Edge Exclusion Tests for Improper Complex Gaussian Graphical Model Selection

Jitendra K. Tugnait, *Life Fellow, IEEE*

***Abstract*—We consider the problem of inferring the conditional independence graph (CIG) of improper complex-valued Gaussian vectors. A p-variate improper complex Gaussian graphical model associated with an undirected graph with p vertices is defined as the family of improper 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 where one first tests for exclusion of each edge from the saturated model, and then infers the CIG. Prior work on proper complex Gaussian graphical models is sparse, while that on improper complex Gaussian graphical models is non-existent. In this paper, we propose and analyze a generalized likelihood ratio test (GLRT) based edge exclusion test statistic for improper complex Gaussian graphical models. The null distribution of the test statistic is specified explicitly to allow analytical calculation of the test threshold. An alternative computationally fast version of the GLRT statistic is also derived. Simulation examples are presented to illustrate the proposed statistic.**

***Index Terms*—Improper complex Gaussian graphical models; undirected graph; generalized likelihood ratio test.**

I. Introduction

GRAPHICAL models provide a powerful tool for analyzing multivariate data [1]–[3]. They are statistical models "embodying a collection of marginal and conditional independences which may be summarized by means of a graph. Such models combine richness in modeling, clarity of interpretation and ease of analysis" [4]. 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 $V = \{1, 2, \cdots, p\} = [1, p]$ is the set of p nodes corresponding to the p random variables x_is, and $\mathcal{E} \subseteq [1, p] \times [1, p]$ 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) of $\mathbf{x}$ where there is no edge between nodes i and j if and only if (iff) x_i and x_j are conditionally independent given the remaining p-2 variables x_ℓ, $\ell \in [1, p]$, $\ell \neq i$, $\ell \neq j$ [2, p. 60]. Thus, edge $\{i, j\} \in \mathcal{E}$ iff x_i and x_j are conditionally dependent, and edge $\{i, j\} \notin \mathcal{E}$ iff x_i and x_j are conditionally independent.

Suppose real-valued $\mathbf{x}$ is multivariate Gaussian with positive-definite covariance matrix $\mathbf{\Sigma}$ and inverse covariance matrix (also known as precision matrix or concentration matrix) $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$. Then Ω_{ij}, the (i, j)-th element of $\mathbf{\Omega}$, is zero iff x_i and x_j are conditionally independent [1, Proposition 5.2], [30, Relation (1.2)]. A CIG $\mathcal{G}$ and real multivariate Gaussian $\mathbf{x}$ specify a real-valued Gaussian graphical model for $\mathbf{x}$ where the distribution of $\mathbf{x}$ obeys the conditional independence restrictions of CIG $\mathcal{G}$ [1, Sec. 5.2], [2, p. 165]. Such models for real-valued $\mathbf{x}$ have been extensively studied, and found to be useful in a wide variety of applications [5]–[11]. For complex-valued $\mathbf{x}$, only the monograph [12], and more recently [13], have studied such models, where $\mathbf{x}$ is assumed to be proper complex. Proper complex-valued graphical models also arise in frequency-domain formulation of graphical modeling of real-valued time series [14]. Such models have been considered in [15] in low-dimensional settings (vector dimension $p \ll N$, number of vector samples), and in [16]–[18] in high-dimensional settings (p comparable to N).

A significant application of graphical modeling of real-valued random vectors has been for analysis of functional magnetic resonance imaging (fMRI) data [19], [20], to provide insights into the functional connectivity of different brain regions [21]. It is stated in [22] that "In fMRI, the measured blood oxygen level dependent (BOLD) signal to detect neural activity is spatially Fourier encoded ... The BOLD fluctuations are measured as a complex-valued fMRI signal over time in the spatial frequency-domain, then the k-space readout is reconstructed with the inverse Fourier transform (IFT). Before the statistical analysis of the fMRI data, the phase portion of the data is generally discarded, despite physiologically useful information contained in the phase." It has been shown in [22], [23] that complex-valued fMRI data yields improved sensitivity in fMRI analysis. It turns out that fMRI data can be accurately modeled as improper complex-valued [24], [25]. Potential application to fMRI related problems is a motivation for investigation of improper complex Gaussian graphical models. Prior work on improper complex Gaussian graphical models is non-existent.

The results of [22], [23] show improvements in performance in detection of neural activity in regions-of-interest when both magnitude and phase fMRI data are used compared to using just the magnitude data. In a different application concerning independent component analysis of fMRI data, where the objective is to separate a linear mixture of independent signals into its independent components, [24], [25] show improved performance when both magnitude and phase fMRI data are used compared to usage of just the magnitude data. Although real-valued fMRI data-based graphical modeling has been investigated in [19], [20], no such

Manuscript received November 16, 2018; revised Feb. 18, 2019, April 3, 2019; accepted May 1, 2019. The associate editor coordinating the review of this paper and approving it for publication was Dr. Yuichi Tanaka.

J.K. Tugnait is with the Department of Electrical & Computer Engineering, 200 Broun Hall, Auburn University, Auburn, AL 36849, USA. Email: tugnajk@auburn.edu .

This work was supported by the National Science Foundation under Grant CCF-1617610.

application involving complex-valued fMRI data has been reported for graphical modeling. One possible reason for lack of such investigations could be that there are no complex-valued fMRI data sets available publicly; all public repositories (such as OpenNEURO https://openneuro.org/public/datasets) provide magnitude-only, i.e., real-valued, data.

Consider complex-valued random vectors $\mathbf{x}$ and $\mathbf{z}$. Let the superscripts $*$, $\top$ and H denote the complex conjugate, transpose and Hermitian (conjugate transpose) operations, respectively, and $\mathbb{E}$ denote the expectation operation. Denote the cross-correlation matrix of $\mathbf{x}$ and $\mathbf{z}$ as $\mathbf{R}_{xz} = \mathbb{E}\{\mathbf{x}\mathbf{z}^H\}$. Random vector $\mathbf{x}$ is said to be circular (or circularly symmetric) if $e^{j\theta}\mathbf{x}$ has the same probability distribution as $\mathbf{x}$ for all real θ [26], [27, p. 53]. A complex-valued random vector $\mathbf{x}$ is said to be proper if $\mathbf{x}$ is uncorrelated with its complex conjugate, i.e., $\mathbb{E}\{\mathbf{x}(\mathbf{x}^*)^H\} = \mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} = \mathbb{E}\{\mathbf{x}\}\mathbb{E}\{\mathbf{x}^\top\}$ [27, p. 35], [28]. A circular $\mathbf{x}$ is proper but converse is not necessarily true [27, p. 53]. A complex zero-mean Gaussian $\mathbf{x}$ is proper if and only if it is circular [27, p. 53].

Edge exclusion tests for real Gaussian graphical models are in [1] (and others), and that for proper complex Gaussian graphical models are in [12], [13]. Prior work on improper complex Gaussian graphical models is non-existent. In this paper we exploit [1, Prop. 5.14] after casting improper complex Gaussian graphical models as real Gaussian graphical models with multiple constraints.

In this paper, we propose and analyze a generalized likelihood ratio test (GLRT) based edge exclusion test statistic for improper complex Gaussian graphical models. In graphical model selection, one needs to decide if a given edge $\{i, j\}$, for $1 \leq i < j \leq p$, in the associated graph $\mathcal{G} = (V, \mathcal{E})$, is in $\mathcal{E}$ or not in $\mathcal{E}$. Since $\Omega_{ij} = 0 \Leftrightarrow \{i, j\} \notin \mathcal{E}$ for real Gaussian graphical models and proper complex Gaussian graphical models, its estimate based on the random vector sample (or a related statistic) can be used to test if $\{i, j\} \notin \mathcal{E}$; hence the term edge exclusion test [29]–[32]. There are $p(p-1)/2$ distinct edges in an undirected graph with p nodes. So for graphical model selection, one has to determine which set of edges out of total $p(p-1)/2$ edges belong to $\mathcal{E}$, which requires multiple testing based on control of overall significance level [31]. In this paper we focus on edge exclusion testing, and do not consider multiple testing issues.

The rest of the paper is organized as follows. In Sec. II-A we review some basic definitions and useful results from [1], [12] concerning simple undirected graphs, that we use later. In Sec. II-B we recall some existing results on edge exclusion tests for selection of real Gaussian graphical models and proper complex Gaussian graphical models. These tests do not apply to improper complex Gaussian graphical models, but their formulation offers a road map for the formulation of the multiple-edge binary hypothesis testing problem formulated in Sec. III, and solved in Sec. IV. In Sec. III-A we review improper complex Gaussian vectors, and in Sec. III-B, we define an improper complex Gaussian graphical model with p nodes in terms of an associated real Gaussian graphical model with $2p$ nodes, using an augmented real vector $\mathbf{y}$ comprised of $\mathrm{real}(\mathbf{x})$ and $\mathrm{imag}(\mathbf{x})$. A binary hypothesis testing problem is formulated in Sec. III-C exploiting the concept of "correct graph" [33], [34], and using the associated real Gaussian graphical model, where we test for exclusion/inclusion of four edges in the real Gaussian graphical model corresponding to a single edge in associated improper complex Gaussian graphical model. A GLRT is derived in Sec. IV-B based on [1, Proposition 5.14], which is discussed in Sec. IV-A. The null distribution of the test statistic is specified explicitly in Sec. IV-C, which allows for analytical calculation of the test threshold. A significant simplification of the GLRT statistic of Sec. IV-A is presented in Sec. V, and it requires $\mathcal{O}(p^3)$ flops compared to $\mathcal{O}(p^5)$ flops needed in Sec. IV-A. Simulation results are presented in Sec. VI.

A. Notation and Terminology

We use $\mathbf{S} \succeq 0$ and $\mathbf{S} \succ 0$ to denote that symmetric (or Hermitian) $\mathbf{S}$ is positive semi-definite and positive definite, respectively. The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. For a square matrix $\mathbf{A}$, $|\mathbf{A}|$ and $\mathrm{etr}(\mathbf{A})$ denote the determinant and the exponential of the trace of $\mathbf{A}$, respectively, i.e., $\mathrm{etr}(\mathbf{A}) = \exp(\mathrm{tr}(\mathbf{A}))$, $[\mathbf{B}_k]_{i:l,j:m}$ denotes the submatrix of the matrix $\mathbf{B}_k$ comprising its rows i through l and columns j through m, $[\mathbf{B}_k]_{ij}$ is its (i, j)-th element, B_{ij} is the (i, j)-th element of $\mathbf{B}$, x_i is the i-th component of $\mathbf{x} \in \mathbb{C}^p$, and $\mathbf{I}$ is the identity matrix. The superscripts $*$, $\top$ and H denote the complex conjugate, transpose and Hermitian (conjugate transpose) operations, respectively, and $\mathbb{E}$ denotes the expectation operation. The real and imaginary parts of $\mathbf{x} \in \mathbb{C}^p$ are denoted by $\mathrm{real}(\mathbf{x})$ and $\mathrm{imag}(\mathbf{x})$, respectively.

For a set V, $|V|$ denotes its cardinality, i.e., the number of elements in V. The set of integers 1 through n is denoted by $[1, n]$. The set $\mathbb{C}_+^{p\times p}$ denotes the set of $p \times p$ positive definite Hermitian matrices, i.e., $\mathbb{C}_+^{p\times p} = \{\mathbf{S} \in \mathbb{C}^{p\times p} \,\big|\, \mathbf{S}^H = \mathbf{S} \succ 0\}$. Similarly, the set $\mathbb{R}_+^{p\times p}$ denotes the set of $p\times p$ positive definite symmetric matrices, i.e., $\mathbb{R}_+^{p\times p} = \{\mathbf{S} \in \mathbb{R}^{p\times p} \,\big|\, \mathbf{S}^\top = \mathbf{S} \succ 0\}$. The notation $y = \mathcal{O}(g(x))$ means that there exists some finite real number $b > 0$ such that $\lim_{x\to\infty} |y/g(x)| \leq b$.

The notation $\mathbf{x} \sim \mathcal{N}_r(\mathbf{m}, \mathbf{\Sigma})$ denotes a real random vector $\mathbf{x}$ that is Gaussian with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$. Also, for complex $\mathbf{x}$, $\mathbf{x} \sim \mathcal{N}_c(\mathbf{m}, \mathbf{\Sigma})$ denotes a proper complex Gaussian vector with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$. For scalar x, $x \sim \mathcal{B}(\alpha, \beta)$ denotes a beta random variable with probability density function (pdf)

$$f_x(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1, \quad (1)$$

where $\Gamma(\alpha)$ denotes the (complete) Gamma function $\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}\,dt$, and we do not distinguish between a random variable and the value taken by it. The abbreviations w.p.1., w.r.t. and i.i.d. stand for with probability one, with respect to, and independent and identically distributed, respectively.

II. Preliminaries and Background

In Sec. II-A, we review some basic definitions and useful results from [1, Chapter 2] and [12, Chapter 5] concerning simple undirected graphs, that we use later in this paper. In Sec. II-B we recall some existing results on edge exclusion tests for selection of real Gaussian graphical models and

proper complex Gaussian graphical models. These tests do not apply to improper complex Gaussian graphical models, but their formulation offers a road map for the formulation of the multiple-edge binary hypothesis testing problem formulated in Sec. III, and solved in Sec. IV.

