A note on identifiability conditions in confirmatory factor analysis

William Leeb*

Abstract

Recently, Chen, Li and Zhang established conditions characterizing asymptotic identifiability of latent factors in confirmatory factor analysis. We give an elementary proof showing that a similar characterization holds non-asymptotically, and prove a related result for identifiability of factor loadings.

1 Introduction

We consider the problem of recovering a low-rank factorization of a large matrix M. We assume that $M = \Theta A^T$, where Θ and A each have K columns for a known value of K much smaller than either dimension of M. We think of the rows of M as labeling members of a population, and the columns of M as labeling attributes. Following the language of factor analysis, we describe the columns of Θ as "latent factors", and the columns of A as "factor loadings". For general background on factor analysis, we refer the reader to [2, 4, 12], and references contained therein.

The factorization $M = \Theta A^T$ is not unique since for any K-by-K invertible matrix B, $M = (\Theta B)(AB^{-T})^T$. In confirmatory factor analysis we are given additional "side information" that specifies the support of each column of A. More precisely, we have a binary matrix Q of the same dimensions as A, where $Q_{jk} = 0$ implies $A_{jk} = 0$. Q is referred to as a "design matrix". The question then arises as to what conditions on Q are enough to ensure uniqueness of M's factorization, up to a rescaling of the columns of Θ and A.

The recent paper [6] provides necessary and sufficient conditions on the matrix Q under which individual columns of Θ (the latent factors) are asymptotically determined up to rescaling, or *identifiable*, under certain assumptions on Θ and A. In this note, we show that a similar characterization applies as well in a non-asymptotic setting. We also provide an elementary proof of a similar characterization of the identifiability of A's columns (the factor loadings), a question which has also attracted interest [15, 13, 1, 16, 5].

The remainder of this note is structured as follows. In Section 2, we describe the precise model and terminology we will be using throughout. In Section 3, we state and prove the main results, namely characterizing when the columns of Θ and A are identifiable within our model. In Section 4, we compare our results to those in [6].

2 Definitions and model description

For positive integers K, N and J, let $\Theta_1, \dots, \Theta_K$ be vectors in $\mathbb{R}^N$ and let $A_1, \dots, A_K$ be vectors in $\mathbb{R}^J$; and define the matrices $\Theta = [\Theta_1, \dots, \Theta_K] \in \mathbb{R}^{N \times K}$ and $A = [A_1, \dots, A_K] \in \mathbb{R}^{J \times K}$. Define the $N \times J$ matrix $M = \Theta A^T$. Additionally, let $Q \in \{0,1\}^{J \times K}$ be a binary matrix with columns $Q_1, \dots, Q_K$.

Remark 2.1. The main results of this paper, namely Theorems 3.1 and 3.2, do not depend on the precise values of N and J (so long as both are sufficiently big), and apply equally well in the doubly-asymptotic setting where Θ and A both have infinitely many rows; that is, we

*School of Mathematics, University of Minnesota

may view $\Theta_1, \ldots, \Theta_K$ and $A_1, \ldots, A_K$ as functions on $\mathbb{Z}_+$, the set of positive integers. This is a common assumption in work on factor analysis, and appears in the work [6] on which the present work is based. Doubly-asymptotic models for related problems also appear in, for example, [3], [17], [8], [11], [9], [14], [7], [10], to give only a partial list. This setting provides a formalization of the "high-dimensional, large-sample" regime where the number of parameters (in our setting, J) is comparable to the number of observations (in our setting, N). Because certain random quantities converge to deterministic limits as N and J grow, this asymptotic model often provides a convenient framework for analyzing statistical phenomena.

