Multi-Channel Factor Analysis for Temporally and Spatially Correlated Time Series

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

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

Dongliang Duan
Electrical Engineering
University of Wyoming
Laramie, USA
dduan@uwyo.edu

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

2023 57th Asilomar Conference on Signals, Systems, and Computers | 979-8-3503-2574-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/IEEECONF59524.2023.10476743

***Abstract*—In Multi-Channel Factor Analysis (MFA), the *spatial* covariance of a multi-channel observation is decomposed into the covariances of latent signal, interference, and noise components. In proposed applications, the observations also have *temporal* correlations which may be of independent interest or may influence the spatial covariance estimates. An extension to MFA is proposed where the common and unique factor series are synthesized using LTI filters with unknown transfer functions. A novel block majorization-minimization procedure for semi-parametric estimation of both spatial and temporal correlations is summarized. Experiments show that the resulting technique for *spatio-temporal* correlation analysis of a multi-channel observation series improves on MFA when the factor series are time-dependent.**

***Index Terms*—factor analysis (FA), multi-channel factor analysis (MFA), space-time, time series**

I. Introduction

Multi-channel factor analysis (MFA) is a technique for the investigation of multi-channel observations that are composed of an unknown signal that is present across all channels, interferences that are confined to individual channels, and anisotropic idiosyncratic noises. In [1], MFA is introduced and several possible applications to communications and array processing are discussed. However, in many of the proposed applications, such as passive radar and multi-point cellular network processing, the observations exhibit *temporal* correlations in addition to the *spatial* correlations treated by MFA. As in [2], [3], the observations are *temporally* and *spatially* correlated time series.

The presence of substantial temporal correlations in the observation series alters the problem of detecting and estimating spatial correlations [4], [5]. Even if the spatial correlations are of primary interest, exploiting the temporal structure of the time series often leads to improved performance. The experiments of Section IV demonstrate this phenomenon in the context of multi-channel factor analysis.

In Section II, a semi-parametric extension to MFA for multi-channel vector-valued time series is proposed, where a non-parametric model for the temporal correlations of the signal and interferences is adjoined to the previous spatial correlation structure of MFA. A three-step block majorization-minimization (MM) procedure [6] is developed to estimate the second-order spatio-temporal structure of the observation series. This is accomplished in the frequency domain by optimizing a Gaussian log-likelihood objective in a fashion analogous to the method of [1] for MFA estimation with independent observations, but with substantial differences in the procedure details which are summarized in Section III. Finally, Section IV contains the results of simulated experiments which assess how the proposed modeling of temporal correlation improves performance.

Notation

Matrices and vectors are written with bold-face uppercase and lowercase symbols respectively, while scalars are denoted with light-face symbols. The identity matrix of size n is $\mathbf{I}_n$. For the matrix $\mathbf{D}$, $\mathbf{D}^\mathsf{T}$ and $\mathbf{D}^\mathsf{H}$ are its transpose and conjugate transpose respectively. 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 discrete Fourier transform of a series $\mathbf{x}[t],\ t \in \mathbb{Z}$ is $\mathcal{F}\{\mathbf{x}\}$ and the discrete convolution of series $\mathbf{x}[t]$ and $\mathbf{y}[t]$ is $(\mathbf{x} * \mathbf{y})[t]$. The function $\delta[t]$ is 1 if $t = 0$ and 0 otherwise.

II. Model

Suppose that a measurement apparatus consists of C heterogeneous sensors, which are presumed to capture differing aspects of some shared underlying phenomenon. At time $t \in \mathbb{Z}$, the cth sensor records n_c scalar inputs which are contained in the channel-c observation vector $\mathbf{x}_c[t] \in \mathbb{R}^{n_c}$. Each time-series of observations made by the C sensors is treated as a distinct *channel* of information, and relevant information about the shared phenomenon may be contained in the second-order *spatio-temporal* properties of the all-channel observation series.

A. Observation Model

The measurements in the cth channel are modeled using three latent series, namely the *signal* $\mathbf{s}_c[t] \in \mathbb{R}^{n_c}$, the interference $\mathbf{i}_c[t]$, and the idiosyncratic noise $\mathbf{u}_c[t]$. At each time, the values of these three latent series in channel c are summed to form the channel-c observation vector,

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

The all-channel observation series $\mathbf{x}[t]$ is obtained by stacking the observations in each in channels $c = 1, \ldots, C$ as

$$\mathbf{x}[t] \equiv \left[\mathbf{x}_1^\mathsf{T}[t] \ \ldots \ \mathbf{x}_C^\mathsf{T}[t]\right]^\mathsf{T}.$$

