Identifiability in Multi-Channel Factor Analysis

Gray Stanton
Statistics
Colorado State University
Fort Collins, USA
gstanton@colostate.edu

David Ramírez
Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid
Madrid, Spain
david.ramirez@uc3m.es

Ignacio Santamaria
Ingeniería de Comunicaciones
Universidad de Cantabria
Santander, Spain
i.santamaria@unican.es

Louis L. Scharf
Mathematics
Colorado State University
Fort Collins, USA
louis.scharf@colostate.edu

Haonan Wang
Statistics
Colorado State University
Fort Collins, USA
wanghn@stat.colostate.edu

***Abstract*—The recently developed Multi-Channel Factor Analysis (MFA) is a method for extracting a latent low-dimensional signal that is present across multiple channels and corrupted by unobserved single-channel interference and idiosyncratic noise. In MFA, only the channel structure and dimensionality of the signal and interference subspaces are specified in advance, which raises the concern that the signal, interference, and noise covariances may not be uniquely determined by the observation model. This paper presents necessary and sufficient conditions on the channel sizes and subspace dimensions to guarantee the *identifiability* of MFA, ensuring that the second-order spatial properties of the latent components can, in principle, be recovered from the multi-channel observations.**

***Index Terms*—factor analysis (FA), identifiability, multi-channel factor analysis (MFA)**

I. Introduction

The aim of Factor Analysis (FA) is to obtain a parsimonious model for the second-order properties of a multivariate observation. The observation is represented as the sum of a *signal*, which lives in an unknown low-dimensional subspace, and idiosyncratic noise. Classical or *single-channel* exploratory Factor Analysis (FA) was originally developed in the field of psychometrics [1], and is now a widely used technique in multivariate data analysis. In array processing, FA is frequently used to analyze the output of uncalibrated systems where the noise variance is anisotropic and unknown [2]–[4].

Of central import to this paper, the recently developed method of *Multi-Channel Factor Analysis* (MFA) [5] extends FA to the *multi-channel* and multi-sensor setting, to allow for the presence of channel-specific *interference*. Similarly to the signal, the channel-specific interferences also live in unknown low-dimensional subspaces, but those subspaces are constrained to lie within the observation spaces for the individual channels. In this fashion, the signal, which has multi-channel effects, can be distinguished from interference, which separately affects distinct channels, as well as idiosyncratic noise. MFA has significant utility for detection and estimation of a weak signal which presents across multiple channels in the presence of channel-specific interference, which is a problem encountered in array processing [6]–[8].

However, to ensure that the second-order model obtained from MFA can be meaningfully interpreted, the decomposition into signal, interference, and noise must be *unique*. Further, it should be possible to give guarantees for the uniqueness of the MFA decomposition using only the channel structure of the observations and the dimensionality of the signal and interference vectors. The problem of the *identifiability of MFA* is that of obtaining such guarantees.

This work presents conditions on the channel sizes and factor numbers which certify the identifiability of MFA, thus allowing for application of MFA to multi-channel signal processing problems without concerns of non-interpretability.

A. Notation

Matrices and vectors are denoted by bold-face uppercase and lowercase symbols respectively, while scalars are denoted with light-face symbols. The zero matrix of size $n \times m$ is $\mathbf{0}_{n,m}$, while the zero vector of size n is $\mathbf{0}_n$. The identity matrix of size $n \times n$ is $\mathbf{I}_n$. For the matrix $\mathbf{D}$, $\mathbf{D}^\mathsf{T}$ is its transpose. The space of real diagonal matrices of size $n \times n$ is $\mathrm{Diag}(n)$, with $\mathrm{Diag}_{\geq 0}(n)$ being the subset with non-negative diagonal entries. The set of positive semidefinite matrices of size n is $\mathrm{PSD}(n)$. The operator $\mathrm{blkdiag}$ applied to a list of matrices constructs the block-diagonal matrix with said matrices as the on-diagonal blocks. The expectation of a random quantity $\mathbf{s}$ is $E[\mathbf{s}]$. The non-negative part of a scalar $a \in \mathbb{R}$ is $(a)_+ \equiv \max\{a, 0\}$.

II. Model

Consider an array of C potentially heterogeneous sensors, each of which produces a real vector-valued observation. The cth sensor provides the observation vector $\mathbf{x}_c \in \mathbb{R}^{n_c}$, which is composed of n_c scalar measurements. To jointly analyze the multi-sensor data, the observation vectors $\mathbf{x}_1, \ldots, \mathbf{x}_C$ are taken to belong to distinct *channels* sensing a common source.

A. Observation Model

To model the channel-c observation $\mathbf{x}_c$, three latent vectors are defined, namely the *signal* $\mathbf{s}_c$, *interference* $\mathbf{i}_c$, and *noise*

$\mathbf{u}_c$, which are specific to channel c and are of size n_c. The channel-c observation is then synthesized as the sum,

$$\mathbf{x}_c = \mathbf{s}_c + \mathbf{i}_c + \mathbf{u}_c. \tag{1}$$

The signal $\mathbf{s}_c$ is as the expression of the shared source within channel c, while $\mathbf{i}_c$ is channel-c-specific influences which are not confined to individual scalar inputs in that channel. The idiosyncratic noise $\mathbf{u}_c$ is then the remaining influences which are particular to individual scalar measurements.

To allow for this parsing of the C observations, the observations for the individual channels are stacked vertically into the all-channel observation as $\mathbf{x} \equiv [\mathbf{x}_1^\mathsf{T} \ \dots \ \mathbf{x}_C^\mathsf{T}]^\mathsf{T}$. The all-channel signal, interference, and noise vectors $\mathbf{s}$, $\mathbf{i}$, and $\mathbf{u}$ are similarly obtained by stacking. The total number of scalar inputs is $n \equiv \sum_{c=1}^{C} n_c$, and the channel sizes are combined into the length-C integer vector $\mathbf{n} \equiv [n_1, \dots, n_c]^\mathsf{T}$. So, $\mathbf{x}, \mathbf{s}, \mathbf{i}$ and $\mathbf{u}$ are vectors in $\mathbb{R}^n$.