A. Simple Undirected Graphs [1, Chapter 2], [12, Chapter 5]

A simple undirected graph $\mathcal{G}$ is a pair $(V, \mathcal{E})$, where V is a finite set of elements called vertices (or nodes), and the set $\mathcal{E} \subseteq V \times V$ of unordered pairs (called edges) of distinct vertices. We denote the edge between vertices α and β of V as $\{\alpha, \beta\}$. The set $V \backslash A$ denotes the set of all elements in V that are not in A.

A graph is said to be complete, or saturated, if $\mathcal{E} = V \times V$, i.e., if any given vertex is connected to every other vertex. A subset $A \subseteq V$ for $\mathcal{G} = (V, \mathcal{E})$ induces the subgraph $\mathcal{G}_A = (A, \mathcal{E}_A)$ where $\mathcal{E}_A = \mathcal{E} \cap (A \times A)$, i.e., $\mathcal{E}_A$ contains exactly those edges in $\mathcal{E}$ which connect vertices from A. A subset is complete if it induces a complete subgraph.

A path of length n in $\mathcal{G}$ is a sequence of distinct vertices $\alpha_0, \alpha_1, \cdots, \alpha_n$, where $\alpha_j \in V$, such that $\{\alpha_{j-1}, \alpha_j\} \in \mathcal{E}$ for all $j = 1, 2, \cdots, n$. Two subsets $A, B \subseteq V$ are said to be separated by $S \subseteq V$ if all paths from A to B go via S, i.e., the paths from a vertex in A to a vertex in B intersect S at some vertex.

Definition 1. Decomposition. A pair (A, B) of subsets of V of a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is said to form a decomposition of $\mathcal{G}$ if $V = A \cup B$, $A \cap B$ is a complete subset of V, and $A \backslash B$ and $B \backslash A$ are separated by $A \cap B$. •

Definition 2. Clique. Let $\mathcal{G} = (V, \mathcal{E})$ be a simple undirected graph. A complete subset of V which is maximal w.r.t. set inclusion is called a clique, i.e.,

$$\begin{aligned}&(C \text{ is complete and } C \subset C' \Rightarrow C' \text{ is not complete})\\ &\quad \Leftrightarrow C \text{ is a clique},\end{aligned}$$

where $C, C' \subseteq V$. •

Definition 3. Decomposability. A simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is said to be decomposable if it is complete, or if there exists a decomposition formed by proper subsets A and B of V into decomposable subgraphs $\mathcal{G}_A$ and $\mathcal{G}_B$. •

Definition 4 is from [1, pp. 14-15], as specialized to this paper.

Definition 4. Perfect sequence and running intersection property. Let $B_1, \cdots, B_m$ be a sequence of subsets of V of a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$. Let

$$\begin{aligned}H_j &= B_1 \cup \cdots \cup B_j\,,\ j \geq 1\\ S_j &= H_{j-1} \cap B_j\,,\ j \geq 2\,.\end{aligned}$$

The sequence $B_1, \cdots, B_m$ is said to be perfect if the following conditions are fulfilled:

(i) for all $i > 1$ there is a $j < i$ such that $S_i \subseteq B_j$;
(ii) the sets S_i are complete for all i.

The condition (i) is known as the running intersection property (RIP). •

We conclude this section with Lemma 1, which is [1, Proposition 2.17] and is also in [12, Chapter 5].

Lemma 1: A simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is decomposable if and only if the cliques of $\mathcal{G}$ can be ordered (or, numbered) to form a perfect sequence. •

1) Markov properties on Undirected Graphs [1, Secs. 3.2, 5.2]: Given $\mathbf{x} \in \mathbb{R}^p$ and associated simple undirected graph $\mathcal{G} = (V, \mathcal{E})$, a probability measure is said to obey the *pairwise Markov property*, relative to $\mathcal{G}$, if for any $\{j, k\} \notin \mathcal{E}$, x_j and x_k are conditionally independent [1, p. 32]. A probability measure is said to obey the *global Markov property*, relative to $\mathcal{G}$, if for any triple (A, B, S) of disjoint subsets of V such that S separates A and B in $\mathcal{G}$, random vectors $\mathbf{x}_A$ and $\mathbf{x}_B$ are conditionally independent given $\mathbf{x}_S$, where, for any $\tilde{V} \subseteq V$, we define $\mathbf{x}_{\tilde{V}} = [x_i]_{i \in \tilde{V}} =$ a column vector of dimension $|\tilde{V}|$, consisting of random variables associated with the nodes in $\tilde{V}$.

For real Gaussian graphical models, the pairwise Markov property implies the global Markov property [1, p. 131].

B. Gaussian Graphical Models: Real, and Proper Complex

A real Gaussian graphical model 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 $\mathcal{E}$. In this paper, for $|V| = p$, we take $V = [1, p]$. For $\mathbf{x} = [x_1\ x_2\ \cdots\ x_p]^\top$, in the corresponding graph $\mathcal{G}$, each variable x_i 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 $\mathcal{G}$. In a CIG $\mathcal{G} = (V, \mathcal{E})$ of $\mathbf{x}$, there is no edge between nodes i and j iff x_i and x_j are conditionally independent given the remaining $p - 2$ variables x_ℓ, $\ell \in [1, p]$, $\ell \neq i$, $\ell \neq j$. Thus, edge $\{i, j\} \in \mathcal{E}$ iff x_i and x_j are conditionally dependent, and edge $\{i, j\} \notin \mathcal{E}$ iff x_i and x_j are conditionally independent. A CIG $\mathcal{G}$ and real multivariate Gaussian $\mathbf{x}$ specify a real-valued Gaussian graphical model for $\mathbf{x}$ where the distribution of $\mathbf{x}$ obeys the conditional independence restrictions of CIG $\mathcal{G}$ [1, Sec. 5.2], [2, p. 165]. Suppose $\mathbf{x}$ has positive-definite covariance matrix $\mathbf{\Sigma}$ with inverse covariance matrix (also known as precision matrix or concentration matrix) $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$. Then Ω_{ij}, the (i, j)-th element of $\mathbf{\Omega}$, is zero iff x_i and x_j are conditionally independent [1, Proposition 5.2]. If $\mathbf{x} \sim \mathcal{N}_r(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} \succ 0$, the probability density function (pdf) of $\mathbf{x}$ is

$$\begin{aligned}f_{\mathbf{x}}(\mathbf{x}) =& \frac{1}{(2\pi)^{p/2}\,|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x}\right)\\ =& \frac{|\mathbf{\Omega}|^{1/2}}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{\Omega}\mathbf{x}\right).\end{aligned} \tag{2}$$

Similarly, a proper complex Gaussian graphical model 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, circularly symmetric (proper), Gaussian random vector, with zero mean and covariance $\mathbf{\Sigma}$, i.e., $\mathbf{x} \sim \mathcal{N}_C(0, \mathbf{\Sigma})$, and we assume that $\mathbf{\Sigma}$ is positive definite. Therefore, $\mathbb{E}\{\mathbf{x}\} = 0$, $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{\Sigma} \in \mathbb{C}^{p \times p}$, and $\mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} = \mathbf{0}$ [27, p. 35]. The pdf

of $\mathbf{x}$ is

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\pi^p |\mathbf{\Sigma}|} \exp\left(-\mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x}\right) = \frac{|\mathbf{\Omega}|}{\pi^p} \exp\left(-\mathbf{x}^H \mathbf{\Omega} \mathbf{x}\right) . \quad (3)$$

It is fairly common to refer to proper complex Gaussian vectors as just complex Gaussian vectors [12]. However, since we are interested in the case where for $\mathbf{x} \in \mathbb{C}^p$, we do not necessarily have $\mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} = 0$, following [27], 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 proper complex Gaussian graphical models is that $\Omega_{ij} = [\mathbf{\Omega}]_{ij}$, the (i,j)-th element of $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$, is zero iff x_i and x_j are conditionally independent [12, Theorem 7.1].

1) Edge Exclusion Tests for real and proper complex Gaussian graphical models: Suppose we are given N i.i.d. observations $\mathbf{x}(0), \mathbf{x}(1), \cdots, \mathbf{x}(N-1)$ of $\mathbf{x}$, where $\mathbf{x}(t) \in \mathbb{R}^p$ for real Gaussian graphical models and $\mathbf{x}(t) \in \mathbb{C}^p$ for proper complex Gaussian graphical models. In graphical model selection, one needs to decide if a given edge $\{i,j\}$, for $1 \leq i < j \leq p$, in the associated graph $\mathcal{G} = (V, \mathcal{E})$, is in $\mathcal{E}$ or not in $\mathcal{E}$. Since $\Omega_{ij} = 0 \Leftrightarrow \{i,j\} \notin \mathcal{E}$, its estimate based on the random vector sample can be used to test if $\{i,j\} \notin \mathcal{E}$; hence the term edge exclusion test [31], [32]. There are $p(p-1)/2$ distinct edges in an undirected graph with p nodes. So for graphical model selection, one has to determine which set of edges out of total $p(p-1)/2$ edges belong to $\mathcal{E}$. If $\mathbf{x}(t) \sim \mathcal{N}_r(\mathbf{0}, \mathbf{\Sigma})$, then the sample covariance matrix is $\hat{\mathbf{\Sigma}} = \frac{1}{N} \sum_{t=0}^{N-1} \mathbf{x}(t)\mathbf{x}^\top(t)$, and an estimate of its inverse is $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}}^{-1}$. In [30], [31],

$$\mathcal{L}_R = \frac{-[\hat{\mathbf{\Omega}}]_{ij}}{\sqrt{[\hat{\mathbf{\Omega}}]_{ii}[\hat{\mathbf{\Omega}}]_{jj}}} \quad (4)$$

(the empirical partial correlation coefficient) is used as a test statistic for the hypothesis testing problem with null hypothesis $\Omega_{ij} = 0 \Leftrightarrow \{i,j\} \notin \mathcal{E}$ versus the alternative $\Omega_{ij} \neq 0 \Leftrightarrow \{i,j\} \in \mathcal{E}$. The null is rejected if (4) is large:

$$|\mathcal{L}_R| \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \quad (5)$$

An alternative formulation in the form a GLRT is in [1, Sec. 5.3.3], which we discuss next.

Consider two undirected graphs $\mathcal{G} = (V, \mathcal{E})$ and $\mathcal{G}' = (V, \mathcal{E}')$ where $|V| = p$, and $\mathcal{G}'$ is a subgraph of $\mathcal{G}$, i.e., $\mathcal{E}'$ is a subset of $\mathcal{E}$. Let $\mathcal{E}$ be saturated, i.e., complete. Now suppose that we remove one edge $\{i,j\}$ from $\mathcal{E}$ to obtain $\mathcal{E}' = \mathcal{E} \backslash \{i,j\}$. Since the associated graph determines if Ω_{ij} is zero or nonzero, we will denote $\mathbf{\Omega}$ as $\mathbf{\Omega}(\mathcal{G})$ to explicitly indicate this dependence. In [1, Sec. 5.3.3], the following hypothesis testing problem is considered

$$\begin{aligned} \mathcal{H}_0 &: \ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}'),\ \mathcal{G}' = (V, \mathcal{E}'),\ \mathcal{E}' = \mathcal{E} \backslash \{i,j\} \\ \mathcal{H}_1 &: \ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}),\ \mathcal{G} = (V, \mathcal{E}),\ \mathcal{E} \text{ is saturated} , \end{aligned} \quad (6)$$

where

$$\mathbb{R}_+(\mathcal{G}) = \Big\{ \mathbf{\Omega} \in \mathbb{R}_+^{p \times p} \,\Big|\, \forall i \neq j \in V \ : \ \{i,j\} \notin \mathcal{E} \Rightarrow \ \Omega_{ij} = 0 \Big\} . \quad (7)$$

The GLRT statistic in [1, Sec. 5.3.3] for problem (6) is

$$\mathcal{L}_{RG} = 1 - \frac{[\hat{\mathbf{\Omega}}]_{ij}^2}{[\hat{\mathbf{\Omega}}]_{ii}[\hat{\mathbf{\Omega}}]_{jj}} , \quad (8)$$

where the null is rejected if (8) is small:

$$\mathcal{L}_{RG} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \tau. \quad (9)$$

We see that the statistics (5) and (9) are equivalent.

