

# New Results on Graphical Modeling of High-Dimensional Dependent Time Series

Jitendra K. Tugnait

Dept. of Electrical & Computer Eng.  
Auburn University, Auburn, AL 36849, USA

**Abstract**—We consider the problem of inferring the conditional independence graph (CIG) of a high-dimensional stationary multivariate Gaussian time series. A sparse-group lasso-based frequency-domain formulation of the problem has been considered in the literature where the objective is to estimate the sparse inverse power spectral density (PSD) of the data via optimization of a sparse-group lasso based penalized log-likelihood cost function that is formulated in the frequency-domain. The CIG is then inferred from the estimated inverse PSD. Optimization in the previous approach was performed using an alternating minimization (AM) approach whose performance depends upon choice of a penalty parameter. In this paper we investigate an alternating direction method of multipliers (ADMM) approach for optimization to mitigate dependence on the penalty parameter. We also investigate selection of the tuning parameters based on Bayesian information criterion, and illustrate our approach using synthetic and real data. Comparisons with the “usual” i.i.d. modeling of time series for graph estimation are also provided.

## I. INTRODUCTION

Graphical models are an important and useful tool for analyzing multivariate data [1]. Graphical modeling is a form of multivariate analysis where one uses graphs to represent models. A central concept is that of conditional independence. Given a collection of random variables, one wishes to assess the relationship between two variables, conditioned on the remaining variables. In graphical models, graphs are used to display the conditional independence structure of the variables.

Consider a graph  $\mathcal{G} = (V, \mathcal{E})$  with a set of  $p$  vertices (nodes)  $V = \{1, 2, \dots, p\} = [p]$ , and a corresponding set of (undirected) edges  $\mathcal{E} \subseteq [p] \times [p]$ . Also consider a stationary (real-valued), zero-mean,  $p$ -dimensional multivariate Gaussian time series  $\mathbf{x}(t)$ ,  $t = 0, \pm 1, \pm 2, \dots$ , with  $i$ th component  $x_i(t)$ , and correlation (covariance) matrix function  $\mathbf{R}_{xx}(\tau) = \mathbb{E}\{\mathbf{x}(t + \tau)\mathbf{x}^T(t)\}$ ,  $\tau = 0, \pm 1, \dots$ . Given  $\{\mathbf{x}(t)\}$ , in the corresponding graph  $\mathcal{G}$ , each component series  $\{x_i(t)\}$  is represented by a node ( $i$  in  $V$ ), and associations between components  $\{x_i(t)\}$  and  $\{x_j(t)\}$  are represented by edges between nodes  $i$  and  $j$  of  $\mathcal{G}$ . In a conditional independence graph (CIG), there is no edge between nodes  $i$  and  $j$  if and only if (iff)  $x_i(t)$  and  $x_j(t)$  are conditionally independent given the remaining  $p-2$  scalar series  $x_\ell(t)$ ,  $\ell \in [p]$ ,  $\ell \neq i$ ,  $\ell \neq j$ . Gaussian graphical models (GGM) are CIGs where  $\{\mathbf{x}(t)\}$  is a multivariate Gaussian sequence.

This work was supported by NSF Grants CCF-1617610 and ECCS-2040536. Author’s email: tugnajk@auburn.edu

A key insight in [2] was to transform the series to the frequency domain and express the graph relationships in the frequency domain. Denote the power spectral density (PSD) matrix of  $\{\mathbf{x}(t)\}$  by  $\mathbf{S}_x(f)$ , where  $\mathbf{S}_x(f) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{xx}(\tau)e^{-j2\pi f\tau}$ , the Fourier transform of  $\mathbf{R}_{xx}(\tau)$ . Here  $f$  is the normalized frequency, in Hz, in the interval  $[0, 1)$  or  $(-0.5, 0.5]$ . In [2], [3] it was shown that conditional independence of two time series components given all other components of the time series, is encoded by zeros in the inverse PSD, that is,  $\{i, j\} \notin \mathcal{E}$  iff the  $(i, j)$ -th element of  $\mathbf{S}_x(f)$ ,  $[\mathbf{S}_x^{-1}(f)]_{ij} = 0$  for every  $f$ .

Graphical models were originally developed for random vectors with multiple independent realizations [4, p. 234], i.e., for time series that is independent and identically distributed (i.i.d.):  $p$ -dimensional  $\mathbf{x}(t)$ ,  $t = 1, 2, \dots$ , with  $\mathbf{x}(t_1)$  independent of  $\mathbf{x}(t_2)$  for  $t_1 \neq t_2$ , and  $\mathbf{x}(t)$  identically distributed for any (integer)  $t$ . Such models have been extensively studied, and found to be useful in a wide variety of applications [5]–[7]. Graphical modeling of real-valued time-dependent data (stationary time series) originated with [3], followed by [2].