Similarly, $\mathbf{s}[t] \equiv [\mathbf{s}_1^\mathsf{T}[t] \ldots \mathbf{s}_C^\mathsf{T}[t]]^\mathsf{T}$, $\mathbf{i}[t] \equiv [\mathbf{i}_1^\mathsf{T}[t] \ldots \mathbf{i}_C^\mathsf{T}[t]]^\mathsf{T}$, and $\mathbf{u}[t] \equiv [\mathbf{u}_1^\mathsf{T}[t] \ldots \mathbf{u}_C^\mathsf{T}[t]]^\mathsf{T}$ are the all-channel latent series. The total number of scalar inputs is $n \equiv \sum_{c=1}^{C} n_c$. At time t, the series values $\mathbf{x}[t], \mathbf{s}[t], \mathbf{i}[t], \mathbf{u}[t]$ are $n \times 1$ vectors.

The core assumption of multi-channel factor analysis is that the signal and interferences live in different *types* of low-dimensional subspaces. At each time t, the all-channel signal $\mathbf{s}[t]$ lies within some subspace of the all-channel observation space $\mathbb{R}^n$, where the dimension of the signal subspace is known to be r_0 but the subspace itself is not known. The signal subspace arises from the sensors' relationship to the underlying phenomenon and so is assumed to be constant across time. If the columns of the *common factor loading matrix* $\mathbf{A} \in \mathbb{R}^{n \times r_0}$ are a basis for the signal space, then $\mathbf{s}[t]$ can be expressed as $\mathbf{Af}[t]$, where the *common factors* $\mathbf{f}[t]$ are a latent time-series of $r_0 \times 1$ vectors.

The channel-c interference $\mathbf{i_c}[t]$ similarly lies within a low-dimensional subspace whose dimension is known to be r_c but is otherwise unknown. However, the channel-c interference space is constrained to be a subspace of the observation space *for channel c only*. As for the signal space, the interference spaces are also taken to be constant across time. If the columns $\mathbf{B}_c \in \mathbb{R}^{n_c \times r_c}$ are a basis for the channel-c interference space, then $\mathbf{i}_c[t] = \mathbf{B}_c \mathbf{g}_c[t]$ where $\mathbf{g}_c[t] \in \mathbb{R}^{r_c}$ is the series of *unique factors* for channel c. To effect the orthogonality of the interference spaces, the *unique factor loading matrix* is $\mathbf{B} = \operatorname{blkdiag}(\mathbf{B}_1, \ldots, \mathbf{B}_C)$. The unique factor series are stacked into $\mathbf{g}[t] = [\mathbf{g}_1^\mathsf{T}[t] \ldots \mathbf{g}_C^\mathsf{T}[t]]^\mathsf{T}$ and so the all-channel interference series can be written as $\mathbf{i}[t] = \mathbf{Bg}[t]$. With these definition, the all-channel observation series is expressed as

$$\begin{aligned} \mathbf{x}[t] &= \mathbf{s}[t] + \mathbf{i}[t] + \mathbf{u}[t] \\ &= \mathbf{Af}[t] + \mathbf{Bg}[t] + \mathbf{u}[t]. \end{aligned} \tag{2}$$

B. Covariance Specification

The latent series are all assumed to have zero mean. As the *temporal* properties of the signal and interference are not known in advance, it is assumed only that they are *wide-sense stationary* and admit Wold representations as

$$\begin{aligned} \mathbf{s}[t] &= \mathbf{A}(\mathbf{K_f} * \mathbf{w}_0)[t], \\ \mathbf{i}_c[t] &= \mathbf{B}_c(\mathbf{K}_{g_c} * \mathbf{w}_c)[t] \quad c = 1, \ldots, C, \end{aligned} \tag{3}$$

where $\mathbf{K}_f[t] \in \mathbb{R}^{r_0 \times r_0}$ and $\mathbf{K}_{g_c}[t] \in \mathbb{R}^{r_c \times r_c}$ are unknown causal time-invariant linear filters which are coloring transformations for the common and channel-c unique factor series respectively. Their matrix transfer functions are $\mathcal{K}_\mathbf{f} = \mathcal{F}\{\mathbf{K_f}\}$ and $\mathcal{K}_{\mathbf{g}_c} = \mathcal{F}\{\mathbf{K}_{\mathbf{g}_c}\}$. The series $\mathbf{w}_k[t], k = 0, .., C$ are white noise processes with identity variance and are uncorrelated across all lags for different values of k. By construction, $\mathbf{s}[t]$ and $\mathbf{i}[t]$ are wide-sense stationary series which are uncorrelated across all lags. Their covariance matrix functions are