The signal $\mathbf{s}$ is assumed to lie within some subspace of $\mathbb{R}^n$, of which only the dimension $r_0 \leq n_c$ is known. Therefore, $\mathbf{s}$ can be written as $\mathbf{s} = \mathbf{A}\mathbf{f}$ for some $\mathbf{A} \in \mathbb{R}^{n \times r_0}$ and $\mathbf{f} \in \mathbb{R}^{r_0}$, where the range of $\mathbf{A}$ is the signal subspace and $\mathbf{f}$ is a vector of latent factors which are *common* across channels. The $n \times r_0$ *common factor loading matrix* $\mathbf{A}$ can be written as $\mathbf{A} = [\mathbf{A}_1^\mathsf{T} \ \dots \ \mathbf{A}_C^\mathsf{T}]^\mathsf{T}$, where the $n_c \times r_0$ submatrix $\mathbf{A}_c$ contains the rows of $\mathbf{A}$ corresponding to $\mathbf{x}_c$ within the stacked $\mathbf{x}$. So, the channel-c signal $\mathbf{s}_c = \mathbf{A}_c\mathbf{f}$ is then the sensing of the shared input $\mathbf{f}$ in channel c, with $\mathbf{A}_c$ controlling how $\mathbf{f}$ is sensed.

In contrast, the channel-c interference component $\mathbf{i}_c$ is assumed to lie within an unknown dimension-r_c subspace of the channel-c observation space. That is, $\mathbf{i}_c$ can be written as $\mathbf{i}_c = \mathbf{B}_c\mathbf{g}_c$ for some $\mathbf{B}_c \in \mathbb{R}^{n_c \times r_c}$ whose range is the interference subspace in channel c and a vector of latent factors $\mathbf{g}_c$ which is *unique* to channel c. The total number of unique factors across all channels is $r \equiv \sum_{c=1}^{C} r_c$. The all-channel *unique factor loading matrix* $\mathbf{B}$ is obtained by *diagonally* stacking the matrices into $\mathbf{B} \equiv \mathrm{blkdiag}(\mathbf{B}_1, \dots, \mathbf{B}_C)$, which is of size $n \times r$. The unique factors are vertically stacked into the length-r vector $\mathbf{g} \equiv [\mathbf{g}_1^\mathsf{T}, \dots, \mathbf{g}_c^\mathsf{T}]^\mathsf{T}$.

With the above definitions, the first-order model for the all-channel observations is

$$\mathbf{x} = \mathbf{A}\mathbf{f} + \mathbf{B}\mathbf{g} + \mathbf{u}. \tag{2}$$

In addition to the channel structure and sizes of the observed data, a key presupposition of MFA is the common factor number r_0 and unique factor numbers $r_1, \dots, r_C$. For simplicity of notation, the factor number are combined into the integer vector $\mathbf{r} \equiv [r_0, r_1, \dots, r_C]^\mathsf{T}$ of length $C + 1$.

B. Covariance Specification

The observation vector $\mathbf{x}$ and the latent vectors $\mathbf{s}, \mathbf{i}$ and $\mathbf{u}$ are taken to be *random* quantities, whose second moments are of interest. In the expressions of the signal and interference in terms of the common and unique factors, $\mathbf{s} = \mathbf{A}\mathbf{f}$ and $\mathbf{i} = \mathbf{B}\mathbf{g}_c$, the loading matrices $\mathbf{A}$ and $\mathbf{B}$ are fixed unknown parameters while the factors $\mathbf{f}$ and $\mathbf{g}$ are random unobserved vectors.

As the focus of MFA is the second-order properties of the multi-channel data, all random quantities are assumed to have mean zero. The second moments of $\mathbf{s}$ and $\mathbf{i}$ are respectively

$$\mathbf{R_{ss}} \equiv \mathbf{A}\mathbf{R_{ff}}\mathbf{A}^\mathsf{T} \quad \text{and} \quad \mathbf{R_{ii}} \equiv \mathbf{B}\mathbf{R_{gg}}\mathbf{B}^\mathsf{T},$$

where $\mathbf{R_{ff}} \equiv E[\mathbf{f}\mathbf{f}^\mathsf{T}]$ and $\mathbf{R_{gg}} \equiv E[\mathbf{g}\mathbf{g}^\mathsf{T}]$. To allow for the desired interpretations of $\mathbf{s}$ and $\mathbf{i}$, factors of different types are assumed to be *uncorrelated*. That is,

$$E[\mathbf{f}\mathbf{g}^\mathsf{T}] = \mathbf{0}_{r_0,r} \quad \text{and} \quad E[\mathbf{g}_c\mathbf{g}_{c'}^\mathsf{T}] = \mathbf{0}_{r_c,r_{c'}} \ \ c \neq c'.$$

This assumption and the structure of $\mathbf{B}$ ensure that $\mathbf{R_{ii}}$ is *block-diagonal*, with C blocks of sizes $n_1 \times n_1$ through $n_C \times n_C$.

For the all-channel noise vector $\mathbf{u}$, the noise components corresponding to distinct scalar inputs are assumed to be uncorrelated but the variances are unconstrained, so

$$\boldsymbol{\Phi} \equiv E[\mathbf{u}\mathbf{u}^\mathsf{T}]$$

is a diagonal covariance matrix. The noise $\mathbf{u}$ is further assumed to be uncorrelated with the latent factors,

$$E[\mathbf{u}\mathbf{f}^\mathsf{T}] = \mathbf{0}_{n,r_0} \quad \text{and} \quad E[\mathbf{u}\mathbf{g}_c^\mathsf{T}] = \mathbf{0}_{n,r_c}, \ \ c = 1, \dots, C.$$

With these specifications on the moments of the latent vectors, the covariance of the all-channel observation is

$$\begin{aligned} \mathbf{R_{xx}} &\equiv \mathbf{R_{ss}} + \mathbf{R_{ii}} + \boldsymbol{\Phi} \\ &= \mathbf{A}\mathbf{R_{ff}}\mathbf{A}^\mathsf{T} + \mathbf{B}\mathbf{R_{gg}}\mathbf{B}^\mathsf{T} + \boldsymbol{\Phi}. \end{aligned} \tag{3}$$

Estimation of the three covariance components enables subsequent analyses of practical interest, such as detecting the existence of a cross-channel signal and predicting the latent vectors $\mathbf{s}, \mathbf{i}$ and $\mathbf{u}$ from the observed $\mathbf{x}$.

