TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

A scalable multi-step least squares method for network identification with unknown disturbance topology

Citation for published version (APA):
Fonken, S. J. M., Ramaswamy, K., & Van den Hof, P. M. J. (2022). A scalable multi-step least squares method for network identification with unknown disturbance topology. *Automatica*, *141*, Article 110295. https://doi.org/10.1016/j.automatica.2022.110295

Document license:
CC BY

DOI:
10.1016/j.automatica.2022.110295

Document status and date:
Published: 01/07/2022

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

Automatica 141 (2022) 110295

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

A scalable multi-step least squares method for network identification with unknown disturbance topology☆

Stefanie J.M. Fonken *, Karthik Raghavan Ramaswamy, Paul M.J. Van den Hof

Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 14 June 2021
Received in revised form 23 November 2021
Accepted 28 January 2022
Available online xxxx

Keywords:
System identification
Dynamic networks
Estimation algorithms
Least squares
Topology estimation

ABSTRACT

Identification methods for dynamic networks typically require prior knowledge of the network and disturbance topology, and often rely on solving poorly scalable non-convex optimization problems. While methods for estimating network topology are available in the literature, less attention has been paid to estimating the disturbance topology, i.e., the (spatial) noise correlation structure and the noise rank in a filtered white noise representation of the disturbance signal. In this work we present an identification method for dynamic networks, in which an estimation of the disturbance topology precedes the identification of the full dynamic network with known network topology. To this end we extend the multi-step Sequential Linear Regression and Weighted Null Space Fitting methods to deal with reduced rank noise, and use these methods to estimate the disturbance topology and the network dynamics in the full measurement situation. As a result, we provide a multi-step least squares algorithm with parallel computation capabilities and that rely only on explicit analytical solutions, thereby avoiding the usual non-convex optimizations involved. Consequently we consistently estimate dynamic networks of Box Jenkins model structure, while keeping the computational burden low. We provide a consistency proof that includes path-based data informativity conditions for allocation of excitation signals in the experimental design. Numerical simulations performed on a dynamic network with reduced rank noise clearly illustrate the potential of this method.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Dynamic networks represent large-scale interconnected systems, and data-driven modeling of dynamic networks has received considerable attention in recent years. These networks can be considered as a set of measurable (node) signals interconnected through linear dynamic systems (the modules), driven by measured external excitation signals and/or unmeasured disturbance signals. Modeling of these networks plays an important role in biological systems (Hagmann et al., 2008; Hickman et al., 2017), economic systems (Materassi & Innocenti, 2010), power networks (Pagani & Aiello, 2013), and many other fields in science and engineering. The challenges addressed in identification of dynamic networks can roughly be divided into three categories. The first is identifying the interconnection structure of the nodes in a dynamic network referred to as network topology detection (Chiuso & Pillonetto, 2012; Materassi & Innocenti, 2010). The second is the identification of a specific module in a network, referred to as local module identification. For this problem closed-loop identification methods have been generalized to the dynamic network situation in Van den Hof, Dankers, Heuberger, and Bombois (2013), formulating the local module identification problem as a multi-input–single-output (MISO) problem. This has been further extended and generalized in e.g., Dankers, Van den Hof, Heuberger, and Bombois (2016), Dankers, Van den Hof, Bombois, and Heuberger (2015), Everitt, Galrinho, and Hjalmarsson (2018), Gevers, Bazanella, and Vian da Silva (2018), Materassi and Salapaka (2020), Ramaswamy, Bottegal, and Van den Hof (2021), Ramaswamy and Van den Hof (2021) and Van den Hof and Ramaswamy (2021). The third challenge is identification of the full network dynamics (Dankers, 2019; Fonken, Ferizbegovic, & Hjalmarsson, 2020; Weerts, Galrinho, Bottegal, Hjalmarsson, & Van den Hof, 2018; Weerts, Van den Hof, & Dankers, 2018b), where the problem is formulated as the identification of a (structured) multi-input–multi-output (MIMO) model.

In this paper we will further explore the development of full network identification methods. While dynamic networks

☆ This project 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). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: s.j.m.fonken@tue.nl (S.J.M. Fonken), k.r.ramaswamy@tue.nl (K.R. Ramaswamy), p.m.j.vandenhof@tue.nl (P.M.J. Van den Hof).

https://doi.org/10.1016/j.automatica.2022.110295
0005-1098/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

increase in complexity and size, and measurement data is becoming increasingly accessible, there is a strong demand for accurate and scalable data driven modeling methods. The joint direct method (Weerts, Van den Hof, & Dankers, 2017; Weerts et al., 2018b) predicts all node signals in the network jointly and achieves consistency and minimum variance properties in the situation that the network and disturbance topology are given a priori and the noise can be of reduced rank. However it strongly relies on solving (constrained) non-convex optimization problems, which seriously limits its scalability to larger networks. There are multi-step convex identification methods available for full network identification, such as the Sequential Linear Regression (SLR) (Dankers, 2019), Sequential Least Squares (SLS) (Weerts et al., 2018) and extensions of Weighted Null Space Fitting (WNSF) (Galrinho, Rojas, & Hjalmarsson, 2019) such as Fonken et al. (2020). Moreover, methods such as the SLR and SLS allow for splitting the MIMO optimization into multiple linear regressions, which contributes to a lower computational burden. The available convex methods are scalable to larger networks, but are limited to particular model structures of the network, and additionally, they do not allow for handling reduced rank noise. Particularly in large-scale network identification, stepping away from the typical assumption that all disturbance signals have their own independent noise source, is an appealing situation that should be supported by an effective estimation algorithm. Handling this situation of reduced-rank noise can substantially reduce the variance of estimated models. However it also introduces the problems of estimating the noise rank and noise correlation structure from data.

All available convex and non-convex methods for network identification require prior knowledge on the topology (i.e. rank and spatial correlation structure of the disturbance model). While in dynamic factor analysis (Deistler, Scherrer, & Anderson, 2015) attention has been paid to the estimation of noise rank, in prediction error identification this does not appear to be included yet in the identification algorithms. For situations where the disturbance topology information is not readily available, it is attractive to develop methods that include estimating this information from data.

