Information Theoretically Optimal Sample Complexity of Learning *Dynamical* Directed Acyclic Graphs

Mishfad Shaikh Veedu*, Deepjyoti Deka †, and Murti V. Salapaka*

Abstract

In this article, the optimal sample complexity of learning the underlying interaction/dependencies of a Linear Dynamical System (LDS) over a Directed Acyclic Graph (DAG) is studied. The sample complexity of learning a DAG's structure is well-studied for static systems, where the samples of nodal states are independent and identically distributed (i.i.d.). However, such a study is less explored for DAGs with dynamical systems, where the nodal states are temporally correlated. We call such a DAG underlying an LDS as *dynamical* DAG (DDAG). In particular, we consider a DDAG where the nodal dynamics are driven by unobserved exogenous noise sources that are wide-sense stationary (WSS) in time but are mutually uncorrelated, and have the same power spectral density (PSD). Inspired by the static settings, a metric and an algorithm based on the PSD matrix of the observed time series are proposed to reconstruct the DDAG. The equal noise PSD assumption can be relaxed such that identifiability conditions for DDAG reconstruction are not violated. For the LDS with WSS (sub) Gaussian exogenous noise sources, it is shown that the optimal sample complexity (or length of state trajectory) needed to learn the DDAG is $n = \Theta(q \log(p/q))$, where p is the number of nodes and q is the maximum number of parents per node. To prove the sample complexity upper bound, a concentration bound for the PSD estimation is derived, under two different sampling strategies. A matching min-max lower bound using generalized Fano's inequality also is provided, thus showing the order optimality of the proposed algorithm.

1 Introduction

Learning the interdependency structure in a network of agents, from passive time series observations, is a salient problem with applications in neuroscience Bower and Beeman (2012), finance Kim et al. (2011), meteorology Ghil et al. (2002), etc. Reconstructing the exact structure with the dependency/causation directions has a wide range of applications. For example, the identification of causation structure among the shares helps in obtaining robust portfolio management in the stock market Kim et al. (2011). Similarly, causal graphs are useful in understanding dynamics and identifying contributing factors of a public epidemic emergency situation Yang et al. (2020).

*Department of Electrical Engineering, University of Minnesota Twin Cities

†Theoretical Division T-5, Los Alamos National Laboratory.

The authors acknowledge support from the Information Science and Technology Institute (ISTI) at Los Alamos National Laboratory. The research work conducted at Los Alamos National Laboratory is done under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy under Contract No. 89233218CNA000001. The first and the third authors acknowledge the support of NSF through Award Number 2030096 and ARPA-E through the Award number DE-AR0001016.

The structure of directed interactions in a network of agents is conveniently represented using directed graphs, with the agents as nodes and the directed interactions as directed edges. If the underlying graph doesn't have cycles, it is called a Directed Acyclic Graph (DAG). In general, it is not possible to reconstruct the exact structure with the direction of different edges. Instead, in many networks it is possible to retrieve only the Markov equivalence graphs, the set of graphs satisfying the same conditional dependence property, from data without any intervention Ghoshal and Honorio (2018). In applications such as finance Kim et al. (2011), climate science Ghil et al. (2002) etc., the agent states, instead of being temporally independent, can evolve over time due to past directed interactions. Such temporal evolution can be represented by a linear dynamical system (LDS). In LDS, the interaction between agent states is captured by a linear time-invariant function. In this paper, we study the identifiability and present the first sample complexity results for learning a DAG of LDS, which we term as *Dynamical* DAG or **DDAG**. This is distinguished from *static* DAG, where the agent states are temporally independent and the DAG does not correspond to temporal dynamics.

1.1 Related Work

Static DAG Learning: The problem of obtaining an upper bound on the sample complexity of learning static DAGs goes back twenty-five years Friedman and Yakhini (1996), Zuk et al. (2006). However, tight characterization of optimal rates for DAG learning is a harder problem compared to undirected networks Gao et al. (2022), primarily due to the order identification step. Identifiability conditions for learning static DAGs with linear interactions and excited by equal variance Gaussian noise were given in Peters and Bühlmann (2014). Several polynomial time algorithms have been proposed for static DAG reconstruction using samples of states at the graph nodes; see Ghoshal and Honorio (2017a,b, 2018); Chen et al. (2019); Gao et al. (2022); Park (2020); Park and Raskutti (2017), and the reference therein. An information-theoretic lower bound on structure estimation was studied in Ghoshal and Honorio (2017a). In Gao et al. (2022), it was shown that the order optimal sample complexity for static Gaussian graphical model with equal variance is $n = \Theta(q \log(p/q))$, where p is the number of nodes and q is the maximum number of parents. The authors showed that the algorithm given in Chen et al. (2019) provides an upper bound that matches a min-max lower bound for the number of samples. However, similar results for DDAGs, with underlying temporal directed interaction between agent states, have not been studied, to the best of our knowledge.

LDS Learning: Graph reconstruction, in general, is challenging in a network of LDS (directed, undirected, or bi-directed), as it involves time-dependencies between collected samples of nodal states. Learning the conditional independence structure in LDS with independent and identically distributed (white) excitation was explored in Basu and Michailidis (2015), Loh and Wainwright (2011), Songsiri and Vandenberghe (2010), Simchowitz et al. (2018), Faradonbeh et al. (2018), and the references therein. However, the methods in the cited papers do not extend to LDS that is excited by WSS (wide-sense stationary) noise, which makes correlations in state samples more pronounced. For LDS with WSS noise, Tank et al. (2015), Dahlhaus (2000), Materassi and Salapaka (2013), and Materassi and Innocenti (2009), estimated the conditional correlation structure, which contains true edges in the network and extra edges between some two-hop neighbors Materassi and Salapaka (2012). A consistent algorithm for the recovery of exact topology in a large class of applications with WSS noise was provided in Talukdar et al. (2020), with the corresponding sample-complexity analysis developed in Doddi et al. (2022), using a neighborhood-based regression framework. However, the developed algorithms and related sample complexity results do not extend to directed graphs and hence exclude DDAG reconstruction. DDAG reconstruction from the time-series data has been explored using the framework of directed mutual information in Quinn et al. (2015) but without rate characterization for learning from finite samples.

Contribution: This article presents an information-theoretically optimal sample complexity analysis for learning a dynamical DAG (DDAG, i.e., a DAG underlying an LDS excited by WSS noise of equal power spectral density), using samples of state trajectories of the corresponding LDS. To the best of our knowledge, this is the first paper to study and prove sample complexity analysis for DDAGs. We consider learning under two sampling scenarios, viz; 1) restart and record, 2) continuous sampling. While the former pertains to samples collected from multiple disjoint (independent) trajectories of state evolution, the latter includes samples from a single but longer trajectory of the state evolution (see Fig. 2). Surprisingly, the results in this article show that the estimation errors are not influenced by the sampling strategy (restart and record or continuous) as long as the collected samples are over a determined threshold given by $n = O(q\log(p/q))$ is obtained, where p is the number of nodes and q is the maximum number of parents per node. We also provide a matching information-theoretic lower-bound, $\max\left(\frac{\log p}{2\beta^2+\beta^4}, \frac{q\log(p/q)}{M^2-1}\right)$, where β and M are system parameters (see Definition 2.4); thus obtaining an order optimal bound $n = \Theta(q\log(p/q))$.

Our learning algorithm relies on first deriving rules for DAG estimation using the Power Spectral Density Matrix (PSDM) of nodal states, inspired by the estimator for static DAGs based on covariance matrices in Chen et al. (2019). Subsequently, the sample complexity associated with learning is derived by obtaining concentration bounds for the PSDM. In this regard, characterization of non-asymptotic bounds of PSDMs for a few spectral estimators have been obtained Fiecas et al. (2019); Veedu et al. (2021); Zhang and Wu (2021) previously. A unified framework of concentration bounds for a general class of PSDM estimators was recently presented in Lamperski (2023). Our concentration bounds of the PSDM are reached using different proof steps, based on Rademacher random variables and symmetrization argument Wainwright (2019).

The rest of the paper is organized as follows. Section 2 introduces the system model and the preliminary definitions for LDS and DDAGs. Section 3 discusses Algorithm 1 and the main results for DDAG reconstruction from PSDM. Section 4 provides a concentration bound for the error in estimating the PSDM and a sample complexity upper bound for DDAG reconstruction using Algorithm 1. Section 5 contains a sample complexity lower bound.

Notations: Bold faced small letters, $\mathbf{x}$ denote vectors; Bold faced capital letters, $\mathbf{A}$ denote matrices; For a time-series, $\mathbf{x}$, $\breve{x}(t)$ denotes the value of $\mathbf{x}$ at time t, $\mathbf{x}(\omega)$ denotes the discrete time Fourier transform of $\mathbf{x}$, $\mathbf{x}(\omega) := \sum_{k=-\infty}^{\infty} \breve{\mathbf{x}}(k)e^{-i\omega k}$, $\omega \in \Omega = [0, 2\pi]$; $diag(v_1, \dots, v_p)$ operator creates a diagonal matrix with diagonal entries $v_1, \dots, v_p$; $\Phi_{\mathbf{x}_{AB}}$ or Φ_{AB} denotes the matrix obtained by selecting rows A and columns B in $\Phi_{\mathbf{x}}$; $\mathbf{A}^*$ denotes conjugate transpose of $\mathbf{A}$. $\|\mathbf{A}\|$ is the spectral norm of $\mathbf{A}$ and $\|\mathbf{v}\|_2$ is the Euclidean norm of vector $\mathbf{v}$.

2 System model of DDAG

We describe different aspects of DDAG (DAG underlying an LDS) and the sampling strategies considered. We begin with some necessary DAG terminologies.

DDAG terminology: The *dynamical* directed acyclic graph (DDAG), is given by $G := (V, E)$, where $V = \{1, \dots, p\}$, and E is edge set of directed edge $i \to j$. A directed path from i to j is a path of the form $i = v_0 \to v_1 \to \dots \to v_\ell \to v_{\ell+1} = j$, where $v_k \in V$ and

$(v_k, v_{k+1}) \in E$ for every $k = 0, \ldots, \ell$. A cycle is a directed path from i to i, which does not exist in DDAG G. For G, $pa(i) := \{j \in V : (j, i) \in E\}$ denotes the parents set and $desc(i)$ denotes the descendants of i, the nodes that have a directed path from i. The set $nd(i) := V \setminus desc(i)$ denotes the non-descendants set and the set $an(i) \subset nd(i)$ denotes the ancestors set, the nodes that have a directed path to i. At set $C \subseteq V$ is called ancestral if for every $i \in C$, $pa(i) \subseteq C$. Figure 2 shows an example DDAG with the node definitions. A node set $C \subseteq V$ is said to be a topological ordering on G if for every $i, j \in C$, $i \in desc(j)$ in G implies $i > j$. $\mathcal{G}_{p,q}$ denotes the family of DDAGs with p nodes and at most q parents per node. Without a loss of generalizability, we use the same terminology for the DDAG and the

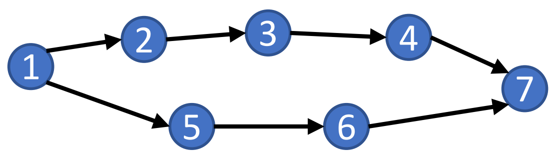

Figure 1: An example DDAG. Node 1 is an ancestor and node 7 is a descendant of every node in the graph. The set $\{1, 2, 5\}$ is an ancestral set but $\{2, 5\}$ is not. $an(3) = \{1, 2\}$, $desc(3) = \{4, 7\}$, $nd(3) = \{1, 2, 5, 6\}$.

underlying DAG.

LDS Model excited by equal PSD WSS noise: For the DDAG $G = (V, E) \in \mathcal{G}_{p,q}$, we consider a linear dynamical system (LDS) with p scalar state variables, corresponding to nodes in V. Node i is equipped with time series measurements, $\{\breve{x}_i(k)\}_{k \in \mathbb{Z}}$, $1 \leq i \leq p$. The LDS evolves according to the linear time-invariant model,

$$\breve{x}_i(k) = \sum_{(i,j) \in E, j \neq i}^{p} (\breve{h}_{ij} \star \breve{x}_j)(k) + \breve{e}_i(k), \; k \in \mathbb{Z}, \tag{1}$$

where transfer function $\breve{h}_{ij} \neq 0$ when directed edge $(i, j) \in E$. The exogenous noise $\{\breve{e}_i(k)\}_{k \in \mathbb{Z}}$, $1 \leq i \leq p$, are zero mean wide sense stationary Gaussian processes, uncorrelated across nodes. Taking the discrete-time Fourier transform (DTFT) of (1) provides the frequency representation for every $\omega \in \Omega = [0, 2\pi]$,

$$\mathbf{x}_i(\omega) = \sum_{(i,j) \in E, j \neq i}^{p} \mathbf{H}_{ij}(\omega) \mathbf{x}_j(\omega) + \mathbf{e}_i(\omega), \; 1 \leq i \leq p, \tag{2}$$

where $\mathbf{x}_i(\omega) = \mathcal{F}\{\breve{x}_i\} := \sum_{k=-\infty}^{\infty} \breve{x}_i(k) e^{-i\omega k}$, $\mathbf{e}_i(\omega) = \mathcal{F}\{\breve{e}_i\}$, and $\mathbf{H}_{ij}(\omega) = \mathcal{F}\{\breve{h}_{ij}\}$. The model in (2) can be represented in the matrix form to obtain the following LDS,

$$\mathbf{x}(\omega) = \mathbf{H}(\omega)\mathbf{x}(\omega) + \mathbf{e}(\omega), \; \forall \omega \in \Omega, \tag{3}$$

where $\mathbf{e}(\omega)$ is the WSS noise. In this article, we are interested in the LDS with $\Phi_{\mathbf{e}}(\omega) = \sigma(\omega) diag(\alpha_1, \ldots, \alpha_p)$, where α_i are known and can be a function of ω. For the simplicity of analysis, henceforth it is assumed that $\Phi_{\mathbf{e}}(\omega) = \sigma(\omega)\mathbf{I}$.

