2021 60th IEEE Conference on Decision and Control (CDC)
December 13-15, 2021. Austin, Texas

Learning local modules in dynamic networks without prior topology information

Venkatakrishnan C. Rajagopal, Karthik R. Ramaswamy and Paul M.J. Van den Hof

Abstract— **Recently different identification methods have been developed for identifying a single module in a dynamic network. In order to select an appropriate predictor model one typically needs prior knowledge on the topology (interconnection structure) of the dynamic network, as well as on the correlation structure of the process disturbances. In this paper we present a new approach that incorporates the estimation of this prior information into the identification, leading to a fully data-driven approach for estimating the dynamics of a local module. The developed algorithm uses non-causal Wiener filters and a series of convex optimizations with parallel computation capabilities to estimate the topology, which subsequently is used to build the appropriate input/output setting for a predictor model in the local direct method under correlated process noise. A regularized kernel-based method is then employed to estimate the dynamic of the target module. This leads to an identification algorithm with attractive statistical properties that is scalable to handle larger-scale networks too. Numerical simulations illustrate the potential of the developed algorithm.**

I. Introduction

A dynamic network can be considered as an interconnection of various subsystems, or alternatively as a structured way of representing the dynamic relationships between observed signals. Dynamic networks find their importance in different fields like robotics, smart grids, transportation systems, oil and gas reservoirs, autonomous vehicles. In data-driven modeling questions, dynamic networks are usually represented by a directed graph where the nodes represent signals that might be measured, and edges representing linear time-invariant system mappings, referred to as modules. The network is externally affected by excitation signals that are under control of the user, and unmeasured disturbance signals (process noise). Over the past decade, increasing attention has been paid to the data-driven modeling of such dynamic networks. The three major problems in the data-driven modeling of dynamic networks include full network identification, topology identification and local module identification, including questions of identifiability.

Local module identification focuses on identifying the dynamics of a single module embedded in a network, which includes the problem of selecting the relevant node signals to be measured. In a prediction error setting this problem was addressed in [1]–[5]. A local direct identification method is presented in [6] that contains a full procedure for minimum variance estimation of a local module for the situation of correlated disturbances, and a variety of options for selecting the measured node signals. However, the current prediction error approach has two limitations: (a) for constructing the appropriate predictor model it requires a prior knowledge of the topology of the network and the correlation structure of the disturbances; and (b) the computational algorithm for estimating the module typically requires solving a non-convex optimization problem which is considered not to be scalable to large dimensions. Recently solutions have become available for problem (b), in the form of regularized kernel-based approaches for identifying a single module ([7], [8]), that build upon the signal selection procedures developed in [6].

In this paper we are going to tackle problem (a) by extending the regularized kernel-based method with a dedicated algorithm for estimating the network topology as well as the noise correlation structure of the disturbances, so as to provide a full data-driven algorithm for learning the dynamics of a single module. For identifying the two topologies (of the network and of the disturbances) we develop a multi-step least squares algorithm with analytical solutions (thus avoiding non-convex optimizations) and exploit a procedure developed in [9] for estimating network topology, based on the sparsity properties of non-causal Wiener filters, and extend it to networks with a non-diagonal disturbance model.

II. System setup

A. Dynamic networks

In this study we follow the dynamic network setting of [1]. To this end, we consider a network of L scalar *internal variables* (referred to as nodes) with dynamics defined by (1).

$$w_j(t) = \sum_{\substack{l=1 \\ l\neq j}}^{L} G_{jl}(q) w_l(t) + r_j(t) + v_j(t) \qquad (1)$$

where, $j \in \{1, \ldots, L\}$ and q^{-1} is the delay operator.

- G_{jl} are stable, strictly proper, rational transfer functions with no poles on the unit circle and are referred to as *modules*,
- There are no self loops in the system i.e. $G_{jj} = 0$.
- $(\mathcal{I} - G(q))^{-1}$ exists, is proper and stable.
- $v_j(t)$ is the process noise acting on the node j, and

The authors are with the Dept. of Electrical Engineering, Eindhoven University of Technology, The Netherlands crvk.93@gmail.com, {k.r.ramaswamy, p.m.j.vandenhof}@tue.nl

This work has received funding from the European Research Council (ERC), Advanced Research Grant SYSDYNET, under the European Union's Horizon 2020 research and innovation programme (grant agreement No 694504).

- $r_j(t)$ is the *external variable* that can be directly manipulated by the user. It may be absent in some nodes.

