Multi-Channel Factor Analysis with Common and Unique Factors

David Ramírez, *Senior Member, IEEE,* Ignacio Santamaria, *Senior Member, IEEE,* Louis L. Scharf, *Life Fellow, IEEE,* and Steven Van Vaerenbergh, *Senior Member, IEEE*

***Abstract*— This work presents a generalization of classical factor analysis (FA). Each of M channels carries measurements that share factors with all other channels, but also contains factors that are unique to the channel. Furthermore, each channel carries an additive noise whose covariance is diagonal, as is usual in factor analysis, but is otherwise unknown. This leads to a problem of multi-channel factor analysis with a specially structured covariance model consisting of shared low-rank components, unique low-rank components, and diagonal components. Under a multivariate normal model for the factors and the noises, a maximum likelihood (ML) method is presented for identifying the covariance model, thereby recovering the loading matrices and factors for the shared and unique components in each of the M MIMO channels. The method consists of a three-step cyclic alternating optimization. Two of the steps have closed-form solutions, and the other is an Expectation-Maximization (EM) step. Numerical results demonstrate the performance of the proposed algorithm and its application passive radar application.**

***Index Terms*— Detection, expectation-maximization (EM) algorithm, factor analysis (FA), maximum likelihood (ML), multi-channel factor analysis (MFA), multiple-input multiple-output (MIMO) channels, multivariate normal (MVN) model, passive radar.**

I. Introduction

Classical factor analysis (FA) was pioneered by Spearman in his seminal paper [1]. Spearman and others applied FA to problems in psychology, and especially to the analysis of the correlation of children's scores across different academic subjects. Later, with the work of Lawley, Anderson, and others [2]–[4], a more rigorous approach was developed, which made FA a well-established technique in multivariate statistics. FA now finds many applications in science and engineering. For instance, in the field of array signal processing, FA has been applied to radio astronomy [5], [6], cognitive radio [7], direction-of-arrival estimation [8]–[10], modal analysis [11], [12], and detection or source enumeration [13]–[16].

In the classical FA model, measurements in a single MIMO channel are modeled as a set of unknown factors scaling the modes of an unknown factor loading matrix, plus a multivariate normal noise of unknown, but diagonal, covariance. The factors are typically treated as multivariate normal, with identity covariance, so that the net effect is to posit a multivariate normal measurement with a structured covariance consisting of an unknown low-rank component to account for the factor loadings plus an unknown diagonal matrix to account for the additive noise. Thus, in second-order FA, the problem is to estimate a low-rank plus diagonal covariance matrix from several multivariate observations. This model has been recently extended in [17] to consider more complicated covariance structures, i.e., not diagonal, but this structure needs to be sparse and known. Common estimation approaches for the FA model are based on the maximum likelihood (ML) criterion. Unfortunately, even under the Gaussian assumption, the maximization problem has no closed-form solution and numerical methods must be employed. A convergent numerical procedure for obtaining the maximum likelihood estimates was first given by Joreskog [18], [19] (cf. Chapter 9 in [3]). Other optimization approaches have been investigated for this problem, ranging from steepest descent [4] and alternating optimization methods [7], [20], [21], to Expectation-Maximization algorithms [22], [23].

The problem of FA analysis has been extended to multiple channels of multivariate observations. To the best of our knowledge, the first generalization was developed by Tucker [24], where he proposed the so-called *inter-battery FA*. In this work, the observations of two channels are composed by linear combinations of *common* factors and independent noises without a particular covariance structure. Additionally, he derived an estimation algorithm based on the least squares (LS) criterion, which was later related to canonical correlation analysis (CCA) by Browne [25]. The works in [26], [27] extend the inter-battery FA model to more than two channels.

The recent work in [17] also proposes a different generalization of FA to several channels, which assumes that the factors at each channel are independent, but the noise covariance matrix is common. Moreover, the number of factors in each channel may be different. A different generalization is presented in [28], and termed group factor analysis (GFA). In GFA, the factors may be common to all channels or to a

D. Ramírez is with the Department of Signal Theory and Communications, University Carlos III of Madrid, Leganés, Spain and with the Gregorio Marañón Health Research Institute, Madrid, Spain (e-mail: david.ramirez@uc3m.es).

I. Santamaria and S. Van Vaerenbergh are with the Department of Communications Engineering, University of Cantabria, Santander, Spain (e-mail: {i.santamaria,steven.vanvaerenbergh}@unican.es).

L. L. Scharf is with the Department of Mathematics, Colorado State University, USA (e-mail: louis.scharf@colostate.edu).

The work of D. Ramírez was supported by the Ministerio de Ciencia, Innovació'n y Universidades, jointly with the European Commission (ERDF), under grants TEC2015-69868-C2-1-R (ADVENTURE), CAIMAN (TEC2017-86921-C2-2-R) and TEC2017-92552-EXP (aMBITION), and The Comunidad de Madrid under grant Y2018/TCS-4705 (PRACTICO-CM). The work of I. Santamaria and S. Van Vaerenbergh was supported by Ministerio de Ciencia, Innovación y Universidades and AEI/FEDER funds of the E.U., under grant TEC2016-75067-C4-4-R (CARMEN). The work of L. Scharf was supported in part by the National Science Foundation under grant CCF-1712788. A short preliminary version of this paper was presented at the 2018 Asilomar Conference on Signals, Systems, and Computers.

subset of them. Other work related to multi-channel factor analysis includes the parallel factor (PARAFAC) analysis model [29] and independent vector analysis (IVA) [30]. Our model for the channel covariance differs significantly from the channel covariance in the PARAFAC model [29]. The multiple channels (i.e., the third dimension in the three-way array) of PARAFAC are obtained from displaced but otherwise identical subarrays, which induces a shift-invariant structure in the loading matrices of the common factors. Further, the noise covariance model in [29] is white. Our model is different is several aspects, namely, we never have rotational invariance, we have both common and unique (or channel-specific) factors, and our model for the noise covariance is diagonal with unknown variances. The standard model in independent vector analysis (IVA) accounts for the dependence of a set of common sources or factors observed through several mixing matrices, but it does not consider channel specific factors or spatially correlated noises [30].

This paper extends the inter-battery FA model to more than two channels and to noise covariance structures that account for additive noise and unique channel factors, which are missing in the original inter-battery analysis of Tucker. This new model is particularly relevant for passive radar since it accounts for leakage of a reference channel transmission into the surveillance channel [21].

The iterative procedure to obtain the maximum likelihood estimates of the multi-channel FA model developed in this paper bears resemblance to ML estimation in the FA model, where there also does not exist a closed-form solution. The iterative procedure consists of three steps. In the first step, a closed-form solution for the loading matrices of the common factors is found. In the second step, one iteration of an EM algorithm returns an estimate of the loading matrices for the uncommon factors. The third step returns a closed-form solution for the estimate of the diagonal noise covariance matrices. We prove that this algorithm converges to a local maximum of the likelihood and demonstrate its performance on several illustrative problems.

A. Outline

The outline of this paper is as follows: Section II summarizes the classical FA model, as well as an ML estimation procedure based on an alternating optimization approach. A brief introduction to inter-battery FA and the proposed generalization are presented in Section III. This section also describes the ML estimation of the unknown parameters. The alternating optimization ML algorithm is derived in Section III-B. Finally, in Section IV the performance of the proposed method is illustrated by means of numerical simulations, and the main conclusions are summarized in Section V.

B. Notation

In this paper, matrices are denoted by bold-faced upper case letters, bold-faced lower case letters are denoted by column vectors, and scalars are denoted by light-face lower case letters. The superscript $(\cdot)^T$ denotes transpose, and the determinant, Frobenius norm and trace of a matrix $\mathbf{A}$ are denoted $\det(\mathbf{A})$, $\|\mathbf{A}\|_F$ and $\operatorname{tr}(\mathbf{A})$, respectively. A real matrix of dimension $M \times N$ is denoted $\mathbf{A} \in \mathbb{R}^{M\times N}$, and $\mathbf{x} \in \mathbb{R}^M$ indicates that $\mathbf{x}$ is a real vector of dimension M. The notation $\mathbf{x} \sim \mathcal{N}_M(\boldsymbol{\mu}, \mathbf{R})$ indicates that $\mathbf{x}$ is an M-dimensional Gaussian random vector of mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$ and $E[\cdot]$ represents the expectation operator. The identity matrix of size $L \times L$ is $\mathbf{I}_L$, $\mathbf{0}_{M\times N}$ denotes the zero matrix of the dimension $M \times N$. We use $\mathbf{A}^{1/2}$ to denote the symmetric square root matrix of the symmetric matrix $\mathbf{A}$. Finally, $\operatorname{diag}(\mathbf{A})$ constructs a diagonal matrix from the diagonal of $\mathbf{A}$, the operator blkdiag denotes block diagonal concatenation of matrices, and $\delta[n]$ denotes the Kronecker delta.