Remark 2.1. *The assumption $\Phi_{\mathbf{e}}(\omega) = \sigma(\omega)\mathbf{I}$ is a restrictive assumption. However, we would like to remark that some form of restriction is required due to the impossibility of DAG* *reconstruction in a general setup due to identifiability issues Shimizu et al. (2006). However, the assumption can be relaxed to incorporate the identifiability conditions on Φ_{e_i} and $\mathbf{H}$ to retrieve the topological ordering, similar to Ghoshal and Honorio (2018). Furthermore, our results on DDAG reconstruction require the equal PSD to hold* only at some known $\omega \in \Omega$ *which is less restrictive.*

The power spectral density matrix (PSDM) of the time-series $\mathbf{x}$ at the angular frequency $\omega \in \Omega$ is given by

$$\Phi_{\mathbf{x}}(\omega) = \mathcal{F}\left\{R_{\mathbf{x}}(t)\right\} = \sum_{k=-\infty}^{\infty} R_{\mathbf{x}}(k)e^{-i\omega k}, \tag{4}$$

where $R_{\mathbf{x}}(k) := \mathbb{E}[\check{\mathbf{x}}(k)\check{\mathbf{x}}^T(0)]$ is the auto-correlation matrix of the time-series $\mathbf{x}$ at lag k. The (i,j)-th entry of $\Phi_{\mathbf{x}}$ is denoted by Φ_{ij}. For the LDS (3), the PSDM is given by

$$\Phi_{\mathbf{x}}(\omega) = (\mathbf{I} - \mathbf{H}(\omega))^{-1}\Phi_{\mathbf{e}}(\omega)((\mathbf{I} - \mathbf{H}(\omega))^{-1})^{*}. \tag{5}$$

Consider the following additional non-restrictive assumptions on the power spectral density and correlation matrix of the LDS states.

Assumption 2.2. *There exists an $M \in \mathbb{R}$ such that $\frac{1}{M} \leq \lambda_{min}(\Phi_{\mathbf{x}}) \leq \lambda_{max}(\Phi_{\mathbf{x}}) \leq M$, where λ_{min} and λ_{max} respectively denote minimum and maximum eigenvalues.*

Assumption 2.3. *The auto-correlation matrix of the time-series $\mathbf{x}$ at lag k, $R_{\mathbf{x}}(k) := \mathbb{E}[\check{\mathbf{x}}(k)\check{\mathbf{x}}^T(0)]$ satisfies $\|R_{\mathbf{x}}(k)\| \leq C\rho^{-|k|}$, for some positive constants $C, \rho \in \mathbb{R}$, $\rho > 1$.*

In the remaining paper, following these assumptions, our interest will be limited to the following family of DDAGs and corresponding LDS.

Definition 2.4. *$\mathcal{H}_{p,q}(\beta, \sigma, M)$ denotes the family of LDS given by (3) such that the corresponding DDAG, $G(V,E) \in \mathcal{G}_{p,q}$ (p nodes with each node having a maximum q parents), with $|\mathbf{H}_{ij}(\omega)| \geq \beta$, $\forall (i,j) \in E, \omega$, $\Phi_{\mathbf{e}}(\omega) = \sigma(\omega)\mathbf{I}$, and $M^{-1} \leq \lambda_{\min}(\Phi_{\mathbf{x}}(\omega)) \leq \lambda_{max}(\Phi_{\mathbf{x}}(\omega)) \leq M$, $\forall \omega \in \Omega$*

Sampling Strategy for LDS states: We consider two sampling settings (see Fig. 2 for details):
i) **restart and record:** The sampling is performed as follows: start recording and stop it after N measurements. For the next trajectory, the procedure is restarted with an independent realization and record for another epoch of N samples; repeat the process another $n-2$ times providing n i.i.d trajectories of N samples each.
ii) **continuous sampling:** Here, a single trajectory of length $n \times N$ is taken consecutively. Then, the observations are divided into n segments, each having N consecutive samples.

The data collected using either strategy is thus grouped into n trajectories of N samples each. For the finite length r^{th} trajectory, $\{\check{x}^r(t)\}_{t=0}^{N-1}$, define

$$\mathbf{x}^r(\omega) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \check{x}^r(k)e^{-i\omega k}, r = 1, \ldots, n,\ \omega \in \Omega. \tag{6}$$

Note that $\mathbf{x}^r(\omega)$, if exists, is a zero mean random variable with the covariance matrix given by

$$\widetilde{\Phi}_{\mathbf{x}}(\omega) := \mathbb{E}\left\{\mathbf{x}^r(\omega)[\mathbf{x}^r(\omega)]^*\right\} = \frac{1}{N}\sum_{k=-(N-1)}^{N-1}(N-|k|)R_{\mathbf{x}}(k)e^{-i\omega k},\ \forall \omega \in \Omega. \tag{7}$$

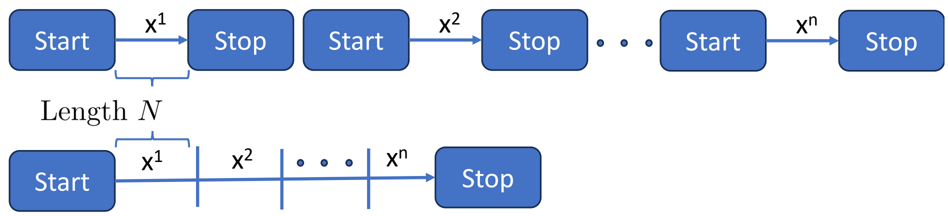

Figure 2: (a) shows restart and record sampling. (b) shows continuous sampling.

For the restart and record setting (unlike for continuous sampling), $\{\mathbf{x}^r(\omega)\}_{r=1}^n$ are i.i.d. Further, as $N \to \infty$, $\widetilde{\Phi}_{\mathbf{x}}(\omega) \to \Phi_{\mathbf{x}}(\omega)$ uniformly in Ω Stoica et al. (2005).

3 Reconstructing DDAGs from PSDM

In this section, we discuss some results on how the PSDM, $\Phi_{\mathbf{x}}$, can be employed to completely reconstruct the DDAG, G when the time series is generated according to (3). Applying these results, inspired from the algorithm for static setting Chen et al. (2019), we propose Algorithm 1 for reconstructing the DDAG, G. First, we prove that conditional PSD (CPSD) of i conditioned on C, defined as

$$f(i, C, \omega) := \Phi_{ii}(\omega) - \Phi_{iC}(\omega)\Phi_{CC}^{-1}(\omega)\Phi_{Ci}(\omega), \tag{8}$$

is a metric sufficient to obtain a topological ordering on the DDAG, G, which aids in the reconstruction of G. Notice that unlike Chen et al. (2019)'s static setting, our algorithm will require the use of CPSD to unveil the dependencies in the DDAG and further affect the sample complexity of learning.

3.1 CPSD and Topological Ordering

Here, we will show that CPSD is the minimum for the nodes that have all the parents included in the conditioning set. We start by proving the result for the source nodes.

Lemma 3.1. *Consider the LDS described by (3). For any $\omega \in \Omega$, let $\alpha^* := \min_{k \in V} \Phi_{kk}(\omega)$. Then $\Phi_{ii}(\omega) = \alpha^*$ if and only if i is a source node.*

Proof. Let $\mathbf{T}(\omega) = (\mathbf{I} - \mathbf{H}(\omega))^{-1}$ and $\Phi_{\mathbf{e}}(\omega) = \sigma(\omega)\mathbf{I}_n$. Then, $\Phi_{\mathbf{x}}(\omega) = \sigma(\omega)\mathbf{T}(\omega)\mathbf{T}^*(\omega)$. By Cayley-Hamilton theorem, there exists constants $a_0(\omega), \dots, a_n(\omega)$ such that $(\mathbf{I} - \mathbf{H}(\omega))^{-1} = \sum_{k=0}^{n-1} a_k(\omega)(\mathbf{I} - \mathbf{H}(\omega))^k$. Using induction, it can be shown that non-diagonal entries of $(\mathbf{I} - \mathbf{H}(\omega))^k$ are zero if and only if there no $k - hop$ path between i and j (almost always). Similarly, (i, i)th entry is 1 if and only if there is no k-hop path between them (almost always).

Then (ignoring (ω)), $\Phi_{ii} = (\Phi_{\mathbf{x}})_{ii} = \sigma \sum_{k=1}^n \mathbf{T}_{ik}\mathbf{T}_{ik} = \sigma(\mathbf{T}_{ii}^2 + \sum_{k \neq i} \mathbf{T}_{ik}^2)$. If i is a source node, then $\mathbf{T}_{ik} = 0$ for every $k \neq i$ and $\mathbf{T}_{ii} = 1$, which implies $\Phi_{ii} = \sigma \mathbf{T}_{ii}^2$. For non-source nodes, $\mathbf{T}_{ik} \neq 0$ for some k, which gives $\Phi_{ii} > \sigma$. $\square$

3.1.1 Conditional PSD deficit

The following is the definition of conditional PSD deficit, which is helpful in proving the subsequent results and in retrieving the DDAG.

Definition 3.2 (The CPSD deficit)**.**

$$\Delta := \min_{\omega\in[-\pi,\pi]} \min_{j\in V} \min_{\substack{C\subsetneq nd(j),\\ Pa(j)\setminus C\neq\emptyset}} f(j,C,\omega)-\sigma(\omega) \tag{9}$$

The following lemma shows that $f(i,C,\omega)$ from Eq. 8 can be used as a metric to obtain a topological ordering of G.

Lemma 3.3. *Consider the DDAG, G, governed by (3). Let $j\in V$ and let $C\subseteq V\setminus\{j\}$ be an ancestral set. Then for every $\omega\in\Omega$,*

$$\begin{aligned} &f(j,C,\omega)=\sigma(\omega) && : \text{ if } Pa(j)\subseteq C,\\ &f(j,C,\omega)\geq\Delta+\sigma(\omega)>\sigma(\omega) && : \text{ if } Pa(j)\nsubseteq C. \end{aligned}$$

Proof. See Appendix A □

Corollary 3.4. *If $Pa(i)\subseteq C$ and $Pa(i)\nsubseteq D$, then $f(i,C,\omega)-f(i,D,\omega)\geq\Delta$.*

Lemma 3.5. $\Delta\geq\beta^2\sigma$.

Proof. Applying (15), $f(j,C,\omega)-\sigma=H_{jD}(\Phi_D-\Phi_{DC}\Phi_{CC}^{-1}\Phi_{CD})H_{jD}^*\geq\sigma|D|\beta^2$. Then

$$\Delta\geq\min_{\omega}\min_{j}|\sigma|D|\beta^2=\sigma|D|\beta^2\overset{(a)}{\geq}\sigma\beta^2,$$

where (a) follows because $|D|\geq1$ if $C\nsubseteq Pa(j)$. □

To determine the DDAG G's structure, first determine a topological ordering of nodes as follows: Beginning with an empty set $\mathcal{S}$, we iteratively add node i in $\mathcal{S}$ where

$$(i,C_i^*)\in\operatorname*{arg\,min}_{\substack{C\subseteq\mathcal{S},|C|\leq q\\ 1\leq j\leq p,\ j\notin\mathcal{S}}} f(j,C,\omega), \tag{10}$$

and f comes from (8). The following Lemma shows that $\mathcal{S}$ is a valid topological ordering w.r.t. G.

