Psychometrika (2025), **90**, 1014–1038
doi:10.1017/psy.2025.1

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

A Generalized Factor Rotation Framework with Customized Regularization

Yongfeng Wu[1], Xiangyi Liao[2] and Qizhai Li[1,3]

[1]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China; [2]Department of Educational Psychology, University of Wisconsin–Madison, Madison, WI, USA; [3]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, China

Corresponding author: Yongfeng Wu; Email: yongfeng@amss.ac.cn

(Received 20 December 2024; accepted 30 December 2024; published online 27 January 2025)

Abstract

Factor rotation is a crucial step in interpreting the results of exploratory factor analysis. Several rotation methods have been developed for simple structure solutions, but their extensions to bi-factor analysis are often not well established. In this article, we propose a mathematical framework that incorporates customized factor structure as a regularization to produce the optimal orthogonal or oblique rotation. We demonstrate the utility of the framework using examples of simple structure rotation and bi-factor rotation. Through detailed simulations, we show that the new method is accurate and robust in recovering the factor structures and latent correlations when bi-factor analysis is applied. The new method is applied to a test data and a Quality of Life survey data. Results show that our method can reveal bi-factor structures that are consistent with the theories.

Keywords: bi-factor analysis; oblique rotation; orthogonal rotation; proximal algorithm; simple structure

1. Introduction

Factor analysis is a popular technique for learning the underlying structure of multivariate data and has wide applications in psychology and the social sciences. Factor analysis suffers from the rotational indeterminacy issue where the loading matrix and the factors can be simultaneously rotated under the same model. The rotation procedure is a crucial step in obtaining an interpretable structure. The prevailing solution seeks the rotation that renders a simple structure (Thurstone, 1947). Roughly speaking, Thurstone's simple structure means that the loading matrix contains as many (approximately) zero entries as possible, so that each observed variable can be explained by only a few factors. Nevertheless, the simple structure is not the only desired solution, especially when the result does not admit a perfect simple structure. The bi-factor analysis, for instance, has become a popular alternative solution for exploring the factor structure (Reise, 2012). A classic approach to the bi-factor analysis is the Schmid–Leiman (SL; Schmid & Leiman, 1957) transformation. However, the SL transformation imposes an unnecessary proportionality constraint, making the loading matrix inevitably rank-deficient (Waller, 2018). Moreover, the SL transformation only produces orthogonal factors, leaving the oblique bi-factor analysis unsolved. Recently, some new strategies have been developed for bi-factor analysis

© The Author(s), 2025. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (https://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

(Abad et al., 2017; Jennrich & Bentler, 2011, 2012), but their performance has not been thoroughly discussed, and none of them has been accepted as the conventional approach.

This article attempts to shed some light on the bi-factor rotation problem, and more generally, on the factor rotation problem. We provide a mathematical framework to formulate and solve the rotation problem (both orthogonal and oblique) in factor analysis. Given any desired property of the factor structure, our framework incorporates it as a penalty term or a constraint, and by solving an optimization problem, it produces a rotation that rotates an initial loading matrix toward the desired property. In the bi-factor analysis, for example, the rotated loading matrix is expected to have a bi-factor structure. Accordingly, our framework takes the bi-factor structure as a constraint and produces a rotation that rotates an initial loading matrix into a bi-factor matrix as much as possible. We also provide a convergent algorithm to solve the optimization problem. Jennrich (2001, 2002) has devised a gradient-based algorithm for optimizing a general rotation criterion function, which generates a sequence of monotone iterates. However, the monotone iterates only guarantee the convergence of some subsequence, and there might exist multiple limiting points. Indeed, the monotone iterates may oscillate indefinitely and generate paths with infinite length (see Absil et al., 2005 for such an example). In contrast, our algorithm guarantees the convergence of the whole iterates.

Although the problem of simple structure rotation has been extensively studied, we still demonstrate the utilization of our framework in solving this problem, partly because it serves as an example of the penalty-type formulation and partly because it provides a new perspective (and a new algorithm) on solving this problem. Thurstone's original concept of simple structure largely concerns the number of zero loadings, but many existing methods maximize some dispersion of factor loadings so that the loadings tend to be either very high or very low. As Nesselroade and Cattell (2013) note, "the position that gives merely a lot of low loadings is different from the exact one that maximizes the number of zero loadings." Moreover, these dispersion-based methods raise the scaling issues, such as sensitivity to outliers and the question of normalization. In contrast, our framework provides a solution on the basis of the count of zero/nonzero loadings, agreeing with the very notion of simple structure.

We emphasize that our framework is not limited to the simple structure or bi-factor rotations. The regularization term in our framework can be customized to represent any subjective or theoretical assumptions about the factor structure, and our framework identifies the optimal rotation solution corresponding to the given assumption. This is an attractive advantage because researchers may have various demands on the exploratory factor analysis (EFA) across different applications. This regularized formulation also provides a perspective to unify the rotation procedure and the penalized estimation. We justify in Section 6.2 that these two seemingly competing procedures are mathematically almost equivalent.

The remainder of this article is organized as follows. Section 2 and Section 3 describe our framework for solving the orthogonal and oblique rotation problems, respectively. In each section, we demonstrate both the simple structure rotation and the bi-factor rotation, along with their algorithms. In Section 4, we conduct a simulation to compare the performance of our framework in the exploratory bi-factor analysis with existing methods. The proposed method is applied to real datasets in Section 5. Section 6 discusses some connections between our framework and other methods. Section 7 concludes this article. Technical derivations and proofs are postponed to the Appendix, which also includes supplementary simulation studies.

2. Orthogonal rotation

2.1. Rotation to simple structure

Let $A \in \mathbb{R}^{p\times k}$ be an initial loading matrix and $T \in \mathbb{R}^{k\times k}$ an orthogonal matrix. The concept of simple structure in factor analysis concerns the search of T such that AT is as simple as possible. If simplicity is defined as the number of zero loadings, a natural choice is to minimize $\|AT\|_0$, where $\|\cdot\|_0$ counts the number of nonzero entries. Unfortunately, AT would not contain many exact zeros in general, especially when the loading matrix is subject to sampling error. In practice, very small loadings are accepted as

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

zeros. In other words, AT would be considered simple if it is close to some matrix with many zeros. Thus, we formulate the objective function as

$$\begin{aligned} &\min_{T,S} \|AT - S\|_{\mathrm{F}}^2 + \rho\|S\|_0, \\ &\text{s.t. } T'T = \mathbf{I}_k, \end{aligned} \tag{2.1}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm, $\rho > 0$ is the tuning parameter, the prime denotes matrix transpose, and $\mathbf{I}_k$ is the k-by-k identity matrix. In practice, we use $\rho = 0.3^2$, and this choice will be explained in Section 6.1.

2.2. Algorithm

Our objective function (2.1) introduces a new parameter S and seems to be more difficult to optimize than the usual criteria that involve only a rotation parameter T. However, we shall show that the introduction of a new parameter not only simplifies the optimization but also broadens its applicability. The key is the separation of the rotation constraint (on T) and the desired property (on S). The usual approaches consider the desired sparsity or its surrogate criterion directly on the rotated loadings, the interlock of which makes the factor rotation problem challenging. In (2.1) these two features are individually applied to T and to S, and they are linked by a simple Frobenius distance function. This makes the optimization with respect to each parameter very simple. While this might suggest applying an alternating minimization algorithm to solve (2.1), this algorithm suffers from the same drawback as the gradient-based algorithm: the sequence of parameters generated by the algorithm is not guaranteed to converge (Powell, 1973). Therefore, a convergent algorithm called the proximal alternating minimization (PAM) algorithm (Attouch et al., 2010) is employed to solve (2.1).

If $X = UDV'$ is the singular value decomposition of X, then $T = UV'$ minimizes $\|X - T\|_{\mathrm{F}}^2$ subject to $T'T = TT' = \mathbf{I}_k$ (Gower & Dijksterhuis, 2004). We denote this projection by $\mathcal{P}_{\mathrm{orth}}(X) = UV'$. Let $\mathcal{H}(X,\kappa) = X \circ I(|X| > \kappa)$ be the entrywise hard-thresholding operator for a matrix X, where $\circ$ is the entrywise matrix product and $I(|X| > \kappa)$ is the entrywise indicator function for whether the absolute value of X entries is greater than a scalar threshold κ. The algorithm is presented in Algorithm 1, which updates T and S alternately using the above operators (see Appendix A.1 for the derivation). The convergence result of this algorithm is summarized in Proposition 1, whose proof is given in Appendix A.2. The algorithm converges to a stationary point[1] for any bounded stepsizes γ_t and η_t. In practice, we choose some small values, such as $\gamma_t = \eta_t = 0.01$. Also worth mentioning is the issue of local minima. Like many other rotation methods, Algorithm 1 may converge to a local minimum because the rotation problem is non-convex. Therefore, it is recommended to run the algorithm with multiple random initializations and choose the one with the smallest objective value.

Proposition 1. *Assume that the sequences of stepsizes γ_t and η_t are bounded away from zero and infinity, that is, there exists some positive numbers $r_+ > r_- > 0$ such that γ_t, η_t belong to (r_-, r_+) for all $t \geq 0$. Then the iterates (T_t, S_t) generated by Algorithm 1 converge to a stationary point of (2.1).*

[1]For a smooth function, a stationary point is a point whose gradient is zero. For a non-smooth function, the concept of gradient is generalized to the subdifferential set, and a stationary point becomes a point whose subdifferential set contains zero. The subdifferential set $\partial f(x)$ of f at x is defined by

$$\partial f(x) = \left\{u \in \mathbb{R}^d : \exists\, x_t \to x, f(x_t) \to f(x) \text{ and } \exists\, u_t \in \widehat{\partial} f(x_t), u_t \to u \text{ as } t \to \infty\right\},$$

where $\widehat{\partial} f(x)$ is the Fréchet subdifferential, defined as

$$\widehat{\partial} f(x) = \left\{u \in \mathbb{R}^d : \liminf_{y \neq x,\, y \to x} \frac{f(y) - f(x) - \langle u, y - x\rangle}{\|y - x\|} \geq 0\right\}.$$

A local minimum must be a stationary point.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

2.3. Bi-factor rotation

In the simple structure rotation, we maximize the degree of simplicity for the loading matrix, so our framework formulates the problem as a penalized optimization. When the loading matrix is desired to satisfy certain restrictions, we formulate it as a constrained problem, which can also be efficiently solved by the PAM algorithm. An example of such a problem is the exploratory bi-factor analysis.

Algorithm 1 PAM for orthogonal simple structure rotation problem (2.1).

Input: Initial loading matrix A, tuning parameter ρ.
Output: Rotation matrix T, sparse loading matrix S.