In the above description of how $\mathbf{x}$ is synthesized from latent factors $\mathbf{f}$ and $\mathbf{g}$ and the noise $\mathbf{u}$, the common factor covariance $\mathbf{R_{ff}}$ is unconstrained while the unique factor covariance $\mathbf{R_{ii}}$ is block-diagonal but otherwise unconstrained. However, without further information about either the loading matrices $\mathbf{A}, \mathbf{B}$ or the factor covariances $\mathbf{R_{ff}}, \mathbf{R_{gg}}$, the pairs $(\mathbf{A}, \mathbf{R_{ff}})$ and $(\mathbf{B}, \mathbf{R_{gg}})$ are non-identifiable using knowledge of $\mathbf{x}$ alone. This follows as any change of basis on the factor spaces which takes $(\mathbf{A}, \mathbf{f})$ to $(\mathbf{A}\mathbf{T}_0, \mathbf{T}_0^{-1}\mathbf{f})$ and $(\mathbf{B}_c, \mathbf{g}_c)$ to $(\mathbf{B}_c\mathbf{T}_c, \mathbf{T}_c^{-1}\mathbf{g}_c)$ leaves $\mathbf{s}$ and $\mathbf{i_c}$ unchanged. In this paper, this indeterminacy is resolved by requiring that the factors $\mathbf{f}$ and $\mathbf{g}$ be unit-scale and uncorrelated, $\mathbf{R_f} = \mathbf{I}_{r_0}$ and $\mathbf{R_{gg}} = \mathbf{I}_r$. Alternative normalizations, such as taking $\mathbf{A}, \mathbf{B}$ to have unit norm columns and $\mathbf{R_{ff}}$ and $\mathbf{R_{ii}}$ to be diagonal with non-increasing diagonal elements, are useful for some analyses.

III. Identifiability of MFA

Without further assumptions on the unobserved signal, interference, and noise, the only information that the observations contain about the statistical properties of the latent vectors is in $\mathbf{R_{xx}}$. If a distinct triple $(\mathbf{R}'_{\mathbf{ss}}, \mathbf{R}'_{\mathbf{ii}}, \boldsymbol{\Phi}')$ of matrices structured as in Section II sums to the same observation covariance $\mathbf{R_{xx}}$, then MFA can yield no meaningful conclusions about $\mathbf{s}$, $\mathbf{i}$, and $\mathbf{u}$. If only one such triple of appropriately structured matrices sums to $\mathbf{R_{xx}}$, then the MFA decomposition of $\mathbf{R_{xx}}$ is *identified* for the specified channel sizes and factor numbers.

Identification of $\mathbf{R_{xx}}$ requires two subproblems to be solvable. First, it must be possible to *uniquely isolate* the noise variance $\boldsymbol{\Phi}$ from the combined signal-and-interference covariance $\mathbf{R_{ss}}+\mathbf{R_{ii}}$. Second, it must be possible to *uniquely separate* $\mathbf{R_{ss}}+\mathbf{R_{ii}}$, which is the noise-free part of the $\mathbf{R_{xx}}$, into the signal covariance $\mathbf{R_{ss}}$ and the interference covariance $\mathbf{R_{ii}}$. The observation covariance $\mathbf{R_{xx}}$ is identified if and only if both subproblems have unique solutions.

However, as $\mathbf{R_{xx}}$ is not known *a priori*, it is instead necessary to determine conditions on the channel sizes $\mathbf{n}$ and factor numbers $\mathbf{r}$ which can guarantee that *almost all* $\mathbf{R_{xx}}$ will admit a unique MFA decomposition. MFA for those channel sizes and factor numbers is then generically *identifiable*. Proofs of the propositions presented here are provided in the journal version of this paper.

Requiring that these subproblems be solvable for *all* possible $\mathbf{R_{xx}}$ is too restrictive, as degenerate cases will always exist. As a simple example, if $\mathbf{A},\mathbf{B}$ are such that $\mathbf{AA}^\mathsf{T}+\mathbf{BB}^\mathsf{T}$ is diagonal, it is impossible to isolate $\boldsymbol{\Phi}$ from $\mathbf{AA}^\mathsf{T}+\mathbf{BB}^\mathsf{T}$. Instead of precisely determining all ways in which identification can break down, we instead find conditions on $\mathbf{n}$ and $\mathbf{r}$ that guarantee that the non-identified $\mathbf{R_{xx}}$ make up a null set and so are ignorable for practical purposes. This type of approach is called *generic identifiability*, and it is used for single-channel FA [9], [10], as well as for low-rank matrix completion [11]. In this paper, a subset of a d-dimensional real vector space is *null* if its image under a linear isomorphism to $\mathbb{R}^d$ has Lebesgue measure zero. A statement is *generically* true if it true for all elements excepting a null subset.

A. Separation of Signal and Interference

For the MFA decomposition to be identified, the noise-free part of the observation covariance, $\mathbf{R_{ss}}+\mathbf{R_{ii}}$, must be uniquely separable into the signal covariance $\mathbf{R_{ss}}$ and the interference covariance $\mathbf{R_{ii}}$. This problem does not arise in exploratory single-channel factor analysis. Although this unique separation problem does not explicitly depend on the loading matrices $\mathbf{A}$ and $\mathbf{B}$, it is useful to frame the problem in terms of the combined loading matrix $[\mathbf{A}\ \mathbf{B}]$, which is of size $n\times(r_0+r)$ and is obtained by horizontally concatenating $\mathbf{A}$ and $\mathbf{B}$. In this setting, unique separability is equivalent to whether

$$[\mathbf{A}\ \mathbf{B}][\mathbf{A}\ \mathbf{B}]^\mathsf{T}=[\tilde{\mathbf{A}}\ \tilde{\mathbf{B}}][\tilde{\mathbf{A}}\ \tilde{\mathbf{B}}]^\mathsf{T} \tag{4}$$

implies that

$$\tilde{\mathbf{A}}=\mathbf{A}\mathbf{Q}_{00},\ \text{and}\ \tilde{\mathbf{B}}_c=\mathbf{B}_c\mathbf{Q}_{cc},\ c=1,\ldots,C, \tag{5}$$

for all $\mathbf{A},\tilde{\mathbf{A}}\in\mathbb{R}^{n\times r_0}$ and $\mathbf{B},\tilde{\mathbf{B}}$ being channel-structured block diagonal as described in Section II, where $\mathbf{Q}_{00}\in\mathbb{R}^{r_0\times r_0}$ and $\mathbf{Q}_{cc}\in\mathbb{R}^{r_c\times r_c}$ are orthogonal matrices of the appropriate sizes.

