

## RESEARCH ARTICLE

# Accelerating $L_1$ -penalized expectation maximization algorithm for latent variable selection in multidimensional two-parameter logistic models

Laixu Shang<sup>1</sup>, Ping-Feng Xu<sup>2\*</sup>, Na Shan<sup>3</sup>, Man-Lai Tang<sup>4</sup>, George To-Sum Ho<sup>5</sup>

**1** School of Mathematics and Statistics, Changchun University of Technology, Changchun, China, **2** Academy for Advanced Interdisciplinary Studies, Northeast Normal University, Changchun, China, **3** School of Psychology & Key Laboratory of Applied Statistics of MOE, Northeast Normal University, Changchun, China, **4** Department of Physics, Astronomy and Mathematics, School of Physics, Engineering & Computer Science, University of Hertfordshire, Hertfordshire, United Kingdom, **5** Department of Supply Chain and Information Management, Hang Seng University of Hong Kong, Hong Kong, China

\* xupf900@gmail.com

**OPEN ACCESS**

**Citation:** Shang L, Xu P-F, Shan N, Tang M-L, Ho GT-S (2023) Accelerating  $L_1$ -penalized expectation maximization algorithm for latent variable selection in multidimensional two-parameter logistic models. PLoS ONE 18(1): e0279918. <https://doi.org/10.1371/journal.pone.0279918>

**Editor:** Mahdi Roozbeh, Semnan University, IRAN, ISLAMIC REPUBLIC OF

**Received:** May 17, 2022

**Accepted:** December 16, 2022

**Published:** January 17, 2023

**Copyright:** © 2023 Shang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting information files](#).

**Funding:** The research of Ping-Feng Xu is supported by the Natural Science Foundation of Jilin Province in China (No. 20210101152JC) and the National Natural Science Foundation of China (No. 11571050). The research of Na Shan is supported by the National Natural Science Foundation of China (No. 11871013). The research of George To-Sum Ho is supported by the

## Abstract

One of the main concerns in multidimensional item response theory (MIRT) is to detect the relationship between observed items and latent traits, which is typically addressed by the exploratory analysis and factor rotation techniques. Recently, an EM-based  $L_1$ -penalized log-likelihood method (EML1) is proposed as a vital alternative to factor rotation. Based on the observed test response data, EML1 can yield a sparse and interpretable estimate of the loading matrix. However, EML1 suffers from high computational burden. In this paper, we consider the coordinate descent algorithm to optimize a new weighted log-likelihood, and consequently propose an improved EML1 (IEML1) which is more than 30 times faster than EML1. The performance of IEML1 is evaluated through simulation studies and an application on a real data set related to the Eysenck Personality Questionnaire is used to demonstrate our methodologies.

## 1 Introduction

Multidimensional item response theory (MIRT) models are widely used to describe the relationship between the designed items and the intrinsic latent traits in psychological and educational tests [1]. Early researches for the estimation of MIRT models are confirmatory, where the relationship between the responses and the latent traits are pre-specified by prior knowledge [2, 3]. Under this setting, parameters are estimated by various methods including marginal maximum likelihood method [4] and Bayesian estimation [5]. However, misspecification of the item-trait relationships in the confirmatory analysis may lead to serious model lack of fit, and consequently, erroneous assessment [6].

To avoid the misfit problem caused by improperly specifying the item-trait relationships, the exploratory item factor analysis (IFA) [4, 7] is usually adopted. The exploratory IFA freely

Research Grants Council of Hong Kong (No. UGC/FDS14/P05/20) and the Big Data Intelligence Centre in The Hong Seng University of Hong Kong. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

estimate the entire item-trait relationships (i.e., the loading matrix) only with some constraints on the covariance of the latent traits. To obtain a simpler loading structure for better interpretation, the factor rotation [8, 9] is adopted, followed by a cut-off. Although the exploratory IFA and rotation techniques are very useful, they can not be utilized without limitations. For some applications, different rotation techniques yield very different or even conflicting loading matrices. Therefore, it can be arduous to select an appropriate rotation or decide which rotation is the best [10]. In addition, different subjective choices of the cut-off value possibly lead to a substantial change in the loading matrix [11].

Recently, regularization has been proposed as a viable alternative to factor rotation, and it can automatically rotate the factors to produce a sparse loadings structure for exploratory IFA [12, 13]. Scharf and Nestler [14] compared factor rotation and regularization in recovering predefined factor loading patterns and concluded that regularization is a suitable alternative to factor rotation for psychometric applications. Regularization has also been applied to produce sparse and more interpretable estimations in many other psychometric fields such as exploratory linear factor analysis [11, 15, 16], the cognitive diagnostic models [17, 18], structural equation modeling [19], and differential item functioning analysis [20, 21].

For MIRT models, Sun et al. [12] proposed a latent variable selection framework to investigate the item-trait relationships by maximizing the  $L_1$ -penalized likelihood [22]. In this framework, one can impose prior knowledge of the item-trait relationships into the estimate of loading matrix to resolve the rotational indeterminacy. Based on the observed test response data, the  $L_1$ -penalized likelihood approach can yield a sparse loading structure by shrinking some loadings towards zero if the corresponding latent traits are not associated with a test item. Consequently, it produces a sparse and interpretable estimation of loading matrix, and it addresses the subjectivity of rotation approach.

Since the marginal likelihood for MIRT involves an integral of unobserved latent variables, Sun et al. [12] carried out the expectation maximization (EM) algorithm [23] to solve the  $L_1$ -penalized optimization problem. We denote this method as EML1 for simplicity. In the E-step of EML1, numerical quadrature by fixed grid points is used to approximate the conditional expectation of the log-likelihood. This results in a naive weighted log-likelihood on augmented data set with size equal to  $N \times G$ , where  $N$  is the total number of subjects and  $G$  is the number of grid points. To optimize the naive weighted  $L_1$ -penalized log-likelihood in the M-step, the coordinate descent algorithm [24] is used, whose computational complexity is  $O(N \times G)$ . However,  $N \times G$  is usually very large, and this consequently leads to high computational burden of the coordinate descent algorithm in the M-step. As shown by Sun et al. [12], EML1 requires several hours for MIRT models with three to four latent traits. Another limitation for EML1 is that it does not update the covariance matrix  $\Sigma$  of latent traits in the EM iteration. Sun et al. [12] proposed a two-stage method. It first computes an estimation of  $\Sigma$  via a constrained exploratory analysis under identification conditions, and then substitutes the estimated  $\Sigma$  into EML1 as a known  $\Sigma$  to estimate discrimination and difficulty parameters. However, our simulation studies show that the estimation of  $\Sigma$  obtained by the two-stage method could be quite inaccurate.

Further development for latent variable selection in MIRT models can be found in [25, 26]. Zhang and Chen [25] proposed a stochastic proximal algorithm for optimizing the  $L_1$ -penalized marginal likelihood. They used the stochastic approximation in the stochastic step, which avoids repeatedly evaluating the numerical integral with respect to the multiple latent traits. However, the choice of several tuning parameters, such as a sequence of step size to ensure convergence and burn-in size, may affect the empirical performance of stochastic proximal algorithm. Xu et al. [26] applied the expectation model selection (EMS) algorithm [27] to minimize the  $L_0$ -penalized log-likelihood (for example, the Bayesian information criterion [28]) for

latent variable selection in MIRT models. In their EMS framework, the model (i.e., structure of loading matrix) and parameters (i.e., item parameters and the covariance matrix of latent traits) are updated simultaneously in each iteration. In the simulation of Xu et al. [26], the EMS algorithm runs significantly faster than EML1, but it still requires about one hour for MIRT with four latent traits.

In this paper, we focus on the classic EM framework of Sun et al. [12] and give an improved EM-based  $L_1$ -penalized marginal likelihood (IEML1) with the M-step's computational complexity being reduced to  $O(2 \times G)$ . The fundamental idea comes from the “artificial data” widely used in the EM algorithm for computing maximum marginal likelihood estimation in the IRT literature [4, 29–32]. In Bock and Aitkin (1981) [29] and Bock et al. (1988) [4], “artificial data” are the expected number of attempts and correct responses to each item in a sample of size  $N$  at a given ability level. Essentially, “artificial data” are used to replace the unobservable statistics in the expected likelihood equation of MIRT models. It should be noted that, the number of “artificial data” is  $G$  but not  $N \times G$ , as “artificial data” correspond to  $G$  ability levels (i.e., grid points in numerical quadrature). As a result, the number of data involved in the weighted log-likelihood obtained in E-step is reduced and the efficiency of the M-step is then improved.

In our IEML1, we use a slightly different artificial data to obtain the weighted complete data log-likelihood [33] which is widely used in generalized linear models with incomplete data. Specifically, we classify the  $N \times G$  augmented data into  $2 \times G$  artificial data  $(z, \theta^{(g)})$ , where  $z$  (equals to 0 or 1) is the response to one item and  $\theta^{(g)}$  is one discrete ability level (i.e., grid point value). Thus, we obtain a new weighted  $L_1$ -penalized log-likelihood based on a total number of  $2 \times G$  artificial data  $(z, \theta^{(g)})$ , which reduces the computational complexity of the M-step to  $O(2 \times G)$  from  $O(N \times G)$ .

In addition, it is crucial to choose the grid points being used in the numerical quadrature of the E-step for both EML1 and IEML1. There are various papers that discuss this issue in non-penalized maximum marginal likelihood estimation in MIRT models [4, 29, 30, 34]. To the best of our knowledge, there is however no discussion about the penalized log-likelihood estimator in the literature. In this paper, we will give a heuristic approach to choose artificial data with larger weights in the new weighted log-likelihood. Based on this heuristic approach, IEML1 needs only a few minutes for MIRT models with five latent traits.

The rest of the article is organized as follows. In Section 2, we introduce the multidimensional two-parameter logistic (M2PL) model as a widely used MIRT model, and review the  $L_1$ -penalized log-likelihood method for latent variable selection in M2PL models. In Section 3, we give an improved EM-based  $L_1$ -penalized log-likelihood method for M2PL models with unknown covariance of latent traits. In Section 4, we conduct simulation studies to compare the performance of IEML1, EML1, the two-stage method [12], a constrained exploratory IFA with hard-threshold (EIFAthr) and a constrained exploratory IFA with optimal threshold (EIFAopt). In Section 5, we apply IEML1 to a real dataset from the Eysenck Personality Questionnaire. A concluding remark is provided in Section 6.

## 2 Latent variable selection in multidimensional two-parameter logistic models

In this section, the M2PL model that is widely used in MIRT is introduced. Furthermore, the  $L_1$ -penalized log-likelihood method for latent variable selection in M2PL models is reviewed.

### 2.1 Multidimensional two-parameter logistic model

Consider a  $J$ -item test that measures  $K$  latent traits of  $N$  subjects. Let  $\mathbf{Y} = (y_{ij})_{N \times J}$  be the dichotomous observed responses to the  $J$  items for all  $N$  subjects, where  $y_{ij} = 1$  represents the correct response of subject  $i$  to item  $j$ , and  $y_{ij} = 0$  represents the wrong response. Let  $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{iK})^T$  be the  $K$ -dimensional latent traits to be measured for subject  $i = 1, \dots, N$ . The relationship between the  $j$ th item response and the  $K$ -dimensional latent traits for subject  $i$  can be expressed by the M2PL model as follows

$$F_j(\boldsymbol{\theta}_i) \equiv P(y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, b_j) = \frac{\exp(\mathbf{a}_j^T \boldsymbol{\theta}_i + b_j)}{1 + \exp(\mathbf{a}_j^T \boldsymbol{\theta}_i + b_j)}, \tag{1}$$

where  $\mathbf{a}_j = (a_{j1}, \dots, a_{jK})^T$  and  $b_j$  are known as the discrimination and difficulty parameters, respectively. The parameter  $a_{jk} \neq 0$  implies that item  $j$  is associated with latent trait  $k$ .  $P(y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, b_j)$  denotes the probability that subject  $i$  correctly responds to the  $j$ th item based on his/her latent traits  $\boldsymbol{\theta}_i$  and item parameters  $\mathbf{a}_j$  and  $b_j$ . For the sake of simplicity, we use the notation  $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_J)^T$ ,  $\mathbf{b} = (b_1, \dots, b_J)^T$ , and  $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)^T$ . The discrimination parameter matrix  $\mathbf{A}$  is also known as the loading matrix, and the corresponding structure is denoted by  $\Lambda = (\lambda_{jk})$  with  $\lambda_{jk} = I(a_{jk} \neq 0)$ .

