Learning Networks from Wide-Sense Stationary Stochastic Processes

Anirudh Rayas, Jiajun Cheng, Rajasekhar Anguluri, Deepjyoti Deka, and Gautam Dasarathy

***Abstract*—Complex networked systems driven by latent inputs are common in fields like neuroscience, finance, and engineering. A key inference problem here is to learn edge connectivity from node outputs (potentials). We focus on systems governed by steady-state linear conservation laws: $X_t = L^* Y_t$, where $X_t, Y_t \in \mathbb{R}^p$ denote inputs and potentials, respectively, and the sparsity pattern of the $p \times p$ Laplacian L^* encodes the edge structure. Assuming X_t to be a wide-sense stationary stochastic process with a known spectral density matrix, we learn the support of L^* from temporally correlated samples of Y_t via an ℓ_1-regularized Whittle's maximum likelihood estimator (MLE). The regularization is particularly useful for learning large-scale networks in the high-dimensional setting where the network size p significantly exceeds the number of samples n.**

We show that the MLE problem is strictly convex, admitting a unique solution. Under a novel mutual incoherence condition and certain sufficient conditions on (n, p, d), we show that the ML estimate recovers the sparsity pattern of L^* with high probability, where d is the maximum degree of the graph underlying L^*. We provide recovery guarantees for L^* in element-wise maximum, Frobenius, and operator norms. Finally, we complement our theoretical results with several simulation studies on synthetic and benchmark datasets, including engineered systems (power and water networks), and real-world datasets from neural systems (such as the human brain).

***Index Terms*—Network topology inference, Conservation laws, ℓ_1-regularized Whittle's likelihood estimator, Spectral precision matrix.**

I. Introduction

Complex networked systems, composed of nodes and edges that connect them are commonly used to model real-world systems in fields such as neuroscience, engineering, climate, and finance [1, 2]. We study networks governed by conservation laws that control edge flows; examples include current in electrical grids, fluids in pipelines, and traffic in transportation systems [3, 4]. In neuroscience, there is growing interest in identifying and understanding conservation laws [5, 6].

Networked systems driven by latent inputs (i.e., nodal injections) generate edge flows that are proportional to differences in node potentials. For example, in electrical networks, nodal current injections induce current flows that are proportional to potential differences between nodes. The overall dynamics of these edge flows are governed by conservation laws. Formally, for a network of size p, these dynamics are described by the balance equation $X = L^* Y$, where $L^* \in \mathbb{R}^{p \times p}$ is a weighted symmetric Laplacian matrix [7]. The off-diagonal entries of L^* capture the edge connectivity structure of the network. Vectors $X, Y \in \mathbb{R}^p$ represent nodal injections and potentials respectively, and in this paper, we treat them as random vectors. Further details on the balance equation are in Section II.

In various practical situations, the network's connectivity is typically not known and needs to be estimated for modeling, management, and control tasks. This involves determining the non-zero elements of the associated Laplacian matrix L^*. Previous methods such as [8] estimate L^* given observations of node injection-potential pairs $\{X, Y\}$ by minimizing an appropriate least squares objective. Such methods critically rely on the ability to observe both injections and potentials simultaneously. However, in various scenarios *node injections are often unobservable*. For instance, in financial or brain networks, nodal injections correspond to economic shocks or unknown stimuli, and these are not observable by the measurement system in place. In these settings, the goal is to estimate L^* with only samples of Y. Indeed, this problem is ill-posed as multiple solutions of X and L^* can satisfy the equation $X = L^* Y$. To address the ill-posedness, we assume we have access to some information about the distribution of X. The challenge of estimating L^* from Y under such assumptions have been previously studied in [9–11].

This line of work relies on the observations of the potentials being independent and identically distributed (i.i.d.). When temporal dependencies exist in the data, such methods are insufficient. In this paper, we adopt a more realistic data model and suppose that the nodal injections (X_t) and potentials (Y_t) are wide-sense stationary processes (WSS). This generalization allows for a more flexible framework for network learning while posing some interesting technical challenges. Before we outline our major contributions, we will first state the problem more formally and outline the challenges it presents.

Structure learning problem: *Given finite samples of node potentials $\{Y_t\}_{t=1}^n$ and assuming the node injections X_t are generated from a WSS process with known spectral density matrix, the goal is to recover the matrix $L^* \in \mathbb{R}^{p \times p}$ such that the estimate $\widehat{L}$ approximately satisfies the balance equation $X_t \approx \widehat{L} Y_t$.*

The structure learning problem stated above assumes that the spectral density matrix for the latent process X_t is known. As discussed earlier, estimating a sparse matrix L^* from observations $\{Y_t\}_{t=1}^n$ alone is fundamentally ill-posed (see Remark 3 for further discussion).

A. Rayas, J. Cheng, and G. Dasarathy are with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA. E-mails: ahrayas@asu.edu, ccheng@asu.edu, gautamd@asu.edu

R. Anguluri is with the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, MD 21250, USA. Email: rajangul@umbc.edu

D. Deka is with the Energy Initiative at Massachusetts Institute of Technology, USA. Email: deepj87@mit.edu

A common approach in related work is to assume access to samples of the latent process X_t [8, 12]. In such a scenario, the spectral density matrix of X_t can be estimated and subsequently L^*. However, access to samples from X_t is unreasonable in many domains such as neuroscience, finance, and biology, where X_t represents unobservable external inputs (e.g., latent external stimuli or economic shocks). An alternative assumption used in latent factor and structural equation models (SEMs) is to assume that the spectral density of X_t is diagonal [13, 14]. However, this assumption is overly restrictive, as real-world exogenous inputs typically exhibit temporal and cross-sectional correlation [15].

To address these limitations, we assume access to the full spectral density matrix of X_t, without imposing diagonality. This standard assumption [16, 17] accommodates correlated latent inputs while still ensuring identifiability of L^*.

Its practical relevance is illustrated in two scenarios. In social networks, Y_t may represent individuals' opinions and X_t their latent beliefs. Though X_t is unobserved, its second-order statistics can be modeled by exploiting homophily (i.e., individuals with similar attributes hold correlated beliefs) [18]. In financial networks, Y_t reflects stock prices driven by investor activity X_t, which are typically unobservable due to privacy concerns. However, many companies release second-order statistical summary information $\mathbb{E}[X_t X_t^\mathsf{T}]$ [19].

Although the structure learning problem can be addressed through a two-step process—first estimating the spectral density of Y_t from $\{Y_t\}_{t=1}^n$, and then estimating L^* from the spectral density of X_t—this approach is statistically inefficient, even when Y_t is i.i.d., this is elaborated in Remark 4 of [9]. To overcome these limitations, we propose a novel single-step estimator for L^* that integrates finite time-series data with constraints imposed by conservation laws. Our method also ensures consistent estimation of L^* in the high-dimensional setting where the number of samples n is significantly smaller than the network size p (i.e., $n \ll p$). This requires that L^* is sparse, which is natural in all of our motivating examples: power grids, social networks, and brain connectivity graphs are inherently sparse, with nodes connected to only a small subset of others. We now provide a high-level overview of our methodology.

Suppose that $\{X_t\}_{t\in\mathbb{Z}}$ is a WSS process with a complex-valued power spectral density matrix $f_X(\omega)$ with $\omega \in [-\pi, \pi]$ (see (3) for a formal definition). The conservation law dictates the spectral density $f_Y(\omega)$ of $\{Y_t\}_{t\in\mathbb{Z}}$ to satisfy $f_X(\omega) = L^* f_Y(\omega)(L^*)^\mathsf{T}$. Given samples from the node potential process $\{Y_t\}_{t=1}^n$ and assuming that $f_X(\omega)$ is known (this is all we know about X), consider the optimization problem:

$$\begin{aligned} \underset{L\in\mathbb{R}^{p\times p}}{\text{maximize}} \quad & \mathcal{L}[\{Y_t\}_{t=1}^n; f_X(\omega)] + \lambda_n \|L\|_1 \\ \text{subject to} \quad & f_X(\omega) = L f_Y(\omega) L^\mathsf{T}, \quad \omega \in [-\pi, \pi], \end{aligned} \tag{1}$$

where $\mathcal{L}[\cdot]$ is an appropriate log-likelihood that measures the fit to observed data, and $\lambda_n \geq 0$ is a regularization parameter. The ℓ_1-norm $\|\cdot\|_1$ (which is the entry-wise absolute sum) helps promote sparsity in our estimate of L^*. Full details of (1) are in Section II. While such optimization problems that target sparse matrix estimation have received considerable attention in the literature (see Sections V and I-B for a brief overview), (1) presents some unique challenges:

i) $\{Y_t\}_{t=1}^n$ is not i.i.d., making standard sample covariance matrix style analyses inapplicable;
ii) it involves a continuum of constraints since $\omega \in [-\pi, \pi]$, rendering (1) an infinite-dimensional optimization problem; and
iii) the constraint is non-convex for arbitrary matrices L, even when considering the symmetry of the Laplacian matrix.

Although a line of work [20–23] addresses challenges of the form (i) and (ii) separately in the context of learning Gaussian graphical models from time-series data, and [9] tackles challenge (iii), no prior work, to the best of our knowledge addresses all three challenges simultaneously. The goal of this paper is to show that despite these challenges, the optimizer of (1) captures the sparsity pattern of L^* with high probability. Thus, the optimizer of (1) is the estimator we seek to recover the sparse matrix L^*. This problem formulation is motivated by several applications where it plays a natural role; here we briefly outline two.

1) Topology learning in power distribution networks: Knowledge of network topology (or structure) enables better fault detection, efficient resource allocation, and better integration of decentralized energy resources, ensuring reliable operation of the power system. However, system operators may lack access to real-time topology information and use nodal voltages or current injections to learn the network topology. A balance equation of the form $X_t = L^* Y_t$, where L^* is the network admittance matrix and injected currents X_t modeled by a WSS process, has been considered in this context [24].

2) Learning sensor to source mapping in the human brain: Learning the mapping from source signals to EEG electrodes is crucial for analyzing brain connections. Many studies [25, 26] suggest a model of the form in (2). Specifically, the Laplacian matrix plays the role of lead-field matrix and the potentials Y_t are the EEG signals. The injections X_t model the latent source signals and are thought to be generated by a vector auto-regressive process (VAR(m)): $X_t = \sum_{k=1}^m A_k x_{t-k} + \epsilon_t$, where ϵ_t could be non-Gaussian; and the integer m and matrices A_k could be known or unknown. Thus, learning the source mapping involves learning L^* from WSS data.

A. Main contributions

1) A novel convex estimator: We propose an ℓ_1-regularized log-likelihood estimator of the form (1) to estimate L^* from finite samples of WSS data $\{Y_t\}_{t=1}^n$. This estimator builds on the Whittle log-likelihood approximation (details in Section II-B). Our first theoretical result establishes that the proposed ℓ_1-regularized estimator is convex in L and under standard conditions, admits a unique minimum even in the high-dimensional regime ($n \ll p$).

Since the Whittle likelihood is closely tied to the likelihood of Gaussian WSS processes, our estimator maximizes an approximate Gaussian likelihood. However, the estimator remains meaningful even for non-Gaussian injections $\{X_t\}_{t\in\mathbb{Z}}$, including stationary linear processes with sub-exponential or

finite fourth-moment error distributions (see the remark on Bregman divergence in Section II-B).

2) Sample complexity and estimation consistency: We provide sufficient conditions on the sample size n of the data $\{Y_t\}_{t=1}^n$ for the estimator to achieve two key properties: *sparsistency*, ensuring the recovery of the sparsity pattern of L^*, and *norm consistency*, providing error bounds in terms of element-wise maximum, Frobenius, and operator norms. Pivotal to our analysis is a novel irrepresentability-like condition on L^*, inspired by similar conditions commonly used in high-dimensional statistics [27, 28]. The sample complexity results are derived for both Gaussian and linear non-Gaussian WSS processes (see Theorem 1 and 2).

3) Experimental validation: We validate our theoretical results with extensive numerical experiments using synthetic and quasi-synthetic data from many benchmark networked systems, as well as a real-world dataset involving the brain network (see Section IV).

B. Related work

1) Structure learning in Gaussian graphical models (GGMs): The graph underlying a GGM can be inferred from the sparsity pattern of the inverse covariance matrix, and numerous papers have focused on learning this pattern from i.i.d. data (see [29] for an overview). Pioneering works like [30, 31] have developed key theoretical concepts for analyzing ℓ_1-regularized likelihood estimators, and our analysis builds on these concepts. Other works like [32, 33] focus on learning Cholesky factors of the inverse covariance matrix, but they lack theoretical guarantees. Survey papers like [34] provide a comprehensive overview of estimators for GGMs in various scenarios, including dynamic and grouped networks, while [35] presents detailed analyses of theoretical frameworks and sample complexity results for these models. However, these approaches face two significant limitations in our context. First, they are primarily designed for i.i.d. data, whereas the problem we address involves time-series data. Second, these methods aim to estimate the inverse covariance matrix, whereas our focus is to estimate the Laplacian L^* directly, bypassing the need to first estimate the inverse covariance matrix.