A typical result in factor analysis (see, e.g., [12]) implies that, for any equally sized real matrices $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{XX}^\mathsf{T}$ equals $\mathbf{YY}^\mathsf{T}$ if and only if $\mathbf{Y}=\mathbf{XQ}$ for some orthogonal $\mathbf{Q}$ of the appropriate size. This is easily seen as $\mathbf{X}$ and $\mathbf{Y}$ share singular values and left singular vectors. Application of this result to $[\mathbf{A}\ \mathbf{B}]$ and $[\tilde{\mathbf{A}}\ \tilde{\mathbf{B}}]$ implies that $[\tilde{\mathbf{A}}\ \tilde{\mathbf{B}}]=[\mathbf{A}\ \mathbf{B}]\mathbf{Q}$, where $\mathbf{Q}$ is an orthogonal matrix which is patterned as

$$\mathbf{Q}=\begin{array}{c} \begin{array}{cccc} r_0 & r_1 & \cdots\cdots & r_C \end{array} \\ \begin{bmatrix} \mathbf{Q}_{00} & \mathbf{Q}_{01} & \cdots & \mathbf{Q}_{0C} \\ \mathbf{Q}_{10} & \mathbf{Q}_{11} & \cdots & \mathbf{Q}_{1C} \\ \vdots & \vdots & & \vdots \\ \mathbf{Q}_{C0} & \mathbf{Q}_{C1} & \cdots & \mathbf{Q}_{CC} \end{bmatrix} \end{array} \begin{array}{c} r_0 \\ r_1 \\ \vdots \\ r_C \end{array}, \tag{6}$$

noting that blocks $\mathbf{Q}_{ij}$ are not themselves orthogonal. Taking $\mathbf{R_{ss}}=\mathbf{AA}^\mathsf{T}$, $\mathbf{R_{ii}}=\mathbf{BB}^\mathsf{T}$ and $\tilde{\mathbf{R}}_\mathbf{ss}=\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\mathsf{T}$, $\tilde{\mathbf{R}}_\mathbf{ii}=\tilde{\mathbf{B}}\tilde{\mathbf{B}}^\mathsf{T}$, (4) implies that $\mathbf{R_{ss}}+\mathbf{R_{ii}}=\tilde{\mathbf{R}}_\mathbf{ss}+\tilde{\mathbf{R}}_\mathbf{ii}$, while (5) implies that $\mathbf{R_{ss}}=\tilde{\mathbf{R}}_\mathbf{ss}$ and $\mathbf{R_{ii}}=\tilde{\mathbf{R}}_\mathbf{ii}$. If the off-diagonal blocks of $\mathbf{Q}$ are required to be zero, then (5) will follow from (4) and so $\mathbf{AA}^\mathsf{T}+\mathbf{BB}^\mathsf{T}$ will be uniquely separable.

This formulation highlights a difference between MFA with r_0 common factors and $r_1,\ldots,r_C$ unique factors and a single-channel factor analysis of same data with r_0+r factors and channel structure ignored. In the latter case, the product $[\mathbf{A}\ \mathbf{B}]\mathbf{Q}$ for any orthogonal $\mathbf{Q}$ yields a valid factor loading matrix, as no distinction is made between factors which influence multiple channels and those which are channel-specific. For MFA, the distinction between these types of factors is imposed by the requirement that $\mathbf{B}$ be block-diagonal with block sizes determined by $n_1,\ldots,n_c$ and $r_1,\ldots,r_C$. Therefore, in MFA, the transformation $[\mathbf{A}\ \mathbf{B}]\mathbf{Q}=[\tilde{\mathbf{A}}\ \tilde{\mathbf{B}}]$ must yield a $\tilde{\mathbf{B}}$ which also has the appropriate block structure. This will clearly be the case if $\mathbf{Q}$ is block diagonal. However, the converse is not true without restrictions on $\mathbf{n}$ and $\mathbf{r}$, even if $\mathbf{A}$ and $\mathbf{B}$ satisfy the lower-triangular conditions of [5, Sec. III].

In fact, if the dimensions of the signal and interference subspaces are too large relative to the channel sizes, such non-separability is typical. The following condition provides an upper bound on $\mathbf{r}$, above which almost all $\mathbf{R_{ss}}+\mathbf{R_{ii}}$ will not be uniquely separable into $\mathbf{R_{ss}}$ and $\mathbf{R_{ii}}$. This can be obtained by application of a theorem for confirmatory FA [13], which evaluates the needed linear constraints on the loading matrix to ensure it is locally determined.

Condition 1. *The channel sizes* $\mathbf{n}$ *and factor numbers* $\mathbf{r}$ *satisfy*

$$r_0r+\frac{1}{2}\sum_{c=1}^{C}r_c(r-r_c)\le\sum_{c=1}^{C}r_c(n-n_c). \tag{7}$$

Proposition 1. *If the channel sizes* $\mathbf{n}$ *and factor numbers* $\mathbf{r}$ *are such that* (4) *implies* $\mathbf{AA}^\mathsf{T}=\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\mathsf{T}$ *and* $\mathbf{BB}^\mathsf{T}=\tilde{\mathbf{B}}\tilde{\mathbf{B}}^\mathsf{T}$ *for almost all MFA loading matrices* $\mathbf{A},\tilde{\mathbf{A}}\in\mathbb{R}^{n\times r_0}$ *and* $\mathbf{B},\tilde{\mathbf{B}}\in\mathbb{R}^{n\times r}$ *patterned as in Section II, then Condition 1 is satisfied.*

The above proposition provides an upper bound on which values of $\mathbf{n}$ and $\mathbf{r}$ could allow the noise-free part of the observation covariance to be generically uniquely separable. On the other hand, the following proposition provides a sufficient condition on $[\mathbf{A}\ \mathbf{B}]$ which ensures that, if $[\mathbf{A}\ \mathbf{B}]\mathbf{Q}$ preserves the channel structure of $\mathbf{B}$, $\mathbf{Q}$ will be block diagonal and yields the unique separability of $\mathbf{AA}^\mathsf{T}+\mathbf{BB}^\mathsf{T}$. It is proven by an extension of the technique of [14], which proves a related result in confirmatory FA, to the block-structured case.

