Supplement to "Joint Maximum Likelihood Estimation for High-dimensional Exploratory Item Factor Analysis"

A Proof of Theorems 1 and 3

Because Theorem 1 is a special case of Theorem 3, it is sufficient to prove Theorem 3. The proof of Theorem 3 is similar to that of Theorem 1 in Davenport et al. (2014). Thus, we only state the main steps and omit the repetitive details. Let $M = \Theta A^\top$ and $M^* = \Theta^* A^{*\top}$, and $\hat{M} = \hat{\Theta}\hat{A}^\top$. Note that l depend on (Θ, A) only through M. Thus, we write $l(M) = l(\Theta, A)$. Let

$$\bar{l}(M) = l(M) - l(\mathbf{0}), \tag{A.1}$$

where $\mathbf{0}$ is an $N \times J$ matrix whose entries are all zero. Then, we have the following lemma from Davenport et al. (2014).

Lemma A.1 (Lemma A.1 of Davenport et al. (2014))**.** *There exist constant C_0 and C_1 such that for all α, r, N, J and n,*

$$P\left(\sup_{\|M\|_* \leq \alpha\sqrt{rNJ}} |\bar{l}(M) - E\bar{l}(M)| \geq C_0 \alpha L_\alpha \sqrt{r}\sqrt{n(N+J) + NJ\log(N+J)}\right) \leq \frac{C_1}{N+J}, \tag{A.2}$$

where $L_\gamma = \sup_{|x|\leq\alpha} \frac{f'(x)}{f(x)(1-f(x))} < \infty$ and $\|\cdot\|_*$ denotes the nuclear norm of a matrix.

Let $\alpha = C^2$ and $r = K$ in the above lemma, we have

$$P\left(\sup_{\|M\|_* \le C^2\sqrt{KNJ}} |\bar{l}(M) - E\bar{l}(M)| \ge C_0 C^2 L_{C^2}\sqrt{K}\sqrt{n(N+J)+NJ\log(N+J)}\right) \le \frac{C_1}{N+J}. \tag{A.3}$$

Define

$$H = \left\{M = (m_{ij})_{1\le i\le N, 1\le j\le J} : m_{ij} = \mathbf{a}_j^\top \boldsymbol{\theta}_i, \|\boldsymbol{\theta}_i\| \le C \text{ and } \|\mathbf{a}_j\| \le C, \text{ for all } i, j\right\}. \tag{A.4}$$

Note that if $M \in H$, then

$$\|M\|_* \le \sqrt{NJ}\sqrt{\text{rank}(M)}\|M\|_\infty \le C^2\sqrt{KNJ}. \tag{A.5}$$

Thus, from (A.3), we further have

$$\begin{aligned} &P\left(\sup_{M\in H} |\bar{l}(M) - E\bar{l}(M)| \ge C_0 C^2 L_{C^2}\sqrt{K}\sqrt{n(N+J)+NJ\log(N+J)}\right) \\ \le &P\left(\sup_{\|M\|_* \le C^2\sqrt{KNJ}} |\bar{l}(M) - E\bar{l}(M)| \ge C_0 C^2 L_{C^2}\sqrt{K}\sqrt{n(N+J)+NJ\log(N+J)}\right) \\ \le &\frac{C_1}{N+J}. \end{aligned} \tag{A.6}$$

We use the following result, which is a slight modification of the last equation on p.210 of Davenport et al. (2014).

$$nD(M^*\|\hat{M}) \le 2\sup_{M\in H} |\bar{l}(M) - E\bar{l}(M)|, \tag{A.7}$$

where $\hat{M} = \hat{\Theta}\hat{A}^\top$, $D(M_1\|M_2)$ denotes the Kullback-Leibler divergence between the joint distribution of $\{Y_{ij}; 1 \le i \le N, 1 \le j \le J\}$ when the model parameters are M_1 and M_2. In addition, we have the following inequality, which is a direct application of Lemma A.2 in Davenport et al. (2014) and the third equation on page 211 of Davenport et al. (2014). For any M_1, M_2 such that $\|M_1\|, \|M_2\|_\infty \le C^2$,

