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

Daniele Alpago, Mattia Zorzi, *Senior Member, IEEE*, and Augusto Ferrante

***Abstract*—In this article, we propose an identification method for latent-variable graphical models associated with 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, maximum entropy, maximum likelihood, reciprocal processes, regularization, system identification.**

I. Introduction

GRAPHICAL models for time-series have been introduced in [1] and [2] and have become an important tool for the analysis and the identification of stochastic processes: among the first papers dealing with these issues, we mention [3], [4]; in [5], complexity criteria for selecting the best graphical model have been introduced; a Bayesian point of view has been discussed in [6]. These models provide a graphical representation of the conditional independence relations among the process' components, thus highlighting the interdependence among such components. As a consequence, graphical models may be used to uncover the topological structure of the system generating the observed data. Observe that the smaller is the number of edges of a graphical model the richer is the information provided by the model on the structure of the system (in the extreme case when the graph is full, it does not provide any information and does not uncover any structure). In other words, sparse graphs correspond to parsimonious models. It may happen, however, that most of the components of the graph are genuinely interconnected thus corresponding to a graph that is almost full and hence essentially uninformative. This may be due to the fact that the data structure is based on the presence of a few latent (i.e., unobserved) variables that explain most of the interdependence among the process components. Thus, by taking into account the presence of latent variables we may unveil a whole hidden structure in the system. A simple example is that of the electrical power consumption of each of the houses in a certain city. It may happen that the strong correlation among all the components is mostly explained by a common correlation with a single unobserved variable such as the external temperature. Graphical models containing also nodes accounting for the effect of the latent variables are called latent-variable graphical models [7], [8].

In this article, we consider the problem of identification of latent-variable graphical models associated with Gaussian autoregressive (AR) processes. This problem, first considered in [8], was effectively solved in [9] by using the alternating direction methods of multipliers (ADMM). Nevertheless, that procedure involves matrix inversions and eigenvalue decompositions which do not scale well with the process dimension.

The purpose of this article is to propose an alternative identification method addressing the numerical complexity of high-order AR processes in latent-variable graphical models. This is achieved by using *reciprocal processes* (see [10], [11] for the general case and [12] for the stationary framework) which, under suitable conditions, well approximate AR processes [13]. The advantage of reciprocal approximation is that the processes are represented by circulant matrices (see [14] and [15]) with the consequent substantial reduction in computational complexity; see also the extensions to the multidimensional case [16], [17], and to the multivariate case [18], [19]. We define a latent-variable graphical structure for reciprocal processes and propose a new identification paradigm in this setting. We also provide a solid argument to justify this paradigm. Some preliminary results in the direction of the proposed paradigm have been obtained in [20]. In that paper, we addressed the identification of graphical models of an observed process with no latent variables. On one hand, this case is much simpler (we just model what we observe) and, on the other, it lacks the possibility to capture the structure induced by the presence of hidden variables that, as previously discussed, can provide a much more interesting model. For more details on the importance of latent variables and their advantages in providing useful models we refer to [21].

The main contributions of this article are the following.

Manuscript received April 14, 2021; accepted July 10, 2021. Date of publication July 16, 2021; date of current version June 30, 2022. This work was supported by the Department of Information Engineering, University of Padova, Italy, through SID Project, a multidimensional, and multivariate moment problem theory for target parameter estimation in automotive radars under Grant ZORZ_SID_1901. Recommended by Associate Editor D. Regruto. *(Corresponding author: Daniele Alpago.)*

The authors are with the Department of Information Engineering, University of Padova, 35131 Padova, Italy (e-mail: daniele.alpago@phd.unipd.it; zorzimat@dei.unipd.it; augusto@dei.unipd.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2021.3097558.

Digital Object Identifier 10.1109/TAC.2021.3097558

1) We establish a new identification paradigm and provide a conceptual argument to show that it relies on both maximum likelihood and maximum entropy principles.
2) We develop the variational analysis of the corresponding optimization problem and prove existence and uniqueness of the solution.
3) We derive an explicit ADMM scheme for the numerical solution of the problem.
4) We provide numerical simulations with a comparison of the ADMM schemes proposed in this article and in [9]. The experiments confirm the following two advantages—expected from the fact that we are working with symmetric, circulant matrices—of our scheme. a) Our algorithm is scalable with respect to the order and the dimension of the AR process, and thus, when such parameters increase, it outperforms the existing one. b) Our algorithm is more robust and stable in that, for high-dimensional instances, the ADMM scheme in [9] fails to converge due to numerical instability. On the contrary, our scheme does indeed converge.

The remainder of the article is organized as follows. In Section II, we set the notation and we recall the fundamental results used throughout the article. In Section II-A, we introduce reciprocal processes and we explain how they are related to AR processes. Latent-variable graphical models associated with reciprocal processes are defined in Section III while, in Section IV, we propose a convex optimization problem for their identification. Section V is devoted to the ADMM formulation of the optimization problem then used in Section VI to validate the proposed approach through numerical experiments. Finally, in Section VII, we draw the conclusion.

II. Notation and Background

We denote by $\mathbb{T} := \{e^{i\theta} : \theta \in [-\pi, \pi]\}$ the unit circle in $\mathbb{C}$. For (matrix-valued) functions defined on $\mathbb{T}$, we omit the dependence on θ when it is clear from the context. With $\mathrm{rank}(G)$ we denote the rank of the matrix G or its normal rank if G is a matrix function defined on $\mathbb{T}$. Given a matrix G, $G^\top$ denotes its transpose, G^* its transpose-conjugate and $\ker(G)$ its kernel. If G is a square matrix, we denote by $\mathrm{diag}(G)$ the vector whose entries are the diagonal elements of G, while $\mathrm{tr}(G)$, $\det(G)$ and G^{-1} denote the trace, the determinant and the inverse of G, respectively. $G > 0$ and $G \geq 0$ denote that G is positive definite and, respectively, positive semidefinite. I_p is the identity matrix of order p. We denote by $\mathcal{H}_p$ the space of square integrable coercive functions on the unit circle taking values in the space of $p \times p$ Hermitian matrices. For any $F \in \mathcal{H}_p$, $\int F$ denotes the integral of F over $[-\pi, \pi]$ with respect to the normalized Lebesgue measure on $\mathbb{T}$.

Given $(x_1, x_2) \subset (a, b) \subset \mathbb{R}$, 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. We consider AR processes and periodic reciprocal processes: both are assumed to be zero-mean; n denotes the order of the AR process and N the period of the reciprocal process. It is always assumed that $N > 2n$ and that N is an even number (the case of N odd is similar). We define the vector space $\mathcal{C} \subset \mathbb{R}^{mN \times mN}$ of the (real) symmetric, block-circulant matrices $\mathbf{C} = \mathrm{circ}\{C_0, C_1, \ldots, C_{\frac{N}{2}-1}, C_{\frac{N}{2}}, C_{\frac{N}{2}-1}^\top, \ldots, C_1^\top\}$, whose first block-column is composed by the $m \times m$ blocks $C_0, C_1, \ldots, C_{\frac{N}{2}-1}, C_{\frac{N}{2}}, C_{\frac{N}{2}-1}^\top, \ldots, C_1^\top$. The space $\mathcal{C}$ is endowed with the inner product $\langle \mathbf{C}, \mathbf{D} \rangle_{\mathcal{C}} := \mathrm{tr}(\mathbf{C}^\top \mathbf{D})$. The *symbol* of $\mathbf{C} \in \mathcal{C}$ is defined as the $m \times m$ pseudo-polynomial

$$\Upsilon(\zeta) := \sum_{k=0}^{N-1} C_k \, \zeta^{-k}, \text{ with } \zeta := e^{i\frac{2\pi}{N}} \text{ and } C_k = C_{N-k}^\top \text{ for } k > \frac{N}{2}. \tag{1}$$

We recall that circulant matrices are diagonalized by Fourier matrices (see e.g., [22, p. 6]) i.e.,

$$\mathbf{C} = \mathbf{F}^* \mathrm{diag}\left\{\Upsilon(\zeta^0), \Upsilon(\zeta^1), \ldots, \Upsilon(\zeta^{N-1})\right\} \mathbf{F} \tag{2}$$

where $\mathbf{F}$ is the (Fourier) unitary $N \times N$ block-matrix whose block in block-row of order $h+1$, $h = 0, \ldots, N-1$, and block-column of order $k+1$, $k = 0, \ldots, N-1$ is $\zeta^{-h \cdot k} I$.

We define the subspace $\mathcal{B} \subseteq \mathcal{C}$ (with the same inner product) of symmetric, banded block-circulant $mN \times mN$ matrices of bandwidth n, containing the matrices of the form

$$\mathbf{B} = \mathrm{circ}\{B_0, B_1, \ldots, B_n, 0, \ldots, 0, B_n^\top, \ldots, B_1^\top\}. \tag{3}$$

Note that, according to definition (1), the symbol of $\mathbf{B} \in \mathcal{B}$ is $\Psi(\zeta) = \sum_{k=-n}^{n} B_k \, \zeta^{-k}$, with $B_{-k} = B_k^\top$. The projection operator $\mathsf{P}_{\mathscr{B}} : \mathcal{C} \to \mathcal{B}$ is defined as $\mathsf{P}_{\mathscr{B}}(\mathbf{C}) := \mathrm{circ}\{C_0, C_1, \ldots, C_n, 0, \ldots, 0, C_n^\top, \ldots, C_1^\top\}$. Given $\Omega = \{(i, j) : i, j = 1, \ldots, m\}$, the projection operator $\mathsf{P}_\Omega : \mathcal{C} \to \mathcal{C}$ is defined such that $\mathsf{P}_\Omega(\mathbf{C})$ is a block-circulant matrix whose blocks have support Ω.

A. AR Processes and Their Reciprocal Approximation

In this section, we recall some facts about periodic Gaussian reciprocal processes, about their identification based on maximum entropy principle, and about their relation to AR processes.

