2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) | 979-8-3503-4452-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/CAMSAP58249.2023.10403521

ESTIMATION OF DIFFERENTIAL GRAPHS FROM TIME-DEPENDENT DATA

Jitendra K. Tugnait

Department of Electrical & Computer Engineering
Auburn University, Auburn, AL 36849, USA. tugnajk@auburn.edu

ABSTRACT

We consider the problem of estimating differences in two time series Gaussian graphical models (TSGGMs) which are known to have similar structure. The TSGGM structure is encoded in its inverse power spectral density (IPSD) just as the vector GGM structure is encoded in its precision (inverse covariance) matrix. Motivated by many applications, in existing works one is interested in estimating the difference in two precision matrices to characterize underlying changes in conditional dependencies of two sets of data comprised of independent and identically distributed observations. In this paper we consider estimation of the difference in two IPSD's to characterize underlying changes in conditional dependencies of two sets of time-dependent data. We analyze a group lasso penalized D-trace loss function approach in the frequency domain for differential graph learning, using Wirtinger calculus. An alternating direction method of multipliers (ADMM) algorithm is presented to optimize the objective function. Theoretical analysis establishing consistency of IPSD difference estimator in high-dimensional settings is presented. We illustrate our approach using a numerical example.

Keywords: Sparse graph learning; differential graph estimation; undirected graph; time series graphs.

1. INTRODUCTION

Graphical models are an important and useful tool for analyzing multivariate data [1]. 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. 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 $x_\ell(t)$, $\ell \in [p]$, $\ell \neq i$, $\ell \neq j$.

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 $\{\boldsymbol{x}(t)\}$ by $\boldsymbol{S}_x(f)$, where $\boldsymbol{S}_x(f) = \sum_{\tau=-\infty}^{\infty} \boldsymbol{R}_{xx}(\tau) e^{-j2\pi f\tau}$. In [2] 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 (IPSD), that is, $\{i, j\} \notin \mathcal{E}$ iff the (i, j)-th element of $\boldsymbol{S}_x^{-1}(f)$ vanishes, i.e., $[\boldsymbol{S}_x^{-1}(f)]_{ij} = 0$ for every f. Hence one can use estimated IPSD of observed time series to infer the associated graph.

Graphical models were originally developed for random vectors (i.i.d. time series) [3, p. 234]. In particular, Gaussian graphical models (GGMs) are CIGs where $\boldsymbol{x}$ is multivariate Gaussian. Suppose $\boldsymbol{x}$ has positive-definite covariance matrix $\boldsymbol{\Sigma}$ with inverse covariance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. Then Ω_{ij}, the (i, j)-th element of $\boldsymbol{\Omega}$, is zero iff x_i and x_j are conditionally independent. Such models have been extensively studied, and found to be useful in a wide variety of applications [4–7]. Graphical modeling of real-valued time-dependent data (stationary time series) originated with [8], followed by [2]. 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 investigated in [9–12], among others.

More recently there has been increasing interest in differential network analysis where one is interested in estimating the difference in two inverse covariance matrices [13–17]. Given observations $\boldsymbol{x}$ and $\boldsymbol{y}$ from two groups of subjects, one is interested in the difference $\boldsymbol{\Delta} = \boldsymbol{\Omega}_y - \boldsymbol{\Omega}_x$, where $\boldsymbol{\Omega}_x = (E\{\boldsymbol{x}\boldsymbol{x}^\top\})^{-1}$ and $\boldsymbol{\Omega}_y = (E\{\boldsymbol{y}\boldsymbol{y}^\top\})^{-1}$. The associated differential graph is $\mathcal{G}_\Delta = (V, \mathcal{E}_\Delta)$ where $\{i, j\} \in \mathcal{E}_\Delta$ iff $[\boldsymbol{\Delta}]_{ij} \neq 0$. It characterizes differences between the GGMs of the two sets of data. We use the term differential graph as in [17, 18] ([13, 14, 16] use the term differential network). As noted in [16], in biostatistics, the differential network/graph describes the changes in conditional dependencies between components under different environmental or genetic conditions. For instance, one may be interested in the differences in the graphical models of healthy and impaired subjects, or models under different disease states, given gene expression data or functional MRI signals [4, 19, 20].

In contrast to the approaches of [13–17], in this paper we address the problem of estimating differences in two time series Gaussian graphical models (TSGGMs) which are known to have similar structure. The TSGGM structure is encoded in its IPSD just as the vector GGM structure is encoded in its precision matrix. We consider estimation of the difference in two IPSD's to characterize underlying changes in conditional dependencies of two sets of time-dependent data $\{\boldsymbol{x}(t)\}_{t=1}^{n_x}$ and $\{\boldsymbol{y}(t)\}_{t=1}^{n_y}$. We analyze a group lasso penalized D-trace loss function approach in the frequency domain for differential graph learning, using Wirtinger calculus [21]. As a preliminary step, we first address the problem of estimation of complex differential graphs, given two complex-valued i.i.d. time series. This problem and the general problem of differential times series graph estimation have not been investigated before. The work of [18] considers time series differential graphs except that in [18] $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are non-stationary ("functional" modeling), and instead of a single record (sample) of $\boldsymbol{x}(t)$, $t = 1, 2, \cdots, n_x$ and $\boldsymbol{y}(t)$, $t = 1, 2, \cdots, n_y$, as in this paper, they assume multiple independent observations of $\boldsymbol{x}(t)$, $t \in \mathcal{T}$, and $\boldsymbol{y}(t)$, $t \in \mathcal{T}$ (a closed subset of

This work is supported by NSF Grant CCF-2308473.

real line).