$$\begin{aligned} \mathbf{\Gamma_{ss}}[t] &= \mathbf{A}(\mathbf{K_f} * \mathbf{K_f}^\dagger)[t]\mathbf{A}^\mathsf{T} \\ \mathbf{\Gamma}_{\mathbf{i}_c\mathbf{i}_c}[t] &= \mathbf{B}_c(\mathbf{K}_{\mathbf{g}_c} * \mathbf{K}_{\mathbf{g}_c}^\dagger)[t]\mathbf{B}_c^\mathsf{T}, \end{aligned}$$

where $\mathbf{K}^\dagger[t] = \mathbf{K}^\mathsf{T}[-t]$. Letting $\mathbf{\Gamma_{ii}}[t]$ be the covariance matrix function of the stacked interferences, we have $\mathbf{\Gamma_{ii}}[t] = \operatorname{blkdiag}(\mathbf{\Gamma}_{\mathbf{i}_1\mathbf{i}_1}[t], \ldots, \mathbf{\Gamma}_{\mathbf{i}_C\mathbf{i}_C}[t])$ as the interference series for different channels are uncorrelated across all lags. Similarly, $\mathbf{s}[t]$ and $\mathbf{i}[t]$ are uncorrelated across all lags. Finally, the noise $\mathbf{u}[t]$ is temporally white with diagonal variance matrix $\mathbf{\Phi}$,

$$\mathbf{\Gamma_{uu}}[t] = \mathbf{\Phi}\delta[t],$$

and is assumed to be uncorrelated with both the signal and interference series across all lags.

With the preceding assumptions, the all-channel observation series $\mathbf{x}[t]$ is zero mean and wide-sense stationary, with covariance matrix function

$$\mathbf{\Gamma_{xx}}[t] = \mathbf{\Gamma_{ss}}[t] + \mathbf{\Gamma_{ii}}[t] + \mathbf{\Gamma_{uu}}[t]. \tag{4}$$

Analysis of the second-order spatio-temporal relationships of the all-channel observations $\mathbf{x}$ is performed in the frequency domain. The additive construction of $\mathbf{x}[t]$ from the latent series implies that the power spectral density $\mathbf{S_{xx}}$ is similarly constructed,

$$\begin{aligned} \mathbf{S_{ss}}(\mathrm{e}^{\mathrm{i}\omega}) &= \mathbf{A}\mathcal{K}_f(\mathrm{e}^{\mathrm{i}\omega})\mathcal{K}_f^\mathsf{T}(\mathrm{e}^{-\mathrm{i}\omega})\mathbf{A}^\mathsf{T}, \\ \mathbf{S_{ii}}(\mathrm{e}^{\mathrm{i}\omega}) &= \mathbf{B}\mathcal{K}_g(\mathrm{e}^{\mathrm{i}\omega})\mathcal{K}_g^\mathsf{T}(\mathrm{e}^{-\mathrm{i}\omega})\mathbf{B}^\mathsf{T}, \\ \mathbf{S_{xx}}(\mathrm{e}^{\mathrm{i}\omega}) &= \mathbf{S_{ss}}(\mathrm{e}^{\mathrm{i}\omega}) + \mathbf{S_{ii}}(\mathrm{e}^{\mathrm{i}\omega}) + \mathbf{\Phi}, \end{aligned} \tag{5}$$

where $\mathbf{B}$ is $\operatorname{blkdiag}(\mathbf{B}_1, \ldots, \mathbf{B}_C)$ and $\mathcal{K}_\mathbf{g}(\mathrm{e}^{\mathrm{i}\omega}) = \operatorname{blkdiag}(\mathcal{K}_{g_1}(\mathrm{e}^{\mathrm{i}\omega}), \ldots, \mathcal{K}_{g_1}(\mathrm{e}^{\mathrm{i}\omega}))$ is similarly block-diagonal.

III. Estimation

A. Approach

To estimate the spatio-temporal correlations of the latent signal, interference, and response series, we assume that each channel is sampled at T equally-spaced times $t = 0, \ldots, T-1$. That is, the data are C discrete time-series $\mathbf{x}_1[t], \ldots, \mathbf{x}_C[t]$ of length T.

As in the case of MFA with temporally independent observations, the estimation of the cross-sectional and temporal covariance parameters is accomplished by maximization of a Gaussian quasi-likelihood. However, when temporal dependence is present, it is significantly simpler to perform estimation in the frequency domain rather than the time domain. Objective functions with the form of a Gaussian log-likelihood in the frequency domain are called *QGML objectives* [7] and have been extensively used in time series analysis. The estimators obtained by minimization of such a QGML objective are *Whittle estimators* [8], which maximize of the principle part of a time-domain Gaussian likelihood [9] and are asymptotically equivalent to the time-domain MLE.

