{\mathbf{x}}), \tag{14}$$

$$\overline{\boldsymbol{\Omega}}\mathbf{G}(\overline{\mathbf{x}}) = \mathbf{0}, \quad \overline{\boldsymbol{\Omega}} \succeq \mathbf{0}. \tag{15}$$

We refer to a matrix $\overline{\boldsymbol{\Omega}}$ satisfying the above conditions (13)–(15) as a Lagrange multipliers matrix and denote by $\Lambda(\overline{\mathbf{x}})$ the set of all such matrices. If the so-called *Slater constraint qualification* is satisfied, that is, there exists $\mathbf{x} \in \mathbb{R}^m_+$ such that the matrix $\mathbf{G}(\mathbf{x})$ is positive definite, then the set $\Lambda(\overline{\mathbf{x}})$ of Lagrange multipliers matrices is nonempty and bounded.

Let us remark that if $\mathbf{G}(\mathbf{x})$ is the FA-mapping, then

$$\partial L(\mathbf{x}, \boldsymbol{\Omega})/\partial x_i = \partial f(\mathbf{x})/\partial x_i + \Omega_{ii}, \quad i = 1, \ldots, p. \tag{16}$$

Moreover, if the corresponding matrix $\boldsymbol{\Sigma}$ is positive definite, then the Slater constraint qualification holds, and hence $\Lambda(\overline{\mathbf{x}})$ is nonempty and bounded. Also if the index set $I(\overline{\mathbf{x}})$ is empty, then conditions (13)–(14) become $\nabla f(\overline{\mathbf{x}}) = -\operatorname{diag}(\overline{\boldsymbol{\Omega}})$.

It is said that the *strict complementarity* condition holds at $\overline{\mathbf{x}}$ if, for some $\overline{\boldsymbol{\Omega}} \in \Lambda(\overline{\mathbf{x}})$, the following two conditions are satisfied:

$$\operatorname{rank}(\overline{\boldsymbol{\Omega}}) + \operatorname{rank}(\mathbf{G}(\overline{\mathbf{x}})) = p, \text{ and} \tag{17}$$

$$\frac{\partial L(\overline{\mathbf{x}}, \overline{\boldsymbol{\Omega}})}{\partial x_i} \neq 0, \quad i \in I(\overline{\mathbf{x}}). \tag{18}$$

Note that because of (14), Condition (18) is equivalent to $\partial L(\overline{\mathbf{x}}, \overline{\boldsymbol{\Omega}})/\partial x_i > 0$, $i \in I(\overline{\mathbf{x}})$. Also, if $I(\overline{\mathbf{x}})$ is empty, then only Condition (17) in the above definition is needed.

We have that if $\overline{\mathbf{x}}$ is nondegenerate, then $\Lambda(\overline{\mathbf{x}})$ is a singleton, that is, the corresponding Lagrange multipliers matrix is unique. Conversely, if $\Lambda(\overline{\mathbf{x}})$ is a singleton and the strict complementarity condition holds, then the point $\overline{\mathbf{x}}$ is nondegenerate.

Let us discuss now second-order optimality conditions. For the sake of simplicity we consider only the case where the set $I(\bar{\mathbf{x}})$ is empty, that is, $\bar{x}_i > 0$ for all $i = 1, \ldots, m$. Let $\overline{\boldsymbol{\Omega}} \in \Lambda(\bar{\mathbf{x}})$ be a Lagrange multipliers matrix. Because of the complementarity condition (15), the matrix $\overline{\boldsymbol{\Omega}}$ can be written in the form $\overline{\boldsymbol{\Omega}} = \boldsymbol{\Xi}_1 \boldsymbol{\Xi}_1'$, where $\boldsymbol{\Xi}_1$ is a matrix of full column rank equal to the rank of $\overline{\boldsymbol{\Omega}}$ and such that $\mathbf{G}(\bar{\mathbf{x}})\boldsymbol{\Xi}_1 = \mathbf{0}$. If the strict complementarity condition (17) holds, then $\operatorname{rank}(\overline{\boldsymbol{\Omega}}) = p - s$, and hence in that case $\boldsymbol{\Xi}_1$ forms a complement of $\mathbf{G}(\bar{\mathbf{x}})$. Otherwise one can choose a complement matrix $\boldsymbol{\Xi}$ (of the matrix $\mathbf{G}(\bar{\mathbf{x}})$) in such a way that $\overline{\boldsymbol{\Omega}} = \boldsymbol{\Xi}_1 \boldsymbol{\Xi}_1'$, where $[\boldsymbol{\Xi}_1, \boldsymbol{\Xi}_2]$ is a partition of $\boldsymbol{\Xi}$.

With the point $\bar{\mathbf{x}}$ is associated the so-called *critical cone* $C(\bar{\mathbf{x}})$, which can be written as follows

$$C(\bar{\mathbf{x}}) = \left\{ \mathbf{h} \in I\!R^m : \begin{array}{ll} \sum_{i=1}^m h_i \boldsymbol{\Xi}_1' \mathbf{A}_i \boldsymbol{\Xi}_1 = \mathbf{0}, & \sum_{i=1}^m h_i \boldsymbol{\Xi}_1' \mathbf{A}_i \boldsymbol{\Xi}_2 = \mathbf{0}, \\ \sum_{i=1}^m h_i \boldsymbol{\Xi}_2' \mathbf{A}_i \boldsymbol{\Xi}_2 \preceq \mathbf{0} & \end{array} \right\}. \tag{19}$$

In particular, if the strict complementarity Condition (17) holds, then

$$C(\bar{\mathbf{x}}) = \left\{ \mathbf{h} \in I\!R^m : \sum_{i=1}^m h_i \boldsymbol{\Xi}' \mathbf{A}_i \boldsymbol{\Xi} = \mathbf{0} \right\}, \tag{20}$$

and hence in that case $C(\bar{\mathbf{x}})$ is a linear space.

