Topology Inference for Networked Dynamical Systems: A Causality and Correlation Perspective

Yushan Li and Jianping He

***Abstract*—Networked dynamical systems (NDSs) have gained considerable attention in recent years, where the networked agents cooperate to accomplish the common task through the interaction topology. In this paper, we focus on the topology inference problem of NDSs. Different from traditional methods, we aim to infer the internal interaction topology from the perspective of node causality and correlation, covering both directed/undirected topology structures and the asymptotic/marginal stabilities of NDSs. Specifically, we propose a causality-based method that takes the noise characteristic into account and asymptotically approaches the real interaction topology structure, with only single round observations over the dynamical process. When the observation number is small, we further design a correlation-based modification to effectively alleviate the influence of noises. We demonstrate the close relation between the proposed method and the traditional Granger estimator and ordinary least square (OLS) estimator in terms of observation rounds and horizon. We further prove the equivalence conditions of the proposed method with the two estimators from the system stability and noise characteristic. The proposed causality-correlation combined method enjoys analogous asymptotic inference performance with Granger estimator of multiple observation rounds, and outperforms OLS estimator in single observation round, verified by extensive simulations.**

I. Introduction

In the last decades, networked dynamical systems (NDSs) have received considerable attention in many realms of science and engineering. The systems are characterized by the locality of information exchange between individual agents (usually described by a topology graph) [1] and the cooperative capability to solve a common task (e.g., optimization) [2]. Despite the popularity of NDSs, the internal interaction structure within the systems is not directly available in many domains, e.g., social networks [3], brain connectivity patterns [4] and multi-robot formation [5]. It arises as a critical problem that how to use the observational data of the system to infer the topology, for better understanding the systems and implementing tasks.

Topology inference, which aims to infer the interaction relationships and their evolution, is essentially a challenging inverse modeling problem. The source of the difficulty comes from multiple aspects [6]. First, direct access to the interaction relation between agents is usually unavailable and sometimes even the model assumptions or prior knowledge are absent. Second, indirect observations over the network dynamics are typically noisy and confounded by overlapping relationships (e.g., temporal and spatial). Last but not least, the inference performance is largely dependent on the observation scale and network dynamics characteristics. Many useful applications benefit from topology inference, for instance, tracing the information flow over a social network [7], group testing and identification of defective items [8], or anomaly detection in communications networks [9].

A large body of researches has been developed to tackle the problem using different methods. For example, [10]–[12] utilize Granger estimator to capture the casual relationships between agents. Spectral decomposition from sample matrix is also a popular tool to estimate the topology [13]–[15]. Kernel-based methods are widely used to identify nonlinear dynamic topology [16]–[18]. Despite the prominent contributions of the pioneering works, there still remain some notable issues. First, in many methods, the observation data is not associated with an actual network, and the topology is inferred to interpret the latent regularity contained in the data [19]–[21]. These methods may not work well for NDSs, especially considering the existence of the actual system structure and dynamics. Second, many effective algorithms are designed for symmetric topology structure. The symmetry brings nice tractability in the inference procedure, but the results are hard to be generalized to account for the directed dependency between agents, e.g., using correlation statistics [22]. Third, the methods are mostly presented in an asymptotic or expected way, requiring a large amount of observation rounds or horizons, (like [12], [13]). When the observation horizon is small, the inference performance can largely degenerate.

Based on the above observations, this paper is motivated to infer the interaction topology of NDSs from the perspective of node causality and correlation, covering both directed and undirected structures. Different from traditional system identification methods, where the main goal is to identify the system's Markov parameters with input/output data (like [23], [24]), our work is oriented from a non-ordinary-least-square (OLS) modeling manner and only noisy observations over the NDS are available. We further explore the relationships between the proposed method and the Granger and OLS estimators. The challenges mainly lie in two aspects. First, only the direct observations over the NDS are available, and the process and observation noises are all unknown. Every adjacent two observations are highly correlated, raising more uncertainty for the actual causality and correlation of system nodes. Second, the network topology is deeply coupled with the system stability. The existence of process noises may cover the real state evolution, making it further hard to split the coupled influences of a system update and noise. The main

The authors are with the Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China. E-mail address: {yushan_li, jphe}@sjtu.edu.cn.

Manuscript 1412 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.

contributions are summarized as follows.

- We investigate the topology inference problem of NDSs, where both the asymptotically and marginally stable interaction matrices are involved. From the perspective of node causality and correlation, we combine the two factors and propose a new topology inference method, applying to both directed and undirected topologies.
- By exploiting the Granger causality characteristic of multiple observation rounds, we propose a causality-based method only using observations from a single round. Considering the inference performance degradation in the small observation horizon, we utilize the correlation measurement as a modification for the original causality-based method, so as to alleviate the influence brought by the small-scale process/observation noises.
- Taking the Granger and OLS estimators as references, we deduce the close relation between their asymptotically solving manners, and prove their equivalent conditions in terms of the system stability, the observation scale and statistical characteristic of process/observation noises. We observe that our proposed method outperforms the OLS estimator, illustrated by extensive simulations.

The remainder of this paper is organized as follows. Section II gives basic preliminaries and describes the problem of interest. The inference method and performance analysis are presented in Section III. Simulation results are shown in Section IV, followed by the concluding remarks and further research issues in Section V. Throughout this paper, the set variable, vector and matrix are expressed in Euclid, lowercase, and uppercase font. Let $\mathbf{0}$ be all-zero matrix in compatible dimensions. Unless otherwise noted, $\|\cdot\|$ and $\|\cdot\|_F$ represent the spectral and Frobenius norm of a matrix, respectively.