2) Graph signal processing (GSP): Recent research in GSP studied sparse inverse covariance estimation problems in GGMs by imposing Laplacian constraints. Both the regularized likelihood and spectral template-based (i.e., using eigenvectors of the sample covariance matrix) techniques are used to learn the Laplacian-constrained inverse covariance matrix [36–38]. However, many papers in this area focus only on estimation consistency or algorithmic convergence, but not on sample complexity. In our problem, the inverse covariance (or spectral density) matrix is represented as a quadratic matrix equation involving products of Laplacian matrices (see (1)), making existing methods in the cited works unsuitable for direct application. In addition, we provide sample complexity guarantees and establish precise rates of convergence for our proposed estimator.

3) Learning network structure from WSS process: Dahlhaus [39] showed that the sparsity pattern of the inverse spectral density (ISD) matrix represents the structure of the graphical model for a Gaussian WSS. Subsequently, many papers (see e.g., [20, 40]) have focused on estimating a sparse ISD matrix. Finally, a few more (see [21–23]) have focused on estimating parameter matrices of latent models (e.g., VAR or state-space) generating the ISD matrix. Our research falls into the latter category, with a parameter matrix that is a Laplacian of a conservation law. However, directly applying these methods often leads to a two-stage approach: first estimating the parameter matrix, followed by a refinement step to identify non-zero entries in L^*. In contrast, our estimator of the form in (1) directly estimates the Laplacian matrix L^*, thus avoiding the statistical inefficiencies inherent in the two-stage approach (see Section I-B1). Related streams of work have addressed latent-variable autoregressive graphical models using sparse + low-rank decompositions of the inverse spectral density [41–43], ARMA factor models using diagonal + low-rank structures [44, 45], and sparse reciprocal graphical models that impose block-circulant patterns [46].

While these approaches provide valuable insights, our problem setting is fundamentally different. We focus on estimating a general sparse Laplacian matrix associated with a conservation law constraint, using a single-step likelihood-based approach in the frequency domain. We do not assume latent-variable factorizations or additional structural constraints such as low-rankness or block-circulant structures. Importantly, we provide theoretical guarantees on the sample complexity required to achieve support recovery and to bound estimation error in matrix norms for this general setting. To the best of our knowledge, these guarantees have not been established in the aforementioned literature.

4) Electric power networks: While there are many motivating examples for this framework, the authors were specifically motivated by the problem of topology learning in power networks. For i.i.d. data, works like [47, 48] infer the sparsity pattern of the Laplacian (associated with a conservation law under linear power flow) by learning the inverse covariance of node potentials and applying algebraic rules. This approach requires minimum cycle length conditions on the network, which we do not need (see Remark 4). Survey papers like [49] provide a good overview of state-of-the-art methods, including the likelihood approaches in [50].

We now contrast this work with a related paper by a subset of the authors [9]. First, the estimator in [9] assumes i.i.d. Gaussian injections X_t, whereas the current work addresses non-i.i.d. X_t and considers a broader class of Gaussian and non-Gaussian WSS processes; we outlined the unique challenges in the discussion following equation (1). Second, our analysis requires a comprehensive examination of Hermitian matrices in the optimization problem, which is more complex than dealing solely with symmetric matrices, as in [9]. Third, we empirically validate the performance of our estimator, particularly regarding sample complexity and error consistency, across a wide range of networked systems, and compare it directly with the estimator proposed in [9].

Notation: Let $\mathbb{Z}, \mathbb{R}$, and $\mathbb{C}$ denote sets of integers, reals, and

complex numbers, respectively. For sets $T_1, T_2 \subset [p] \times [p]$, denote by $A_{T_1T_2}$ the submatrix of A with rows and columns indexed by T_1 and T_2. If $T_1 = T_2$, we denote the submatrix by A_{T_1}. For a matrix $A = [A_{i,j}]$, $\|A\|_F$ and $\|A\|_2$ denote the Frobenius and the operator norm; $\|A\|_\infty \triangleq \max_{i,j} |A_{ij}|$ and $\|A\|_{1,\text{off}} = \sum_{i \neq j} |A_{ij}|$. The ℓ_∞-matrix norm of A is defined as $\nu_A = |||A|||_\infty \triangleq \max_{j=1,\ldots,p} \sum_{j=1}^p |A_{ij}|$. We use $\text{vec}(A)$ to denote the p^2-vector formed by stacking the columns of A and $\Gamma(A) = (I \otimes A)$ to denote the Kronecker product of A with the identity matrix I. For two symmetric positive definite matrices A_1 and A_2, $A_1 \succ A_2$ means $A_1 - A_2$ is positive definite. We define $\text{sign}(A_{ij}) = +1$ if $A_{ij} > 0$ and $\text{sign}(A_{ij}) = -1$ if $A_{ij} < 0$. For two-real valued functions $f(\cdot)$ and $g(\cdot)$, we write $f(n) = \mathcal{O}(g(n))$ if $f(n) \leq cg(n)$ and $f(n) = \Omega(g(n))$ if $f(n) \geq c'g(n)$ for constants $c, c' > 0$.

Organization of the paper: In Section II, we define the structure learning problem and propose the modified ℓ_1-regularized Whittle likelihood estimator for learning a network structure from WSS data. Section III establishes the convexity of the proposed estimator and provides guarantees for support recovery and norm consistency for both Gaussian and non-Gaussian node injections X_t. In Section IV, we evaluate the performance of our estimator on synthetic, benchmark, and real-world datasets. Section V emphasizes the parallels that our structure learning framework shares by drawing connections to other learning problems in the literature. Finally, Section VI concludes with a summary and outlines future directions. Proofs of theoretical results and additional experimental details are provided in the supplementary material. Throughout, we use *estimation* and *learning* interchangeably, as well as *network* and *graph.*

II. Preliminaries and Problem Setup

For directed graph $\mathcal{G} = ([p], E)$, where the node set is defined as $[p] \triangleq \{1, 2, \ldots, p\}$ and the edge set is $E \subseteq [p] \times [p]$, let $\mathcal{D}$ denote the $p \times |E|$ *incidence matrix*. Each column of $\mathcal{D}$ corresponds to an edge (i, j) and is populated with zeros except at the i-th and j-th positions, where it takes the values -1 and $+1$, respectively. Suppose $X \in \mathbb{R}^p$ denotes the vector of *node injections*. The basic *conservation law* is given by: $\mathcal{D}f + X = 0$, where $f \in \mathbb{R}^{|E|}$ is the vector of *edge flows*. This law states that the sum of flows over the edges incident to a vertex equals the injected flow at that vertex. In other words, edge and injected flows are conserved.

In physical systems, edge flows are determined by *potentials* $Y \in \mathbb{R}^p$ at the vertices. Under natural linearity assumptions, the edge flow on the (i, j)-th edge is proportional to $Y_j - Y_i$. For all edges, $f = -\mathcal{D}^\mathsf{T} Y$. Substituting this edge flow relation in the basic conservation law yields the *balance equation*:

$$X - L^* Y = 0, \tag{2}$$

where $L^* \triangleq \mathcal{D}\mathcal{D}^\mathsf{T}$ is the $p \times p$ real-valued symmetric Laplacian matrix. A typical system satisfying (2) is an electrical network with unit resistances, where Y represents voltage potentials, f edge currents, and X injected currents. For examples involving hydraulic, social, and transportation systems, see [3, 4].

A. Structure learning problem

The sparsity pattern (locations of zero and non-zero entries) of L^* reflects the edge connectivity of the underlying network. Specifically, $(i, j) \in E$ if and only if $L^*_{ij} \neq 0$. Our goal is to learn the unknown edge set E (or the sparsity pattern of L^*) from data collected at the nodes of the graph.

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a zero-mean p-dimensional vector-valued WSS process, where, for each $t \in \mathbb{Z}$, $X_t = (X_{t1}, \ldots, X_{tp})^\mathsf{T} \in \mathbb{R}^p$. The auto-covariance function of this process is $\Phi_X(l) \triangleq \mathbb{E}[X_t X_{t-l}^\mathsf{T}]$, for all $t \in \mathbb{Z}$ and $l \in \mathbb{Z}$ is the lag parameter. We assume that $\Phi_X(l) \succ 0$. Because $\{X_t\}_{t \in \mathbb{Z}}$ is WSS, it holds that $\|\Phi_X(l)\|_2 < \infty$. Hence, the power spectral density (PSD) function of $\{X_t\}_{t \in \mathbb{Z}}$ exists and is defined via the discrete-time Fourier transform of $\Phi_X(l)$:

$$f_X(\omega) \triangleq \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \Phi_X(l) e^{-\mathrm{i}\, l\omega}, \quad \omega \in [-\pi, \pi], \tag{3}$$

where $\mathrm{i} = \sqrt{-1}$ and $f_X(\omega) \in \mathbb{C}^{p \times p}$ is a Hermitian positive definite matrix. Let $\Theta_X(\omega) \triangleq f_X^{-1}(\omega)$ be the inverse PSD.

Let $\{Y_t\}_{t \in \mathbb{Z}}$ be generated per the balance equation in (2). We want to obtain a sparse estimate of L^* using the finite time-series potential data $\{Y_t\}_{t=1}^n$ and only the nodal injection's inverse PSD matrix $\Theta_X(\omega)$; see Remark 3. We emphasize that our processes need not be Gaussian. A major challenge in developing maximum-likelihood parameter estimates from time-series data is obtaining tractable likelihood formulas. Whittle [51] developed a good approximation for the Gaussian case, and the later work extended this approach to other cases. Following [20], we provide likelihood approximations for $\{Y_t\}_{t=1}^n$.

B. Modified Whittle's likelihood approximation

Suppose that L^* is invertible (see Remark 1), the equation in (2) simplifies to $Y_t = (L^*)^{-1} X_t$. Due to this linear relationship, $\{Y_t\}_{t \in \mathbb{Z}}$ is also a WSS process with the auto-covariance matrix:

$$\Phi_Y(l) \triangleq \mathbb{E}[Y_t, Y_{t-l}^\mathsf{T}] = (L^*)^{-1} \Phi_X(l) (L^*)^{-1},$$

and the PSD matrix:

$$f_Y(\omega) \triangleq \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \Phi_Y(l) e^{-\mathrm{i}\, l\omega} = (L^*)^{-1} f_X(\omega) (L^*)^{-1}, \tag{4}$$

where $\omega \in [-\pi, \pi]$. Finally, define the inverse PSD matrix:

$$\Theta_Y(\omega) \triangleq f_Y^{-1}(\omega) = L^* \Theta_X(\omega) L^*. \tag{5}$$

For now assume that $\{Y_t\}_{t \in \mathbb{Z}}$ is a WSS Gaussian process. We will relax this assumption later. Define $\omega_j = 2\pi j/n$ and denote $\mathcal{F}_n = \{\omega_0, \ldots, \omega_{n-1}\}$ to be the set of Fourier frequencies. The discrete Fourier transform (DFT) of $\{Y_t\}_{t=1}^n$ is then given by $d_j = \frac{1}{\sqrt{n}} \sum_{t=1}^n Y_t e^{-\mathrm{i}\, t\omega_j} \in \mathbb{C}^p$. Observe that DFT is a linear transformation; hence, d_js are complex-valued multivariate Gaussian with the inverse covariance $\Theta_Y(\omega_j) \in \mathbb{C}^{p \times p}$.

The log-likelihood of the finite-time series data $\{Y_t\}_{t=1}^n$ as per the *Whittle approximation* [51] (see Remark 2 for justification and benefits of the frequency-domain formulation) is given by

$$\frac{1}{2} \sum_{j \in \mathcal{F}_n} \left[\log\det(\Theta_Y(\omega_j)) - \text{Tr}(\Theta_Y(\omega_j) d_j d_j^\dagger) \right], \tag{6}$$

where † is the conjugate transpose and we dropped the constants in the approximation that do not depend on L^*. Expression in (6) resembles the log-likelihood formula for i.i.d. $\{Y_t\}_{t=1}^n$. Thus, we can view $\hat{f}_j \triangleq \hat{f}(\omega_j) = d_j d_j^\dagger$ as playing the role of sample covariance for the spectral density matrix $f_Y(\omega_j)$.

The log-likelihood in (6) requires modifications to serve as a suitable objective function in $\mathcal{L}[\cdot]$ in (1). First, for $\widehat{L}$ to have better statistical performance, the spectral density estimate $\hat{f}_j$, which has a high variance (see [52, Proposition 10.3.2]), needs to be smoothed.

We use the *averaged periodogram* [52]:

$$P_j \triangleq P(\omega_j) = \frac{1}{2\pi(2m+1)} \sum_{|k|\leq m} d(\omega_{j+k}) d^\dagger(\omega_{j+k}), \quad (7)$$

where $\omega_j \in \mathcal{F}_n$ and $P_j \in \mathbb{C}^{p\times p}$. The bandwidth m regulates the bias and variance of P_j [52], which in turn impacts the estimation consistency results for L^* in Theorem 1 and 2. For a theoretical discussion on periodograms consult [52].

