2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Sparse Plus Low-rank Identification of Latent-variable Graphical ARMA Models

Junyao You, Chengpu Yu

Abstract— **This paper deals with the identification of graphical autoregressive moving-average (ARMA) models with latent variables. Combining sparse structural characteristics of the graphical model with low-rank modeling of the latent variables, a sparse plus low-rank based iterative identification algorithm is proposed. The topological information embedded in the sparse AR dynamics is estimated from a regularized Yule-Walker optimization problem, which is then treated as prior graphical structure constraint. The latent-variable plus MA part is identified by solving a convex constrained trace norm minimization problem. Based on the MA part estimate and the structural constraint, the graphical AR estimates are updated by the sparse plus low-rank optimization framework and are then used for the update of the latent-variable plus MA part. The effectiveness of the proposed method is illustrated through a simulation study.**

I. Introduction

Graphical models can deal with the network topology identification problem, whose edges describe conditional dependence relations among the model. In many applications throughout science and engineering [1]–[4], there may exist unmeasured (latent) variables, which make it difficult to infer graphical structure. Ancestral graphs can be used to represent independence relations among observed variables in the presence of latent confounders [5], and the fast causal inference is a common approach for causal discovery. For tree-structured latent graphical models, [6] provided consistent and computationally efficient algorithms for learning minimal latent trees and [7] provided a method that can detect the precise location of latent variables. The problem of estimating the conditional dependency structure of latent-variable AR graphical models was considered in [8], which introduces two regularizers inducing sparsity and low-rank on the corresponding components of the observed inverse spectral density matrix. As the generalization of the covariance extension approach [9], this identification paradigm was successfully applied to neuroimaging datasets [10] and further modified by constructing a confidence set constraint to simplify the regularization parameter selection procedure [11]. The robustness and scalability of the identification procedure was improved by using reciprocal processes to approximate AR processes [12]. The work in [13], [14] retrieved the sparse plus low-rank graphical structure from a different Bayesian view [15], and the corresponding non-parametric Gaussian regression approaches were proposed.

A host of literature address the latent variables through some factor analysis technique [16]–[19]. The robust estimation of static factor models was studied in [17], which exploits the decomposition of the covariance matrix as the sum of a low-rank matrix whose rank determines the number of factors and the diagonal matrix modeling the idiosyncratic noise. The estimation uncertainty of the sample covariance was also considered and dealt with by setting a divergence constraint. This identification idea has been generalized to dynamic AR, MA and ARMA factor model learning, where the AR dynamic is estimated by means of moments matching and the MA dynamic factor analysis is performed with the low-rank plus diagonal decomposition of the spectral density matrix [18], [19].

In this paper, we focus on the identification of latent-variable graphical ARMA models, which connects the recent research on the MA dynamic factor analysis [19] to graphical model selection and extends the latent-variable graphical AR model identification results in [20]. We address the following problems:

- A regularized Yule-Walker optimization model is proposed for sparse AR dynamics identification, which is reformulated as a vector form and solved by cross-validated LASSO.
- A modified low-rank plus diagonal decomposition method for the latent-variable and MA dynamics learning is formulated by taking into account of a convex measure of estimation uncertainty.
- Based on the sparsity of the interdependence between observed variables and the low-rank property of the spectrum corresponding to the latent-variable part, a sparse plus low-rank optimization framework is presented to update the graphical AR structure. By alternatingly estimating the latent plus MA part and the graphical AR part, finer results can be obtained in terms of estimation accuracy.

II. Notation

In this paper, the set of real column vectors of size $n \times 1$ is denoted by $\mathbb{R}^n$. For $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{y}_i$ represents its i-th entry, $\|\boldsymbol{y}\|_1$ is the l_1 norm defined as $\sum_1^n |\boldsymbol{y}_i|$, and $\|\boldsymbol{y}\|_2$ denotes the Euclidean norm defined as $\sqrt{\sum_1^n \boldsymbol{y}_i^2}$. The set of real matrices of size $m \times n$ is denoted by $\mathbb{R}^{m\times n}$ and the set of real symmetric matrices of order n is represented as $\mathbb{S}^n$. For a matrix $\boldsymbol{X}$, $[\boldsymbol{X}]_{ij}$ or $\boldsymbol{X}_{ij}$ represents its ij-th entry, $\|\boldsymbol{X}\|_0$ denotes the number of its non-zero entries,

This work was supported by the National Key Research and Development Project under Grant 2020YFC1512503, and the National Natural Science Foundation of China (Grant No. 61991414, 62088101)

J. You is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: yjy804521297@163.com).

C. Yu is with the Chongqing Innovation Center, Beijing Institute of Technology, Chongqing 401147, China (e-mail: yuchengpu@bit.edu.cn).

$\|\boldsymbol{X}\|_1$ is the l_1 norm defined as $\sum_{i,j}|\boldsymbol{X}_{ij}|$, $\|\boldsymbol{X}\|_F$ stands for the Frobenius norm, and $\mathrm{rank}(\boldsymbol{X})$ denotes its rank. $\mathrm{supp}(\boldsymbol{X})$ is the support set defined as $\{(i,j) : \boldsymbol{X}_{ij} \neq 0\}$, and $|\mathrm{supp}(\boldsymbol{X})|$ represents the dimension of the set. For a square matrix $\boldsymbol{X}$, its trace is denoted by $\mathrm{tr}(\boldsymbol{X})$. If $\boldsymbol{X}$ is positive definite (semi-definite), we write $\boldsymbol{X} > 0$ ($\boldsymbol{X} \geq 0$). Given $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n] \in \mathbb{R}^{m\times n}$ with $\boldsymbol{X}_j \in \mathbb{R}^m$ representing the j-th column, define the vector $\boldsymbol{Y} \in \mathbb{R}^{mn}$ as $\boldsymbol{Y} = \mathrm{vec}(\boldsymbol{X}) := [\boldsymbol{X}_1^{\mathrm{T}}, \boldsymbol{X}_2^{\mathrm{T}}, \cdots, \boldsymbol{X}_n^{\mathrm{T}}]^{\mathrm{T}}$. The inner product of square matrices $\boldsymbol{A}, \boldsymbol{B}$ is defined as $\langle \boldsymbol{A}, \boldsymbol{B}\rangle := \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B})$. $\boldsymbol{I}_n \in \mathbb{S}^n$ denotes the identity matrix. The symbol $\mathbb{M}^{n,p}$ is the set of matrices $\boldsymbol{Q} := [\boldsymbol{Q}_0\ \boldsymbol{Q}_1\ \cdots\ \boldsymbol{Q}_p]$ with $\boldsymbol{Q}_0 \in \mathbb{S}^n$ and $\boldsymbol{Q}_j \in \mathbb{R}^{n\times n}$ $(j = 1, \cdots, p)$. The mapping $\mathcal{D}$ from $\mathbb{S}^{n(p+1)}$ to $\mathbb{M}^{n,p}$ is defined as follows. If the matrix $\boldsymbol{S} \in \mathbb{S}^{n(p+1)}$ is partitioned as