Notation: For a set V, $|V|$ denotes its cardinality. Given $\boldsymbol{A} \in \mathbb{C}^{p\times p}$, we use $\phi_{\min}(\boldsymbol{A})$, $\phi_{\max}(\boldsymbol{A})$, $|\boldsymbol{A}|$ and $\mathrm{tr}(\boldsymbol{A})$ to denote the minimum eigenvalue, maximum eigenvalue, determinant and trace of $\boldsymbol{A}$, respectively. $[\boldsymbol{B}]_{ij}$ denotes the (i,j)-th element of $\boldsymbol{B}$, and so does B_{ij}. $\boldsymbol{I}$ is the identity matrix. The symbol $\otimes$ denotes the Kronecker product. The superscripts $*$ and H denote the complex conjugate and the Hermitian (conjugate transpose) operations, respectively. For $\boldsymbol{B} \in \mathbb{C}^{p\times q}$, we define $\|\boldsymbol{B}\| = \sqrt{\phi_{\max}(\boldsymbol{B}^H\boldsymbol{B})}$, $\|\boldsymbol{B}\|_F = \sqrt{\mathrm{tr}(\boldsymbol{B}^H\boldsymbol{B})}$, $\|\boldsymbol{B}\|_1 = \sum_{i,j}|B_{ij}|$ and $\|\boldsymbol{B}\|_\infty = \max_{i,j}|B_{ij}|$. The notation $\boldsymbol{x} \sim \mathcal{N}_c(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a random vector $\boldsymbol{x}$ that is circularly symmetric (proper) complex Gaussian with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$. Similarly, $\boldsymbol{x} \sim \mathcal{N}_r(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a random vector $\boldsymbol{x}$ that is real-valued Gaussian with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$.

2. COMPLEX DIFFERENTIAL GRAPHS

We first recall a formulation of [13–16] for real-valued data. Let $\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{x} \sim \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{\Sigma}_{0x})$ and suppose we are given i.i.d. samples $\boldsymbol{x}(t), t = 1, 2, \cdots, n_x$, of $\boldsymbol{x}$, and similarly given i.i.d. samples $\boldsymbol{y}(t)$, $t = 1, 2, \cdots, n_y$, of independent $\boldsymbol{y} \in \mathbb{R}^p$, $\boldsymbol{y} \sim \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{\Sigma}_{0y})$. Form the sample covariance estimates $\hat{\boldsymbol{\Sigma}}_x = \frac{1}{n_x}\sum_{t=1}^{n_x}\boldsymbol{x}(t)\boldsymbol{x}^\top(t)$ and $\hat{\boldsymbol{\Sigma}}_y = \frac{1}{n_y}\sum_{t=1}^{n_y}\boldsymbol{y}(t)\boldsymbol{y}^\top(t)$. In [13–16] one seeks to estimate $\boldsymbol{\Delta}_0 = \boldsymbol{\Omega}_{0y} - \boldsymbol{\Omega}_{0x}$ and graph $\mathcal{G}_\Delta = (V, \mathcal{E}_\Delta)$, based on $\hat{\boldsymbol{\Sigma}}_x$ and $\hat{\boldsymbol{\Sigma}}_y$, where $\boldsymbol{\Omega}_{0y} = \boldsymbol{\Sigma}_{0y}^{-1}$, $\boldsymbol{\Omega}_{0x} = \boldsymbol{\Sigma}_{0x}^{-1}$. In [14] (see also [15, Sec. 2.1]), the following convex D-trace loss function is used

$$L_r(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y) = \frac{1}{2}\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_x\boldsymbol{\Delta}\hat{\boldsymbol{\Sigma}}_y\boldsymbol{\Delta}^\top) - \mathrm{tr}(\boldsymbol{\Delta}(\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}_y)) \quad (1)$$

where D-trace refers to difference-in-trace loss function, a term coined in [22] in the context of graphical model estimation. The function $L_r(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_{0x}, \boldsymbol{\Sigma}_{0y})$ is strictly convex in $\boldsymbol{\Delta}$ and has a unique minimum at $\boldsymbol{\Delta}_0 = \boldsymbol{\Omega}_{0y} - \boldsymbol{\Omega}_{0x}$ [14, 15]. When one uses sample covariances, $\boldsymbol{\Delta}$ is estimated by minimizing a lasso-penalized D-trace loss function [13–16] (group-lasso in [17]).

2.1. Proper Complex Gaussian Vectors

Consider complex-valued $\boldsymbol{x} \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{\Sigma}_{0x})$ and $\boldsymbol{y} \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{\Sigma}_{0y})$ with $\boldsymbol{\Sigma}_{0x} \succ \boldsymbol{0}$, $\boldsymbol{\Sigma}_{0y} \succ \boldsymbol{0}$. We need to estimate $\boldsymbol{\Delta}_0 = \boldsymbol{\Omega}_{0y} - \boldsymbol{\Omega}_{0x}$. Consider the real-valued cost

$$\begin{aligned} L(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y) = &\frac{1}{2}\Big(\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_x\boldsymbol{\Delta}\hat{\boldsymbol{\Sigma}}_y\boldsymbol{\Delta}^H) + \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_x^*\boldsymbol{\Delta}^*\hat{\boldsymbol{\Sigma}}_y^*\boldsymbol{\Delta}^\top)\Big) \\ &- \mathrm{tr}\Big(\boldsymbol{\Delta}(\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}_y) + \boldsymbol{\Delta}^*(\hat{\boldsymbol{\Sigma}}_x^* - \hat{\boldsymbol{\Sigma}}_y^*)\Big) \end{aligned} \quad (2)$$