Definition 2.1. *For a subset $S \subset \{1, \ldots, K\}$, we define $\mathcal{R}(S) \subset \mathbb{Z}_+$ to be the set of indices $j \in \mathbb{Z}_+$ such that $Q_k(j) = 1$ whenever $k \in S$, and $Q_k(j) = 0$ whenever $k \notin S$.*

We introduce some additional notation. For a vector $\mathbf{x} \in \mathbb{R}^I$ and a subset $\mathcal{R} \subset \{1, \ldots, I\}$, we will denote by $\mathbf{x}(\mathcal{R})$ the restriction of $\mathbf{x}$ to $\mathcal{R}$. If in addition $S \subset \{1, \ldots, K\}$ and B is a $I \times K$ matrix, we will denote by $B_{[\mathcal{R},S]}$ the submatrix of B with rows from $\mathcal{R}$ and columns from S. We will also use colon notation to denote ranges of indices; for example, $\mathbf{x}(1:n)$ denotes the subvector consisting of the first n entries of $\mathbf{x}$; $B_{[:,S]}$ denotes the submatrix of B with column indices in S; and so forth.

We now describe our assumptions on Θ, A and Q.

Model Assumptions

1. The columns of Θ are linearly independent.
2. If $S \subset \{1, \ldots, K\}$ and $\mathcal{R}(S)$ is non-empty, then the columns of the submatrix $A_{[\mathcal{R}(S),S]}$ are linearly independent.
3. For any $k = 1, \ldots, K$, $A_k(j) = 0$ whenever $Q_k(j) = 0$.
4. There is a constant $C > 0$ such that

$$\sup_{\substack{1 \le k \le K \\ i \ge 1}} |\Theta_k(i)| < C, \qquad \sup_{\substack{1 \le k \le K \\ j \ge 1}} |A_k(j)| < C. \tag{1}$$

With the model described, we now define *identifiability* of the latent factors and factor loadings.

Definition 2.2. *The latent factor Θ_k is* identifiable *if for any decomposition $M = \widetilde{\Theta}\widetilde{A}^T$ satisfying assumptions 1 – 4, Θ_k and $\widetilde{\Theta}_k$ are linearly dependent (that is, one is a scalar multiple of the other). Similarly, the factor loading A_k is* identifiable *if for any decomposition $M = \widetilde{\Theta}\widetilde{A}^T$ satisfying assumptions 1 – 4, A_k and $\widetilde{A}_k$ are linearly dependent.*

Remark 2.2. [6] defines identifiability of Θ_k to mean that the angle between $\Theta_k(1:N)$ and $\widetilde{\Theta}_k(1:N)$ converges to 0 as $N \to \infty$, which is a weaker notion than the one we employ. We will compare the two definitions in Section 4.2.

Remark 2.3. Assumptions analogous to 1 and 2 are found in [6]. In Section 4.1, we will show that assumptions 1 and 2 are weaker than those found in [6].

Remark 2.4. Assumption 4 is slightly different than the boundedness assumption from [6]. Because we assume the supremum is strictly less than C, sufficiently small perturbations are permitted without violating the bound. This simplifies some of the analysis without changing the essential properties of the model.

Before stating the main results, we introduce the key concept of *masking*, defined as follows.

Definition 2.3. *We say k'* masks *k if $\operatorname{supp}(Q_{k'}) \subset \operatorname{supp}(Q_k)$.*

3 Main results

We provide necessary and sufficient conditions on the design matrix Q which characterize when the latent factors Θ_k and the factor loadings A_k are identifiable. Theorem 3.1 addresses identifiability of Θ_k; the identifiability condition is similar to the one in [6], but makes sense in a non-asymptotic setting; and the proof of Theorem 3.1 amounts to a tightening of the proof of Proposition 8 in [6]. Theorem 3.2 characterizes identifiability of A_k, and appears to be new.

Theorem 3.1. *For each k, Θ_k is identifiable if and only if k does not mask any $k' \neq k$, or equivalently if*

$$\{k\} = \bigcap_{\substack{S \subset \{1,\ldots,K\} \\ k \in S \\ \mathcal{R}(S) \text{ non-empty}}} S. \tag{2}$$

Theorem 3.2. *Suppose* $\mathrm{supp}(Q_k)$ *is non-empty for all k. Then for each k, A_k is identifiable if and only if no $k' \neq k$ masks k.*

Remark 3.1. The identifiability conditions in Theorems 3.1 and 3.2 may be efficiently checked for any given design matrix Q, simply by verifying that $\mathrm{supp}(Q_k)$ does not lie entirely within $\mathrm{supp}(Q_{k'})$ when $k' \neq k$.

