Consistency of Sparse-Group Lasso Graphical Model Selection for 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 p-variate Gaussian time series graphical model associated with an undirected graph with p vertices is defined as the family of time series that obey the conditional independence restrictions implied by the edge set of the graph. A sparse-group lasso-based frequency-domain formulation of the problem has been considered in the literature where the objective is to estimate the 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. In this paper we establish sufficient conditions for consistency of the inverse PSD estimator resulting from the sparse-group graphical lasso-based approach.**

I. Introduction

Graphical interaction models ("graphical models," in short) 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) $V = \{1, 2, \cdots, p\} = [p]$, 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 $\boldsymbol{x}(t)$, $t = 0, \pm 1, \pm 2, \cdots$, with ith component $x_i(t)$, and correlation (covariance) matrix function $\boldsymbol{R}_{xx}(\tau) = \mathbb{E}\{\boldsymbol{x}(t+\tau)\boldsymbol{x}^T(t)\}$, $\tau = 0, \pm 1, \cdots$. Given $\{\boldsymbol{x}(t)\}$, 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 $x_\ell(t)$, $\ell \in [p]$, $\ell \neq i$, $\ell \neq j$. Thus, edge $\{i, j\} \in \mathcal{E}$ iff time series components $x_i(t)$ and $x_j(t)$ are conditionally dependent, and edge $\{i, j\} \notin \mathcal{E}$ iff $x_i(t)$ and $x_j(t)$ are conditionally independent. Gaussian graphical models (GGM) are CIGs where $\{\boldsymbol{x}(t)\}$ is a multivariate Gaussian sequence.

A key insight in [2], [3] 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 $\{\boldsymbol{x}(t)\}$ by $\boldsymbol{S}_x(f)$, where $\boldsymbol{S}_x(f) = \sum_{\tau=-\infty}^{\infty} \boldsymbol{R}_{xx}(\tau) e^{-j2\pi f\tau}$, the Fourier transform of $\boldsymbol{R}_{xx}(\tau)$. 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 $\boldsymbol{S}_x(f)$, $[\boldsymbol{S}_x^{-1}(f)]_{ij} = 0$ 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 $\boldsymbol{x}(t)$, $t = 1, 2, \cdots$, with $\boldsymbol{x}(t_1)$ independent of $\boldsymbol{x}(t_2)$ for $t_1 \neq t_2$, and $\boldsymbol{x}(t)$ identically distributed for any (integer) t. Such models have been extensively studied, and found to be useful in a wide variety of applications [5]–[10]. Graphical modeling of real-valued time-dependent data (stationary time series) originated with [2], followed by [3].

A sparse-group lasso-based frequency-domain formulation of the problem was investigated in [11] where the objective was to estimate the inverse power spectral density (PSD) of the data via optimization of a sparse-group lasso based penalized log-likelihood cost function that was formulated in the frequency-domain. The CIG is then inferred from the estimated inverse PSD. In this paper we establish sufficient conditions for consistency of the approach of [11]. Only the computational aspects of this problem were addressed in [11] where simulation comparisons with [12] were also provided; [11] significantly outperforms [12]. Further comparisons with [12] are in [11].

Notation: Given $\boldsymbol{A} \in \mathbb{C}^{p\times p}$, we use $\phi_{\min}(\boldsymbol{A})$, $\phi_{\max}(\boldsymbol{A})$, $|\mathbf{A}|$, tr($\boldsymbol{A}$) and etr($\boldsymbol{A}$) to denote the minimum eigenvalue, maximum eigenvalue, determinant, trace, and exponential of trace of $\boldsymbol{A}$, respectively. The Kronecker product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is denotes by $\boldsymbol{A} \otimes \boldsymbol{B}$. For $\boldsymbol{B} \in \mathbb{C}^{p\times q}$, we define $\|\boldsymbol{B}\| = \sqrt{\phi_{\max}(\boldsymbol{B}^H\boldsymbol{B})}$, $\|\boldsymbol{B}\|_F = \sqrt{\text{tr}(\boldsymbol{B}^H\boldsymbol{B})}$ and $\|\boldsymbol{B}\|_1 = \sum_{i,j} |B_{ij}|$ where B_{ij} is the (i, j)-th element of $\boldsymbol{B}$, also denoted by $[\boldsymbol{B}]_{ij}$. Given $\boldsymbol{A} \in \mathbb{C}^{p\times p}$, $\boldsymbol{A}^+ = \text{diag}(\boldsymbol{A})$ is a diagonal matrix with the same diagonal as $\boldsymbol{A}$, and $\boldsymbol{A}^- = \boldsymbol{A} - \boldsymbol{A}^+$ is $\boldsymbol{A}$ with all its diagonal elements set to zero. We use $\boldsymbol{A}^{-*}$ for $(\boldsymbol{A}^*)^{-1}$, the inverse of complex conjugate of $\boldsymbol{A}$, and $\boldsymbol{A}^{-\top}$ for $(\boldsymbol{A}^\top)^{-1}$. The notation $\boldsymbol{y}_n = \mathcal{O}_P(\boldsymbol{x}_n)$ for random $\boldsymbol{y}_n, \boldsymbol{x}_n \in \mathbb{C}^p$ means that for any $\varepsilon > 0$, there exists $0 < M < \infty$ such that $P(\|\boldsymbol{y}_n\| \leq M\|\boldsymbol{x}_n\|) \geq 1 - \varepsilon\ \forall n \geq 1$.

II. Penalized Log-Likelihood

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

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

where

$$f_m = m/n, \quad m = 0, 1, \cdots, n-1. \tag{2}$$

It is established in [13] that the set of complex-valued random vectors $\{\boldsymbol{d}_x(f_m)\}_{m=0}^{n/2}$ is a sufficient statistic for any inference problem based on dataset $\{\boldsymbol{x}(t)\}_{t=0}^{n-1}$. Suppose $\boldsymbol{S}_x(f_m)$ is locally smooth (a standard

