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
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 0 be all-zero matrix in compatible dimensions. Unless otherwise noted, ||·|| and ||·||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 $j$ will use information from $i$. 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 $N_i = \{j \in \mathcal{V} : a_{ij} > 0\}$ as the in-neighbor set of $i$, and $d_i = |N_i|$ 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
\[ x_t = Wx_{t-1} + \theta_{t-1}, \]
\[ y_t = x_t + \nu_t, \tag{1} \]
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 $\theta_t$ and $\nu$ represent the i.i.d process and observation noises, satisfying $\theta_t \sim N(0, \sigma_{\theta}^2 I)$ and $\nu_t \sim N(0, \sigma_{\nu}^2 I)$. Denote the spectral radius of $W$ by $\rho(W)$. We present asymptotically stable matrix class $S_a$ and the (strictly) marginally stable matrix $S_m$ as follows:
\[ S_a = \{Z \in \mathbb{R}^{n \times n}, \rho(Z) < 1\}, \]
\[ S_m = \{Z \in \mathbb{R}^{n \times n}, \rho(Z) = 1\} \quad \text{and the geometric multiplicity of eigenvalue 1 equals to one}. \tag{2} \]

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 $i \neq j$,

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

where the auxiliary parameter $\gamma$ 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 $W \in S_m$. A typical matrix in $S_a$ 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

\[ \lim_{t \to \infty} W^t = \begin{cases} 
0, & \text{if } W \in S_a \\
W^\infty, & \text{if } W \in S_m, \tag{5} 
\end{cases} \]

where 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

\[ \lim_{t \to \infty} y_t = \begin{cases} 
\sum_{m=1}^{\infty} W^{m-1} \theta_{t-m} + v_\infty, & \text{if } W \in S_a, \\
W^\infty x_0 + \sum_{m=1}^{\infty} W^{m-1} \theta_{t-m} + v_\infty, & \text{if } W \in S_m. \tag{7} \end{cases} \]

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

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

where $\rho_i^0 = \sqrt{\sum_{t=1}^{T} (x_t^i - \bar{x}_i^i)^2}$ is the sample standard deviations of $\{x_t^i\}$ and $\bar{x}_i = \sum_{t=0}^{T} x_t^i / T$, $\forall i \in \mathcal{V}$. Note that $\rho_{ij}^0$ 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

\[
R_0(t) = \mathbb{E}[x_t x_t^T] , \quad R_1(t) = \mathbb{E}[x_t x_{t-1}^T].
\] (9)

Since \( R_1(t) = \mathbb{E}[W \cdot x_{t-1} x_{t-1}^T + \theta_t x_{t-1}^T] = W R_0(t - 1), \) one obtains the Granger estimator [12] as

- **Granger estimator:**
  \[
  \hat{W}_G = R_1^+(t)(R_0^+(t - 1))^{-1}.
  \] (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

\[
\phi_M : \{y_t\}_{t=1}^T \rightarrow W.
\] (11)

In this paper, we focus on finding the mapping mechanisms \( \phi_M \) 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

\[
X_T^+ = [x_0, x_2, \cdots, x_{T-1}], \quad X_T^- = [x_1, x_2, \cdots, x_T],
\]

\[
Y_T^- = [y_0, y_2, \cdots, y_{T-1}], \quad Y_T^+ = [y_1, y_2, \cdots, y_T],
\]

\[
\Theta_T = [\theta_0, \theta_1, \cdots, \theta_{T-1}], \quad \Upsilon_T = [\upsilon_1, \upsilon_2, \cdots, \upsilon_T].
\] (12)

Then, the whole evolution process is compactly written as

\[
X_T^+ = W X_T^- + \Theta_T, \quad Y_T^+ = Q X_T^+ + \Upsilon_T.
\] (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 \( R_0(t) \) and \( R_1(t) \), we define the following sample covariance matrix and its one-lag version as

\[
\Sigma_0(T) = \frac{1}{T}(Y_T^-)(Y_T^-)^T, \quad \Sigma_1(T) = \frac{1}{T}(Y_T^+)(Y_T^-)^T.
\] (14)

Before demonstrating the relationship between \( R_0^+/R_1^+ \) and \( \Sigma_0/\Sigma_1 \), we first present the asymptotic performance of \( \Sigma_0/\Sigma_1 \).