In M2PL models, several general assumptions are adopted. The latent traits  $\boldsymbol{\theta}_i$ ,  $i = 1, \dots, N$ , are assumed to be independent and identically distributed, and follow a  $K$ -dimensional normal distribution  $N(0, \Sigma)$  with zero mean vector and covariance matrix  $\Sigma = (\sigma_{kk'})_{K \times K}$ . Furthermore, the local independence assumption is assumed, that is, given the latent traits  $\boldsymbol{\theta}_i$ ,  $y_{i1}, \dots, y_{iJ}$  are conditional independent.

To guarantee the parameter identification and resolve the rotational indeterminacy for M2PL models, some constraints should be imposed. To identify the scale of the latent traits, we assume the variances of all latent trait are unity, i.e.,  $\sigma_{kk} = 1$  for  $k = 1, \dots, K$ . Dealing with the rotational indeterminacy issue requires additional constraints on the loading matrix  $\mathbf{A}$ . We adopt the constraints used by Sun et al. [12] and Xu et al. [26], that is, each of the first  $K$  items is associated with only one latent trait separately, i.e.,  $a_{jj} \neq 0$  and  $a_{jk} = 0$  for  $1 \leq j \neq k \leq K$ . In practice, the constraint on  $\mathbf{A}$  should be determined according to priori knowledge of the item and the entire study.

### 2.2 Latent variable selection based on $L_1$ -penalized method

The response function for M2PL model in Eq (1) takes a logistic regression form, where  $y_{ij}$  acts as the response, the latent traits  $\boldsymbol{\theta}_i$  as the covariates,  $\mathbf{a}_j$  and  $b_j$  as the regression coefficients and intercept, respectively. We are interested in exploring the subset of the latent traits related to each item, that is, to find all non-zero  $a_{jk}$ s. This can be viewed as variable selection problem in a statistical sense.

Under the local independence assumption, the likelihood function of the complete data  $(\mathbf{Y}, \boldsymbol{\Theta})$  for M2PL model can be expressed as follow

$$L(\mathbf{A}, \mathbf{b}, \Sigma | \mathbf{Y}, \boldsymbol{\Theta}) = \prod_{i=1}^N \varphi(\boldsymbol{\theta}_i | \Sigma) \prod_{j=1}^J F_j(\boldsymbol{\theta}_i)^{y_{ij}} [1 - F_j(\boldsymbol{\theta}_i)]^{1-y_{ij}}, \tag{2}$$

where  $\varphi(\boldsymbol{\theta}_i | \Sigma)$  is the density function of latent trait  $\boldsymbol{\theta}_i$ . The log-likelihood function of observed

data  $Y$  can be written as

$$l(\mathbf{A}, \mathbf{b}, \Sigma | \mathbf{Y}) = \sum_{i=1}^N \log \left[ \int_{\theta_i} \varphi(\theta_i | \Sigma) \prod_{j=1}^J F_j(\theta_i)^{y_{ij}} [1 - F_j(\theta_i)]^{1-y_{ij}} d\theta_i \right]. \tag{3}$$

To investigate the item-trait relationships, Sun et al. [12] applied the  $L_1$ -penalized marginal log-likelihood method to obtain the sparse estimate of  $\mathbf{A}$  for latent variable selection in M2PL model. They carried out the EM algorithm [23] with coordinate descent algorithm [24] to solve the  $L_1$ -penalized optimization problem. However, the covariance matrix  $\Sigma$  of latent traits is assumed to be known and is not realistic in real-world applications.

Instead, we will treat  $\Sigma$  as an unknown parameter and update it in each EM iteration. For this purpose, the  $L_1$ -penalized optimization problem including  $\Sigma$  is represented as

$$(\hat{\mathbf{A}}_\eta, \hat{\mathbf{b}}_\eta, \hat{\Sigma}_\eta) = \arg \max_{\mathbf{A}, \mathbf{b}, \Sigma} l(\mathbf{A}, \mathbf{b}, \Sigma | \mathbf{Y}) - \eta \|\mathbf{A}\|_1 \tag{4}$$

where  $\|\mathbf{A}\|_1 = \sum_{j=1}^J \sum_{k=1}^K |a_{jk}|$  denotes the entry-wise  $L_1$  norm of  $\mathbf{A}$ . The tuning parameter  $\eta > 0$  controls the sparsity of  $\mathbf{A}$ . Larger value of  $\eta$  results in a more sparse estimate of  $\mathbf{A}$ . The tuning parameter is always chosen by cross validation or certain information criteria. In this paper, we employ the Bayesian information criterion (BIC) as described by Sun et al. [12].

### 3 Implementation of the EM algorithm

Due to the presence of the unobserved variable (e.g., the latent traits  $\Theta$ ), the parameter estimates in Eq (4) can not be directly obtained. Sun et al. [12] carried out EML1 to optimize Eq (4) with a known  $\Sigma$ . Similarly, we first give a naive implementation of the EM algorithm to optimize Eq (4) with an unknown  $\Sigma$ . Then, we give an efficient implementation with the M-step’s computational complexity being reduced to  $O(2 \times G)$ , where  $G$  is the number of grid points. Lastly, we will give a heuristic approach to choose grid points being used in the numerical quadrature in the E-step.

#### 3.1 A naive implementation of the EM algorithm

The EM algorithm iteratively executes the expectation step (E-step) and maximization step (M-step) until certain convergence criterion is satisfied. Specifically, the E-step is to compute the  $Q$ -function, i.e., the conditional expectation of the  $L_1$ -penalized complete log-likelihood with respect to the posterior distribution of latent traits  $\Theta$ . The M-step is to maximize the  $Q$ -function. Let  $\Psi = (\mathbf{A}, \mathbf{b}, \Sigma)$  be the set of model parameters, and  $\Psi^{(t)} = (\mathbf{A}^{(t)}, \mathbf{b}^{(t)}, \Sigma^{(t)})$  be the parameters in the  $t$ th iteration. The  $(t + 1)$ th iteration is described as follows.

**3.1.1 E-step.** In the E-step of the  $(t + 1)$ th iteration, under the current parameters  $\Psi^{(t)}$ , we compute the  $Q$ -function involving a  $\Sigma$ -term as follows

$$\begin{aligned} Q(\Psi | \Psi^{(t)}) &= E\{\log L(\Psi | \mathbf{Y}, \Theta) | \mathbf{Y}, \Psi^{(t)}\} - \eta \|\mathbf{A}\|_1 \\ &= Q_0(\Sigma | \Psi^{(t)}) + \sum_{j=1}^J Q_j(\mathbf{a}_j, \mathbf{b}_j | \Psi^{(t)}), \end{aligned} \tag{5}$$

where  $Q_0$  is

$$Q_0(\Sigma | \Psi^{(t)}) = \sum_{i=1}^N E\{\log \varphi(\theta_i | \Sigma) | y_i, \Psi^{(t)}\}$$

and for  $j = 1, \dots, J$ ,  $Q_j$  is

$$Q_j(\mathbf{a}_j, \mathbf{b}_j | \Psi^{(t)}) = \sum_{i=1}^N E\{y_{ij} \log(F_j(\boldsymbol{\theta}_i)) + (1 - y_{ij}) \log(1 - F_j(\boldsymbol{\theta}_i)) | \mathbf{y}_i, \Psi^{(t)}\} - \eta \|\mathbf{a}_j\|_1,$$

where  $\|\mathbf{a}_j\|_1 = \sum_{k=1}^K |a_{jk}|$  denotes the  $L_1$ -norm of vector  $\mathbf{a}_j$ . The conditional expectations in  $Q_0$  and each  $Q_j$  are computed with respect to the posterior distribution of  $\boldsymbol{\theta}_i$  as follows

$$p(\boldsymbol{\theta}_i | \mathbf{y}_i, \Psi^{(t)}) \propto \prod_{j=1}^J [F_j^{(t)}(\boldsymbol{\theta}_i)]^{y_{ij}} [1 - F_j^{(t)}(\boldsymbol{\theta}_i)]^{1-y_{ij}} \cdot \varphi(\boldsymbol{\theta}_i | \Sigma^{(t)}),$$

where  $F_j^{(t)}(\boldsymbol{\theta}_i) = P(y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j^{(t)}, \mathbf{b}_j^{(t)})$ ,  $\mathbf{a}_j^{(t)}$  is the  $j$ th row of  $\mathbf{A}^{(t)}$ , and  $\mathbf{b}_j^{(t)}$  is the  $j$ th element in  $\mathbf{b}^{(t)}$ .

Note that the conditional expectations in  $Q_0$  and each  $Q_j$  do not have closed-form solutions. It is usually approximated using the Gaussian-Hermite quadrature [4, 29] and Monte Carlo integration [35]. For simplicity, we approximate these conditional expectations by summations following Sun et al. [12]. Specifically, we choose fixed grid points  $\mathcal{G} \subseteq [-4, 4]^K$  and the posterior distribution of  $\boldsymbol{\theta}_i$  is then approximated by

$$\tilde{p}(\boldsymbol{\theta}_i | \mathbf{y}_i, \Psi^{(t)}) = \begin{cases} C^{-1} \times \prod_{j=1}^J [F_j^{(t)}(\boldsymbol{\theta}_i)]^{y_{ij}} [1 - F_j^{(t)}(\boldsymbol{\theta}_i)]^{1-y_{ij}} \cdot \varphi(\boldsymbol{\theta}_i | \Sigma^{(t)}) & \text{if } \boldsymbol{\theta}_i \in \mathcal{G}, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where  $C = \sum_{\boldsymbol{\theta}_i' \in \mathcal{G}} \prod_{j=1}^J [F_j^{(t)}(\boldsymbol{\theta}_i')]^{y_{ij}} [1 - F_j^{(t)}(\boldsymbol{\theta}_i')]^{1-y_{ij}} \cdot \varphi(\boldsymbol{\theta}_i' | \Sigma^{(t)})$  serves as a normalizing factor. Thus,  $Q_0$  can be approximated by

$$\tilde{Q}_0(\Sigma | \Psi^{(t)}) = \sum_{i=1}^N \sum_{\boldsymbol{\theta}_i \in \mathcal{G}} \log \varphi(\boldsymbol{\theta}_i | \Sigma) \cdot \tilde{p}(\boldsymbol{\theta}_i | \mathbf{y}_i, \Psi^{(t)}) \tag{7}$$

and  $Q_j$  for  $j = 1, \dots, J$  is approximated by

$$\begin{aligned} \tilde{Q}_j(\mathbf{a}_j, \mathbf{b}_j | \Psi^{(t)}) &= \sum_{i=1}^N \sum_{\boldsymbol{\theta}_i \in \mathcal{G}} [y_{ij} \log(F_j(\boldsymbol{\theta}_i)) + (1 - y_{ij}) \log(1 - F_j(\boldsymbol{\theta}_i))] \cdot \tilde{p}(\boldsymbol{\theta}_i | \mathbf{y}_i, \Psi^{(t)}) - \eta \|\mathbf{a}_j\|_1. \end{aligned} \tag{8}$$

Hence, the Q-function can be approximated by

$$\tilde{Q}(\Psi | \Psi^{(t)}) = \tilde{Q}_0(\Sigma | \Psi^{(t)}) + \sum_{j=1}^J \tilde{Q}_j(\mathbf{a}_j, \mathbf{b}_j | \Psi^{(t)}). \tag{9}$$

**3.1.2 M-step.** In the M-step of the  $(t + 1)$ th iteration, we maximize the approximation of Q-function obtained by E-step

$$\Psi^{(t+1)} = (\mathbf{A}^{(t+1)}, \mathbf{b}^{(t+1)}, \Sigma^{(t+1)}) = \arg \max_{\Psi} \tilde{Q}(\Psi | \Psi^{(t)}), \tag{10}$$

subject to  $\Sigma \succ 0$  and  $\text{diag}(\Sigma) = \mathbf{1}$ , where  $\Sigma \succ 0$  denotes that  $\Sigma$  is a positive definite matrix, and  $\text{diag}(\Sigma) = \mathbf{1}$  denotes that all the diagonal entries of  $\Sigma$  are unity.