Let $\{\mathbf{y}(k), k = 1, 2, \ldots, N\}$ be an m-dimensional Gaussian stationary stochastic process defined on a finite interval $[1, N]$. For $k = 1, \ldots, N$, we have $\mathbf{y}(k) := [\boldsymbol{y}_1(k) \ \ldots \ \boldsymbol{y}_m(k)]^\top \in \mathbb{R}^m$, therefore the process is completely characterized by the random vector $\mathbf{y} := [\boldsymbol{y}_1(1) \ \ldots \ \boldsymbol{y}_m(1) \ \ldots\ldots \ \boldsymbol{y}_1(N) \ \ldots \ \boldsymbol{y}_m(N)]^\top \in \mathbb{R}^{mN}$. In [13], it is proved 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 $\mathbf{\Sigma}$ of $\mathbf{y}$ is symmetric block-circulant: $\mathbf{\Sigma} = \mathrm{circ}\{\Sigma_0, \Sigma_1, \ldots, \Sigma_{\frac{N}{2}}, \ldots, \Sigma_1^\top\}$, where $\Sigma_{i-j} := \mathbb{E}[\mathbf{y}(i)\mathbf{y}(j)^\top]$, $i, j = 1, \ldots, N$, are the covariance lags of the process so that $\Sigma_k = \Sigma_{N-k}^\top$ for $k > N/2$. 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 from the random variables in $(t_1, t_2)^c$, given the $2n$ boundary values $\mathbf{y}(t_1 - n + 1), \ldots, \mathbf{y}(t_1), \mathbf{y}(t_2), \ldots, \mathbf{y}(t_2 + n - 1)$, where the sums $t - k$ and $t + k$ are understood to be modulo N.

The following result has been proved in [13, Th. 3.3]:

Theorem 1: A nonsingular $mN \times mN$-dimensional matrix $\mathbf{\Sigma}$ is the covariance matrix of a periodic reciprocal process of order n if and only if $\mathbf{\Sigma}^{-1} \in \mathcal{B}$.

As detailed in [13], given the estimates $\hat{\Sigma}_0 \dots \hat{\Sigma}_n$ of the first $n+1$ covariance lags of the reciprocal process, the covariance $\mathbf{\Sigma_y}$ of the the *maximum entropy* reciprocal process with such covariance lags is obtained by solving the dual problem

$$\begin{aligned} \underset{\mathbf{X}\in\mathcal{B}}{\text{argmin}} \quad & -\log\det \mathbf{X} + \langle \mathbf{X},\, \hat{\mathbf{\Sigma}}\rangle_{\mathcal{C}} \\ \text{subject to} \quad & \mathbf{X} > 0 \end{aligned} \tag{4}$$

where $\hat{\Sigma} := \text{circ}\{\hat{\Sigma}_0, \hat{\Sigma}_1, \dots, \hat{\Sigma}_n, 0, \dots, 0, \hat{\Sigma}_n^\top, \dots, \hat{\Sigma}_1^\top\} \in \mathcal{B}$. The solution to (4) is $\mathbf{\Sigma}_{\mathbf{y}}^{-1}$.

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

$$\sum_{k=0}^{n} B_k\, \mathsf{y}(t-k) = \mathbf{e}(t), \qquad \mathbf{e}(t) \sim \mathcal{N}(0, I_m), \quad t \in \mathbb{Z} \tag{5}$$

and let $R_k := \mathbb{E}[\mathsf{y}(t)\mathsf{y}(t-k)^\top]$, $k \in \mathbb{Z}$, be its kth covariance lag. The spectrum of y is

$$\Phi(e^{i\theta}) = \sum_{k=-\infty}^{\infty} R_k\, e^{-i\theta k}, \qquad R_{-k} = R_k^\top, \;\; \theta \in [-\pi, \pi]. \tag{6}$$

This spectrum is associated with an infinite block-Toeplitz covariance matrix whose first block-row is $[R_0 \;\; R_1 \;\; \dots]$. Given the estimates $\hat{R}_0, \dots, \hat{R}_n$ of the first $n+1$ covariance lags, Burg in [23] proposed a method to obtain the maximum entropy extension of the sequence, i.e., the maximum entropy block-Toeplitz infinite covariance matrix. This problem has been extensively studied and generalized in the recent years: the maximum entropy estimator, indeed, belongs to a class of tunable high estimators proposed in [24] and [25]; the latter has been extended to the multivariate [26]–[28] and the multidimensional case [16], [29]; moreover, all those estimators can be seen in a unified way by means of the beta and tau divergence family [30], [31]; it is also worth noting that some of those paradigms have been also extended to the case in which cepstral matching is needed [32], moreover, it has been proved the convergence of a nonlinear mapping used to compute the optimal estimator [33]. We recall that, Toeplitz matrices of sufficiently large size can be approximated arbitrarily well by circulant matrices [34, Lemma 4.2]. This means that the AR model, specified by the Toeplitz matrix, can be approximated by the reciprocal model, specified by the block-circulant covariance matrix, see [13] for further details. In view of [14, Th. 3.1], for N sufficiently large, we conclude that the reciprocal process associated with the solution of (4) can be interpreted as an approximation of the AR process solution of the Burg's problem.

The reciprocal approximation has also an interesting interpretation in the frequency domain. Indeed, it corresponds to sample the spectrum (6) of the AR process 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

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

Notice that, being an approximation, in general $\Sigma_k \neq R_k$. They will match as $N \to \infty$, see [20] for more details about the frequency-domain interpretation of the reciprocal approximation of AR processes.

Remark 1: The estimation techniques based on matching some covariance lags (or some moments) such as the one proposed by Burg and the one in this article, may be viewed as an alternative to the spectral methods based on the periodogram. The advantage of the former is that, in general, they perform much better than the latter in terms of spectral resolution.

III. Latent-Variable Reciprocal Graphical Models

The aim of this section is to define latent-variable graphical models for Gaussian reciprocal processes and to establish the corresponding algebraic conditions on the associated covariance matrix. We start by recalling the concept of graphical model of a Gaussian random vector.

The graphical model associated with a Gaussian random vector $\mathbf{z}$ is an undirected graph $\mathcal{G} = (V, E)$ constructed by taking $V = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_m\}$, where $\boldsymbol{z}_i$ are the components of $\mathbf{z}$, and selecting the edges $E \subset V \times V$ by stipulating that an edge is not present between nodes $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ if and only if $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are conditionally independent given all the other $\boldsymbol{z}_k$, $k \neq i, j$. By recalling that this happens if and only if the entry in position (i, j) of the inverse of the covariance matrix Σ of $\mathbf{z}$ vanishes, we have the following relations:

$$(\boldsymbol{z}_i, \boldsymbol{z}_j) \notin E \iff \boldsymbol{z}_i \perp \boldsymbol{z}_j \,|\, \{z_k\}_{k\neq i,j} \iff (\Sigma^{-1})_{ij} = 0. \tag{7}$$

Hence, the graph $\mathcal{G}$ associated with $\mathbf{z}$ is completely determined by the sparsity pattern of the inverse covariance matrix Σ^{-1}, see [35] for further details.

If $\mathbf{z}$ is a random process defined on the interval $[1, N]$ (hence, in particular, if $\mathbf{z}$ is a reciprocal process) the corresponding graphical model is defined similarly by taking again $V = \{\mathbf{z}_1, \dots, \mathbf{z}_m\}$, where $\mathbf{z}_i := [\boldsymbol{z}_i(1) \dots \boldsymbol{z}_i(N)]^\top$, $i = 1, \dots, m$, are the components of $\mathbf{z}$, and defining $E \subset V \times V$ by the rule

$$(\mathbf{z}_i, \mathbf{z}_j) \notin E \iff \forall t_1, t_2 = 1, \dots, N,\; \boldsymbol{z}_i(t_1) \perp \boldsymbol{z}_j(t_2) \,| \, \{\mathbf{z}_k\}_{k\neq i,j}. \tag{8}$$

The aim of this section is to establish the counterpart of the last equivalence in (7) for latent-variable graphical models in the reciprocal processes setting. This will be crucial for the mathematical formulation of our identification problem. To this end let

$$\mathbf{z} := [\mathbf{y}^\top \, \mathbf{x}^\top]^\top \tag{9}$$

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 $\mathbf{\Sigma_z}$ of $\mathbf{z}$ and its inverse can

be partitioned as

$$\mathbf{\Sigma_z} = \left[\begin{array}{c|c} \mathbf{\Sigma_y} & \mathbf{\Sigma_{yx}} \\ \hline \mathbf{\Sigma_{yx}^\top} & \mathbf{\Sigma_x} \end{array}\right], \qquad \mathbf{\Sigma_z^{-1}} = \left[\begin{array}{c|c} \mathbf{S} & \mathbf{A} \\ \hline \mathbf{A}^\top & \mathbf{R} \end{array}\right] \tag{10}$$

where $\mathbf{\Sigma_y} \in \mathcal{C}$ and $\mathbf{\Sigma_x} \in \mathcal{C}_l$ 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 $l \times l$. By considering the Schur complement, we obtain the relation

$$\mathbf{\Sigma_y^{-1}} = \mathbf{S} - \mathbf{L} \tag{11}$$

where $\mathbf{S} > 0$ 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, $\mathbf{\Sigma_y^{-1}} \in \mathcal{B}$ we assume both $\mathbf{S} \in \mathcal{B}$ and $\mathbf{L} \in \mathcal{B}$, e.g.,

$$\mathbf{S} = \text{circ}\{S_0, S_1, \ldots, S_n, 0, \ldots, 0, S_n^\top, \ldots, S_1^\top\}. \tag{12}$$

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. 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 $\mathbf{y}_i := [\boldsymbol{y}_i(1) \ldots \boldsymbol{y}_i(N)]^\top$, $i = 1, \ldots, m$, be the ith component of the process $\mathbf{y}$ and let $\mathbf{x}_j := [\boldsymbol{x}_j(1) \ldots \boldsymbol{x}_j(N)]^\top$, $j = 1, \ldots, l$, be the jth 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 $k \in [1, N]$. Let $\Omega \subseteq \{(i,j) : i, j = 1, \ldots, m\}$ be the common support of the blocks $S_0, S_1, \ldots, S_n$ of $\mathbf{S}$, namely

$$\Omega^c := \{(i,j) : (S_k)_{ij} = (S_k)_{ji} = 0, \quad \forall k = 0, \ldots, n\}. \tag{13}$$

Notice that Ω contains all the pairs (i, i), $i = 1, \ldots, m$, because of the positivity of $\mathbf{S}$.

The following crucial result establishes the equivalence between the sparsity pattern of $\mathbf{S}$, intended as in (13), and the conditional independence relations among observed and latent variables. It may be viewed as the equivalent, in the latent-variable reciprocal setting, of the last equivalence in (7).

Proposition 1: Let E be the set of edges of the graphical model associated with the process $\mathbf{z}$ defined in (9). Then

$$(\mathbf{y}_i, \mathbf{y}_j) \notin E \iff (i,j) \in \Omega^c \tag{14}$$

with Ω^c defined in (13).