II. CLASSICAL FA

In single-channel (or classical) factor analysis (FA) [2]–[4], the real-valued observations $\mathbf{x} \in \mathbb{R}^n$ are modeled as[1]

$$\mathbf{x} = \mathbf{H}\mathbf{f} + \mathbf{e}, \tag{1}$$

where $\mathbf{f} \in \mathbb{R}^p$ contains the p factors, and $\mathbf{H} \in \mathbb{R}^{n\times p}$ is the factor loading matrix; p is usually much smaller than n. The n-dimensional noise vector $\mathbf{e}$ is typically assumed zero mean, Gaussian distributed and its components are independent, i.e., $\mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma}$ is diagonal. In classical FA the factors $\mathbf{f}$ are assumed to be zero-mean Gaussian with identity covariance. As a consequence, the measurements $\mathbf{x}$ are zero-mean Gaussian with an $n \times n$ covariance matrix

$$\mathbf{R} = \mathbf{H}\mathbf{H}^T + \boldsymbol{\Sigma}. \tag{2}$$

That is, the covariance matrix is a non-negative definite rank-p matrix plus a diagonal covariance $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. The FA model implies that, conditioned on the factors, the observations are uncorrelated, and hence the common factors explain all the dependence structure among the observations.

The invariances of this model determine the identifiability of this second-order model. The covariance model of (1) is invariant to the transformation $\mathbf{H} \longrightarrow \mathbf{H}\mathbf{Q}\gamma, \mathbf{f} \longrightarrow \gamma^{-1}\mathbf{Q}^T\mathbf{f}$, where $\mathbf{Q}$ is any orthogonal matrix, and $\gamma \neq 0$. The model is therefore unique only up to equivalence class of frames $\mathbf{H}$, denoted by the subspace $\langle\mathbf{H}\rangle$, which is a point on a Grassmannian manifold of dimension p. Moreover, any estimation procedure provides unique solutions only when the number of factors p is sufficiently small in comparison to the dimension of the ambient space. A model is said to be generically identified if we can find a unique FA factorization as in (2) for almost every pair of matrices $(\mathbf{H}, \boldsymbol{\Sigma})$ viewed as points in a parameter space of dimension $(np + n)$ [31]. The non-identifiable models therefore should live in a set of zero Lebesgue measure. According to this definition, it was proven in [32] that a necessary and sufficient condition for a FA model to be generically identified is

$$(n - p)^2 - (n + p) > 0. \tag{3}$$

[1]To simplify the exposition, the case of real-valued channels is considered throughout this work, but its extension to the complex-valued case is straightforward.

Other definitions of identifiability are possible. In [17], a model is considered identifiable if the corresponding Fisher information matrix is nonsingular. Using this definition, it is shown in [17] that (3) is a necessary (but not sufficient) condition for the uniqueness of the solution.

A. ML estimation in the FA model

Maximum likelihood is the most common principle for estimation in factor analysis. However, since it is not possible to find the ML estimates of $(\mathbf{H}, \mathbf{\Sigma})$ in closed form, solutions based on iterative procedures have been typically proposed. These include numerical procedures by Joreskog based on first-order or second-order derivatives [18], [19], alternating optimization methods [7], [20], [21], and Expectation-Maximization type algorithms [20], [22], [23].

In our experience, alternating optimization methods are preferable for moderate-size problems. For instance, the alternating optimization approach in [21] operates as follows. It starts with the likelihood function for N observations, $\mathbf{x}[1], \ldots, \mathbf{x}[N]$, which is to be maximized with respect to the factor loading matrix $\mathbf{H}$ and the diagonal noise covariance matrix $\mathbf{\Sigma}$.

The likelihood of the observations is

$$l(\mathbf{H}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{nN/2} \det^{N/2}(\mathbf{R})} \exp\left[-\frac{N}{2} \operatorname{tr}(\mathbf{R}^{-1}\mathbf{S})\right], \quad (4)$$

where

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}[n]\mathbf{x}^T[n], \quad (5)$$

is the sample covariance matrix. The ML estimation problem can be re-formulated as

$$\underset{\mathbf{H},\mathbf{\Sigma}}{\text{maximize}} \quad \mathcal{L}(\mathbf{H}, \mathbf{\Sigma}), \quad (6)$$

where the objective function is

$$\mathcal{L}(\mathbf{H}, \mathbf{\Sigma}) = -\log\det(\mathbf{H}\mathbf{H}^T + \mathbf{\Sigma}) - \operatorname{tr}\left[(\mathbf{H}\mathbf{H}^T + \mathbf{\Sigma})^{-1}\mathbf{S}\right].$$

There is no closed-form solution to the problem (6), but it is possible to find a local maximum of likelihood by assuming $\mathbf{\Sigma}$ known and maximizing with respect to $\mathbf{H}$, then assuming $\mathbf{H}$ known and maximizing with respect to $\mathbf{\Sigma}$, etc. Concretely, defining the noise-whitened sample covariance matrix

$$\tilde{\mathbf{S}} = \hat{\mathbf{\Sigma}}^{-1/2}\mathbf{S}\hat{\mathbf{\Sigma}}^{-1/2}, \quad (7)$$

and its eigenvalue decomposition

$$\tilde{\mathbf{S}} = \tilde{\mathbf{W}} \operatorname{diag}\left(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n\right) \tilde{\mathbf{W}}^T, \quad (8)$$

with $\tilde{\lambda}_i \geq \tilde{\lambda}_{i+1}$, the ML estimate of $\mathbf{H}$ is [21]

$$\hat{\mathbf{H}} = \hat{\mathbf{\Sigma}}^{1/2}\tilde{\mathbf{W}}\tilde{\mathbf{D}}^{1/2}\mathbf{Q}, \quad (9)$$

where $\hat{\mathbf{\Sigma}}$ is the previous estimate of $\mathbf{\Sigma}$, $\tilde{\mathbf{D}} = \operatorname{diag}\left(\tilde{d}_1, \ldots, \tilde{d}_p, 0, \ldots, 0\right)$, with $\tilde{d}_i = \max(\tilde{\lambda}_i - 1, 0)$, and $\mathbf{Q}$ is any orthogonal matrix. Now, given the above estimate of $\mathbf{H}$, $\hat{\mathbf{H}}$, the ML estimate of $\mathbf{\Sigma}$ is [21]

$$\hat{\mathbf{\Sigma}} = \operatorname{diag}\left(\mathbf{S} - \hat{\mathbf{H}}\hat{\mathbf{H}}^T\right). \quad (10)$$

Since each step of the above procedure may not decrease the value of the cost function, this alternating algorithm is ensured to attain a local maximum of the likelihood.

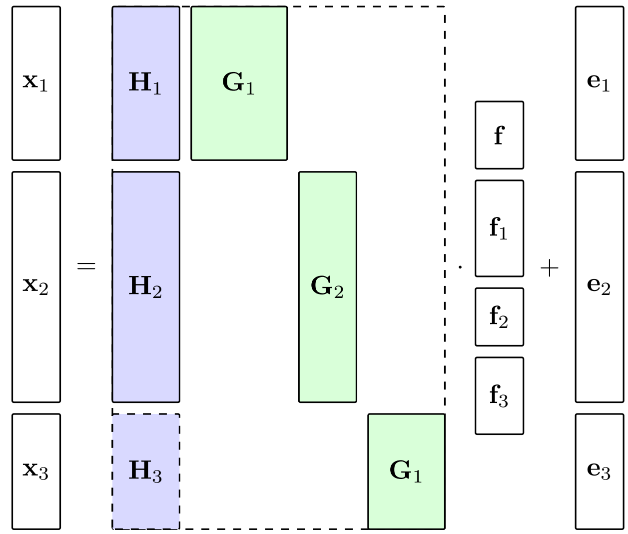


Fig. 1: Diagram of the proposed multi-channel factor analysis model for $M = 3$ channels. The observations are represented by $\mathbf{x}_i$; The loading matrices for the common factors $\mathbf{f}$ and for the unique factors $\mathbf{f}_i$ are depicted by $\mathbf{H}_i$ (blue) and $\mathbf{G}_i$ (green), respectively; $\mathbf{e}_i$ is the channel noise, for $i = 1, \ldots, M$.

III. Multi-channel Factor Analysis

The first generalization of FA to more than one channel was introduced by Tucker in the fifties [24]. This generalization, known as inter-battery FA, aims at extracting factors $\mathbf{f}$ common to two sets of variables (or batteries), and is based on the generative model

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{H}_1\mathbf{f} + \mathbf{e}_1, \\ \mathbf{x}_2 &= \mathbf{H}_2\mathbf{f} + \mathbf{e}_2, \end{aligned} \quad (11)$$

where the covariance matrices of the noise vectors does not have any further structure besides being arbitrary positive definite matrices. In the work of Tucker, a solution is proposed based on a least squares (LS) criterion, which results in the singular value decomposition (SVD) of the sample cross-correlation matrix between the two data sets. Interestingly, a few decades later, Browne presented a connection between the inter-battery FA and canonical correlation analysis (CCA) in [25]. The extension of the inter-battery FA model to more than two channels was developed in [26], [27].