It can be easily seen from Eq (9) that  $\tilde{Q}$  can be factorized as the summation of  $\tilde{Q}_0$  involving  $\Sigma$  and  $\tilde{Q}_j$  involving  $(\mathbf{a}_j, b_j)$ . Thus, the maximization problem in Eq (10) can be decomposed to maximizing  $\tilde{Q}_0$  and maximizing penalized  $\tilde{Q}_j$  separately, that is,

$$\Sigma^{(t+1)} = \arg \max_{\Sigma} \tilde{Q}_0(\Sigma|\Psi^{(t)}) \text{ s.t. } \Sigma \succ 0 \text{ and } \text{diag}(\Sigma) = \mathbf{1}, \tag{11}$$

and for  $j = 1, \dots, J$ ,

$$(\mathbf{a}_j^{(t+1)}, b_j^{(t+1)}) = \arg \max_{\mathbf{a}_j, b_j} \tilde{Q}_j(\mathbf{a}_j, b_j|\Psi^{(t)}), \tag{12}$$

For maximization problem (11),  $\tilde{Q}_0$  can be represented as

$$\tilde{Q}_0(\Sigma|\Psi^{(t)}) = -\frac{1}{2} \{NK\log(2\pi) + N\log \det\Sigma + N\text{tr}[\Sigma^{-1}S^*]\},$$

where  $\text{tr}[\cdot]$  denotes the trace operator of a matrix, where

$$S^* = N^{-1} \sum_{i=1}^N \sum_{\theta_i \in \mathcal{G}} \theta_i \theta_i^T \cdot \tilde{p}(\theta_i|y_i, \Psi^{(t)}). \tag{13}$$

Therefore, the optimization problem in (11) is known as a semi-definite programming problem in convex optimization. We can obtain the  $\Sigma^{(t+1)}$  in the same way as Zhang et al. [36] by applying a proximal gradient descent algorithm [37]. It is noteworthy that in the EM algorithm used by Sun et al. [12],  $Q_0$  is a constant and thus need not be optimized, as  $\Sigma$  is assumed to be known.

For maximization problem (12), it is noted that  $\tilde{Q}_j$  in Eq (8) can be regarded as the weighted  $L_1$ -penalized log-likelihood in logistic regression with naive augmented data  $(y_{ij}, \theta_i)$  and weights  $\tilde{p}(\theta_i|y_i, \mathbf{A}^{(t)}, \mathbf{b}^{(t)})$ , where  $\theta_i \in \mathcal{G}$ . Hence, the maximization problem in (Eq 12) is equivalent to the variable selection in logistic regression based on the  $L_1$ -penalized likelihood. Several existing methods such as the coordinate decent algorithm [24] can be directly used.

After solving the maximization problems in Eqs (11) and (12), it is straightforward to obtain the parameter estimates of  $\Sigma^{(t+1)}$ ,  $\mathbf{A}^{(t+1)} = (\mathbf{a}_1^{(t+1)}, \dots, \mathbf{a}_J^{(t+1)})^T$  and  $\mathbf{b}^{(t+1)} = (b_1^{(t+1)}, \dots, b_J^{(t+1)})^T$  for the next iteration.

We call the implementation described in this subsection the naive version since the M-step suffers from a high computational burden. It should be noted that the computational complexity of the coordinate descent algorithm for maximization problem (12) in the M-step is proportional to the sample size of the data set used in the logistic regression [24]. In (12), the sample size (i.e.,  $N \times G$ ) of the naive augmented data set  $\{(y_{ij}, \theta_i) | i = 1, \dots, N, \text{ and } \theta_i \in \mathcal{G}\}$  is usually large, where  $G$  is the number of quadrature grid points in  $\mathcal{G}$ . For example, if  $N = 1000$ ,  $K = 3$  and 11 quadrature grid points are used in each latent trait dimension, then  $G = 1331$  and  $N \times G = 1.331 \times 10^6$ . This leads to a heavy computational burden for maximizing (12) in the M-step. As a result, the EML1 developed by Sun et al. [12] is computationally expensive.

### 3.2 An improved EM-based $L_1$ -penalized likelihood method

In this subsection, motivated by the idea about “artificial data” widely used in maximum marginal likelihood estimation in the IRT literature [30], we will derive another form of weighted log-likelihood based on a new artificial data set with size  $2 \times G$ . Therefore, the computational complexity of the M-step is reduced to  $O(2 \times G)$  from  $O(N \times G)$ .

As described in Section 3.1.1, we use the same set  $\mathcal{G}$  of fixed grid points for all  $\theta_i$ s to approximate the conditional expectation. Let  $\mathcal{G} = \{\theta^{(g)}, g = 1, \dots, G\}$  with  $\theta^{(g)}$  representing a discrete ability level, and  $\tilde{p}(\theta^{(g)}|\mathbf{y}_i, \Psi^{(t)})$  denote the value of  $\tilde{p}(\theta_i|\mathbf{y}_i, \Psi^{(t)})$  at  $\theta_i = \theta^{(g)}$ . Using the traditional “artificial data” described in Baker and Kim [30], we can write  $\tilde{Q}_j$  as

$$\begin{aligned} & \tilde{Q}_j(\mathbf{a}_j, b_j|\Psi^{(t)}) \\ &= \sum_{i=1}^N \sum_{g=1}^G [y_{ij} \log(F_j(\theta^{(g)})) + (1 - y_{ij}) \log(1 - F_j(\theta^{(g)}))] \cdot \tilde{p}(\theta^{(g)}|\mathbf{y}_i, \Psi^{(t)}) - \eta \|\mathbf{a}_j\|_1 \\ &= \sum_{g=1}^G [\log(F_j(\theta^{(g)})) \cdot r_{jg}^{(t)} + \log(1 - F_j(\theta^{(g)})) \cdot (f_g^{(t)} - r_{jg}^{(t)})] - \eta \|\mathbf{a}_j\|_1, \end{aligned} \tag{14}$$

where  $f_g^{(t)} = \sum_{i=1}^N \tilde{p}(\theta^{(g)}|\mathbf{y}_i, \Psi^{(t)})$  is the “expected sample size” at ability level  $\theta^{(g)}$ , and  $r_{jg}^{(t)} = \sum_{i=1}^N y_{ij} \tilde{p}(\theta^{(g)}|\mathbf{y}_i, \Psi^{(t)})$  is the “expected frequency” of correct response to item  $j$  at ability  $\theta^{(g)}$ . Note that, in the IRT literature,  $f_g^{(t)}$  and  $r_{jg}^{(t)}$  are known as “artificial data”, and they are applied to replace the unobservable sufficient statistics in the complete data likelihood equation in the E-step of the EM algorithm for computing maximum marginal likelihood estimation [30–32]. If  $\eta = 0$ , differentiating Eq (14), we can obtain a likelihood equation involving the traditional “artificial data”, which can be solved by standard optimization methods [30, 32].

For  $L_1$ -penalized log-likelihood estimation, we should maximize Eq (14) for  $\eta > 0$ . Although the coordinate descent algorithm [24] can be applied to maximize Eq (14), some technical details are needed. In this paper, from a novel perspective, we will view  $\tilde{Q}_j(\mathbf{a}_j, b_j|\Psi^{(t)})$  as a weighted  $L_1$ -penalized log-likelihood of logistic regression based on our new artificial data inspired by Ibrahim 1990 [33] and maximize  $\tilde{Q}_j(\mathbf{a}_j, b_j|\Psi^{(t)})$  by applying the efficient R package `glmnet` [24].

Specifically, we group the  $N \times G$  naive augmented data in Eq (8) into  $2 \times G$  new artificial data  $(z, \theta^{(g)})$ , where  $z$  (equals to 0 or 1) is the response to item  $j$  and  $\theta^{(g)}$  is a discrete ability level. Thus,  $\tilde{Q}_j$  in Eq (8) can be rewritten as

$$\begin{aligned} & \tilde{Q}_j(\mathbf{a}_j, b_j|\Psi^{(t)}) \\ &= \sum_{i=1}^N \sum_{g=1}^G \sum_{z=0,1} I(y_{ij} = z) [z \log(F_j(\theta^{(g)})) + (1 - z) \log(1 - F_j(\theta^{(g)}))] \cdot \tilde{p}(\theta^{(g)}|\mathbf{y}_i, \Psi^{(t)}) - \eta \|\mathbf{a}_j\|_1 \\ &= \sum_{g=1}^G \sum_{z=0,1} [z \log(F_j(\theta^{(g)})) + (1 - z) \log(1 - F_j(\theta^{(g)}))] \cdot \sum_{i=1}^N I(y_{ij} = z) \tilde{p}(\theta^{(g)}|\mathbf{y}_i, \Psi^{(t)}) - \eta \|\mathbf{a}_j\|_1 \\ &= \sum_{g=1}^G \sum_{z=0,1} [z \log(F_j(\theta^{(g)})) + (1 - z) \log(1 - F_j(\theta^{(g)}))] \cdot w_j^{(t)}(z, \theta^{(g)}) - \eta \|\mathbf{a}_j\|_1, \end{aligned} \tag{15}$$

where  $w_j^{(t)}(z, \theta^{(g)}) = \sum_{i=1}^N I(y_{ij} = z) \tilde{p}(\theta^{(g)}|\mathbf{y}_i, \Psi^{(t)})$  is the “expected frequency” of correct or incorrect response to item  $j$  at ability  $\theta^{(g)}$ . The second equality in Eq (15) holds since  $z$  and  $F_j(\theta^{(g)})$  do not depend on  $y_{ij}$  and the order of the summation is interchanged. Thus, we obtain a new form of weighted  $L_1$ -penalized log-likelihood of logistic regression in the last line of Eq (15) based on the new artificial data  $(z, \theta^{(g)})$  with a weight  $w_j^{(t)}(z, \theta^{(g)})$ . Note that  $w_j^{(t)}(1, \theta^{(g)}) = r_{jg}^{(t)}$  and  $w_j^{(t)}(0, \theta^{(g)}) = f_g^{(t)} - r_{jg}^{(t)}$ , so the traditional “artificial data” can be viewed as weights for our new artificial data  $(z, \theta^{(g)})$ .



Since Eq (15) is a weighted  $L_1$ -penalized log-likelihood of logistic regression, it can be optimized directly via the efficient R package `glmnet` [24]. This is an advantage of using Eq (15) instead of Eq (14). Moreover, the size of the new artificial data set  $\{(z, \theta^{(g)}) | z = 0, 1, \text{ and } \theta^{(g)} \in \mathcal{G}\}$  involved in Eq (15) is  $2 \times G$ , which is substantially smaller than  $N \times G$ . This significantly reduces the computational burden for optimizing  $\tilde{Q}_j$  in the M-step. We call this version of EM as the improved EML1 (IEML1). Since the computational complexity of the coordinate descent algorithm is  $O(M)$  where  $M$  is the sample size of data involved in penalized log-likelihood [24], the computational complexity of M-step of IEML1 is reduced to  $O(2 \times G)$  from  $O(N \times G)$ .