Lemma 3.6. *$\mathcal{S}$ is a valid topological ordering with respect to G.*

Proof. In the first step, $C=\emptyset$ and $f(i,C,\omega)=\Phi_{ii}$. By Lemma 3.1 and Lemma 3.3, $\Phi_{ii}=\sigma$ if i is a source node, where as $\Phi_{ii}\geq\sigma+\Delta$ if i is not a source node. Thus, the first node in the ordered set $\mathcal{S}$, $\mathcal{S}_1$, is always a source node. Induction assumption: Nodes $\mathcal{S}_1$ to $\mathcal{S}_n$ in $\mathcal{S}$ follow topological order. For the $n+1$, by Lemma 3.3, for every $C\subseteq\mathcal{S}$, $f(k,C,\omega)$ is minimum for $k\in V\setminus\mathcal{S}$ if and only if $Pa(k)\subseteq C$. Thus nodes $\mathcal{S}_1$ to $\mathcal{S}_{n+1}$ follow a topological order, which proves the result. □

Identification of the parents: Parents of a node are identified from the ordered set by applying Corollary 3.4. Let $D=C\setminus\{k\}$. As shown in Corollary 3.4, if $Pa(i)\subseteq C$ and $k\in Pa(i)$, then $f(i,C,\omega)-f(i,D,\omega)\geq\Delta$. Thus, from the set $\mathcal{S}$, for every node $\mathcal{S}_i$, one can eliminate nodes $\mathcal{S}_1,\dots,\mathcal{S}_{i-1}$ by checking if the difference is greater than Δ. If the difference is greater than Δ for some $\mathcal{S}_k$, then $\mathcal{S}_k$ is a parent of $\mathcal{S}_i$. That is,

Lemma 3.7. *Let* (i, C_i^*) *be a solution of* (10) *and let*

$$P_i := \{j \in C_i^* \mid |f(i, C_i^*, \omega) - f(i, C_i^* \setminus j, \omega)| \geq \Delta\}.$$

Then, $Pa(i) = P_i$.

Applying the above procedure and Lemma 3.3-Lemma 3.7, one can formulate Algorithm 1 to obtain the ordering of the DDAG, G and eventually reconstruct the DDAG exactly. In Algorithm 1, $\widehat{f}(i, C, \omega) := \widehat{\Phi}_{ii}(\omega) - \widehat{\Phi}_{iC}(\omega)\widehat{\Phi}_{CC}^{-1}(\omega)\widehat{\Phi}_{Ci}(\omega)$, an empirical estimate of $f(i, C, \omega)$, is employed instead of f and $\gamma = \Delta/2$. The following lemma proves that if the empirical estimate $\widehat{f}(\cdot)$ is close enough to the original $f(\cdot)$, then Algorithm 1 reconstructs the DDAG exactly.

Algorithm 1 Ordering algorithm

Input: Estimated PSDM, $\widehat{\Phi}(\omega)$: Δ, $z = e^{j\omega}$, $\omega \in (-\pi, \pi]$
Output: $\widehat{G}$

1. Initialize the ordering, $\mathcal{S} \leftarrow ()$
2. For $i = 1, \ldots, p$
   (a) Compute $\left(j^*, C_j^*\right) \in \underset{\substack{C \subseteq \mathcal{S}, |C| \leq q \\ 1 \leq j \leq p,\ j \notin \mathcal{S}}}{\arg\min} \widehat{f}(j, C, \omega)$
   (b) $\mathcal{S} \leftarrow (\mathcal{S},\ j^*)$
3. $\widehat{G} = (V, \widehat{E}), V \leftarrow \{1, \ldots, p\}, \widehat{E} \leftarrow \emptyset$
4. For $i = 1, \ldots, p$
   (a) Parents of i, $P_i := \left\{j \in C_i^* \,\middle|\, |\widehat{f}(\mathcal{S}_i, C_i^*, \omega) - \widehat{f}(\mathcal{S}_i, C_i^* \setminus j, \omega)| \geq \gamma\right\}$
   (b) For $k \in P_i$
      i. Do $\widehat{E} \leftarrow \widehat{E} \cup (k, i)$
5. Return $\widehat{G}$

Lemma 3.8. *If* $|\widehat{f}(i, C, \omega) - f(i, C, \omega)| < \Delta/4$, *for every* i, ω, C, *then Algorithm 1 reconstructs the DAG,* G *successfully. That is,* $G = \widehat{G}$.

Proof. See Appendix B ☐

Therefore, it suffices to derive the conditions under which $|f(\cdot) - \widehat{f}(\cdot)| < \Delta/4$. In the following section, we derive a concentration bound to guarantee a small error, which in turn is applied in obtaining the upper bound on the sample complexity of estimating G.

4 Finite Sample Analysis of Reconstructing DDAGs

In Lemma 3.8, it was shown that the DDAG, G, can be reconstructed exactly if the error in estimating $f(i, C, \omega)$ (given by (8)) is small enough. In this section, a concentration bound on the error in estimating $\Phi_{\mathbf{x}}$ from finite data is obtained, which is used later to obtain a concentration bound on the error in estimating the metric f.

Recall that we consider n state trajectories (see Fig. 2 for the two sampling strategies) with each trajectory being of length N samples, i.e $\left\{\{\breve{\mathbf{x}}^r(k)\}_{k=0}^{N-1}\right\}_{r=1}^{n}$. The DFTs for each trajectory, $\mathbf{x}^r(\omega)$, is a complex Gaussian with mean zero and covariance matrix, $\widetilde{\Phi}_x(\omega))$, i.e., for every $r=1,\dots,n$, $\mathbf{x}^r(\omega)\sim\mathcal{N}(0,\widetilde{\Phi}_{\mathbf{x}}(\omega))$, as given in Eq. 7. Increasing N ensures that $\widetilde{\Phi}_{\mathbf{x}}$ is close to $\Phi_{\mathbf{x}}$. To estimate the PSDM, we thus rely on the spectrogram method and estimate $\widetilde{\Phi}_x(\omega))$ using finite n samples for $\mathbf{x}^r$.

4.1 Non-Asymptotic Estimation Error in Spectrogram Method

Let $\widehat{\Phi}_x(\omega) := \frac{1}{n}\sum_{r=1}^{n}\mathbf{x}^r(\omega)[\mathbf{x}^r(\omega)]^*$. Let the estimation error in estimating $\Phi_{\mathbf{x}}$ by $\widehat{\Phi}_x(\omega)$ be $Q := \widehat{\Phi}_x(\omega)-\Phi_x(\omega)$.

Applying the triangle inequality, $\|Q\|\leq\|Q_{approx}\|+\|Q_2\|$, where $Q_{approx} := \widetilde{\Phi}_x(\omega)-\Phi_x(\omega)$ and $Q_2 := \widehat{\Phi}_x(\omega)-\widetilde{\Phi}_x(\omega)$. Note that $\widetilde{\Phi}_x(\omega)$ is the covariance of the DFT of each trajectory. To bound the estimation error, we bound both Q_{approx} and Q_2.

The following Lemma shows that Q_{approx} is small if N (length of each trajectory) is large.

Lemma 4.1 (Lemma 5.1, Doddi et al. (2022)). *Consider an LDS given by (3) that satisfies Assumption 2.3. Let $Q_{approx}=\widetilde{\Phi}_x(\omega)-\Phi_x(\omega)$ where $\widetilde{\Phi}_x(\omega)$ is given in Eq. 7. Then $\|Q_{approx}\|<\varepsilon_1$ if $N>\frac{2C\rho^{-1}}{(1-\rho^{-1})^2\varepsilon_1}$.*

The next step is to characterize Q_2, the error in estimating $\widetilde{\Phi}_{\mathbf{x}}$ using $\widehat{\Phi}_x(\omega)$. Since $\mathbb{E}\left(\mathbf{x}^r(\omega)[\mathbf{x}^r(\omega)]^*\right)=\widetilde{\Phi}_x(\omega))$, $\widehat{\Phi}_x(\omega)$ is an unbiased estimator of $\widetilde{\Phi}_x(\omega)$. The following theorem provides a concentration bound on $\|Q_2\|$. The concentration bound is applicable under two sampling scenarios; the restart and record setting and continuous sampling setting, as shown in Fig. 2.

Theorem 4.2. *Suppose $\{\breve{x}^r(k)\}_{k=0}^{N-1}$, $1\leq r\leq n$ be the time series measurements obtained from an LDS governed by (3), satisfying Assumption 2.2. Then*

$$\mathbb{P}\left(\left\|\widehat{\Phi}_x(\omega)-\widetilde{\Phi}_x(\omega))\right\|\geq\epsilon\right)\leq\exp\left(-\frac{\epsilon^2 n}{128M^2}+6p\right),\forall\omega\in[-\pi,\pi]. \tag{11}$$

Proof. See Appendix C □

By combining Lemma 4.1 and Theorem 4.2, the following corollary is obtained, which gives a concentration bound on the estimation error, $\|Q\|$.

Corollary 4.3. *Consider an LDS governed by (3) that satisfies Assumptions 2.2 and 2.3. Let $\{\breve{x}^r(k)\}_{k=0}^{N-1}$, $1\leq r\leq n$ be the time series measurements obtained for the LDS. Suppose that $N>\frac{2C\rho^{-1}}{(1-\rho^{-1})^2\varepsilon_1}$, where $0<\varepsilon_1$. Let $0<\varepsilon_1,\varepsilon_2<\varepsilon$ be such that $\varepsilon_2=\varepsilon-\varepsilon_1$. Then $\forall\omega\in\Omega$,*

$$\mathbb{P}\left(\left\|\Phi_{xx}(\omega)-\widehat{\Phi}_{xx}(\omega))\right\|\geq\varepsilon\right)\leq\exp\left(-\frac{\varepsilon_2^2 n}{128M^2}+6p\right), \tag{12}$$

$$\mathbb{P}\left(\left\|\Phi_{CC}(\omega)-\widehat{\Phi}_{CC}(\omega))\right\|\geq\varepsilon\right)\leq\exp\left(-\frac{\varepsilon_2^2 n}{128M^2}+6q\right),\ \textit{and} \tag{13}$$

$$\mathbb{P}\left(\left\|\Phi_{ii}(\omega)-\widehat{\Phi}_{ii}(\omega))\right\|\geq\varepsilon\right)\leq\exp\left(-\frac{\varepsilon_2^2 n}{128M^2}+6\right). \tag{14}$$

4.2 Sample Complexity Bounds: Upper Bound

In the previous subsection, concentration bounds on the estimation errors in PSDM were obtained. Here, a concentration bound on the error in estimating f is obtained, which is used to obtain a concentration bound in reconstructing the DDAG, G. The following result provides a concentration bound on $|f - \widehat{f}|$.

Lemma 4.4. *Consider an LDS governed by (3) that satisfies Assumptions 2.2 and 2.3. Let $\{\breve{x}^r(k)\}_{k=0}^{N-1}$, $1 \le r \le n$ be the time series measurements obtained for the LDS. Suppose that $N > \frac{2C\rho^{-1}}{(1-\rho^{-1})^2\varepsilon_1}$, where $0 < \varepsilon_1$. Let $0 < \varepsilon_1, \varepsilon_2 < \varepsilon$ be such that $\varepsilon_2 = \varepsilon - \varepsilon_1$. Then there exists a $c_0 \in \mathbb{R}$ such that, for any $\omega \in \Omega$,*

$$\mathbb{P}\left(|f(i,C,\omega) - \widehat{f}(i,C,\omega) \ge \varepsilon\right) \le c_0 e^{\left(-\frac{\varepsilon_2^2 n}{10368 M^6} + 6(q+1)\right)},$$

where q is the maximum number of parents any node has in G.

Proof. See Appendix D $\square$

Based on Lemma 4.4, the following upper bound on the probability of error in estimating G can be obtained.

Theorem 4.5. *Suppose $q \le p/2$. Consider an LDS that belongs to $\mathcal{H}_{p,q}(\beta,\sigma,M)$ (Definition 2.4) that satisfies Assumptions 2.2 and 2.3. Let $\{\breve{x}^r(k)\}_{k=0}^{N-1}$, $1 \le r \le n$ be the time series measurements of the LDS and let $\widehat{G}$ be the DDAG reconstructed by Algorithm 1. Suppose that $N > \frac{2C\rho^{-1}}{(1-\rho^{-1})^2\varepsilon_1}$, where $0 < \varepsilon_1 < \Delta/4$. Let $0 < \varepsilon_1, \varepsilon_2 < \varepsilon < \Delta/4$ be such that $\varepsilon_2 = \varepsilon - \varepsilon_1$. Then $\mathbb{P}\left(G(\omega) \ne \widehat{G}(\omega)\right) \le \delta$ if*

$$n \gtrsim \frac{M^6\left(q\log(p/q) - \log\delta\right)}{\epsilon_2^2}.$$

Proof. Applying the bound in Lemma 4.4,

$$\begin{aligned}
\mathbb{P}\left(G(\omega) \ne \hat{G}(\omega)\right) &= \mathbb{P}\left(\bigcup_{k\in V, C\subseteq V\setminus\{k\}, |C|\le q} \left\{|f(i,C,\omega) - \widehat{f}(i,C,\omega)| > \epsilon\right\}\right)\\
&\overset{(a)}{\le} \sum_{k\in V, C\subseteq V\setminus\{k\}, |C|\le q} \mathbb{P}\left(\left\{|f(i,C,\omega) - \widehat{f}(i,C,\omega)| > \epsilon\right\}\right)\\
&\overset{(b)}{\le} p\left[\binom{(p-1)}{1} + \dots + \binom{(p-1)}{q}\right] c_0 e^{\left(-\frac{\epsilon^2 n}{10368M^6} + 6(q+1)\right)}\\
&\overset{(c)}{\lesssim} c_1 p \times q \times (p/q)^q e^{\left(-\frac{\epsilon^2 n}{10368M^6} + 6(q+1)\right)}\\
&\approx \exp\left(\log p + \log q + q\log(p/q) + \left(-\frac{\epsilon^2 n}{10368M^6} + 6(q+1)\right)\right)\\
&\approx \exp\left(\log p + \log q + q\log(p/q) + 6q - \frac{\epsilon^2 n}{M^6}\right)\\
&\lesssim \exp\left(q\log(p/q) - \frac{\epsilon^2 n}{M^6}\right) < \delta,
\end{aligned}$$

where (a) follows by union bound, (b) follows since $|V| = p$ and there are $\binom{p-1}{k}$ number of combinations with $|C| = k$. (c) follows by applying Stirling's approximation, $\binom{n}{k} \leq (ne/k)^k$. Thus, $n \gtrsim \frac{M^6(q\log(p/q) - \log\delta)}{\epsilon^2}$. By selecting the threshold γ in the algorithm appropriately, we can get a sample complexity $n \gtrsim \frac{M^6 q \log(p/q)}{\Delta^2}$. □

In the following section, a matching lower bound is derived.