$$\|M_1 - M_2\|_F^2 \le 8\beta_C NJD(M_1 \| M_2), \tag{A.8}$$

with $\beta_C = \sup_{|x|\le C^2} \frac{f(x)(1-f(x))}{f'(x)^2}$. According to Assumption A2, $\beta_C < \infty$. Combining (A.6), (A.7) and (A.8), we can see that with probability $1 - \frac{C_1}{N+J}$,

$$\|\hat{M} - M^*\|_F^2 \le \frac{16\beta_C NJ}{n} \times C_0 C^2 L_{C^2}\sqrt{K}\sqrt{n(N+J) + NJ\log(N+J)}. \tag{A.9}$$

Rearranging the terms, we have

$$\frac{1}{NJ}\|\hat{M} - M^*\|_F^2 \le 16\beta_C C_0 C^2 L_{C^2}\sqrt{\frac{K(J+N)}{n}} \times \sqrt{1 + \frac{NJ\log(N+J)}{n(N+J)}}. \tag{A.10}$$

For $n \ge (N+J)\log(NJ)$, we further have

$$\frac{NJ\log(N+J)}{n(N+J)} \le \frac{NJ}{(N+J)^2} \le \frac{1}{4}. \tag{A.11}$$

Combine the above equation with (A.10) and note that K is assumed fixed, we complete the proof.

B On Rotational Invariance

We summarize the phenomenon of rotational invariance by the following proposition.

Proposition B.1 (Rotational Invariance)**.** *Suppose that A3 is satisfied. For any $K \times K$ orthogonal matrix Q (i.e., $Q^\top Q = I_{K\times K}$), $\tilde{\Theta}^* = \Theta^* Q$ and $\tilde{A}^* = A^* Q$ satisfy $\tilde{\Theta}^*(\tilde{A}^*)^\top = \Theta^*(A^*)^\top$ and the constraints* (7) - (9)*.*

Proof of Proposition B.1. First, because Q is an orthogonal matrix, we can see that

$$\tilde{\Theta}^*\tilde{A}^{*\top} = \Theta^* Q Q^\top A^{*\top} = \Theta^* A^{*\top}. \tag{B.1}$$

We proceed to verify (7)-(9). We have

$$\mathbf{1}_N^\top \tilde{\Theta}^* = \mathbf{1}_N^\top \Theta^* Q = \mathbf{0}_K^\top Q = \mathbf{0}_K, \tag{B.2}$$

$$\frac{1}{N}\tilde{\Theta}^{*\top}\tilde{\Theta}^* = \frac{1}{N}(\Theta^* Q)^\top \Theta^* Q = \frac{1}{N} Q^\top \Theta^{*\top}\Theta^* Q = Q^\top Q = I_K. \tag{B.3}$$

Thus, (7)-(9) are verified. □

C Proof of Theorems 2 and 4

Because Theorem 2 is a special case of Theorem 4, we will only present the proof for the latter one. In fact, Theorem 4 is implied by the following lemma.

Lemma C.1. *Suppose that assumptions A1 - A5 are satisfied. Further assume that* $n \geq (N+J)\log(JN)$*. Then there exists a constant* C_4 *which does not depend on* N *and* J*, such that*

$$\min_Q \left\{ \frac{1}{J}\|A^* - \tilde{A}Q\|_F^2 : Q^\top Q = I_{K\times K} \right\} \leq C_4 \sqrt{\frac{J+N}{n}} \tag{C.1}$$

is satisfied with probability at least $1 - C_1/(N+J)$*, where* $\tilde{A}$ *is the standardized version of* $\hat{A}$*.*

Since $n \geq (N+J)\log(JN)$, when N, J grow to infinity simultaneously, the right hand side of (C.1) converges to zero and the probability of (C.1) being satisfied converges to one. It implies that the left hand side of (C.1) converges to zero in probability, which completes the proof of Theorem 4. In what follows, we prove Lemma C.1.