Proof: By the conditional independence characterization (7), $(i,j) \in \Omega^c$ if and only if

$$\forall t_1, t_2 \in [1,N], \boldsymbol{y}_i(t_1) \perp \boldsymbol{y}_j(t_2) \mid \{\mathbf{x}, \mathbf{y}_h, \boldsymbol{y}_i(s_1), \boldsymbol{y}_j(s_2) : h \neq i,j \\ s_1 \neq t_1, s_2 \neq t_2, \}. \tag{15}$$

We need to show that this is equivalent to

$$\forall t_1, t_2 \in [1,N], \ \boldsymbol{y}_i(t_1) \perp \boldsymbol{y}_j(t_2) \mid \{\mathbf{x}, \mathbf{y}_h : h \neq i,j\}. \tag{16}$$

The proof exploits basic results of the theory of Hilbert spaces of second-order random variables, see for instance [36, Ch. 2]. First of all,

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

be the error affecting the projection of $[\mathbf{y}_i^\top \ \mathbf{y}_j^\top]^\top$ onto the subspace generated by $\{\mathbf{y}_h(s),\, h \neq i,j,\, s = 1, \ldots, N, \mathbf{x}\}$, for any $t_1, t_2 \in [1, N]$ and for any $(i,j) \in \Omega^c$. It can be showed that ϵ is a zero-mean, Gaussian, random vector. Accordingly, proving (16) is equivalent to prove that

$$\mathbb{E}\left[\boldsymbol{\epsilon}_i\, \boldsymbol{\epsilon}_j^\top\right] = 0 \qquad \Longleftrightarrow \qquad \boldsymbol{\epsilon}_i(t_1) \perp \boldsymbol{\epsilon}_j(t_2) \tag{17}$$

for any $t_1, t_2 \in [1, N]$ and for any $(i,j) \in \Omega^c$, [36]. 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

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

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

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

where $\mathbf{\Sigma_{\bar{z}_1}}$ and $\mathbf{\Sigma_{\bar{z}_2}}$ 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

$$\mathbf{\Sigma_{\bar{z}}^{-1}} = \Pi\,\Sigma_{\mathbf{z}}^{-1}\,\Pi^\top = \left[\begin{array}{c|c} \bar{\mathbf{S}} & * \\ \hline * & * \end{array}\right]$$

where $\bar{\mathbf{S}} := (\mathbf{\Sigma_{\bar{z}_1}} - \mathbf{\Sigma_{\bar{z}_1\bar{z}_2}}\mathbf{\Sigma_{\bar{z}_2}^{-1}}\mathbf{\Sigma_{\bar{z}_2\bar{z}_1}})^{-1}$ is a permuted version of matrix $\mathbf{S}$, according to the permutation matrix Π. By construction, the Schur complement formula applied on $\mathbf{\Sigma_{\bar{z}}}$ gives

$$\Sigma_\epsilon = \mathbf{\Sigma_{\bar{z}_1}} - \mathbf{\Sigma_{\bar{z}_1\bar{z}_2}}\mathbf{\Sigma_{\bar{z}_2}^{-1}}\mathbf{\Sigma_{\bar{z}_2\bar{z}_1}} = \bar{\mathbf{S}}^{-1} \tag{18}$$

that relates the covariance matrix Σ_ϵ of the projection error ϵ to the covariance matrix $\mathbf{\Sigma_{\bar{z}}}$ of $\bar{\mathbf{z}}$. Condition (13) is equivalent to say that $\bar{\mathbf{S}}$, and therefore $\bar{\mathbf{S}}^{-1}$, is block-diagonal. Accordingly, by (18), $\mathbf{\Sigma_\epsilon}$ is block-diagonal, i.e., $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\epsilon}_j$ are independent, which is equivalent to (17) as we wanted to prove.

We refer to (11), together with the group-sparsity (13) and $\text{rank}(L) \ll m$, as *sparse plus low-rank decomposition* of $\mathbf{\Sigma_y^{-1}}$. The latter is the reciprocal-processes analogous of the inverse spectrum decomposition that holds for AR processes: a summary of the classical results on latent-variable graphical models describing AR Gaussian processes can be found in Appendix A.

Thanks to Proposition 1, we can conclude that $\mathbf{\Sigma_z^{-1}}$ in (10) together with (13) define an undirected graph for the Gaussian random vector $\mathbf{z}$ which admits a two-layer structure.

1) The nodes in the upper-layer represent the l variables of the latent-process $\mathbf{x}_1, \ldots, \mathbf{x}_l$ while the nodes in the bottom-layer represent the m variables of the observed process $\mathbf{y}_1, \ldots, \mathbf{y}_m$.

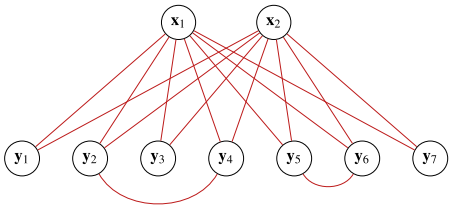

Fig. 1. Example of a latent-variable graphical model: $\mathbf{x}_1$, $\mathbf{x}_2$ are the latent-variables and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_7$ are the manifest variables.

2) The edges are given by the entries of the concentration matrix $\mathbf{\Sigma}_{\mathbf{z}}^{-1}$. In particular, the edge (i, j), between two vectors $\mathbf{y}_i$ and $\mathbf{y}_j$, $i \neq j$, is described by

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

Example 1: Consider the case in which $N = 2, m = 7, l = 2$, and suppose that the graphical model associated with the vector $\mathbf{z}$ is the one depicted in Fig. 1 . In this case, the concentration matrix of vector $\mathbf{z}$ has the structure (10) with

$$\mathbf{S} = \left[\begin{array}{c|c} S_0 & S_1^\top \\ \hline S_1 & S_0 \end{array}\right], \quad S_0,\ S_1 \in \mathbb{R}^{7\times 7}$$

and $\mathbf{R}$ 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 hold for the edge between $\mathbf{y}_5$ and $\mathbf{y}_6$. Thus, $\Omega = \{(i, i) : i = 1, \ldots, 7\} \cup \{(2, 4),\ (5, 6)\}$.

IV. Identification of Latent-Variable Reciprocal Graphical Models

In this section, we propose a procedure for the identification of latent-variable reciprocal graphical models from data generated by latent-variable AR graphical models. Therefore, the identified model must be understood as an approximation of the true generating model. The idea is to consider a regularized version of Problem (4). The regularization term, which is derived by leveraging on condition (11) and Proposition 1, induces the solution with a structure of a latent-variable reciprocal graphical model. Although the regularized problem provides a solution with the desired properties, there may exist several such models, each being a solution corresponding to a different cost function. The selection of a particular solution is, so far, mainly based on heuristic arguments. In Section IV-A, we show that maximum entropy and maximum likelihood principles are actually ingrained in the chosen selection rule, thus providing foundations for our choice.

Consider an m-dimensional AR Gaussian process y. Assume that T observations $\mathsf{y}(1), \ldots, \mathsf{y}(T)$ are available, and let

$$\hat{R}_k = \frac{1}{T}\sum_{t=k}^{T} \mathsf{y}(t)\mathsf{y}(t-k)^\top, \qquad k = 0, 1, \ldots, n. \tag{19}$$

We assume that

$$\hat{\mathbf{\Sigma}}_{\mathsf{y}} := \mathrm{circ}\{\hat{R}_0, \hat{R}_1, \ldots, \hat{R}_n, 0, \ldots, 0, \hat{R}_n^\top, \ldots, \hat{R}_1^\top\} \tag{20}$$

is positive definite. In the case that $\hat{\mathbf{\Sigma}}_{\mathsf{y}}$ is not positive definite, we can consider a positive definite banded block-circulant matrix sufficiently close to $\hat{\mathbf{\Sigma}}_{\mathsf{y}}$ obtained by solving a structured covariance estimation problem, see [37], [38].

Recalling that a latent-variable graphical model of a reciprocal process is characterized by (11), we can formulate the following identification problem.

Problem 1: Set $\Sigma_k := \hat{R}_k$, $k = 0, 1, \ldots, n$, with $\hat{R}_k$ given by (19). Compute the blocks $\Sigma_{n+1}, \ldots, \Sigma_{\frac{N}{2}}$ of the block-circulant covariance matrix $\mathbf{\Sigma}_{\mathbf{y}} = \mathrm{circ}\{\Sigma_0, \Sigma_1, \ldots, \Sigma_{\frac{N}{2}-1}, \Sigma_{\frac{N}{2}}, \Sigma_{\frac{N}{2}-1}^\top, \ldots, \Sigma_1^\top\}$ such that $\mathbf{\Sigma}_{\mathbf{y}}^{-1} = \mathbf{S} - \mathbf{L}$, where $\mathbf{S} > 0$ and $\mathbf{L} \geq 0$ belong to $\mathcal{B}$ with $S_0, \ldots, S_n$ having the smallest possible common support Ω, as in (13), 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 $\mathbf{\Sigma}_{\mathbf{y}}$ solving Problem 1 is the covariance of the reciprocal process $\mathbf{y}$ approximating the observed process y. The solution of Problem 1 actually requires the identification of a reciprocal process which could be achieved, in principle, by solving the maximum entropy dual problem (4) recalled in Section II-A. However, the support Ω is not known in advance, thus it has to be estimated from the data. In order to do that, we consider the following regularizer proposed in [39]:

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

In [39], the function $h_\infty(\cdot)$ was applied to the coefficients of the power spectral density, while here we apply the same function to the blocks of $\mathbf{S}$. The reason why this regularizer is expected to induce the group desired sparsity on $\mathbf{S}$ is the same of that discussed in [39]. The numerical simulations at the end of the paper confirm this fact. The trace (as a tractable proxy of the nuclear norm) is used instead for inducing low-rank of $\mathbf{L}$. Therefore, the paradigm for the estimation of the sparse plus low-rank decomposition of the concentration matrix $\mathbf{\Sigma}_{\mathbf{y}}^{-1}$ now directly follows from (4) by setting $\mathbf{X} = \mathbf{S} - \mathbf{L}$, with $\mathbf{L} \geq 0$, and by adding the regularizers just introduced:

$$\begin{aligned} \underset{\mathbf{S},\mathbf{L}\in\mathcal{B}}{\text{argmin}} \quad & -\log\det(\mathbf{S}-\mathbf{L}) + \langle\hat{\mathbf{\Sigma}}_{\mathsf{y}},\ \mathbf{S}-\mathbf{L}\rangle_{\mathcal{C}} \\ & + \lambda_S\, h_\infty(\mathbf{S}) + \lambda_L \mathrm{tr}(\mathbf{L}) \\ \text{subject to} \quad & \mathbf{S}-\mathbf{L} > 0, \quad \mathbf{L} \geq 0 \end{aligned} \tag{21}$$

where $\lambda_L,\ \lambda_S > 0$ are the two regularization parameters and $\hat{\mathbf{\Sigma}}_{\mathsf{y}}$ is given by (20). Notice that Problem (21) is precisely the optimization-program form of Problem 1 we want to solve. The next proposition ensures that Problem 1 does actually admit a solution. In what follows, we will express Problem (21) in terms of $\mathbf{X} := \mathbf{S} - \mathbf{L}$ and $\mathbf{L}$.

