GRAPHICAL LASSO FOR HIGH-DIMENSIONAL COMPLEX GAUSSIAN GRAPHICAL MODEL SELECTION

Jitendra K. Tugnait

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

ABSTRACT

We consider the problem of inferring the conditional independence graph (CIG) of both proper and improper, complex-valued, high-dimensional multivariate Gaussian vectors. A p-variate complex Gaussian graphical model (CGGM) associated with an undirected graph with p vertices is defined as the family of complex Gaussian distributions that obey the conditional independence restrictions implied by the edge set of the graph. For real random vectors, considerable body of work exists, whereas that on proper complex Gaussian graphical models (PCGGMs) is sparse, while that on ICGGMs is non-existent. In this paper, we present a graphical lasso based penalized log-likelihood approach for both PCGGMs and ICGGMs. An alternating minimization algorithm is used to optimize the objective functions. Numerical examples illustrate the proposed algorithms.

Keywords: Complex Gaussian graphical models; improper complex Gaussian graphical models; undirected graph; graphical lasso.

1. INTRODUCTION

Graphical models provide a powerful tool for analyzing multivariate data [1–3]. In a typical setting of an undirected graphical model, the conditional dependency structure among p (real-valued) random variables $x_1, x_1, \cdots, x_p$, ($\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_p]^\top$), is represented using an undirected graph $\mathcal{G} = (V, \mathcal{E})$, where $V = \{1, 2, \cdots, p\} = [p]$ is the set of p nodes corresponding to the p random variables x_is, and $\mathcal{E} \subseteq [p] \times [p]$ is the set of undirected edges describing conditional dependencies among the components of $\mathbf{x}$. The graph $\mathcal{G}$ then is a conditional independence graph (CIG) where there is no edge between nodes i and j (i.e., $\{i, j\} \notin \mathcal{E}$) if and only if (iff) x_i and x_j are conditionally independent given the remaining p-2 variables x_ℓ, $\ell \in [p]$, $\ell \neq i$, $\ell \neq j$.

Real-valued Gaussian graphical models (RGGMs) are CIGs where $\mathbf{x}$ is multivariate Gaussian. Suppose $\mathbf{x}$ has positive-definite covariance matrix $\boldsymbol{\Sigma}$ with inverse covariance matrix (also known as precision matrix or concentration 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 for real-valued $\mathbf{x}$ have been extensively studied, and found to be useful in a wide variety of applications [4–9]. Given N samples of $\mathbf{x}$, prior work can be classified into two categories: *low-dimensional setting* where $p \ll N$ and p is "small," and *high-dimensional setting* where $p \gg 1$ and/or N is of the order of p. In low-dimensions a focus is on edge exclusion tests [1, 2, 24–26], to decide which set of edges out of total $p(p-1)/2$ edges are in $\mathcal{E}$. In high-dimensions, the edge exclusion testing approach is not feasible or practical, so one estimates $\boldsymbol{\Omega}$ directly under some sparsity constraints; see [5, 8, 9, 29, 30] for RGGMs.

For complex-valued $\mathbf{x}$, only the monograph [10], and more recently [11], have studied such models, where $\mathbf{x}$ is assumed to be proper complex and of low dimension. In the context of frequency-domain formulation of graphical modeling of real-valued time series in high-dimensional settings, proper complex-valued graphical models have been considered in [12, 14] using a neighborhood regression scheme, and in the form of penalized log-likelihood (graphical lasso) in [13]. A significant application of graphical modeling of real-valued random vectors has been for analysis of fMRI data [15], to provide insights into the functional connectivity of different brain regions [16]. It has been shown in [17,18] that complex-valued fMRI data yields improved sensitivity in fMRI analysis compared to real-valued fMRI data where the phase information, although collected, is simply discarded. It turns our that fMRI data is improper complex-valued [19, 20]. Prior work on ICGGMs is non-existent.

A complex-valued random vector $\mathbf{x}$ is said to be circular if $e^{j\theta}\mathbf{x}$ has the same probability distribution as $\mathbf{x}$ for all real θ [21], [22, p. 53]. A complex-valued random vector $\mathbf{x}$ is said to be proper if $E\{\mathbf{x}\mathbf{x}^\top\} = E\{\mathbf{x}\}E\{\mathbf{x}^\top\}$ [22, p. 35], [23]. A circular $\mathbf{x}$ is proper but converse is not necessarily true. A complex zero-mean Gaussian $\mathbf{x}$ is proper if and only if it is circular [22, p. 53].

Relation to Prior Work: Prior work on ICGGMs is non-existent. PCGGMs are considered implicitly (in the context of real time series) in [12–14]. As in [5,13], we use penalized log-likelihood as objective function, but solve it via alternating minimization based on variable splitting and penalty, unlike [5, 13], who use ADMM (alternating direction method of multipliers) [33].

In this paper, we present a graphical lasso based penalized log-likelihood approach for both PCGGMs and ICGGMs. We review some existing results in Sec. 2, and develop necessary and sufficient conditions for x_i and x_j to be conditionally independent in ICGGMs in Sec. 3. In Sec. 4, we propose penalized log-likelihood based objective functions for PCGGMs based on graphical lasso, and for ICGGMs based on sparse group graphical lasso, respectively. An alternating minimization algorithm is used to optimize the objective functions. Numerical examples in Sec. 5 illustrate the proposed algorithms.