This work was supported by the National Science Foundation under Grants CCF-1617610 and ECCS-2040536. E-mail: tugnajk@auburn.edu

assumption in PSD estimation), so that $\boldsymbol{S}_x(f_m)$ is (approximately) constant over $K = 2m_t + 1$ consecutive frequency points f_m's. Pick

$$\tilde{f}_k = \frac{(k-1)K + m_t + 1}{N}, \qquad k = 1, 2, \cdots, M, \tag{3}$$

$$M = \left\lfloor \frac{\frac{N}{2} - m_t - 1}{K} \right\rfloor, \tag{4}$$

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

$$\mathbf{S}_x(\tilde{f}_{k,\ell}) = \mathbf{S}_x(\tilde{f}_k) \text{ for } \ell = -m_t, -m_t+1, \cdots, m_t, \tag{5}$$

where

$$\tilde{f}_{k,\ell} = \frac{(k-1)K + m_t + 1 + \ell}{n}. \tag{6}$$

It is known ([14, Theorem 4.4.1]) that asymptotically (as $n \to \infty$), $\boldsymbol{d}_x(f_m)$, $m = 1, 2, \cdots, (n/2) - 1$, ($N$ even), are independent proper (i.e., circularly symmetric), complex Gaussian $\mathcal{N}_c(\mathbf{0}, \boldsymbol{S}_x(f_m))$ random vectors, respectively; $\boldsymbol{x}(t)$ need not be Gaussian but must satisfy some regularity conditions [13]. Then the joint probability density of the sufficient statistic, for large n, is

$$f_{\boldsymbol{D}}(\boldsymbol{D}) = \prod_{k=1}^{M} \frac{\exp\left(-\mathrm{tr}(\tilde{\mathbf{D}}(\tilde{f}_k)\mathbf{S}_x^{-1}(\tilde{f}_k))\right)}{\pi^{Kp}\left|\mathbf{S}_x(\tilde{f}_k)\right|^K} = \prod_{k=1}^{M} f_{\check{\mathbf{D}}(\tilde{f}_k)}(\check{\mathbf{D}}(\tilde{f}_k)) \tag{7}$$

where

$$\check{\boldsymbol{D}}(\tilde{f}_k) = \left[\boldsymbol{d}_x(\tilde{f}_{k,-m_t})\ \boldsymbol{d}_x(\tilde{f}_{k,-m_t+1})\ \cdots\ \boldsymbol{d}_x(\tilde{f}_{k,m_t})\right]^H \tag{8}$$

$$\hat{\boldsymbol{S}}_k = \frac{1}{K} \underbrace{\sum_{\ell=-m_t}^{m_t} \boldsymbol{d}_x(\tilde{f}_{k,\ell})\boldsymbol{d}_x^H(\tilde{f}_{k,\ell})}_{=:\check{\boldsymbol{D}}(\tilde{f}_k)} \tag{9}$$

and $\hat{\boldsymbol{S}}_k$ represents PSD estimator at frequency $\tilde{f}_k$ using unweighted frequency-domain smoothing.

In the high-dimension case of $K < p(p-1)/2$ (# of unknowns in $\boldsymbol{S}_x^{-1}(\tilde{f}_k)$)), one may need to use penalty terms to enforce sparsity and to make the problem well-conditioned. We wish to estimate inverse PSD matrix $\boldsymbol{\Phi}_k := \boldsymbol{S}_x^{-1}(\tilde{f}_k)$. In terms of $\boldsymbol{\Phi}_k$, we have the log-likelihood [11]

$$\ln f_{\boldsymbol{D}}(\boldsymbol{D}) \propto -G(\{\boldsymbol{\Phi}\}) \tag{10}$$

$$:= \sum_{k=1}^{M} \left[(\ln|\boldsymbol{\Phi}_k| + \ln|\boldsymbol{\Phi}_k^*|) - \mathrm{tr}\left(\hat{\boldsymbol{S}}_k\boldsymbol{\Phi}_k + \hat{\boldsymbol{S}}_k^*\boldsymbol{\Phi}_k^*\right)\right] \tag{11}$$

where the first expression in (11) follows by specifying the pdf of $\boldsymbol{D}$ in terms of joint pdf of $\boldsymbol{D}$ and $\boldsymbol{D}^*$ (correct way to handle complex variates [15]). Imposing a sparse-group sparsity constraint [5], [16], minimize a penalized version of negative log-likelihood w.r.t. $\{\boldsymbol{\Phi}\}$

$$L_{SGL}(\{\boldsymbol{\Phi}\}) = -\ln f_{\boldsymbol{D}}(\boldsymbol{D}) + P(\{\boldsymbol{\Phi}\}), \tag{12}$$

$$P(\{\boldsymbol{\Phi}\}) := \bar{\lambda}_1 \sum_{k=1}^{M} \sum_{\substack{i,j=1 \\ i\neq j}}^{p} \left|[\boldsymbol{\Phi}_k]_{ij}\right| + \bar{\lambda}_2 \sum_{\substack{i,j=1 \\ i\neq j}}^{p} \sqrt{\sum_{k=1}^{M} \left|[\boldsymbol{\Phi}_k]_{ij}\right|^2} \tag{13}$$

where $\bar{\lambda}_1$, $\bar{\lambda}_2 \geq 0$ are tuning parameters. Computational aspects of this problem were addressed in [11] where simulation comparisons with [12] were also provided; [11] significantly outperforms [12].

In this paper we analyze properties of the minimizer $\{\hat{\boldsymbol{\Phi}}\}$ of $L_{GL}(\{\boldsymbol{\Phi}\})$.

III. Consistency

Define $p \times (pM)$ matrix $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} = [\boldsymbol{\Phi}_1\ \boldsymbol{\Phi}_2\ \cdots\ \boldsymbol{\Phi}_M] \tag{14}$$

With $0 \leq \alpha \leq 1$, re-express the objective function (12) as

$$L_{SGL}(\boldsymbol{\Omega}) = G(\{\boldsymbol{\Phi}\}) + \alpha\lambda_n \sum_{k=1}^{M_n} \sum_{\substack{i,j=1 \\ i\neq j}}^{p_n} \left|[\boldsymbol{\Phi}_k]_{ij}\right| + (1-\alpha)\lambda_n \sum_{\substack{i,j=1 \\ i\neq j}}^{p_n} \sqrt{\sum_{k=1}^{M_n} \left|[\boldsymbol{\Phi}_k]_{ij}\right|^2} \tag{15}$$

where we now allow p, M, K (see (3)), and λ to be functions of sample size n, denoted as p_n, M_n, K_n and λ_n, respectively. Note that $K_nM_n \approx n/2$. Pick $K_n = a_1n^\gamma$ and $M_n = a_2n^{1-\gamma}$ for some $0.5 < \gamma < 1$ so that $M_n/K_n \to 0$ as $n \to \infty$. We have rewritten $\bar{\lambda}_1$ and $\bar{\lambda}_1$ as $\alpha\lambda_n$ and $(1-\alpha)\lambda_n$, respectively, following [16]. Parameter $\lambda > 0$ is a penalty (tuning) parameter used to control sparsity, and $0 \leq \alpha \leq 1$ yields a convex combination of lasso and group lasso penalties ($\alpha = 0$ gives the group-lasso fit while $\alpha = 1$ yields the lasso fit). In (15), an ℓ_1 penalty term is applied to each off-diagonal element of $\boldsymbol{\Phi}_k$ via $\alpha\lambda_n\left|[\boldsymbol{\Phi}_k]_{ij}\right|$ (lasso), and to the off-block-diagonal group of M_n terms via $(1-\alpha)\lambda_n\sqrt{\sum_{k=1}^{M_n}|[\boldsymbol{\Phi}_k]_{ij}|^2}$ (group lasso). The function $L_{SGL}(\boldsymbol{\Omega})$ is strictly convex in $\boldsymbol{\Omega}$ for $\boldsymbol{\Phi}_k \succ \mathbf{0}$ $\forall k$.