Collecting (1) for all the nodes, we get the expression (2)[1].

$$\begin{bmatrix} w_1 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G_{12} & \dots & G_{1L} \\ G_{21} & 0 & \dots & G_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ G_{L1} & G_{L2} & \dots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_L \end{bmatrix} + \begin{bmatrix} r_1 \\ \vdots \\ r_L \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_L \end{bmatrix}. \tag{2}$$

Written as a matrix equation we get

$$w(t) = G(q)w(t) + r(t) + v(t) \tag{3}$$

The process noise $v(t)$ is modelled as a stationary stochastic process with rational spectral density $\Phi_v(\omega)$, such that there exists a white noise process $e(t)$ with covariance matrix $\Lambda > 0$, such that $v(t) = H(q)e(t)$, where H is a stable, monic and minimum phase transfer function matrix. The white noise process $e(t)$ is referred to as the innovation signal of the network (G, H). The process noise $v(t)$ may be correlated resulting in non-zero entries along the off-diagonal terms in the power spectral density $\Phi_v(\omega)$ and the noise model $H(q)$. This situation is referred to as a *dynamic network with correlated noise*.

III. Problem setting and approach

One of the modules in the network G_{ji} is assigned to be the target module of which the dynamics needs to be identified from (a selection of) N measurements of the measured node signals. While there are prediction error methods that explicitly exploit the presence of measured excitation signals r, the local direct method of identification [6], [10] is built on the basis of a predictor model

$$w_{\mathcal{Y}}(t) = \bar{G}(q)w_{\mathcal{D}}(t) + \breve{v}(t) \tag{4}$$

where $\bar{G}(q)$ is the transfer function matrix and $\mathcal{Y}$, $\mathcal{D}$ are selections of node signals that are chosen on the basis of the presumed topology of the network, and the correlation structure of the disturbance signals. Typical elements in the selection of these sets of node signals include ([3], [6]):

- All paths from node w_i to node w_j and all loops around node w_j that do not pass through G_{ji} should be "blocked" by a measured node in $w_{\mathcal{D}}$, and
- There are restrictions on the presence of confounding variables, i.e. disturbance signals that have paths to nodes in $w_{\mathcal{D}}$ and $w_{\mathcal{Y}}$ through unmeasured nodes only.

Under these conditions the local direct method can provide a consistent and minimum variance estimate of the target module. However these conditions can only be satisfied if the topology of the network, i.e. the binary structure of $G(q)$, is known as well as the correlation structure of the disturbances, reflected by the binary structure of $H(q)$. For an accurate estimation of the target module, we will now exploit the following strategy consisting of three steps:

[1]time and frequency dependencies are dropped for convenience

1) Step 1: Select the node signals that carry information on the estimation problem of the target module, by estimating the "locality" of the output node by generalizing the Wiener filter based approach from [9] to situations with non-diagonal noise models, and remove (immerse) the remaining node signals from the network;
2) Step 2: Estimate the topology (i.e. the structure in G and H) of the immersed network, by applying a model structure selection procedure based on the AIC selection criterion and a multi-step least squares algorithm with analytical solutions and with parallel computation capabilities.
3) Step 3: Select an appropriate predictor model and estimate the target module dynamics by exploiting an Empirical Bayes method, in which the target module is parametrically parametrized while the remaining nuisance modules are modelled as Gaussian processes. This leads to an attractive and scalable estimation algorithm with a limited number of parameters to be estimated while achieving small mean-squared errors.

IV. Locality detection

A. Graph aspects

Before we identify the nodal and noise interconnection structure, we need to identify the nodes that contain information about the output node. In this study, we refer to such a set of nodes as the locality of the output node. The locality of a node can be seen as the Markov blanket [11] of the node in a network with correlated process noise. The Markov blanket of a node is the set of nodes that contain all the information about the node and is defined for Bayesian networks. Similarly, we need to define *locality* for networks with correlated process noise. To this end, we define the following sets.

Definition 1: Given a dynamic network (G, H), the children of a node j are defined by $\mathcal{C}_j := \{i | G_{ij} \neq 0\}$; its parents are defined by $\mathcal{P}_j := \{i | G_{ji} \neq 0\}$, and its noise confounders by $\mathcal{V}_j := \{i \mid \Phi_{v_j v_i} \neq 0\}$.

Usual practice in literature is to represent the topology of the network by an adjacency matrix [1]. However, since in this study, we identify both the network and the noise topology, we formally redefine them as follows.