2. PRELIMINARIES AND BACKGROUND

2.1. Real-Valued GGM (RGGM)

An RGGM associated with a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is defined as the family of p-variate real-valued Gaussian random vectors $\mathbf{x} \in \mathbb{R}^p$, $p = |V|$, that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$. We take $V = [p]$ and $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_p]^\top$, and in the corresponding graph $\mathcal{G}$, each variable x_i

This work was supported by NSF Grant CCF-1617610.

is represented by a node (i in V), and associations between variables x_i and x_j are represented by edges between nodes i and j of $\mathcal{G}$. Edge $\{i,j\} \in \mathcal{E}$ iff x_i and x_j are conditionally dependent given the remaining p-2 variables.

RGGMs are CIGs where $\mathbf{x}$ is real multivariate Gaussian. Suppose $\mathbf{x}$ is zero-mean with positive-definite covariance matrix $\boldsymbol{\Sigma} \succ 0$, denoted by $\mathbf{x} \sim \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma})$, then its probability density function (pdf) is

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{e^{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}}}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}} = \frac{|\boldsymbol{\Omega}|^{1/2}}{(2\pi)^{p/2}} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{x}\mathbf{x}^\top \boldsymbol{\Omega})} \tag{1}$$

where we do not distinguish between a random vector/matrix and the values taken by them in our notation (for simplicity). It is known that Ω_{ij}, the (i,j)-th element of $\boldsymbol{\Omega}$ $(= \boldsymbol{\Sigma}^{-1})$, is zero iff x_i and x_j are conditionally independent [1, Proposition 5.2].

2.2. Proper Complex-Valued GGM (PCGGM)

Similarly, a PCGGM associated with an undirected graph $\mathcal{G} = (V, \mathcal{E})$ is defined as the family of p-variate proper complex-valued Gaussian vectors $\mathbf{x} \in \mathbb{C}^p$, $p = |V|$, that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$. We take $\mathbf{x}$ to be a complex, proper, Gaussian random vector, with zero mean and covariance $\boldsymbol{\Sigma}$, i.e., $\mathbf{x} \sim \mathcal{N}_c(0, \boldsymbol{\Sigma})$, and we assume that $\boldsymbol{\Sigma}$ is positive definite. Therefore, $\mathbb{E}\{\mathbf{x}\} = 0$, $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \boldsymbol{\Sigma} \in \mathbb{C}^{p\times p}$, and $\mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} = \mathbf{0}$ [22, p. 35]. The pdf of $\mathbf{x}$ is

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{e^{-\mathbf{x}^H \boldsymbol{\Sigma}^{-1}\mathbf{x}}}{\pi^p\,|\boldsymbol{\Sigma}|} = \frac{|\boldsymbol{\Omega}|}{\pi^p} e^{-\mathrm{tr}(\mathbf{x}\mathbf{x}^H \boldsymbol{\Omega})} \tag{2}$$

It is fairly common to refer to proper complex Gaussian vectors as just complex Gaussian vectors [10]. However, since we are also interested in the case where for $\mathbf{x} \in \mathbb{C}^p$, we do not necessarily have $\mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} = \mathbf{0}$, following [22], we explicitly distinguish between proper complex Gaussian vectors ($\mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} = \mathbf{0}$) and improper complex Gaussian vectors ($\mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} \neq \mathbf{0}$). A key result for PCGGMs is that $\Omega_{ij} = [\boldsymbol{\Omega}]_{ij}$, the (i,j)-th element of $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$, is zero iff x_i and x_j are conditionally independent [10, Theorem 7.1].

2.3. Improper Complex Gaussian Vectors

Given $\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_i \in \mathbb{C}^p$, with real part $\mathbf{x}_r$ and imaginary part $\mathbf{x}_i$, define the augmented complex vector $\mathbf{y}$ and the real vector $\mathbf{z}$ as

$$\mathbf{y} = \begin{bmatrix} \mathbf{x}^\top & \mathbf{x}^H \end{bmatrix}^\top, \quad \mathbf{z} = \begin{bmatrix} \mathbf{x}_r^\top & \mathbf{x}_i^\top \end{bmatrix}^\top. \tag{3}$$

The pdf of an improper complex Gaussian $\mathbf{x}$ is defined in terms of that of the augmented $\mathbf{z}$ or $\mathbf{y}$ [22, Sec. 2.3.1]. Assume $\mathbb{E}\{\mathbf{x}\} = \mathbf{0}$, and define $\mathbf{R}_{uv} = \mathbb{E}\{\mathbf{u}\mathbf{v}^\top\}$ for (zero-mean) $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$, and define the covariance matrix $\mathbf{R}_{uv} = \mathbb{E}\{\mathbf{u}\mathbf{v}^H\}$, and the complementary covariance matrix $\tilde{\mathbf{R}}_{uv} = \mathbb{E}\{\mathbf{u}\mathbf{v}^\top\}$ [22, Sec. 2.2], for zero-mean $\mathbf{u}, \mathbf{v} \in \mathbb{C}^p$. Then we have $\mathbf{z} \sim \mathcal{N}_r(\mathbf{0}, \mathbf{R}_{zz})$ where

$$\mathbf{R}_{zz} = \begin{bmatrix} \mathbf{R}_{x_r x_r} & \mathbf{R}_{x_r x_i} \\ \mathbf{R}_{x_i x_r} & \mathbf{R}_{x_i x_i} \end{bmatrix}, \; \mathbf{R}_{yy} = \begin{bmatrix} \mathbf{R}_{xx} & \tilde{\mathbf{R}}_{xx} \\ \tilde{\mathbf{R}}_{xx}^* & \mathbf{R}_{xx}^* \end{bmatrix} = \mathbf{R}_{yy}^H. \tag{4}$$