Second, substituting P_j given by (7) in (6) results in an approximate likelihood that is analytically intractable because of the double summation that appears within the $\mathrm{Tr}[\cdot]$ operator. We address this by further approximating the likelihood in (6) as suggested by [20]. The idea here is to consider the likelihood in the neighborhood of a frequency ω_j, where $j \in \mathcal{F}_n$. Thus, for $j-m \leq l \leq j+m$, a reasonable likelihood near ω_j is

$$\frac{1}{2} \sum_{l=j-m}^{j+m} \left[\log\det(\Theta_Y(\omega_l)) - \mathrm{Tr}(\Theta_Y(\omega_l) d_l d_l^\dagger) \right]. \quad (8)$$

This local likelihood could be simplified by assuming $\Theta_X(\omega)$ is a smooth function of $\omega \in [-\pi, \pi]$. Thus, $\Theta_X(\omega_l)$ is constant for the frequencies neighboring ω_j. This smoothness assumption along with the relationship in (5) implies $\Theta_Y(\omega_j) = \Theta_Y(\omega_l)$, for all $j-m \leq l \leq j+m$. Consequently, (8) simplifies to

$$\frac{(2m+1)}{2} \left[\log\det(\Theta_Y(\omega_j)) - \mathrm{Tr}(\Theta_Y(\omega_j) P_j) \right], \quad (9)$$

which we call the modified Whittle's approximate likelihood for the Gaussian node potentials $\{Y_t\}_{t=1}^n$.

The modified (per frequency) likelihood in (9) is valid even if $\{Y_t\}_{t=1}^n$ is non-Gaussian. This is because as $n \to \infty$, the DFT vectors d_j converge to a complex-valued multivariate Gaussian with inverse covariance $\Theta_Y(\omega_j)$, per [52, Propositions 11.7.4 and 11.7.3]. Thus, the likelihood either in (6) or in (9) remains applicable for non-Gaussian $\{Y_t\}_{t\in Z}$. However, this standard justification relies on n being large and might not be appropriate for smaller n. A more robust theoretical justification can be given using Bregman divergences, which we discuss next.

The Bregman divergence between $p \times p$ Hermitian matrices A and B is $D_\phi(A;B) \triangleq \phi(A) - \phi(B) - \langle \nabla\phi(B), A-B \rangle$, where $\phi(\cdot)$ is a differentiable, strictly convex function mapping matrices to reals [31, 53]. The log-det Bregman divergence is a special case for $\phi(\cdot) = \log\det[\cdot]$. Thus, for $A \succ 0$ and $B \succ 0$ (either real or complex-valued matrices), we have,

$$D_\phi(A;B) = -\log\det(A) + \log\det(B) + \mathrm{Tr}(B^{-1}(A-B)).$$

Let $A = \Theta_Y(\omega)$; and $B = \Theta_Y^*(\omega)$ be the true inverse spectral density matrix with $f_Y^* = {\Theta_Y^*}^{-1}$. We drop terms that do not depend on $\Theta_Y(\omega)$ in $D_\phi(A;B)$ and note that $D_\phi(A;B)$ is proportional to $-\log|\Theta_Y(\omega)| + \mathrm{Tr}(f_Y^*(\omega)\Theta_Y(\omega))$. Finally, replacing $f_Y^*(\omega)$ in this expression with the periodogram estimator $P(\omega)$ gives us the negative of the modified likelihood given in (9).

In view of the foregoing discussion, we see that our modified approximate likelihood function in (9) is a good candidate for the loss function $\mathcal{L}[\cdot]$ in (1) even for non-Gaussian $\{Y_t\}_{t\in\mathbb{Z}}$.

Remark 1. *(Inverse of L^*). The invertibility assumption is necessary for identifying L^* from the time series data $\{Y_t\}_{t=1}^n$. However, L^* is not invertible because it has single or multiple zero eigenvalues. A workaround is to use the reduced-order Laplacian, which is obtained by removing k rows and columns from L^* (see [54]), or to perturb the diagonal of L^* with a small positive quantity. In power networks, this perturbation corresponds to adding shunt impedance (self-loops in graph theory) at the nodes. We assume that one of the approaches is in place and that L^* is invertible.*

Remark 2. *(Frequency-domain approach): Frequency-domain methods are increasingly used for multivariate time series due to their computational efficiency [20, 21, 55–59]. For a stationary univariate process with n samples, the Whittle approximation reduces the $O(n^3)$ cost of likelihood evaluation to $O(n\log n)$ via fast Fourier transforms [60]. In the multivariate case, with n samples and a $p \times p$ spectral density matrix, this computational advantage becomes even more critical, thus justifying the choice of a frequency-domain formulation.*

III. Convexity and Statistical Guarantees

Using the modified Whittle's approximate likelihood in (9), we first introduce our ℓ_1-regularized estimator as a convex optimization problem. We then present our main results that theoretically characterize the performance of this estimator when $\{X_t\}_{t\in\mathbb{Z}}$ is Gaussian and more generally a linear process. Complete proofs are in the Appendix.

The invertibility assumption (see Remark 1) and the diagonal dominance property of L^* imply that L^* is a symmetric positive definite matrix. Recall that $f^{-1}(\omega) = \Theta(\omega)$, for $\omega \in [-\pi, \pi]$. Given these conditions and the likelihood formula in (9), the optimization problem in (1) modifies to:

$$\hat{L}_j = \underset{L\succ 0}{\arg\min}\ \mathrm{Tr}(\Theta_Y(\omega_j)P_j) - \log\det(\Theta_Y(\omega_j)) + \lambda_n \|L\|_{1,\mathrm{off}}$$
$$\text{subject to}\ \ \Theta_Y(\omega_j) = L\Theta_X(\omega_j)L^\mathsf{T}, \quad (10)$$

where $j = \{0, \ldots, n-1\}$, $\lambda_n > 0$, and $\|L\|_{1,\mathrm{off}} = \sum_{i\neq j} |L_{ij}|$ is the ℓ_1-norm (see Remark 5 for more discussion on this choice) applied to the off-diagonals of $L \in \mathbb{R}^{p\times p}$. Note that the constraint in (1) is stated in terms of the density matrix $f(\omega)$. But note that the constraint in (10) is in terms of the inverse matrix $f^{-1}(\omega) = \Theta(\omega)$.

Let $D_j \in \mathbb{C}^{p\times p}$ be the unique Hermitian positive-definite square root of $\Theta_X(\omega_j)$ satisfying $D_j^2 = \Theta_X(\omega_j)$. Then substituting $\Theta_Y(\omega_j) = LD_j^2L^\mathsf{T}$ and $L = L^\mathsf{T}$ in the cost function of (10), followed by an application of the cyclic

property of the trace, results in the following unconstrained estimator:

$$\widehat{L}_j = \arg\min_{L\succ 0} \operatorname{Tr}(D_j L P_j L D_j) - \log\det(L^2) + \lambda_n \|L\|_{1,\text{off}}. \quad (11)$$

We dropped constants that bear no effect on the optimization problem. In summary, for $\omega_j \in \mathcal{F}_n$, we propose a point-wise estimator $\widehat{L}_j$ via (11). While the true Laplacian L^* is fixed and does not vary with frequency, our estimator $\widehat{L}_j$ is defined at each ω_j. Theorems 1 and 2 show that $\widehat{L}_j$ satisfies the same statistical guarantees with respect to L^* for all $\omega_j \in \mathcal{F}_n$. Therefore, any $\widehat{L}_j$ can be chosen as a candidate estimator for L^*. This per-frequency formulation aligns with recent methods such as [20, 56, 59], which also estimate spectral quantities locally at each frequency, in contrast to approaches that penalize across all frequencies [21, 61, 62]. Hereafter, we refer to P_j and $\widehat{L}_j$ as P and $\widehat{L}$, respectively, since our results hold for all $\omega_j \in \mathcal{F}_n$. Finally, we use $P_1 = \Re(P)$ and $P_2 = \Im(P)$ to denote the real and imaginary parts of the periodogram P and Ψ_1, Ψ_2 to denote the real and imaginary parts of D^2 respectively.

The following lemma establishes two crucial properties of (11): (i) the objective function is strictly convex in L and (ii) $\widehat{L}$ is unique. The proof of this lemma is in Appendix A.

***Lemma* 1.** *For any $\lambda_n > 0$ and $L \succ 0$, if all the diagonals of the averaged periodogram $P_{ii} > 0$, then (i) the ℓ_1-regularized Whittle likelihood estimator in (11) is strictly convex and (ii) $\widehat{L}$ in (11) is the unique minima satisfying the sub-gradient condition $2\Psi_1\widehat{L}P_1 - 2\Psi_2\widehat{L}P_2 - 2\widehat{L}^{-1} + \lambda_n\widehat{Z} = 0$, where $\widehat{Z}$ belong to the sub-gradient $\partial\|L\|_{1,off}$ evaluated at $\widehat{L}$.*

Establishing strict convexity of the objective function in (11) is non-trivial and crucial to derive sample complexity and estimation consistency results discussed in Section III-C. Furthermore, this strict convexity enforces the existence of unique minima even in the high-dimensional regime ($n \ll p$), where the Hessian of the objective function is rank deficient. The key ingredient in establishing such minima is the coercivity of the objective function (discussed later). The combination of convexity, coercivity, and separable property of the ℓ_1-regularizer also facilitates the development of efficient coordinate descent algorithms, which we leave for future research.

Remark 3. *(Identifiability of L^*) The matrix L^* is identifiable under two conditions: (i) the spectral density matrix Φ_X or its inverse Θ_X is known, and (ii) L^* is constrained to be symmetric and positive definite (PD). Under these assumptions, L^* has a unique closed-form expression in terms of Φ_X and Φ_Y, since the relation $\Phi_X = L^*\Phi_Y L^{*\top}$ admits a unique PD factorization. However, identifiability fails when these assumptions are relaxed. Suppose L^* is symmetric but not PD. Then, multiple symmetric square roots of Φ_X may exist, and therefore L^* may not have a unique representation in terms of Φ_X and Φ_Y, leading to a loss of identifiability. Now, if L^* is non-symmetric, and Φ_X is diagonal, then L^* is indistinguishable from L^*U for any orthogonal matrix U. Lastly, if Φ_X is unknown, then multiple pairs of L^* and Φ_X can yield the same Φ_Y, and therefore L^* is not identifiable.*

Remark 4. *(Advantage of directly estimating L^*) The estimator in (11) directly estimates L^* subject to the constraint $\Theta_Y = L^*\Theta_X L^*$. In contrast, prior methods (see for e.g., [48]) learn the network structure by first estimating the ISD matrix Θ_Y corresponding to $\{Y_t\}_{t=1}^n$ and then perform a post-processing step of applying algebraic rules to recover the support of L^*. Ref. [9] explains in great detail why this top-stage procedure is inferior to direct estimation in terms of sample complexity for the i.i.d. setting (see Fig. 1 in Ref. [9]). Mutatis mutandis, the same reasoning applies to our problem setup.*

Remark 5. *(Choosing ℓ_1-regularization) The ℓ_1-regularization is used to estimate a sparse matrix $\widehat{L}_j$. Popular applications include sparse linear regression, where it achieves both asymptotic support recovery [63, 64] and finite-sample recovery under conditions such as mutual incoherence [27, 65]. In contrast, convex alternatives such as ridge regression do not induce sparsity [66]. Iterative ℓ_2-based methods like broken adaptive ridge (BAR) regression [67] can recover support asymptotically only when both the number of samples and iterations tend to infinity. Non-convex penalties such as the smoothly clipped absolute deviation (SCAD) and minimax concave penalty (MCP) relax mutual incoherence assumptions [68, 69], but are difficult to optimize due to non-convexity, sensitivity to tuning, and initialization. Given these trade-offs, we choose the ℓ_1-penalty for its balance of theoretical guarantees and computational tractability.*

A. Statement of main results

This section features two main results. The first one concerns the theoretical characterization of the convex estimator in (11) when $\{X_t\}_{t\in\mathbb{Z}}$ is a Gaussian time series. And the second one gives such a characterization when $\{X_t\}_{t\in\mathbb{Z}}$ is a non-Gaussian linear process. At a high level our result for the Gaussian setting states that as long as the time domain samples n scales as $\Omega(d^3 \log p)$, the estimate $\widehat{L}$ correctly recovers the true support and is close to L^* (measured in Frobenius and operator norms) with high probability. Here d is the maximum degree of the graph underlying L^*. In the linear process setting, such a performance is guaranteed if n scales as $\Omega(d^3(\log p)^{4+\rho})$ for sub-exponential families with parameter ρ and $\Omega(d^3p^2)$ for distributions with finite fourth moment, respectively.

Our main results rely on three assumptions. These type of assumptions, but not identical, appeared in the literature of ℓ_1-constrained least squares problem [27, 70] and in the literature of ℓ_1-regularized inverse-covariance and spectral density estimation [20, 31]. Define the edge set $\mathcal{E}(L^*) = \{(i,j) : L^*_{ij} \neq 0, \text{for all } i \neq j\}$. Let $E = \{\mathcal{E}(L^*) \cup (1,1) \ldots \cup (p,p)\}$ be the augmented edge set including edges for the diagonal elements of L^*. Let E^c be the set complement of E.

[A1] Mutual incoherence condition: Let Γ^* be the Hessian of the log-determinant in (11):

$$\Gamma^* \triangleq \nabla^2_L \log\det(L)|_{L=L^*} = L^{*-1} \otimes L^{*-1}. \quad (12)$$

We say that L^* satisfies the mutual incoherence condition if $\left\|\left|\Gamma^*_{E^cE}{\Gamma^*_{EE}}^{-1}\right|\right\|_\infty \leq 1-\alpha$, for some $\alpha \in (0,1]$.