1: Initialize an orthogonal matrix T_0 and let $S_0 = \mathcal{H}(AT_0, \sqrt{\rho})$.
2: **for** $t = 0, 1, \ldots$ **do**
3: Take $\gamma_t > 0$ and compute

$$T_{t+1} = \mathcal{P}_{\text{orth}}(A'S_t + \gamma_t T_t). \tag{2.2}$$

4: Take $\eta_t > 0$ and compute

$$S_{t+1} = \mathcal{H}\Big(\frac{AT_{t+1} + \eta_t S_t}{1 + \eta_t}, \sqrt{\rho/(1+\eta_t)}\Big). \tag{2.3}$$

5: **end for**

A bi-factor model has a loading matrix of the form

$$\Lambda = \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ * & 0 & * \\ * & 0 & * \\ * & 0 & * \end{pmatrix}.$$

Formally speaking, the loading matrix has a column of free parameters, and besides this column, it has at most one free parameter in each row. The factor corresponding to the free column is called a general factor, and the remaining factors are called group factors. Exploratory bi-factor analysis (Jennrich & Bentler, 2011; Reise, 2012) can uncover the bi-factor structure and estimate the loadings simultaneously, unlike the confirmatory bi-factor analysis that requires the bi-factor structure to be specified in advance. Let $\mathcal{S}_{\text{bi}}$ denote the set of matrices with bi-factor structure. Our method performs exploratory bi-factor analysis by solving

$$\begin{aligned} &\min_{T,S} \|AT - S\|_{\text{F}}^2, \\ &\text{s.t. } T'T = \mathbf{I}_k, S \in \mathcal{S}_{\text{bi}}. \end{aligned} \tag{2.4}$$

The algorithm for solving (2.4) is very similar to Algorithm 1 and is presented in Algorithm 2. It involves the projection operator onto bi-factor matrices $\mathcal{P}_{\text{bi}}(X) := \arg\min_{S \in \mathcal{S}_{\text{bi}}} \|S - X\|_{\text{F}}$. The evaluation of this projection requires solving a simple combinatorial optimization to find the column of general factor loadings and then keeping the largest (in absolute value) entry in each row of the remaining columns as the group factor loadings (see Appendix A.3 for a detailed description).

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Algorithm 2 PAM for orthogonal bi-factor rotation problem (2.4).

Input: Initial loading matrix A.

Output: Rotation matrix T, bi-factor loading matrix S.

1: Initialize an orthogonal matrix T_0 and let $S_0 = \mathcal{P}_{\text{bi}}(AT_0)$.

2: **for** $t = 0, 1, \ldots$ **do**

3: Take $\gamma_t > 0$ and compute

$$T_{t+1} = \mathcal{P}_{\text{orth}}(A'S_t + \gamma_t T_t). \tag{2.5}$$

4: Take $\eta_t > 0$ and compute

$$S_{t+1} = \mathcal{P}_{\text{bi}}\Big(\frac{AT_{t+1} + \eta_t S_t}{1 + \eta_t}\Big). \tag{2.6}$$

5: **end for**

3. Oblique rotation

3.1. Rotation to simple structure

In the orthogonal rotation case, the factors are uncorrelated, and the rotation matrix is restricted to be an orthogonal matrix. When the factors are allowed to be correlated, the oblique rotation problem arises. This section provides a counterpart of our framework to solve the oblique rotation problem, which is similar to the orthogonal case but has some subtle yet crucial differences.

If one is primarily interested in the loading matrix estimation, the oblique version of the simple structure rotation problem might be formulated as

$$\begin{aligned} &\min_{T,S} \|A(T')^{-1} - S\|_{\text{F}}^2 + \rho\|S\|_0, \\ &\text{s.t. } \text{diag}(T'T) = \mathbf{I}_k, \end{aligned} \tag{3.1}$$

where $\text{diag}(\cdot)$ keeps the diagonal part of a matrix and assigns zeros to the off-diagonal part, as (3.1) finds the rotated loading matrix $A(T')^{-1}$ that is closest to a hypothesized simple loading matrix S. However, in the oblique factor analysis, the factor correlation matrix also needs to be estimated, and the rotation T is responsible for both the loading matrix $A(T')^{-1}$ and the factor correlation matrix $\Phi = T'T$. In order to get an overall better estimation of the loading matrix and the factor correlation matrix, we propose to formulate the oblique rotation problem as

$$\begin{aligned} &\min_{T,S} \|A - ST'\|_{\text{F}}^2 + \rho\|S\|_0, \\ &\text{s.t. } \text{diag}(T'T) = \mathbf{I}_k. \end{aligned} \tag{3.2}$$

When $\|A - ST'\|_{\text{F}}$ is minimized, the reproduced covariance matrix $A(T')^{-1}\Phi T^{-1}A' + \Psi^2 = AA' + \Psi^2$ will be close to the hypothesized covariance matrix $ST'TS' + \Psi^2$ with a simple loading matrix S, where Ψ^2 is a diagonal matrix of uniqueness. As the covariance matrix is governed by the loading matrix and factor correlation matrix, achieving closeness in the covariance matrix facilitates a balanced estimation of these two parameters. The rotated loading matrix $A(T')^{-1}$ will be approximately simple since $A(T')^{-1} \approx ST'(T')^{-1} = S$; the factor correlation matrix $\Phi = T'T$ is suitable for the hypothesized simple structure since $ST'TS' + \Psi^2$ is close to the optimal covariance matrix $AA' + \Psi^2$. The advantages of (3.2) in estimating the correlation matrix will be numerically illustrated in Section 4 and Appendix A.7.

This nuance does not appear in the orthogonal rotation problem because the orthogonal factor correlation matrix is invariant and the rotation is only responsible for the loading matrix. In effect, for an orthogonal matrix T, $\|A - ST'\|_{\text{F}} = \|(A - ST')T\|_{\text{F}} = \|AT - S\|_{\text{F}}$, so the two formulations are equivalent.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

As for the algorithm, although the alternating minimization algorithm and the PAM algorithm are conceptually applicable to (3.2), they do not lead to a practical algorithm because the iteration steps for problem (3.2) do not bear explicit updating formulas with these algorithms. We employ the proximal alternating linearized minimization (PALM) algorithm (Bolte et al., 2014) to solve (3.2). The resulting algorithm is reported in Algorithm 3. The projection onto oblique rotation matrices $\mathcal{P}_{\text{oblq}}(X) = X\{\text{diag}(X'X)\}^{-1/2}$ rescales each column of matrix X to unit length. We use an orthogonal matrix as initialization because, in general, it has empirically better performance than an oblique one. In practice, we choose the values of γ and η slightly above one, such as $\gamma = \eta = 1.01$. This algorithm is also convergent, as recorded in Proposition 2 with proof given in Appendix A.5. It is again recommended to run the algorithm multiple times to alleviate the issue of local minima.

Algorithm 3 PALM for oblique simple structure rotation problem (3.2).

Input: Initial loading matrix A, tuning parameter ρ.

Output: Rotation matrix T, sparse loading matrix S.

1: Initialize an orthogonal matrix T_0 and let $S_0 = \mathcal{H}(A(T_0')^{-1}, \sqrt{\rho})$.

2: **for** $t = 0, 1, \ldots$ **do**

3: Take $\gamma > 1$ and compute

$$T_{t+1} = \mathcal{P}_{\text{oblq}}\Big(T_t + \frac{(A' - T_t S_t')S_t}{\gamma \|S_t' S_t\|_{\text{F}}}\Big). \tag{3.3}$$

4: Take $\eta > 1$ and compute

$$S_{t+1} = \mathcal{H}\Big(S_t + \frac{(A - S_t T_{t+1}')T_{t+1}}{\eta \|T_{t+1}' T_{t+1}\|_{\text{F}}}, \sqrt{\rho/(\eta \|T_{t+1}' T_{t+1}\|_{\text{F}})}\Big). \tag{3.4}$$

5: **end for**

Proposition 2. *Assume* $\inf_{t\geq 0} \|S_t\|_{\text{F}} > 0$ *and* $\sup_{t\geq 0} \|S_t\|_{\text{F}} < \infty$, *then the iterates* (T_t, S_t) *generated by Algorithm 3 converge to a stationary point of (3.2).*

3.2. Bi-factor rotation

As in the orthogonal case, formulation (3.2) and the PALM algorithm can be used for other purposes in oblique rotation. We continue to demonstrate the exploratory bi-factor analysis because it highlights some new issues in the oblique case. We shall show that the bi-factor model suffers from what we would call group-factor indeterminacy when the factors are allowed to be correlated. This indeterminacy can be suppressed if we restrict all the group factors to be uncorrelated with the general factor.

The group-factor indeterminacy is illustrated as follows. Let $\Lambda \in \mathbb{R}^{p\times k}$ be a bi-factor loading matrix and $\boldsymbol{F} \in \mathbb{R}^k$ the corresponding factors. Without loss of generality, we let the first component in $\boldsymbol{F}$ be the general factor. Let $\Phi = (\phi_{ij}) \in \mathbb{R}^{k\times k}$ be the factor correlation matrix. Construct a transformation matrix $\Gamma \in \mathbb{R}^{k\times k}$ as

$$\Gamma = \begin{pmatrix} 1 & & & \\ d_2 & c_2 & & \\ d_3 & & c_3 & \\ \vdots & & & \ddots \\ d_k & & & & c_k \end{pmatrix} \tag{3.5}$$

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

that has zeros at locations other than the main diagonal and the first column. Matrix Γ has an inverse

$$\Gamma^{-1} = \begin{pmatrix} 1 & & & \\ -d_2/c_2 & 1/c_2 & & \\ -d_3/c_3 & & 1/c_3 & \\ \vdots & & & \ddots \\ -d_k/c_k & & & 1/c_k \end{pmatrix}$$

whenever it exists. The transformed factor $\widetilde{\boldsymbol{F}} = \Gamma \boldsymbol{F}$ has a covariance matrix $\Gamma\Phi\Gamma'$, whose diagonal elements are $d_r^2 + 2c_r\phi_{1r}d_r + c_r^2$ (except for the first one). We set $d_r = -c_r\phi_{1r} \pm \sqrt{1 - c_r^2(1-\phi_{1r}^2)}$ for all $2 \le r \le k$, so that the transformed factor is standardized. The c_r can be any nonzero number between $-1/\sqrt{1-\phi_{1r}^2}$ and $1/\sqrt{1-\phi_{1r}^2}$. The transformed loading matrix $\widetilde{\Lambda} = \Lambda\Gamma^{-1}$ has the same bi-factor structure as Λ. Thus, we have constructed a different oblique bi-factor representation $\widetilde{\Lambda}\widetilde{\boldsymbol{F}} = \Lambda \boldsymbol{F}$. Intuitively, each cluster of items indicated by the group factors forms a micro factor model with two common factors (the general factor and the corresponding group factor). Under the oblique factor case, the group factor can be rotated toward or against the general factor within this two-factor model (we should not rotate the general factor because it is shared by other clusters). This explains how we construct the transformation matrix Γ, and we call this phenomenon group-factor indeterminacy of the oblique bi-factor model. This result is not new and has been disclosed by Jennrich and Bentler (2012). A natural yet putative strategy to resolve this indeterminacy is to restrict the group factors to be uncorrelated with the general factor. We call such a bi-factor representation a semi-oblique bi-factor model. A fully oblique bi-factor representation can be transformed into a semi-oblique one using Γ with $c_r = 1/\sqrt{1-\phi_{1r}^2}$ and $d_r = -c_r\phi_{1r}$ for all $2 \le r \le k$.