Definition 2 (Topology): For a dynamic network (G, H), the topologies $\mathcal{T}_G$ and $\mathcal{T}_H$ are defined as:

$$\mathcal{T}_G(j,i) = \begin{cases} 1, & G_{ji} \neq 0 \\ 0, & G_{ji} = 0 \end{cases} \quad \text{and } \mathcal{T}_H(j,i) = \begin{cases} 1, & H_{ji} \neq 0 \\ 0, & H_{ji} = 0 \end{cases}.$$

With the help of the defined quantities, we now define the locality of the output node j.

Definition 3 (Locality of a node): Let j be a node in (3). The locality of the node j is defined by

$$\mathcal{B}_j := \{i \mid i = j \ \vee \ i \in \mathcal{P}_j \ \vee \ i \in \mathcal{C}_j \ \vee \ i \in \mathcal{P}(\mathcal{C}_j) \ \vee \ i \in \mathcal{V}_j \ \vee \ i \in \mathcal{V}(\mathcal{C}_j) \ \vee \ i \in \mathcal{P}(\mathcal{V}_j) \ \vee \ i \in \mathcal{P}(\mathcal{V}(\mathcal{C}_j))\}.$$

The process of removing a set of nodes and their signals from the network while preserving the second-order properties of the remaining nodes and signals is called immersion

(refer to [12] for details). Based on the way we define the locality, immersing the remaining nodes (i.e. nodes other than the nodes corresponding to set $\mathcal{B}_j$) in the network does not affect the output node or the interconnection structure of the output node. As a result, a smaller network for which we need to estimate the topology is obtained, which is referred to as the local network and is denoted by $(\tilde{G}, \tilde{H})$.

B. Identifying the Locality

Locality of the output node is analogous to the Markov blanket defined for Bayesian Networks. In [9], it has been shown that a multivariate non-causal Wiener filter [13] computed by projecting the node j on to the remaining nodal signals can detect the Markov blanket of the corresponding node by analyzing the sparsity of the Wiener filter. However, this sparsity result was developed for networks with a diagonal noise model only. In this section, we extend the results of [9] to networks with non-diagonal noise model.

Theorem 1 (Sparsity of the non-causal Wiener filter): Define $w := \big(w_1, \ldots, w_L\big)^T$ to be the nodal signals obtained from (3) and let the non-causal Wiener filter estimate of w_j on the basis of all other node signals be given by $\hat{w}_j(t) = \sum_{i \neq j} W_{ji}(q) w_i(t)$. Then $W_{ji}(z) \neq 0$ implies $w_i \in \mathcal{B}_j$. ■

For a proof, see [14]. As a result of Theorem 1, in order to identify the locality of j, we simply need to compute the multivariate non-causal Wiener filter by projecting the node j on the remaining nodes in the network and group the nodes that have a non-zero Wiener filter entry. Although there is no strict guarantee that in this way we detect *all* nodes in the locality, examples under which a true link between two nodes remains undetected are pathological [9]. In a practical implementation a threshold ρ needs to be defined where $\|W_{ji}\|_\infty > \rho$ is taken as evidence for the corresponding Wiener filter entry to be non-zero. The value of ρ can be chosen based on the signals w_j and w_i. For ease of computation, the Wiener filter is implemented as a non-causal FIR filter of length F, see e.g. [15].

V. Topology estimation

In this section, we develop a method to identify the network and noise topology $(\mathcal{T}_{\tilde{G}}, \mathcal{T}_{\tilde{H}})$ of the local network $(\tilde{G}, \tilde{H})$ obtained after immersing the nodes that are absent in the locality of the output node, using only analytical solutions and thereby making it simple and easy to apply. Numerous solutions are provided in the literature for network topology estimation considering the noise topology to be known and diagonal. In this paper, we address the situation of having an unknown noise topology that can be non-diagonal. We handle this by following a new multi-step strategy:

1) First, we estimate the innovation process using a high-order (non-parametric) model, similar to the identification algorithm presented in [16] for ARMAX networks.
2) Next we treat this innovation estimate as an additionally available input signal and simultaneously estimate the network and noise topology.

A. Innovation estimation

The network equation of the local network is described as,

$$w_{\mathcal{B}}(t) = \tilde{G} w_{\mathcal{B}}(t) + \tilde{H} \xi_{\mathcal{B}}(t). \tag{5}$$

In (5), $w_{\mathcal{B}}$ are the nodal signals of $\mathcal{B}_j$, $\tilde{G}$ is a hollow, strictly proper transfer function matrix, and $\tilde{H}$ is square, monic, stable and minimum phase, $\xi_{\mathcal{B}}(t)$ is a white noise vector (see [6] for more details).