Now consider proper complex Gaussian graphical models with N i.i.d. observations $\mathbf{x}(0), \mathbf{x}(1), \cdots, \mathbf{x}(N-1)$ of $\mathbf{x} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{\Sigma})$. Now the hypothesis testing problem is

$$\begin{aligned} \mathcal{H}_0 &: \ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{C}_+(\mathcal{G}'),\ \mathcal{G}' = (V, \mathcal{E}'),\ \mathcal{E}' = \mathcal{E} \backslash \{i,j\} \\ \mathcal{H}_1 &: \ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{C}_+(\mathcal{G}),\ \mathcal{G} = (V, \mathcal{E}),\ \mathcal{E} \text{ is saturated} , \end{aligned} \quad (10)$$

where

$$\mathbb{C}_+(\mathcal{G}) = \Big\{ \mathbf{\Omega} \in \mathbb{C}_+^{p \times p} \,\Big|\, \forall i \neq j \in V \ : \ \{i,j\} \notin \mathcal{E} \Rightarrow \ \Omega_{ij} = 0 \Big\} . \quad (11)$$

The sample covariance matrix in this case is $\hat{\mathbf{\Sigma}} = \frac{1}{N} \sum_{t=0}^{N-1} \mathbf{x}(t)\mathbf{x}^H(t)$, and an estimate of its inverse is $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}}^{-1}$. Exploiting [12, Theorem 7.1], it is shown in [13] that a GLRT statistic for the problem (10) is

$$\mathcal{L}_{CG} = 1 - \frac{\left|[\hat{\mathbf{\Omega}}]_{ij}\right|^2}{[\hat{\mathbf{\Omega}}]_{ii}[\hat{\mathbf{\Omega}}]_{jj}}, \quad (12)$$

where the null is rejected if (12) is small:

$$\mathcal{L}_{CG} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \tau. \quad (13)$$

III. System Model

We first review improper complex Gaussian vectors in Sec. III-A, and then, in Sec. III-B, we define an improper complex Gaussian graphical model with p nodes in terms of an associated real Gaussian graphical model with $2p$ nodes, using an augmented real vector $\mathbf{y}$ comprised of $\mathrm{real}(\mathbf{x})$ and $\mathrm{imag}(\mathbf{x})$. A binary hypothesis testing problem is formulated in Sec. III-C exploiting the concept of "correct graph" [33], [34], and using the associated real Gaussian graphical model, where we test for exclusion/inclusion of four edges in the real Gaussian graphical model corresponding to a single edge in associated improper complex Gaussian graphical model.

A. 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 $\mathbf{z}$ and the real vector $\mathbf{y}$ as

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_i \end{bmatrix}. \tag{14}$$

The pdf of an improper complex Gaussian $\mathbf{x}$ is defined in terms of that of the augmented $\mathbf{y}$ or $\mathbf{z}$ [27, 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) $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$. Then we have $\mathbf{y} \sim \mathcal{N}_r(\mathbf{0}, \mathbf{R}_{zz})$ where

$$\mathbf{R}_{yy} = \begin{bmatrix} \mathbf{R}_{x_r x_r} & \mathbf{R}_{x_r x_i} \\ \mathbf{R}_{x_i x_r} & \mathbf{R}_{x_i x_i} \end{bmatrix}. \tag{15}$$

The augmented vectors $\mathbf{z}$ and $\mathbf{y}$ are related via

$$\mathbf{z} = \boldsymbol{\mathcal{T}}\mathbf{y}, \tag{16}$$

where

$$\boldsymbol{\mathcal{T}} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \in \mathbb{C}^{(2p)\times(2p)} \tag{17}$$

and $\boldsymbol{\mathcal{T}}$ is full-rank. For complex vectors, define the covariance matrix $\mathbf{R}_{uv} = \mathbb{E}\{\mathbf{u}\mathbf{v}^H\}$, and the complementary covariance matrix $\tilde{\mathbf{R}}_{uv} = \mathbb{E}\{\mathbf{u}\mathbf{v}^\top\}$ [27, Sec. 2.2], for zero-mean $\mathbf{u}, \mathbf{v} \in \mathbb{C}^p$. Then we have

$$\mathbf{R}_{zz} = \begin{bmatrix} \mathbf{R}_{xx} & \tilde{\mathbf{R}}_{xx} \\ \tilde{\mathbf{R}}_{xx}^* & \mathbf{R}_{xx}^* \end{bmatrix} = \mathbf{R}_{zz}^H. \tag{18}$$

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

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\pi)^{2p/2}\,|\mathbf{R}_{yy}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{R}_{yy}^{-1}\mathbf{y}\right). \tag{19}$$

Using (16), one can also express (19) as [27, Sec. 2.3.1]

$$\begin{aligned} f_{\mathbf{x}}(\mathbf{x}) &:= f_{\mathbf{z}}(\mathbf{z}) \\ &= \frac{1}{\pi^p\,|\mathbf{R}_{zz}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{z}^H \mathbf{R}_{zz}^{-1}\mathbf{z}\right). \end{aligned} \tag{20}$$

For proper $\mathbf{x}$, $\tilde{\mathbf{R}}_{xx} = \mathbf{0}$, and (20) reduces to (3).

In this paper, in order to exploit some developments in [1] pertaining to real Gaussian graphical models, we will use the representation $\mathbf{y}$ in (14) and its pdf (19) for improper complex Gaussian $\mathbf{x}$.

B. Improper Complex Gaussian Graphical Model

An improper complex Gaussian graphical model 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 $\mathbf{x} \in \mathbb{C}^p$, $p = |V|$, $V = [1, p]$, that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$. Similar to real Gaussian and proper complex Gaussian graphical models, the CIG $\mathcal{G}$ and multivariate complex-valued improper Gaussian $\mathbf{x}$ specify an improper complex Gaussian graphical model for $\mathbf{x}$ where the distribution of $\mathbf{x}$ obeys the conditional independence restrictions of CIG $\mathcal{G}$.

However, as discussed in Sec. III-A, since an improper x_i 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 improper complex Gaussian graphical model, in fact, each node corresponds to two random variables $\mathrm{real}(x_j) =: x_{jr}$ and $\mathrm{imag}(x_j) =: x_{ji}$. In the notation of (20), conditional independence of improper x_j and x_k is equivalent to

$$f_{x_j, x_k \,|\, \mathbf{x}_{\tilde{V}}}(x_j, x_k \,|\, \mathbf{x}_{\tilde{V}}) = f_{x_j \,|\, \mathbf{x}_{\tilde{V}}}(x_j \,|\, \mathbf{x}_{\tilde{V}}) f_{x_k \,|\, \mathbf{x}_{\tilde{V}}}(x_k \,|\, \mathbf{x}_{\tilde{V}}) \tag{21}$$

where $\tilde{V} = V \backslash \{j, k\}$. Express the left-side of (21) as

$$\begin{aligned} &f_{x_j, x_k \,|\, \mathbf{x}_{\tilde{V}}}(x_j, x_k \,|\, \mathbf{x}_{\tilde{V}}) \\ &\qquad = f_{x_{jr}, x_{ji}, x_{kr}, x_{ki} \,|\, \mathbf{x}_{\tilde{V}}}(x_{jr}, x_{ji}, x_{kr}, x_{ki} \,|\, \mathbf{x}_{\tilde{V}}) \end{aligned} \tag{22}$$

and similarly, express the right-side of (21) as (l is either j or k)

$$f_{x_l \,|\, \mathbf{x}_{\tilde{V}}}(x_l \,|\, \mathbf{x}_{\tilde{V}}) = f_{x_{lr}, x_{li} \,|\, \mathbf{x}_{\tilde{V}}}(x_{lr}, x_{li} \,|\, \mathbf{x}_{\tilde{V}}). \tag{23}$$

Since

$$\begin{aligned} &f_{x_{jr}, x_{kr} \,|\, \mathbf{x}_{\tilde{V}}}(x_{jr}, x_{kr} \,|\, \mathbf{x}_{\tilde{V}}) \\ &= \int\int f_{x_{jr}, x_{ji}, x_{kr}, x_{ki} \,|\, \mathbf{x}_{\tilde{V}}}(x_{jr}, x_{ji}, x_{kr}, x_{ki} \,|\, \mathbf{x}_{\tilde{V}})\, dx_{ji}\, dx_{ki} \end{aligned} \tag{24}$$

and

$$f_{x_{lr} \,|\, \mathbf{x}_{\tilde{V}}}(x_{lr} \,|\, \mathbf{x}_{\tilde{V}}) = \int f_{x_{lr}, x_{li} \,|\, \mathbf{x}_{\tilde{V}}}(x_{lr}, x_{li} \,|\, \mathbf{x}_{\tilde{V}})\, dx_{li}, \tag{25}$$

(21)-(25) imply that

$$\begin{aligned} &f_{x_{jr}, x_{kr} \,|\, \mathbf{x}_{\tilde{V}}}(x_{jr}, x_{kr} \,|\, \mathbf{x}_{\tilde{V}}) \\ &\qquad = f_{x_{jr} \,|\, \mathbf{x}_{\tilde{V}}}(x_{lr} \,|\, \mathbf{x}_{\tilde{V}}) f_{x_{kr} \,|\, \mathbf{x}_{\tilde{V}}}(x_{lr} \,|\, \mathbf{x}_{\tilde{V}}). \end{aligned} \tag{26}$$

Proceeding similarly, it then follows that

$$f_{u, v \,|\, \mathbf{x}_{\tilde{V}}}(u, v \,|\, \mathbf{x}_{\tilde{V}}) = f_{u \,|\, \mathbf{x}_{\tilde{V}}}(u \,|\, \mathbf{x}_{\tilde{V}}) f_{v \,|\, \mathbf{x}_{\tilde{V}}}(v \,|\, \mathbf{x}_{\tilde{V}}) \tag{27}$$

for any real scalars u, v that satisfy

$$u \in \{\mathrm{real}(x_j),\, \mathrm{imag}(x_j)\},\ v \in \{\mathrm{real}(x_k),\, \mathrm{imag}(x_k)\}. \tag{28}$$

In order to exploit some results in [1] pertaining to real Gaussian graphical models for testing the validity of (27)-(28) for a given improper complex Gaussian graphical model, we will use the representation $\mathbf{y}$ in (14), and exploit a larger real Gaussian graphical model corresponding to the given improper complex Gaussian graphical model $\mathcal{G} = (V, \mathcal{E})$. Consider an real Gaussian graphical model $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ associated with improper complex Gaussian graphical model $\mathcal{G} = (V, \mathcal{E})$, where $\bar{V} = [1, 2p]$, vertex j for $1 \le j \le p$ represents the real part of improper x_j, $\mathrm{real}(x_j)$, and vertex $j + p$ represents $\mathrm{imag}(x_j)$. For a given edge $\{j, k\}$, define the set of four edges $\bar{\mathcal{E}}^{(jk)}$ as

$$\bar{\mathcal{E}}^{(jk)} = \big\{\{j, k\}, \{j+p, k\}, \{j, k+p\}, \{j+p, k+p\}\big\}. \tag{29}$$

If the edge $\{j, k\} \notin \mathcal{E}$, then we have edges $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$, implying relations (27)-(28) for all four possible values of pair (u, v). Clearly, (21), i.e., $\{j, k\} \notin \mathcal{E}$, implies (27)-(28) for all four possible values of pairs (u, v), which is equivalent to $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$. What about the converse? Here we invoke the Markov properties on undirected graphs, discussed in Sec. II-A1, following [1]. Consider subsets $A = \{j, j+p\}$, $B = \{k, k+p\}$ and $S = \bar{V} \backslash (A \cup B)$ of $\bar{V}$ for $j \ne k$, $j, k \in$

$[1,p]$. These three sets are disjoint, and S separates A and B when $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$. Since pairwise Markov property implies global Markov property for real Gaussian graphical models ([1, p. 131], Sec. II-A1), $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$ implies that $\mathbf{x}_A$ and $\mathbf{x}_B$ are conditionally independent given $\mathbf{x}_S$, which, in terms of the original improper complex Gaussian graphical model $\mathcal{G} = (V, \mathcal{E})$, implies that $\{j,k\} \notin \mathcal{E}$.

Thus, we have established that given improper complex Gaussian graphical model $\mathcal{G} = (V, \mathcal{E})$ with $\mathbf{x} \in \mathbb{C}^p$, and the associated real Gaussian graphical model $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ with $\mathbf{y} = [\mathrm{real}(\mathbf{x}^\top)\ \mathrm{imag}(\mathbf{x}^\top)]^\top \in \mathbb{R}^{2p}$,

$$\{j,k\} \notin \mathcal{E} \Leftrightarrow \bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset\,. \tag{30}$$

Let $\mathbf{R}_{yy} = \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\} \succ \mathbf{0}$ and $\bar{\mathbf{\Omega}} = \mathbf{R}_{yy}^{-1}$. Then, by [1, Prop. 5.2] pertaining to real Gaussian graphical models,

$$\{j,k\} \notin \mathcal{E} \Leftrightarrow [\bar{\mathbf{\Omega}}]_{\ell m} = 0\ \forall \{\ell, m\} \in \bar{\mathcal{E}}^{(jk)}\,. \tag{31}$$

We state this result as Lemma 2.

Lemma 2: Consider an improper complex Gaussian graphical model $\mathcal{G} = (V, \mathcal{E})$ with $\mathbf{x} \in \mathbb{C}^p$ and $V = [1,p]$, and the associated real Gaussian graphical model $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ with $\mathbf{y} = [\mathrm{real}(\mathbf{x}^\top)\ \mathrm{imag}(\mathbf{x}^\top)]^\top \in \mathbb{R}^{2p}$, $\bar{V} = [1,2p]$, where vertex j for $1 \le j \le p$ represents the real part of improper x_j, $\mathrm{real}(x_j)$, and vertex $j+p$ represents $\mathrm{imag}(x_j)$. Assume that $\mathbf{R}_{yy} = \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\} \succ \mathbf{0}$. 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 $\bar{\mathbf{\Omega}} = \mathbf{R}_{yy}^{-1}$ and $\bar{\mathcal{E}}^{(jk)}$ is defined in (29). •

Lemma 2 is the counterpart of [1, Prop. 5.2] pertaining to real Gaussian graphical models, and of [12, Theorem 7.1] pertaining to proper complex Gaussian graphical models.