In the frequency domain, the key datum is the *all-channel periodogram*, which, for times $t = 0, \ldots, T-1$ and radial frequencies $\omega_k = 2\pi k/T$ for $k = 0, \ldots, \lfloor T/2 \rfloor \equiv K$, is computed from the observations $\mathbf{x}[0], \ldots, \mathbf{x}[T-1]$ as

$$\mathbf{I}_T[k] = \frac{1}{T}\big(\textstyle\sum_{t=0}^{T-1} \mathbf{x}[t]\mathrm{e}^{-\mathrm{i}t\omega_k}\big)\big(\textstyle\sum_{t=0}^{T-1} \mathbf{x}[t]\mathrm{e}^{-\mathrm{i}t\omega_k}\big)^{\mathsf{H}}.$$

The finite-sample QGML objective then is

$$\ell(\mathbf{I}_T; \mathbf{S}_{\mathbf{xx}}) \equiv \frac{1}{K+1}\sum_{k=0}^{K} \log\det \mathbf{S}_{\mathbf{xx}}(\omega_k) + \operatorname{tr} \mathbf{S}_{\mathbf{xx}}^{-1}(\omega_k)\mathbf{I}_T[k]. \tag{6}$$

To obtain the desired correlation estimates, we can minimize $\ell(\mathbf{I}_T; \mathbf{S}_{\mathbf{xx}})$ over the class of observation power spectral densities $\mathbf{S}_{\mathbf{xx}}$ realizable by (5) for some parameter values $(\mathbf{A}, \mathbf{B}, \mathcal{K}_{\mathbf{f}}, \mathcal{K}_{\mathbf{g}})$. However, there are both computational and theoretical issues which must be addressed. The first issue is that $\mathbf{S}_{\mathbf{xx}}$ is unchanged under change of *scale* and *orthonormal bases* of the factor spaces as well as *frequency-varying unitary transformation* of the transfer functions. The second issue is that the space of functional parameters $\mathcal{K}_{\mathbf{f}}(\mathrm{e}^{\mathrm{i}\omega})$ and $\mathcal{K}_{\mathbf{g}}(\mathrm{e}^{\mathrm{i}\omega})$ is too large to optimize over directly.

B. Invariants and Equivalence Class Representatives

In writing the signal processes $\mathbf{s}[t]$ as the time-varying weighted combination of the columns of $\mathbf{A}$, where the weight process $\mathbf{f}[t]$ is obtained by passing white noise through the linear filter with transfer function $\mathcal{K}_{\mathbf{f}}(\mathrm{e}^{\mathrm{i}\omega})$, we introduce two indeterminacies into our synthesis. First, any time-constant change of basis of the *factor* space which takes $\mathbf{A} \mapsto \mathbf{AT}$ and $\mathbf{f}[t] \mapsto \mathbf{T}^{-1}\mathbf{f}[t]$ leaves $\mathbf{s}[t]$ fixed. Second, if $\mathbf{U}(\mathrm{e}^{\mathrm{i}\omega})$ is a Hermitian-even unitary $r_0 \times r_0$ matrix function, then $\mathbf{f}'[t]$ synthesized as $\{\mathbf{f}'\} = (\mathcal{K}_{\mathbf{f}}\mathbf{U})(\mathrm{e}^{\mathrm{i}\omega})\{\mathbf{w}_0\}$ would have the same second-order statistical properties as $\mathbf{f}[t]$. However, we can remove these indeterminacies and similar ones for the interference process by restricting the set of loading matrices and transfer functions while not reducing the realizable correlations of the latent series.

This is accomplished by selecting unique representatives of each class of $(\mathbf{A}, \mathbf{B}, \mathcal{K}_{\mathbf{f}}, \mathcal{K}_{\mathbf{g}})$ which yield the same $\mathbf{S}_{\mathbf{xx}}$. However, the minimizing values themselves are only arbitrary representatives of entire equivalence classes. Therefore, second-order quantities such as $\mathbf{\Gamma}_{\mathbf{ss}}[t], \mathbf{\Gamma}_{\mathbf{ii}}[t]$ and $\mathbf{S}_{\mathbf{xx}}(\mathrm{e}^{\mathrm{i}\omega})$ are used when *interpreting* the obtained results, as they are equal for any choice of representative. Details of the unique-representative constraints are provided in the journal version of this paper.