The topology estimation literature shows a variety of available methods to estimate the topology, such as Wiener filter based methods (Materassi & Innocenti, 2010; Materassi & Salapaka, 2012; Materassi, Salapaka, & Giarrè, 2011), Bayesian model selection techniques (Chiuso & Pillonetto, 2012; Shi, Bottegal, & Van den Hof, 2019; Wasserman, 2000), or methods that infer the topology from parametric estimates (Bolstad, Van Veen, & Nowak, 2011; Dankers, Van den Hof, Heuberger, & Bombois, 2012; Yuan, Stan, Warnick, & Gonçalves, 2011). While the main focus of topology detection literature has been on estimating network topology in the situation of a diagonal disturbance spectrum $\Phi_v(\omega)$, extensions towards nondiagonal spectra have been presented in Bombois and Hjalmarsson (2021), Dimovska and Materassi (2017) and Veedu and Salapaka (2020). In Veedu and Salapaka (2020) network topology and the non-zero pattern in the disturbance spectrum are estimated jointly. In this paper we assume that we do not know the disturbance topology a priori, but we assume that the network topology is known e.g., from its underlying physics, which is commonly the case for engineered systems. In the situation that the network topology is not known beforehand, it is possible to use any of the above cited methods to estimate it. We allow the process noise to be spatially correlated, i.e. the disturbance spectrum $\Phi_v(\omega)$ is not necessarily diagonal. Additionally the noise is allowed to be of reduced rank, i.e. $\Phi_v(\omega)$ can be singular.

The objective is to develop a multi-step convex algorithm that estimates the disturbance topology and the dynamic modules in the network for general model structures including the Box Jenkins (BJ) structure, while adhering to computational algorithms that are scalable, while achieving favorable properties in terms of low experiment cost, consistency and reduced variance of the network estimates.

To this end we develop a multi-step algorithm to identify the network dynamics. In the first step the noise rank and the nonzero pattern in the corresponding disturbance model (noise shaping filter) are estimated. This is done through a (nonparametric) high-order ARX model, inspired by the SLR method (Dankers, 2019). Next, this information is used to develop a multi-step convex algorithm that can accurately identify the dynamics of the network in the situation of reduced rank noise and for a very general Box Jenkins model structure, thereby combining the recently introduced multi-step convex identification methods SLR (Dankers, 2019) and WNSF (Fonken et al., 2020; Galrinho et al., 2019) and extending them to the described situation.

The paper proceeds with a definition of the considered dynamic network setup in Section 2. In Section 3 we present a new method for estimating the disturbance topology from data, followed in Section 4 by a multi-step identification algorithm that exploits the prior estimated disturbance topology. Section 5 presents the consistency analysis of the method, including graph-based conditions for data informativity. Results of numerical simulations are provided in Section 6, followed by conclusions in Section 7. The consistency proofs are collected in Appendix.

2. Dynamic networks

Following the setting of Van den Hof et al. (2013) a dynamic network is defined by L nodes or internal variables $w_j(t)$, $j = 1, \dots, L$, that are scalar-valued measured signals. The underlying network is linear time invariant (LTI), and the nodes of the network can be expressed as

$$w_j(t) = \sum_{l \in \mathcal{N}_j} G^0_{jl}(q) w_l(t) + \sum_{k \in \mathcal{R}_j} R^0_{jk}(q) r_k(t) + v_j(t), \tag{1}$$

where

- q^{-1} the delay operator, i.e. $q^{-1} w_j(t) = w_j(t-1)$,
- $\mathcal{N}_j$ defines the set of indices of measured node signals w_l, $l \neq j$, for which $G^0_{jl}(q) \neq 0$, where $G^0_{jl}(q)$ is a strictly proper rational transfer function,
- $\mathcal{R}_j$ defines the set of indices of measured external excitation signals r_k, for which $R^0_{jk}(q) \neq 0$, where $R^0_{jk}(q)$ is a known proper rational transfer function,
- $v_j(t)$ is unmeasured process noise, where the disturbance vector $v = [v_1 \cdots v_L]^\top$ is modeled as a wide sense stationary stochastic process represented by $v(t) = H^0(q)e(t)$. The $e = [e_1 \cdots e_p]^\top$ is a white noise process of dimension $p \leq L$ with covariance matrix $\Lambda^0 > 0$. $H^0(q)$ is a rational transfer function matrix.

The full network expression, with omitted q and t, is

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G^0_{12} & \cdots & G^0_{1L} \\ G^0_{21} & 0 & \ddots & G^0_{2L} \\ \vdots & \ddots & \ddots & \vdots \\ G^0_{L1} & G^0_{L2} & \cdots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + R^0 \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_K \end{bmatrix} + H^0 \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix} \tag{2}$$

with the matrix notation given by

$$w = G^0 w + R^0 r + H^0 e, \tag{3a}$$

$$w = (I - G^0)^{-1}(R^0 r + H^0 e), \tag{3b}$$

where we assume that the inverse $(I - G^0)^{-1}$ exists and the network is well-posed, as used in Van den Hof et al. (2013).

In the situation $p < L$, i.e. when the noise is of reduced rank or singular, the disturbance model H^0 is a non-square matrix, i.e.

- $H^0 \in \mathbb{R}^{L\times p}(z)$ is stable and has a stable left inverse $H^\dagger$ that satisfies $H^\dagger H = I \in \mathbb{R}^{p\times p}$;

For a unique representation of reduced rank spectra that can be used to construct a predictor we can adopt a result from Weerts et al. (2018b) where the disturbance term is equivalently written as $\breve{H}^0\breve{e}$ with $\breve{H}^0$ square.

Lemma 1 (*Weerts et al., 2018b*). *Consider an L-dimensional disturbance process v with rank p. Then the disturbance signals v can be reordered in such a way that the following unique representations result:*

$$\begin{bmatrix} v_a \\ v_b \end{bmatrix} = H^0 e = \breve{H}^0\breve{e} \quad \text{with}$$

$$H^0 = \begin{bmatrix} H_a^0 \\ H_b^0 \end{bmatrix},\ \breve{H}^0 = \begin{bmatrix} H_a^0 & 0 \\ H_b^0 - \Gamma^0 & I \end{bmatrix},\ \breve{e} = \begin{bmatrix} \breve{e}_a \\ \breve{e}_b \end{bmatrix} = \begin{bmatrix} e \\ \Gamma^0 e \end{bmatrix} \tag{4}$$

and $\Gamma^0 = lim_{z\to\infty} H_b^0(z)$

such that

- *$H_a^0 \in \mathbb{R}^{p\times p}(z)$ is a monic full rank rational transfer function matrix;*
- *$H_b^0 \in \mathbb{R}^{(L-p)\times p}(z)$ is a stable proper rational transfer function matrix.*
- *The covariance matrix of $\breve{e}$ is given by,*

$$\breve{\Lambda}^0 = \begin{bmatrix} I \\ \Gamma^0 \end{bmatrix} \Lambda^0 \begin{bmatrix} I \\ \Gamma^0 \end{bmatrix}^\top = \begin{bmatrix} \Lambda^0 & \Lambda^0\Gamma^{0\top} \\ \Gamma^0\Lambda^0 & \Gamma^0\Lambda^0\Gamma^{0\top} \end{bmatrix}, \tag{5}$$

  where $\Lambda^0 \in \mathbb{R}^{p\times p}$ has rank p.
- *If additionally H_a^0 is minimum phase then $\breve{H}^0$ is monic, stable and minimum phase.* [1] □

The result of the reordering of signals as indicated in the Lemma is that the first p components of the reordered signal constitute a full rank p process.