Even given the indeterminacy and the putative restriction, we can still formulate and solve the oblique exploratory bi-factor analysis with

$$\begin{aligned} &\min_{T,S} \|A - ST'\|_{\mathrm{F}}^2, \\ &\text{s.t. } \operatorname{diag}(T'T) = \mathbf{I}_k,\ S \in \mathcal{S}_{\mathrm{bi}}. \end{aligned} \tag{3.6}$$

It should be clarified that the fully oblique bi-factor models are not considered false or invalid. They are different representations of the equivalent bi-factor models. In effect, this indeterminacy does not affect the feasibility $S \in \mathcal{S}_{\mathrm{bi}}$ and the objective value $\|A - ST'\|_{\mathrm{F}}^2$. Consequently, (3.6) has a continuum of optimal solutions that correspond to a single model, and we transform the final result into a semi-oblique one to provide a unique representation.[2] Hence, Algorithm 4 solves the oblique bi-factor problem with the PALM algorithm, followed by a partial orthogonalization step.

[2]It is possible to formulate the optimization problem that constrains the rotation T to be semi-oblique, but the algorithm would become impractical because the set of semi-oblique rotations does not have a simple projection solution.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Algorithm 4 PALM for oblique bi-factor problem (3.6).

Input: Initial loading matrix A.

Output: Rotation matrix T, bi-factor loading matrix S.

1: Initialize an orthogonal matrix T_0 and let $S_0 = \mathcal{P}_{\text{bi}}(A(T_0')^{-1})$.

2: **for** $t = 0, 1, \ldots$ **do**

3: Take $\gamma > 1$ and compute

$$T_{t+1} = \mathcal{P}_{\text{oblq}}\Big(T_t + \frac{(A' - T_t S_t')S_t}{\gamma \|S_t' S_t\|_{\text{F}}}\Big). \tag{3.7}$$

4: Take $\eta > 1$ and compute

$$S_{t+1} = \mathcal{P}_{\text{bi}}\Big(S_t + \frac{(A - S_t T_{t+1}')T_{t+1}}{\eta \|T_{t+1}' T_{t+1}\|_{\text{F}}}\Big). \tag{3.8}$$

5: **end for**

6: Partially orthogonalize T_∞ and S_∞ to $T_\infty \Gamma'$ and $S_\infty \Gamma^{-1}$ with Γ in (3.5), $c_r = 1/\sqrt{1 - \phi_{1r}^2}$, $d_r = -c_r \phi_{1r}$ for all $2 \le r \le k$, and $\Phi = T_\infty' T_\infty$, provided that the general factor is rearranged to the first column.

4. Simulation

We have proposed a general framework to formulate and solve the rotation problem in factor analysis and demonstrated it through the examples of simple structure rotation and exploratory bi-factor analysis. Because the simple structure rotation problem has been extensively studied and mature solutions have been developed, our simulation experience shows considerable similarity between our method and the existing popular methods in terms of numerical performance (see Appendix A.6 for the simulation results). Here, we exhibit the simulation results for the exploratory bi-factor analysis. The R code for running the simulations and conducting the analyses are available in the online supplemental materials.

We compare our proposed method with eight existing methods: (a) the SL (the Schmid-Leiman procedure; Schmid & Leiman, 1957); (b) the SLt (the SL followed by a partially specified target rotation; Reise et al., 2010); (c) the SLi (the SL followed by iterated target rotations; Abad et al., 2017); (d) the DSL (the Direct Schmid-Leiman method; Waller, 2018); (e) the DBF (the Direct Bi-Factor method; Waller, 2018); (f) the PEBI (the orthogonal or oblique Pure Exploratory BI-factor analysis; Lorenzo-Seva & Ferrando, 2019); (g) the BQ.orth (the orthogonal Bi-Quartimin method; Jennrich & Bentler, 2011); and (h) the BQ.oblq (the oblique Bi-Quartimin method; Jennrich & Bentler, 2012). We basically follow the simulation settings from Abad et al. (2017) and Giordano and Waller (2020). Specifically, we consider a total of 22 items clustered into four groups, with four, five, six, and seven items in each group. We examine four types of bi-factor structures: (a) the independent cluster (IC) structure that is a perfect bi-factor loading matrix; (b) the independent cluster basis (ICB) structure that contains cross-loadings; (c) the independent cluster pure (ICP) structure where some items have nonzero loadings only on the general factor; and (d) the independent cluster basis pure (ICBP) structure that contains both cross-loadings and pure items. In our simulation, the group factor loadings take either high or low values. When they take high values, they are randomly selected from the interval $[0.6, 0.9]$; for the low value case, they are selected from the range $[0.3, 0.6]$. When cross-loadings are present, the last item in each cluster has a cross-loading of 0.4 on the next group factor. For pure items, the item in the middle position of each cluster has a loading of 0.01 on the corresponding group factor. The loadings on the general factor are randomly selected so that every item has communality no greater than 0.81. When necessary, some rows of loading vector are rescaled to prevent excessively large communality caused by the cross-loadings. Finally, every loading entry is randomly assigned a positive or negative sign. An example of simulated loading matrices is presented in Table 1.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Table 1. Examples of the loading matrix under the four types of bi-factor structure

Independent cluster (IC)				Independent cluster basis (ICB)				Independent cluster pure (ICP)				Independent cluster basis pure (ICBP)			
.21	.68			−.31	−.69			−.25	−.73			−.14	.82		
.23	−.60			.41	−.74			−.32	*.01*			−.44	*.01*		
−.03	.75			−.23	.85			.49	.75			.11	.75		
.32	−.60			.29	.75	*−.39*		−.24	−.85			−.10	−.75	*.39*	
.26		−.62		−.15		.84		−.19		.84		.16		−.61	
.01		−.85		.38		−.82		−.17		.76		.49		−.64	
.08		−.63		−.09		−.78		.68		*−.01*		−.32		*−.01*	
−.23		.69		−.08		−.82		.15		−.70		.53		−.62	
.16		−.85		−.06		−.73	*−.40*	−.33		.60		.28		−.63	*−.40*
−.21			−.64	.14			−.71	.51			.64	.40			.73
−.26			−.65	−.24			−.85	.18			−.81	.61			−.63
−.11			−.73	.07			.78	−.30			*.01*	−.10			*.01*
.25			.85	−.50			−.74	−.38			.67	−.03			−.84
.20			−.84	−.01			−.84	−.08			.70	−.37			.64
−.10			−.82	.28			−.73	*−.40*	.18			−.83	.34			.73	*−.40*
.05			.77	−.03			.62	−.29			.84	−.21			.65
.13			.74	.03			.73	−.09			−.83	.58			−.62
−.03			.70	−.32			−.65	−.50			−.72	.25			−.73
−.15			−.65	−.12			−.79	.30			*.01*	−.47			*−.01*
.11			−.74	.24			.71	−.04			−.85	.18			−.69
.30			.66	−.30			−.67	.29			−.77	.28			−.79
.26			.80	.01	*.40*		−.71	−.25			−.61	.10	*.40*		.72

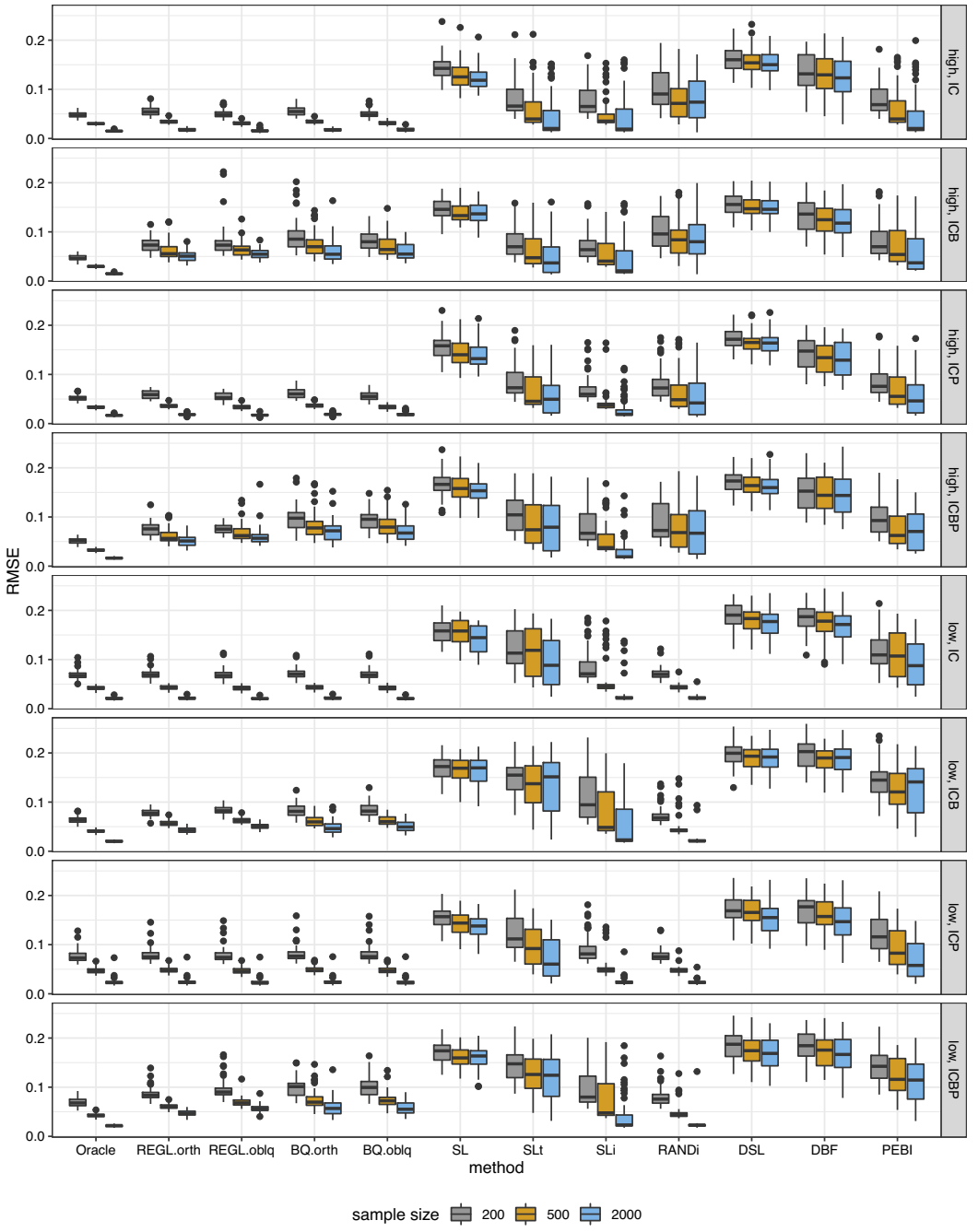


Figure 1. Boxplot of the estimation error for loading matrix under orthogonal bi-factor models.