Proof of Lemma C.1. Let $H_1 = \tilde{\Theta}\tilde{A}^\top - \Theta^* A^{*\top}$ and $H_2 = \mathbf{1}_N \tilde{\mathbf{d}}^\top - \mathbf{1}_N \mathbf{d}^{*\top}$. Observe that

$$\begin{aligned}\|\hat{\Theta}\hat{A}^\top + \mathbf{1}_N\hat{\mathbf{d}} - \Theta^* A^{*\top} - \mathbf{1}_N \mathbf{d}^{*\top}\|_F^2 = \|H_1 + H_2\|_F^2 &= \|H_1\|_F^2 + \|H_2\|_F^2 + 2tr(H_1^\top H_2) \\ &= \|H_1\|_F^2 + \|H_2\|_F^2.\end{aligned} \tag{C.2}$$

The first equation is due to Proposition B.1. The last equation is due to $\tilde{\Theta}^\top \mathbf{1}_N = \mathbf{0}_K$ and $\Theta^{*\top}\mathbf{1}_N = \mathbf{0}_K$ under Assumption A3, where we write $\mathbf{0}_K$ for a K dimensional column vector whose entries are all 0's. Combine the above equation with (12), we have

$$\|\tilde{\Theta}\tilde{A}^\top - \Theta^* A^{*\top}\|_F^2 = \|H_1\|_F^2 \leq C_2 NJ\sqrt{\frac{J+N}{n}}, \tag{C.3}$$

probability at least $1 - C_1/(N+J)$ according to Theorem 1 and 3. Define the event

$$E_1 = \{\|\tilde{\Theta}\tilde{A}^\top - \Theta^* A^{*\top}\|_F^2 \leq C_2 NJ\sqrt{\frac{J+N}{n}}\}. \tag{C.4}$$

We will focus our analysis on E_1.

For the next step, we show that the column spaces of $\tilde{\Theta}$ and Θ^* are close to each other in the sine angle sense. For an $m \times n$ matrix H, we write $\sigma_1(H) \geq \sigma_2(H) \geq \ldots \geq \sigma_{\min(m,n)}(H)$ for the singular values of H in a descending order. Note that $\tilde{\Theta}^\top\tilde{\Theta} = \Theta^*\Theta^{*\top} = NI_K$, where I_K denotes the $K \times K$ identity matrix. Thus, $\sigma_1(\tilde{\Theta}) = \ldots = \sigma_K(\tilde{\Theta}) = \sigma_1(\Theta^*) = \ldots = \sigma_K(\Theta^*) = \sqrt{N}$. Under Assumption A4, we have $\sigma_K(A^*) \geq \sqrt{J}C_3$. Thus we have,

$$\sigma_K(\Theta^* A^{*\top}) \geq \sigma_K(\Theta^*)\sigma_K(A^*) \geq \sqrt{NJ}C_3. \tag{C.5}$$

See (Hogben, 2006, Chapter 17-8) for more details of the above inequality. Denote $M^* = \Theta^*(A^*)^\top$. Let $M^* = U^*\Sigma^* V^{*\top}$ be the reduced singular decomposition of M^*, where $V^* = (v_1^*, \ldots, v_K^*)$ is an orthonormal matrix, $\Sigma_{K\times K} = \mathrm{diag}(\sigma_1^*, \sigma_2^*, \ldots, \sigma_K^*)$, and $U^*_{N\times K}$ be the left singular matrix. We first show that the column space of Θ^* is the same as that of U^*. We know that

$$\Theta^*(A^*)^\top = U^*\Sigma^* V^{*\top}. \tag{C.6}$$

According to our discussions above, for large enough N and J, $\sigma_K^* \geq C_3\sqrt{NJ} > 0$. Thus, both Θ^* and A^* are of full rank. Therefore, by multiplying $A^*(A^{*\top}A^*)^{-1}$ on both sides we have

$$\Theta^* = U^*\Sigma^*V^{*\top}A^*(A^{*\top}A^*)^{-1}. \tag{C.7}$$

Note that $\Sigma^*V^{*\top}A^*(A^{*\top}A^*)^{-1}$ is also of full rank (rank K). Thus, Θ^* and U^* have the same column space, for which we denote it as $\mathcal{R}(U^*)$. Under the event E_1 and by Weyl's perturbation theorem (see, e.g. Stewart and Sun, 1990), we have