Since $\mathbf{z} \sim \mathcal{N}_r(\mathbf{0}, \mathbf{R}_{zz})$, its pdf is given by (assuming $\mathbf{R}_{zz} \succ \mathbf{0}$)

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{1}{(2\pi)^{2p/2}\,|\mathbf{R}_{zz}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{z}^\top \mathbf{R}_{zz}^{-1}\mathbf{z}\right). \tag{5}$$

One can also express (5) as [22, Sec. 2.3.1]

$$f_{\mathbf{x}}(\mathbf{x}) := f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{\pi^p\,|\mathbf{R}_{yy}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}^H \mathbf{R}_{yy}^{-1}\mathbf{y}\right). \tag{6}$$

For proper $\mathbf{x}$, $\tilde{\mathbf{R}}_{xx} = \mathbf{0}$, and (6) reduces to

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{e^{-\frac{1}{2}\mathbf{x}^H \mathbf{R}_{xx}^{-1}\mathbf{x} - \frac{1}{2}\left(\mathbf{x}^H \mathbf{R}_{xx}^{-1}\mathbf{x}\right)^*}}{\pi^p\,|\mathbf{R}_{xx}|^{1/2}\,|\mathbf{R}_{xx}^*|^{1/2}}. \tag{7}$$

Since $\mathbf{R}_{xx} = \mathbf{R}_{xx}^H$, $|\mathbf{R}_{xx}| = |\mathbf{R}_{xx}^*|$, $(\mathbf{x}^H \mathbf{R}_{xx}^{-1}\mathbf{x})^* = \mathbf{x}^H \mathbf{R}_{xx}^{-1}\mathbf{x}$, for proper $\mathbf{x}$, (7) has the familiar form used in (2).

3. IMPROPER COMPLEX GAUSSIAN GRAPHICAL MODEL

An ICGGM associated with a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is defined as the family of p-variate complex-valued improper Gaussian random vectors $\mathbf{x} \in \mathbb{C}^p$, $p = |V|$, $V = [p]$, that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$. However, since an improper x_j is specified in terms of two random variables (real and imaginary parts of the random variable, or the variable and its complex conjugate), in an ICGGM, in fact, each node corresponds to two random variables $\mathrm{real}(x_j)$ and $\mathrm{imag}(x_j)$. In the notation of (6), conditional independence of improper x_j and x_k is equivalent to

$$f_{x_j, x_k \,|\, \mathbf{x}_{\tilde{V}}}(x_j, x_k \,|\, \mathbf{x}_{\tilde{V}}) = f_{x_j \,|\, \mathbf{x}_{\tilde{V}}}(x_j \,|\, \mathbf{x}_{\tilde{V}}) f_{x_k \,|\, \mathbf{x}_{\tilde{V}}}(x_k \,|\, \mathbf{x}_{\tilde{V}}) \tag{8}$$

where $\tilde{V} = V \backslash \{j,k\}$ and $\mathbf{x}_{\tilde{V}} = [x_i]_{i \in \tilde{V}}$ =a column vector of dimension $|\tilde{V}|$, composed of components of $\mathbf{x}$ associated with nodes in $\tilde{V}$. By integrating out the unwanted variables on both sides of (8), we then have

$$f_{u,v \,|\, \mathbf{x}_{\tilde{V}}}(u, v \,|\, \mathbf{x}_{\tilde{V}}) = f_{u \,|\, \mathbf{x}_{\tilde{V}}}(u \,|\, \mathbf{x}_{\tilde{V}}) f_{v \,|\, \mathbf{x}_{\tilde{V}}}(v \,|\, \mathbf{x}_{\tilde{V}}) \tag{9}$$

for any real scalars u, v that satisfy $u \in \{\mathrm{real}(x_j),\ \mathrm{imag}(x_j)\},\ v \in \{\mathrm{real}(x_k),\ \mathrm{imag}(x_k)\}$.

In order to exploit some results in [1] pertaining to real-valued GGMs (RGGMs) for testing the validity of (9) for a given ICGGM, we will use the representation $\mathbf{z}$ in (3), and exploit a larger RGGM corresponding to the given ICGGM $\mathcal{G} = (V, \mathcal{E})$. Consider an RGGM $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ associated with ICGGM $\mathcal{G} = (V, \mathcal{E})$, where $\bar{V} = [2p]$, vertix j for $1 \le j \le p$ represents the real part of improper x_j, $\mathrm{real}(x_j)$, and vertix $j+p$ represents $\mathrm{imag}(x_j)$. For a given edge $\{j,k\}$, define the set of four edges $\bar{\mathcal{E}}^{(jk)}$ as

$$\bar{\mathcal{E}}^{(jk)} = \big\{\{j,k\}, \{j+p,k\}, \{j,k+p\}, \{j+p,k+p\}\big\}. \tag{10}$$

If the edge $\{j,k\} \notin \mathcal{E}$, then we have edges $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$, implying relations (9) for all four possible values of pair (u,v). Since pairwise Markov property implies global Markov property for RGGMs ([1, p. 131]), we can establish that given ICGGM $\mathcal{G} = (V, \mathcal{E})$ with $\mathbf{x} \in \mathbb{C}^p$, and the associated RGGM $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ with $\mathbf{z} = [\mathrm{real}(\mathbf{x}^\top)\ \mathrm{imag}(\mathbf{x}^\top)]^\top \in \mathbb{R}^{2p}$,

$$\{j,k\} \notin \mathcal{E} \;\Leftrightarrow\; \bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset. \tag{11}$$

Let $\mathbf{R}_{zz} = \mathbb{E}\{\mathbf{z}\mathbf{z}^\top\} \succ \mathbf{0}$ and $\bar{\boldsymbol{\Omega}} = \mathbf{R}_{zz}^{-1}$. We are able to prove Lemma 1.