We assume that the data generating network satisfies the following properties.

Assumption 1.

a. The network is well-posed, i.e. all principle minors of $\left(I - G^0(\infty)\right)$ are nonzero (Araki & Saeki, 1983).
b. $(I - G^0)^{-1}$ is stable and causal.
c. All elements in $G(q)$ are strictly proper.
d. H^0 is stable and has a stable left inverse.
e. $\breve{H}^0$ is square, monic and minimum phase.
f. The topology of G^0 and R^0, and the non-zero elements of R^0 are fixed and known.
g. The matrix R^0 has a block diagonal structure: $R^0 = diag(R_a^0, R_b^0)$ in the situation of ordered nodes as meant in (4).
h. Measurements of all node signals w and all present excitation signals r are available.
i. The standard regularity conditions on the data are satisfied that are required for consistency results of the prediction error identification method.[2]

[1] It has recently been pointed out in Cao, Picci, and Lindquist (2021) that this excluded the situation where the (deterministic) mapping from v_a to v_b is unstable.

[2] See Ljung (1999) page 249. This includes the property that $e(t)$ has bounded moments of order higher than 4.

The two main steps of the identification method that will be developed in this paper are

- Estimating the disturbance topology, i.e. the noise rank and the zero pattern in the disturbance model.
- Estimating the dynamical components in the network for a given network and disturbance topology, while using a parametric BJ model structure.

In the next section we first focus on the disturbance topology estimation method, followed by the developed identification method in the section thereafter.

3. Disturbance topology estimation

Before we can use a unique disturbance model that is structured according to $\breve{H}^0$ in (4), we need to estimate the noise rank p and we need to be able to reorder the node signals in such a way that a noise representation as in (4) can be used. This step is necessary as the unstructured disturbance model H^0 is non-unique in the situation $p < L$. Therefore the disturbance topology estimation is performed in two main steps:

- Step 1: Estimating the noise rank, and reordering the signals to the situation of Lemma 1.
- Step 2: Estimating the structure of the disturbance model $\breve{H}^0$.

3.1. Step 1: Estimating noise rank p and reordering of nodes

For estimating the noise rank p, we are going to estimate the covariance matrix $\breve{\Lambda}^0$(5) of innovation signal $\breve{e}$, which through its rank p can provide us access to the correct noise rank.

An estimate of the covariance matrix is obtained by estimating a high-order (nonparametric) ARX model on the basis of measured signals w, r, and by using the residual (predictor error) of this estimated model as an estimate of the white noise term $\breve{e}$.

A parametrized ARX model is chosen according to

$$\breve{A}(q,\zeta) = I + \breve{A}_1 q^{-1} + \cdots + \breve{A}_n q^{-n} \tag{6}$$

$$\breve{B}(q,\zeta) = \breve{B}_0 + \breve{B}_1 q^{-1} + \cdots \breve{B}_{n-1} q^{-(n-1)} \tag{7}$$

while all coefficients of $\breve{A}_k, \breve{B}_k$ are vectorized and collected in the parameter vector ζ. The one-step-ahead predictor Ljung (1999), defined as

$$\hat{w}(t|t-1;\zeta) := \mathbb{E}\{w(t)|w^{t-1}, r^t\}, \tag{8}$$

where w^{t-1} and r^t are defined according to $w^{t-1} := \{w(0), w(1), \ldots, w(t-1)\}$ and $r^t := \{r(0), r(1), \ldots, r(t)\}$, is given by

$$\hat{w}(t|t-1,\zeta) = \left(I - \breve{A}(q,\zeta)\right)w(t) + \breve{B}(q,\zeta)r(t) \tag{9}$$

$$= \varphi(t)\zeta \tag{10}$$

with $\varphi(t)$ composed of the appropriate terms in w and r.

Note that for an actual network with representation $G^0, \breve{H}^0, R^0$, the one-step predictor will be given by

$$\hat{w}(t|t-1) = \left(I - (\breve{H}^0(q))^{-1}(I - G^0(q))\right)w(t) + (\breve{H}^0(q))^{-1}R^0(q)r(t). \tag{11}$$

This implies that the polynomial predictor model (9) can only accurately approximate the rational filters that are present in (11) if the ARX order n is chosen very high. The ARX model is estimated according to $\hat{\zeta}_N^n = \arg\min_\zeta \frac{1}{N}\sum_{t=1}^N \varepsilon^T(t,\zeta)\varepsilon(t,\zeta)$, with $\varepsilon(t,\zeta) = w(t) - \hat{w}(t|t-1;\zeta)$, leading to the analytical solution

$$\hat{\zeta}_N^n = \left[\frac{1}{N}\sum_{t=1}^N \varphi(t)\varphi^\top(t)\right]^{-1} \frac{1}{N}\sum_{t=1}^N \varphi(t)w(t). \tag{12}$$

Since the network identifiability conditions of Weerts, Van den Hof, and Dankers (2018a) are satisfied for the considered model set, the sample estimate

$$\hat{\Lambda} := \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t, \hat{\zeta}_N^n)\varepsilon^\top(t, \hat{\zeta}_N^n), \tag{13}$$

will then, under mild regularity conditions, be a consistent estimate of the noise covariance $\breve{\Lambda}^0$. The rank p of the noise process can then be estimated through a rank test on $\hat{\Lambda}$, e.g., through a singular value decomposition. Alternatively, other matrix factorizations or information based criteria can be applied for estimating the rank, see e.g., Camba-Méndez and Kapetanios (2009). When $\hat{\Lambda}$ and the estimated rank $\hat{p} < L$ have been determined, the L signals can be reordered through a permutation matrix Π such that the first $\hat{p}$ components of the permuted noise vector have a rank $\hat{p}$ covariance matrix, i.e. $\begin{bmatrix} I_{\hat{p}} & 0 \end{bmatrix} \Pi^\top \hat{\Lambda} \Pi \begin{bmatrix} I_{\hat{p}} & 0 \end{bmatrix}^\top$ has rank $\hat{p}$.

Remark 1. Since the polynomials $\breve{A}(\zeta)$ and $\breve{B}(\zeta)$ are fully parametrized with independent parameters on each polynomial entry, the MIMO least squares optimization that leads to the solution (12) can also be decomposed in L separate linear regressions that minimize the residual $\varepsilon_j(t, \zeta)$ separately for each j, which is computationally attractive since the computations can be performed in parallel or sequentially.

