An alternating optimization algorithm for two-channel factor analysis with common and uncommon factors

David Ramírez
University Carlos III of Madrid, Spain
Gregorio Marañón Health Research Institute, Spain

Ignacio Santamaria and
Steven Van Vaerenbergh
University of Cantabria, Spain

Louis L. Scharf
Colorado State University, USA

Abstract—An alternating optimization algorithm is presented and analyzed for identifying low-rank signal components, known in factor analysis terminology as common factors, that are correlated across two multiple-input multiple-output (MIMO) channels. The additive noise model at each of the MIMO channels consists of white uncorrelated noises of unequal variances plus a low-rank structured interference that is not correlated across the two channels. The low-rank components at each channel represent uncommon or channel-specific factors.

Keywords—Factor Analysis, Expectation-Maximization, Maximum Likelihood, MIMO channels, Multivariate Normal Model.

I. Introduction

Standard second-order factor analysis aims to identify a low-rank plus diagonal covariance model from multiple realizations of a multivariate measurement. The interpretation is that the low-rank covariance component models a set of common factors that are mapped to measurements through a factor loading matrix [1]–[3]. The diagonal covariance component models additive noise, or errors in measurements. This problem may be extended to multiple channels of multivariate measurements, as in the inter-battery work of Tucker [4] and Browne [5].

In [6], multi-channel factor analysis has been applied to the problem of detecting correlation between two MIMO channels, and applied to passive radar. Several models of the additive noise are analyzed in [6], one of them leading to canonical correlation analysis (CCA). In this paper, we further generalize these results by allowing for low-rank plus diagonal correlation models for the additive noise in two MIMO channels. The resulting problem of identifying common factors between the two MIMO channels leads to a maximum likelihood problem. This problem is solved with an alternating optimization algorithm, consisting of three steps per iteration. Two of the steps have closed form solutions, and the other uses an Expectation-Maximization (EM) step. Iteration of these steps produces a convergent improvement algorithm. Simulation results demonstrate the efficacy of the method for identifying and tracking both common and uncommon factors.

II. Two-channel Factor Analysis

A. Single-channel FA

In the classical single-channel factor analysis (FA) problem [1]–[3], an n-dimensional real-valued data vector $\mathbf{x}$ is modeled as

$$\mathbf{x} = \mathbf{Hf} + \mathbf{e},$$

where $\mathbf{f}$ is a p-dimensional vector of factors (usually p is much smaller than n), $\mathbf{H} \in \mathbb{R}^{n \times p}$ is the factor loading matrix, and the n-dimensional error vector $\mathbf{e}$ is distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal matrix. That is, the observed variables are independent given the factors. In addition, the factors $\mathbf{f}$ are assumed to be distributed as $\mathcal{N}(\mathbf{0}, \mathbf{I})$. According to this model, the observations $\mathbf{x}$ follow a multivariate normal distribution with zero mean and covariance matrix $\mathbf{R} = \mathbf{HH}^T + \boldsymbol{\Sigma}$: a rank-$p$ matrix plus a diagonal matrix. This second-order model is invariant to the transformation $\mathbf{H} \longrightarrow \mathbf{HQ}\gamma, \mathbf{f} \longrightarrow \gamma^{-1}\mathbf{Q}^T\mathbf{f}$, with $\mathbf{Q}$ orthogonal. Any inference procedure to identify $\mathbf{H}$ is therefore a procedure to identify an equivalence class of frames, or a subspace, on a Grassmannian manifold. Notice also that forcing the noise variances to be identical reduces the model to probabilistic or Bayesian principal component analysis (PCA) [7].

Under a multivariate normal model for the factors and the noises, a maximum likelihood estimate of $\mathbf{R}$ is typically sought. Since a closed-form solution for this problem does not exist, various iterative procedures have been proposed to this end, including steepest descent [3], alternating optimization [8] and Expectation-Maximization algorithms [9].

B. Two-channel FA

We now describe a two-channel factor analysis (2FA) problem with both common and uncommon factors. To simplify the exposition we consider the case of two real-valued channels that have the same dimension n. The extensions to the complex-valued multi-channel case, or to the case where the two channels have different input dimensions are being investigated.