Proposition 1: For every dynamic network of the form (5), there exists a network representation,

$$Q w_{\mathcal{B}}(t) = P w_{\mathcal{B}}(t) + \xi_{\mathcal{B}}(t), \tag{6}$$

where, Q is a diagonal and monic transfer function matrix, and P is a hollow, strictly proper transfer function matrix. ■

For a proof, see [14]. Every entry present in Q and P can be represented as an independent impulse response where its length ℓ_{ARX} is chosen sufficiently long to capture all the dynamics. By approximating Q and P with (high-order) polynomials, we get an ARX representation of the network in (6) where each row of Q and P can be identified independently in a parallel MISO setup. Parameterizing the coefficients of the impulse response of modules at each m^{th} row and representing the parameter vector as η^m_{ARX}, the parameter vector can be estimated using the prediction error method [17]. Considering the one-step ahead predictor [17],

$$\hat{w}_m(t|t-1; \eta^m_{ARX}) = \varphi^\top_m(t) \eta^m_{ARX} \tag{7}$$

with $\varphi^\top_m(t)$ representing the data matrix and the resulting prediction error $\varepsilon_m(t, \eta^m_{ARX}) = w_m(t) - \hat{w}_m(t|t-1; \eta^m_{ARX}) = w_m(t) - \varphi^\top_m(t) \eta^m_{ARX}$, the parameter vector can be consistently estimated [18] by minimizing the sum of squared prediction errors. This will result in a closed form solution,

$$\hat{\eta}^m_{ARX} = \left[\frac{1}{N} \sum_{t=1}^{N} \varphi_m(t) \varphi^\top_m(t) + R^m \right]^{-1} \cdot \frac{1}{N} \sum_{t=1}^{N} \varphi_m(t) w_m(t),$$

where $R^m = \text{diag}(R^m_1, \ldots, R^m_{|\mathcal{B}_j|})$ is the regularization matrix to handle the excessive variance of the estimate. The regularization term R^m_k for $k \in \mathcal{B}_j$ is chosen to be a modified Tuned/Correlated (TC) kernel [19] as it enforces stability. The modified TC kernel has the following structure

$$R^m_k = \text{diag}(1, \frac{1}{\alpha_k}, \ldots, \frac{1}{\alpha_k^{\ell_{ARX}-1}}). \tag{8}$$

Here, the coefficients α_k represent the decay rate of the impulse response of the corresponding module that are estimated by cross-validation [19] and ℓ_{ARX} is the length of the impulse response of the modules. An estimate of the innovation $\xi_{\mathcal{B}}(t)$ is obtained as $\hat{\xi}_{\mathcal{B}}(t) = \text{vec}\{\varepsilon_m(t, \hat{\eta}^m_{ARX})\}_{m=1,\cdots|\mathcal{B}_j|}$.

B. Structure selection

For estimating the local topology $(\mathcal{T}_{\tilde{G}}, \mathcal{T}_{\tilde{H}})$, we re-write (5) as,

$$w_{\mathcal{B}}(t) = \tilde{G} w_{\mathcal{B}}(t) + \bar{H} \xi_{\mathcal{B}}(t) + \xi_{\mathcal{B}}(t) \tag{9}$$

with $\bar{H} = (\tilde{H} - \mathcal{I})$ consisting of strictly proper modules. Motivated by the fact that $\hat{\xi}_{\mathcal{B}}(t) \to \xi_{\mathcal{B}}(t)$ w.p. 1 due to consistent estimation in the previous step, we now replace the term $\bar{H}\xi_{\mathcal{B}}(t)$ by $\bar{H}\hat{\xi}_{\mathcal{B}}(t)$, so that we can use $(w_{\mathcal{B}}, \hat{\xi}_{\mathcal{B}})$ as measured inputs in our topology estimation step.