It is noteworthy that, for  $y_i = y_{i'}$  with the same response pattern, the posterior distribution of  $\theta_i$  is the same as that of  $\theta_{i'}$ , i.e.,  $\tilde{p}(\theta^{(g)} | y_i, \Psi^{(t)}) = \tilde{p}(\theta^{(g)} | y_{i'}, \Psi^{(t)})$ . When the sample size  $N$  is large, the item response vectors  $y_1, \dots, y_N$  can be grouped into distinct response patterns, and then the summation in computing  $w_j^{(t)}(z, \theta^{(g)})$  is not over  $N$ , but over the number of distinct patterns, which will greatly reduce the computational time [30].

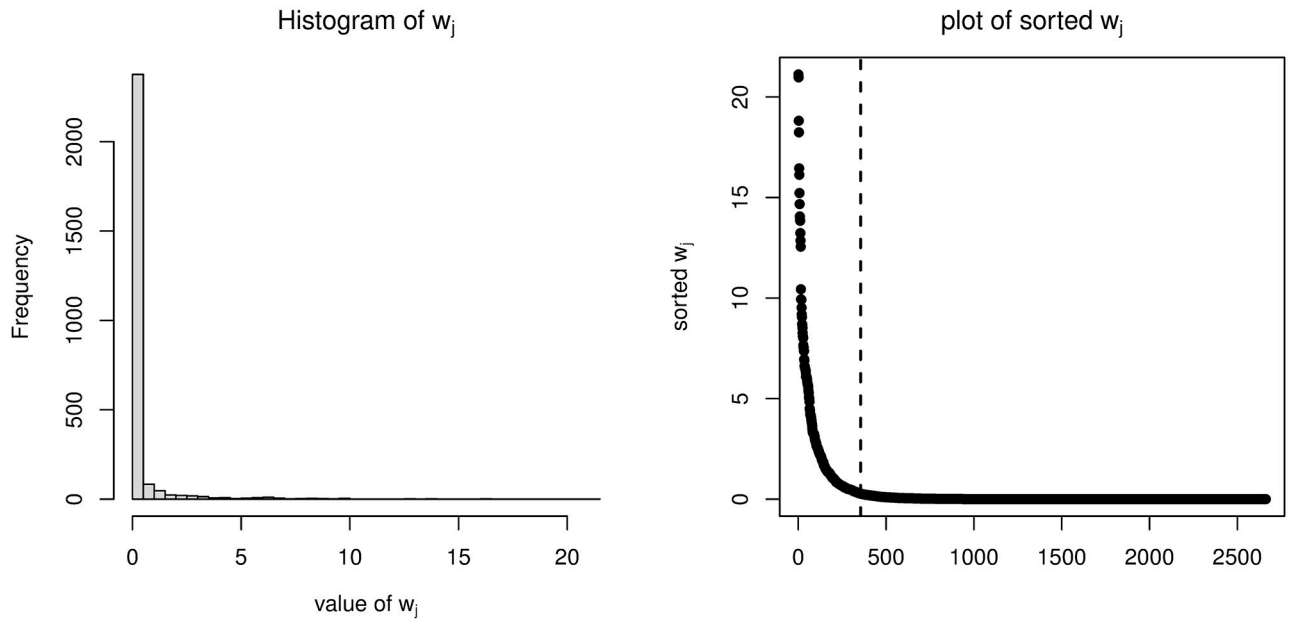
It should be noted that any fixed quadrature grid points set, such as Gaussian-Hermite quadrature points set, will result in the same weighted  $L_1$ -penalized log-likelihood as in Eq (15). However, neither the adaptive Gaussian-Hermite quadrature [34] nor the Monte Carlo integration [35] will result in Eq (15) since the adaptive Gaussian-Hermite quadrature requires different adaptive quadrature grid points for different  $\theta_i$  while the Monte Carlo integration usually draws different Monte Carlo samples for different  $\theta_i$ .

### 3.3 Heuristic approach for choosing grid points

In the new weighted log-likelihood in Eq (15), the more artificial data  $(z, \theta^{(g)})$  are used, the more accurate the approximation of  $\tilde{Q}_j$  is; but, the more computational burden IEML1 has. To reduce the computational burden of IEML1 without sacrificing too much accuracy, we will give a heuristic approach for choosing a few grid points used to compute  $\tilde{Q}_j$ .

Let us consider a motivating example based on a M2PL model with item discrimination parameter matrix  $A_1$  with  $K = 3$  and  $J = 40$ , which is given in Table A in S1 Appendix. The grid point set  $\mathcal{G} = \mathcal{S}^3$ , where  $\mathcal{S}$  denotes a set of equally spaced 11 grid points on the interval  $[-4, 4]$ . Therefore, the size of our new artificial data set used in Eq (15) is  $2 \times 11^3 = 2662$ . Based on one iteration of the EM algorithm for one simulated data set, we calculate the weights of the new artificial data  $(z, \theta^{(g)}) \in \{0, 1\} \times \mathcal{G}$  and then sort them in descending order.

Fig 1 (left) gives the histogram of all weights, which shows that most of the weights are very small and only a few of them are relatively large. Fig 1 (right) gives the plot of the sorted weights, in which the top 355 sorted weights are bounded by the dashed line. The sum of the top 355 weights constitutes 95.9% of the sum of all the 2662 weights. This suggests that only a few  $(z, \theta^{(g)})$  contribute significantly to  $\tilde{Q}_j$ . Furthermore, Fig 2 presents scatter plots of our artificial data  $(z, \theta^{(g)})$ , in which the darker the color of  $(z, \theta^{(g)})$ , the greater the weight  $w_j^{(t)}(z, \theta^{(g)})$ . It can be seen roughly that most  $(z, \theta^{(g)})$  with greater weights are included in  $\{0, 1\} \times [-2.4, 2.4]^3$ . In fact, artificial data with the top 355 sorted weights in Fig 1 (right) are all in  $\{0, 1\} \times [-2.4, 2.4]^3$ . These observations suggest that we should use a reduced grid point set  $\mathcal{G}^*$  with each dimension consisting of 7 equally spaced grid points on the interval  $[-2.4, 2.4]$ . Thus, the size of the corresponding reduced artificial data set is  $2 \times 7^3 = 686$ . In this way, only 686 artificial data are required in the new weighted log-likelihood in Eq (15). Our simulation studies show that IEML1 with this reduced artificial data set performs well in terms of correctly selected latent variables and computing time.



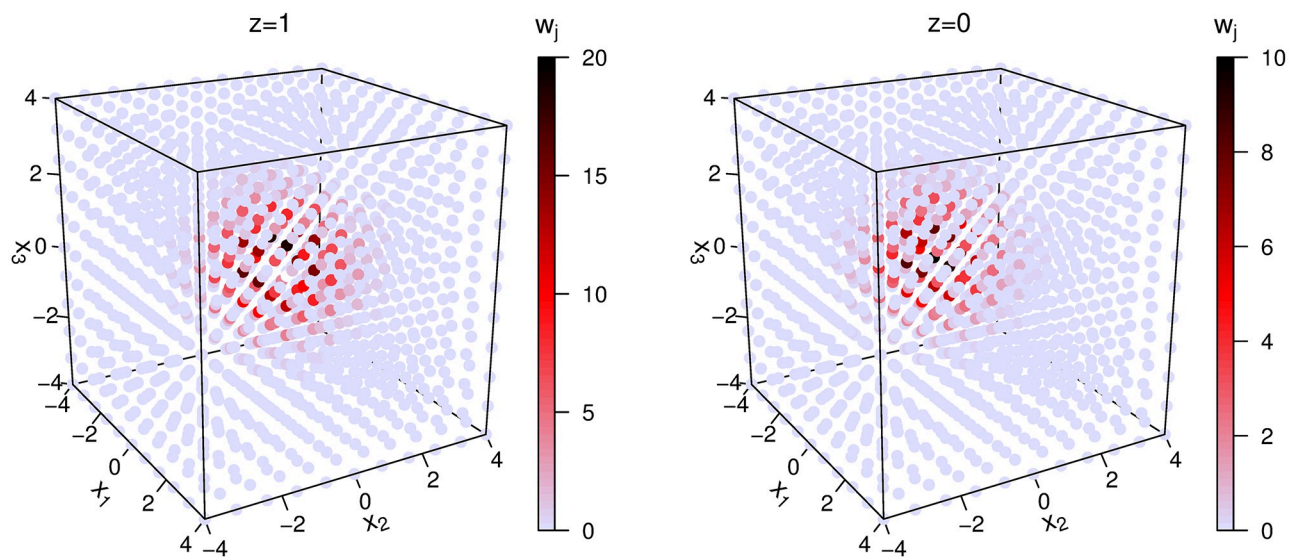
**Fig 1. Histogram of  $w_j$  (left column) and plot of sorted  $w_j$  (right column).**

<https://doi.org/10.1371/journal.pone.0279918.g001>

In the literature, Xu et al. [26] gives a similar approach to choose the naive augmented data  $(y_{ij}, \theta_i)$  with larger weight  $\tilde{p}(\theta_i | y_i, \Psi^{(t)})$  for computing Eq (8). In this paper, we however choose our new artificial data  $(z, \theta^{(g)})$  with larger weight  $w_j^{(t)}(z, \theta^{(g)})$  to compute Eq (15).

### 4 Simulation studies

In this section, we conduct simulation studies to evaluate and compare the performance of our IEML1, the EML1 proposed by Sun et al. [12] and the constrained exploratory IFAs with hard-



**Fig 2. Scatter plots of the grid points with the weights  $w_j$  under  $z = 1$  (left column) and  $z = 0$  (right column).**

<https://doi.org/10.1371/journal.pone.0279918.g002>

threshold and optimal threshold. In all methods, we use the same identification constraints described in subsection 2.1 to resolve the rotational indeterminacy. In addition, we also give simulation studies to show the performance of the heuristic approach for choosing grid points. The R codes of the IEML1 method are provided in [S4 Appendix](#).

Here, we consider three M2PL models with the item number  $J$  equal to 40. Three true discrimination parameter matrices  $A_1, A_2$  and  $A_3$  with  $K = 3, 4, 5$  are shown in Tables A, C and E in [S1 Appendix](#), respectively. The corresponding difficulty parameters  $b_1, b_2$  and  $b_3$  are listed in Tables B, D and F in [S1 Appendix](#). The non-zero discrimination parameters are generated from the identically independent uniform distribution  $U(0.5, 2)$ . The true difficulty parameters are generated from the standard normal distribution. The diagonal elements of the true covariance matrix  $\Sigma$  of the latent traits are setting to be unity with all off-diagonals being 0.1.

For parameter identification, we constrain items 1, 10, 19 to be related only to latent traits 1, 2, 3 respectively for  $K = 3$ , that is,  $(a_1, a_{10}, a_{19})^T$  in  $A_1$  was fixed as diagonal matrix in each EM iteration. Similarly, items 1, 7, 13, 19 are related only to latent traits 1, 2, 3, 4 respectively for  $K = 4$  and items 1, 5, 9, 13, 17 are related only to latent traits 1, 2, 3, 4, 5 respectively for  $K = 5$ .

