

A note on identifiability conditions in confirmatory factor analysis

William Leeb*

Abstract

Recently, Chen, Li and Zhang have established simple conditions characterizing when latent factors are asymptotically identifiable within a certain model of confirmatory factor analysis. In this note, we prove a related characterization of when factor loadings are identifiable, under slightly weaker modeling assumptions and a slightly stronger definition of identifiability. We also check that the previously-established characterization for latent factors still holds in our modified model.

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”. In typical statistical applications, M is the expected value of a randomly observed matrix with independent entries. For example, row i of M may be of the form $\Lambda \mathbf{z}_i + \boldsymbol{\varepsilon}_i$, where \mathbf{z}_i is a random vector in \mathbb{R}^K , and $\boldsymbol{\varepsilon}_i$ is a random noise vector with known covariance. For general background on factor analysis, we refer the reader to [2, 3, 6], 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 $A_{jk} = 0$. 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 [5] provides necessary and sufficient conditions on the matrix Q under which individual columns of Θ are asymptotically determined (up to rescaling), or *identifiable*, under certain assumptions on Θ and A . In this note, we provide an elementary proof of a similar characterization of the identifiability of A ’s columns, a question which has also attracted interest [8, 7, 1, 9, 4]. Because our model is slightly different from the one in [5], we also verify that their characterization of the identifiability of Θ ’s columns remains valid in our model.

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 relate the assumptions of our model to the model employed in [5].

2 Definitions and model description

For a fixed value of K , we assume we have bounded, real-valued functions $\Theta_1, \dots, \Theta_K$ and A_1, \dots, A_K defined on \mathbb{Z}_+ , the set of positive integers. We define the matrices $\Theta = [\Theta_1, \dots, \Theta_K] \in \mathbb{R}^{\mathbb{Z}_+ \times K}$ and $A = [A_1, \dots, A_K] \in \mathbb{R}^{\mathbb{Z}_+ \times K}$. We will let $M = \Theta A^T$; note that though M is infinite in size, each entry of M is an inner product between two K -dimensional vectors. Additionally, we assume access to a binary matrix $Q \in \{0, 1\}^{\mathbb{Z}_+ \times K}$ with columns Q_1, \dots, Q_K .

Before listing the precise conditions on the matrices Θ , A and Q , we introduce several definitions that we will use throughout the note.

Definition 2.1. A subset $\Delta \subset \mathbb{Z}_+$ is negligible if

$$\lim_{J \rightarrow \infty} \frac{|\Delta \cap \{1, \dots, J\}|}{J} = 0. \quad (1)$$

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

Remark 2.1. The definition of negligible depends crucially on applying the usual ordering of \mathbb{Z}_+ . Indeed, if Δ is any infinite subset of \mathbb{Z}_+ , we can always reorder \mathbb{Z}_+ so that $|\Delta \cap \{1, \dots, J\}|/J$ 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.

Remark 2.2. It is easy to see that if $\Delta \subset \mathbb{Z}_+$ is negligible, so too is any subset of Δ . Furthermore, the union of finitely many negligible sets is also negligible.

*School of Mathematics, University of Minnesota

Definition 2.2. For a subset $S \subset \{1, \dots, K\}$, 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 function $\mathbf{x} : \mathbb{Z}_+ \rightarrow \mathbb{R}$ and a subset $\mathcal{R} \subset \mathbb{Z}_+$, we will denote by $\mathbf{x}(\mathcal{R})$ the restriction of \mathbf{x} to \mathcal{R} . If in addition $S \subset \{1, \dots, K\}$, we will also denote by $A_{[\mathcal{R}, S]}$ the submatrix of A with rows in \mathcal{R} and columns from S .