We follow proof technique of [17] which deals with i.i.d time series models and lasso penalty, to establish our main result, Theorem 1. Assume

(A1) Define the true edge set of the graph by $\mathcal{E}_0$, implying that $\mathcal{E}_0 = \{\{i,j\} : [\boldsymbol{S}_0^{-1}(f)]_{ij} \not\equiv 0,\ i \neq j,\ 0 \leq f \leq 0.5\}$ where $\boldsymbol{S}_0(f)$ denotes the true PSD of $\boldsymbol{x}(t)$. (We also use $\boldsymbol{\Phi}_{0k}$ for $\boldsymbol{S}_0^{-1}(\tilde{f}_k)$ where $\tilde{f}_k$ is as in (3), and use $\boldsymbol{\Omega}_0$ to denote the true value of $\boldsymbol{\Omega}$). Assume that $\mathrm{card}(\mathcal{E}_0) = |(\mathcal{E}_0)| \leq s_{n0}$.

(A2) The minimum and maximum eigenvalues of $p_n \times p_n$ PSD $\boldsymbol{S}_0(f) \succ \mathbf{0}$ satisfy

$$0 < \beta_{\min} \leq \min_{f\in[0,0.5]} \phi_{\min}(\boldsymbol{S}_0(f)) \leq \max_{f\in[0,0.5]} \phi_{\max}(\boldsymbol{S}_0(f)) \leq \beta_{\max} < \infty\,.$$

Here $\beta_{\min}$ and $\beta_{\max}$ are not functions of n (or p_n).

Let $\hat{\boldsymbol{\Omega}}_\lambda = \arg\min_{\boldsymbol{\Omega}\,:\,\boldsymbol{\Phi}_k\succ\mathbf{0}} L_{SGL}(\boldsymbol{\Omega})$. Theorem 1 whose proof is given in Sec. IV, establishes consistency of $\hat{\boldsymbol{\Omega}}_\lambda$.

Theorem 1 (Consistency). For $\tau > 2$, let

$$C_0 = 80 \max_{\ell,f}([\boldsymbol{S}_0(f)]_{\ell\ell})\sqrt{\frac{2\ln(16p_n^\tau M_n)}{\ln(p_n)}}\,. \tag{16}$$

Given real numbers $\delta_1 \in (0,1)$, $\delta_2 > 0$ and $C_1 > 0$, let

$$R = C_2C_0/\beta_{\min}^2, \quad C_2 = (4 + C_1 + \delta_2)(1+\delta_1)^2\,, \tag{17}$$

$$r_n = \sqrt{\frac{M_n(p_n + s_{n0})\ln(p_n)}{K_n}}, \quad C_2r_n = o(1)\,, \tag{18}$$

$$N_1 = 2\ln(16p_n^\tau M_n), \tag{19}$$

$$N_2 = \arg\min\left\{n\,:\, r_n \leq \frac{\delta_1\beta_{\min}}{C_2C_0}\right\}\,. \tag{20}$$

Suppose the regularization parameter λ_n and $\alpha \in [0,1]$ satisfy

$$2C_0\sqrt{\frac{\ln(p_n)}{K_n}} \le \frac{\lambda_n}{\sqrt{M_n}} \le \frac{C_1C_0}{1+\alpha(\sqrt{M_n}-1)}\sqrt{\left(1+\frac{p_n}{s_{n0}}\right)\frac{\ln(p_n)}{K_n}}\,. \quad (21)$$

Then if the sample size is such that $K_n > \max\{N_1, N_2\}$ and assumptions (A1)-(A2) hold true, $\hat{\mathbf{\Omega}}_\lambda$ satisfies

$$\|\hat{\mathbf{\Omega}}_\lambda - \mathbf{\Omega}_0\|_F \le Rr_n \quad (22)$$

with probability greater than $1 - 1/p_n^{\tau-2}$. A sufficient condition for the lower bound in (21) to be less than the upper bound for every $\alpha \in [0,1]$ is $C_1 = 2(1+\alpha(\sqrt{M_n}-1))$. In terms of rate of convergence, $\|\hat{\mathbf{\Omega}}_\lambda - \mathbf{\Omega}_0\|_F = \mathcal{O}_P(C_1 r_n)$ •

Remark. If $\alpha = 0$, then C_1 is a constant, and therefore, $\|\hat{\mathbf{\Omega}}_\lambda - \mathbf{\Omega}_0\|_F = \mathcal{O}_P(r_n)$. If $\alpha > 0$, then $C_1 = \mathcal{O}(\sqrt{M_n})$, therefore, $\|\hat{\mathbf{\Omega}}_\lambda - \mathbf{\Omega}_0\|_F = \mathcal{O}_P\left(\sqrt{M_n}\, r_n\right)$. As note before, since $K_nM_n \approx n/2$, if one picks $K_n = a_1 n^\gamma$ and $M_n = a_2 n^{1-\gamma}$ for some $0 < \gamma < 1$, then we must have $\frac{2}{3} < \gamma < 1$ so that $M_n^2/K_n \to 0$ as $n \to \infty$, ensuring $C_2 r_n = o(1)$. In fact, one needs $\sqrt{\frac{M_n^2(p_n+s_{n0})\ln(p_n)}{K_n}} \to 0$. □

IV. Proof of Theorem 1

Our proof relies on the method of [17] which deals with i.i.d time series models and lasso penalty, and our prior results in [18] dealing with complex Gaussian vectors (not time series). From now on we use the term "with high probability" (w.h.p.) to denote with probability greater than $1 - 1/p_n^{\tau-2}$. First we need several auxiliary results.