**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

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

where \( |z|_m = \max\{|z|^2|, t = 1, \cdots, T, j = 1, \cdots, n\} \). Specifically, if \( |z|_m < \infty, \forall T \in \mathbb{N}^+, \) then

\[
\Pr\left\{ \lim_{t \to \infty} \frac{1}{T} \Theta_T Z_T^T = 0 \right\} = 1.
\] (16)

**Proof.** To ease notation, let \( \Xi(T) = \frac{1}{T} \Theta_T Z_T^T = \sum_{t=1}^T \theta_t z_{t-1}^T \).

Consider an element-wise analysis for \( \Xi(T) = \frac{1}{T} \Theta_T Z_T^T_{ij} \), which is calculated by

\[
\Xi_{ij}(T) = \frac{1}{T} \sum_{t=1}^T \theta_t z_{t-1}^j.
\] (17)

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

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

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

where \( |z|_{\text{max}} = \max\{|z|^2|, t = 0, 1, \cdots, T-1\} \). By the famous Chebyshev inequality, given arbitrary \( \epsilon > 0 \), we have

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

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

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

\[
\lim_{T \to \infty} \Pr\{ |\Xi_{ij}(T)| < \epsilon \} = 1.
\] (21)

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

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 \(y_t\) are directly available, for every two adjacent observations, it follows that
\[
y_t = W x_{t-1} + \theta_{t-1} + v_t
= W y_{t-1} - W v_{t-1} + \theta_{t-1} + v_t
= W y_{t-1} + \omega_t,
\] (22)
where \(\omega_t = -W v_{t-1} + \theta_{t-1} + v_t\), satisfying \(N(0, \sigma_\omega^2 W W^T + \sigma_\theta^2 I + \sigma_v^2 I)\), which is highly auto-correlated. Besides, \(\omega_t\) is independent of all \(\{x_i\}_{i < t}\) and \(\{v_i\}_{i < t-1}\). 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
\[
E(\omega_t y_{t-1}^T) = E(-W v_{t-1} y_{t-1}^T) = -\sigma_v^2 W.
\] (23)

**Theorem 1.** If \(W \in S_a\), we have
\[
W = \Sigma_1(\infty)(\Sigma_0(\infty) - \sigma_v^2 I)^{-1},
\] (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
\[
\Sigma_1(T) = \frac{1}{T}(Y_T^T Y_T)^{-1} = \frac{1}{T} \sum_{t=1}^{T} y_t y_t^T
= W \sum_{t=1}^{T} y_{t-1} y_{t-1}^T + \frac{1}{T} \sum_{t=1}^{T} (\theta_t + v_t - W v_{t-1}) y_t^T.
\] (25)

Note that when \(W \in S_a\), it follows that \(\lim_{T \to \infty} \|x_t\| = 0\) by (7). Since \(\theta_{t-1}\) and \(v_t\) are independent of \(y_t\), applying Lemma 1 on \(\Sigma_1(T)\), it yields that
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \theta_t y_{t-1}^T = 0, \quad \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} v_t y_{t-1}^T = 0.
\] (26)
Recalling \(v_t\) is independent of \(x_{t-1}\), it follows that
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} v_{t-1} y_{t-1}^T = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} v_{t-1} (x_{t-1} + v_{t-1})^T = \sigma_v^2 I.
\]
Then, one infers that
\[
\lim_{T \to \infty} \Sigma_1(T) = W \left( \lim_{T \to \infty} \Sigma_0(T) - \sigma_v^2 I \right).
\] (27)
Therefore, \(W = \Sigma_1(\infty) (\Sigma_0(\infty) - \sigma_v^2 I)^{-1}\) holds. Hence, the proof is completed. \(\square\)

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:**
\[
\hat{W}_c = \Sigma_1(T)(\Sigma_0(T) - \sigma_v^2 I)^{-1}.
\] (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 \(W \in S_a\) is the precondition, and this is to ensure \(\|x_T\| < \infty\) strictly holds. Even if \(W \in S_m\), we have \(E(x_T) = W x_0\), i.e., \(\|x_T\|\) 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 \(S_m\) 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 an 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 \(\sigma_v^2 I\) to filter the influence of observation noises is not a appropriate choice, for \(\sigma_v^2 I\) 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
\[
S_0(T) = \frac{1}{T} \sum_{t=0}^{T-1} \tilde{y}_t (\tilde{y}_t^\top)^T, \quad S_1(T) = \frac{1}{T} \sum_{t=0}^{T-1} \tilde{y}_t^\top (\tilde{y}_t^\top)^T,
\] (29)
where the elements of \(\tilde{y}_t\) and \(\tilde{y}_t^\top\) are given by
\[
[\tilde{y}_t]_i = [y_t - \frac{Y_T^\top 1_T}{T}]_i / \rho_i, \quad [\tilde{y}_t^\top]_i = [y_t - \frac{Y_T^\top 1_T}{T}]_i / \rho_i^2,
\] (30)
where \(1_T \in \mathbb{R}^T\) is the all-one vector and
\[
\rho_i = \sqrt{\frac{1}{T} \sum_{t=0}^{T-1} (y_t^\top - \frac{Y_T^\top 1_T}{T})_i^2}, \quad \rho_i^2 = \sqrt{\frac{1}{T} \sum_{t=0}^{T-1} (y_t^\top - \frac{Y_T^\top 1_T}{T})_i^2}.
\]
Finally, the correlation-based modified causality estimator is designed as