We propose the following generalization of inter-battery FA analysis. We consider $M \geq 2$ channels with noise covariance matrices that have further structure to account for the existence in each channel of a factor component that is unique to the channel. The generative model is

$$\mathbf{x}_i = \mathbf{H}_i\mathbf{f} + \mathbf{G}_i\mathbf{f}_i + \mathbf{e}_i, \quad i = 1, \ldots, M, \quad (12)$$

where $\mathbf{H}_i \in \mathbb{R}^{n_i \times p}$ is the loading matrix in channel i for the common factors $\mathbf{f}$ and $\mathbf{G}_i \in \mathbb{R}^{n_i \times p_i}$ is the loading matrix in channel i for the unique factors $\mathbf{f}_i$; $\mathbf{e}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbf{\Sigma}_i)$ is the noise in channel i. This model is illustrated in Fig. 1 for $M = 3$. The noise covariance matrices, $\mathbf{\Sigma}_i$, are diagonal and the noises at different channels are uncorrelated: $E[\mathbf{e}_i\mathbf{e}_j^T] = \mathbf{\Sigma}_i\delta[i-j]$. Moreover, common and specific factors are uncorrelated: $E[\mathbf{f}\,\mathbf{f}_i^T] = \mathbf{0}_{p \times p_i}, \forall i$, and $E[\mathbf{f}_i\mathbf{f}_j^T] = \mathbf{0}_{p_i \times p_j}$, for $i \neq j$.

In this multi-channel generative model, the common factors explain the inter-channel dependence structure, whereas the unique factors explain the intra-channel dependence structure. Further, conditioned on both the common and unique factors, the multi-channel observations are uncorrelated. This structure makes our model different from other multi-channel models assumed in PARAFAC [29], IVA [30] or multiset CCA [33], [34].

As with classical FA, only the subspaces $\langle \mathbf{H} \rangle$ and $\langle \mathbf{G}_i \rangle$ can be identified. Thus, without loss of generality, we consider the factors to be normalized as follows: $E[\mathbf{f}\,\mathbf{f}^T] = \mathbf{I}_p$, $E[\mathbf{f}_i\mathbf{f}_i^T] = \mathbf{I}_{p_i}$.

Assuming a multivariate normal model for common and uncommon factors, the composite vector $\mathbf{x} = [\mathbf{x}_1^T \cdots \mathbf{x}_M^T]^T$ is distributed as $\mathcal{N}_n(\mathbf{0}, \mathbf{R})$ with a structured covariance matrix that is

$$\mathbf{R} = \mathbf{H}\mathbf{H}^T + \mathbf{E}. \tag{13}$$

Here, the composite loading matrix is $\mathbf{H} = [\mathbf{H}_1^T \cdots \mathbf{H}_M^T]^T \in \mathbb{R}^{n \times p}$, with $n = \sum_{i=1}^M n_i$. The common-factors-plus-noise covariance matrix is

$$\mathbf{E} = \mathbf{G}\mathbf{G}^T + \mathbf{\Sigma}, \tag{14}$$

where the composite loading matrix for the uncommon factors and the composite noise covariance matrix are, respectively,

$$\mathbf{G} = \text{blkdiag}[\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_M] \tag{15}$$

and

$$\mathbf{\Sigma} = \text{blkdiag}[\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \ldots, \mathbf{\Sigma}_M]. \tag{16}$$

In these equations, $\mathbf{G}_i \in \mathbb{R}^{n_i \times p_i}$ and $\mathbf{\Sigma}_i \in \mathbb{R}^{n_i \times n_i}$.

As in single-channel FA, the identification of a MFA model is not unique without a constraint on the number of parameters to be identified. In the following, we present a necessary condition on the largest number of common and specific factors that yield a unique solution. To do so, we need to count how many knowns and unknowns the model has. The number of knowns is given by the number of different elements of the sample covariance matrix, which are $n(n+1)/2$. The number of unknowns is slightly more involved to compute. Let us start with the number of unknowns in $\mathbf{E}_i$, which are those of the classical FA model, i.e., $n_i + n_i p_i - p_i(p_i - 1)/2$. Finally, since the number of unknowns in $\mathbf{H}\mathbf{H}^T$ are $np - p(p-1)/2$, for the solution to be unique it is required that

$$\frac{n(n+1)}{2} - np + \frac{p(p-1)}{2} - \sum_{i=1}^{M} \left(n_i + n_i p_i - \frac{p_i(p_i-1)}{2} \right) > 0. \tag{17}$$

Additionally, after removing the common factors each single-channel FA model must be also identifiable, which adds the following conditions

$$(n_i - p_i)^2 - (n_i + p_i) > 0, \qquad i = 1, \ldots, M. \tag{18}$$

A. ML estimation in the MFA model

In this section, we present the ML estimation of the unknown parameters in the MFA model. In particular, assuming that N observations of each channel are available, $\mathbf{x}_i[1], \ldots, \mathbf{x}_i[N], i = 1, \ldots, M$, the goal is to estimate the composite common-factor loading matrix $\mathbf{H}$, the composite channel-specific loading matrix $\mathbf{G}$, and the composite diagonal noise covariance matrix $\mathbf{\Sigma}$ that maximize the likelihood. Hence, the ML estimates of $\mathbf{H}, \mathbf{G}$, and $\mathbf{\Sigma}$ are obtained by solving the maximization problem

$$\underset{\mathbf{H},\mathbf{G},\mathbf{\Sigma}}{\text{maximize}} \quad \mathcal{L}(\mathbf{H}, \mathbf{G}, \mathbf{\Sigma}), \tag{19}$$

where the objective function is

$$\mathcal{L}(\mathbf{H}, \mathbf{G}, \mathbf{\Sigma}) = -\log\det(\mathbf{R}) - \text{tr}\left(\mathbf{R}^{-1}\mathbf{S}\right), \tag{20}$$

with $\mathbf{R}$ given in (13) and the sample covariance matrix given in (5) with the vector of multi-channel observations $\mathbf{x}[n] = [\mathbf{x}_1^T[n] \cdots \mathbf{x}_M^T[n]]^T$.

The maximization problem in (19) does not have a closed-form solution. In this work, we propose therefore to use an alternating optimization approach, as described in the following section.

B. Alternating Optimization Algorithm

We propose a cyclic alternating-optimization algorithm for maximizing the log-likelihood function in (20). The procedure applies three steps in a cyclic fashion. At each of the three steps, a subset of variables is optimized while the remaining variables are fixed at previously estimated values. The fixed parameters in each step are denoted with a hat, whereas the parameters to be optimized are denoted without a hat. That is, $\mathcal{L}(\mathbf{H}, \hat{\mathbf{G}}, \hat{\mathbf{\Sigma}})$ is the objective function for fixed values of $\mathbf{G}$ and $\mathbf{\Sigma}$.

a) Step 1: Estimation of $\mathbf{H}$*:* The first step of the proposed method consists in estimating $\mathbf{H}$, assuming that $\mathbf{G}$ and $\mathbf{\Sigma}$ are fixed. Thus, the optimization problem is

$$(\mathcal{P}_1) \quad \underset{\mathbf{H}}{\text{maximize}} \quad \mathcal{L}\left(\mathbf{H}, \hat{\mathbf{G}}, \hat{\mathbf{\Sigma}}\right). \tag{21}$$

With $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{G}}$ fixed at their previously estimated values, $\hat{\mathbf{E}} = \hat{\mathbf{G}}\hat{\mathbf{G}}^T + \hat{\mathbf{\Sigma}}$ is fixed. The whitened sample covariance matrix and its eigenvalue decomposition are

$$\tilde{\mathbf{S}} = \hat{\mathbf{E}}^{-1/2}\mathbf{S}\hat{\mathbf{E}}^{-1/2} = \tilde{\mathbf{W}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{W}}^T \tag{22}$$

where $\tilde{\mathbf{\Lambda}} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n)$ with $\tilde{\lambda}_i \geq \tilde{\lambda}_{i+1}$.

