

A Scalable Strategy for the Identification of Latent-variable Graphical Models

Daniele Alpagò, Mattia Zorzi, Augusto Ferrante

Abstract—In this paper we propose an identification method for latent-variable graphical models associated to autoregressive (AR) Gaussian stationary processes. The identification procedure exploits the approximation of AR processes through stationary reciprocal processes thus benefiting of the numerical advantages of dealing with block-circulant matrices. These advantages become more and more significant as the order of the process gets large. We show how the identification can be cast in a regularized convex program and we present numerical examples that compares the performances of the proposed method with the existing ones.

Index Terms—Latent-variable graphical models, Reciprocal processes, Maximum likelihood, Maximum entropy, Regularization, System identification.

I. INTRODUCTION

THE ideas behind graphical models have their origins in several scientific areas, such as statistical physics and genetics back at the beginning of the last century. However, only recent developments of such ideas allowed to employ graphical models in identification problems involving high dimensional data [1], [2], [3], [4], [5], [6], [7], [8]. In this direction, particularly useful are *sparse graphical models*, i.e. graphs with few edges that describe the interactions between a large number of variables. Such models have become very popular in the literature in the recent years because, beside giving a concise representation of the phenomenon under scrutiny, sparsity implies a limited number of model's parameters thus avoiding overfitting in the identification procedure.

Although the latter is a desirable property, enforcing sparsity in the identification procedure is not always the best choice, as it may prevent a sufficiently rich description of the underlying phenomenon. Indeed, in many practical situations, the presence of a few common, hidden behaviors between the variables of interest explaining the most part of the interactions between the observed variables can be crucial. The fact that a sparse graphical model is not able to describe the essential features of this kind of phenomena motivates the introduction of the so-called *latent-variable graphical models*. The latter consist in a two-layer graph where the conditional dependence relations between the observed variables are mainly due to the latent variables (i.e. variables not accessible to observations): each latent-variable (on the top-layer) is connected to the majority

of the observed variables (on the bottom-layer), making the latter a sparse subgraph. Since the number of latent-variables is small, the overall graph has a reduced number of edges. In the simplest possible setting, one can associate this kind of models to a Gaussian random vector [8]. The particular graphical structure translates in a sparse plus low-rank decomposition of its concentration matrix. In [8] the identification of the sparse and the low-rank part of the concentration matrix has been cast in a regularized maximum-likelihood optimization problem. A dynamic version of this problem, i.e. the identification of latent-variable graphical models for AR Gaussian processes, has been considered in [9] where the problem has been shown to be strictly connected to a maximum-entropy problem. As showed in [10], this identification problem can be effectively solved by an ADMM-type algorithm. The optimization procedure, however, involves the inversion and the eigenvalue decomposition of matrices whose dimension is proportional to the product of the order of the process by the dimension of the process, making the procedure numerically critical when the order of the AR process is high, as it happens, for example, when the AR process is an approximation of an ARMA one. In this paper we consider the problem of identifying latent-variable graphical models for stationary Gaussian reciprocal processes. The latter are periodic stationary processes [11], [12], [13], [14], [15], [16], [17] and they have been proven to be a worthy approximation of Gaussian AR processes, provided that the period N is sufficiently large [18], [16]. We will show that the proposed identification procedure is in fact an approximation of the maximum entropy and maximum likelihood identification paradigms proposed for the classical AR processes. The fact that stationary reciprocal process can be modeled by means of block-circulant matrices represents a big numerical advantage as the inversion and the eigenvalue decomposition of such matrices can be performed robustly [19] making the proposed procedure attractive also for the identification AR processes of high order and hence for ARMA processes.

The paper is organized as follows: In Section II we fix the notation and we recall the fundamental results used in the rest of the paper. In Section III we introduce reciprocal processes and we explain how they are related to AR processes. In Section IV we characterize graphical models associated to reciprocal processes while, in Section V, we propose a convex optimization problem for the identification of such models. Section VI is devoted to the ADMM formulation of the optimization problem and Section VII reports numerical experiments concerning the implementation of the proposed procedure. Finally, in Section VIII we draw the conclusions.

D. Alpagò, M. Zorzi and A. Ferrante are with the Department of Information Engineering, University of Padova, Padova, Italy; email: alpagodani@dei.unipd.it (D. Alpagò) zorzimat@dei.unipd.it (M. Zorzi) augusto@dei.unipd.it (A. Ferrante)

II. NOTATION AND BACKGROUND

In this paper we will deal both with real matrices and with matrix-valued functions defined on the unit-circle $\mathbb{T} := \{e^{i\theta} : \theta \in [-\pi, \pi]\}$. For such functions we will omit the dependence on θ when it is clear from the context, i.e. we will write F in place of $F(e^{i\theta})$. The rank of a matrix G is denoted by $\text{rank}(G)$ while the (normal) rank of any $\mathbb{C}^{p \times p}$ -valued analytic function F defined on \mathbb{T} , is defined as

$$\text{rank}(F) := \max_{\theta \in [-\pi, \pi]} \text{rank}(F(e^{i\theta})). \quad (1)$$

In the same fashion, the following notations will be used indifferently in the case that G is a $\mathbb{C}^{p \times p}$ -valued function defined on \mathbb{T} or a square constant matrix: G^\top denotes the transpose of G , G^* its transpose-conjugate and $\text{diag}(G) \in \mathbb{C}^p$ denotes the vector whose entries are the diagonal elements of G . $\ker(G)$ indicates the kernel of G . $G > 0$ and $G \geq 0$ denote that G is a positive definite and, respectively, positive semidefinite. $\text{tr}(G)$, $\det(G)$ and G^{-1} denote the trace of G , the determinant of G and its inverse, respectively. I_p denotes the identity matrix of order p .

We define the cone

$$\mathcal{S}_p := \{F \in \mathcal{H}_p : \Phi - \alpha I_p \geq 0 \text{ a.e. on } \mathbb{T}, \text{ for some } \alpha > 0\},$$

where \mathcal{H}_p is the space of square integrable coercive functions defined on the unit circle and taking values in the space of $p \times p$ Hermitian matrices. For any $F \in \mathcal{H}_p$ we will use equivalently the notations

$$\int_{-\pi}^{\pi} F(e^{i\theta}) \frac{d\theta}{2\pi}, \quad \int F$$

for the integral of F over $[-\pi, \pi]$ with respect to the normalized Lebesgue measure on \mathbb{T} . We define also the family of matrix pseudo-polynomials

$$\mathcal{Q}_{p,n} := \left\{ \sum_{k=-n}^n Q_k e^{i\theta k}, \quad Q_{-k} = Q_k^\top \in \mathbb{R}^{p \times p} \right\}.$$

For any sub-interval $(x_1, x_2) := \{x : x_1 < x < x_2\}$ of an interval $(a, b) \subset \mathbb{R}$, we denote with $(x_1, x_2)^c$ the complement set of (x_1, x_2) in (a, b) . $\mathbb{E}[\cdot]$ denotes the expectation operator.

In this paper we will always consider AR processes of order n and reciprocal processes of period N , i.e. completely specified in a finite interval of length N . All such processes are understood with zero mean throughout the paper. It will be always assumed that $N > 2n$ and that N is an even number. The case with N odd can be dealt in a similar way. We define the vector space $\mathcal{C} \subset \mathbb{R}^{mN \times mN}$ of the (real) symmetric, block-circulant matrices

$$\mathbf{C} = \text{circ}\{C_0, C_1, \dots, C_{\frac{N}{2}-1}, C_{\frac{N}{2}}, C_{\frac{N}{2}-1}^\top, \dots, C_1^\top\},$$

whose first block-column is composed by the $m \times m$ blocks $C_0, C_1, \dots, C_{\frac{N}{2}-1}, C_{\frac{N}{2}}, C_{\frac{N}{2}-1}^\top, \dots, C_1^\top$. The space \mathcal{C} is endowed with the inner product $\langle \mathbf{C}, \mathbf{D} \rangle_{\mathcal{C}} := \text{tr}(\mathbf{C}^\top \mathbf{D})$. The symbol of the block-circulant matrix $\mathbf{C} \in \mathcal{C}$ is defined as the $m \times m$ pseudo-polynomial

$$\Phi(\zeta) := \sum_{k=0}^{N-1} C_k \zeta^{-k}, \quad \text{with} \quad C_k = C_{N-k}^\top \text{ for } k > \frac{N}{2}, \quad (2)$$

where $\zeta := e^{i\frac{2\pi}{N}}$ is the N -th root of unity.

Proposition 1: Let \mathbf{C} be a block-circulant matrix with symbol $\Phi(\zeta)$ defined by (2). Then

$$\mathbf{C} = \mathbf{F}^* \text{diag}\{\Phi(\zeta^0), \Phi(\zeta^1), \dots, \Phi(\zeta^{N-1})\} \mathbf{F}, \quad (3)$$

where \mathbf{F} is the (Fourier) unitary block-matrix

$$\mathbf{F} = \frac{1}{\sqrt{N}} \begin{bmatrix} \zeta^{-0 \cdot 0} I & \zeta^{-0 \cdot 1} & \dots & \zeta^{-0 \cdot (N-1)} I \\ \zeta^{-1 \cdot 0} I & \zeta^{-1 \cdot 1} I & \dots & \zeta^{-1 \cdot (N-1)} I \\ \vdots & \vdots & \ddots & \vdots \\ \zeta^{-(N-1) \cdot 0} I & \zeta^{-(N-1) \cdot 1} I & \dots & \zeta^{-(N-1) \cdot (N-1)} I \end{bmatrix}.$$

This is a classical result in the scalar case; technical details for the block-circulant case can be found, for instance, in [20, page 6]. We define the subspace $\mathcal{B} \subseteq \mathcal{C}$ of symmetric, banded block-circulant $mN \times mN$ matrices of bandwidth n , with $N > 2n$, containing the matrices of the form

$$\mathbf{B} = \text{circ}\{B_0, B_1, \dots, B_n, 0, \dots, 0, B_n^\top, \dots, B_1^\top\}, \quad (4)$$

that inherits the inner product defined on \mathcal{C} . Note that, according to definition (2), the symbol of a banded matrix $\mathbf{B} \in \mathcal{B}$ is

$$\Psi(\zeta) = \sum_{k=-n}^n B_k \zeta^{-k}, \quad B_{-k} = B_k^\top.$$

The projection operator $P_{\mathcal{B}} : \mathcal{C} \rightarrow \mathcal{B}$ is defined as

$$P_{\mathcal{B}}(\mathbf{C}) := \text{circ}\{C_0, C_1, \dots, C_n, 0, \dots, 0, C_n^\top, \dots, C_1^\top\}.$$

Given $\Omega = \{(i, j) : i, j = 1, \dots, m\}$, the projection operator $P_{\Omega} : \mathcal{C} \rightarrow \mathcal{C}$ is defined such that $P_{\Omega}(\mathbf{C})$ is a block-circulant matrix whose blocks have support Ω .

III. RECIPROCAL PROCESSES

Let $\{\mathbf{y}(k), k = 1, 2, \dots, N\}$, be an m -dimensional Gaussian stationary stochastic process defined on a finite interval $[1, N]$. For $k = 1, \dots, N$, we have $\mathbf{y}(k) := [y_1(k) \dots y_m(k)]^\top \in \mathbb{R}^m$, therefore the process is completely characterized by the random vector $\mathbf{y} := [y_1(1) \dots y_m(1) \dots y_1(N) \dots y_m(N)]^\top \in \mathbb{R}^{mN}$. In [11] it has been shown that \mathbf{y} is a restriction of a wide-sense stationary periodic process of period N defined on the whole integer line \mathbb{Z} if and only if the $mN \times mN$ covariance matrix Σ of \mathbf{y} is symmetric block-circulant:

$$\Sigma = \text{circ}\{\Sigma_0, \Sigma_1, \dots, \Sigma_{\frac{N}{2}}, \dots, \Sigma_1^\top\}, \quad (5)$$

where $\mathbb{E}[\mathbf{y}(i)\mathbf{y}(j)^\top] = \Sigma_{i-j}$, $i, j = 1, \dots, N$, are the covariance lags of the process such that $\Sigma_k = \Sigma_{N-k}^\top$ for $k > N/2$. In view of the above equivalence, we will denote with \mathbf{y} both the wide-sense stationary periodic process defined in the whole line \mathbb{Z} and its restriction, depending on the context. A particular class of stationary periodic processes is represented by reciprocal processes.

Definition 1: \mathbf{y} is a reciprocal process of order n on $[1, N]$ if, for all $t_1, t_2 \in [1, N]$, the random variables of the process in the interval $(t_1, t_2) \subset [1, N]$ are conditionally independent to the random variables in $(t_1, t_2)^c$, given the $2n$ boundary values $\mathbf{y}(t_1 - n + 1), \dots, \mathbf{y}(t_1), \mathbf{y}(t_2), \dots, \mathbf{y}(t_2 + n - 1)$, where the sums $t - k$ and $t + k$ are to be understood modulo N .

The following result has been proved in [11, Theorem 3.3]: it states that a reciprocal process is completely specified by a block-circulant matrix whose inverse has a banded structure.

Theorem 1: A non-singular $mN \times mN$ -dimensional matrix Σ is the covariance matrix of a periodic reciprocal process of order n if and only if its inverse is a positive definite symmetric block-circulant matrix which is banded of bandwidth n , namely $\Sigma^{-1} \in \mathcal{B}$.