Graphical modeling of real-valued time-dependent data (stationary time series) originated with [3], followed by [2]. To test whether  $\{i, j\} \notin \mathcal{E}$ , [2] suggested a test based on the maximum of nonparametrically estimated partial coherence over  $f \in [0, 0.5]$ . Edge exclusion tests in this context are also given in [8]–[11]. When  $p$  is large, it may not be feasible to test all  $p(p-1)/2$  edges. Nonparametric approaches for graphical modeling of real time series in high-dimensional settings ( $p$  is large and/or sample size  $N$  is of the order of  $p$ ) have been formulated in the form of penalized log-likelihood in frequency-domain in [12].

In this paper we investigate a penalized log-likelihood approach, as in [13]. An alternating minimization (AM) based solution to this problem is in [13] where simulation comparisons with [12] were also provided; [13] significantly outperforms [12]. An analysis of the properties of the minimizer of the objective function of [13] is given in [14]. In this paper we investigate an alternating direction method of multipliers (ADMM) approach for optimization. While the sparse-group lasso-based penalized log-likelihood formulation of the problem is as in [13], the solution in this paper is via ADMM whose performance does not depend upon a penalty parameter used in variable splitting solution, unlike the AM method of [13].

*Notation:*  $|\mathbf{A}|$  and  $\text{tr}(\mathbf{A})$  denote the determinant and the

trace of the square matrix  $\mathbf{A}$ , respectively.  $[\mathbf{B}]_{ij}$  denotes the  $(i, j)$ -th element of  $\mathbf{B}$ , and so does  $B_{ij}$ .  $\mathbf{I}$  is the identity matrix. The superscripts  $*$  and  $H$  denote the complex conjugate and the Hermitian (conjugate transpose) operations, respectively. The notation  $\mathbf{x} \sim \mathcal{N}_c(\mathbf{m}, \mathbf{\Sigma})$  denotes a random vector  $\mathbf{x}$  that is circularly symmetric (proper) complex Gaussian with mean  $\mathbf{m}$  and covariance  $\mathbf{\Sigma}$ , similarly,  $\mathbf{x} \sim \mathcal{N}_r(\mathbf{m}, \mathbf{\Sigma})$  denotes a random vector  $\mathbf{x}$  that is real-valued Gaussian with mean  $\mathbf{m}$  and covariance  $\mathbf{\Sigma}$ .

## II. SUFFICIENT STATISTIC AND PENALIZED LOG-LIKELIHOOD

Given  $\mathbf{x}(t)$  for  $t = 0, 1, 2, \dots, n-1$ . Define the (normalized) DFT  $\mathbf{d}_x(f_m)$  of  $\mathbf{x}(t)$ , ( $j = \sqrt{-1}$ ),

$$\mathbf{d}_x(f_m) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \mathbf{x}(t) \exp(-j2\pi f_m t) \quad (1)$$

$$f_m = \frac{m}{n}, \quad m = 0, 1, \dots, n-1. \quad (2)$$

It is established in [11] that the set of complex-valued random vectors  $\{\mathbf{d}_x(f_m)\}_{m=0}^{n/2}$  is a sufficient statistic for any inference problem based on dataset  $\{\mathbf{x}(t)\}_{t=0}^{n-1}$ . Suppose  $\mathbf{S}_x(f_k)$  is locally smooth (a standard assumption in PSD estimation), so that  $\mathbf{S}_x(f_k)$  is (approximately) constant over  $K = 2m_t + 1$  consecutive frequency points  $f_m$ s. Pick  $M = \lfloor (\frac{n}{2} - m_t - 1)/K \rfloor$  and

$$\tilde{f}_k = \frac{(k-1)K + m_t + 1}{n}, \quad k = 1, 2, \dots, M, \quad (3)$$

yielding  $M$  equally spaced frequencies  $\tilde{f}_k$  in the interval  $(0, 0.5)$ . By local smoothness

$$\mathbf{S}_x(\tilde{f}_{k,\ell}) = \mathbf{S}_x(\tilde{f}_k) \text{ for } \ell = -m_t, -m_t + 1, \dots, m_t, \quad (4)$$

$$\text{where } \tilde{f}_{k,\ell} = \frac{(k-1)K + m_t + 1 + \ell}{n}. \quad (5)$$