From the original result of Anderson [35], the ML estimate of $\mathbf{H}$ is

$$\hat{\mathbf{H}} = \hat{\mathbf{E}}^{1/2}\tilde{\mathbf{W}}\tilde{\mathbf{D}}^{1/2}\mathbf{Q}, \tag{23}$$

where $\mathbf{Q}$ is an arbitrary orthogonal matrix and $\tilde{\mathbf{D}} = \text{diag}(d_1, \ldots, d_p, 0, \ldots, 0)$, with $d_i = \max(\tilde{\lambda}_i - 1, 0)$ and p the number of common factors. A consequence of this result is that the value of $\mathbf{H}\mathbf{H}^T$ that maximizes the likelihood is

$$\hat{\mathbf{H}}\hat{\mathbf{H}}^T = \hat{\mathbf{E}}^{1/2}\tilde{\mathbf{W}}\tilde{\mathbf{D}}\tilde{\mathbf{W}}^T\hat{\mathbf{E}}^{1/2}.$$

b) Step 2: Estimation of $\mathbf{G}$*:* In this step the channel-specific loading matrices, $\mathbf{G}_i$, for fixed $\mathbf{H}$ and $\boldsymbol{\Sigma}$, are estimated. The optimization problem is

$$(\mathcal{P}_2) \quad \underset{\mathbf{G}}{\text{maximize}} \quad \mathcal{L}\left(\hat{\mathbf{H}}, \mathbf{G}, \hat{\boldsymbol{\Sigma}}\right). \tag{24}$$

There is no closed-form solution to $(\mathcal{P}_2)$. Thus, we propose to apply an EM approach that is guaranteed to improve likelihood. Interestingly, we will show that the EM step amounts to removing the loading matrix for the common factors $\hat{\mathbf{H}}$ from the corresponding block in the diagonal of the sample covariance. Then apply Anderson's result [35].

Here is the idea. After marginalization with respect to the Gaussian factors $\mathbf{f}$, the second-order MVN model for the measurement $\mathbf{x}$ is $\mathbf{x} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{H}\mathbf{H}^T + \mathbf{G}\mathbf{G}^T + \boldsymbol{\Sigma})$. The problem is to find joint ML estimates for $\mathbf{H}, \mathbf{G}, \boldsymbol{\Sigma}$ in this model. We might say we started with the joint distribution for $\{\mathbf{X}, \mathbf{f}\}$, with $\mathbf{X} = [\mathbf{x}[1], \ldots, \mathbf{x}[N]]$ normally distributed, conditioned on $\mathbf{f}$, and $\mathbf{f}$ normal. The distribution of $\mathbf{f}$ is conjugate with respect to the conditional distribution of $\mathbf{X}$, so the marginalization of the joint distribution of $\{\mathbf{X}, \mathbf{f}\}$ is easy, producing the second-order normal distribution $\mathcal{N}_n(\mathbf{0}, \mathbf{H}\mathbf{H}^T + \mathbf{G}\mathbf{G}^T + \boldsymbol{\Sigma})$. But the maximization of the likelihood in this second-order model with respect to $\mathbf{G}$ is intractable, even in an alternating maximization with $\mathbf{H}$ and $\boldsymbol{\Sigma}$ fixed at their most recent estimates.

So, we replace the model $\mathcal{N}_n(\mathbf{0}, \mathbf{H}\mathbf{H}^T + \mathbf{G}\mathbf{G}^T + \boldsymbol{\Sigma})$ with the conditional first-order model $\mathcal{N}_n(\mathbf{H}\mathbf{f}, \mathbf{G}\mathbf{G}^T + \boldsymbol{\Sigma})$, treating $\mathbf{f}$ as an unmeasured random effect, and proceed with an EM algorithm in this first-order model to actually find the $\hat{\mathbf{G}}$ that maximizes likelihood in the second-order model $\mathbf{x} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{H}\mathbf{H}^T + \mathbf{G}\mathbf{G}^T + \boldsymbol{\Sigma})$, for fixed $\{\mathbf{H}, \boldsymbol{\Sigma}\}$.

To this end, let us begin by writing the log-likelihood function in (24) as

$$\mathcal{L}(\mathbf{G}) = \sum_{n=1}^{N} \log p(\mathbf{x}[n] | \mathbf{G}) \tag{25}$$

where we have omitted the notational dependence on $\hat{\mathbf{H}}$ and $\hat{\boldsymbol{\Sigma}}$, which are considered fixed in this step, and the equality is up to constant terms that do not modify the optimization problem. Let us denote the estimate of the loading matrix for the unique factors obtained in the previous iteration as $\hat{\mathbf{G}} = \text{blkdiag}[\hat{\mathbf{G}}_1, \ldots, \hat{\mathbf{G}}_M]$. The factors $\mathbf{f}[n]$ are considered unmeasured random effects, or hidden data, so the augmented data is $(\mathbf{x}[n], \mathbf{f}[n])$ [20], [22], [23]. For this choice of the augmented data set, in the E-step of the algorithm we get a lower-bound on $\mathcal{L}(\mathbf{G})$ as

$$\mathcal{L}(\mathbf{G}) \geq Q(\mathbf{G}, \hat{\mathbf{G}}) + \text{const.}, \tag{26}$$

where

$$Q(\mathbf{G}, \hat{\mathbf{G}}) = \sum_{n=1}^{N} E\left[\log p(\mathbf{x}[n] | \mathbf{f}[n], \mathbf{G})\right]. \tag{27}$$

The conditional distribution of $\mathbf{x}[n]$ given $\mathbf{f}[n]$ is

$$\mathbf{x}[n] | \mathbf{f}[n] \sim \mathcal{N}_n\left(\hat{\mathbf{H}}\mathbf{f}[n], \boldsymbol{\Psi}\right), \tag{28}$$

where we have defined $\boldsymbol{\Psi} = \mathbf{G}\mathbf{G}^T + \hat{\boldsymbol{\Sigma}}$. Then, the expectation in (27) becomes (up to constant terms)

$$E\left[\log p(\mathbf{x}[k] | \mathbf{f}[k], \mathbf{G})\right] = -\frac{1}{2}\log\det(\boldsymbol{\Psi}) - \frac{1}{2}\mathbf{x}[n]^T \boldsymbol{\Psi}^{-1}\mathbf{x}[n] + \mathbf{x}[n]^T \boldsymbol{\Psi}^{-1}\hat{\mathbf{H}}\hat{\mathbf{f}}[n] - \frac{1}{2}\text{tr}\left(\hat{\mathbf{H}}^T \boldsymbol{\Psi}^{-1}\hat{\mathbf{H}}\mathbf{C}[n]\right), \tag{29}$$

where

$$\hat{\mathbf{f}}[n] = E\left[\mathbf{f}[n] | \mathbf{x}[n]\right] = \mathbf{W}\mathbf{x}[n], \tag{30}$$

is the expected value of the factors, which is the minimum mean squared estimator (MMSE) of $\mathbf{f}[n]$ given $\mathbf{x}[n]$, and

$$\begin{aligned}\mathbf{C}[n] &= E\left[\mathbf{f}[n]\mathbf{f}[n]^T | \mathbf{x}[n]\right] \\ &= \left(\mathbf{I} + \hat{\mathbf{H}}^T \hat{\mathbf{E}}^{-1}\hat{\mathbf{H}}\right)^{-1} + \mathbf{W}\mathbf{x}[n]\mathbf{x}^T[n]\mathbf{W}^T,\end{aligned} \tag{31}$$

is the second order moment of the factors given the observations. In (30) and (31) the MMSE matrix $\mathbf{W}$ is

$$\mathbf{W} = \hat{\mathbf{H}}^T\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^T + \hat{\mathbf{E}}\right)^{-1}, \tag{32}$$

where

$$\hat{\mathbf{E}} = \hat{\mathbf{G}}\hat{\mathbf{G}}^T + \hat{\boldsymbol{\Sigma}}. \tag{33}$$

Plugging now (30) and (31) into (27) yields

$$\begin{aligned}Q(\mathbf{G}, \hat{\mathbf{G}}) &= -\frac{N}{2}\log\det(\boldsymbol{\Psi}) \\ &\quad - \frac{N}{2}\text{tr}\left[\boldsymbol{\Psi}^{-1}\left(\left(\mathbf{I} - 2\hat{\mathbf{H}}\mathbf{W}\right)\mathbf{S} + \hat{\mathbf{H}}\bar{\mathbf{C}}\hat{\mathbf{H}}^T\right)\right],\end{aligned} \tag{34}$$

where

$$\bar{\mathbf{C}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{C}[n] = \left(\mathbf{I} + \hat{\mathbf{H}}^T\hat{\mathbf{E}}^{-1}\hat{\mathbf{H}}\right)^{-1} + \mathbf{W}\mathbf{S}\mathbf{W}^T.$$

Exploiting the block-diagonal structure of $\boldsymbol{\Psi}$, $Q(\mathbf{G}, \hat{\mathbf{G}})$ can be written as

$$\begin{aligned}Q(\mathbf{G}, \hat{\mathbf{G}}) = -\frac{N}{2}\sum_{i=1}^{M}\Bigg[&\log\det\left(\mathbf{G}_i\mathbf{G}_i^T + \hat{\boldsymbol{\Sigma}}_i\right) \\ &+ \text{tr}\left(\left(\mathbf{G}_i\mathbf{G}_i^T + \hat{\boldsymbol{\Sigma}}_i\right)^{-1}\mathbf{T}_i\right)\Bigg],\end{aligned} \tag{35}$$

where $\mathbf{T}_i$ is the ith block of the appropriate dimensions in the diagonal of

$$\mathbf{T} = \left(\mathbf{I} - 2\hat{\mathbf{H}}\mathbf{W}\right)\mathbf{S} + \hat{\mathbf{H}}\bar{\mathbf{C}}\hat{\mathbf{H}}^T. \tag{36}$$

The following Lemma allows a further simplification of the expected log-likelihood function.