C. Binary Hypothesis Testing

Suppose we are given N i.i.d. observations $\mathbf{x}(0), \mathbf{x}(1), \cdots, \mathbf{x}(N-1)$ of zero-mean improper $\mathbf{x} \in \mathbb{C}^p$ which obeys the conditional independence restriction implied by the associated graph $\mathcal{G} = (V, \mathcal{E})$, $V = [1,p]$. We need to decide if for any given $j,k \in V$, $j \neq k$, edge $\{j,k\} \in \mathcal{E}$ or $\{j,k\} \notin \mathcal{E}$. As in Lemma 2, consider the associated real Gaussian graphical model $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ and $\mathbf{y} \in \mathbb{R}^{2p}$. Then by (30), we need to test if $\bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}} = \emptyset$. In a binary hypothesis testing framework, we compare two competing models: $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ and $\bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}')$ with $\bar{\mathcal{E}}' = \bar{\mathcal{E}} \backslash \bar{\mathcal{E}}^{(jk)}$. What should $\bar{\mathcal{E}}$ be in the real Gaussian graphical model, i.e., what should $\mathcal{E}$ be in the original improper complex Gaussian graphical model? Note that we do not know the true edge set $\mathcal{E}$. To this end, we first recall the concept of a *correct* graph from [33], [34], where it is defined for time series graphical models, but it applies here as well.

Definition 5. Correct Graph [33, Definition 2], [34, Definition 1]. Let $\mathcal{G} = (V, \mathcal{E})$, $V = [1,p]$, denote the true graphical model for $\mathbf{x}$ (in $\mathbb{R}^p$ or $\mathbb{C}^p$). Then $\tilde{\mathcal{G}} = (V, \tilde{\mathcal{E}})$ is correct for $\mathcal{G} = (V, \mathcal{E})$ if $\mathcal{E} \subseteq \tilde{\mathcal{E}}$. •

Thus, if $\{j,k\} \in \mathcal{E}$, then we must have $\{j,k\} \in \tilde{\mathcal{E}}$, and if $\{j,k\} \notin \tilde{\mathcal{E}}$, then $\{j,k\} \notin \mathcal{E}$. Note that a saturated (complete) graph $\mathcal{G} = (V, \mathcal{E}_s)$ is a correct graph for any graphical model, where $\mathcal{E}_s$ denotes the set of all possible edges with vertices in V.

This definition of a correct graph is central to the approach of [34], patterned after that of [33], to graphical modeling of real time series, and it applies in our case as well. To test if edge $\{j,k\} \notin \mathcal{E}$, we consider if edge $\{j,k\} \notin \mathcal{E}_s$ (null: correct model), or $\{j,k\} \in \mathcal{E}_s$ (alternative: incorrect model), where $(V, \mathcal{E}_s)$ is correct for any $(V, \mathcal{E})$. There are $L = p(p-2)/2$ edges in $(V, \mathcal{E}_s)$. Number edge $\{j,k\}$ as $i = (j-1)p + k$, $1 \le j < k \le p$. Let $(V, \mathcal{E}^i)$ denote the graph where $\mathcal{E}^i = \mathcal{E}_s \backslash \{j,k\}$, with $i = (j-1)p+k$, i.e., only one edge is missing from the saturated graph. Lemma 3 is a restatement of [34, Proposition 2].

Lemma 3. If the graph $(V, \mathcal{E}^i)$ is correct for edges corresponding to $i = i_1, i_2, \cdots, i_s$, and incorrect for all others, then the graphical model $\mathcal{G} = (V, \mathcal{E})$ for $\mathbf{x}$ is the graph with only edges $\{i_1, i_2, \cdots, i_s\}$ missing. •

Motivated by Lemma 3 and (6), and using Lemma 2, we consider the following hypothesis testing problem for testing the exclusion of single edge $\{j,k\}$ from $\mathcal{E}$:

$$\begin{aligned} \mathcal{H}_0 :&\ \ \bar{\mathbf{\Omega}} = \bar{\mathbf{\Omega}}(\bar{\mathcal{G}}) \in \bar{\mathbb{R}}_+(\bar{\mathcal{G}}'),\ \bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}'),\ \bar{\mathcal{E}}' = \bar{\mathcal{E}}_s \backslash \bar{\mathcal{E}}^{(jk)} \\ \mathcal{H}_1 :&\ \ \bar{\mathbf{\Omega}} = \bar{\mathbf{\Omega}}(\bar{\mathcal{G}}) \in \bar{\mathbb{R}}_+(\bar{\mathcal{G}}),\ \bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s),\ \bar{\mathcal{E}}_s \text{ is saturated}\,, \end{aligned} \tag{32}$$

where

$$\bar{\mathbb{R}}_+(\bar{\mathcal{G}}) = \Big\{ \bar{\mathbf{\Omega}} \in \mathbb{R}_+^{(2p)\times(2p)} \,\Big|\, \forall j \neq k \in V\ :\ \bar{\mathcal{E}}^{(jk)} \cap \bar{\mathcal{E}}_s = \emptyset \Rightarrow\ \bar{\Omega}_{\ell m} = 0\ \forall \{\ell, m\} \in \mathcal{E}^{(jk)} \Big\}\,. \tag{33}$$

It follows from Lemmas 2 and 3 that if $\mathcal{H}_0$ in (32) is true then $\{j,k\} \notin \mathcal{E}$, else $\{j,k\} \in \mathcal{E}$. In the next section we derive a GLRT for testing problem (32).

D. Relation to Prior Work

As noted in Sec. I, real Gaussian graphical models associated with undirected graphs have been extensively investigated. These models are also called covariance selection models [1], [35], or Gaussian Markov random fields (GMRF) models [36]. Learning a Gaussian graphical model corresponds to learning which entries of the inverse covariance matrix are zero, and also estimating its nonzero entries. Dempster [35] appears to be the first to address this problem. One may classify prior work into two broad categories: low-dimensional models [1], [2], [7], [8], [19], [29]–[32], [35] where $p \ll N$, and high-dimensional models [3], [5], [6], [9]–[11], [20], [36], [37] where p is of the order of N (or larger).

Work on high-dimensional models relies crucially on the assumption that the non-zero elements of the inverse covariance matrix are sparse. Since p is of the order of N (or larger), the sample covariance matrix may not be invertible or may be quite ill-conditioned. Furthermore, testing for each possible edge for inclusion/exclusion in the graph is not practically feasible. So an emphasis is on devising approaches for estimating $\mathbf{\Omega}$ in such a way that the resulting estimate of $\mathbf{\Omega}$ is sparse. In [3], [9] a penalized regression (pseudo-likelihood) approach is used, whereas in [6], a penalized log-likelihood function is used with lasso-related penalties. Similar approaches include [10], [11]. Furthermore, these approaches, and in general,

high-dimensional statistical methods, lack classical measures of uncertainty and statistical significance, such as confidence intervals and p-values, needed for classical hypothesis testing and estimator quality quantification.

This paper is concerned with the low-dimensional case, with no sparsity assumption or restriction. A focus in graphical model selection has been to devise statistical tests for exclusion/inclusion of edges in the graph [1], [2], [29]–[32]. For a single edge, the edge exclusion test of [2, Sec. 6.8] is the same as that of [1, Sec. 5.3.3] (see (9)), but in order to calculate the test threshold to yield a specified significance level α, [2, Sec. 6.8] uses an asymptotic χ^2-distribution under the null hypothesis, whereas [1, Sec. 5.3.3] specifies an exact distribution under the null hypothesis. As noted in Sec. II-B1, an equivalent test statistic has also been used in [30], [31]. As already noted in Sec. I, there are $p(p-1)/2$ distinct edges in an undirected graph with p nodes. Therefore, for graphical model selection, one has to determine which set of edges out of total $p(p-1)/2$ edges belong to $\mathcal{E}$, which requires multiple testing based on control of overall significance level [31] over all tested edges. In this paper we focus on edge exclusion testing, and do not consider multiple testing issues which, in principle, can be handled in a similar manner as in [31] (and others [38, Sec. 10.7]) using various existing approaches such as Bonferroni and Holm methods for controlling the family-wise error rate at level α [31], and Benjamini-Hochberg method for control of false discovery rate at level α [31], [38]. All these multiple testing approach depend crucially on the underlying single edge exclusion tests, and moreover, in general, they can control only an upper bound on the overall α. Note that [30], [31] also use an asymptotic distribution under the null hypothesis in their approaches. In this paper, we derive an exact distribution under the null hypothesis for single edge exclusion test for improper complex Gaussian graphical models, similar to [1, Sec. 5.3.3] but unlike the asymptotic distributions of [2], [30], [31], pertaining to real Gaussian graphical models.

Our formulation (32) of edge exclusion testing for improper complex Gaussian graphical models leads to testing for joint exclusion of four edges in an augmented real Gaussian graphical model. The single edge exclusion test of [1, Sec. 5.3.3] can be used to test these four edges but only in a multiple testing set-up where we test each of the four edges separately, one-by-one, whereas in our formulation we have a single test. Our test presented in Sec. IV yields the specified significance level α exactly whereas, multiple testing based on (9) can control only an upper bound on α, resulting in a loss in test power. We illustrate this aspect later in Sec. VI via simulations.

Existing work on high-dimensional models ([3], [6], [9]–[11] and others) considers real graphical models only, whereas we investigate complex Gaussian graphical models. These high-dimensional model-based approaches do not consider classical hypothesis testing involving significance levels and p-values, whereas we do so in this paper. Finally, an assumption of sparsity of connected edges is crucial to existing work on high-dimensional models, whereas we do not require it in this paper. On the other hand, our approach needs $N \gg p$ whereas high-dimensional model-based approaches allow p to be less than N.

IV. GLRT for Edge Exclusion in improper complex Gaussian graphical models

Here a GLRT is derived in Sec. IV-B for the hypothesis testing problem (32), based on [1, Proposition 5.14] discussed in Sec. IV-A. The null distribution of the test statistic is specified explicitly in Sec. IV-C, which allows for analytical calculation of the test threshold.

Our objective is to derive the GLRT for the binary hypothesis testing problem specified in (32), given N i.i.d. observations $\mathbf{x}(0), \mathbf{x}(1), \cdots, \mathbf{x}(N-1)$ of zero-mean improper $\mathbf{x} \in \mathbb{C}^p$ which obeys the conditional independence restriction implied by the associated graph $\mathcal{G} = (V, \mathcal{E})$, $V = [1, p]$. Given $\mathbf{x}(t)$, construct $\mathbf{y}(t)$ as in (14), and let

$$\mathbf{Y} = [\mathbf{y}(0)\ \mathbf{y}(1)\ \cdots\ \mathbf{y}(N-1)]^\top \in \mathbb{R}^{N\times(2p)}\,. \tag{34}$$

By assumption, $\mathbb{E}\{\mathbf{x}(t)\} = \mathbf{0}$, hence, $\mathbb{E}\{\mathbf{y}(t)\} = \mathbf{0}$, $\mathbf{R}_{yy} = \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\} \succ \mathbf{0}$ and $\bar{\mathbf{\Omega}} = \mathbf{R}_{yy}^{-1}$. Then the joint pdf of $\mathbf{Y}$ is

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{Y}) &= \prod_{t=0}^{N-1} f_{\mathbf{y}(t)}(\mathbf{y}(t)) \\ &= \prod_{t=0}^{N-1} \frac{1}{(2\pi)^{2p/2}\,|\mathbf{R}_{yy}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{R}_{yy}^{-1}\mathbf{y}\right) \\ &= \frac{|\bar{\mathbf{\Omega}}|^{N/2}}{(2\pi)^{Np}} \mathrm{etr}(-\frac{1}{2}\bar{\mathbf{\Omega}}\mathbf{Y}^\top\mathbf{Y})\,. \end{aligned} \tag{35}$$

The unknown in (35) is $\bar{\mathbf{\Omega}}$ which satisfies the respective restrictions in (32) under the two hypotheses. The GLRT for problem (32) is given by

$$\mathcal{L}(\mathbf{Y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{36}$$

where

$$\mathcal{L}(\mathbf{Y}) = \frac{\sup_{\bar{\mathbf{\Omega}}(\bar{\mathcal{G}})\in\bar{\mathbb{R}}_+(\bar{\mathcal{G}})}\ f_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}|\mathcal{H}_1)}{\sup_{\bar{\mathbf{\Omega}}(\bar{\mathcal{G}})\in\bar{\mathbb{R}}_+(\bar{\mathcal{G}}')}\ f_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}|\mathcal{H}_0)}\,. \tag{37}$$

The solution to (36)-(37) will be obtained by exploiting [1, Proposition 5.14], which we recall next.