Lemma 1. Consider an ICGGM $\mathcal{G} = (V, \mathcal{E})$ with $\mathbf{x} \in \mathbb{C}^p$ and $V = [p]$, and the associated RGGM $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ with $\mathbf{z} = [\mathrm{real}(\mathbf{x}^\top)\ \mathrm{imag}(\mathbf{x}^\top)]^\top \in \mathbb{R}^{2p}$, $\bar{V} = [2p]$, where vertix j

for $1 \le j \le p$ represents the real part of improper x_j, $\mathrm{real}(x_j)$, and vertix $j+p$ represents $\mathrm{imag}(x_j)$. Assume that $\mathbf{R}_{zz} = \mathbb{E}\{\mathbf{z}\mathbf{z}^\top\} \succ \mathbf{0}$. Then $\forall j, k \in V$, $j \neq k$, $\{j,k\} \notin \mathcal{E}$, i.e., x_j and x_k are conditionally independent given $\mathbf{x}_{V\setminus\{j,k\}}$, iff $[\bar{\mathbf{\Omega}}]_{\ell m} = 0 \,\forall\, \{\ell, m\} \in \bar{\mathcal{E}}^{(jk)}$, where $\bar{\mathbf{\Omega}} = \mathbf{R}_{zz}^{-1}$ and $\bar{\mathcal{E}}^{(jk)}$ is defined in (10). •

Lemma 1 is the counterpart of [1, Prop. 5.2] pertaining to RGGMs, and of [10, Theorem 7.1] pertaining to PCGGMs.

4. GRAPHICAL LASSO FOR HIGH-DIMENSIONAL CGGMS

Given N samples $\mathbf{x}(t)$, $t = 0, 1, \cdots, N-1$, denoted by $\mathbf{X}$, define $\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{t=0}^{n-1} \mathbf{x}(t)\mathbf{x}^H(t)$, and similarly $\hat{\bar{\mathbf{\Sigma}}} = \frac{1}{n}\sum_{t=0}^{n-1} \mathbf{z}(t)\mathbf{z}^\top(t)$. Here we investigate the case where $p \gg 1$ and/or N is of the order of p. When p is large, it may not be feasible to test all $p(p-1)/2$ edges. When p is large in comparison to N, $\hat{\mathbf{\Sigma}}$ or $\hat{\bar{\mathbf{\Sigma}}}$ may be ill-conditioned. As noted in [5] (and by others, e.g., [8, 29]) in the context of real-valued random vectors, one may need to use penalty terms to enforce sparsity and to make the problem well-conditioned. Consider the likelihood function (2) or (7) for $\mathbf{x}$, and (5) for $\mathbf{z}$. We wish to estimate inverse covariance $\mathbf{\Omega}$ or $\bar{\mathbf{\Omega}}$. Given N i.i.d. realizations, we have the log-likelihood (up to some constants)

$$\ln f_{\mathbf{X}}(\mathbf{X}) = \frac{N}{2}\Big(\ln|\mathbf{\Omega}| + \ln|\mathbf{\Omega}^*| - \mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega}) - \mathrm{tr}(\hat{\mathbf{\Sigma}}^*\mathbf{\Omega}^*)\Big) \quad (12)$$

$$= N\Big(\ln|\mathbf{\Omega}| - \mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega})\Big), \quad (13)$$

$$\ln f_{\mathbf{Z}}(\mathbf{Z}) = N\ln|\bar{\mathbf{\Omega}}| - \mathrm{tr}(\mathbf{Z}^\top\mathbf{Z}\bar{\mathbf{\Omega}}). \quad (14)$$

As an alternative to edge exclusion tests, one possible solution is to maximize the log-likelihood (12) w.r.t. $\mathbf{\Omega}$ (or $\bar{\mathbf{\Omega}}$), and then threshold elements of estimated $\hat{\mathbf{\Omega}}$ (or $\hat{\bar{\mathbf{\Omega}}}$) to zero or nonzero. If $[\hat{\mathbf{\Omega}}]_{ij} = 0$, the edge $\{i,j\} \notin \mathcal{E}$. If $[\hat{\bar{\mathbf{\Omega}}}]_{ij} = 0\, \forall\{i,j\} \in \bar{\mathcal{E}}^{(jk)}$, the edge $\{i,j\} \notin \mathcal{E}$. In general, one would obtain inverse covariance estimates with no elements that are exactly equal to zero, and if one resorts to element-by-element thresholding, choice of threshold level is unclear.

4.1. PCGGMs

A solution, following [8] (also [29, 30]), is to impose a sparsity constraint. Instead of maximizing $\ln f_{\mathbf{X}}(\mathbf{X})$, maximize a penalized version w.r.t. Hermitian $\mathbf{\Omega} \succ \mathbf{0}$

$$L_{PC}(\mathbf{X}) = \ln f_{\mathbf{X}}(\mathbf{X}) - \lambda\,\|\mathbf{\Omega}\|_{1d}\,, \quad (15)$$

$$\|\mathbf{\Omega}\|_{1d} = \sum_{\substack{l,m=1\\ l\neq m}}^{p} \Big|[\mathbf{\Omega}]_{lm}\Big| = \ell_1 \text{ norm of } \mathbf{\Omega} \text{ without diagonal terms}$$