It is known ([15, Theorem 4.4.1]) that asymptotically (as  $n \rightarrow \infty$ ),  $\mathbf{d}_x(f_m)$ ,  $m = 1, 2, \dots, (n/2)-1$ , ( $n$  even), are independent proper (i.e., circularly symmetric), complex Gaussian  $\mathcal{N}_c(\mathbf{0}, \mathbf{S}_x(f_n))$  random vectors, respectively;  $\mathbf{x}(t)$  need not be Gaussian but must satisfy some regularity conditions [11]. Then the joint probability density of the sufficient statistic, for large  $n$ , is

$$f_D(\mathbf{D}) = \prod_{k=1}^M \frac{1}{\pi^{Kp} |\mathbf{S}_x(\tilde{f}_k)|^K} \exp\left(-\text{tr}(\tilde{\mathbf{D}}(\tilde{f}_k) \mathbf{S}_x^{-1}(\tilde{f}_k))\right) \quad (6)$$

$$= \prod_{k=1}^M f_{\tilde{\mathbf{D}}(\tilde{f}_k)}(\tilde{\mathbf{D}}(\tilde{f}_k)) \quad (7)$$

where  $\tilde{\mathbf{D}}(\tilde{f}_k) = [\mathbf{d}_x(\tilde{f}_{k,-m_t}) \mathbf{d}_x(\tilde{f}_{k,-m_t+1}) \dots \mathbf{d}_x(\tilde{f}_{k,m_t})]^H$ ,  $\tilde{\mathbf{D}}(\tilde{f}_k) = \sum_{\ell=-m_t}^{m_t} \mathbf{d}_x(\tilde{f}_{k,\ell}) \mathbf{d}_x^H(\tilde{f}_{k,\ell}) =: K \hat{\mathbf{S}}_k$ , and the PSD estimator using unweighted frequency-domain smoothing is

$$\hat{\mathbf{S}}_k = \frac{1}{K} \sum_{\ell=-m_t}^{m_t} \mathbf{d}_x(\tilde{f}_{k,\ell}) \mathbf{d}_x^H(\tilde{f}_{k,\ell}). \quad (8)$$

In the high-dimension case of  $K < p(p-1)/2$  (# of unknowns in  $\mathbf{S}_x^{-1}(\tilde{f}_k)$ ), one may need to use penalty terms to enforce sparsity and to make the problem well-conditioned. We wish to estimate inverse PSD matrix  $\mathbf{\Phi}_k := \mathbf{S}_x^{-1}(\tilde{f}_k)$ . In terms of  $\mathbf{\Phi}_k$ , we have the log-likelihood [13]

$$\begin{aligned} \ln f_D(\mathbf{D}) &\propto -G(\{\mathbf{\Phi}\}, \{\mathbf{\Phi}^*\}) \\ &:= \sum_{k=1}^M \frac{1}{2} \left[ (\ln |\mathbf{\Phi}_k| + \ln |\mathbf{\Phi}_k^*|) - \text{tr} \left( \hat{\mathbf{S}}_k \mathbf{\Phi}_k + \hat{\mathbf{S}}_k^* \mathbf{\Phi}_k^* \right) \right] \end{aligned} \quad (9)$$

where the first expression in (10) follows by specifying the pdf of  $\mathbf{D}$  in terms of joint pdf of  $\mathbf{D}$  and  $\mathbf{D}^*$  (correct way to handle complex variates [16]). Imposing a sparse-group sparsity constraint [5], [17], [18], minimize a penalized version of negative log-likelihood w.r.t.  $\{\mathbf{\Phi}\}$

$$L_{SGL}(\{\mathbf{\Phi}\}) = G(\{\mathbf{\Phi}\}, \{\mathbf{\Phi}^*\}) + P(\{\mathbf{\Phi}\}), \quad (11)$$

$$P(\{\mathbf{\Phi}\}) = \lambda_1 \sum_{k=1}^M \sum_{i \neq j}^p \left| [\mathbf{\Phi}_k]_{ij} \right| + \lambda_2 \sum_{i \neq j}^p \|\mathbf{\Phi}^{(ij)}\| \quad (12)$$

$$\text{where } \mathbf{\Phi}^{(ij)} := [[\mathbf{\Phi}_1]_{ij} [\mathbf{\Phi}_2]_{ij} \dots [\mathbf{\Phi}_M]_{ij}]^T \in \mathbb{C}^M \quad (13)$$

and  $\lambda_1, \lambda_2 \geq 0$  are tuning parameters. An alternating minimization (AM) based solution to this problem is in [13] where simulation comparisons with [12] were also provided; [13] significantly outperforms [12]. An analysis of the properties of the minimizer  $\{\hat{\mathbf{\Phi}}\}$  of  $L_{SGL}(\{\mathbf{\Phi}\})$  is given in [14]. Following [14], we will set  $\lambda_1 = \alpha\lambda$  and  $\lambda_2 = (1-\alpha)\lambda$  with  $\lambda > 0$ , and  $\alpha \in [0, 1]$  providing a convex combination of lasso and group-lasso penalties [17], [18].