Similar to the previous step, parameterizing the coefficients of the impulse response of the modules in $\tilde{G}$ and $\bar{H}$ at each m^{th} row of (9) and representing the parameter vector as η^m_{FIR}, the parameter vector of modules in each row can be estimated independently in parallel using the prediction error method [17] with one-step ahead predictor $\hat{w}_m(t|t-1;\eta^m_{FIR}) = \bar{\varphi}^\top_m(t)\eta^m_{FIR}$ where $\bar{\varphi}^\top_m$ is the data matrix. The parameter vector can be consistently estimated [18] by minimizing the sum of squared prediction errors $\varepsilon_m(t,\eta^m_{ARX}) = w_m(t) - \hat{w}_m(t|t-1;\eta^m_{ARX}) = w_m(t) - \bar{\varphi}^\top_m(t)\eta^m_{FIR}$. This will result in a closed form solution,

$$\hat{\eta}^m_{FIR} = \left[\frac{1}{N}\sum_{t=1}^{N}\bar{\varphi}_m(t)\bar{\varphi}^\top_m(t)\right]^{-1} \cdot \frac{1}{N}\sum_{t=1}^{N}\bar{\varphi}_m(t)w_m(t). \quad (10)$$

To identify the local topology $(\mathcal{T}_{\tilde{G}}, \mathcal{T}_{\bar{H}})$, we need to define a criterion that operates on the basis of data and provides an indirect measure of how close a certain candidate topology structure is to the true topology structure of the network in (9). In this regard, we consider a modified representation of the *Akaike Information Criterion (AIC)* [17]

$$J_{AIC}(\hat{\eta}^m_{FIR}) = \log\left(\frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\hat{\eta}^m_{FIR})\right) + \frac{2N_G\ell_{FIR}}{N}.$$

Here, N_G is the total number of nodes that is used as input for predicting node w_m in the candidate structure, ℓ_{FIR} is the length of the impulse response of the modules and is chosen long enough to sufficiently represent the dynamics of the network.

Having defined the selection criteria, we estimate $\hat{\eta}^m_{FIR}$ in an iterative manner for different topologies of $\tilde{G}$ and $\bar{H}$. The estimate $\hat{\eta}^m_{FIR}$ obtained from (10) is used to compute $\epsilon_m(t,\hat{\eta}^m_{FIR})$ which in turn is used to compute the cost of the selection criterion for this candidate structure. The ideal structure is the one that minimizes the selection criterion and to find it, a total $2^{|\mathcal{B}_j|}$ combinations need to be tested out. Testing these combinations becomes computationally infeasible as $|\mathcal{B}_j|$ increases. Developing search algorithms to reduce the number of combinations to test is a non-trivial problem. In the numerical simulation of Section VII a Focus search algorithm has been implemented, see [14] for more details.

VI. Identification algorithm

We now identify the target module which is one of the modules in $\tilde{G}$, based on the identified local topology $(\hat{\mathcal{T}}_{\tilde{G}}, \hat{\mathcal{T}}_{\bar{H}})$. For this we resort to the already established direct method for local module identification [6]. As indicated in Section III, this requires choosing the node signals in a predictor model (4) such that conditions on parallel paths and loops and on the absence of confounding variables are satisfied. An appropriate choice for a predictor model can be made based on the estimated topology $(\hat{\mathcal{T}}_{\tilde{G}}, \hat{\mathcal{T}}_{\bar{H}})$, resulting in either a MISO or a MIMO predictor model, in which the target module to be identified is embedded.

Identifying this MISO or MIMO model becomes computationally expensive as the number of modules in the predictor model increases. Also, parameterizing all the modules results in an explosion of nuisance parameters that affect the variance of the target module estimate, and complicates model order selection. As a solution, we consider the regularized kernel based methods, *Empirical Bayes Direct Method* (EBDM) of [7] for MISO structures and *Empirical Bayes Local Direct Method* (EBLDM) of [8] for MIMO structures. The problem of model order selection step is simplified by opting for a non-parametric impulse response model for all modules except the target module. To reduce the number of parameters, each impulse response is modelled as a Gaussian Process (GP) governed by a *Stable spline (SS) kernel* [19], being parametrized by only two hyperparameters that govern the SS kernel. The parameters are identified using an Empirical Bayes approach [20] by maximizing the marginal likelihood of data which inherently minimizes the mean square error of the estimation problem. This further reduces the variance of estimated parameters of the target module.

VII. Numerical simulation

A. Toplogy estimation

To highlight the effectiveness of the topology estimation algorithm, we generate 50 random stable networks (modules of 2nd order are randomly generated) with the network topology as shown in Figure 1. In this 10-node network, the process noise (v_1, v_8), (v_4, v_5), and (v_7, v_9) are pairwise correlated, while e is a Gaussian white noise process with $cov(e) = I$. We first identify the locality of the node 7, and