with $\hat{\boldsymbol{\Sigma}}_x = \frac{1}{n_x}\sum_{t=1}^{n_x}\boldsymbol{x}(t)\boldsymbol{x}^H(t)$ and $\hat{\boldsymbol{\Sigma}}_y = \frac{1}{n_y}\sum_{t=1}^{n_y}\boldsymbol{y}(t)\boldsymbol{y}^H(t)$. Using Wirtinger calculus, we find $\boldsymbol{0} = \frac{\partial L}{\partial \boldsymbol{\Delta}^*} = \hat{\boldsymbol{\Sigma}}_x\boldsymbol{\Delta}\hat{\boldsymbol{\Sigma}}_y - (\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}_y)$, implying $L(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_{0x}, \boldsymbol{\Sigma}_{0y})$ has a minimum at $\boldsymbol{\Delta}_0 = \boldsymbol{\Omega}_{0y} - \boldsymbol{\Omega}_{0x}$. Define $\boldsymbol{\theta} = [\mathrm{vec}(\boldsymbol{\Delta})^\top\ \mathrm{vec}(\boldsymbol{\Delta})^H]^\top \in \mathbb{C}^{2p^2}$. Then using $\mathrm{tr}(\boldsymbol{A}^\top\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}^\top) = \mathrm{vec}(\boldsymbol{A})^\top(\boldsymbol{D}\otimes\boldsymbol{B})\mathrm{vec}(\boldsymbol{C})$, we have $L(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_{0x}, \boldsymbol{\Sigma}_{0y}) = \frac{1}{2}\boldsymbol{\theta}^H\mathcal{H}\boldsymbol{\theta} - \boldsymbol{\theta}^H\boldsymbol{b}$ where

$$\mathcal{H} = \begin{bmatrix} \boldsymbol{\Sigma}_{0y}^* \otimes \boldsymbol{\Sigma}_{0x} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{0y} \otimes \boldsymbol{\Sigma}_{0x}^* \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} \mathrm{vec}(\boldsymbol{\Sigma}_{0x} - \boldsymbol{\Sigma}_{0y}) \\ \mathrm{vec}(\boldsymbol{\Sigma}_{0x}^* - \boldsymbol{\Sigma}_{0y}^*) \end{bmatrix}.$$

As in the real case, $L(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_{0x}, \boldsymbol{\Sigma}_{0y})$ is strictly convex in $\boldsymbol{\Delta}$ and the minimum at $\boldsymbol{\Delta}_0$ is unique since the Hessian $\mathcal{H} \succ \boldsymbol{0}$. For $\lambda > 0$, define the lasso-penalized D-trace loss

$$L_\lambda(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y) = L(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y) + \lambda\sum_{k,\ell=1}^{p}|\Delta_{k\ell}|. \quad (3)$$

We seek $\hat{\boldsymbol{\Delta}} = \arg\min_{\boldsymbol{\Delta}} L_\lambda(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y)$.

2.2. Optimization

Similar to [15] (also [14]), we use an alternating direction method of multipliers (ADMM) approach [23] with variable splitting. Using variable splitting, consider

$$\min_{\boldsymbol{\Delta}, \boldsymbol{W}}\Big\{L(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y) + \lambda\sum_{k,\ell=1}^{p}|W_{k\ell}|\Big\} \text{ subject to } \boldsymbol{\Delta} = \boldsymbol{W}. \quad (4)$$

The scaled augmented Lagrangian for this problem is [23]

$$L_\rho = L(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y) + \lambda\sum_{k,\ell=1}^{p}|W_{k\ell}| + \frac{\rho}{2}\|\boldsymbol{\Delta} - \boldsymbol{W} + \boldsymbol{U}\|_F^2 \quad (5)$$

where $\boldsymbol{U}$ is the dual variable, and $\rho > 0$ is the penalty parameter. Given the ith iteration results $\boldsymbol{\Delta}^{(i)}, \boldsymbol{W}^{(i)}, \boldsymbol{U}^{(i)}$, in the $(i+1)$st iteration, the algorithm executes the following 3 updates:

(a) $\boldsymbol{\Delta}^{(i+1)} \leftarrow \arg\min_{\boldsymbol{\Delta}} L_a(\boldsymbol{\Delta})$, $L_a(\boldsymbol{\Delta}) := L(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y) + \frac{\rho}{2}\|\boldsymbol{\Delta} - \boldsymbol{W}^{(i)} + \boldsymbol{U}^{(i)}\|_F^2$

(b) $\boldsymbol{W}^{(i+1)} \leftarrow \arg\min_{\boldsymbol{W}} L_b(\boldsymbol{W})$, $L_b(\boldsymbol{W}) := \lambda\sum_{k,\ell=1}^{p}|W_{k\ell}| + \frac{\rho}{2}\|\boldsymbol{\Delta}^{(i+1)} - \boldsymbol{W} + \boldsymbol{U}^{(i)}\|_F^2$

(c) $\boldsymbol{U}^{(i+1)} \leftarrow \boldsymbol{U}^{(i)} + \Big(\boldsymbol{\Delta}^{(i+1)} - \boldsymbol{W}^{(i+1)}\Big)$