Two sample size (i.e.,  $N = 500, 1000$ ) are considered. For each setting, we draw 100 independent data sets for each M2PL model. We obtain results by IEML1 and EML1 and evaluate their results in terms of computation efficiency, correct rate (CR) for the latent variable selection and accuracy of the parameter estimation. The computation efficiency is measured by the average CPU time over 100 independent runs. The CR for the latent variable selection is defined by the recovery of the loading structure  $\Lambda = (\lambda_{jk})$  as follows:

$$CR = \frac{1}{K(J - K)} \sum_{\substack{1 \leq j \leq J, 1 \leq k \leq K, \\ a_k \text{ is not fixed for identification}}} I(\hat{\lambda}_{jk} = \lambda_{jk}),$$

where  $\hat{\Lambda} = (\hat{\lambda}_{jk})$  is an estimate of the true loading structure  $\Lambda$ . The following mean squared error (MSE) is used to measure the accuracy of the parameter estimation:

$$MSE(a_{jk}) = \frac{1}{S} \sum_{s=1}^S (\hat{a}_{jk}^s - a_{jk})^2,$$

where  $\hat{a}_{jk}^s$  denotes the estimate of  $a_{jk}$  from the  $s$ th replication and  $S = 100$  is the number of data sets. The MSE of each  $b_j$  in  $\mathbf{b}$  and  $\sigma_{kk'}$  in  $\Sigma$  is calculated similarly to that of  $a_{jk}$ .

### 4.1 Computational efficiency

We first compare computational efficiency of IEML1 and EML1. To make a fair comparison, the covariance of latent traits  $\Sigma$  is assumed to be known for both methods in this subsection.

In this study, we consider M2PL with  $A_1$ . We use the fixed grid point set  $\mathcal{G} = \mathcal{S} \times \mathcal{S} \times \mathcal{S}$ , where  $\mathcal{S}$  is the set of equally spaced 11 grid points on the interval  $[4, 4]$ . In each M-step, the maximization problem in (12) is solved by the R-package `glmnet` for both methods. Due to tedious computing time of EML1, we only run the two methods on 10 data sets. For each replication, the initial value of  $(\mathbf{a}_1, \mathbf{a}_{10}, \mathbf{a}_{19})^T$  is set as identity matrix, and other initial values in  $\mathbf{A}$  are set as  $1/J = 0.025$ . The initial value of  $\mathbf{b}$  is set as the zero vector. The candidate tuning parameters are given as  $(0.10, 0.09, \dots, 0.01) \times N$ , and we choose the best tuning parameter by Bayesian information criterion as described by Sun et al. [12].

The average CPU time (in seconds) for IEML1 and EML1 are given in [Table 1](#). From [Table 1](#), IEML1 runs at least 30 times faster than EML1. Moreover, IEML1 and EML1 yield

**Table 1. The average CPU time in seconds for IEML1 and EML1 under  $K = 3$  and  $J = 40$ .**

	IEML1	EML1
$N = 500$	91	3019
$N = 1000$	114	6553

<https://doi.org/10.1371/journal.pone.0279918.t001>

comparable results with the absolute error no more than  $10^{-13}$ . It numerically verifies that two methods are equivalent.

## 4.2 Simulation for the unknown $\Sigma$ case

In this subsection, we compare our IEML1 with a two-stage method proposed by Sun et al. [12], a constrained exploratory IFA with hard threshold (EIFAthr) and a constrained exploratory IFA with optimal threshold (EIFAopt). In the EIFAthr, all parameters are estimated via a constrained exploratory analysis satisfying the identification conditions, and then the estimated discrimination parameters that smaller than a given threshold are truncated to be zero. In the simulation studies, several thresholds, i.e., 0.30, 0.35, . . . , 0.70, are used, and the corresponding EIFAthr are denoted by EIFA0.30, EIFA0.35, . . . , EIFA0.70, respectively. In EIFAthr, it is subjective to preset a threshold, while in EIFAopt we further choose the optimal truncated estimates corresponding to the optimal threshold with minimum BIC value from several given thresholds (e.g., 0.30, 0.35, . . . , 0.70 used in EIFAthr) in a data-driven manner.

For IEML1, the initial value of  $\Sigma$  is set to be an identity matrix. For other three methods, a constrained exploratory IFA is adopted to estimate  $\Sigma$  first by R-package `mirt` with the setting being “method = EM” and the same grid points are set as in subsection 4.1.

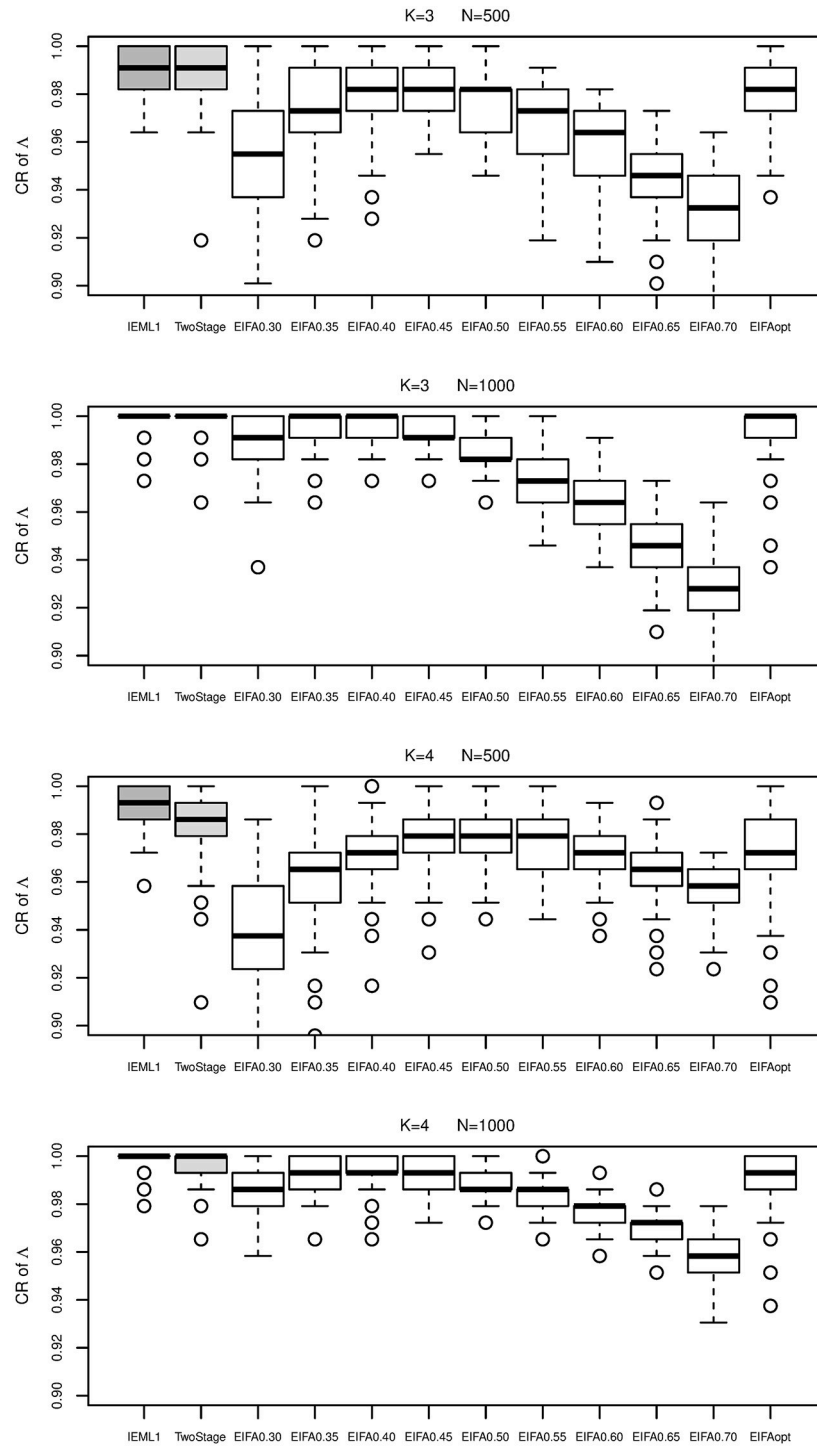
We consider M2PL models with  $A_1$  and  $A_2$  in this study. To compare the latent variable selection performance of all methods, the boxplots of CR are displayed in Fig 3. From Fig 3, IEML1 performs the best and then followed by the two-stage method. As we expect, different hard thresholds leads to different estimates and the resulting different CR, and it would be difficult to choose a best hard threshold in practices. EIFAopt performs better than EIFAthr. As complements to CR, the false negative rate (FNR), false positive rate (FPR) and precision are reported in S2 Appendix. The boxplots of these metrics show that our IEML1 has very good performance overall.

Fig 4 presents boxplots of the MSE of  $A$  obtained by all methods. From Fig 4, IEML1 and the two-stage method perform similarly, and better than EIFAthr and EIFAopt. We can see that larger threshold leads to smaller median of MSE, but some very large MSEs in EIFAthr.

Figs 5 and 6 show boxplots of the MSE of  $b$  and  $\Sigma$  obtained by all methods. Note that, EIFAthr and EIFAopt obtain the same estimates of  $b$  and  $\Sigma$ , and consequently, they produce the same MSE of  $b$  and  $\Sigma$ . Therefore, their boxplots of  $b$  and  $\Sigma$  are the same and they are represented by “EIFA” in Figs 5 and 6. We can see that all methods obtain very similar estimates of  $b$ . IEML1 gives significant better estimates of  $\Sigma$  than other methods.

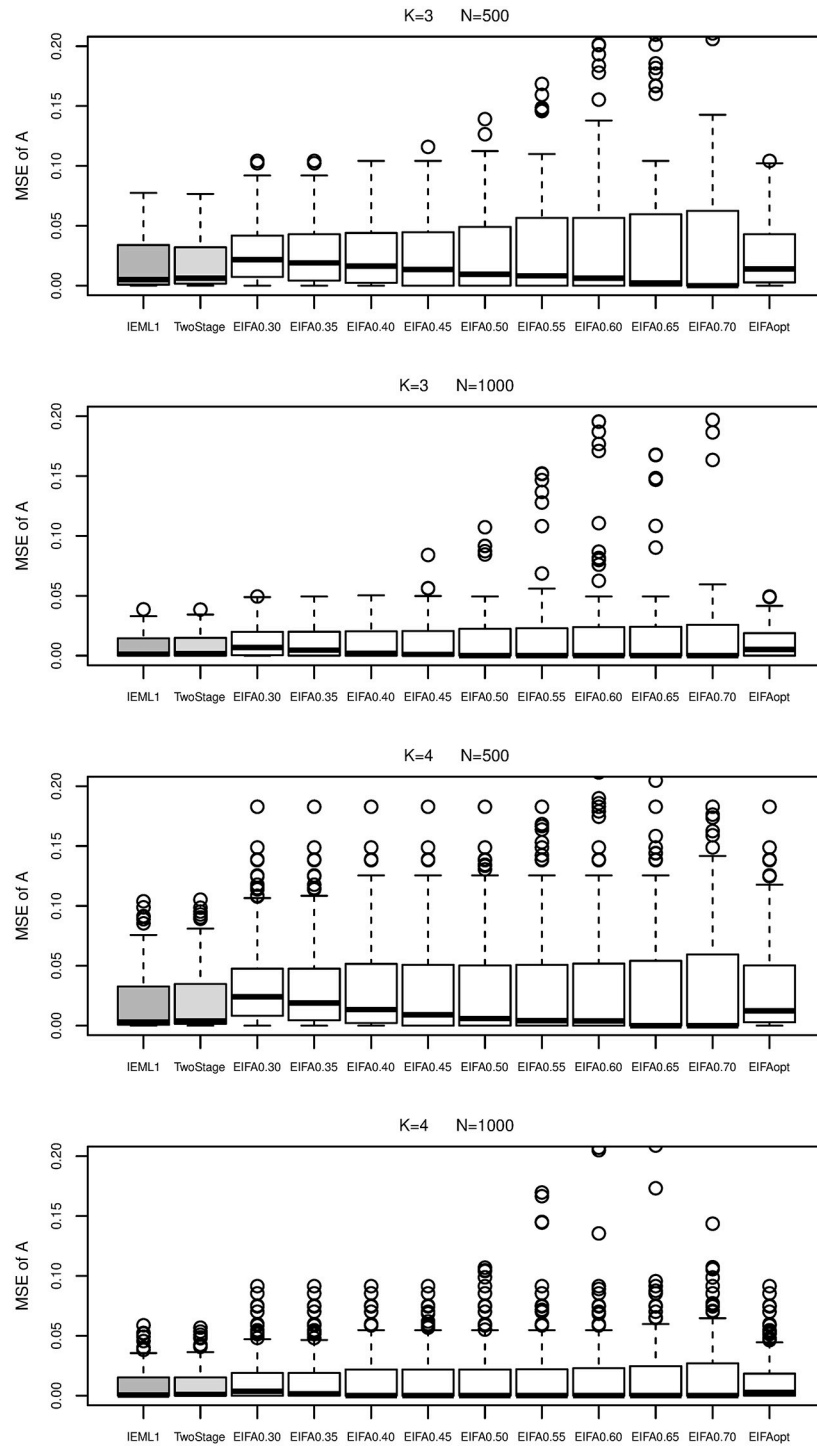
## 4.3 Evaluation on heuristic approach for choosing grid points