Consider the $m \times m$ matrix $\mathbf{H}(\bar{\mathbf{x}}, \boldsymbol{\Omega})$ with typical elements

$$[\mathbf{H}(\bar{\mathbf{x}}, \boldsymbol{\Omega})]_{ij} := 2\operatorname{tr}\left[\boldsymbol{\Omega} \mathbf{A}_i \overline{\mathbf{A}}^{\dagger} \mathbf{A}_j\right], \qquad i, j = 1, \ldots, m, \tag{21}$$

where $\overline{\mathbf{A}} := \mathbf{G}(\bar{\mathbf{x}})$. Under the Slater constraint qualification, the conditions

$$\sup_{\boldsymbol{\Omega} \in \Lambda(\bar{\mathbf{x}})} \mathbf{h}'\big(\nabla^2 f(\bar{\mathbf{x}}) + \mathbf{H}(\bar{\mathbf{x}}, \boldsymbol{\Omega})\big)\mathbf{h} \geq 0, \quad \text{for all } \mathbf{h} \in C(\bar{\mathbf{x}}), \tag{22}$$

are necessary, and conditions

$$\sup_{\boldsymbol{\Omega} \in \Lambda(\bar{\mathbf{x}})} \mathbf{h}'\big(\nabla^2 f(\bar{\mathbf{x}}) + \mathbf{H}(\bar{\mathbf{x}}, \boldsymbol{\Omega})\big)\mathbf{h} > 0, \quad \text{for all } \mathbf{h} \in C(\bar{\mathbf{x}}) \setminus \{\mathbf{0}\}, \tag{23}$$

are sufficient for local optimality of $\bar{\mathbf{x}}$. Note that the only difference between the second-order necessary conditions (22) and the corresponding sufficient conditions (23) is that the strict inequality sign is used in (23). The above second-order conditions can be also applied in cases where the constraint mapping $\mathbf{G}(\mathbf{x})$ is not necessarily linear. In such cases the function $f(\bar{\mathbf{x}})$ in (22) and (23) should be replaced by the Lagrangian $L(\bar{\mathbf{x}}, \boldsymbol{\Omega})$.

Recall that if the nondegeneracy and strict complementarity conditions hold, then $\Lambda(\bar{\mathbf{x}}) = \{\overline{\boldsymbol{\Omega}}\}$ is a singleton and $C(\bar{\mathbf{x}})$ is a linear space. In that case the second-order sufficient conditions (23) mean that the matrix $\nabla^2 f(\bar{\mathbf{x}}) + \mathbf{H}(\bar{\mathbf{x}}, \overline{\boldsymbol{\Omega}})$ is positive definite over the space $C(\bar{\mathbf{x}})$. In the case of the FA-mapping, the matrix $\mathbf{H}(\bar{\mathbf{x}}, \boldsymbol{\Omega})$ becomes

$$\mathbf{H}(\bar{\mathbf{x}}, \boldsymbol{\Omega}) = 2\,\boldsymbol{\Omega} * \overline{\mathbf{A}}^{\dagger}. \tag{24}$$

Consider the MRFA function

$$f_{\text{mrfa}}(\mathbf{x}) := \sum_{i=r+1}^{p} \lambda_i(\boldsymbol{\Sigma} - \mathbf{X}). \tag{25}$$

This function is differentiable at $\overline{\mathbf{x}}$ if and only if

$$\lambda_r(\boldsymbol{\Sigma} - \overline{\mathbf{X}}) > \lambda_{r+1}(\boldsymbol{\Sigma} - \overline{\mathbf{X}}), \tag{26}$$

in which case

$$\nabla f_{\text{mrfa}}(\overline{\mathbf{x}}) = -\operatorname{diag}\left(\sum_{i=r+1}^{p} \mathbf{e}_i \mathbf{e}_i'\right) = -\sum_{i=r+1}^{p} \mathbf{e}_i * \mathbf{e}_i, \tag{27}$$

and

$$\nabla^2 f_{\text{mrfa}}(\overline{\mathbf{x}}) = 2 \sum_{(i,j)\in\mathcal{I}} \frac{(\mathbf{e}_i * \mathbf{e}_j)(\mathbf{e}_i * \mathbf{e}_j)'}{\lambda_j - \lambda_i}, \tag{28}$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_p$ is a set of orthonormal eigenvectors of $\boldsymbol{\Sigma} - \overline{\mathbf{X}}$ corresponding to the eigenvalues $\lambda_i = \lambda_i(\boldsymbol{\Sigma} - \overline{\mathbf{X}})$, and

$$\mathcal{I} := \{(i, j) : i = 1, \ldots, r, j = r + 1, \ldots, p\}. \tag{29}$$

Note that the function $f_{\text{mrfa}}(\cdot)$ is concave and the Hessian matrix $\nabla^2 f_{\text{mrfa}}(\overline{\mathbf{x}})$ is negative semidefinite.

