MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

Bayesian factor analysis and model selection

K. Hirose, S. Kawano
S. Konishi, M. Ichikawa

MHF 2008-2

(Received January 15, 2008)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Bayesian Factor Analysis and Model Selection

Kei Hirose[1], Shuichi Kawano[1],
Sadanori Konishi[2] and Masanori Ichikawa[3]

[1] *Graduate School of Mathematics, Kyushu University,*
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.

[2] *Faculty of Mathematics, Kyushu University,*
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.

[3] *Tokyo University of Foreign Studies,*
3-11-1 Asahi-cho, Fuchu-shi, Tokyo 183-8534, Japan.

E-mail: ma207048@math.kyushu-u.ac.jp, s-kawano@math.kyushu-u.ac.jp,
konishi@math.kyushu-u.ac.jp, Ichikawa.M@tufs.ac.jp.

Abstract

Factor analysis provides a useful tool for exploring the covariance structure among a set of observed random variables by construction of a smaller number of random variables called common factors. In maximum likelihood factor analysis, the estimates of unique or error variances can turn out to be zero or negative, which makes no sense from a statistical point of view. In order to overcome the problem of these so-called improper solutions, we use a Bayesian approach by specifying a prior distribution for the variances of specific factors, i.e., we introduce a prior distribution for the parameters to prevent the occurrence of improper solutions. Crucial aspects of Bayesian factor analysis include the choice of adjusted parameters, in particular, the hyper-parameters for the prior distribution and also choosing an appropriate number of factors. The choice of these parameters can be viewed as a model selection and evaluation problem. We derive a model selection criterion for a Bayesian factor analysis model. Monte Carlo simulations are conducted to investigate the efficiency of the proposed procedures. A real data example is also given to illustrate our procedures.

Key Words: Bayesian approach, EM algorithm, Factor analysis, Model selection criterion.

1 Introduction

Factor analysis is one of the most popular methods of multivariate statistical analysis, used in the social and behavioral sciences to explore the covariance structure among observed variables in terms of a smaller number of latent variables. In maximum likelihood factor analysis, the estimates of unique variances sometimes turn out to be zero or negative. Such estimates are known as improper solutions, and many researchers have studied these inappropriate estimates both from a theoretical point of view and also by means of numerical examples (see, e.g., Jöreskog (1967), van Driel (1978) and Sato (1987)). In this

paper we take a Bayesian approach to this problem by specifying a prior distribution for the variances of specific factors. In Bayesian factor analysis, the choice of a prior distribution is a fundamental issue. Some prior distributions for Bayesian factor analysis have been proposed in the literature. Martin and Mcdonald (1975) used a prior distribution for the elements of unique variances. Press (1982) used a natural conjugate prior distribution. Akaike (1987) introduced a natural prior distribution for factor loadings to avoid the occurrence of improper solutions. In this paper we introduce a prior distribution that is the modification of the prior distribution given by Akaike (1987).

Another important aspect of Bayesian factor analysis is the choice of adjusted parameters including the hyper-parameters of the prior distribution and the number of factors; however in most previous studies, these parameters have been assumed to be subjectively given. Selection of these parameters can be viewed as a model selection and evaluation problem. In this paper we derive a model selection criterion from a Bayesian viewpoint (Konishi *et al.*, 2004) for a Bayesian factor analysis model. The proposed method enables us to choose the appropriate number of factors objectively and also to prevent the occurrence of the improper solutions.

The remainder of this paper is organized as follows: Section 2 describes the orthogonal factor analysis model and improper solutions. In Section 3, we introduce a proper prior distribution based on the prior distribution given by Akaike (1987). In Section 4, we provide the parameter estimation in Bayesian factor analysis by using EM algorithm. Section 5 derives the model selection criterion for Bayesian factor analysis. Section 6 presents numerical results for both artificial and real datasets. We summarize our work in Section 7.

2 Factor analysis model

2.1 Model

Let $\boldsymbol{x} = (x_1, \cdots, x_p)'$ be a p-dimensional observable random vector with mean vector $\boldsymbol{\mu}$ and variance covariance matrix Σ. The fundamental equation for the orthogonal factor analysis model is

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

where Λ is a $p \times k$ $(k < p)$ matrix of factor loadings, and $\boldsymbol{f} = (f_1, \cdots, f_k)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)'$ are unobservable random vectors. The elements of $\boldsymbol{f}$ and $\boldsymbol{\varepsilon}$ are called common

factors and specific factors, respectively. It is assumed that $\boldsymbol{f}$ and $\boldsymbol{\varepsilon}$ are independent, with $E(\boldsymbol{f}) = \mathbf{0},\ E(\boldsymbol{\varepsilon}) = \mathbf{0},\ E(\boldsymbol{f}\boldsymbol{f}') = I_k$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Psi$, where I_k is the identity matrix of order k and Ψ is a $p \times p$ diagonal matrix with i-th diagonal element ψ_i^2. Under these assumptions, the variance-covariance matrix of $\boldsymbol{x}$ is given by

$$\Sigma = \Lambda\Lambda' + \Psi. \tag{2}$$

Assume that the common factors $\boldsymbol{f}$ and the specific factors $\boldsymbol{\varepsilon}$ have multivariate normal distributions given by

$$\boldsymbol{f} \sim N_k(\mathbf{0}, I_k), \quad \boldsymbol{\varepsilon} \sim N_p(\mathbf{0}, \Psi). \tag{3}$$

Suppose that we have a random sample of N observations $\boldsymbol{X}_N = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)'$ from the multivariate normal population $N(\boldsymbol{\mu}, \Sigma)$ with $\Sigma = \Lambda\Lambda' + \Psi$. Then the log-likelihood function is given by

$$\log f(\boldsymbol{X}_N|\Lambda, \Psi) = -\frac{N}{2}\left[p\log(2\pi) + \log|\Sigma| + \mathrm{tr}(\Sigma^{-1}S)\right], \tag{4}$$

where $f(\boldsymbol{X}_N|\Lambda, \Psi)$ is the likelihood function and S is the sample variance-covariance matrix

$$S = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})'. \tag{5}$$

Here $\bar{\boldsymbol{x}}$ is the sample mean vector. The maximum likelihood estimates in factor analysis model are obtained by maximizing Equation (4).

2.2 Improper solutions

In practical situations, however, estimates of the specific variances sometimes turn out to be zero or negative. Such estimates have been called improper solutions. van Driel (1978) categorized the causes of improper solutions into the following three types:

(i) sampling fluctuation,
(ii) there exist no appropriate factor analysis models for extraction of beneficial information from the data,
(iii) indefiniteness of the model.

He observed that situations of type (i) and (iii) can result in interpretable factor analysis models whereas situations of type (ii) produce only non-interpretable models.

In order to avoid the occurrence of improper solutions, many researchers have proposed various ways to deal with the problem, for example, estimating the parameters (A) under

the condition that the specific variances are bounded below 0.005 (see Jöreskog (1967)), (B) after eliminating variables which cause the improper solutions and (C) by utilizing a Bayesian framework. Some problems remain, however, with approaches (A) and (B). With approach (A), the variances of specific factors are provided subjectively, whereas they should be estimated. With approach (B), we often obtain inappropriate estimates even when variables that cause the improper solutions are eliminated. In this paper, we focus our attention on approach (C) i.e. on a method that estimates the parameters included in the factor analysis model with the help of Bayesian techniques, since approaches (A) and (B) have the disadvantages mentioned above. In the Bayesian approach we need to specify a prior distribution for the parameters $\Psi = \mathrm{diag}(\psi_1^2, \cdots, \psi_p^2)$. This requirement is discussed in the next section.