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

Model Assumptions

1. The columns of Θ are linearly independent.
2. If $S \subset \{1, \dots, K\}$ and $\mathcal{R}(S)$ is non-negligible, then the columns of the submatrix $A_{[\mathcal{R}(S), S]}$ are linearly independent off of any negligible subset of $\mathcal{R}(S)$. That is, if $\mathcal{R}' \subset \mathcal{R}(S)$ and $\mathcal{R}(S) \setminus \mathcal{R}'$ is negligible, then $A_k(\mathcal{R}')$, $k \in S$, are linearly independent.
3. For any $k = 1, \dots, K$, $A_k(j) = 0$ whenever $Q_k(j) = 0$, except possibly on a negligible set of indices j . That is, $\text{supp}(A_k) \setminus \text{supp}(Q_k)$ is negligible.
4. There is a constant $C > 0$ such that

$$\sup_{\substack{1 \leq k \leq K \\ i \geq 1}} |\Theta_k(i)| < C, \quad \sup_{\substack{1 \leq k \leq K \\ j \geq 1}} |A_k(j)| < C. \quad (2)$$

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

Definition 2.3. The latent factor Θ_k is identifiable if for any decomposition $M = \tilde{\Theta} \tilde{A}^T$ satisfying assumptions 1 – 4, Θ_k and $\tilde{\Theta}_k$ are linearly dependent. Similarly, the factor loading A_k is identifiable if for any decomposition $M = \tilde{\Theta} \tilde{A}^T$ satisfying assumptions 1 – 4, A_k and \tilde{A}_k are linearly dependent.

Remark 2.3. [5] defines identifiability of Θ_k to mean that the angle between $\Theta_k(1 : N)$ and $\tilde{\Theta}_k(1 : N)$ converges to 0 as $N \rightarrow \infty$, which is a weaker notion than the one we employ. In particular, the definition from [5] permits Θ_k and $\tilde{\Theta}_k$ to differ (modulo a global rescaling) on negligible subsets of \mathbb{Z}_+ , which our definition prohibits.

Remark 2.4. Assumptions analogous to 1 and 2 are found in [5]. In Section 4, we will show that assumptions 1 and 2 are strictly weaker than those found in [5].

Remark 2.5. Assumption 3 permits the constraints defined by Q to be violated on a negligible set, whereas the corresponding assumption from [5] does not allow this.

Remark 2.6. Assumption 4 is slightly different than the boundedness assumption from [5]. 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 concept of *masking*, defined as follows.

Definition 2.4. We say k' masks k if $\text{supp}(Q_{k'}) \setminus \text{supp}(Q_k)$ is negligible.

In other words, k' masks k if the support of $Q_{k'}$ is contained in the support of Q_k , up to a negligible set.

3 Main results

We provide necessary and sufficient conditions on Q which characterize when Θ_k and A_k are identifiable. Theorem 3.1 addresses identifiability of Θ_k ; the identifiability condition is equivalent to the one in [5], and our proof amounts to a tightening of the proof of Proposition 8 in [5]. 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 other k' , or equivalently if

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

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

Technical lemmas

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

Proof. Let $\mathcal{T} = \text{supp}(Q_{k'}) \cap \text{supp}(Q_k)$. Then $\text{supp}(Q_{k'}) \setminus \mathcal{T}$ is negligible, and since $\text{supp}(Q_{k'})$ is not negligible, neither is \mathcal{T} . Any index $j \in \mathcal{T}$ is contained in $\mathcal{R}(S)$ for some $S \subset \{1, \dots, K\}$ with $k, k' \in S$. Since there are only finitely many possible S , some such $\mathcal{R}(S)$ must be non-negligible, since their union covers \mathcal{T} . 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 $\tilde{A} = [A_1, \dots, A_{k-1}, A_k + \epsilon A_{k'}, A_{k+1}, \dots, A_K]$ satisfies assumption 2.

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

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

Next, suppose $1 \notin S$. Then by definition $\text{supp}(Q_1)$ is disjoint from $\mathcal{R}(S)$, and hence also from the subset \mathcal{R}' . Since $\text{supp}(A_1) \setminus \text{supp}(Q_1)$ is negligible, $\Delta \equiv \mathcal{R}' \cap \text{supp}(A_1)$ is also negligible since it is a subset of $\text{supp}(A_1) \setminus \text{supp}(Q_1)$. If we define $\mathcal{R}'' \equiv \mathcal{R}' \setminus \text{supp}(A_1) = \mathcal{R}' \setminus \Delta$, then $\mathcal{R}(S) \setminus \mathcal{R}'' = (\mathcal{R}(S) \setminus \mathcal{R}') \cup \Delta$ is negligible, and $A_1(j) = 0$ for $j \in \mathcal{R}''$. Consequently, $A_K(\mathcal{R}'') = A_K(\mathcal{R}'') + \epsilon A_1(\mathcal{R}'')$, and since $A_k(\mathcal{R}'')$, $k \in S$, are linearly independent, the same is true after replacing $A_K(\mathcal{R}'')$ by $A_K(\mathcal{R}'') + \epsilon A_1(\mathcal{R}'')$. But since these vectors are linearly independent on \mathcal{R}'' , they are linearly independent on the larger set \mathcal{R}' . \square