3. Asymptotics of Factor Analysis Models

Consider the following optimization problem:

$$\operatorname*{Min}_{\mathbf{x}\in\mathbb{R}^m_+} \phi(\mathbf{x}, \mathbf{z}) \quad \text{subject to } \mathbf{Z} - \mathbf{X} \succeq \mathbf{0}, \tag{30}$$

where $\mathbf{X}$ is a $p \times p$ diagonal matrix, $\mathbf{x} := \operatorname{diag}(\mathbf{X})$, $\mathbf{Z}$ is a $p \times p$ symmetric matrix, $\mathbf{z} := \operatorname{vec}(\mathbf{Z})$ and $\phi(\mathbf{x}, \mathbf{z})$ is a continuous real valued function. We denote $\boldsymbol{\sigma} := \operatorname{vec}(\boldsymbol{\Sigma})$ and assume that for $\mathbf{Z} = \boldsymbol{\Sigma}$ the function $\phi(\cdot, \boldsymbol{\sigma})$ coincides with the function $f(\cdot)$ used in the previous section. Consequently, for $\mathbf{Z} = \boldsymbol{\Sigma}$ and the FA-mapping the above problem (30) coincides with the problem (6). By $\vartheta(\mathbf{z})$ and $\overline{\mathbf{x}}(\mathbf{z})$ we denote the optimal value and an optimal solution, respectively, of problem (30).

Now let $\mathbf{S}$ be the sample covariance matrix based on a sample of size n, and $\mathbf{s} := \operatorname{vec} \mathbf{S}$. For $\mathbf{Z} = \mathbf{S}$ we refer to (30) as the *sample* FA-problem, and we refer to (6) as the true (or population) FA-problem. In particular, for the function

$$\phi_{\text{mrfa}}(\mathbf{x}, \mathbf{z}) := \sum_{i=r+1}^{p} \lambda_i(\mathbf{Z} - \mathbf{X}) \tag{31}$$

we refer to (30) as the true (or population) MRFA problem for $\mathbf{Z} = \boldsymbol{\Sigma}$, and the sample MRFA problem for $\mathbf{Z} = \mathbf{S}$. The optimal value $\widehat{\vartheta} = \vartheta(\mathbf{s})$ and an optimal solution $\hat{\boldsymbol{\psi}} = \overline{\mathbf{x}}(\mathbf{s})$ of the sample problem give estimators of their true (population) counterparts $\vartheta_0 = \vartheta(\boldsymbol{\sigma})$ and $\boldsymbol{\psi}_0 = \overline{\mathbf{x}}(\boldsymbol{\sigma})$, respectively.

In this section we investigate asymptotic properties of $\widehat{\vartheta} = \vartheta(\mathbf{s})$ and $\hat{\boldsymbol{\psi}} = \overline{\mathbf{x}}(\mathbf{s})$. We assume throughout the paper that the population covariance matrix $\boldsymbol{\Sigma}$ is *nonsingular* and hence is positive definite. Clearly the asymptotic properties of the sample estimators $\widehat{\vartheta}$ and $\hat{\boldsymbol{\psi}}$ are closely related to the continuity and differentiablity properties of the functions $\vartheta(\cdot)$ and $\overline{\mathbf{x}}(\cdot)$. We refer to Bonnans and Shapiro (2000) for a rigorous derivation of the following properties of $\vartheta(\cdot)$ and $\overline{\mathbf{x}}(\cdot)$.

Since for all $\mathbf{S}$ in a neighborhood of $\boldsymbol{\Sigma}$ the set of feasible $\mathbf{x}$ of the problem (30) is bounded and $\phi(\cdot, \cdot)$ is continuous, we have that the optimal value function $\vartheta(\cdot)$ is continuous at $\boldsymbol{\sigma}$. Since

$\mathbf{S}$ is a consistent estimator of $\boldsymbol{\Sigma}$, it follows that $\widehat{\vartheta}$ is a consistent estimator of ϑ_0. Moreover, if $\boldsymbol{\psi}_0 = \bar{\mathbf{x}}(\boldsymbol{\sigma})$ is the *unique* optimal solution of the true problem (6), then $\bar{\mathbf{x}}(\mathbf{z})$ converges to $\bar{\mathbf{x}}(\boldsymbol{\sigma})$ as $\mathbf{z} \to \boldsymbol{\sigma}$, and hence $\hat{\boldsymbol{\psi}}$ is a consistent estimator of $\boldsymbol{\psi}_0$. By $\hat{\boldsymbol{\Psi}}$ and $\boldsymbol{\Psi}_0$ we denote the corresponding diagonal matrices, that is, $\hat{\boldsymbol{\psi}} = \operatorname{diag}(\hat{\boldsymbol{\Psi}})$ and $\boldsymbol{\psi}_0 = \operatorname{diag}(\boldsymbol{\Psi}_0)$.

Proposition 1. Suppose that the true problem has a unique optimal solution $\boldsymbol{\psi}_0 = \bar{\mathbf{x}}(\boldsymbol{\sigma})$ to which corresponds a unique Lagrange multipliers matrix $\overline{\boldsymbol{\Omega}}$, and that the function $\phi(\cdot, \cdot)$ is continuously differentiable in a neighborhood of $(\boldsymbol{\psi}_0, \boldsymbol{\sigma})$. Then the optimal value function $\vartheta(\cdot)$ is differentiable at $\boldsymbol{\sigma}$ and

$$\nabla\vartheta(\boldsymbol{\sigma}) = \nabla_z L(\boldsymbol{\psi}_0, \boldsymbol{\sigma}, \overline{\boldsymbol{\Omega}}), \tag{32}$$

where $L(\mathbf{x}, \mathbf{z}, \boldsymbol{\Omega}) := \phi(\mathbf{x}, \mathbf{z}) - \operatorname{tr}\left[\boldsymbol{\Omega}\mathbf{G}(\mathbf{x})\right]$.