3 Bayesian approach

In the Bayesian procedure, we introduce a prior distribution for factor loadings Λ and specific variances Ψ, and estimate parameters from a posterior distribution:

$$\begin{aligned} f(\Lambda, \Psi | \boldsymbol{X}_N) &= \frac{f(\boldsymbol{X}_N | \Lambda, \Psi)\pi(\Lambda, \Psi)}{\int\int f(\boldsymbol{X}_N | \Lambda, \Psi)\pi(\Lambda, \Psi) d\Lambda d\Psi} \\ &\propto f(\boldsymbol{X}_N | \Lambda, \Psi)\pi(\Lambda, \Psi), \end{aligned} \tag{6}$$

where $\pi(\Lambda, \Psi)$ is the density function of the prior distribution of Λ and Ψ. In this paper we estimate parameters by obtaining a maximization (mode) of the posterior distribution given by Equation (6). This technique can be regarded as a maximum penalized likelihood approach. The penalized log-likelihood function is

$$l_\rho(\Lambda, \Psi) = \log f(\boldsymbol{X}_N | \Lambda, \Psi) - \rho \frac{N}{2} H(\Lambda, \Psi), \tag{7}$$

where ρ is a hyper-parameter and $H(\Lambda, \Psi)$ is a penalty term. We observe that the second term of the right side of Equation (7) corresponds to the prior distribution $\pi(\Lambda, \Psi)$.

A fundamental issue in the maximum penalized likelihood method is the choice of the penalty term $H(\Lambda, \Psi)$ (i.e. the choice of a prior distribution) because estimation of parameters depends on this term.

3.1 Prior distributions

The selection of a precise and potent prior distribution to prevent the occurrence of improper solutions is a fundamental issue in Bayesian factor analysis. The prior dis-

tribution presented by Akaike (1987) provides us with a guideline. In this subsection we investigate the nature of improper solutions and the relationship between the prior distributions proposed by Martin and Mcdonald (1975) and that of Akaike (1987), and introduce a prior distribution according to the basic idea given by Akaike (1987).

For convenience, let us consider the quantity (see Equation (4.14) of Lawley and Maxwell (1971))

$$q(\Lambda, \Psi) = \log|\Sigma| + \mathrm{tr}(\Sigma^{-1}S) - \log|S| - p. \tag{8}$$

The maximum likelihood estimates are obtained by minimizing Equation (8) instead of maximizing Equation (4). In maximum likelihood factor analysis, Equation (8) is rewritten as follows. (see Equation (4.16) of Lawley and Maxwell (1971))

$$q_k(\Psi) = \sum_{i=k+1}^{p} (\theta_i - \log\theta_i) + (p-k). \tag{9}$$

Here $\theta_1, \cdots, \theta_p$ are the eigenvalues of $\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}$, and it is assumed that $\theta_1 > \cdots > \theta_p$. From Equation (9) the maximum likelihood estimates of the covariance matrix of specific factors are obtained when the values of $\theta_{k+1}, \cdots, \theta_p$ are as close to one as possible.

Akaike (1987) showed that since Equation (9) is sensitive only to the behavior of smaller eigenvalues of $\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}$, the values of $\theta_1, \cdots, \theta_k$ sometimes go to infinity because the diagonal elements of Ψ go to zero, which causes improper solutions. To avoid the occurrence of improper solutions, we should perform the analysis under the restriction that the values of $\theta_1, \cdots, \theta_k$ are not too large.

For this reason, we add a penalty term $\rho \sum_{i=1}^{k} \theta_i$ to Equation (8) and consider minimizing the following equation with respect to Λ and Ψ.

$$q^*(\Lambda, \Psi) \;=\; q(\Lambda, \Psi) + \rho \sum_{i=1}^{k} \theta_i. \tag{10}$$

The additive term prevents the occurrence of improper solutions because it does not allow the values of $\theta_1, \cdots, \theta_k$ to be infinite. If we introduce the constraint

$$\Lambda' \Psi^{-1} \Lambda = \Delta, \tag{11}$$

where Δ is a diagonal matrix (see, e.g., Anderson (2003)), the penalty term $\rho \sum_{i=1}^{k} \theta_i$ is equal to $\rho \mathrm{tr}(\Lambda' \Psi^{-1} \Lambda + I_k)$. This term leads to a prior distribution proposed by Akaike (1987), which is given by

$$K \exp\left\{ -\frac{N}{2} \rho \ \mathrm{tr}\ (\Psi^{-\frac{1}{2}} \Lambda \Lambda' \Psi^{-\frac{1}{2}}) \right\}, \tag{12}$$

where K denotes the normalizing constant. However, this prior distribution cannot be correctly normalized because the normalizing constant K diverges.

We propose adding a penalty term $\rho \sum_{i=1}^{p} \theta_i$ to Equation (8) and then the estimates are obtained by minimizing

$$\begin{aligned} q^{**}(\Lambda, \Psi) &= q(\Lambda, \Psi) + \rho \sum_{i=1}^{p} \theta_i & (13) \\ &= q(\Lambda, \Psi) + \rho \mathrm{tr}\ (\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}). & (14) \end{aligned}$$

It is not unreasonable to add the term $\sum_{i=k+1}^{p} \theta_i$ to $\sum_{i=1}^{k} \theta_i$ since the values of $\theta_{k+1}, \cdots, \theta_p$ are close to one and so can be ignored in comparison to the large values of $\theta_1, \cdots, \theta_k$. Minimizing Equation (14) is equivalent to maximizing Equation

$$\log f(\boldsymbol{X}_N | \Lambda, \Psi) - \frac{N\rho}{2} \mathrm{tr}(\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}). \quad (15)$$

The prior distribution is then given by

$$\begin{aligned} & K \exp\left\{ -\frac{N\rho}{2} \mathrm{tr}(\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}) \right\} & (16) \\ = \ & K \prod_{i=1}^{p} \exp\left\{ -\frac{N\rho s_{ii}}{2} \psi_i^{-2} \right\}, & (17) \end{aligned}$$

where s_{ii} is the i-th diagonal elements of the sample variance-covariance matrix S. We assume that the inverses of the elements of Ψ have an exponential distribution and then the normalizing constant K is given by $\prod_{i=1}^{p} N\rho s_{ii}/2$.

Our proposed prior distribution and the prior distribution proposed by Martin and Mcdonald (1975) are closely related. The prior distribution proposed by Martin and Mcdonald (1975) is given by

$$K \prod_{i=1}^{p} \exp\left\{ -\frac{N\alpha_i}{2} \psi_i^{-2} \right\}, \quad (18)$$

where $\alpha_1, \cdots, \alpha_p$ are hyper-parameters of the prior distribution. However, it is difficult to specify these hyper-parameters because p is large. Martin and Mcdonald (1975) recommended restricting these hyper-parameters by requiring that

$$\alpha_1 = \cdots = \alpha_p = s_{ii}\alpha. \quad (19)$$

If we assume Equation (19), the proposed prior distribution given by (17) coincides with the prior distribution given by Martin and Mcdonald (1975). They subjectively selected a

hyper-parameter α which controls the trade off between the log-likelihood and the penalty term.

For a complete specification of the Bayesian model, it is necessary to define a prior distribution for Λ. However, we have no prior convictions about Λ in exploratory factor analysis because Λ has a rotational indeterminacy. For this reason Martin and Mcdonald (1975) suggested not assigning a prior distribution to the elements of Λ. In this paper we also do not assign a prior distribution to Λ.

The proposed prior distribution given by (17) has two advantages. Firstly, the prior distribution avoids the occurrence of improper solutions naturally because the additive term $\rho \sum_{i=1}^{k} \theta_i$ does not allow the values of $\theta_1, \cdots, \theta_k$ to be infinite. Secondly, we can consider the model selection criterion from a Bayesian viewpoint (Konishi *et al.* (2004)) with this prior distribution since the normalizing constant K is not infinite. While the prior distribution proposed by Akaike (1987) prevents the occurrence of improper solutions, we cannot apply a Bayesian model selection criterion because K is infinite.