Let $\hat{\Sigma}_0, \dots, \hat{\Sigma}_n$ be given estimates of the first $n+1$ covariance lags $\Sigma_0, \dots, \Sigma_n$ of the underlying reciprocal process. In view of Theorem 1, the identification of a reciprocal process can be formulated as the following matrix completion problem.

Problem 1: Given the $n+1$ estimates $\hat{\Sigma}_0, \dots, \hat{\Sigma}_n$, compute a sequence $\Sigma_{n+1}, \dots, \Sigma_{\frac{N}{2}}$, in such a way to form a symmetric, positive definite block-circulant matrix

$$\Sigma = \text{circ}\{\hat{\Sigma}_0, \dots, \hat{\Sigma}_n, \Sigma_{n+1}, \dots, \Sigma_{\frac{N}{2}}, \dots, \Sigma_{n+1}^\top, \hat{\Sigma}_n^\top, \dots, \hat{\Sigma}_1^\top\},$$

with $\Sigma^{-1} \in \mathcal{B}$.

It has been shown in [16], [11] that a particular solution to Problem 1 is the one which solves the following maximum entropy problem:

$$\begin{aligned} & \underset{\Sigma \in \mathcal{C}}{\text{argmax}} && \log \det \Sigma \\ & \text{subject to} && \Sigma > 0 \\ & && P_{\mathcal{B}}(\Sigma - \hat{\Sigma}) = 0. \end{aligned} \quad (6)$$

whose dual problem has been proven to be

$$\begin{aligned} & \underset{\mathbf{X} \in \mathcal{B}}{\text{argmin}} && -\log \det \mathbf{X} + \langle \mathbf{X}, \hat{\Sigma} \rangle_{\mathcal{C}} \\ & \text{subject to} && \mathbf{X} > 0 \end{aligned} \quad (7)$$

where $\hat{\Sigma} \in \mathcal{B}$ is the symmetric, banded block-circulant matrix of bandwidth n ,

$$\hat{\Sigma} = \text{circ}\{\hat{\Sigma}_0, \hat{\Sigma}_1, \dots, \hat{\Sigma}_n, 0, \dots, 0, \hat{\Sigma}_n^\top, \dots, \hat{\Sigma}_1^\top\},$$

containing the covariance lags estimated from the data and the optimal value of dual variable \mathbf{X} is indeed equal to Σ^{-1} , i.e. the inverse of the solution of (6). Strong duality between (6) and (7) implies that (6) and (7) are equivalent. In what follows we assume that $\hat{\Sigma} > 0$ as it is a necessary condition for Problem (6) to be feasible. In the case that $\hat{\Sigma}$ is not positive definite, we can consider a positive definite banded block-circulant matrix sufficiently close to $\hat{\Sigma}$ which can be obtained by solving a structured covariance estimation problem, see [21], [22].

AR approximation: Next we recall how reciprocal processes can be seen as an approximation of autoregressive (AR) processes. More precisely, let $\mathbf{y} := \{\mathbf{y}(t) : t \in \mathbb{Z}\}$ be an m -dimensional, AR, full-rank, Gaussian wide-sense stationary process of order n ,

$$\sum_{k=0}^n B_k \mathbf{y}(t-k) = \mathbf{e}(t), \quad \mathbf{e}(t) \sim \mathcal{N}(0, I_m), \quad t \in \mathbb{Z}, \quad (8)$$

and let $R_k := \mathbb{E}[\mathbf{y}(t)\mathbf{y}(t-k)^\top]$, $k \in \mathbb{Z}$, be its k -th covariance lag. The spectrum of \mathbf{y} is the Fourier transform of the sequence R_k with $k \in \mathbb{Z}$, i.e.

$$\Phi(e^{i\theta}) = \sum_{k=-\infty}^{\infty} R_k e^{-i\theta k}, \quad R_{-k} = R_k^\top, \quad \theta \in [-\pi, \pi]. \quad (9)$$

Suppose now that T observations $y(1), \dots, y(T)$ of the process \mathbf{y} are available, and let

$$\hat{R}_k = \frac{1}{T} \sum_{t=k}^T y(t)y(t-k)^\top, \quad k = 0, 1, \dots, n, \quad (10)$$

be estimates of the first $n+1$ covariance lags R_0, \dots, R_n . The identification of such a process can be cast to a *covariance extension problem*.

Problem 2: Given $n+1$ estimates $\hat{R}_0, \dots, \hat{R}_n$, complete them with a sequence R_{n+1}, R_{n+2}, \dots in such a way that the Fourier transform of the extended (infinite) sequence is a power spectral density.

A particular solution of Problem 2 is the one proposed by J. P. Burg in [23]: choose R_{n+1}, R_{n+2}, \dots maximizing the *entropy rate* of the process, i.e. that solves the following optimization problem

$$\begin{aligned} & \underset{\Phi \in \mathcal{S}_m}{\text{argmax}} && \int \log \det \Phi \\ & \text{subject to} && \int e^{i\theta k} \Phi = \hat{R}_k, \quad k = 0, 1, \dots, n. \end{aligned} \quad (11)$$

The dual of (11) has been shown to be, see for instance [24]:

$$\begin{aligned} & \underset{\Phi^{-1} \in \mathcal{Q}_{m,n}}{\text{argmin}} && \int -\log \det \Phi^{-1} + \langle \Phi^{-1}, \hat{\Phi} \rangle \\ & \text{subject to} && \hat{\Phi} > 0 \end{aligned} \quad (12)$$

where

$$\hat{\Phi}(e^{i\theta}) = \sum_{k=-n}^n \hat{R}_k e^{-i\theta k}, \quad \hat{R}_{-k} = \hat{R}_k^\top, \quad (13)$$

is the truncated periodogram of the process \mathbf{y} . These kind of problems have been extensively studied and generalized in the recent years, see for instance [25], [26], [27], [28], [29], [30], [31].

We recall that, for $N \rightarrow \infty$, Toeplitz matrices can be approximated arbitrarily well by circulant matrices [32, Lemma 4.2]; hence, for $N \rightarrow \infty$, Problem 1 consists in searching a completion that leads to an infinite positive definite block-Toeplitz covariance matrix, i.e. such that the Fourier transform of the resulting extended sequence is a power spectral density. By Theorem 3.1 in [18], for $N \rightarrow \infty$, Problem (6) is the classical Burg's maximum entropy problem whose solution is an AR process of order n . In light of this observation, we can understand the reciprocal process associated to the solution of (7) as an approximation of the AR process solution of the Burg's maximum entropy problem (11). In the following sections we will exploit this approximation for the identification of latent-variable AR graphical models.

The reciprocal approximation just explained has also an interesting interpretation in the frequency domain. Indeed, it corresponds to sampling the spectrum (9) of the AR process \mathbf{y} , over the interval $[-\pi, \pi]$, with sample period $2\pi/N$, thus obtaining the *symbol* of the covariance matrix of the corresponding reciprocal process:

$$\Phi(\zeta) = \sum_{k=0}^{N-1} \Sigma_k \zeta^{-k}, \quad \Sigma_k = \Sigma_{N-k}^\top \text{ for } k > \frac{N}{2}.$$

Figure 1 illustrates this relation. According to Proposition 1, the covariance matrix Σ of the reciprocal process \mathbf{y} that approximates \mathbf{y} writes as

$$\Sigma = \mathbf{F}^* \text{circ}\{\Phi(\zeta^0), \Phi(\zeta^1), \dots, \Phi(\zeta^{N-1})\} \mathbf{F}, \quad (14)$$

hence, its inverse

$$\Sigma^{-1} = \mathbf{F}^* \text{circ}\{\Phi(\zeta^0)^{-1}, \Phi(\zeta^1)^{-1}, \dots, \Phi(\zeta^{N-1})^{-1}\} \mathbf{F}, \quad (15)$$

can be robustly computed by inverting the N blocks $\Phi(\zeta^0), \Phi(\zeta^1), \dots, \Phi(\zeta^{N-1})$, all of size $m \times m$. As a final remark, we recall that eigenvalues and eigenvectors of circulant matrices can be robustly computed as well, thanks to the availability of closed-form formulas, see for instance [32].

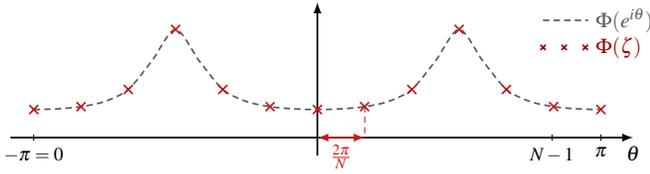


Fig. 1: Spectrum $\Phi(e^{i\theta})$ and its sampled version $\Phi(\zeta)$ with $N = 12$ samples.

As highlighted by the frequency-domain interpretation, the goodness of the approximation depends on the regularity of the spectrum: the larger is the rate of variation of the spectrum, the larger N has to be chosen in order to get a good approximation of the AR process. The frequency-domain interpretation makes even more explicit the relationship between Burg's maximum entropy problem (11) and Problem (6): provided that the number of samples N is sufficiently large, by sampling the spectrum solution of (11) we obtain an approximation of the matrix Σ solution of (6); viceversa, the symbol of Σ can be extended over the whole interval $[-\pi, \pi]$ in order to approximate the solution of (11). Figure 2 summarizes this bi-directional relationships.

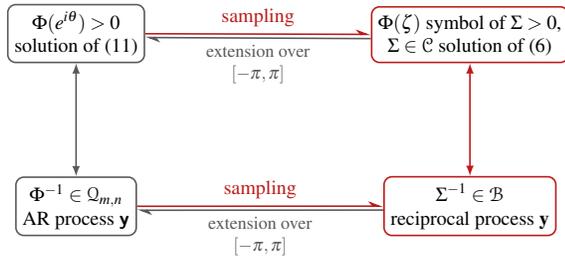


Fig. 2: Schematic representation of the reciprocal approximation of the AR process in terms of solutions of problems (11) and (6).

IV. GRAPHICAL MODELS

Consider a Gaussian random vector $\mathbf{x} \sim \mathcal{N}(0, \Sigma)$ taking values in \mathbb{R}^m , where $\Sigma = \Sigma^\top > 0$ so that the concentration matrix $K = [k_{ij}] := \Sigma^{-1}$ is well-defined. If we denote the components of \mathbf{x} as x_1, \dots, x_m , for any $i \neq j$, we have that

x_i is *conditionally independent* from x_j given the remaining random variables x_k , $k \neq i, j$, i.e.

$$x_i \perp\!\!\!\perp x_j \mid \{x_k\}_{k \neq i, j}, \quad (16)$$

if and only if the element k_{ij} in position (i, j) of the concentration matrix K is equal to zero. Formally,

$$x_i \perp\!\!\!\perp x_j \mid \{x_k\}_{k \neq i, j} \iff k_{ij} = 0. \quad (17)$$

The previous relation allows to construct an undirected graph $\mathcal{G} = (V, E)$, with $V = \{1, \dots, m\}$ and $E \subset V \times V$, associated to the random vector \mathbf{x} by taking the components x_1, \dots, x_m of \mathbf{x} as nodes and such that the edges reflect the conditional dependence relations between the random variables, i.e.

$$(i, j) \notin E \iff x_i \perp\!\!\!\perp x_j \mid \{x_k\}_{k \neq i, j}. \quad (18)$$

The graph \mathcal{G} is called the *graphical model* associated to \mathbf{x} and it gives a visual representation of the conditional dependence relations between the components of \mathbf{x} . Observe that \mathcal{G} is completely characterized by the sparsity pattern of the concentration matrix of the random vector.

A characterization of conditional independence can be given also in the dynamic setting. In particular, we consider an m -dimensional, Gaussian, wide-sense stationary AR process \mathbf{x} described by a model like (8). For any index set $I \subset V$, define

$$\mathcal{X}_I := \text{span}\{x_j(t) : j \in I, t \in \mathbb{Z}\},$$

as the closure of the set containing all the finite linear combinations of the variables $x_j(t)$. For any $i \neq j$, the notation

$$\mathcal{X}_{\{i\}} \perp\!\!\!\perp \mathcal{X}_{\{j\}} \mid \mathcal{X}_{V \setminus \{i, j\}}$$

generalizes (16) and it means that for all t_1, t_2 , $x_i(t_1)$ and $x_j(t_2)$ are conditionally independent given the space linearly generated by $\{x_k(t), k \in V \setminus \{i, j\}, t \in \mathbb{Z}\}$. One can prove that

$$\mathcal{X}_{\{i\}} \perp\!\!\!\perp \mathcal{X}_{\{j\}} \mid \mathcal{X}_{V \setminus \{i, j\}} \iff [\Phi(e^{i\theta})^{-1}]_{ij} = 0, \quad (19)$$

for any $\theta \in [-\pi, \pi]$, see [2], [3], which is the natural generalization of (17). Accordingly, we can construct the undirected graph $\mathcal{G} = (V, E)$ representing the conditional dependence relations between the components of the process \mathbf{x} by defining the set of edges as follows:

$$(i, j) \notin E \iff \mathcal{X}_{\{i\}} \perp\!\!\!\perp \mathcal{X}_{\{j\}} \mid \mathcal{X}_{V \setminus \{i, j\}}. \quad (20)$$

In this framework, the graph \mathcal{G} is completely characterized by the sparsity pattern of the inverse power spectral density of the process. Identification of sparse graphical models of reciprocal processes have been studied in [33].