Note that for the FA-mapping $\mathbf{G}(\mathbf{x})$ we have that

$$\nabla_z L(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\Omega}) = \nabla_z \phi(\mathbf{x}, \boldsymbol{\sigma}) - \operatorname{vec}(\boldsymbol{\Omega}). \tag{33}$$

Moreover, consider the MRFA function defined in (31) and suppose that Condition (26) holds. Then ϕ_{mrfa} is continuously differentiable in a neighborhood of $(\boldsymbol{\psi}_0, \boldsymbol{\sigma})$ and

$$\nabla_z \phi_{\text{mrfa}}(\boldsymbol{\psi}_0, \boldsymbol{\sigma}) = \operatorname{vec}\left(\sum_{i=r+1}^{p} \mathbf{e}_i \mathbf{e}_i'\right), \tag{34}$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_p$ is a set of orthonormal eigenvectors of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$.

Suppose now that the population distribution has fourth order moments, and hence by the Central Limit Theorem $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ converges in distribution to multivariate normal $N(\mathbf{0}, \boldsymbol{\Gamma})$, denoted $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}) \Rightarrow N(\mathbf{0}, \boldsymbol{\Gamma})$. Note that since $\mathbf{s}$ has at most $p(p+1)/2$ distinct elements, the rank of the $p^2 \times p^2$ covariance matrix $\boldsymbol{\Gamma}$ is less than or equal to $p(p+1)/2$. In particular, if the population distribution is normal, then $\boldsymbol{\Gamma} = 2\mathbf{M}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$, where $\mathbf{M}_p$ is a symmetric idempotent matrix of rank $p(p+1)/2$ with element in row ij and column kl given by

$$M_p(ij, kl) = \tfrac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \tag{35}$$

where $\delta_{ik} = 1$ if $i = k$, and $\delta_{ik} = 0$ if $i \neq k$ (Browne, 1974).

By using the Delta method (e.g., Rao, 1973) and employing Proposition 1 we obtain the following result.

Proposition 2. Suppose that the true problem has unique optimal solution $\boldsymbol{\psi}_0 = \operatorname{diag}(\boldsymbol{\Psi}_0)$ to which corresponds a unique Lagrange multipliers matrix $\overline{\boldsymbol{\Omega}}$, the function $\phi(\cdot, \cdot)$ is continuously differentiable in a neighborhood of $(\boldsymbol{\psi}_0, \boldsymbol{\sigma})$, and $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ converges in distribution to $N(\mathbf{0}, \boldsymbol{\Gamma})$. Then $n^{1/2}[\widehat{\vartheta} - \vartheta_0]$ converges in distribution to $N(0, \sigma_\vartheta^2)$, where $\sigma_\vartheta^2 = [\nabla\vartheta(\boldsymbol{\sigma})]'\boldsymbol{\Gamma}[\nabla\vartheta(\boldsymbol{\sigma})]$. In particular, in the case of MRFA and if the population distribution is normal and condition (26) holds, then

$$\sigma_\vartheta^2 = 2\operatorname{tr}\left[\left(\sum_{i=r+1}^{p} \mathbf{e}_i \mathbf{e}_i' - \overline{\boldsymbol{\Omega}}\right)\boldsymbol{\Sigma}\left(\sum_{i=r+1}^{p} \mathbf{e}_i \mathbf{e}_i' - \overline{\boldsymbol{\Omega}}\right)\boldsymbol{\Sigma}\right], \tag{36}$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_p$ is a set of orthonormal eigenvectors of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$.

Let us discuss now second-order derivatives of the optimal value function $\vartheta(\cdot)$. It turns out that first-order asymptotics of $\hat{\boldsymbol{\psi}}$ are closely related to a second-order analysis of $\vartheta(\cdot)$. For the sake of simplicity we assume in the remainder of this section that all components of the

optimal solution $\boldsymbol{\psi}_0 = \bar{\mathbf{x}}(\boldsymbol{\sigma})$, of the true problem, are *positive*. Suppose that the function $\phi(\cdot, \cdot)$ is twice continuously differentiable in a neighborhood of $(\bar{\mathbf{x}}(\boldsymbol{\sigma}), \boldsymbol{\sigma})$, and the nondegeneracy and strict complementarity conditions (for the true problem) hold at the point $\bar{\mathbf{x}}(\boldsymbol{\sigma})$. Consider the following optimization problem

$$\begin{array}{ll} \text{Min} & \left\{ q(\mathbf{h}) := \kappa(\mathbf{h}, \boldsymbol{\delta}) + \operatorname{tr}\left[\overline{\boldsymbol{\Omega}}(\boldsymbol{\Delta} - \mathbf{H})(\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0)^{\dagger}(\boldsymbol{\Delta} - \mathbf{H}) \right] \right\} \\ \text{subject to} & \boldsymbol{\Xi}'\boldsymbol{\Delta}\boldsymbol{\Xi} - \boldsymbol{\Xi}'\mathbf{H}\boldsymbol{\Xi} = \mathbf{0}, \end{array} \tag{37}$$

depending on $p \times p$ symmetric matrix $\boldsymbol{\Delta}$. Here $\mathbf{H}$ is $p \times p$ diagonal matrix, $\mathbf{h} = \operatorname{diag}(\mathbf{H})$ is the corresponding vector, $\boldsymbol{\delta} = \operatorname{vec}(\boldsymbol{\Delta})$, $\boldsymbol{\Xi}$ is a complement of the matrix $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$, and

$$\kappa(\mathbf{h}, \boldsymbol{\delta}) := \tfrac{1}{2}\mathbf{h}'\nabla^2_{xx}\phi(\boldsymbol{\psi}_0, \boldsymbol{\sigma})\mathbf{h} + \mathbf{h}'\nabla^2_{xz}\phi(\boldsymbol{\psi}_0, \boldsymbol{\sigma})\boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}'\nabla^2_{zz}\phi(\boldsymbol{\psi}_0, \boldsymbol{\sigma})\boldsymbol{\delta}. \tag{38}$$