Lemma 3.5. *Suppose k' masks $k \neq k'$, and let $\epsilon \in \mathbb{R}$. Then $\tilde{A} = [A_1, \dots, A_{k-1}, A_k + \epsilon A_{k'}, A_{k+1}, \dots, A_K]$ satisfies assumption 3.*

Proof. Without loss of generality, suppose $k' = K$ masks $k = 1$. The sets $\text{supp}(A_1) \setminus \text{supp}(Q_1)$, $\text{supp}(A_K) \setminus \text{supp}(Q_K)$, and $\text{supp}(Q_K) \setminus \text{supp}(Q_1)$ are all negligible. We have

$$[\text{supp}(A_1) \cup \text{supp}(A_K)] \setminus \text{supp}(Q_1) = [\text{supp}(A_1) \setminus \text{supp}(Q_1)] \cup [\text{supp}(A_K) \setminus \text{supp}(Q_1)]. \quad (4)$$

Since $\text{supp}(A_K) \setminus \text{supp}(Q_1)$ is the union of $\text{supp}(A_K) \cap [\text{supp}(Q_K) \setminus \text{supp}(Q_1)]$ and $[\text{supp}(A_K) \setminus \text{supp}(Q_K)] \setminus \text{supp}(Q_1)$, both of which are negligible, it follows that (4) is negligible. Since $\text{supp}(A_1 + \epsilon A_K) \subset \text{supp}(A_1) \cup \text{supp}(A_K)$, we are done. \square

The next two lemmas relate the concept of masking to the identifiability condition from [5].

Lemma 3.6. *Suppose k does not mask any $k' \neq k$. Then*

$$\{k\} = \bigcap_{\substack{S \subset \{1, \dots, K\} \\ k \in S \\ \mathcal{R}(S) \text{ non-negligible}}} S. \quad (5)$$

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-negligible. Indeed, $\text{supp}(Q_k)$ must be non-negligible, since otherwise k would mask every k' . But each $j \in \text{supp}(Q_k)$ is contained in $\mathcal{R}(S)$ for some subset $S \subset \{1, \dots, K\}$ with $k \in S$; and since there are only finitely many such subsets S , at least one such $\mathcal{R}(S)$ must be non-negligible. Consequently, the right side of (5) is non-empty, and obviously contains k .

To show the reverse inclusion, take any $k' \neq k$. Since k does not mask k' , $\text{supp}(Q_k) \setminus \text{supp}(Q_{k'})$ is non-negligible. Each $j \in \text{supp}(Q_k) \setminus \text{supp}(Q_{k'})$ is contained in some $\mathcal{R}(S)$, where $S \subset \{1, \dots, K\}$ contains k but not k' . Since there are only finitely many such S , there must exist some $S \subset \{1, \dots, K\}$ which contains k but not k' and for which $\mathcal{R}(S)$ is non-negligible, implying that k' is not contained in the right side of (5). \square

The converse to Lemma 3.6 is also true:

Lemma 3.7. *Suppose (5) holds. Then k does not mask any $k' \neq k$.*

Proof. Without loss of generality, suppose $k = K$. If $\text{supp}(Q_K)$ were negligible, then for any $S \subset \{1, \dots, K\}$ containing K , $\mathcal{R}(S) \subset \text{supp}(Q_K)$ would also be negligible, and the right side of (5) would be empty; a contradiction. Consequently, $\text{supp}(Q_K)$ must not be negligible.

For contradiction, suppose without loss of generality that K masks 1. Take any $S \subset \{1, \dots, K\}$ containing K but not 1. Since $\mathcal{R}(S) \subset \text{supp}(Q_K) \setminus \text{supp}(Q_1)$, $\mathcal{R}(S)$ is negligible, and so S is not included in the right side of (5). Therefore, the only S included on the right side of (5) contain both K and 1. But then 1 is also in the intersection, a contradiction. \square

Remark 3.1. Lemmas 3.6 and 3.7 together show that the identifiability condition of k not masking any other $k' \neq k$ is identical to the identifiability condition (3) from Theorem 1 in [5].