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{0,0} & \boldsymbol{S}_{0,1} & \cdots & \boldsymbol{S}_{0,p} \\ \boldsymbol{S}_{0,1}^{\mathrm{T}} & \boldsymbol{S}_{1,1} & \cdots & \boldsymbol{S}_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S}_{0,p}^{\mathrm{T}} & \boldsymbol{S}_{1,p}^{\mathrm{T}} & \cdots & \boldsymbol{S}_{p,p} \end{bmatrix},$$

where $\boldsymbol{S}_{i,i} \in \mathbb{S}^n$ and $\boldsymbol{S}_{i,j} \in \mathbb{R}^{n\times n}$ $(i \neq j)$ denote the sub-blocks of $\boldsymbol{S}$, then

$$\mathcal{D}(\boldsymbol{S}) := [\mathcal{D}(\boldsymbol{S})_0\ \mathcal{D}(\boldsymbol{S})_1\ \cdots\ \mathcal{D}(\boldsymbol{S})_p] \in \mathbb{M}^{n,p},$$
$$\mathcal{D}(\boldsymbol{S})_0 = \sum_{v=0}^{p} \boldsymbol{S}_{v,v},\ \mathcal{D}(\boldsymbol{S})_j = 2\sum_{v=0}^{p-j} \boldsymbol{S}_{v,v+j},\ j = 1, \cdots, p.$$

Denote the expectation operator by $\mathbb{E}$ and use $\otimes$ to represent the Kronecker product. $F^*(z) = F(z^{-1})^{\mathrm{T}}$ represents the complex conjugate transpose, where the operator is taken as $z = e^{\mathrm{j}\omega}$ for $\omega \in [-\pi, \pi]$.

III. Problem Formulation

Consider the graphical ARMA models with latent variables:

$$\sum_{j=0}^{p_1} \boldsymbol{A}_j \boldsymbol{y}(t-j) = \boldsymbol{W}_L \boldsymbol{x}(t) + \sum_{i=0}^{p_2} \boldsymbol{W}_{D,i} \boldsymbol{w}(t-i), \tag{1}$$
$$\boldsymbol{A}_0 = \boldsymbol{I}_n,\ \boldsymbol{x}(t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_l),\ \boldsymbol{w}(t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n),$$

where $\boldsymbol{W}_L \in \mathbb{R}^{n\times l}$ $(l \leq n)$, the AR parameter matrices $\boldsymbol{A}_j \in \mathbb{R}^{n\times n}$ $(j = 1, 2, \cdots, p_1)$ are sparse, and the MA parameter matrices $\boldsymbol{W}_{D,i} \in \mathbb{R}^{n\times n}$ $(i = 1, 2, \cdots, p_2)$ are diagonal. The processes $\boldsymbol{x}(t) \in \mathbb{R}^l$ and $\boldsymbol{w}(t) \in \mathbb{R}^n$ are independent white Gaussian noise vectors, that is $\forall t_1, t_2$, $\mathbb{E}\boldsymbol{x}(t_1)\boldsymbol{w}(t_2)^{\mathrm{T}} = 0$, and $\boldsymbol{y}(t) \in \mathbb{R}^n$ is the measurable system output.

Remark 1 Model (1) allows for latent variables $\boldsymbol{x}_k(t)$ $(k = 1, 2, \cdots, l)$ to produce the conditional independence among observed variables $\boldsymbol{y}_q(t)(q = 1, 2, \cdots, n)$, which can be regarded as the extension of graphical ARMA models [21]–[23] or latent-variable graphical AR models [8], [10], [11].

Remark 2 Model (1) can also be related to factor models [16]–[19]. Let $\boldsymbol{A}(z)$ and $\boldsymbol{W}_D(z)$ denote the matrix polynomials consisting of the AR and MA coefficients, respectively. Then $\boldsymbol{A}(z)^{-1}\boldsymbol{W}_L$ is the factor loading transfer matrix polynomial; $\boldsymbol{x}(t)$ is the process describing unobservable l factors; $\boldsymbol{A}(z)^{-1}\boldsymbol{W}_L\boldsymbol{x}(t)$ represents the latent variable and $\boldsymbol{A}(z)^{-1}\boldsymbol{W}_D(z)\boldsymbol{w}(t)$ is the idiosyncratic noise.

By separating the observable variables and latent variables, the sub-blocks of the power spectrum $\boldsymbol{\Phi}(z)$ of Model (1) have the following relationship:

$$\begin{aligned} \boldsymbol{A}(z)\boldsymbol{\Phi}_y(z)\boldsymbol{A}^*(z) &= \boldsymbol{W}_L\boldsymbol{\Phi}_x\boldsymbol{W}_L^{\mathrm{T}} + \boldsymbol{W}_D(z)\boldsymbol{\Phi}_w\boldsymbol{W}_D^*(z) \\ &= \boldsymbol{\Phi}_{W_L} + \boldsymbol{\Phi}_{W_D}(z), \\ \boldsymbol{\Phi}_{W_L} &:= \boldsymbol{W}_L\boldsymbol{W}_L^{\mathrm{T}},\ \boldsymbol{\Phi}_{W_D}(z) := \boldsymbol{W}_D(z)\boldsymbol{W}_D^*(z), \end{aligned} \tag{2}$$

where $\boldsymbol{\Phi}_y(z)$ is the spectral density matrix corresponding to the observable output $\boldsymbol{y}(t)$, $\boldsymbol{\Phi}_x = \boldsymbol{I}_l$ and $\boldsymbol{\Phi}_w = \boldsymbol{I}_n$ are spectrums of white Gaussian processes $\boldsymbol{x}(t)$ and $\boldsymbol{w}(t)$, respectively. Note that the latent-variable spectrum $\boldsymbol{\Phi}_{W_L} \geq 0$ is a low-rank matrix of size $n \times n$ and rank l, and the MA spectrum $\boldsymbol{\Phi}_{W_D}(z)$ is diagonal. By exploiting block matrix inversion formula and Schurs complement representation [24], the inverse spectrum $\boldsymbol{\Phi}_y(z)^{-1}$ associated with the observed variables has a sparse plus low-rank structure [25]:

$$\boldsymbol{\Phi}_y(z)^{-1} = \boldsymbol{\Phi}_S(z) - \boldsymbol{\Phi}_L(z), \tag{3}$$

$$\begin{aligned} \boldsymbol{\Phi}_S(z) &= \boldsymbol{A}^*(z)[\boldsymbol{W}_D(z)\boldsymbol{\Phi}_\omega\boldsymbol{W}_D^*(z)]^{-1}\boldsymbol{A}(z) \\ &= \boldsymbol{A}^*(z)\boldsymbol{\Phi}_{W_D}(z)^{-1}\boldsymbol{A}(z), \\ \boldsymbol{\Phi}_L(z) &= \boldsymbol{\Psi}^*(z)\boldsymbol{\Lambda}(z)^{-1}\boldsymbol{\Psi}(z), \\ \boldsymbol{\Psi}(z) &= \boldsymbol{W}_L^{\mathrm{T}}[\boldsymbol{W}_D(z)\boldsymbol{\Phi}_\omega\boldsymbol{W}_D^*(z)]^{-1}\boldsymbol{A}(z) \\ &= \boldsymbol{W}_L^{\mathrm{T}}\boldsymbol{\Phi}_{W_D}(z)^{-1}\boldsymbol{A}(z), \\ \boldsymbol{\Lambda}(z) &= \boldsymbol{\Phi}_x^{-1} + \boldsymbol{W}_L^{\mathrm{T}}[\boldsymbol{W}_D(z)\boldsymbol{\Phi}_\omega\boldsymbol{W}_D^*(z)]^{-1}\boldsymbol{W}_L \\ &= \boldsymbol{I}_l + \boldsymbol{W}_L^{\mathrm{T}}\boldsymbol{\Phi}_{W_D}(z)^{-1}\boldsymbol{W}_L. \end{aligned} \tag{4}$$