Note that although the complement matrix $\boldsymbol{\Xi}$ is not unique, the corresponding equations in (37) do not depend on a particular choice of $\boldsymbol{\Xi}$. For example, one can use a complement matrix formed from the eigenvectors of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$ corresponding to its zero eigenvalue.

The objective function of the above optimization problem (37) is quadratic and the constraints are linear in $\mathbf{h}$. Therefore, the optimal value of (37), considered as a function of $\boldsymbol{\delta}$, can be written as the quadratic form $\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}$ for some $p^2 \times p^2$ symmetric matrix $\mathbf{Q}$. It follows then that $\nabla^2\vartheta(\boldsymbol{\sigma}) = 2\mathbf{Q}$.

Moreover, let $\bar{\mathbf{h}}(\boldsymbol{\delta})$ be the optimal solution of (37). We have that $\bar{\mathbf{h}}(\boldsymbol{\delta})$ is linear, and hence can be written as $\bar{\mathbf{h}}(\boldsymbol{\delta}) = \mathbf{J}\boldsymbol{\delta}$ for some $p \times p^2$ matrix $\mathbf{J}$. It follows then that the Jacobian matrix $\nabla\bar{\mathbf{x}}(\boldsymbol{\sigma}) = \mathbf{J}$.

Proposition 3. Suppose that the true problem has a unique optimal solution $\boldsymbol{\psi}_0 = \operatorname{diag}(\boldsymbol{\Psi}_0)$, with all diagonal elements of $\boldsymbol{\Psi}_0$ being positive, the nondegeneracy and strict complementarity conditions (for the true problem) hold at the point $\bar{\mathbf{x}}(\boldsymbol{\sigma})$, the function $\phi(\cdot, \cdot)$ is twice continuously differentiable in a neighborhood of $(\boldsymbol{\psi}_0, \boldsymbol{\sigma})$, and $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ converges in distribution to $N(\mathbf{0}, \boldsymbol{\Gamma})$.

Then $n^{1/2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ converges in distribution to $N(\mathbf{0}, \mathbf{J}\boldsymbol{\Gamma}\mathbf{J}')$, where $\mathbf{J} = \nabla\bar{\mathbf{x}}(\boldsymbol{\sigma})$.

It is possible to use the above results in order to derive asymptotics of individual eigenvalues $\hat{\lambda}_k = \lambda_k(\mathbf{S} - \hat{\boldsymbol{\Psi}})$ viewed as estimators of $\lambda_k = \lambda_k(\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0)$. Consider the $p^2 \times p$ matrix

$$\mathbf{K}_p := [\boldsymbol{\eta}_1\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_p\boldsymbol{\eta}_p']', \tag{39}$$

where $\boldsymbol{\eta}_i$ is the i-th column vector of the $p \times p$ identity matrix. Note that for any $p \times p$ diagonal matrix $\mathbf{H}$ the equation

$$\operatorname{vec}(\mathbf{H}) = \mathbf{K}_p[\operatorname{diag}(\mathbf{H})]$$

holds. Let s be the rank of the matrix $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$. Then $\lambda_{s+1} = \cdots = \lambda_p = 0$. We say that the eigenvalue λ_k has multiplicity one if $\lambda_{k-1} > \lambda_k > \lambda_{k+1}$.

Proposition 4. Suppose that the true problem has a unique optimal solution $\boldsymbol{\psi}_0 = \operatorname{diag}(\boldsymbol{\Psi}_0)$, with all diagonal elements of $\boldsymbol{\Psi}_0$ being positive, the nondegeneracy and strict complementarity conditions (for the true problem) hold at the point $\bar{\mathbf{x}}(\boldsymbol{\sigma})$, the function $\phi(\cdot, \cdot)$ is twice continuously differentiable in a neighborhood of $(\boldsymbol{\psi}_0, \boldsymbol{\sigma})$, $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ converges in distribution to $N(\mathbf{0}, \boldsymbol{\Gamma})$, and let s be the rank of the matrix $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$.

Then $\hat{\lambda}_{s+1} = \cdots = \hat{\lambda}_p = 0$ for all $\mathbf{S}$ in a neighborhood of $\boldsymbol{\Sigma}$. Moreover, if an eigenvalue λ_k of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$ has multiplicity one, then $n^{1/2}(\hat{\lambda}_k - \lambda_k)$ converges in distribution to $N(0, \sigma_k^2)$, where

$$\sigma_k^2 = [\operatorname{vec}(\mathbf{e}_k\mathbf{e}_k')]'\left[\mathbf{I}_{p^2} - \mathbf{K}_p\mathbf{J}\right]\boldsymbol{\Gamma}\left[\mathbf{I}_{p^2} - \mathbf{J}'\mathbf{K}_p'\right][\operatorname{vec}(\mathbf{e}_k\mathbf{e}_k')]. \tag{40}$$

Now let us calculate the matrices $\mathbf{Q}$ and $\mathbf{J}$ in the case of MRFA. For the MRFA problem we have that under the condition (26) the function $\phi(\cdot,\cdot)$ is twice continuously differentiable at $(\bar{\mathbf{x}}(\boldsymbol{\sigma}), \boldsymbol{\sigma})$, and

$$\kappa(\mathbf{h}, \boldsymbol{\delta}) = \sum_{(i,j)\in\mathcal{I}} \frac{[\mathbf{e}_i'(\boldsymbol{\Delta} - \mathbf{H})\mathbf{e}_j]^2}{\lambda_j - \lambda_i}, \tag{41}$$

where $\mathcal{I}$ is defined in (29), $\lambda_1 \geq \cdots \geq \lambda_p$ are the eigenvalues and $\mathbf{e}_1, \ldots, \mathbf{e}_p$ is a set of orthonormal eigenvectors of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$, respectively.