In many practical situations there is the presence of a few common, latent, behaviors between the variables of interest that are responsible of the most part of the interactions between the observed variables and that cannot be captured by considering only a sparse model structure. This leads to a particular type of graphical models called *latent-variable graphical models* or *sparse plus low-rank graphical models*, [8]. Such models admit a two-layer graphical structure in which the nodes in the upper layer stand for the (few) latent-variables, while the nodes in the bottom layer represent the observed variables.

Latent-variable graphical models associated to Gaussian random vectors have been considered in [8] and then generalized in [9] to AR stochastic processes. The latter, say $\mathbf{z} := \{\mathbf{z}(t), t \in \mathbb{Z}\}$, is assumed to be of the form $\mathbf{z} = [\mathbf{y}^\top \mathbf{x}^\top]^\top$ where \mathbf{y} is the \mathbb{R}^m -valued process containing the observed variables while \mathbf{x} is the process containing l latent variables. Let $\Phi_{\mathbf{y}}$ denotes the spectral density of \mathbf{y} . Under the assumptions that $l \ll m$ and the dependence relations among the observed variables are mostly through the latent variables, we have the decomposition

$$\Phi_{\mathbf{y}}^{-1} = \Gamma - \Lambda, \quad (21)$$

where $\Gamma > 0$ is sparse and its support reflects the conditional dependencies among the observed variables, while $\Lambda \geq 0$ is low-rank and its rank equals the number l of latent-variables.

We are now ready to extend the previous results for Gaussian reciprocal processes. Let $\mathbf{z} := [\mathbf{y}^\top \mathbf{x}^\top]^\top$ be a Gaussian, periodic, reciprocal process of order n defined on the interval $[1, N]$, where \mathbf{y} plays the role of the m -dimensional observed process and \mathbf{x} is the l -dimensional latent process, respectively. The covariance matrix $\Sigma_{\mathbf{z}}$ of \mathbf{z} and its inverse can be partitioned as

$$\Sigma_{\mathbf{z}} = \left[\begin{array}{c|c} \Sigma_{\mathbf{y}} & \Sigma_{\mathbf{y}\mathbf{x}} \\ \hline \Sigma_{\mathbf{y}\mathbf{x}}^\top & \Sigma_{\mathbf{x}} \end{array} \right], \quad \Sigma_{\mathbf{z}}^{-1} = \left[\begin{array}{c|c} \mathbf{S} & \mathbf{A} \\ \hline \mathbf{A}^\top & \mathbf{R} \end{array} \right], \quad (22)$$

where $\Sigma_{\mathbf{y}} \in \mathcal{C}$ and $\Sigma_{\mathbf{x}} \in \mathcal{C}_l$ are the covariance matrices of \mathbf{y} and \mathbf{x} , respectively. Here, \mathcal{C}_l denotes the vector space of block-circulant, symmetric matrices as \mathcal{C} , except that the blocks have dimension $l \times l$. Applying the Schur complement, we obtain the relation

$$\Sigma_{\mathbf{y}}^{-1} = \mathbf{S} - \mathbf{L}, \quad (23)$$

where $\mathbf{S} > 0$ is the concentration matrix of process \mathbf{y} conditioned on \mathbf{x} , and $\mathbf{L} \geq 0$ is defined as $\mathbf{L} := \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top$. In order to ensure that, according to Theorem 1, $\Sigma_{\mathbf{y}}^{-1} \in \mathcal{B}$ we assume both \mathbf{S} and \mathbf{L} to be symmetric, block-circulant, banded of bandwidth n , i.e.

$$\begin{aligned} \mathbf{S} &= \text{circ}\{S_0, S_1, \dots, S_n, 0, \dots, 0, S_n^\top, \dots, S_1^\top\}, \\ \mathbf{L} &= \text{circ}\{L_0, L_1, \dots, L_n, 0, \dots, 0, L_n^\top, \dots, L_1^\top\}. \end{aligned} \quad (24)$$

By construction, the matrix \mathbf{L} has rank equal to the number of latent variables l , therefore under the assumption that $l \ll m$, it is a low-rank matrix. If \mathbf{S} is a sparse matrix, then we will refer to (23) as *sparse plus low-rank decomposition* of $\Sigma_{\mathbf{y}}^{-1}$ which is the analogue of (21) for reciprocal processes. It remains to show that an appropriate sparsity pattern of \mathbf{S} reflects that the dependence relations among observed variables are mostly through the few latent variables. For this purpose, let $\mathbf{y}_i := [y_i(1) \dots y_i(N)]^\top$, $i = 1, \dots, m$, be the i -th component of the process \mathbf{y} and let $\mathbf{x}_j := [x_j(1) \dots x_j(N)]^\top$, $j = 1, \dots, l$, be the j -th component of the process \mathbf{x} . Although the components of the reciprocal processes are defined for any $k \in \mathbb{Z}$, by periodicity it is sufficient to impose conditional independence only for $k \in [1, N]$. We assume that the blocks S_0, S_1, \dots, S_n of \mathbf{S} have common support $\Omega \subseteq \{(i, j) : i, j = 1, \dots, m\}$ namely,

$$(S_k)_{ij} = (S_k)_{ji} = 0, \quad k = 0, \dots, n, \quad \forall (i, j) \in \Omega^c, \quad (25)$$

where Ω is the set of pairs that contains all the (i, i) , $i = 1, \dots, m$. By property (17), equation (25) is equivalent to

$$\begin{aligned} y_i(t_1) \perp y_j(t_2) \mid \{y_h(s), h \neq i, j, s = 1, \dots, N, \\ y_i(s_1), s_1 \neq t_1, y_j(s_2), s_2 \neq t_2, \mathbf{x}\}, \end{aligned} \quad (26)$$

for any $t_1, t_2 \in [1, N]$ and for any pair $(i, j) \in \Omega^c$.

Proposition 2: Condition (26) is equivalent to

$$y_i(t_1) \perp y_j(t_2) \mid \{y_h(s), h \neq i, j, s = 1, \dots, N, \mathbf{x}\} \quad (27)$$

for any $t_1, t_2 \in [1, N]$ and for any $(i, j) \in \Omega^c$.

Proof: The proof exploits basic results of the theory of Hilbert spaces of second-order random variables, see for instance [34, Chapter 2]. First of all, let

$$\boldsymbol{\varepsilon} := \begin{bmatrix} \boldsymbol{\varepsilon}_i \\ \boldsymbol{\varepsilon}_j \end{bmatrix} = \begin{bmatrix} \mathbf{y}_i \\ \mathbf{y}_j \end{bmatrix} - \mathbb{E} \left[\begin{bmatrix} \mathbf{y}_i \\ \mathbf{y}_j \end{bmatrix} \mid \mathbf{y}_h(s), h \neq i, j, s = 1, \dots, N, \mathbf{x} \right]$$

denotes the error affecting the projection of $[\mathbf{y}_i^\top \mathbf{y}_j^\top]^\top$ onto the subspace generated by $\{\mathbf{y}_h(s), h \neq i, j, s = 1, \dots, N, \mathbf{x}\}$, for any $t_1, t_2 \in [1, N]$ and for any $(i, j) \in \Omega^c$. It can be shown that $\boldsymbol{\varepsilon}$ is a zero-mean, Gaussian, random vector. Accordingly, proving (27) is equivalent to prove that

$$\mathbb{E} [\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^\top] = 0 \quad \iff \quad \boldsymbol{\varepsilon}_i(t_1) \perp \boldsymbol{\varepsilon}_j(t_2) \quad (28)$$

for any $t_1, t_2 \in [1, N]$ and for any $(i, j) \in \Omega^c$, [34]. Let now Π be a permutation matrix that permutes the rows of $\mathbf{z} = [\mathbf{y}^\top \mathbf{x}^\top]^\top$ in order to obtain

$$\bar{\mathbf{z}} := \Pi \mathbf{z} = \begin{bmatrix} \mathbf{y}_i \\ \mathbf{y}_j \\ \mathbf{y}_{h \neq i, j} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \end{bmatrix},$$

where $\mathbf{y}_{h \neq i, j}$ is the vector containing the random variables $\mathbf{y}_h(s)$, $h \neq i, j$, $s = 1, \dots, N$. We partition the covariance matrix $\Sigma_{\bar{\mathbf{z}}}$ of $\bar{\mathbf{z}}$ as

$$\Sigma_{\bar{\mathbf{z}}} = \left[\begin{array}{c|c} \Sigma_{\bar{\mathbf{z}}_1} & \Sigma_{\bar{\mathbf{z}}_1 \bar{\mathbf{z}}_2} \\ \hline \Sigma_{\bar{\mathbf{z}}_2 \bar{\mathbf{z}}_1} & \Sigma_{\bar{\mathbf{z}}_2} \end{array} \right],$$

where $\Sigma_{\bar{\mathbf{z}}_1}$ and $\Sigma_{\bar{\mathbf{z}}_2}$ are the covariance matrices of $\bar{\mathbf{z}}_1$ and $\bar{\mathbf{z}}_2$, respectively. It is well known that its inverse can be partitioned conformably as

$$\Sigma_{\bar{\mathbf{z}}}^{-1} = \Pi \Sigma_{\mathbf{z}}^{-1} \Pi^\top = \begin{bmatrix} \bar{\mathbf{S}} & * \\ * & * \end{bmatrix},$$

where

$$\bar{\mathbf{S}} := (\Sigma_{\bar{\mathbf{z}}_1} - \Sigma_{\bar{\mathbf{z}}_1 \bar{\mathbf{z}}_2} \Sigma_{\bar{\mathbf{z}}_2}^{-1} \Sigma_{\bar{\mathbf{z}}_2 \bar{\mathbf{z}}_1})^{-1} \quad (29)$$

is a permuted version of matrix \mathbf{S} , according to the permutation matrix Π . By construction, the Schur complement formula applied on $\Sigma_{\bar{\mathbf{z}}}$ gives

$$\Sigma_{\boldsymbol{\varepsilon}} = \Sigma_{\bar{\mathbf{z}}_1} - \Sigma_{\bar{\mathbf{z}}_1 \bar{\mathbf{z}}_2} \Sigma_{\bar{\mathbf{z}}_2}^{-1} \Sigma_{\bar{\mathbf{z}}_2 \bar{\mathbf{z}}_1} = \bar{\mathbf{S}}^{-1}, \quad (30)$$

that relates the covariance matrix $\Sigma_{\boldsymbol{\varepsilon}}$ of the projection error $\boldsymbol{\varepsilon}$ to the covariance matrix $\Sigma_{\bar{\mathbf{z}}}$ of $\bar{\mathbf{z}}$. Condition (25) is equivalent to say that $\bar{\mathbf{S}}$, and therefore $\bar{\mathbf{S}}^{-1}$, is block-diagonal. Accordingly, by (30), $\Sigma_{\boldsymbol{\varepsilon}}$ is block-diagonal, i.e. $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_j$ are independent, which is equivalent to (28) as we wanted to prove. \square

The above result reflects the fact that the random variables $\{y_i(s_1), y_j(s_2), s_1 \neq t_1, s_2 \neq t_2\}$ do not play any role in the

conditioning (26). Moreover, the group-sparsity condition (25) translates in the fact that the conditional dependence relations between the observed variables are mainly due to the few latent variables. Accordingly, condition (27) represents the reciprocal counterpart of condition (17). We conclude that $\Sigma_{\mathbf{z}}^{-1}$ in (22) together with (25) define an undirected graph for the Gaussian random vector \mathbf{z} which admits a two-layer structure where

- The nodes in the upper-layer represent the l variables of the latent-process $\mathbf{x}_1, \dots, \mathbf{x}_l$ while the nodes in the bottom-layer represent the m variables of the observed process $\mathbf{y}_1, \dots, \mathbf{y}_m$.
- The edges are given by the entries of the concentration matrix $\Sigma_{\mathbf{z}}^{-1}$. In particular, the edge (i, j) , between two vectors \mathbf{y}_i and \mathbf{y}_j , $i \neq j$, is described by

$$[(S_0)_{ij} (S_1)_{ij} \dots (S_n)_{ij} 0 \dots 0 (S_n)_{ji} \dots (S_1)_{ji}].$$

Example 1: Consider the case in which $N = 2$, $m = 7$, $l = 2$, and suppose that the graphical model associated to the vector \mathbf{z} is the one depicted in Figure 3.

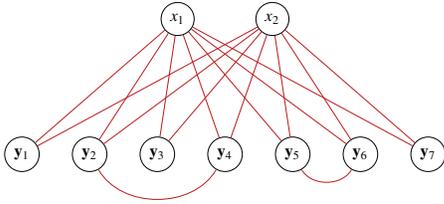


Fig. 3: Example of a latent-variable graphical model: x_1, x_2 are the latent-variables and y_1, y_2, \dots, y_7 are the manifest variables.

In this case, the concentration matrix of vector \mathbf{z} will have the structure (22) with

$$\mathbf{S} = \begin{bmatrix} S_0 & S_1^\top \\ S_1 & S_0 \end{bmatrix}, \quad S_0, S_1 \in \mathbb{R}^{7 \times 7},$$

and R is a 2×2 matrix. The presence of an edge between \mathbf{y}_2 and \mathbf{y}_4 implies that at least one of the two elements $(S_0)_{24}$ and $(S_1)_{24}$ is different from zero. Similar arguments holds for the edge between \mathbf{y}_5 and \mathbf{y}_6 . Thus, $\Omega = \{(i, i) : i = 1, \dots, 7\} \cup \{(2, 4), (5, 6)\}$.