II. Preliminaries and Problem Formulation

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph that models the networked system, where $\mathcal{V} = \{1, \cdots, n\}$ is the finite set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of interaction edges. An edge $(i, j) \in \mathcal{E}$ indicates that i will use information from j. The adjacency matrix $A = [a_{ij}]_{N\times N}$ of $\mathcal{G}$ is defined such that $a_{ij} > 0$ if (i, j) exists, and $a_{ij} = 0$ otherwise. Denote $\mathcal{N}_i = \{j \in \mathcal{V} : a_{ij} > 0\}$ as the in-neighbor set of i, and $d_i = |\mathcal{N}_i|$ as its in-degree. A directed path is a sequence of nodes $\{r_1, r_2, \cdots, r_j\}$ such that $(r_{i+1}, r_i) \in \mathcal{E}$, $i = 1, 2, \cdots, j-1$. A directed graph has a (directed) spanning tree if there exists at least a node having a directed path to all other nodes. $\mathcal{G}$ must have a spanning tree to guarantee that at least one node's information can reach all other nodes.

A. System Model

Consider the following networked dynamical model

$$\begin{aligned} x_t &= W x_{t-1} + \theta_{t-1}, \\ y_t &= x_t + \upsilon_t, \end{aligned} \tag{1}$$

where x_t and y_t represents the system state and corresponding observation at time t ($t = 1, 2, \cdots, T$), $W \in \mathbb{R}^{n\times n}$ is the unknown interaction matrix related to the adjacent matrix A, and θ_t and υ represent the i.i.d process and observation noises, satisfying $\theta_t \sim N(0, \sigma_\theta^2 I)$ and $\upsilon_t \sim N(0, \sigma_\upsilon^2 I)$.

Denote the spectral radius of W by $\rho(W)$. We present asymptotically stable matrix class $\mathcal{S}_a$ and the (strictly) marginally stable matrix $\mathcal{S}_m$ as follows:

$$\begin{aligned} \mathcal{S}_a =& \{Z \in \mathbb{R}^{n\times n}, \rho(Z) < 1\}, \\ \mathcal{S}_m =& \{Z \in \mathbb{R}^{n\times n}, \rho(Z) = 1 \text{ and the geometric} \\ & \text{multiplicity of eigenvalue 1 equals to one}\}. \end{aligned} \tag{2}$$

In terms of the setup of W, some useful and popular choices are the Laplacian and the Metropolis rules, which are defined as follows [25]. For $i \neq j$,

$$w_{ij} = \begin{cases} \gamma a_{ij}/\max\{d_i, i \in \mathcal{V}\}, & \text{by Laplacian rule,} \\ a_{ij}/\max\{d_i, d_j\}, & \text{by Metropolis rule,} \end{cases} \tag{3}$$

where the auxiliary parameter γ satisfies $0 < \gamma \leq 1$. For both rules, the self-weights are given by

$$w_{ii} = 1 - \sum_{j\neq i} w_{ij}. \tag{4}$$

Note that if W is specified by either one of the two rules, then $W \in \mathcal{S}_m$. A typical matrix in $\mathcal{S}_a$ can be directly obtained via multiplying (3) and (4) by a factor $0 < \alpha < 1$, which is common in adaptive diffusion networks [26]. Considering different stabilities, it holds that

$$\lim_{t\to\infty} W^t = \begin{cases} \mathbf{0}, & \text{if } W \in \mathcal{S}_a \\ W^\infty, & \text{if } W \in \mathcal{S}_m, \end{cases} \tag{5}$$

where $\mathbf{0}$ represents all-zero matrix in compatible dimensions and $\|W^\infty\| < \infty$. In a recursive form, (1) is rewritten as

$$x_t = W^t x_0 + \sum_{m=1}^{t} W^{m-1} \theta_{t-m}. \tag{6}$$

Then, when $t \to \infty$, the observation satisfies

$$\lim_{t\to\infty} y_t = \begin{cases} \displaystyle\sum_{m=1}^{\infty} W^{m-1}\theta_{t-m} + \upsilon_\infty, \text{if } W \in \mathcal{S}_a, \\ W^\infty x_0 + \displaystyle\sum_{m=1}^{\infty} W^{m-1}\theta_{t-m} + \upsilon_\infty, \text{if } W \in \mathcal{S}_m, \end{cases} \tag{7}$$

B. Node Causality and Correlation

After sufficient exchange of information, the system states become highly correlated. A common metric quantifying the adjacency a_{ij} is the Pearson correlation coefficient, given by

$$\rho_{ij}^0 = \sum_{t=1}^{T} \frac{(x_t^i - \bar{x}^i)}{\rho_i^0} \frac{(x_t^j - \bar{x}^j)}{\rho_j^0}, \tag{8}$$

where $\rho_i^0 = \sqrt{\sum_{t=1}^{T} (x_t^i - \bar{x}^i)^2}$ is the sample standard deviations of $\{x_i(t)\}$ and $\bar{x}^i = \sum_{t=0}^{T} x_t^i / T$, $\forall i \in \mathcal{V}$. Note that ρ_{ij}^0 directly describes the (linear) correlation between two nodes. Given a probability of false alarms, one can test whether a link exists between two nodes. However, the directionality (i.e., causality) of the links cannot be revealed by the correlation coefficient due to its symmetry. To capture the node causality, Granger causality is widely adopted [10].

Define the autocorrelation and one-lag autocorrelation matrices of x_t as

$$R_0^x(t)=\mathbb{E}\left[x_t x_t^{\mathsf{T}}\right], R_1^x(t)=\mathbb{E}\left[x_t x_{t-1}^{\mathsf{T}}\right]. \tag{9}$$

Since $R_1^x(t)=\mathbb{E}[Wx_{t-1}x_{t-1}^{\mathsf{T}}+\theta_t x_{t-1}^{\mathsf{T}}]=WR_0^x(t-1)$, one obtains the Granger estimator [12] as

- **Granger estimator**:

$$\hat{W}_g=R_1^x(t)(R_0^x(t-1))^{-1}. \tag{10}$$

Note that (10) can be interpreted as finding the coefficients $\{w_{ij}\}$ that provide the best linear prediction of x_t given the past state x_{t-1}.