5 Sample Complexity Bounds: Lower Bound

The lower bound for reconstructing DDAG is derived using information-theoretic techniques, in particular Fano's inequality, and by restricting the interested family of graphs to a finite set. Notice that, except for a couple of non-trivial facts in dynamic setup, this is a direct extension of the lower bound for the static case provided in Gao et al. (2022). The approach is to construct restricted ensembles of graphical models and then to lower bound the probability of error using Generalized Fano's inequality. Theorem 5.1 below provides the lower bound. For completeness, the proof is provided in the Appendix E.2.

Theorem 5.1. *Suppose $q \leq p/2$. If*

$$n \leq (1-2\delta)\max\left(\frac{\log p}{2\beta^2+\beta^4}, \frac{q\log(p/q)}{M^2-1}\right)$$

then

$$\inf_{\widehat{G}} \max_{H \in \mathcal{H}_{p,q}(\beta,\sigma,M)} \mathbb{P}\{(G(H) \neq \widehat{G})\} \geq \delta,$$

That is, if $n \leq (1-2\delta)\max\left(\frac{\log p}{2\beta^2+\beta^4}, \frac{q\log(p/q)}{M^2-1}\right)$, then any given estimator fails to reconstruct the DDAG with probability greater than δ. The lower-bound provides the fundamental limit on reconstructing the DDAG from finite samples.

Theorem 5.1 provides the lower bound $\Omega\left(\frac{\log p}{2\beta^2+\beta^4} \bigvee \frac{q\log(p/q)}{M^2-1}\right)$. Notice that the upper bound in Theorem 4.5 is $O(q\log\frac{p}{q})$. Thus we obtain a matching order bound when the lower bound is dominated by the second term.

Conclusion

In this article, we characterized the optimal sample complexity for structure identification with directions in linear dynamical networks. Inspired by the static setting, a metric and an algorithm were proposed based on the power spectral density matrix to exactly reconstruct the DAG. It is shown that the optimal sample complexity is $n = \Theta(q\log(p/q))$. For the upper bound characterization, we obtained a tight concentration bound for power spectral density matrix. An information-theoretic min-max lower bound also was provided for (sub) Gaussian linear dynamical sysytems. It was shown that the upper-bound is order optimal with respect to the lower bound.

Appendices

A Proof of Lemma 3.3

Let $C \subseteq V \setminus \{j\}$ be an ancestral set and let $D = nd(j) \setminus C$. Then,

$$\mathbf{x}_j(\omega) = H_{jC}(\omega)X_C(\omega) + H_{jD}(\omega)X_D(\omega) + e_j(\omega).$$

Applying $\Phi_{e_jC} = \Phi_{e_jD} = 0$, we obtain $\Phi_{jC}(\omega) = H_{jC}(\omega)\Phi_{CC}(\omega) + H_{jD}(\omega)\Phi_{DC}(\omega)$ and

$$\Phi_j(\omega) = H_{jC}(\omega)\Phi_C H_{jC}(\omega)^* + H_{jD}(\omega)\Phi_{DC}(\omega)H_{jC}^*(\omega) + H_{jC}(\omega)\Phi_{CD}(\omega)H_{jD}^*(\omega) + H_{jD}(\omega)\Phi_D(\omega)H_{jD}^*(\omega) + \Phi_{e_je_j}(\omega).$$

Then

$$f(j, C, \omega) = \Phi_j - \Phi_{jC}\Phi_C^{-1}\Phi_{Cj} = \Phi_{e_je_j} + H_{jD}(\Phi_D - \Phi_{DC}\Phi_{CC}^{-1}\Phi_{CD})H_{jD}^*.$$

Notice that when $Pa(j) \subseteq C$, $H_{jD} = 0$, and $f(j, C, \omega) = \Phi_{e_je_j}$, which shows the first part.

To prove the second part, suppose $Pa(j) \cap D \neq \emptyset$. We need to show that $H_{jD}(\Phi_D - \Phi_{DC}\Phi_{CC}^{-1}\Phi_{CD})\mathbf{H}_{jD}^* > 0$. Let $A = nd(j) = C \cup D$ and $B = desc(j) \cup \{j\}$. From Talukdar et al. (2018); Veedu et al. (2021),

$$\begin{aligned}
\Phi_{AA}^{-1} =& \mathbf{S} + \mathbf{L}, \text{ where} \\
&\mathbf{S} = (\mathbf{I}_A - \mathbf{H}_{AA}^*)\Phi_{e_A}^{-1}(\mathbf{I}_A - \mathbf{H}_{AA}), \\
&\mathbf{L} = \mathbf{H}_{BA}\Phi_{e_B}^{-1}\mathbf{H}_{BA} - \Psi^*\Lambda^{-1}\Psi, \\
&\Psi = \mathbf{H}_{AB}^*\Phi_{e_A}^{-1}(\mathbf{I} - \mathbf{H}_{AA}) + (\mathbf{I} - \mathbf{H}_{BB}^*)\Phi_{e_B}^{-1}\mathbf{H}_{BA}, \text{ and} \\
&\Lambda = \mathbf{H}_{AB}^*\Phi_{e_A}^{-1}\mathbf{H}_{AB} + (\mathbf{I} - \mathbf{H}_{BB}^*)\Phi_{e_B}^{-1}(\mathbf{I} - \mathbf{H}_{BB}).
\end{aligned}$$

Notice that since B is the set of descendants of j, $\mathbf{H}_{AB} = 0$, as cycles can be formed otherwise. Then, $\mathbf{L} = 0$ and $\Phi_{AA}^{-1} = (\mathbf{I}_A - \mathbf{H}_{AA}^*)\Phi_{e_A}^{-1}(\mathbf{I}_A - \mathbf{H}_{AA})$.

$$\Phi_{AA}^{-1} = \begin{bmatrix} \Phi_{DD} & \Phi_{DC} \\ \Phi_{CD} & \Phi_{CC} \end{bmatrix}^{-1} = \begin{bmatrix} K_{DD} & K_{DC} \\ K_{CD} & K_{CC} \end{bmatrix} = \frac{1}{\sigma}\left(\mathbf{I} - \mathbf{H}_{AA}^*\right)\left(\mathbf{I} - \mathbf{H}_{AA}\right)$$

By Scur's complement, $(\Phi_D - \Phi_{DC}\Phi_{CC}^{-1}\Phi_{CD})^{-1} = K_{DD} = \frac{1}{\sigma}(\mathbf{I}_D - \mathbf{H}_{DD}^* - \mathbf{H}_{DD} + (\mathbf{H}_{AA}^*\mathbf{H}_{AA})_{D\times D})$. Moreover,

$$\mathbf{H}_{AA} = \begin{bmatrix} \mathbf{H}_{DD} & \mathbf{H}_{DC} \\ \mathbf{H}_{CD} & \mathbf{H}_{CC} \end{bmatrix} \text{ and } (\mathbf{H}_{AA}^*\mathbf{H}_{AA})_{D\times D} = \mathbf{H}_{DD}^*\mathbf{H}_{DD} + \mathbf{H}_{CD}^*\mathbf{H}_{CD}.$$

Since C is ancestral, $\mathbf{H}_{CD} = 0$ and

$$K_{DD} = \frac{1}{\sigma}(\mathbf{I}_D - \mathbf{H}_{DD})^*(\mathbf{I}_D - \mathbf{H}_{DD}).$$

Since G is a DAG, the rows and columns of $\mathbf{H}$ can be rearranged to obtain a lower triangular matrix with zeros on the diagonal. Thus eigenvalues of $(\mathbf{I}_D - \mathbf{H}_{DD})$ and its inverse are all ones. Hence minimum eigenvalue of K_{DD}^{-1} is greater than σ. Applying Rayleigh Ritz theorem on $\mathbf{H}_{jD}K_{DD}^{-1}\mathbf{H}_{jD}^*$, we have

$$\mathbf{H}_{jD}(\Phi_D - \Phi_{DC}\Phi_{CC}^{-1}\Phi_{CD})\mathbf{H}_{jD}^* = \mathbf{H}_{jD}K_{DD}^{-1}\mathbf{H}_{jD}^* \geq \sigma|D|\beta^2 \tag{15}$$

which is strictly greater than zero if D is non-empty. $\square$

B Proof of Lemma 3.8

The proof is done in two steps. First, we show that $\mathcal{S}$ in Algorithm 1 is a topological ordering. Then, we show that step (4) in Algorithm 1 can identify the parents of every node in G. The first step is shown via induction. Since $|\widehat{f}(i, C, \omega) - f(i, C, \omega)| < \Delta/4$ for empty set, $|\Phi_{ii} - \widehat{\Phi}_{ii}| < \Delta/4$ for every i. Recall from Lemma 3.3 that $\Phi_{ii} - \Phi_{jj} > \Delta$ if i is a source node and j is a non-source node. Then, $\widehat{\Phi}_{jj} \geq \Phi_{jj} - \Delta/4 \geq \Phi_{ii} + 3\Delta/4 \geq \widehat{\Phi}_{ii} + \Delta/2$. Thus, $i \in \arg\min_{1 \leq k \leq p} \widehat{\Phi}_{kk}$ if and only if $i \in \arg\min_{1 \leq k \leq p} \Phi_{kk}$ and thus $\mathcal{S}_1$ is always a source node.

For the induction step, assume that $\mathcal{S}_1, \ldots, \mathcal{S}_n$ forms a correct topologically ordered set w.r.t. G. Let $C \subseteq \mathcal{S}(1:n)$. If $Pa(i) \subseteq C$ and $Pa(j) \nsubseteq C$, then by applying Lemma 3.3, $\widehat{f}(j, C, \omega) > f(j, C, \omega) - \Delta/4 \geq \sigma + 3\Delta/4 = f(i, C, \omega) + 3\Delta/4 \geq \widehat{f}(i, C, \omega) + \Delta/2$. Thus, $i \in \arg\min_{k \in V \setminus \mathcal{S}} \widehat{f}(k, C, \omega)$ if and only if $i \in \arg\min_{k \in V \setminus \mathcal{S}} f(k, C, \omega)$ and thus $(\mathcal{S}, \mathcal{S}_{n+1})$ forms a topological order w.r.t. G, by Lemma 3.6.

To prove the second step, let $C \subseteq \mathcal{S}(1:i)$. Since $\mathcal{S}(1:i)$ is a valid topological ordering, $Pa(i) \subseteq \mathcal{S}(1:i-1)$. Let $k \in Pa(i)$ and let $D = C \setminus \{k\}$. Then, as shown in Corollary 3.4 $f(i, C, \omega) - f(i, D, \omega) \geq \Delta$, and

$$\begin{aligned}
\Delta \leq |f(i, C, \omega) - f(i, D, \omega)| &\leq |f(i, C, \omega) - \widehat{f}(i, C, \omega)| + |\widehat{f}(i, C, \omega) - \widehat{f}(i, D, \omega)| \\
&\quad + |\widehat{f}(i, D, \omega) - f(i, D, \omega)| \\
&< \Delta/4 + \Delta/4 + |\widehat{f}(i, C, \omega) - \widehat{f}(i, D, \omega)| \\
\implies |\widehat{f}(i, C, \omega) - \widehat{f}(i, D, \omega)| > \Delta/2.
\end{aligned}$$

Suppose $k \notin Pa(i)$ but $k \in \mathcal{S}(1:i)$. Then, for $D = C \setminus \{k\}$, $f(i, C, \omega) - f(i, D, \omega) = 0$. Repeating the same series of inequalities above by exchanging f and $\widehat{f}$, we obtain $|\widehat{f}(i, C, \omega) - \widehat{f}(i, D, \omega)| < \Delta/2$.