4 Estimation

4.1 EM algorithm

We use a numerical iterative algorithm to obtain the penalized maximum likelihood estimates. Since the solutions cannot be expressed in closed form, we use some iterative procedure. In maximum likelihood factor analysis, some numerical algorithms have been proposed by earlier researchers (see, e.g., Jöreskog (1967) and Jennrich and Robinson (1969)). Rubin and Thayer (1982) proposed using EM algorithms for maximum likelihood factor analysis. The advantage of EM algorithms is that even if the likelihood of the initial points is not convex, the estimates converge to a maximum of the likelihood estimates. Furthermore, the EM algorithm can easily be adapted to apply to the penalized maximum likelihood.

We give the expectation and maximization steps for the Bayesian factor analysis model within a general framework for EM algorithms. We consider the common factors $\boldsymbol{f}$ as a latent variable, and maximize the complete-data log-likelihood using a posterior distribution for the latent variable.

The iterative procedure is given by

$$\begin{aligned}\hat{\Lambda} &= (S\Psi^{-1}\Lambda)(I_k + M^{-1}\Lambda'\Psi^{-1}S\Psi^{-1}\Lambda)^{-1},\\ \hat{\Psi} &= \mathrm{diag}\left[S - 2S\Psi^{-1}\Lambda M^{-1}\hat{\Lambda}' + \hat{\Lambda}M^{-1}\hat{\Lambda}' + \hat{\Lambda}M^{-1}\Lambda'\Psi^{-1}S\Psi^{-1}\Lambda M^{-1}\hat{\Lambda}' + \rho S\right].\end{aligned}$$

The derivation of the procedure is detailed in Appendix A.

4.2 Constraint on factor loading

It is well known that all factor loadings obtained from the initial loadings by orthogonal transformation have the same ability to reproduce the covariance matrix, so the analysis should proceed by imposing conditions on the factor loadings. In maximum likelihood factor analysis, it is convenient to use the constraint given by Equation of (11). However, this constraint is not appropriate for the EM algorithm and the model selection criterion in a Bayesian context. For this reason, we use the constraint on factor loadings given by the following equation (Anderson and Rubin (1956)).

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 & 0 & \cdots & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 & \cdots & 0 & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{k-1,1} & \lambda_{k-1,2} & \lambda_{k-1,3} & \cdots & \lambda_{k-1,k-1} & 0 \\ \lambda_{k,1} & \lambda_{k,2} & \lambda_{k,3} & \cdots & \lambda_{k,k-1} & \lambda_{k,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{p,1} & \lambda_{p,2} & \lambda_{p,3} & \cdots & \lambda_{p,k-1} & \lambda_{p,k} \end{pmatrix}. \tag{20}$$

5 Model selection criterion

The generalized Bayesian information criterion (GBIC), proposed by Konishi *et al.* (2004), enables us to choose adjusted parameters including a hyper-parameter ρ and also the number of factors k by extending the Bayesian information criterion (BIC) proposed by Schwarz (1978). The basic idea of BIC is to select a model from a set of candidate models by maximizing the posterior probability. BIC only deals with models estimated by the maximum likelihood method, whereas the model selection criterion GBIC also applies to models estimated by the maximum penalized likelihood method.

Suppose that $\boldsymbol{\theta}$ is the parameter of a Bayesian factor analysis. The model selection

criterion GBIC for Bayesian factor analysis is given by

$$\begin{aligned}\text{GBIC} &= -p^* \log(2\pi) + p^* \log N + \log |J_\rho(\hat{\boldsymbol{\theta}}_{p^*})| + N \left\{ p \log(2\pi) + \log |\hat{\Sigma}| + \text{tr}(\hat{\Sigma}^{-1} S) \right\} \\ &\quad -2 \sum_{i=1}^{p} \log \left(\frac{N \rho s_{ii}}{2} \right) + N\rho \sum_{i=1}^{p} (s_{ii} \hat{\psi}_i^{-2}), \end{aligned} \tag{21}$$

where p^* is the number of parameters, $\hat{\boldsymbol{\theta}}_{p^*}$, $\hat{\Sigma}$ and $\hat{\psi}_i^2$ are the posterior modes of parameters and $J_\rho(\hat{\boldsymbol{\theta}}_{p^*})$ is

$$J_\rho(\hat{\boldsymbol{\theta}}_{p^*}) = -\frac{1}{N} \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \left\{ \log f(\boldsymbol{X}_N | \boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta} | \rho) \right\} \Big|_{\hat{\boldsymbol{\theta}}_{p^*}} \right]. \tag{22}$$

Here $\log f(\boldsymbol{X}_N | \boldsymbol{\theta})$ is a log-likelihood function given by Equation (4) and $\pi(\boldsymbol{\theta}|\rho)$ is the density of the prior distribution of Equation (17). We choose optimum values of the hyper-parameter ρ and the number of factors k which minimize the value of the model selection criterion given by Equation (21). The derivation of GBIC is detailed in Appendix B.

6 Numerical Examples

6.1 Simulation results

For our first simulation study we consider two factors for a seven dimensional problem generating fifty observations: $k = 2,\ p = 7$ and $N = 50$. The datasets are generated by using $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, 0.64\, I_7)$, where I_7 is the identity matrix of order 7 and

$$\Lambda' = \begin{pmatrix} 0.6 & 0.0 & 0.6 & 0.6 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.6 & 0.0 & 0.0 & 0.6 & 0.6 & 0.6 \end{pmatrix}. \tag{23}$$

When the 100 different datasets were generated by Equation (1), improper solutions were obtained for 30 out of 100 datasets. We investigate one of the datasets that causes improper solutions. First, maximum likelihood estimates of factor loadings and specific variances are given by

$$\hat{\Lambda}' = \begin{pmatrix} 0.397 & 0.340 & 0.530 & 0.977 & 0.027 & -0.070 & 0.187 \\ 0.109 & 0.791 & 0.316 & -0.203 & 0.589 & 0.671 & 0.552 \end{pmatrix}, \tag{24}$$

$$\text{diag}\hat{\Psi} = \begin{pmatrix} 0.831 & 0.259 & 0.619 & 0.000 & 0.652 & 0.545 & 0.660 \end{pmatrix}. \tag{25}$$

The estimate of ψ_4^2 goes to zero, which is the inappropriate estimate.

Table 1: The values of ρ corresponding to the minimum values of GBIC for each number of factors

	$k=1$	$k=2$	$k=3$
ρ	0.0316	0.0251	0.0199
GBIC	982.84	974.73	977.43

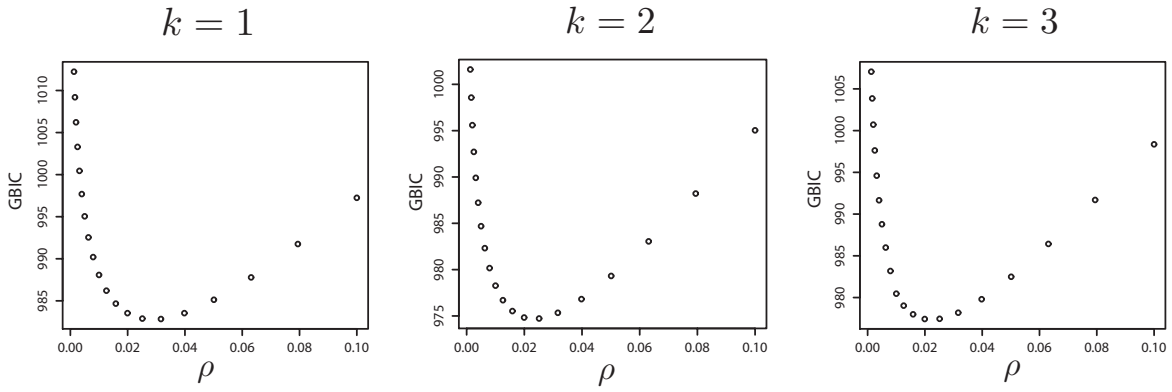

Figure 1: The model selection criterion GBIC for varying values of the hyper-parameter ρ. On the left side $k=1$, in the center $k=2$ and on the right side $k=3$