Update (a): Differentiate $L_a(\boldsymbol{\Delta})$ w.r.t. $\boldsymbol{\Delta}^*$ to obtain $\boldsymbol{0} = \frac{\partial L_a(\boldsymbol{\Delta})}{\partial \boldsymbol{\Delta}^*} = \hat{\boldsymbol{\Sigma}}_x\boldsymbol{\Delta}\hat{\boldsymbol{\Sigma}}_y - (\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}_y) + \frac{\rho}{2}(\boldsymbol{\Delta} - \boldsymbol{W}^{(i)} + \boldsymbol{U}^{(i)})$. Hence,

$$(\hat{\boldsymbol{\Sigma}}_y^* \otimes \hat{\boldsymbol{\Sigma}}_x + \frac{\rho}{2}\boldsymbol{I})\mathrm{vec}(\boldsymbol{\Delta}) = \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}_y + \frac{\rho}{2}(\boldsymbol{W}^{(i)} - \boldsymbol{U}^{(i)})) \quad (6)$$

where we used $\mathrm{vec}(\boldsymbol{A}\boldsymbol{Y}\boldsymbol{B}) = (\boldsymbol{B}^\top \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{Y})$. Direct matrix inversion solution of (6) requires inversion of a $p^2 \times p^2$ matrix. A computationally cheaper solution follows as in [14, 15], but for Hermitian matrices. Carry out eigendecomposition of $\hat{\boldsymbol{\Sigma}}_x$ and $\hat{\boldsymbol{\Sigma}}_y$ as $\hat{\boldsymbol{\Sigma}}_x = \boldsymbol{Q}_x\boldsymbol{D}_x\boldsymbol{Q}_x^H$, $\boldsymbol{Q}_x\boldsymbol{Q}_x^H = \boldsymbol{I}$ and $\hat{\boldsymbol{\Sigma}}_y = \boldsymbol{Q}_y\boldsymbol{D}_y\boldsymbol{Q}_y^H$, $\boldsymbol{Q}_y\boldsymbol{Q}_y^H = \boldsymbol{I}$, where $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ are diagonal matrices. Then $\hat{\boldsymbol{\Delta}}$ that minimizes $L_a(\boldsymbol{\Delta})$ is given by

$$\hat{\boldsymbol{\Delta}} = \boldsymbol{Q}_x\Big[\boldsymbol{D} \circ [\boldsymbol{Q}_x^H\big(\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}_y + \frac{\rho}{2}(\boldsymbol{W}^{(i)} - \boldsymbol{U}^{(i)})\big)\boldsymbol{Q}_y]\Big]\boldsymbol{Q}_y^H \quad (7)$$

where the symbol $\circ$ denotes the Hadamard product and $\boldsymbol{D} \in \mathbb{R}^{p\times p}$ organizes the diagonal of $(\boldsymbol{D}_y \otimes \boldsymbol{D}_x + \frac{\rho}{2}\boldsymbol{I})^{-1}$ in a matrix with $D_{jk} = 1/([\boldsymbol{D}_x]_{jj}[\boldsymbol{D}_y]_{kk} + \frac{\rho}{2})$. Note that the eigendecomposition of $\hat{\boldsymbol{\Sigma}}_x$ and $\hat{\boldsymbol{\Sigma}}_y$ has to be done only once. Thus

$$\boldsymbol{\Delta}^{(i+1)} = \boldsymbol{Q}_x\Big[\boldsymbol{D} \circ [\boldsymbol{Q}_x^H p\big(\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}_y + \frac{\rho}{2}(\boldsymbol{W}^{(i)} - \boldsymbol{U}^{(i)})\big)\boldsymbol{Q}_y]\Big]\boldsymbol{Q}_y^H \quad (8)$$

Update (b): Here we have the lasso solution [12, Lemma 1]

$$W_{k\ell}^{(i+1)} = \Big(1 - \frac{(\lambda/\rho)}{|[\boldsymbol{\Delta}^{(i+1)} + \boldsymbol{U}^{(i)}]_{k\ell}|}\Big)_+[\boldsymbol{\Delta}^{(i+1)} + \boldsymbol{U}^{(i)}]_{k\ell} \quad (9)$$

where $(a)_+ = \max(0, a)$. It results from separable optimization of $L_b(\boldsymbol{W}) = \sum_{k,\ell=1}^{p} \{\lambda|W_{k\ell}| + \frac{\rho}{2}|\Delta_{k\ell}^{(i+1)} - W_{k\ell} + U_{k\ell}^{(i)}|^2\}$.

Convergence. A stopping (convergence) criterion following [23, Sec. 3.3.1] can be devised. The stopping criterion is based on primal and dual residuals being small where, in our case, at $(i+1)$st iteration, the primal residual is given by $\boldsymbol{\Delta}^{(i+1)} - \boldsymbol{W}^{(i+1)}$ and the dual residual by $\rho(\boldsymbol{W}^{(i+1)} - \boldsymbol{W}^{(i)})$. Convergence criterion is met when the norms of these residuals are below some threshold. The objective function $L_\lambda(\boldsymbol{\Delta}, \hat{\boldsymbol{\Sigma}}_x, \hat{\boldsymbol{\Sigma}}_y)$, given by (3), is strictly convex. It is also closed, proper and lower semi-continuous. Hence, for any fixed $\rho > 0$, the ADMM algorithm is guaranteed to converge [23, Sec. 3.2], in the sense that we have primal residual convergence to 0, dual residual convergence to 0, and objective function convergence to the optimal value.