Lemma 1: The matrices $\mathbf{T} = \left(\mathbf{I} - 2\hat{\mathbf{H}}\mathbf{W}\right)\mathbf{S} + \hat{\mathbf{H}}\bar{\mathbf{C}}\hat{\mathbf{H}}^T$ and $\mathbf{P} = \mathbf{S} - \hat{\mathbf{H}}\hat{\mathbf{H}}^T$ are identical.

Proof: Using (23), $\mathbf{T}$ can be written as

$$\mathbf{T} = \hat{\mathbf{E}}^{1/2}\hat{\mathbf{W}}\boldsymbol{\Delta}\hat{\mathbf{W}}^T\hat{\mathbf{E}}^{1/2}, \tag{37}$$

where

$$\boldsymbol{\Delta} = \text{diag}\left(\delta_1, \ldots, \delta_p, \tilde{\lambda}_{p+1}, \ldots, \tilde{\lambda}_n\right), \tag{38}$$

with $\delta_j = \min(\tilde{\lambda}_j, 1)$. On the other hand, substituting $\hat{\mathbf{H}}\hat{\mathbf{H}}^T = \hat{\mathbf{E}}^{1/2}\tilde{\mathbf{W}}\tilde{\mathbf{D}}\tilde{\mathbf{W}}^T\hat{\mathbf{E}}^{1/2}$ into $\mathbf{P} = \mathbf{S} - \hat{\mathbf{H}}\hat{\mathbf{H}}^T$, we obtain

$$\begin{aligned}\mathbf{P} &= \hat{\mathbf{E}}^{1/2}\hat{\mathbf{W}}\left(\tilde{\mathbf{\Lambda}} - \tilde{\mathbf{D}}\right)\hat{\mathbf{W}}^T\hat{\mathbf{E}}^{1/2} \\ &= \hat{\mathbf{E}}^{1/2}\hat{\mathbf{W}}\mathbf{\Delta}\hat{\mathbf{W}}^T\hat{\mathbf{E}}^{1/2} = \mathbf{T},\end{aligned} \tag{39}$$

which proves the lemma.

□

From this result, we finally find that the expected log-likelihood function can be written as

$$\begin{aligned}Q(\mathbf{G}, \hat{\mathbf{G}}) = -\frac{N}{2}\sum_{i=1}^{M} &\log\det\left(\mathbf{G}_i\mathbf{G}_i^T + \hat{\mathbf{\Sigma}}_i\right) \\ &+ \operatorname{tr}\left[\left(\mathbf{G}_i\mathbf{G}_i^T + \hat{\mathbf{\Sigma}}_i\right)^{-1}\mathbf{P}_i\right],\end{aligned} \tag{40}$$

and

$$\mathbf{P}_i = \mathbf{S}_i - \hat{\mathbf{H}}_i\hat{\mathbf{H}}_i^T \tag{41}$$

where $\mathbf{S}_i$ is the sample covariance matrix of the ith channel and $\hat{\mathbf{H}}_i$ is the (fixed) loading matrix for the common factors derived in the previous step.

The interesting point from this derivation is that in the maximization step, the problem can be decoupled into M standard FA problems. To obtain the low-rank component of $\mathbf{S}_i - \hat{\mathbf{H}}_i\hat{\mathbf{H}}_i^T$ that models the loading matrix for the unique factors we proceed as follows. The whitened version of $\mathbf{P}_i$ is defined as $\tilde{\mathbf{P}}_i = \hat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{P}_i\hat{\mathbf{\Sigma}}_i^{-1/2}$, where $\hat{\mathbf{\Sigma}}_i$ is the fixed noise covariance matrix of the ith channel. Using the eigenvalue decomposition of $\tilde{\mathbf{P}}_i$, which is given by

$$\tilde{\mathbf{P}}_i = \tilde{\mathbf{W}}_i\tilde{\mathbf{\Lambda}}_i\tilde{\mathbf{W}}_i^T, \tag{42}$$

where $\tilde{\mathbf{\Lambda}}_i = \operatorname{diag}(\tilde{\lambda}_{i,1}, \ldots, \tilde{\lambda}_{i,n_i})$ with $\tilde{\lambda}_{i,j} \geq \tilde{\lambda}_{i,j+1}$, the value of $\mathbf{G}_i$ that optimizes the likelihood is found from the fundamental result of Anderson [35]:

$$\hat{\mathbf{G}}_i = \hat{\mathbf{\Sigma}}_i^{1/2}\tilde{\mathbf{W}}_i\tilde{\mathbf{D}}_i^{1/2}\mathbf{Q}_i. \tag{43}$$

The matrix $\mathbf{Q}_i$ is an arbitrary orthogonal matrix and $\tilde{\mathbf{D}}_i = \operatorname{diag}(d_{i,1}, \ldots, d_{i,p_i}, 0, \ldots, 0)$, with $d_{i,j} = \max(\tilde{\lambda}_{i,j} - 1, 0)$ and p_i the number of specific factors in the ith channel.

c) Step 3: Estimation of $\mathbf{\Sigma}$: The last step of the proposed algorithm is to estimate $\mathbf{\Sigma}$ as the solution to the optimization problem

$$(\mathcal{P}_3) \quad \underset{\mathbf{\Sigma}}{\text{maximize}} \quad \mathcal{L}\left(\hat{\mathbf{H}}, \hat{\mathbf{G}}, \mathbf{\Sigma}\right). \tag{44}$$

To solve $(\mathcal{P}_3)$ we take the constrained gradient of $\mathcal{L}(\hat{\mathbf{H}}, \hat{\mathbf{G}}, \mathbf{\Sigma})$ with respect to $\mathbf{\Sigma}$, a gradient that takes into account the diagonal structure of $\mathbf{\Sigma}$. This gradient can be found in [4], and is given by

$$\operatorname{diag}\left[\mathbf{R}^{-1}(\mathbf{R} - \mathbf{S})\mathbf{R}^{-1}\right]. \tag{45}$$

Now, setting (45) equal to zero, the ML estimate of $\mathbf{\Sigma}$ is the solution to

$$\operatorname{diag}\left[\mathbf{\Sigma}^{-1}(\mathbf{R} - \mathbf{S})\mathbf{\Sigma}^{-1}\right] = \mathbf{0}, \tag{46}$$

where we have applied the matrix inversion lemma to $\mathbf{R}$.

Algorithm 1 ML-MFA algorithm.

1: **Initialize**: $k = 0, \hat{\mathbf{\Sigma}}^{(0)} = \mathbf{I}_n$ and $\hat{\mathbf{G}}_i^{(0)} = \mathbf{0}_{n_i \times p_i}$
2: **repeat**
3: $k = k + 1$
4: Step 1: Estimate $\hat{\mathbf{H}}^{(k)}$ according to (23)
5: Cancel out the effect of the common factors using (41)
6: Step 2: Estimate $\hat{\mathbf{G}}_i^{(k)}$ following (43)
7: Step 3: Estimate $\hat{\mathbf{\Sigma}}_i^{(k)}$ as in (48)
8: **until** convergence

The solution is

$$\hat{\mathbf{\Sigma}} = \operatorname{diag}\left(\mathbf{S} - \hat{\mathbf{H}}\hat{\mathbf{H}}^T - \hat{\mathbf{G}}\hat{\mathbf{G}}^T\right), \tag{47}$$

or, equivalently,

$$\hat{\mathbf{\Sigma}}_i = \operatorname{diag}\left(\mathbf{P}_i - \hat{\mathbf{G}}_i\hat{\mathbf{G}}_i^T\right). \tag{48}$$

The non-negativity of the diagonal elements of $\mathbf{\Sigma}$ has not been imposed. However, taking into account the solution for $\hat{\mathbf{G}}$, it is easy to show that the elements of $\hat{\mathbf{\Sigma}}_i$ are positive.