$$|\sigma_K^* - \hat{\sigma}_K| \leq \|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_2 \leq \|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_F,$$

where $\hat{\sigma}_K = \sigma_K(\hat{\Theta}\hat{A}^\top)$, $\|\cdot\|_2$ denotes the spectral norm of a matrix and the second inequality is due to the relationship between matrix spectral norm and matrix Frobenius norm. Thus, when event E_1 happens and for sufficiently large N and J,

$$\begin{aligned}\hat{\sigma}_K &\geq \sigma_K^* - \|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_F \\ &\geq \sqrt{NJ}\left(C_3 - C_2\left(\frac{J+N}{n}\right)^{\frac{1}{4}}\right) \\ &> 0,\end{aligned} \tag{C.8}$$

which is because $\|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_F$ is of order $o(\sqrt{NJ})$ when $n \geq (N+J)\log(JN)$ and N and J grow to infinity. Then following the same proof above, we also have that $\tilde{\Theta}$ and $\hat{U}$ have the same column space, where we $\hat{U}$ is left singular matrix of $\hat{\Theta}\hat{A}^\top$. That is, $\hat{\Theta}\hat{A}^\top = \hat{U}\hat{\Sigma}\hat{V}^\top$ and $\hat{\Sigma}$ is a $K \times K$ diagonal matrix. We write the column space of $\hat{U}$ as $\mathcal{R}(\hat{U})$. Following the Modified Davis-Kahan-Wedin sine theorem (Theorem 20) in O'Rourke et al. (2018), we have

$$\sin\angle(\mathcal{R}(U^*), \mathcal{R}(\hat{U})) \leq 2\frac{\|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_2}{\sigma_K^*}.$$

Because of the relationship between the matrix spectral norm and Frobenius norm and (C.5), we have

$$\sin\angle(\mathcal{R}(U^*),\mathcal{R}(\hat{U})) \leq 2\frac{\|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_2}{\sigma_K^*} \leq 2\frac{\|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_F}{\sigma_K^*} \leq 2\frac{\|\hat{\Theta}\hat{A}^\top - \Theta^*(A^*)^\top\|_F}{C_3\sqrt{NJ}}. \tag{C.9}$$

On the event E_1, we have

$$\sin\angle(\mathcal{R}(U^*),\mathcal{R}(\hat{U})) \leq 2C_2^{1/2}C_3^{-1}\left(\frac{J+N}{n}\right)^{\frac{1}{4}}. \tag{C.10}$$

On the other hand, using theory in canonical angles between column spaces of matrices (Hogben, 2006, Chapter 15-2), we know that the cosine angles between the column space of $\tilde{\Theta}$ and Θ^* are the singular values of $\frac{1}{N}\tilde{\Theta}^\top\Theta^*$. This fact together with (C.10) gives the following inequality on the event E_1,

$$\sqrt{1 - 4C_2C_3^{-2}\left(\frac{J+N}{n}\right)^{\frac{1}{2}}} \leq \frac{1}{N}\sigma_K(\tilde{\Theta}^\top\Theta^*) \leq \frac{1}{N}\sigma_1(\tilde{\Theta}^\top\Theta^*) \leq 1. \tag{C.11}$$

Now we choose a $K\times K$ orthogonal matrix Q that is close to $\frac{1}{N}\tilde{\Theta}^\top\Theta^*$. According to Theorem 4.1 of Higham (1989), the best choice of the orthogonal matrix Q that approximates $\frac{1}{N}\tilde{\Theta}^\top\Theta^*$ is

$$Q = B(B^\top B)^{-1/2}, \tag{C.12}$$

where $B = \frac{1}{N}\tilde{\Theta}^\top\Theta^*$, and

$$\|Q - \frac{1}{N}\tilde{\Theta}^\top\Theta^*\|_F = \sqrt{\sum_{i=1}^K [1 - \frac{1}{N}\sigma_i(\tilde{\Theta}^\top\Theta^*)]^2}. \tag{C.13}$$

Combine the above equation with (C.11), we have

$$\|Q - \frac{1}{N}\tilde{\Theta}^\top\Theta^*\|_F \leq \sqrt{K}4C_2C_3^{-2}\left(\frac{J+N}{n}\right)^{1/4}. \tag{C.14}$$