Remark 2. The resulting estimation scheme will generally not provide us with consistent estimates of the ARX model. This is not only due to the fact that typically the order n of the ARX model would need to go to infinity, but also to the fact that the solution for $\hat{\zeta}_N^n$ is non-unique in the situation $p < L$. However, this latter non-uniqueness does not affect the uniqueness and whiteness of the residual $\varepsilon(t, \hat{\zeta}_N^n)$ since, according to the projection theorem, every solution for $\hat{\zeta}_N^n$ determines the same predictor (Deistler, Anderson, Filler, Zinner, & Chen, 2010). The estimate $\hat{\Lambda}$ is therefore consistent, i.e. $\hat{\Lambda} = cov(\breve{e})$ w.p. 1 as $n, N \to \infty$.

Remark 3. Although a correct estimation of the noise rank p cannot be guaranteed, consistency results for estimating p would be possible when applying information-based criteria for rank estimation, e.g., based on the BIC criterion (Camba-Méndez & Kapetanios, 2009). In the next steps of our approach it will be assumed that a correct estimation of p has been obtained.

After reordering the node signals as described above, we can now adhere to a network representation with a unique disturbance model according to the structure in Lemma 1, where $\breve{H}^0$ can be parametrized by the transfer function matrices H_a and H_b.

3.2. Step 2: Estimating the noise correlation structure

In the second step we are going to estimate which entries in our disturbance model are nonzero. To this end we extend the SLR method (Dankers, 2019) to the situation of reduced rank noise and show how the noise correlation structure can be obtained.

3.2.1. Step 2.1: Refining the nonparametric ARX model

With the noise rank p available and the nodes being ordered, we have gained additional information on $\breve{H}^0$ (4), namely the last $L - p$ columns are now known. Now, we perform the same approach of identification using high order ARX modeling as in the previous step, but by utilizing the known entries in $\breve{H}^0$, leading to refined estimates of $\breve{A}(\hat{\zeta}_N^n)$ and $\breve{B}(\hat{\zeta}_N^n)$. In the analysis results of Section 5.1 it shown that the known entries in $\breve{H}^0$ can simply be mapped to known entries in the parametrized polynomial $\breve{B}(\zeta)$, and therefore can simply be taken into account in the least squares problem (12). In Section 5.1 it is shown that this leads to consistent estimates $\hat{\zeta}_N^n$ for $n, N \to \infty$.

3.2.2. Step 2.2: Predictor model with reconstructed innovation input

In this step we are going to use the estimated nonparametric ARX model to reconstruct the innovation signal. This allows us to use the reconstructed innovation signal as a measured input in the predictor model that will be used for estimating the structure of the disturbance model.

If there exists a parameter ζ^0 such that the ARX model $(\breve{A}(\zeta^0), \breve{B}(\zeta^0))$ captures the dynamics of the network, then it follows from Weerts et al. (2018b) that

$$\varepsilon(t, \zeta^0) = \begin{bmatrix} I \\ \Gamma^0 \end{bmatrix} e(t). \tag{14}$$

We can accordingly decompose $\varepsilon(t, \zeta)$ as

$$\varepsilon(t, \zeta) = \begin{bmatrix} \varepsilon_a(t, \zeta) \\ \varepsilon_b(t, \zeta) \end{bmatrix} \tag{15}$$

while the consistency property of $\hat{\zeta}_N^n$ implies that

$$\begin{aligned} &\varepsilon_a(t, \hat{\zeta}_N^n) \to e(t) \qquad \text{w.p. 1 as } N \to \infty \,\forall t, \\ &\varepsilon_b(t, \hat{\zeta}_N^n) \to \Gamma^0 e(t) \quad \text{w.p. 1 as } N \to \infty \,\forall t. \end{aligned} \tag{16}$$

We will refer to $\varepsilon(t, \hat{\zeta}_N^n)$ as the "reconstructed innovation".

For a network with ordered nodes we evaluate a new one-step-ahead predictor

$$\hat{w}(t|t-1) := \mathbb{E}\{w(t)|w^{t-1}, r^t, e^{t-1}\} \tag{17}$$

that includes the innovation signal $e^{t-1} := \{e(0), e(1), \cdots, e(t-1)\}$ in the expectation. Then it follows that

$$\hat{w}(t|t-1) = G^0(q)w(t) + (\breve{H}^0(q) - I)\breve{e}(t) + R^0(q)r(t), \tag{18}$$

where

$$(\breve{H}^0 - I)\breve{e} = \left(\begin{bmatrix} H_a^0 & 0 \\ H_b^0 - \Gamma^0 & I \end{bmatrix} - I\right)\breve{e} = \begin{bmatrix} H_a^0 - I \\ H_b^0 - \Gamma^0 \end{bmatrix} e = \bar{H}^0 e. \tag{19}$$

This motivates the use of the following parametrized predictor model per node:

$$\begin{aligned} &\hat{w}_j(t|t-1, \eta_j) = \\ &\sum_{l \in \mathcal{N}_j} G_{jl}(\eta_j) w_l + \sum_{s \in \mathcal{V}_j} \bar{H}_{js}(\eta_j)\varepsilon_{a_s}(\hat{\zeta}_N^n) + \sum_{k \in \mathcal{R}_j} R_{jk} r_k, \end{aligned} \tag{20}$$

where the terms $G(\eta)$ and $\bar{H}(\eta)$ are parametrized versions of G^0 and $\bar{H}^0$ respectively, and $\varepsilon_a(\hat{\zeta}_N^n)$ is an estimate of the noise signal $e(t)$. $G_{jl}(\eta) = \sum_{k=1}^{n} g_k^{jl} q^{-k}$ and $\bar{H}_{js}(\eta) = \sum_{k=1}^{n} h_k^{js} q^{-k}$ are parametrized as strictly proper polynomials of order n, the term $\sum_{k \in \mathcal{R}_j} R_{jk} r_k(t)$ is known, the sets $\mathcal{N}_j$ and $\mathcal{R}_j$ are known from the topology of G^0 and R^0, and $\mathcal{V}_j$ defines the set of indices of noise signals for which noise dynamics is present in the disturbance model. This leads to an ARX model, like in Step 1, but now with the reconstructed innovation $\varepsilon_a(t, \hat{\zeta}_N^n)$ added as external predictor input signal, and the coefficients of the unknown polynomials collected in the parameter vector η. It is our next objective now to determine the sets $\mathcal{V}_j$ for $j = 1, \ldots, L$. To this end we follow two approaches namely the structure selection approach and the Glasso approach, which will be presented next.