where $[\mathbf{\Omega}]_{lm}$ denotes the (l,m)-th element of $\mathbf{\Omega} \in \mathbb{C}^{p\times p}$, and $\lambda \ge 0$ is a tuning parameter. This is the lasso penalty, leading to the term graphical lasso [8, 29]. Notice that unlike the formulations in [8,29,30] where $\mathbf{\Omega}$ is real-valued, we have complex-valued $\mathbf{\Omega}$. So we resort to Wirtinger calculus (complex differential calculus) [22, Appendix 2], [31] coupled with corresponding definition of subdifferential/subgradients [32], to minimize convex $-L_{PC}(\mathbf{X})$ w.r.t. complex $\mathbf{\Omega}$ using the necessary and sufficient KKT conditions for a global optimum. Several approaches, such as ADMM [33, Sec. 6.5], or AM (alternating minimization) methods [34] based on variable splitting and penalty techniques, are possible for minimization. Some recent results [35] suggest that when both approaches are applicable, ADMM which requires dual variables, is inferior to the AM method which is a primal-only method, in terms of computational complexity and accuracy.

Consider minimization of $-\ln f_{\mathbf{X}}(\mathbf{X}) + \lambda\,\|\mathbf{\Omega}\|_1$ subject to $\mathbf{\Omega} \succ \mathbf{0}$. By variable splitting, we reformulate as

$$\min_{\mathbf{\Omega},\mathbf{W}}\Big\{\frac{N}{2}\big[\mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega} + \hat{\mathbf{\Sigma}}^*\mathbf{\Omega}^*) - (\ln(|\mathbf{\Omega}|) + \ln(|\mathbf{\Omega}^*|))\big] + \lambda\,\|\mathbf{W}\|_{1d}\Big\} \quad (16)$$

$$\text{subject to } \mathbf{W} = \mathbf{\Omega} \succ 0\,. \quad (17)$$

Using the penalty method, consider the *relaxed* problem ($\rho > 0$ is "large")

$$\min_{\mathbf{\Omega},\mathbf{W}}\Big\{\frac{N}{2}\mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega} + \hat{\mathbf{\Sigma}}^*\mathbf{\Omega}^*) - \frac{N}{2}(\ln(|\mathbf{\Omega}|) + \ln(|\mathbf{\Omega}^*|)) + \lambda\,\|\mathbf{W}\|_{1d} + \frac{\rho}{2}\|\mathbf{\Omega} - \mathbf{W}\|_F^2\,.\Big\} \quad (18)$$

Given the results $\mathbf{\Omega}^{(i)}, \mathbf{W}^{(i)}$, of the ith iteration, in the $(i+1)$st iteration, an AM algorithm executes the following two updates:

(a) $\mathbf{\Omega}^{(i+1)} \leftarrow \arg\min_{\mathbf{\Omega}} J_a(\mathbf{\Omega}, \mathbf{\Omega}^*)$, $J_a(\mathbf{\Omega}, \mathbf{\Omega}^*) := \frac{N}{2}\mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega} + \hat{\mathbf{\Sigma}}^*\mathbf{\Omega}^*) - \frac{N}{2}(\ln(|\mathbf{\Omega}|) + \ln(|\mathbf{\Omega}^*|)) + \frac{\rho}{2}\|\mathbf{\Omega} - \mathbf{W}^{(i)}\|_F^2$

(b) $\mathbf{W}^{(i+1)} \leftarrow \arg\min_{\mathbf{W}} J_b(\mathbf{W})$, $J_b(\mathbf{W}) := \lambda\,\|\mathbf{W}\|_{1d} + \frac{\rho}{2}\|\mathbf{\Omega}^{(i+1)} - \mathbf{W}\|_F^2$

A necessary and sufficient condition for a global optimum in update (a) is that the gradient of $J_a(\mathbf{\Omega}, \mathbf{\Omega}^*)$ w.r.t. $\mathbf{\Omega}^*$, given by (20), vanishes, with $\mathbf{\Omega} = \mathbf{\Omega}^H \succ \mathbf{0}$:

$$\mathbf{0} = \frac{\partial J_a(\mathbf{\Omega}, \mathbf{\Omega}^*)}{\partial \mathbf{\Omega}^*} = \frac{N}{2}\hat{\mathbf{\Sigma}}^H - \frac{N}{2}(\mathbf{\Omega}^H)^{-1} + \frac{\rho}{2}(\mathbf{\Omega} - \mathbf{W}^{(i)}) \quad (19)$$

$$= \frac{N}{2}\hat{\mathbf{\Sigma}} - \frac{N}{2}\mathbf{\Omega}^{-1} + \frac{\rho}{2}(\mathbf{\Omega} - \mathbf{W}^{(i)})\,. \quad (20)$$

The solution to (20) follows as in [33, Sec. 6.5]. Let $\mathbf{V}\mathbf{D}\mathbf{V}^H$ denote the eigen-decomposition of the matrix $\hat{\mathbf{\Sigma}} - (\rho/N)\mathbf{W}^{(i)}$. Then $\mathbf{\Omega}^{(i+1)} = \mathbf{V}\tilde{\mathbf{D}}\mathbf{V}^H$ where $\tilde{\mathbf{D}}$ is the diagonal matrix with mth diagonal element $\tilde{\mathbf{D}}_{mm} = (N/(2\rho))(-\mathbf{D}_{mm} + \sqrt{\mathbf{D}_{mm}^2 + 4\rho/N})$. Notice that $J_b(\mathbf{W})$ is completely separable w.r.t. each pair (j,k) of matrix elements, i.e., solve $W_{jk}^{(i+1)} \leftarrow \arg\min_{W_{jk}} J_{bjk}(W_{jk})$, for each (j,k), where $J_{bjk}(W_{jk}) := \lambda\, 1_{j\neq k}|W_{jk}| + (\rho/2)|[\mathbf{\Omega}^{(i+1)}]_{jk} - W_{jk}|^2$. A necessary and sufficient condition for a global optimum in update (b) is that the subdifferential of $J_{bjk}(W_{jk})$ w.r.t. W_{jk}^*, given by (21), must contain 0:

$$0 \in \partial J_{bjk}(W_{jk}) = 1_{j\neq k}\frac{\lambda_1}{2}t + \frac{1}{2}(W_{jk} - [\mathbf{\Omega}^{(i+1)}]_{jk}), \quad (21)$$

$$t = \begin{cases} W_{jk}/|W_{jk}| & \text{if } W_{jk} \neq 0 \\ \in \{u : |u| \le 1,\ u \in \mathbb{C}\} & \text{if } W_{jk} = 0 \end{cases} \quad (22)$$

This leads to the soft-thresholding solution ($(a)_+ := \max(0, a)$)

$$[\mathbf{W}^{(i+1)}]_{jk} = \begin{cases} [\mathbf{\Omega}^{(i+1)}]_{jj} & \text{if } j = k \\ S([\mathbf{\Omega}^{(i+1)}]_{jk}, \lambda_1/\rho) & \text{if } j \neq k \end{cases} \quad (23)$$

where

$$S([\mathbf{\Omega}^{(i+1)}]_{jk}, \lambda_1/\rho) := \left(1 - \frac{\lambda_1}{\rho|[\mathbf{\Omega}^{(i+1)}]_{jk}|}\right)_+ [\mathbf{\Omega}^{(i+1)}]_{jk}\,. \quad (24)$$

By [34, Theorem 3.7], the iterative solution sequence $\mathbf{\Omega}^{(i)}, \mathbf{W}^{(i)}$ generated by the AM method converges to a global minimum.

4.2. ICGGMs

In view of (11), we propose using a sparse-group lasso penalty to maximize $L_{IC}(\mathbf{Z})$ w.r.t. $\bar{\mathbf{\Omega}}$

$$L_{IC}(\mathbf{Z}) = \ln f_{\mathbf{Z}}(\mathbf{Z}) - P(\bar{\mathbf{\Omega}})\,, \tag{25}$$

$$P(\bar{\mathbf{\Omega}}) := \lambda_1 \sum_{\substack{l,m=1\\ l\neq m}}^{2p} |[\bar{\mathbf{\Omega}}]_{lm}| + \lambda_2 \sum_{\substack{l,m=1\\ l\neq m}}^{p} \|\mathbf{v}_{lm}(\bar{\mathbf{\Omega}})\|_2\,, \tag{26}$$

$$\mathbf{v}_{lm}(\bar{\mathbf{\Omega}}) := \left[[\bar{\mathbf{\Omega}}]_{lm}\ [\bar{\mathbf{\Omega}}]_{(l+p)m}\ [\bar{\mathbf{\Omega}}]_{l(m+p)}\ [\bar{\mathbf{\Omega}}]_{(l+p)(m+p)}\right]^\top \tag{27}$$

where $\lambda_1,\ \lambda_2 \geq 0$ are tuning parameters. In (25), an ℓ_1 penalty term is applied to each off-diagonal element of $\bar{\mathbf{\Omega}}$, and to the group of four in (11). Recall that now we are dealing with real-valued augmented random vector $\mathbf{z} \in \mathbb{R}^{2p}$. Using variable splitting and the penalty method, consider

$$\min_{\bar{\mathbf{\Omega}},\mathbf{W}} \left\{ N\mathrm{tr}(\hat{\bar{\mathbf{\Sigma}}}\bar{\mathbf{\Omega}}) - N\ln(|\bar{\mathbf{\Omega}}|) + P(\mathbf{W}) + \frac{\rho}{2}\|\bar{\mathbf{\Omega}} - \mathbf{W}\|_F^2 \right\} \tag{28}$$

involving an AM algorithm with the following two updates:

(a) $\bar{\mathbf{\Omega}}^{(i+1)} \leftarrow \arg\min_{\bar{\mathbf{\Omega}}} J_a(\bar{\mathbf{\Omega}})$, $\quad J_a(\bar{\mathbf{\Omega}}) := N\mathrm{tr}(\hat{\bar{\mathbf{\Sigma}}}\bar{\mathbf{\Omega}}) - N\ln(|\bar{\mathbf{\Omega}}|) + \frac{\rho}{2}\|\bar{\mathbf{\Omega}} - \mathbf{W}^{(i)}\|_F^2$

(b) $\mathbf{W}^{(i+1)} \leftarrow \arg\min_{\mathbf{W}} J_b(\mathbf{W})$, $\quad J_b(\mathbf{W}) := P(\mathbf{W}) + \frac{\rho}{2}\|\bar{\mathbf{\Omega}}^{(i+1)} - \mathbf{W}\|_F^2$