Proposition 2. *For MFA loading matrices* $\mathbf{A} \in \mathbb{R}^{n\times r_0}$ *and* $\mathbf{B} \in \mathbb{R}^{n\times r}$ *patterned as in Section II, suppose that, after possibly renumbering the channels, the submatrices* $\mathbf{M}_1, \ldots, \mathbf{M}_C$ *of* $[\mathbf{A}\ \mathbf{B}]$ *have full column rank, where* $\mathbf{M}_c$ *is*

$$\mathbf{M}_c = \begin{bmatrix} \mathbf{A}_{<c} & \mathbf{B}_{<c} \\ \mathbf{A}_{>c} & \mathbf{0} \end{bmatrix},\ \mathbf{M}_1 = [\mathbf{A}_2^\mathsf{T}\ \ldots\ \mathbf{A}_C^\mathsf{T}]^\mathsf{T}, \tag{8}$$

with $\mathbf{A}_{<c} = [\mathbf{A}_1^\mathsf{T} \ldots \mathbf{A}_{c-1}^\mathsf{T}]^\mathsf{T}$, $\mathbf{A}_{>c} = [\mathbf{A}_{c+1}^\mathsf{T} \ldots \mathbf{A}_C^\mathsf{T}]^\mathsf{T}$ *and* $\mathbf{B}_{<c} = \operatorname{blkdiag}(\mathbf{B}_1, \ldots, \mathbf{B}_{c-1})$. *Then any orthogonal* $\mathbf{Q}$ *patterned as* (6) *with* $[\mathbf{A}\ \mathbf{B}]\mathbf{Q} = [\tilde{\mathbf{A}}\ \tilde{\mathbf{B}}]$ *for some MFA loading matrices* $\tilde{\mathbf{A}} \in \mathbb{R}^{n\times r_0}$ *and* $\tilde{\mathbf{B}} \in \mathbb{R}^{n\times r}$ *must have* $\mathbf{Q}_{ij} = \mathbf{0}$ *for all* $i \neq j$.

Although framed in terms of the loading matrices $\mathbf{A}$ and $\mathbf{B}$, Proposition 2 also provides a result about $\mathbf{R_{ss}} + \mathbf{R_{ii}}$ itself. To see this, suppose $\mathbf{A}, \mathbf{A}' \in \mathbb{R}^{n\times r_0}$ and $\mathbf{B}, \mathbf{B}' \in \mathbb{R}^{n\times r}$ are MFA loading matrices such that $[\mathbf{A}\ \mathbf{B}][\mathbf{A}\ \mathbf{B}]^\mathsf{T} = \mathbf{R_{ss}} + \mathbf{R_{ii}} = [\mathbf{A}'\ \mathbf{B}'][\mathbf{A}'\ \mathbf{B}']^\mathsf{T}$. If the conclusion of Proposition 2 is true for $\mathbf{A}$ and $\mathbf{B}$, then it is similarly true for $\mathbf{A}'$ and $\mathbf{B}'$, which follows from the fact that $[\mathbf{A}'\ \mathbf{B}']$ equals $[\mathbf{A}\ \mathbf{B}]\mathbf{Q}'$ for some $\mathbf{Q}'$ and the product of block-diagonal matrices is block-diagonal. So, if $\mathbf{R_{ss}} + \mathbf{R_{ii}}$ equals $\mathbf{A}\mathbf{A}^\mathsf{T} + \mathbf{B}\mathbf{B}^\mathsf{T}$ for any $\mathbf{A}$ and $\mathbf{B}$ such that the hypothesis of Proposition 2 applies, then $\mathbf{R_{ss}} + \mathbf{R_{ii}}$ can be uniquely separated into $\mathbf{R_{ss}}$ and $\mathbf{R_{ii}}$.

However, as $\mathbf{R_{ss}} + \mathbf{R_{ii}}$ is not known, it is instead desirable to establish conditions on the channel sizes $\mathbf{n}$ and factor numbers $\mathbf{r}$ which ensure that Proposition 2 applies *generically*. The following conditions, which are established by investigating the structure of the submatrices $\mathbf{M}_1, \ldots, \mathbf{M}_C$, suffice to ensure that Proposition 2 applies and hence that $\mathbf{R_{ss}} + \mathbf{R_{ii}}$ is uniquely separable in almost all cases. Although Condition 2 as stated depends on the channel numbering, Proposition 3 shows that the choice of channel numbering does not in fact affect the generic separability.

Condition 2. *The channel sizes* $\mathbf{n}$ *and factor numbers* $\mathbf{r}$ *satisfy*

$$r_0 + \sum_{k=1}^{c-1} r_k \leq n - n_c, \tag{9}$$

for all $c = 1, \ldots, C$.

Proposition 3. (Generic Separability of $\mathbf{R_{ss}} + \mathbf{R_{ii}}$) *If Condition 2 is satisfied for some channel ordering, then the hypothesis of Proposition 2 is satisfied for* $\mathbf{A} \in \mathbb{R}^{n\times r_0}$ *and* $\mathbf{B} \in \mathbb{R}^{n\times r}$ *patterned as in Section II, excepting a null set of* $\mathbf{A}$ *and* $\mathbf{B}$.