V. IDENTIFICATION OF LATENT-VARIABLE RECIPROCAL GRAPHICAL MODELS

The problem of identifying a latent-variable graphical model associated to a Gaussian random vector has been firstly considered in [8] where the solution is obtained by solving a regularized maximum likelihood problem. In [9] the problem has been extended to a dynamic setting, by considering an AR Gaussian process. More precisely, in [9] a regularized version of Problem (12) that relies on the sparse plus low-

rank decomposition of the inverse of the observed spectrum in (21), has been considered:

$$\begin{aligned} & \underset{\Gamma, \Lambda \in \mathcal{Q}_{m,n}}{\operatorname{argmin}} \int -\log \det(\Gamma - \Lambda) + \langle \Gamma - \Lambda, \hat{\Phi}_{\mathbf{y}} \rangle \\ & \quad + \gamma_S \phi_1(\Gamma) + \gamma_L \phi_*(\Lambda) \\ \text{subject to} \quad & \Gamma - \Lambda > 0 \\ & \Lambda \geq 0. \end{aligned} \quad (31)$$

Here, $\gamma_S, \gamma_L > 0$ are the regularization parameters that balance the effects of the two regularizers ϕ_1 and ϕ_* inducing sparsity and low-rank on Γ and Λ , respectively, while $\hat{\Phi}_{\mathbf{y}}$ is the truncated periodogram of the observed process \mathbf{y} . In this section we propose a procedure for the identification of a latent-variable graphical model associated to an AR Gaussian process that exploits the approximation of an AR process through a reciprocal process in the sense explained in Section III. Recalling that a latent-variable graphical model of a reciprocal process is characterized by (23), the system identification problem can be stated as follows.

Problem 3: Consider an m -dimensional AR process \mathbf{y} and let $\hat{R}_0, \dots, \hat{R}_n$ be the estimates of the first $n+1$ covariance lags of \mathbf{y} computed as in (10). Set $\Sigma_0 := \hat{R}_0, \dots, \Sigma_n := \hat{R}_n$. Compute the blocks $\Sigma_{n+1}, \dots, \Sigma_{\frac{N}{2}}$ of the block-circulant covariance matrix $\Sigma_{\mathbf{y}} = \operatorname{circ}\{\Sigma_0, \Sigma_1, \dots, \Sigma_{\frac{N}{2}-1}, \Sigma_{\frac{N}{2}}, \Sigma_{\frac{N}{2}-1}^\top, \dots, \Sigma_1^\top\}$ such that $\Sigma_{\mathbf{y}}^{-1} = \mathbf{S} - \mathbf{L}$, where $\mathbf{S} > 0$ and $\mathbf{L} \geq 0$ are as in (24) with S_0, \dots, S_n having the smallest possible common support Ω , as in (25), and the rank of \mathbf{L} is as small as possible.

We stress the fact that only samples of the observed processes are available. Clearly, the matrix $\Sigma_{\mathbf{y}}$ solving Problem 3 is the covariance of the reciprocal process \mathbf{y} approximating the observed process \mathbf{y} . Since we are going to identify a model for a reciprocal process, we can exploit the maximum entropy dual problem (7) recalled in Section III. It is worth noting that the support Ω is not known in advance, thus it has to be estimated from the data. In order to do that, inspired by [4], we consider the following regularizer

$$h_\infty(\mathbf{S}) = \sum_{k>h} \max \left\{ |(S_0)_{hk}|, 2 \max_{j=1, \dots, n} |(S_j)_{hk}|, 2 \max_{j=1, \dots, n} |(S_j)_{kh}| \right\}.$$

The latter is a generalization of the ℓ^∞ -norm used to induce group-sparsity on vectors, and it is used to enforce on \mathbf{S} the group sparsity in (25). The trace (as a tractable proxy of the nuclear norm) is used instead for inducing low-rankness in \mathbf{L} . Therefore, the paradigm for the estimation of the sparse plus low-rank decomposition of the concentration matrix $\Sigma_{\mathbf{y}}^{-1}$ now directly follows from (7) by setting $\mathbf{X} = \mathbf{S} - \mathbf{L}$, with $\mathbf{L} \geq 0$, and by adding the regularizers just introduced:

$$\begin{aligned} & \underset{\mathbf{S}, \mathbf{L} \in \mathcal{B}}{\operatorname{argmin}} \quad -\log \det(\mathbf{S} - \mathbf{L}) + \langle \hat{\Sigma}_{\mathbf{y}}, \mathbf{S} - \mathbf{L} \rangle_{\mathcal{C}} \\ & \quad + \lambda_S h_\infty(\mathbf{S}) + \lambda_L \operatorname{tr}(\mathbf{L}) \\ \text{subject to} \quad & \mathbf{S} - \mathbf{L} > 0, \quad \mathbf{L} \geq 0 \end{aligned} \quad (32)$$

where $\lambda_S, \lambda_L > 0$ are the two regularization parameters and

$$\hat{\Sigma}_{\mathbf{y}} = \operatorname{circ}\{\hat{R}_0, \hat{R}_1, \dots, \hat{R}_n, 0, \dots, 0, \hat{R}_n^\top, \dots, \hat{R}_1^\top\}$$

is the symmetric, banded block-circulant matrix of bandwidth n , containing the covariance lags estimated from the

observations. As a further motivation, observe that Problem (32) is precisely the reciprocal counterpart of Problem (31) considered in [9]. By replacing \mathbf{S} with $\mathbf{X} := \mathbf{S} - \mathbf{L}$, it becomes

$$\begin{aligned} & \underset{\mathbf{X}, \mathbf{L} \in \mathcal{B}}{\operatorname{argmin}} \quad -\log \det(\mathbf{X}) + \operatorname{tr}(\hat{\Sigma}_{\mathbf{y}} \mathbf{X}) + \lambda_S h_\infty(\mathbf{X} + \mathbf{L}) + \lambda_L \operatorname{tr}(\mathbf{L}) \\ & \text{subject to} \quad \mathbf{X} > 0, \quad \mathbf{L} \geq 0. \end{aligned} \quad (33)$$

We address the previous constrained optimization problem using the Lagrange multipliers theory. In doing that we add a new dummy variable \mathbf{Y}

$$\begin{aligned} & \underset{\substack{\mathbf{X} \in \mathcal{C} \\ \mathbf{Y}, \mathbf{L} \in \mathcal{B}}}{\operatorname{argmin}} \quad -\log \det(\mathbf{X}) + \operatorname{tr}(\hat{\Sigma}_{\mathbf{y}} \mathbf{X}) + \lambda_S h_\infty(\mathbf{Y}) + \lambda_L \operatorname{tr}(\mathbf{L}) \\ & \text{subject to} \quad \mathbf{X} > 0, \quad \mathbf{L} \geq 0 \\ & \quad \quad \quad \mathbf{Y} = \mathbf{X} + \mathbf{L}. \end{aligned} \quad (34)$$

The Lagrangian function for this problem is

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Y}, \mathbf{L}, \mathbf{V}, \mathbf{Z}) = & -\log \det(\mathbf{X}) + \langle \hat{\Sigma}_{\mathbf{y}}, \mathbf{X} \rangle_{\mathcal{C}} + \lambda_S h_\infty(\mathbf{Y}) \\ & + \lambda_L \operatorname{tr}(\mathbf{L}) - \langle \mathbf{V}, \mathbf{L} \rangle_{\mathcal{C}} + \langle \mathbf{Z}, \mathbf{X} + \mathbf{L} - \mathbf{Y} \rangle_{\mathcal{C}} \end{aligned} \quad (35)$$

where, $\mathbf{V} \in \mathcal{B}$, because $\mathbf{L} \in \mathcal{B}$, and $\mathbf{V} \geq 0$, while $\mathbf{Z} \in \mathcal{C}$. After simple computations we have

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Y}, \mathbf{L}, \mathbf{V}, \mathbf{Z}) = & -\log \det(\mathbf{X}) + \langle \hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}, \mathbf{X} \rangle_{\mathcal{C}} \\ & + \langle \lambda_L I_{mN} - \mathbf{V} + \mathbf{Z}, \mathbf{L} \rangle_{\mathcal{C}} \\ & + \lambda_S h_\infty(\mathbf{Y}) - \langle \mathbf{Z}, \mathbf{Y} \rangle_{\mathcal{C}}. \end{aligned}$$

The dual objective function is the infimum over \mathbf{X} , \mathbf{Y} and \mathbf{L} of the Lagrangian. The unique term on \mathcal{L} that depends on \mathbf{Y} is $\lambda_S h_\infty(\mathbf{Y}) - \langle \mathbf{Z}, \mathbf{Y} \rangle_{\mathcal{C}}$. The latter is bounded below if and only if

$$\operatorname{diag}(\mathbf{Z}_j) = 0, \quad j = 0, \dots, n, \quad (36)$$

$$2|(Z_0)_{kh}| + \sum_{j=1}^n |(Z_j)_{kh}| + |(Z_j)_{hk}| \leq \frac{\lambda_S}{N}, \quad k > h, \quad (37)$$

in which case the infimum is zero. Accordingly,

$$\inf_{\mathbf{Y}} \mathcal{L} = \begin{cases} -\log \det(\mathbf{X}) + \langle \hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}, \mathbf{X} \rangle_{\mathcal{C}} + \langle \lambda_L I_{mN} - \mathbf{V} + \mathbf{Z}, \mathbf{L} \rangle_{\mathcal{C}} \\ \text{if (36), (37) hold,} \\ -\infty \quad \text{otherwise.} \end{cases}$$

The only term that depends on \mathbf{L} is $\langle \lambda_L I_{mN} - \mathbf{V} + \mathbf{Z}, \mathbf{L} \rangle_{\mathcal{C}}$. Recalling that $\mathbf{L}, \mathbf{V} \in \mathcal{B}$, by using the linearity of the projection operator $\mathcal{P}_{\mathcal{B}}$, we have that

$$\langle \lambda_L I_{mN} - \mathbf{V} + \mathbf{Z}, \mathbf{L} \rangle_{\mathcal{C}} = \langle \lambda_L I_{mN} - \mathbf{V} + \mathcal{P}_{\mathcal{B}}(\mathbf{Z}), \mathbf{L} \rangle_{\mathcal{C}} \quad (38)$$

which is linear in \mathbf{L} , and therefore it is bounded below if and only if

$$\lambda_L I_{mN} - \mathbf{V} + \mathcal{P}_{\mathcal{B}}(\mathbf{Z}) = 0. \quad (39)$$

In this case, the minimum of (38) is zero. Accordingly,

$$\inf_{\mathbf{Y}, \mathbf{L}} \mathcal{L} = \begin{cases} -\log \det(\mathbf{X}) + \langle \hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}, \mathbf{X} \rangle_{\mathcal{C}} \\ \text{if (36), (37), (39) hold,} \\ -\infty \quad \text{otherwise.} \end{cases}$$

If (36), (37), (39) hold, it remains to minimize the strictly convex function

$$\bar{\mathcal{L}}(\mathbf{X}) := \inf_{\mathbf{Y}, \mathbf{L}} \mathcal{L} = -\log \det(\mathbf{X}) + \langle \hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}, \mathbf{X} \rangle_{\mathcal{C}}$$

over the cone of the symmetric, positive definite, banded block-circulant matrices. Observe that, for any $\mathbf{Z} \in \mathcal{C}$, any $\hat{\Sigma}_{\mathbf{y}} \in \mathcal{B}$, and for any sequence $\mathbf{X}_k > 0$ converging to a singular matrix,

$$\lim_{k \rightarrow \infty} \bar{\mathcal{L}}(\mathbf{X}_k) = \infty.$$

Accordingly, we can assume that the solution lies in the interior of the cone so that a necessary and sufficient condition for \mathbf{X}_o to be a minimum point for $\bar{\mathcal{L}}$ is that its first Gateaux derivative computed at $\mathbf{X} = \mathbf{X}_o$ is equal to zero in every direction $\delta \mathbf{X}$, namely

$$\delta \bar{\mathcal{L}}(\mathbf{X}_o; \delta \mathbf{X}) = \operatorname{tr} [(-\mathbf{X}_o^{-1} + \hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}) \delta \mathbf{X}] = 0, \quad \forall \delta \mathbf{X} \in \mathcal{C}. \quad (40)$$

Notice that $\bar{\mathcal{L}}$ is bounded below if and only if

$$\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z} > 0, \quad (41)$$

therefore condition (40) is satisfied if and only if $\mathbf{X}_o = (\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z})^{-1}$. Hence,

$$\inf_{\mathbf{Y}, \mathbf{L}, \mathbf{X}} \mathcal{L} = \begin{cases} \log \det(\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}) + mN, \\ \text{if (36), (37), (39), (41) hold,} \\ -\infty \quad \text{otherwise.} \end{cases}$$

Therefore, the dual problem of (32) is

$$\begin{aligned} & \underset{\mathbf{V} \in \mathcal{B}, \mathbf{Z} \in \mathcal{C}}{\operatorname{argmin}} \quad -\log \det(\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}) - mN \\ & \text{subject to} \quad \mathbf{V} \geq 0, \text{ (36), (37), (39), (41)}. \end{aligned} \quad (42)$$

Notice that we can remove the variable \mathbf{V} . Indeed, recalling that $\mathbf{V} \geq 0$, the constraint (39) becomes $\lambda_L I_{mN} + \mathcal{P}_{\mathcal{B}}(\mathbf{Z}) = \mathbf{V} \geq 0$. Accordingly, the dual problem takes the form

$$\begin{aligned} & \underset{\mathbf{Z} \in \mathcal{C}}{\operatorname{argmin}} \quad -\log \det(\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}) - mN \\ & \text{subject to} \quad \text{(36), (37), (41)} \\ & \quad \quad \quad \lambda_L I_{mN} + \mathcal{P}_{\mathcal{B}}(\mathbf{Z}) \geq 0. \end{aligned} \quad (43)$$

Proposition 3: Under the assumption that $\hat{\Sigma}_{\mathbf{y}} \in \mathcal{B}$ and $\hat{\Sigma}_{\mathbf{y}} > 0$, Problem (43) admits a unique solution.