Thus, from the set $\mathcal{S}$, for every node $\mathcal{S}_i$, one can check nodes $\mathcal{S}_1, \ldots, \mathcal{S}_{i-1}$ and verify if the difference of including and excluding the node is greater than $\Delta/2$. If the difference is greater than $\Delta/2$ for some k, then k is a parent of i, and if not, then the node is not a parent of i. That is, let $C_i = \{\mathcal{S}_1, \ldots, \mathcal{S}_{i-1}\}$, $i > 1$, and let

$$\widehat{P}_i := \left\{ j \in C_i \,\middle|\, |\widehat{f}(\mathcal{S}_i, C_i, \omega) - \widehat{f}(\mathcal{S}_i, C_i \setminus \{j\}, \omega)| > \Delta/2 \right\}.$$

Then, $Pa(i) = \widehat{P}_i$. □

C Proof of Theorem 4.2

By the variational form of spectral norm Horn and Johnson (2012),

$$\|Q\| = \sup_{v \in \mathbb{C}^p, \|v\|=1} |v^* Q v|,$$

where the max is taken over a p-dimensional unit complex sphere, $\mathbb{S}^p := \{v \in \mathbb{C}^p : \|v\|_2 = 1\}$. The first step here is to reduce supremum to finite maximization using finite covers of a unit ball, which is done using a δ cover. A δ-cover of a set $\mathcal{A}$ is a set $v^1, \ldots, v^m$ such that for every $v \in \mathcal{A}$, there exists an $i \in 1, \ldots, m$ such that $\|v^i - v\|_2 \leq \delta$. The following Lemma is obtained by extending example 5.8 in Wainwright (2019) to the complex field.

Lemma C.1. *Let $v^1, \ldots, v^m$ be a δ-covering of the unit sphere $\mathbb{S}^p$. Then there exists such a covering with $m \leq (1 + 2/\delta)^{2p}$ vectors.*

Proof. The proof follows by extending (5.9) in Wainwright (2019), to the complex field. $\square$

Let $v \in \mathbb{S}^p$ and let v^j be such that $v = v^j + \Delta$, where $\|\Delta\| \leq \delta$. Then, $v^*Qv = (v^j)^*Qv^j + 2\Re\{\Delta^*Qv^j\} + \Delta^*Q\Delta$. Applying triangle inequality,

$$\begin{aligned} |v^*Qv| &\leq |(v^j)^*Qv^j| + 2\|\Delta\|\|Q\|\|v^j\| + |\Delta\|^2\|Q\| \\ &\leq |(v^j)^*Qv^j| + 2\delta\|Q\| + \delta^2\|Q\| \\ &\leq |(v^j)^*Qv^j| + \frac{1}{2}\|Q\| \text{ for } \delta \leq 0.22474. \end{aligned}$$

Thus,

$$\begin{aligned} &\|Q\| = \max_{v \in \mathbb{S}^p} |v^*Qv| \leq \max_{j=1,\ldots,m} |(v^j)^*Qv^j| + \frac{1}{2}\|Q\| \text{ and} \\ &\|Q\| \leq 2 \max_{j=1,\ldots,m} |(v^j)^*Qv^j| \end{aligned}$$

Next, we find an upper bound for $\mathbb{E}\left[e^{\lambda\|Q\|}\right]$, which is treated with Chernoff-type bounding technique to obtain the desired result.

$$\begin{aligned} \mathbb{E}\left[e^{\lambda\|Q\|}\right] &\leq \mathbb{E}\left[\exp\left(2\lambda \max_{j=1,\ldots,m} |(v^j)^*Qv^j|\right)\right] \\ &\leq \sum_{j=1}^{m} \mathbb{E}\left[e^{2\lambda(v^j)^*Qv^j}\right] + \mathbb{E}\left[e^{-2\lambda(v^j)^*Qv^j}\right] \end{aligned} \tag{16}$$

Next, we complete the proof for the restart and record sampling and the continuous sampling separately.

C.1 Restart and Record Sampling

Under the restart and record sampling settings, for any given $\omega \in \Omega$, $\{\mathbf{x}^r(\omega)\}_{r=1}^n$ is i.i.d. Thus

$$\begin{aligned} \mathbb{E}\left[\exp\left(t(v^j)^*Qv^j\right)\right] &= \mathbb{E}\left[\exp\left(t(v^j)^*(\widehat{\Phi}_x(\omega) - \widetilde{\Phi}_x(\omega)))v^j\right)\right] \\ &= \mathbb{E}\left[\exp\left(\frac{t}{n}\sum_{r=1}^{n}(v^j)^*\mathbf{x}^r(\omega)[\mathbf{x}^r(\omega)]^*v^j - (v^j)^*\widetilde{\Phi}_x(\omega)v^j\right)\right] \\ &= \prod_{r=1}^{n}\mathbb{E}\left[\exp\left(\frac{t}{n}(v^j)^*\mathbf{x}^r(\omega)[\mathbf{x}^r(\omega)]^*v^j - (v^j)^*\widetilde{\Phi}_x(\omega)v^j\right)\right] \\ &= \left(\mathbb{E}\left[\exp\left(\frac{t}{n}(v^j)^*\mathbf{x}^1(\omega)[\mathbf{x}^1(\omega)]^*v^j - (v^j)^*\widetilde{\Phi}_x(\omega)v^j\right)\right]\right)^n \\ &= \left(\mathbb{E}\left[\exp\left(\frac{t}{n}|v^*\mathbf{x}^r(\omega)|^2 - v^*\widetilde{\Phi}_x(\omega)v\right)\right]\right)^n \end{aligned}$$

Let $\varepsilon \in \{-1, +1\}$ be a Rademacher variable independent of $\mathbf{x}^r$. It can be shown that Proposition 4.11 in Wainwright (2019) will hold for complex numbers also. Then

$$\mathbb{E}_{\mathbf{x}^r(\omega)}\left[\exp\left(\frac{t}{n}|v^*\mathbf{x}^r(\omega)|^2 - v^*\widetilde{\Phi}_x(\omega)v\right)\right] \leq \mathbb{E}_{\mathbf{x}^r(\omega),\varepsilon}\left[\exp\left(\frac{2t\varepsilon}{n}|v^*\mathbf{x}^r(\omega)|^2\right)\right] \tag{17}$$

$$= \sum_{k=0}^{\infty}\frac{(2t/n)^{2k}}{2k!}\mathbb{E}\left[|v^*\mathbf{x}^r(\omega)|^{4k}\right] \tag{18}$$

Recall that $\widetilde{\Phi}_{\mathbf{x}}$ is a positive definite matrix and $v^*\mathbf{x}^r \sim N(\mathbf{0}, \eta)$, where $\eta = v^*\widetilde{\Phi}_{\mathbf{x}}v \leq \lambda_{max}(\widetilde{\Phi}_{\mathbf{x}}) \leq M$. The even moments of $y \sim N(0, \eta)$ is given by $\mathbb{E}\{y^{2k}\} = \eta^{2k}(2k-1)!! = \frac{(2k)!}{2^k k!}\eta^{2k}$. Then

$$\mathbb{E}\left[|v^*\mathbf{x}^r(\omega)|^{4k}\right] \leq \frac{(4k)!}{2^{2k}(2k)!}M^{2k}.$$

Therefore using the inequality $(4k)! \leq 2^{2k}[(2k)!]^2$,

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}^r(\omega)}\left[\exp\left(\frac{t}{n}|v^*\mathbf{x}^r(\omega)|^2 - v^*\widetilde{\Phi}_x(\omega)v\right)\right] &\leq 1 + \sum_{k=1}^{\infty}\frac{(2t/n)^{2k}}{2k!}\frac{(4k)!}{2^{2k}(2k)!}M^{2k} \\
&\leq 1 + \sum_{k=1}^{\infty}\frac{(2t/n)^{2k}}{2k!}\frac{2^{2k}[(2k)!]^2}{2^{2k}(2k)!}M^{2k} \\
&= 1 + \sum_{k=1}^{\infty}\left(\frac{2Mt}{n}\right)^{2k} = \frac{1}{1-\left(\frac{2Mt}{n}\right)^2} \\
&\leq \exp\left(\frac{8M^2t^2}{n^2}\right)
\end{aligned}$$

whenever $\frac{2Mt}{n} < 3/4$, where the final inequality follows by applying $1 - x \geq e^{-2x}$ for $x \in [0, 3/4]$ (to be precise 0.77). Thus,

$$\mathbb{E}\left[\exp\left(t(v^j)^*Qv^j\right)\right] \leq \exp\left(\frac{8M^2t^2}{n}\right),\ \forall |t| \leq \frac{3n}{8M}.$$

Applying Lemma C.1 and the bound $2m \leq 2(1 + 2/0.22474)^{2p} \leq 2e^{4.6p} \leq e^{5p+0.693} \leq e^{6p}$,

$$\begin{aligned}
\text{From (16)},\quad \mathbb{E}\left[e^{\lambda\|Q\|}\right] &\leq \mathbb{E}\left[\exp\left(2\lambda\max_{j=1,\dots,m}|(v^j)^*Qv^j|\right)\right] \\
&\leq \sum_{j=1}^{m}\mathbb{E}\left[e^{2\lambda(v^j)^*Qv^j}\right] + \mathbb{E}\left[e^{-2\lambda(v^j)^*Qv^j}\right] \\
&\leq 2m\exp\left(\frac{32M^2\lambda^2}{n}\right) \\
&\leq \exp\left(\frac{32M^2\lambda^2}{n} + 6p\right),\ \forall |\lambda| \leq \frac{3n}{16M}.
\end{aligned}$$

Applying Chernoff-type bounding approach,

$$\mathbb{P}\left(\|Q\| \geq t\right) \leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda\|Q\|}\right] \leq \exp\left(-\lambda t + \frac{32M^2\lambda^2}{n} + 6p\right),\ \forall |\lambda| \leq \frac{3n}{16M}.$$

The tightest bound is given by $g^*(t) := \inf_{|\lambda|\le\frac{3n}{16M}} \left\{-\lambda t + \frac{32M^2\lambda^2}{n} + 6p\right\}$, where the objective is convex. Taking derivative w.r.t. λ and equating to zero, $\lambda^* = \frac{tn}{64M^2}$ and $g^* = -\frac{t^2 n}{64M^2} + \frac{32M^2}{n}\frac{t^2n^2}{64^2M^4} + 6p = 6p - \frac{t^2 n}{128M^2}$, if t is such that $t \le 12M$, which is reasonable as we can always pick $M \ge 1$.

Thus, $\mathbb{P}\left(\|\mathbf{Q}\| \ge t\right) \le \exp\left(-\frac{t^2 n}{128M^2} + 6p\right)$ The theorem statement follows. $\square$

C.2 Continuous Sampling

In the continuous sampling setting, the samples $\check{\mathbf{x}}(0), \dots, \check{\mathbf{x}}(N-1), \check{\mathbf{x}}(N), \dots, \check{\mathbf{x}}(2N-1), \dots, \check{\mathbf{x}}((n-1)N), \dots, \check{\mathbf{x}}(nN-1)$ are sampled continuously and are correlated with each other. Thus, $\mathbf{x}^r(\omega)$ and $\mathbf{x}^s(\omega)$, $r \neq s$, $1 \le r, s \le n$, can be correlated, in contrast to the restart and record (RR) setting, where the $\mathbf{x}^r(\omega)$ and $\mathbf{x}^s(\omega)$, $r \neq s$ are i.i.d. For any given $\omega \in \Omega$, let $\mathbf{x}(\omega) := [[\mathbf{x}^1(\omega)]^T, [\mathbf{x}^2(\omega)]^T, \dots, [\mathbf{x}^n(\omega)]^T]^T \in \mathbb{C}^{pn\times 1}$ be the vectorized form of $\{\mathbf{x}^r(\omega)\}_{r=1}^n$ and let $\mathcal{C}(\omega) := \mathbb{E}\{\mathbf{x}(\omega)\mathbf{x}^*(\omega)\}$ be the covariance matrix of $\mathbf{x}(\omega)$. Under the RR setting, $\mathcal{C}(\omega) \in \mathbb{C}^{pn\times nn}$ will be a block-diagonal matrix (of block size $p\times p$), whereas in the continuous sampling, the non-block-diagonal entries of $\mathcal{C}(\omega)$ can be non-zero. However, the vector $\mathbf{x}(\omega)$, with correlated entries, can be written as a linear transformation of i.i.d. vector $\mathbf{w} \in \mathbb{C}^{pn\times 1}$ with unit variance, i.e. $\mathbf{x}(\omega) = \mathcal{C}^{1/2}(\omega)\mathbf{w}$, where $\mathcal{C}^{1/2}$ is the square-root of $\mathcal{C}$. As shown in Appendix E.1, when $\{\check{\mathbf{e}}(k)\}_{k=1}^n$ in the linear time-invariant model (1) are Gaussian, $\mathbf{x}^r(\omega)$ and thus $\mathbf{x}(\omega)$ are Gaussian distributed. In this case, a candidate is $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{pn})$. It can be verified that $\mathbb{E}\{\mathbf{x}(\omega)\mathbf{x}^*(\omega)\} = \mathcal{C}^{1/2}(\omega)\mathbb{E}\{\mathbf{w}\mathbf{w}^*\}\mathcal{C}^{1/2}(\omega) = \mathcal{C}(\omega)$. Notice that the covariance matrix $\mathcal{C}(\omega)$ is a block matrix, defined as

$$\mathcal{C} = \begin{bmatrix} \mathcal{C}^{11} & \mathcal{C}^{12} & \dots & \mathcal{C}^{1n} \\ \mathcal{C}^{21} & \mathcal{C}^{22} & \dots & \mathcal{C}^{2n} \\ \vdots & & & \\ \mathcal{C}^{n1} & \mathcal{C}^{n2} & \dots & \mathcal{C}^{nn} \end{bmatrix}, \text{where } \mathcal{C}^{rs}(\omega) \in \mathbb{C}^{p\times p},\ 1 \le r, s \le n,$$

where the entries of $\mathcal{C}^{rs}(\omega)$ is given by $\mathbb{E}\left\{\mathbf{x}^r(\omega)[\mathbf{x}^s(\omega)]^*\right\}$. Recall that

$$\mathbf{x}^r(\omega) = \frac{1}{\sqrt{N}} \sum_{\ell=0}^{N-1} \check{x}((r-1)N + \ell)e^{-i\omega\ell}.$$