To prevent the occurrence of improper solutions, the model is estimated by using the EM algorithm with a prior distribution given by Equation (17). We choose the adjusted parameters including hyper-parameter of prior distribution and the number of factors using the model selection criterion given by (21). Figure 1 plots the model selection criterion against the hyper-parameter ρ. Table 1 gives the minimum value of GBIC and the value of the hyper-parameter ρ estimated by using GBIC when the number of factors is fixed. Table 1 shows that the value of GBIC is a minimum when $k=2$ and $\rho=0.0251$, and the estimates of factor loadings Λ and specific variances Ψ are given by

$$\hat{\Lambda}' = \begin{pmatrix} 0.407 & 0.344 & 0.595 & 0.811 & 0.065 & -0.090 & 0.151 \\ 0.106 & 0.750 & 0.297 & 0.155 & 0.581 & 0.689 & 0.577 \end{pmatrix}, \tag{26}$$

$$\mathrm{diag}\hat{\Psi} = \begin{pmatrix} 0.848 & 0.345 & 0.583 & 0.344 & 0.683 & 0.542 & 0.669 \end{pmatrix}. \tag{27}$$

Unlike the maximum likelihood estimates, the estimate of ψ_4^2 obtained by the proposed method does not go to zero. As a result, the model selection criterion results in choice of an appropriate number of factors and adjusted hyper-parameter ρ to prevent the occurrence of improper solutions.

Table 2: Four types of datasets

	dataset (a)			dataset (b)			dataset (c)			dataset (d)		
i	Λ		diag(Ψ)	Λ		diag(Ψ)	Λ		diag(Ψ)	Λ		diag(Ψ)
1	0.6	0.0	0.64	0.9	0.0	0.19	0.9	0.0	0.19	0.9	0.0	0.19
2	0.0	0.6	0.64	0.0	0.6	0.64	0.0	0.8	0.36	0.0	0.8	0.36
3	0.6	0.0	0.64	0.6	0.0	0.64	0.7	0.0	0.51	0.7	0.0	0.51
4	0.6	0.0	0.64	0.6	0.0	0.64	0.6	0.0	0.64	0.6	0.0	0.64
5	0.0	0.6	0.64	0.0	0.6	0.64	0.0	0.6	0.64	0.0	0.7	0.51
6	0.0	0.6	0.64	0.0	0.6	0.64	0.0	0.6	0.64	0.0	0.5	0.75
7	0.0	0.6	0.64	0.0	0.6	0.64	0.0	0.6	0.64	0.0	0.9	0.19

6.2 Numerical comparison

In a second simulation study we investigate the efficiency of the proposed procedure by comparing various model selection criteria. We use four types of factor loadings Λ and specific variances Ψ, which are given in Table 2, and two variants for the number of observations, $N = 30$ and $N = 50$.

Each dataset was generated 100 times and improper solutions were obtained as shown:

$N = 30$: (a): 57 times, (b): 49 times, (c): 48 times, (d): 50 times,
$N = 50$: (a): 30 times, (b): 44 times, (c): 37 times, (d): 28 times.

We counted the number of times that each k-factor model achieved the lowest value of model selection criteria, which are summarized in Table 3. For example, AIC selected the one factor model 34 times out of 100 datasets at dataset (a) when $N = 30$. We compared the model selection criteria AIC, BIC obtained by maximum likelihood estimates with GBIC obtained by maximum penalized likelihood estimates. AIC and BIC are given by

$$\begin{aligned} \text{AIC} &= -2\log f(\boldsymbol{X}_N|\hat{\Lambda},\hat{\Psi}) + 2p^*, \\ \text{BIC} &= -2\log f(\boldsymbol{X}_N|\hat{\Lambda},\hat{\Psi}) + p^*\log N, \end{aligned}$$

where $\hat{\Lambda}$ and $\hat{\Psi}$ are the maximum likelihood estimates and p^* is the number of parameters, which is given by $p(k+1) - k(k-1)/2$. Table 3 shows that, for each dataset, the model selection criterion GBIC performs well in the sense that our proposed procedure most often selects an optimal number of factors.

Table 3: Comparisons of model selection criteria for simulated datasets.

dataset (a):

N	k	AIC	BIC	GBIC
$N = 30$	1	34	80	26
	2	56	19	67
	3	10	1	7
$N = 50$	1	9	59	9
	2	77	41	86
	3	14	0	5

dataset (b):

N	k	AIC	BIC	GBIC
$N = 30$	1	13	61	18
	2	75	38	76
	3	12	1	6
$N = 50$	1	0	26	2
	2	85	74	97
	3	15	0	1

dataset (c):

N	k	AIC	BIC	GBIC
$N = 30$	1	3	25	6
	2	83	74	94
	3	14	1	0
$N = 50$	1	0	26	1
	2	85	74	98
	3	15	0	1

dataset (d):

N	k	AIC	BIC	GBIC
$N = 30$	1	1	16	3
	2	85	83	96
	3	14	1	1
$N = 50$	1	0	3	0
	2	87	97	98
	3	13	0	2

6.3 Job application dataset

This dataset is illustrated in Kendall (1980). There are 48 applicants for a certain job, who have been scored on $p = 15$ variables regarding their acceptability. The variables are

(1) Form of letter application,	(2) Appearance,	(3) Academic ability,
(4) Likeability,	(5) Self confidence,	(6) Lucidity,
(7) Honesty,	(8) Salesmanship,	(9) Experience,
(10) Drive,	(11) Ambition,	(12) Grasp,
(13) Potential,	(14) Keenness to join,	(15) Suitability.

We compared the model selection criteria AIC, BIC for maximum likelihood estimates with GBIC for penalized maximum likelihood estimates, and the results are shown in Table 4. The model selection criterion GBIC selected 4 factors; we obtained improper solutions for this case with maximum likelihood factor analysis. AIC and BIC also selected a model that resulted in improper solutions.

It is important to identify the cause of the improper solutions. The maximum likelihood estimates of Ψ and the standard deviation $\hat{\sigma}_{\psi_i^2}$ of $\sqrt{N}\psi_i^2/s_{ii}$ (see Equation (5.50) of Lawley and Maxwell (1971)) for $k = 2$ to 6 are shown in Table 5. We were not able to calculate the standard deviation $\hat{\sigma}_{\psi_i^2}$ at $k = 5$ and $k = 6$. For $k = 4$, the maximum likelihood estimates $\hat{\psi}_1^2$, $\hat{\psi}_3^2$, $\hat{\psi}_7^2$, $\hat{\psi}_{13}^2$, $\hat{\psi}_{14}^2$ were less than the corresponding estimates for $k = 3$. These results for the estimates of specific variances seem to suggest that we have identified some new common factors. The standard deviation $\hat{\sigma}_{\psi_i^2}$ of $\sqrt{N}\psi_i^2/s_{ii}$ is not especially large when $k = 4$. For these reasons, the improper solutions might be due to sampling fluctuations rather than indefiniteness of the model.

The estimates of factor loadings Λ and specific factors Ψ obtained using the proposed method are given in Table 6. The estimates of factor loadings Λ are obtained by Varimax rotation. Table 6 shows that our proposed model can prevent the occurrence of improper solutions. Moreover, in earlier research, Press and Shigemasu (1989) and Fokoue (2004) also suggested four factors for this dataset. For these reasons, we claim that, at least in these cases, our approach results in efficient model estimates.