The solution to update (a) is as in Sec. 4.1. Let $\mathbf{VDV}^H$ denote the eigen-decomposition of the matrix $\hat{\bar{\mathbf{\Sigma}}} - (\rho/N)\mathbf{W}^{(i)}$. Then $\bar{\mathbf{\Omega}}^{(i+1)} = \mathbf{V}\tilde{\mathbf{D}}\mathbf{V}^H$ where $\tilde{\mathbf{D}}$ is the diagonal matrix with mth diagonal element $\tilde{\mathbf{D}}_{mm} = (N/(2\rho))(-\mathbf{D}_{mm} + \sqrt{\mathbf{D}_{mm}^2 + 4\rho/N})$. Following [36] (see also [5]), the solution to update (b) is given by

$$[\mathbf{W}^{(i+1)}]_{jk} = \begin{cases} [\bar{\mathbf{\Omega}}^{(i+1)}]_{jj} & \text{if } j = k \\ S([\bar{\mathbf{\Omega}}^{(i+1)}]_{jk}, \lambda_1/\rho)\left(1 - \frac{\lambda_2}{\rho\|\mathbf{v}_{jk}(\mathbf{S}(\bar{\mathbf{\Omega}}^{(i+1)}, \lambda_1/\rho))\|_2}\right)_+ & \text{if } j \neq k \end{cases}$$

where $[\mathbf{S}(\mathbf{\Omega}, \alpha)]_{jm} = S([\mathbf{\Omega}]_{jm}, \alpha)$, defined in (24).

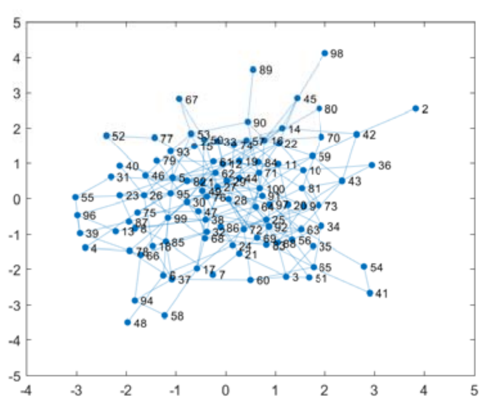

Fig. 1. *A typical realization of true CIG, p =100.*

5. SIMULATIONS

We start with $p \times p$ $\check{\mathbf{\Omega}}$ with all diagonal elements set to 1 and off-diagonal elements equal to 0.5. With probability q and independently, we set off-diagonal elements in the upper triangle of $\check{\mathbf{\Omega}}$ to zero (taking care to set the corresponding elements in lower triangle also to zero so that the resulting matrix $\hat{\mathbf{\Omega}}$ is symmetric). Now set $\mathbf{\Omega} = \hat{\mathbf{\Omega}} + \beta\mathbf{I}$ with β picked to make $\mathbf{\Omega}$ positive definite. This choice of $\mathbf{\Omega}$ is similar to one of the examples in [37]. Then approximately $100q\%$ entries of $\mathbf{\Omega}$ are null. With $\mathbf{\Phi\Phi}^H = \mathbf{\Omega}^{-1}$, we generate $\mathbf{x} = \mathbf{\Phi w}$ with $\mathbf{w} \in \mathbb{C}^p$ as zero-mean, improper Gaussian with independent components. Let $\mathbf{v} \sim \mathcal{N}_c(0, \mathbf{I})$. Then we set w_l, the lth component of $\mathbf{w}$, as $w_l = \mathrm{real}(v_l) + j(0.9 \times \mathrm{real}(v_l) + 0.2 \times \mathrm{imag}(v_l))$, yielding improper complex Gaussian $\mathbf{w}$. We generate N i.i.d. observations from $\mathbf{x}$, with $q = 0.95$ to have approximately 95% entries of $\bar{\mathbf{\Omega}}$ (corresponding to $\mathbf{z} \in \mathbb{R}^{2p}$), as null, leading to a sparse graph with only 5% of connected edges.

A typical realization of the graph is in Fig. 1 where only the nodes which are connected to at least one other node, are shown. The ROC curves are in Fig. 2 based on 100 runs. While the true graph is an ICGGM, one can pretend that it can be modeled as a PCGGM and then also use the method of Sec. 4.1. Results of both methods, Sec. 4.1 and Sec. 4.2, are shown in Fig. 2. To generate the ROC, having fitted $\bar{\mathbf{\Omega}}$ to the data, one compares against a threshold η to determine if a given edge $\{i, j\}$ is connected or not: If $\|\mathbf{v}_{ij}(\bar{\mathbf{\Omega}})\|_2 \geq \eta$, then $\{i, j\} \in \mathcal{E}$, else $\{i, j\} \notin \mathcal{E}$, and for PCGGM and $\mathbf{\Omega}$, we pick $\{i, j\} \in \mathcal{E}$ if $|[\mathbf{\Omega}]_{ij}| = |\Omega_{ij}| \geq \eta$, else $\{i, j\} \notin \mathcal{E}$. It is seen from Fig. 2 that using impropriety properties does significantly improve detection performance.

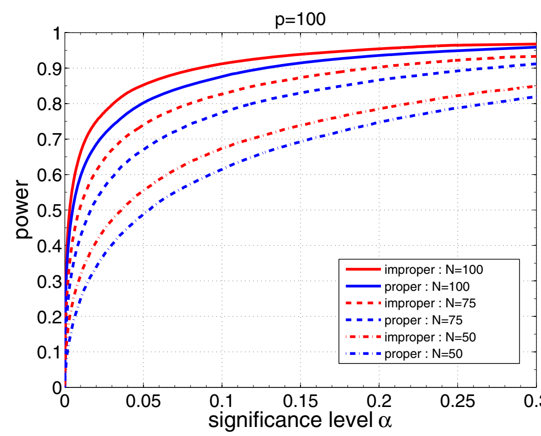

Fig. 2. *Improper Gaussian, p =100, N =50, 75 or 100. ROC curves based on treating the data as (i) proper complex Gaussian, using (18) with λ =0.03, ρ =10, (labeled "proper"), or (ii) improper complex Gaussian, using (28) with λ_1 =0.03, λ_2 =0.1, ρ =10, (labeled "improper"). Based on 100 runs.*

6. CONCLUSIONS

We considered the problem of inferring the conditional independence graph of complex-valued, both proper and improper, multivariate Gaussian vectors in high dimensions. We first developed necessary and sufficient conditions on elements of the inverse covariance matrix $\bar{\mathbf{\Omega}}$ of the real $2p$-vector associated with improper complex $\mathbf{x}$, for x_i and x_j to be conditionally independent. Then we proposed penalized log-likelihood based objective functions for PCGGMs based on graphical lasso, and for ICGGMs based on sparse group graphical lasso, respectively. An alternating minimization algorithm was used to optimize the objective functions. Numerical examples were presented to illustrate the proposed algorithms.