As presented in the motivating example in Section 3.3, most of the grid points with larger weights are distributed in the cube  $[-2.4, 2.4]^3$ . Intuitively, the grid points for each latent trait dimension can be drawn from the interval  $[-2.4, 2.4]$ . In this subsection, we generate three grid point sets denoted by Grid11, Grid7 and Grid5 and compare the performance of IEML1 based on these three grid point sets via simulation study. Specifically, Grid11, Grid7 and Grid5 are three  $K$ -ary Cartesian power, where 11, 7 and 5 equally spaced grid points on the intervals  $[-4, 4]$ ,  $[-2.4, 2.4]$  and  $[-2.4, 2.4]$  in each latent trait dimension, respectively.



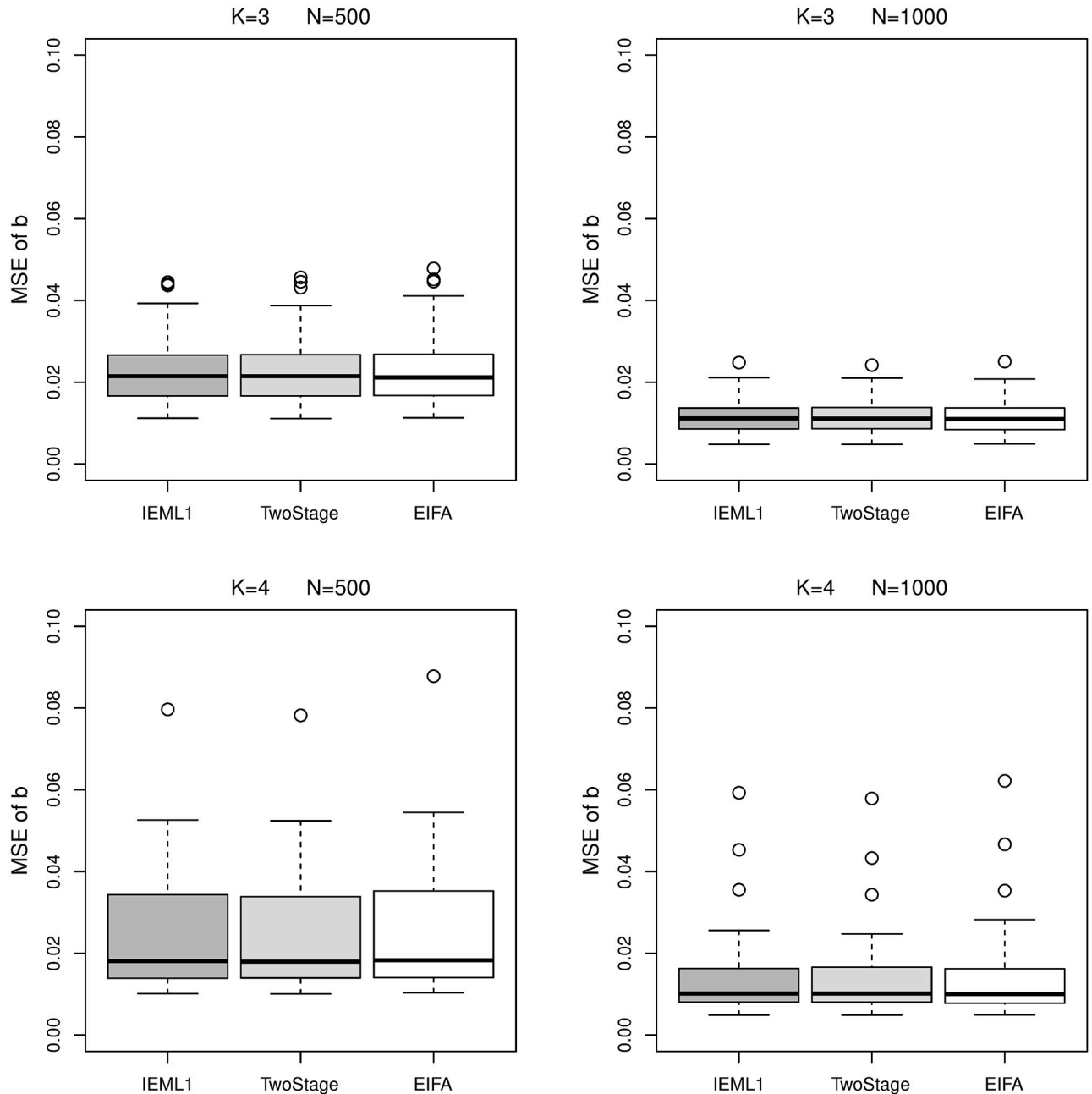
**Fig 3. Boxplots of the correct rate of  $\lambda$  obtained by IEML1 (dark gray boxes), two-stage (light gray boxes), EIFAthr and EIFAopt (white boxes) for  $K = 3$  and  $4$  under sample size  $N = 500$  and  $1000$ .**

<https://doi.org/10.1371/journal.pone.0279918.g003>



**Fig 4. Boxplots of the MSE of  $A$**  obtained by IEMML1 (dark gray boxes), two-stage (light gray boxes), EIFAthr and EIFAOpt (white boxes) for  $K = 3$  and  $4$  under sample size  $N = 500$  and  $1000$ .

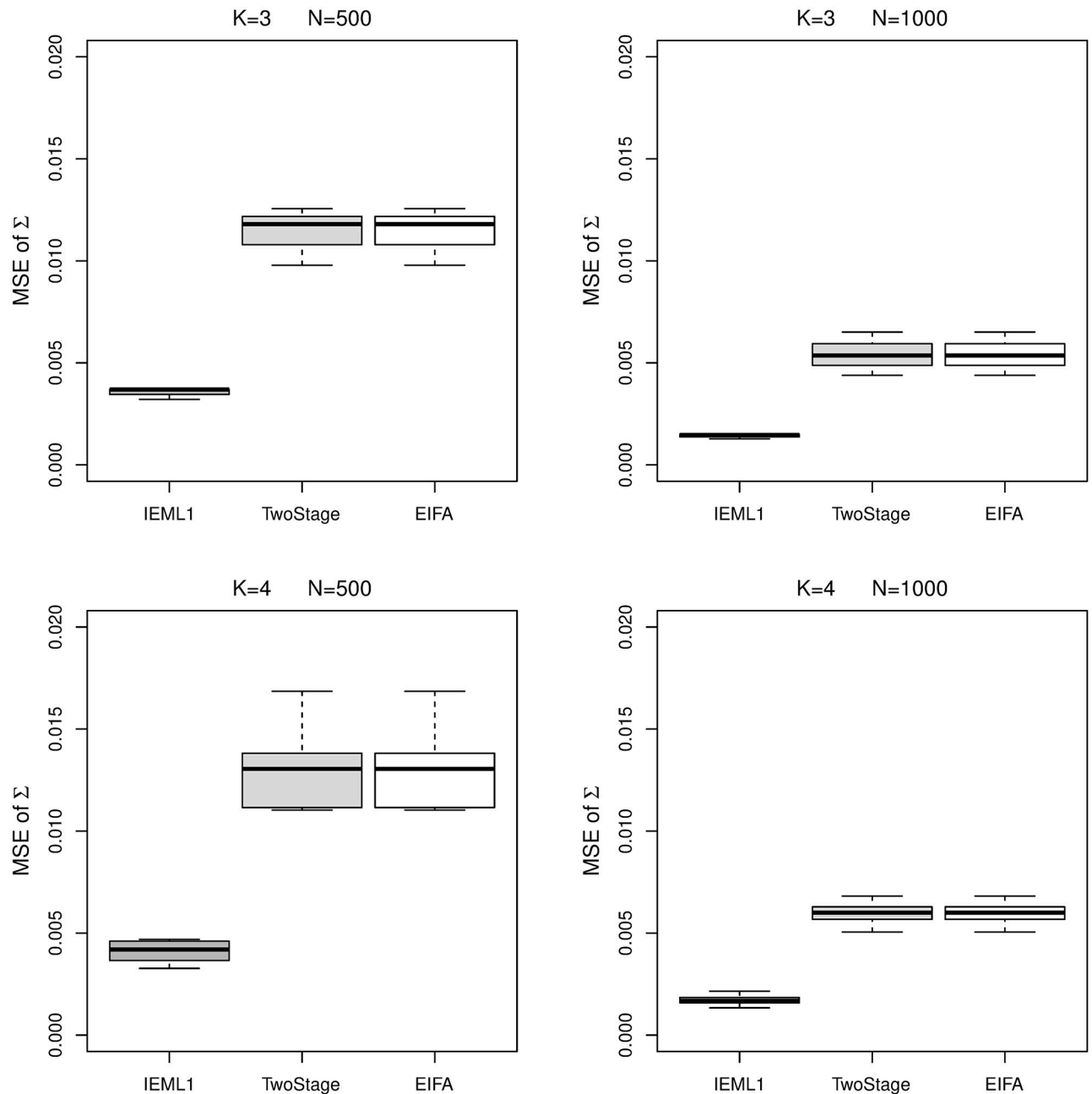
<https://doi.org/10.1371/journal.pone.0279918.g004>



**Fig 5. Boxplots of the MSE of  $\mathbf{b}$  obtained by IEML1 (dark gray boxes), two-stage (light gray boxes), EIFAthr and EIFAopt (white boxes) for  $K = 3$  and  $4$  under sample size  $N = 500$  and  $1000$ .**

<https://doi.org/10.1371/journal.pone.0279918.g005>

Fig 7 summarizes the boxplots of CRs and MSE of parameter estimates by IEML1 for all cases. From Fig 7, we obtain very similar results when Grid11, Grid7 and Grid5 are used in IEML1. Table 2 shows the average CPU time for all cases. The computing time increases with the sample size and the number of latent traits. The simulation studies show that IEML1 can give quite good results in several minutes if Grid5 is used for M2PL with  $K \leq 5$  latent traits.