Let $\mathbf{I}_r = \left[\mathbf{0}|\dots|\mathbf{0}|\mathbf{I}_{p\times p}|\dots|\mathbf{0}\right] \in \mathbb{R}^{p\times np}$ be such that r^{th} block is identity matrix. Then $\mathbf{x}^r(\omega) = \mathbf{I}_r\mathbf{x}(\omega)$. The estimated PSDM is then given by

$$\widehat{\Phi}_{\mathbf{x}}(\omega) = \frac{1}{n}\sum_{r=1}^{n} \mathbf{x}^r(\omega)[\mathbf{x}^r(\omega)]^* = \frac{1}{n}\sum_{r=1}^{n} \mathbf{I}_r\mathbf{x}(\omega)\mathbf{x}^*(\omega)\mathbf{I}_r^*.$$

Substituting $\mathbf{x}(\omega) = \mathcal{C}^{1/2}(\omega)\mathbf{w}$, and letting $\mathbf{B}(\omega) := (\mathcal{C}^{1/2})^*(\omega)\sum_{r=1}^n \mathbf{I}_r^* u u^* \mathbf{I}_r \mathcal{C}^{1/2}(\omega)$

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(tu^*\mathbf{Q}u\right)\right] &= \mathbb{E}\left[\exp\left(\frac{t}{n}\sum_{r=1}^{n}u^*\mathbf{x}^r(\omega)[\mathbf{x}^r(\omega)]^*u - u^*\widetilde{\Phi}_x(\omega)u\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{n}\sum_{r=1}^{n}\left[\mathbf{x}^*(\omega)\mathbf{I}_r^*uu^*\mathbf{I}_r\mathbf{x}(\omega) - \mathbb{E}\left\{\mathbf{x}^*(\omega)\mathbf{I}_r^*uu^*\mathbf{I}_r\mathbf{x}(\omega)\right\}\right]\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{n}\sum_{r=1}^{n}\left[\mathbf{w}^*(\mathcal{C}^{1/2})^*(\omega)\mathbf{I}_r^*uu^*\mathbf{I}_r\mathcal{C}^{1/2}(\omega)\mathbf{w} - \mathbb{E}\left\{\mathbf{w}^*(\mathcal{C}^{1/2})^*(\omega)\mathbf{I}_r^*uu^*\mathbf{I}_r\mathcal{C}^{1/2}(\omega)\mathbf{w}\right\}\right]\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{n}\left[\mathbf{w}^*(\mathcal{C}^{1/2})^*(\omega)\sum_{r=1}^{n}(\mathbf{I}_r^*uu^*\mathbf{I}_r)\mathcal{C}^{1/2}(\omega)\mathbf{w} - \mathbb{E}\left\{\mathbf{w}^*(\mathcal{C}^{1/2})^*(\omega)\sum_{r=1}^{n}(\mathbf{I}_r^*uu^*\mathbf{I}_r)\mathcal{C}^{1/2}(\omega)\mathbf{w}\right\}\right]\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{n}\left[\mathbf{w}^*\mathbf{B}(\omega)\mathbf{w} - \mathbb{E}\left\{\mathbf{w}^*\mathbf{B}(\omega)\mathbf{w}\right\}\right]\right)\right]
\end{aligned}
$$

Notice that $\mathbf{I}_r^*u = \left[0,\ldots,0,u^T,,\ldots,0\right]^T$ is a column vector,

$$
\mathbf{I}_r^*uu^*\mathbf{I}_r = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ & \cdots & & \cdots & \mathbf{0} \\ \vdots & \cdots & \underbrace{uu^*}_{(r,r)^{th}\ block} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \end{bmatrix} \text{ and } \sum_{r=1}^{n}\mathbf{I}_r^*uu^*\mathbf{I}_r = \begin{bmatrix} uu^* & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & uu^* & \ldots & \mathbf{0} \\ \vdots & \ddots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & uu^* \end{bmatrix},
$$

i.e., $rank(\mathbf{I}_r^*uu^*\mathbf{I}_r) = 1$ and $rank(\mathbf{B}(\omega)) \leq n$. Let $\mathbf{B}(\omega) = \mathbf{U}(\omega)\Lambda(\omega)\mathbf{U}^*(\omega)$ be the eigen value decomposition of $\mathbf{B}(\omega)$, where $\Lambda = diag(\lambda_1,\ldots,\lambda_n)$. Consequently, omitting ω from the notations,

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(tu^*\mathbf{Q}u\right)\right] &= \mathbb{E}\left[\exp\left(\frac{t}{n}\left[\mathbf{w}^*\mathbf{B}\mathbf{w} - \mathbb{E}\left\{\mathbf{w}^*\mathbf{B}\mathbf{w}\right\}\right]\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{n}\left[\mathbf{w}^*\mathbf{U}\Lambda\mathbf{U}^*\mathbf{w} - \mathbb{E}\left\{\mathbf{w}^*\mathbf{U}\Lambda\mathbf{U}^*\mathbf{w}\right\}\right]\right)\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\exp\left(\frac{t}{n}\left[\mathbf{w}^*\Lambda\mathbf{w} - \mathbb{E}\left\{\mathbf{w}^*\Lambda\mathbf{w}\right\}\right]\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{n}\sum_{i=1}^{n}\lambda_i\left[w_i^2 - \mathbb{E}\left\{w_i^2\right\}\right]\right)\right] \\
&= \prod_{i=1}^{n}\mathbb{E}\left[\exp\left(\frac{t\lambda_i}{n}\left[w_i^2 - \mathbb{E}\left\{w_i^2\right\}\right]\right)\right],
\end{aligned}
$$

where (a) follows because $\mathbf{w}$ is invariant under unitary transformations Cui et al. (2019). Let $\varepsilon \in \{+1,-1\}$ be a uniform random variable independent of $\mathbf{w}$. Similar to (17), we can now apply the Rademacher random variable trick.

$$\begin{aligned}
\mathbb{E}_{w_i}\left[\exp\left(\lambda\left[w_i^2-\mathbb{E}\left\{w_i^2\right\}\right]\right)\right] &\leq \mathbb{E}_{w_i,\varepsilon}\left[\exp\left(2\lambda\varepsilon w_i^2\right)\right] \\
&= \sum_{k=0}^{\infty}\frac{(2\lambda)^{2k}}{(2k)!}\mathbb{E}\left[w_i^{4k}\right] \\
&\leq \sum_{k=0}^{\infty}\frac{(2\lambda)^{2k}}{2k!}\frac{(4k)!}{(2k)!2^{2k}} \\
&\leq \sum_{k=0}^{\infty}(2\lambda)^{2k} = \frac{1}{1-4\lambda^2} \leq \exp(8\lambda^2),
\end{aligned}$$

for every $|\lambda| < 3/8$. Thus, (with the substitution $\lambda = t\lambda_i/n$ and the upperbound $\lambda_i \leq \|\mathbf{B}\|$)

$$\begin{aligned}
\mathbb{E}\left[\exp\left(tu^*\mathbf{Q}u\right)\right] &= \prod_{i=1}^{n}\mathbb{E}\left[\exp\left(\frac{t\lambda_i}{n}\left[w_i^2-\mathbb{E}\left\{w_i^2\right\}\right]\right)\right] \\
&\leq \prod_{i=1}^{n}\exp\left(\frac{8t^2\lambda_i^2}{n^2}\right) \\
&= \exp\left(\frac{8t^2}{n^2}\sum_{i=1}^{n}\lambda_i^2\right) \\
&\leq \exp\left(\frac{8t^2}{n}\|\mathcal{C}\|^2\right),\ \forall\ |t| \leq \frac{3n}{8\|\mathcal{C}\|},
\end{aligned}$$

where we have used $\|B\| \leq \|\mathcal{C}\|$ in the final equality. Now, combining this with the $\delta-$cover argument,

$$\begin{aligned}
\mathbb{E}\left[e^{t\|Q\|}\right] &\leq \sum_{j=1}^{m}\mathbb{E}\left[e^{2t(v^j)^*Qv^j}\right]+\mathbb{E}\left[e^{-2t(v^j)^*Qv^j}\right] \\
&\leq 2m\exp\left(\frac{32\|\mathcal{C}\|^2t^2}{n}\right) \\
&\leq \exp\left(\frac{32\|\mathcal{C}\|^2t^2}{n}+6p\right),\ \forall|t| \leq \frac{3n}{16\|\mathcal{C}\|}.
\end{aligned}$$

Finally, applying Chernoff bound, $\mathbb{P}\left(\|\mathbf{Q}\| \geq t\right) \leq \exp\left(-\frac{t^2n}{128\|\mathcal{C}\|^2}+6p\right)$.

C.2.1 Tight upper bound for $\|\mathcal{C}\|$

An explicit expression for $\mathcal{C}$ is given as follows:

$$\begin{aligned}
\mathcal{C}^{rs}(\omega) := \mathbb{E}\left\{\mathbf{x}^r(\omega)[\mathbf{x}^s(\omega)]^*\right\} &= \frac{1}{N}\sum_{\ell=0}^{N-1}\sum_{k=0}^{N-1}\mathbb{E}\left\{\breve{x}((r-1)N+\ell)[\breve{x}((s-1)N+k)]^T\right\}e^{-i\omega(\ell-k)} \\
&= \frac{1}{N}\sum_{\ell=0}^{N-1}\sum_{k=0}^{N-1}R_{\breve{\mathbf{x}}}((r-s)N+\ell-k)e^{-i\omega(\ell-k)} \\
&= \frac{1}{N}\sum_{\tau=-N+1}^{N-1}(N-|\tau|)R_{\breve{\mathbf{x}}}((r-s)N+\tau)e^{-i\omega\tau} \\
&= \sum_{\tau=-N+1}^{N-1}\left(1-\frac{|\tau|}{N}\right)R_{\breve{\mathbf{x}}}((r-s)N+\tau)e^{-i\omega\tau} \\
&= \sum_{\tau=-N+1}^{N-1}\left(1-\frac{|\tau|}{N}\right)e^{-i\omega\tau}R_{\breve{\mathbf{x}}}((r-s)N+\tau).
\end{aligned}$$

Let $\alpha_\tau = e^{-i\omega\tau}\left(1-\frac{|\tau|}{N}\right)$. Then

$$\mathcal{C} = \sum_{\tau=-N+1}^{N-1}\alpha_\tau\begin{bmatrix} R_{\breve{\mathbf{x}}}(\tau) & R_{\breve{\mathbf{x}}}(-N+\tau) & \dots & R_{\breve{\mathbf{x}}}((1-n)N+\tau) \\ R_{\breve{\mathbf{x}}}(N+\tau) & R_{\breve{\mathbf{x}}}(\tau) & \dots & R_{\breve{\mathbf{x}}}((2-n)N+\tau) \\ \vdots & \vdots & \ddots & \\ R_{\breve{\mathbf{x}}}((n-1)N+\tau) & R_{\breve{\mathbf{x}}}((n-2)N+\tau) & \dots & R_{\breve{\mathbf{x}}}(\tau)\end{bmatrix}.$$

Notice that $\breve{g}(\tau) := 1-|\tau|/N$ is a triangle function, and the Fourier transform of $\breve{g}(\tau)$, $g(\omega)$ has the property that $|g(\omega)| \le 1$. Then for any $u \in \mathbb{C}^{np}$ such that $\|u\|_2 \le 1$,

$$\begin{aligned}
u^*\mathcal{C}u &= \sum_{i,j=1}^{n}[u^i]^*\sum_{\tau=-N+1}^{N-1}\alpha_\tau R_{\breve{\mathbf{x}}}((i-j)N+\tau)u^j \\
&\overset{(a)}{\le} \mathcal{F}^\tau\{u\}\mathcal{F}^\tau\{R_{\mathbf{x}}(tN+\tau)\}\mathcal{F}^\tau\{u\} \\
&\overset{(b)}{\le} \|\Phi_{\mathbf{x}}(\omega)\| \le M,
\end{aligned}$$

where (a) follows by taking Fourier transform with respect to τ ($\mathcal{F}^\tau$ denotes Fourier transform with respect to the variable τ) and (b) since $\|\mathcal{F}^\tau\{u\}\|_2 \le 1$. Thus, $\mathbb{P}(\|\mathbf{Q}\| \ge t) \le \exp\left(-\frac{t^2n}{128M^2}+6p\right)$, similar to the restart and record case.