7. REFERENCES

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

[2] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*. New York: Wiley, 1990.

[3] A.P. Dawid and S.L. Lauritzen, “Hyper Markov laws in the statistical analysis of decomposable graphical models,” *Ann. Statist.*, vol. 2, no. 3, pp. 1272-1317, 1993.

[4] A.O. Hero, III, and B. Rajaratnam, “Foundational principles for large-scale inference: Illustrations through correlation mining,” *Proc. IEEE*, vol. 64, pp. 93-110, Jan. 2016.

[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, 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, P. London, M. Fazel, D. Witten and S.I. Lee, “Node-based learning of multiple Gaussian graphical models,” *J. Machine Learning Research*, vol. 15, 2014.

[10] H.H. Andersen, M. Hojbjerre, D. Sorensen, and P.S. Eriksen, *Linear and Graphical Models for the Multivariate Complex Normal Distribution*, Lecture Notes in Statistics, vol. 101. New York: Springer-Verlag, 1995.

[11] J.K. Tugnait, “An edge exclusion test for complex Gaussian graphical model selection," in *Proc. 2018 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 678-682, Freiburg, Germany, June 10-13, 2018.

[12] A. Jung, R. Heckel, H. Bölcskei, and F. Hlawatsch, “Compressive nonparametric graphical model selection for time series,” in *Proc. IEEE ICASSP-2014*, Florence, Italy, May 2014.

[13] 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.

[14] A. Jung, “Learning the conditional independence structure of stationary time series: A multitask learning approach,” *IEEE Trans. Signal Process.*, vol. 63, pp. 5677-5690, Nov. 2015.

[15] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, M.S. Lehéricy, J. Doyon and H. Benali, “Partial correlation for functional brain interactivity investigation in functional MRI,” *NeuroImage*, vol. 32, no. 1, pp. 228-237, 2006.

[16] O. Sporns, *Networks of the Brain*. MIT Press, 2010.

[17] M.C. Kociuba and D.B. Rowe, “Complex-valued time-series correlation increases sensitivity in fMRI analyis,” *Magnetic Resonance Imaging*, vol. 34, pp. 765-770, 2016.

[18] D.B. Rowe, “Magnitude and phase signal detection in complex-valued fMRI data,” *Magnetic Resonance in Medicine*, vol. 62, pp. 1356-7, 2009.

[19] T. Adali and V.D. Calhoun, “Complex ICA of brain imaging data,” *IEEE Signal Process. Mag.* vol. 24, pp. 136-139, 2007.

[20] M.-C. Yu, Q.-H. Lin, L.D. Kuang, X.-F. Gong, F. Cong and V.D. Calhoun, “ICA of full complex-valued fMRI data using phase information of spatial maps,” *J. Neuroscience Methods*, vol. 249, pp. 75-91, 2015.

[21] B. Picinbono, “On circularity,” *IEEE Trans. Signal Proc.*, vol. 42, pp. 3473-3482, Dec. 1994.

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

[23] F.D. Neeser and J.L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Inf. Theory*, vol. 39, pp. 1293-1302, July 1993.

[24] P.W.F. Smith and J. Whittaker, “Edge exclusion tests for graphical Gaussian models,” in *Learning in Graphical Models*, M.I. Jordan (Ed), MIT Press, 1998.

[25] M. Drton and M.D. Perlman, “Model selection for Gaussian concentration graphs,” *Biometrika*, vol. 91, pp. 591-602, 2004.

[26] M. Drton and M.D. Perlman, “Multiple testing and error control in Gaussian graphical model selection,” *Statistical Science*, vol. 22, pp. 430-439, 2007.

[27] Y. Matsuda, “A test statistic for graphical modelling of multivariate time series,” *Biometrika*, vol. 93, pp. 399-409, 2006.

[28] 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.

[29] 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.

[30] O. Banerjee, L.E. Ghaoui and A. d'Aspremont, “Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data,” *J. Machine Learning Research*, vol. 9, pp. 485-516, 2008.

[31] A. Hjorungnes and D. Gesbert, “Complex-valued matrix differentiation: techniques and key results,” *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2740-2746, June 2007.

[32] E. Ollila, “Direction of arrival estimation using robust complex Lasso,” in *Proc. 2016 10th European Conf. Antennas Propag. (EuCAP)*, Davos, April 2016, pp. 1-5.

[33] 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.

[34] A. Beck, “On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes,” *SIAM J. Optimization*, vol. 25, No. 1, pp. 185-209, 2015.

[35] P. Zheng, T. Askham, S.L. Brunton, J.N. Kutz and A.V. Aravkin, “Sparse relaxed regularized regression: SR3,” *arXiv:1807.05411v2 [stat.ML]*, 18 Jul 2018.

[36] 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.

[37] 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.