C. Problem of Interest

Given the observation sequence $\{y_t\}_{t=1}^T$ adhering to (1), the goal of this paper is to infer the unknown interaction matrix W. Mathematically, one aims to find an associated mapping

$$\phi_{\mathcal{M}}:\{y_t\}_{t=1}^T\rightarrow W. \tag{11}$$

In this paper, we focus on finding the mapping mechanisms $\phi_{\mathcal{M}}$ from the perspective of node causality and correlation. Since only a single observation round of system (1) is available where the observations are corrupted by unknown noises and the input noises are also unknown, the key challenge is how to deal with the observation noises and formulate an interpretable inference model. Borrowing the idea of node causality and correlation, we are able to design an efficient inference method to improve the inference performance. Specifically, we further explore the relationships between the proposed method and the Granger estimator and the OLS estimator.

III. Causality and Correlation based Method

In this section, we first utilize the causality described by the Granger estimator (10), design a causality-based inference method for a single observation round setting, and provide their equivalent conditions. Then, considering when the observation size is small, we use node correlation to revise the proposed method. Finally, we elaborate on its close relationships with the Granger and OLS estimators.

To ease notation, we organize the input/noise data as

$$\begin{aligned}
&X_T^-=[x_0,x_2,\cdots,x_{T-1}],\ X_T^+=[x_1,x_2,\cdots,x_T],\\
&Y_T^-=[y_0,y_2,\cdots,y_{T-1}],\ Y_T^+=[y_1,y_2,\cdots,y_T],\\
&\Theta_T=[\theta_0,\theta_1\cdots,\ \theta_{T-1}],\ \Upsilon_T=[\upsilon_1,\upsilon_2\cdots,\upsilon_T].
\end{aligned} \tag{12}$$

Then, the whole evolution process is compactly written as

$$X_T^+=WX_T^-+\Theta_T,\ Y_T^+=QX_T^++\Upsilon_T. \tag{13}$$

A. Causality-based Inference Method

Although the Granger estimator presents a direct and analytic expression for inferring W, it is based on observations over multiple process rounds and the observation noises are often ignored. It cannot be directly applied in single observation round case. Nevertheless, it provides beneficial modeling ideas from the perspective of node causality. Similar with $R_0^x(t)$ and $R_1^x(t)$, we define the following sample covariance matrix and its one-lag version as

$$\Sigma_0(T)=\frac{1}{T}(Y_T^-)(Y_T^-)^{\mathsf{T}},\ \Sigma_1(T)=\frac{1}{T}(Y_T^+)(Y_T^-)^{\mathsf{T}}. \tag{14}$$

Before demonstrating the relationship between R_0^x/R_1^x and Σ_0/Σ_1, we first present the asymptotic performance of Σ_0/Σ_1.

Lemma 1. *Given arbitrary $Z_T\in\mathbb{R}^{n\times T}$ and noise matrix $\Theta_T\in\mathbb{R}^{n\times T}$ with i.i.d. Gaussian entries, we have*

$$\Pr\left\{\left|[\frac{1}{T}\Theta_T Z_T^{\mathsf{T}}]_{ij}\right|<\epsilon\right\}\geq 1-\frac{(|z|_m)^2\sigma^2}{T\epsilon^2}, \tag{15}$$

where $|z|_m=\max\{|z_t^j|,t=1,\cdots,T,\ j=1,\cdots,n\}$. Specifically, if $|z|_m<\infty,\forall T\in\mathbb{N}^+$, then

$$\Pr\left\{\lim_{T\rightarrow\infty}\frac{1}{T}\Theta_T Z_T^{\mathsf{T}}=\mathbf{0}\right\}=1. \tag{16}$$

Proof. To ease notation, let $\Xi(T)=\frac{1}{T}\Theta_T Z_T^{\mathsf{T}}=\sum_{t=1}^{T}\theta_{t-1}z_{t-1}^{\mathsf{T}}$. Consider an element-wise analysis for $\Xi_{ij}(T)=\frac{1}{T}[\Theta_T Z_T^{\mathsf{T}}]_{ij}$, which is calculated by

$$\Xi_{ij}(T)=\frac{1}{T}\sum_{t=1}^{T}\theta_{t-1}^i z_{t-1}^j. \tag{17}$$

Since $\theta_t\sim\mathcal{N}(0,\sigma^2)$, it follows that

$$\mathbb{E}[\Xi_{ij}(T)]=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\theta_{t-1}^i]z_{t-1}^j=0, \tag{18}$$

$$\mathbb{D}[\Xi_{ij}(T)]=\sum_{t=1}^{T}(\frac{x_{t-1}^j}{T})^2\sigma^2\leq\frac{(|z|_{\max}^j)^2\sigma^2}{T}, \tag{19}$$

where $|z|_{\max}^j=\max\{|z_t^j|,t=0,1,\cdots,T-1\}$. By the famous Chebyshev inequality, given arbitrary $\epsilon>0$, we have

$$\Pr\{|\Xi_{ij}(T)|<\epsilon\}\geq 1-\frac{\mathbb{D}[\Xi_{ij}(T)]}{\epsilon^2}\geq 1-\frac{(|z|_{\max}^j)^2\sigma^2}{T\epsilon^2}. \tag{20}$$

Consequently, $\forall i,j\in\mathcal{V}$, $\Pr\{|\Xi_{ij}(T)|<\epsilon\}\geq 1-\frac{(|z|_{\max}^j)^2\sigma^2}{T\epsilon^2}\geq 1-\frac{(|z|_{\max})^2\sigma^2}{T\epsilon^2}$, which completes the first statement.