- **Correlation-modified estimator:**
\[
\hat{W}_m = S_1(T) S_0^{-1}(T).
\] (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 quantify 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 \(W \in S_a\), (31) will achieve better inference...
accuracy that the OLS estimator. The reason is that $W \in S_a$ 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 $W \in S_a$, we have

$$\Sigma_0(\infty) = R_0^e(\infty) + \sigma^2 \Sigma, \Sigma_1(\infty) = R_1^y(\infty),$$

(32)

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

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

$$y_t^T = (W^t x_0 + \eta_t)(W^t x_0 + \eta_t)^T,$$

(33)

where $\eta_t = \sum_{m=1}^{t} W^{m-1} \theta_{t-m} + Y_T$ and $\eta_0 = v_0$. Then, (33) is expanded as

$$y_t^T = W^t x_0 x_0^T (W^t)^T + \sum_{m=1}^{t} W^{m-1} \theta_{t-m} \eta_{t-m} (W^t)^T + \eta_t \eta_t^T.$$  

(35)

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

$$\frac{1}{T} \sum_{t=0}^{T-1} Q_1^T = W^\infty x_0 x_0^T W^\infty.$$

(36)

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

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

(37)

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

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

(38)

Finally, focus on the calculation of $\frac{1}{T} \sum_{t=0}^{T-1} Q_4^T$. Since $\mathbb{E}(\theta_t \theta_t^T) = \sigma^2 \delta$ and $\mathbb{E}(v_t v_t^T) = \sigma^2 I$, one can divide $Q_4^T$ as

$$Q_4^T = \sum_{m=0}^{t} \left( W^{m-1} \theta_{t-m} (W^{m-1} \theta_{t-m})^T + \sum_{m=1}^{\infty} W^{m-1} \theta_{t-m} (W^{m-1} \theta_{t-m})^T + v_t v_t^T \right).$$

(39)

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

$$\theta_t(m) = W^{m-1} \theta_{t-m}, \theta_t(m) = \sum_{m=0}^{t} W^{m-1} \theta_{t-m}. $$

As $W \in S_a$, one can infer that $\lim_{T \to \infty} ||\theta_t(m)|| < \infty$. Therefore, by the famous Lebesgue’s dominated convergence theorem and Lemma 1, it follows that with probability one

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \left( \sum_{m=0}^{t} \theta_t(m) \theta_t(m)^T \right)^T = 0.$$  

(41)

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

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \left( \sum_{m=0}^{t} \theta_t(m) \theta_t(m)^T \right)^T = 0.$$  

(42)

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

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} v_t v_t^T = 0.$$  

(43)

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

$$\lim_{T \to \infty} \Sigma_0(T) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} (Q_1^T + Q_2^T + Q_3^T + Q_4^T)$$

$$= \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} (Q_1^T + Q_3^T)$$

$$= W^\infty x_0 x_0^T W^\infty + \sum_{t=0}^{\infty} W^t v_t v_t^T + \sigma^2 I$$

$$= R_0^e(\infty) + \sigma^2 I.$$  

(44)
The proof of \( \lim_{T \to \infty} \Sigma_1(T) = R_1^2(\infty) \) is likewise and omitted here. The proof is completed. \( \square \)

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

\[
P_1: \min W \sum_{t=1}^T \|y_t - W y_{t-1}\|^2. \quad (45)
\]

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, \( P_1 \) is transformed to

\[
P_2: \min W \sum_{t=1}^T \|y_t - W y_{t-1}\|^2 + \beta ||W||_F^2, \quad (46)
\]

where \( \beta \) 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 \( P_1 \). If \( \sigma_v \neq 0 \), then estimator (28) is equivalent with the solution of \( P_2 \) with \( \beta = -\sigma_v^2 \).

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

\[
\min W \|Y_T^+ - W Y_T^-\|^2_F. \quad (47)
\]

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

- **OLS estimator:**

  \[ \hat{W}_o = Y_T^+ (Y_T^-)^T (Y_T^- (Y_T^-)^T)^{-1} = \Sigma_1(T) \Sigma_0^{-1}(T). \quad (48) \]

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

  \[
  \min W \|Y_T^+ - W Y_T^-\|^2_F + \beta ||W||_F^2, \quad (49)
  \]

  whose optimal solution is given by

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

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

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 \( \beta > 0 \).

**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 \( W \in S_6 \) and \( W \in S_m \) 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 \( \sigma_o^2 = 1 \) and \( \sigma_v^2 = 1 \). To evaluate the inference accuracy, we use the following index as

\[ E_r = ||\hat{W} - W||. \quad (51) \]

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 fair comparison, we use all $R_k(t)(R_k(t-1)-\sigma_k^2)^{-1}T_{t=1}$ 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 $W \in S_m$, 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 confirms 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