The first work considering a two-channel FA problem is the so-called inter-battery method proposed by L. Tucker in the fifties [4]. The goal was to extract factors $\mathbf{f}$ common to two domains of variables (or batteries), according to the generative model

$$\mathbf{x}_1 = \mathbf{H}_1\mathbf{f} + \mathbf{e}_1,$$
$$\mathbf{x}_2 = \mathbf{H}_2\mathbf{f} + \mathbf{e}_2.$$

Tucker proposed a least squares solution that essentially performed the singular value decomposition of the cross-correlation matrix between the two data sets. In the original

inter-battery factor analysis problem, the covariances of the noise terms specific to each data set had no particular structure, i.e., $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma}_i$ being an arbitrary positive definite matrix for $i = 1, 2$. The connection between these problems and CCA was shown later by Browne in [5].

The generative model that we consider in this paper is

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{H}_1 \mathbf{f} + \mathbf{G}_1 \mathbf{f}_1 + \mathbf{e}_1, \\ \mathbf{x}_2 &= \mathbf{H}_2 \mathbf{f} + \mathbf{G}_2 \mathbf{f}_2 + \mathbf{e}_2, \end{aligned} \quad (1)$$

where $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{R}^{n \times p}$ are the loading matrices for the p common factors in vector $\mathbf{f}$; $\mathbf{f}_i$ are p_i-dimensional vectors containing the channel-specific factors; $\mathbf{G}_i \in \mathbb{R}^{n \times p_i}$ are the loading matrices for the specific factors, $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ are the noises with diagonal covariance matrices, and $E[\mathbf{e}_1 \mathbf{e}_2^T] = \mathbf{0}$. Without loss of generality, we also assume that both the common and uncommon factors are normalized such that $E[\mathbf{f} \mathbf{f}^T] = \mathbf{I}_p$, $E[\mathbf{f}_i \mathbf{f}_i^T] = \mathbf{I}_{p_i}$, and $E[\mathbf{f}_1 \mathbf{f}_2^T] = \mathbf{0}_{p_1 \times p_2}$. The difference between our generative model in (1) and Tucker's inter-battery FA model, or other group FA or sparse FA problems in the literature [10], is that we also consider a low-rank-plus-diagonal decomposition for the covariance matrix of the specific (unshared) factors, which allows us to model structured interference at each of the two channels.

Under a multivariate normal model for both the common and uncommon factors, the composite vector $\mathbf{x} = [\mathbf{x}_1^T \, \mathbf{x}_2^T]^T$ is distributed as $\mathcal{N}(\mathbf{0}, \mathbf{R})$ with a structured covariance matrix given by

$$\mathbf{R} = \mathbf{H}\mathbf{H}^T + \underbrace{\begin{bmatrix} \mathbf{G}_1 \mathbf{G}_1^T + \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \mathbf{G}_2^T + \boldsymbol{\Sigma}_2 \end{bmatrix}}_{\mathbf{E}}, \quad (2)$$

where the rank-p composite loading matrix is $\mathbf{H} = [\mathbf{H}_1^T \, \mathbf{H}_2^T]^T$. The composite matrices for the uncommon factors and noises are, respectively

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0}_{n \times p_2} \\ \mathbf{0}_{n \times p_1} & \mathbf{G}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & \boldsymbol{\Sigma}_2 \end{bmatrix}.$$

This covariance model is an element of the cone $\mathcal{R} = \{\mathbf{R} : \gamma \mathbf{R} \succ \mathbf{0}, \gamma > 0\}$.

C. ML approach

Given a set of M observations, $\mathbf{x}[1], \ldots, \mathbf{x}[M]$, the goal of 2FA is to find the common factor loading matrix $\mathbf{H}$, the channel specific loading matrices $\mathbf{G}$, and the diagonal noise covariance matrices $\boldsymbol{\Sigma}$ that best model the sample covariance of the observations, which is given by

$$\mathbf{S} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}[m] \mathbf{x}^T[m].$$

The ML estimates of $\boldsymbol{\Theta} = (\mathbf{H}, \mathbf{G}, \boldsymbol{\Sigma})$ are obtained by solving the following optimization problem

$$\underset{\boldsymbol{\Theta}}{\text{maximize}} \quad \mathcal{L}(\boldsymbol{\Theta}) = \log \det(\mathbf{R}^{-1} \mathbf{S}) - \text{tr}(\mathbf{R}^{-1} \mathbf{S}), \quad (3)$$

where $\mathbf{R}$ is given in (2). As $\mathbf{R} \in \mathcal{R}$, any $\hat{\mathbf{R}}$ that can compete for the title *maximum likelihood* satisfies the constraint $\text{tr}(\mathbf{R}^{-1} \mathbf{S}) = 2n$. So the maximum likelihood problem is to maximize the product of eigenvalues of $\mathbf{R}^{-1} \mathbf{S}$ under the constraint that their sum is $2n$ [6].