Let $\Delta = Q - \frac{1}{N}\tilde{\Theta}^\top\Theta^*$. Now we show that A^* and $\tilde{A}Q$ are close to each other. We have

$$
\begin{aligned}
\|A^* - \tilde{A}Q\|_F &= \|A^* - \tilde{A}(\frac{1}{N}\tilde{\Theta}^\top\Theta^* + \Delta)\|_F && \text{(C.15)}\\
&= \|A^* - \frac{1}{N}\hat{A}\hat{\Theta}^\top\Theta^* + \tilde{A}\Delta\|_F && \text{(C.16)}\\
&\le \|A^* - \frac{1}{N}A^*\Theta^{*\top}\Theta^*\|_F + \|\frac{1}{N}(A^*\Theta^{*\top} - \frac{1}{N}\hat{A}\hat{\Theta}^\top)\Theta^*\|_F + \|\tilde{A}\Delta\|_F && \text{(C.17)}\\
&\le \|A^* - A^*\|_F + \|\frac{1}{N}(A^*\Theta^{*\top} - \frac{1}{N}\hat{A}\hat{\Theta}^\top)\|_F\|\Theta^*\|_F + \|\tilde{A}\Delta\|_F && \text{(C.18)}\\
&\le C_2\left(\frac{N+J}{n}\right)^{1/4}\sqrt{\frac{J}{N}}\|\Theta^*\|_F + \|\tilde{A}\Delta\|_F && \text{(C.19)}\\
&\le C_2\left(\frac{N+J}{n}\right)^{1/4}\sqrt{J}\sqrt{C^2-1} + \sqrt{K}4C_2C_3^{-2}\left(\frac{J+N}{n}\right)^{1/4}\|\tilde{A}\|_F. && \text{(C.20)}
\end{aligned}
$$

On the other hand, we have

$$\tilde{A}\tilde{\Theta}^\top = \hat{A}\hat{\Theta}^\top. \tag{C.21}$$

Multiplying $\tilde{\Theta}(\tilde{\Theta}^\top\tilde{\Theta})^{-1}$ on both side gives,

$$\tilde{A} = \hat{A}\hat{\Theta}^\top\tilde{\Theta}(\tilde{\Theta}^\top\tilde{\Theta})^{-1} = \frac{1}{N}\hat{A}\hat{\Theta}^\top\tilde{\Theta}. \tag{C.22}$$

Thus,

$$\|\tilde{A}\|_F \le \frac{1}{N}\|\hat{A}\|_F\|\hat{\Theta}\|_F\|\tilde{\Theta}\|_F \le \frac{1}{N}\sqrt{J}C\sqrt{N}\sqrt{C^2-1}\sqrt{N}\sqrt{K} = \sqrt{J}C\sqrt{C^2-1}. \tag{C.23}$$

Combining the above equation with (C.20) completes the proof. $\square$

D Proof of Corollary 1

Corollary 1 is implied by Theorem 1 through the facts that

$$
\begin{aligned}
&\frac{\sum_{i=1}^{N}\sum_{j=1}^{J}\Big(f(\hat{d}_j+\hat{\mathbf{a}}_j^\top\hat{\boldsymbol{\theta}}_i)-f(d_j^*+(\mathbf{a}_j^*)^\top\boldsymbol{\theta}_i^*)\Big)^2}{NJ}\\
\leq&\left(\sup_{|x|\leq C^2}|f'(x)|\right)^2\times\frac{1}{NJ}\|\hat{\Theta}\hat{A}^\top+\mathbf{1}_N\hat{\mathbf{d}}^\top-\Theta^*(A^*)^\top-\mathbf{1}_N\mathbf{d}^{*\top}\|_F^2
\end{aligned}
\tag{D.1}
$$

and that $\big(\sup_{|x|\leq C^2}|f'(x)|\big)^2$ is finite according to condition A2.