3.2.3. Structure selection

For a particular choice of $\mathcal{V}_j$ we evaluate the residual $\varepsilon_j(t, \hat{\eta}^n_{N_j}) := w_j(t) - \hat{w}_j(t|t-1, \hat{\eta}^n_{N_j})$ where $\hat{\eta}^n_{N_j}$ is the estimated parameter that minimizes the quadratic criterion $\frac{1}{N}\sum_{t=1}^{N} \varepsilon_j^2(t, \eta_j)$, and that is obtained through an analytical solution, similar to (12). We test this residual with possible combinations in set $\mathcal{V}_j$ and employ model selection techniques such as AIC, BIC and Cross-validation (CV) on the obtained estimates $\hat{\eta}^n_{N_j}$ (Yuan et al., 2011), of which the BIC provides a consistent estimate (Kass & Raftery, 1995; Schwarz, 1978). Because we use ARX models to estimate η, model selection techniques such as AIC, BIC and CV are convex. Additionally, since we derive the disturbance topology per node, we have to test at most 2^L possible sets $\mathcal{V}_j$ for L nodes. This results in a lower computational burden compared to when we detect the topology in a MIMO setting, where we would have to test at most 2^{L^2-L} possible sets $\mathcal{V}_j$ simultaneously for all j (Yuan et al., 2011). However, for large networks these model selection techniques can still become computationally heavy.

3.2.4. Sparse estimation with Glasso

For each node j, a Glasso (Group Lasso) estimate is computed by minimizing the following cost function over η_j for a fully parametrized disturbance model with p white noise inputs:

$$\min_{\eta_j} \left\{ \frac{1}{2} \sum_{t=1}^{N} (w_j(t) - \hat{w}_j(t|t-1, \eta_j))^2 + \lambda_j \cdot \|\eta_j\|_2 \right\} \tag{21}$$

with the one-step-ahead predictor (20), and η_j being the vector of parameters related to the modules G_{ji} for $i \in \mathcal{N}_j$, and related to the modules $\bar{H}_{js}$ for $s = 1, \ldots, p$; λ_j is the tuning parameter (penalization factor) of Glasso. The tuning of λ_j is described in the numerical illustrations in Section 6.

The right hand side of (21) is a mixed l_1/l_2 norm. The Glasso estimate is a convex extension to lasso that penalizes groups of estimated parameters (Yuan & Lin, 2006), imposing sparsity at group level. Within a group, it does not yield sparsity (Bach, Jenatton, Mairal, Obozinski, et al., 2011). If an appropriate penalization factor is chosen, only the dynamic modules that are actually present in the data generating network remain while the non-present terms are forced to 0, thus providing an estimate of the structure of $\bar{H}$.

With either of the methods of Sections Section 3.2.3 or 3.2.4 the structure $\mathcal{V}_j$ of the disturbance model can be estimated entirely with convex and thus scalable methods, employing non-parametric (high-order ARX-) models. This structural information can be effectively used in the actual estimation of parametric dynamic models in the next Section.

Remark 4. It is possible to add regularization when estimating the high-order ARX models presented in this section to guarantee stability of the estimates.

4. Estimating parametric network models

The next step in our identification procedure is

- Step 3: Estimating a parametric network model.

While in Step 1 and 2 high-order (nonparametric) models of the same model order n are used, and thus providing estimates with relatively high variance, in this step a parametric model is estimated from data where we exploit a very flexible Box–Jenkins model structure. In Step 3 we extend the WNSF method (Galrinho et al., 2019), and its application to dynamic networks in Fonken et al. (2020), to the reduced rank noise case such that we are able to obtain parametric models $G(\theta)$ and $H(\theta)$. The WNSF is in itself a three step method that starts with a high-order model before estimating the parametric model.

4.1. Step 3.1: Refining the nonparametric model

By fixing the correctly estimated disturbance topology obtained in the previous section we obtain consistent estimates of η_j using one-step-ahead predictor (20) defined in (17), leading to a high-order ARX model with structured disturbance model. The conditions for consistency of $\hat{\eta}^n_{j_N}$ are derived in Section 5. By employing the structured disturbance model we reduce the variance of $\hat{\eta}^n_{j_N}$, while the model order n remains the same.

Using the consistent estimate $\hat{\eta}^n_{j_N}$, we update the reconstructed innovation. Subsequently, we again update the high-order ARX model by replacing $\varepsilon_a(\check{\zeta}^n_{j_N})$ with the updated reconstructed innovation $\varepsilon_a(\hat{\eta}^n_{j_N})$ in (20), and use this updated predictor to re-estimate η_j. This latter estimate can be seen as the starting high-order model for the WNSF method. At this point we still have a high variance on the estimates of η but negligible bias if model order n throughout all the steps is chosen sufficiently large. In the next step we reduce the variance by reducing the number of parameters to estimate, where we will make the step from a high-order (nonparametric) model to a parametric model.

4.2. Step 3.2: Parametric model estimate

On the basis of the nonparametric model estimate characterized by $\hat{\eta}^n_{j_N}$ we are now going to estimate a parametric model of the dynamic network by utilizing a Box Jenkins model structure:

$$\begin{aligned}
G_{jl}(q,\theta) &= \frac{l_1^{jl} q^{-1} + \cdots + l_{m_l}^{jl} q^{-m_l}}{1 + f_1^{jl} q^{-1} + \cdots + f_{m_f}^{jl} q^{-m_f}}, \\
H_{jj}(q,\theta) &= \frac{1 + c_1^{jj} q^{-1} + \cdots + c_{m_c}^{jj} q^{-m_c}}{1 + d_1^{jj} q^{-1} + \cdots + d_{m_d}^{jj} q^{-m_d}}, \\
H_{js}(q,\theta) &= \frac{c_1^{js} q^{-1} + \cdots + c_{m_c}^{js} q^{-m_c}}{1 + d_1^{js} q^{-1} + \cdots + d_{m_d}^{js} q^{-m_d}}, \quad s \neq j
\end{aligned} \tag{22}$$

that can be rewritten as

$$G_{jl}(q,\theta) = \frac{L_{jl}(q,\theta)}{F_{jl}(q,\theta)}, \quad H_{js}(q,\theta) = \frac{C_{js}(q,\theta)}{D_{js}(q,\theta)}. \tag{23}$$

From $G_{jl}(\hat{\eta}^n_{j_N})$ and $\bar{H}_{js}(\hat{\eta}^n_{j_N})$ that are obtained in the previous step through the predictor (20), we can derive a related estimate of $H^0(q)$ according to (19) leading to $H(\hat{\eta}^n_N) = \bar{H}(\hat{\eta}^n_N) + \begin{bmatrix} I \\ \Gamma(\hat{\eta}^n_N) \end{bmatrix}$, with $\Gamma(\hat{\eta}^n_N)$ an estimate of the direct feedthrough term Γ^0 of H_b^0, and that based on the relation $\check{e}_b(t) = \Gamma^0 \check{e}_a(t)$ from (4), can be given by

$$\Gamma(\hat{\eta}^n_N) = \left(\frac{1}{N}\sum_{t=1}^{N} \varepsilon_b(\hat{\eta}^n_N)\varepsilon_a^\top(\hat{\eta}^n_N)\right)\left(\frac{1}{N}\sum_{t=1}^{N} \varepsilon_a(\hat{\eta}^n_N)\varepsilon_a^\top(\hat{\eta}^n_N)\right)^{-1}. \tag{24}$$