Proof: Define $f(\mathbf{Z}) := \log \det(\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z})$. Let

$$\mathcal{Q} := \{\mathbf{Z} \in \mathcal{C} \mid \text{(36), (37), (41) and } \lambda_L I_{mN} + \mathcal{P}_{\mathcal{B}}(\mathbf{Z}) \geq 0 \text{ hold}\}$$

be the set of constraints of Problem (43). First of all, notice that constraints (36) and (37) ensure that \mathcal{Q} is a bounded subset of \mathcal{C} . Indeed, the entries of any $\mathbf{Z} \in \mathcal{Q}$ are bounded by λ_S/N so that $\|\mathbf{Z}\|_{\mathcal{C}} < \infty$ for any $\mathbf{Z} \in \mathcal{Q}$. Let now $(\mathbf{Z}^{(k)})_{k \in \mathbb{N}}$ be a generic sequence of elements of \mathcal{Q} converging to some $\bar{\mathbf{Z}} \in \mathcal{C}$, such that $\hat{\Sigma}_{\mathbf{y}} + \bar{\mathbf{Z}} \geq 0$ is singular. Then

$$\lim_{k \rightarrow \infty} -\log \det(\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}^{(k)}) = +\infty,$$

and therefore $\mathbf{Z}^{(k)}$ is not an infimizing sequence. Hence, we can restrict the research of the minimum to the closed subset of \mathcal{Q} defined by

$$\bar{\mathcal{Q}} := \{\mathbf{Z} \in \mathcal{C} \mid \hat{\Sigma}_{\mathbf{y}} + \mathbf{Z} \geq \varepsilon I_{mN}, (36), (37) \\ \text{and } \lambda_L I_{mN} + \mathbf{P}_{\mathcal{B}}(\mathbf{Z}) \geq 0 \text{ hold}\}$$

with $\varepsilon > 0$ small enough. By what we have shown till now, the function f is continuous on the compact set $\bar{\mathcal{Q}}$ and therefore it admits at least one minimum point. Since f is strictly convex, the minimum is also unique. \square

Proposition 4: Under the assumption that $\hat{\Sigma}_{\mathbf{y}} \in \mathcal{B}$ and $\hat{\Sigma}_{\mathbf{y}} > 0$, Problem (33) admits a solution $(\mathbf{X}_o, \mathbf{L}_o)$ and \mathbf{X}_o is unique.

Proof: Notice that Problem (33) is a strictly feasible convex optimization problem (for instance, pick $\mathbf{X} = I_{mN}$ and $\mathbf{L} = 0$). Accordingly, Slater's condition holds, hence strong duality holds between (33) and its dual. The strong duality between problems (33) and (43) and the existence of a unique optimum \mathbf{Z}_o for the dual problem (43), imply that there exists a unique $\mathbf{X}_o \in \mathcal{B}$ so that $\mathbf{X}_o = (\hat{\Sigma}_{\mathbf{y}} + \mathbf{Z}_o)^{-1}$ which solves the primal problem (33).

It remains to show that there exists an $\mathbf{L}_o \in \mathcal{B}$ that solves the optimization problem

$$\begin{aligned} \underset{\mathbf{L} \in \mathcal{B}}{\operatorname{argmin}} \quad & \lambda_S h_{\infty}(\mathbf{X}_o + \mathbf{L}) + \lambda_L \operatorname{tr}(\mathbf{L}) \\ \text{subject to} \quad & \mathbf{L} \geq 0. \end{aligned} \quad (44)$$

Notice that, the objective function in (44) is continuous. Since $\mathbf{L} = 0$ is a feasible point, the problem is equivalent to find $\mathbf{L} \in \mathcal{B}$ that minimizes $\lambda_S h_{\infty}(\mathbf{X}_o + \mathbf{L}) + \lambda_L \operatorname{tr}(\mathbf{L})$ over the set

$$\mathcal{K} := \left\{ \mathbf{L} \in \mathcal{B} \mid \mathbf{L} \geq 0, \lambda_S h_{\infty}(\mathbf{X}_o + \mathbf{L}) + \lambda_L \operatorname{tr}(\mathbf{L}) \leq \lambda_S h_{\infty}(\mathbf{X}_o) \right\}.$$

It is easy to see that \mathcal{K} is a closed and bounded (and thus compact) subset of \mathcal{B} . Hence, by Weierstrass' Theorem, Problem (44) admits a solution \mathbf{L}_o . At this point we can conclude that the primal problem (33) admits a solution $(\mathbf{X}_o, \mathbf{L}_o)$. \square

A. Interpretations

In the remaining of this section we will show how Problem (32) can be interpreted either as a regularized maximum-likelihood problem or as a dual of a maximum entropy problem.

Maximum likelihood interpretation: The reciprocal approximation of AR processes illustrated in Section III allows to interpret Problem (32) as a regularized (conditional) maximum likelihood problem. Indeed, in the following we will show that the fitting function in (32), i.e.

$$-\log \det(\mathbf{S} - \mathbf{L}) + \operatorname{tr}(\hat{\Sigma}_{\mathbf{y}}(\mathbf{S} - \mathbf{L})), \quad (45)$$

is the approximation of the (conditional) negative log-likelihood of the AR process (8) that should be understood in the sense explained in Section III. Following [35], consider the observed AR process \mathbf{y} whose spectrum is denoted by $\Phi_{\mathbf{y}}$, and suppose that T observations $y(1), \dots, y(T)$ of the process are available. The conditional likelihood of the process \mathbf{y} is defined as the likelihood function associated to the

conditional distribution of $y(n+1), y(n+2), \dots, y(n+T)$ given $y(1), \dots, y(n)$. Let

$$\mathbf{T}_n := \operatorname{Toep}\{\hat{R}_0, \hat{R}_1, \dots, \hat{R}_n\}$$

be the block-Toeplitz matrix having in the first rows the estimates of the first $n+1$ covariance lags of the process $\hat{R}_0, \hat{R}_1, \dots, \hat{R}_n$ computed as in (10). For T large enough, the conditional negative log-likelihood function of the AR process can be well approximated by

$$\ell(B) := -(T-n) \log \det B_0 + \frac{T-n}{2} \operatorname{tr}(B \mathbf{T}_n B^{\top})$$

where $B := [B_0 \ B_1 \ \dots \ B_n]$ is the $(n+1)m$ -dimensional vector containing the coefficients of the process. Applying Jensen's formula, it turns out that

$$\log \det B_0 = \frac{1}{2} \int \log \det \Phi_{\mathbf{y}}(e^{i\theta}),$$

moreover, if $\hat{\Phi}_{\mathbf{y}}(e^{i\theta})$ is the truncated periodogram of the AR process in (13), it is easy to see that

$$\int \hat{\Phi}_{\mathbf{y}}(e^{i\theta}) e^{i\theta k} = \hat{R}_{-k} = \hat{R}_k^{\top}.$$

Accordingly, the approximated conditional negative log-likelihood can be rewritten as

$$\ell(B) = \frac{T-n}{2} \int \log \det \Phi_{\mathbf{y}}(e^{i\theta}) + \operatorname{tr} \left[\hat{\Phi}_{\mathbf{y}}(e^{i\theta}) \Phi_{\mathbf{y}}(e^{i\theta})^{-1} \right]. \quad (46)$$

A natural way to approximate (46) is to approximate the integral with a finite sum, i.e. to discretize the interval $[-\pi, \pi]$. This is precisely the frequency interpretation of the reciprocal approximation explained in Section III that consists in sampling the spectrum of the process to obtain the corresponding symbol (see Figure 1). In fact, considering as sample frequency $\Delta\theta = 2\pi/N$, the Backward Euler approximation leads to the discrete approximation

$$\ell(B) \simeq \frac{T-n}{2} \frac{\Delta\theta}{2\pi} \sum_{k=0}^{N-1} \log \det \Phi_{\mathbf{y}}(e^{i\theta_k}) + \operatorname{tr} \left[\hat{\Phi}_{\mathbf{y}}(e^{i\theta_k}) \Phi_{\mathbf{y}}(e^{i\theta_k})^{-1} \right].$$

where $\theta_k = k\Delta\theta - \pi$. The conditional log-likelihood can now be rewritten straightforward in terms of symbols as

$$\ell(B) \simeq \frac{T-n}{2N} \left[\sum_{k=0}^{N-1} \log \det \Phi_{\mathbf{y}}(\zeta^k) + \operatorname{tr} \sum_{k=0}^{N-1} \hat{\Phi}_{\mathbf{y}}(\zeta^k) \Phi_{\mathbf{y}}(\zeta^k)^{-1} \right].$$

Observe now that $\Phi_{\mathbf{y}}(\zeta)$ is precisely the symbol of the block-circulant covariance matrix $\Sigma_{\mathbf{y}}$ of the reciprocal process \mathbf{y} approximating the process \mathbf{y} and $\hat{\Phi}_{\mathbf{y}}(\zeta)$ is the symbol of the block-circulant matrix $\hat{\Sigma}_{\mathbf{y}}$ in Problem (32). Accordingly, from Proposition 1, it follows that

$$\ell(B) \simeq \frac{T-n}{2N} \left[-\log \det \Sigma_{\mathbf{y}}^{-1} + \operatorname{tr}(\hat{\Sigma}_{\mathbf{y}} \Sigma_{\mathbf{y}}^{-1}) \right].$$

Since $\Sigma_{\mathbf{y}}^{-1} = \mathbf{S} - \mathbf{L}$, this is precisely (up to a scaling factor) equal to (45).

Maximum entropy interpretation: We will show that Problem (32) can be interpreted a regularized version of the dual of a maximum entropy problem, see [36] for a general overview of these problems. Consider the regularized solution $(\mathbf{S}_o, \mathbf{L}_o)$ of (32) and let Ω be the support of \mathbf{S}_o , i.e. \mathbf{S}_o satisfies (25). Since $\mathbf{L}_o \in \mathcal{B}$ is so that $\mathbf{L}_o \geq 0$ and $\text{rank } \mathbf{L}_o = lN \ll mN$, there exists

$$\mathbf{G} = \text{circ}\{G_0, G_1, \dots, G_n, 0, \dots, 0\}$$

such that $G_k \in \mathbb{R}^{m \times l}$ and $\mathbf{L}_o = \mathbf{G}^\top \mathbf{G}$. Accordingly, we can consider a modified version of Problem (32) where the regularizers are replaced by the corresponding hard-constraints $\mathbf{S} \in \mathcal{V}_\Omega$ and $\mathbf{L} \in \mathcal{V}_G$, where $\mathcal{V}_\Omega := \{\mathbf{S} \in \mathcal{C} : \text{P}_{\Omega^c}(\mathbf{S}) = 0\}$ and $\mathcal{V}_G := \{\mathbf{G}^\top (I_N \otimes H) \mathbf{G} : H \in \mathbb{R}^{l \times l}, H = H^\top\}$ is such that $\mathcal{V}_G \subseteq \mathcal{B}$. Thus, the resulting problem is

$$\begin{aligned} & \underset{\mathbf{S}, \mathbf{L} \in \mathcal{B}}{\text{argmin}} && -\log \det(\mathbf{S} - \mathbf{L}) + \langle \hat{\Sigma}_y, \mathbf{S} - \mathbf{L} \rangle_{\mathcal{C}} \\ & \text{subject to} && \mathbf{S} - \mathbf{L} > 0, \mathbf{L} \geq 0, \\ & && \mathbf{S} \in \mathcal{V}_\Omega, \mathbf{L} \in \mathcal{V}_G. \end{aligned} \quad (47)$$

Proposition 5: The primal of Problem (47) is

$$\begin{aligned} & \underset{\Sigma_y \in \mathcal{C}}{\text{argmax}} && \log \det \Sigma_y \\ & \text{subject to} && \text{P}_{\Omega} \text{P}_{\mathcal{B}}(\Sigma_y - \hat{\Sigma}_y) = 0, \\ & && \mathbf{E}^* \mathbf{G}(\Sigma_y - \hat{\Sigma}_y) \mathbf{G}^\top \mathbf{E} \geq 0, \end{aligned} \quad (48)$$

where $\mathbf{E}^* := \frac{1}{\sqrt{N}} [I_l \ 0 \ \dots \ 0]$.