3. DIFFERENTIAL TIME SERIES GRAPHS

We will address the problem via a frequency-domain formulation. For simplicity, we take $n_x = n_y = n$. Given $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ for $t = 1, 2, \cdots, n$, define their respective (normalized) DFTs

$$\boldsymbol{d}_x(f_m) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{x}(t) \exp\left(-j2\pi f_m(t-1)\right), \tag{10}$$

$$\boldsymbol{d}_y(f_m) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{y}(t) \exp\left(-j2\pi f_m(t-1)\right), \tag{11}$$

$$f_m = \frac{m}{n},\ m = 0, 1, \cdots, n-1. \tag{12}$$

The set of complex-valued random vectors $\{\boldsymbol{d}_x(f_m), \boldsymbol{d}_y(f_m)\}_{m=0}^{n/2}$ is a sufficient statistic for any inference problem based on data set $\{\boldsymbol{x}(t), \boldsymbol{y}(t)\}_{t=1}^{n}$ [12]. Suppose $\boldsymbol{S}_x(f_k)$ is locally smooth, so that $\boldsymbol{S}_x(f_k)$ is (approx.) constant over $K = 2m_t + 1$ consecutive frequency points f_m's, $m_t > 0$. Pick $M = \lfloor(\frac{n}{2} - m_t - 1)/K\rfloor$ and $\tilde{f}_k = ((k-1)K + m_t + 1)/n$ for $k = 1, 2, \cdots, M$ yielding M equally spaced frequencies $\tilde{f}_k$ in the interval $(0, 0.5)$. It turns out that for "large" n, the DFT $\boldsymbol{d}_x(f_m)$ is a complex-valued proper (i.e., circularly symmetric) Gaussian vector $\sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{S}_x(f_m))$, and (mutually) independent for $m = 1, 2, \cdots, (n/2)-1$, ($n$ even) [24, Theorem 4.4.1], though not identically distributed, so long as the autocorrelation function of $\boldsymbol{x}(t)$ is absolutely summable. Similar comments apply to $\boldsymbol{d}_y(f_m)$. By local smoothness, $\mathbf{S}_x(\tilde{f}_{k,\ell}) = \mathbf{S}_x(\tilde{f}_k)$ for $\ell = -m_t, -m_t+1, \cdots, m_t$, where $\tilde{f}_{k,\ell} = \frac{(k-1)K+m_t+1+\ell}{n}$. Define

$$\hat{\boldsymbol{S}}_{xk} = \frac{1}{K} \sum_{\ell=-m_t}^{m_t} \boldsymbol{d}_x(\tilde{f}_{k,\ell}) \boldsymbol{d}_x^H(\tilde{f}_{k,\ell}) \tag{13}$$

$$\hat{\boldsymbol{S}}_{yk} = \frac{1}{K} \sum_{\ell=-m_t}^{m_t} \boldsymbol{d}_y(\tilde{f}_{k,\ell}) \boldsymbol{d}_y^H(\tilde{f}_{k,\ell}) \tag{14}$$

where $\hat{\boldsymbol{S}}_{xk}$ and $\hat{\boldsymbol{S}}_{yk}$ represent PSD estimators at frequency $\tilde{f}_k$ using unweighted frequency-domain smoothing [24]. By local smoothness, $\boldsymbol{d}_x(\tilde{f}_{k,\ell}) \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{S}_x(\tilde{f}_k),)$ and $\boldsymbol{d}_y(\tilde{f}_{k,\ell}) \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{S}_y(\tilde{f}_k),)$.

Henceforth we will denote the true values of $\boldsymbol{S}_x(\tilde{f}_k)$ and $\boldsymbol{S}_y(\tilde{f}_k)$ as $\boldsymbol{S}_{0xk}$ and $\boldsymbol{S}_{0yk}$, respectively, with their respective sample estimates $\hat{\boldsymbol{S}}_{xk}$ and $\hat{\boldsymbol{S}}_{yk}$, $k = 1, 2, \cdots, M$. Let $\boldsymbol{\Delta}_{0k} = \boldsymbol{S}_{0yk}^{-1} - \boldsymbol{S}_{0xk}^{-1}$. By local smoothness assumption, we apply the approach of Sec. 2.1 to estimate the differences $\boldsymbol{\Delta}_{0k}$, $k = 1, 2, \cdots, M$, together with a group-lasso penalty to enforce a common edgeset across the M frequency points. Define

$$\tilde{\boldsymbol{\Delta}} = [\boldsymbol{\Delta}_1,\ \boldsymbol{\Delta}_2,\ \cdots,\ \boldsymbol{\Delta}_M] \in \mathbb{C}^{p^2 M}, \tag{15}$$

$$\boldsymbol{\Delta}^{(ij)} = \big[[\boldsymbol{\Delta}_1]_{ij},\ [\boldsymbol{\Delta}_2]_{ij},\ \cdots,\ [\boldsymbol{\Delta}_M]_{ij}\big] \in \mathbb{C}^M. \tag{16}$$

We propose to estimate $\boldsymbol{\Delta}_k$'s by minimizing

$$L_f(\tilde{\boldsymbol{\Delta}}) = \sum_{k=1}^{M} L(\boldsymbol{\Delta}_k, \hat{\boldsymbol{S}}_{xk}, \hat{\boldsymbol{S}}_{yk}) + \lambda \sum_{i,j=1}^{p} \|\boldsymbol{\Delta}^{(ij)}\| \tag{17}$$

In the estimated differential graph, $\{i, j\} \in \mathcal{E}_{\hat{\Delta}} \Leftrightarrow \|\hat{\boldsymbol{\Delta}}^{(ij)}\| \neq 0$.