Table 4: The values of AIC, BIC obtained by maximum likelihood estimates and ρ corresponding to the minimum values of GBIC for varying numbers of factors for Kendall's dataset. The right side column of "improper solutions?" indicates whether the maximum likelihood estimates are proper or improper.

k	AIC	BIC	GBIC	ρ	improper solutions?
1	1660.54	1716.68	1741.90	0.01585	proper
2	1605.39	1687.72	1708.77	0.01259	proper
3	1562.46	1669.11	1689.79	0.01259	proper
4	1532.43	1661.54	1682.31	0.01000	improper
5	1525.34	1675.04	1697.00	0.00794	improper
6	1521.25	1689.66	1716.17	0.00794	improper
7	1510.83	1696.08	1730.84	0.00631	improper
8	1510.96	1711.18	1749.20	0.00631	improper
9	1519.14	1732.46	1764.02	0.00631	improper
10	1528.35	1752.90	1780.21	0.00631	improper

Table 5: Maximum likelihood estimates of specific variances and the standard derivations of $\sqrt{N}\psi_i^2/s_{ii}$ for $k = 2$ to 6 using Kendall's dataset.

	$k=2$		$k=3$		$k=4$		$k=5$		$k=6$	
i	$\hat{\psi}_i^2$	$\hat{\sigma}_{\psi_i^2}$	$\hat{\psi}_i^2$	$\hat{\sigma}_{\psi_i^2}$	$\hat{\psi}_i^2$	$\hat{\sigma}_{\psi_i^2}$	$\hat{\psi}_i^2$	$\hat{\sigma}_{\psi_i^2}$	$\hat{\psi}_i^2$	$\hat{\sigma}_{\psi_i^2}$
1	0.546	0.834	0.535	0.816	0.444	0.757	0.467	—	0.394	—
2	0.717	0.779	0.701	0.777	0.688	0.777	0.668	—	0.625	—
3	0.951	0.449	0.945	0.473	0.523	0.779	0.508	—	0.494	—
4	0.741	0.772	0.000	0.717	0.199	0.566	0.196	—	0.200	—
5	0.139	0.340	0.109	0.291	0.112	0.249	0.113	—	0.128	—
6	0.191	0.382	0.196	0.381	0.194	0.350	−0.072	—	−0.056	—
7	0.795	0.757	0.445	0.771	0.341	0.800	0.294	—	0.254	—
8	0.171	0.346	0.144	0.304	0.133	0.263	0.132	—	0.145	—
9	0.366	0.766	0.360	0.747	0.360	0.756	0.360	—	0.365	—
10	0.247	0.460	0.238	0.447	0.225	0.403	0.232	—	0.020	—
11	0.178	0.361	0.157	0.325	0.140	0.271	0.113	—	0.043	—
12	0.192	0.377	0.204	0.390	0.153	0.285	0.163	—	0.152	—
13	0.208	0.404	0.183	0.357	0.089	0.195	0.080	—	0.088	—
14	0.600	0.779	0.420	0.671	−0.000	0.001	0.051	—	0.064	—
15	0.190	0.553	0.188	0.534	0.250	0.569	0.242	—	0.264	—

Table 6: The estimates of factor loading Λ and specific variances Ψ obtained by the proposed method for $k = 4$ using Kendall's dataset.

i	factor 1	factor 2	factor 3	factor 4	specific variances
1	0.717	0.130	−0.107	0.118	0.453
2	0.154	0.449	0.131	0.255	0.703
3	0.116	0.072	0.735	−0.018	0.451
4	0.242	0.226	−0.053	0.848	0.178
5	−0.092	0.915	−0.083	0.149	0.135
6	0.120	0.837	0.063	0.303	0.200
7	−0.211	0.248	−0.019	0.740	0.356
8	0.238	0.893	−0.072	0.084	0.144
9	0.777	0.090	0.180	−0.051	0.363
10	0.386	0.767	−0.052	0.174	0.240
11	0.180	0.899	−0.056	0.107	0.154
12	0.267	0.790	0.180	0.348	0.161
13	0.343	0.730	0.261	0.428	0.109
14	0.366	0.430	−0.509	0.549	0.130
15	0.781	0.362	0.103	0.059	0.254

7 Concluding Remarks

In maximum likelihood factor analysis models, there arise situations in which the estimates of unique variances go to zero or become negative. To avoid the occurrence of such improper solutions, we used a Bayesian approach by specifying a proper prior distribution for variances of specific factors. In practice, an optimal choice of the number of factors is also of importance for exploring the covariance structure. We derived the model selection and evaluation criterion from a Bayesian point of view, and used it to choose adjusted parameters that include a hyper-parameter for a prior distribution and the number of factors. Monte Carlo simulations and a real data example showed that the proposed procedure performs well in various situations. We would recommend our factor analysis modeling procedure based on the proper prior distribution of exponential type and the Bayesian model selection criterion.

Appendix

Appendix A: The derivation of the EM algorithm for Bayesian factor analysis

To apply the EM algorithm, we consider the common factors $\boldsymbol{f}_n$ to be missing data and maximize the complete-data penalized log-likelihood. The complete-data penalized log-likelihood is given by

$$l_\rho^C = \sum_{n=1}^{N} \log\{f(\boldsymbol{x}_n, \boldsymbol{f}_n)\pi(\boldsymbol{\theta}|\rho)\}, \tag{28}$$

where $f(\boldsymbol{x}_n, \boldsymbol{f}_n)$ is the density of the complete-data distribution and $\pi(\boldsymbol{\theta}|\rho)$ is the prior distribution given by Equation (17).

To derive the posterior mean of the log-likelihood, we use the conditional distribution of common factors $\boldsymbol{f}$ given the observed $\boldsymbol{x}$. It is well known that the conditional distribution is given (see, e.g., Anderson (2003)) by

$$\boldsymbol{f}|\boldsymbol{x} \sim N(M^{-1}\Lambda'\Psi^{-1}\boldsymbol{x}, M^{-1}), \tag{29}$$

where $M = \Lambda'\Psi^{-1}\Lambda + I_k$. Therefore the values for $E[\boldsymbol{F}_n|\boldsymbol{x}_n]$ and $E[\boldsymbol{F}_n\boldsymbol{F}_n'|\boldsymbol{x}_n]$ in the E-step are

$$E[\boldsymbol{F}_n|\boldsymbol{x}_n] = M^{-1}\Lambda'\Psi^{-1}\boldsymbol{x}_n, \tag{30}$$

$$E[\boldsymbol{F}_n\boldsymbol{F}_n'|\boldsymbol{x}_n] = M^{-1} + E[\boldsymbol{F}_n|\boldsymbol{x}_n]\ E[\boldsymbol{F}_n|\boldsymbol{x}_n]'. \tag{31}$$