The incoherence condition on L^* controls the influence of irrelevant variables (elements of the Hessian matrix restricted

to $E^c \times E$ on relevant ones (elements restricted to $E \times E$). The α-incoherence assumption, commonly used in the literature, has been validated for various graphs like chain and grid graphs [31]. While α-incoherence in [20, 31] is imposed on the inverse covariance or spectral density matrix, we enforce it on L^*. A similar condition has also been explored in [9]. We note that mutual incoherence is sufficient but not strictly necessary for support recovery [1]. Non-convex penalties such as SCAD and MCP achieve support recovery without requiring incoherence [68, 69]. Although these non-convex regularizers introduce challenges related to optimization (see Remark 5), we view them as a promising direction for future work.

[A2] Bounding temporal dependence: $\{Y_t\}_{t\in\mathbb{Z}}$ has short range dependence: $\sum_{l=-\infty}^{\infty} \|\Phi_Y(l)\|_\infty < \infty$. Thus, the auto-correlation function $\Phi_Y(l)$ decreases quickly as the time lag l increases, leading to negligible temporal dependence between samples that are far apart in time.

This mild assumption holds if the nodal injections $\{X_t\}_{t\in\mathbb{Z}}$ exhibits short range dependence: $\sum_{l=-\infty}^{\infty} \|\Phi_X(l)\|_\infty < \infty$. In fact, $\sum_{l=-\infty}^{\infty} \|\Phi_Y(l)\|_\infty = \sum_{l=-\infty}^{\infty} \|L^{*-1}\Phi_X(l)L^{*-1}\|_\infty \leq \nu^2_{L^{*-1}} \sum_{l=-\infty}^{\infty} \|\Phi_X(l)\|_\infty < \infty$, where $\nu_{L^{*-1}}$ is the ℓ_∞-matrix norm of L^{*-1}. Notice that in real systems like power networks, injections typically are short-range dependent processes [15].

[A3] Condition number bound on the Hessian: The condition number $\kappa(\Gamma^*)$ of the Hessian matrix in (12) satisfies:

$$\kappa(\Gamma^*) \triangleq |||\Gamma^*|||_\infty \left|\left|\left|\Gamma^{*-1}\right|\right|\right|_\infty \leq \frac{1}{4d\nu_{D_j^2}\|\Theta_Y^{-1}(\omega_j)\|_\infty C_\alpha}, \quad (13)$$

where $C_\alpha = 1 + \frac{24}{\alpha}$, $\alpha \in (0, 1]$, $\omega_j \in \mathcal{F}_n$, and d is the maximum degree of the graph underlying L^*. Bounding $\kappa(\Gamma^*)$ to derive estimation consistency results is standard in the high-dimensional graphical model literature [71, 72].

1) Structure learning with Gaussian injections: Let $\{X_t\}_{t\in\mathbb{Z}}$ in (2) be a WSS Gaussian process. Consequently, $\{Y_t\}_{t\in\mathbb{Z}}$, a linear transformation of X_t, is also a WSS Gaussian process. Under this assumption, Theorem 1 provides sufficient conditions on the number of samples n of Y_t required so that the estimator $\widehat{L}$ in (11) exactly recovers the sparsity structure of L^* and achieves norm and sign consistency. Here, sign consistency is defined as $\text{sign}(\widehat{L}_{ij}) = \text{sign}(L^*_{ij})$, for all $(i, j) \in E$. We recall that $\nu_A = |||A|||_\infty \triangleq \max_{j=1,\ldots,p} \sum_{j=1}^{p} |A_{ij}|$.

Define the two model-dependent quantities:

$$\Omega_n(\Theta_Y^{-1}) = \max_{r\geq 1, s\leq p} \sum_{|l|<n} |l||\Phi_{Y,rs}(l)| \quad (14)$$

$$L_n(\Theta_Y^{-1}) = \max_{r\geq 1, s\leq p} \sum_{|l|>n} |\Phi_{Y,rs}(l)|. \quad (15)$$

These quantities play a crucial role in the norm consistency bounds presented in Theorem 1 and Theorem 2 (see Remark 6).

Below is an informal version of the main theorem. A formal statement and a proof with all numerical and model-dependent constants are in Appendix A. We define $|L^*_{\min}| \triangleq \min_{(i,j)\in E} |L^*_{ij}|$ to be the minimum absolute value of the non-zero entries in L^*. We use $x \gtrsim y$ to denote $x \geq cy$, where the constant c is independent of model parameters and dimensions.

***Theorem* 1.** *Let the node injections X_t be a WSS Gaussian time series. Consider any Fourier frequency $\omega_j \in [-\pi, \pi]$. Suppose that assumptions in **[A1-A3]** hold. Define $\alpha > 0$ and $C_\alpha = 1 + 24/\alpha$. Let $\lambda_n = 96\nu_{D^2}\nu_{L^*}\delta_{\Theta_Y^{-1}}(m, n, p)/\alpha$ and the bandwidth parameter $m \gtrsim \left|\left|\left|\Theta_Y^{-1}\right|\right|\right|^2_\infty \zeta^2 d^2 \log p$, where $\zeta = \max\{\nu_{\Gamma^{*-1}}\nu_{L^{*-1}}\nu_{L^*}\nu_{D^2}C^2_\alpha, \nu^2_{\Gamma^{*-1}}\nu^3_{L^{*-1}}\nu_{L^*}\nu_{D^2}C^2_\alpha\}$.*

If the sample size $n \gtrsim \Omega_n(\Theta_Y^{-1})\zeta md$. Then with probability greater than $1 - 1/p^{\tau-2}$, for some $\tau > 2$, we have

(a) $\widehat{L}_j$ exactly recovers the sparsity structure i.e., $[\widehat{L}_j]_{E^c} = 0$.

(b) The estimate $\widehat{L}_j$ which is the solution of (11) satisfies

$$\|\widehat{L}_j - L^*\|_\infty \leq 8\nu'\delta_{\Theta_Y^{-1}}(m, n, p). \quad (16)$$

(c) $\widehat{L}_j$ satisfies sign consistency if:

$$|L^*_{\min}(E)| \geq 8\nu'\delta_{\Theta_Y^{-1}}(m, n, p), \quad (17)$$

where, $\nu' = \nu_{\Gamma^{*-1}}\nu_{D^2}\nu_{L^*}C_\alpha$ and

$$\delta_{\Theta_Y^{-1}}(m, n, p) = \sqrt{\frac{\tau \log p}{m}} + \frac{m + \frac{1}{2\pi}}{n}\Omega_n(\Theta_Y^{-1}) + \frac{1}{2\pi}L_n(\Theta_Y^{-1}).$$

Some remarks are in order. Assume that ζ and $\left|\left|\left|\Theta_Y^{-1}\right|\right|\right|_\infty$ are independent of (n, p, d) and that we are in the high-dimensional regime where $\log p/n \to 0$ as $(n, p) \to \infty$. Under assumptions in Theorem 1, and when $n = \Omega(d^3 \log p)$, with high probability: (a) The support of $\widehat{L}$ is contained within L^*; meaning there are no false negatives. Furthermore, when $(m/n)\Omega_n(\Theta_Y^{-1}) \to 0$ as $(m, n) \to \infty$, part (b) asserts that the element-wise ℓ_∞-norm, $\|\widehat{L} - L^*\|_\infty$, vanishes asymptotically (see Remark 6 for further discussion on the asymptotic decay of the error norm). Finally, part (c) establishes the sign consistency of $\widehat{L}$. Crucial is the requirement of $|L^*_{\min}| = \Omega\left(\delta_{\Theta_Y^{-1}}(m, n, p)\right)$, which limits the minimum value (in absolute) of the nonzero entries in L^*. This condition parallels the familiar *beta-min* condition in the LASSO literature (see [20, 27, 31]). Finally, since each estimate $\widehat{L}_j$ for $j = 1, \ldots, n-1$ satisfies the same statistical guarantees with high probability, any $\widehat{L}_j$ can be selected as a candidate estimator for L^*.

The error bound $\delta_{\Theta_Y^{-1}}$ in Theorem 1 quantifies the deviation of the estimator $\widehat{L}_j$ from the true Laplacian L^* in the element-wise ℓ_∞-norm. It has two components: the first term, $\sqrt{\log p/m}$, captures the leading statistical error, while the second, involving Ω_n and L_n, accounts for temporal and contemporaneous dependencies in the data. As defined in equations (14) and (15), these terms vanish under i.i.d. data and increase with stronger temporal dependence in the data.

We also emphasize the strength of the above result. Although $\widehat{L}_j$ is derived from the Whittle approximation, Theorem 1 ensures support recovery and norm consistency. Prior works such as [73] have studied the discrepancy between the Gaussian and Whittle likelihoods. While formally quantifying this approximation error is beyond the scope of the present work, we view it as a valuable direction for future research.

[1] It is nearly necessary for sign selection consistency, but not for support recovery; see [70]

We state a corollary to Theorem 1 that gives error-consistency rates for $\widehat{L}$ in the Frobenius and operator norms. Let $\mathcal{E}(L^*) = \{(i,j) : L^*_{ij} \neq 0, \text{for all } i \neq j\}$ be the edge set.

Corollary 1. *Let $s = |\mathcal{E}(L^*)|$ be the cardinality of the edge set $\mathcal{E}(L^*)$. Under the hypothesis as in Theorem 1, with probability greater than $1 - \frac{1}{p^{\tau-2}}$, the estimator $\widehat{L}$ defined in (11) satisfies*

$$\|\widehat{L} - L^*\|_F \leq 8\nu'(\sqrt{s+p})\delta_{\Theta_Y^{-1}}(m,n,p) \quad \text{and}$$
$$\|\widehat{L} - L^*\|_2 \leq 8\nu' \min\{d, \sqrt{s+p}\}\delta_{\Theta_Y^{-1}}(m,n,p),$$

where ν' and $\delta_{\Theta_Y^{-1}}(m,n,p)$ are defined in Theorem 1.

Proof sketch:. Both the Frobenius and operator norm bounds follow by applying standard matrix norm inequalities to the ℓ_∞ consistency bound in part (b) of Theorem 1. Importantly, $s+p$ is the bound on the maximum number of non-zero entries in L^*, where s is the total number of off-diagonal non-zeros in L^*. Complete details are in Appendix A.

2) Structure learning for non-Gaussian injections : We consider a class of WSS processes $\{X_t\}_{t\in\mathbb{Z}}$ that are not necessarily Gaussian. Examples include Vector Auto Regressive (VAR(p)) and Vector Auto Regressive Moving Average (VARMA (p,q)) models with non-Gaussian noise terms. Such models, and many others, belong to a family of linear WSS processes with absolute summable coefficients:

$$X_t = \sum_{l=0}^{\infty} A_l \epsilon_{t-l}, \tag{18}$$

where $A_l \in \mathbb{R}^{p\times p}$ is known and $\epsilon_t \in \mathbb{R}^p$ is a zero mean i.i.d. process with tails possibly heavier than Gaussian. The absolute summability $\sum_{l=0}^{\infty} |A_l(i,j)| < \infty$ ensures stationarity for all $i,j \in \{1,\ldots,p\}$ [74]. We assume that ϵ_{kl} (for all $k \in [p]$), the k-th component of $\epsilon_l \in \mathbb{R}^p$, is given by one of the distributions below:

[B1] Sub-Gaussian: There exists $\sigma > 0$ such that for all $\eta > 0$, we have $\mathbb{P}[|\epsilon_{kl}| > \eta] \leq 2\exp(-\frac{\eta^2}{2\sigma^2})$.

[B2] Generalized sub-exponential with parameter $\rho > 0$: There exists constants a and b such that for all $\eta > 0$: $\mathbb{P}[|\epsilon_{kl}| > \eta^\rho] \leq a\exp(-b\eta)$.

[B3] Distributions with finite 4^{th} moment: There exists a constant $M > 0$ such that $\mathbb{E}[\epsilon_{kl}^4] \leq M < \infty$.

We need additional notation. Let $n_k = \Omega(d^3\mathcal{T}_k)$ represent the family of sample sizes indexed by $k = \{1,2,3\}$, where $\mathcal{T}_1 = \log p$ correspond to the distribution in **[B1]**, $\mathcal{T}_2 = (\log p)^{4+4\rho}$ in **[B2]**, and $\mathcal{T}_3 = p^2$ in **[B3]**.

***Theorem* 2.** *Let X_t be given by (18) and $Y_t = L^{*-1}X_t$. Fix $\omega_j \in [-\pi,\pi]$. Let $n_k = \Omega(d^3\mathcal{T}_k)$, where $k = \{1,2,3\}$. Then for some $\tau > 2$, with probability greater than $1 - 1/p^{\tau-2}$:*

(a) $\widehat{L}$ exactly recovers the sparsity structure i.e., $\widehat{L}_{E^c} = 0$.

(b) The ℓ_∞ bound of the error satisfies:

$$\|\widehat{L} - L^*\|_\infty = \mathcal{O}(\delta^{(k)}_{\Theta_Y^{-1}}(n,m,p)). \tag{19}$$

(c) $\widehat{L}$ satisfies sign consistency if:

$$|L^*_{\min}(E)| = \Omega(\delta^{(k)}_{\Theta_Y^{-1}}(n,m,p)), \tag{20}$$

where $\delta^{(k)}_{\Theta_Y^{-1}}(n,m,p)$ for $k = \{1,2,3\}$ is given by

$$\delta^{(1)}_{\Theta_Y^{-1}}(n,m,p) = \left|\left|\left|\Theta_Y^{-1}\right|\right|\right|_\infty \frac{(\tau\log p)^{1/2}}{\sqrt{m}} + \triangle(n,m,\Theta_Y^{-1})$$
$$\delta^{(2)}_{\Theta_Y^{-1}}(n,m,p) = \left|\left|\left|\Theta_Y^{-1}\right|\right|\right|_\infty \frac{(\tau\log p)^{2+2\rho}}{\sqrt{m}} + \triangle(n,m,\Theta_Y^{-1})$$
$$\delta^{(3)}_{\Theta_Y^{-1}}(n,m,p) = \left|\left|\left|\Theta_Y^{-1}\right|\right|\right|_\infty \frac{p^{1+\tau}}{\sqrt{m}} + \triangle(n,m,\Theta_Y^{-1}),$$

where $\triangle(n,m,\Theta_Y^{-1}) = \frac{m+\frac{1}{2\pi}}{n}\Omega_n(\Theta_Y^{-1}) + \frac{1}{2\pi}L_n(\Theta_Y^{-1})$.