Proof: We derive the dual of Problem (48). Observing that $\mathbf{E} = \mathbf{F}^* \mathbf{1}$ where $\mathbf{1} := \frac{1}{\sqrt{N}} [I_l \ I_l \ \dots \ I_l]^\top$, the Lagrangian of Problem (48) writes as

$$\begin{aligned} \mathcal{L}(\Sigma_y, \mathbf{W}, H) &= \log \det \Sigma_y + \langle \text{P}_{\Omega \cup \mathcal{B}}(\hat{\Sigma}_y - \Sigma_y), \mathbf{W} \rangle_{\mathcal{C}} \\ &+ \langle \mathbf{1}^\top \mathbf{F} \mathbf{G}(\Sigma_y - \hat{\Sigma}_y) \mathbf{G}^\top \mathbf{F}^* \mathbf{1}, H \rangle_{\mathcal{C}}, \end{aligned}$$

where $\mathbf{W} \in \mathcal{C}$, $H \in \mathbb{R}^{l \times l}$ is a positive semidefinite symmetric matrix, and $\text{P}_{\Omega \cup \mathcal{B}}(\mathbf{S}) = \text{P}_{\Omega} \text{P}_{\mathcal{B}}(\mathbf{S})$. The last term of the Lagrangian can be rewritten as

$$\begin{aligned} & \text{tr} \left[\mathbf{F}(\Sigma_y - \hat{\Sigma}_y) \mathbf{F}^* \mathbf{F} \mathbf{G}^\top \mathbf{F}^* \mathbf{1} \mathbf{H} \mathbf{1}^\top \mathbf{F} \mathbf{G} \mathbf{F}^* \right] \\ &= \text{tr} \left[\mathbf{F}(\Sigma_y - \hat{\Sigma}_y) \mathbf{F}^* \mathbf{F} \mathbf{G}^\top \mathbf{F}^* (I_N \otimes H) \mathbf{F} \mathbf{G} \mathbf{F}^* \right] \\ &= \text{tr} \left[(\Sigma_y - \hat{\Sigma}_y) \mathbf{G}^\top (I_N \otimes H) \mathbf{G} \right], \end{aligned}$$

where we have exploited the fact that $\mathbf{F}(\Sigma_y - \hat{\Sigma}_y) \mathbf{F}^*$ and $\mathbf{F} \mathbf{G} \mathbf{F}^*$ are block-diagonal matrices and the fact that $\mathbf{F}^* (I_N \otimes H) \mathbf{F} = I_N \otimes H$. Accordingly,

$$\begin{aligned} \mathcal{L}(\Sigma_y, \mathbf{W}, H) &= \log \det \Sigma_y + \langle \hat{\Sigma}_y - \Sigma_y, \text{P}_{\Omega \cup \mathcal{B}}(\mathbf{W}) \rangle_{\mathcal{C}} \\ &+ \langle \Sigma_y - \hat{\Sigma}_y, \mathbf{G}^\top (I_N \otimes H) \mathbf{G} \rangle_{\mathcal{C}} \\ &= \log \det \Sigma_y + \langle \hat{\Sigma}_y - \Sigma_y, \mathbf{S} - \mathbf{L} \rangle_{\mathcal{C}}, \end{aligned}$$

where $\mathbf{S} := \text{P}_{\Omega \cup \mathcal{B}}(\mathbf{W})$ belongs to \mathcal{V}_Ω and $\mathbf{L} := \mathbf{G}^\top (I_N \otimes H) \mathbf{G} \geq 0$ belongs to $\mathcal{V}_G \subseteq \mathcal{B}$, i.e. they satisfy all the constraints in (47). Similar arguments as the ones used to prove formula (40), allow us to assert that a necessary and sufficient condition for Σ_o to be a minimum point for \mathcal{L} is that its first Gateaux

derivative computed at $\Sigma_y = \Sigma_o$ is equal to zero in every direction $\delta \Sigma$, namely

$$\delta \mathcal{L}(\Sigma_o; \delta \Sigma) = \text{tr} [(\Sigma_o^{-1} - \mathbf{S} + \mathbf{L}) \delta \Sigma] = 0, \quad \forall \delta \Sigma \in \mathcal{C}.$$

By assumption we have that $\mathbf{S} - \mathbf{L} > 0$ thus, the substitution of the optimum $\Sigma_o = (\mathbf{S} - \mathbf{L})^{-1}$ in the Lagrangian \mathcal{L} leads precisely to the objective function in (47). \square

Some observations on the two constraints of (48) are in order. The first constraint $\text{P}_{\Omega} \text{P}_{\mathcal{B}}(\Sigma_y - \hat{\Sigma}_y) = 0$ fixes the entries corresponding to the indexes in Ω of the first $n+1$ lags of the reciprocal process. Concerning the second constraint, let $\Psi(\zeta)$ and $\Phi_y(\zeta)$ be the symbols of \mathbf{G} and Σ_y , respectively. By Proposition 1, we have that

$$\mathbf{E}^* \mathbf{G} \Sigma_y \mathbf{G}^\top \mathbf{E} = \frac{1}{N} \sum_{k=0}^{N-1} \Psi(\zeta^k) \Phi_y(\zeta^k) \Psi(\zeta^k)^*, \quad (49)$$

which is the covariance of the output of the $m \times l$ filter $\Psi(\zeta) = \sum_{k=0}^n G_k \zeta^{-k}$ fed with the reciprocal process \mathbf{y} . Accordingly, the second constraint in (48) states that the covariance matrix of the process at the output of the filter is lower-bounded by $\mathbf{E}^* \mathbf{G} \hat{\Sigma}_y \mathbf{G}^\top \mathbf{E}$. We conclude that Problem (48) can be seen as the reciprocal counterpart of the maximum entropy problem [9],

$$\begin{aligned} & \underset{\Phi_y \in \mathcal{S}_m}{\text{argmax}} && \int \log \det \Phi_y \\ & \text{subject to} && \left(\int e^{i\theta k} \Phi_y - \hat{R}_k \right)_{pq} = 0, \quad \begin{matrix} k=0,1,\dots,n \\ (p,q) \in \Omega \end{matrix} \\ & && \int \Psi(\Phi_y - \hat{\Phi}_y) \Psi^* \geq 0, \end{aligned} \quad (50)$$

where $\Psi(e^{i\theta}) = \sum_{k=0}^n G_k e^{-i\theta k}$. Indeed, the second constraint in (50) can be approximated with the backward Euler approximation with sample frequency $\Delta\theta = 2\pi/N$ obtaining (49).

VI. ALTERNATING DIRECTION METHOD OF MULTIPLIERS

The solution of Problem (43) requires the joint enforcement of the constraints (36), (37) and $\lambda_L I_{mN} + \text{P}_{\mathcal{B}}(\mathbf{Z}) \geq 0$, which may be a difficult task. In this section we will use the alternating direction methods of multipliers (ADMM) [37] to solve Problem (43) by showing that the constraints can be separated and each one can be enforced in an alternating way. First of all observe that, by defining the variable $\mathbf{P} := \lambda_L I_{mN} + \text{P}_{\mathcal{B}}(\mathbf{Z})$, Problem (43) rewrites as

$$\begin{aligned} & \underset{\mathbf{Z}, \mathbf{P} \in \mathcal{C}}{\text{argmin}} && -\log \det(\hat{\Sigma}_y + \mathbf{Z}) - mN \\ & \text{subject to} && (36), (37) \\ & && \mathbf{P} = \lambda_L I_{mN} + \text{P}_{\mathcal{B}}(\mathbf{Z}) \\ & && \mathbf{P} \geq 0. \end{aligned} \quad (51)$$

where we have omitted the domain of the objective function $\hat{\Sigma}_y + \mathbf{Z} > 0$ since it will be checked in the stepsize-choice stage of the algorithm. The augmented Lagrangian for the problem is

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{Z}, \mathbf{P}, \mathbf{M}) &= -\log \det(\hat{\Sigma}_y + \mathbf{Z}) - \langle \mathbf{M}, \mathbf{P} - \lambda_L I_{mN} - \text{P}_{\mathcal{B}}(\mathbf{Z}) \rangle_{\mathcal{C}} \\ &+ \frac{\rho}{2} \|\mathbf{P} - \lambda_L I_{mN} - \text{P}_{\mathcal{B}}(\mathbf{Z})\|_{\mathcal{C}}^2 \end{aligned}$$

where $\rho > 0$ is the penalty term and $\mathbf{M} \in \mathcal{C}$ is the Lagrange multiplier associated to the equality constraint on \mathbf{P} . Accordingly, the ADMM updates are the following:

1) The \mathbf{Z} -minimization step

$$\begin{aligned} \mathbf{Z}^{k+1} = \operatorname{argmin}_{\mathbf{Z} \in \mathcal{C}} \quad & \mathcal{L}_\rho(\mathbf{Z}, \mathbf{P}^k, \mathbf{M}^k) \\ \text{subject to} \quad & \mathbf{Z} \in \mathcal{Z}. \end{aligned} \quad (52)$$

2) The \mathbf{P} -minimization step

$$\begin{aligned} \mathbf{P}^{k+1} = \operatorname{argmin}_{\mathbf{P} \in \mathcal{C}} \quad & \mathcal{L}_\rho(\mathbf{Z}^{k+1}, \mathbf{P}, \mathbf{M}^k) \\ \text{subject to} \quad & \mathbf{P} \geq 0. \end{aligned} \quad (53)$$

3) Dual variable update

$$\mathbf{M}^{k+1} = \mathbf{M}^k - \rho \left(\mathbf{P}^{k+1} - \lambda_L I_{mN} - \mathbf{P}_{\mathcal{B}}(\mathbf{Z}^{k+1}) \right). \quad (54)$$

where $\mathcal{Z} := \{\mathbf{Z} \in \mathcal{C} : (36), (37)\}$ and we have considered a constant value of ρ in order to simplify the notation. We will discuss later how to update ρ to get a faster convergence. Updates 1) and 2) are not in an implementable format. The \mathbf{Z} -update step (52) is equivalent to the minimization of

$$\begin{aligned} \mathcal{J}(\mathbf{Z}) := & -\log \det(\hat{\Sigma}_y + \mathbf{Z}) + \frac{\rho}{2} \|\mathbf{P}_{\mathcal{B}}(\mathbf{Z})\|_{\mathcal{C}}^2 \\ & + \left\langle \mathbf{M}^k - \rho (\mathbf{P}^k - \lambda_L I_{mN}), \mathbf{P}_{\mathcal{B}}(\mathbf{Z}) \right\rangle_{\mathcal{C}}, \end{aligned}$$

over the set \mathcal{Z} , which has no closed-form solution, as noticed in [4] where the solution is approximated by a projective-gradient step. Following the same lines, the new \mathbf{Z} -update step starts from a known feasible point $\mathbf{Z}^0 = \bar{\mathbf{Z}}$ and continue the iterations following the update rule

$$\mathbf{Z}^{k+1} = \mathbf{P}_{\mathcal{Z}} \left(\mathbf{Z}^k - t_k \nabla \mathcal{J}(\mathbf{Z}^k) \right) \quad (55)$$

where

$$\nabla \mathcal{J}(\mathbf{Z}^k) = -(\hat{\Sigma}_y + \mathbf{Z}^k)^{-1} + \mathbf{P}_{\mathcal{B}}(\mathbf{M}^k) + \rho \mathbf{P}_{\mathcal{B}} \left(\mathbf{Z}^k - \mathbf{P}^k + \lambda_L I_{mN} \right)$$

is the gradient of the cost-function \mathcal{J} computed in \mathbf{Z}^k , t_k is the stepsize founded by the Armijo condition, and $\mathbf{P}_{\mathcal{Z}}$ is the projection operator onto the constraints space \mathcal{Z} .

The optimization problem involved in the \mathbf{P} -update step (53) is equivalent to minimize the functional

$$\mathcal{J}(\mathbf{P}) := \frac{\rho}{2} \|\mathbf{P}\|_{\mathcal{C}}^2 - \left\langle \mathbf{P}, \mathbf{M}^k + \rho \left(\lambda_L I_{mN} + \mathbf{P}_{\mathcal{B}}(\mathbf{Z}^{k+1}) \right) \right\rangle_{\mathcal{C}}$$

over all $\mathbf{P} \geq 0$. Since \mathcal{J} is a quadratic functional of \mathbf{P} the minimization of \mathcal{J} over the whole vector space \mathcal{C} admits the closed form solution

$$\mathbf{P}_o = \frac{1}{\rho} \mathbf{M}^k + \lambda_L I_{mN} + \mathbf{P}_{\mathcal{B}}(\mathbf{Z}^{k+1})$$

but it is not a positive semidefinite matrix in general. Accordingly, in order to find our solution, we have to find the positive semidefinite block-circulant matrix that better approximates \mathbf{P}_o in the norm induced by the scalar product on \mathcal{C} (i.e. the Frobenius norm on \mathcal{C}). Recall the following well-known result.