D Proof of Lemma 4.4

Notice that $\|A\mathbf{x}\|_2 \le \|A\|\|\mathbf{x}\|_2$ for every matrix A and vector $\mathbf{x}$. Applying this identity with $\mathbf{x} = [1, 0\dots, 0]$, $\|\Phi_{Ci}\|_2 \le \|\Phi_{AA}\|$, where $A = [k,\ C]$. Then, applying CBS inequality for complex vectors, $|x^*Ay| \leq \|x\|_2\|Ay\|_2 \leq \|x\|_2\|A\|\|y\|_2$, the error can be upper bounded as

$$\begin{aligned}
|f(i,C,\omega) - \widehat{f}(i,C,\omega)| &= |(\Phi_{ii} - \Phi_{iC}\Phi_{CC}^{-1}\Phi_{Ci}) - (\widehat{\Phi}_{ii} - \widehat{\Phi}_{iC}\widehat{\Phi}_{CC}^{-1}\widehat{\Phi}_{Ci})| \\
&= |(\Phi_{ii} - \widehat{\Phi}_{ii}) + (\widehat{\Phi}_{iC}\widehat{\Phi}_{CC}^{-1}\widehat{\Phi}_{Ci} - \Phi_{iC}\Phi_{CC}^{-1}\Phi_{Ci})| \\
&\leq |\Phi_{ii} - \widehat{\Phi}_{ii}| + |\widehat{\Phi}_{iC}(\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1})\widehat{\Phi}_{Ci})| \\
&\qquad + |(\widehat{\Phi}_{iC} - \Phi_{iC})\Phi_{CC}^{-1}\widehat{\Phi}_{Ci})| + |\Phi_{iC}\Phi_{CC}^{-1}(\widehat{\Phi}_{Ci} - \Phi_{Ci})| \\
&\leq |\Phi_{ii} - \widehat{\Phi}_{ii}| + \|\widehat{\Phi}_{iC}\|_2\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci}\|_2 \\
&\qquad + \|\widehat{\Phi}_{iC} - \Phi_{iC}\|_2\|\Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci}\|_2 + \|\Phi_{iC}\|_2\|\Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 \\
&\leq |\Phi_{ii} - \widehat{\Phi}_{ii}| + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci}\|_2^2 \\
&\qquad + M\|\widehat{\Phi}_{iC} - \Phi_{iC}\|_2\|\widehat{\Phi}_{Ci}\|_2 + M^2\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 \\
&\leq |\Phi_{ii} - \widehat{\Phi}_{ii}| + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\Phi_{Ci}\|_2^2 \\
&\qquad + M\|\widehat{\Phi}_{iC} - \Phi_{iC}\|_2\left(\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 + \|\Phi_{Ci}\|_2\right) + M^2\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 \\
&\leq |\Phi_{ii} - \widehat{\Phi}_{ii}| + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|M^2 \\
&\qquad + M\|\widehat{\Phi}_{iC} - \Phi_{iC}\|_2\left(\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 + M\right) + M^2\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 \\
&= |\Phi_{ii} - \widehat{\Phi}_{ii}| + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|M^2 \\
&\qquad + M\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + M^2\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 + M^2\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 \\
&\leq |\Phi_{ii} - \widehat{\Phi}_{ii}| + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + M^2\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\| \\
&\qquad + M\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + 2M^2\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2.
\end{aligned}$$

The above expression can be bounded above if we can bound the three errors, $\|\widehat{\Phi}_{ii} - \Phi_{ii}\| = \epsilon_i$, $\|\widehat{\Phi}_{AA} - \Phi_{AA}\| = \epsilon_A$, and $\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\| = \epsilon_{Cinv}$. Simplifying the above expression,

$$\begin{aligned}
|f(i,C,\omega) - \widehat{f}(i,C,\omega)| &\leq |\Phi_{ii} - \widehat{\Phi}_{ii}| + \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + M^2\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\| \\
&\qquad + M\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2^2 + 2M^2\|\widehat{\Phi}_{Ci} - \Phi_{Ci}\|_2 \\
&\leq \epsilon_i + \epsilon_{Cinv}\epsilon_A^2 + 2M^2\epsilon_{Cinv} + M\epsilon_A^2 + 2M^2\epsilon_A \\
&\leq \epsilon_i + \epsilon_{Cinv}(\epsilon_A^2 + 2M^2) + 3M^2\epsilon_A \\
&\leq \epsilon_i + 3M^2\epsilon_{Cinv} + 3M^2\epsilon_A
\end{aligned}$$

Pick $\epsilon_i = 3M^2\epsilon_{Cinv} = 3M^2\epsilon_A = \epsilon/3$. Then $|f(i,C,\omega) - \widehat{f}(i,C,\omega)| < \epsilon$. From Section 5.8 in Horn and Johnson (2012),

$$\begin{aligned}
\|\Phi_{CC} - \widehat{\Phi}_{CC}\| &\leq \|\Phi_{CC}\|\|\Phi_{CC}^{-1}\|^{-1}\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|\frac{M^2}{1 - M^2\frac{\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|}{\|\Phi_{CC}^{-1}\|}}, \\
&\leq \frac{M^4\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|}{1 - M\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\|} \leq \epsilon \implies \|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\| \leq \frac{\epsilon}{M^4 + M\epsilon} \leq \frac{\epsilon}{M^4}.
\end{aligned}$$

Therefore, to guarantee that $\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\| < \epsilon$, it is sufficient to guarantee that $\|\widehat{\Phi}_{CC} - \Phi_{CC}\| < \epsilon$ since $M \geq 1$. Rewriting Corollary 4.3,

$$\mathbb{P}\left(|\Phi_{ii} - \widehat{\Phi}_{ii}| \geq \epsilon\right) \leq e^{-\frac{\epsilon^2 n}{128M^2}+6}, \tag{19}$$

$$\mathbb{P}\left(\|\Phi_{AA} - \widehat{\Phi}_{AA}\| > \epsilon\right) \leq e^{-\frac{\epsilon^2 n}{128M^2}+6(q+1)}, \text{ and} \tag{20}$$

$$\mathbb{P}\left(\|\Phi_{CC} - \widehat{\Phi}_{CC}\| > \epsilon\right) \leq e^{-\frac{\epsilon^2 n}{128M^2}+6q}, \quad \forall \epsilon \geq 0. \tag{21}$$

Plugging these bounds in the above expressions gives the concentration upper bound

$$\begin{aligned}
\mathbb{P}\left(|f(i,C,\omega) - \widehat{f}(i,C,\omega) \geq \epsilon\right) &\leq \mathbb{P}\left(|\Phi_{ii} - \widehat{\Phi}_{ii}| \geq \epsilon/3\right) + \mathbb{P}\left(\|\widehat{\Phi}_{CC}^{-1} - \Phi_{CC}^{-1}\| \geq \epsilon/(9M^2)\right) \\
&\quad + \mathbb{P}\left(\|\widehat{\Phi}_{AA} - \Phi_{AA}\| \geq \epsilon/(9M^2)\right) \\
&\leq \mathbb{P}\left(|\Phi_{ii} - \widehat{\Phi}_{ii}| \geq \epsilon/3\right) + \mathbb{P}\left(\|\widehat{\Phi}_{CC} - \Phi_{CC}\| \geq \epsilon M^2/(9)\right) \\
&\quad + \mathbb{P}\left(\|\widehat{\Phi}_{AA} - \Phi_{AA}\| \geq \epsilon/(9M^2)\right) \\
&\leq e^{\left(-\frac{\epsilon^2 n}{1152M^2}+6\right)} + e^{\left(-\frac{\epsilon^2 M^2 n}{10368}+6q\right)} + e^{\left(-\frac{\epsilon^2 n}{10368M^6}+6(q+1)\right)} \\
&\leq c_0 e^{\left(-\frac{\epsilon^2 n}{10368M^6}+6(q+1)\right)}.
\end{aligned}$$

□

E Lower bound: Proof of Theorem 5.1

E.1 Density function in frequency domain

Consider an autoregressive (AR) model

$$\breve{\mathbf{x}}(k) = \sum_{l=0}^{T_1} \breve{\mathbf{H}}(l)\breve{\mathbf{x}}(k-l) + \breve{\mathbf{e}}(k), \ \forall k \in \mathbb{Z}, \tag{22}$$

where $\breve{\mathbf{e}}(k) = [\breve{\mathbf{e}}_1(k), \ldots, \breve{\mathbf{e}}_n(k)]^T$, $\widetilde{\mathbf{x}}(k) = [\breve{\mathbf{x}}_1(k), \ldots, \breve{\mathbf{x}}_n(k)]^T$, $\breve{\mathbf{e}}_i$ is a stochastic process such that the Fourier transform $\mathbf{e}(\omega)$ exists. To understand the problem, let us assume $\breve{\mathbf{e}}(k) \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ for $k = 0, \ldots, T_2$, i.i.d. and zero otherwise. i.e., $\breve{\mathbf{x}}(k)$ is non zero only for $k =$

$0, \ldots, T_1 + T_2$. Then

$$
\begin{aligned}
\mathbf{e}(\omega) &= \sum_{k=0}^{T_2} \breve{\mathbf{e}}(k) e^{-j\omega k} = \sum_{k=0}^{T_2} \breve{\mathbf{e}}(k) \cos(\omega k) - j \sum_{k=0}^{T_2} \breve{\mathbf{e}}(k) \sin(\omega k). \\
\implies cov(e(\omega)) &= \mathbb{E}\{e(\omega) e^*(\omega)\} \\
&= \mathbb{E}\left\{ \left(\sum_{k_1=0}^{T_2} \breve{\mathbf{e}}(k_1) e^{-j\omega k_1} \right) \left(\sum_{k_2=0}^{T_2} \breve{\mathbf{e}}(k_2) e^{-j\omega k_2} \right)^* \right\} \\
&= \mathbb{E}\left\{ \left(\sum_{k_1=0}^{T_2} \sum_{k_2=0}^{T_2} \breve{\mathbf{e}}(k_1) [\breve{\mathbf{e}}(k_2)]^* e^{-j\omega k_1} e^{j\omega k_2} \right) \right\} \\
&= \mathbb{E}\left\{ \left(\sum_{k=0}^{T_2} \breve{\mathbf{e}}(k) [\breve{\mathbf{e}}(k)]^* \right) \right\} \\
&= \sum_{k=0}^{T_2} \mathbb{E}\left\{ \breve{\mathbf{e}}(k) [\breve{\mathbf{e}}(k)]^* \right\} \\
&= \sigma^2 \mathbf{I} \sum_{k=0}^{T_2} 1.
\end{aligned}
$$

Thus, $\mathbf{e}(\omega) \sim N\left(0, (T_2+1)\sigma^2 \mathbf{I}\right)$.

Now, consider the general LDS (3), with $\breve{e}(k) \sim \mathcal{N}(0, \sigma_k \mathbf{I}),\ k \in \mathbb{Z}$. Then as shown above, $\mathbf{e}(\omega) \sim N(0, \Phi_{\mathbf{e}}(\omega))$, where $\Phi_{\mathbf{e}}(\omega) = \sum_{k \in \mathbb{Z}} \sigma_k \mathbf{I}$. It follows that $\mathbf{x}(\omega) = (I - \mathbf{H}(\omega))^{-1} \mathbf{e}(\omega)$ and $\mathbf{x}(\omega) \sim N(0, \Phi_{\mathbf{x}})$, where $\Phi_{\mathbf{x}}(\omega) = (I - \mathbf{H}(\omega))^{-1} \Phi_{\mathbf{e}}(\omega) ((I - \mathbf{H}(\omega))^{-1})^*$. Thus, the density function $f^{(\omega)}$ is Gaussian, where the covariance matrix is the power spectral density matrix, a function of ω.

E.2 Proof of lowerbound

The proof is based on Generalized Fano's inequality.

Lemma E.1 (Generalized Fano's method)**.** *Gao et al. (2022) Consider a class of observational distribution $\mathcal{F}$ and a subclass $\mathcal{F}' = \{F_1, \ldots, F_r\} \subseteq \mathcal{F}$ with r distributions and the estimators $\widehat{\theta}$. Then*

$$
\inf_{\widehat{\theta}} \max_{F \in \mathcal{F}} \mathbb{E}\{\mathbb{I}(\theta(F) \neq \widehat{\theta})\} \geq \frac{\alpha_r}{2} \left(1 - \frac{n\beta_r + \log 2}{\log r} \right),
$$

where n is the number of samples,

$$
\begin{aligned}
\alpha_r &:= \max_{k \neq j} \mathbb{I}(\theta(F_k) \neq \theta(F_j)), \\
\beta_r &:= \max_{k \neq j} KL(F_k || F_j),
\end{aligned}
$$

with $KL(P||Q) := \mathbb{E}_P\left[\log \frac{P}{Q}\right] = \mathbb{E}_P[\log P] - \mathbb{E}_P[\log Q]$ being the KL divergence.