Proposition 2: The dual of Problem (21) is the constrained optimization problem

$$\begin{aligned} & \mathrm{argmin}_{\mathbf{Z}\in\mathcal{C}} \quad -\log\det(\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z}) - mN \\ & \text{subject to} \quad \hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z} > 0 \end{aligned} \tag{22a}$$

$$\lambda_L I_{mN} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}) \geq 0 \tag{22b}$$

$$\operatorname{diag}(Z_j) = 0, \quad j = 0, \ldots, n \tag{22c}$$

$$2|(Z_0)_{kh}| + \sum_{j=1}^{n} |(Z_j)_{kh}| + |(Z_j)_{hk}| \leq \frac{\lambda_S}{N}, \quad k > h. \tag{22d}$$

Moreover, if $\hat{\mathbf{\Sigma}}_{\mathsf{y}} \in \mathcal{B}$ and $\hat{\mathbf{\Sigma}}_{\mathsf{y}} > 0$ hold, Problem (21) admits a solution $(\mathbf{S}_o, \mathbf{L}_o)$ where $\mathbf{X}_o = \mathbf{S}_o - \mathbf{L}_o$ is unique.

Proof: The proof leverages on strong duality between (21) and its dual. We first prove that the latter has a unique solution exploiting Lagrange-multipliers theory and then conclude by establishing the existence of a solution for (21).

More formally, note that Problem (21) is a strictly feasible convex optimization problem (for instance, pick $\mathbf{S} = I_{mN}$ and $\mathbf{L} = 0$) therefore Slater's condition, hence strong duality between (21) and its dual, holds. We start the derivation of the dual problem by adding an auxiliary variable $\mathbf{Y}$ to Problem (21)

$$\begin{aligned} \underset{\substack{\mathbf{X}\in\mathcal{C}\\ \mathbf{Y},\mathbf{L}\in\mathcal{B}}}{\operatorname{argmin}} \quad & -\log\det(\mathbf{X}) + \operatorname{tr}(\hat{\mathbf{\Sigma}}_{\mathsf{y}}\,\mathbf{X}) + \lambda_S\, h_\infty(\mathbf{Y}) + \lambda_L \operatorname{tr}(\mathbf{L}) \\ \text{subject to} \quad & \mathbf{X} > 0, \quad \mathbf{L} \geq 0 \\ & \mathbf{Y} = \mathbf{X} + \mathbf{L}. \end{aligned} \tag{23}$$

The Lagrangian function for this problem is

$$\begin{aligned} \mathcal{L}(\mathbf{X},\mathbf{Y},\mathbf{L},\mathbf{V},\mathbf{Z}) = & -\log\det(\mathbf{X}) + \langle \hat{\mathbf{\Sigma}}_{\mathsf{y}},\, \mathbf{X}\rangle_{\mathcal{C}} + \lambda_S\, h_\infty(\mathbf{Y}) \\ & + \lambda_L \operatorname{tr}(\mathbf{L}) - \langle \mathbf{V},\, \mathbf{L}\rangle_{\mathcal{C}} + \langle \mathbf{Z}, \mathbf{X} + \mathbf{L} - \mathbf{Y}\rangle_{\mathcal{C}} \end{aligned} \tag{24}$$

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

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

The dual objective function is the infimum over $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{L}$ of the Lagrangian. The unique term in $\mathcal{L}$ 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 (22c) and (22d) hold, in which case the infimum is zero. Accordingly

$$\inf_{\mathbf{Y}} \mathcal{L} = \begin{cases} -\log\det(\mathbf{X}) + \langle \hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z},\, \mathbf{X}\rangle_{\mathcal{C}} + \langle \lambda_L I_{mN} - \mathbf{V} + \mathbf{Z},\, \mathbf{L}\rangle_{\mathcal{C}} \\ \text{if} \quad (22c),\ (22d) \text{ hold} \\ \\ -\infty \quad \text{otherwise.} \end{cases}$$

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 $\mathsf{P}_{\mathscr{B}}$, we have that

$$\langle \lambda_L I_{mN} - \mathbf{V} + \mathbf{Z},\, \mathbf{L}\rangle_{\mathcal{C}} = \langle \lambda_L I_{mN} - \mathbf{V} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}),\, \mathbf{L}\rangle_{\mathcal{C}} \tag{25}$$

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

$$\lambda_L I_{mN} - \mathbf{V} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}) = 0. \tag{26}$$

In this case, the minimum of (25) is zero. Accordingly

$$\inf_{\mathbf{Y},\mathbf{L}} \mathcal{L} = \begin{cases} -\log\det(\mathbf{X}) + \langle \hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z},\, \mathbf{X}\rangle_{\mathcal{C}} \\ \text{if} \quad (22c),\ (22d),\ (26) \text{ hold} \\ \\ -\infty \quad \text{otherwise.} \end{cases}$$

If $(22c)$, $(22d)$, (26) hold, it remains to minimize the strictly convex function

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

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

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

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

$$\delta\bar{\mathcal{L}}(\mathbf{X}_o; \delta\mathbf{X}) = \operatorname{tr}\left[\left(-\mathbf{X}_o^{-1} + \hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z}\right)\delta\mathbf{X}\right] = 0 \qquad \forall\, \delta\mathbf{X} \in \mathcal{C}. \tag{27}$$

Notice that $\bar{\mathcal{L}}$ is bounded below if and only if (22a) holds, therefore condition (27) is satisfied if and only if $\mathbf{X}_o = (\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z})^{-1}$. Hence

$$\inf_{\mathbf{Y},\mathbf{L},\mathbf{X}} \mathcal{L} = \begin{cases} \log\det(\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z}) + mN \\ \text{if } (22c),\ (22d),\ (26),\ (22a) \text{ hold} \\ \\ -\infty \qquad \text{otherwise.} \end{cases}$$

Therefore, the dual problem of (21) is

$$\begin{aligned} \operatorname{argmin}_{\mathbf{V}\in\mathcal{B},\, \mathbf{Z}\in\mathcal{C}} \quad & -\log\det(\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z}) - mN \\ \text{subject to} \quad & V \geq 0,\ (22c),\ (22d),\ (26),\ (22a). \end{aligned} \tag{28}$$

Notice that we can remove the variable $\mathbf{V}$. Indeed, recalling that $\mathbf{V} \geq 0$, the constraint (26) becomes $\lambda_L I_{mN} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}) = \mathbf{V} \geq 0$, and the dual problem takes precisely the form (22) as we wanted to show.

We now prove that Problem (22) admits indeed a unique solution provided that $\hat{\mathbf{\Sigma}}_{\mathsf{y}} \in \mathcal{B}$ and $\hat{\mathbf{\Sigma}}_{\mathsf{y}} > 0$. To this end, define $f(\mathbf{Z}) := \log\det(\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z})$ and let

$$\begin{aligned} \mathcal{Q} := \{ & \mathbf{Z} \in \mathcal{C} \,|\, (22c),\ (22d),\ (22a) \text{ and } \lambda_L I_{mN} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}) \\ & \geq 0 \text{ hold}\} \end{aligned}$$

be the set of constraints of Problem (22). First of all, notice that constraints (22c) and (22d) 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 $\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \bar{\mathbf{Z}} \geq 0$ is singular. Then

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

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

$$\bar{\Omega} := \{\mathbf{Z} \in \mathcal{C} \,|\, \hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z} \geq \epsilon I_{mN},\ (22c),\ (22d) \text{ and } \lambda_L I_{mN} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}) \geq 0 \text{ hold}\}$$

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

The strong duality between problems (21) and (22) and the existence of a unique optimum $\mathbf{Z}_o$ for the dual problem (22), imply that there exists a unique $\mathbf{X}_o \in \mathcal{B}$ so that $\mathbf{X}_o = (\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z}_o)^{-1}$ which solves the primal problem (21). It remains to show that there exists an $\mathbf{L}_o \in \mathcal{B}$ that solves the optimization problem

$$\begin{aligned} \underset{\mathbf{L}\in\mathcal{B}}{\text{argmin}} \quad & \lambda_S\, h_\infty(\mathbf{X}_o + \mathbf{L}) + \lambda_L \mathrm{tr}(\mathbf{L}) \\ \text{subject to} \quad & \mathbf{L} \geq 0. \end{aligned} \tag{29}$$

The objective function in (29) is continuous and since $\mathbf{L} = 0$ is a feasible point, the problem is equivalent to find $\mathbf{L} \in \mathcal{B}$ that minimizes $\lambda_S\, h_\infty(\mathbf{X}_o + \mathbf{L}) + \lambda_L \mathrm{tr}(\mathbf{L})$ over the set

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

It is easy to see that $\mathcal{K}$ is a closed and bounded (thus compact) subset of $\mathcal{B}$. Hence, by Weierstrass' Theorem, Problem (29) admits a solution $\mathbf{L}_o$. At this point, we can conclude that the primal problem (21) admits a solution $(\mathbf{S}_o, \mathbf{L}_o)$.

As a concluding remark, observe that the regularized problem used to identify a latent-variable AR graphical model relies on the sparse plus low-rank decomposition of the inverse of the observed process' spectrum (see Appendix A)

$$\begin{aligned} \underset{\Gamma,\,\Lambda\in\mathcal{Q}_{m,n}}{\text{argmin}} \quad & \int -\log\det(\Gamma - \Lambda) + \langle \Gamma - \Lambda,\, \hat{\Phi}_{\mathsf{y}} \rangle \\ & + \gamma_S\, \phi_1(\Gamma) + \gamma_L\, \phi_*(\Lambda) \\ \text{subject to} \quad & \Gamma - \Lambda > 0 \\ & \Lambda \geq 0. \end{aligned} \tag{30}$$

Here, the domain of the optimization is the family of matrix pseudo-polynomials $\mathcal{Q}_{m,n} := \{\sum_{k=-n}^{n} Q_k\, e^{i\theta k},\ \ Q_{-k} = Q_k^\top \in \mathbb{R}^{m\times m}\}$ while $\gamma_S,\, \gamma_L > 0$ are the regularization parameters that balance the effects of the two regularizers ϕ_1 and ϕ_* inducing sparsity and low-rank of Γ and Λ, respectively. Being

$$\hat{\Phi}(e^{i\theta}) = \sum_{k=-n}^{n} \hat{R}_k\, e^{-i\theta k}, \qquad \hat{R}_{-k} = \hat{R}_k^\top \tag{31}$$

the truncated periodogram of the observed process y, we conclude that Problem (21) is the reciprocal counterpart of Problem (30) considered in [8].