Remark 3 Due to the relationship [21], [22], [26]:

$$\mathcal{X}_{\{k\}} \perp\!\!\!\perp \mathcal{X}_{\{q\}} | \mathcal{X}_{\{\mathcal{V}_{n+l}\setminus\{k,q\}\}} \Leftrightarrow [\boldsymbol{\Phi}(z)^{-1}]_{kq} = 0, \forall |z| \leq 1,$$
$$\mathcal{X}_{\mathcal{I}} := \mathrm{span}\{\boldsymbol{z}_i(t) : i \in \mathcal{I};\ t \in \mathbb{Z}\},\ \mathcal{I} \subset \mathcal{V}_{n+l},\ \boldsymbol{z} := [\boldsymbol{y}^{\mathrm{T}}, \boldsymbol{x}^{\mathrm{T}}]^{\mathrm{T}},$$

the sparse pattern of $\boldsymbol{\Phi}_S$ reflects the existence of few edges among the observable nodes of the graph $\mathcal{G}(\mathcal{V}_{n+l}, \mathcal{E}_{n+l})$, and also characterizes the conditional dependence relations among the observable variables [8]. Owing to the diagonal matrices $\boldsymbol{W}_{D,i}(i = 1, 2, \cdots, p_2)$, the sparsity of $\boldsymbol{\Phi}_S$ computed from (4) is essentially inherited by $\boldsymbol{A}^*(z)\boldsymbol{A}(z)$. Since the rank of $\boldsymbol{\Phi}_L$ coincides with l, $\mathrm{rank}(\boldsymbol{\Phi}_L)$ can measure the number of latent variables chosen to model the statistical conditional dependencies of the data [8].

The identification objective is to estimate the graphical structure $\mathcal{E}_{n+l}$, the number of latent variables, and the unknown parameters $\boldsymbol{A}_j$, $\boldsymbol{\Phi}_{W_L}$, $\boldsymbol{\Phi}_{W_D}$ according to the observations $\boldsymbol{y}(t)$.

IV. Main Results

By virtue of (3), it is plausible to handel the identification for Model (1) under a sparse plus low-rank decomposition optimization framework. However, due to the existence of the MA part, the optimization problem cannot be reformulated as a matricial form like those in [8], [11] to solve and it is difficult to decouple the AR and MA parameters.

Inspired by dynamic factor analysis [18], [19], we can perform the identification of Model (1) in two parts. Once the graphical AR part is determined, the latent-variable coefficient and the MA part can be estimated by low-rank plus diagonal decomposition according to (2). Different from the scalar AR dynamics estimation in factor models, the matrix form of AR parameters in our problem setting leads to the failure of model conversion for data disentanglement (see [18], Eq. (6)). Besides, since the graphical AR part identification involves the graph topology selection, the common maximum likelihood estimator is infeasible. Therefore, we aim to seek for new relationship among the AR, MA parameters and latent variables to design efficient identification method.

A. Preliminary Identification of Sparse AR Dynamics

According to the generalized Yule-Walker equation [27], we have

$$\boldsymbol{R}_k = -\sum_{j=1}^{p_1} \boldsymbol{R}_{k+j}\boldsymbol{A}_j^{\mathrm{T}},\ k < -p_2,$$

where $\boldsymbol{R}_k = \mathbb{E}\boldsymbol{y}(t+k)\boldsymbol{y}(t)^{\mathrm{T}}$ is the autocovariance sequence. Since $\boldsymbol{y}(t)$ is real, we have $\boldsymbol{R}_{-k} = \boldsymbol{R}_k^{\mathrm{T}}$. Define the parameter matrix consisting of the AR parameters as

$$\boldsymbol{\theta}_A := \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \cdots \boldsymbol{A}_{p_1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{p_1 n \times n}.$$

It follows that $\Gamma\boldsymbol{\theta}_A = \Xi$ with

$$\Gamma := \begin{bmatrix} \boldsymbol{R}_{p_2}^{\mathrm{T}} & \boldsymbol{R}_{p_2-1}^{\mathrm{T}} & \cdots \boldsymbol{R}_{p_2-p_1+1}^{\mathrm{T}} \\ \boldsymbol{R}_{p_2+1}^{\mathrm{T}} & \boldsymbol{R}_{p_2}^{\mathrm{T}} & \cdots \boldsymbol{R}_{p_2-p_1+2}^{\mathrm{T}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{R}_{p_1+p_2-1}^{\mathrm{T}} & \boldsymbol{R}_{p_1+p_2-2}^{\mathrm{T}} \cdots & \boldsymbol{R}_{p_2}^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{p_1 n \times p_1 n}, \quad (5)$$

$$\Xi := \begin{bmatrix} -\boldsymbol{R}_{p_2+1} & -\boldsymbol{R}_{p_2+2} \cdots -\boldsymbol{R}_{p_1+p_2} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{p_1 n \times n}. \quad (6)$$

In practice, we can obtain the sample estimate of the autocovariance matrix $\hat{\boldsymbol{R}}_k = \frac{1}{N}\sum_{t=1}^{N-k} \boldsymbol{y}(t+k)\boldsymbol{y}(t)^{\mathrm{T}}$ from N samples, and the corresponding matrix estimates of (5) and (6) can be denoted by $\hat{\Gamma}$ and $\hat{\Xi}$. Inspired by the compressive sensing theory [28], the identification of the sparse parameter matrix $\boldsymbol{\theta}_A$ can be expressed as an optimization problem

$$\min_{\boldsymbol{\theta}_A} \|\boldsymbol{\theta}_A\|_0, \quad \text{s.t.} \quad \|\hat{\Xi} - \hat{\Gamma}\boldsymbol{\theta}_A\|_F^2 \le \varepsilon, \quad (7)$$

where ε is the error tolerance. However, it is a non-convex problem and NP-hard. By employing the l_1 norm as the surrogate for l_0 norm [29] and introducing a nonnegative parameter λ to balance the tradeoff between the l_1 norm and Frobenius norm, Problem (7) can be transformed into a tractable convex relaxation form:

$$\min_{\boldsymbol{\theta}_A} \frac{1}{2}\|\hat{\Xi} - \hat{\Gamma}\boldsymbol{\theta}_A\|_F^2 + \lambda\|\boldsymbol{\theta}_A\|_1. \quad (8)$$