Remark 6. *(Asymptotic decay rate of the error $\|\widehat{L} - L^*\|_\infty$) The model-dependent quantities $\Omega_n(\Theta_Y^{-1})$ and $L_n(\Theta_Y^{-1})$, as defined in (14) and (15), are critical for bounding the element-wise ℓ_∞-norm of the error $\|\widehat{L} - L^*\|_\infty$ in Theorems 1 and 2. We examine conditions under which this error vanishes asymptotically. Specifically, by definition in (15), the quantity $(\sqrt{\log p/m}, L_n(\Theta_Y^{-1})) \to 0$ as $(m,n) \to \infty$. Furthermore, if $(m/n)\Omega_n(\Theta_Y^{-1}) \to 0$ as $(m,n) \to \infty$, then the error norm vanishes asymptotically. This condition holds in scenarios where the autocovariance function $\Phi_Y(l)$ exhibits a geometric decay rate or if $\{Y_t\}_{t\in\mathbb{Z}}$ is a VAR(d) process or other stationary processes with strong mixing conditions (see Proposition 3.4 in [55]). As a consequence, the condition $(m/n)\Omega_n(\Theta_Y^{-1}) \to 0$ as $(m,n) \to \infty$ holds for a wide range of stationary processes, leading to asymptotic decay of the error norm.*

B. Outline of technical analysis for main results

We summarize the key techniques used to prove Theorems 1 and 2. Complete details are in Appendix A. We leverage the primal-dual witness (PDW) method—a general technique used to derive statistical guarantees for sparse convex estimators [27, 31]. Before detailing the PDW method, we state differences in our proof approach compared to the cited literature. First, our analysis is in the frequency domain, this accounts for temporal dependencies from the WSS process, requiring careful treatment of the Hermitian matrices P_j and D^2 in (11). Second, unlike most literature where the objective function's dependence on the optimization variable L is linear, our objective function in (11) has a quadratic dependence. This distinction in the frequency domain necessitates stricter control of the Hessian matrix Γ^* via our assumption **[A3]**.

In the PDW method, we construct an optimal primal-dual pair $(\widetilde{L}, \widetilde{Z})$ that satisfies the zero sub-gradient condition of the problem in (11). (i) The primal $\widetilde{L}$ is constrained to have the correct signed support E of the true Laplacian matrix L^* and (ii) The dual $\widetilde{Z}$ is the sub-gradient of $\|L\|_{1,\text{off}}$ evaluated at $\widetilde{L}$. If the dual $\widetilde{Z}$ satisfies the strict dual feasibility condition $\|\widetilde{Z}_{E^c}\|_\infty < 1$. Then the dual acts as a witness to certify that $\widetilde{L} = \widehat{L}$ and $\widetilde{L}$ is indeed the unique global optimum.

C. The primal-dual construction and supporting lemmata

We construct an optimal primal-dual pair $(\widetilde{L}, \widetilde{Z})$. Lemma 2 gives conditions under which this construction succeeds. First, we determine $\widetilde{L}$ by solving the restricted problem:

$$\widetilde{L} \triangleq \underset{L\succ 0, L_{E^c}=0}{\arg\min} \ \text{Tr}(DLPLD) - \log\det(L^2) + \lambda_n\|L\|_{1,\text{off}}. \tag{21}$$

Notice that $\widetilde{L} \succ 0$ and $\widetilde{L}_{E^c} = 0$. We choose the dual $\widetilde{Z} \in \|\widetilde{L}\|_{1,\text{off}}$ to satisfy the zero sub-gradient condition of (21) by setting $\lambda_n \widetilde{Z}_{ij} = -2[\Psi_1 \widetilde{L} P_1]_{ij} + 2[\Psi_2 \widetilde{L} P_2]_{ij} + 2[\widetilde{L}^{-1}]_{ij}$, for all $(i,j) \in E^c$, where P_1 (resp. Ψ_1) and P_2 (resp. Ψ_1) are the real and imaginary parts of P (resp. D). Therefore, the pair $(\widetilde{L}, \widetilde{Z})$ satisfies the zero sub-gradient condition of the restricted problem in (21).

We verify the strict dual feasibility condition: $|\widetilde{Z}_{ij}| < 1$, for any $(i,j) \in E^c$. We introduce three quantities. First, $W \triangleq P - \Theta_Y^{-1}$ quantifies the error between the averaged periodogram P and the true spectral density matrix Θ_Y^{-1}. Second, let $\Delta \triangleq \widetilde{L} - L^*$ be the measure of distortion between $\widetilde{L}$ given by (21) and the true Laplacian matrix L^*. The final quantity $R(\Delta)$ captures higher order terms in the Taylor expansion of the gradient $\nabla \log\det(\widetilde{L})$ centered around L^*. In fact, expand $\nabla \log\det(\widetilde{L}) = \widetilde{L}^{-1} = L^{*-1} + L^{*-1}\Delta L^{*-1} + \widetilde{L}^{-1} - L^{*-1} - L^{*-1}\Delta L^{*-1}$, and then define $\widetilde{L}^{-1} - L^{*-1} - L^{*-1}\Delta L^{*-1} = R(\Delta)$.

The following lemma establishes the sufficient conditions for ensuring strict dual feasibility.

Lemma 2. *(Conditions for strict-dual-feasibility) Let $\lambda_n > 0$ and α be defined as in **[A1]**. Suppose that $\max\{2\nu_{D^2}(d\|\Delta\|_\infty + \nu_{L^*})\|W\|_\infty, \|R(\Delta)\|_\infty, 2\nu_{D^2} d\|\Delta\|_\infty \|\Theta_Y^{-1}\|_\infty\} \leq \frac{\alpha\lambda_n}{24}$. Then the dual vector $\widetilde{Z}_{E^c}$ satisfies $\|\widetilde{Z}_{E^c}\|_\infty < 1$, and hence, $\widetilde{L} = \widehat{L}$.*

Proof sketch: Express the sub-gradient condition in Lemma 1 in a vectorized form as a function of $R(\Delta)$, $W = P - \Theta_Y^{-1}$, and Θ_Y^{-1}. We decompose the vectorized sub-gradient condition into two linear equations corresponding to the edge set E and its complement E^c. An expression for $\widetilde{Z}_{E^c}$ is obtained as a function of $R(\Delta), W$ and Θ_Y^{-1}. We finish the proof by utilizing the mutual incoherence condition stated in **[A1]**.

The following results provide us with dimension and model complexity dependent bounds on the remainder term $R(\Delta)$. The proof, adapted from [31, lemma 5], relies on matrix expansion techniques; see Appendix A for details.

Lemma 3. *Suppose that the ℓ_∞-norm $\|\Delta\|_\infty \leq 1/(3\nu_{L^{*-1}} d)$, then $\|R(\Delta)\|_\infty \leq \frac{3}{2} d\|\Delta\|_\infty^2 \nu_{L^{*-1}}^3$.*

The result below provides a sufficient condition under which the ℓ_∞-bound on Δ in Lemma 3 holds. Full proof in Appendix A.

Lemma 4. *Define $r \triangleq 8\nu_{\Gamma^{*-1}}[\nu_{D^2}\nu_{L^*}\|W\|_\infty + \lambda_n/4]$ and suppose $r \leq \min\{1/(3\nu_{L^{*-1}} d), 1/(6\nu_{\Gamma^{*-1}}\nu_{L^{*-1}}^3 d)\}$. Then we have the element-wise ℓ_∞-bound: $\|\Delta\|_\infty = \|\widetilde{L} - L^*\|_\infty \leq r$.*

Proof sketch: Since $\widetilde{L}_{E^c} = L^*_{E^c} = 0$, we note $\|\Delta\|_\infty = \|\Delta_E\|_\infty$, where $\Delta_E = \widetilde{L}_E - L^*_E$ and it is the solution of the sub-gradient associated with the restricted problem in (21). We construct a continuous function $F : \mathbb{R}^{|E|} \to \mathbb{R}^{|E|}$ with two properties: (i) it has a unique fixed point Δ_E and (ii) On invoking assumption **[A3]**, F is a contraction—specifically, $F(\mathbb{B}_r) \subseteq \mathbb{B}_r$, where $\mathcal{B}_r = \{A \in \mathbb{R}^{|\mathbb{E}|} : \|A\|_\infty \leq r\}$ and r. The proof follows by invoking Brower's fixed point theorem [75] and exploiting the unique fixed point property of F to show that $\Delta_E \in \mathcal{B}_r$, and hence, $\|\Delta\|_\infty \leq r$.

Remark 7. *A consequence of assumption **[A3]** is the lower bound on the norm of the Hessian $\|\|\Gamma^*\|\|_\infty$. This implies that the curvature at the true minimum L^* is lower bounded. This bound on the curvature is specific to our problem and helps in attaining control on the distortion parameter Δ, as demonstrated in Lemma 4.*

IV. Simulations

We report the results of multiple simulations to validate our theoretical claims. The results in Theorems 1 and 2 involve several constants, along with the dimensional parameters (n, m, d, p). Therefore, we do not expect the theoretical results to capture the nuanced behavior of the simulations in every detail. However, we observe that the learning performance of the estimator in (11) improves as the rescaled sample size $n/(d^3 \log(p))$ increases, and that the error norm decreases with increasing $n/\log p$. Additionally, the experimental results are also influenced by the choice of the regularization λ_n. We ran the experiments using CVXPY 1.2, an open-source Python package. The reproducible code for generating simulation results in this paper is publicly available at https://tinyurl.com/LNSWSSP.

A. Setup and accuracy evaluation metrics

Our experiments assess the finite-sample performance of the proposed estimator for two families of stochastic injections $\{X_t\}_{t\in\mathbb{Z}}$, namely, vector autoregressive (VAR (1)) and vector autoregressive moving average (VARMA (2,2)) processes. These processes not only satisfy our technical assumptions but are also widely used for empirical studies.

(i) VAR(1) process: Here the injections $\{X_t\}_{t\in\mathbb{Z}}$ satisfy $X_t = AX_{t-1} + \epsilon_t$ where $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $A = 0.7I_p$. The PSD matrix of this process, for $z = e^{-i\omega}$ and $\omega \in [-\pi, \pi]$, is

$$f_X(\omega) = \frac{1}{2\pi}\left(I_p - Az\right)\left(I_p - Az^{-1}\right)^{-1}.$$

(ii) VARMA(2,2) process: We let $X_t = A_1 X_{t-1} + A_2 X_{t-2} + \epsilon_t + B_1\epsilon_{t-1} + B_2\epsilon_{t-2}$ where $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. The PSD matrix of this process, for $z = e^{-i\omega}$ and with $\omega \in [-\pi, \pi]$, is [52]

$$f_X(\omega) = \frac{1}{2\pi}\mathcal{A}(z)\mathcal{B}(z)\mathcal{B}^\dagger(z)(\mathcal{A}^{-1}(z^{-1}))^\dagger, \tag{22}$$

where $\mathcal{A}(z) = I_p - \sum_{t=1}^2 A_t z^t$ and $\mathcal{B}(z) = I_p - \sum_{t=1}^2 B_t z^t$. We set $A_1 = 0.4I_p$ and $A_2 = 0.2I_p$. Furthermore, $B_1 = 1.5(I_5 + J_5)$ and $B_2 = 0.75(I_5 + J_5)$, where $J_k \in \mathbb{R}^{k\times k}$ is the matrix of all ones.

For the above processes, we assume that the nodal observation data $\{Y_t\}_{t\in\mathbb{Z}}$ satisfy $Y_t = L^{*-1}X_t$, where we consider L^* for synthetic, benchmark, and real-world networks (discussed later). The periodogram of $\{Y_t\}_{t=1}^n$ at frequency ω_j is then computed as $P(\omega_j) = \frac{1}{2\pi(2m+1)}\sum_{|k|\leq m} d(\omega_{j+k})d(\omega_{j+k})^\dagger$. For simplicity, we set the centering frequency as $\omega_j = 0$. However, our numerical and theoretical analysis applies to any non-zero Fourier frequency. Further, the bandwidth parameter $m = \sqrt{n}$, which is theoretically justified because we consider the regime $m/n \to 0$ as $(m,n) \to \infty$ where the periodogram is asymptotically unbiased (see Remark 6 and [55]).