A. Foundations of Problem (21)

Next, we show how Problem (21) can be viewed both as a regularized maximum-likelihood problem and as a relaxed version of a maximum entropy paradigm.

1) Maximum Likelihood: In the following, we show that the fitting function in (21), i.e.,

$$-\log\det(\mathbf{S} - \mathbf{L}) + \mathrm{tr}\left(\hat{\mathbf{\Sigma}}_{\mathsf{y}}\,(\mathbf{S} - \mathbf{L})\right) \tag{32}$$

is the approximation, in the sense explained in Section II-A, of the (conditional) negative log-likelihood of the AR process (5). Following [3], consider the observed AR process y whose spectrum is denoted by Φ_{y}, and suppose that T observations $\mathsf{y}(1), \ldots, \mathsf{y}(T)$ of the process are available. The conditional likelihood of the process y is defined as the likelihood function associated with the conditional distribution of $\mathsf{y}(n+1), \mathsf{y}(n+2), \ldots, \mathsf{y}(n+T)$ given $\mathsf{y}(1), \ldots, \mathsf{y}(n)$. Let $\mathbf{T}_n := \mathrm{Toepl}\{\hat{R}_0,\, \hat{R}_1, \ldots,\, \hat{R}_n\}$ be the block-Toeplitz matrix having in the first block-row the estimates of the first $n+1$ covariance lags of the process $\hat{R}_0, \hat{R}_1, \ldots, \hat{R}_n$ computed as in (19). For T large enough, the conditional negative log-likelihood function of the AR process can be well approximated by

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

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

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

moreover, given the truncated periodogram $\hat{\Phi}_{\mathsf{y}}(e^{i\theta})$ defined in (31), it is easy to see that

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

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

$$\ell(B) = \frac{T-n}{2} \int \log\det \Phi_{\mathsf{y}}(e^{i\theta}) + \mathrm{tr}\left[\hat{\Phi}_{\mathsf{y}}(e^{i\theta})\, \Phi_{\mathsf{y}}(e^{i\theta})^{-1}\right]. \tag{33}$$

A natural way to approximate (33) is to approximate the integral with a finite sum, i.e., to discretize the interval $[-\pi, \pi]$. This is the frequency interpretation of the reciprocal approximation explained in Section II-A that consists in sampling the spectrum of the process to obtain the corresponding symbol. In fact, considering as sample frequency $\Delta\theta = 2\pi/N$, the backward Euler approximation leads to the discrete approximation

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

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

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

Observe now that $\Phi_{\mathsf{y}}(\zeta)$ is precisely the symbol of the block-circulant covariance matrix $\mathbf{\Sigma}_{\mathbf{y}}$ of the reciprocal process $\mathbf{y}$ approximating the process y and $\hat{\Phi}_{\mathsf{y}}(\zeta)$ is the symbol of the block-circulant matrix $\hat{\mathbf{\Sigma}}_{\mathsf{y}}$ in Problem (21). Accordingly, from (2), it follows that

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

Since $\mathbf{\Sigma}_{\mathbf{y}}^{-1} = \mathbf{S} - \mathbf{L}$, this is precisely (up to a scaling factor) equal to (32).

2) Maximum Entropy: We now show that Problem (21) can be interpreted as a relaxed version of a maximum entropy problem for latent-variable reciprocal graphical models. Consider the regularized solution $(\mathbf{S}_o, \mathbf{L}_o)$ of (21) and let Ω be the support of $\mathbf{S}_o$, i.e., $\mathbf{S}_o$ satisfies (13). Since $\mathbf{L}_o \in \mathcal{B}$ is so that $\mathbf{L}_o \geq 0$ and $\operatorname{rank} \mathbf{L}_o = lN \ll mN$, there exists $\mathbf{G} = \operatorname{circ}\{G_0, G_1, \ldots, G_n, 0, \ldots, 0\}$ 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 (21) where the regularizers are replaced by the corresponding hard-constraints $\mathbf{S} \in \mathcal{V}_\Omega$ and $\mathbf{L} \in \mathcal{V}_G$, where $\mathcal{V}_\Omega := \{\mathbf{S} \in \mathcal{C} : \mathsf{P}_{\Omega^c}(\mathbf{S}) = 0\}$ 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 $\mathcal{V}_G \subseteq \mathcal{B}$. The resulting problem is

$$\begin{aligned} \underset{\mathbf{S},\mathbf{L}\in\mathcal{B}}{\operatorname{argmin}} \quad & -\log\det(\mathbf{S}-\mathbf{L}) + \langle \hat{\mathbf{\Sigma}}_{\mathsf{y}},\, \mathbf{S}-\mathbf{L}\rangle_{\mathcal{C}} \\ \text{subject to} \quad & \mathbf{S}-\mathbf{L} > 0,\ \mathbf{L} \geq 0 \\ & \mathbf{S} \in \mathcal{V}_\Omega,\ \mathbf{L} \in \mathcal{V}_G. \end{aligned} \tag{34}$$

The next proposition proves that the primal formulation of the sharp-constrained version of Problem (21) stated in (34) is indeed a maximimum entropy problem for latent variable reciprocal graphical models.

Proposition 3: The primal of Problem (34) is

$$\begin{aligned} \underset{\mathbf{\Sigma}_{\mathbf{y}}\in\mathcal{C}}{\operatorname{argmax}} \quad & \log\det \mathbf{\Sigma}_{\mathbf{y}} \\ \text{subject to} \quad & \mathsf{P}_\Omega \mathsf{P}_{\mathcal{B}}(\mathbf{\Sigma}_{\mathbf{y}} - \hat{\mathbf{\Sigma}}_{\mathsf{y}}) = 0 \\ & \mathbf{E}^*\, \mathbf{G}(\mathbf{\Sigma}_{\mathbf{y}} - \hat{\mathbf{\Sigma}}_{\mathsf{y}})\mathbf{G}^\top \mathbf{E} \geq 0 \end{aligned} \tag{35}$$

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

Proof: We derive the dual of Problem (35). Observing that $\mathbf{E} = \mathbf{F}^*\mathbf{1}$ where $\mathbf{1} := \frac{1}{\sqrt{N}}[I_l \quad I_l \quad \cdots \quad I_l]^\top$, the Lagrangian of Problem (35) writes as

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

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

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

where we have exploited the fact that $\mathbf{F}(\mathbf{\Sigma}_{\mathbf{y}} - \hat{\mathbf{\Sigma}}_{\mathsf{y}})\mathbf{F}^*$ 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

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

where $\mathbf{S} := \mathsf{P}_{\Omega\cup\mathcal{B}}(\mathbf{W})$ belongs to $\mathcal{V}_\Omega$ 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 (34).

Similar arguments as the one used to prove formula (27), allow us to assert that a necessary and sufficient condition for $\mathbf{\Sigma}_o$ to be a minimum point for $\mathcal{L}$ is that its first Gateaux derivative computed at $\mathbf{\Sigma}_{\mathbf{y}} = \mathbf{\Sigma}_o$ is equal to zero in every direction $\delta\mathbf{\Sigma}$, namely

$$\delta\mathcal{L}(\mathbf{\Sigma}_o; \delta\mathbf{\Sigma}) = \operatorname{tr}\left[\left(\mathbf{\Sigma}_o^{-1} - \mathbf{S} + \mathbf{L} \right) \delta\mathbf{\Sigma} \right] = 0 \qquad \forall\, \delta\mathbf{\Sigma} \in \mathcal{C}.$$

By assumption $\mathbf{S} - \mathbf{L} > 0$ thus the substitution of the optimum $\mathbf{\Sigma}_o = (\mathbf{S}-\mathbf{L})^{-1}$ in the Lagrangian $\mathcal{L}$ leads precisely to the objective function in (34).

Some observations on the two constraints of (35) are in order. The first constraint $\mathsf{P}_\Omega \mathsf{P}_{\mathcal{B}}(\mathbf{\Sigma}_{\mathbf{y}} - \hat{\mathbf{\Sigma}}_{\mathsf{y}}) = 0$ fixes the entries corresponding to the indexes in Ω of the first $n+1$ lags of the reciprocal process. For the second constraint, let $\Psi(\zeta)$ and $\Phi_{\mathbf{y}}(\zeta)$ be the symbols of $\mathbf{G}$ and $\mathbf{\Sigma}_{\mathbf{y}}$, respectively. By (2), we have that

$$\mathbf{E}^*\, \mathbf{G}\, \mathbf{\Sigma}_{\mathbf{y}}\, \mathbf{G}^\top \mathbf{E} = \frac{1}{N} \sum_{k=0}^{N-1} \Psi(\zeta^k)\, \Phi_{\mathbf{y}}(\zeta^k)\, \Psi(\zeta^k)^* \tag{36}$$

which is the covariance of the output of the $l \times m$ filter $\Psi(\zeta) = \sum_{k=0}^{n} G_k\, \zeta^{-k}$ fed with the reciprocal process $\mathbf{y}$. Accordingly, the second constraint in (35) states that the covariance matrix of the process at the output of the filter is lower bounded by $\mathbf{E}^*\, \mathbf{G}\, \hat{\mathbf{\Sigma}}_{\mathbf{y}}\, \mathbf{G}^\top \mathbf{E}$. We conclude that Problem (35) can be seen as the reciprocal counterpart of the maximum entropy problem considered in [8]

$$\begin{aligned} \operatorname{argmax}_{\Phi_{\mathbf{y}}\in\mathcal{S}_m} \quad & \int \log\det \Phi_{\mathbf{y}} \\ \text{subject to} \quad & \left(\int e^{i\theta k}\, \Phi_{\mathbf{y}} - \hat{R}_k \right)_{pq} = 0, \quad \substack{k=0,1,\ldots,n \\ (p,q)\in\Omega} \\ & \int \Psi\, (\Phi_{\mathbf{y}} - \hat{\Phi}_{\mathbf{y}})\, \Psi^* \geq 0 \end{aligned} \tag{37}$$

where $\Psi(e^{i\theta}) = \sum_{k=0}^{n} G_k\, e^{-i\theta k}$ and $\mathcal{S}_m = \{F \in \mathcal{H}_p : F - \alpha\, I_p \geq 0$ a.e. on $\mathbb{T}$, for some $\alpha > 0\}$. Indeed, the second constraint in (37) can be approximated with the backward Euler approximation with sample frequency $\Delta\theta = 2\pi/N$ obtaining (36).