3.1. Optimization

As in Sec. 2.2, we use an ADMM approach. Using variable splitting, the scaled augmented Lagrangian for this problem is

$$L_\rho(\tilde{\boldsymbol{\Delta}}, \tilde{\boldsymbol{W}}, \tilde{\boldsymbol{U}}) = L_f(\tilde{\boldsymbol{\Delta}}) + \lambda \sum_{i,j=1}^{p} \|\boldsymbol{W}^{(ij)}\| + \frac{\rho}{2} \sum_{k=1}^{M} \|\boldsymbol{\Delta}_k - \boldsymbol{W}_k + \boldsymbol{U}_k\|_F^2 \tag{18}$$

where $\tilde{\boldsymbol{W}} = [\boldsymbol{W}_1\ \cdots\ \boldsymbol{W}_M]$, $\tilde{\boldsymbol{U}} = [\boldsymbol{U}_1\ \cdots\ \boldsymbol{U}_M]$ is the dual variable, and $\rho > 0$ is the penalty parameter. Given the results $\tilde{\boldsymbol{\Delta}}^{(m)}, \tilde{\boldsymbol{W}}^{(m)}, \tilde{\boldsymbol{U}}^{(m)}$ of the mth iteration, in the $(m+1)$st iteration, an ADMM algorithm executes the following three updates:

(a) $\tilde{\boldsymbol{\Delta}}^{(m+1)} \leftarrow \arg\min_{\tilde{\boldsymbol{\Delta}}} \sum_{k=1}^{M} L_{ak}(\boldsymbol{\Delta}_k)$, $L_{ak}(\boldsymbol{\Delta}_k) = L(\boldsymbol{\Delta}_k, \hat{\boldsymbol{S}}_{xk}, \hat{\boldsymbol{S}}_{yk}) + \frac{\rho}{2}\|\boldsymbol{\Delta}_k - \boldsymbol{W}_k^{(m)} + \boldsymbol{U}_k^{(m)}\|_F^2$.

(b) $\tilde{\boldsymbol{W}}^{(m+1)} \leftarrow \arg\min_{\tilde{\boldsymbol{W}}} L_b(\tilde{\boldsymbol{W}})$, $L_b(\tilde{\boldsymbol{W}}) = \lambda \sum_{i,j=1}^{p} \|\boldsymbol{W}^{(ij)}\| + \frac{\rho}{2} \sum_{k=1}^{M} \|\boldsymbol{\Delta}_k^{(m+1)} - \boldsymbol{W}_k + \boldsymbol{U}_k^{(m)}\|_F^2$

(c) $\tilde{\boldsymbol{U}}^{(m+1)} \leftarrow \tilde{\boldsymbol{U}}^{(m)} + \tilde{\boldsymbol{\Delta}}^{(m+1)} - \tilde{\boldsymbol{W}}^{(m+1)}$.

Update (a): Optimization in step (a) is separable in $\boldsymbol{\Delta}_k$, and the solution discussed in Sec. 2.2 applies. Therefore, with notational changes, for $k = 1, 2, \cdots, M$, we have

$$\boldsymbol{\Delta}_k^{(m+1)} = \boldsymbol{Q}_{xk}\Big[\boldsymbol{D}^{(k)} \circ \big[\boldsymbol{Q}_{xk}^H(\hat{\boldsymbol{S}}_{xk} - \hat{\boldsymbol{S}}_{yk} + \frac{\rho}{2}(\boldsymbol{W}_k^{(m)} - \boldsymbol{U}_k^{(m)}))\boldsymbol{Q}_{yk}\big]\Big]\boldsymbol{Q}_{yk}^H \tag{19}$$

where we have the eigendecomposition of $\hat{\boldsymbol{S}}_{xk}$ and $\hat{\boldsymbol{S}}_{yk}$ as $\hat{\boldsymbol{S}}_{xk} = \boldsymbol{Q}_{xk}\boldsymbol{D}_{xk}\boldsymbol{Q}_{xk}^H$ and $\hat{\boldsymbol{S}}_{yk} = \boldsymbol{Q}_{yk}\boldsymbol{D}_{yk}\boldsymbol{Q}_{yk}^H$, and $\boldsymbol{D}^{(k)} \in \mathbb{R}^{p\times p}$ organizes the diagonal of $(\boldsymbol{D}_{yk} \otimes \boldsymbol{D}_{xk} + \frac{\rho}{2}\boldsymbol{I})^{-1}$ in a matrix with $[\boldsymbol{D}^{(k)}]_{ij} = 1/([\boldsymbol{D}_{xk}]_{ii}[\boldsymbol{D}_{yk}]_{ii} + \frac{\rho}{2})$.

Update (b): Optimization in step (b) is separable in $\boldsymbol{W}^{(ij)}$, and the group lasso solution [12, Lemma 1] applies. With $\boldsymbol{A}_k = \boldsymbol{\Delta}_k^{(m+1)} + \boldsymbol{U}_k^{(m)}$, and for $k = 1, \cdots, M$ and $i, j = 1, \cdots, p$,

$$[\boldsymbol{W}_k^{(m+1)}]_{ij} = \left(1 - \frac{\lambda}{\rho\|\boldsymbol{A}^{(ij)}\|}\right)_+ [\boldsymbol{A}_k]_{ij}, \tag{20}$$

$$\text{where } \boldsymbol{A}^{(ij)} = \big[[\boldsymbol{A}_1]_{ij}, \cdots, [\boldsymbol{A}_M]_{ij}\big] \in \mathbb{C}^M. \tag{21}$$

4. THEORETICAL ANALYSIS

Here we analyze consistency of $\hat{\boldsymbol{\Delta}}$ by following the approach of [26]. The i.i.d. results of [14,15,22] follow the method of [25] which

requires an irrepresentability condition that we do not impose. Define the true differential edgeset

$$\mathcal{E}_{\Delta_0} = \Big\{ \{i,j\} \ : \ [\boldsymbol{S}_{0y}^{-1}(f) - \boldsymbol{S}_{0x}^{-1}(f)]_{ij} \not\equiv 0, \\ i \neq j, \ 0 \leq f \leq 0.5 \Big\}, \quad s = |\mathcal{E}_{\Delta_0}| \,. \tag{22}$$

Let $\boldsymbol{R}_{xx}(\tau) = \mathbb{E}\{\boldsymbol{x}(t+\tau)\boldsymbol{x}^T(t)\}$ and $\boldsymbol{R}_{yy}(\tau) = \mathbb{E}\{\boldsymbol{y}(t+\tau)\boldsymbol{y}^T(t)\}$. In the rest of this section, we allow p, $K = 2m_t + 1$, M, s and λ to be a functions of sample size n, denoted as p_n, K_n, M_n, s_n and λ_n, respectively.