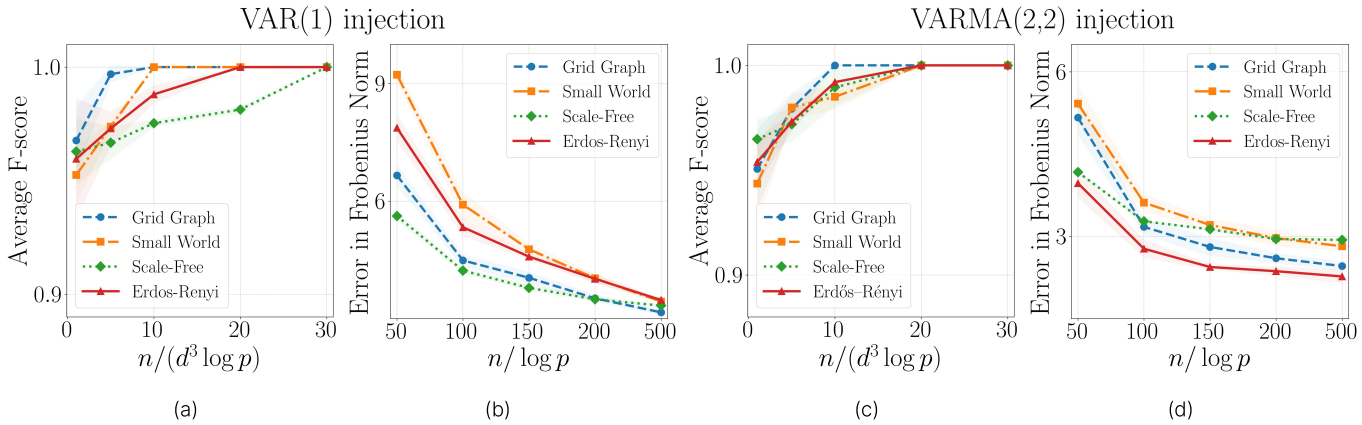

Fig. 1. We evaluate the support recovery metric (F-score) and the Frobenius norm error for synthetic random networks under VAR(1) and VARMA(2,2) stochastic injections. Synthetic networks of size $p = 30$ are examined, with results averaged over 50 independent trials. Solid curves represent mean performance, while shaded regions around each curve indicate one-sigma standard deviations. The random networks analyzed include grid, small-world, scale-free, and Erdős-Rényi, with maximum degrees $d = \{4, 3, 9, 4\}$, respectively. Panels (a,b) present the average F-score and Frobenius norm error versus rescaled sample size for VAR(1) injection, while panels (c,d) display the same metrics for VARMA(2,2) injection. The rescaled sample size for the F-score is $n/(d^3 \log p)$, and for the Frobenius norm error, it is $n/\log p$, based on asymptotic convergence rates in Theorem 1. Notably, rescaling the sample size to $n/(d^3 \log p)$ aligns all curves on top of each other as predicted by Theorem 1.

We consider sparsistency (the ability to recover the correct edge structure) and norm-consistency (the Frobenius norm of the deviation between $\widehat{L}$ and L^*) metrics to evaluate the estimation performance. We assess sparsistency via the F-score: F-score $= 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN}) \in [0, 1]$, where TP (true positives) is the number of correctly detected edges, FP (false positives) is the number of non-existent edges detected, and FN (false negatives) is the number of actual edges not detected. The higher the F-score, the better the performance of the estimator in learning the true structure, with F-score $= 1$ signifying perfect structure recovery.

B. Synthetic networks

We present simulations evaluating the performance of the proposed estimator on synthetic random networks. All synthetic networks have a fixed size of $p = 30$. The random networks examined in Figure 1 include Erdős-Rényi, Small-World (Watts-Strogatz model), and Scale-Free (Barabási-Albert model) networks, with maximum degrees $d = \{4, 3, 9\}$, respectively. Additionally, a synthetic grid graph ($d = 4$) is constructed by connecting each node to its fourth-nearest neighbor.

For details on constructing the Laplacian matrix L^* for the synthetic random networks, we refer the readers to [76] and the GitHub repository[2]. Once L^* is obtained, we ensure its positive definiteness by adding a small diagonal perturbation of 0.1 (positive definiteness by diagonal perturbation follows from the Gershgorin circle theorem). This perturbed matrix is no longer a Laplacian in the strict sense. However, this perturbation is acceptable since our estimation task focuses only on recovering the sparsity pattern of L^* and not its spectral properties. In Figure 1, we plot the average F-score and the average Frobenius norm of the error (averaged over 50 independent trials) versus rescaled sample size under VAR(1) and VARMA(2,2) injections. Panels (a-b) depict these metrics for VAR(1) injection, while panels (c-d) show results for VARMA(2,2). The rescaled sample size is $n/(d^3 \log p)$ for F-score and $n/\log p$ for Frobenius norm error, based on asymptotic convergence rates in Theorem 1. As shown in panels (a) and (c), the F-score increases with $n/(d^3 \log p)$, achieving perfect structure recovery, as predicted by Theorem 1. This causes all plots in panels (a) and (c) to align on top of each other. Panels (b) and (d) demonstrate similar behavior for the Frobenius norm error metric, where the error norm decreases with an increase in $n/\log p$.

In Figure 2, we compare F-scores for i.i.d., VAR(1), and VARMA(2,2) injections on an Erdős-Rényi network with size $p = 30$ and maximum degree $d = 4$. The results indicate that fewer samples are needed to achieve perfect structure recovery (that is, F-score $= 1$) with i.i.d. injections compared to injections of VAR (1) and VARMA (2,2). This trend aligns with theoretical expectations: structure recovery under i.i.d. injections requires $n = \mathcal{O}(d^2 \log p)$ samples (see [9]), compared to the higher sample complexity of $n = \mathcal{O}(d^3 \log p)$ for VAR(1) and VARMA(2,2) (see Theorem 1).

Finally, we comment on obtaining the regularization parameter λ_n for experiments in Figure 1 and 2. We apply the extended Bayesian information criterion (EBIC) [77] to select λ_n. The EBIC is given by:

$$\text{EBIC}_\gamma(\widehat{L}) = -2\mathcal{L}_n(\widehat{L}) + |\widehat{E}| \log n + 4\gamma |\widehat{E}| \log p, \qquad (23)$$

where $\mathcal{L}_n(\widehat{L})$ is the log-likelihood in (11), $\widehat{E} = E(\widehat{L})$ represents the edge set of the candidate graph $\widehat{L}$, and $\gamma \in [0, 1]$ is a tuning parameter that influences the penalization. Higher values of γ lead to sparser networks. The optimal regularization parameter is $\lambda_n = \arg\min_{\lambda > 0} \text{EBIC}_\gamma(\widehat{L})$.

The results in Figure 1 and Figure 2 are for $\gamma = 0.4$. In Figure 3, we fix a sample size $n = 1000$ and plot the regularization path for both the F-score and Frobenius norm

[2] https://github.com/psjayadev/Predicting-Links-Conserved-Networks

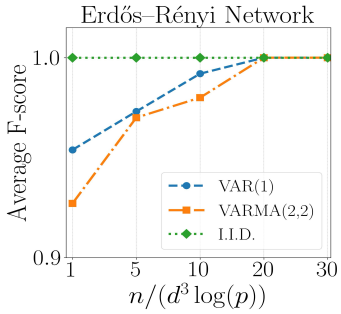

Fig. 2. Average F-score comparison for $\{X_t\}_{t\in\mathbb{Z}}$ governed by i.i.d., VAR(1), and VARMA(2,2) processes versus rescaled sample size for an Erdős-Rényi network ($p = 30$, $d = 4$). Perfect structure recovery under VAR(1) and VARMA(2,2) injections requires more samples than under i.i.d. injections.

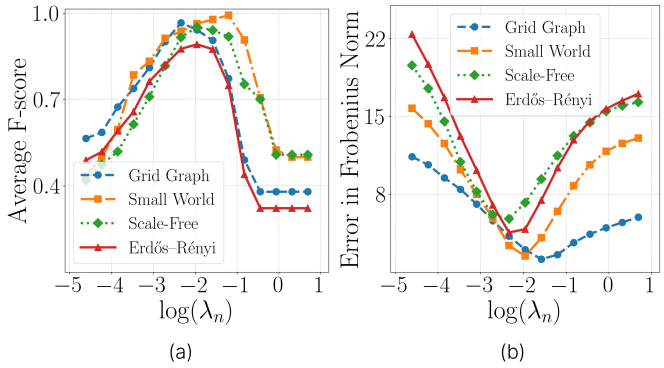

Fig. 3. For a fixed sample size $n = 1000$, we plot (a) Regularization path for F-score and (b) regularization path for Frobenius norm error, both on a linear-log scale. All networks have $p = 30$ nodes, with maximum degrees as follows: grid ($d = 4$), small-world ($d = 3$), scale-free ($d = 9$), and Erdős–Rényi ($d = 4$).

error across various network types. Notably, we observe that for a class of random networks, and the fixed sample size $n = 1000$ the value $\log(\lambda_n) \approx -2$ simultaneously maximizes both the F-score and minimizes the Frobenius norm error.

C. Benchmark networks

For $\{X_t\}_{t\in\mathbb{Z}}$ governed by the VARMA(2,2) process, we evaluate the performance of our estimator on three benchmark networks: the power distribution network, water network, and the brain network. Each network has an associated ground truth matrix $L^* = A + \epsilon I_p$, where A is the adjacency matrix that defines the edge structure of the network, $\epsilon = \{2, 2, 3\}$ for the power, water, and brain networks, respectively, and I_p is the p-dimensional identity matrix. This diagonal perturbation ensures that L^* is positive definite while preserving its sparsity pattern and thus does not affect the structure learning objective.

1) Power distribution network: We consider the IEEE 33-bus power distribution network whose raw data files are publicly available[3]. An adjacency matrix A can be constructed from this dataset. The network corresponding to A consists of 33 buses and 32 branches (edges) with maximum degree $d = 3$.

2) Water distribution network: We examine the Bellingham water distribution network, using data sourced from the database described in [78]. The raw data files are publicly accessible[4]. The ground truth adjacency matrix A, containing 121 nodes and 162 edges with maximum degree $d = 6$, is generated by loading the raw data files into the WNTR simulator[5]. Complete details on obtaining the adjacency matrix are provided in [79].

3) Brain network: The ground truth adjacency matrix A for this study is publicly accessible[6], with the detailed methodology regarding its construction described in [80]. The matrix A is a 90×90 matrix (i.e., 90 nodes), where each row and column corresponds to a specific region of interest (ROI) in the brain, as defined by the Automated Anatomical Labeling (AAL) atlas. From 88 patient-derived connectivity matrices found in the database, one was selected (filename: S001.csv) for numerical analyses. The selected network consists of 90 nodes, 141 edges and maximum degree $d = 7$.

Figure 4 compares the performance of the proposed single-step Whittle likelihood estimator with a two-step baseline method (square root). The matrix L^* is an IEEE 33-bus power distribution network and X_t is a Gaussian VAR(1) stochastic injection with diagonal auto-covariance: $\Phi_X(l) = \rho^{|l|}I$, with $\rho = 0.1$ and $l = \{1, \ldots, n-1\}$. The single-step approach estimates L^* from samples of Y_t as described in earlier experiments.

In contrast, the two-step procedure first estimates the inverse spectral density matrix $\Theta_Y(\omega)$ from samples of Y_t and then computes its positive definite square root to estimate L^*. In this experiment, we fix the frequency at $\omega = 0$, where $\Theta_Y(0) = L^{*2}K.I$, where K is some constant and I is the identity matrix. In more general settings where $\Theta_X(\omega)$ is non-diagonal, the baseline would compute $\widehat{L} = \widehat{\Theta}_Y\Theta_X^{-1/2}$.

Panels (a) and (b) show the average F-score and Frobenius norm error, respectively, as functions of sample size n, averaged over 50 trials. The single-step estimator recovers the structure with fewer samples and achieves lower error compared to the two-step approach, thereby highlighting its superior performance over the baseline approach. As Θ_Y has degree d^2 (presence of two-hop neighbors) versus d for L^*, Theorem 1 implies that the two-step method requires $O(d^6 \log p)$ samples as compared to $O(d^3 \log p)$ for the proposed approach.

Figure 5 shows the F-score and element-wise ℓ_∞-norm of the error versus the rescaled sample size. For benchmark networks with varying sizes p and maximum degrees d, there is a sharp increase in the F-score when the sample size is $n/(d^3 \log p) \approx 1$, thus validating the sample complexity of $n = \mathcal{O}(d^3 \log p)$ as suggested by Theorem 1. This sharp increase in F-score is consistent across different benchmark networks with differing size p and maximum degree d. Similarly, across the benchmark networks, the element-wise ℓ_∞-norm of the error decreases sharply at $n/(\log p) \approx 1$.

[3] https://www.mathworks.com/matlabcentral/fileexchange/73127-ieee-33-bus-system

[4] https://www.uky.edu/WDST/index.html

[5] https://github.com/USEPA/WNTR

[6] https://osf.io/yw5vf/

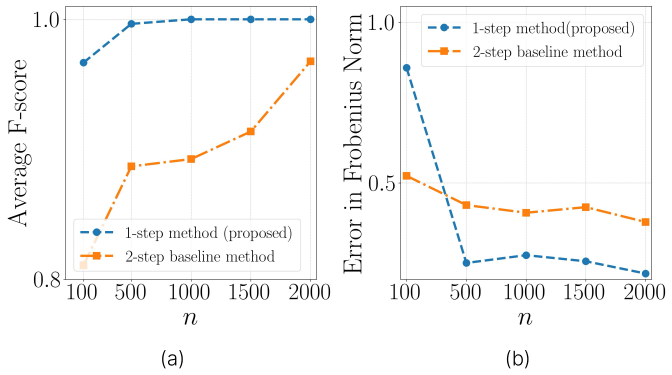

Fig. 4. Performance comparison between the proposed single-step Whittle likelihood estimator and a two-step baseline method on the IEEE 33-bus power distribution network under VAR(1) stochastic injection with diagonal auto-covariance structure $\Phi_X(l) = \rho^{|l|} I$ ($\rho = 0.1$). Panel (a) shows the average F-score versus sample size n, and panel (b) shows the average Frobenius norm error versus n. The single-step estimator achieves perfect structure recovery with fewer samples and exhibits faster error decay compared to the two-step approach, thereby signifying better performance. All results are averaged over 50 independent trials.