Remark 3.2. Clearly, if all Θ_k are identifiable – that is, the entire matrix Θ is identifiable – then so too is the entire matrix A; and vice versa. Theorems 3.1 and 3.2 together provide more granular information on the relationship between Θ and A. Specifically, if Θ_k is *not* identifiable, Theorem 3.1 states that there must be some $k' \neq k$ that is masked by k. Theorem 3.2, in turn, tells us that $A_{k'}$ is not identifiable. That is, knowing which column of Θ is not identifiable automatically tells us what which column of A is not identifiable. The same reasoning applies in reverse as well: knowing which column of A is not identifiable automatically tells us what which column of Θ is not identifiable.

3.1 Technical lemmas

Lemma 3.3. *Suppose* $\mathrm{supp}(Q_{k'})$ *is not empty. If k' masks $k \neq k'$, then A_k and $A_{k'}$ are linearly independent.*

Proof. Take any index $j \in \mathrm{supp}(Q_{k'})$, and let $S = \{k'' : Q_{k''}(j) = 1\}$; then j is contained in $\mathcal{R}(S)$, and $k, k' \in S$. From assumption 2 $A_{[\mathcal{R}(S),S]}$ has linearly independent columns; in particular $A_k(\mathcal{R}(S))$ and $A_{k'}(\mathcal{R}(S))$ are independent, and hence so too are $A_{k'}$ and A_k. $\square$

Lemma 3.4. *Suppose k' masks $k \neq k'$, and let $\epsilon \in \mathbb{R}$. Then $\widetilde{A} = [A_1, \ldots, A_{k-1}, A_k + \epsilon A_{k'}, A_{k+1}, \ldots, A_K]$ satisfies assumptions 2 and 3.*

Proof. Assumption 3 is immediate, since the support of $A_k + \epsilon A_{k'}$ is still contained in $\mathrm{supp}(Q_k)$, because $\mathrm{supp}(A_{k'}) \subset \mathrm{supp}(A_k)$.

We now show that assumption 2 holds. Without loss of generality, take $k = K$ and $k' = 1$; so 1 masks K. Take any subset $S \subset \{1, \ldots, K\}$, with $\mathcal{R}(S)$ non-empty. We will show that the columns of $\widetilde{A}_{[\mathcal{R}(S),S]}$ are linearly independent. This follows immediately from assumption 2 for A if $K \notin S$; so assume $K \in S$.

First suppose $1 \in S$. From assumption 2 the vectors $A_k(\mathcal{R}(S))$, $k \in S$, are linearly independent. Since 1 and K are in S, linear independence is preserved after replacing $A_K(\mathcal{R}(S))$ with $A_K(\mathcal{R}(S)) + \epsilon A_1(\mathcal{R}(S))$.

Next, suppose $1 \notin S$. Then by definition $\mathrm{supp}(Q_1)$ is disjoint from $\mathcal{R}(S)$, so $A_1(j) = 0$ for $j \in \mathcal{R}(S)$. Consequently, $A_K(\mathcal{R}(S)) = A_K(\mathcal{R}(S)) + \epsilon A_1(\mathcal{R}(S))$, and since $A_k(\mathcal{R}(S))$, $k \in S$, are linearly independent, the same is true after replacing $A_K(\mathcal{R}(S))$ by $A_K(\mathcal{R}(S)) + \epsilon A_1(\mathcal{R}(S))$. $\square$

Lemma 3.5. *Suppose k does not mask any $k' \neq k$. Then*

$$\{k\} = \bigcap_{\substack{S\subset\{1,\dots,K\} \\ k\in S \\ \mathcal{R}(S) \text{ non-empty}}} S. \tag{3}$$

Proof. Because k does not mask any other k', there must exist some subset $S \subset \{1,\dots,K\}$ containing k with $\mathcal{R}(S)$ non-empty. Indeed, $\mathrm{supp}(Q_k)$ must be non-empty, since otherwise k would mask every k'. But each $j \in \mathrm{supp}(Q_k)$ is contained in $\mathcal{R}(S)$, where $S = \{k'' : Q_{k''}(j) = 1\}$; and $k \in S$. Consequently, the right side of (3) is non-empty, and obviously contains k.