Next, if $|z|_m<\infty,\forall T\in\mathbb{N}^+$, when $T\rightarrow\infty$, it yields that

$$\lim_{T\rightarrow\infty}\Pr\{|\Xi_{ij}(T)|<\epsilon\}=1. \tag{21}$$

Finally, in the matrix form, (21) is equivalent to $\Pr\{\lim_{T\rightarrow\infty}\Xi(T)=\mathbf{0}\}=1$. The proof is completed. ☐

Lemma 1 illustrates the independence of sample matrix on the noise matrix in single observation round. The result (16) also applies to linear transform $B\Theta$, where $B\in\mathbb{R}^{n\times n}$ and

$\|B\| < \infty$. Since only $\{y_t\}_{t=0}^T$ are directly available, for every two adjacent observations, it follows that

$$\begin{aligned} y_t &= Wx_{t-1} + \theta_{t-1} + \upsilon_t \\ &= Wy_{t-1} - W\upsilon_{t-1} + \theta_{t-1} + \upsilon_t \\ &= Wy_{t-1} + \omega_t, \end{aligned} \tag{22}$$

where $\omega_t = -W\upsilon_{t-1} + \theta_{t-1} + \upsilon_t$, satisfying $N(0, \sigma_\upsilon^2 WW^\mathsf{T} + \sigma_\upsilon^2 I + \sigma_\theta^2 I)$, which is highly auto-correlated. Besides, ω_t is independent of all $\{x_{t'}\}_{t'<t}$ and $\{\theta_{t'}\}_{t'<t-1}$. Note that (22) only represents the quantitative relationship between adjacent observations, not a causal dynamical process. Based on the observations, we present the following theorem.

Considering the causal representation of y_t, we have

$$\mathbb{E}(\omega_t y_{t-1}^\mathsf{T}) = \mathbb{E}(-W\upsilon_{t-1} y_{t-1}^\mathsf{T}) = -\sigma_\upsilon^2 W. \tag{23}$$

Theorem 1. *If $W \in \mathcal{S}_a$, we have*

$$W = \Sigma_1(\infty)(\Sigma_0(\infty) - \sigma_\upsilon^2 I)^{-1}, \tag{24}$$

where $\Sigma_1(\infty) = \lim_{T\to\infty} \Sigma_1(T)$ and $\Sigma_0(\infty) = \lim_{T\to\infty} \Sigma_0(T)$.

Proof. Substitute (22) into $\Sigma_1(T)$ and it follows that

$$\Sigma_1(T) = \frac{1}{T}(Y_T^+)(Y_T^-)^\mathsf{T} = \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1}^\mathsf{T}$$
$$= \frac{W}{T}\sum_{t=1}^{T} y_{t-1}y_{t-1}^\mathsf{T} + \frac{1}{T}\sum_{t=1}^{T}\left(\theta_{t-1} + \upsilon_t - W\upsilon_{t-1}\right) y_{t-1}^\mathsf{T}. \tag{25}$$

Note that when $W \in \mathcal{S}_a$, it follows that $\lim_{t\to\infty}\|x_t\| = 0$ by (7). Since θ_{t-1} and υ_t are independent of y_t, applying Lemma 1 on $\Sigma_1(T)$, it yields that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\theta_{t-1}y_{t-1}^\mathsf{T} = \mathbf{0},\ \ \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\upsilon_t y_{t-1}^\mathsf{T} = \mathbf{0}. \tag{26}$$

Recalling υ_{t-1} is independent of x_{t-1}, it follows that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\upsilon_{t-1}y_{t-1}^\mathsf{T} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\upsilon_{t-1}(x_{t-1}+\upsilon_{t-1})^\mathsf{T} = \sigma_\upsilon^2 I.$$

Then, one infers that

$$\lim_{T\to\infty}\Sigma_1(T) = W\left(\lim_{T\to\infty}\Sigma_0(T) - \sigma_\upsilon^2 I\right). \tag{27}$$

Therefore, $W = \Sigma_1(\infty)\left(\Sigma_0(\infty) - \sigma_\upsilon^2 I\right)^{-1}$ holds. Hence, the proof is completed. □

Different from Granger estimator (10), Theorem 1 relaxes the dependence on multiple observation rounds, and presents a causality-based estimator for single observation round, while taking the observation noises into consideration. Under finite horizon T, the causality-based estimator is given by

- **Causality-based estimator**:

$$\hat{W}_c = \Sigma_1(T)(\Sigma_0(T) - \sigma_\upsilon^2 I)^{-1}. \tag{28}$$

Remark 1. *In principle, the interaction matrix can be estimated with arbitrarily large precision as observation horizon increases. Note that in Theorem 1 it is demonstrated that $W \in \mathcal{S}_a$ is the precondition, and this is to ensure $\|x_\infty\| < \infty$ strictly holds. Even if $W \in \mathcal{S}_m$, we have $\mathbb{E}(x_t) = W^\infty x_0$, i.e., $\|x_\infty\|$ will not go infinity with large probability (as long as the noise variance is not very large). In practice, given observations of finite horizon of $\mathcal{S}_m$ cases, (28) is still a feasible method and thus has better adaptability.*

B. Correlation-based Modification Design

Like Granger estimator (10), the proposed causality-based estimator (28) is a asymptotic solving manner, i.e., it will approximate the real W as $T \to \infty$. When T is small, the performance of (28) may largely degenerate. The main cause is that directly using $\sigma_\upsilon^2 I$ to filter the influence of observation noises is not a appropriate choice, for $\sigma_\upsilon^2 I$ is a meaningful statistical characteristic in the sense of large noise samples.