We now turn to the proofs of Theorems 3.1 and 3.2.

Proof of Theorem 3.1

As noted, this proof is an adaptation of the proof of Proposition 8 in [5] to our model. First, suppose, without loss of generality, that K masks 1. We write:

$$M = \Theta_1 A_1^T + \Theta_2 A_2^T + \dots + \Theta_K A_K^T = \Theta_1 (A_1 + \epsilon A_K)^T + \Theta_2 A_2^T + \dots + (\Theta_K - \epsilon \Theta_1) A_K^T, \quad (6)$$

where ϵ is sufficiently small so as to not violate assumption 4. From Lemmas 3.4 and 3.5, 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, suppose component K does not mask any other component $k \neq K$. Then $\text{supp}(Q_K)$ is non-negligible, since otherwise it would mask every k . Each $j \in \text{supp}(Q_K)$ is contained in $\mathcal{R}(S)$ for some $S \subset \{1, \dots, K\}$ with $K \in S$; since there are only finitely many such S , at least one such $\mathcal{R}(S)$ must be non-negligible.

Suppose $M = \tilde{\Theta}\tilde{A}^T$ is another factorization of M satisfying the model assumptions 1 – 4. We will show that Θ_K and $\tilde{\Theta}_K$ are linearly dependent. For $1 \leq k \leq K$, define $\Delta_k = \text{supp}(A_k) \setminus \text{supp}(Q_k)$ and $\tilde{\Delta}_k = \text{supp}(\tilde{A}_k) \setminus \text{supp}(Q_k)$. Then all the sets Δ_k and $\tilde{\Delta}_k$ are negligible, and hence so is their union $\Delta \equiv \Delta_1 \cup \dots \cup \Delta_K \cup \tilde{\Delta}_1 \cup \dots \cup \tilde{\Delta}_K$.

Take any S containing K with $\mathcal{R}(S)$ non-negligible. Define $\mathcal{R}'_S \equiv \mathcal{R}(S) \setminus \Delta$. 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

$$M_{[:,\mathcal{R}'_S]} = \Theta_{[:,S]}(A_{[\mathcal{R}'_S,S]})^T. \quad (7)$$

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 $V_S \equiv \text{span}\{M_j : j \in \mathcal{R}(S) \setminus \Delta\}$, then $V_S = \text{span}\{\Theta_k : k \in S\}$ and $\dim(V_S) = |S|$.

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

$$\Theta_K \in V_{S_K} = \bigcap_{\substack{S \subset \{1, \dots, K\} \\ K \in S \\ \mathcal{R}(S) \text{ non-negligible}}} V_S \quad (8)$$

where S_K is the intersection of all sets S with $K \in S$ and $\mathcal{R}(S)$ non-negligible. But because K does not mask any $k \neq K$, Lemma 3.6 implies that $S_K = \{K\}$, and so $V_{S_K} = \text{span}\{\Theta_K\}$. But the exact same argument with $\tilde{\Theta}$ and \tilde{A} in place of Θ and A also shows $V_{S_K} = \text{span}\{\tilde{\Theta}_K\}$. Consequently, Θ_K and $\tilde{\Theta}_K$ are linearly dependent.

Proof of Theorem 3.2

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

$$M = \Theta_1 A_1^T + \Theta_2 A_2^T + \dots + \Theta_K A_K^T = (\Theta_1 - \epsilon \Theta_K) A_1^T + \Theta_2 A_2^T + \dots + \Theta_K (A_K + \epsilon A_1)^T, \quad (9)$$