Lemma 1 below is specialization of [19, Lemma 1] to Gaussian random vectors. It follows from [19, Lemma 1] after setting the sub-Gaussian parameter σ in [19, Lemma 1] to 1.

Lemma 1. Consider a zero-mean Gaussian random vector $\boldsymbol{z} \in \mathbb{R}^p$ with covariance $\boldsymbol{R} \succ \mathbf{0}$. Given n i.i.d. samples $\boldsymbol{z}(t)$, $t = 0, 1, \cdots, n-1$, of $\mathbf{z}$, let $\hat{\boldsymbol{R}} = (1/n)\sum_{t=0}^{n-1} \boldsymbol{z}\boldsymbol{z}^\top$ denote the sample covariance matrix. Then $\hat{\boldsymbol{R}}$ satisfies the tail bound

$$P\left(\left|[\hat{\boldsymbol{R}} - \boldsymbol{R}]_{ij}\right| > \delta\right) \le 4\exp\left(-\frac{n\delta^2}{3200\max_i(R_{ii}^2)}\right) \quad (23)$$

for all $\delta \in (0, 40\max_i(R_{ii}))$ •

Exploiting Lemma 1, we have Lemma 2 regarding $\hat{\boldsymbol{S}}_k$. We denote $\boldsymbol{S}_0(\tilde{f}_k)$ as $\boldsymbol{S}_{0k}$ in this section.

Lemma 2. Under Assumption (A2), $\hat{\boldsymbol{S}}_k$ satisfies the tail bound

$$P\Big(\max_{k,q,l}\Big|[\hat{\boldsymbol{S}}_k - \boldsymbol{S}_{0k}]_{ql}\Big| > C_0\sqrt{\frac{\ln(p_n)}{K_n}}\Big) \le \frac{1}{p_n^{\tau-2}} \quad (24)$$

for $\tau > 2$, if the sample size $n > N_1$, where C_0 is defined in (16) and N_1 is defined in (19). •

Proof. We will use $\boldsymbol{d}_k(\ell)$ for $\boldsymbol{d}_x(\tilde{f}_{k,\ell})$. Notice that $\hat{\boldsymbol{S}}_k$ is the covariance estimate based on K_n i.i.d. samples $\boldsymbol{d}_k(\ell)$ of $\boldsymbol{d}_k \sim \mathcal{N}_c(\mathbf{0}, \boldsymbol{S}_{0k})$. Define $\boldsymbol{d}_{kr} = \mathrm{Re}(\boldsymbol{d}_k)$, $\boldsymbol{d}_{ki} = \mathrm{Im}(\boldsymbol{d}_k)$ and $\boldsymbol{z}_k = [\boldsymbol{d}_{kr}^\top \quad \boldsymbol{d}_{ki}^\top]^\top \in \mathbb{R}^{2p_n}$. Then with $\boldsymbol{R}_{y_1y_2} := \mathbb{E}\{\boldsymbol{y}_1\boldsymbol{y}_2^\top\}$ and $\hat{\boldsymbol{R}}_{y_1y_2} := (1/K_n)\sum_\ell \boldsymbol{y}_1(\ell)\boldsymbol{y}_2^\top(\ell)$, we have

$$\hat{\boldsymbol{S}}_k = \hat{\boldsymbol{R}}_{d_{kr}d_{kr}} + \hat{\boldsymbol{R}}_{d_{ki}d_{ki}} + j\left(-\hat{\boldsymbol{R}}_{d_{kr}d_{ki}} + \hat{\boldsymbol{R}}_{d_{ki}d_{kr}}\right) \quad (25)$$

and

$$\boldsymbol{S}_{0k} = \boldsymbol{R}_{d_{kr}d_{kr}} + \boldsymbol{R}_{d_{ki}d_{ki}} + j\left(-\boldsymbol{R}_{d_{kr}d_{ki}} + \boldsymbol{R}_{d_{ki}d_{kr}}\right) \quad (26)$$

Also, $\boldsymbol{z}_k \sim \mathcal{N}_r(\mathbf{0}, \boldsymbol{R}_{0k})$. Since $\boldsymbol{d}_k(\ell)$ is proper, a little algebra leads to

$$[\boldsymbol{S}_{0k}]_{qq} = 2[\boldsymbol{R}_{0k}]_{qq},\ q = 1, 2, \cdots, p_n \quad (27)$$

Denote the estimate of $\boldsymbol{R}_{0k}$ based on K_n samples as $\hat{\boldsymbol{R}}_{zzk}$. By Lemma 1 and applying the union bound over all $M_n(2p_n)^2$ entries of $\hat{\boldsymbol{R}}_{zzk} - \boldsymbol{R}_{0k}$, $k = 1, 2, \cdots, K_n$, we have

$$P\left(\max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{zzk} - \boldsymbol{R}_{0k}]_{ql}\right| > \frac{\delta}{4}\right) \le P_{tb} = 4M_n(2p_n)^2\exp\left(-\frac{K_n\delta^2/16}{3200\max_{k,i}([\boldsymbol{R}_{0k}]_{ii}^2)}\right) \quad (28)$$

$$= 16M_np_n^2\exp\left(-\frac{K_n\delta^2/16}{800\max_{k,i}([\boldsymbol{S}_{0k}]_{ii}^2)}\right), \quad (29)$$

for all $\delta/4 \in (0, 40\max_{k,i}([\boldsymbol{R}_{0k}]_{ii}))$, equivalently, for all $\delta \in (0, c_*^{-1})$, where in the last step above we have used (27), and

$$c_* = (80\max_{k,i}([\boldsymbol{S}_{0k}]_{ii}))^{-1}. \quad (30)$$

It follows from (25)-(26) that

$$\max_{k,q,l}\left|[\hat{\boldsymbol{S}}_k - \boldsymbol{S}_{0k}]_{ql}\right| \le \max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{d_{kr}d_{kr}} - \boldsymbol{R}_{d_{kr}d_{kr}}]_{ql}\right| + \max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{d_{kr}d_{ki}} - \boldsymbol{R}_{d_{kr}d_{ki}}]_{ql}\right| + \max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{d_{ki}d_{kr}} - \boldsymbol{R}_{d_{ki}d_{kr}}]_{ql}\right| + \max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{d_{ki}d_{ki}} - \boldsymbol{R}_{d_{ki}d_{ki}}]_{ql}\right| \le 4\max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{zzk} - \boldsymbol{R}_{0k}]_{ql}\right|$$

implying

$$\Big\{\max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{zzk} - \boldsymbol{R}_{0k}]_{ql}\right| \le \frac{\delta}{4}\Big\} \subseteq \Big\{\max_{k,q,l}\left|[\hat{\boldsymbol{S}}_k - \boldsymbol{S}_{0k}]_{ql}\right| \le \delta\Big\}. \quad (31)$$