A. GLRT for Decomposable real Gaussian graphical models [1, Sec. 5.3.3]

Consider $\mathbf{x} \in \mathbb{R}^p$ and the associated simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ which is assumed to be decomposable. Let $\mathcal{G}_0 = (V, \mathcal{E}_0)$ denote a decomposable submodel of $\mathcal{G}$ with k (≥ 1) edges less than $\mathcal{G}$. By [1, Lemma 2.21], there is a sequence $\mathcal{G}_0 \subset \cdots \subset \mathcal{G}_k = \mathcal{G}$ of graphs that are decomposable and differ by one edge only. Let e_i denote the edge that is in $\mathcal{G}_i$ but not in $\mathcal{G}_{i-1}$. Consider the hypothesis testing problem

$$\begin{aligned} \mathcal{H}_0 :&\ \ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}_0),\ \mathcal{G}_0 = (V, \mathcal{E}_0) \\ \mathcal{H}_1 :&\ \ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}),\ \mathcal{G} = (V, \mathcal{E})\,, \end{aligned} \tag{38}$$

given N i.i.d. observations $\mathbf{x}(t)$, $t = 0, 1, \cdots, N-1$, of zero-mean $\mathbf{x}$. Define

$$\mathbf{X} = [\mathbf{x}(0)\ \mathbf{x}(1)\ \cdots\ \mathbf{x}(N-1)]^\top \in \mathbb{R}^{N\times p}\,. \tag{39}$$

The GLRT for problem (38) is given by

$$\mathcal{L}(\mathbf{X}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{40}$$

where

$$\mathcal{L}(\mathbf{X}) = \frac{\sup_{\mathbf{\Omega}(\mathcal{G})\in\mathbb{R}_+(\mathcal{G})} f_{\mathbf{X}|\mathcal{H}_1}(\mathbf{X}|\mathcal{H}_1)}{\sup_{\mathbf{\Omega}(\mathcal{G})\in\mathbb{R}_+(\mathcal{G}_0)} f_{\mathbf{X}|\mathcal{H}_0}(\mathbf{X}|\mathcal{H}_0)}\,. \tag{41}$$

Under the above set-up, the solution (40)-(41) to problem (38) is specified in [1, Proposition 5.14], which we now state as Lemma 4 in the notation of this paper.

Lemma 4: The GLRT for the test (38) is given by (40) where

$$\mathcal{L}(\mathbf{X}) = \left(\frac{|\hat{\mathbf{\Omega}}_0|}{|\hat{\mathbf{\Omega}}|}\right)^{-N/2}, \tag{42}$$

$\hat{\mathbf{\Omega}}_0$ maximizes $f_{\mathbf{X}|\mathcal{H}_0}(\mathbf{X}|\mathcal{H}_0)$ under the constraint $\mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G}_0)$, and $\hat{\mathbf{\Omega}}$ maximizes $f_{\mathbf{X}|\mathcal{H}_1}(\mathbf{X}|\mathcal{H}_1)$ under the constraint $\mathbf{\Omega}(\mathcal{G}) \in \mathbb{R}_+(\mathcal{G})$. Under $\mathcal{H}_0$, $\mathcal{L}^{-2/N}(\mathbf{X}) \sim B = \prod_{i=1}^k B_i$ where B_is are mutually independent beta random variables $\mathcal{B}((N-|C_i^*|+1)/2, 1/2)$ and C_i^* is the unique clique of $\mathcal{G}_i$ that contains the edge e_i. •

We should note that in [1, Proposition 5.14], both mean and inverse covariance are unknown, whereas in our case, we have zero mean **x**. This results in an extra degree of freedom in estimating the covariance matrix, hence $B_i \sim \mathcal{B}((N-|C_i^*|+1)/2, 1/2)$ in Lemma 4, instead of $B_i \sim \mathcal{B}((N-|C_i^*|)/2, 1/2)$ as in [1, Proposition 5.14].

B. Proposed GLRT

The solution to (36)-(37) will be obtained by exploiting Lemma 4. To this end, we need to establish that all the requirements of Lemma 4, as applied to (36)-(37), are fulfilled.

Consider the graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s)$ under $\mathcal{H}_1$. Since it is complete, it is decomposable (by Definition 3). Starting with $\bar{\mathcal{G}}$, we remove one edge at a time to obtain $\bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}_s \backslash \bar{\mathcal{E}}^{(jk)})$ under $\mathcal{H}_0$. We go through the following steps, one edge at a time, to verify Lemma 1 and find the associated cliques and separators.

(i) We have a complete, hence decomposable, graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s)$, with a single clique $\bar{V}$

(ii) Remove edge $\{j,k\}$ from $\bar{\mathcal{G}}$ to obtain $\bar{\mathcal{G}}_1 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \{j,k\})$. Define $B_1 = \bar{V}\backslash\{k\}$ and $B_2 = \bar{V}\backslash\{j\}$. With $m = 2$ in Definition 4, we have

$$\begin{aligned} &H_1 = B_1 = \bar{V}\backslash\{k\}, \quad H_2 = B_1 \cup B_2 = \bar{V} \\ &S_2 = H_1 \cap B_2 = \bar{V}\backslash\{j,k\}\,. \end{aligned} \tag{43}$$

The sequence of subsets B_1, B_2 of $\bar{V}$ is perfect since S_2 is complete and $S_2 \subset B_1$. The sets B_1 and B_2 are the cliques of $\bar{\mathcal{G}}_1$, which is decomposable by Lemma 1.

(iii) Now remove edge $\{j, k+p\}$ from $\bar{\mathcal{G}}_1$ to obtain $\bar{\mathcal{G}}_2 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \{\{j,k\},\{j,k+p\}\})$. Define $B_1 = \bar{V}\backslash\{j\}$ and $B_2 = \bar{V}\backslash\{k,k+p\}$. With $m = 2$ in Definition 4, we have

$$\begin{aligned} &H_1 = B_1 = \bar{V}\backslash\{j\}, \quad H_2 = B_1 \cup B_2 = \bar{V} \\ &S_2 = H_1 \cap B_2 = \bar{V}\backslash\{j,k,k+p\}\,. \end{aligned} \tag{44}$$

The sequence of subsets B_1, B_2 of $\bar{V}$ is perfect since S_2 is complete and $S_2 \subset B_1$. The sets B_1 and B_2 are the cliques of $\bar{\mathcal{G}}_2$, which is, therefore, decomposable by Lemma 1.

(iv) Now remove edge $\{j+p, k\}$ from $\bar{\mathcal{G}}_2$ to obtain $\bar{\mathcal{G}}_3 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \{\{j,k\},\{j,k+p\},\{j+p,k\}\})$. Define $B_1 = \bar{V}\backslash\{j, j+p\}$, $B_2 = \bar{V}\backslash\{j,k\}$ and $B_3 = \bar{V}\backslash\{k,k+p\}$. With $m = 3$ in Definition 4, we have

$$\begin{aligned} &H_1 = B_1 = \bar{V}\backslash\{j,j+p\},\ H_2 = B_1 \cup B_2 = \bar{V}\backslash\{j\} \\ &H_3 = B_1 \cup B_2 \cup B_3 = \bar{V} \\ &S_2 = H_1 \cap B_2 = \bar{V}\backslash\{j,j+p,k\} \\ &S_3 = H_2 \cap B_3 = \bar{V}\backslash\{j,k,k+p\}\,. \end{aligned} \tag{45}$$

The separator sets S_2 and S_3 are complete, $S_2 \subset B_1$, and $S_3 \subset B_2$, hence the sequence of subsets B_1, B_2, B_3 of $\bar{V}$ is a perfect sequence. The sets B_1, B_2 and B_3 are all possible cliques of $\bar{\mathcal{G}}_3$, which is, therefore, decomposable by Lemma 1.

(v) Finally, remove edge $\{j+p, k+p\}$ from $\bar{\mathcal{G}}_3$ to obtain $\bar{\mathcal{G}}_4 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \bar{\mathcal{E}}^{(jk)})$. Define $B_1 = \bar{V}\backslash\{j, j+p\}$ and $B_2 = \bar{V}\backslash\{k,k+p\}$. With $m = 2$ in Definition 4, we have

$$\begin{aligned} &H_1 = B_1 = \bar{V}\backslash\{j,j+p\}, \quad H_2 = B_1 \cup B_2 = \bar{V} \\ &S_2 = H_1 \cap B_2 = \bar{V}\backslash\{j,k,j+p,k+p\}\,. \end{aligned} \tag{46}$$

Since S_2 is complete and $S_2 \subset B_1$, the sequence B_1, B_2 is perfect. The sets B_1 and B_2 are all possible cliques of $\bar{\mathcal{G}}_4$, which is, therefore, decomposable by Lemma 1.

With the above results, we are ready to invoke Lemma 4 (i.e., [1, Proposition 5.14]). Define

$$\begin{aligned} &\mathcal{G}_4 := \bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}}_s) \\ &\mathcal{G}_3 := \bar{\mathcal{G}}_1 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \{j,k\}) \\ &\mathcal{G}_2 := \bar{\mathcal{G}}_2 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \{\{j,k\},\{j,k+p\}\}), \\ &\mathcal{G}_1 := \bar{\mathcal{G}}_3 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \{\{j,k\},\{j,k+p\},\{j+p,k\}\}) \\ &\mathcal{G}_0 := \bar{\mathcal{G}}_4 = (\bar{V}, \bar{\mathcal{E}}_s \backslash \bar{\mathcal{E}}^{(jk)})\,. \end{aligned} \tag{47}$$

With reference to (32), then we have a sequence of graphs $\bar{\mathcal{G}}' = \mathcal{G}_0 \subset \mathcal{G}_1 \subset \cdots \subset \mathcal{G}_4 = \bar{\mathcal{G}}$ that are decomposable, and differ by one edge only. With e_i as in Lemma 4, we have

$$\begin{aligned} &e_1 = \{j+p, k+p\}, \quad e_2 = \{j+p, k\} \\ &e_3 = \{j, k+p\}, \quad e_4 = \{j,k\}\,. \end{aligned} \tag{48}$$

Then with C_i^* as in Lemma 4, we have

$$\begin{aligned} &C_1^* = \bar{V}\backslash\{j,k\}, \quad C_2^* = \bar{V}\backslash\{j\} \\ &C_3^* = \bar{V}\backslash\{k\}, \quad C_4^* = \bar{V}\,. \end{aligned} \tag{49}$$

Thus, $|C_1^*| = 2p-2$, $|C_2^*| = 2p-1$, $|C_3^*| = 2p-1$ and $|C_4^*| = 2p$.

We also need the determinants of the ML estimates of $\bar{\mathbf{\Omega}}$ under the two hypotheses. Define the sample covariance matrix (we now use $\bar{\mathbf{\Sigma}} = \mathbf{R}_{yy}$)

$$\hat{\bar{\mathbf{\Sigma}}} = \frac{1}{N}\mathbf{Y}^\top\mathbf{Y} = \frac{1}{N}\sum_{t=0}^{N-1}\mathbf{y}(t)\mathbf{y}^\top(t)\,. \tag{50}$$

For $\mathbf{y}(t)$ associated with $\bar{\mathcal{G}}$ or $\bar{\mathcal{G}}'$, for $\tilde{V} \subseteq V$, we define $\mathbf{y}_{\tilde{V}}(t) = [y_i(t)]_{i\in\tilde{V}}$ = a column vector of dimension $|\tilde{V}|$. Thus, $\mathbf{y}(t) = \mathbf{y}_V(t)$. Under $\mathcal{H}_1$, the graph $\bar{\mathcal{G}}$ is saturated, hence by [1, Theorem 5.1], the ML estimate $\hat{\bar{\mathbf{\Omega}}}$ of $\bar{\mathbf{\Omega}} = \bar{\mathbf{\Sigma}}^{-1}$ is given by (noting that the mean of $\mathbf{y}(t)$ is zero)

$$\hat{\bar{\mathbf{\Omega}}} = \hat{\bar{\mathbf{\Sigma}}}^{-1} \Rightarrow |\hat{\bar{\mathbf{\Omega}}}| = 1/|\hat{\bar{\mathbf{\Sigma}}}|\,. \tag{51}$$

Under $\mathcal{H}_0$, the graph $\bar{\mathcal{G}}' = \mathcal{G}_0$ is decomposable with cliques $\bar{V}\backslash\{j, j+p\}$ and $\bar{V}\backslash\{k, k+p\}$, and separator $\bar{V}\backslash\{j,k,j+p,k+p\}$ (see step (v) of the procedure detailed above for finding the decomposable graphs and associated cliques). By [1, Proposition 5.9], the determinant of the ML estimate $\hat{\bar{\mathbf{\Omega}}}_0$ of $\bar{\mathbf{\Omega}} = \bar{\mathbf{\Sigma}}^{-1}$ under $\mathcal{H}_0$ is determined by the cliques and separators of the graph $\bar{\mathcal{G}}' = (\bar{V}, \bar{\mathcal{E}}') = \mathcal{G}_0$. Noting that the mean of $\mathbf{z}(t)$ is zero, by [1, Proposition 5.9], we have

$$|\hat{\bar{\mathbf{\Omega}}}_0| = \frac{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,k,j+p,k+p\}}|}{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,j+p\}}||\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{k,k+p\}}|} \tag{52}$$

where

$$\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,k,j+p,k+p\}} = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}_{\bar{V}\backslash\{j,k,j+p,k+p\}}(t) \times \mathbf{y}^{\top}_{\bar{V}\backslash\{j,k,j+p,k+p\}}(t) \tag{53}$$

$$\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,j+p\}} = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}_{\bar{V}\backslash\{j,j+p\}}(t)\mathbf{y}^{\top}_{\bar{V}\backslash\{j,j+p\}}(t) \tag{54}$$

$$\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{k,k+p\}} = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}_{\bar{V}\backslash\{k,k+p\}}(t)\mathbf{y}^{\top}_{\bar{V}\backslash\{k,k+p\}}(t)\,. \tag{55}$$

We now state the result of application of Lemma 4 to the hypothesis testing problem (32) as Theorem 1.