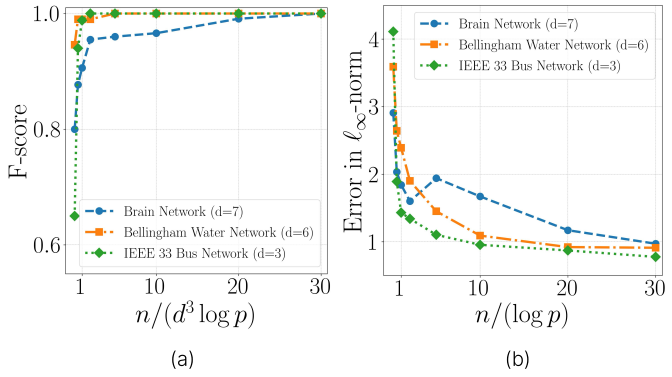

Fig. 5. (a) F-score versus rescaled sample size $(n/(d^3 \log p))$ across different benchmark networks. (b) Element-wise ℓ_∞-norm of the error versus rescaled sample size $(n/\log p)$ for the same networks. Both panels compare the human brain structural connectivity network (size $p = 90$), the Bellingham water network ($p = 120$), and the IEEE 33 bus power distribution network ($p = 33$).

D. Real world brain network

We aim to estimate the brain networks for the control and autism groups using fMRI data (obtained under resting-state conditions) from the Autism Brain Imaging Data Exchange (ABIDE) dataset[7]. The pre-processed dataset is accessible[8], we refer to [14] for more details. For each subject, we have access to 249 samples of time series measurements across 90 anatomical regions of interest (ROIs) that result in a data matrix, $\{Y_t\}_{t=1}^{249} \in \mathbb{R}^{90}$. We collect such measurements for 86 subjects (46 from the autism group and 40 from the control group), from https://github.com/jitkomut/cvxsem.

Using this dataset, we estimate a common brain network for each group: one for the control group (among 40 subjects) and one for the autism group (among 46 subjects). The common networks are constructed by identifying the *statistically significant edges* (to be defined later) present across subjects in each group. While our goal is to evaluate the common brain network estimates against the ground truth using metrics like the F-score and Frobenius norm, this is not possible since the true network L^* is unknown for both groups. Instead, we analyze the relative similarities and differences between the estimated common networks for the control and autism groups.

We begin the experiment by modeling the autocovariance matrix of the noise $\{X_t\}$ as $\Phi_X(l) = \rho^{|l|} I_p$ with $\rho = 0.1$, $l = \{1, \ldots, 248\}$ and I_p is the p-dimensional identity matrix. The noise $\{X_t\}$ is therefore a WSS process. The PSD matrix $f_X(\omega) = D^2$ is computed as the Fourier transform of the autocovariance function $\Phi_X(l)$ at $\omega = 0$. Our estimator is then applied with regularization $\lambda_n = 0.23$ (tuned via grid search) across all 86 subjects. The common brain networks for each group are then constructed by retaining the statistically significant edges, that is, the edges that appear in over 90% of the subjects.

Figure 6 (a,b) illustrates the sparsity pattern of the estimated common adjacency matrix for the control group ($\widehat{L}_C$) and the autism group ($\widehat{L}_A$) brain networks. Each colored point in Figure 6 (a,b) represents a statistically significant edge. We observe that the estimated adjacency matrix for both groups exhibits sparsity as proposed in [81, 82]. In Figure 6 (c), we plot the difference matrix $\widehat{L}_C - \widehat{L}_A$ to highlight control-specific connections, indicating more connections in the control group than in the autism group. Furthermore, we identify connections that are unique to each group as well as shared across groups. Figure 6 (d) displays a bar plot of the distribution of the group-specific and shared connections, showing that while both groups share numerous connections, the control group exhibits greater connectivity, suggesting a denser network compared to the autism group. This sparsity trend persists for values of λ_n between 0.1 and 0.23. For values below 0.1, the estimated networks become too dense to support any meaningful conclusions. Similarly, for values above 0.23, the networks become overly sparse and lack interpretability. At $\lambda_n = 0.23$, the estimated network recovers several connections reported in the literature.

In Appendix C, we list all estimated neural connections present only in the control group. Table II links these control-specific connections to well-established cognitive functions, including social interaction, face and image recognition, working memory, and language comprehension. Each of these findings is supported by prior neuroscience literature cited in Table II.

V. Parallels with other structure learning problems

In this section, we loop back to emphasize the generality of the network learning framework considered in this paper. Towards this, we present four examples here that fit well into the framework presented in (1). It is worth noting that many of these assume that $\{Y_t\}_{t\in\mathbb{Z}}$ is i.i.d.; so $f_Y(\omega)$ is constant. However, we allow for $\{Y_t\}_{t\in\mathbb{Z}}$ to be a WSS process (which subsumes the i.i.d. case); that is, we do not require $f_Y(\omega)$ to be a constant.

1) Graph signal processing (GSP) extends classical signal processing by analyzing signals supported on a graph. For

[7] https://fcon_1000.projects.nitrc.org/indi/abide/
[8] http://preprocessed-connectomes-project.org/abide/

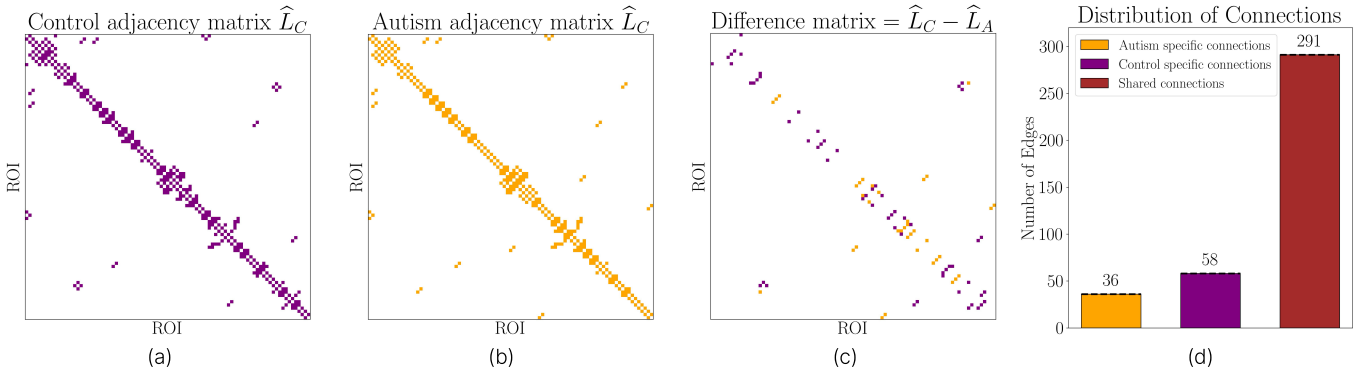

Fig. 6. The results here are obtained using a fixed regularization parameter of $\lambda_n = 0.23$. Each dot in the heatmaps represents a statistically significant edge, i.e., an edge present in more than 90% of the subjects. Panels (a) and (b) display the heatmaps of the estimated common adjacency matrices for the control group ($\widehat{L}_C$) and autism group ($\widehat{L}_A$), respectively, while panel (c) illustrates the difference matrix, $\widehat{L}_C - \widehat{L}_A$. This difference matrix captures both control-specific and autism-specific connections. Panel (d) provides a bar plot representing the distribution of connections, detailing the number of group-specific and shared connections. The bar plot indicates that the control adjacency matrix is denser than that of the autism group.

random signals, a simple generative model is $Y_t = H(\alpha)X_t$. Here X_t is white noise and $H(\alpha) = \sum_{k=0}^{K-1} \alpha_k S^k$ is the graph filter for a given α_k and K. The shift matrix S (e.g., adjacency or Laplacian) encodes the edge connectivity of the graph. [17] discusses several methods to infer sparsity pattern of S from finitely many observations of Y_t for a variety of loss functions $\mathcal{L}[\cdot]$. Note that when $K \to \infty$, $\alpha_k = 1$, and $S = L - I$, we have[9] $H(\alpha) = (I - S)^{-1} = L^{-1}$. Thus, $f_Y(\omega) = H(\alpha) f_X(\omega) H(\alpha)^{\mathsf{T}}$ becomes the constraint in our learning problem in (10).

2) Structural equation models (SEMs) are used to model cause-and-effect relationships between variables, allowing us to infer the causal structure of systems in medicine, economics, and social sciences. Networks generated by SEMs, including directed acyclic graphs are of great interest [29].

A random vector $Y_t \in \mathbb{R}^p$ follows linear SEM if $Y_t = B^{\mathsf{T}} Y_t + X_t$. The path (or autoregressive) matrix B is upper triangular—a structure essential for modeling causal relationships. Therefore, we can take $L = I - B^{\mathsf{T}}$ in (10) to reproduce this problem setup. However, our theoretical results need to be suitably adapted to handle a non-symmetric matrix L needed for SEMs, and we leave this for future work.

3) Cholesky decomposition for correlation networks: Let $Y_t \sim \mathcal{N}(0, \Sigma)$. The sparsity pattern of Σ or the inverse $\Omega = \Sigma^{-1}$ allows us to construct the correlation and partial correlation networks, respectively [83]. Learning sparse covariance or inverse covariance matrices has been well-studied (see Section I-B).

However, for a clear statistical interpretation, one wants to learn the underlying Cholesky matrices T or W, where $\Sigma = T D_1 T^{\mathsf{T}}$ or $\Omega = W D_2 W^{\mathsf{T}}$. The sparse triangular matrices T and W can be learned using our framework in (11) by letting $f_X(\omega) = D$ and $L^* = W^{-1}$. However, our approach is more general and does not constrain L^* to be triangular.

4) Factor analysis (FA) is a statistical method that discovers latent structures within high-dimensional data and is used in Finance and Psychology. The fundamental FA equation is $\tilde{X}_t = \Lambda Y_t + \Phi U_t$. Here Y_t and U_t are called the common and unique factors; and Λ (loading) and Φ (diagonal) are parametric matrices [84, Chapter 5]. Assuming the contribution from the unique factor is known, define $X_t \triangleq \tilde{X}_t - \Phi U_t = \Lambda Y_t$, where Λ plays the role of L^*. Then by treating $\tilde{X}_t$ as a latent random signal, we can use the estimator in (10) to learn Λ.

[9]The invertibility of the Laplacian matrix L is discussed in Remark 1.

VI. Conclusion and Future Work

We study the structure learning problem in systems obeying conservation laws under wide-sense stationary (WSS) stochastic injections. This problem appears in domains like power, the human brain, finance, and social networks. We propose a novel ℓ_1-regularized (approximate) Whittle likelihood estimator to solve the network learning problem for WSS injections that include Gaussian and a few classes of non-Gaussian processes. Our theoretical analysis demonstrates that the estimator is convex and has a unique minimum in the high-dimensional regime. We establish sample complexity guarantees for recovering the sparsity structure of L^*, along with norm-consistency bounds (that is, estimation error computed using element-wise maximum, Frobenius, and operator norms). We validate our theoretical results on synthetic, benchmark, and real-world networks under VAR(1) and VARMA(2,2) injections.

We identify three significant future extensions. First, deriving minimax lower bounds to establish the statistical optimality of our estimator building upon the tools developed in [85]. Second, the work in [86] showed that incorporating diagonal dominance and non-positive off-diagonal constraints of Laplacian matrices could improve the estimation performance for precision matrices modeled as Laplacians. Thus, it would be interesting to exploit such constraints into the estimator in (10), and also to relax the symmetry assumption. Non-symmetric Laplacian matrices model directional flows and appear in many fields like transportation, hydrodynamics, and neuronal networks; see [3].

Finally, we could broaden the class of distributions considered for the nodal injection process X_t. Although we model X_t as a WSS process, non-stationarity often arises in applications such as task-based fMRI signals in neuroscience [87] and stock

market data, which is frequently modeled by Brownian or Lévy processes [88, 89]. Characterizing sample complexity results for non-stationary processes is challenging and much work needs to be done.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation (NSF) award CCF-2048223 and the National Institutes of Health (NIH) under the award 1R01GM140468-01. D. Deka acknowledges the funding provided by LANL's Directed Research and Development (LDRD) project: "High-Performance Artificial Intelligence" (20230771DI).

REFERENCES

[1] S. H. Strogatz, "Exploring complex networks," *nature*, vol. 410, no. 6825, pp. 268–276, 2001.

[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[3] A. van der Schaft, "Modeling of physical network systems," *Systems & Control Letters*, vol. 101, pp. 21–27, 2017.