Inspired by the correlation measurement (8), an alternative way to alleviate the influence of observation noises is to implement correlation coefficient calculation, which also directly represents the linear correlation between nodes. Then, we define the following correlation-based sample matrix and its one-lag version as

$$S_0(T) = \frac{1}{T}\sum_{t=0}^{T-1}\tilde{y}_t^-(\tilde{y}_t^-)^\mathsf{T},\ S_1(T) = \frac{1}{T}\sum_{t=1}^{T}\tilde{y}_t^+(\tilde{y}_{t-1}^-)^\mathsf{T}, \tag{29}$$

where the elements of $\tilde{y}_t^-$ and $\tilde{y}_t^+$ are given by

$$[\tilde{y}_t^-]_i = [y_t - \frac{Y_T^-\mathbf{1}_T}{T}]_i/\rho_i^-,\ [\tilde{y}_t^+]_i = [y_t - \frac{Y_T^+\mathbf{1}_T}{T}]_i/\rho_i^+, \tag{30}$$

where $\mathbf{1}_T \in \mathbb{R}^T$ is the all-one vector and

$$\rho_i^- = \sqrt{\sum_{t=0}^{T-1}(y_t^i - [\frac{Y_T^-\mathbf{1}_T}{T}]_i)^2}, \rho_i^+ = \sqrt{\sum_{t=1}^{T}(y_t^i - [\frac{Y_T^+\mathbf{1}_T}{T}]_i)^2}.$$

Finally, the correlation-based modified causality estimator is designed as

- **Correlation-modified estimator**:

$$\hat{W}_s = S_1(T)S_0^{-1}(T). \tag{31}$$

Remark 2. *The three estimators (10), (28) and (31) approximate W from different angles. From a pure statistical viewpoint, (10) implements the inference over multiple rounds of observations at a single moment, while (28) and (31) do that over a sequence of observations in a single observation round, which is more common in practice. Compared with estimator (28), (31) is a modified version of (28) for small horizon T.*

The main merit of (31) lies that it subtly takes the noise filtering and the internal node correlation in the NDS into account at the same time by the correlation coefficient calculation. . In statistics, it can be seen as normalization operation to quantity the observations in the same measurement space. In Simulation section, we will demonstrate that correlation-based modification improves the inference performance of estimator (28) in small sample cases, and obtains results that are generally no worse than that by the OLS method. Specifically, if $W \in \mathcal{S}_a$, (31) will achieve better inference accuracy that the OLS estimator. The reason is that $W \in \mathcal{S}_a$ incurs that the system state will always converge zero, which indicates the system is mainly driven by noises regardless of the initial states. Under this situation, (31) enhances the correlation and causality between two observations.

C. Relationships between Different Methods

In this part, we demonstrate the mutual relation between the causality-based estimator and the Granger, OLS estimators.

Theorem 2. *If $W \in \mathcal{S}_a$, we have*

$$\Sigma_0(\infty) = R_0^x(\infty) + \sigma_\upsilon^2 I,\ \Sigma_1(\infty) = R_1^y(\infty), \tag{32}$$

where $\Sigma_0(\infty) = \lim\limits_{T\to\infty} \Sigma_0(T)$ *and* $\Sigma_1(\infty) = \lim\limits_{T\to\infty} \Sigma_1(T)$.

Proof. Without losing generality, we first consider $\Sigma_0(T)$. Substituting the expanded form (6) of x_t into $y_t y_t^\mathsf{T}$, one obtains

$$y_t y_t^\mathsf{T} = (W^t x_0 + \eta_t)(W^t x_0 + \eta_t)^\mathsf{T}, \tag{33}$$

where $\eta_t = \sum\limits_{m=1}^{t} W^{m-1}\theta_{t-m} + \upsilon_t$ and $\eta_0 = \upsilon_0$. Then, (33) is expanded as

$$y_t y_t^\mathsf{T} = \underbrace{W^t x_0 x_0^\mathsf{T} (W^t)^\mathsf{T}}_{Q_1^t} + \underbrace{W^t x_0 \eta_t^\mathsf{T}}_{Q_2^t} + \underbrace{\eta_t x_0^\mathsf{T} (W^t)^\mathsf{T}}_{Q_3^t} + \underbrace{\eta_t \eta_t^\mathsf{T}}_{Q_4^t}. \tag{34}$$

First, consider taking the average of all $\{Q_1^t\}_{t=0}^{T-1}$.

$$\frac{1}{T}\sum_{t=0}^{T-1} Q_1^t = \frac{1}{T}\sum_{t=0}^{T-1} W^t x_0 x_0^\mathsf{T} (W^t)^\mathsf{T} \tag{35}$$

Note that when $W \in \mathcal{S}_a \cup \mathcal{S}_m$, $W^\infty x_0$ converge to a constant vector. Therefore, it yields that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} Q_1^t = W^\infty x_0 x_0^\mathsf{T} (W^\infty)^\mathsf{T}. \tag{36}$$

Then, consider $\frac{1}{T}\sum\limits_{t=0}^{T-1} Q_2^t = \frac{1}{T}\sum\limits_{t=0}^{T-1} W^t x_0 \eta_t^\mathsf{T}$. Since η_t is a typical linear combination of Gaussian noises $\{\theta_m\}_{m=0}^{t-1}$ and υ_t, and is independent of $W^t x_0$, by Lemma 1, one infers that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} Q_2^t = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} (W^t x_0)\eta_t^\mathsf{T} = \mathbf{0}. \tag{37}$$