Hence, using (29), we have

$$P\Big(\max_{k,q,l}\left|[\hat{\boldsymbol{S}}_k - \boldsymbol{S}_{0k}]_{ql}\right| \le \delta\Big) \ge P\Big(\max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{zzk} - \boldsymbol{R}_{0k}]_{ql}\right| \le \frac{\delta}{4}\Big) = 1 - P\Big(\max_{k,q,l}\left|[\hat{\boldsymbol{R}}_{zzk} - \boldsymbol{R}_{0k}]_{ql}\right| > \frac{\delta}{4}\Big) \ge 1 - P_{tb}\,.$$

Thus

$$P\Big(\max_{k,q,l}\left|[\hat{\boldsymbol{S}}_k - \boldsymbol{S}_{0k}]_{ql}\right| > \delta\Big) \le P_{tb}\,. \quad (32)$$

Now with $C_0' = c_*^{-1}\sqrt{\frac{2\ln(16p_n^\tau M_n)}{\ln(p_n)}}$, pick δ to satisfy

$$\delta = C_0'\sqrt{\frac{\ln(p_n)}{K_n}} \in (0, c_*^{-1}) \ \Rightarrow\ c_*\delta = c_*C_0'\sqrt{\frac{\ln(p_n)}{K_n}} \in (0,1)\,. \quad (33)$$

Let us pick

$$c_*\delta = \sqrt{\frac{2\ln(16p_n^\tau M_n)}{K_n}} = \sqrt{N_1/K_n} < 1 \quad (34)$$

for $K_n > N_1$, thereby satisfying $\delta \in (0, c_*^{-1})$. Using (30) in (29), we have

$$P_{tb} = 16M_np_n^2/\exp(c_*^2\delta^2K_n/2) = 16M_np_n^2/\exp(\ln(16p_n^\tau M_n)) = \frac{1}{p_n^{\tau-2}}\,. \quad (35)$$

Finally, since $\max_{\ell,f}[\boldsymbol{S}_0(f)]_{\ell\ell} \ge \max_{k,i}[\boldsymbol{S}_{0k}]_{ii}$, hence, $C_0 \ge C_0'$, bound (24) holds. This completes the proof. ■

Lemma 3 deals with a Taylor series expansion using Wirtinger calculus; its proof is omitted for lack of space.

Lemma 3. For $\mathbf{\Phi}_k = \mathbf{\Phi}_k^H \succ \mathbf{0}$, define a real scalar function

$$c(\mathbf{\Phi}_k, \mathbf{\Phi}_k^*) = \ln|\mathbf{\Phi}_k| + \ln|\mathbf{\Phi}_k^*|\,. \quad (36)$$

Let $\boldsymbol{\Phi}_k = \boldsymbol{\Phi}_{0k} + \boldsymbol{\Gamma}_k$ with $\boldsymbol{\Phi}_{0k} = \boldsymbol{\Phi}_{0k}^H \succ \mathbf{0}$ and $\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k^H$. Then using Wirtinger calculus, the Taylor series expansion of $c(\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_k^*)$ is given by

$$\begin{aligned} c(\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_k^*) &= c(\boldsymbol{\Phi}_{0k}, \boldsymbol{\Phi}_{0k}^*) + \mathrm{tr}(\boldsymbol{\Phi}_{0k}^{-1}\boldsymbol{\Gamma}_k + \boldsymbol{\Phi}_{0k}^{-*}\boldsymbol{\Gamma}_k^*) \\ &\quad - \frac{1}{2}(\mathrm{vec}(\boldsymbol{\Gamma}_k))^H(\boldsymbol{\Phi}_{0k}^{-*} \otimes \boldsymbol{\Phi}_{0k}^{-1})\mathrm{vec}(\boldsymbol{\Gamma}_k) \\ &\quad - \frac{1}{2}(\mathrm{vec}(\boldsymbol{\Gamma}_k^*))^H(\boldsymbol{\Phi}_{0k}^{-1} \otimes \boldsymbol{\Phi}_{0k}^{-*})\mathrm{vec}(\boldsymbol{\Gamma}_k^*) + \text{h.o.t.} \end{aligned} \tag{37}$$

where h.o.t. stands for higher-order terms in $\boldsymbol{\Gamma}_k$ and $\boldsymbol{\Gamma}_k^*$. •

Lemma 3 regarding Taylor series expansion immediately leads to Lemma 4 regarding Taylor series with integral remainder, needed to follow the proof of [17] pertaining to the real-valued case.

Lemma 4. With $c(\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_k^*)$ and $\boldsymbol{\Phi}_k = \boldsymbol{\Phi}_{0k} + \boldsymbol{\Gamma}_k$ as in Lemma 3, the Taylor series expansion of $c(\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_k^*)$ in integral remainder form is given by (v is real)

$$\begin{aligned} c(\boldsymbol{\Phi}_k, \boldsymbol{\Phi}_k^*) &= c(\boldsymbol{\Phi}_{0k}, \boldsymbol{\Phi}_{0k}^*) + \mathrm{tr}(\boldsymbol{\Phi}_{0k}^{-1}\boldsymbol{\Gamma}_k + \boldsymbol{\Phi}_{0k}^{-*}\boldsymbol{\Gamma}_k^*) \\ &\quad - \boldsymbol{g}^H(\boldsymbol{\Gamma}_k)\left(\int_0^1 (1-v)\boldsymbol{H}(\boldsymbol{\Phi}_{0k}, \boldsymbol{\Gamma}_k, v)\, dv\right)\boldsymbol{g}(\boldsymbol{\Gamma}_k) \end{aligned} \tag{38}$$

where

$$\boldsymbol{g}(\boldsymbol{\Gamma}_k) = \begin{bmatrix} \mathrm{vec}(\boldsymbol{\Gamma}_k) \\ \mathrm{vec}(\boldsymbol{\Gamma}_k^*) \end{bmatrix}, \quad \boldsymbol{H}(\boldsymbol{\Phi}_{0k}, \boldsymbol{\Gamma}_k, v) = \begin{bmatrix} \boldsymbol{H}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_{22} \end{bmatrix} \tag{39}$$

$$\boldsymbol{H}_{11} = (\boldsymbol{\Phi}_{0k} + v\boldsymbol{\Gamma}_k)^{-*} \otimes (\boldsymbol{\Phi}_{0k} + v\boldsymbol{\Gamma}_k)^{-1} \tag{40}$$

and

$$\boldsymbol{H}_{22} = (\boldsymbol{\Phi}_{0k} + v\boldsymbol{\Gamma}_k)^{-1} \otimes (\boldsymbol{\Phi}_{0k} + v\boldsymbol{\Gamma}_k)^{-*} \quad \bullet \tag{41}$$

We now turn to the proof of Theorem 1.