To simplify the solution, we stack the columns of the matrix $\boldsymbol{\theta}_A$ below each other to construct a sparse vector $\mathrm{vec}(\boldsymbol{\theta}_A) \in \mathbb{R}^{p_1 n^2}$, and reformulate Problem (8) as

$$\min_{\mathrm{vec}(\boldsymbol{\theta}_A)} \frac{1}{2}\|\mathrm{vec}(\hat{\Xi}) - (\boldsymbol{I}_n \otimes \hat{\Gamma})\mathrm{vec}(\boldsymbol{\theta}_A)\|_2^2 + \lambda\|\mathrm{vec}(\boldsymbol{\theta}_A)\|_1,$$

which can be directly solved by LASSO with the regularization coefficient λ selected by cross validation [30]. Then the AR parameter estimates $\hat{\boldsymbol{A}}_j$ $(j = 1, \cdots, p_1)$ can be recovered from its optimal solution.

B. Joint Estimation of Latent-variable and MA Spectrums

Denote the sampled covariance and spectrum of the AR process $\boldsymbol{y}_{AR}(t) := \hat{\boldsymbol{A}}(z)\boldsymbol{y}(t)$ as $\hat{\boldsymbol{R}}_k^{AR}$ and $\hat{\boldsymbol{\Phi}}_{AR}(z)$, respectively. According to (2), true spectrum $\boldsymbol{\Phi}_{AR}(z)$ of the AR process can be decomposed as a low-rank plus diagonal structure $\boldsymbol{\Phi}_{AR}(z) = \boldsymbol{\Phi}_{W_L} + \boldsymbol{\Phi}_{W_D}(z)$, which has been studied in [19]. We modify the proposed optimization framework with a Frobenius norm convex constraint instead of the divergence constraint to measure the difference between true spectrum and its estimate, which can avoid the integral operation, and thereby construct a more tractable form:

$$\begin{aligned} \min_{\boldsymbol{\Phi}_{W_L}, \boldsymbol{\Phi}_{W_D}} \quad & \mathrm{tr}\boldsymbol{\Phi}_{W_L}, \\ \text{s.t.} \quad & \boldsymbol{\Phi}_{W_L} + \boldsymbol{\Phi}_{W_D} \ge 0, \ \text{a.e. } \boldsymbol{\Phi}_{W_L}, \boldsymbol{\Phi}_{W_D} \ge 0, \\ & \boldsymbol{\Phi}_{W_D} \ \text{diagonal}, \\ & \|\boldsymbol{\Phi}_{W_L} + \boldsymbol{\Phi}_{W_D} - \hat{\boldsymbol{\Phi}}_{AR}\|_F \le \delta. \end{aligned} \quad (9)$$

Specifically, the last constraint imposes that the true spectrum $\boldsymbol{\Phi}_{AR}$ belongs to a set 'centered' in the nominal spectrum estimate $\hat{\boldsymbol{\Phi}}_{AR}$ with prescribed tolerance δ. The matrix parametrization for $\boldsymbol{\Phi}_{W_D}$ can be represented as

$$\begin{aligned} \boldsymbol{\Phi}_{W_D}(e^{\mathrm{j}\omega}) &= \Delta(e^{\mathrm{j}\omega})\boldsymbol{D}\Delta^*(e^{\mathrm{j}\omega}), \\ \Delta(e^{\mathrm{j}\omega}) &:= [\boldsymbol{I}_n, e^{-\mathrm{j}\omega}\boldsymbol{I}_n, \cdots, e^{-\mathrm{j}p_2\omega}\boldsymbol{I}_n] \in \mathbb{R}^{n \times n(p_2+1)}, \\ \boldsymbol{D} &= \boldsymbol{\theta}_n\boldsymbol{\theta}_n^{\mathrm{T}} \in \mathbb{S}^{n(p_2+1)}, \\ \boldsymbol{\theta}_n &:= [\boldsymbol{W}_{D,0}^{\mathrm{T}}, \boldsymbol{W}_{D,1}^{\mathrm{T}}, \cdots, \boldsymbol{W}_{D,p_2}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n(p_2+1) \times n}. \end{aligned}$$

Then the constraint $\boldsymbol{\Phi}_{W_D} \ge 0$ is equivalent to $\boldsymbol{D} \ge 0$. The condition '$\boldsymbol{\Phi}_{W_D}$ diagonal' can be expressed as $\mathrm{ofd}_B(\mathcal{D}(\boldsymbol{D})) = 0$ with

$$\begin{aligned} \mathrm{ofd}_B(\mathcal{D}(\boldsymbol{D})) &:= [\mathrm{ofd}(\mathcal{D}(\boldsymbol{D})_0) \ \cdots \ \mathrm{ofd}(\mathcal{D}(\boldsymbol{D})_{p_2})], \\ \mathrm{ofd}(\mathcal{D}(\boldsymbol{D})_j) &:= \begin{bmatrix} 0 & [\mathcal{D}(\boldsymbol{D})_j]_{12} \cdots [\mathcal{D}(\boldsymbol{D})_j]_{1n} \\ [\mathcal{D}(\boldsymbol{D})_j]_{21} & 0 & \cdots [\mathcal{D}(\boldsymbol{D})_j]_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ [\mathcal{D}(\boldsymbol{D})_j]_{n1} & [\mathcal{D}(\boldsymbol{D})_j]_{n2} \cdots & 0 \end{bmatrix}, \\ & j = 0, \cdots, p_2. \end{aligned}$$

By expanding the dimension of $\boldsymbol{\Phi}_{W_L}$ to construct a block diagonal low-rank matrix $\boldsymbol{L} \in \mathbb{S}^{n(p_2+1)}$ satisfying $\sum_{v=0}^{p_2} \boldsymbol{L}_{v,v} = \boldsymbol{\Phi}_{W_L}$ and $\boldsymbol{L}_{u,v} = \boldsymbol{0}$ $(u \ne v)$, the conditions $\boldsymbol{\Phi}_{W_L} \ge 0$ and $\boldsymbol{\Phi}_{W_L} + \boldsymbol{\Phi}_{W_D} \ge 0$ can be rewritten as $\boldsymbol{L} \ge 0$ and $\boldsymbol{L} + \boldsymbol{D} \ge 0$, respectively. The objective function $\mathrm{tr}\boldsymbol{\Phi}_{W_L}$ is equal to $\mathrm{tr}\boldsymbol{L}$. The Frobenius norm constraint is thereby split into