Lemma 1: Let $A \in \mathbb{C}^{n \times n}$ be an Hermitian matrix whose eigenvalue decomposition is given by $A = U^* \Lambda U$, with

$$U^* U = U U^* = I \quad \text{and} \quad \Lambda = \operatorname{diag}\{\lambda_1, \dots, \lambda_n\}.$$

Then, the positive semidefinite matrix that better approximates A in the Frobenius norm is the projection of A onto the cone of positive semidefinite matrices \mathcal{P}^+ , namely

$$\mathbf{P}_{\mathcal{P}^+}(A) := \operatorname{argmin}_{X \geq 0} \|X - A\|_F = U^* \operatorname{diag}\{\gamma_1^{\rho}, \dots, \gamma_n^{\rho}\} U,$$

where

$$\gamma_i^{\rho} = \begin{cases} \lambda_{ii}, & \text{if } \lambda_{ii} \geq 0, \\ 0, & \text{if } \lambda_{ii} < 0. \end{cases}$$

The following proposition ensures that the projection of a symmetric, block-circulant matrix onto the cone of positive semi-definite matrices is still block-circulant.

Proposition 6: Let \mathbf{C} be a symmetric, block-circulant matrix

$$\mathbf{C} = \mathbf{F}^* \operatorname{diag}\{C(\zeta^0), C(\zeta^1), \dots, C(\zeta^{N-1})\} \mathbf{F},$$

and let $C(\zeta^k) = V_k \Lambda_k V_k^*$ with

$$V_k^* V_k = V_k V_k^* = I_m \quad \text{and} \quad \Lambda_k = \operatorname{diag}\{\lambda_{k1}, \dots, \lambda_{km}\},$$

being the eigen-decomposition of the (Hermitian) block $C(\zeta^k)$, for $k = 0, \dots, N-1$. Then the eigen-decomposition of \mathbf{C} can be written as

$$\mathbf{C} = \mathbf{W}^* \Lambda \mathbf{W}, \quad \mathbf{W} = \mathbf{V}^* \mathbf{F},$$

where $\mathbf{V} = \operatorname{diag}\{V_0, \dots, V_{N-1}\}$ and $\Lambda = \operatorname{diag}\{\Lambda_0, \dots, \Lambda_{N-1}\}$. Then,

$$\mathbf{P}_{\mathcal{C}^+}(\mathbf{C}) := \operatorname{argmin}_{\mathbf{X} \geq 0} \|\mathbf{X} - \mathbf{C}\|_{\mathcal{C}} = \mathbf{W}^* \operatorname{diag}\{\Gamma_0, \dots, \Gamma_{N-1}\} \mathbf{W}$$

where $\Gamma_k = \operatorname{diag}\{\gamma_{k1}, \dots, \gamma_{km}\}$ and

$$\gamma_{ki} = \begin{cases} \lambda_{ki}, & \text{if } \lambda_{ki} \geq 0, \\ 0, & \text{if } \lambda_{ki} < 0. \end{cases}$$

for $k = 0, \dots, N-1$.

Proof: The result follows from applying Lemma 1 with $U = \mathbf{W}$. Of course,

$$\mathbf{P}_{\mathcal{C}^+}(\mathbf{C}) = \mathbf{F}^* \operatorname{diag}\{V_0 \Gamma_0 V_0^*, \dots, V_{N-1} \Gamma_{N-1} V_{N-1}^*\} \mathbf{F}$$

is a block-circulant matrix because it is block-diagonalized by the Fourier-block matrix. \square

According to Proposition 6, the positive semidefinite block-circulant matrix that better approximates \mathbf{P}_o in the \mathcal{C} norm is the projection of \mathbf{P}_o onto the cone of the symmetric, positive semidefinite, block-circulant matrices \mathcal{C}^+ , that is

$$\mathbf{P}^{k+1} = \mathbf{P}_{\mathcal{C}^+}(\mathbf{P}_o) = \mathbf{P}_{\mathcal{C}^+} \left(\frac{1}{\rho} \mathbf{M}^k + \lambda_L I_{mN} + \mathbf{P}_{\mathcal{B}}(\mathbf{Z}^{k+1}) \right). \quad (56)$$

We conclude that the ADMM algorithm for the estimation of the sparse and the low-rank component of the inverse of the covariance matrix of the reciprocal process consists in the following updates

$$\begin{aligned} \mathbf{Z}^{k+1} &= \mathbf{P}_{\mathcal{Z}} \left[\mathbf{Z}^k - t_k \nabla \mathcal{J}(\mathbf{Z}^k) \right], \\ \mathbf{P}^{k+1} &= \mathbf{P}_{\mathcal{C}^+} \left[\frac{1}{\rho^k} \mathbf{M}^k + \lambda_L I_{mN} + \mathbf{P}_{\mathcal{B}}(\mathbf{Z}^{k+1}) \right], \\ \mathbf{M}^{k+1} &= \mathbf{M}^k - \rho^k \left[\mathbf{P}^{k+1} - \lambda_L I_{mN} - \mathbf{P}_{\mathcal{B}}(\mathbf{Z}^{k+1}) \right]. \end{aligned} \quad (57)$$

A typical update for ρ is $\rho^{k+1} = \alpha \rho^k$, with $\alpha > 1$ being a certain growth coefficient that needs to be properly tuned. Notice that the matrices involved in (57) are all symmetric and block-circulant. Accordingly, as explained in Section III, the introduction of the reciprocal approximation allows to obtain a robust identification procedure even in the case when n is large. Indeed, relations (14) and (15), allow to compute inverse matrices and eigenvalues in a robust way. Moreover, it is worth noting from (15), that the dimensions of the matrices whose eigenvalues must be computed in the optimization procedure, depend only on m , hence the identification algorithm we are proposing scales with respect to n gaining robustness in the results even if the order of the AR process is large.

Following [37], the basic stopping criterium for the algorithm is based on the primal and dual residuals of the optimality conditions that respectively measure the satisfaction of the inequality constraint $\mathbf{P} \geq 0$ and the distance between two successive iterates of the variable \mathbf{P} . More precisely, the primal residual at iteration $k+1$ is defined as

$$r^{k+1} := \mathbf{P}^{k+1} - \lambda_L I_{mN} - \mathbf{P}_{\mathcal{B}}(\mathbf{Z}^{k+1}),$$

while the dual residual turns out to be

$$s^{k+1} := \mathbf{P}_{\mathcal{B}^c}(\mathbf{M}^k) - \rho^k [\mathbf{P}^{k+1} - \mathbf{P}_{\mathcal{B}}(\mathbf{P}^k)].$$

It is reasonable that the primal and dual residual must be small, that is

$$\|r^k\|_e \leq \varepsilon^p \quad \text{and} \quad \|s^k\|_e \leq \varepsilon^d,$$

where $\varepsilon^p > 0$ and $\varepsilon^d > 0$ are feasibility tolerances for the primal and dual feasibility conditions. The latter are defined as

$$\begin{aligned} \varepsilon^p &:= mN \varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max \left\{ \lambda_L \sqrt{mN}, \|\mathbf{Z}^k\|_e, \|\mathbf{P}^k\|_e \right\}, \\ \varepsilon^d &:= mN \varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \|\mathbf{M}^k\|_e. \end{aligned}$$

Here, ε^{abs} and ε^{rel} are predefined absolute and relative tolerances for the problem. Accordingly, the algorithm converges if all the conditions

$$\|r^k\|_e \leq \varepsilon^p, \quad \|s^k\|_e \leq \varepsilon^d, \quad \rho^k = \rho_{\max} \quad (58)$$

hold true, where $\rho_{\max} > 0$ is the maximum value allowed for the penalty parameter ρ^k , selected by the user.

VII. NUMERICAL EXAMPLES

In this section we compare the performances of our method to which we will refer to as *approximated algorithm* with the one proposed in [10] for the solution of Problem (31), which will be referred to as *exact algorithm*. In particular we will show how the two algorithms behave considering both the case in which the observed process has low dimension and the case in which we have an high dimensional observed process.

Low-dimensional case: Synthetic data are generated from the AR latent-variable model of order $n = 8$,

$$\mathbf{y}(t) = \sum_{k=1}^n A_k \mathbf{y}(t-k) + \boldsymbol{\eta}(t), \quad (59)$$

with $m = 20$ observed variables and $l = 1$ latent variables. Here, $\boldsymbol{\eta}(t)$ is white Gaussian noise with variance

$\mathbb{E}[\boldsymbol{\eta}(t)^\top \boldsymbol{\eta}(t)] = 21.14$ and $T = 1000$ samples have been used to compute the estimated covariance lags \hat{R}_k , $k = 0, \dots, n$. Figure 4 (center) reports the sparsity pattern of the underlying model, randomly generated so that the non-zero elements represents the 5% of the total elements. For the approxi-

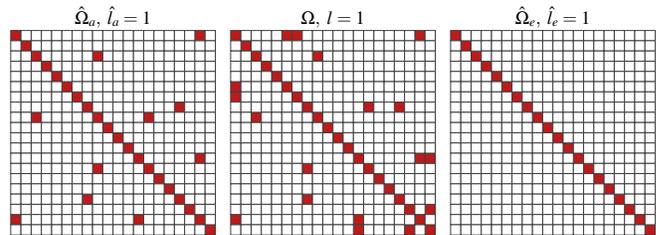


Fig. 4: Sparsity pattern estimated by the approximated algorithm with $\alpha = 1.007$, $\lambda_S = 95$, $\lambda_L = 5.4$ (left), true sparsity pattern (center), sparsity pattern estimated by the exact algorithm with $\alpha = 1.002$, $\gamma_S = 2.6$, $\gamma_L = 2.95$ (right). The red squares indicate the conditional dependent pairs while the white squares indicates the conditional independent pairs. \hat{l}_a , l and \hat{l}_e denote the number of latent variables.

mated algorithm we have considered $N = 30$ samples of the spectrum. In both the ADMM implementations we have set $\varepsilon^{\text{abs}} = 10^{-5}$ and $\varepsilon^{\text{rel}} = 10^{-4}$ while $\rho_{\max} = 10^4$. In order to tune the update of the penalty term ρ in the ADMM, we have ran both the algorithms for different values of the growth coefficient $\alpha \in [1.001, 1.1]$. More precisely, for each value of α , a 5×5 grid of candidate estimated models has been produced, corresponding to five linearly spaced values of the regularization parameters $\lambda_S \in [60, 130]$ and $\lambda_L \in [3, 7.8]$ for the approximated algorithm, and five linearly spaced values of $\gamma_S \in [1.42, 2.6]$ and $\gamma_L \in [2.425, 2.95]$ for the exact algorithm. The values of the regularization parameters that identify the grids have been selected so that the estimated models capture a range of features as complete as possible: from a very sparse model with a relatively high rank, to a quasi-full model with the lowest rank possible. Figure 5 shows the supports and the ranks estimated by the approximated algorithm corresponding to the different values of λ_S and λ_L . For both methods the value of α that gives the better performances, i.e. that guarantees the minimum gap between $\varepsilon^p / \varepsilon^d$ and the primal/dual residual at the final iteration, respectively, has been selected. Accordingly, we have chosen $\alpha = 1.007$ for the approximated algorithm while $\alpha = 1.002$ has been chosen for the exact algorithm. Let $\mathbf{s}(h)$ and $\boldsymbol{\varepsilon}(h)$ denote the vectors containing the dual residual and its feasibility tolerance for the model $h = 1, \dots, 25$ respectively. Figure 6 displays the (logarithm of the) averages

$$\mu_s = \frac{1}{25} \sum_{h=1}^{25} \mathbf{s}(h), \quad \mu_\varepsilon = \frac{1}{25} \sum_{h=1}^{25} \boldsymbol{\varepsilon}(h),$$

obtained by our method with $\alpha = 1.007$ (left) and by the exact method for $\alpha = 1.002$ (right). For both algorithms, the primal residual always satisfies the condition in the stopping criterium (58) therefore there is no need to displaying it. We observe that the exact algorithm does not converge for any value of α we have considered. Indeed, the plot in Figure 6 (right) clearly shows that the mean dual-residual

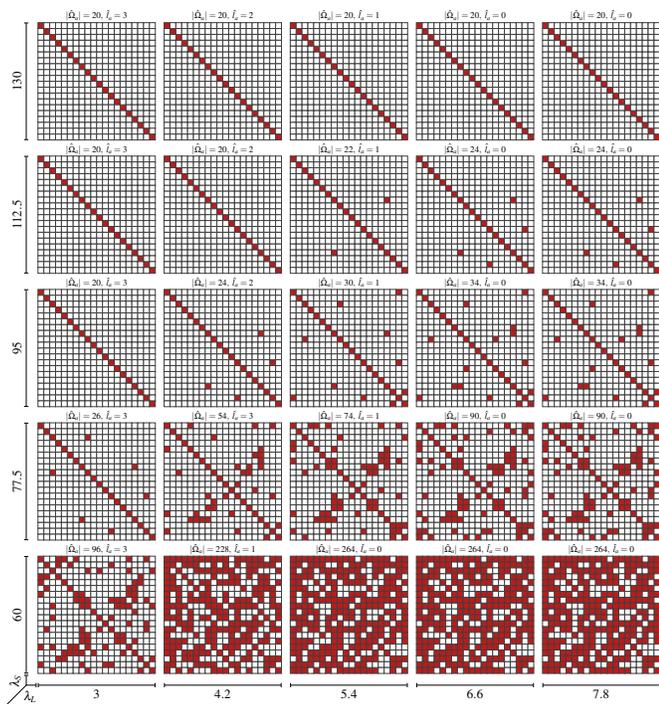


Fig. 5: Supports and ranks estimated by the approximated algorithm for $\lambda_S \in [60, 130]$ and $\lambda_L \in [3, 7.8]$. The growth coefficient is set $\alpha = 1.007$.