Given the simulated loading matrix Λ, we generate the population correlation matrix

$$R = \Lambda\Phi\Lambda' + \Psi^2,$$

where Φ is the factor correlation matrix and Ψ^2 is a diagonal matrix of uniqueness, chosen to constrain the diagonal elements of R to one. In the orthogonal bi-factor case, Φ will be an identity matrix. In

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

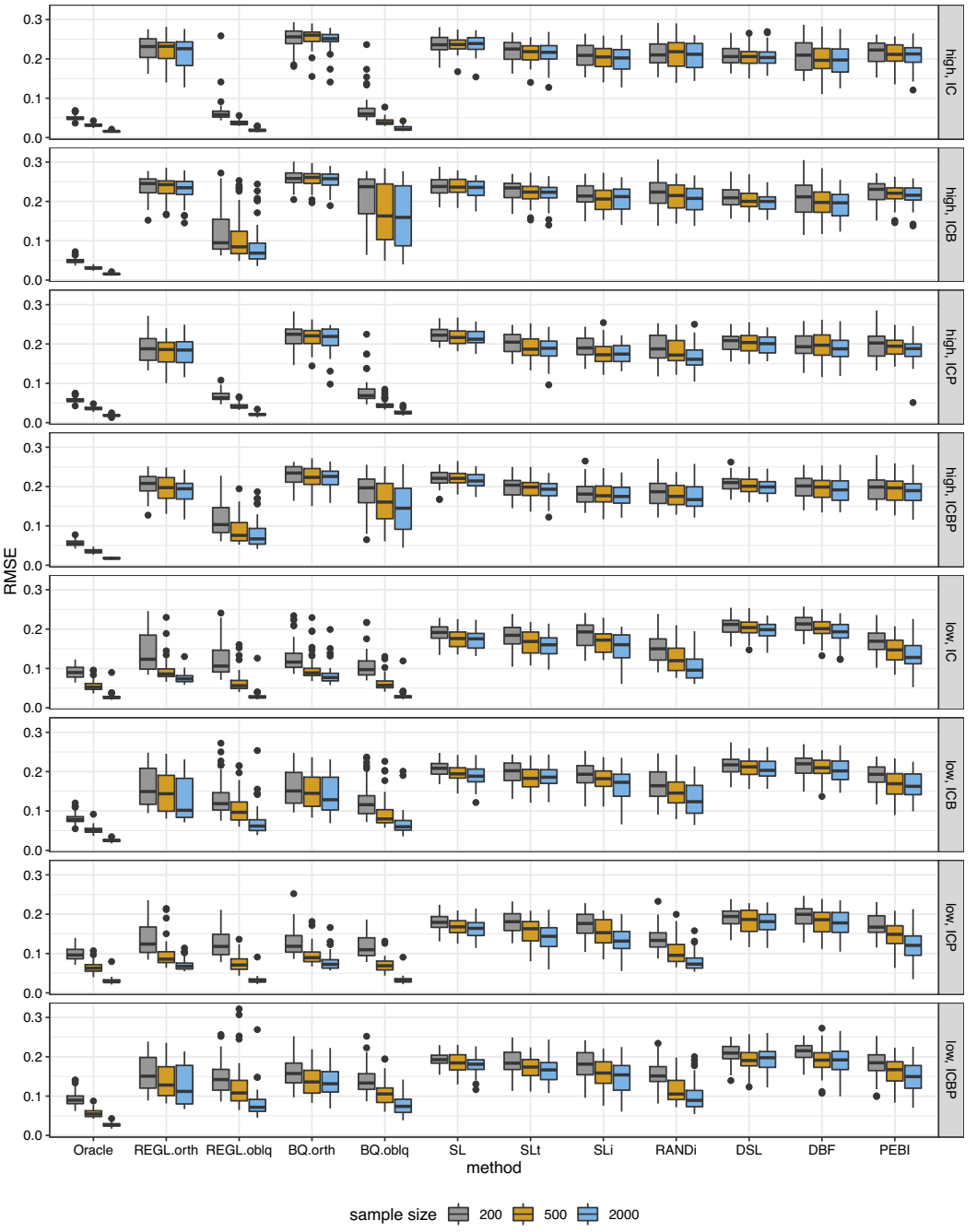


Figure 2. Boxplot of the estimation error for loading matrix under semi-oblique bi-factor models.

the semi-oblique case, the correlations among group factors are randomly selected from the interval $[0.2, 0.6]$. If the generated correlation matrix is not positive definite, we re-generate a new one until it is positive definite. The final step generates the data from a multivariate normal distribution with zero mean and covariance matrix R, with a sample size $N \in \{200, 500, 2000\}$. The simulation is replicated 50 times for each scenario.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

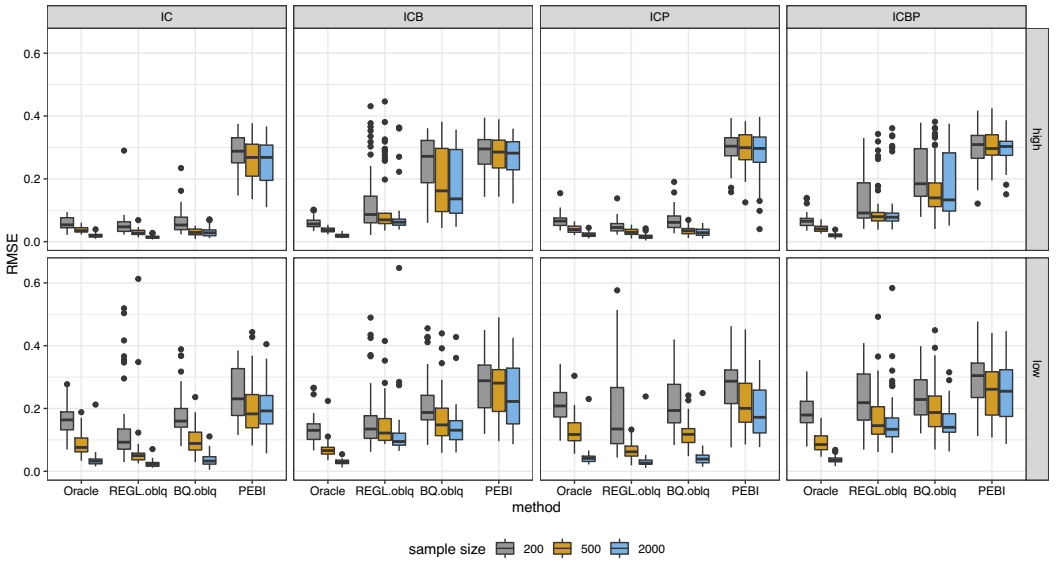


Figure 3. Boxplot of the estimation error for factor correlation matrix under semi-oblique bi-factor models.

The accuracy of the rotation methods is evaluated by the root mean squared error (RMSE) between the population and estimated bi-factor loading matrices:

$$\mathrm{RMSE}(\Lambda,\widehat{\Lambda}) = \frac{1}{\sqrt{pk}}\|\Lambda-\widehat{\Lambda}\|_{\mathrm{F}}, \tag{4.1}$$

after the estimated factors are aligned and orientated with respect to the population factors. The initial loading matrix is extracted using the maximum likelihood method. The results of our regularized rotation methods (REGL.orth and REGL.oblq) and the competitors are shown in Figures 1, 2, and 3. We also include an oracle method that rotates (orthogonally or obliquely) the initial loading matrix toward the true loading matrix Λ, that is, the oracle orthogonal rotation minimizes $\|AT-\Lambda\|_{\mathrm{F}}$ and the oracle oblique rotation minimizes $\|A(T')^{-1}-\Lambda\|_{\mathrm{F}}$. This oracle method can be considered the optimal rotation for loading estimation and is used as a reference.

Under the orthogonal bi-factor models (Figure 1), our methods and the Bi-Quartimin methods perform best, but our methods are slightly better at recovering the ICB and ICBP structures. Since the orthogonal model is a special case of the oblique model, the oblique versions of these two methods have almost the same performance as their orthogonal counterparts. The SLi method has comparable results to the best ones. We contend that the success of SLi should attribute to the iterating steps. To illustrate this, we replace the SL initialization in the SLi procedure with a random bi-factor loading initialization, and the resultant RANDi method remains successful. Note also the similarity between its iterative spirit and the PAM algorithm. All the other SL-based methods fail to recover the loading matrix.

As for the oblique bi-factor models (Figure 2), our proposed oblique rotation method and the oblique Bi-Quartimin method are the only successful methods, while the oblique PEBI method and all other orthogonal methods fail to provide reasonable estimates. Compared to the Bi-Quartimin method, our method again has an advantage in recovering the ICB and ICBP structures, indicating its robustness. This superiority is more evident when the group loadings take high values. Moreover, our method provides better estimation of the factor correlation matrix than the Bi-Quartimin method, as shown in Figure 3. The accuracy of factor correlation estimation is measured by

$$\mathrm{RMSE}(\Phi,\widehat{\Phi}) = \frac{1}{\sqrt{k^2}}\|\Phi-\widehat{\Phi}\|_{\mathrm{F}}.$$

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Interestingly, our method can even outperform the oracle method when the underlying model is strictly a bi-factor model. This is because the oracle method minimizes $\|A(T')^{-1} - \Lambda\|_F$ which emphasizes the discrepancy of the loading matrix, and the estimated rotation does not necessarily produce an optimal factor correlation matrix. In contrast, our method minimizes $\|A - ST'\|_F$ which is compatible with the discrepancy of the covariance matrix, resulting in a rotation that balances the estimation of the loading matrix and the factor correlation matrix. Hence, our method can better estimate the factor correlation matrix even though it does not use the true loading matrix Λ. A detailed comparison of the two formulations in estimating the loading matrix and the factor correlation matrix is given in Appendix A.7.

5. Real data examples

5.1. Holzinger's fourteen tests data

We now apply the proposed exploratory bi-factor analysis approach to Holzinger's fourteen tests data. This data were used by Holzinger and Swineford (1937) to illustrate bi-factor analysis. The correlation matrix was provided in Holzinger and Swineford (1937), and their preliminary analysis divided the fourteen tests into four groups to reflect spatial, mental speed, motor speed, and verbal factors (see Table IV in Holzinger & Swineford, 1937). This bi-factor structure is consistently recovered by our orthogonal and oblique bi-factor analyses, as shown in Table 2, except that the oblique bi-factor model has two crossing loadings. The estimated factor correlation matrix in the oblique model is

$$\Phi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & .77 & .37 & .67 \\ 0 & .77 & 1 & .46 & .36 \\ 0 & .37 & .46 & 1 & .17 \\ 0 & .67 & .36 & .17 & 1 \end{pmatrix}.$$

Since both orthogonal and oblique bi-factor analyses recover the desired structure, determining which model is more appropriate may depend on domain knowledge.