C. Initialization and convergence

This algorithm for ML multi-channel Factor Analysis, or ML-MFA, is initialized at $\hat{\mathbf{\Sigma}}^{(0)} = \mathbf{I}_n$ and $\hat{\mathbf{G}}_i^{(0)} = \mathbf{0}_{n_i \times p_i}$, respectively. A smarter initialization of $\mathbf{\Sigma}$, which could achieve faster convergence for small signal-to-noise ratios [4], is $\hat{\mathbf{\Sigma}}^{(0)} = \operatorname{diag}(\mathbf{S})$. Then, the ML-MFA algorithm is presented in Algorithm 1.

In the theorem to follow, Theorem 1, we prove the convergence of the proposed algorithm to a local maximum of (19).

Theorem 1: Denote by

$$\{\hat{\mathbf{\Theta}}^{(k)}\} = \{\hat{\mathbf{H}}^{(k)}, \hat{\mathbf{G}}^{(k)}, \hat{\mathbf{\Sigma}}^{(k)}\} \tag{49}$$

the sequence of solutions generated by Algorithm 1. This sequence converges to a stationary point $\hat{\mathbf{\Theta}}^*$ of (19). That is, the algorithm converges to a local maximum of the second-order MVN likelihood function of (19).

Proof: It is easy to see that Problems $(\mathcal{P}_1)$ and $(\mathcal{P}_3)$ have closed-form solutions that do not decrease the value of the likelihood function. Moreover, [36] shows that, under mild conditions, a step of the EM algorithm does not decrease the likelihood. Then, since the set of structured covariance matrices given by (13) with finite trace is closed and bounded, and the likelihood function is bounded above, the sequence $\{\hat{\mathbf{\Theta}}^{(k)}\}$ converges to a limit point $\hat{\mathbf{\Theta}}^*$, which is a stationary point of (19). □

IV. Numerical results

A. Demonstrating convergence

In the first example, we demonstrate the convergence of the proposed alternating optimization method by considering $M = 3$ channels of dimensions $n_1 = 20, n_2 = 15$, and $n_3 = 10$. The number of observations is $N = 100$, and the number of common and unique factors are, respectively, $p = 2$ and $p_1 =$

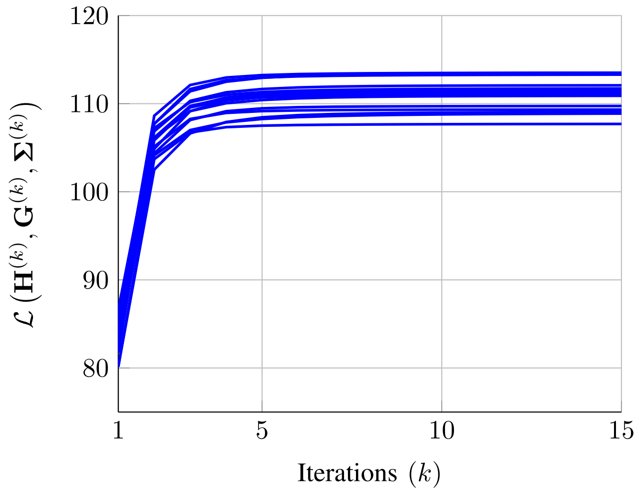


Fig. 2: Convergence of the ML-MFA algorithm.

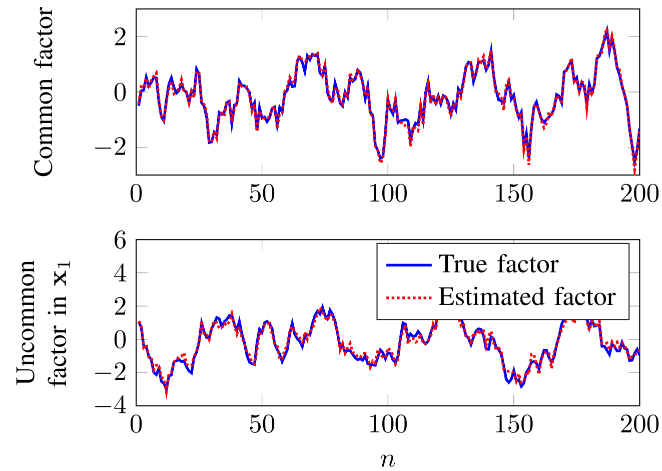


Fig. 3: Estimated and true factors.

$4, p_2 = 3$, and $p_3 = 2$. Moreover, the power ratio explained by the common, unique, and noise components for the ith channel with respect to the total power are given by

$$\eta_c = \operatorname{tr}(\mathbf{H}_i\mathbf{H}_i^T)/\operatorname{tr}(\mathbf{R}_i) = 0.3, \tag{50}$$

$$\eta_s = \operatorname{tr}(\mathbf{G}_i\mathbf{G}_i^T)/\operatorname{tr}(\mathbf{R}_i) = 0.3, \tag{51}$$

$$\eta_n = \operatorname{tr}(\mathbf{\Sigma}_i)/\operatorname{tr}(\mathbf{R}_i) = 0.4, \tag{52}$$

where $\mathbf{R}_i$ is the covariance matrix of the ith channel:

$$\mathbf{R}_i = \mathbf{H}_i\mathbf{H}_i^T + \mathbf{G}_i\mathbf{G}_i^T + \mathbf{\Sigma}_i. \tag{53}$$

Note, that for simplicity, the power ratios for all channels are identical. It is straightforward to extend the model to unequal power ratios.

The results for this example are shown in Figure 2, where the convergence curves for 15 runs of the proposed method are plotted. The loading and covariance matrices are randomly generated. That is, the model is different in each run. Consequently the value of likelihood achieved varies from run-to-run.

B. Estimating the common and unique factors

In the second example, the identification of the composite covariance matrix for all channels is used in uncoupled MMSE estimates of common and unique factors:

$$\hat{\mathbf{f}}[n] = \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\mathbf{x}[n], \tag{54}$$

$$\hat{\mathbf{f}}_i[n] = \hat{\mathbf{G}}_i^T\hat{\mathbf{R}}_i^{-1}\mathbf{x}_i[n], \tag{55}$$

where $\hat{\mathbf{R}}_i$ is the ith block of $\hat{\mathbf{R}}$. The results are shown in Figure 3 for an experiment with $p = p_i = 1$ factors, which are now AR(1) signals,[2] and $N = 1000$. The remaining parameters of the measurement model are those in the previous example. Note that in this example $p = p_i = 1$ factors are considered to avoid the subspace ambiguity, which is now reduced only to a sign change. As can be seen in Figure 3, the estimated factors in this scenario are nearly identical to the true factors.

[2]The temporal correlation induced by the AR(1) model is only used for visualization purposes and not exploited in the estimation algorithm, which still considers independent and identically distributed observations.

C. Mean-squared error of the estimated covariance model

The next example compares the performance of the MFA method and the group factor analysis (GFA) model proposed in [28]. GFA learns a structured sparse FA model so that the factor loading matrix is group-wise sparse. Sparsity in GFA is enforced by assuming independent gamma distributions as the precision parameter of the prior distribution for the elements of the loading matrix, and approximate inference is performed using the mean-field variational approximation. A final point to mention is that, while MFA needs an estimate of the number of common and unique factors, GFA only needs to know the total number of factors, and the variational optimization procedure finds the most adequate group-wise sparse structure for the multi-channel loading matrix. As a figure of merit, we use the normalized mean-squared error in the estimate of $\mathbf{R}$, which is defined as

$$\text{NMSE} = E\left[\frac{\left\|\mathbf{R} - \hat{\mathbf{R}}\right\|_F^2}{\|\mathbf{R}\|_F^2}\right],$$

estimated by averaging 1000 Monte Carlo trials for each value of N.

We generate data according to the proposed MFA model with $M = 3$ channels of dimensions $n_1 = 20, n_2 = 15$, and $n_3 = 10$. The number of common factors is $p = 2$ and the number of unique factors in each channel are $p_1 = 4, p_2 = 3$, and $p_3 = 2$. The power ratios are

$$\eta_c = 0.3, \qquad \eta_s = 0.3, \qquad \eta_n = 0.4.$$

Fig. 4 shows the NMSE for the MFA and the GFA models as a function of the sample size N. For the MFA we use the correct number of common and unique factors, while for the GFA we use the correct number of total factors $p + p_1 + p_2 + p_3 = 11$. As Fig. 4 shows, the gain obtained by enforcing the right sparsity structure in the composite loading matrix (cf. Fig. 1) increases with the number of samples.