Like in single-channel FA using the ML method, Problem (3) does not admit a closed-form solution. Therefore, one needs to resort to numerical methods to maximize the log-likehood function like the one proposed in this paper.

III. ALTERNATING OPTIMIZATION ALGORITHM

In this section we propose a sequential alternating optimization approach to maximize the log-likehood function that, at each step, optimizes a subset of the 2FA variables while considering the other variables fixed. To clarify the notation, the fixed parameters at a particular iteration will be denoted with a hat, whereas the parameters without hat denote variables to be optimized. For instance, $\mathcal{L}(\hat{\mathbf{H}}, \mathbf{G}, \hat{\boldsymbol{\Sigma}})$ denotes the function to be maximized for fixed values of $\hat{\mathbf{H}}$ and $\hat{\boldsymbol{\Sigma}}$. A given iteration of the algorithm applies three sequential steps which are described in the following.

A. Step 1: Estimation of $\mathbf{H}$

In the first step of the optimization procedure, we consider fixed $\hat{\boldsymbol{\Sigma}}$ and $\hat{\mathbf{G}}$ and solve the problem

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

Given $\hat{\mathbf{E}} = \hat{\mathbf{G}} \hat{\mathbf{G}}^T + \hat{\boldsymbol{\Sigma}}$, the ML estimate of $\mathbf{H}\mathbf{H}^T$ can be obtained in closed-form using a result by Anderson [11] (cf. also Theorem 9.4.1 in [2] and [6]). Let us define the noise-whitened sample covariance matrix $\tilde{\mathbf{S}} = \hat{\mathbf{E}}^{-1/2} \mathbf{S} \hat{\mathbf{E}}^{-1/2}$ with eigenvalue (EV) decomposition $\tilde{\mathbf{S}} = \tilde{\mathbf{W}} \tilde{\boldsymbol{\Lambda}} \tilde{\mathbf{W}}^T$ where $\tilde{\boldsymbol{\Lambda}} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{2n})$ with $\tilde{\lambda}_i \geq \tilde{\lambda}_{i+1}$. The value of $\mathbf{H}\mathbf{H}^T$ that maximizes the likelihood is

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

where $\tilde{\mathbf{D}} = \text{diag}(d_1, \ldots, d_p, 0 \ldots 0)$, with $d_i = \max(\tilde{\lambda}_i - 1, 0)$ and p the number of common factors. Notice finally that an estimate of the loading matrix can be trivially obtained up to an arbitrary orthogonal matrix $\mathbf{Q}$ as

$$\hat{\mathbf{H}} = \hat{\mathbf{E}}^{1/2} \tilde{\mathbf{W}} \tilde{\mathbf{D}}^{1/2} \mathbf{Q}. \quad (4)$$

B. Step 2: Estimation of $\mathbf{G}$

In the second step of the alternating optimization approach we consider the estimation of the channel-specific loading matrices for fixed $\hat{\mathbf{H}}, \hat{\boldsymbol{\Sigma}}$. The problem to solve is

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

Unfortunately, the solution of $\mathcal{P}_2$ cannot be obtained in closed-form. As an alternative to get new estimates of $\{\hat{\mathbf{G}}_i\}$ that are guaranteed to increase the log-likelihood function, we use the Expectation-Maximization (EM) algorithm, which was originally proposed by Rubin and Thayer in [9] and later extended to mixtures of factor analyzers in [12].

Let us define the simplified model $\mathbf{x} = \mathbf{H}\mathbf{f} + \mathbf{u}$, where $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{E})$ combines the specific factors and noises, and $\mathbf{E}$ is therefore a block diagonal matrix. Given $\hat{\mathbf{H}}$ and a previous estimate of $\hat{\mathbf{E}}$ with blocks $\hat{\mathbf{E}}_i = \hat{\mathbf{G}}_i \hat{\mathbf{G}}_i^T + \hat{\boldsymbol{\Sigma}}_i$, the E-step

estimates the common factors $\hat{\mathbf{f}}[m]$ for each data point $\mathbf{x}[m]$ as follows