Define

$$M_0 = \max \Big\{ \max_{f\in[0,0.5]} \max_{ij} |[\boldsymbol{S}_{0x}(f)]_{ij}| \,, \\ \max_{f\in[0,0.5]} \max_{ij} |[\boldsymbol{S}_{0y}(f)]_{ij}| \Big\}, \tag{23}$$

$$\phi_{0,min} = \min \big\{ \min_{f\in[0,0.5]} \phi_{\min}(\boldsymbol{S}_{0x}(f)) \,, \min_{f\in[0,0.5]} \phi_{\min}(\boldsymbol{S}_{0y}(f)) \big\}, \tag{24}$$

$$M_d = \max_{f\in[0,0.5]} \max_{ij} \Big| [\boldsymbol{S}_{0y}^{-1}(f) - \boldsymbol{S}_{0x}^{-1}(f)]_{ij} \Big| \,, \tag{25}$$

$$C_0 = 80 \max_{\ell,f} \Big\{ [\boldsymbol{S}_{0x}(f)]_{\ell\ell}, [\boldsymbol{S}_{0y}(f)]_{\ell\ell} \Big\} \\ \times \sqrt{2\big(\ln(16 p_n^{\tau} M_n)/\ln(p_n)\big)} \tag{26}$$

where $\tau > 2$.

Let $\hat{\tilde{\boldsymbol{\Delta}}} = \arg\min_{\tilde{\boldsymbol{\Delta}}} L_f(\tilde{\boldsymbol{\Delta}})$.

Theorem 1 : Assume that $\sum_{\tau=-\infty}^{\infty} |[\boldsymbol{R}_{xx}(\tau)]_{k\ell}| < \infty$ and $\sum_{\tau=-\infty}^{\infty} |[\boldsymbol{R}_{yy}(\tau)]_{k\ell}| < \infty$ for every $k, \ell \in [p]$, Under (22), if

$$\lambda_n \geq 2\sqrt{M_n}(3M_0 s_n M_d + 2) C_0 \sqrt{\frac{\ln(p_n)}{K_n}} \,, \tag{27}$$

$$K_n > \Big(\frac{96 M_n M_0 s_n}{\phi_{0,min}}\Big)^2 C_0^2 \ln(p_n) \,, \tag{28}$$

then with probability $> 1 - 2/p_n^{\tau-2}$, for any $\tau > 2$, we have

$$\|\hat{\tilde{\boldsymbol{\Delta}}} - \tilde{\boldsymbol{\Delta}}_0\|_F \leq \frac{12\sqrt{s_n}\,\lambda_n}{\phi_{0,min}} \quad \bullet \tag{29}$$

The proof of Theorem 1 is omitted for lack of space.

Remark 1: *Convergence Rate*. If M_0, M_d, $\phi_{0,min}$ and C_0 stay bounded with increasing sample size n, we have $\|\hat{\tilde{\boldsymbol{\Delta}}} - \tilde{\boldsymbol{\Delta}}_0\|_F = \mathcal{O}_P(s_n^{1.5}\sqrt{M_n \ln(p_n)/K_n})$. Therefore, for $\|\hat{\tilde{\boldsymbol{\Delta}}} - \tilde{\boldsymbol{\Delta}}_0\|_F \to 0$ as $n \to \infty$, we must have $s_n^{1.5}\sqrt{M_n \ln(p_n)/K_n} \to 0$. $\square$

5. NUMERICAL EXAMPLE

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 x-data are generated using a vector autoregressive (VAR) model of order 3. Consider community q, $q = 1, 2, \cdots, 16$. Then $\boldsymbol{x}^{(q)}(t) \in \mathbb{R}^8$ is generated as $\boldsymbol{x}^{(q)}(t) = \sum_{i=1}^{3} \boldsymbol{A}_i^{(q)} \boldsymbol{x}^{(q)}(t-i) + \boldsymbol{w}^{(q)}(t)$. Only 15% of entries of $\boldsymbol{A}_i^{(q)}$'s are nonzero and the nonzero elements are independently and uniformly distributed over $[-0.6, 0.6]$. We then check if the VAR(3) model is stable with all eigenvalues of the companion matrix ≤ 0.95 in magnitude; if not, we re-draw randomly till this condition is fulfilled. The overall data $\boldsymbol{x}(t)$ is given by $\boldsymbol{x}(t) = [\,\boldsymbol{x}^{(1)\top}(t) \ \cdots \ \boldsymbol{x}^{(16)\top}(t)\,]^\top \in \mathbb{R}^p$ with $\boldsymbol{w}^{(q)}(t)$ as i.i.d. zero-mean Gaussian with identity covariance matrix. To generate y-data, we randomly eliminate one of the 16 clusters of $\boldsymbol{x}(t)$ and replace it with an independently generated $\boldsymbol{y}^{(q)}(t)$ mimicking generation of $\boldsymbol{x}^{(q)}(t)$. First 100 samples are discarded to eliminate transients. This set-up leads to a differential time series graph with one cluster difference (8^2 edges out of 128^2 edges). We generate $n = n_x = n_y$ observations for $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$, with $n \in \{512, 2048, 4096\}$.