C. Functional Basis Expansion

A flexible model for the transfer functions $\mathcal{K}_{\mathbf{f}}$ and $\mathcal{K}_{\mathbf{g}}$ of the latent factor series is obtained by approximating the functional parameters with their expansions relative to a chosen collection of basis functions. Significant computational advantages are obtained if *local basis functions* are used, as then certain matrices used in the estimation procedure will be sparse. Therefore, we employ the frequently-used *B-spline basis*. Let $B_{m,\ell,\mathbf{p}}(\omega)$ denote the mth B-spline basis function of order ℓ with knot vector $\mathbf{p}$, $0 = p_0 \le p_1 \le \ldots, \le p_{M+\ell} = \pi$. The functions $B_{m,\ell,\mathbf{p}}$ are uniquely defined up to normalization and can be computed via the Cox-de Boor [10] recursion.

With this choice of basis, the unknown transfer functions $\mathcal{K}_{\mathbf{f}}$ and $\mathcal{K}_{\mathbf{g}}$ are approximated as

$$\mathcal{K}_{\mathbf{f},M}[k] = \textstyle\sum_{m=1}^{M} \mathbf{L}_m B_{m,\ell,\mathbf{p}}(\omega_k),$$
$$\mathcal{K}_{\mathbf{g},M}[k] = \textstyle\sum_{m=1}^{M} \mathbf{R}_m B_{m,\ell,\mathbf{p}}(\omega_k),$$

where the $\mathbf{L}_m \in \mathbb{C}^{r_0 \times r_0}$ are the lower-triangular matrix coefficients for the common factor transfer function and $\mathbf{R}_{m,c} \in \mathbb{C}^{r_c \times r_c}$ are the lower-triangular matrix coefficients for the channel-c unique factor transfer function with $\mathbf{R}_m = \operatorname{blkdiag}(\mathbf{R}_{m,1}, \ldots, \mathbf{R}_{m,C})$.

D. Optimization Problem

The QGML estimators of the MFA parameters based on the observations at times $t = 0, \ldots, T-1$ and using M basis functions are obtained by solving

$$\min_{(\mathbf{A},\mathbf{B},\mathbf{\Phi},\{\mathbf{L}_m\},\{\mathbf{R}_m\})} \ell(\mathbf{I}_T; \mathbf{A}, \mathbf{B}, \mathbf{\Phi}, \{\mathbf{L}_m\}, \{\mathbf{R}_m\}) \tag{7}$$

where $\ell(\mathbf{I}_T; \mathbf{A}, \mathbf{B}, \mathbf{\Phi}, \{\mathbf{L}_m\}, \{\mathbf{R}_m\})$ is (6) with

$$\mathbf{S}_{\mathbf{xx}}(\omega_k) = \mathbf{A}\mathcal{K}_{\mathbf{f},M}[k]\mathcal{K}_{\mathbf{f},M}^{\mathsf{H}}[k]\mathbf{A} + \mathbf{B}\mathcal{K}_{\mathbf{g},M}[k]\mathcal{K}_{\mathbf{g},M}^{\mathsf{H}}[k]\mathbf{B} + \mathbf{\Phi}.$$

As is the case for MFA with independent observations, optimization of ℓ can be accomplished by an *alternating block Majorization-Minimization* (MM) procedure [6], where the parameters blocks $\mathbf{\Phi}$, $(\mathbf{A}, \mathbf{B})$, and $(\{\mathbf{L}_m\}, \{\mathbf{R}_m\})$ are updated cyclically with the other two blocks held fixed. For notational convenience, let $\mathbf{C}$ be the block matrix $[\mathbf{A}\ \ \mathbf{B}]$, let $\mathcal{K}[k]$ be $\operatorname{blkdiag}(\mathcal{K}_{\mathbf{f},M}[k], \mathcal{K}_{\mathbf{g},M}[k])$, and let $\bar{\mathbf{I}}_T$ be the frequency-average periodogram $(K+1)^{-1}\sum_{k=0}^{K} \mathbf{I}_T[k]$. The objective can be separated into $\ell(\mathbf{I}_T) = \ell_{T,1}(\bar{\mathbf{I}}_T) + \ell_{T_2}(\mathbf{I}_T)$, which are

$$\ell_{T,1}(\bar{\mathbf{I}}_T; \mathbf{\Phi}) = \textstyle\sum_{m=1}^{n} \log \Phi_{mm} + [\bar{\mathbf{I}}_T]_{mm}/\Phi_{mm},$$
$$\ell_{T,2}(\mathbf{I}_T; \mathbf{A}, \mathbf{B}, \mathbf{\Phi}, \{\mathbf{L}_m\}, \{\mathbf{R}_m\}) = \textstyle\sum_{k=0}^{K} \big[\log\det \mathbf{\Xi}[k]$$
$$- \operatorname{tr}(\mathbf{\Xi}^{-1}[k]\mathcal{K}^{\mathsf{H}}[k]\mathbf{C}^{\mathsf{T}}\mathbf{\Phi}^{-1}\mathbf{I}_T[k]\mathbf{\Phi}^{-1}\mathbf{C}\mathcal{K}[k])\big] \cdot (K+1)^{-1}$$