$$\hat{\mathbf{f}}[m] = E[\mathbf{f}|\mathbf{x}[m]] = \mathbf{W}\mathbf{x}[m],$$

where $\mathbf{W} = \hat{\mathbf{H}}^T \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^T + \hat{\mathbf{E}} \right)^{-1}$, and computes the correlation matrix of the factors as

$$\mathbf{C}[m] = E[\mathbf{f}\,\mathbf{f}^T|\mathbf{x}[m]] = \mathbf{I} - \mathbf{W}\hat{\mathbf{H}} + \mathbf{W}\mathbf{x}[m]\mathbf{x}^T[m]\mathbf{W}^T. \quad (5)$$

Using the estimate of Step 1, the M-step first cancels the effect of the common factors as

$$\mathbf{S}_e = \text{blkdiag}\left[\mathbf{S} - \hat{\mathbf{H}}\hat{\mathbf{H}}^T \right] = \begin{bmatrix} \mathbf{S}_{e,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{e,2} \end{bmatrix},$$

where $\text{blkdiag}(\cdot)$ creates a block-diagonal matrix with blocks $\{\mathbf{S}_{e,i}\}$ of dimension $n \times n$. Then, we next obtain new estimates $\{\hat{\mathbf{G}}_i\}$ from the blocks $\{\mathbf{S}_{e,i}\}$ applying again Anderson's result [11]. Repeating the procedure followed for the common loading matrix in Subsection III-A, we define

$$\tilde{\mathbf{S}}_{e,i} = \hat{\mathbf{\Sigma}}_i^{-1/2} \mathbf{S}_{e,i} \hat{\mathbf{\Sigma}}_i^{-1/2},$$

with eigenvalue decomposition $\tilde{\mathbf{S}}_{e,i} = \tilde{\mathbf{W}}_i \tilde{\mathbf{\Lambda}}_i \tilde{\mathbf{W}}_i^T$, $\tilde{\mathbf{\Lambda}}_i = \text{diag}(\tilde{\lambda}_{i,1}, \ldots, \tilde{\lambda}_{i,n})$ and $\tilde{\lambda}_{i,j} \geq \tilde{\lambda}_{i,j+1}$. The new estimate of $\mathbf{G}_i$ (up to a unitary rotation) is

$$\hat{\mathbf{G}}_i = \hat{\mathbf{\Sigma}}_i^{1/2} \tilde{\mathbf{W}}_i \tilde{\mathbf{D}}_i^{1/2}, \quad (6)$$

where $\tilde{\mathbf{D}}_i = \text{diag}\left(d_{i,1}, \ldots, d_{i,p_i}, 0 \ldots 0 \right)$, with $d_{i,j} = \max(\tilde{\lambda}_{i,j} - 1, 0)$ and p_i the number of specific factors in the ith channel.

We could apply several iterations of the EM algorithm to get a minimizer of $(\mathcal{P}_2)$. However, it is known that the EM-algorithm might have slow numerical convergence. For this reason in the proposed implementation we chose to apply a single EM iteration in Step 2 that ensures an increase in the likelihood.

C. Step 3: Estimation of $\mathbf{\Sigma}$

In the last step of our cyclic procedure we solve

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

The constrained gradient of $\mathcal{L}(\hat{\mathbf{H}}, \hat{\mathbf{G}}, \mathbf{\Sigma})$ with respect to $\mathbf{\Sigma}$ is given by [3]

$$\text{diag}\left[\mathbf{R}^{-1} \left(\mathbf{R} - \mathbf{S} \right) \mathbf{R}^{-1} \right]. \quad (7)$$

Setting (7) equal to zero and applying the matrix inversion lemma to $\mathbf{R}^{-1}$, the ML estimate of $\mathbf{\Sigma}$ is the solution to

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

which yields

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

or, equivalently,

$$\hat{\mathbf{\Sigma}}_i = \text{diag}\left(\mathbf{S}_{e,i} - \hat{\mathbf{G}}_i\hat{\mathbf{G}}_i^T \right).$$

Algorithm 1: ML-2FA algorithm.