$$\begin{cases} \|\sum_{v=0}^{p_2}(\boldsymbol{L}_{v,v} + \boldsymbol{D}_{v,v}) - \hat{\boldsymbol{R}}_0^{AR}\|_F \le \delta_0, \\ \|\sum_{v=0}^{p_2-1}(\boldsymbol{L}_{v+1,v} + \boldsymbol{D}_{v+1,v}) - \hat{\boldsymbol{R}}_1^{AR}\|_F \le \delta_1, \\ \qquad \vdots \\ \|\boldsymbol{L}_{p_2,0} + \boldsymbol{D}_{p_2,0} - \hat{\boldsymbol{R}}_{p_2}^{AR}\|_F \le \delta_{p_2}, \end{cases}$$

where the tolerance parameters $\delta_0, \cdots, \delta_{p_2}$ can be estimated by a resampling-based method [19]. Given a sequence of $\hat{\boldsymbol{y}}_{AR}(1), \cdots, \hat{\boldsymbol{y}}_{AR}(N)$ generated from $\hat{\boldsymbol{\Phi}}_{AR}(z)$, we can obtain the resampling covariance estimate $\hat{\boldsymbol{R}}_{r,k}^{AR}$ and the random variable $\|\hat{\boldsymbol{R}}_k^{AR} - \hat{\boldsymbol{R}}_{r,k}^{AR}\|_F$ whose distribution can be empirically approximated by the Monte Carlo method. Then δ_{k,α_k} can be computed numerically according to $\Pr(\|\hat{\boldsymbol{R}}_k^{AR} - \hat{\boldsymbol{R}}_{r,k}^{AR}\|_F \le \delta_{k,\alpha_k}) = \alpha_k$ for $k = 0, 1, \cdots, p_2$ by choosing a desired probability $\alpha_k \in (0,1)$.

As a consequence, a matrix reparametrization form of the optimization problem (9) is given by

$$\begin{aligned} \min_{\boldsymbol{L},\boldsymbol{D}} \ & \mathrm{tr}\boldsymbol{L}, \\ \text{s.t. } & \boldsymbol{L} + \boldsymbol{D} \ge 0, \ \text{a.e. } \boldsymbol{L}, \boldsymbol{D} \ge 0, \\ & \mathrm{ofd}_B(\mathcal{D}(\boldsymbol{D})) = 0, \\ & \|\sum_{v=0}^{p_2-k} (\boldsymbol{L}_{v+k,v} + \boldsymbol{D}_{v+k,v}) - \hat{\boldsymbol{R}}_k^{AR}\|_F \le \delta_k, \ k = 0, \cdots, p_2. \end{aligned} \tag{10}$$

Based on the optimal solutions $\hat{\boldsymbol{L}}$ and $\hat{\boldsymbol{D}}$, the latent-variable and MA spectrums can be recovered as $\hat{\boldsymbol{\Phi}}_{W_L} = \sum_{v=0}^{p_2} \hat{\boldsymbol{L}}_{v,v}$ and $\hat{\boldsymbol{\Phi}}_{W_D} = \Delta\hat{\boldsymbol{D}}\Delta^*$, respectively, and the estimated number of latent variables is $\hat{l} = \mathrm{rank}(\hat{\boldsymbol{\Phi}}_{W_L})$.

C. Iterative Update of Graphical AR Estimates

According to Remark 3, the sparse AR dynamics estimated by (8) or its vector form can determine the graph topology $\mathcal{E}_n$ for the observable variables. The accuracy of the AR dynamics obtained from this sparse regularized Yule-Walker optimization framework depends on the accuracy of the sampled autocovariance matrix $\hat{\boldsymbol{R}}_k$. If the number of sampling data N is large enough, the estimates $\hat{\boldsymbol{R}}_k$ are close to true values, and we can obtain an efficient latent-variable graphical ARMA model estimate by following a two-stage identification method with stage 1 described in Sec. IV-A and stage 2 shown in Sec. IV-B. However, in the limited sampling data case, redundant links may be selected due to inaccurate sparse AR entry estimates, so that the accurate graph topology cannot be obtained. Taking account of this case, we can make use of the two-stage identified results as prior information to update the graphical AR part for the further reduction of the estimation error.

Given the MA spectrum estimate $\hat{\boldsymbol{\Phi}}_{W_D}(z)$ and sampled output spectrum $\hat{\boldsymbol{\Phi}}_y(z)$, we can construct an optimization problem with respect to the graphical AR part based on (2):

$$\begin{aligned} \min_{\boldsymbol{A}(z)} \ & \phi_*[\boldsymbol{A}(z)\hat{\boldsymbol{\Phi}}_y(z)\boldsymbol{A}^*(z) - \hat{\boldsymbol{\Phi}}_{W_D}(z)] \\ & + \gamma\phi_1[\boldsymbol{A}^*(z)\boldsymbol{A}(z)], \end{aligned} \tag{11}$$

where $\gamma > 0$, ϕ_* and ϕ_1 are penalty functions inducing low-rank structure and sparsity, respectively. The low-rank term is further written as

$$\begin{aligned} & \phi_*[\boldsymbol{A}(z)\hat{\boldsymbol{\Phi}}_y(z)\boldsymbol{A}^*(z) - \hat{\boldsymbol{\Phi}}_{W_D}(z)] \\ &= \mathrm{tr}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \boldsymbol{A}(e^{\mathrm{j}\omega})\hat{\boldsymbol{\Phi}}_y(e^{\mathrm{j}\omega})\boldsymbol{A}^*(e^{\mathrm{j}\omega}) - \hat{\boldsymbol{\Phi}}_{W_D}(e^{\mathrm{j}\omega})d\omega\right) \\ &= \mathrm{tr}([\boldsymbol{I}_n \ \boldsymbol{\theta}_A^{\mathrm{T}}]\hat{\boldsymbol{K}}[\boldsymbol{I}_n; \boldsymbol{\theta}_A]) - \mathrm{tr}(\hat{\boldsymbol{D}}) \\ &= \mathrm{tr}(\hat{\boldsymbol{K}}\boldsymbol{X}) - \mathrm{tr}(\hat{\boldsymbol{D}}), \end{aligned} \tag{12}$$

where

$$\begin{aligned} \hat{\boldsymbol{K}} &:= \begin{bmatrix} \hat{\boldsymbol{R}}_0 & \hat{\boldsymbol{R}}_1 & \cdots & \hat{\boldsymbol{R}}_{p_1} \\ \hat{\boldsymbol{R}}_1^{\mathrm{T}} & \hat{\boldsymbol{R}}_0 & \cdots & \hat{\boldsymbol{R}}_{p_1-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{R}}_{p_1}^{\mathrm{T}} & \hat{\boldsymbol{R}}_{p_1-1}^{\mathrm{T}} & \cdots & \hat{\boldsymbol{R}}_0 \end{bmatrix} \in \mathbb{S}^{n(p_1+1)}, \\ \boldsymbol{X} &= [\boldsymbol{I}_n; \boldsymbol{\theta}_A][\boldsymbol{I}_n \ \boldsymbol{\theta}_A^{\mathrm{T}}] \in \mathbb{S}^{n(p_1+1)}. \end{aligned}$$

As introduced in [8], [11], [31], the sparsity term can be defined as

$$\begin{aligned} \phi_1[\boldsymbol{A}^*(z)\boldsymbol{A}(z)] &= h_\infty(\boldsymbol{Q}) \\ &= \sum_{j>i} \max_{k=0,\cdots,p_1} \left\{ |[\boldsymbol{Q}_k]_{ij}|, |[\boldsymbol{Q}_k]_{ji}| \right\} \end{aligned} \tag{13}$$

with $\boldsymbol{Q}_0 := \sum_{v=0}^{p_1} \boldsymbol{X}_{v,v}$ and $\boldsymbol{Q}_k := 2\sum_{v=0}^{p_1-k} \boldsymbol{X}_{v,v+k}$ for $k = 1, \cdots, p_1$.