Proof of Theorem 1. Let $\boldsymbol{\Omega} = \boldsymbol{\Omega}_0 + \boldsymbol{\Delta}$ where

$$\boldsymbol{\Delta} = [\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2\ \cdots\ \boldsymbol{\Gamma}_{M_n}] \tag{42}$$

$$\boldsymbol{\Gamma}_k = \boldsymbol{\Phi}_k - \boldsymbol{\Phi}_{0k},\ k = 1, 2, \cdots, M_n, \tag{43}$$

and $\boldsymbol{\Phi}_k$, $\boldsymbol{\Phi}_{0k}$ are both Hermitian positive-definite, implying $\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k^H$. Let

$$Q(\boldsymbol{\Omega}) := \mathcal{L}_{SGL}(\boldsymbol{\Omega}) - \mathcal{L}_{SGL}(\boldsymbol{\Omega}_0)\,. \tag{44}$$

The estimate $\hat{\boldsymbol{\Omega}}_\lambda$, denoted by $\hat{\boldsymbol{\Omega}}$ hereafter suppressing dependence upon λ, minimizes $Q(\boldsymbol{\Omega})$, or equivalently, $\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0$ minimizes $G(\boldsymbol{\Delta}) := Q(\boldsymbol{\Omega}_0 + \boldsymbol{\Delta})$. We will follow the method of proof of [17, Theorem 1] pertaining to real-valued i.i.d. time series. Consider the set

$$\Theta_n(R) := \left\{\boldsymbol{\Delta} \,:\, \boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k^H\ \forall k,\ \|\boldsymbol{\Delta}\|_F = Rr_n\right\} \tag{45}$$

where R and r_n are as in (17) and (18), respectively. Observe that $G(\boldsymbol{\Delta})$ is a convex function of $\boldsymbol{\Delta}$, and

$$G(\hat{\boldsymbol{\Delta}}) \le G(\mathbf{0}) = 0\,. \tag{46}$$

Therefore, if we can show that

$$\inf_{\boldsymbol{\Delta}}\{G(\boldsymbol{\Delta}) \,:\, \boldsymbol{\Delta} \in \Theta_n(R)\} \,>\, 0\,, \tag{47}$$

the minimizer $\hat{\boldsymbol{\Delta}}$ must be inside the sphere defined by $\Theta_n(R)$, and hence

$$\|\hat{\boldsymbol{\Delta}}\|_F \le Rr_n\,. \tag{48}$$

Using Lemma 4 we rewrite $G(\boldsymbol{\Delta})$ as

$$G(\boldsymbol{\Delta}) = \sum_{k=1}^{M_n}(A_{1k} + A_{2k} + A_{3k}) + A_4\,, \tag{49}$$

where, noting that $\boldsymbol{\Phi}_k^{-1} = \boldsymbol{S}_k$,

$$A_{1k} = \boldsymbol{g}^H(\boldsymbol{\Gamma}_k)\left(\int_0^1 (1-v)\boldsymbol{H}(\boldsymbol{\Phi}_{0k}, \boldsymbol{\Gamma}_k, v)\, dv\right)\boldsymbol{g}(\boldsymbol{\Gamma}_k)\,, \tag{50}$$

$$A_{2k} = \mathrm{tr}\left((\hat{\boldsymbol{S}} - \boldsymbol{S}_{0k})\boldsymbol{\Gamma}_k + (\hat{\boldsymbol{S}} - \boldsymbol{S}_{0k})^*\boldsymbol{\Gamma}_k^*\right)\,, \tag{51}$$

$$A_{3k} = \bar{\lambda}_1(\|\boldsymbol{\Phi}_{0k}^- + \boldsymbol{\Gamma}_k^-\|_1 - \|\boldsymbol{\Phi}_{0k}^-\|_1)\,, \tag{52}$$

$$A_4 = \bar{\lambda}_2\sum_{i\ne j}^{p_n}(\|\boldsymbol{\Omega}_0^{(ij)} + \boldsymbol{\Delta}^{(ij)}\|_F - \|\boldsymbol{\Omega}_0^{(ij)}\|_F)\,, \tag{53}$$

$$\boldsymbol{\Omega}_0^{(ij)} := [[\boldsymbol{\Phi}_{01}]_{ij}\ \cdots\ [\boldsymbol{\Phi}_{0M_n}]_{ij}]^\top \in \mathbb{C}^{M_n}\,, \tag{54}$$

$$\boldsymbol{\Delta}^{(ij)} := [[\boldsymbol{\Gamma}_1]_{ij}\ \cdots\ [\boldsymbol{\Gamma}_{M_n}]_{ij}]^\top \in \mathbb{C}^{M_n}\,. \tag{55}$$

Define

$$d_{1n} := \sqrt{\frac{\ln(p_n)}{K_n}},\ \ d_{2n} := d_{1n}\sqrt{(p_n + s_{n0})}\,. \tag{56}$$

Similar to proof of [18, Theorem 1] (see (42) therein), we deduce that

$$A_{1k} \ge \|\boldsymbol{\Gamma}_k\|_F^2\left(\beta_{\min}^{-1} + \|\boldsymbol{\Gamma}_k\|_F\right)^{-2}\,. \tag{57}$$

But $\|\boldsymbol{\Gamma}_k\|_F \le \|\boldsymbol{\Delta}\|_F = Rr_n$. Hence

$$A_{1k} \ge \|\boldsymbol{\Gamma}_k\|_F^2\left(\beta_{\min}^{-1} + Rr_n\right)^{-2} \tag{58}$$

and

$$A_1 = \sum_{k=1}^{M_n} A_{1k} \ge \frac{\sum_{k=1}^{M_n}\|\boldsymbol{\Gamma}_k\|_F^2}{\left(\beta_{\min}^{-1} + Rr_n\right)^2} = \frac{\|\boldsymbol{\Delta}\|_F^2}{\left(\beta_{\min}^{-1} + Rr_n\right)^2} \tag{59}$$

Similar to the proof of [18, Theorem 1] (see (48) therein), we deduce that w.h.p.