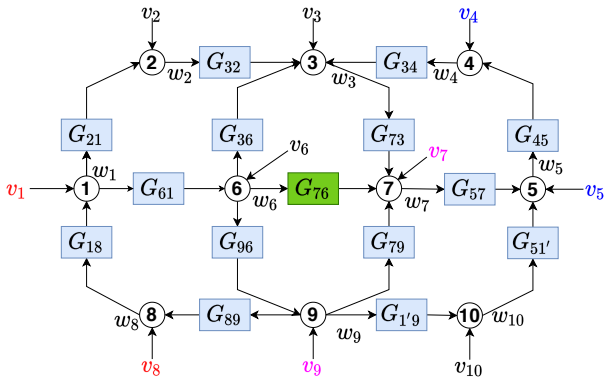

Fig. 1: 10-Node dynamic network with target module G_{76}. The noise processes (v_1, v_8), (v_4, v_5), and (v_7, v_9) are correlated.

immerse the remaining nodes. The Wiener filter length F is commonly chosen to be 20 across all the different networks; the threshold value ρ is chosen after visually examining the obtained Wiener filters. Following this, we identify the local topology with $\ell_{ARX} = 50$. The behaviour of this algorithm

is studied by comparing the *True Positive Rate (TPR)* and the *False Positive Rate (FPR)* over different choices of ℓ_{FIR}. The expressions for TPR and FPR are as follows

$$\text{TPR} = \frac{\text{TP}}{\text{P}}, \quad \text{FPR} = \frac{\text{FP}}{\text{A}}.$$

Here, P refers to the number of present interconnections in the network, TP refers to the number of instances of present interconnection identified by the algorithm, A refers to the number of absent interconnections in the network, and FP refers to the number of instances of absent interconnections falsely identified by the algorithm.

The obtained TPR vs FPR for $N = 1000$ over different $\ell_{FIR} = (10, 20, 40, 80)$ is shown in Figure 2. The quiver represent the direction of increasing ℓ_{FIR}. For a comparison with results for different selection criteria, the reader is referred to [14]. It can be seen from this Figure that as ℓ_{FIR} increases, the curves get closer to $(1, 0)$ until $\ell_{FIR} = 40$, being the typical length of the actual impulse responses in the network.

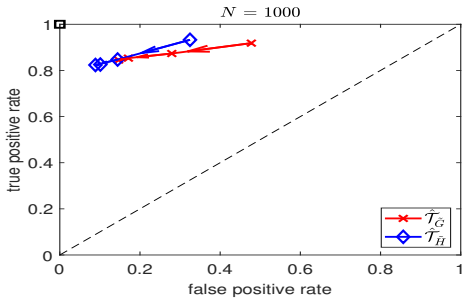

Fig. 2: TPR and FPR for nodal topology ($\hat{\mathcal{T}}_{\tilde{G}}$) and noise topology ($\hat{\mathcal{T}}_{\tilde{H}}$) using AIC as selection criterion, over different $\ell_{FIR} = (10, 20, 40, 80)$ for $N = 1000$. The quiver represents the direction of increasing ℓ_{FIR}.

B. Target module identification

To highlight the effectiveness of the estimation algorithm, we consider one of the 50 randomly generated networks for evaluating the performance of the complete algorithm from a parameter identification perspective. To prevent issues that might arise due to lack of sufficient excitation for target module identification, we add an additional white noise source with unit power to each node in the network. We run 50 independent Monte Carlo (MC) experiments with $N = 1000$ for different realizations of the noise source $e(t)$ while keeping the dynamics of the network fixed. The network dynamics for the MC simulations can be found in [14]. The target module is given by,

$$G_{76} = \frac{0.1050q^{-1} - 0.3465q^{-2}}{1 + 0.0480q^{-1} - 0.2534q^{-2}} = \frac{b_1 q^{-1} - b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}}.$$

For the first step of finding the locality of w_7, the non-causal Wiener filter length is chosen to be 20 while threshold ρ for identifying the locality is chosen to be 0.18. The true locality for this network is $\mathcal{B}_7 = \{3, 4, 5, 6, 9, 10\}$ and only this set of nodes should be non-causally Wiener correlated

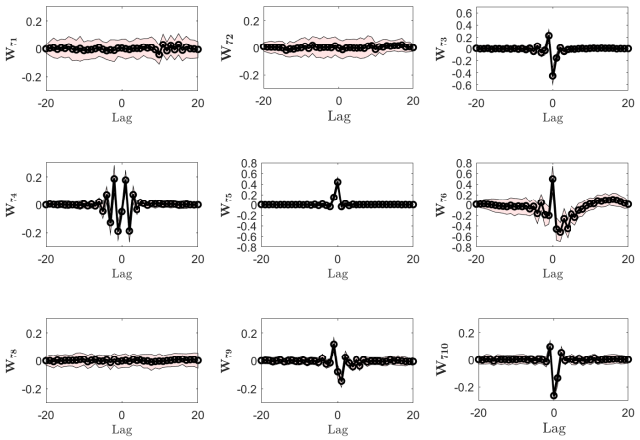

Fig. 3: Non-causal Wiener filter estimate $\hat{W}_7$ obtained by projecting w_7 on the remaining nodes in the network. Mean (black) and standard deviation (red) over 50 MC simulations.