where ϵ is sufficiently small so as to not violate assumption 4. From Lemmas 3.4 and 3.5, assumptions 2 and 3 and are still satisfied by $A_1, A_2, \dots, A_K + \epsilon A_1$. 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 $M = \tilde{\Theta}\tilde{A}^T$ is another factorization within the same model, and that \tilde{A}_K and A_K are linearly independent. Define $\tilde{\Delta}_k = \text{supp}(\tilde{A}_k) \setminus \text{supp}(Q_k)$, $1 \leq k \leq K$, and $\tilde{\Delta}_K \equiv \text{supp}(\tilde{A}_K) \setminus \text{supp}(Q_K)$. From assumption 3, the sets $\Delta_1, \dots, \Delta_K$ and $\tilde{\Delta}_K$ are negligible, and so too is their union $\Delta \equiv \tilde{\Delta}_K \cup \Delta_1 \cup \dots \cup \Delta_K$. Furthermore, if $\mathcal{R}_K \equiv \text{supp}(Q_K)^c \setminus \Delta$ is the set of roots of Q_K minus Δ , then A_K and \tilde{A}_K are both zero on \mathcal{R}_K .

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

$$\sum_{k=1}^{K-1} c_k A_k(\mathcal{R}_K) = \mathbf{0}. \quad (10)$$

Suppose, without loss of generality, that $c_1 \neq 0$. We will show that 1 masks K . Suppose not; then $\mathcal{T} \equiv \text{supp}(Q_1) \cap \mathcal{R}_K$ is non-negligible. Indeed, we have $\mathcal{T} = [\text{supp}(Q_1) \setminus \text{supp}(Q_K)] \setminus \Delta$, where $\text{supp}(Q_1) \setminus \text{supp}(Q_K)$ is non-negligible and Δ is negligible.

Each $j \in \mathcal{T}$ is contained in $\mathcal{R}(S)$ for some $S \subset \{1, \dots, K-1\}$ with $1 \in S$. Since there are finitely many such subsets S , some such set $\mathcal{R}(S)$ must be non-negligible, since their union covers the non-negligible set \mathcal{T} . By definition, $\mathcal{R}(S)$ is a subset of the set of roots of Q_K ; hence $\mathcal{R}' \equiv \mathcal{R}(S) \setminus \Delta$ is a subset of \mathcal{R}_K .

Since \mathcal{R}' excludes Δ , if $k \notin S$ and $j \in \mathcal{R}'$ then $A_k(j) = 0$. Hence from (10)

$$\sum_{k \in S} c_k A_k(\mathcal{R}') = \mathbf{0}. \quad (11)$$

But by assumption 2, the columns of $A_{[\mathcal{R}',S]}$ 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 Remarks on the model assumptions

We conclude with several additional remarks on the model assumptions, and their relationship to the model from [5].

4.1 Assumptions 1 and 2

In [5], assumption 1 is replaced by the assumption that the minimum singular value of the matrix $\Theta_{[1:n,1:K]}/\sqrt{n}$ converges to a positive limit as $n \rightarrow \infty$; and an analogous assumption is made in place of assumption 2. We show that the assumptions in [5] imply assumptions 1 and 2.

Proposition 4.1. *Suppose all entries of $B \in \mathbb{R}^{z+ \times K}$ are bounded by $C > 0$, and*

$$\lim_{n \rightarrow \infty} \frac{\sigma_K(B_{[1:n,1:K]})}{\sqrt{n}} > 0, \quad (12)$$

where σ_K denotes the K^{th} singular value. Then the columns of B are linearly independent off of any negligible set.

Proof. Suppose for contradiction that there is a negligible subset $\Delta \subset \mathbb{Z}_+$ so that $B_1(\mathcal{R}), \dots, B_K(\mathcal{R})$ are linearly dependent, where $\mathcal{R} = \mathbb{Z}_+ \setminus \Delta$.

Define $B^{(n)} = B_{[1:n, 1:K]}/\sqrt{n}$, $\mathcal{R}_n = \mathcal{R} \cap \{1, \dots, n\}$, and $\Delta_n = \Delta \cap \{1, \dots, n\}$. We can partition $B^{(n)}$ into $B^{(n)}(\mathcal{R}_n)$ and $B^{(n)}(\Delta_n)$, the restrictions to those two sets of rows. Since the columns of $B^{(n)}(\mathcal{R}_n)$ are linearly dependent, there is a unit vector $\mathbf{x} \in \mathbb{R}^K$ so that $B^{(n)}(\mathcal{R}_n)\mathbf{x} = \mathbf{0}$. Then the norm squared of $B^{(n)}\mathbf{x}$ is

$$\|B^{(n)}\mathbf{x}\|^2 = \|B^{(n)}(\Delta_n)\mathbf{x}\|^2 \leq \|B^{(n)}(\Delta_n)\|_F^2 \leq C^2 K \frac{|\Delta_n|}{n}, \quad (13)$$

which converges to 0 because Δ is negligible. Consequently, the smallest singular value of $B^{(n)}$ becomes arbitrarily small as $n \rightarrow \infty$, contradicting (12). \square