where $\mathbf{\Xi}[k]$ is an abbreviation of

$$\mathbf{\Xi}[k; \mathbf{A}, \mathbf{B}, \mathbf{\Phi}, \{\mathbf{L}_m\}, \{\mathbf{R}_m\}] \equiv \mathbf{I}_r + \mathcal{K}^{\mathsf{H}}[k]\mathbf{C}^{\mathsf{T}}\mathbf{\Phi}^{-1}\mathbf{C}\mathcal{K}[k].$$

The first quantity, $\ell_{T,1}$, by itself would be simple to optimize, while the second quantity, $\ell_{T,2}$, is the average of frequency-dependent terms, each of which admits a majorizing surrogate function. The surrogate functions used follow from the fact that, in the kth term of $\ell_{T,2}$, each block of parameters enter the matrices $\mathbf{\Xi}$ and $\mathcal{K}^{\mathsf{H}}[k]\mathbf{C}^{\mathsf{T}}\mathbf{\Phi}^{-1}\mathbf{I}_T[k]\mathbf{\Phi}^{-1}\mathbf{C}\mathcal{K}[k]$ only at most quadratically. Hence, the solution of (7) is obtained by iterative closed-form solution of linear or quadratic subproblems. The sth iteration consists of

1) (Optimization for $\mathbf{\Phi}$) Using the previous iteration's parameter values, the noise variance is updated by extracting the diagonal of a frequency-averaged transformation of the periodogram.
2) (Optimization for $\mathbf{A}, \mathbf{B}$) Holding the other parameters fixed, solving a non-frequency-dependent matrix least-squares problem gives updates for $\mathbf{A}$ and $\mathbf{B}$.

3) (Optimization for $\{\mathbf{L}_m\}$, $\{\mathbf{R}_m\}$) Mapping $\mathbf{I}_T[k]$ onto the factor space, the objective is majorized by a positive definite quadratic form in $\{\mathbf{L}_m\}$ and $\{\mathbf{R}_m\}$. Use of a local basis ensures that the form can be minimized efficiently with a banded least-squares solver.
4) (Enforcement of Unique-Representative Constraints) In the previous steps, the requirements on $(\mathbf{A}, \mathbf{B}, \mathcal{K}_{\mathbf{f}}, \mathcal{K}_{\mathbf{g}})$ which select a unique representative, as mentioned in Section III-B, are not enforced. However, we can transform the obtained values to satisfy the constraints without changing the objective value.

Details of the procedure and convergence results are provided in the journal version of this paper.

IV. Experiments

In this section, we examine how the proposed MFA extension to model the temporal correlations in the latent series improves model quality. In particular, the quality of the estimates for the lag-zero *spatial correlation* matrix $\mathbf{R} = \mathbf{\Gamma}_{\mathbf{ss}}[0] + \mathbf{\Gamma}_{\mathbf{ii}}[0] + \mathbf{\Phi}$ under varying conditions is explored in Section IV-A. The proposed MFA extension can be directly compared to MFA with the assumption of *independent* observations as presented in [1]. This allows the evaluation of how modeling the temporal correlations in the latent series in addition to the their spatial correlations improves performance.

In the following experiments, the relative contribution of the signal, interference, and noise series to the channel-c observation are quantified by their lag-zero power ratios,

$$\eta_{\mathbf{s},c} = \operatorname{tr}(\mathbf{\Gamma}_{\mathbf{s}_c\mathbf{s}_c}[0]) / \operatorname{tr}(\mathbf{R}_c),$$
$$\eta_{\mathbf{i},c} = \operatorname{tr}(\mathbf{\Gamma}_{\mathbf{i}_c\mathbf{i}_c}[0]) / \operatorname{tr}(\mathbf{R}_c),$$
$$\eta_{\mathbf{u},c} = \operatorname{tr}(\mathbf{\Phi}_c) / \operatorname{tr}(\mathbf{R}_c).$$

Here, $\mathbf{R}_c = \mathbf{\Gamma}_{\mathbf{x}_c\mathbf{x}_c}[0]$ is the zero-lag observation covariance within channel c and $\mathbf{\Gamma}_{\mathbf{s}_c\mathbf{s}_c}[0]$ and $\mathbf{\Gamma}_{\mathbf{i}_c\mathbf{i}_c}[0]$ are respectively the $n_c \times n_c$ blocks on the main diagonal of the lag-zero signal and interference covariances in channel c. All experiments use cubic B-splines $(M = 6, \ell = 4)$ with equidistant knots.