E Standardization of CJMLE Solution

We provide the procedure for standardizing a set of parameters $(\Theta, A, \mathbf{d})$ to $(\tilde{\Theta}, \tilde{A}, \tilde{\mathbf{d}})$, so that

1. $\tilde{\Theta}\tilde{A}^\top+\mathbf{1}_N\tilde{\mathbf{d}}^\top=\Theta A^\top+\mathbf{1}_N\mathbf{d}^\top$,
2. $\mathbf{1}_N^\top\tilde{\Theta}_{[k]}=0$,
3. $\frac{1}{N}(\tilde{\Theta}_{[k]})^\top\tilde{\Theta}_{[k]}=1$,
4. $(\tilde{\Theta}_{[k]})^\top\tilde{\Theta}_{[k']}=0, \quad k,k'=1,...,K, k\neq k'$.

This is achieved by the following steps.

1. Set $\tilde{\mathbf{d}}=\mathbf{d}+\frac{1}{N}A\Theta^\top\mathbf{1}_N$.
2. Apply singular value decomposition to matrix $\Theta-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top\Theta$ and obtain

$$\Theta-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top\Theta=UDV^\top,$$

   where U is a $N\times K$ matrix containing all the left singular vectors, D is a $K\times K$ diagonal matrix with the diagonal entries being the singular values, and V is a $K\times K$ matrix, containing all the right singular vectors.

3. Set $\tilde{\Theta} = \sqrt{N} U$ and $\tilde{A} = \frac{1}{\sqrt{N}} AVD$.

F A SVD-based Algorithm

In what follows, we propose a fast algorithm for obtaining a good starting point for Algorithm 1. This algorithm is based on a method proposed in Chatterjee (2015). In the description of the algorithm below, we assume $N \geq J$. Modifications are needed when $J > N$.

Algorithm F.1 (SVD Algorithm for Starting Point)**.**

1. *(Initialization) Input responses y_{ij}, nonmissing response indicator ω_{ij}, dimension K of latent space, and tolerance ϵ.*

2. *Compute $\hat{p} = (\sum_{i=1}^{N} \sum_{j=1}^{J} \omega_{ij})/(NJ)$ as the proportion of observed responses.*

3. *Let $X = (x_{ij})_{N \times J}$, where*

$$x_{ij} = \begin{cases} 2y_{ij} - 1, & \textit{if } \omega_{ij} = 1, \\ 0, & \textit{otherwise.} \end{cases}$$

4. *Apply singular value decomposition to matrix X and obtain $X = \sum_{j=1}^{J} \sigma_j \mathbf{u}_j \mathbf{v}_j^\top$, where $\sigma_1 \geq \cdots \geq \sigma_J$ are the singular values and $\mathbf{u}_j$s and $\mathbf{v}_j$s are the left and right singular vectors.*

5. *Let*

$$\tilde{X} = (\tilde{x}_{ij})_{N \times J} = \sum_{j: \sigma_j \geq 2\sqrt{N\hat{p}}} \sigma_j \mathbf{u}_j \mathbf{v}_j^\top.$$

6. *Let* $M = (m_{ij})_{N \times J} =$, *where*

$$m_{ij} = \begin{cases} f^{-1}(\epsilon) & \textit{if } \tilde{x}_{ij} < -1+\epsilon, \\ f^{-1}(0.5(x_{ij}+1)) & \textit{if } -1+\epsilon \leq \tilde{x}_{ij} \leq 1-\epsilon, \\ f^{-1}(1-\epsilon) & \textit{if } \tilde{x}_{ij} > 1-\epsilon \end{cases}$$

7. *Set* $\mathbf{d}^{(0)} = (d_1^{(0)}, ..., d_J^{(0)})$, *where* $d_j^{(0)} = (\sum_{i=1}^N m_{ij})/N$.

8. *Apply singular value decomposition to matrix* $\tilde{M} = (m_{ij} - d_j^{(0)})_{N \times J}$ *and obtain* $\tilde{M} = \sum_{j=1}^J \tilde{\sigma}_j \tilde{\mathbf{u}}_j \tilde{\mathbf{v}}_j^\top$, *where* $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_J$ *are the singular values and* $\tilde{\mathbf{u}}_j$*s and* $\tilde{\mathbf{v}}_j$*s are the left and right singular vectors.*

9. *Set* $\Theta_{[k]}^{(0)} = \sqrt{N}\tilde{\mathbf{u}}_k$ *and* $A_{[k]}^{(0)} = \tilde{\sigma}_k \tilde{\mathbf{v}}_k / \sqrt{N}$, $k = 1, ..., K$.