V. Alternating Direction Method of Multipliers

In this section, we address the numerical solution of Problem (22) through an ADMM algorithm [40]. We outline the derivation of the ADMM updates for the solution of (22) referring to Appendix B for further details.

The joint enforcement of the constraints (22c), (22d), and $\lambda_L\, I_{mN} + \mathsf{P}_{\mathcal{B}}(\mathbf{Z}) \geq 0$ in solving Problem (22), might be a difficult requirement. In the following, we show how such constraints can be decoupled and enforced in an alternating way. First of all observe that, by defining the variable $\mathbf{P} := \lambda_L I_{mN} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z})$, Problem (22) rewrites as

$$\begin{aligned} \underset{\mathbf{Z},\mathbf{P}\in\mathcal{C}}{\text{argmin}} \quad & -\log\det(\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z}) - mN \\ \text{subject to} \quad & (22c),\ (22d) \\ & \mathbf{P} = \lambda_L I_{mN} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}) \\ & \mathbf{P} \geq 0 \end{aligned} \tag{38}$$

where the domain $\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z} > 0$ of the objective function has been omitted, since it is checked in the stepsize-choice stage of the algorithm. Notice that Problem (38) is convex by construction, being the (minimization form of the) dual of Problem (21). The augmented Lagrangian for the problem is

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

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

1) The $\mathbf{Z}$-minimization step

$$\begin{aligned} \mathbf{Z}^{k+1} = \underset{\mathbf{Z}\in\mathcal{C}}{\text{argmin}} \quad & \mathcal{L}_\rho(\mathbf{Z},\mathbf{P}^k,\mathbf{M}^k) \\ \text{subject to} \quad & \mathbf{Z} \in \mathcal{Z}. \end{aligned} \tag{39}$$

2) The $\mathbf{P}$-minimization step

$$\begin{aligned} \mathbf{P}^{k+1} = \underset{\mathbf{P}\in\mathcal{C}}{\text{argmin}} \quad & \mathcal{L}_\rho(\mathbf{Z}^{k+1},\mathbf{P},\mathbf{M}^k) \\ \text{subject to} \quad & \mathbf{P} \geq 0. \end{aligned} \tag{40}$$

3) Dual variable update

$$\mathbf{M}^{k+1} = \mathbf{M}^k - \rho\left(\mathbf{P}^{k+1} - \lambda_L I_{mN} - \mathsf{P}_{\mathscr{B}}(\mathbf{Z}^{k+1})\right) \tag{41}$$

where $\mathcal{Z} := \{\mathbf{Z} \in \mathcal{C} :\ (22c),\ (22d)\}$ and we have considered a constant value of ρ in order simplify the notation. Updates 1) and 2) are not in an implementable format. The $\mathbf{Z}$-update step (39) is indeed equivalent to the minimization of

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

over the set $\mathcal{Z}$, which has no closed-form solution as noticed in [39]. We approximate it with a projective-gradient step: additional details on the gradient-iteration step can be found in Appendix B. As for the $\mathbf{P}$-minimization step, we notice that (40) is equivalent to minimize the functional

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

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

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

which, however, is not going to be positive semidefinite in general. Accordingly, in order to find an appropriate solution, we search for 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 $\mathcal{C}$). The details of the derivation of the $\mathbf{P}$-update step are reported in Appendix B.

We conclude that the ADMM for Problem (22) consists in the following updates:

$$\begin{aligned} \mathbf{Z}^{k+1} &= \mathsf{P}_{\mathcal{Z}}\left[\mathbf{Z}^k - t_k\, \nabla\mathcal{J}(\mathbf{Z}^k)\right] \\ \mathbf{P}^{k+1} &= \mathsf{P}_{\mathcal{C}^+}\left[\frac{1}{\rho^k}\, \mathbf{M}^k + \lambda_L I_{mN} + \mathsf{P}_{\mathscr{B}}(\mathbf{Z}^{k+1})\right] \\ \mathbf{M}^{k+1} &= \mathbf{M}^k - \rho^k\left[\mathbf{P}^{k+1} - \lambda_L I_{mN} - \mathsf{P}_{\mathscr{B}}(\mathbf{Z}^{k+1})\right] \end{aligned} \tag{42}$$

where $\nabla\mathcal{J}(\mathbf{Z}^k)$ 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 $\mathsf{P}_{\mathcal{Z}}$ is the projection operator onto the constraints space $\mathcal{Z}$. 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.

Remark 2: Notice that the matrices involved in (42) are all symmetric and block-circulant. Accordingly, as mentioned in Section I, the reciprocal approximation allows to obtain a robust identification procedure even in the case when n is large. Indeed, as detailed, e.g., in [34], we can compute inverse matrices and eigenvalues in a robust way. Moreover, in view of (2), 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 [40], 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

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

while the dual residual turns out to be

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

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

$$\|r^k\|_{\mathfrak{C}} \leq \epsilon^{\mathrm{p}} \qquad \text{and} \qquad \|s^k\|_{\mathfrak{C}} \leq \epsilon^{\mathrm{d}}$$

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

$$\epsilon^{\mathrm{p}} := mN\,\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}} \max\left\{\lambda_L\sqrt{mN},\ \|\mathbf{Z}^k\|_{\mathfrak{C}},\ \|\mathbf{P}^k\|_{\mathfrak{C}}\right\}$$

$$\epsilon^{\mathrm{d}} := mN\,\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}}\,\|\mathbf{M}^k\|_{\mathfrak{C}}.$$

The intuition behind such definitions goes as follows. The ϵ^{abs} part of the tolerances is accounting for the dimension of the problem while the ϵ^{rel} part wants to take into consideration possible scalings of the constraints. One can employ different norms in the tolerances by slightly adjusting the definitions of ϵ^{p} and ϵ^{d} [41]. Here, ϵ^{abs} and ϵ^{rel} are predefined absolute and relative tolerances for the problem. Accordingly, the algorithm converges if all the conditions

$$\|r^k\|_{\mathfrak{C}} \leq \epsilon^{\mathrm{p}}, \qquad \|s^k\|_{\mathfrak{C}} \leq \epsilon^{\mathrm{d}}, \qquad \rho^k = \rho_{\max} \tag{43}$$

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

VI. Numerical Examples

In this section, we compare the performances of our method to which we will refer to as *approximated algorithm* with the method proposed in [8] and [9] for the solution of Problem (30), which will be referred to as *exact algorithm*. In particular, we analyze how the two algorithms behave considering both the case in which the observed process has low dimension and the case in which we have a high dimensional observed process. We remark that [8] and [9] consider the same identification problem and propose the same solution. The difference is that [9] presents an ADMM technique to address the numerical implementation (while the simulations in [8] are implemented in CVX). Therefore, for the comparison to be fair we compare our method with the ADMM algorithm and refer just to [9] when speaking about the exact algorithm.

1) Low-Dimensional Case: Synthetic data are generated from the AR latent-variable model of order $n = 8$

$$\mathsf{y}(t) = \sum_{k=1}^{n} A_k\,\mathsf{y}(t-k) + \boldsymbol{\eta}(t) \tag{44}$$

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 $\hat{R}_k$, $k = 0, \ldots, n$. Fig. 2 (center) reports the sparsity pattern of the underlying model, randomly generated so that the nonzero elements represents the 5% of the total elements.

For the approximated algorithm we have considered $N = 30$ samples of the spectrum. In both the ADMM implementations,

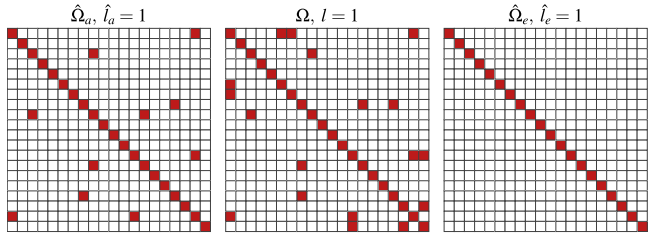

Fig. 2. Sparsity pattern estimated by the approximated algorithm with $\alpha = 1.007$, $\lambda_S = 95$, $\lambda_L = 5.4$ (left), true sparsity pattern (center), sparsity pattern estimated by the exact algorithm with $\alpha = 1.002$, $\gamma_S = 2.6$, $\gamma_L = 2.95$ (right). The red squares indicate 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.

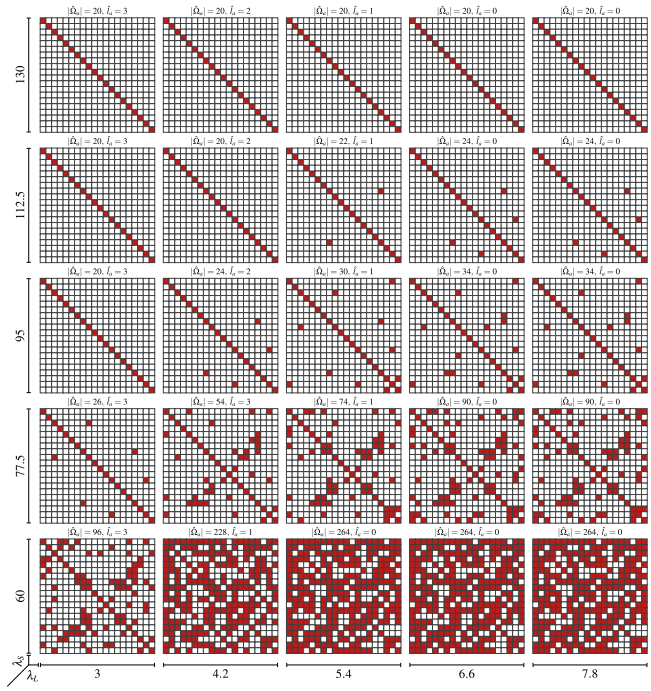

Fig. 3. Supports and ranks estimated by the approximated algorithm for $\lambda_S \in [60\,130]$ and $\lambda_L \in [3, 7.8]$. The growth coefficient is set to $\alpha = 1.007$.

we have set $\epsilon^{\mathrm{abs}} = 10^{-5}$, $\epsilon^{\mathrm{rel}} = 10^{-4}$ and $\rho_{\max} = 10^4$. 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. Fig. 3 shows the supports and the ranks estimated by the approximated algorithm corresponding to the different values of λ_S and λ_L.