Let us consider the following matrices. Let $\boldsymbol{\Xi}$ be such that $\boldsymbol{\Xi}\boldsymbol{\Xi}' = \overline{\boldsymbol{\Omega}}$. Note that since by the first-order optimality condition (15) we have that $\overline{\boldsymbol{\Omega}}(\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0) = \mathbf{0}$, and because of the strict complementarity condition (17), it follows that $\boldsymbol{\Xi}$ is a complement of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$. Note, however, that $\boldsymbol{\Xi}$ is not an arbitrary complement of $\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0$ since it should satisfy the equation $\boldsymbol{\Xi}\boldsymbol{\Xi}' = \overline{\boldsymbol{\Omega}}$. Define $\boldsymbol{\Phi} := (\boldsymbol{\Sigma} - \boldsymbol{\Psi}_0)^\dagger$, $\mathbf{B} := (\boldsymbol{\Xi}' \otimes \boldsymbol{\Xi}')\mathbf{K}_p$ and $\mathbf{N}$ is a complement of the matrix $\mathbf{B}$, i.e., $\mathbf{N}$ is a $p^2 \times [p^2 - (p-s)^2]$ matrix such that $\mathbf{BN} = \mathbf{0}$,

$$\mathbf{Y} := \sum_{(i,j)\in\mathcal{I}} \frac{(\mathbf{e}_i \otimes \mathbf{e}_j)(\mathbf{e}_i \otimes \mathbf{e}_j)'}{\lambda_j - \lambda_i},$$

and $\mathbf{C} := (\mathbf{K}_p'\mathbf{Y}\mathbf{K}_p + \boldsymbol{\Phi} * \overline{\boldsymbol{\Omega}})$.

Then, after some lengthy algebra, the Jacobian matrix $\mathbf{J}$ can be written in the form

$$\mathbf{J} = \mathbf{B}^\dagger(\boldsymbol{\Xi}' \otimes \boldsymbol{\Xi}') - \mathbf{N}(\mathbf{N}'\mathbf{C}\mathbf{N})^{-1}\mathbf{N}'[\mathbf{C}\mathbf{B}^\dagger(\boldsymbol{\Xi}' \otimes \boldsymbol{\Xi}') - \mathbf{K}_p'(\boldsymbol{\Phi} \times \overline{\boldsymbol{\Omega}}) - \mathbf{K}_p'\mathbf{Y}], \tag{42}$$

and the matrix $\mathbf{Q}$ as

$$\mathbf{Q} = \left(\mathbf{K}_p\mathbf{J} - \mathbf{I}_{p^2}\right)'(\mathbf{Y} + \boldsymbol{\Phi} \otimes \overline{\boldsymbol{\Omega}})\left(\mathbf{K}_p\mathbf{J} - \mathbf{I}_{p^2}\right). \tag{43}$$

The above derivations are similar to derivations related to MTFA presented in Shapiro and ten Berge (2000).

Let us give a brief summary of the developed (asymptotic) statistical inference of the MRFA. The result of Proposition 2 shows that asymptotically (under the corresponding regularity conditions) $\widehat{\vartheta}$ has a normal distribution with mean ϑ_0 and variance $n^{-1}\sigma_\vartheta^2$, where σ_ϑ^2 is given in (36). This result is based on the first-order expansion of the optimal value function $\vartheta(\cdot)$ at $\boldsymbol{\sigma}$. By considering the second-order Taylor expansion

$$\vartheta(\mathbf{s}) = \vartheta(\boldsymbol{\sigma}) + [\nabla\vartheta(\boldsymbol{\sigma})]'(\mathbf{s} - \boldsymbol{\sigma}) + \tfrac{1}{2}(\mathbf{s} - \boldsymbol{\sigma})'[\nabla^2\vartheta(\boldsymbol{\sigma})](\mathbf{s} - \boldsymbol{\sigma}) + o(\|\mathbf{s} - \boldsymbol{\sigma}\|^2) \tag{44}$$

of the optimal value function, the bias of $\widehat{\vartheta} = \vartheta(\mathbf{s})$ can be approximated. Since $\mathbf{s}$ is an unbiased estimator of $\boldsymbol{\sigma}$, that is, the expected value of $\mathbf{s} - \boldsymbol{\sigma}$ is $\mathbf{0}$, we have that (under the regularity conditions of Proposition 3) the bias of $\widehat{\vartheta} = \vartheta(\mathbf{s})$ is approximated by

$$\tfrac{1}{2}n^{-1}\mathrm{tr}[\boldsymbol{\Gamma}\nabla^2\vartheta(\boldsymbol{\sigma})] = n^{-1}\mathrm{tr}[\boldsymbol{\Gamma}\mathbf{Q}]. \tag{45}$$

The result of Proposition 4 shows that (under the corresponding regularity conditions, and in particular the nondegeneracy condition) the eigenvalues $\hat{\lambda}_{s+1}, \ldots, \hat{\lambda}_p$ of the sample reduced matrix are zeros for all $\mathbf{S}$ sufficiently close to $\boldsymbol{\Sigma}$. Therefore, if $r = s$, then $\vartheta(\mathbf{s})$ is zero for all $\mathbf{s}$ sufficiently close to $\boldsymbol{\sigma}$, and hence the asymptotic bias of $\widehat{\vartheta}$ is zero.