The average of all $\{Q_3^t\}_{t=0}^{T-1}$ is likewise, i.e.,

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} Q_3^t = \mathbf{0}. \tag{38}$$

Finally, focus on the calculation of $\frac{1}{T}\sum\limits_{t=0}^{T-1} Q_4^t$. Since $\mathbb{E}(\theta_t\theta_t^\mathsf{T}) = \sigma_\theta^2 I$ and $\mathbb{E}(\upsilon_t\upsilon_t^\mathsf{T}) = \sigma_\upsilon^2 I$, one can divide Q_4^t as

$$\begin{aligned} Q_4^t = &\sum_{m=0}^{t}\left(W^{m-1}\theta_{t-m}(\upsilon_t + \sum_{\substack{m_1=0,\\ m_1\neq m}}^{t} \theta_{t-m_1}^\mathsf{T} (W^{m_1-1})^\mathsf{T}) \right) \\ &+ \upsilon_t \left(\sum_{m_1=0}^{t} \theta_{t-m_1}^\mathsf{T} (W^{m_1-1})^\mathsf{T} \right) \\ &+ \sum_{m=0}^{t} W^{m-1}\theta_{t-m}\theta_{t-m}^\mathsf{T}(W^{m-1})^\mathsf{T} + \upsilon_t\upsilon_t^\mathsf{T}. \end{aligned} \tag{39}$$

Consider the first term in Q_4^t. For simple expression, define

$$\theta_t^a(m) = W^{m-1}\theta_{t-m},\ \theta_t^b(m) = \sum_{\substack{m_1=0,\\ m_1\neq m}}^{t} W^{m_1-1}\theta_{t-m_1}. \tag{40}$$

As $W \in \mathcal{S}_a$, one can infer that $\lim\limits_{t\to\infty} \|\theta_t^b(m)\| < \infty$. Therefore, by the famous Lebesgue's dominated convergence theorem and Lemma 1, it follows that with probability one

$$\begin{aligned} &\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{m=0}^{t} \theta_t^a(m)(\theta_t^b(m))^\mathsf{T} \\ =&\sum_{t=0}^{\infty} \lim_{T\to\infty} \frac{\sum\limits_{m=0}^{t} \theta_t^a(m)(\theta_t^b(m))^\mathsf{T}}{T} = \mathbf{0}. \end{aligned} \tag{41}$$

Likewise, for the second term in (39), it also holds that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \upsilon_t \left(\sum_{m_1=0}^{t} \theta_{t-m_1}^\mathsf{T}(W^{m_1-1})^\mathsf{T} \right) = \mathbf{0}. \tag{42}$$

As for the last two parts in (39), recalling $\mathbb{D}[\theta_t] = \sigma_\theta^2 I$ and $\mathbb{D}[\upsilon_t] = \sigma_\upsilon^2 I$, and one infers that

$$\begin{aligned} &\lim_{T\to\infty} \frac{\sum\limits_{t=0}^{T-1}\sum\limits_{m=0}^{t} W^{m-1}\theta_{t-m}\theta_{t-m}^\mathsf{T}(W^{m-1})^\mathsf{T} + \sum\limits_{t=0}^{T-1}\upsilon_t\upsilon_t^\mathsf{T}}{T} \\ =&\sum_{t=0}^{\infty} W^t \lim_{T\to\infty}\left(\sum_{m=0}^{T-t} \theta_{t-m}\theta_{t-m}^\mathsf{T}/T \right)(W^t)^\mathsf{T} + \sigma_\upsilon^2 I \\ =&\sigma_\theta^2 \sum_{t=0}^{\infty} W^t (W^t)^\mathsf{T} + \sigma_\upsilon^2 I = \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1} Q_4^t. \end{aligned} \tag{43}$$

In summary, taking (36)-(38) and (43) into $\lim\limits_{T\to\infty}\Sigma_0(T)$, it yields that

$$\begin{aligned} \lim_{T\to\infty}\Sigma_0(T) &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}(Q_1^t + Q_2^t + Q_3^t + Q_4^t) \\ &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}(Q_1^t + Q_4^t) \\ &= W^\infty x_0 x_0^\mathsf{T}(W^\infty)^\mathsf{T} + \sigma_\theta^2\sum_{t=0}^{\infty} W^t(W^t)^\mathsf{T} + \sigma_\upsilon^2 I \\ &= R_0^x(\infty) + \sigma_\upsilon^2 I. \end{aligned} \tag{44}$$

The proof of $\lim_{T\to\infty} \Sigma_1(T) = R_1^x(\infty)$ is likewise and omitted here. The proof is completed. □

Theorem 2 demonstrates the equivalent condition between estimators (10) and (28). It reveals that the expected state covariance matrix of $T \to \infty$ is identical with the sample covariance matrix along all the single time horizon, which is an interesting result that describes the relation between multiple and single observation rounds.

Next, from the perspective of OLS method, to infer the W from $\{y_t\}_{t=1}^T$ is to solving the following problem

$$\mathbf{P_1}: \quad \min_W \sum_{t=1}^T \|y_t - W y_{t-1}\|^2. \tag{45}$$

Besides, in optimization field, it is commonly to use some modification on the objective function of the OLS problem to improve the solving performance, typically by adding a regularization term. Taking $\|W\|_F$ as the regularizator, $\mathbf{P_1}$ is transformed to

$$\mathbf{P_2}: \ \min_W \sum_{t=1}^T \|y_t - W y_{t-1}\|^2 + \beta \|W\|_F^2, \tag{46}$$

where β is a regularization parameter. Then, we present the following theorem to demonstrate the relationship between the proposed causality-based method and the LS method.