In the M-step, the expectation of the complete log-likelihood with respect to the distributions of Equation (28) is

$$\begin{aligned} E[l_\rho^C] &= -\sum_{n=1}^{N}\Bigg[\frac{p}{2}\log(2\pi)+\frac{p}{2}\log|\Psi|+\frac{1}{2}\mathrm{tr}\left\{E[\boldsymbol{F}_n\boldsymbol{F}_n']\right\}+\frac{1}{2}\mathrm{tr}\left\{\Psi^{-1}\boldsymbol{x}_n\boldsymbol{x}_n'\right\} \\ &\quad -E[\boldsymbol{F}_n]'\Lambda'\Psi^{-1}\boldsymbol{x}_n+\frac{1}{2}\mathrm{tr}(\Lambda'\Psi^{-1}\Lambda E[\boldsymbol{F}_n\boldsymbol{F}_n'])+\frac{\rho}{2}\mathrm{tr}(\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}})\Bigg], \end{aligned} \tag{32}$$

where $E[\boldsymbol{F}_n]=E[\boldsymbol{F}_n|\boldsymbol{x}_n]$ and $E[\boldsymbol{F}_n\boldsymbol{F}_n']=E[\boldsymbol{F}_n\boldsymbol{F}_n'|\boldsymbol{x}_n]$. Under these assumptions, the new parameter estimates are obtained by maximizing $E[l_\rho^C]$ with respect to Λ and Ψ, resulting in

$$\hat{\Lambda} = \left\{\sum_{n=1}^{N}\boldsymbol{x}_nE[\boldsymbol{F}_n]'\right\}\left\{\sum_{n=1}^{N}E[\boldsymbol{F}_n\boldsymbol{F}_n']\right\}^{-1}, \tag{33}$$

$$\hat{\Psi} = \frac{1}{N}\mathrm{diag}\left[\sum_{n=1}^{N}\left\{\boldsymbol{x}_n\boldsymbol{x}_n'-2\boldsymbol{x}_nE[\boldsymbol{F}_n]'\hat{\Lambda}'+\hat{\Lambda}E[\boldsymbol{F}_n\boldsymbol{F}_n']\hat{\Lambda}'\right\}+N\rho S\right]. \tag{34}$$

The maximum likelihood estimates are given by calculating the conditional distribution of Equation (30) and (31) and the revised estimates are given by Equation (33) and (34). We can construct an efficient algorithm by substituting (30) and (31) into (33) and (34) , and then the new parameters are

$$\begin{aligned} \hat{\Lambda} &= (S_N\Psi^{-1}\Lambda)(I_k+M^{-1}\Lambda'\Psi^{-1}S_N\Psi^{-1}\Lambda)^{-1}, \\ \hat{\Psi} &= \mathrm{diag}\left[S_N-2S_N\Psi^{-1}\Lambda M^{-1}\hat{\Lambda}'+\hat{L}M^{-1}\hat{\Lambda}'+\hat{\Lambda}M^{-1}\Lambda'\Psi^{-1}S_N\Psi^{-1}\Lambda M^{-1}\hat{\Lambda}'+\rho S\right]. \end{aligned}$$

The two equations are iterated until the estimates of the parameters converge.

Appendix B: The derivation of GBIC for Bayesian factor analysis model

This section derives the model selection criterion for a Bayesian factor analysis model. Suppose that $\boldsymbol{\theta}$ is the parameter and p^* is the number of parameters. Considering the constraint on factor loadings given by Equation (20), the number of parameters of p^* is $p(k+1)-k(k-1)/2$, and the parameter $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta}=(\boldsymbol{\lambda}_{\cdot 1}',\boldsymbol{\lambda}_{\cdot 2}',\cdots,\boldsymbol{\lambda}_{\cdot k}',\mathrm{diag}(\Psi)')', \tag{35}$$

where $\boldsymbol{\lambda}_{\cdot i}=(\lambda_{i,i},\lambda_{i+1,i},\cdots,\lambda_{p,i})$.

Let us consider selecting a model from a set of candidate models $M_1, \cdots, M_r$. The model M_t has the probability density $f_t(\boldsymbol{x}|\boldsymbol{\theta}_t)$, and $\boldsymbol{\theta}_t$ has a prior density $\pi_t(\boldsymbol{\theta}_t|\rho_t)$, where ρ_t is a hyper-parameter. The Bayesian procedure for selecting a model is to choose the model with the largest posterior probability, which is given by

$$\begin{aligned} P(M_t|\boldsymbol{x}) &\propto P(M_t) \int \prod_{t=1}^{r} f(\boldsymbol{x}_n|\boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t|\rho_t)d\boldsymbol{\theta}_t \qquad (36) \\ &=: P(M_t)p_t(\boldsymbol{x}_n|\rho_t), \qquad (37) \end{aligned}$$

where $P(M_t)$ is the prior probability for model M_t. If it is assumed that the prior probability $P(M_t)$ is the same for all models, it follows that the model that maximizes the marginal likelihood $p_t(\boldsymbol{x}_n|\rho_t)$ of the data must be selected. The model selection criterion GBIC (Konishi *et al.* (2004)) is obtained by minimizing $-2\log p_t(\boldsymbol{x}_n|\rho_t)$ with the use of Laplace approximation (Tierney and Kadane (1986)), and GBIC is written as follows (see Equation (10) of Konishi *et al.* (2004)) :

$$\text{GBIC} = -p^*\log(2\pi) + p^*\log N + \log|J_\rho(\hat{\boldsymbol{\theta}}_{p^*})| - 2\left\{\log f(\boldsymbol{X}_N|\hat{\boldsymbol{\theta}}_{p^*}) + \log\pi(\hat{\boldsymbol{\theta}}_{p^*}|\rho)\right\}, \quad (38)$$

where $J_\rho(\hat{\boldsymbol{\theta}}_{p^*})$ is given by

$$J_\rho(\hat{\boldsymbol{\theta}}_{p^*}) = -\frac{1}{N}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left\{\log f(\boldsymbol{X}_N|\boldsymbol{\theta}) + \log\pi(\boldsymbol{\theta}|\rho)\right\}\Big|_{\hat{\boldsymbol{\theta}}_{p^*}}\right], \quad (39)$$

$\log f(\boldsymbol{X}_N|\boldsymbol{\theta})$ is a log-likelihood function and $\pi(\boldsymbol{\theta}|\rho)$ is the density of the prior distribution, which is written as follows (see Equations (4) and (17)) :

$$\begin{aligned} \log f(\boldsymbol{X}_N|\boldsymbol{\theta}) &= -\frac{N}{2}\left[p\log(2\pi) + \log|\Sigma| - \text{tr}\left\{\Sigma^{-1}S\right\}\right], \qquad (40) \\ \pi(\boldsymbol{\theta}|\rho) &= \prod_{i=1}^{p}\frac{N\rho s_{ii}}{2}\exp\left\{-\frac{N\rho s_{ii}}{2}\psi_i^{-2}\right\}. \qquad (41) \end{aligned}$$

By substituting Equation (40) and (41) into (38), we obtain a model selection criterion for Bayesian factor analysis which can be expressed as

$$\begin{aligned} \text{GBIC} &= -p^*\log(2\pi) + p^*\log N + \log|J_\rho(\hat{\boldsymbol{\theta}}_{p^*})| + N\left\{p\log(2\pi) + \log|\hat{\Sigma}| + \text{tr}(\hat{\Sigma}^{-1}S)\right\} \\ &\quad -2\sum_{i=1}^{p}\log\left(\frac{N\rho s_{ii}}{2}\right) + N\rho\sum_{i=1}^{p}(s_{ii}\hat{\psi}_i^{-2}). \end{aligned}$$