Proposition 3 combined with Proposition 2 gives *sufficient* conditions on the channel sizes and factor numbers to allow for generic unique separability of $\mathbf{R_{ss}} + \mathbf{R_{ii}}$. Conversely, Proposition 1 gives *necessary* conditions on $\mathbf{n}$ and $\mathbf{r}$ for the same conclusion. To assess the size of the gap between the two sets of conditions, note that if the inequalities in (9) hold with equality, the necessary condition (7) will also hold with equality. So, for some values of $\mathbf{n}$ and $\mathbf{r}$, the necessary and sufficient conditions coincide. This provides reason to believe that the gap between the two conditions is not too large, which is supported by the quantitative comparisons in Section IV.

B. Isolation of Noise

For an MFA observation covariance $\mathbf{R_{xx}}$ obtained by (3) to be identified, the diagonal noise variance $\mathbf{\Phi}$ must be uniquely isolable from the low-rank portion of the observation covariance $\mathbf{R_{ss}} + \mathbf{R_{ii}}$. A problem of this type has been studied extensively in the literature on single-channel FA, as it is the crux of the identifiability problem for exploratory single-channel FA [9]. In MFA however, the multi-channel aspect of the observations alters the noise variance isolation problem, as the low-rank part of the observation covariance has additional structure not present in exploratory single-channel FA. This prevents direct application of previous results, but the technique of [10] used to prove the generic identifiability of the noise variance in single-channel FA can be generalized to MFA.

In single-channel FA, the central criterion for determining identifiability of the noise variances is

$$\phi(n, r, \rho) = \frac{r(r+1)}{2} - \frac{\rho(\rho+1)}{2} - \rho(r-\rho) - n,$$

for observation dimension n, factor number r, and $\rho \in \mathbb{N}$. The inequality $\phi(n, r, 2r-n) > 0$ is equivalent to the Ledermann bound [15] for $n > 6$, which provides the threshold for identifiability in the single-channel case [9], [10].

For MFA, unique isolation of the noise variances depends on a similar criterion,

$$\psi(\mathbf{n}, \mathbf{r}, \boldsymbol{\rho}) = n + \phi(n, r_0, \rho_0) + \sum_{c=1}^{C} \phi(n_c, r_c, \rho_c) + r_c(r_0 - \rho_0),$$

for non-negative integer vector $\boldsymbol{\rho} = [\rho_0, \rho_1, \ldots, \rho_C]^\mathsf{T}$. It can be seen that ψ is not a function of the total number of factors $r_0 + r$ alone, but instead depends on how those factors are distributed. So, the channel structure influences whether $\mathbf{\Phi}$ can be uniquely isolated from $\mathbf{R_{ss}} + \mathbf{R_{ii}}$, even though $\mathbf{\Phi}$ itself is not channel-structured.

Unlike in the single-channel case, the isolation of $\mathbf{\Phi}$ requires that the minimum value of ψ be positive over a class of $\mathbf{n}', \mathbf{r}'$ reduced from the original $\mathbf{n}, \mathbf{r}$, due to the constraints of the channel structure on the interference component. The following condition sets out the required class of reduced $\mathbf{n}', \mathbf{r}'$, over which $\psi(\mathbf{n}', \mathbf{r}', \boldsymbol{\rho}) > 0$ for all valid $\boldsymbol{\rho}$ and for all members of the class ensures that $\mathbf{\Phi}$ can be uniquely isolated. The set of reductions and valid $\boldsymbol{\rho}$ is M, defined below.

Condition 3. *The channel sizes* $\mathbf{n}$ *and factor numbers* $\mathbf{r}$ *satisfy* $r_0 + r \leq n$. *In addition, let* ψ^* *be the smallest criterion value over possible MFA reductions,*

$$\psi^* = \min_{(\mathbf{n}', \mathbf{r}', \boldsymbol{\rho}) \in M} \psi(\mathbf{n}', \mathbf{r}', \boldsymbol{\rho}) \tag{10}$$

where $M \subset \mathbb{N}^C \times \mathbb{N}^{C+1} \times \mathbb{N}^{C+1}$ *contains* $(\mathbf{n}', \mathbf{r}', \boldsymbol{\rho})$ *satisfying*

$$\begin{aligned} & n'_c \leq n_c,\ c = 1, \ldots, C, \\ & r'_c = (r_c - (n_c - n'_c))_+,\ c = 1, \ldots, C, \\ & r'_0 = [r_0 - \textstyle\sum_{c=1}^{C} (n_c - n'_c - r_c)_+]_+, \\ & \rho_c \leq \min\{r'_c, 2(r'_0 + r'_c) - n'_c\},\ c = 1, \ldots, C, \\ & \rho_0 = \min\{r_0, 2r'_0 + \textstyle\sum_{c=1}^{C} 2r'_c - \rho_c - n'_c\}, \end{aligned} \tag{11}$$

and $\sum_{c=1}^{C} n'_c > 0$. *Either* $\psi^* > 0$ *or* M *is empty.*

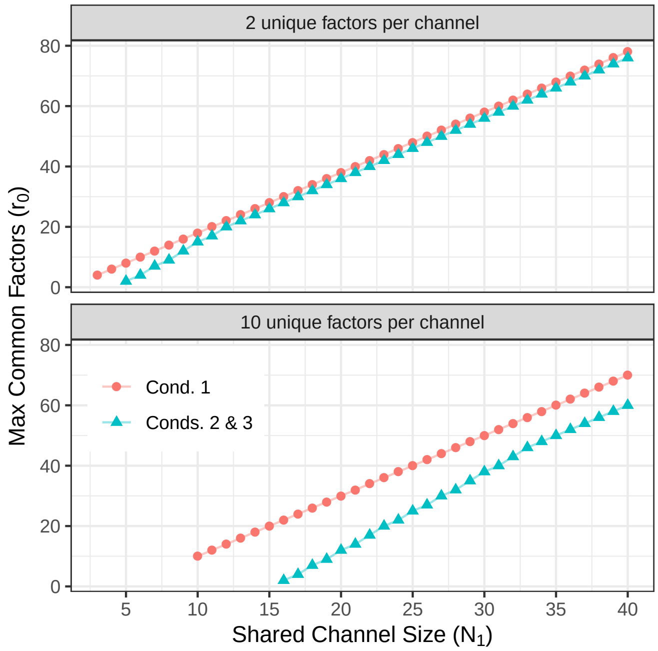

Fig. 1. Comparison of maximum identifiable common factor number r_0 under two sets of conditions, for varying channel sizes. Channel structure depicted is three homogeneous channels, with the number of unique factors per channel indicated by the facet heading.