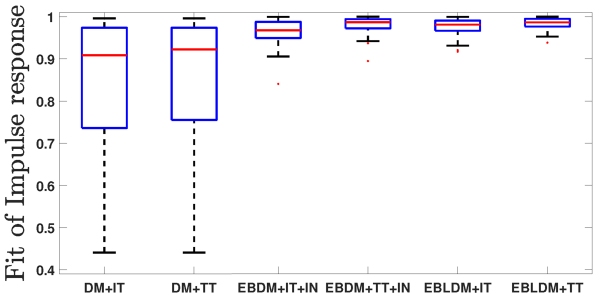

Fig. 4: Box plot of the fit of the impulse response of $\hat{G}_{76}$ obtained over 50 MC simulations for different identification methods

(i.e. $\|W_{7\star}\|_\infty > \rho$) to the node 7. The mean and standard deviation plots in Figure 3 confirm the sparsity results from Theorem 1.

Next, the local topology is identified by choosing $\ell_{ARX} = 50$ and $\ell_{FIR} = 40$ for the lengths of impulse response and the AIC selection criterion. This leads to a TPR of 0.8417 and FPR of 0.1175 over 50 MC runs.

To evaluate the performance of the identified topology ('IT') in estimating the target module G_{76}, we compare the following local module identification strategies:

1) **EBLDM+TT:** This is the method developed in [8]. We use the *Full input case* algorithm of local direct method [6] for predictor model selection considering the true local topology ('TT'). This will lead to a MIMO predictor model $\{w_6, w_3, w_9\} \rightarrow \{w_7, w_9\}$. We use this estimator as an upper bound of the performance to identify the target module.;
2) **EBLDM+IT:** This is the same estimator as the previous one, with the difference that we use the identified local topology ('IT') for obtaining the MIMO predictor model;
3) **DM+TT:** This is the standard direct method introduced in [1] for systems with uncorrelated disturbances, lead-

ing to a MISO predictor model $\{w_6, w_3, w_9\} \rightarrow \{w_7\}$;

4) **DM+IT:** This is the same estimator as the previous one, with the difference that the predictor model is formed with the identified topology;
5) **EBDM+TT+IN:** This is the method developed in [7] for uncorrelated disturbances. However in order to deal with correlated disturbances in the network, a MISO noise model is estimated by using the innovations estimates $\hat{\xi}_s$ as predictor input signals, and using the true topology of the network;
6) **EBDM+IT+IN:** This is same as the previous one but using the identified topology to form the predictor model.

The true model orders of all the modules in the network are assumed to be known. To evaluate the performance of the methods, we use the standard goodness-of-Fit metric,

$$\text{Fit} = 1 - \frac{\|g_{76} - \hat{g}_{76}\|_2}{\|g_{76} - \bar{g}_{76}\|_2}$$

where, g_{76} is the true value of the impulse response of G_{76}, $\hat{g}_{76}$ is the impulse response of the estimated $\hat{G}_{76}$ and $\bar{g}_{76}$ is the sample mean of g_{76}. The box plot showing the fit of the impulse response of G_{76} for the different methods is shown in Figure 4. It can be observed that the fit of the impulse response for the methods in which the identified topology is used, is similar to those where the true topology is used, implying that the use of the identified topology essentially preserves the performance. In Figure 4, we also see that the EBLDM and the EBDM have significantly better fit compared to the DM. The reduction in variance is attributed to the regularization approach followed in both the EBLDM and EBDM. Therefore, the developed approach for no prior topology information has an on par performance compared to the approaches that use the known true topology.

Remark 1: The example considered has a network of moderate size. However, when the size of the network increases, the locality of the output node does not necessarily expand. Additionally, even if the locality of the output node expands, the local topology estimation run in parallel due to its MISO formulation and the estimation steps have analytical solutions requiring no optimization. As a result, the developed algorithm will be less complex, faster and scalable to larger networks.

Remark 2: Step 2 acts as a stand-alone topology estimation procedure for the entire network when the set $\mathcal{B}_j$ is chosen to consist of all nodes in the network.