Theorem 3. *If $\sigma_v = 0$, then estimator (28) is equivalent with the solution of $\boldsymbol{P_1}$. If $\sigma_v \neq 0$, then estimator (28) is equivalent with the solution of $\boldsymbol{P_2}$ with $\beta = -\sigma_v^2$.*

Proof. Recalling $Y_T^- = [y_0, y_2, \cdots, y_{T-1}]$ and $Y_T^+ = [y_1, y_2, \cdots, y_T]$, the objective function of $\mathbf{P_1}$ is rewritten as

$$\min_W \|Y_T^+ - W Y_T^-\|_F^2. \tag{47}$$

Then, by finding the derivative, one obtains that the optimal solution is given by

- **OLS estimator**:

$$\hat{W}_o = Y_T^+ (Y_T^-)^\mathsf{T} (Y_T^- (Y_T^-)^\mathsf{T})^{-1} = \Sigma_1(T)\Sigma_0^{-1}(T). \tag{48}$$

Apparently, $\hat{W}_o = \hat{W}_c$ when $\sigma_v = 0$. For $\mathbf{P_2}$, its objective function is rewritten as

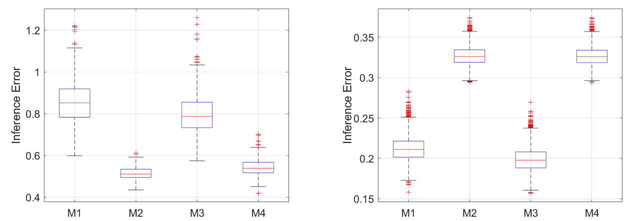
$$\min_W \|Y_T^+ - W Y_T^-\|_F^2 + \beta \|W\|_F^2, \tag{49}$$

whose optimal solution is given by

$$\hat{W}_r = \Sigma_1(T)(\Sigma_0(T) + \beta I)^{-1}. \tag{50}$$

It is straightforward that $\hat{W}_r = \hat{W}_c$ when $\beta = -\sigma_v^2$. Hence, the proof is completed. □

Theorem 3 reveals the close relationship of the proposed causality-based method with least square methods. On the one hand, it illustrates a new interpretation for using LS methods to infer the interaction topology from the perspective of node causality and correlation. On the other hand, it provides the idea about how to set a reasonable regularization term and parameters for the LS problem modeling when both the input and output data are corrupted. Specifically, in the latter case, $\beta = -\sigma_v^2$ essentially is not a typical regularization but a de-regularization form, which is quite different from normal regularization situations where $\beta > 0$.



(a) 500 samples. (b) 5000 samples.

Fig. 1. Inference error distribution with different observations scales.

Remark 3. *In summary, the original Granger estimator applies to cases of multiple observation rounds, while the OLS estimator and our proposed method can be directly used in single observation cases. All the three methods approximate the real interaction matrix asymptotically, where our proposed causality-based method outperforms the OLS estimator. Besides, when the observation horizon is small, the proposed correlation-based modification design can achieve inference results that are no worse than using the OLS estimator.*

IV. SIMULATION

In this section, we conduct extensive simulations to demonstrate the effectiveness of our proposed methods, by comparing with the classical Granger estimator and ordinary least method. Then, we will elaborate the merits and demerits of the methods, as well as their applicabilities in different cases.

A. Simulation Setup

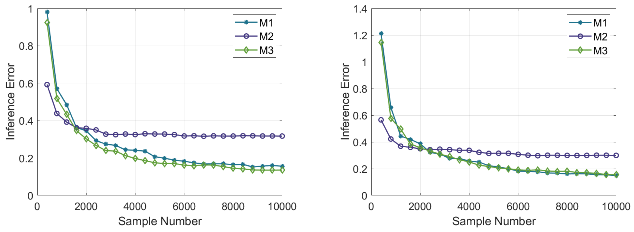
The most critical components are the adjacent matrix A and the interaction matrix W. We randomly generate a directed topology structure of 20 nodes, and the weight of W is designed by Laplacian rule. Both $W \in \mathcal{S}_a$ and $W \in \mathcal{S}_m$ are considered. For generality, the initial states of all agents are randomly selected from the interval $[-100, 100]$, and the variance of the process and observation noise satisfy $\sigma_\theta^2 = 1$ and $\sigma_v^2 = 1$. To evaluate the inference accuracy, we use the following index as

$$E_r = \|\hat{W} - W\|. \tag{51}$$

For simple expression, hereafter we denote the Granger estimator (10), the OLS estimator (48), the proposed causality-based method (28) and the correlation-based method (31) by M1, M2, M3 and M4, respectively.

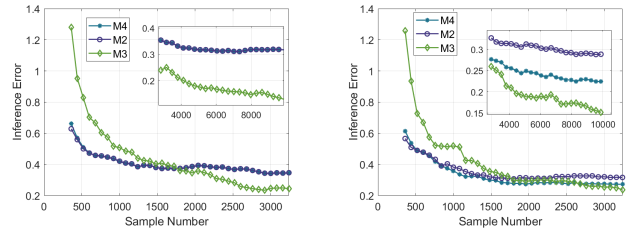
B. Results and Analysis

First, we conduct two groups of contrast experiments for the statistical analysis of M1-M3, where the interaction processes are implemented 500 and 5000 time iterations (denoted by T) and the same process rounds (denoted by r), respectively. For



(a) $W \in \mathcal{S}_m$ (b) $W \in \mathcal{S}_a$

Fig. 2. Asymptotic performance of the three methods. The same topology structure and initial system state are set for two kinds of experiments.