10. *(Output) Output* $\Theta^{(0)}$, $A^{(0)}$, *and* $\mathbf{d}^{(0)}$ *as the starting point for Algorithm 1.*

The tolerance ϵ is a positive constant that is close to 0. A default value $\epsilon = 0.01$ is used in the analysis of the paper.

G Additional Results for Real Data Analysis

In what follows, we provide the estimated loading parameters from the real data analysis in Section 5. Specifically, the estimated factor loadings under the Geomin rotation are presented in Table 1. In addition, in Table 2, the standardized factor loadings are presented, which are obtained by scaling the estimated factor loadings in Table 1 by 1.7. Thanks to the connection between the probit and logistic link functions, the standardized factor loadings should be close to the estimated factor loadings from a probit IFA model.

Item#	F1	F2	F3
1	0.19	2.12	0.41
2	0.27	1.39	−0.10
3	0.49	1.24	0.58
4	0.19	1.45	0.78
5	0.82	1.29	0.07
6	0.38	1.12	0.23
7	−0.35	1.90	0.01
8	0.31	0.85	−0.23
9	−0.67	1.15	0.49
10	−0.11	1.56	0.71
11	0.18	0.85	−0.10
12	−0.11	1.81	0.33
13	0.01	0.46	0.48
14	0.89	0.97	0.29
15	−0.21	1.09	−0.98
16	0.48	1.16	−0.52
17	0.04	0.64	−0.01
18	−0.11	1.20	−0.56
19	0.11	0.58	−0.30
20	−0.01	1.67	−0.20
21	−0.54	2.11	−0.59
22	−0.53	1.91	−0.58
23	−0.63	1.55	0.02
24	0.18	1.15	−0.48
25	0.15	0.76	−0.06
26	−0.46	1.18	0.06
27	0.35	1.80	0.14
28	0.95	1.04	0.42
29	−0.18	0.48	0.44
30	0.11	1.30	−0.24
31	−0.17	1.23	−0.06
32	−0.21	0.58	−0.26
33	0.34	−0.15	−0.24
34	2.80	−0.09	0.51
35	4.02	−0.23	−0.02
36	2.09	0.16	−0.06
37	2.17	−0.25	−0.40
38	1.64	0.08	0.07
39	2.08	−0.68	−0.40
40	1.00	0.27	−0.57
41	1.87	−0.14	−0.07
42	2.91	0.51	−0.05
43	0.93	−0.04	−0.04
44	3.12	−1.02	0.21
45	1.18	1.00	0.67
46	0.49	0.01	−0.09
47	0.79	0.41	−0.32
48	0.93	0.63	0.20
49	0.43	−0.01	0.12
50	2.55	0.12	−0.09
51	1.98	−0.06	0.03
52	3.67	0.17	0.13
53	3.91	0.73	−0.10
54	1.88	−0.13	0.03
55	2.74	0.20	−0.04
56	0.31	0.60	2.28
57	0.11	0.32	1.58
58	0.46	−1.20	2.12
59	−0.03	0.58	1.64
60	−0.01	−0.29	1.78
61	−0.03	0.33	2.11
62	0.41	−0.10	1.67
63	−0.50	0.00	2.04
64	−0.33	−0.71	2.76
65	−0.09	−0.26	1.37
66	−0.28	0.40	1.88
67	0.13	−0.11	0.86
68	0.02	0.30	0.58
69	0.14	0.36	1.25
70	0.01	0.53	1.35
71	0.72	−0.21	1.18
72	−0.15	0.68	0.77
73	−0.28	−0.70	2.22
74	−0.26	0.19	1.97
75	−0.22	0.62	1.55
76	−0.25	−0.53	1.81
77	0.84	0.32	1.50
78	0.32	0.01	1.32
79	0.45	0.49	1.01
.	.	.	.

Table 1: Results of fitting a three-factor model to the EPQ-R data: The original estimated factor loadings (after Geomin rotation).