To optimize  $L_{SGL}(\{\mathbf{\Phi}\})$ , using variable splitting, one may reformulate as in [13]:

$$\min_{\{\mathbf{\Phi}\}, \{\mathbf{W}\}} \left\{ G(\{\mathbf{\Phi}\}, \{\mathbf{\Phi}^*\}) + P(\{\mathbf{W}\}) \right\} \quad (14)$$

subject to  $\mathbf{W}_k = \mathbf{\Phi}_k \succ \mathbf{0}$ ,  $k = 1, 2, \dots, M$ , where  $\{\mathbf{\Phi}\} = \{\mathbf{\Phi}_k, k = 1, 2, \dots, M\}$  and  $\{\mathbf{W}\} = \{\mathbf{W}_k, k = 1, 2, \dots, M\}$ . In the AM method, using the penalty method, [13] considers the *relaxed* problem ( $\rho > 0$  is “large”)

$$\min_{\{\mathbf{\Phi}\}, \{\mathbf{W}\}} \left\{ G(\{\mathbf{\Phi}\}, \{\mathbf{\Phi}^*\}) + P(\{\mathbf{W}\}) + \frac{\rho}{2} \sum_{k=1}^M \|\mathbf{\Phi}_k - \mathbf{W}_k\|_F^2 \right\}. \quad (15)$$

The result depends on  $\rho$  and strictly speaking, one must have  $\rho \rightarrow \infty$  which can make the problem numerically ill-conditioned.

## III. ADMM FOR SPARSE-GROUP GRAPHICAL LASSO

In ADMM, we consider the scaled augmented Lagrangian for this problem [5], [19]

$$\begin{aligned} L_\rho(\{\mathbf{\Phi}\}, \{\mathbf{W}\}, \{\mathbf{U}\}) &= G(\{\mathbf{\Phi}\}, \{\mathbf{\Phi}^*\}) + P(\{\mathbf{W}\}) \\ &\quad + \frac{\rho}{2} \sum_{k=1}^M \|\mathbf{\Phi}_k - \mathbf{W}_k + \mathbf{U}_k\|_F^2 \end{aligned} \quad (16)$$

where  $\{\mathbf{U}\} = \{\mathbf{U}_k, k = 1, 2, \dots, M\}$  are dual variables, and  $\rho > 0$  is the “penalty parameter” [19].

### A. ADMM Algorithm

Given the results  $\{\Phi^{(m)}\}, \{W^{(m)}\}, \{U^{(m)}\}$  of the  $m$ th iteration, in the  $(m+1)$ st iteration, an ADMM algorithm executes the following three updates:

- (a)  $\{\Phi^{(m+1)}\} \leftarrow \arg \min_{\{\Phi\}} L_\rho(\{\Phi\}, \{W^{(m)}\}, \{U^{(m)}\})$
- (b)  $\{W^{(m+1)}\} \leftarrow \arg \min_{\{W\}} L_\rho(\{\Phi^{(m+1)}\}, \{W\}, \{U^{(m)}\})$
- (c)  $\{U^{(m+1)}\} \leftarrow \{U^{(m)}\} + (\{\Phi^{(m+1)}\} - \{W^{(m+1)}\})$

1) *Update (a)*: Notice that  $L_\rho(\{\Phi\}, \{W^{(m)}\}, \{U^{(m)}\})$  is separable in  $k$  with  $L_\rho(\{\Phi\}, \{W^{(m)}\}, \{U^{(m)}\}) = \sum_{k=1}^M \frac{1}{2} L_{\rho k}(\Phi_k, W_k^{(m)}, U_k^{(m)})$  up to some terms not dependent upon  $\Phi_k$ 's, where

$$L_{\rho k}(\Phi_k, W_k^{(m)}, U_k^{(m)}) := \ln |\Phi_k| + \ln |\Phi_k^*| - \text{tr}(\hat{S}_k \Phi_k + \hat{S}_k^* \Phi_k^*) + \rho \|\Phi_k - W_k^{(m)} + U_k^{(m)}\|_F^2. \quad (17)$$

As in [19, Sec. 6.5] but accounting for complex-valued vectors/matrices in this paper compared to real-valued vectors/matrices in [5], and therefore using the CR or Wirtinger calculus as in [13], the solution to  $\arg \min_{\Phi_k} L_{\rho k}(\Phi_k, W_k^{(m)}, U_k^{(m)})$  is as follows. Let  $V D V^H$  denote the eigen-decomposition of the matrix  $(\hat{S}_k - \rho W_k^{(m)} + \rho U_k^{(m)})$ . Then  $\Phi_k^{(m+1)} = V \tilde{D} V^H$  where  $\tilde{D}$  is the diagonal matrix with  $\ell$ th diagonal element

$$\tilde{D}_{\ell\ell} = \frac{1}{2\rho} \left( -D_{\ell\ell} + \sqrt{|D_{\ell\ell}|^2 + 4\rho} \right) \quad (18)$$

By construction  $\tilde{D}_{\ell\ell} > 0$  for any  $\rho > 0$ , hence,  $\Phi_k^{(m+1)} = V \tilde{D} V^H \succ 0$ .