This is admittedly a rigged experiment, as the measurements are generated from a model matched to the MFA structure assuming that the exact number of common factors, p, and

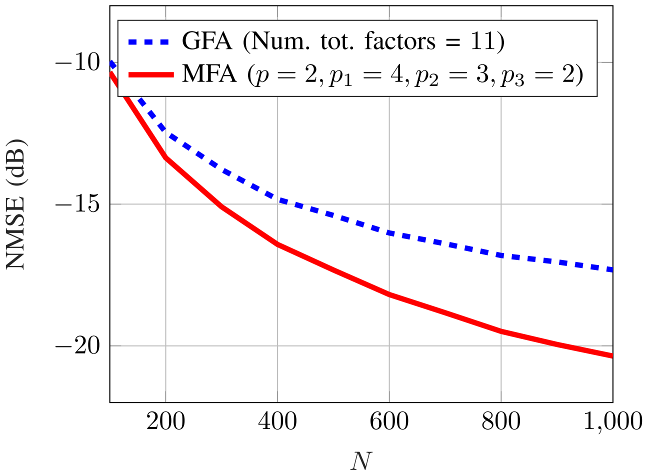
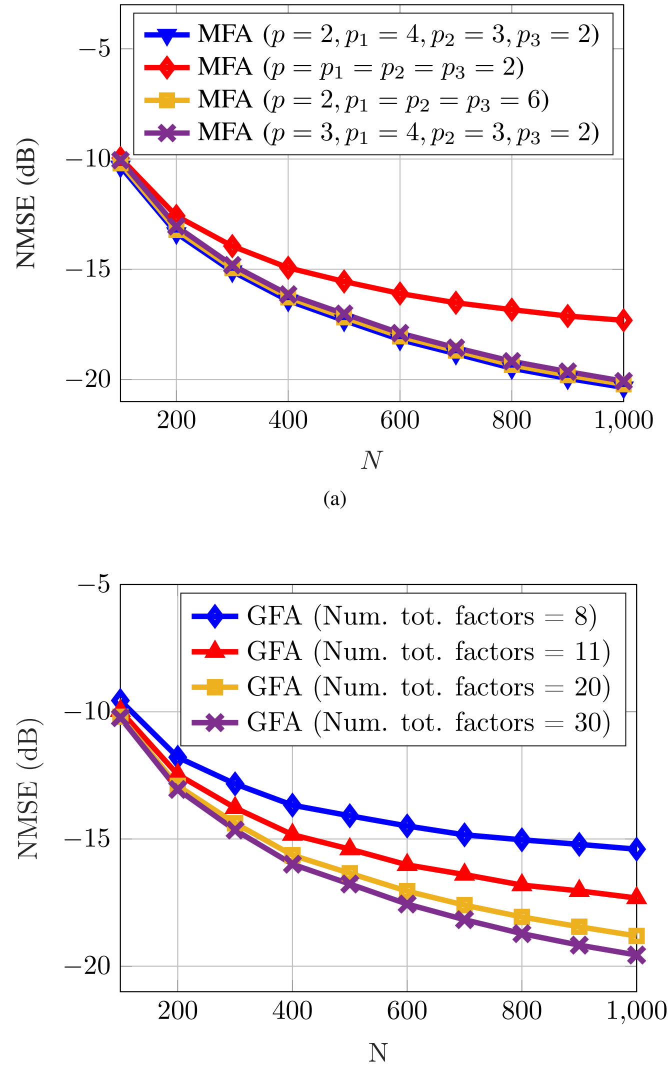


Fig. 4: NMSE in the estimate of $\mathbf{R}$ for the MFA and GFA models.

the exact number of unique factors, p_i, are known. In the next example we evaluate the NMSE performance of MFA and GFA against mismatched models. Let us recall that the true number of common factors is $p = 2$, and the unique factors for the 3 channels are $p_1 = 4$, $p_2 = 3$, and $p_3 = 2$, respectively. Fig. 5a shows that the performance of MFA is rather insensitive to an overestimation of either the number of common or unique factors. However, MFA is sensitive to underestimation of the number of unique factors. The GFA model experiences a similar behavior, as Fig. 5b shows: it is robust against an overestimation of the total number of factors present in the true model, but sensitive to under-estimation of the number of factors. In fact, this example suggests that GFA is more sensitive to model order underestimation than is MFA.



(a)

(b)

Fig. 5: Robustness of a) MFA and b) GFA against mismatched models.

D. Application to Passive Radar

A passive radar is equipped with both surveillance and reference antenna arrays [37]. The detection problem is to test $\mathcal{H}_1$: target present vs $\mathcal{H}_0$: target absent:

$$\begin{aligned} \mathcal{H}_1 &: \begin{cases} \mathbf{x}_1[n] = \mathbf{H}_1\mathbf{f}[n] + \mathbf{G}_1\mathbf{f}_1[n] + \mathbf{e}_1[n], \\ \mathbf{x}_2[n] = \mathbf{H}_2\mathbf{f}[n] + \mathbf{G}_2\mathbf{f}_2[n] + \mathbf{e}_2[n], \end{cases} \\ \mathcal{H}_0 &: \begin{cases} \mathbf{x}_1[n] = \mathbf{G}_1\mathbf{f}_1[n] + \mathbf{e}_1[n], \\ \mathbf{x}_2[n] = \mathbf{H}_2\mathbf{f}[n] + \mathbf{G}_2\mathbf{f}_2[n] + \mathbf{e}_2[n]. \end{cases} \end{aligned} \tag{56}$$

Here, $\mathbf{x}_1[n]$ and $\mathbf{x}_2[n]$ are respectively the surveillance and reference observations, $\mathbf{f}[n]$ is the unknown signal transmitted by the opportunistic illuminators, and $\mathbf{H}_1$ and $\mathbf{H}_2$ correspond to the channels between the illuminators and the surveillance and reference antennas. The factor $\mathbf{f}[n]$ is common when there is a target present to reflect the direct path signal, and the scanning surveillance channel comes into synchrony with the reference channel. The factors $\mathbf{f}_1[n]$ and $\mathbf{f}_2[n]$, and their channels $\mathbf{G}_1$ and $\mathbf{G}_2$, model the local interference at the surveillance and reference antenna arrays. Local interference in the surveillance channel models the direct path signal to the surveillance channel, and local interference in the reference channel allows for the modeling of multipath from the transmitter. We assume that the number of common and unique factors is known, which is not unrealistic for this application.

In [21], the model in (56) has been studied under different assumptions on the composite covariance matrix for the surveillance and direct channels. One of these assumptions is that there is no channel specific interference in the surveillance and reference channels, and that the covariances for the noises $\mathbf{e}_1$ and $\mathbf{e}_2$ are positive definite, but not diagonal. In this case, [21] derived the generalized likelihood ratio test (GLRT),

$$T(\mathbf{x}[1], \ldots, \mathbf{x}[N]) = \prod_{i=1}^{p} \frac{1}{1 - k_i^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta, \tag{57}$$

where η is a properly selected threshold and k_i is the ith canonical correlation between the surveillance and reference channels. That is, k_i is the ith singular value of $\mathbf{C} = \mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}$, with $\mathbf{S}_{ij}$ the ijth block of $\mathbf{S}$. The statistic may be termed a coherence statistic, as $\mathbf{C}$ is a coherence matrix. The statistic $1 - 1/T$ makes the coherence interpretation more clear:

$$1 - 1/T = 1 - \prod_{i=1}^{p}(1 - k_i^2). \tag{58}$$

Let us compare the GLRT in (57) with the GLRT statistic for the problem (56)

$$\mathscr{G}(\mathbf{x}[1], \ldots, \mathbf{x}[N]) = \frac{\max\limits_{\mathbf{H}_1,\mathbf{H}_2,\mathbf{G},\mathbf{\Sigma}} l(\mathbf{H}_1, \mathbf{H}_2, \mathbf{G}, \mathbf{\Sigma})}{\max\limits_{\mathbf{H}_2,\mathbf{G},\mathbf{\Sigma}} l(\mathbf{H}_2, \mathbf{G}, \mathbf{\Sigma})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{59}$$

Here $l(\cdot)$ is the Gaussian likelihood function. The maximization in the numerator is identical to that in Section III and we can therefore solve it using the ML-MFA algorithm. Under $\mathcal{H}_0$, the measurement model is

$$\begin{aligned} \mathbf{x}_1[n] &= \mathbf{G}_1\mathbf{f}_1[n] + \mathbf{e}_1[n], \\ \mathbf{x}_2[n] &= \mathbf{H}_2\mathbf{f}[n] + \mathbf{G}_2\mathbf{f}_2[n] + \mathbf{e}_2[n], \end{aligned} \tag{60}$$

which is equivalent to two (independent) FA problems. Thus, the computation of the generalized likelihood under the null hypothesis $\mathcal{H}_0$ may be carried out by solving two FA problems as in Section II.