Following the WNSF approach, we are now going to fit the parametric Box Jenkins model to the nonparametric model estimated from Step 3.1, by solving for θ in the equations

$$\begin{aligned}
F_{jl}(\theta) G_{jl}(\hat{\eta}^n_N) - L_{jl}(\theta) &= 0\ , \\
D_{js}(\theta) H_{js}(\hat{\eta}^n_N) - C_{js}(\theta) &= 0.
\end{aligned} \tag{25}$$

However, since these equations cannot be solved exactly, an optimization problem is formulated (Galrinho et al., 2019) that comes down to minimizing the quadratic residual vector on the Eqs. (25) by solving (in node-wise notation):

$$\min_{\theta_j} \|\hat{\eta}^n_{j_N} - Q_j(\hat{\eta}^n_{j_N})\theta_j\|_2 \tag{26}$$

where

$$Q_j(\eta) = \begin{bmatrix} Q_j^g & 0 \\ 0 & Q_j^h \end{bmatrix}, \tag{27}$$

with Q_j^g and Q_j^h diagonal matrices with entries

$$\begin{aligned} Q_j^{g^{jl}}(\eta) &= \left[-\mathcal{T}_{n\times m_f}[G_{jl}(\eta)] \quad \bar{I}_{n\times m_l}\right], \\ Q_j^{h^{js}}(\eta) &= \left[-\mathcal{T}_{n\times m_d}[H_{js}(\eta)] \quad \bar{I}_{n\times m_c}\right], \end{aligned} \tag{28}$$

with model orders $m_i, i \in \{l, f, c, d\}$ according to (22), the top left corner of $\bar{I}_{n\times m}$ is $I_{m\times m}$ and has zeros otherwise, and $\mathcal{T}_{n\times m}[X_{ji}(q)]$ is a lower triangular Toeplitz matrix where the first column is $\begin{bmatrix} x_0^{ji} & \cdots & x_{n-1}^{ji}\end{bmatrix}^\top$ with $X_{ji}(q) = \sum_{k=0}^{\infty} x_k^{ji} q^{-k}$.

The problem (26) is solved in first instance through the analytical least squares solution

$$\hat{\theta}_{j_N}^{[0]} = \left(Q_j^\top(\hat{\eta}_{j_N}^n)Q_j(\hat{\eta}_{j_N}^n)\right)^{-1} Q_j^\top(\hat{\eta}_{j_N}^n)\hat{\eta}_{j_N}^n. \tag{29}$$

However, a parameter estimate with smaller variance can be achieved if a weighted least squares criterion is applied.[3] This is introduced in the next step.

4.3. Step 3.3: Re-estimation of parametric model

In this step we reduce the variance further by re-estimating the obtained parametric models $G(\theta)$ and $H(\theta)$ defined in (23). For a statistical optimal solution of (26), instead of the standard least squares problem (26), a weighted least squares problem should be solved, where the optimal weight is given by the inverse of the covariance matrix of the residual $\hat{\eta}_{j_N}^n - Q_j(\hat{\eta}_{j_N}^n)\theta_j^0$, with θ_j^0 the actual network coefficients related to node w_j. This is not directly applicable since θ_j^0 is unknown. However it can be shown (Galrinho et al., 2019) that

$$\hat{\eta}_{j_N}^n - Q_j(\hat{\eta}_{j_N}^n)\theta_j^0 = T_j(\theta_j^0)(\hat{\eta}_{j_N}^n - \eta_j^{n0}), \tag{30}$$

with η_j^{n0} the real network coefficients related to the η-parametrized ARX model and $T_j(\theta)$ a block diagonal matrix with the denominator polynomials as entries

$$\begin{aligned} T_j^{g^{jl}}(\theta) &= \mathcal{T}_{n\times n}[F_{jl}(\theta)], \\ T_j^{h^{js}}(\theta) &= \mathcal{T}_{n\times n}[D_{js}(\theta)], \end{aligned} \tag{31}$$

where $\mathcal{T}_{n\times n}[X_{ji}(q)]$ is a lower triangular Toeplitz matrix where the first column is $\begin{bmatrix} 1 & x_1^{ji} & \cdots & x_m^{ji} & 0_{n-m-1}\end{bmatrix}^\top$ with $X_{ji}(q) = 1 + \sum_{k=1}^{\infty} x_k^{ji} q^{-k}$.

Result (30) motivates the use of a weighted least estimator with weighting matrix

$$W_j = T_j^{-1}(\theta_j^0)(P_{\hat{\eta}_{j_N}^n})^{-1}T_j^{-T}(\theta_j^0)$$

with $P_{\hat{\eta}_{j_N}^n}$ the covariance matrix of the nonparametric model. This can be implemented in an iterative scheme according to

$$\begin{aligned} &\hat{\theta}_{j_N}^{[k+1]} = \\ &\left(Q_j^\top(\hat{\eta}_{j_N}^n)W_j(\hat{\theta}_{j_N}^{[k]})Q_j(\hat{\eta}_{j_N}^n)\right)^{-1} Q_j^\top(\hat{\eta}_{j_N}^n)W_j(\hat{\theta}_{j_N}^{[k]})\hat{\eta}_{j_N}^n. \end{aligned} \tag{32}$$

For consistency of the estimates of parameter vector θ we refer to the proof in the WNSF method (Galrinho et al., 2019), with the actual model orders m_i with $i = f, l, c, d$ (22) known.

Remark 5. Because in this final step we correct for the variance due to the modeling error (30), the final estimate will have a reduced variance.

[3] As an alternative we can consider a weighted least squares criterion to obtain $\hat{\theta}_{j_N}^{[0]}$ (29), with the covariance matrix of the nonparametric model as weight.

Throughout the presented steps we split the MIMO optimization into L linear regressions that rely on explicit analytical solutions, and that allows for parallel computing. The Algorithm is given as follows.