Remark 4.1. It is not difficult to see that the converse to Proposition 4.1 is false. For example, we may take $\text{supp}(B_1)$ to be the even integers, and $\text{supp}(B_2)$ to be the 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 off of any negligible set.

Take any large n and $m < n$. Define $\mathcal{T}_n = \{1, \dots, m\}$ and $\mathcal{R}_n = \{m+1, \dots, n\}$, and partition $B^{(n)} \equiv B_{[1:n, 1:2]}/\sqrt{n}$ into $B^{(n)}(\Delta_n)$ and $B^{(n)}(\mathcal{R}_n)$. Then the Frobenius norm squared of $B^{(n)}$ may be bounded above:

$$\|B^{(n)}\|_F^2 = \|B^{(n)}(\mathcal{T}_n)\|_F^2 + \|B^{(n)}(\mathcal{R}_n)\|_F^2 \leq \frac{2m}{n} + \frac{1}{m^2} \frac{2(n-m)}{n}. \quad (14)$$

Choosing $m = O(\sqrt{n})$ shows that the norm of $B^{(n)}$ converges to 0 as $n \rightarrow \infty$, and so condition (12) is violated.

In other words, assumptions 1 and 2 are strictly weaker than the corresponding assumptions from [5].

4.2 Assumption 3

The model in [5] requires that if $Q_k(j) = 0$, then $A_k(j) = 0$ as well. In our model, specifically assumption 3, we permit this constraint to be violated on a negligible set of indices j . This is a necessary relaxation in order for Theorems 3.1 and 3.2 to be true, due to our using a stronger notion of identifiability than in [5] (see Remark 2.3).

To see this, consider a model with $K = 2$. Suppose $\text{supp}(A_1) = \text{supp}(Q_1) = 2\mathbb{Z}_+$, the even integers; and $\text{supp}(A_2) = \text{supp}(Q_2) = 4\mathbb{Z}_+ \cup \{1\}$. Then 2 masks 1, since $\text{supp}(Q_2) \setminus \text{supp}(Q_1) = \{1\}$, which is negligible. However, if $M = \tilde{\Theta}\tilde{A}^T$, we would have $\tilde{A}_1 = \alpha A_1 + \beta A_2$, for some $\alpha, \beta \in \mathbb{R}$; and so $\tilde{A}_1(1) = \beta A_2(1)$. Consequently, if our model required that $\tilde{A}_k(j) = 0$ for *all* j with $Q_k(j) = 0$, then A would be identifiable (as would Θ), even though 2 masks 1. In order for Theorems 3.1 and 3.2 to be correct under our stronger notion of identifiability, we must therefore relax the constraints imposed by the model.

Acknowledgements

I am grateful to Xiaou Li for discussing her work from [5]. 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] David J. Bartholomew, Martin Knott, and Iriini Moustaki. *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley & Sons, 2011.
- [4] Xin Bing, Florentina Bunea, Yang Ning, and Marten Wegkamp. Adaptive estimation in structured factor models with applications to overlapping clustering. *Annals of Statistics*, 2019.
- [5] Yunxiao Chen, Xiaou Li, and Siliang Zhang. Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association*, 2019.
- [6] Ian Jolliffe. *Principal Component Analysis*. Wiley Online Library, 2002.
- [7] Tjalling C. Koopmans and Olav Reiersøl. The identification of structural characteristics. *The Annals of Mathematical Statistics*, 21(2):165–181, 1950.
- [8] Olav Reiersøl. On the identifiability of parameters in Thurston’s multiple factor analysis. *Psychometrika*, 15(2):121–149, 1950.
- [9] Alexander Shapiro. Identifiability of factor analysis: Some results and open problems. *Linear Algebra and its Applications*, 70:1–7, 1985.