[4] A. Bressan, S. Čanić, M. Garavello, M. Herty, and B. Piccoli, "Flows on networks: recent results and perspectives," *EMS Surveys in Mathematical Sciences*, vol. 1, pp. 47–111, 2014.

[5] H. U. Voss and N. D. Schiff, "Searching for conservation laws in brain dynamics—bold flux and source imaging," *Entropy*, vol. 16, no. 7, pp. 3689–3709, 2014.

[6] B. Podobnik, M. Jusup, Z. Tiganj, W.-X. Wang, J. M. Buldú, and H. E. Stanley, "Biological conservation law as an emerging functionality in dynamical neuronal networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 45, pp. 11,826–11,831, 2017.

[7] F. R. Chung and F. C. Graham, *Spectral graph theory*. American Mathematical Soc., 1997, no. 92.

[8] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Network topology inference from non-stationary graph signals," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 5870–5874.

[9] A. Rayas, R. Anguluri, and G. Dasarathy, "Learning the Structure of Large Networked Systems Obeying Conservation Laws," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 14,637–14,650.

[10] D. Deka, S. Talukdar, M. Chertkov, and M. V. Salapaka, "Graphical models in meshed distribution grids: Topology estimation, change detection & limitations," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4299–4310, 2020.

[11] R. Anguluri, G. Dasarathy, O. Kosut, and L. Sankar, "Grid topology identification with hidden nodes via structured norm minimization," *IEEE Control Systems Letters*, vol. 6, pp. 1244–1249, 2021.

[12] S. Segarra, A. G. Marques, M. Goyal, and S. Rey-Escudero, "Network topology inference from input-output diffusion pairs," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2018, pp. 508–512.

[13] G. Park, S. J. Moon, S. Park, and J.-J. Jeon, "Learning a high-dimensional linear structural equation model via l1-regularized regression," *Journal of Machine Learning Research*, vol. 22, no. 102, pp. 1–41, 2021.

[14] A. Pruttiakaravanich and J. Songsiri, "Convex formulation for regularized estimation of structural equation models," *Signal Processing*, vol. 166, 2020.

[15] H. Doddi, D. Deka, S. Talukdar, and M. Salapaka, "Efficient and passive learning of networked dynamical systems driven by non-white exogenous inputs," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 9982–9997.

[16] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Identifying the topology of undirected networks from diffused non-stationary graph signals," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 171–189, 2021.

[17] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.

[18] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[19] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3675–3687, 2017.

[20] N. Deb, A. Kuceyeski, and S. Basu, "Regularized estimation of sparse spectral precision matrices," *arXiv preprint arXiv:2401.11128*, 2024.

[21] A. Dallakyan, R. Kim, and M. Pourahmadi, "Time series graphical Lasso and sparse VAR estimation," *Computational Statistics & Data Analysis*, vol. 176, 2022.

[22] S. Basu and G. Michailidis, "Regularized estimation in sparse high-dimensional time series models," *Annals of Statistics*, pp. 1535–1567, 2015.

[23] H. Doddi, D. Deka, S. Talukdar, and M. V. Salapaka, "Learning networked linear dynamical systems under non-white excitation from a single trajectory." *CoRR*, 2021.

[24] H. Doddi, S. Talukdar, D. Deka, and M. Salapaka, "Exact topology learning in a network of cyclostationary processes," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 4968–4973.

[25] S. Ranciati, A. Roverato, and A. Luati, "Fused graphical Lasso for brain networks with symmetries," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 70, no. 5, pp. 1299–1322, 2021.

[26] R. P. Monti, P. Hellyer, D. Sharp, R. Leech, C. Anagnostopoulos, and G. Montana, "Estimating time-varying brain connectivity networks from functional MRI time series," *NeuroImage*, vol. 103, pp. 427–443, 2014.

[27] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (LASSO)," *IEEE transactions on information theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

[28] S. A. Van de Geer, "High-dimensional generalized linear models and the Lasso," *The Annals of Statistics*, vol. 36, no. 2, pp. 614–645, 2008.

[29] M. Drton and M. H. Maathuis, "Structure learning in graphical modeling," *Annual Review of Statistics and Its Application*, vol. 4, no. Volume 4, 2017, pp. 365–393, 2017.

[30] M. Yuan and Y. Lin, "Model selection and estimation in the Gaussian graphical model," *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.

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

[32] A. Dallakyan and M. Pourahmadi, "Fused-Lasso regularized Cholesky factors of large nonstationary covariance matrices of replicated time series," *Journal of Computational and Graphical Statistics*, vol. 32, no. 1, pp. 157–170, 2023.

[33] C. Chang and R. S. Tsay, "Estimation of covariance matrix via the sparse Cholesky factor with Lasso," *Journal of Statistical Planning and Inference*, vol. 140, no. 12, pp. 3858–3873, 2010.

[34] K. Tsai, O. Koyejo, and M. Kolar, "Joint gaussian graphical model estimation: A survey," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 14, no. 6, 2022.

[35] L. Chen, "Estimation of graphical models: An overview of selected topics," *International Statistical Review*, vol. 92, no. 2, pp. 194–245, 2024.

[36] J. Ying, J. V. de Miranda Cardoso, and D. P. Palomar, "Does the ℓ_1 norm learn a sparse graph under Laplacian constrained graphical models?" *arXiv preprint arXiv:2006.14925*, 2020.

[37] S. Kumar, J. Ying, J. V. de Miranda Cardoso, and D. Palomar, "Structured graph learning via laplacian spectral constraints," *Advances in neural information processing systems*, vol. 32, 2019.

[38] J. Ying, X. Han, R. Zhou, X. Wang, and H. C. So, "Network topology inference with sparsity and laplacian constraints," in *2023 IEEE 11th International Conference on Information, Communication and Networks (ICICN)*. IEEE, 2023, pp. 283–288.

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

[40] C. Baek, M.C. Düker, and V. Pipiras, "Local Whittle estimation of high-dimensional long-run variance and precision matrices," *arXiv preprint arXiv:2105.13342*, 2021.

[41] M. Zorzi and R. Sepulchre, "Ar identification of latent-variable graphical models," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2327–2340, 2015.

[42] M. Zorzi, "Empirical bayesian learning in ar graphical models," *Automatica*, vol. 109, p. 108516, 2019.

[43] F. Crescente, L. Falconi, F. Rozzi, A. Ferrante, and M. Zorzi, "Learning ar factor models," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 274–279.

[44] L. Falconi, A. Ferrante, and M. Zorzi, "A robust approach to arma factor modeling," *IEEE Transactions on Automatic Control*, vol. 69, no. 2, pp. 828–841, 2023.

[45] M. Zorzi, "On the identification of arma graphical models," *IEEE Transactions on Automatic Control*, 2024.

[46] D. Alpago, M. Zorzi, and A. Ferrante, "Identification of sparse reciprocal graphical models," *IEEE Control Systems Letters*, vol. 2, no. 4, pp. 659–664, 2018.

[47] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1061–1074, 2018.

[48] D. Deka, S. Talukdar, M. Chertkov, and M. V. Salapaka, "Graphical models in meshed distribution grids: Topology estimation, change detection & limitations," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4299–4310, 2020.

[49] D. Deka, V. Kekatos, and G. Cavraro, "Learning distribution grid topologies: A tutorial," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 999–1013, 2023.

[50] S. Grotas, Y. Yakoby, I. Gera, and T. Routtenberg, "Power systems topology and state estimation by graph blind source separation," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2036–2051, 2019.

[51] P. Whittle, "Estimation and information in stationary time series," *Arkiv för matematik*, vol. 2, no. 5, pp. 423–434, 1953.

[52] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*. Springer science & business media, 2009.

[53] I. S. Dhillon and J. A. Tropp, "Matrix nearness problems with Bregman divergences," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 4, pp. 1120–1146, 2008.

[54] F. Dorfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 1, pp. 150–163, 2012.

[55] Y. Sun, Y. Li, A. Kuceyeski, and S. Basu, "Large spectral density matrix estimation by thresholding," *arXiv preprint arXiv:1812.00532*, 2018.

[56] M. Fiecas, C. Leng, W. Liu, and Y. Yu, "Spectral analysis of high-dimensional time series," *Electronic Journal of Statistics*, vol. 13, no. 2, pp. 4079 – 4101, 2019.

[57] C. Baek, M.-C. Düker, and V. Pipiras, "Local whittle estimation of high-dimensional long-run variance and precision matrices," *The Annals of Statistics*, vol. 51, no. 6, pp. 2386–2414, 2023.

[58] S. Basu and S. Subba Rao, "Graphical models for nonstationary time series," *The Annals of Statistics*, vol. 51, no. 4, pp. 1453–1483, 2023.

[59] J. Krampe and E. Paparoditis, "Frequency domain statistical inference for high-dimensional time series," *Journal of the American Statistical Association*, no. just-accepted, pp. 1–22, 2025.

[60] C. Hurvich, "Whittle's approximation to the likelihood function," *Lecture Notes (New York University Stern School of Business, 2002)*, 2002.

[61] A. Jung, G. Hannak, and N. Goertz, "Graphical Lasso based model selection for time series," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1781–1785, 2015.

[62] C. Baek, M.-C. Düker, and V. Pipiras, "Thresholding and graphical local whittle estimation," *arXiv preprint arXiv:2105.13342*, 2021.

[63] F. Bunea, A. Tsybakov, and M. Wegkamp, "Sparsity oracle inequalities for the Lasso," *Electronic Journal of Statistics*, vol. 1, pp. 169 – 194, 2007.

[64] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *The Annals of Statistics*, vol. 37, no. 4, pp. 1705 – 1732, 2009.

[65] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, "A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers," *Statistical Science*, vol. 27, no. 4, pp. 538 – 557, 2012.

[66] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

[67] L. Dai, K. Chen, Z. Sun, Z. Liu, and G. Li, "Broken adaptive ridge regression and its asymptotic properties," *Journal of multivariate analysis*, vol. 168, pp. 334–351, 2018.

[68] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[69] P.-L. Loh and M. J. Wainwright, "Support recovery without incoherence: A case for nonconvex regularization," *The Annals of Statistics*, vol. 45, no. 6, pp. 2455 – 2482, 2017.

[70] P. Zhao and B. Yu, "On model selection consistency of Lasso," *The Journal of Machine Learning Research*, vol. 7, pp. 2541–2563, 2006.

[71] T. Cai, W. Liu, and X. Luo, "A constrained ℓ_1-minimization approach to sparse precision matrix estimation," *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 594–607, 2011.

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

[73] S. S. Rao and J. Yang, "Reconciling the gaussian and whittle likelihood with an application to estimation in the frequency domain," *The Annals of Statistics*, vol. 49, no. 5, pp. 2774–2802, 2021.

[74] M. Rosenblatt, *Stationary sequences and random fields*. Springer Science & Business Media, 2012.

[75] R. B. Kellogg, T.Y. Li, and J. Yorke, "A constructive proof of the Brouwer fixed-point theorem and computational results," *SIAM Journal on Numerical Analysis*, vol. 13, no. 4, pp. 473–483, 1976.

[76] S. Jayadev, S. Narasimhan, and N. Bhatt, "Learning conserved networks from flows," *arXiv preprint arXiv:1905.08716*, 2019.

[77] J. Chen and Z. Chen, "Extended Bayesian information criteria for model selection with large model spaces," *Biometrika*, vol. 95, no. 3, pp. 759–771, 2008.

[78] E. Hernadez, S. Hoagland, and L. Ormsbee, "Water distribution database for research applications," in *World Environmental and Water Resources Congress*, 2016, pp. 465–474.

[79] F. Seccamonte, "Bilevel optimization in learning and control with applications to network flow estimation," Ph.D. dissertation, UC Santa Barbara, 2023.

[80] A. Škoch, B. Rehák Bučková, J. Mareš, J. Tintěra, P. Sanda, L. Jajcay, J. Horáček, F. Španiel, and J. Hlinka, "Human brain

structural connectivity matrices–ready for modelling," *Scientific Data*, vol. 9, no. 1, 2022.

[81] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex," *PLoS biology*, vol. 6, no. 7, 2008.

[82] D. S. Bassett and E. Bullmore, "Small-world brain networks," *The neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.

[83] M. Pourahmadi, "Covariance Estimation: The GLM and Regularization Perspectives," *Statistical Science*, vol. 26, pp. 369 – 387, 2011.

[84] N. Trendafilov and M. Gallo, *Multivariate data analysis on matrix manifolds*. Springer, 2021.

[85] J. Ying, J. V. de Miranda Cardoso, and D. Palomar, "Minimax estimation of Laplacian constrained precision matrices," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 3736–3744.

[86] S. Kumar, J. Ying, J. V. d. M. Cardoso, and D. P. Palomar, "A unified framework for structured graph learning via spectral constraints," *Journal of Machine Learning Research*, vol. 21, no. 22, pp. 1–60, 2020.

[87] M. G. Preti, T. A. Bolton, and D. Van De Ville, "The dynamic functional connectome: State-of-the-art and perspectives," *Neuroimage*, vol. 160, pp. 41–54, 2017.

[88] C. Peng and C. Simon, "Financial modeling with geometric brownian motion," *Open Journal of Business and Management*, vol. 12, no. 2, pp. 1240–1250, 2024.

[89] S. Engelke, J. Ivanovs, and J. D. Thøstesen, "Lévy graphical models," *arXiv preprint arXiv:2410.19952*, 2024.