Algorithm 1. Algorithm for full network identification in dynamic networks, including disturbance topology detection

Inputs: $w(t)$, $r(t)$, $R^0(q)$, model orders $m_i, i \in \{l, f, c, d\}$, network topology.
Output: Disturbance topology, $\hat{\theta}_N$.
Disturbance topology detection

1. Estimate noise rank p based on the reconstructed innovation $\varepsilon(t, \hat{\zeta}_N^n)$ (15), and if $p < L$ order the nodes.
2. 2.1 Obtain consistent estimate $\hat{\zeta}_N^n$ with least squares solution (36), where the nodes are ordered and by utilizing the estimated noise rank p.
   2.2 Use the reconstructed innovation $\varepsilon_a(t, \hat{\zeta}_N^n)$ as measured input in the one-step-ahead predictor (20) defined in (17) to estimate the noise correlation structure. We use
   (i) Structure selection with AIC, BIC and CV,
   (ii) Glasso,
   applied to estimate $\hat{\eta}_{j_N}^n$ that is obtained with least squares solution (37).

Estimating parametric network models

3. 3.1 Refine the nonparametric ARX model and obtain consistent estimate $\hat{\eta}_N^n$ with one-step-ahead predictor (20), where the estimated disturbance topology is fixed and update the reconstructed innovation to $\varepsilon_a(t, \hat{\eta}_N^n)$ to re-estimate $\hat{\eta}_N^n$.
   3.2 Reduce the nonparametric ARX model to a parametric model and obtain initial estimate $\hat{\theta}_N^{[0]}$ by (29).
   3.3 Re-estimate $\hat{\theta}_{j_N}^{[k+1]}$ with (32), where we update the weighting matrix $W_j(\hat{\theta}_{j_N}^{[k]})$ in each iteration.

We continue to iterate until we have reached the convergence criterion $\frac{\|\hat{\theta}_N^{[k]} - \hat{\theta}_N^{[k-1]}\|}{\|\hat{\theta}_N^{[k-1]}\|} < 0.0001$. This convergence criterion is also used in the simulation results in Section 6. In the next Section we derive the conditions required for consistency of estimates $\hat{\zeta}_{j_N}^n$ and $\hat{\eta}_{j_N}^n$.

5. Theoretical analyses

From here on we consider $n = n(N)$ i.e. the model order n increases as the data length N increases, while with increasing N, n/N tends to 0 with a particular rate (Galrinho et al., 2019; Ljung & Wahlberg, 1992).

Next we derive the conditions under which the estimates $\hat{\zeta}_N^n$ and $\hat{\eta}_N^n$, and consequently the reconstructed innovation are consistent.

5.1. Consistency of $\hat{\zeta}_N^n$ in Step 2.1: Refining the nonparametric model

With the noise rank p available and the nodes ordered we gained structural information on the unique noise model $\breve{H}^0(q)$ (4), namely we know that for the reduced noise rank case $p < L$

the last $L-p$ columns in $\breve{H}^0(q)$ are $\begin{bmatrix}0 & I\end{bmatrix}^\top$. Moreover, taking the inverse of $\breve{H}^0(q)$ does not affect the last $L-p$ columns since

$$(\breve{H}^0)^{-1} = \begin{bmatrix} (H_a^0)^{-1} & 0 \\ -\big(H_b^0 - \Gamma^0\big)(H_a^0)^{-1} & I \end{bmatrix}. \tag{33}$$

As a result the term $(\breve{H}^0(q))^{-1}R^0(q)$ in the one-step predictor (11), has the following structure

$$(\breve{H}^0)^{-1}R^0 = \begin{bmatrix} (H_a^0)^{-1}R_a^0 & 0 \\ -\big(H_b^0 - \Gamma^0\big)(H_a^0)^{-1}R_a^0 & R_b^0 \end{bmatrix}, \tag{34}$$

with the second block column consisting of known terms only. This allows in the parametrization of the predictor (9) to replace the square polynomial $\breve{B}(\zeta)$ with a non-square polynomial $B(\zeta)$, leading to

$$\begin{aligned}\hat{w}(t|t-1,\zeta) &= \big(I - \breve{A}(\zeta)\big)w(t) + B(\zeta)r_a(t) + \begin{bmatrix}0\\ R_b^0\end{bmatrix} r_b(t) \\ &= \varphi(t)\zeta + \begin{bmatrix}0\\ R_b^0\end{bmatrix} r_b(t),\end{aligned} \tag{35}$$

with $\varphi(t)$ composed of the appropriate terms in w and r_a.

Note that for an actual network with representation $G^0, \breve{H}^0, R^0$, the one-step predictor is still given by (11), but now the predictor model (35) can use the known external excitation signals $r_b(t)$. The ARX model is estimated according to $\hat{\zeta}_N^n = \arg\min_\zeta \frac{1}{N}\sum_{t=1}^N \varepsilon^T(t,\zeta)\varepsilon(t,\zeta)$, with $\varepsilon(t,\theta) = w(t) - \hat{w}(t|t-1;\zeta)$, leading to the analytical solution:

$$\hat{\zeta}_N^n = \left[\frac{1}{N}\sum_{t=1}^N \varphi(t)\varphi^\top(t)\right]^{-1} \frac{1}{N}\sum_{t=1}^N \varphi(t)\left[w(t) - \begin{bmatrix}0\\ R_b^0\end{bmatrix} r_b(t)\right]. \tag{36}$$

Note that Remark 1 holds and therefore predictor (35) can be decomposed in separate predictors for each node. The conditions for consistency are formulated in Proposition 1 and the proof is added in the Appendix.

Proposition 1 (*Consistency $\hat{\zeta}_N^n$*). *Consider a dynamic network that satisfies Assumption 1. Additionally, consider the one-step-ahead predictor (35). Then the transfer function matrices $(\breve{H}^0(q))^{-1}(I - G^0(q))$ and $(\breve{H}^0(q))^{-1}\begin{bmatrix}R_a^0(q)^\top & 0\end{bmatrix}^\top$ are consistently estimated with the analytical solution (12), if the following conditions hold:*

(1) The external excitation $r(t)$ is uncorrelated to the noise $e(t)$.
(2) The spectral density of $\kappa(t) = \begin{bmatrix}r_a(t)^\top & w(t)^\top\end{bmatrix}^\top$, $\Phi_\kappa(\omega) > 0$ for a sufficiently high number of frequencies ω.
(3) $\breve{A}(q,\zeta)$ and $B(q,\zeta)$ are of high order, such that $n \to \infty$.

Remark 6. Condition (1) and (2) of Proposition 1 are given for all signals present in the network. These conditions remain unchanged when we convert from a MIMO predictor to L linear regressions. Therefore the proof also holds for a predictor assessed per node.

Proof. See Appendix.

5.2. Consistency of $\hat{\eta}_N^n$ in Step 3.1: Refining the nonparametric model

A refined nonparametric model is estimated by exploiting the information on the noise topology in the form of a structured polynomial model $B(\eta_j)$ for $H_{js}(\eta_j)$ in the predictor (20), leading to the analytical solution

$$\hat{\eta}_N^n = \left[\frac{1}{N}\sum_{t=1}^N \varphi(t)\varphi^\top(t)\right]^{-1} \frac{1}{N}\sum_{t=1}^N \varphi(t)\big[w(t) - R^0 r(t)\big]. \tag{37}$$

with $\varphi(t)$ composed of the appropriate terms in w and $\varepsilon(\hat{\eta}_N^n)$.