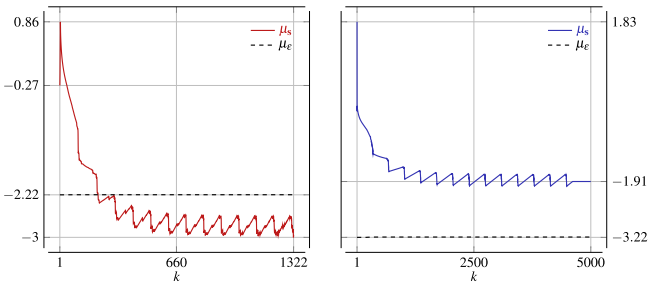

Fig. 4. Logarithm of the average dual residual for the approximated method (left) and for the exact method (right). The dashed lines correspond to the logarithm of the associated average feasibility tolerances.

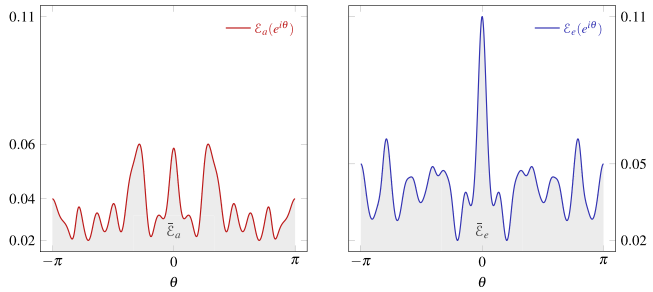

Fig. 5. Relative errors in the estimated spectra—approximated algorithm (left), exact algorithm (right).

For both methods the value of α that gives the better performances, i.e., that guarantees the minimum gap between ϵ^{p} / ϵ^{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 $\epsilon(h)$ denote the vectors containing the dual residual and its feasibility tolerance for the model $h = 1, \ldots, 25$ respectively. Fig. 4 displays the (logarithm of the) averages

$$\mu_{\mathbf{s}} = \frac{1}{25}\sum_{h=1}^{25} \mathbf{s}(h), \qquad \mu_{\epsilon} = \frac{1}{25}\sum_{h=1}^{25} \epsilon(h)$$

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 (43) therefore there is no need to display it. We observe that the exact algorithm does not converge for any value of α we have considered. Indeed, the plot in Fig. 4 (right) clearly shows that the mean dual-residual $\mu_{\mathbf{s}}$ 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.

Fig. 2 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 spectrum Φ_{y} (of the observed process) obtained by the solutions of problems (30) and (21), respectively. The squared-estimation errors for the two algorithms are depicted in Fig. 5; the corresponding mean values over $[-\pi, \pi]$ are

$$\mathcal{E}_a := \frac{\|\Phi_{\mathsf{y}} - \hat{\Phi}_a\|_F^2}{\|\Phi_{\mathsf{y}}\|_F^2} \qquad \bar{\mathcal{E}}_a := \int \mathcal{E}_a(e^{i\theta}) = 0.0358$$

$$\mathcal{E}_e := \frac{\|\Phi_{\mathsf{y}} - \hat{\Phi}_e\|_F^2}{\|\Phi_{\mathsf{y}}\|_F^2} \qquad \bar{\mathcal{E}}_e := \int \mathcal{E}_e(e^{i\theta}) = 0.0443.$$

TABLE I
SUMMARY OF THE PERFORMANCES OF THE TWO ALGORITHMS FOR $\lambda_S = 100,\ 146.25,\ 350$ AND $\gamma_S = 0.7,\ 0.826,\ 1.7$

$\hat{l}_a$	$\lvert\hat{\Omega}_a\rvert$	$\lvert\Omega - \hat{\Omega}_a\rvert$	$\mathcal{E}_a$	time [s]	$\hat{l}_e$	$\lvert\hat{\Omega}_e\rvert$	$\lvert\Omega - \hat{\Omega}_e\rvert$	$\mathcal{E}_e$	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

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 1.4 GHz MacBook Air.

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.

Similar results have been obtained by comparing the two algorithms in the case $l = 2$.

1) High-Dimensional Case: We consider now an AR latent-variable model as in (44) where we have $m = 80$ observed variables and $l = 1$ latent variable, $n = 24$ and the variance of the noise is $\mathbb{E}[\boldsymbol{\eta}(t)^\top \boldsymbol{\eta}(t)] = 87.1$. The number of samples used to estimate the covariance lags R_k is $T = 15\,000$. The number of conditionally dependent pairs in the true model is 158 so that the cardinality of the true support is $|\Omega| = 396$. Table I compares the performances of our approximated algorithm with the exact algorithm proposed in [9] for different values of the sparsity regularization parameters λ_S and γ_S. These parameters have been chosen for both algorithms in order to sweep the various possible degrees of sparsity (from almost totally sparse to almost full) while keeping the same numbers of estimated latent variables. The notation $|\Omega - \hat{\Omega}|$ indicates the error on the sparsity pattern in terms of number of misclassified entries.

The numerical evidence empirically highlights the improvement in scalability obtained from the introduced approximation. From a more theoretical point of view, a comparison with the ADMM implementation of the algorithm in [9] goes as follows: the analogous gradient computation for the $\mathbf{Z}$-update step (which is the computational bottleneck of the algorithm) has complexity $O(m^3 n^3)$. The circulant approximation allows instead for a computational complexity of $O(mN \log(mN))$, which is a drastic improvement especially if we account for the fact that the period N of the reciprocal process is typically much smaller than the order n of the corresponding AR process.

Both algorithms estimate the correct number of latent variables. While the approximated algorithm gives a result very close to the true one (highlighted in red in Table I), the exact algorithm makes a pretty high error in the reconstruction of the sparsity pattern, even if the cardinality of the true support has been correctly estimated. This is due to the fact that the higher the order of the process n is, the lower the accuracy in the computation of eigenvalues and inverse matrices by the exact algorithm. Table I 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 IV. This kind of scenario agrees with what we have discussed in Section IV: high-order AR process are quite challenging instances for the exact procedure proposed in [9]; in this case, the reciprocal approximation leads to remarkable benefits in the performances of the identification procedure.

VII. Conclusion

In this article, an identification paradigm for latent-variable graphical models associated with reciprocal processes has been presented. It has been showed 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 [9] where no approximation is introduced. The numerical examples have showed that for high-order AR processes reciprocal approximation gives substantial improvements in terms of robustness and scalability of the identification procedure.

Appendix A

A. Graphical Models for Gaussian Processes

This Appendix is devoted to recall the basic concepts related to graphical models and latent-variable graphical models associated with Gaussian processes, [8]. Let $\mathsf{z} := \{\mathsf{z}(t),\, t \in \mathbb{Z}\}$, assumed to be of the form $\mathsf{z} = [\mathsf{y}^\top\, \mathsf{x}^\top]^\top$ where y is the $\mathbb{R}^m$-valued process containing the observed variables while x is the $\mathbb{R}^l$-valued process containing the l latent variables. We assume z to be a purely nondeterministic, Gaussian, stationary process. Let Φ_{y} denote the spectral density of y. Provided that $l \ll m$, Φ_{y}^{-1} admits the decomposition

$$\Phi_{\mathsf{y}}(e^{i\theta})^{-1} = \Gamma(e^{i\theta}) - \Lambda(e^{i\theta}) \tag{45}$$

where $\Lambda \geq 0$ is low-rank with rank equal to the number l of latent-variables and $\Gamma > 0$, whose support reflects the conditional dependencies among the observed variables. More precisely, let $V = \{1 \dots m\}$. For $I \subset V$ we define $\mathcal{V}_I := \operatorname{span}\{\mathsf{x}(t), \mathsf{y}_j(t) : j \in I,\, t \in \mathbb{Z}\}$ as the closure of the set containing all the finite linear combinations of the variables in $\mathsf{x}(t)$ and $\mathsf{y}_j(t)$, $j \in I$. Thus, for any $i \neq j$, the notation

$$\mathcal{V}_{\{i\}} \perp \mathcal{V}_{\{j\}} \,|\, \mathcal{V}_{V\setminus\{i,j\}}$$

means that $\mathsf{y}_i(t_1)$ and $\mathsf{y}_j(t_2)$ are conditionally independent given the space linearly generated by $\{\mathsf{x}(t), \mathsf{y}_k(t) : k \in V \setminus \{i, j\}, t \in \mathbb{Z}\}$, for all t_1, t_2. Then, we have

$$\mathcal{V}_{\{i\}} \perp \mathcal{V}_{\{j\}} \,|\, \mathcal{V}_{V\setminus\{i,j\}} \iff [\Gamma(e^{i\theta})]_{ij} = 0. \tag{46}$$

The sparse plus low-rank decomposition in (45) corresponds to a two-layer graphical model: the nodes in the upper layer stand for the (few) latent variables in x; the nodes in the bottom layer represent the observed variables in y whose conditional dependence relations are few.

B. ADMM Formulation

In this Appendix, we want to provide some additional details about the ADMM implementation. This serves as a justification for the update equations which are only reported in the Section V.

The $\mathbf{Z}$*-update step* (39) is equivalent to the minimization of

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

over the set $\mathcal{Z}$, which has no closed-form solution. As noticed in [39], the solution can be 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:

$$\mathbf{Z}^{k+1} = \mathsf{P}_{\mathcal{Z}}\left(\mathbf{Z}^k - t_k\,\nabla\mathcal{J}(\mathbf{Z}^k)\right) \tag{47}$$

where

$$\begin{aligned}\nabla\mathcal{J}(\mathbf{Z}^k) = & -(\hat{\mathbf{\Sigma}}_{\mathsf{y}} + \mathbf{Z}^k)^{-1} + \mathsf{P}_{\mathscr{B}}(\mathbf{M}^k) \\ & + \rho\,\mathsf{P}_{\mathscr{B}}\left(\mathbf{Z}^k - \mathbf{P}^k + \lambda_L I_{mN}\right)\end{aligned}$$

which is precisely the $\mathbf{Z}$-update step in (42).

The $\mathbf{P}$*-update step* (40) involves an optimization problem which is equivalent to minimize the functional

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

over all $\mathbf{P} \geq 0$. As mentioned in Section V, the solution

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

of the minimization of $\mathcal{J}$ over $\mathcal{C}$ is not, in general, positive semidefinite. The following proposition ensures that the projection of a symmetric, block-circulant matrix onto the cone of positive semidefinite matrices is still block-circulant.