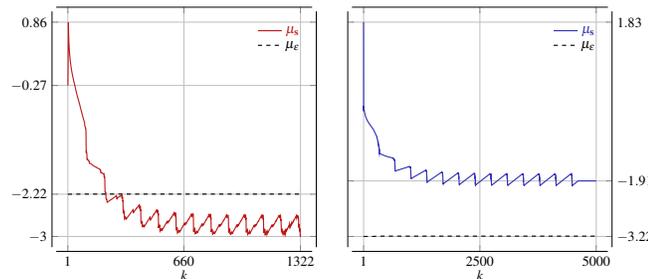


Fig. 6: Logarithm of the average dual residual for the approximated method (left) and for the exact method (right). The dashed lines correspond to the logarithm of the associated average feasibility tolerances.

μ_s stays significantly above the threshold μ_ϵ . The optimal values of the regularization parameters have then been selected by cross-validation, using a test data set of 500 samples. Figure 4 compares the optimal sparsity pattern provided by the approximated algorithm $\hat{\Omega}_a$ (left), corresponding to $\lambda_S = 95$ and $\lambda_L = 5.4$, and the optimal sparsity pattern estimated by the exact algorithm $\hat{\Omega}_e$ (right) corresponding to $\gamma_S = 2.6$ and $\gamma_L = 2.95$, together with the estimates of the number of latent variables, \hat{l}_a and \hat{l}_e , respectively. Notice that both algorithms estimates the correct number of latent variables but only the approximated one produces an estimate of the sparsity pattern comparable with the true one. Let $\hat{\Phi}_e$ and $\hat{\Phi}_a$ be the estimates of the spectra of the true observed process Φ_y obtained by the solutions of problems (31) and (32), respectively. According to Figure 2, $\hat{\Phi}_a$ is the extension over the whole interval $[-\pi, \pi]$ of the symbol of the estimated covariance matrix of the reciprocal

process y approximating y . The squared-estimation errors for the two algorithms are depicted in Figure 7; the corresponding mean values over $[-\pi, \pi]$ are

$$\begin{aligned} \mathcal{E}_a &:= \frac{\|\Phi_y - \hat{\Phi}_a\|_F^2}{\|\Phi_y\|_F^2}, & \bar{\mathcal{E}}_a &:= \int \mathcal{E}_a(e^{i\theta}) = 0.0358, \\ \mathcal{E}_e &:= \frac{\|\Phi_y - \hat{\Phi}_e\|_F^2}{\|\Phi_y\|_F^2}, & \bar{\mathcal{E}}_e &:= \int \mathcal{E}_e(e^{i\theta}) = 0.0443. \end{aligned}$$

The approximated algorithm performs better both in terms of the mean value and in terms of the height of the peaks of the relative error.

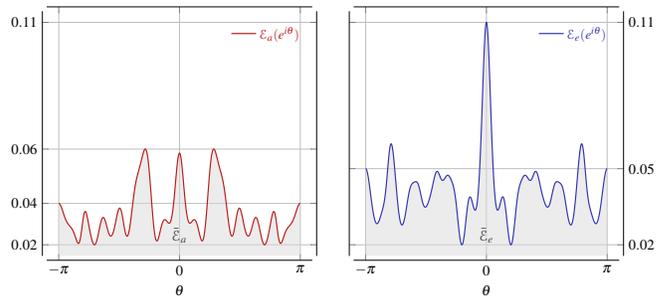


Fig. 7: Relative errors in the estimated spectra: approximated algorithm (left), exact algorithm (right).

High-dimensional case: We consider now an AR latent-variable model as in (59) where we have $m = 80$ observed variables and $l = 1$ latent variable, $n = 24$ and the variance of the noise is $\mathbb{E}[\eta(t)^\top \eta(t)] = 87.1$. The number of samples used to estimate the covariance lags R_k is $T = 15000$. The number of conditionally dependent pairs in the true model is 158 so that the cardinality of the true support is $|\Omega| = 396$. Table 8 compares the performances of our approximated algorithm with the exact algorithm proposed in [10] for different values of the sparsity regularization parameters λ_S and γ_S , that have been chosen in order to have approximately the same variety on the results. The notation $|\Omega - \hat{\Omega}|$ indicates the error on the sparsity pattern in terms of number of misclassified entries. Both algorithms estimate the correct number of latent

\hat{l}_a	$ \hat{\Omega}_a $	$ \Omega - \hat{\Omega}_a $	\mathcal{E}_a	time [s]	\hat{l}_e	$ \hat{\Omega}_e $	$ \Omega - \hat{\Omega}_e $	\mathcal{E}_e	time [s]
1	80	316	0.2944	16609	1	88	308	0.5611	34875
1	394	10	0.3170	16953	1	396	156	0.5168	34681
1	1064	668	0.1902	17352	1	994	706	0.5701	35072

Fig. 8: Summary of the performances of the two algorithm for $\lambda_S = 100, 146.25, 350$ and $\gamma_S = 0.7, 0.826, 1.7$. The values of the low-rank regularization parameters are $\lambda_L = 8.6875$ for the approximated algorithm (left) and $\gamma_L = 2.3$ for the exact algorithm (right). These results have been obtained on a 2014 1.4GHz MacBook Air.

variables, but the approximated algorithm gives a result very close to the true one (highlighted in red in Figure 8) while for the exact algorithm, even if the cardinality of the true support has been correctly estimated, the error in the reconstruction of the sparsity pattern is quite high. This is due to the fact that the higher is the order of the process n , the less accurate

is the computation of eigenvalues and inverse matrices by the exact algorithm. Figure 8 shows that such an issue is avoided in the approximated version, thanks to the availability of closed-form formulas for the computation of the eigenvalues of block-circulant matrices. Moreover, we see that the run time of the exact algorithm is about twice the run time of the approximated one. This confirms the fact that the approximated algorithm scales with the order n of the AR process we are approximating as suggested in Section V. This kind of scenario agrees with what we have discussed in Section V: high-order AR processes are quite challenging instances for the exact procedure proposed in [10]; in this case, the reciprocal approximation leads to remarkable benefits in the performances of the identification procedure.

VIII. CONCLUSIONS

In this paper an identification paradigm for latent-variable graphical models associated to reciprocal processes has been presented. It has been shown that the proposed paradigm is theoretically strongly sustained, being an approximation of the corresponding problem for AR processes both in a maximum likelihood and in a maximum entropy sense. The performances of the proposed method have been compared with the approach proposed in [10] where no approximation is introduced. The numerical examples have shown that for high-order AR processes reciprocal approximation gives substantial improvements in terms of robustness and scalability of the identification procedure.

REFERENCES

- [1] S. Lauritzen, *Graphical Models*. Oxford, U.K.: Oxford university press, 1996.
- [2] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, no. 2, pp. 157–172, 2000.
- [3] E. Avventi, A. G. Lindquist, and B. Wahlberg, "Arma identification of graphical models," *IEEE Transactions on Automatic Control*, vol. 58, pp. 1167–1178, May 2013.
- [4] J. Songsiri and L. Vandenberghe, "Topology selection in graphical models of autoregressive processes," *J. Mach. Learn. Res.*, vol. 11, pp. 2671–2705, 2010.
- [5] J. Songsiri, J. Dahl, and L. Vandenberghe, "Maximum-likelihood estimation of autoregressive models with conditional independence constraints," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1701–1704, April 2009.
- [6] S. Maanan, B. Dumitrescu, and C. Giurcaneanu, "Conditional independence graphs for multivariate autoregressive models by convex optimization: Efficient algorithms," *Signal Processing*, vol. 133, pp. 1–11, 2016.
- [7] S. Maanan, B. Dumitrescu, and C. Giurcaneanu, "Maximum entropy expectation-maximization algorithm for fitting latent-variable graphical models to multivariate time series," *Entropy*, vol. 20, p. 76, 01 2018.
- [8] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *Ann. Statist.*, vol. 40, pp. 1935–1967, 08 2012.
- [9] M. Zorzi and R. Sepulchre, "AR identification of latent-variable graphical models," *IEEE Transactions on Automatic Control*, vol. 61, pp. 2327–2340, Sept 2016.
- [10] R. Ligeois, B. Mishra, M. Zorzi, and R. Sepulchre, "Sparse plus low-rank autoregressive identification in neuroimaging time series," in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 3965–3970, Dec 2015.
- [11] F. P. Carli, A. Ferrante, M. Pavon, and G. Picci, "A maximum entropy solution of the covariance extension problem for reciprocal processes," *IEEE Transactions on Automatic Control*, vol. 56, pp. 1999–2012, Sept 2011.
- [12] B. C. Levy and A. Ferrante, "Characterization of stationary discrete-time gaussian reciprocal processes over a finite interval," *SIAM Journal on Matrix Analysis and Applications*, vol. 24, no. 2, pp. 334–355, 2002.
- [13] B. C. Levy, "Regular and reciprocal multivariate stationary gaussian reciprocal processes over z are necessarily markov," *J. Math. Syst. Est. Control*, vol. 2, no. 2, pp. 134–154, 1992.
- [14] B. C. Levy, R. Frezza, and A. J. Krener, "Modeling and estimation of discrete-time gaussian reciprocal processes," *IEEE Transactions on Automatic Control*, vol. 35, pp. 1013–1023, Sep 1990.
- [15] A. Ringh, J. Karlsson, and A. Lindquist, "Multidimensional rational covariance extension with applications to spectral estimation and image compression," *SIAM Journal on Control and Optimization*, vol. 54, no. 4, pp. 1950–1982, 2016.
- [16] A. G. Lindquist and G. Picci, "The circulant rational covariance extension problem: The complete solution," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2848–2861, Nov 2013.
- [17] A. Chiuso, A. Ferrante, and G. Picci, "Reciprocal realization and modeling of textured images," in *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 6059–6064, Dec 2005.
- [18] F. Carli, A. Ferrante, M. Pavon, and G. Picci, "An efficient algorithm for maximum entropy extension of block-circulant covariance matrices," *Linear Algebra and its Applications*, vol. 439, no. 8, pp. 2309 – 2329, 2013.
- [19] A. Ringh and J. Karlsson, "A fast solver for the circulant rational covariance extension problem," in *Control Conference (ECC), 2015 European*, pp. 727–733, IEEE, 2015.
- [20] D. Alpago, "On the identification of sparse plus low-rank graphical models," Master's thesis, University of Padova, Dept. of Information Engineering, Padova, Italy, 2017.
- [21] A. Ferrante, M. Pavon, and M. Zorzi, "A maximum entropy enhancement for a family of high-resolution spectral estimators," *IEEE Transactions on Automatic Control*, vol. 57, pp. 318–329, Feb 2012.
- [22] M. Zorzi and A. Ferrante, "On the estimation of structured covariance matrices," *Automatica*, vol. 48, no. 9, pp. 2145 – 2151, 2012.
- [23] J. Burg, *Maximum entropy spectral analysis*. PhD thesis, Stanford University, Dept. of Geophysics, Stanford, CA, 1975.
- [24] M. Zorzi, "An interpretation of the dual problem of the THREE-like approaches," *Automatica*, vol. 62, pp. 87 – 92, 2015.
- [25] M. Zorzi, "A new family of high-resolution multivariate spectral estimators," *IEEE Transactions on Automatic Control*, vol. 59, pp. 892–904, April 2014.
- [26] A. Ringh, J. Karlsson, and A. Lindquist, "Multidimensional rational covariance extension with applications to spectral estimation and image compression," *SIAM Journal on Control and Optimization*, vol. 54, no. 4, pp. 1950–1982, 2016.
- [27] T. Georgiou and A. Lindquist, "Kullback-leibler approximation of spectral density functions," *Information Theory, IEEE Transactions on*, vol. 49, pp. 2910 – 2917, 12 2003.
- [28] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Transactions on Automatic Control*, vol. 57, pp. 2561–2575, Oct 2012.
- [29] T. Georgiou, "Relative entropy and the multivariable multidimensional moment problem," *IEEE Trans. on Information Theory*, vol. 52, no. 3, pp. 1052–1066, 2006.
- [30] C. Byrnes, T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. on Signal Processing*, vol. 48, no. 11, pp. 3189–3205, 2000.
- [31] C. I. Byrnes, P. Enqvist, and A. Lindquist, "Identifiability and well-posedness of shaping filter parametrizations: A global analysis approach," *SIAM Journal on Control and Optimization*, vol. 41, no. 1, pp. 23–59, 2002.
- [32] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [33] D. Alpago, M. Zorzi, and A. Ferrante, "Identification of sparse reciprocal graphical models," *IEEE Control Systems Letters*, vol. 2, pp. 659–664, Oct 2018.
- [34] A. G. Lindquist and G. Picci, *Linear Stochastic Systems: a Geometric Approach to Modeling, Estimation and Identification*. Berlin, Germany: Springer, 2015.
- [35] J. Songsiri, J. Dahl, and L. Vandenberghe, "Graphical models of autoregressive processes," *Convex optimization in signal processing and communications*, pp. 89–116, 2010.
- [36] M. Pavon and A. Ferrante, "On the geometry of maximum entropy problems," *SIAM Review*, vol. 55, no. 3, pp. 415–439, 2013.
- [37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, Jan. 2011.