To show the reverse inclusion, take any $k' \neq k$. Since k does not mask k', $\mathrm{supp}(Q_k) \setminus \mathrm{supp}(Q_{k'})$ is non-empty. Each $j \in \mathrm{supp}(Q_k) \setminus \mathrm{supp}(Q_{k'})$ is contained in $\mathcal{R}(S)$, where $S = \{k'' : Q_{k''}(j) = 1\}$ contains k but not k', implying that k' is not contained in the right side of (3). □

The converse to Lemma 3.5 is also true:

Lemma 3.6. *Suppose* (3) *holds. Then k does not mask any $k' \neq k$.*

Proof. Without loss of generality, suppose $k = K$. If $\mathrm{supp}(Q_K)$ were empty (i.e. $Q_K(j) = 0$ for all j), then for any $S \subset \{1,\dots,K\}$ containing K, $\mathcal{R}(S) \subset \mathrm{supp}(Q_K)$ would also be empty, and the right side of (3) would be empty; a contradiction. Consequently, $\mathrm{supp}(Q_K)$ must be non-empty.

For contradiction, suppose without loss of generality that K masks 1; then $\mathrm{supp}(Q_K) \setminus \mathrm{supp}(Q_1)$ is empty. If $S \subset \{1,\dots,K\}$ contains K but not 1, then $\mathcal{R}(S) \subset \mathrm{supp}(Q_K) \setminus \mathrm{supp}(Q_1)$, so $\mathcal{R}(S)$ is also empty and S is not included in the right side of (3). Therefore, the only S included on the right side of (3) contain both K and 1. But then 1 is also in the intersection, a contradiction. □

3.2 Proof of Theorem 3.1

First, suppose, without loss of generality, that K masks 1. We write:

$$M = \Theta_1 A_1^T + \Theta_2 A_2^T + \cdots + \Theta_K A_K^T = \Theta_1 (A_1 + \epsilon A_K)^T + \Theta_2 A_2^T + \cdots + (\Theta_K - \epsilon\Theta_1) A_K^T, \tag{4}$$

where ϵ is sufficiently small so as to not violate assumption 4. From Lemma 3.4, assumptions 2 and 3 and are still satisfied by $A_1 + \epsilon A_K, A_2, \dots, A_K$. Assumption 1 still holds if we replace Θ_K by $\Theta_K - \epsilon\Theta_1$. Since assumption 1 implies $\Theta_K - \epsilon\Theta_1$ and Θ_K are linearly independent, Θ_K is not identifiable.

For the other direction, assume that component K does not mask any other component $k \neq K$. Suppose $M = \widetilde{\Theta}\widetilde{A}^T$ is another factorization of M satisfying the model assumptions 1 – 4. We will show that Θ_K and $\widetilde{\Theta}_K$ are linearly dependent.

Observe that $\mathrm{supp}(Q_K)$ is non-empty, since otherwise it would mask every k. Each $j \in \mathrm{supp}(Q_K)$ is contained in $\mathcal{R}(S)$, where $S = \{k : Q_k(j) = 1\}$. Then if $k \notin S$ and $j \in \mathcal{R}(S)$, we must have $A_k(j) = 0$. Consequently, if $j \in \mathcal{R}(S)$, $M_j(i) = \sum_{k=1}^K \Theta_k(i) A_k(j) = \sum_{k\in S} \Theta_k(i) A_k(j)$, and so we may write

$$M_{[:,\mathcal{R}(S)]} = \Theta_{[:,S]} (A_{[\mathcal{R}(S),S]})^T. \tag{5}$$

By assumption 2, $A_{[\mathcal{R}(S),S]}$ has linearly independent columns, and since Θ has linearly independent columns, the column space of $M_{[:,\mathcal{R}(S)]}$ has dimension $|S|$. Consequently, if we define $V_S \equiv \mathrm{span}\{M_j : j \in \mathcal{R}(S)\}$, then $V_S = \mathrm{span}\{\Theta_k : k \in S\}$ and $\dim(V_S) = |S|$.