Proposition 4: Let $\mathbf{C}$ be a symmetric, block-circulant matrix

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

and let $C(\zeta^k) = V_k \Lambda_k V_k^*$ with $V_k^* V_k = V_k V_k^* = I_m$ and $\Lambda_k = \operatorname{diag}\{\lambda_{k1}, \dots, \lambda_{km}\}$, 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

$\mathbf{C} = \mathbf{W}^*\mathbf{\Lambda}\mathbf{W}$, $\mathbf{W} = \mathbf{V}^*\mathbf{F}$, where $\mathbf{V} = \text{diag}\{V_0, \ldots, V_{N-1}\}$ and $\mathbf{\Lambda} = \text{diag}\{\Lambda_0, \ldots, \Lambda_{N-1}\}$. Then

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

where $\Gamma_k = \text{diag}\{\gamma_{k1}, \ldots, \gamma_{km}\}$ and

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

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

Proof: It is well known that the best positive semidefinite approximation in Frobenious norm of an Hermitian matrix $A \in \mathbb{C}^{n\times n}$, is the matrix obtained by setting the negative eigenvalues of A to zero in its eigenvalue decomposition. Moreover, $\mathsf{P}_{\mathcal{C}^+}(\mathbf{C}) = \mathbf{F}^*\text{diag}\{V_0\Gamma_0V_0^*, \ldots, V_{N-1}\Gamma_{N-1}V_{N-1}^*\}\mathbf{F}$ is a block-circulant matrix because it is block-diagonalized by the Fourier-block matrix.

According to Proposition 4, 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 $\mathcal{C}^+$, from which follows the $\mathbf{P}$-update step in (42).

REFERENCES

[1] D. R. Brillinger, "Remarks concerning graphical models for time series and point processes," *Braz. Rev. Econometrics*, vol. 16, no. 1, pp. 1–23, 1996.

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

[3] J. Songsiri, J. Dahl, and L. Vandenberghe, "Graphical models of autoregressive processes," in *Proc. Convex Optim. Signal Process. Commun.*, 2010, pp. 89–116.

[4] E. Avventi, A. G. Lindquist, and B. Wahlberg, "ARMA identification of graphical models," *IEEE Trans. Autom. Control*, vol. 58, no. 5, pp. 1167–1178, May 2013.

[5] S. Maanan, B. Dumitrescu, and C. Giurcaneanu, "Conditional independence graphs for multivariate autoregressive models by convex optimization: Efficient algorithms," *Signal Process.*, vol. 133, pp. 122–134, 2017.

[6] M. Zorzi, "Empirical Bayesian learning in AR graphical models," *Automatica*, vol. 109, 2019, Art. no. 108516.

[7] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *Ann. Statist.*, vol. 40, pp. 1935–1967, 2012.

[8] M. Zorzi and R. Sepulchre, "AR identification of latent-variable graphical models," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2327–2340, Sep. 2016.

[9] R. Liégeois, B. Mishra, M. Zorzi, and R. Sepulchre, "Sparse plus low-rank autoregressive identification in neuroimaging time series," in *Proc. IEEE Conf. Decis. Control*, 2015, pp. 3965–3970.

[10] B. C. Levy, "Regular and reciprocal multivariate stationary Gaussian reciprocal processes over Z are necessarily Markov," *J. Math. Syst. Estimation Control*, vol. 2, no. 2, pp. 134–154, 1992.

[11] B. C. Levy, R. Frezza, and A. J. Krener, "Modeling and estimation of discrete-time Gaussian reciprocal processes," *IEEE Trans. Autom. Control*, vol. 35, no. 9, pp. 1013–1023, Sep. 1990.

[12] B. C. Levy and A. Ferrante, "Characterization of stationary discrete-time Gaussian reciprocal processes over a finite interval," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 2, pp. 334–355, 2002.

[13] F. P. Carli, A. Ferrante, M. Pavon, and G. Picci, "A maximum entropy solution of the covariance extension problem for reciprocal processes," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 1999–2012, Sep. 2011.

[14] F. Carli, A. Ferrante, M. Pavon, and G. Picci, "An efficient algorithm for maximum entropy extension of block-circulant covariance matrices," *Linear Algebra Appl.*, vol. 439, no. 8, pp. 2309–2329, 2013.

[15] A. G. Lindquist and G. Picci, "The circulant rational covariance extension problem: The complete solution," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2848–2861, Nov. 2013.

[16] A. Ringh, J. Karlsson, and A. Lindquist, "Multidimensional rational covariance extension with applications to spectral estimation and image compression," *SIAM J. Control Optim.*, vol. 54, no. 4, pp. 1950–1982, 2016.

[17] A. Ringh and J. Karlsson, "A fast solver for the circulant rational covariance extension problem," in *Proc. IEEE Eur. Control Conf.*, 2015, pp. 727–733.

[18] A. Lindquist and G. Picci, "Modeling of stationary periodic time series by ARMA representations," in *Optimization and its Applications in Control and Data Sciences*, B. Goldengorin, Ed., Berlin, Germany: Springer, 2016, pp. 281–314.

[19] A. Lindquist, C. Masiero, and G. Picci, "On the multivariate circulant rational covariance extension problem," in *Proc. IEEE Conf. Decis. Control*, 2013, pp. 7155–7161.

[20] D. Alpago, M. Zorzi, and A. Ferrante, "Identification of sparse reciprocal graphical models," *IEEE Contr. Syst. Lett.*, vol. 2, no. 4, pp. 659–664, Oct. 2018.

[21] V. Ciccone, A. Ferrante, and M. Zorzi, "Learning latent variable dynamic graphical models by confidence sets selection," *IEEE Trans. Autom. Control*, vol. 65, no. 12, pp. 5130–5143, Dec. 2020.

[22] D. Alpago, "On the identification of sparse plus low-rank graphical models," Master's thesis, Dept. Inf. Eng., Univ. Padova, Padova, Italy, 2017.

[23] J. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Dept. Geophys., Stanford Univ., Stanford, CA, USA, 1975.

[24] C. Byrnes, T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3189–3205, Nov. 2000.

[25] T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.

[26] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2561–2575, Oct. 2012.

[27] B. Zhu and G. Baggio, "On the existence of a solution to a spectral estimation problem á la Byrnes-Georgiou-Lindquist," *IEEE Trans. Autom. Control*, vol. 64, no. 2, pp. 820–825, Feb. 2019.

[28] M. Pavon and A. Ferrante, "On the geometry of maximum entropy problems," *SIAM Rev.*, vol. 55, no. 3, pp. 415–439, 2013.

[29] T. Georgiou, "Relative entropy and the multivariable multidimensional moment problem," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

[30] M. Zorzi, "A new family of high-resolution multivariate spectral estimators," *IEEE Trans. Autom. Control*, vol. 59, no. 4, pp. 892–904, Apr. 2014.

[31] M. Zorzi, "Multivariate spectral estimation based on the concept of optimal prediction," *IEEE Trans. Autom. Control*, vol. 60, no. 6, pp. 1647–1652, Jun. 2015.

[32] C. I. Byrnes, P. Enqvist, and A. Lindquist, "Identifiability and well-posedness of shaping filter parametrizations: A global analysis approach," *SIAM J. Control Optim.*, vol. 41, no. 1, pp. 23–59, 2002.

[33] G. Baggio, "Further results on the convergence of the Pavon-Ferrante algorithm for spectral estimation," *IEEE Trans. Autom. Control*, vol. 63, no. 10, pp. 3510–3515, Oct. 2018.

[34] R. M. Gray, "Toeplitz and circulant matrices: A review," *Found. Trends Commun. Inf. Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[35] S. Lauritzen, *Graphical Models*. Oxford, U.K.: Oxford Univ. Press, 1996.

[36] A. G. Lindquist and G. Picci, *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*. Berlin, Germany: Springer, 2015.

[37] A. Ferrante, M. Pavon, and M. Zorzi, "A maximum entropy enhancement for a family of high-resolution spectral estimators," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 318–329, Feb. 2012.

[38] M. Zorzi and A. Ferrante, "On the estimation of structured covariance matrices," *Automatica*, vol. 48, no. 9, pp. 2145–2151, 2012.

[39] J. Songsiri and L. Vandenberghe, "Topology selection in graphical models of autoregressive processes," *J. Mach. Learn. Res.*, vol. 11, pp. 2671–2705, 2010.

[40] 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.

[41] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, "OSQP: An operator splitting solver for quadratic programs," in *Proc. Int. Conf. Control*, 2017, pp. 339–339.

Daniele Alpago received the M.S. degree (cum laude) in automation engineering in 2017 from the University of Padova, Padova, Italy, where he is currently working toward the Ph.D degree with the Univesity of Padova, Department of Information Engineering.

His research interests include system identification, spectral estimation, and optimal mass transport.

Mattia Zorzi (Senior Member, IEEE) received the M.S. degree in automation engineering and the Ph.D. degree in information engineering from the University of Padova, Padova, Italy, in 2009 and 2013, respectively.

He held Postdoctoral appointment with the Department of Electrical Engineering and Computer Science, University of Liege, Liege, Belgium, and with the Human Inspired Technology Research Centre, University of Paodva. In 2011, he held visiting positions with the Department of Electrical and Computer Engineering, University of California, Davis, CA, USA, and from 2013 to 2014, he was with the Department of Engineering, University of Cambridge, Cambridge, U.K. He is currently working as an Associate Professor with the Department of Information Engineering, University of Padova. His research interests include machine learning, robust estimation, and identification theory.

Since 2019 and 2021, Dr. Zorzi has been an Associate Editor for the IEEE CONTROL SYSTEMS LETTERS AND AUTOMATICA. Since 2017, he has been an Associate Editor on the IEEE CONTROL SYSTEM SOCIETY CONFERENCE EDITORIAL BOARD and in many other international conferences. He is a Member of the IFAC Technical Committee on Modelling, Identification, and Signal Processing.

Augusto Ferrante was born in Piove di Sacco, Italy, on August 5, 1967. He received the Laurea degree (cum laude) in electrical engineering and the Ph.D. degree in control systems engineering from the University of Padova, Padova, Italy, in 1991 and 1995, respectively.

He has been a Faculty Member with the Colleges of Engineering with the University of Udine, Udine, Italy, and with the Politecnico di Milano, Milan, Italy. He is currently a Professor with the Department of Information Engineering, University of Padova. His research interests include linear systems, spectral estimation, optimal control and optimal filtering, quantum control, and stochastic realization.