5.2. Quality of life data

When applied to another data set, our methods demonstrate the necessity of oblique bi-factor analysis. Chen et al. (2006) have applied a confirmatory bi-factor analysis to a Quality of Life data set. This data set contained 403 observations for 17 items answered on a 5-point Likert scale from 1 (all of the time) to 5 (never), with high scores on the scale indicating a high quality of life. These items were hypothesized to reflect a common general factor (Quality of Life) and four group factors (Cognition, Vitality, Mental Health, and Disease Worry). We apply both the proposed orthogonal and oblique exploratory bi-factor analyses to this data set, with results shown in Table 3. In the orthogonal bi-factor case, not all loadings on the group factor are significant for the third hypothesized cluster (Mental Health), consistent with the published studies (Abad et al., 2017; Chen et al., 2006; Jennrich & Bentler, 2011). Additionally, we identify a possible cross-loading for the "pep" item, which is also reported by Abad et al. (2017). The results become promising in the oblique case, where our method produces a bi-factor structure consistent with the hypothesized structure, except for a potential cross-loading for the "nerv" item. The estimated factor correlation matrix is

$$\Phi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & .51 & .65 & .45 \\ 0 & .51 & 1 & .64 & .51 \\ 0 & .65 & .64 & 1 & .65 \\ 0 & .45 & .51 & .65 & 1 \end{pmatrix},$$

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Table 2. Exploratory bi-factor rotation results of the proposed methods applied to the fourteen tests data (loadings ≥ .20 in absolute value are bolded)

Item	Orthogonal bi-factor loadings					Oblique bi-factor loadings				
T1	**.56**	**.28**	.17	.03	−.08	**.21**	**.38**	**.33**	.01	−.11
T2	**.80**	**.33**	−.10	−.04	.07	**.37**	**.94**	−.12	.01	−.08
T3.4	**.59**	**.56**	.04	.01	.14	.05	**.68**	.12	−.01	.08
T6	**.64**	.04	**.39**	.14	−.01	**.41**	−.19	**.69**	.11	.10
T28	**.51**	.10	**.35**	.00	−.12	**.29**	−.04	**.64**	−.05	−.07
T29	**.61**	−.05	**.35**	−.05	−.00	**.47**	−.15	**.64**	−.09	.08
T32	.12	.15	−.12	**.41**	.13	−.06	.13	**−.25**	**.47**	.15
T34	**.41**	−.09	.02	**.60**	−.14	**.29**	.04	−.02	**.69**	−.13
T35	.12	.12	.13	**.49**	−.12	−.06	−.14	.18	**.50**	−.05
T36a	**.54**	−.16	.04	**.37**	.05	**.46**	−.02	.06	**.45**	.10
T13	**.62**	.06	−.05	.03	**.53**	**.42**	.13	−.05	.07	**.61**
T18	**.63**	−.13	−.09	−.04	**.56**	**.56**	.05	−.11	.03	**.65**
T25b	**.43**	.14	.02	−.04	**.61**	**.24**	−.08	.08	−.04	**.76**
T77	**.45**	.07	.02	−.07	**.58**	**.31**	−.10	.08	−.06	**.72**

and the group factors have moderate correlations. This might explain the failure of the orthogonal bi-factor models to recover the hypothesized structure.

6. Discussion

6.1. Relation to other methods

The rotation to simple structure has been a classic problem in factor analysis, and a number of methods have been proposed in the literature. Although we develop the solution from a different perspective, it is mathematically related to some existing methods. We discuss the connection to Jennrich's (2004) component loss function (CLF) method and Kiers's (1994) simplimax method.

Jennrich (2004) has investigated a class of rotation criteria based on the CLF including the family of right constant CLF, to which we now demonstrate our method is equivalent. Let $\Lambda = AT$ be the rotated loading matrix with entries λ_{ir}. The CLF method finds the rotation T that minimizes the component loss criterion $Q(\Lambda)$, defined as

$$Q(\Lambda) = \sum_i \sum_r h(\lambda_{ir}^2)$$

with some component loss function $h(\cdot)$. One particular CLF is the right constant function

$$h(\lambda^2) = \begin{cases} (\lambda/b)^2, & \text{if } |\lambda| \le b, \\ 1, & \text{if } |\lambda| > b. \end{cases}$$

The component loss criterion with this right constant function is equivalent to (2.1) for $\rho = b^2$. To see this, we rewrite (2.1) as a partial minimization problem

$$\min_T \Big\{ \min_S \big(\|AT - S\|_{\mathrm{F}}^2 + \rho \|S\|_0 \big) \Big\}. \tag{6.1}$$

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Table 3. Exploratory bi-factor rotation results of the proposed methods applied to the quality of life data (loadings ≥ .20 in absolute value are bolded)

Item	Orthogonal bi-factor loadings					Oblique bi-factor loadings				
diff	**.56**	**.64**	−.01	−.08	−.04	.16	**.78**	−.01	.11	−.03
slo	**.46**	**.47**	.07	−.03	.03	.19	**.58**	.08	−.03	.05
con	**.53**	**.67**	.04	.07	.04	**.28**	**.86**	−.01	−.09	.04
for	**.44**	**.67**	.02	.01	−.01	.17	**.86**	−.03	−.08	−.03
dcon	**.57**	**.61**	−.00	−.01	−.02	**.23**	**.76**	−.02	.08	−.01
tired	**.67**	.04	**.52**	−.02	−.06	**.42**	−.00	**.76**	.03	−.07
ener	**.56**	−.00	**.38**	.01	.04	**.37**	−.03	**.54**	.02	.06
worn	**.66**	.07	**.54**	−.11	.03	**.35**	.04	**.79**	−.06	.05
pep	**.66**	−.01	**.43**	**.21**	.00	**.60**	−.01	**.56**	.03	−.03
calm	**.73**	−.01	−.04	**.41**	.02	**.72**	.03	−.10	**.45**	.01
blue	**.83**	−.04	−.10	.05	−.03	**.48**	−.10	−.03	**.74**	.05
hap	**.66**	−.05	.09	**.38**	−.01	**.68**	−.02	.08	**.34**	−.03
nerv	**.68**	.19	−.04	.01	.02	**.35**	**.21**	−.00	**.40**	.09
down	**.82**	.01	−.08	−.08	−.16	**.35**	−.08	.05	**.85**	−.10
afr	**.69**	.00	−.09	−.08	**.55**	**.35**	−.01	−.08	.08	**.82**
frust	**.69**	.00	.09	.01	**.42**	**.44**	−.00	.13	.01	**.60**
wor	**.61**	−.00	.00	.06	**.54**	**.43**	.02	−.03	−.10	**.76**

The inner minimization problem over S has an explicit solution $S = S(T) := \mathcal{H}(AT, \sqrt{\rho})$. Thus, if we let $\Lambda = AT$, (6.1) becomes

$$\begin{aligned} &\min_T \ \|AT - S(T)\|_{\mathrm{F}}^2 + \rho\|S(T)\|_0 \\ &\quad = \min_T \ \sum_i \sum_r \left\{\lambda_{ir}^2 I(\lambda_{ir}^2 \le \rho) + \rho I(\lambda_{ir}^2 > \rho)\right\}, \end{aligned}$$

which is exactly $b^2 Q(\Lambda)$ with the right constant CLF and $\rho = b^2$. This equivalence has several consequences. First, the desirable properties of the right constant CLF provided by Jennrich (2004) directly apply to our method, such as the ability to recover perfect simple structure or Thurstone's simple structure whenever they exist. Second, it suggests choosing the tuning parameter ρ as the square of the threshold b, such as $\rho = 0.3^2$. Finally, our method provides a natural justification and interpretation for the right constant CLF, and we also offer a simple and convergent algorithm for the equivalent methods.

Another related method is the simplimax rotation (Kiers, 1994). Although the simplimax method is proposed for oblique rotation, the idea can be analogously applied to the orthogonal rotation. Given a number m, the simplimax maximizes the simplicity by minimizing the sum of m smallest squared entries of the rotated loading matrix. It is derived from the formulation:

$$\begin{aligned} &\min_{T,S} \ \|AT - S\|_{\mathrm{F}}^2, \\ &\text{s.t. } \|S\|_0 \le pk - m, \end{aligned} \tag{6.2}$$

where T is either an orthogonal or an oblique rotation matrix. Thus, the simplimax can be viewed as a constrained version of (2.1), while (2.1) is a penalized version. The PAM algorithm is still applicable to the constrained problem (6.2), in which (2.3) is replaced by a truncation operation that sets the

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Table 4. Classification of estimation procedures

Objective function	Output	
	AT or $A(T')^{-1}$	S
Likelihood/squared loss/etc.		Penalized estimation/CFA
approximated loss function	EFA/REGL	REGL

m smallest (in absolute value) entries of $(AT_{t+1} + \eta_t S_t)/(1 + \eta_t)$ to zero. The constrained version (6.2) is appropriate when the number of zero loadings is pre-specified, while the penalized version (2.1) is appropriate when the threshold for small loadings is given, which is typically the case.

In developing the approach of exploratory bi-factor analysis, Jennrich and Bentler (2011, 2012) proposed constructing the rotation criterion as an index that measures the departure of a loading matrix from a bi-factor structure. If such departure is measured by the Frobenius distance to the set of bi-factor loading matrices, one naturally derives our formulation (2.4). Formulation (3.6) measures a distance not directly under the loading matrix scale but under the covariance matrix scale. This scale has the benefit of balancing the estimation accuracy for the loading matrix and the factor correlation matrix, hence the better performance shown in Figure 3.

6.2. *Bridging simple structure rotation and penalized estimation*

Penalized estimation is a popular technique in statistics and machine learning for incorporating prior knowledge about parameters. When the parameters are assumed to be sparse, sparsity-promoting penalties such as the Lasso (Tibshirani, 1996) are incorporated into the loss functions (e.g., the likelihood function) to produce sparse estimates. This technique has been introduced to the EFA for the estimation of sparse loadings and has been suggested as an alternative to the factor rotation procedure because it produces sparser loadings (Hirose & Yamamoto, 2015; Scharf & Nestler, 2019). We now show via our framework that the factor rotation procedure and the penalized estimation are two sides of the same coin. Their relation is summarized in Table 4. In the REGL formulation (2.1), the optimal value for S given a fixed T is $\mathcal{H}(AT, \sqrt{\rho})$, and the optimal value for T given a fixed S is $\mathcal{P}_{\text{orth}}(A'S) = \arg\min_{\{T:T'T=\mathbf{I}_k\}} \|AT - S\|_{\text{F}}$. Thus, the penalized estimate S is a truncated matrix of the rotated loading AT, and the rotated loading AT is an untruncated version of S. The difference between the rotation procedure and penalized estimation is merely a matter of choosing the output between AT and S.