A. Spatial Correlation Estimation

To understand how the presence of non-trivial temporal correlations in the latent series affect estimation of the zero-lag spatial correlation matrix $\mathbf{R} = \mathbf{\Gamma}_{\mathbf{xx}}[0]$, we simulate $\mathbf{x}[t]$ using temporally correlated latent factor series. Under mild moment conditions, the estimation procedure in [1], which assumes that the observations are temporally *independent*, still allows for consistent estimation of $\mathbf{R}$ even when the observations are temporally dependent. Therefore, the estimates $\hat{\mathbf{R}}_{\text{ind}}$ obtained using the method of [1] can be compared to the estimated $\hat{\mathbf{R}}_{\text{dep}} = \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathsf{T}} + \hat{\mathbf{B}}\hat{\mathbf{B}}^{\mathsf{T}} + \hat{\mathbf{\Phi}}$ obtained from the procedure of Section III-D.

To make this comparison, we replicate the experimental settings of Experiment 3 in [1, Sec. IV.c], where $C = 4$ channels of sizes $N_1 = 6, N_2 = 8, N_3 = 10$ and $N_4 = 12$, with $r_0 = 2$ common factors and $r_c = 1$ unique factor for each channel. The proportion of variance explained by signal, interference, and noise components is equal across channels,

$$\eta_{\mathbf{s},c} = 0.1 \quad \eta_{\mathbf{i},c} = 0.5 \quad \eta_{\mathbf{u},c} = 0.4 \;\; c = 1, \ldots, C.$$

The latent factors are independent realizations of a Gaussian $AR(1)$ process with AR coefficient ϕ and error variance σ^2. The factor loadings and lag-zero factor covariances are normalized so that the lag-zero covariance of $\mathbf{f}[t]$ and $\mathbf{g}_c[t]$ $c = 1, \ldots, C$ is the identity. The measure of spatial covariance estimation quality is the Normalized Mean-Squared Error (NMSE), defined as

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

which we estimate by averaging 400 Monte-Carlo trials at each simulation setting.

In Figures 1a and 1b, we compare the achieved NMSE for varying number of samples T when the latent factor exhibit temporal correlations, $\phi = 0.9, \sigma = 0.435$ versus the setting where the latent factors are uncorrelated across time, $\phi = 0.0, \sigma = 1.0$. Three estimators of $\mathbf{R}$ are compared, namely the estimated $\hat{\mathbf{R}}_{\text{dep}}$ based on modeling the temporal correlations, the estimated $\hat{\mathbf{R}}_{\text{ind}}$ based on the assumption of independent observations, and the non-parametric sample covariance estimate $\hat{\mathbf{R}}_{\text{samp}} = T^{-1}\sum_{t=0}^{T} \mathbf{x}[t]\mathbf{x}^{\mathsf{T}}[t]$.

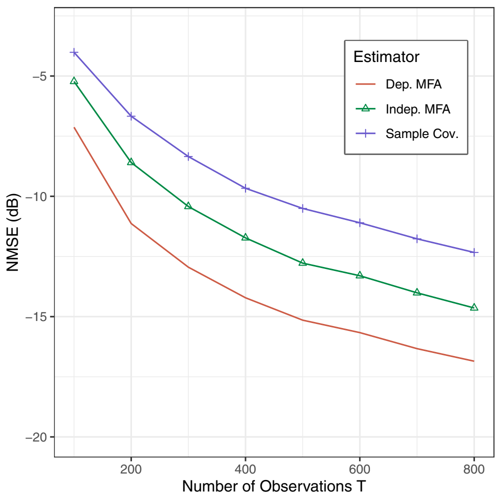
Figure 1a demonstrates that $\hat{\mathbf{R}}_{\text{dep}}$ exhibits improved performance for estimation of the spatial covariance when non-zero temporal correlation in the latent factors is present, and that the performance gap between $\hat{\mathbf{R}}_{\text{dep}}$ and $\hat{\mathbf{R}}_{\text{ind}}$ increases as the number of samples increases. Conversely, when the latent factors are temporally uncorrelated, Figure 1b shows that $\hat{\mathbf{R}}_{\text{dep}}$ and $\hat{\mathbf{R}}_{\text{ind}}$ perform similarly except for small sample sizes.

To further compare how varying levels of temporal dependence affect the estimation of the cross-sectional parameters, Figure 1c compares the achieved NMSE with $T = 400$ samples and varying $AR(1)$ coefficient. For low levels of temporal dependence of the latent factors, the performances of $\hat{\mathbf{R}}_{\text{ind}}$ and $\hat{\mathbf{R}}_{\text{dep}}$ are comparable. However, for high levels of temporal dependence, $\hat{\mathbf{R}}_{\text{dep}}$ has a significant advantage. For a fixed sample length, increasing the temporal dependence reduces the amount of information present and so all estimators perform worse when the temporal correlation is high.