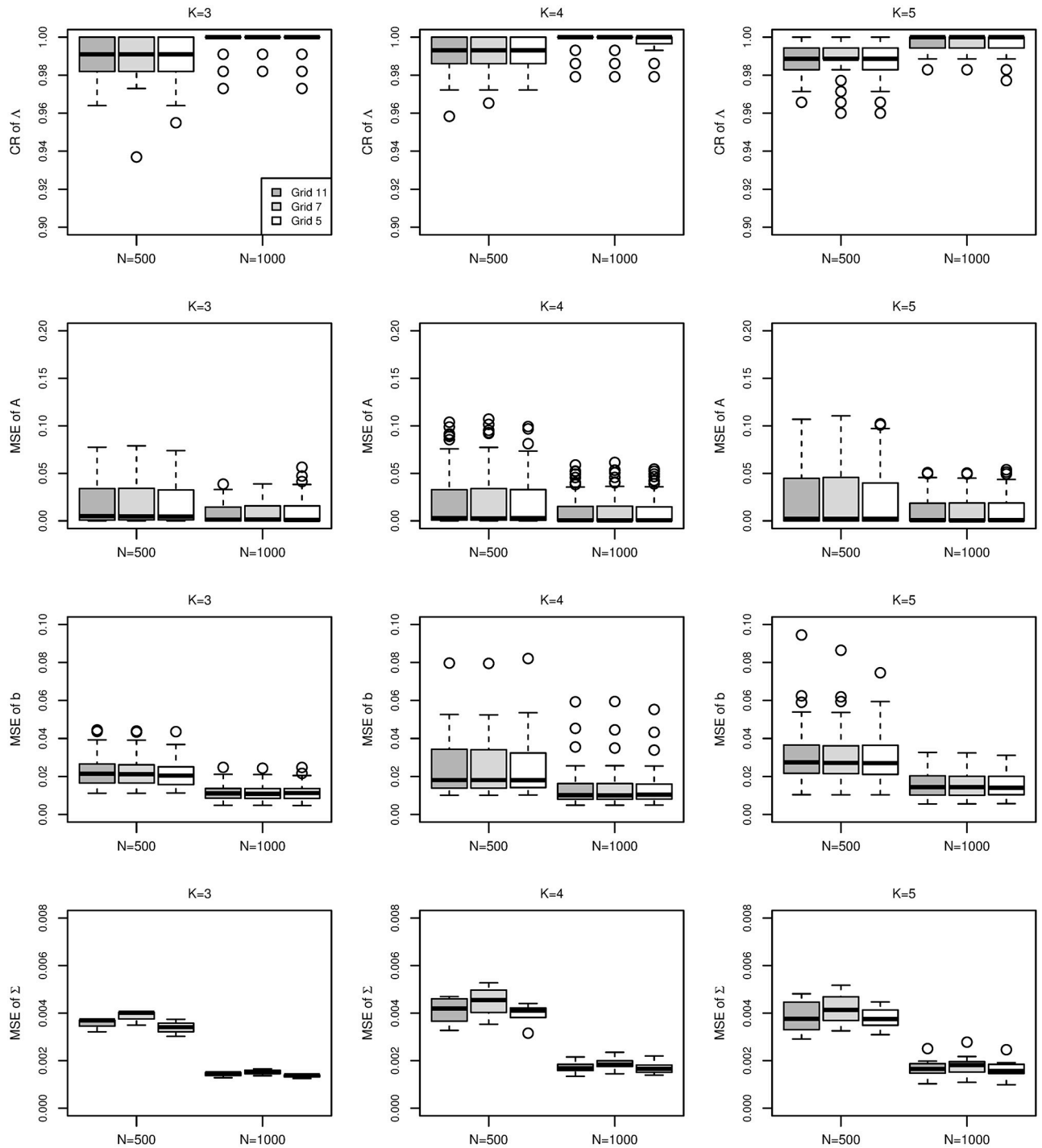


**Fig 6.** Boxplots of the MSE of  $\Sigma$  obtained by IEMML1 (dark gray boxes), two-stage (light gray boxes), EIFAthr and EIFAopt (white boxes) for  $K = 3$  and 4 under sample size  $N = 500$  and 1000.

<https://doi.org/10.1371/journal.pone.0279918.g006>

In fact, we also try to use grid point set Grid3 in which each dimension uses three grid points equally spaced in interval  $[-2.4, 2.4]$ . But the numerical quadrature with Grid3 is not good enough to approximate the conditional expectation in the E-step. It should be noted that IEMML1 may depend on the initial values. In all simulation studies, we use the initial values similarly as described for  $A_1$  in subsection 4.1. These initial values result in quite good results and they are good enough for practical users in real data applications.





**Fig 7. Boxplots of the correct rate of  $\lambda$  (row 1), the MSE of  $A$  (row 2), the MSE of  $b$  (row 3) and the MSE of  $\Sigma$  (row 4) for  $K = 3$  (column 1), 4 (column 2) and 5 (column 3) under sample size  $N = 500$  and 1000. The dark gray boxes, light gray boxes and white boxes represent the results via 11, 7 and 5 grid points per dimension respectively.**

<https://doi.org/10.1371/journal.pone.0279918.g007>

Table 2. The average CPU time in seconds of IEML1 with Grid11, Grid7 and Grid5 under  $K = 3, 4, 5$  with sample size  $N = 500, 1000$ .

	$K = 3$		$K = 4$		$K = 5$	
	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
Grid 11	108	134	768	961	9230	14804
Grid 7	68	66	193	287	953	1155
Grid 5	38	53	80	86	209	267

<https://doi.org/10.1371/journal.pone.0279918.t002>

## 5 Real data analysis

In this section, we analyze a data set of the Eysenck Personality Questionnaire given in Eysenck and Barrett [38]. The data set includes 754 Canadian females' responses (after eliminating subjects with missing data) to 69 dichotomous items, where items 1–25 consist of the psychoticism (P), items 26–46 consist of the extraversion (E) and items 47–69 consist of the neuroticism (N). This data set was also analyzed in Xu et al. [26]. In order to guarantee the psychometric properties of the items, we select those items whose corrected item-total correlation values are greater than 0.2 [39]. The selected items and their original indices are listed in Table 3, with 10, 19 and 23 items corresponding to P, E and N respectively. Items marked by asterisk correspond to negatively worded items whose original scores have been reversed.

In the analysis, we designate two items related to each factor for identifiability. Based on the meaning of the items and previous research, we specify items 1 and 9 to P, items 14 and 15 to E, items 32 and 34 to N. We employ the IEML1 to estimate the loading structure and then compute the observed BIC under each candidate tuning parameters in  $(0.040, 0.038, 0.036, \dots, 0.002) \times N$ , where  $N$  denotes the sample size 754. The minimal BIC value is 38902.46 corresponding to  $\eta = 0.02 \times N$ . The parameter estimates of  $\mathbf{A}$  and  $\mathbf{b}$  are given in Table 4, and the estimate of  $\Sigma$  is

$$\hat{\Sigma} = \begin{pmatrix} 1.000 & 0.121 & -0.018 \\ 0.121 & 1.000 & -0.246 \\ -0.018 & -0.246 & 1.000 \end{pmatrix}.$$

From the results, most items are found to remain associated with only one single trait while some items related to more than one trait. Most of these findings are sensible. For example, item 19 ('Would you call yourself happy-go-lucky?') designed for extraversion is also related to neuroticism which reflects individuals' emotional stability. Item 49 ('Do you often feel lonely?') is also related to extraversion whose characteristics are enjoying going out and socializing. In addition, it is reasonable that item 30 ('Does your mood often go up and down?') and item 40 ('Would you call yourself tense or 'highly-strung?') are related to both neuroticism and psychoticism.

## 6 Concluding remarks

In this paper, we obtain a new weighted log-likelihood based on a new artificial data set for M2PL models, and consequently we propose IEML1 to optimize the  $L_1$ -penalized log-likelihood for latent variable selection. We give a heuristic approach for choosing the quadrature points used in numerical quadrature in the E-step, which reduces the computational burden of IEML1 significantly. There are three advantages of IEML1 over EML1, the two-stage method, EIFAthr and EIFAopt. First, the computational complexity of M-step in IEML1 is reduced to  $O(2 \times G)$  from  $O(N \times G)$ . In our simulation studies, IEML1 needs a few minutes for M2PL

**Table 3. The eysenck personality questionnaire items.**

1*	P6	Would being in debt worry you?
2	P22	Would you take drugs which may have strange or dangerous effects?
3	P33	Do you enjoy practical jokes that can sometimes really hurt people?
4*	P37	Do good manners and cleanliness matter much to you?
5	P46	Do people who drive carefully annoy you?
6	P50	Do most things taste the same to you?
7*	P57	Do you like to arrive at appointments in plenty of time?
8	P67	Do you think people spend too much time safeguarding their future with savings and insurances?
9	P74	When you catch a train do you often arrive at the last minute?
10	P83	Would you like other people to be afraid of you?
11	E5	Are you a talkative person?
12	E10	Are you rather lively?
13	E14	Can you usually let yourself go and enjoy yourself at a lively party?
14	E17	Do you enjoy meeting new people?
15*	E21	Do you tend to keep in the background on social occasions?
16	E25	Do you like going out a lot?
17*	E29	Do you prefer reading to meeting people?
18	E32	Do you have many friends?
19	E36	Would you call yourself happy-go-lucky?
20	E40	Do you usually take the initiative in making new friends?
21*	E42	Are you mostly quiet when you are with other people?
22	E45	Can you easily get some life into a rather dull party?
23	E49	Do you like telling jokes and funny stories to your friends?
24	E52	Do you like mixing with people?
25	E60	Do you like doing things in which you have to act quickly?
26	E64	Do you often take on more activities than you have time for?
27	E70	Can you get a party going?
28	E82	Do you like plenty of bustle and excitement around you?
29	E86	Do other people think of you as being very lively?
30	N3	Does your mood often go up and down?
31	N7	Do you ever feel "just miserable" for no reason?
32	N12	Do you often worry about things you should not have done or said?
33	N15	Are you an irritable person?
34	N19	Are your feelings easily hurt?
35	N23	Do you often feel "fed-up"?
36	N27	Are you often troubled about feelings of guilt?
37	N31	Would you call yourself a nervous person?
38	N34	Are you a worrier?
39	N38	Do you worry about awful things that might happen?
40	N41	Would you call yourself tense or "highly-strung"?
41	N47	Do you worry about your health?
42	N54	Do you suffer from sleeplessness?
43	N58	Have you often felt listless and tired for no reason?
44	N62	Do you often feel life is very dull?
45	N66	Do you worry a lot about your looks?
46	N68	Have you ever wished that you were dead?
47	N72	Do you worry too long after an embarrassing experience?
48	N75	Do you suffer from "nerves"?

*(Continued)*

**Table 3.** (Continued)

49	N77	Do you often feel lonely?
50	N80	Are you easily hurt when people find fault with you or the work you do?
51	N84	Are you sometimes bubbling over with energy and sometimes very sluggish?
52	N88	Are you touchy about some things?

<https://doi.org/10.1371/journal.pone.0279918.t003>

models with no more than five latent traits. Second, IEML1 updates covariance matrix  $\Sigma$  of latent traits and gives a more accurate estimate of  $\Sigma$ . Third, IEML1 outperforms the two-stage method, EIFAthr and EIFAopt in terms of CR of the latent variable selection and the MSE for the parameter estimates.