We have that (under the assumptions of Proposition 3) the diagonal elements of the sample MRFA solution $\hat{\boldsymbol{\Psi}}$ asymptotically have a multivariate normal distribution with the covariance matrix $n^{-1}\mathbf{J}\boldsymbol{\Gamma}\mathbf{J}'$.

3.1. A Numerical Example

Example 2: Seven intelligence tests. To gain an impression of the accuracy of the asymptotic theory in finite samples, we have analyzed an empirical correlation matrix of seven intelligence tests, treated as the population covariance matrix. Under the hypothesis of multivariate normality, 500 simulation runs of the MRFA procedure were performed with sample sizes $n = 500$, $n = 1000$, and $n = 5000$, respectively. Each of the sample covariance matrices (obtained in the 1500 simulation runs) was analyzed with MRFA, using one, two, and three factors. The results are reported in Tables 5, 6, 7, and 8.

The first row in each subtable refers to population values for the *proportion* of explained common variance (ECV), the amount of unexplained common variance (UCV), and the asymptotic sample bias of UCV and its asymptotic variance calculated by the theoretical formulas. The second row always refers to the respective average values over 500 simulation runs. The first two elements are the average of the 500 ECV proportions, and of the 500 ECV values, respectively. The third element is the average of 500 asymptotic bias values, when each sample is treated as if it were the population. The fourth element is the sampling variance of the UCV, and hence of the

TABLE 5.
Simulation results for $r = 1$

$n = 500, r = 1$	ECV	UCV	Bias	Variance
Population	.8336	.7713	.0636	.0104
Sample	.8207	.8516	.0591	.0076
Sampling bias	−0.0129	.0802		
$n = 1000, r = 1$	ECV	UCV	Bias	Variance
Population	.8336	.7713	.0318	.0052
Sample	.8266	.8129	.0361	.0040
Sampling bias	−0.0070	.0416		
$n = 5000, r = 1$	ECV	UCV	Bias	Variance
Population	.8336	.7713	.0064	.0010
Sample	.8325	.7789	.0075	.0010
Sampling bias	−.0010	.0076		

TABLE 6.
Simulation results for $r = 2$

$n = 500, r = 2$	ECV	UCV	Bias	Variance
Population	.9248	.3489	.0177	.0062
Sample	.9164	.3956	.0490	.0033
Sampling bias	−.0084	.0467		
$n = 1000, r = 2$	Ecv	UCV	Bias	Variance
Population	.9248	.3489	.0088	.0031
Sample	.9219	.3667	.0235	.0017
Sampling bias	−.0029	.0178		
$n = 5000, r = 2$	ECV	UCV	Bias	Variance
Population	.9248	.3489	.0018	.0006
Sample	.9247	.3505	.0038	.0006
Sampling bias	−.0001	.0016		

TABLE 7.
Simulation results for $r = 3$

$n = 500, r = 3$	ECV	UCV	Bias	Variance
Population	.9940	.0286	−.0055	.0023
Sample	.9863	.0662	.0230	.0008
Sampling bias	−.0076	.0377		
$n = 1000, r = 3$	ECV	UCV	Bias	Variance
Population	.9940	.0286	−.0027	.0011
Sample	.9895	.0506	.0242	.0005
Sampling bias	−.0045	.0220		
$n = 5000, r = 3$	ECV	UCV	Bias	Variance
Population	.9940	.0286	−.0005	.0002
Sample	.9925	.0355	.0017	.0002
Sampling bias	−.0015	.0070		

TABLE 8.
Simulation results for $r = 4$

$n = 500, r = 4$	ECV	UCV	Bias	Variance
Population	1.0000	.0000	.0000	.0000
Sample	.9985	.0078	.0095	.0002
Sampling bias	−.0015	.0078		
$n = 1000, r = 4$	ECV	UCV	Bias	Variance
Population	1.0000	.0000	.0000	.0000
Sample	.9989	.0059	−.0025	.0001
Sampling bias	−.0011	.0059		
$n = 5000, r = 4$	ECV	UCV	Bias	Variance
Population	1.0000	.0000	.0000	.0000
Sample	.9994	.0031	.0012	.0000
Sampling bias	−.0006	.0031		

UCV bias. The third row gives the differences between the ECV and UCV values of rows two (sample averages) and those of row 1 (population values).

It is clear that UCV will decrease as r, the number of factors extracted, increases. In fact, when $r = 4$, UCV reaches zero for the first time, which means that the population minimal reduced rank is 4. Because this is a case above the Ledermann bound, the solution (unique variances) is non-unique (Shapiro, 1985). Also, the asymptotic bias is zero in this case, see Proposition 4.