While one may argue that our sparse estimates S in (2.1) and (3.2) are not exactly the penalized likelihood estimations discussed in the literature, the difference is peripheral. The classic penalized estimation minimizes a loss function (e.g., the likelihood function or the squared loss) $d(\widehat{\Sigma}, S\Phi S' + \Psi^2)$, which represents the discrepancy between the sample covariance matrix $\widehat{\Sigma}$ and the model covariance matrix $S\Phi S' + \Psi^2$, along with a penalty term for S. In contrast, we minimize $\|A - ST'\|_{\text{F}}^2$ plus a penalty term. The loss function $d(\widehat{\Sigma}, S\Phi S' + \Psi^2)$ can be quadratically approximated (up to constant scaling and shifting) by $\|AA' - S\Phi S'\|_{\text{F}}^2$, as $AA' + \widehat{\Psi}^2$ is the minimizer of the loss function. Taking the "square root" of the matrices AA' and $S\Phi S' = ST'TS'$, we can further approximate $\|AA' - S\Phi S'\|_{\text{F}}^2$ by $\|A - ST'\|_{\text{F}}^2$. Therefore, our formulas (2.1) and (3.2) are approximated penalized loss functions, and our sparse estimates S are approximate solutions to the standard penalized estimation. Ideally, it would be best to solve the penalized loss function directly because the initial loading matrix A is only an intermediate estimator. We frame the problem within the rotation paradigm partly because factor rotation procedures have historically been central to the EFA and partly because the rotation of a given matrix presents an algebraic problem of independent interest. Moreover, employing a penalized loss function introduces challenges in solving the optimization problem, as our algorithm is neither practical nor necessarily convergent when applied to it. The exploration of efficient algorithms for minimizing the penalized loss function in factor analysis remains a topic for future research.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

We can further draw a connection with the confirmatory factor analysis (CFA). In the CFA, the likelihood function (or a loss function in general) is minimized under the restriction where certain loading entries are pre-specified with fixed values or subject to equality constraints. This approach constitutes constrained estimation. Similar to the penalized formulation, our REGL framework provides a mathematical correspondence between the rotation procedure and the constrained estimation. The constrained estimates produced by our framework are again approximate solutions for CFA estimation by substituting the standard loss function with $\|A - ST'\|_{\mathrm{F}}^2$. Although the exploratory and confirmatory factor analyses are usually considered distinct disciplines with different objectives, our framework reveals a mathematical connection between their estimates.

7. Conclusion

We have proposed a general framework to solve the rotation problem in factor analysis. The problem is formulated as either a penalized or a constrained optimization, depending on the type of rotation purpose. This regularized formulation can incorporate any desired assumptions about the factor structure, and the optimization process finds the optimal rotation based on these assumptions. The optimization problem is solved using simple and convergent algorithms. This framework is applicable to both orthogonal and oblique rotations. We illustrate the penalized and constrained formulations using examples of simple structure rotation and bi-factor rotation, respectively.

Simulation studies show that, for exploratory bi-factor analysis, our method performs better than most other methods under many conditions, and mostly equally well as Bi-Quartimin, except in the conditions ICB and ICBP, where it performs better than Bi-Quartimin. When applied to real data sets, our method uncovers bi-factor structures that are consistent with the hypothesized theory. Our framework also provides insight into the mathematical relationship among exploratory factor rotation, penalized estimation, and confirmatory factor analysis.

Finally, we want to point out that we are essentially providing a numerical iterative approach to an algebraic problem: given a matrix A, how to find its approximate factorization ST' where S and T satisfy certain structures or properties. This matrix factorization framework has potential applications beyond factor analysis and could be valuable in other fields. For example, in dictionary learning (Rubinstein et al., 2010; Zhai et al., 2020), the problem of sparse representation modeling can be approached in a manner similar to (2.1) for orthogonal dictionaries or (3.2) for general dictionaries. The potential connections and extensions of our algebraic framework to related problems may be explored in future work. A reviewer raised a question regarding the performance of formulation (3.1) for the oblique rotation problem. We have not yet identified a convergent algorithm for solving (3.1), although some preliminary findings are presented in Appendix A.7. A comprehensive and systematic study of (3.1) represents another promising direction for future research.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2025.1.

Acknowledgments. The authors wish to thank Dan Bolt for his helpful feedback on drafts of this article. The authors also kindly thank the editor, the associate editor, and two anonymous referees for their insightful comments and suggestions that greatly improve this article.

Funding statement. Q.L. is partially supported by the National Nature Science Foundation of China (Grant Nos. 12325110 and 12288201) and the CAS Project for Young Scientists in Basic Research (YSBR-034).

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Abad, F. J., Garcia-Garzon, E., Garrido, L. E., & Barrada, J. R. (2017). Iteration of partially specified target matrices: Application to the bi-factor case. *Multivariate Behavioral Research*, *52*(4), 416–429.

Absil, P.-A., Mahony, R., & Andrews, B. (2005). Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, *16*(2), 531–547.

Attouch, H., Bolte, J., Redont, P., & Soubeyran, A. (2010). Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality. *Mathematics of Operations Research*, *35*(2), 438–457.

Bolte, J., Sabach, S., & Teboulle, M. (2014). Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, *146*(1), 459–494.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225.

Giordano, C., & Waller, N. G. (2020). Recovering bifactor models: A comparison of seven methods. *Psychological Methods*, *25*(2), 143.

Gower, J. C., & Dijksterhuis, G. B. (2004). *Procrustes problems*. OUP Oxford. https://doi.org/10.1093/acprof:oso/9780198510581.001.0001

Hirose, K., & Yamamoto, M. (2015). Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing*, *25*(5), 863–875.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*(1), 41–54.

Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, *66*(2), 289–306.

Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, *67*(1), 7–19.

Jennrich, R. I. (2004). Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika*, *69*(2), 257–273.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*(4), 537–549.

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*, *77*(3), 442–454.

Kiers, H. A. (1994). Simplimax: Oblique rotation to an optimal target with simple structure. *Psychometrika*, *59*(4), 567–579.

Lorenzo-Seva, U., & Ferrando, P. J. (2019). A general approach for fitting pure exploratory bifactor models. *Multivariate Behavioral Research*, *54*(1), 15–30.

Nesselroade, J. R., & Cattell, R. B. (2013). *Handbook of multivariate experimental psychology*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4613-0893-5

Powell, M. J. (1973). On search directions for minimization algorithms. *Mathematical Programming*, *4*, 193–201.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559.

Rubinstein, R., Bruckstein, A. M., & Elad, M. (2010). Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, *98*(6), 1045–1057.

Scharf, F., & Nestler, S. (2019). Should regularization replace simple structure rotation in exploratory factor analysis? *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 576–590.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 53–61.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. University of Chicago Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Waller, N. G. (2018). Direct Schmid–Leiman transformations and rank-deficient loadings matrices. *Psychometrika*, *83*(4), 858–870.

Zhai, Y., Yang, Z., Liao, Z., Wright, J., & Ma, Y. (2020). Complete dictionary learning via ℓ^4-norm maximization over the orthogonal group. *The Journal of Machine Learning Research*, *21*(1), 6622–6689.

1. Appendix

A.1. Derivation of Algorithm 1

For the problem (2.1), the PAM algorithm (Attouch et al., 2010) reads

$$\begin{cases} T_{t+1} = \arg\min\limits_{T'T=\mathbf{I}_k} \left\{ \|AT - S_t\|_{\mathrm{F}}^2 + \rho\|S_t\|_0 + \gamma_t \|T - T_t\|_{\mathrm{F}}^2 \right\}, & \text{(A.1)} \\ S_{t+1} = \arg\min\limits_{S} \left\{ \|AT_{t+1} - S\|_{\mathrm{F}}^2 + \rho\|S\|_0 + \eta_t \|S - S_t\|_{\mathrm{F}}^2 \right\}. & \text{(A.2)} \end{cases}$$

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

Step (A.1) is a proximal version of Procrustes problem, and can be converted to a standard one:

$$\begin{aligned}
&\arg\min_{T'T=\mathbf{I}_k} \|AT-S_t\|_{\mathrm{F}}^2+\rho\|S_t\|_0+\gamma_t\|T-T_t\|_{\mathrm{F}}^2\\
&=\arg\min_{T'T=\mathbf{I}_k} \operatorname{trace}\big[(T'A'-S_t')(AT-S_t)\big]+\gamma_t\operatorname{trace}\big[(T'-T_t')(T-T_t)\big]\\
&=\arg\min_{T'T=\mathbf{I}_k} \operatorname{trace}(-2S_t'AT)+\gamma_t\operatorname{trace}(-2T_t'T)\\
&=\arg\min_{T'T=\mathbf{I}_k} \operatorname{trace}\big[-2(S_t'A+\gamma_tT_t')T\big]\\
&=\arg\min_{T'T=\mathbf{I}_k} \|T-(A'S_t+\gamma_tT_t)\|_{\mathrm{F}}^2.
\end{aligned}$$

Thus $T_{t+1}=\mathcal{P}_{\mathrm{orth}}(A'S_t+\gamma_tT_t)$. Step (A.2) can be rearranged to

$$\begin{aligned}
&\arg\min_S \|AT_{t+1}-S\|_{\mathrm{F}}^2+\rho\|S\|_0+\eta_t\|S-S_t\|_{\mathrm{F}}^2\\
&=\arg\min_S \left\|S-\frac{AT_{t+1}+\eta_tS_t}{1+\eta_t}\right\|_{\mathrm{F}}^2+\frac{\rho}{1+\eta_t}\|S\|_0,
\end{aligned}$$

which can be optimized entry-wisely. In general, function $(x-t)^2+\kappa\|x\|_0$ is minimized at $x=0$ if $|t|<\sqrt{\kappa}$ and at $x=t$ otherwise, that is, the minimizer is $x=\mathcal{H}(t,\sqrt{\kappa})$. Hence,

$$S_{t+1}=\mathcal{H}\Big(\frac{AT_{t+1}+\eta_tS_t}{1+\eta_t},\sqrt{\rho/(1+\eta_t)}\Big).$$

A.2. Proof of Proposition 1

This convergence result is a direct application of Theorem 3.2 in Attouch et al. (2010). In particular, Theorem 3.2 of Attouch et al. (2010) requires that the objective function has the Kurdyka–Łojasiewicz property and that its smooth part has a Lipschitz continuous gradient. A function $f(x)$ is called Lipschitz continuous if there exists a constant $L\geq 0$ such that $\|f(x)-f(y)\|\leq L\|x-y\|$ for all x and y. The constant L is called the Lipschitz modulus. The function $\|AT-S\|_{\mathrm{F}}^2$ in (2.1) is a quadratic function of T and S, so it is continuously differentiable and has a Lipschitz continuous gradient on bounded subsets. Theorem 3 in Bolte et al. (2014) shows that semi-algebraic functions must have the Kurdyka–Łojasiewicz property (both definitions can be found in Attouch et al., 2010 or Bolte et al., 2014). Semi-algebraic functions are ubiquitous. In particular, polynomial functions, the Stiefel manifold (i.e., the set of orthogonal matrices), and $\|\cdot\|_0$ are all semi-algebraic (see the Appendix in Bolte et al., 2014). Hence, the objective function (2.1) has the Kurdyka–Łojasiewicz property.