Because the Θ_k are linearly independent and $V_S = \mathrm{span}\{\Theta_k : k \in S\}$, we have $V_S \cap V_{S'} = V_{S\cap S'}$. Consequently

$$\Theta_K \in V_{S_K} = \bigcap_{\substack{S\subset\{1,\dots,K\} \\ K\in S \\ \mathcal{R}(S) \text{ non-empty}}} V_S \tag{6}$$

where S_K is the intersection of all sets S with $K \in S$ and $\mathcal{R}(S)$ non-empty. But because K does not mask any $k \neq K$, Lemma 3.5 implies that $S_K = \{K\}$, and so $V_{S_K} = \text{span}\{\Theta_K\}$. But the exact same argument with $\widetilde{\Theta}$ and $\widetilde{A}$ in place of Θ and A also shows $V_{S_K} = \text{span}\{\widetilde{\Theta}_K\}$. Consequently, Θ_K and $\widetilde{\Theta}_K$ are linearly dependent.

3.3 Proof of Theorem 3.2

First, let us suppose without loss of generality that $k = K$ is masked by $k' = 1$. We write

$$M = \Theta_1 A_1^T + \Theta_2 A_2^T + \cdots + \Theta_K A_K^T = (\Theta_1 - \epsilon\Theta_K)A_1^T + \Theta_2 A_2^T + \cdots + \Theta_K (A_K + \epsilon A_1)^T, \quad (7)$$

where ϵ is sufficiently small so as to not violate assumption 4. From Lemma 3.4, assumptions 2 and 3 and are still satisfied by $A_1, A_2, \ldots, A_K + \epsilon A_1$. Assumption 1 still holds if we replace Θ_1 by $\Theta_1 - \epsilon\Theta_K$. From Lemma 3.3, $A_K + \epsilon A_1$ and A_K are linearly independent. Consequently, A_K is not identifiable.

For the other implication, suppose $M = \widetilde{\Theta}\widetilde{A}^T$ is another factorization within the same model, and that $\widetilde{A}_K$ and A_K are linearly independent. Let $\mathcal{R}_K = \text{supp}(Q_K)^c$ be the set of roots of Q_K; then A_K and $\widetilde{A}_K$ are both zero on $\mathcal{R}_K$.

Since the column space of $\widetilde{A}$ is contained in the column space of A, $\widetilde{A}_K$ is in the span of $A_1, \ldots, A_K$. Therefore, there are coefficients $c_1, \ldots, c_{K-1}$, not all zero, so that

$$\sum_{k=1}^{K-1} c_k A_k(\mathcal{R}_K) = \mathbf{0}. \quad (8)$$

Suppose, without loss of generality, that $c_1 \neq 0$. We will show that 1 masks K. Suppose not; then $\text{supp}(Q_1) \cap \mathcal{R}_K = \text{supp}(Q_1) \setminus \text{supp}(Q_K)$ is non-empty. Take any $j \in \text{supp}(Q_1) \cap \mathcal{R}_K$; then j is contained in $\mathcal{R}(S)$, where $S = \{k : Q_k(j) = 1\}$. Since $K \notin S$, $\mathcal{R}(S) \subset \mathcal{R}_K$. Furthermore, if $k \notin S$ and $j' \in \mathcal{R}(S)$ then $A_k(j') = 0$. Hence from (8)

$$\sum_{k \in S} c_k A_k(\mathcal{R}(S)) = \mathbf{0}. \quad (9)$$

But by assumption 2, the columns of $A_{[\mathcal{R}(S),S]}$ are linearly independent; so we must have $c_k = 0$ for all $k \in S$. Since $1 \in S$, this contradicts that $c_1 \neq 0$.