2) *Update (b)*: Update  $\{W_k^{(m+1)}\}_{k=1}^M$  as the minimizer w.r.t.  $\{W\}_{k=1}^M$  of

$$\frac{\rho}{2} \sum_{k=1}^M \|W_k - (\Phi_k^{(m+1)} + U_k^{(m)})\|_F^2 + P(\{W\}). \quad (19)$$

The solution is given by [13] (which follows real-valued results of [17]). Define  $(b)_+ := \max(0, b)$ , soft-thresholding operator  $S(b, \beta) := (1 - \beta/|b|)_+ b$ , and vector operator  $[S(\mathbf{a}, \beta)]_j = S(a_j, \beta)$ ,  $a_j = [\mathbf{a}]_j$ . Let  $\mathbf{A}_k = \Phi_k^{(m+1)} + U_k^{(m)}$ . The solution to minimization of (19) is

$$[\hat{W}_k]_{ij} = \begin{cases} [\mathbf{A}_k]_{ii}, & \text{if } i = j \\ S([\mathbf{A}_k]_{ij}, \frac{\alpha \lambda_{ij}}{\rho}) \left( 1 - \frac{(1-\alpha)\lambda_{ij}}{\rho \|S(\mathbf{A}_k^{(ij)}, \alpha \lambda_{ij}/\rho)\|} \right)_+, & \text{if } i \neq j \end{cases}$$

3) *Update (c)*: For the scale Lagrangian formulation of ADMM [19], for  $k = 1, 2, \dots, M$ , update  $U_k^{(m+1)} = U_k^{(m)} + (\Phi_k^{(m+1)} - W_k^{(m+1)})$ .

4) *Algorithm Outline*:

- (i) Initialize the variables:  $\Phi_k^{(0)} = I_p$ ,  $W_k^{(0)} = U_k^{(0)} = \mathbf{0}$  for  $k = 1, 2, \dots, M$ . Pick scalar  $\rho > 0$ .
- (ii) Until convergence, for  $m = 1, 2, \dots$ , do steps (iii)-(v):
- (iii) For  $k = 1, 2, \dots, M$ , update  $\Phi_k^{(m+1)}$  as in Sec. III-A.1.
- (iv) Update  $\{W_k^{(m+1)}\}_{k=1}^M$  as in Sec. III-A.2
- (v) For  $k = 1, 2, \dots, M$ , update  $U_k^{(m+1)}$  as in Sec. III-A.3.

(vi) Denote the converged estimates as  $\hat{\Phi}_k$ ,  $k = 1, \dots, M$ . Edge selection:

$$\text{If } \|\hat{\Phi}^{(ij)}\| > 0, \text{ then } \{i, j\} \in \mathcal{E}, \text{ else } \{i, j\} \notin \mathcal{E}. \quad (20)$$

### B. BIC for selection of $\lambda$ and $\alpha$

Let  $\hat{\Phi}_k$ ,  $k = 1, \dots, M$ , denote the converged estimates. Given  $n$  and choice of  $K$  and  $M$ , the Bayesian information criterion (BIC) is given by

$$\text{BIC}(\lambda, \alpha) = 2K \sum_{k=1}^M \left( -\ln |\hat{\Phi}_k| + \text{tr}(\hat{S}_k \hat{\Phi}_k) \right) + \ln(2KM) \sum_{k=1}^M (\# \text{ of nonzero elements in } \hat{\Phi}_k) \quad (21)$$

where  $2KM$  are total number of real-valued measurements in frequency-domain and  $2K$  are number of real-valued measurements per frequency point, with total  $M$  frequencies in  $(0, \pi)$ . Pick  $\alpha$  and  $\lambda$  to minimize BIC. We use BIC to first select  $\lambda$  over a grid of values with fixed  $\alpha$ , and then with selected  $\lambda$ , we search over  $\alpha$  values in  $[0, 1]$ . This sequential search computationally less demanding than a two-dimensional search.