Theorem 3.2 in Attouch et al. (2010) implies that any bounded sequence converges to a stationary point. We now show that our sequence is bounded. The orthogonal matrix T_t is bounded. The iteration (2.3) implies that

$$\begin{aligned}
\|S_{t+1}\|_{\mathrm{F}}&=\left\|\mathcal{H}\Big(\frac{AT_{t+1}+\eta_tS_t}{1+\eta_t},\sqrt{\rho/(1+\eta_t)}\Big)\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{AT_{t+1}+\eta_tS_t}{1+\eta_t}\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{AT_{t+1}}{1+\eta_t}\right\|_{\mathrm{F}}+\left\|\frac{\eta_tS_t}{1+\eta_t}\right\|_{\mathrm{F}}\\
&=\frac{1}{1+\eta_t}\|A\|_{\mathrm{F}}+\frac{\eta_t}{1+\eta_t}\|S_t\|_{\mathrm{F}}\\
&\leq\frac{1}{1+\eta_t}\max\{\|A\|_{\mathrm{F}},\|S_t\|_{\mathrm{F}}\}+\frac{\eta_t}{1+\eta_t}\max\{\|A\|_{\mathrm{F}},\|S_t\|_{\mathrm{F}}\}\\
&=\max\{\|A\|_{\mathrm{F}},\|S_t\|_{\mathrm{F}}\}.
\end{aligned}$$

This is valid for any t. Therefore,

$$\begin{aligned}
\|S_{t+1}\|_{\mathrm{F}}&\leq\max\{\|A\|_{\mathrm{F}},\|S_t\|_{\mathrm{F}}\}\\
&\leq\max\big\{\|A\|_{\mathrm{F}},\max\{\|A\|_{\mathrm{F}},\|S_{t-1}\|_{\mathrm{F}}\}\big\}\\
&=\max\{\|A\|_{\mathrm{F}},\|S_{t-1}\|_{\mathrm{F}}\}\\
&\leq\cdots\\
&\leq\max\{\|A\|_{\mathrm{F}},\|S_0\|_{\mathrm{F}}\}.
\end{aligned}$$

Hence our sequence (T_t,S_t) is bounded and thus converges to a stationary point.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

A.3. Projection onto bi-factor loadings

Algorithms 2 and 4 involve projecting a matrix onto the set of bi-factor loading matrices $\mathcal{P}_{\text{bi}}(X) = \arg\min_{S\in\mathcal{S}_{\text{bi}}} \|S-X\|_{\text{F}}$. To find this projection, one needs to identify which column of S corresponds to the general factor. Suppose that the rth column of S represents the general factor loadings; then this column is free and should be equal to the rth column of X in order to minimize the Frobenius distance. After that, simple algebra shows that the group factor loadings in S should be identical (in both value and location) to the largest (in absolute value) entry in each row of the remaining X sub-matrix. By letting every column be a potential general factor loading, we obtain k candidates for the bi-factor loading matrices. The projection will be the one with the largest Frobenius norm, because in this case $\|S-X\|_{\text{F}}^2 = \|X\|_{\text{F}}^2 - \|S\|_{\text{F}}^2$ for those candidate matrices S.

A.4. Derivation of Algorithm 3

For the minimization of $H(T,S)+f(T)+g(S)$, the PALM algorithm (Bolte et al., 2014) reads

$$\begin{cases} T_{t+1} = \arg\min_{T} \left\{ \langle T-T_t, \nabla_T H(T_t,S_t)\rangle + \frac{c_t}{2}\|T-T_t\|_{\text{F}}^2 + f(T)\right\}, & \text{(A.3)} \\ S_{t+1} = \arg\min_{S} \left\{ \langle S-S_t, \nabla_S H(T_{t+1},S_t)\rangle + \frac{d_t}{2}\|S-S_t\|_{\text{F}}^2 + g(S)\right\}. & \text{(A.4)} \end{cases}$$

In our case (3.2), $H(T,S) = \|A-ST'\|_{\text{F}}^2$, $g(S) = \rho\|S\|_0$, and $f(T)$ is the indicator function for the constraint $\text{diag}(T'T) = \mathbf{I}_k$, i.e., $f(T) = 0$ if T belongs to the constraint, and $f(T) = +\infty$ otherwise. We have

$$\nabla_T H(T,S) = -2(A'-TS')S, \text{ and } \nabla_S H(T,S) = -2(A-ST')T.$$

For these functions,

$$\|\nabla_T H(T_1,S) - \nabla_T H(T_2,S)\|_{\text{F}} = \|2(T_1-T_2)S'S\|_{\text{F}} \le 2\|T_1-T_2\|_{\text{F}}\|S'S\|_{\text{F}}$$

and

$$\|\nabla_S H(T,S_1) - \nabla_S H(T,S_2)\|_{\text{F}} = \|2(S_1-S_2)T'T\|_{\text{F}} \le 2\|S_1-S_2\|_{\text{F}}\|T'T\|_{\text{F}},$$

so they are Lipschitz continuous with moduli $L_1(S) = 2\|S'S\|_{\text{F}}$ and $L_2(T) = 2\|T'T\|_{\text{F}}$, respectively. The PALM algorithm requires that $c_t = \gamma L_1(S_t)$ and $d_t = \eta L_2(T_{t+1})$ for some $\gamma > 1, \eta > 1$. Thus, (A.3) becomes

$$\begin{aligned} &\arg\min_{\text{diag}(T'T)=\mathbf{I}_k} \langle T-T_t, -2(A'-T_tS_t')S_t\rangle + \gamma\|S_t'S_t\|_{\text{F}}\|T-T_t\|_{\text{F}}^2 \\ &= \arg\min_{\text{diag}(T'T)=\mathbf{I}_k} \left\| T - T_t - \frac{(A'-T_tS_t')S_t}{\gamma\|S_t'S_t\|_{\text{F}}}\right\|_{\text{F}}^2. \end{aligned}$$

Therefore,

$$T_{t+1} = \mathcal{P}_{\text{oblq}}\left(T_t + \frac{(A'-T_tS_t')S_t}{\gamma\|S_t'S_t\|_{\text{F}}}\right).$$

Step (A.4) becomes

$$\begin{aligned} &\arg\min_{S} \langle S-S_t, -2(A-S_tT_{t+1}')T_{t+1}\rangle + \eta\|T_{t+1}'T_{t+1}\|_{\text{F}}\|S-S_t\|_{\text{F}}^2 + \rho\|S\|_0 \\ &= \arg\min_{S} \left\| S - S_t - \frac{(A-S_tT_{t+1}')T_{t+1}}{\eta\|T_{t+1}'T_{t+1}\|_{\text{F}}}\right\|_{\text{F}}^2 + \frac{\rho}{\eta\|T_{t+1}'T_{t+1}\|_{\text{F}}}\|S\|_0. \end{aligned}$$

Hence

$$S_{t+1} = \mathcal{H}\left(S_t + \frac{(A-S_tT_{t+1}')T_{t+1}}{\eta\|T_{t+1}'T_{t+1}\|_{\text{F}}}, \sqrt{\rho/(\eta\|T_{t+1}'T_{t+1}\|_{\text{F}})}\right).$$

A.5. Proof of Proposition 2

This convergence result is a direct application of Theorem 1 in Bolte et al. (2014). Specifically, the constraint $\text{diag}(T'T) = \mathbf{I}_k$ is defined by a series of polynomial equations, so it is semi-algebraic by definition. Similar to the proof of Proposition 1, the other terms in (3.2) are also semi-algebraic. Therefore, (3.2) has the Kurdyka–Łojasiewicz property. The Assumptions 1 and 2 required in Bolte et al. (2014) can be easily verified to be true for (3.2). Thus, the bounded iterates (T_t,S_t) converge to a stationary point.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

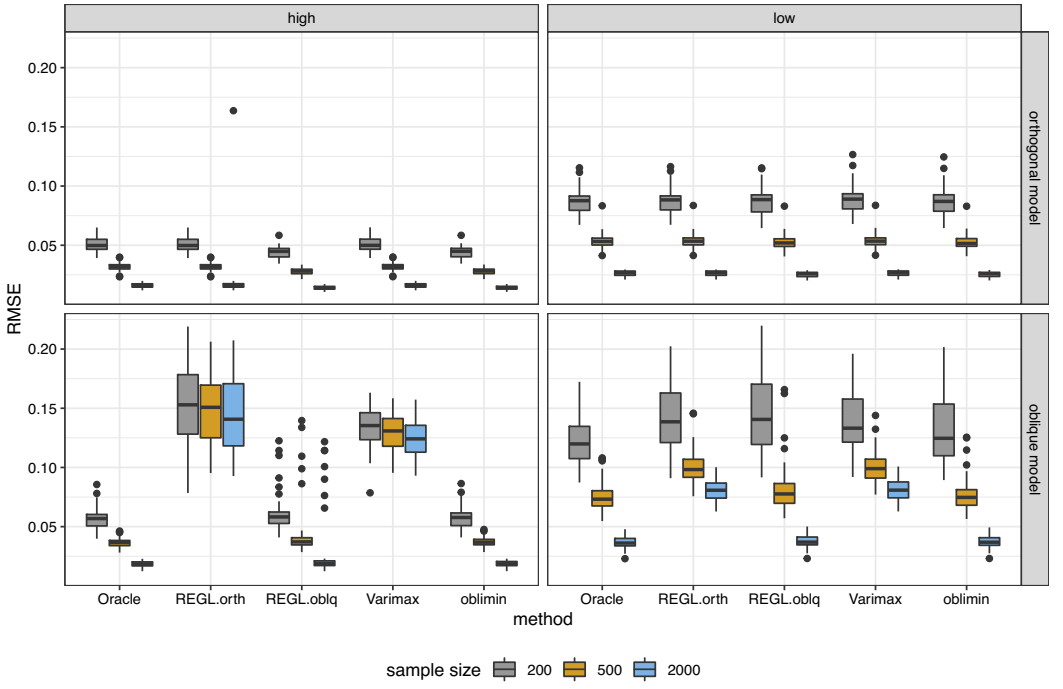


Figure A.1. Boxplot of the estimation error for loading matrix under factor models with perfect simple structure.

A.6. Simulation on simple structure rotation

This section provides the simulation results of the proposed method and existing methods for recovering simple structure in EFA. The simulation setting is very similar to that of the bi-factor rotation simulation in Section 4. There are a total of 22 items, and they admit a factor model with a perfect simple structure. There are four latent common factors, and these common factors have nonzero loadings on four, five, six, and seven items, respectively. The nonzero loadings are randomly selected from the intervals $[0.3, 0.6]$ or $[0.6, 0.9]$, and then each entry is randomly assigned a positive or negative sign. The population correlation matrix is generated as

$$R = \Lambda \Phi \Lambda' + \Psi^2,$$

where Φ is the factor correlation matrix and Ψ^2 is a diagonal matrix of uniqueness, chosen to constrain the diagonal elements of R to one. In the orthogonal model, Φ will be an identity matrix. In the oblique model, the correlations among common factors are randomly selected from the interval $[0.2, 0.6]$. If the generated correlation matrix is not positive definite, we re-generate a new one until it is positive definite. The final step generates the data from a multivariate normal distribution with zero mean and covariance matrix R, with a sample size $N \in \{200, 500, 2000\}$. The simulation is replicated 50 times for each scenario. The estimation accuracy is evaluated by the RMSE (4.1).