We refer to the sparse AR dynamics estimates obtained from (8) as $\hat{\boldsymbol{A}}_j^{\mathrm{ini}}$ $(j = 1, \cdots, p_1)$ and place restrictions on the sparsity pattern of $\boldsymbol{X}$ by utilizing the corresponding graphical structure, that is

$$\boldsymbol{X}_{0,j} \in \mathrm{supp}(\hat{\boldsymbol{A}}_j^{\mathrm{ini}}), \ j = 1, \cdots, p_1. \tag{14}$$

Let $\rho := \frac{\gamma}{1+\gamma} \in (0,1)$. By virtue of (12)–(14), we can obtain a constrained matricial reformulation of (11):

$$\begin{aligned} \min_{\boldsymbol{X}} \ & (1-\rho)[\mathrm{tr}(\hat{\boldsymbol{K}}\boldsymbol{X}) - \mathrm{tr}(\hat{\boldsymbol{D}})] + \rho h_\infty(\boldsymbol{Q}), \\ \text{s.t. } & \boldsymbol{X}_{0,j} \in \mathrm{supp}(\hat{\boldsymbol{A}}_j^{\mathrm{ini}}), \ j = 1, \cdots, p_1. \end{aligned} \tag{15}$$

Given a string of penalty parameters ρ chosen from the range $(0,1)$ via cross-validation, the graphical AR part can be updated with AR parameter estimates $\hat{\boldsymbol{A}}_1, \cdots, \hat{\boldsymbol{A}}_{p_1}$ computed from the optimal solutions $\hat{\boldsymbol{X}}$ by spectral decomposition and the graph topology estimates $\hat{\mathcal{E}}_n$ determined from the sparsity pattern of $\hat{\boldsymbol{A}}^*(z)\hat{\boldsymbol{A}}(z)$. Note that $\hat{\mathcal{E}}_n$ can be more accurate after normalization and filtering [23]. The optimization model (15) exploits the acquired MA part estimate and preliminary AR graphical structure, which can reduce the number of parameters to be identified and further improve the estimate accuracy for the graphical AR part. In practice, we can take the two-stage estimates obtained from (8) and (10) as the initial values, and estimate the AR graphical structure with (15) and the latent plus MA dynamics with (10) in an alternative way [18].

Let k denote the iterative number. Then the parameter estimates of the k-th iteration are denoted by $\hat{\boldsymbol{A}}_k(z)$, $\hat{\boldsymbol{\Phi}}_{W_L,k}$, $\hat{\boldsymbol{\Phi}}_{W_D,k}(z)$, $\hat{l}_k$, $\hat{\boldsymbol{\theta}}_{A,k}$, $\hat{\boldsymbol{L}}_k$ and $\hat{\boldsymbol{D}}_k$. Inspired by [8], we define a model fitness function as

$$\begin{aligned} f &= \mathbb{D}(\hat{\boldsymbol{\Phi}}_y^{NP} \| \hat{\boldsymbol{\Phi}}_{y,k}^{P}) \times p_k, \\ \mathbb{D}(\hat{\boldsymbol{\Phi}}_y^{NP} \| \hat{\boldsymbol{\Phi}}_{y,k}^{P}) &:= \frac{1}{2}\Big\{ \frac{1}{2\pi}\int_{-\pi}^{\pi} \Big[\log\det\Big(\Big(\hat{\boldsymbol{\Phi}}_y^{NP}\Big)^{-1}\hat{\boldsymbol{\Phi}}_{y,k}^{P}\Big) \\ &\quad + \Big\langle \hat{\boldsymbol{\Phi}}_y^{NP}, \Big(\hat{\boldsymbol{\Phi}}_{y,k}^{P}\Big)^{-1} \Big\rangle \Big] d\omega - n \Big\}, \\ p_k &= |\mathrm{supp}(\hat{\boldsymbol{A}}_k^*(z)\hat{\boldsymbol{A}}_k(z))| - n + n\hat{l}_k, \end{aligned}$$

where $\hat{\boldsymbol{\Phi}}_y^{NP}$ is the smoothed non-parametric spectral estimate of the process $\boldsymbol{y}(t)$, and $\hat{\boldsymbol{\Phi}}_{y,k}^{P}$ is the k-th parametric spectral estimate computed by

$$\hat{\boldsymbol{\Phi}}_{y,k}^{P}(z) = \hat{\boldsymbol{A}}_k(z)^{-1}[\hat{\boldsymbol{\Phi}}_{W_L,k-1} + \hat{\boldsymbol{\Phi}}_{W_D,k-1}(z)]\hat{\boldsymbol{A}}_k^*(z)^{-1}.$$

The best value of the penalty parameter ρ in (15) and the corresponding graphical AR topology estimate $\hat{\mathcal{E}}_n$ will be selected according to the lowest score f.

Define the mean square difference between identified parameters of two sequential iterations as $e_k := \|\hat{\boldsymbol{\theta}}_{A,k} - \hat{\boldsymbol{\theta}}_{A,k-1}\|_F^2/p_1n^2 + \|\hat{\boldsymbol{L}}_k - \hat{\boldsymbol{L}}_{k-1}\|_F^2/[n(p_2+1)]^2 + \|\hat{\boldsymbol{D}}_k - \hat{\boldsymbol{D}}_{k-1}\|_F^2/n(p_2+1)$. If $e_k - e_{k-1} \le \varepsilon$, where ε takes a small positive value, stop the iteration.

V. Simulation Example

Consider the latent-variable ARMA graphical model with $n = 10$ observable variables and $l = 1$ latent variables:

$$\boldsymbol{y}(t) + \boldsymbol{A}_1\boldsymbol{y}(t-1) + \boldsymbol{A}_2\boldsymbol{y}(t-2) = \boldsymbol{W}_L\boldsymbol{x}(t) + \boldsymbol{W}_{D,0}\boldsymbol{\omega}(t) + \boldsymbol{W}_{D,1}\boldsymbol{\omega}(t-1), \quad (16)$$

where $\boldsymbol{A}_j \in \mathbb{R}^{10\times10}$ $(j = 1,2)$ are sparse matrices with non-zero entries $[\boldsymbol{A}_1]_{53} = [\boldsymbol{A}_1]_{67} = [\boldsymbol{A}_1]_{8,10} = [\boldsymbol{A}_2]_{12} = [\boldsymbol{A}_2]_{35} = [\boldsymbol{A}_2]_{96} = 0.5$, the entries of $\boldsymbol{W}_L \in \mathbb{R}^{10\times1}$ are randomly generated within $(0,1)$ and $\boldsymbol{W}_{D,i} \in \mathbb{R}^{10\times10}$ $(i = 0,1)$ are randomly generated such that the zeros of the transfer functions $[\boldsymbol{W}_D]_{vv}(z)$ $(v = 1,\cdots,n)$ are within the circle with radius 0.95 on the complex plane.