Theorem 1: The GLRT for the test (32) is given by (36) where

$$\mathcal{L}(\mathbf{Y}) = \left[\frac{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,k,j+p,k+p\}}|\,|\hat{\bar{\mathbf{\Sigma}}}|}{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,j+p\}}|\,|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{k,k+p\}}|}\right]^{-N/2}. \tag{56}$$

Under $\mathcal{H}_0$, $\mathcal{L}^{-2/N}(\mathbf{Y}) \sim B = \prod_{i=1}^4 B_i$ where B_is are mutually independent beta random variables, with $B_1 \sim \mathcal{B}((N-2p+3)/2, 1/2), B_2 \sim \mathcal{B}((N-2p+2)/2, 1/2), B_3 \sim \mathcal{B}((N-2p+2)/2, 1/2)$, and $B_4 \sim \mathcal{B}((N-2p+1)/2, 1/2)$. •

C. Threshold Selection

We now turn to calculating the pdf of the random variable B in Theorem 1, under $\mathcal{H}_0$. We need the following results from [39] and [40].

Lemma 5 [39, Corollary 3.1.2]: If independent beta random variables $X_1 \sim \mathcal{B}(\alpha_1, \beta_1)$ and $X_2 \sim \mathcal{B}(\alpha_2, \beta_2)$, then $Y = X_1X_2 \sim \mathcal{B}(\alpha_1, \beta_1 + \beta_2)$ if $\alpha_2 = \alpha_1 + \beta_1$. •

Lemma 6 [40]: If independent exponential random variables $X_1 \sim \exp(\lambda_1)$ and $X_2 \sim \exp(\lambda_2)$, then $Y = X_1 + X_2$ has the density function

$$f_Y(y) = \frac{\lambda_1\lambda_2}{\lambda_2 - \lambda_1}\left(e^{-\lambda_1 y} - e^{-\lambda_2 y}\right),\ y \geq 0,$$

if $\lambda_1 \neq \lambda_2$. •

Consider $B = \prod_{i=1}^4 B_i$ in Theorem 1. By Lemma 5, $B_1B_2 \sim \mathcal{B}((N-2p+1)/2, 1)$ and $B_3B_4 \sim \mathcal{B}((N-2p+2)/2, 1)$. Consider log-likelihood ratio $\ln(\mathcal{L}(\mathbf{Y}))$. By Theorem 1, $\ln(\mathcal{L}(\mathbf{Y})) \sim Y_1 + Y_2$ where Y_1 and Y_2 are independent, $Y_1 = -(N/2)\ln(B_1B_2) \sim \exp(1 - (2p-1)/N)$ and $Y_2 = -(N/2)\ln(B_3B_4) \sim \exp(1-(2p-2)/N)$, based on the fact that if $X \sim \mathcal{B}(\alpha, 1)$, then $-m\ln(X) \sim \exp(\alpha/m)$. Using Lemma 6, under $\mathcal{H}_0$, $Y = \ln(\mathcal{L}(\mathbf{Y}))$ has the pdf ($y \geq 0$)

$$f_Y(y) = N\left(1 - \frac{2p-1}{N}\right)\left(1 - \frac{2p-2}{N}\right) \times e^{-y}\left[e^{(2p-1)y/N} - e^{(2p-2)y/N}\right]. \tag{57}$$

If $N \gg p$, using $e^x = 1 + x + \mathcal{O}(x^2)$, we have $e^{(2p-1)y/N} - e^{(2p-2)y/N} = (y/N) + \mathcal{O}(p^2/N^2)$, leading to

$$f_Y(y) = ye^{-y},\ \ y \geq 0. \tag{58}$$

For $N \gg p$, $\lambda_1 \approx \lambda_2 \approx 1$. The result (58) is exact for $\lambda_1 = \lambda_2 = 1$.

V. GLRT Simplification

A significant simplification of statistic $\mathcal{L}(\mathbf{Y})$ of Theorem 1 is possible to yield a computationally fast statistic, just as [13, Theorem 1] simplifies [12, Theorem 7.6]. Assume that $N \geq 2p$, so that $\hat{\bar{\mathbf{\Sigma}}}^{-1}$ exists w.p.1.

Define the expectations

$$\bar{\mathbf{\Sigma}}_{\bar{V}} = \mathbb{E}\{\mathbf{y}_{\bar{V}}(t)\mathbf{y}^{\top}_{\bar{V}}(t)\} = \bar{\mathbf{\Sigma}} = \bar{\mathbf{\Omega}}^{-1} \tag{59}$$

$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,k,j+p,k+p\}} = \mathbb{E}\{\mathbf{y}_{\bar{V}\backslash\{j,k,j+p,k+p\}}(t) \times \mathbf{y}^{\top}_{\bar{V}\backslash\{j,k,j+p,k+p\}}(t)\} \tag{60}$$

$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,j+p\}} = \mathbb{E}\{\mathbf{y}_{\bar{V}\backslash\{j,j+p\}}(t)\mathbf{y}^{\top}_{\bar{V}\backslash\{j,j+p\}}(t)\} \tag{61}$$

$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{k,k+p\}} = \mathbb{E}\{\mathbf{y}_{\bar{V}\backslash\{k,k+p\}}(t)\mathbf{y}^{\top}_{\bar{V}\backslash\{k,k+p\}}(t)\}\,. \tag{62}$$

In general, for subsets A and B of $\bar{V}$, we use the notation

$$\bar{\mathbf{\Sigma}}_{A,B} = \mathbb{E}\{\mathbf{y}_A(t)\mathbf{y}^{\top}_B(t)\},\ \bar{\mathbf{\Sigma}}_A = \mathbb{E}\{\mathbf{y}_A(t)\mathbf{y}^{\top}_A(t)\}\,. \tag{63}$$

Lemma 7: For graph $\bar{\mathcal{G}} = (\bar{V}, \bar{\mathcal{E}})$ and associated $\mathbf{y} \in \mathbb{R}^{2p}$ considered in Lemma 2, we have

$$\frac{|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,k,j+p,k+p\}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}}|}{|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,j+p\}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{k,k+p\}}|} = \frac{|\bar{\mathbf{\Omega}}^{-1}_{(j)}|}{|[\bar{\mathbf{\Omega}}^{-1}_{(jk)}]_{1:2,1:2}|} \tag{64}$$

where, with $\bar{\Omega}_{\ell m}$ denoting the (l,m)-th entry of $\bar{\mathbf{\Omega}}$,

$$\bar{\mathbf{\Omega}}_{(j)} := \begin{bmatrix} \bar{\Omega}_{jj} & \bar{\Omega}_{j(j+p)} \\ \bar{\Omega}_{(j+p)j} & \bar{\Omega}_{(j+p)(j+p)} \end{bmatrix} \tag{65}$$

$$\bar{\mathbf{\Omega}}_{(jk)} := \begin{bmatrix} \bar{\Omega}_{jj} & \bar{\Omega}_{j(j+p)} & \bar{\Omega}_{jk} & \bar{\Omega}_{j(k+p)} \\ \bar{\Omega}_{(j+p)j} & \bar{\Omega}_{(j+p)(j+p)} & \bar{\Omega}_{(j+p)k} & \bar{\Omega}_{(j+p)(k+p)} \\ \bar{\Omega}_{kj} & \bar{\Omega}_{k(j+p)} & \bar{\Omega}_{kk} & \bar{\Omega}_{k(k+p)} \\ \bar{\Omega}_{(k+p)j} & \bar{\Omega}_{(k+p)(j+p)} & \bar{\Omega}_{(k+p)k} & \bar{\Omega}_{(k+p)(k+p)} \end{bmatrix} \tag{66}$$

and $[\bar{\mathbf{\Omega}}^{-1}_{(jk)}]_{1:2,1:2}$ is a 2×2 submatrix of $\bar{\mathbf{\Omega}}^{-1}_{(jk)}$ comprising its first two rows and first two columns. •

Proof: We are interested in determinants of certain matrices. Any elementary row or column operation on a matrix does not change its determinant. Just for this proof, in dealing with $\bar{\mathbf{\Sigma}}$

and $\bar{\mathbf{\Omega}}$, we move the elements associated with the four vertices $\{j, k, j+p, k+p\}$ to the top left corner of the respective matrices, in the order $j, j+p, k, k+p$. The so modified matrices are still denoted by $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{\Omega}}$.

Consider the correspondence of the following conformably partitioned matrices

$$\bar{\mathbf{\Omega}} = \begin{bmatrix} \bar{\mathbf{\Omega}}_{(jk)} & \vdots & * \\ \cdots & \cdots & \cdots \\ * & \vdots & * \end{bmatrix} = \bar{\mathbf{\Sigma}}^{-1} = \begin{bmatrix} \bar{\mathbf{\Sigma}}_{(jk)} & \vdots & * \\ \cdots & \cdots & \cdots \\ * & \vdots & * \end{bmatrix}^{-1} \tag{67}$$

where $\bar{\mathbf{\Sigma}}_{(jk)}$ is defined using $\bar{\mathbf{\Sigma}}$ just as $\bar{\mathbf{\Omega}}_{(jk)}$ is defined from $\bar{\mathbf{\Omega}}$ in (66). For ease of notation, redefine some subsets of $\bar{V}$ as

$$\begin{aligned} V_{-(jk)} &:= \bar{V}\backslash\{j,k,j+p,k+p\} \\ V_{(jk)} &:= \{j,k,j+p,k+p\} \\ V_{-(j)} &:= \bar{V}\backslash\{j,j+p\}, \quad V_{(j)} := \{j,j+p\} \\ V_{-(k)} &:= \bar{V}\backslash\{k,k+p\}, \quad V_{(k)} := \{k,k+p\}\,. \end{aligned} \tag{68}$$

In this notation, we rewrite (59)-(62) as

$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,k,j+p,k+p\}} = \bar{\mathbf{\Sigma}}_{V_{-(jk)}} \tag{69}$$

$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,j+p\}} = \bar{\mathbf{\Sigma}}_{V_{-(j)}} \tag{70}$$

$$\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{k,k+p\}} = \bar{\mathbf{\Sigma}}_{V_{-(k)}}\,. \tag{71}$$

From [1, Eqn. (B.2)] and (67), we have (recall the notation in (63))

$$\bar{\mathbf{\Omega}}^{-1}_{(jk)} = \bar{\mathbf{\Sigma}}_{V_{(jk)}} - \bar{\mathbf{\Sigma}}_{V_{(jk)},V_{-(jk)}} \bar{\mathbf{\Sigma}}^{-1}_{V_{-(jk)}} \bar{\mathbf{\Sigma}}_{V_{-(jk)},V_{(jk)}}. \tag{72}$$

By [1, Proposition C.5], the right-side of (72) is the conditional covariance

$$\begin{aligned} \bar{\mathbf{\Sigma}}_{V_{(jk)}|V_{-(jk)}} &= \mathbb{E}\{\mathbf{z}_{V_{(jk)}}(t)\mathbf{z}^{\top}_{V_{(jk)}}(t)|\mathbf{z}_{V_{-(jk)}}(t)\} \\ &= \bar{\mathbf{\Sigma}}_{V_{(jk)}} - \bar{\mathbf{\Sigma}}_{V_{(jk)},V_{-(jk)}} \bar{\mathbf{\Sigma}}^{-1}_{V_{-(jk)}} \bar{\mathbf{\Sigma}}_{V_{-(jk)},V_{(jk)}}. \end{aligned} \tag{73}$$

Thus

$$\bar{\mathbf{\Omega}}^{-1}_{(jk)} = \bar{\mathbf{\Sigma}}_{V_{(jk)}|V_{-(jk)}}. \tag{74}$$

Similarly,

$$\bar{\mathbf{\Omega}}^{-1}_{(j)} = \bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(j)}}. \tag{75}$$

By [1, Eqn. (B.1)] concerning determinant of partitioned matrices, we have

$$\begin{aligned} |\bar{\mathbf{\Sigma}}_{\bar{V}}| &= |\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,j+p\}}|\,|\bar{\mathbf{\Sigma}}_{\{j,j+p\}|\bar{V}\backslash\{j,j+p\}}| \\ &= |\bar{\mathbf{\Sigma}}_{V_{-(j)}}|\,|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(j)}}|\,. \end{aligned} \tag{76}$$

Similarly, we have

$$|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{k,k+p\}}| = |\bar{\mathbf{\Sigma}}_{V_{-(k)}}| = |\bar{\mathbf{\Sigma}}_{V_{-(jk)}}|\,|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(jk)}}|\,. \tag{77}$$

Using notation (68), (76) and (77), we have

$$\begin{aligned} \frac{|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,k,j+p,k+p\}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}}|}{|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{j,j+p\}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}\backslash\{k,k+p\}}|} &= \frac{|\bar{\mathbf{\Sigma}}_{V_{-(jk)}}|\,|\bar{\mathbf{\Sigma}}_{\bar{V}}|}{|\bar{\mathbf{\Sigma}}_{V_{-(j)}}|\,|\bar{\mathbf{\Sigma}}_{V_{-(k)}}|} \\ &= \frac{|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(j)}}|}{|\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(jk)}}|}\,. \end{aligned} \tag{78}$$

Since $\bar{\mathbf{\Omega}}^{-1}_{(jk)} = \bar{\mathbf{\Sigma}}_{V_{(jk)}|V_{-(jk)}}$ by (74), we have

$$\bar{\mathbf{\Sigma}}_{V_{(j)}|V_{-(jk)}} = [\bar{\mathbf{\Omega}}^{-1}_{(jk)}]_{1:2,1:2}\,. \tag{79}$$

Using (75), (78) and (79), we have the desired result. ■

Assume that $N \geq 2p$, so that $\hat{\bar{\mathbf{\Sigma}}}^{-1}$ exists w.p.1. Lemma 8 then follows just as Lemma 7, by manipulations of $\hat{\bar{\mathbf{\Sigma}}}^{-1}$ and submatrices of $\hat{\bar{\mathbf{\Sigma}}} = \hat{\bar{\mathbf{\Sigma}}}_{\bar{V}}$.