Initialize: $k = 0, \hat{\mathbf{G}}^k = \mathbf{0}, \hat{\mathbf{\Sigma}}^k = \mathbf{I}$
repeat
  $k = k + 1$
  Step 1: Estimate $\hat{\mathbf{H}}^k$ according to (4)
  Cancel out the effect of the common factors
    $\mathbf{S}_{e,i} = \mathbf{S}_i - \hat{\mathbf{H}}_i\hat{\mathbf{H}}_i^T$
  Step 2: Estimate $\hat{\mathbf{G}}_i^k$ using (6)
  Step 3: Estimate $\hat{\mathbf{\Sigma}}_i^k$ as
    $\hat{\mathbf{\Sigma}}_i^k = \text{diag}\left(\mathbf{S}_{e,i} - \hat{\mathbf{G}}_i^k[\hat{\mathbf{G}}_i^k]^T \right)$
until *convergence*

D. Initialization and convergence

In this section we discuss the initialization and convergence of the proposed algorithm to which we refer as ML two-channel Factor Analysis or ML-2FA algorithm. Some implementation aspects of the proposed approach are also discussed.

The following theorem guarantees the convergence of the ML-2FA algorithm to a (possibly local) minimum of the original problem (3).

Theorem 1: Let $\{\hat{\mathbf{\Theta}}^k\} = \{\hat{\mathbf{H}}^k, \hat{\mathbf{G}}^k, \hat{\mathbf{\Sigma}}^k\}$ be a sequence generated by the ML-2FA Steps 1-3. Then, the sequence converges to a stationary point $\mathbf{\Theta}^*$ of the constrained optimization problem (3). That is, a (possibly local) maximum of the likelihood function is obtained.

Proof: Problems $\mathcal{P}_1$ and $\mathcal{P}_3$ have closed-form solutions that necessarily increase the global likelihood function. On the other hand, under mild conditions we can resort to the well-known convergence result by Wu [13] to show that the application of the EM-step also increases the likelihood. Since the feasible set (structured covariance matrices as in (2) with finite trace) is a closed and bounded set, and the likelihood function is bounded above, then the sequence $\{\hat{\mathbf{\Theta}}^k\}$ converges to a limit point $\mathbf{\Theta}^*$ which is a stationary point of the original likelihood function with the required structure. ∎

For the initialization of the ML-2FA algorithm we use the following procedure. The specific loading matrices are initialized as $\hat{\mathbf{G}}^0 = \mathbf{0}$, whereas the noise covariance matrices are initialized as $\hat{\mathbf{\Sigma}}^0 = \mathbf{I}$. An alternative to the initialization of $\mathbf{\Sigma}$, which results in faster convergence for small signal-to-noise ratios, is $\hat{\mathbf{\Sigma}}^0 = \text{diag}(\mathbf{S})$. The ML-2FA algorithm is finally summarized in Algorithm 1.

IV. SIMULATION RESULTS

We consider a toy example where the two-channel data is explained by 1 common factor and 1 specific factor for each channel ($p = p_1 = p_2 = 1$). The factors are generated independently as AR(1) signals. Although the factors for this example are not temporally white, this information is not exploited in our ML approach, which assumes i.i.d. observations. The proportion of the total power explained by the common, specific and noise components for the ith channel are given
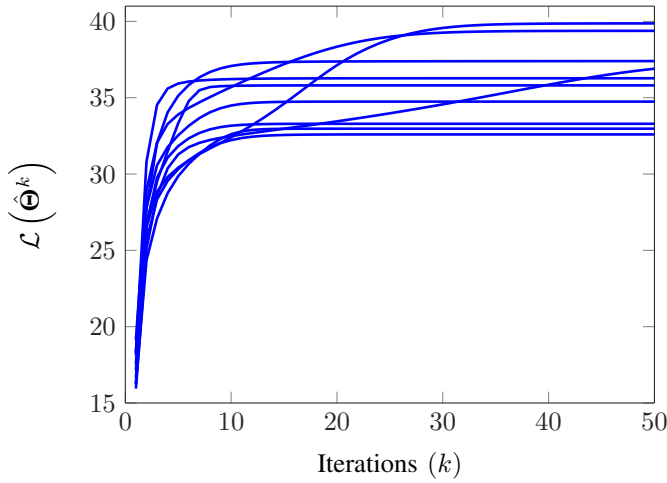
Fig. 1: Convergence of the ML-2FA algorithm.