4 Discussion

We conclude with a discussion comparing our work to [6]. In this section, we will treat Θ_k and A_k as functions on $\mathbb{Z}_+$, rather than finite-length vectors, since this is the setting used in [6]. As noted in Remark 2.1, Theorems 3.1 and 3.2 are valid in this doubly-asymptotic model.

To aid the discussion, it is convenient to define the following notion.

Definition 4.1. *A subset* $\Delta \subset \mathbb{Z}_+$ *is* negligible *if*

$$\lim_{N\to\infty} \frac{|\Delta \cap \{1, \ldots, N\}|}{N} = 0. \quad (10)$$

In other words, Δ is negligible if the fraction of entries it contains from $\{1, \ldots, N\}$ vanishes as N grows.

Remark 4.1. Any finite subset of $\mathbb{Z}_+$ is negligible. Furthermore, the definition of negligible depends crucially on the ordering of $\mathbb{Z}_+$. Indeed, if Δ is any infinite subset of $\mathbb{Z}_+$, we can always reorder $\mathbb{Z}_+$ so that $|\Delta \cap \{1, \ldots, N\}|/N$ converges to a positive number, by interlacing the elements of Δ and $\mathbb{Z}_+ \setminus \Delta$. Similarly, we can reorder $\mathbb{Z}_+$ so that arbitrarily large gaps occur between the elements of Δ, making Δ negligible under that ordering.

4.1 Assumptions 1 and 2

In [6], assumption 1 is replaced by the assumption that the limsup of the minimum singular values of the matrices $\Theta_{[1:N,1:K]}/\sqrt{N}$ is positive as $N \to \infty$; an analogous assumption is made in place of assumption 2. The assumptions in [6] imply assumptions 1 and 2. Indeed, suppose $B = [B_1, \ldots, B_K]$, where each B_k is a bounded function on $\mathbb{Z}_+$; and suppose that

$$\limsup_{n\to\infty} \frac{\sigma_{\min}(B_{[1:n,1:K]})}{\sqrt{n}} > 0, \tag{11}$$

where $\sigma_{\min}$ denotes the smallest singular value. Then $B_1, \ldots, B_K$ are linearly independent, since for sufficiently large n the minimum singular value of $B_{[1:n,1:K]}$ must be positive.

It is not difficult to see that the converse statement is false; that is, assumptions 1 and 2 do not imply the corresponding assumptions from [6]. For example, we may take $\text{supp}(B_1)$ to be the positive even integers, and $\text{supp}(B_2)$ to be the positive odd integers; and define $B_1(2i) = 1/2i$ and $B_2(2i-1) = 1/(2i-1)$. Then B_1 and B_2 are linearly independent. Take any large n and $m < n$. Define $\mathcal{T}_n = \{1, \ldots, m\}$ and $\mathcal{R}_n = \{m+1, \ldots, n\}$, and partition $B^{(n)} \equiv B_{[1:n,1:2]}/\sqrt{n}$ into $B^{(n)}(\mathcal{T}_n)$ and $B^{(n)}(\mathcal{R}_n)$. Then the squared Frobenius norm of $B^{(n)}$ may be bounded above:

$$\|B^{(n)}\|_F^2 = \|B^{(n)}(\mathcal{T}_n)\|_F^2 + \|B^{(n)}(\mathcal{R}_n)\|_F^2 \leq \frac{2m}{n} + \frac{1}{m^2}\frac{2(n-m)}{n}. \tag{12}$$

Choosing $m = O(\sqrt{n})$ shows that the norm of $B^{(n)}$ converges to 0 as $n \to \infty$, and so condition (11) is violated.

4.2 Identifiability

As noted in Remark 2.2, [6] employs a weaker notion of identifiability of Θ_k than the one we use in the present work. In particular, the definition from [6] permits Θ_k and $\widetilde{\Theta}_k$ to differ (modulo a global rescaling) on negligible subsets of $\mathbb{Z}_+$.