To evaluate the performance of the statistics $\mathscr{G}$ and T, let us construct the following experiment. The noise covariance matrices are generated as $\mathbf{\Sigma}_i = \mathrm{diag}(\sigma_{i,1}^2, \ldots, \sigma_{i,n_i}^2)$ with $\sigma_{i,j}^2$ uniformly distributed between 0 and 1, and the elements of the common and uncommon loading matrices are generated as independent complex normals with zero mean and unit variance; the common loading matrices are scaled to achieve the desired SINR. The surveillance and reference channels are both equipped with $n_i = 10$ antennas, the number of antennas at the illuminator is $p = 3$, the interferers have both $p_i = 1$ antenna, and the number of available samples is $N = 200$. The results for this scenario are shown in Figure 6. This figure shows the probability of missed detection (p_m) for fixed probability of false alarm $p_{fa} = 10^{-3}$ and varying signal-to-interference-plus-noise ratio (SINR), which is defined as

$$\text{SINR (dB)} = 10\log_{10}\left(\frac{\mathrm{tr}(\mathbf{H}_i\mathbf{H}_i^H)}{\mathrm{tr}(\mathbf{G}_i\mathbf{G}_i^H + \mathbf{\Sigma}_i)}\right). \tag{61}$$

As we can see, the proposed detector of (59) outperforms the detector T of (57) because it exploits the additional structure induced by the low-rank interferers, which is to say the statistic $\mathscr{G}$ is matched to the measurement model and the statistic T is mismatched.

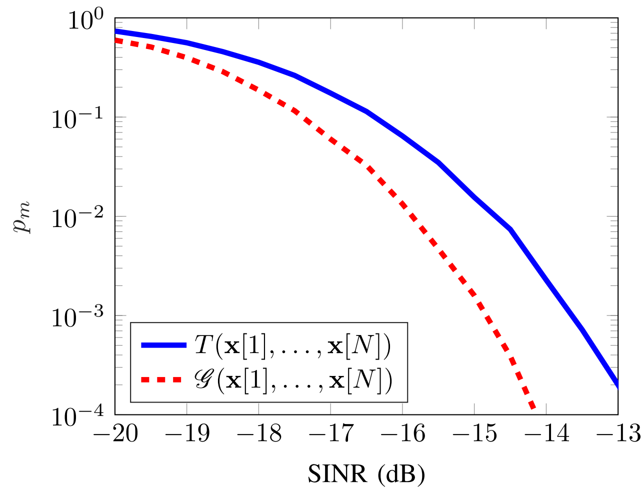


Fig. 6: Probability of missed detection (p_m) for varying SINR and fixed probability of false alarm $p_{fa} = 10^{-3}$.

V. Conclusions

This paper reports an extension to factor analysis (FA) for several MIMO channels that share factors and also contain unique factors. One important application of these results is to the problem of target detection in passive radar. Compared to other multi-channel generalizations of FA, such as inter-battery FA, the model proposed in this paper allows for shared *and* unique factors in each channel. The net of this model is to produce a multivariate Gaussian distribution for the set of MIMO channels in which a composite covariance matrix is structured in a very special way. The maximum likelihood problem is to identify this structured covariance matrix from a sequence of multi-channel measurements. Since there is no closed form solution, we report an iterative algorithm, consisting of a sequence of three steps, all of which increase the log-likelihood function. The performance of the algorithm is demonstrated with numerical experiments on illustrative problems. In the theory developed here, the number of factors must be known for each channel, suggesting further refinements for order determinations in each channel.

References

[1] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, pp. 72–101, 1904.

[2] B. S. Everitt, *An Introduction to Latent Variable Models*. Chapman and Hall, London, 1984.

[3] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. New York: Academic, 1979.

[4] D. N. Lawley and A. E. Maxwell, *Factor Analysis as a Statistical Method*. American Elsevier Pub. Co., 1971.

[5] A.-J. Boonstra and A.-J. Van der Veen, "Gain calibration methods for radio telescope arrays," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 25–38, 2003.

[6] A. J. van der Veen, A. Leshem, and A. J. Boonstra, "Array signal processing for radio astronomy," *Experimental Astronomy*, vol. 17, no. 1, pp. 231–249, Jun. 2004.

[7] D. Ramírez, G. Vázquez-Vilar, R. López-Valcarce, J. Vía, and I. Santamaría, "Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, Aug. 2011.

[8] M. Pesavento and A. B. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1310–1324, Jul. 2001.

[9] B. Friedlander and A. J. Weiss, "Direction finding using noise covariance modeling," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1557–1567, Jul. 1995.

[10] Q. Wu and K. M. Wong, "UN-MUSIC and UN-CLE: Aan application of generakized correlation analysis to the estimation of the directionsl of arrival in unknown correlated noise," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2331–2343, Sept. 1994.

[11] T. Z. Kou and A. A. L. Beex, "Parameter identification of exponential signals in colored environments," in *IFAC Symposium on Identification and System Parameters Identification*, Beijing, China, 1988, pp. 1023–1028.

[12] B. Peeters and G. D. Roeck, "Reference-based stochastic subspace identification for output-only modal analysis," *Mech. Systems and Signal Process.*, vol. 13, no. 6, pp. 855–878, 1999.

[13] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. on Acoustics, Speech and Signal Process.*, vol. ASSP-33, no. 2, p. 387, Apr. 1985.

[14] P. Stoica and M. Cedervall, “Detection tests for array processing in unknown correlated noise fields,” *IEEE Trans. Signal Process.*, vol. 45, no. 9, pp. 2351–2362, Sep. 1997.

[15] A. Leshem and A.-J. Van der Veen, “Multichannel detection of Gaussian signals with uncalibrated receivers,” *IEEE Signal Process. Lett.*, vol. 8, no. 4, pp. 120–122, Apr. 2001.

[16] E. Byrne and P. Schniter, “Adaptive detection of structured signals in low-rank interference,” Tech. Rep., 2018.

[17] A. M. Sardarabadi and A. J. van der Veen, “Complex factor analysis and extensions,” *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 954–967, Feb. 2018.

[18] K. G. Joreskog, “Testing a simple structure hypotheses in factor analysis,” *Psychometrika*, vol. 31, pp. 165–178, 1966.

[19] ——, “Some contributions to maximum likelihood factor analysis,” *Psychometrika*, vol. 32, pp. 443–482, 1967.

[20] J.-H. Zhao, P. L. H. Yu, and Q. Jiang, “ML estimation for factor analysis: EM or non-EM?” *Stat. Comp.*, vol. 18, no. 2, pp. 109–123, 2008.

[21] I. Santamaría, L. Scharf, J. Via, Y. Wang, and H. Wang, “Passive detection of correlated subspace signals in two MIMO channels,” *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1752–1764, Mar. 2017.

[22] D. Rubin and D. Thayer, “EM algorithms for ML factor analysis,” *Psychometrika*, vol. 47, no. 1, pp. 69–76, 1982.

[23] Z. Ghahramani and G. E. Hinton, “The EM algorithm for mixtures of factor analyzers,” in *Technical Report CRG-TR-96-1, University of Toronto*, 1996.

[24] L. R. Tucker, “An inter-battery method of factor analysis,” *Psychometrika*, vol. 23, no. 2, pp. 111–135, 1958.

[25] M. W. Browne, “The maximum likelihood solution in inter-battery factor analysis,” *Bristish Journal Math. Stat. Psyc.*, vol. 32, pp. 75–86, 1979.

[26] R. P. McDonald, “Three common factor models for groups of variables,” *Psychometrika*, vol. 35, no. 1, pp. 111–128, Mar. 1970.

[27] M. W. Browne, “Factor analysis of multiple batteries by maximum likelihood,” *Bristish Journal Math. Stat. Psyc.*, vol. 33, no. 2, pp. 184–199, Nov. 1980.

[28] A. Klami, S. Virtanen, E. Leppäaho, and S. Kaski, “Group factor analysis,” *IEEE Trans. Neural Netw. and Learning Syst.*, vol. 26, no. 9, pp. 2136–2147, Sept. 2015.

[29] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, “Parallel factor analysis in sensor array processing,” *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, 2000.

[30] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adali, “Independent vector analysis: Identification conditions and performance bounds,” *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, 2014.

[31] A. Shapiro, “Identifiability of factor analysis: Some results and open problems,” *Linear Algebra Appl.*, vol. 70, pp. 1–7, 1985.

[32] P. A. Bekker and J. M. F. ten Berge, “Generic global identification in factor analysis,” *Linear Algebra Appl.*, vol. 264, pp. 255–263, 1997.

[33] J. R. Kettenring, “Canonical analysis of several sets of variables,” *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[34] J. Vía, I. Santamaría, and J. Pérez, “A learning algorithm for adaptive canonical correlation analysis of several data sets,” *Neural Networks*, vol. 20, no. 1, pp. 139–152, Jan. 2007.

[35] T. W. Anderson, “Asymptotic theory for principal component analysis,” *Ann. Math. Statist.*, vol. 34, no. 1, pp. 122–148, Mar. 1963.

[36] C. F. J. Wu, “On the convergence properties of the EM algorithm,” *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

[37] D. E. Hack, L. K. Patton, B. Himed, and M. A. Saville, “Detection in passive MIMO radar networks,” *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2999–3012, 2014.