The current study will be extended in the following directions for future research. First, we will generalize IEML1 to multidimensional three-parameter (or four parameter) logistic models that give much attention in recent years. Second, other numerical integration such as Gaussian-Hermite quadrature [4, 29] and adaptive Gaussian-Hermite quadrature [34] can be adopted in the E-step of IEML1. Gaussian-Hermite quadrature uses the same fixed grid point set for each individual and can be easily adopted in the framework of IEML1. However, further simulation results are needed. Compared to the Gaussian-Hermite quadrature, the adaptive Gaussian-Hermite quadrature produces an accurate fast converging solution with as few as

**Table 4.** The parameter estimates by the IEML1 algorithm for the real data.

	A			b		A			b
1	0.723	0.000	0.000	-1.943	27	0.000	1.839	0.000	0.745
2	0.659	0.000	0.000	-2.181	28	0.000	1.311	0.000	1.479
3	1.097	0.000	0.000	-3.510	29	0.000	1.699	0.000	1.469
4	0.966	0.000	0.000	-3.175	30	0.321	0.000	1.167	0.769
5	1.177	0.000	0.000	-1.779	31	0.000	0.000	0.890	0.911
6	0.854	0.000	0.000	-3.093	32	0.000	0.000	1.321	2.082
7	1.689	0.000	0.000	-1.952	33	0.000	0.000	0.887	-1.173
8	0.515	0.000	0.000	-1.079	34	0.000	0.000	1.356	1.264
9	2.030	0.000	0.000	-1.718	35	0.000	-0.153	1.430	0.399
10	1.075	0.000	0.000	-3.316	36	0.000	0.000	1.473	0.393
11	0.000	1.542	0.000	1.128	37	0.000	0.000	1.228	-0.562
12	0.000	1.683	0.000	2.283	38	-0.392	0.000	1.679	1.284
13	0.000	1.397	0.000	1.931	39	0.000	0.000	1.197	0.350
14	0.000	1.324	0.000	2.872	40	0.422	0.000	1.475	-1.440
15	0.000	1.696	0.000	0.702	41	-0.248	0.188	0.751	0.311
16	0.336	0.840	0.000	1.294	42	0.255	0.000	0.784	-0.408
17	0.000	1.143	0.000	2.073	43	0.000	0.000	1.168	0.777
18	0.000	1.200	0.000	2.075	44	0.000	-0.247	1.023	-1.214
19	0.000	0.787	-0.463	-0.133	45	0.000	0.225	1.105	0.587
20	0.000	1.617	0.000	0.716	46	0.000	-0.128	0.663	0.057
21	0.000	1.831	0.000	0.983	47	-0.380	-0.253	1.568	0.898
22	0.000	1.962	0.000	-0.066	48	0.000	0.000	1.749	-0.878
23	0.000	0.970	0.000	2.025	49	0.000	-0.220	1.199	-0.327
24	0.000	2.180	0.000	3.455	50	0.000	-0.188	1.126	0.997
25	0.358	0.631	-0.232	0.541	51	0.000	0.000	0.912	1.889
26	0.245	0.478	0.000	0.584	52	0.000	0.000	0.829	2.085

<https://doi.org/10.1371/journal.pone.0279918.t004>

two points per dimension for estimation of MIRT models [34]. Therefore, the adaptive Gaussian-Hermite quadrature is also potential to be used in penalized likelihood estimation for MIRT models although it is impossible to get our new weighted log-likelihood in Eq (15) due to applying different grid point set for different individual. Third, we will accelerate IEML1 by parallel computing technique for medium-to-large scale variable selection, as [40] produced larger gains in performance for MIRT estimation by applying the parallel computing technique. Fourth, the new weighted log-likelihood on the new artificial data proposed in this paper will be applied to the EMS in [26] to reduce the computational complexity for the MS-step.

## Supporting information

**S1 Appendix. True discrimination and difficulty parameters in simulation studies.**

(PDF)

**S2 Appendix. FNR, FPR and precision of the loading structure in the simulation for the unknown  $\Sigma$  case.**

(PDF)

**S3 Appendix. Data sets of the study.**

(PDF)

**S4 Appendix. R codes of IEML1.**

(PDF)

## Author Contributions

**Conceptualization:** Ping-Feng Xu.

**Formal analysis:** Laixu Shang.

**Funding acquisition:** Ping-Feng Xu, Na Shan, George To-Sum Ho.

**Methodology:** Laixu Shang, Ping-Feng Xu, Na Shan, Man-Lai Tang.

**Resources:** Ping-Feng Xu.

**Software:** Laixu Shang.

**Supervision:** Ping-Feng Xu.

**Visualization:** Laixu Shang.

**Writing – original draft:** Laixu Shang.

**Writing – review & editing:** Ping-Feng Xu, Na Shan, Man-Lai Tang, George To-Sum Ho.

## References

1. Reckase MD. Multidimensional Item Response Theory. 1st ed. New York: Springer; 2009.
2. Janssen R, De Boeck P. Confirmatory analyses of componential test structure using multidimensional item response theory. *Multivariate Behavioral Research*. 1999; 34(2): 245–268. <https://doi.org/10.1207/S15327906Mb340205> PMID: 26753937
3. Mckinley R. Confirmatory analysis of test structure using multidimensional item response theory. *ETS Research Report Series*. 1989; 2: i–40.
4. Bock RD, Gibbons R, Muraki E. Full-information item factor analysis. *Applied Psychological Measurement*. 1988; 12(3): 261–280. <https://doi.org/10.1177/014662168801200305>
5. Béguin AA, Glas CAW. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*. 2001; 66(4): 541–561. <https://doi.org/10.1007/BF02296195>

6. da Silva MA, Liu R, Huggins-Manley AC, Bazán JL. Incorporating the Q-matrix into multidimensional item response theory models. *Educational and Psychological Measurement*. 2019; 79(4): 665–687. <https://doi.org/10.1177/0013164418814898> PMID: 32655178
7. Cai L. High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*. 2010; 75(1): 33–57. <https://doi.org/10.1007/s11336-009-9136-x>
8. Benaards CA, Jennrich RI. Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*. 2005; 65(5): 676–696. <https://doi.org/10.1177/0013164404272507>
9. Browne MW. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*. 2001; 36(1): 111–150. [https://doi.org/10.1207/S15327906MBR3601\\_05](https://doi.org/10.1207/S15327906MBR3601_05)
10. Sass DA, Schmitt TA. A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*. 2010; 45(1): 73–103. <https://doi.org/10.1080/00273170903504810> PMID: 26789085
11. Jin S, Moustaki I, Yang-Wallentin F. Approximated penalized maximum likelihood for exploratory factor analysis: An orthogonal case. *Psychometrika*. 2018; 83(3): 628–649. <https://doi.org/10.1007/s11336-018-9623-z> PMID: 29876715
12. Sun J, Chen Y, Liu J, Ying Z, Xin T. Latent variable selection for multidimensional item response theory models via  $L_1$  regularization. *Psychometrika*. 2016; 81(4): 921–939. <https://doi.org/10.1007/s11336-016-9529-6>
13. Hui FKC, Tanaka E, Warton DI. Order selection and sparsity in latent variable models via the ordered factor LASSO. *Biometrics*. 2018; 74(4): 1311–1319. <https://doi.org/10.1111/biom.12888> PMID: 29750847
14. Scharf F, Nestler S. Should regularization replace simple structure rotation in exploratory factor analysis? *Structural Equation Modeling: A Multidisciplinary Journal*. 2019; 26(4): 576–590. <https://doi.org/10.1080/10705511.2018.1558060>
15. Hirose K, Konishi S. Variable selection via the weighted group lasso for factor analysis models. *The Canadian Journal of Statistics*. 2012; 40(2): 345–361. <https://doi.org/10.1002/cjs.11129>
16. Hirose K, Yamamoto M. Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing*. 2015; 25(5): 863–875. <https://doi.org/10.1007/s11222-014-9458-0>
17. Chen Y, Liu J, Xu G, Ying Z. Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*. 2015; 110(510): 850–866. <https://doi.org/10.1080/01621459.2014.934827> PMID: 26294801
18. Liu J, Kang HA. Q-matrix learning via latent variable selection and identifiability. In: von Davier M, Lee YS, editors. *Handbook of Diagnostic Classification Models*. Cham: Springer; 2019. pp. 247–263.
19. Huang PH, Chen H, Weng LJ. A penalized likelihood method for structural equation modeling. *Psychometrika*. 2017; 82(2): 329–354. <https://doi.org/10.1007/s11336-017-9566-9> PMID: 28417228
20. Magis D, Tuerlinckx F, De Boeck P. Detection of differential item functioning using the lasso approach. *Journal of Educational and Behavioral Statistics*. 2015; 40(2): 111–135. <https://doi.org/10.3102/1076998614559747>
21. Tutz G, Schauberger G. A penalty approach to differential item functioning in Rasch models. *Psychometrika*. 2015; 80(1): 21–43. <https://doi.org/10.1007/s11336-013-9377-6> PMID: 24297435
22. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*. 1996; 58(1): 267–288.
23. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*. 1977; 39(1): 1–38.
24. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*. 2010; 33(1): 1–22. <https://doi.org/10.18637/jss.v033.i01> PMID: 20808728
25. Zhang S, Chen Y. Computation for latent variable model estimation: A unified stochastic proximal framework. *Psychometrika*. 2022; 87(4): 1473–1502. <https://doi.org/10.1007/s11336-022-09863-9> PMID: 35524934
26. Xu PF, Shang L, Zheng QZ, Shan N, Tang ML. Latent variable selection in multidimensional item response theory models using the expectation model selection algorithm. *British Journal of Mathematical and Statistical Psychology*. 2022; 75(2): 363–394. <https://doi.org/10.1111/bmsp.12261> PMID: 34918834
27. Jiang J, Nguyen T, Rao JS. The E-MS algorithm: Model selection with incomplete data. *Journal of the American Statistical Association*. 2015; 110(511): 1136–1147. <https://doi.org/10.1080/01621459.2014.948545> PMID: 26783375

28. Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978; 6(2): 461–464. <https://doi.org/10.1214/aos/1176344136>
29. Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*. 1981; 46(4): 443–459. <https://doi.org/10.1007/BF02293801>
30. Baker FB, Kim SH. *Item Response Theory: Parameter Estimation Techniques*. 2nd ed. Boca Raton: CRC press; 2004.
31. Zheng C, Meng X, Guo S, Liu Z. Expectation-maximization-maximization: A feasible MLE algorithm for the three-parameter logistic model based on a mixture modeling reformulation. *Frontiers in Psychology*. 2018; 8:2302. <https://doi.org/10.3389/fpsyg.2017.02302> PMID: 29354089
32. Chen P, Wang C. Using EM algorithm for finite mixtures and reformed supplemented EM for MIRT calibration. *Psychometrika*. 2021; 86(1): 299–326. <https://doi.org/10.1007/s11336-021-09745-6> PMID: 33591556
33. Ibrahim JG. Incomplete data in generalized linear models. *Journal of the American Statistical Association*. 1990; 85(411): 765–769. <https://doi.org/10.1080/01621459.1990.10474938>
34. Schilling S, Bock RD. High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*. 2005; 70(3): 533–555.
35. Meng XL, Schilling S. Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*. 1996; 91(435): 1254–1267. <https://doi.org/10.1080/01621459.1996.10476995>
36. Zhang S, Chen Y, Liu Y. An improved stochastic EM algorithm for large-scale full-information item factor analysis. *British Journal of Mathematical and Statistical Psychology*. 2020; 73(1): 44–71. <https://doi.org/10.1111/bmsp.12153> PMID: 30511445
37. Parikh N, Boyd S. Proximal algorithms. *Foundations and Trends in Optimization*. 2014; 1(3): 127–239. <https://doi.org/10.1561/2400000003>
38. Eysenck S, Barrett P. Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences*. 2013; 54(4): 485–489. <https://doi.org/10.1016/j.paid.2012.09.022>
39. Kline P. *A Handbook of Test Construction: Introduction to Psychometric Design*. New York: Methuen; 1986.
40. von Davier M. New results on an improved parallel EM algorithm for estimating generalized latent variable models. In van der Ark LA, Wiberg M, Culpepper SA, Douglas JA, Wang WC, editors. *Quantitative Psychology*. Cham: Springer; 2017. pp. 1–8. [https://doi.org/10.1007/978-3-319-56294-0\\_1](https://doi.org/10.1007/978-3-319-56294-0_1)