Corollary E.2. *Consider subclass of graphs $\mathcal{G}' = \{G_1, \ldots, G_r\} \subseteq \mathcal{G}_{p,q}$, and let $\mathbf{H}^i$ be the distribution corresponding to a distinct $G_i \in \mathcal{G}'$. Then, any estimator $\widehat{G} := \bigcup_{\omega \in \Omega} \widehat{G}(\omega)$ of G_i*

is δ *unreliable,*

$$\inf_{\widehat{G}} \sup_{G_i \in \mathcal{G}'} \mathbb{P}\{(G(\mathbf{H}^i) \neq \widehat{G}\} \geq \delta,$$

if

$$n \leq \frac{(1-2\delta)\log r - \log 2}{\beta_r}$$

Therefore, building a lower bound involves finding a subclass that has 1) small β_r and 2) large r. First, we can find an upper bound of r by upper bounding the number of directed graphs possible with at most q parents. Overall, p^2 number of possible positions are there and at most pq many edges. The number of possible ways to choose k edges is $\binom{p^2}{k}$. Thus $r = \sum_{k=1}^{pq} \binom{p^2}{k} \leq pq\binom{p^2}{pq} \lesssim pq(p/q)^{pq}$. Therefore, $\log r \lesssim \log(pq) + pq\log(p/q) \lesssim pq\log(p/q)$ Similarly, it can be shown that $\log r \gtrsim pq\log(p/q)$ Gao et al. (2022).

Consider the LDS (3), with $\breve{e}(k) \sim N(0, \sigma_k \mathbf{I})$, i.i.d. across time. Then as shown in Appendix E.1, $\mathbf{e}(\omega) \sim N\left(0, \Phi_{\mathbf{e}}(\omega)\right)$ and $\mathbf{x}(\omega) \sim N(0, \Phi_{\mathbf{x}})$, where $\Phi_{\mathbf{e}}(\omega) = \sum_{k\in\mathbb{Z}} \sigma_k \mathbf{I}$ and $\Phi_{\mathbf{x}}(\omega) = (I - \mathbf{H}(\omega))^{-1}\Phi_{\mathbf{e}}(\omega)((I-\mathbf{H}(\omega))^{-1})^*$.

Ensemble A: Consider all possible DAGs in $\mathcal{G}'$ with i.i.d. Gaussian exogenous distribution such that $\Phi_{\mathbf{e}}(\omega)$ exists. For the two distributions, F_k and F_j such that for any $\omega \in \Omega$, $F_k(\omega) \sim \mathcal{N}(\mathbf{0}, \Phi^{(k)}(\omega))$ and $F_j(\omega) \sim \mathcal{N}(\mathbf{0}, \Phi^{(j)}(\omega))$,

$$\begin{aligned}
KL(F_k(\omega)||F_j(\omega)) &= \frac{1}{2}\left(\mathbb{E}_{F_j(\omega)}[\mathbf{x}^*(\omega)[\Phi^{(k)}(\omega)]^{-1}\mathbf{x}(\omega)] - \mathbb{E}_{F_j}[\mathbf{x}^*(\omega)[\Phi^{(j)}(\omega)]^{-1}\mathbf{x}(\omega)]\right) \\
&= \frac{1}{2}\left(\mathbb{E}_{F_j(\omega)}[tr(\mathbf{x}(\omega)[\Phi^{(k)}(\omega)]^{-1}\mathbf{x}^*(\omega))] - p\right) \\
&= \frac{1}{2}\left(tr([\Phi^{(k)}(\omega)]^{-1}\Phi^{(j)}(\omega)) - p\right) \\
&\leq \frac{1}{2}\left(\sqrt{\|[\Phi^{(k)}(\omega)]^{-1}\|_F^2\|\Phi^{(j)}(\omega)\|_F^2} - p\right) \\
&\leq \frac{1}{2}\left(pM^2 - p\right) \leq (M^2-1)p
\end{aligned}$$

Therefore one of the lower bounds is

$$\inf_{\widehat{G}} \sup_{G \in \mathcal{G}'} \mathbb{P}\{(G \neq \widehat{G})\} \geq \delta,$$

if

$$\begin{aligned}
n &\leq \frac{(1-2\delta)pq\log(p/q)) - \log 2}{(M^2-1)p} \\
&\lesssim \frac{q\log(p/q)}{M^2-1}.
\end{aligned}$$

Ensemble B: Here, we consider graphs in $\mathcal{H}_{p,q}(\beta,\sigma,M)$ (recall the definition from 3.1.1) with a single edge $u \to v$ with $H_{vu}(\omega) = \beta$ for every $\omega \in \Omega$, i.e. constant matrix. For LDS with i.i.d. Gaussian noise with PSD matrix $\Phi_{\mathbf{x}}$ that satisfies this condition, $\mathbf{H}$ is such that $H_{vu} \neq 0$ and $H_{ij} = 0$ otherwise. Here, the total number of graphs, $r = 2\binom{p}{2} = (p^2-p) \approx p^2$

Notice that $[\Phi_{\mathbf{x}}^{-1}]_{ij} = (\mathbf{I}_{ij} - H_{ij} - H_{ji}^* + \sum_{k=1}^{p} H_{ki}^* H_{kj})/\sigma$. Thus (ignoring ω), $[\Phi_{\mathbf{x}}^{-1}]_{uv} = \frac{-H_{vu}^* + \sum_{k=1}^{p} H_{ku}^* H_{kv}}{\sigma} = \frac{-H_{vu}^*}{\sigma} = -\beta/\sigma$ and $[\Phi_{\mathbf{x}}^{-1}]_{ij} = 0$ if $i, j \neq u, v$. Then,

$$
\begin{aligned}
\mathbf{x}^* \Phi_{\mathbf{x}}^{-1} \mathbf{x} &= \sum_{ij} x_i^* [\Phi_{\mathbf{x}}^{-1}]_{ij} x_j \\
&= \frac{1}{\sigma} \left[\sum_{i=1}^{p} |x_i|^2 (1 + \sum_k |H_{ki}|^2) + x_u^* [\Phi_{\mathbf{x}}^{-1}]_{uv} x_v + x_v^* [\Phi_{\mathbf{x}}^{-1}]_{vu} x_u \right] \\
&= \frac{1}{\sigma} \left[\sum_{i \neq u} |x_i|^2 + (1 + |H_{vu}|^2) |x_u|^2 - 2\beta \Re\{x_u^* x_v\} \right] \\
&= \frac{1}{\sigma} \left[\sum_{i} |x_i|^2 + \beta^2 |x_u|^2 - 2\beta \Re\{x_u^* x_v\} \right] \\
&= \frac{1}{\sigma} \left[\sum_{i \neq v} |x_i|^2 + |x_v - \beta x_u|^2 \right]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
KL(F^{uv} || F^{jk}) &= \mathbb{E}_{F^{uv}} \left[\log F^{uv} - \log F^{jk} \right] \\
&= \frac{1}{2\sigma} \mathbb{E}_{F^{uv}} \left[|x_v|^2 + |x_v - \beta x_u|^2 - |x_k|^2 - |x_k - \beta x_j|^2 \right] \\
&= \frac{1}{2\sigma} \left[\beta^2 \sigma + \mathbb{E}_{F^{uv}} \left(\beta^2 |x_j|^2 - 2\beta \Re\{x_j^* x_k\} \right) \right]
\end{aligned}
$$

Considering all the cases of (u, v) vs (j, k) it can be shown that $KL(F^{uv} || F^{jk}) \leq \beta^2 + \beta^4/2$ Gao et al. (2022). Thus, $n \gtrsim \frac{\log p}{\beta^2 + \beta^4/2}$ gives the second lower bound. The lower bound follows by combining ensembles A and B. □

References

S. Basu and G. Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, pages 1535–1567, 2015.

J. M. Bower and D. Beeman. *The book of GENESIS: exploring realistic neural models with the GEneral NEural SImulation System*. Springer Science & Business Media, 2012.

W. Chen, M. Drton, and Y. S. Wang. On causal discovery with an equal-variance assumption. *Biometrika*, 106(4):973–980, 2019.

W. Cui, X. Zhang, and Y. Liu. Covariance matrix estimation from linearly-correlated gaussian samples. *IEEE Transactions on Signal Processing*, 67(8):2187–2195, 2019.

R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51:157–172, 2000.

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*, pages 9982–9997. PMLR, 2022.

M. K. S. Faradonbeh, A. Tewari, and G. Michailidis. Finite time identification in unstable linear systems. *Automatica*, 96:342–353, 2018.

M. Fiecas, C. Leng, W. Liu, and Y. Yu. Spectral analysis of high-dimensional time series. *Electronic Journal of Statistics*, 13(2):4079–4101, 2019.

N. Friedman and Z. Yakhini. On the sample complexity of learning bayesian networks. In *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pages 274–282, 1996.

M. Gao, W. M. Tai, and B. Aragam. Optimal estimation of gaussian dag models. In *International Conference on Artificial Intelligence and Statistics*, pages 8738–8757. PMLR, 2022.

M. Ghil, M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou. Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40(1):3–1–3–41, 2002. doi: 10.1029/2000RG000092. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000RG000092.

A. Ghoshal and J. Honorio. Information-theoretic limits of bayesian network structure learning. In *Artificial Intelligence and Statistics*, pages 767–775. PMLR, 2017a.

A. Ghoshal and J. Honorio. Learning identifiable gaussian bayesian networks in polynomial time and sample complexity. *Advances in Neural Information Processing Systems*, 30, 2017b.

A. Ghoshal and J. Honorio. Learning linear structural equation models in polynomial time and sample complexity. In A. Storkey and F. Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1466–1475. PMLR, 09–11 Apr 2018. URL https://proceedings.mlr.press/v84/ghoshal18a.html.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, USA, 2nd edition, 2012. ISBN 0521548233.

Y. S. Kim, S. T. Rachev, M. L. Bianchi, I. Mitov, and F. J. Fabozzi. Time series analysis for financial market meltdowns. *Journal of Banking & Finance*, 35(8):1879–1891, 2011.

A. Lamperski. Non-asymptotic pointwise and worst-case bounds for classical spectrum estimators. *arXiv preprint arXiv:2303.11908*, 2023.

P.-L. Loh and M. J. Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Advances in neural information processing systems*, 24, 2011.

D. Materassi and G. Innocenti. Unveiling the connectivity structure of financial networks via high-frequency analysis. *Physica A: Statistical Mechanics and its Applications*, 388(18): 3866–3878, 2009. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2009.06.003. URL https://www.sciencedirect.com/science/article/pii/S0378437109004324.

D. Materassi and M. V. Salapaka. On the problem of reconstructing an unknown topology via locality properties of the wiener filter. *IEEE Transactions on Automatic Control*, 57 (7):1765–1777, July 2012. ISSN 0018-9286. doi: 10.1109/TAC.2012.2183170.

D. Materassi and M. V. Salapaka. Reconstruction of directed acyclic networks of dynamical systems. In *2013 American Control Conference*, pages 4687–4692. IEEE, 2013.

G. Park. Identifiability of additive noise models using conditional variances. *The Journal of Machine Learning Research*, 21(1):2896–2929, 2020.

G. Park and G. Raskutti. Learning quadratic variance function (qvf) dag models via overdispersion scoring (ods). *J. Mach. Learn. Res.*, 18(224):1–44, 2017.

J. Peters and P. Bühlmann. Identifiability of gaussian structural equation models with equal error variances. *Biometrika*, 101(1):219–228, 2014.

C. J. Quinn, N. Kiyavash, and T. P. Coleman. Directed information graphs. *IEEE Transactions on Information Theory*, 61(12):6887–6909, 2015. doi: 10.1109/TIT.2015.2478440.

S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, and M. Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pages 439–473. PMLR, 2018.

J. Songsiri and L. Vandenberghe. Topology selection in graphical models of autoregressive processes. *The Journal of Machine Learning Research*, 11:2671–2705, 2010.

P. Stoica, R. L. Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

S. Talukdar, D. Deka, M. Chertkov, and M. Salapaka. Topology learning of radial dynamical systems with latent nodes. In *2018 Annual American Control Conference (ACC)*, pages 1096–1101. IEEE, 2018.

S. Talukdar, D. Deka, H. Doddi, D. Materassi, M. Chertkov, and M. V. Salapaka. Physics informed topology learning in networks of linear dynamical systems. *Automatica*, 112: 108705, 2020. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2019.108705.

A. Tank, N. J. Foti, and E. B. Fox. Bayesian structure learning for stationary time series. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 872–881, 2015.

M. S. Veedu, D. Harish, and M. V. Salapaka. Topology learning of linear dynamical systems with latent nodes using matrix decomposition. *IEEE Transactions on Automatic Control*, Early Access:1–1, 2021. doi: 10.1109/TAC.2021.3124979.

M. J. Wainwright. *Random matrices and covariance estimation*, page 159–193. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. doi: 10.1017/9781108627771.006.

X. Yang, T. Xu, P. Jia, H. Xia, L. Guo, L. Zhang, and K. Ye. Transportation, germs, culture: a dynamic graph model of covid-19 outbreak. *Quantitative Biology*, 8:238–244, 2020.

D. Zhang and W. B. Wu. Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics*, 49(1):233–254, 2021.

O. Zuk, S. Margel, and E. Domany. On the number of samples needed to learn the correct structure of a bayesian network. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 560–567, 2006.