Lemma 8: An alternative expression for the test statistic of Theorem 1 is given by

$$\frac{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,k,j+p,k+p\}}|\,|\hat{\bar{\mathbf{\Sigma}}}|}{|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{j,j+p\}}|\,|\hat{\bar{\mathbf{\Sigma}}}_{\bar{V}\backslash\{k,k+p\}}|} = \frac{|\hat{\bar{\mathbf{\Omega}}}^{-1}_{(j)}|}{|[\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jk)}]_{1:2,1:2}|} \tag{80}$$

where, with $\hat{\bar{\Omega}}_{\ell m}$ denoting the (l,m)-th entry of $\hat{\bar{\mathbf{\Omega}}} = \hat{\bar{\mathbf{\Sigma}}}^{-1}$,

$$\hat{\bar{\mathbf{\Omega}}}_{(j)} := \begin{bmatrix} \hat{\bar{\Omega}}_{jj} & \hat{\bar{\Omega}}_{j(j+p)} \\ \hat{\bar{\Omega}}_{(j+p)j} & \hat{\bar{\Omega}}_{(j+p)(j+p)} \end{bmatrix} \tag{81}$$

$$\hat{\bar{\mathbf{\Omega}}}_{(jk)} := \begin{bmatrix} \hat{\bar{\Omega}}_{jj} & \hat{\bar{\Omega}}_{j(j+p)} & \hat{\bar{\Omega}}_{jk} & \hat{\bar{\Omega}}_{j(k+p)} \\ \hat{\bar{\Omega}}_{(j+p)j} & \hat{\bar{\Omega}}_{(j+p)(j+p)} & \hat{\bar{\Omega}}_{(j+p)k} & \hat{\bar{\Omega}}_{(j+p)(k+p)} \\ \hat{\bar{\Omega}}_{kj} & \hat{\bar{\Omega}}_{k(j+p)} & \hat{\bar{\Omega}}_{kk} & \hat{\bar{\Omega}}_{k(k+p)} \\ \hat{\bar{\Omega}}_{(k+p)j} & \hat{\bar{\Omega}}_{(k+p)(j+p)} & \hat{\bar{\Omega}}_{(k+p)k} & \hat{\bar{\Omega}}_{(k+p)(k+p)} \end{bmatrix} \tag{82}$$

and $[\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jk)}]_{1:2,1:2}$ is a 2×2 submatrix of $\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jk)}$ in its top left corner. •

Using log-likelihood we have the following GLRT

$$\begin{aligned} Y &:= \ln(\mathcal{L}(\mathbf{Y})) \\ &= \frac{N}{2}\left[\ln\left(|[\hat{\bar{\mathbf{\Omega}}}^{-1}_{(jk)}]_{1:2,1:2}|\right) - \ln\left(|\hat{\bar{\mathbf{\Omega}}}^{-1}_{(j)}|\right)\right] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1 \end{aligned} \tag{83}$$

where τ_1 is picked to achieve a specified probability of false alarm (significance level) P_{fa}. We summarize the above results in Theorem 2.

Theorem 2. The GLRT for the test (32) is given by (83), where $\hat{\bar{\mathbf{\Omega}}}_{(j)}$ and $\hat{\bar{\mathbf{\Omega}}}_{(jk)}$ are given by (81) and (82), respectively. Under $\mathcal{H}_0$, the test statistic Y has pdf (57), which can be approximated by (58) for $N \gg p$. •

Remark 1: Suppose that we need to carry out edge exclusion tests for all $p(p-1)/2$ edges in $\mathcal{E}$. Then it is established easily that the computational complexity of our proposed alternative statistic is $\mathcal{O}(p^3)$ versus $\mathcal{O}(p^5)$ for the original expression of Theorem 1. □

VI. Numerical Experiments

Given $\mathbf{x}(t) \in \mathbb{C}^p$, there are $p(p-1)/2$ unordered pair of vertices in the associated CIG that may or may not be connected. So we have to perform at least $p(p-1)/2$ binary hypothesis tests (a given edge is missing from the graph is the null, and the complete graph is the alternative hypothesis). Thus, we have a multiple testing problem where the main issue is how to control the overall significance level. Instead, we will use the ROC (receiver operating characteristic) averaged over all edges, as a performance measure. The ROC curve is the trade-off between the false-alarm rate P_{fa} and the detection probability P_d. Here we follow [34, Sec. VII] where trade-off between average type I (false-alarm rate) and type II

(miss probability) errors (over all edges) has been used as a performance measure (in the context of time series).

We consider two models for generating the synthetic data.

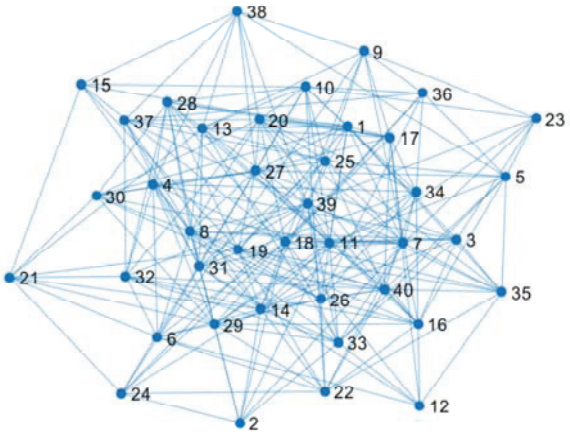

Fig. 1: *A typical realization of true CIG: model 1, p =40, k =5*

Model 1. 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]. Since an off-diagonal element of $\mathbf{\Omega}$ is zero with probability q, approximately $100q\%$ of the entries of $\mathbf{\Omega}$ are null. A typical realization of the graph (for $\mathbf{x}$) is in Fig. 1 for $p = 40$ and $q = 0.7$, where only the nodes that are connected to at least one another node, are shown. With $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{\Omega}^{-1}$, we generate $\mathbf{x} = \mathbf{\Phi}\mathbf{w}$ 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 $w_l = \text{real}(v_l)+j(0.9\times\text{real}(v_l)+0.2\times\text{imag}(v_l))$, yielding improper complex Gaussian $\mathbf{w}$. We generate N i.i.d. observations from $\mathbf{x}$, with $q = 0.7$.

Model 2. Following [34, Appendix], initially let $\mathbf{A} \in \mathbb{R}^{p\times p}$ have all zero entries. Then, for a fixed k, elements in position (i, j) of $\mathbf{A}$ for which $(i + j)_{mod\, k} = 1$, are randomly sampled from $\mathcal{N}_r(0, 1)$ (standard Gaussian) distribution. Set $\mathbf{\Omega} = \mathbf{A}^\top\mathbf{A} + \beta\mathbf{I}$ with $\beta = 0.01\lambda_{max}$, where λ_{max} denotes the maximum eigenvalue of $\mathbf{A}^\top\mathbf{A}$. With $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{\Omega}^{-1}$, we generate $\mathbf{x} = \mathbf{\Phi}\mathbf{w}$, with all other details as for Model 1. The choice of k controls the sparsity: choosing $k = 5$ makes 80% entries of the $\mathbf{\Phi}$ matrix zero for $p = 40$, while choosing $k = 20$ makes 94% entries of the $\mathbf{\Phi}$ matrix zero. A typical realization of the graph (for $\mathbf{x}$) is in Fig. 2 for $p = 40$ and $k = 5$. In simulations we selected k randomly and uniformly over $5 \le k \le 20$ in each run.

We apply the GLRT (83) coupled with Theorem 2 and Lemma 2, and pdf (57), to compute the test threshold for a given significance level α, to test each of $p(p-1)/2$ edges.

The simulation results are shown in Figs. 3 and 4 for model 1, and in Figs. 5 and 6 for model 2, based on 1000 runs and

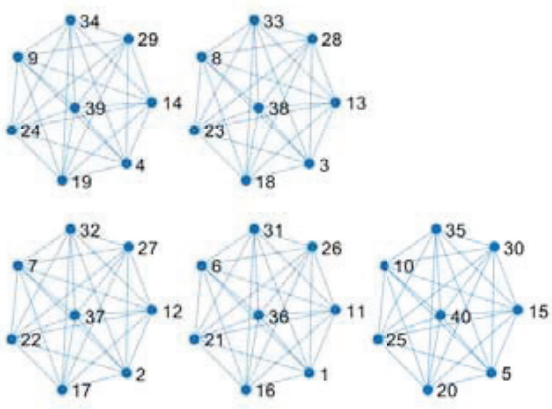

Fig. 2: *A typical realization of true CIG: model 2, p =40*

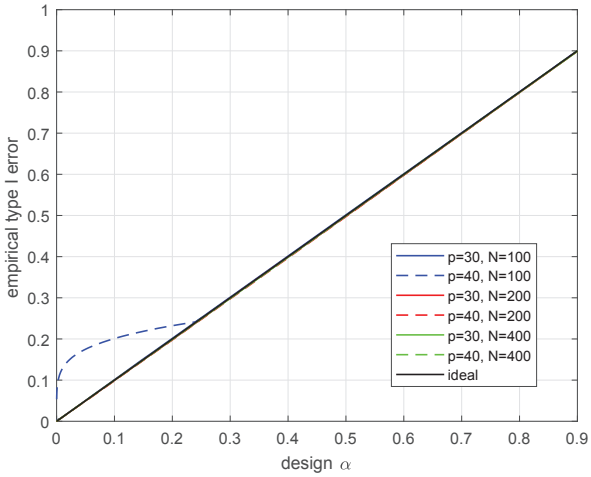

Fig. 3: *Model 1, improper Gaussian, p =30 or 40, N =100, 200 or 400. Empirical significance level vs. design significance level.*

averaged over all edges, with $N = 100$, 200 and 400, and $p = 30$ and 40 ($\mathbf{y} \in \mathbb{R}^{60}$ or $\mathbb{R}^{80}$), using a different randomly generated model in each run. It is seen from Figs. 3 and 5 that empirical α tracks the design α quite well, except for the case of $p = 40$, $N = 100$, where the 80×80 data correlation matrix of augmented $\mathbf{y}$ is estimated from 100 samples. The estimated correlation matrix, though of full rank w.p.1, is ill-conditioned, and its inverse needed to implement (83) suffers from this ill-conditioning.

The ROC curves are shown in Figs. 4 and 6 for models 1 and 2, respectively, and they show that the performance improves with increasing sample size N and decreasing number of nodes p, for a given value of P_{fa} (i.e., significance level or type I error).