VIII. Conclusions

A novel approach for identifying a module embedded in a dynamic network effectively and efficiently without any prior topology information has been developed. The approach incorporates the estimation of the required prior knowledge on the network and noise topology into the identification framework, leading to a unified and complete data-driven approach for local module identification. Numerical simulations performed on fifty 10-node network examples with correlated process disturbances show promising results for the network and noise topology estimation, while the target module estimate has on par performance compared to approaches that use the known true topology.

References

[1] P. M. J. Van den Hof, A. G. Dankers, P. S. C. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods - basic methods for consistent module estimates," *Automatica*, vol. 49, no. 10, pp. 2994–3006, 2013.

[2] A. G. Dankers, P. M. J. Van den Hof, X. Bombois, and P. S. C. Heuberger, "Errors-in-variables identification in dynamic networks – consistency results for an instrumental variable approach," *Automatica*, vol. 62, pp. 39–50, 2015.

[3] A. G. Dankers, P. M. J. Van den Hof, P. S. C. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods: Predictor input selection," *IEEE Trans. on Automatic Control*, vol. 61, no. 4, pp. 937–952, 2016.

[4] M. Gevers, A. Bazanella, and G. Vian da Silva, "A practical method for the consistent identification of a module in a dynamical network," *IFAC-PapersOnLine*, vol. 51-15, pp. 862–867, 2018, proc. 18th IFAC Symp. System Identif. (SYSID2018).

[5] H. H. M. Weerts, J. Linder, M. Enqvist, and P. M. J. Van den Hof, "Abstractions of linear dynamic networks for input selection in local module identification," *Automatica*, vol. 117, no. 108975, July 2020.

[6] K. R. Ramaswamy and P. M. J. Van den Hof, "A local direct method for module identification in dynamic networks with correlated noise," *IEEE Trans. Automatic Control*, vol. 66, November 2021, to appear. [Online]. Available: https://arxiv.org/abs/1908.00976

[7] K. R. Ramaswamy, G. Bottegal, and P. M. J. Van den Hof, "Learning linear models in a dynamic network using regularized kernel-based methods," *Automatica*, vol. 129, no. 109591, July 2021.

[8] V. C. Rajagopal, K. R. Ramaswamy, and P. M. J. Van den Hof, "A regularized kernel-based method for learning a module in a dynamic network with correlated noise," in *Proc. 59th IEEE Conf. on Decision and Control (CDC)*. Jeju Island, Korea: IEEE, 2020, pp. 4348–4353.

[9] D. Materassi and M. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the Wiener filter," *IEEE Trans. Automatic Control*, vol. 57, no. 7, pp. 1765–1777, 2012.

[10] P. M. J. Van den Hof, K. R. Ramaswamy, A. G. Dankers, and G. Bottegal, "Local module identification in dynamic networks with correlated noise: the full input case," in *Proc. 58th IEEE Conf. on Decision and Control (CDC)*, Nice, France, 2019, pp. 5494–5499.

[11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[12] A. Dankers, P. M. J. Van den Hof, X. Bombois, and P. S. C. Heuberger, "Identification of dynamic models in complex networks with prediction error methods: Predictor input selection," *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 937–952, April 2016.

[13] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. New Jersey: Prentice Hall, 2000.

[14] V. C. Rajagopal, "Learning local modules in dynamic networks without prior topology information," Master's thesis, Eindhoven University of Technology, October 2020.

[15] S. Talukdar, D. Deka, B. Lundstrom, M. Chertkov, and M. V. Salapaka, "Learning exact topology of a loopy power grid from ambient dynamics," in *Proceedings of the Eighth International Conference on Future Energy Systems*, ser. e-Energy '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 222–227.

[16] H. H. M. Weerts, M. Galrinho, G. Bottegal, H. Hjalmarsson, and P. M. J. Van den Hof, "A sequential least squares algorithm for armax dynamic network identification," *IFAC-PapersOnLine*, vol. 51-15, pp. 844–849, 2018, proc. 18th IFAC Symp. System Identification.

[17] L. Ljung, *System Identification : Theory for the User*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.

[18] L. Ljung and B. Wahlberg, "Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra," *Advances in Applied Probability*, vol. 24, no. 2, pp. 412–440, 1992.

[19] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and gaussian processes—revisited," *Automatica*, vol. 48, no. 8, pp. 1525 – 1535, 2012.

[20] J. S. Maritz and T. Lwin, *Empirical Bayes Methods*. Chapman and Hall, 1989.