## IV. NUMERICAL EXAMPLES

### A. Synthetic Data

Consider  $p = 128$ , 16 clusters (communities) of 8 nodes each, where nodes within a community are not connected to any nodes in other communities. Within any community of 8 nodes, the data are generated using a vector autoregressive (VAR) model of order 3. Consider community  $q$ ,  $q = 1, 2, \dots, 16$ . Then  $\mathbf{x}^{(q)}(t) \in \mathbb{R}^8$  is generated as

$$\mathbf{x}^{(q)}(t) = \sum_{i=1}^3 \mathbf{A}_i^{(q)} \mathbf{x}^{(q)}(t-i) + \mathbf{w}^{(q)}(t)$$

Only 10% of entries of  $\mathbf{A}_i^{(q)}$ 's are nonzero and the nonzero elements are independently and uniformly distributed over  $[-0.8, 0.8]$ . We then check if the VAR(3) model is stable with all eigenvalues of the companion matrix  $\leq 0.95$  in magnitude; if not, we re-draw randomly till this condition is fulfilled. The overall data  $\mathbf{x}(t)$  is given by  $\mathbf{x}(t) = [\mathbf{x}^{(1)\top}(t) \dots \mathbf{x}^{(16)\top}(t)]^\top \in \mathbb{R}^p$  with  $\mathbf{w}^{(q)}(t)$  as i.i.d. zero-mean Gaussian with identity covariance matrix. First 100 samples are discarded to eliminate transients. This set-up leads to approximately 3.5% connected edges.

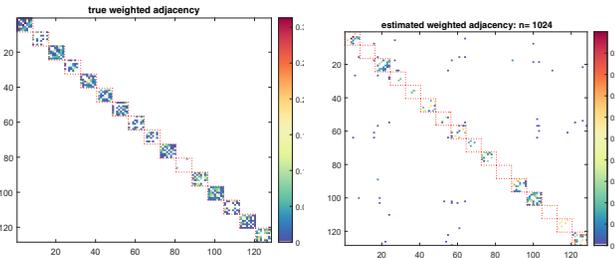
Simulation results are shown in Table I where we used  $M = 4$  for all samples sizes and  $K = 15, 31, 63, 127, 255$  for  $n = 128, 256, 512, 1024, 2048$ , respectively. All ADMM approaches used variable penalty parameter  $\rho$ , as in [19, Sec. 3.4.1], and the stopping (convergence) criterion following [19, Sec. 3.3.1]. The performance measure is  $F_1$ -score for efficacy in edge detection. The  $F_1$ -score is defined as  $F_1 = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$  where  $\text{precision} = |\hat{\mathcal{E}} \cap \mathcal{E}_0| / |\hat{\mathcal{E}}|$ ,  $\text{recall} = |\hat{\mathcal{E}} \cap \mathcal{E}_0| / |\mathcal{E}_0|$ , and  $\mathcal{E}_0$  and  $\hat{\mathcal{E}}$  denote the true and estimated edge sets, respectively. The conventional

Model:	VAR(3) model: number of nodes $p=128$				
sample size $n$	128	256	512	1024	2048
Approach	$F_1$ score ( $\pm\sigma$ )				
i.i.d. modeling	0.1388 $\pm$ 0.0295	0.1649 $\pm$ 0.0434	0.1609 $\pm$ 0.0405	0.2864 $\pm$ 0.0509	0.3017 $\pm$ 0.0301
Proposed: exhaustive	0.3215 $\pm$ 0.0323	0.4256 $\pm$ 0.0368	0.5356 $\pm$ 0.0298	0.6036 $\pm$ 0.0293	0.6861 $\pm$ 0.0236
Proposed: BIC	0.2519 $\pm$ 0.0426	0.2610 $\pm$ 0.0492	0.4942 $\pm$ 0.0555	0.6102 $\pm$ 0.0290	0.6615 $\pm$ 0.0306

TABLE I: “i.i.d. modeling” stands for the “conventional” ADMM lasso approach ([19, Sec. 6.4]) that models data as i.i.d., “Proposed: exhaustive” denotes our proposed ADMM approach where  $(\alpha, \lambda)$  were selected via exhaustive search over a 2-dimensional grid to maximize the  $F_1$ -score, and “Proposed: BIC” denotes our proposed ADMM approach where  $(\alpha, \lambda)$  were selected via BIC.

i.i.d. modeling approach exploits only the sample covariance  $\frac{1}{n} \sum_{t=0}^{n-1} \mathbf{x}(t)\mathbf{x}^\top(t)$  whereas the proposed approaches exploits the entire correlation function (equivalently PSD), and thus, can deliver better performance. In Table I, the label “i.i.d. modeling” stands for the conventional ADMM lasso approach ([19, Sec. 6.4]) that models data as i.i.d., “Proposed: exhaustive” denotes our proposed ADMM approach where  $(\alpha, \lambda)$  were selected via exhaustive search over a 2-dimensional grid to maximize the  $F_1$ -score, and “Proposed: BIC” denotes our proposed ADMM approach where  $(\alpha, \lambda)$  were selected via BIC.