Item#	F1	F2	F3	Item#	F1	F2	F3
1	0.11	1.24	0.24	**41**	1.10	−0.08	−0.04
2	0.16	0.81	−0.06	**42**	1.71	0.30	−0.03
3	0.29	0.73	0.34	**43**	0.55	−0.02	−0.02
4	0.11	0.85	0.46	**44**	1.83	−0.60	0.12
5	0.48	0.76	0.04	**45**	0.69	0.59	0.40
6	0.22	0.66	0.13	**46**	0.29	0.01	−0.05
7	−0.21	1.11	0.01	**47**	0.47	0.24	−0.19
8	0.18	0.50	−0.14	**48**	0.55	0.37	0.12
9	−0.39	0.67	0.29	**49**	0.25	−0.01	0.07
10	−0.07	0.92	0.42	**50**	1.50	0.07	−0.06
11	0.10	0.50	−0.06	**51**	1.16	−0.04	0.02
12	−0.06	1.06	0.19	**52**	2.15	0.10	0.07
13	0.00	0.27	0.28	**53**	2.30	0.43	−0.06
14	0.53	0.57	0.17	**54**	1.11	−0.08	0.02
15	−0.12	0.64	−0.57	**55**	1.61	0.12	−0.03
16	0.28	0.68	−0.30	**56**	0.18	0.35	1.34
17	0.02	0.37	0.00	**57**	0.07	0.19	0.93
18	−0.06	0.71	−0.33	**58**	0.27	−0.70	1.25
19	0.06	0.34	−0.18	**59**	−0.01	0.34	0.96
20	0.00	0.98	−0.12	**60**	0.00	−0.17	1.04
21	−0.32	1.24	−0.35	**61**	−0.02	0.19	1.24
22	−0.31	1.12	−0.34	**62**	0.24	−0.06	0.98
23	−0.37	0.91	0.01	**63**	−0.29	0.00	1.20
24	0.11	0.68	−0.28	**64**	−0.19	−0.41	1.62
25	0.09	0.45	−0.03	**65**	−0.06	−0.16	0.81
26	−0.27	0.69	0.03	**66**	−0.17	0.23	1.11
27	0.20	1.06	0.08	**67**	0.07	−0.07	0.51
28	0.56	0.61	0.25	**68**	0.01	0.18	0.34
29	−0.11	0.28	0.26	**69**	0.08	0.21	0.74
30	0.06	0.76	−0.14	**70**	0.00	0.31	0.79
31	−0.10	0.72	−0.03	**71**	0.42	−0.12	0.69
32	−0.12	0.34	−0.15	**72**	−0.09	0.40	0.45
33	0.20	−0.09	−0.14	**73**	−0.17	−0.41	1.31
34	1.64	−0.05	0.30	**74**	−0.15	0.11	1.16
35	2.36	−0.13	−0.01	**75**	−0.13	0.36	0.91
36	1.23	0.09	−0.04	**76**	−0.15	−0.31	1.06
37	1.28	−0.15	−0.23	**77**	0.50	0.19	0.88
38	0.96	0.04	0.04	**78**	0.19	0.01	0.78
39	1.22	−0.40	−0.24	**79**	0.27	0.29	0.59
40	0.59	0.16	−0.34	.	.	.	.

Table 2: Results of fitting a three-factor model to the EPQ-R data: The standardized estimated factor loadings (after Geomin rotation).

References

Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1):177–214.

Davenport, M. A., Plan, Y., van den Berg, E., and Wootters, M. (2014). 1-bit matrix completion. *Information and Inference*, 3(3):189–223.

Higham, N. J. (1989). Matrix nearness problems and applications. In Gover, M. J. C. and Barnett, S., editors, *Applications of matrix theory*, pages 1–27. Oxford University Press, Oxford, UK.

Hogben, L. (2006). *Handbook of linear algebra*. CRC Press, Boca Raton, FL.

O'Rourke, S., Vu, V., and Wang, K. (2018). Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications*, 540:26–59.

Stewart, G. and Sun, J. (1990). *Matrix perturbation theory*. Academic Press, Cambridge, MA.