Let the the sample size be $N = 1000$, $N = 2000$, and $N = 5000$, respectively. The preliminary AR dynamics is estimated with the vector form of (8) solved by using 10-fold cross-validated LASSO, and the graph topology estimates $\hat{\mathcal{E}}_n$ for the observable variables corresponding to the minimum cross-validated mean squared error are depicted in Fig. 1. As N increases, the number of selected spurious links is reduced, and the estimated graph topology $\hat{\mathcal{E}}_n$ is closer to the true graph topology $\mathcal{E}_n$. The proposed sparse plus low-rank based iterative identification algorithm is applied to update the preliminary estimates with δ_{k,α_k} $(k = 0,1)$ computed from 200 Monte Carlo experiments for $\alpha_k = 0.5$. For convenience, we only give the graph topology estimates $\hat{\mathcal{E}}_n$ for different regularization parameters ρ and the corresponding fitness function scores f in the case of $N = 1000$ (shown in Fig. 2). Compared with that of the preliminary case, the iterative graph topology estimates are more sparse with redundant edges further reduced. The final selected $\hat{\mathcal{E}}_n$ is marked in red for $\rho = 0.3$ with the minimum score $f = 9.88$, which is identical to true $\mathcal{E}_n$. In the cases of larger sample size, accurate graph topology estimates can also be obtained based on more accurate preliminary information. The preliminary and iteratively updated non-zero entry estimates in $\hat{\boldsymbol{A}}_j$ $(j = 1,2)$ are tabulated in Table I. Note that some redundant estimates of zero entries are not listed. We define the AR parameter estimation error including these estimates as $\delta_A := \|\boldsymbol{\theta}_A - \hat{\boldsymbol{\theta}}_A\|_F/\|\boldsymbol{\theta}_A\|_F$ and list it for each case in the last column of Table I. The sparse plus low-rank based iterative procedure can improve the estimation precision, and the estimation precision increases with the increment of the amount of sampled data.

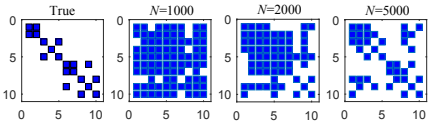

Fig. 1. Preliminary graph topology $\hat{\mathcal{E}}_n$ against the sample size N

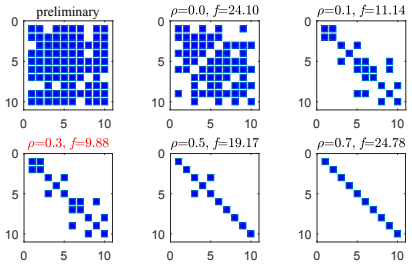

Fig. 2. Estimated graph topology $\hat{\mathcal{E}}_n$ against the regularization parameter ρ for $N = 1000$

To provide empirical estimates of the latent-variable and MA part, we perform 20 Monte Carlo runs for the sparse plus low-rank based iterative identification of Model (16) with the graphical AR part fixed but $\boldsymbol{W}_L$, $\boldsymbol{W}_{D,i}$ $(i = 0,1)$ randomly generated as described before, and $N = 5000$ samples are taken for each Monte Carlo experiment. The box-plots of the estimation errors $\delta_{\Phi_{WL}} := \|\boldsymbol{\Phi}_{W_L} - \hat{\boldsymbol{\Phi}}_{W_L}\|_F/\|\boldsymbol{\Phi}_{W_L}\|_F$ and $\delta_{\Phi_{WD}}(e^{\mathrm{j}\omega}) := \|\boldsymbol{\Phi}_{W_D}(e^{\mathrm{j}\omega}) - \hat{\boldsymbol{\Phi}}_{W_D}(e^{\mathrm{j}\omega})\|_F/\|\boldsymbol{\Phi}_{W_D}(e^{\mathrm{j}\omega})\|_F$ for $\omega \in [0,\pi]$ are shown in Fig. 3, indicating that the latent-variable and MA spectrums can be recovered with negligible numerical errors of the order 10^{-1}. The box-plot of the singular values σ_j $(j = 1,\cdots,10)$ of $\hat{\boldsymbol{\Phi}}_{W_L}$ are shown in Fig. 4, revealing that the number of latent variables can be accurately recovered as $\hat{l} = 1$.

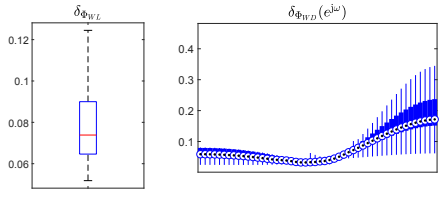

Fig. 3. Box-plot of the estimation errors $\delta_{\Phi_{WL}}$ and $\delta_{\Phi_{WD}}(e^{\mathrm{j}\omega})$

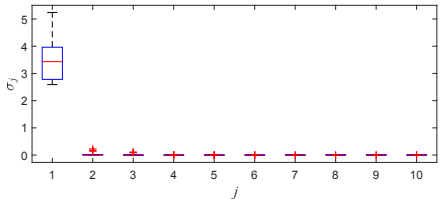

Fig. 4. Box-plot of the singular values of $\hat{\boldsymbol{\Phi}}_{W_L}$

TABLE I

THE SPARSE AR PARAMETER ESTIMATES OBTAINED FROM PRELIMINARY TWO-STAGE AND SPARSE PLUS LOW-RANK BASED ITERATIVE METHODS AND ESTIMATION ERRORS δ_A FOR DIFFERENT SAMPLE SIZE N

	AR parameter matrices	$\hat{\boldsymbol{A}}_1$	$\hat{\boldsymbol{A}}_1$	$\hat{\boldsymbol{A}}_1$	$\hat{\boldsymbol{A}}_2$	$\hat{\boldsymbol{A}}_2$	$\hat{\boldsymbol{A}}_2$	δ_A
	Non-zero positions	(5,3)	(6,7)	(8,10)	(1,2)	(3,5)	(9,6)	
N=1000	Two-stage estimates	0.3465	0.3409	0.4258	0.3699	0.4406	0.4996	32.23%
	Iteratively updated estimates	0.4841	0.4771	0.5361	0.4741	0.4686	0.5197	5.24%
N=2000	Two-stage estimates	0.5066	0.4074	0.4250	0.5200	0.4815	0.4783	11.93%
	Iteratively updated estimates	0.5062	0.5072	0.4924	0.5433	0.4912	0.4814	4.04%
N=5000	Two-stage estimates	0.4466	0.4189	0.2752	0.4423	0.4958	0.5046	20.87%
	Iteratively updated estimates	0.4848	0.5009	0.4925	0.4861	0.5101	0.5164	2.39%
	True values	0.5	0.5	0.5	0.5	0.5	0.5	

VI. CONCLUSIONS

An identification approach for latent-variable graphical ARMA models has been developed in this paper, which alternatingly estimates the graphical AR part by sparse plus low-rank optimization and the latent MA part through low-rank plus diagonal decomposition technique based on the preliminary regularized Yule-Walker estimates. Simulation results have demonstrated the efficacy of the proposed method. The identification procedure is sub-optimal and theoretical properties of the sequential strategy, such as convergence, are open problems which can be investigated for future research. In addition, the practical use of the work will be further explored. For example, we can apply the proposed approach to neural dynamics learning.