As noted in Remark 4.1, any finite set is negligible, and any infinite subset may be made negligible or non-negligible by reordering $\mathbb{Z}_+$. Consequently, the definition of identifiability employed in [6] depends on the ordering of $\mathbb{Z}_+$. By contrast, the stronger notion of identifiability of Θ_k employed in the present work does not depend on a specified ordering.

4.3 Condition (2)

Condition (2) from Theorem 3.1 may be easily verified for any specified matrix Q. A similar condition appears in Theorem 3.1 from [6], which we may state as follows:

$$\{k\} = \bigcap_{\substack{S\subset\{1,\ldots,K\} \\ k\in S \\ \mathcal{R}(S) \text{ non-negligible}}} S. \tag{13}$$

The right side of (13) is the intersection of all subsets $S \subset \{1, \ldots, K\}$ containing k where $\mathcal{R}(S)$ are *non-negligible*; by contrast, condition (2) from Theorem 3.1 is the intersection of all such S with $\mathcal{R}(S)$ that are *non-empty*. While the latter condition may be verified for finite-sized matrices M, the condition that $\mathcal{R}(S)$ is non-negligible is an asymptotic condition, which is not determinable for a finite sized matrix. Furthermore, as noted in Remark 4.1, it depends on the ordering of the indices in $\mathbb{Z}_+$. While conceptually similar to (13), the condition (2) given in Theorem 3.1 is more suitable in practical settings as it is well-defined non-asymptotically.

Acknowledgements

I am grateful to Xiaoou Li for discussing her work from [6], and to the reviewers for their helpful comments. I acknowledge support from NSF award IIS-1837992 and BSF award 2018230.

References

[1] Theodore W. Anderson and Herman Rubin. Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, pages 111–150, 1956.

[2] Theodore Wilbur Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, 1962.

[3] Jushan Bai and Kunpeng Li. Statistical analysis of factor models of high dimension. *Annals of Statistics*, 40(1):436–465, 2012.

[4] David J. Bartholomew, Martin Knott, and Irini Moustaki. *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley & Sons, 2011.

[5] Xin Bing, Florentina Bunea, Yang Ning, and Marten Wegkamp. Adaptive estimation in structured factor models with applications to overlapping clustering. *Annals of Statistics*, 2019.

[6] Yunxiao Chen, Xiaoou Li, and Siliang Zhang. Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association*, 2019.

[7] Edgar Dobriban, William Leeb, and Amit Singer. Optimal prediction in the linearly transformed spiked model. *Annals of Statistics*, 48(1):491–513, 2020.

[8] Edgar Dobriban and Art B. Owen. Deterministic parallel analysis: an improved method for selecting factors and principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(1):163–183, 2019.

[9] David L. Donoho, Matan Gavish, and Iain M Johnstone. Optimal shrinkage of eigenvalues in the spiked covariance model. *Annals of Statistics*, 46(6), 2018.

[10] David Hong, Laura Balzano, and Jeffrey A. Fessler. Asymptotic performance of PCA for high-dimensional heteroscedastic data. *Journal of Multivariate Analysis*, 167:435–452, 2018.

[11] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

[12] Ian Jolliffe. *Principal Component Analysis*. Wiley Online Library, 2002.

[13] Tjalling C. Koopmans and Olav Reiersøl. The identification of structural characteristics. *The Annals of Mathematical Statistics*, 21(2):165–181, 1950.

[14] William Leeb and Elad Romanov. Optimal spectral shrinkage and PCA with heteroscedastic noise. *IEEE Transactions on Information Theory*, available online, 2021.

[15] Olav Reiersøl. On the identifiability of parameters in Thurston's multiple factor analysis. *Psychometrika*, 15(2):121–149, 1950.

[16] Alexander Shapiro. Identifiability of factor analysis: Some results and open problems. *Linear Algebra and its Applications*, 70:1–7, 1985.

[17] James H. Stock and Mark W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002.