We compare our proposed methods with the conventional rotation methods, namely, the Varimax for orthogonal rotation and the oblimin for oblique rotation. The results are shown in Figures A.1 and A.2. The oracle method rotates the initial loading matrix toward the true loading matrix, so it is optimal in the sense of minimizing the RMSE of loading matrix estimation. Under the orthogonal model, all the methods achieve nearly optimal performance. Under the oblique model, these oblique methods (REGL.oblq and oblimin) are as good as the oracle method. Nonetheless, as shown in Figure A.2, our method has better performance in estimating the factor correlations. Our method exhibits outliers in Figures A.1 and A.2 because it gets trapped in local minima in these cases.

A.7. Comparing two formulations of oblique rotation

This section investigates the performance of two different formulations of the oblique rotation problem through simulations. The simulation settings are exactly the same as those in Section 4. The two formulations are

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

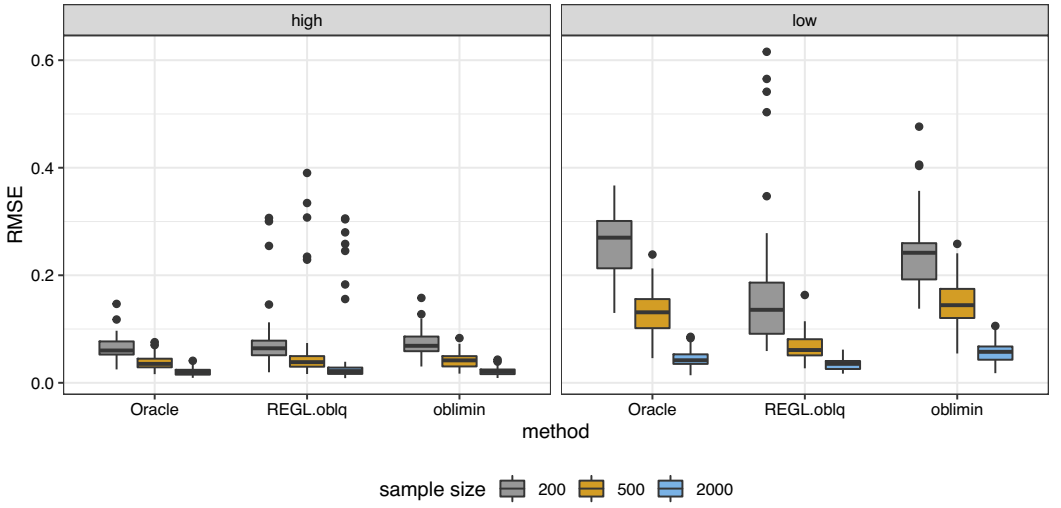


Figure A.2. Boxplot of the estimation error for factor correlation matrix under oblique factor models with perfect simple structure.

$$\begin{aligned} &\min_{T,S} \|A(T')^{-1} - S\|_{\mathrm{F}}^2, \\ &\text{s.t. } \mathrm{diag}(T'T) = \mathbf{I}_k, S \in \mathcal{S}_{\mathrm{bi}}, \end{aligned} \tag{A.5}$$

and

$$\begin{aligned} &\min_{T,S} \|A - ST'\|_{\mathrm{F}}^2, \\ &\text{s.t. } \mathrm{diag}(T'T) = \mathbf{I}_k, S \in \mathcal{S}_{\mathrm{bi}}. \end{aligned} \tag{A.6}$$

(A.5) is solved by the alternating minimization algorithm and the updating of T given a fixed S is achieved by Jennrich (2002)'s gradient projection algorithm. (A.6) is solved by Algorithm 4. Both solutions are partially orthogonalized to the semi-oblique bi-factor results. The two corresponding oracle methods are

$$\begin{aligned} &\min_{T} \|A(T')^{-1} - \Lambda\|_{\mathrm{F}}^2, \\ &\text{s.t. } \mathrm{diag}(T'T) = \mathbf{I}_k, \end{aligned} \tag{A.7}$$

and

$$\begin{aligned} &\min_{T} \|A - \Lambda T'\|_{\mathrm{F}}^2, \\ &\text{s.t. } \mathrm{diag}(T'T) = \mathbf{I}_k, \end{aligned} \tag{A.8}$$

where Λ is the true loading matrix. (A.7) and (A.8) are solved by the gradient projection algorithm.

The results of the four methods in estimating the loading matrix and factor correlation matrix are shown in Figures A.3 and A.4. The LoadingOracle (A.7) always has the optimal loading matrix estimation, as it minimizes the loading matrix RMSE (4.1). The CovOracle (A.8) produces comparable results for loading matrix estimation when the loadings are high but performs slightly worse when the loadings are low. The extremely poor results of CovOracle under ICP and ICBP conditions arises from a nearly singular estimate of the matrix T, with which $A(T')^{-1}$ will be extremely biased. A reversed trend is observed for the estimation of the factor correlation matrix. While LoadingOracle and CovOracle yield similar results when the loadings are high, CovOracle demonstrates superior estimation of the factor correlation matrix when the loadings are low. This outcome highlights the benefit of minimizing the covariance distance, even though it sacrifices some accuracy in the loading matrix estimation.

The pattern follows for the two practical methods, CovREGL (A.6) and LoadingREGL (A.5). These methods perform comparably in estimating the loading matrix when the loadings are high, but CovREGL shows slightly worse performance when the loadings are low. For factor correlation matrix estimation, the two methods are comparable when the loadings

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

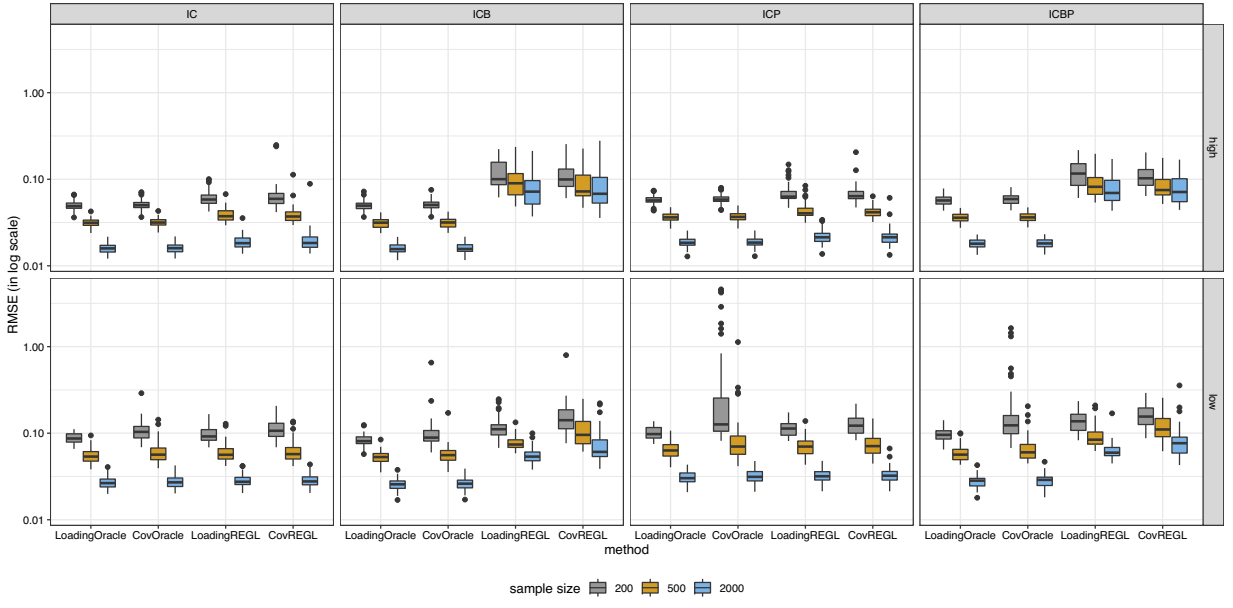


Figure A.3. Boxplot of the estimation error for loading matrix under semi-oblique bi-factor models.

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press

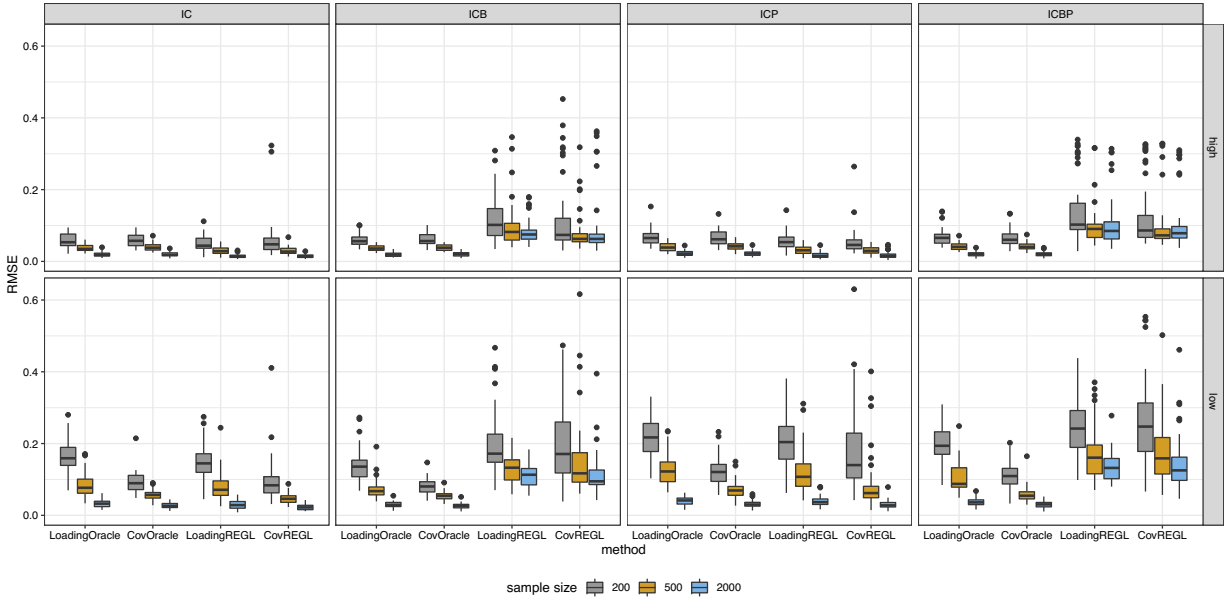


Figure A.4. Boxplot of the estimation error for factor correlation matrix under semi-oblique bi-factor models.

are high, but CovREGL outperforms LoadingREGL when the loadings are low, particularly under IC and ICP conditions. However, under ICB and ICBP conditions, the advantage of CovREGL over LoadingREGL in estimating the correlation matrix diminishes. This deterioration is likely due to model misspecification, as $\Lambda \notin \mathcal{S}_{\text{bi}}$ under these conditions. Surprisingly, CovREGL occasionally outperforms its oracle counterpart under IC and ICP conditions. In particular, the issue of extreme bias observed with CovOracle is no longer present with CovREGL.

Cite this article: Wu, Y., Liao, X. and Li, Q. (2025). A Generalized Factor Rotation Framework with Customized Regularization. *Psychometrika*, 1014–1038. https://doi.org/10.1017/psy.2025.1

https://doi.org/10.1017/psy.2025.1 Published online by Cambridge University Press