$$|A_{2k}| \le 2C_0\big(\|\boldsymbol{\Gamma}_k^-\|_1\, d_{1n} + \|\boldsymbol{\Gamma}_k^+\|_F\, d_{2n}\big)\,. \tag{60}$$

Hence with $A_2 = \sum_{k=1}^{M_n} A_{2k}$,

$$|A_2| \le \sum_{k=1}^{M_n}|A_{2k}| \le 2C_0\sum_{k=1}^{M_n}\Big(d_{1n}\|\boldsymbol{\Gamma}_k^-\|_1 + d_{2n}\|\boldsymbol{\Gamma}_k^+\|_F\Big)\,. \tag{61}$$

We now derive an alternative bound on A_2. We have w.h.p.

$$\begin{aligned} |A_2| &\le 2\sum_{i,j=1}^{p_n}\sum_{k=1}^{M_n}\big|[(\hat{\boldsymbol{S}} - \boldsymbol{S}_{0k}]_{ij}\big|\cdot\big|[\boldsymbol{\Gamma}_k]_{ij}\big| && (62)\\ &\le 2C_0\, d_{1n}\sum_{i,j=1}^{p_n}\sum_{k=1}^{M_n}\big|[\boldsymbol{\Gamma}_k]_{ij}\big| && (63)\\ &\le 2C_0\, d_{1n}\sum_{i,j=1}^{p_n}\left(\sqrt{M_n}\|\boldsymbol{\Delta}^{(ij)}\|_F\right) && (64)\\ &= 2\sqrt{M_n}C_0\, d_{1n}\big(\|\tilde{\boldsymbol{\Delta}}^-\|_1 + \tilde{\boldsymbol{\Delta}}^+\|_1\big) && (65) \end{aligned}$$

where $\tilde{\boldsymbol{\Delta}} \in \mathbb{R}^{p_n\times p_n}$ has its (i,j)th element $\tilde{\Delta}_{ij} = \|\boldsymbol{\Delta}^{(ij)}\|_F$.

We now bound A_{3k}. Let $\mathcal{E}_0^c$ denote the complement of $\mathcal{E}_0$, given by $\mathcal{E}_0^c = \{\{i,j\} \,:\, [\boldsymbol{S}_0^{-1}(f)]_{ij} \equiv 0,\ i \ne j,\ 0 \le f \le 0.5\}$. For an index set $\boldsymbol{B}$ and a matrix $\boldsymbol{C} \in \mathbb{C}^{p\times p}$, we write $\boldsymbol{C}_{\boldsymbol{B}}$ to denote a matrix in $\mathbb{C}^{p\times p}$ such that $[\boldsymbol{C}_{\boldsymbol{B}}]_{ij} = C_{ij}$ if $(i,j) \in \boldsymbol{B}$, and $[\boldsymbol{C}_{\boldsymbol{B}}]_{ij} = 0$ if $(i,j) \notin \boldsymbol{B}$. Then $\boldsymbol{\Gamma}_k^- = \boldsymbol{\Gamma}_{k\mathcal{E}_0}^- + \boldsymbol{\Gamma}_{k\mathcal{E}_0^c}^-$, and $\|\boldsymbol{\Gamma}_k^-\|_1 = \|\boldsymbol{\Gamma}_{k\mathcal{E}_0}^-\|_1 + \|\boldsymbol{\Gamma}_{k\mathcal{E}_0^c}^-\|_1$. We have

$$\begin{aligned} A_{3k} &= \bar{\lambda}_1(\|\boldsymbol{\Phi}_{0k}^- + \boldsymbol{\Gamma}_k^-\|_1 - \|\boldsymbol{\Phi}_{0k}^-\|_1) \\ &= \bar{\lambda}_1(\|\boldsymbol{\Phi}_{0k}^- + \boldsymbol{\Gamma}_{k\mathcal{E}_0}^-\|_1 + \boldsymbol{\Gamma}_{k\mathcal{E}_0^c}^-\|_1 - \|\boldsymbol{\Phi}_{0k}^-\|_1) \\ &\ge \bar{\lambda}_1(\|\boldsymbol{\Gamma}_{k\mathcal{E}_0^c}^-\|_1 - \|\boldsymbol{\Gamma}_{k\mathcal{E}_0}^-\|_1) \end{aligned} \tag{66}$$

leading to ($A_3 = \sum_{k=1}^{M_n} A_{3k}$)

$$A_3 \geq \bar{\lambda}_1 \sum_{k=1}^{M_n} (\|\mathbf{\Gamma}^-_{k\mathcal{E}_0^c}\|_1 - \|\mathbf{\Gamma}^-_{k\mathcal{E}_0}\|_1)\,. \tag{67}$$

Similarly,

$$A_4 \geq \bar{\lambda}_2 (\|\tilde{\mathbf{\Delta}}^-_{\mathcal{E}_0^c}\|_1 - \|\tilde{\mathbf{\Delta}}^-_{\mathcal{E}_0}\|_1)\,. \tag{68}$$

By Cauchy-Schwartz inequality, $\|\mathbf{\Gamma}^-_{k\mathcal{E}_0}\|_1 \leq \sqrt{s_{n0}}\|\mathbf{\Gamma}^-_{k\mathcal{E}_0}\|_F \leq \sqrt{s_{n0}}\|\mathbf{\Gamma}_k\|_F$, hence

$$\sum_{k=1}^{M_n} \|\mathbf{\Gamma}^-_{k\mathcal{E}_0}\|_1 \leq \sqrt{M_n s_{n0}}\|\mathbf{\Delta}\|_F\,. \tag{69}$$

Set $\|\mathbf{\Gamma}^-_k\|_1 = \|\mathbf{\Gamma}^-_{k\mathcal{E}_0}\|_1 + \|\mathbf{\Gamma}^-_{k\mathcal{E}_0^c}\|_1$ in A_2 of (61) to deduce that w.h.p.

$$\begin{aligned}
\alpha A_2 + A_3 &\geq \alpha(\lambda_n - 2C_0 d_{1n}) \sum_{k=1}^{M_n} \|\mathbf{\Gamma}^-_{k\mathcal{E}_0^c}\|_1 \\
&\quad - \alpha(2C_0 d_{1n} + \lambda_n) \sum_{k=1}^{M_n} \|\mathbf{\Gamma}^-_{k\mathcal{E}_0}\|_1 - \alpha 2C_0 d_{2n} \sum_{k=1}^{M_n} \|\mathbf{\Gamma}^+_k\|_F \\
&\geq -\alpha\Big((2C_0 d_{1n} + \lambda_n)\sqrt{s_{n0}} - 2C_0 d_{2n}\Big)\sqrt{M_n}\|\mathbf{\Delta}\|_F
\end{aligned} \tag{70}$$

where we have used the fact that $\lambda_n \geq 2C_0 d_{1n}$ and $\sum_{k=1}^{M_n} \|\mathbf{\Gamma}^+_k\|_F \leq \sqrt{M_n}\|\mathbf{\Delta}\|_F$. Now use A_2 of (65) to deduce that w.h.p.