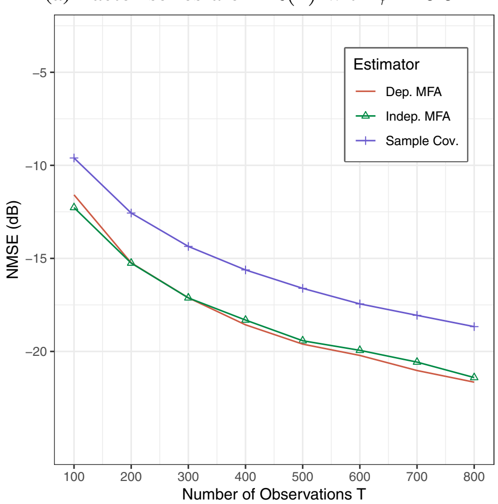
V. Discussion

Many array processing problems center around combining multiple observation series in distinct channels, with the goal of recovering the commonalities across the different channels when both channel-specific interference and idiosyncratic noise are present. In Section II, an extension to MFA that models the multi-channel spatio-temporal correlations in the observation series is proposed, which augments the spatial covariance parameters with unknown transfer functions for the latent factors. Section III describes a novel majorization-minimization estimation method for the covariance parameters.

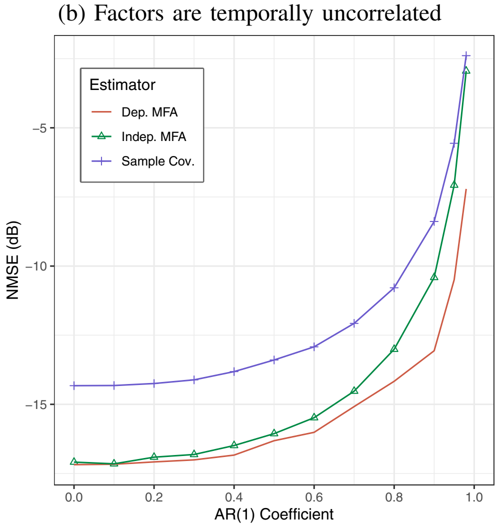
The experiments of Section IV demonstrate that modeling the second order temporal properties of the observations



(a) Factor series are $AR(1)$ with $\phi = 0.9$



(b) Factors are temporally uncorrelated



(c) Varying temporal correlation with $T = 400$

Fig. 1: Normalized mean-squared error achieved by estimators of zero-lag cross-sectional covariance $\mathbf{R}$ with varying T at different temporal correlations (1a, 1b) and varying temporal correlations with fixed T (1c).

improves performance when the observations have strong temporal correlations. As array processing problems often involve both spatially and temporally correlated observations, the proposed extension to MFA is a step forward to attaining the envisioned function of multi-channel factor analysis.

Acknowledgments

This work was supported in part by National Science Foundation grants DMS-1923142, CNS-1932413, and DMS-2123761.

References

[1] 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.

[2] N. Klausner, M. R. Azimi-Sadjadi, and L. L. Scharf, “Detection of spatially correlated time series from a network of sensor arrays,” *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1396–1407, 2014.

[3] A. Rosuel, P. Vallet, P. Loubaton, and X. Mestre, “On the detection of low-rank signal in the presence of spatially uncorrelated noise: A frequency domain approach,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 4458–4473, 2021.

[4] D. Ramírez, J. Via, I. Santamaría, and L. L. Scharf, “Detection of spatially correlated Gaussian time series,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5006–5015, 2010.

[5] J. Manco-Vásquez, M. Lazaro-Gredilla, D. Ramírez, J. Vía, and I. Santamaria, “Bayesian multiantenna sensing for cognitive radio,” in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 77–80, 2012.

[6] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Transactions on Signal Processing*, vol. 65, pp. 794–816, 2017.

[7] M. Taniguchi and Y. Kakizawa, *Asymptotic Theory of Statistical Inference for Time Series*. Springer New York, 2000.

[8] P. Whittle, “The analysis of multiple stationary time series,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 15, no. 1, pp. 125–139, 1953.

[9] K. Dzhaparidze and A. Yaglom, “Spectrum parameter estimation in time series analysis,” vol. 4 of *Developments in Statistics*, pp. 1–96, Elsevier, 1983.

[10] C. de Boor, “A practical guide to splines,” in *Applied Mathematical Sciences*, 1978.