The matrix $J_\rho(\boldsymbol{\theta})$ consists of the elements of $\frac{\partial^2 l_\rho}{\partial\lambda_{ab}\partial\lambda_{cd}}$, $\frac{\partial^2 l_\rho}{\partial\psi_i^2\partial\lambda_{cd}}$ and $\frac{\partial^2 l_\rho}{\partial\psi_i^2\partial\psi_j^2}$, which are given

by

$$
\begin{aligned}
\frac{\partial^2 l_\rho}{\partial\lambda_{ab}\partial\lambda_{cd}} &= N\Big\{ (\Sigma^{-1})_{(a,c)}(\Lambda'\Sigma^{-1}\Lambda)_{(b,d)} + (\Sigma^{-1}\Lambda)_{(a,d)}(\Sigma^{-1}\Lambda)_{(c,b)} \\
&\quad -(\Sigma^{-1}S\Sigma^{-1})_{(a,c)}(\Lambda'\Sigma^{-1}\Lambda)_{(b,d)} - (\Sigma^{-1}\Lambda)_{(a,d)}(\Sigma^{-1}S\Sigma^{-1}\Lambda)_{(c,b)} \\
&\quad -(\Sigma^{-1})_{(a,c)}(\Lambda'\Sigma^{-1}S\Sigma^{-1}\Lambda)_{(b,d)} - (\Sigma^{-1}S\Sigma^{-1}\Lambda)_{(a,d)}(\Sigma^{-1}\Lambda)_{(c,b)} \\
&\quad -(\Sigma^{-1})_{(a,c)}(I_k)_{(b,d)} + (\Sigma^{-1}S\Sigma^{-1})_{(a,c)}(I_k)_{(b,d)} \Big\}, \\
\frac{\partial^2 l_\rho}{\partial\psi_i^2\partial\lambda_{cd}} &= N\Big\{ (\Sigma^{-1})_{(c,i)}(\Sigma^{-1}\Lambda)_{(i,d)} - (\Sigma^{-1}S\Sigma^{-1})_{(c,i)}(\Sigma^{-1}\Lambda)_{(i,d)} \\
&\quad -(\Sigma^{-1})_{(c,i)}(\Sigma^{-1}S\Sigma^{-1}\Lambda)_{(i,d)} \Big\}, \\
\frac{\partial^2 l_\rho}{\partial\psi_i^2\partial\psi_j^2} &= \frac{N}{2}\Big\{ (\Sigma^{-1})^2_{(i,j)} - 2(\Sigma^{-1})_{(i,j)}(\Sigma^{-1}S\Sigma^{-1})_{(i,j)} - 2\rho(\Psi^{-3})_{(i,j)}S_{(i,i)} \Big\},
\end{aligned}
$$

where $(X)_{(\alpha,\beta)}$ is the (α,β)-th element of the matrix X.

References

[1] Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317-332.

[2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd.ed.). Wiley, New York.

[3] Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis, In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **5**, (pp. 111-150). Berkeley: University of California Press.

[4] Fokoue, E. (2004). Stochastic determination of the intrinsic structure in Bayesian factor analysis. *Technical Report 17*.

[5] Jennrich, R. I. and Robinson, S. M. (1969). A Newton-Raphson algorithm for maximum likelihood factor analysis. *Psychometrika*, **34**, 111-123.

[6] Jöreskog, K. G. (1967). Some contribution to maximum likelihood factor analysis. *Psychometrika*, **32**, 443-482.

[7] Kendall, M. G. (1980). *Multivariate Analysis* (2nd.ed.). Charles Griffin.

[8] Konishi, S., Ando, T. and Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, **91**, 27-43.

[9] Lawley, D. N. and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method* (2nd ed.). London: Butterworths. *Proceedings of the Royal Society of Edinburgh*, **60**, 64-82.

[10] Martin, J. K. and McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, **40**, 505-517.

[11] Press, S. J. (1982). *Applied Multivariate analysis: using Bayesian and frequentist methods of inference*. Krieger.

[12] Press, S. J. and Shigemasu, K. (1989). Bayesian inference in factor analysis. *ASA Proc. Social Statistics Section*, 292-294.

[13] Rubin, D. B. and Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, **47**, 69-76.

[14] Sato, M. (1987). Pragmatic treatment of improper solutions in factor analysis. *Annals of the Institute of Statistical Mathematics*, **39**, 443-455.

[15] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461-464.

[16] Tierny, L. and Kanade, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.*, **81**, 82-6.

[17] van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, **43**, 225-243.

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF2005-1 Hideki KOSAKI
Matrix trace inequalities related to uncertainty principle

MHF2005-2 Masahisa TABATA
Discrepancy between theory and real computation on the stability of some finite element schemes

MHF2005-3 Yuko ARAKI & Sadanori KONISHI
Functional regression modeling via regularized basis expansions and model selection

MHF2005-4 Yuko ARAKI & Sadanori KONISHI
Functional discriminant analysis via regularized basis expansions

MHF2005-5 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Point configurations, Cremona transformations and the elliptic difference Painlevé equations

MHF2005-6 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Construction of hypergeometric solutions to the q - Painlevé equations

MHF2005-7 Hiroki MASUDA
Simple estimators for non-linear Markovian trend from sampled data:
I. ergodic cases

MHF2005-8 Hiroki MASUDA & Nakahiro YOSHIDA
Edgeworth expansion for a class of Ornstein-Uhlenbeck-based models

MHF2005-9 Masayuki UCHIDA
Approximate martingale estimating functions under small perturbations of dynamical systems

MHF2005-10 Ryo MATSUZAKI & Masayuki UCHIDA
One-step estimators for diffusion processes with small dispersion parameters from discrete observations

MHF2005-11 Junichi MATSUKUBO, Ryo MATSUZAKI & Masayuki UCHIDA
Estimation for a discretely observed small diffusion process with a linear drift

MHF2005-12 Masayuki UCHIDA & Nakahiro YOSHIDA
AIC for ergodic diffusion processes from discrete observations

MHF2005-13 Hiromichi GOTO & Kenji KAJIWARA
Generating function related to the Okamoto polynomials for the Painlevé IV equation

MHF2005-14 Masato KIMURA & Shin-ichi NAGATA
Precise asymptotic behaviour of the first eigenvalue of Sturm-Liouville problems with large drift

MHF2005-15 Daisuke TAGAMI & Masahisa TABATA
Numerical computations of a melting glass convection in the furnace

MHF2005-16 Raimundas VIDŪNAS
Normalized Leonard pairs and Askey-Wilson relations

MHF2005-17 Raimundas VIDŪNAS
Askey-Wilson relations and Leonard pairs

MHF2005-18 Kenji KAJIWARA & Atsushi MUKAIHIRA
Soliton solutions for the non-autonomous discrete-time Toda lattice equation

MHF2005-19 Yuu HARIYA
Construction of Gibbs measures for 1-dimensional continuum fields

MHF2005-20 Yuu HARIYA
Integration by parts formulae for the Wiener measure restricted to subsets in $\mathbb{R}^d$

MHF2005-21 Yuu HARIYA
A time-change approach to Kotani's extension of Yor's formula

MHF2005-22 Tadahisa FUNAKI, Yuu HARIYA & Mark YOR
Wiener integrals for centered powers of Bessel processes, I

MHF2005-23 Masahisa TABATA & Satoshi KAIZU
Finite element schemes for two-fluids flow problems

MHF2005-24 Ken-ichi MARUNO & Yasuhiro OHTA
Determinant form of dark soliton solutions of the discrete nonlinear Schrödinger equation

MHF2005-25 Alexander V. KITAEV & Raimundas VIDŪNAS
Quadratic transformations of the sixth Painlevé equation

MHF2005-26 Toru FUJII & Sadanori KONISHI
Nonlinear regression modeling via regularized wavelets and smoothing parameter selection

MHF2005-27 Shuichi INOKUCHI, Kazumasa HONDA, Hyen Yeal LEE, Tatsuro SATO, Yoshihiro MIZOGUCHI & Yasuo KAWAHARA
On reversible cellular automata with finite cell array

MHF2005-28 Toru KOMATSU
Cyclic cubic field with explicit Artin symbols

MHF2005-29 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Kaori NAGATOU
A computational approach to constructive a priori and a posteriori error estimates for finite element approximations of bi-harmonic problems

MHF2005-30 Kaori NAGATOU, Kouji HASHIMOTO & Mitsuhiro T. NAKAO
Numerical verification of stationary solutions for Navier-Stokes problems

MHF2005-31 Hidefumi KAWASAKI
A duality theorem for a three-phase partition problem

MHF2005-32 Hidefumi KAWASAKI
A duality theorem based on triangles separating three convex sets

MHF2005-33 Takeaki FUCHIKAMI & Hidefumi KAWASAKI
An explicit formula of the Shapley value for a cooperative game induced from the conjugate point