The conventional i.i.d. modeling approach estimates the (sparse) precision matrix  $\Omega = (E\{\mathbf{x}(t)\mathbf{x}^\top(t)\})^{-1}$ : there is an edge  $\{i, j\} \in \mathcal{E}$ . For a typical Monte Carlo run, we show the estimated weighted adjacency matrices resulting from the conventional approach and from the proposed:BIC approach in Figs. 1 and 2 respectively. Fig. 1 shows true and estimated  $\{i, j\} \in \mathcal{E}$  as edge weights, whereas Fig. 2 shows true  $\sqrt{\sum_{k=1}^M |[\Phi_k]_{ij}|^2}$  and estimated  $\sqrt{\sum_{k=1}^M |[\hat{\Phi}_k]_{ij}|^2}$  as edge weights.

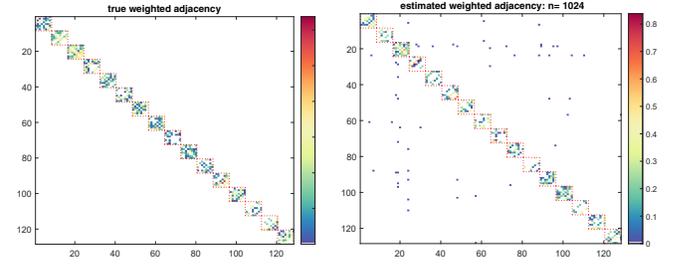


(a) True  $|\Omega_{ij}|$  as edge weight. (b) Estimated  $|\Omega_{ij}|$  as edge weight.

Fig. 1: IID modeling weighted adjacency matrices. The red squares (in dotted lines) show the communities – they are not part of the adjacency matrices.

### B. Real data: Financial Time Series

We consider daily share prices (at close of the day) of 97 stocks in S&P 100 index from Jan. 1, 2013 through Jan. 1, 2018, yielding 1259 samples. If  $y_m(t)$  is share price of  $m$ th stock on day  $t$ , we consider (as is conventional in such studies)  $x_m(t) = \ln(y_m(t)/y_m(t-1))$  as the time series to analyze, yielding  $n = 1258$  and  $p = 97$ . These 97 stocks are



(a) True  $\sqrt{\sum_{k=1}^M |[\Phi_k]_{ij}|^2}$  as edge weight. (b) Estimated  $\sqrt{\sum_{k=1}^M |[\hat{\Phi}_k]_{ij}|^2}$  as edge weight.

Fig. 2: Weighted adjacency matrices for dependent time series modeling:  $M = 4$ . The red squares (in dotted lines) show the communities – they are not part of the adjacency matrices.

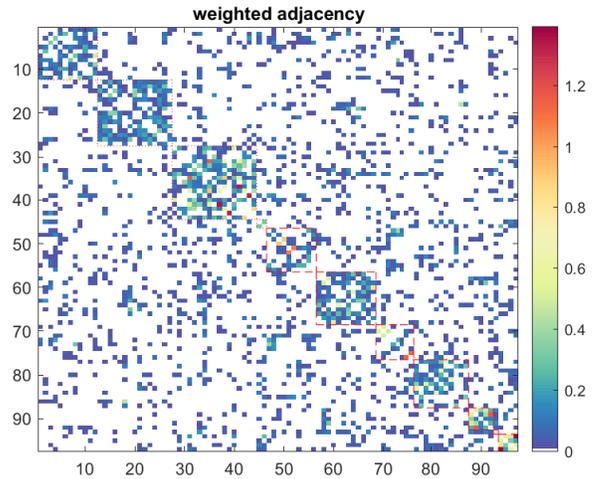
classified into 11 sectors and we order the nodes to group them as information technology (nodes 1-12), health care (13-27), financials (28-44), real estate (45-46), consumer discretionary (47-56), industrials (57-68), communication services (69-76), consumer staples (77-87), energy (88-92), materials (93), utilities (94-97). The weighted adjacency matrices resulting from the conventional i.i.d. modeling approach and the proposed approach with  $M = 4$ , are shown in Fig. 3. In both cases we used BIC to determine the tuning parameters. While the ground truth is unknown, the dependent time series based proposed approach yields sparser, more interpretable CIG (429 edges for the proposed approach versus 1285 edges for conventional modeling) which also conforms better with the sector classification according to the Global Industry Classification Standard.