A. Applying Existing Methods

For a given edge in our improper complex Gaussian graphical model, we test for joint exclusion of four edges in an augmented real Gaussian graphical model. As noted in Sec. III-D, the single edge exclusion test (9) of [1, Sec. 5.3.3] can

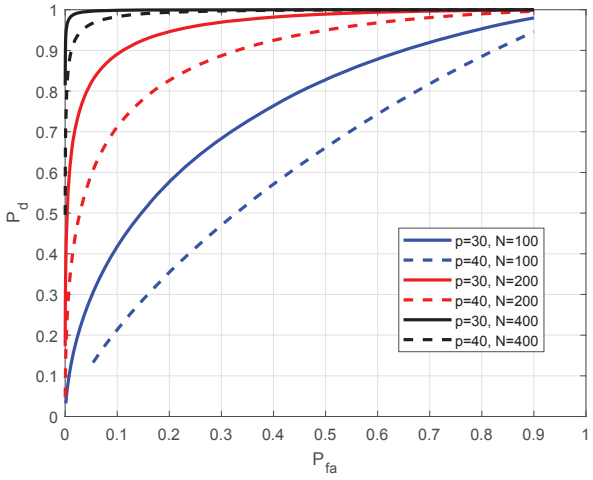

Fig. 4: *Model 1, improper Gaussian, p =30 or 40, N =100, 200 or 400. ROC curves*

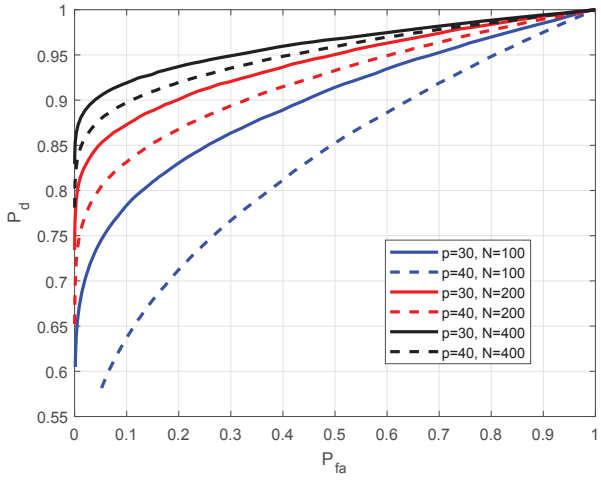

Fig. 6: *Model 2, improper Gaussian, p =30 or 40, N =100, 200 or 400. ROC curves*

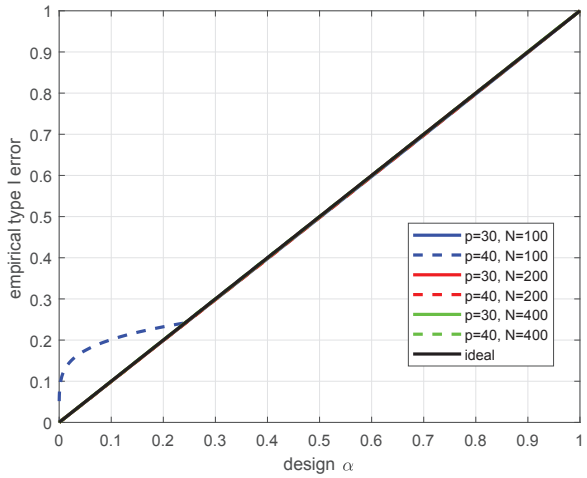

Fig. 5: *Model 2, improper Gaussian, p =30 or 40, N =100, 200 or 400. Empirical significance level vs. design significance level.*

be used to test these four edges but only in a multiple testing set-up where one tests each of the four edges separately, one-by-one. For zero-mean $\mathbf{x}$, under the null hypothesis, $\mathcal{L}_{RG}$ in (9) is distributed as $\mathcal{B}((N-2p+1)/2, 1/2)$. We applied multiple testing using (9) to data generated by model 1, based on two different methods of threshold selection:

(a) Bonferroni method ([31] and [38, Sec. 10.7]) where, to achieve significance level α as in our proposed approach, we pick significance level $\alpha' = \alpha/4$ for each of four applications of the test (9). The overall significance level then has an upper bound $4\alpha' = \alpha$.

(b) A naive approach where we set $\alpha' = \alpha$ for each of four applications of the test (9), and ignore the overall significance level which would be higher than α.

We also applied test (13) for proper complex Gaussian graphical models to data generated by model 1, by arbitrarily assuming that $\mathbf{x}$ is proper complex Gaussian. As shown in [13], under the null hypothesis, $-N\ln(\mathcal{L}_{CG})$ is an exponential random variable with mean $N/(N-p+1)$, provided $\mathbf{x}$ is proper complex Gaussian. Here we do not need multiple testing.

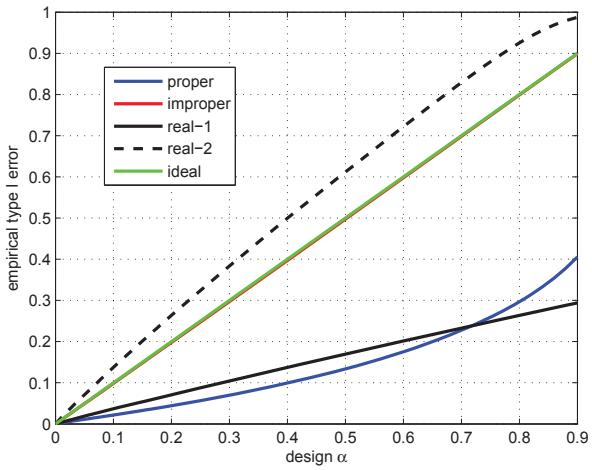

Fig. 7: *Model 1, improper Gaussian, $p = 40$, $N = 200$. Empirical significance level vs. design significance level. Labels "proper" and "improper" refer to tests (13) and (83), respectively, and labels "real-1" and "real-2" refer to test (9) with threshold selected as in items (a) and (b), respectively, of Sec. VI-A.*

The simulation results based on 1000 Monte Carlo runs, averaged over all edges, are shown in Figs. 7-9 for model 1 with $p = 40$ and $N = 200$. In these figures, labels "proper" and "improper" refer to tests (13) and (83), respectively, and labels "real-1" and "real-2" refer to test (9) with threshold selected as in items (a) and (b), respectively, in the preceding paragraph. It is seen from Fig. 7 that only the proposed approach delivers the design significance level. Test (13) cannot do so as it incorrectly assumes that $\mathbf{x}$ is proper. Test (9) based on Bonferroni method of multiple testing keeps the

empirical significance level α below the upper bound of design value, as expected, but there is a wide gap between the design value and the empirical significance level. Test (9) based on selecting $\alpha' = \alpha$ and ignoring the overall significance level, yields an empirical significance level α that far exceeds the design value.

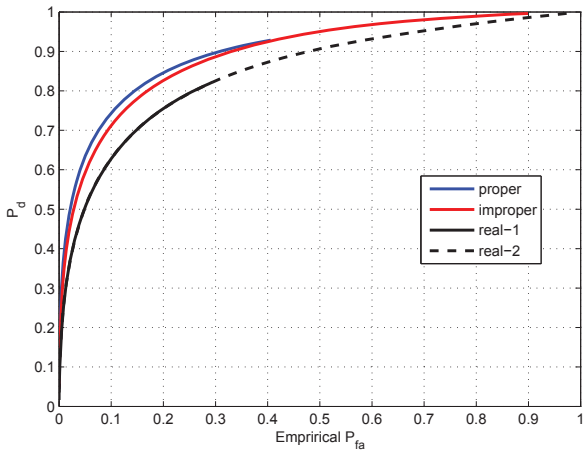

Fig. 8: *Model 1, improper Gaussian,* $p = 40,\ N = 200$. P_d *vs empirical* P_{fa}. *The curve labels are as for Fig. 7.*

Because there is a discrepancy between empirical and design P_{fa} (level α) for all tests except for the proposed test (83), we show two ROC curves, P_d versus empirical P_{fa} in Fig. 8, and P_d versus design P_{fa} in Fig. 9. The performance of test (13) shown in Fig. 8 cannot be achieved since the test threshold corresponding to empirical P_{fa} cannot be determined, either analytically or via simulations. If one uses the test threshold corresponding to design P_{fa} which can be obtained, the performance shown in Fig. 9 for test (13) is significantly worse than that of the proposed test (83). Turning to test (9), with Bonferroni method based threshold selection, its performance (labeled "real-1") shown in Fig. 9 is significantly worse than that of the proposed test (83)). Its performance shown in Fig. 8 is better than that in Fig. 9, but the former cannot be achieved as the test threshold corresponding to empirical P_{fa} cannot be determined, either analytically or via simulations. Performance of test (9) implemented via naive selection of test threshold (curves labeled "real-2") shown in Fig. 9 appears to be quite good but as seen from Fig. 7, empirical P_{fa} corresponding to the design P_{fa} significantly exceeds the design value, and this fact accounts for its apparent good performance in Fig. 9. Finally, its performance shown in Fig. 8 cannot be achieved as the test threshold corresponding to empirical P_{fa} cannot be determined. Thus, only the proposed test (83) and test (9) with Bonferroni method based threshold selection, can be implemented in practice, and comparing these two tests, as shown in Fig. 9, (83) significantly outperforms (9) in that for a given design P_{fa}, power of (83) is much higher than that of (9).

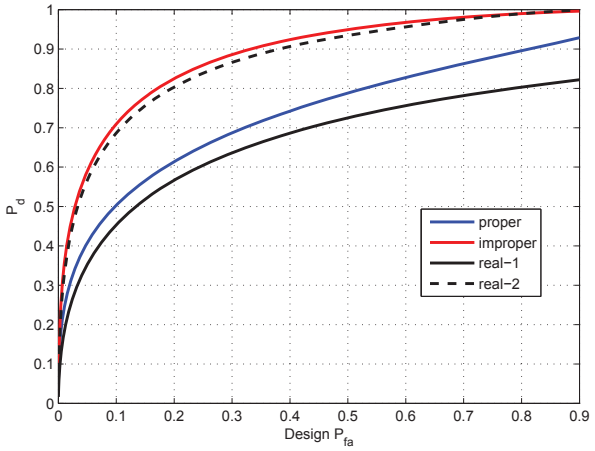

Fig. 9: *Model 1, improper Gaussian,* $p = 40,\ N = 200$. P_d *vs design* P_{fa}. *The curve labels are as for Fig. 7.*

VII. Conclusions

We considered the problem of inferring the conditional independence graph of improper complex-valued multivariate Gaussian vectors. For real random vectors, considerable body of work exists where one first tests for exclusion of each edge from the saturated model, and then infers the CIG. Existing work on proper complex Gaussian graphical models is sparse, while that on improper complex Gaussian graphical models is non-existent. We proposed and analyzed a GLRT-based edge exclusion test statistic for improper complex Gaussian graphical models. The test statistic was also expressed in an alternative form, where for p-dimensional random vectors, the alternative statistic reduces the computational complexity from $\mathcal{O}(p^5)$ to $\mathcal{O}(p^3)$. Simulation examples were presented to illustrate the proposed statistic.

VIII. Acknowledgment

I would like to thank three reviewers for their helpful comments and suggestions that significantly improved the paper.

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] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional data*. Berlin: Springer, 2011.
[4] 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.
[5] A.O. Hero, III, and B. Rajaratnam, "Foundational principles for large-scale inference: Illustrations through correlation mining," *Proc. IEEE*, vol. 64, no. 104, pp. 93-110, Jan. 2016.
[6] 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.
[7] N. Friedman, "Inferring cellular networks using probabilistic graphical models," *Science*, vol 303, pp. 799-805, 2004.
[8] S.L. Lauritzen and N.A. Sheehan, "Graphical models for genetic analyses," *Statistical Science*, vol. 18, pp. 489-514, 2003.

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

[10] 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.

[11] 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.

[12] 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.

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

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

[15] J.K. Tugnait, "An edge exclusion test for graphical modeling of multivariate time series," in *Proc. 2018 52nd Annual Conf. on Information Sciences & Systems (CISS)*, Princeton University, Princeton, NJ, March 21-23, 2018, pp. 1-6.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] S. Ryali, T. Chen, K. Supekar and V. Menon, "Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty," *NeuroImage*, vol. 59, pp. 3852-3861, 2012.

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

[22] 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.

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

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

[25] 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.

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

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

[28] 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.

[29] P.W.F. Smith and J. Whittaker, "Edge exclusion tests for graphical Gaussian models," in *Learning in Graphical Models*, M.I. Jordan (Ed), pp. 555-574, Cambridge, MA: MIT Press, 1998.

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

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

[32] M.F. Salgueiro, P.W.F. Smith and J.W. McDonald, "Power of edge exclusion tests in graphical Gaussian models," *Biometrika*, vol. 92, pp. 173-182, March 2005.

[33] Y. Matsuda, "A test statistic for graphical modelling of multivariate time series," *Biometrika*, vol. 93, no. 2, pp. 399-409, 2006.

[34] 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.

[35] A.P. Dempster, "Covariance selection," *Biometrics*, vol. 28, pp. 157-175, 1972.

[36] J. Duchi, S. Gould and D. Koller, "Projected subgradient methods for learning sparse Gaussians," in *Proc. 24th Conf. Uncertainty Artificial Intelligence (UAI'08)*, Helsinki, Finland, July 2008, pp. 153-160.

[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.

[38] L. Wasserman, *All of Statistics*, New York, NY: Springer, 2004.

[39] D.K. Nagar and E. Zarrazola, "Distributions of the product and the quotient of independent Kummer-beta variables," *Scientiae Mathematicae Japonicae Online*, e-2004, pp. 45-53, 2004.

[40] M. Bibinger, "Notes on the sum and maximum of independent exponentially distributed random variables with different scale parameters," arXiv:1307.3945v1 [math.PR], July 2013.

Jitendra K. Tugnait (M'79–SM'93–F'94–LF'16) received the B.Sc.(Hons.) degree in electronics and electrical communication engineering from the Punjab Engineering College, Chandigarh, India in 1971, the M.S. and the E.E. degrees from Syracuse University, Syracuse, NY and the Ph.D. degree from the University of Illinois, Urbana-Champaign in 1973, 1974, and 1978, respectively, all in electrical engineering.

From 1978 to 1982 he was an Assistant Professor of Electrical and Computer Engineering at the University of Iowa, Iowa City, IA. He was with the Long Range Research Division of the Exxon Production Research Company, Houston, TX, from June 1982 to Sept. 1989. He joined the Department of Electrical & Computer Engineering, Auburn University, Auburn, AL, in September 1989 as a Professor, where he is now James B. Davis Professor. His current research interests are in statistical signal processing, wireless physical and secure communications, and multiple target tracking.

Dr. Tugnait has served as Associate Editor of the IEEE Transactions on Automatic Control, the IEEE Transactions on Signal Processing, IEEE Signal Processing Letters, and the IEEE Transactions on Wireless Communications, and as Senior Area Editor of the IEEE Transactions on Signal Processing, and Senior Editor of IEEE Wireless Communications Letters.