MHF2005-34 Hideki MURAKAWA
A regularization of a reaction-diffusion system approximation to the two-phase Stefan problem

MHF2006-1 Masahisa TABATA
Numerical simulation of Rayleigh-Taylor problems by an energy-stable finite element scheme

MHF2006-2 Ken-ichi MARUNO & G R W QUISPEL
Construction of integrals of higher-order mappings

MHF2006-3 Setsuo TANIGUCHI
On the Jacobi field approach to stochastic oscillatory integrals with quadratic phase function

MHF2006-4 Kouji HASHIMOTO, Kaori NAGATOU & Mitsuhiro T. NAKAO
A computational approach to constructive a priori error estimate for finite element approximations of bi-harmonic problems in nonconvex polygonal domains

MHF2006-5 Hidefumi KAWASAKI
A duality theory based on triangular cylinders separating three convex sets in R^n

MHF2006-6 Raimundas VIDŪNAS
Uniform convergence of hypergeometric series

MHF2006-7 Yuji KODAMA & Ken-ichi MARUNO
N-Soliton solutions to the DKP equation and Weyl group actions

MHF2006-8 Toru KOMATSU
Potentially generic polynomial

MHF2006-9 Toru KOMATSU
Generic sextic polynomial related to the subfield problem of a cubic polynomial

MHF2006-10 Shu TEZUKA & Anargyros PAPAGEORGIOU
Exact cubature for a class of functions of maximum effective dimension

MHF2006-11 Shu TEZUKA
On high-discrepancy sequences

MHF2006-12 Raimundas VIDŪNAS
Detecting persistent regimes in the North Atlantic Oscillation time series

MHF2006-13 Toru KOMATSU
Tamely Eisenstein field with prime power discriminant

MHF2006-14 Nalini JOSHI, Kenji KAJIWARA & Marta MAZZOCCO
Generating function associated with the Hankel determinant formula for the solutions of the Painlevé IV equation

MHF2006-15 Raimundas VIDŪNAS
Darboux evaluations of algebraic Gauss hypergeometric functions

MHF2006-16 Masato KIMURA & Isao WAKANO
New mathematical approach to the energy release rate in crack extension

MHF2006-17 Toru KOMATSU
Arithmetic of the splitting field of Alexander polynomial

MHF2006-18 Hiroki MASUDA
Likelihood estimation of stable Lévy processes from discrete data

MHF2006-19 Hiroshi KAWABI & Michael RÖCKNER
Essential self-adjointness of Dirichlet operators on a path space with Gibbs measures via an SPDE approach

MHF2006-20 Masahisa TABATA
Energy stable finite element schemes and their applications to two-fluid flow problems

MHF2006-21 Yuzuru INAHAMA & Hiroshi KAWABI
Asymptotic expansions for the Laplace approximations for Itô functionals of Brownian rough paths

MHF2006-22 Yoshiyuki KAGEI
Resolvent estimates for the linearized compressible Navier-Stokes equation in an infinite layer

MHF2006-23 Yoshiyuki KAGEI
Asymptotic behavior of the semigroup associated with the linearized compressible Navier-Stokes equation in an infinite layer

MHF2006-24 Akihiro MIKODA, Shuichi INOKUCHI, Yoshihiro MIZOGUCHI & Mitsuhiko FUJIO
The number of orbits of box-ball systems

MHF2006-25 Toru FUJII & Sadanori KONISHI
Multi-class logistic discrimination via wavelet-based functionalization and model selection criteria

MHF2006-26 Taro HAMAMOTO, Kenji KAJIWARA & Nicholas S. WITTE
Hypergeometric solutions to the q-Painlevé equation of type $(A_1 + A_1')^{(1)}$

MHF2006-27 Hiroshi KAWABI & Tomohiro MIYOKAWA
The Littlewood-Paley-Stein inequality for diffusion processes on general metric spaces

MHF2006-28 Hiroki MASUDA
Notes on estimating inverse-Gaussian and gamma subordinators under high-frequency sampling

MHF2006-29 Setsuo TANIGUCHI
The heat semigroup and kernel associated with certain non-commutative harmonic oscillators

MHF2006-30 Setsuo TANIGUCHI
Stochastic analysis and the KdV equation

MHF2006-31 Masato KIMURA, Hideki KOMURA, Masayasu MIMURA, Hidenori MIYOSHI, Takeshi TAKAISHI & Daishin UEYAMA
Quantitative study of adaptive mesh FEM with localization index of pattern

MHF2007-1 Taro HAMAMOTO & Kenji KAJIWARA
Hypergeometric solutions to the q-Painlevé equation of type $A_4^{(1)}$

MHF2007-2 Kouji HASHIMOTO, Kenta KOBAYASHI & Mitsuhiro T. NAKAO
Verified numerical computation of solutions for the stationary Navier-Stokes equation in nonconvex polygonal domains

MHF2007-3 Kenji KAJIWARA, Marta MAZZOCCO & Yasuhiro OHTA
A remark on the Hankel determinant formula for solutions of the Toda equation

MHF2007-4 Jun-ichi SATO & Hidefumi KAWASAKI
Discrete fixed point theorems and their application to Nash equilibrium

MHF2007-5 Mitsuhiro T. NAKAO & Kouji HASHIMOTO
Constructive error estimates of finite element approximations for non-coercive elliptic problems and its applications

MHF2007-6 Kouji HASHIMOTO
A preconditioned method for saddle point problems

MHF2007-7 Christopher MALON, Seiichi UCHIDA & Masakazu SUZUKI
Mathematical symbol recognition with support vector machines

MHF2007-8 Kenta KOBAYASHI
On the global uniqueness of Stokes' wave of extreme form

MHF2007-9 Kenta KOBAYASHI
A constructive a priori error estimation for finite element discretizations in a non-convex domain using singular functions

MHF2007-10 Myoungnyoun KIM, Mitsuhiro T. NAKAO, Yoshitaka WATANABE & Takaaki NISHIDA
A numerical verification method of bifurcating solutions for 3-dimensional Rayleigh-Bénard problems

MHF2007-11 Yoshiyuki KAGEI
Large time behavior of solutions to the compressible Navier-Stokes equation in an infinite layer

MHF2007-12 Takashi YANAGAWA, Satoshi AOKI and Tetsuji OHYAMA
Human finger vein images are diverse and its patterns are useful for personal identification

MHF2007-13 Masahisa TABATA
Finite element schemes based on energy-stable approximation for two-fluid flow problems with surface tension

MHF2007-14 Mitsuhiro T. NAKAO & Takehiko KINOSHITA
Some remarks on the behaviour of the finite element solution in nonsmooth domains

MHF2007-15 Yoshiyuki KAGEI & Takumi NUKUMIZU
Asymptotic behavior of solutions to the compressible Navier-Stokes equation in a cylindrical domain

MHF2007-16 Shuichi INOKUCHI, Yoshihiro MIZOGUCHI, Hyen Yeal LEE & Yasuo KAWAHARA
Periodic Behaviors of Quantum Cellular Automata

MHF2007-17 Makoto HIROTA& Yasuhide FUKUMOTO
Energy of hydrodynamic and magnetohydrodynamic waves with point and continuous spectra

MHF2007-18 Mitsunori KAYANO& Sadanori KONISHI
Functional principal component analysis via regularized Gaussian basis expansions and its application to unbalanced data

MHF2007-19 Mitsunori KAYANO, Koji DOZONO & Sadanori KONISHI
Functional Cluster Analysis via Orthonormalized Gaussian Basis Expansions and Its Application

MHF2008-1 Jun-ichi SATO
An application of a discrete fixed point theorem to the Cournot model

MHF2008-2 Kei HIROSE, Shuichi KAWANO, Sadanori KONISHI & Masanori ICHIKAWA
Bayesian factor analysis and model selection