Fig. 2: Estimated and true factors.

respectively by

$$\eta_c = \operatorname{tr}(\mathbf{H}_i \mathbf{H}_i^T) / \operatorname{tr}(\mathbf{R}_i),$$
$$\eta_s = \operatorname{tr}(\mathbf{G}_i \mathbf{G}_i^T) / \operatorname{tr}(\mathbf{R}_i),$$
$$\eta_n = \operatorname{tr}(\mathbf{\Sigma}_i) / \operatorname{tr}(\mathbf{R}_i),$$

where $\mathbf{R}_i = \mathbf{H}_i \mathbf{H}_i^T + \mathbf{G}_i \mathbf{G}_i^T + \mathbf{\Sigma}_i$ is the covariance matrix of the ith channel. In our example we used $\eta_c = 0.3$, $\eta_s = 0.3$ and $\eta_n = 0.4$ for the two channels. The dimensionality of channels $\mathbf{x}_1$ and $\mathbf{x}_2$ is 20 and 15, respectively; and the sample covariance is estimated from $M = 200$ observations.

Fig. 1 shows the convergence curves for 10 different runs of the ML-2FA algorithm. In each run we changed all model parameters. As we observe, for most of the runs the proposed algorithm converges to a stationary point of the likelihood function in less than 10 iterations.

The common and specific factors using the obtained estimates of the loading and noise covariance matrices can be estimated as

$$\hat{\mathbf{f}}[m] = (\mathbf{H}^*)^T \left(\mathbf{H}^* (\mathbf{H}^*)^T + \mathbf{E}^* \right)^{-1} \mathbf{x}[m],$$
$$\hat{\mathbf{f}}_i[m] = (\mathbf{G}_i^*)^T \left(\mathbf{G}_i^* (\mathbf{G}_i^*)^T + \mathbf{H}_i^* (\mathbf{H}_i^*)^T + \mathbf{\Sigma}_i^* \right)^{-1} \mathbf{x}_i[m],$$

respectively The results obtained in a simulation are shown in Fig. 2. For this example we only have one common and one uncommon factor at each MIMO channel, therefore, the factors can be recovered without ambiguity. In examples with multiple factors we would just recover a mixture of the factors by an orthogonal matrix.

It has been also assumed in this work that the number of factors is known. In practice, p, p_1 and p_2 should be estimated. To this end, we might use a similar method to the one proposed in [14]. The order-estimation problem for 2FA, as well as more realistic applications in wireless communications and industrial monitoring problems, will be considered in a future work.

V. CONCLUSIONS

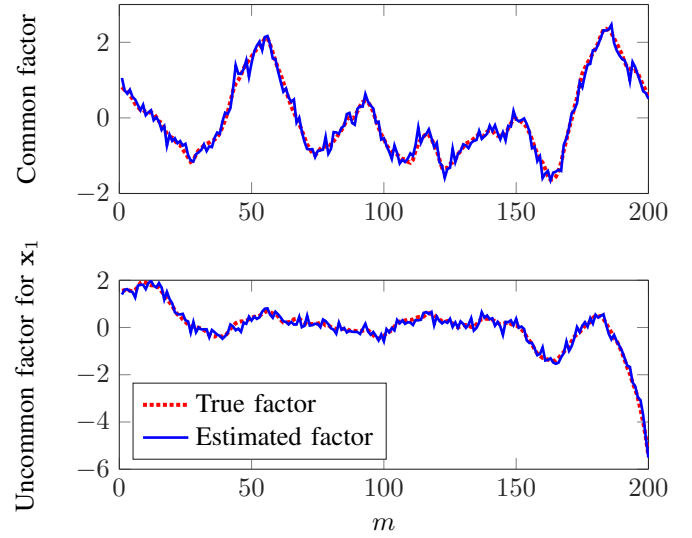A general model for measurements at two MIMO channels accounts for low-rank signal components in each channel that are correlated across the channels, low-rank interference components that are not correlated across the channels, and additive uncorrelated noises in each channel that have unequal variances across sensors in the MIMO channel. In this paper we have presented and analyzed an iterative optimization algorithm that identifies the common and uncommon low-rank signal components for this MFA problem.

REFERENCES

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

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

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

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

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

[6] I. Santamaria, 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.

[7] C. M. Bishop, "Bayesian PCA," in *Neural Inf. Process. Systems*, 1999, pp. 382–388.

[8] 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.

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

[10] 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.

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

[12] 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.

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

[14] Y. Song, P. Schreier, D. Ramírez, and T. Hasija, "Canonical correlation analysis of high-dimensional data with very small sample support," *Signal Process.*, vol. 128, pp. 449–458, Nov. 2016.