It is obvious that the asymptotic bias (elements 1,3 of the subtables) will decrease with increasing sample size n. Comparing the theoretical asymptotic bias with the average bias found by the simulation runs (elements (3,2) of the tables), it can be seen that the theoretical asymptotic bias tends to slightly underestimate the bias encountered in simulations. The latter bias is consistently positive (for $r = 4$ this is necessarily the case). The size of the bias, however, is remarkably small. Specifically, the counterpart of the UCV, the ECV, expressed as a proportion of the common variance, has a small negative bias in samples, usually of less than one percent. For this data set, it seems that UCV bias is hardly a problem at all. It can be noted that the theoretical asymptotic bias seems reasonably accurate in samples sizes of 5000, when $r = 1$ or $r = 2$, but

TABLE 9.
The theoretical asymptotic covariance matrix of the sample unique variances ($r = 2$, $n = 5000$)

0.00019	−0.00001	−0.00003	−0.00001	−0.00006	−0.00008	0.00001
−0.00001	0.00042	−0.00008	−0.00027	−0.00066	0.00013	0.00047
−0.00003	−0.00008	0.00028	0.00012	0.00004	−0.00009	0.00001
−0.00001	−0.00027	0.00012	0.00040	0.00063	−0.00010	−0.00039
−0.00006	−0.00066	0.00004	0.00063	0.00277	−0.00005	−0.00169
−0.00008	0.00013	−0.00009	−0.00010	−0.00005	0.00030	0.00007
0.00001	0.00047	0.00001	−0.00039	−0.00169	0.00007	0.00131

TABLE 10.
The estimated by simulation asymptotic covariance matrix of the sample unique variances ($r = 2$, 500 runs of samples of size $n = 5000$)

0.00021	−0.00001	−0.00005	−0.00002	−0.00004	−0.00009	−0.00004
−0.00001	0.00059	−0.00005	−0.00033	−0.00079	0.00013	0.00049
−0.00005	−0.00005	0.00022	0.00006	0.00008	−0.00005	−0.00002
−0.00002	−0.00033	0.00006	0.00040	0.00072	−0.00007	−0.00039
−0.00004	−0.00079	0.00008	0.00072	0.00398	−0.00013	−0.00206
−0.00009	0.00013	−0.00005	−0.00007	−0.00013	0.00027	0.00013
−0.00004	0.00049	−0.00002	−0.00039	−0.00206	0.00013	0.00143

not with $r = 3$ or $r = 4$. The theoretical asymptotic variance estimate (elements 1,4 compared to elements 2,4) seems fully accurate throughout the samples with size 5000.

The theoretical asymptotic covariance matrix of the sample unique variance has also been compared to its simulation estimate. Again, for $r = 1$ or 2, and $n = 5000$, the theoretical values of the asymptotic covariances of the sample unique variances were quite accurate. Tables 9 and 10 give the corresponding theoretical and simulation results for $r = 2$ and $n = 5000$. However, for $r = 3$ and $r = 4$, much larger samples were needed to reach the same degree of accuracy.

For $r = 4$, the poor accuracy of the theoretical asymptotic estimates is not surprising, because the assumption of a unique solution is violated. It is yet to be clarified why the estimate was so inaccurate when $r = 3$, where the population solution does appear to be unique. A possible explanation may rest in the fact that the sample solutions for $r = 3$ often display Heywood cases (the population solution does not). Heywood cases in samples do not invalidate the asymptotic theory but they do detract from the accuracy of the MRFA program we have used. This remains a matter of further investigation.

References

Bentler, P.M., & Woodward, J.A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika, 45*, 249–267.

Bonnans, J.F., & Shapiro, A. (2000). *Perturbation analysis of optimization problems*. New York, NY: Springer-Verlag.

Browne, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistitical Journal, 8*, 1–24.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor analytic research*. New York, NY: Cambridge University Press.

Guttman, L. (1958). To what extent can communalities reduce rank? *Psychometrika, 23*, 297–308.

Harman, H.H. (1967). *Modern factor analysis*. Chicago, IL: The University of Chicago Press.

Harman, H.H. & Jones, W.H. (1966). Factor analysis by minimizing residuals (minres). *Psychometrika, 31*, 351–368.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 31*, 443–482.

Ledermann, W. (1937). On the rank of reduced correlation matrices in multiple factor analysis. *Psychometrika, 2*, 85–93.

Rao, C.R. (1973). *Linear statistical inference and its applications*. New York, NY: Wiley.

Saigal, R., Vandenberghe, L. & Wolkowicz, H. (Eds.). (2000). *Semidefinite programming and applications handbook.* Boston, MA: Kluwer Academic Publishers.

Schutz, R.E. (1958). Factorial validity of the Holzinger-Crowder Uni-Factor Tests. *Educational and Psychological Measurement, 18*, 873–875.

Shapiro, A. (1982). Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika, 47*, 187–199.

Shapiro, A. (1985). Identifiability of factor analysis: Some results and open problems. *Linear Algebra and its Applications, 70*, 1–7.

Shapiro, A. & ten Berge, J.M.F. (2000). The asymptotic bias of Minimum Trace Factor Analysis, with applications to the greatest lower bound to reliability. *Psychometrika, 65*, 413–425.

ten Berge, J.M.F. (1998). Some recent developments in factor analysis and the search for proper communalities. In A. Rizzi, M. Vichi, & H.H. Bock (Eds), *Advances in data science and classification* (pp. 325–334). Heidelberg, Germany: Springer.

ten Berge, J.M.F. (2000). Linking reliability and factor analysis:recent developments in some classical psychometric problems. In S.E. Hampson (Ed.), *Advances in personality psychology, Vol. 1* (pp. 138–156). London, England: Routledge.

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

ten Berge, J.M.F., Snijders, T.A.B. & Zegers, F.E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. *Psychometrika, 46*, 201–213.