REFERENCES

[1] L. Du, H. A. Wang, J. Zhang, et al, "Adaptive structured sparse multiview canonical correlation analysis for multimodal brain imaging association identification," Sci. China Inf. Sci, vol. 66, no. 4, pp. 142106:1-142106:16, 2023.

[2] W. F. Shi, S. Q. Xu, T. L. Fan, et al, "Cost effective approach to identify multiple influential spreaders based on the cycle structure in networks," Sci. China Inf. Sci, vol. 66, pp. 192203:1-192203:10, 2023.

[3] C. F. Hu, H. Hu, H. W. Lin, et al, "Isogeometric analysis-based topological optimization for heterogeneous parametric porous structures," J Syst Sci Complex, vol. 36, pp. 29-52, 2023.

[4] R. H. Yang, L. Liu, G. Feng, "An overview of recent advances in distributed coordination of mlti-agent systems," Unmanned Systems, vol. 10, no. 3, pp. 307-325, 2022.

[5] T. Richardson, and P. Spirtes, "Ancestral graph Markov models," The Annals of Statistics, vol. 30, no. 4, pp. 962-1030, 2002.

[6] M. J. Choi, V. Y. F. Tan, A. Anandkumar, et al, "Learning latent tree graphical models," Journal of Machine Learning Research, vol. 12, pp. 1771-1812, 2011.

[7] F. Sepehr, and D. Materassi, "An algorithm to learn polytree networks with hidden nodes," Neural Information Processing Systems, 2019.

[8] M. Zorzi, and R. Sepulchre, "AR identification of latent-variable graphical models," IEEE Transactions on Automatic Control, vol. 61, no. 9, pp. 2327-2340, 2016.

[9] V. Chandrasekaran, P. Parrilo, and A. Willsky, "Latent variable graphical model selection via convex optimization," The Annals of Statistics, vol. 40, no. 8, pp. 1935-1967, 2012.

[10] R. Liégeois, B. Mishra, M. Zorzi, et al, Sparse plus low-rank autoregressive identification in neuroimaging time series. In Proc. *2015 54th IEEE Conference on Decision and Control (CDC)*, Osaka, Japan, 2015, pp. 3965-3970.

[11] V. Ciccone, A. Ferrante, and M. Zorzi, "Learning latent variable dynamic graphical models by confidence sets selection," IEEE Transactions on Automatic Control, vol. 65, no. 12, pp. 5130-5143, 2020.

[12] D. Alpago, M. Zorzi, and A. Ferrante, "A scalable strategy for the identification of latent-variable graphical models," IEEE Transactions on Automatic Control, vol. 67, no. 7, pp. 3349-3362, 2022.

[13] M. Zorzi, and A. Chiuso, "A Bayesian approach to sparse plus low rank network identification," In Proc. *2015 IEEE 54th Annual Conference on Decision and Control (CDC)*, Osaka, Japan, 2015, pp. 7386-7391.

[14] M. Zorzi, and A. Chiuso, "Sparse plus low rank network identification: A nonparametric approach," Automatica, vol. 76, pp. 355-366, 2017.

[15] W. M. Fu, J. H. Qin, Q. Ling, et al, "Trust-region based stochastic variational inference for distributed and asynchronous networks," J Syst Sci Complex, vol. 35, pp. 2062-2076, 2022.

[16] J. Z. Pan, and Q. W. Yao, "Modelling multiple time series via common factors," Biometrika, vol. 95, no. 2, pp. 365-379, 2008.

[17] V. Ciccone, A. Ferrante, and M. Zorzi, "Factor models with real data: A robust estimation of the number of factors," IEEE Transactions on Automatic Control, vol. 64 no. 6, pp. 2412-2425, 2019.

[18] F. Crescente, L. Falconi, F. Rozzi, A. Ferrante, et al, "Learning AR factor models," In Proc. *2020 59th IEEE Conference on Decision and Control (CDC)*, Jeju, Korea (South), 2020, pp. 274-279.

[19] L. Falconi, A. Ferrante, and M. Zorzi, "A robust approach to ARMA factor modeling," arXiv:2107.03873, 2021.

[20] You, J. Y., & Yu, C. P. (2023). Sparse plus low-rank identification for dynamical latent-variable graphical AR models. ArXiv:.

[21] E. Avventi, A. Lindquist, and B. Wahlberg, "ARMA identification of graphical models," IEEE Transactions on Automatic Control, vol. 58, no. 5, pp. 1167-1178, 2013.

[22] D. Alpago, M. Zorzi, and A. Ferrante, "Data-driven Link Prediction over Graphical Models," IEEE Transactions on Automatic Control, doi: 10.1109/TAC.2021.3137157, 2021.

[23] J. Y. You, C. P. Yu, J. Sun, et al, "Generalized maximum entropy based identification of graphical ARMA models," Automatica, vol. 141, no. 110319, pp. 1-9, 2022.

[24] R. A. Horn, and C. R. Johnson, *Matrix Analysis (2nd ed).* New York, USA: Cambridge University Press, 2012.

[25] M. S. Veedu, H. Doddi, and M. V. Salapaka, "Topology learning of linear dynamical systems with latent nodes using matrix decomposition," IEEE Transactions on Automatic Control, vol. 67 no. 11, pp. 5746-5761, 2022.

[26] R. Dahlhaus, "Graphical interaction models for multivariate time series," Metrika, vol. 51, no. 2, pp. 157-172, 2000.

[27] B. Friedlander, and B. Porat, "The modified Yule-Walker method of ARMA spectral estimation," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-20 no. 2, pp. 158-173, 1984.

[28] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing.* New York, USA: Springer-Verlag, 2010.

[29] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," IEEE Transactions on Information Theory, vol. 52, no. 3, pp. 1030-1051, 2006.

[30] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), vol. 58, no. 1, pp. 267-288, 1996.

[31] J. Songsiri, and L. Vandenberghe, "Topology selection in graphical models of autoregressive processes," Journal of Machine Learning Research, vol. 11, pp. 2671-2705, 2010.