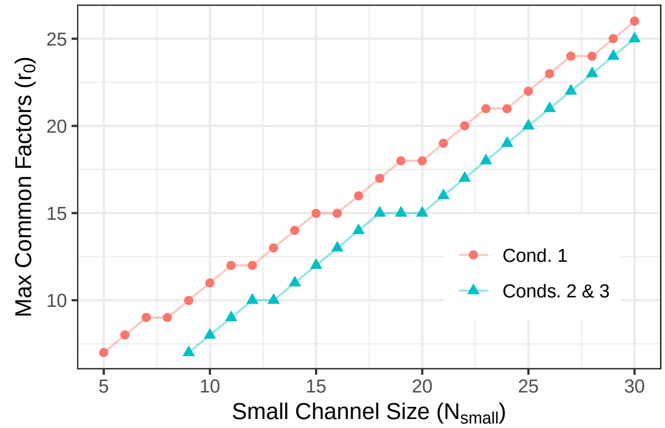

Fig. 2. Comparison of maximum identifiable common factor number r_0 under two sets of conditions, with two heterogeneous channels. Large channel size is fixed at $n_{\text{big}} = 30$ with $r_{\text{big}} = 15$ unique factors. Small channel size varies, with fixed number of unique factors $r_{\text{small}} = 5$.

As in [10], $(\mathbf{A}, \mathbf{B}, \mathbf{\Phi})$ is said to have identified noise variances if $\mathbf{A}\mathbf{A}^\top + \mathbf{B}\mathbf{B}^\top + \mathbf{\Phi} = \tilde{\mathbf{A}}\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\tilde{\mathbf{B}}^\top + \tilde{\mathbf{\Phi}}$ implies that $\mathbf{\Phi} = \tilde{\mathbf{\Phi}}$, where $(\mathbf{A}, \mathbf{B}, \mathbf{\Phi})$ and $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{\Phi}})$ are as defined in Section II with equal channel sizes and factor numbers. The following Proposition shows that Condition 3 is sufficient for $(\mathbf{A}, \mathbf{B}, \mathbf{\Phi})$ to generically have identifiable noise variances. So, when $\mathbf{n}$ and $\mathbf{r}$ satisfy Condition 3, the noise variances can be uniquely isolated from $\mathbf{R}_{\mathbf{ss}} + \mathbf{R}_{\mathbf{ii}}$ except for a null set.

Proposition 4. (Isolation of $\mathbf{\Phi}$) *If Condition 3 is met,* $(\mathbf{A}, \mathbf{B}, \mathbf{\Phi})$ *have identified noise variances except for a null subset of* $(\mathbf{A}, \mathbf{B}, \mathbf{\Phi})$.

IV. Discussion

This section explores the quantitative relations between Conditions $1-4$ and their associated identifiability results. For single-channel FA, the intuitive understanding of identifiability is that taking the number of factors to be substantially smaller than the total number of observations will ensure the uniqueness of the FA decomposition. For multi-channel FA, however, the channel structure of the data and the inclusion of two types of latent factors renders intuitive assessment of identifiability more challenging. To improve this intuition, Section IV-A discusses the identifiability in the special case where the channels are equally sized with equal unique factor numbers, while Section IV-B discusses how substantial inequality between channels influences identifiability. Finally, Section IV-C describes the relationship between the two identifiability subproblems.

A. Homogeneous Channels

A common measurement setup to which MFA is applicable is the case of C *homogeneous* multivariate sensors. That is, the C observation vectors $\mathbf{x}_1, \dots, \mathbf{x}_C$ are obtain from distinct but otherwise identical sensors, and so the channels associated with each sensor have equal dimensions, $n_1 = \dots = n_C$. If the channels are treated as interchangeable, an additional simplifying assumption is that the *unique* factor numbers associated with each channel are similarly identical, $r_1 = \dots = r_C$. In this case, identifiability of MFA depends only on the number of channels C, the number of common factors r_0, the shared number of unique factors r_1, and the shared channel size n_1. In this case, Conditions 1 and 2 have simpler form, as given in the following Corollary.

Corollary 1. *For C channels with $n_1 = \dots = n_c$ and $r_1 = \dots = r_C$ satisfying $r_1 \leq n_1$, Condition 1 is*

$$r_0 \leq (C-1)(n_1 - r_1/2).$$

and $r_1 \leq \frac{1}{2}(2n_1 + 1 - \sqrt{8n_1 + 1})$.

Condition 3, however, is not simplified by the assumption of homogeneity, as the class of reduced $\mathbf{n}', \mathbf{r}'$ quantified over in (10) includes non-equal channel sizes and factor numbers.

For $C = 3$ channels with $r_1 = 2, 5, 10$ unique factors per channel, Figure 1 compares the maximum r_0 satisfying Condition 1 and satisfying both Conditions 2 and 3. As Proposition 1 shows that Condition 1 must be met for $\mathbf{R}_{\mathbf{ss}} + \mathbf{R}_{\mathbf{ii}}$ to be separable, MFA with (n_1, r_0) above the top line (indicated with circles) in Figure 1 cannot be identifiable. Conditions 2 and 3 together imply that both identification subproblems are generically solvable, and so MFA with (n_1, r_0) below the bottom line will be generically identifiable.