(a) $W \in \mathcal{S}_a$ (b) $W \in \mathcal{S}_m$

Fig. 3. Asymptotic performance of the three methods. The same topology structure and initial system state are set for two kinds of experiments.

a fair comparison, we use all $R_1^y(t)(R_0^y(t-1)-\sigma_v^2)^{-1}\}_{t=1}^T$ and make the statistics for M1, while all the observations $\{y_t(r)\}$ in the single round are used to infer a topology for M2 and M3. As Fig. 1 shows, when the observation scale is small, M2 obtains best inference accuracy and becomes the worst in large observation scale. A notable point is that M3 always perform better than M1 and obtain best inference accuracy in large observation scale. Another point is that the variance of the error distribution by M2 is always smaller than the other two, which illustrates the solving stability of least square method. In either cases of $W\mathcal{S}_m$, the performance of M4 is generally the same as M2.

Next, we directly present the asymptotic inference performance of the methods in terms of sample size, covering both marginal and asymptotic cases, as shown in Fig. 2. Note that the accuracy magnitudes of M1 and M3 are very close, which verifies the conclusion of Theorem 1 that the inference result from single observation round becomes equivalent with that from multiple observation rounds when $T \to \infty$. Comparing the results of M2 and M3, it is found that M3 behaves worse than M2 when in small observation scale, but outperforms M2 asymptotically, while the accuracy of M3 remains stable as the data size increase. The main reason is that the statistical characteristic of observation noises will matter a lot when the observation scale is large, which is not taken into account by ordinary least square method.

Finally, the performance of M2-M3 are evaluated under the same single observation round, as demonstrated in Fig. 3. Fig. 3(a) shows that, for a marginal stable NDS, M4 has almost the same inference performance as M2. However, for a asymptotic stable NDS, M4 outperforms M2 asymptotically, but is still worse than M3, as shown in Fig. 3(b). In summary, M3 applies to large observation scales, while M4 applies to other situations with no worse performance than M2.

V. Conclusions

In this paper, we investigate the topology inference problem for NDSs from the perspective of node causality and correlation. First, we propose the causality-based estimator for single observation round, which has analogous asymptotic performance of the Granger estimator for multiple observation rounds. Then, we design a correlation-based modification to alleviate the inference performance degradation when the observation horizon is small, and achieve inference results that are no worse than using OLS estimator. Finally, we demonstrate the close relationships between our proposed method and the Granger and OLS estimators in asymptotically solving manners, and prove the equivalent conditions between the methods in terms of the system stability, the observation scale and statistical characteristic of process/observation noises. Extensive simulations with different observation scales are conducted to demonstrate the effectiveness of the proposed method. Future directions include extending the method to the dynamic topology and nonlinear dynamical system cases.

References

[1] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215-233, 2007.

[2] M. Nokleby and W. U. Bajwa, "Stochastic optimization from distributed streaming data in rate-limited networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 1, pp. 152–167, 2018.

[3] A. Ahmed and E. P. Xing, "Recovering time-varying networks of dependencies in social and biological studies," *Proceedings of the National Academy of Sciences*, vol. 106, no. 29, pp. 11 878–11 883, 2009.

[4] 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.

[5] C. Liu, J. He, S. Zhu, and C. Chen, "Dynamic topology inference via external observation for multi-robot formation control," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*. IEEE, 2019, pp. 1–6.

[6] I. Brugere, B. Gallagher, and T. Y. Berger-Wolf, "Network structure inference, a survey: Motivations, methods, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 1–39, 2018.

[7] S. Mahdizadehaghdam, H. Wang, H. Krim, and L. Dai, "Information diffusion of topic propagation in social media," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 569–581, 2016.

[8] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 248–262, 2012.

[9] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 50–66, 2013.

[10] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[11] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality," *Proceedings of the National Academy of Sciences*, vol. 101, no. 26, pp. 9849–9854, 2004.

Manuscript 1412 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.

[12] A. Santos, V. Matta, and A. H. Sayed, "Local tomography of large networks under the low-observability regime," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 587–613, 2020.

[13] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

[14] M. T. Schaub, S. Segarra, and H.-T. Wai, "Spectral partitioning of time-varying networks with unobserved edges," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019, pp. 4938–4942.

[15] Y. Zhu, M. T. Schaub, A. Jadbabaie, and S. Segarra, "Network inference from consensus dynamics with unknown parameters," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 300–315, 2020.

[16] G. Karanikolas, G. B. Giannakis, K. Slavakis, and R. M. Leahy, "Multi-kernel based nonlinear models for connectivity identification of brain networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 6315–6319.

[17] G. V. Karanikolas, O. Sporns, and G. B. Giannakis, "Multi-kernel change detection for dynamic functional connectivity graphs," in *Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 1555–1559.

[18] S. Wang, E. D. Herzog, I. Z. Kiss, W. J. Schwartz, G. Bloch, M. Sebek, D. Granados-Fuentes, L. Wang, and J.-S. Li, "Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, pp. 9300–9305, 2018.

[19] N. Shahid, N. Perraudin, V. Kalofolias, G. Puy, and P. Vandergheynst, "Fast robust pca on graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 740–756, 2016.

[20] M. Onuki, S. Ono, M. Yamagishi, and Y. Tanaka, "Graph signal denoising via trilateral filter on graph spectral domain," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 137–148, 2016.

[21] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under laplacian and structural constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.

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

[23] A. Tsiamis and G. J. Pappas, "Finite sample analysis of stochastic system identification," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 3648–3654.

[24] S. Oymak and N. Ozay, "Non-asymptotic identification of lti systems from a single trajectory," in *IEEE American control conference (ACC)*. IEEE, 2019, pp. 5655–5661.

[25] A. H. Sayed *et al.*, "Adaptation, learning, and optimization over networks," *Foundations and Trends® in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.

[26] V. Matta and A. H. Sayed, "Consistent tomography under partial observations over adaptive networks," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 622–646, 2019.

Manuscript 1412 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.