$$\begin{aligned}
(1-\alpha)A_2 + A_4 &\geq (1-\alpha)\Big((\lambda_n - 2C_0\sqrt{M_n} d_{1n})\|\tilde{\mathbf{\Delta}}^-_{\mathcal{E}_0^c}\|_1 \\
&\quad - (2C_0\sqrt{M_n} d_{1n} + \lambda_n)\|\tilde{\mathbf{\Delta}}^-_{\mathcal{E}_0}\|_1 - 2C_0\sqrt{M_n p_n} d_{1n}\|\mathbf{\Delta}\|_F\Big) \\
&\geq -(1-\alpha)\|\mathbf{\Delta}\|_F\Big(\lambda_n\sqrt{s_{n0}} + 2C_0\sqrt{M_n} d_{1n}\big(\sqrt{s_{n0}} + \sqrt{p_n}\big)\Big)
\end{aligned} \tag{71}$$

where we have used the facts that $\lambda_n \geq 2C_0\sqrt{M_n} d_{1n}$, and $\|\tilde{\mathbf{\Delta}}^-_{\mathcal{E}_0}\|_1 \leq \sqrt{s_{n0}}\|\tilde{\mathbf{\Delta}}^-_{\mathcal{E}_0}\|_F \leq \sqrt{s_{n0}}\|\mathbf{\Delta}\|_F$ (by Cauchy-Schwartz inequality).

Since $r_n = \sqrt{M_n} d_{2n} > \sqrt{M_n s_{n0}} d_{1n}$, w.h.p. we have

$$\begin{aligned}
A_2 + A_3 + A_4 &\geq -\|\mathbf{\Delta}\|_F\Big(\alpha\big(4C_0 r_n + \lambda_n\sqrt{M_n s_{n0}}\big) \\
&\quad + (1-\alpha)\big(\lambda_n\sqrt{s_{n0}} + 4C_0 r_n\big)\Big) \\
&\geq -\|\mathbf{\Delta}\|_F\Big(4C_0 r_n + \lambda_n\sqrt{s_{n0}}(\alpha\sqrt{M_n} + (1-\alpha))\Big) \\
&\geq -\|\mathbf{\Delta}\|_F\Big((4 + C_1)C_0 r_n\Big)
\end{aligned} \tag{72}$$

where we have used the fact that, by (18) and (21), $\lambda_n\sqrt{s_{n0}}(\alpha\sqrt{M_n} + (1-\alpha)) \leq C_1 C_0 r_n$. Using (49), (59) and (72), and $\|\mathbf{\Delta}\|_F = R r_n$, we have w.h.p.

$$G(\mathbf{\Delta}) \geq \|\mathbf{\Delta}\|_F^2\Big[\big(\beta_{\min}^{-1} + R r_n\big)^{-2} - \frac{(4+C_1)C_0}{R}\Big]\,. \tag{73}$$

For $n \geq N_2$, if we pick R as specified in (17), we obtain $R r_n \leq R r_{N_2} \leq \delta_1/\beta_{\min}$. Then

$$\begin{aligned}
\frac{1}{(\beta_{\min}^{-1} + R r_n)^2} &\geq \frac{\beta_{\min}^2}{(1+\delta_1)^2} = \frac{(4 + C_1 + \delta_2)C_0}{R} \\
&> \frac{(4+C_1)C_0}{R}\,,
\end{aligned}$$

implying $G(\mathbf{\Delta}) > 0$ w.h.p. This proves the desired result. ■

V. Conclusions

Graphical modeling of dependent Gaussian time series was considered. A sparse-group lasso-based frequency-domain formulation of the problem has previously been investigated in [11] where the objective was to estimate the 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 graphical model is then inferred from the estimated inverse PSD. In this paper we established sufficient conditions for consistency of the inverse PSD estimator resulting from the sparse-group graphical lasso-based approach.

References

[1] S.L. Lauritzen, *Graphical models*. Oxford, UK: Oxford Univ. Press, 1996.

[2] 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.

[3] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, pp. 157-172, 2000.

[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] N. Friedman, "Inferring cellular networks using probabilistic graphical models," *Science*, vol 303, pp. 799-805, 2004.

[7] S.L. Lauritzen and N.A. Sheehan, "Graphical models for genetic analyses," *Statistical Science*, vol. 18, pp. 489-514, 2003.

[8] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the Lasso," *Ann. Statist.*, vol. 34, no. 3, pp. 1436-1462, 2006.

[9] K. Mohan, M.J.Y. Chung, S. Han, D. Witten, S.I. lee and M. Fazel, "Structured learning of Gaussian graphical models," in *Proc. NIPS 2012*, Lake Tahoe, Dec. 2012, pp. 620-628.

[10] K. Mohan, P. London, M. Fazel, D. Witten and S.I. Lee, "Node-based learning of multiple Gaussian graphical models," *J. Machine Learning Research*, vol. 15, pp. 445-488, 2014.

[11] 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.

[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, "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.

[14] D.R. Brillinger, *Time Series: Data Analysis and Theory*, Expanded edition. New York: McGraw Hill, 1981.

[15] P.J. Schreier and L.L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge, UK: Cambridge Univ. Press, 2010.

[16] N. Simon, J. Friedman, T. Hastie and R. Tibshirani, "A sparse-group lasso," *J. Computational Graphical Statistics*, vol. 22, pp. 231-245, 2013.

[17] A.J. Rothman, P.J. Bickel, E. Levina and J. Zhu, "Sparse permutation invariant covariance estimation," *Electronic J. Statistics*, vol. 2, pp. 494-515, 2008.

[18] J.K. Tugnait, "On sparse complex Gaussian graphical model selection," in *Proc. 2019 IEEE Intern. Workshop on Machine Learning for Signal Processing (MLSP 2019)*, Pittsburgh, PA, Oct. 13-16, 2019.

[19] P. Ravikumar, M.J. Wainwright, G. Raskutti and B. Yu, "High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence," *Electronic J. Statistics*, vol. 5, pp. 935-980, 2011.