## V. CONCLUSIONS

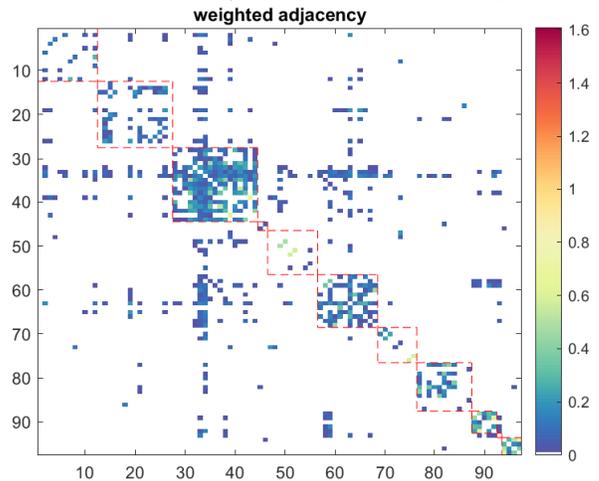
Graphical modeling of dependent Gaussian time series was considered. A sparse-group lasso-based frequency-domain formulation of the problem has been considered in [13] using an alternating minimization (AM) approach whose performance depends upon choice of a penalty parameter. In this paper we investigate an ADMM approach for optimization of sparse-group lasso-based penalized log-likelihood formulation of the problem. The frequency-domain formulation results in consideration of optimization w.r.t. complex-values variables using Wirtinger calculus. The approach was illustrated via synthetic and real data examples.

REFERENCES

- [1] S.L. Lauritzen, *Graphical models*. Oxford, UK: Oxford Univ. Press, 1996.
- [2] R. Dahlhaus, “Graphical interaction models for multivariate time series,” *Metrika*, vol. 51, pp. 157-172, 2000.
- [3] D.R. Brillinger, “Remarks concerning graphical models of times series and point processes,” *Revista de Econometria (Brazilian Rev. Econometr.)*, vol. 16, pp. 1-23, 1996.
- [4] M. Eichler, “Graphical modelling of multivariate time series,” *Probability Theory and Related Fields*, vol. 153, issue 1-2, pp. 233-268, June 2012.
- [5] P. Danaher, P. Wang and D.M. Witten, “The joint graphical lasso for inverse covariance estimation across multiple classes,” *J. Royal Statistical Society, Series B (Methodological)*, vol. 76, pp. 373-397, 2014.
- [6] S.L. Lauritzen and N.A. Sheehan, “Graphical models for genetic analyses,” *Statistical Science*, vol. 18, pp. 489-514, 2003.
- [7] N. Meinshausen and P. Bühlmann, “High-dimensional graphs and variable selection with the Lasso,” *Ann. Statist.*, vol. 34, no. 3, pp. 1436-1462, 2006.
- [8] Y. Matsuda, “A test statistic for graphical modelling of multivariate time series,” *Biometrika*, vol. 93, no. 2, pp. 399-409, 2006.
- [9] R.J. Wolstenholme and A.T. Walden, “An efficient approach to graphical modeling of time series,” *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3266-3276, June 15, 2015.
- [10] J.K. Tugnait, “An edge exclusion test for graphical modeling of multivariate time series,” in *Proc. 52nd Annual Conf. Information Sciences & Systems*, Princeton University, Princeton, NJ, March 21-23, 2018, pp. 1-6.
- [11] J.K. Tugnait, “Edge exclusion tests for graphical model selection: Complex Gaussian vectors and time series,” *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 5062-5077, Oct. 1, 2019.
- [12] A. Jung, G. Hannak and N. Goertz, “Graphical LASSO based model selection for time series,” *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1781-1785, Oct. 2015.
- [13] J.K. Tugnait, “Graphical modeling of high-dimensional time series,” in *Proc. 52nd Asilomar Conference on Signals, Systems and Computers*, pp. 840-844, Pacific Grove, CA, Oct. 29 - Oct. 31, 2018.
- [14] J.K. Tugnait, “Consistency of sparse-group lasso graphical model selection for time series,” in *Proc. 54th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1-4, 2020.
- [15] D.R. Brillinger, *Time Series: Data Analysis and Theory*, Expanded edition. New York: McGraw Hill, 1981.
- [16] P.J. Schreier and L.L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge, UK: Cambridge Univ. Press, 2010.
- [17] J. Friedman, T. Hastie and R. Tibshirani, “A note on the group lasso and a sparse group lasso,” *arXiv:1001.0736v1 [math.ST]*, 5 Jan 2010.
- [18] N. Simon, J. Friedman, T. Hastie and R. Tibshirani, “A sparse-group lasso,” *J. Computational Graphical Statistics*, vol. 22, pp. 231-245, 2013.
- [19] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2010.



(a) Estimated  $|\Omega_{ij}|$  as edge weight; 1285 edges.



(b) Estimated  $\sqrt{\sum_{k=1}^M |[\hat{\Phi}_k]_{ij}|^2}$  as edge weight; 429 edges.

Fig. 3: Weighted adjacency matrices for financial time series. The red squares (in dashed lines) show the 11 sectors: information technology (nodes 1-12), health care (13-27), financials (28-44), real estate (45-46), consumer discretionary (47-56), industrials (57-68), communication services (69-76), consumer staples (77-87), energy (88-92), materials (93), utilities (94-97) – they are not part of the adjacency matrices.