Simulation results based on 100 runs are shown in Fig. 1. By changing the penalty parameter λ and determining the resulting edges over 100 runs, we calculated the true positive rate (TPR) which calculates true edges correctly detected ($\|\hat{\boldsymbol{\Delta}}^{(ij)}\| \neq 0$ and $\|\tilde{\boldsymbol{\Delta}}_0^{(ij)}\| \neq 0$), and false positive rate 1-TNR (where TNR is the true negative rate) which are the edges $\{i,j\}$ for which $\|\hat{\boldsymbol{\Delta}}^{(ij)}\| \neq 0$ but $\|\tilde{\boldsymbol{\Delta}}_0^{(ij)}\| = 0$. (Estimated $\hat{\boldsymbol{\Delta}}_k$'s are not necessarily Hermitian. We use $(\hat{\boldsymbol{\Delta}}_k + \hat{\boldsymbol{\Delta}}_k^H)/2$ as the Hermitian estimate.) The receiver operating characteristic (ROC) is shown in Fig. 1 for our proposed approach (labeled "DTS") as well as for an approach that assumes the data is i.i.d. (labeled "IID"), based on [15] which minimizes lasso-penalized (1) based on difference of precision matrices. For the proposed approach we used $M = 2$ and $K = 127, 511, 1023$ for $n = 512, 2048, 4096$, respectively. It is seen from Fig. 1 that our approach significantly outperforms the IID approach, yielding much higher TPR for a given 1-TNR.

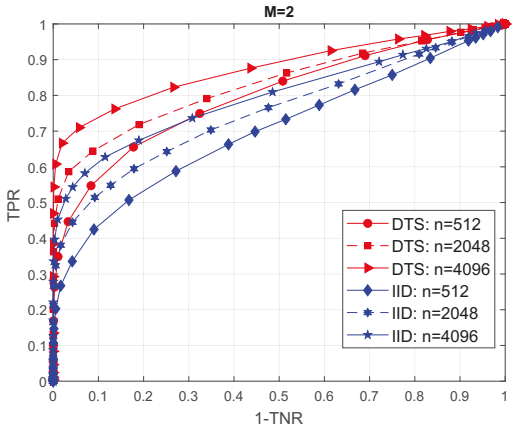

Fig. 1: ROC curves: DTS is the proposed approach and IID is the approach of [15]. TPR=true positive rate, TNR=true negative rate

6. CONCLUSIONS

We addressed the problem of estimating differences in two time series Gaussian graphical models (TSGGMs) which are known to have similar structure, via estimation of the difference in their IPSD's. We analyzed a group lasso penalized D-trace loss function approach in the frequency domain. An ADMM algorithm was presented to optimize the convex objective function. Theoretical analysis establishing consistency of the estimator of IPSD difference in high-dimensional settings was performed. We illustrated our approach via a numerical example.

7. 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] M. Eichler, "Graphical modelling of multivariate time series," *Probability Theory and Related Fields*, vol. 153, issue 1-2, pp. 233-268, June 2012.

[4] 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.

[5] J. Friedman, T. Hastie and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432-441, July 2008.

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

[7] 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.

[8] 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.

[9] A. Jung, "Learning the conditional independence structure of stationary time series: A multitask learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5677-5690, Nov. 1, 2015.

[10] 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.

[11] 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, pp. 589-593.

[12] J.K. Tugnait, "On sparse high-dimensional graphical model learning for dependent time series," *Signal Processing*, vol. 197, pp. 1-18, Aug. 2022, Article 108539.

[13] Y. Wu, T. Li, X. Liu and L.| Chen, "Differential network inference via the fused D-trace loss with cross variables," *Electronic J. Statistics*, vol. 14, pp. 1269-1301, 2020.

[14] H. Yuan, R. Xi, C. Chen and M. Deng, "Differential network analysis via lasso penalized D-trace loss," *Biometrika*, vol. 104, pp. 755-770, 2017.

[15] B. Jiang, X. Wang and C. Leng, "A direct approach for sparse quadratic discriminant analysis," *J. Machine Learning Research*, vol. 19, pp. 1-37, 2018.

[16] Z. Tang, Z. Yu and C. Wang, "A fast iteraive algorithm for high-dimensional differential network," *Computational Statistics*, vol. 35, pp. 95-109, 2020.

[17] J.K. Tugnait, "Estimation of high-dimensional differential graphs from multi-attribute data," in *Proc. 2023 IEEE Intern. Conf. Acoustics, Speech & Signal Processing (ICASSP 2023)*, pp. 1-5, Rhodes Island, Greece, June 4-9, 2023

[18] B. Zhao, Y.S. Wang and M. Kolar, "FuDGE: A method to estimate a functional differential graph in a high-dimensional setting," *J. Machine Learning Research*, vol. 23, pp. 1-82, 2022.

[19] S.D. Zhao, T.T. Cai and H. Li, "Direct estimation of differential networks," *Biometrika*, vol. 101, pp. 253-268, June 2014.

[20] E. Belilovsky, G. Varoquaux and M.B. Blaschko, "Hypothesis testing for differences in Gaussian graphical models: Applications to brain connectivity," *Advances in Neural Information Processing Systems (NIPS 2016)*, vol. 29, Dec. 2016.

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

[22] T. Zhang and H. Zou, "Sparse precision matrix estimation via lasso penalized D-trace loss," *Biometrika*, vol. 101, pp. 103-120, 2014.

[23] 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.

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

[25] 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.

[26] S.N. Negahban, P. Ravikumar, M.J. Wainwright and B. Yu, "A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers," *Statistical Science*, vol. 27, No. 4, pp. 538-557, 2012.