B. Heterogeneous Channels

In the case where the channels are of different sizes or have differing numbers of unique factors, the identifiability of MFA

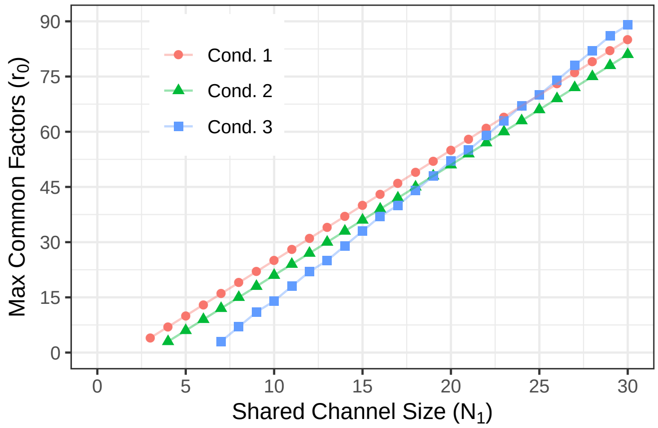

Fig. 3. Comparison of maximum identifiable common factor number r_0 under Conditions $1-3$, for $C=4$ homogeneous channels each with shared channel size ranging from $n_1=3$ to $n_1=40$ and $r_1=3$ unique factors.

depends on all of $\mathbf{n}$ and $\mathbf{r}$ rather than any unidimensional summaries thereof. As MFA is equally applicable to datasets obtained from heterogeneous sensors, understanding identifiability for channels of differing sizes is practically relevant.

In Figure 2, the maximal identifiable r_0 under the same two sets of conditions as defined in Section IV-A are plotted, for $C=2$ heterogenous channels. The larger channel has fixed size $n_{\text{big}}=30$ and $r_{\text{big}}=15$ unique factors. The smaller channel size n_{small} ranges from $n_{\text{small}}=3$ to $n_{\text{small}}=30$ with $r_{\text{small}}=5$ unique factors. It can be seen that the heterogeneity in the channel sizes affects MFA identifiability in a step-wise fashion, as different inequalities become binding at different levels of channel size imbalance.

C. Partial Identifiability

As Figure 3 illustrates, there are channel sizes and factor numbers for which one of the two identifiability subproblems has a unique solution but not the other. If Condition 3 is satisfied but Condition 1 is violated, the noise variances can be uniquely isolated from the combined signal-and-interference covariance $\mathbf{R_{ss}}+\mathbf{R_{ii}}$ but the latter term does not uniquely determine $(\mathbf{R_{ss}},\mathbf{R_{ii}})$. In this case, the identified $\mathbf{\Phi}$ can be interpreted, allowing estimation of some quantities of interest such as the relative contribution of idiosyncratic noise to the overall observation variability. The other case, where the noise-free part of the MFA covariance can be uniquely separated into $(\mathbf{R_{ss}},\mathbf{R_{ii}})$ but the idiosyncratic noise $\mathbf{\Phi}$ is not guaranteed to be isolable, is of less practical relevance as MFA is most applicable to the uncalibrated case where $\mathbf{\Phi}$ is to be estimated.

V. Conclusion

Multi-channel factor analysis decomposes a multivariate, multi-channel observation into *signal*, *interference*, and *noise*. For MFA to be practically useful, it must be possible to certify the uniqueness of the associated covariance decomposition using only what is specified *a priori*, namely the channel sizes and dimensions of the signal and interference subspaces. The identifiability question for MFA is divides into two questions: can the signal and interference covariances be uniquely *separated* in the absence of noise, and can the idiosyncratic noise variances can be uniquely *isolated* from the systemic part of the observation covariance?

The main results presented in this paper ensure that the two questions lead to unique solutions in the *generic* sense (that is, excepting a null set of degenerate cases). The relationships between the various identifiability conditions are explored in Section IV, which also illustrates that MFA is identifiable for common and unique factor numbers which are reasonable relative to the channel sizes. In sum, the results of this paper provide theoretical support for MFA and justify its application to real-world array processing problems.

Acknowledgments

This work was supported in part by National Science Foundation grants DMS-1923142, CNS-1932413, and DMS-2123761. The work of I. Santamaria was funded by MCIN/ AEI /10.13039/501100011033, under grant PID2022-137099NB-C43 (MADDIE). The work of D. Ramírez was partially supported by MCIN/AEI/10.13039/501100011033/FEDER, UE, under grant PID2021-123182OB-I00 (EPiCENTER) and by the Office of Naval Research (ONR) Global under contract N62909-23-1-2002.

References

[1] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904. Place: US Publisher: Univ of Illinois Press.

[2] M. Pesavento and A. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1310–1324, 2001.

[3] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Estimation of sensor array signal model parameters using factor analysis," in *27th European Signal Processing Conference (EUSIPCO)*, 2019.

[4] D. Ramírez, I. Santamaría, and L. Scharf, *Coherence in Signal Processing and Machine Learning*. Springer Cham, 2022.

[5] D. Ramírez, I. Santamaria, L. L. Scharf, and S. Van Vaerenbergh, "Multi-channel factor analysis with common and unique factors," *IEEE Transactions on Signal Processing*, vol. 68, pp. 113–126, 2020.

[6] D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multiple-channel signal detection," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2049–2057, 1995.

[7] S. Bose and A. O. Steinhardt, "Optimum array detector for a weak signal in unknown noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 3, pp. 911–922, 1996.

[8] H. Wu, W. Xiao, and H. Ren, "Automatic time picking for weak seismic phase in the strong noise and interference environment: An hybrid method based on array similarity," *Sensors*, vol. 22, no. 24, 2022.

[9] A. Shapiro, "Identifiability of factor analysis: Some results and open problems," *Linear Algebra and its Applications*, vol. 70, pp. 1–7, 1985.

[10] P. A. Bekker and J. M. ten Berge, "Generic global identification in factor analysis," *Linear Algebra and its Applications*, vol. 264, pp. 255–263, 1997. Sixth Special Issue on Linear Algebra and Statistics.

[11] F. Király and R. Tomioka, "A combinatorial algebraic approach for the identifiability of low-rank matrix completion," *29th International Conference on Machine Learning (ICML)*, vol. 1, 06 2012.

[12] T. W. Anderson and H. Rubin, "Statistical inference in factor analysis," *University of California Press*, pp. 111–150, 1956.

[13] P. A. Bekker, "A note on the identification of restricted factor loading matrices.," *Psychometrika*, vol. 51, pp. 607–611, 1986.

[14] J. E. Dunn, "A note on a sufficiency condition for uniqueness of a restricted factor matrix," *Psychometrika*, vol. 38, pp. 141–143, 1973.

[15] W. Ledermann, "On the rank of the reduced correlational matrix in multiple-factor analysis," *Psychometrika*, vol. 2, pp. 85–93, 1937.