The conditions for consistency are formulated in Proposition 2.

Proposition 2 (*Consistency $\hat{\eta}_N^n$*). *Consider a dynamic network that satisfies Assumption 1 and Proposition 1, and assume the disturbance topology is estimated correctly. Additionally, consider the one-step-ahead predictor (20) for all j. Then the transfer function matrices of $G^0(q)$ and $\breve{H}^0(q) - I$ are consistently estimated with the analytical solution $\hat{\eta}_N^n$ (37), if the following conditions hold:*

(1) For all j, the spectral density $\Phi_{\bar{\kappa}}(\omega)$ of $\bar{\kappa}(t) := \begin{bmatrix}w_{\{\mathcal{N}_j\}}(t)^\top & e_{\{\mathcal{V}_j\}}(t)^\top\end{bmatrix}^\top$, satisfies $\Phi_{\bar{\kappa}}(\omega) > 0$ for a sufficiently high number of frequencies ω.
(2) The data generating system is in the model set, i.e. there exists a η_0 such that $G(q,\eta_0) = G^0(q)$ and $\bar{H}(q,\eta_0) = \breve{H}^0(q) - I$.

Proof. See Appendix.

With consistent estimate $\hat{\eta}_N^n$ we can update the reconstructed innovation $\varepsilon(t,\hat{\eta}_N^n) = \begin{bmatrix}\varepsilon_a(t,\hat{\eta}_N^n)^\top & \varepsilon_b(t,\hat{\eta}_N^n)^\top\end{bmatrix}^\top$ consistently for each time step $t = 1,\dots,N$

$$\varepsilon(t,\hat{\eta}_N^n) \to \breve{e}(t) \qquad \text{w.p. 1 as } N \to \infty\ \forall t, \tag{38}$$

where the innovation is reconstructed per node according to $\varepsilon_j(t,\eta) = w_j(t) - \hat{w}_j(t|t-1,\eta)$ using one-step-ahead predictor (20).

Remark 7. Note that Condition (2) of Proposition 2 incorporates the condition that the noise rank p is chosen correctly, and the disturbance model is flexible enough to represent the exact disturbance topology of the network.

Following the line of reasoning in Van den Hof and Ramaswamy (2020), the spectral conditions in Propositions 1 and 2, which are actually data informativity conditions, can generically be replaced by path-based conditions on the graph of the network model set.

5.3. Generic data informativity conditions

Condition (2) of Proposition 1 and Condition (1) of Proposition 2 is a spectral data informativity condition on internal node signals in w, and it is difficult to interpret it for an experimenter. In this section we replace the spectral condition with a path-based data informativity condition in a generic sense,[4] i.e. independent of the numerical values of the network dynamics. By doing so we can evaluate if data informativity is satisfied based on the network and disturbance topology, and the properties of the external signals. Next we formulate the conditions in terms of properties and locations of the external signals analogous to Lemma 1 and Proposition 1 from Van den Hof and Ramaswamy (2020), by means of vertex-disjoint paths from external signals to internal node signals, where two paths are vertex-disjoint if they have no nodes in common, including their start and end nodes (Van der Woude, 1991). The consequences are illustrated in a 6-node example.

5.3.1. Vertex-disjoint paths

The generic version of Condition (2) of Proposition 1 is given in Proposition 3.

Proposition 3. *The spectrum condition $\Phi_\kappa(\omega) > 0$ for $\kappa(t) = \begin{bmatrix}r_a(t)^\top & w(t)^\top\end{bmatrix}^\top$ in Condition (2) of Proposition 1 is generically satisfied if there are L vertex-disjoint paths from $\begin{bmatrix}r_b(t)^\top & e(t)^\top\end{bmatrix}^\top$ to $w(t)$.*

Proof. See Appendix.

[4] Genericity is considered in the sense that the corresponding property holds for almost all models in the model set, possibly excluding a set of measure 0.

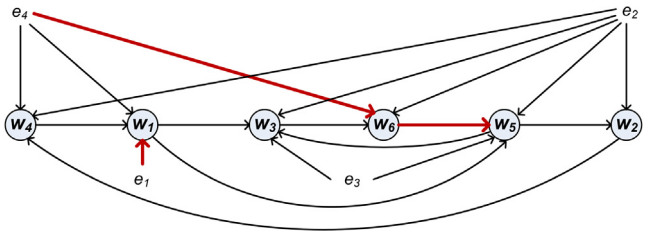

Fig. 1. 6-node dynamic network with reduced rank noise that has rank $p = 4$, no $r(t)$ signals are shown. The arrows represent the edges for which $G_{ji}^0 \neq 0$ and $H_{ji}^0 \neq 0$, where the arrows indicated in red are examples of the two vertex disjoint paths needed to satisfy Proposition 4 for output $w_3(t)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Proposition 3 gives a sufficient generic path-based condition that requires to have external excitation signals at certain locations in the network, combining data informativity conditions with identifiability (Van den Hof & Ramaswamy, 2020).

The set $\mathcal{V}$ denotes the set of indices of all the disturbing noise signals, where $\mathcal{V}_j$ is a subset of $\mathcal{V}$. For the generic condition for Condition (1) of Proposition 2 we introduce notation $e_{\{\mathcal{X}_j\}}(t)$, where $\mathcal{X}_j$ is the set of indices of all the disturbing noise signals excluding indices that are already present in set $\mathcal{V}_j$, i.e. $\mathcal{X}_j = \mathcal{V}/\mathcal{V}_j$.

Proposition 4. *The spectrum condition $\Phi_{\bar{\kappa}}(\omega) > 0$ for $\bar{\kappa}(t) = \begin{bmatrix} w_{\{\mathcal{N}_j\}}(t)^\top & e_{\{\mathcal{V}_j\}}(t)^\top \end{bmatrix}^\top$ in Condition (1) of Proposition 2 is generically satisfied if there are Cardinal$\{\mathcal{N}_j\}$ vertex-disjoint paths from $\begin{bmatrix} r(t)^\top & e_{\{\mathcal{X}_j\}}(t)^\top \end{bmatrix}^\top$ to $w_{\{\mathcal{N}_j\}}(t)$.*

Proof. See Appendix.

Proposition 4 gives a sufficient generic path based condition that requires external excitation signals at certain locations such that $\Phi_{\bar{\kappa}}(\omega) > 0$ for a sufficiently high number of frequencies.

Remark 8. If we want to identify only the jth row of the network (or only part of the network), we can consider the predictor in Proposition 2 only for node j and satisfy the conditions in Propositions 2 and 4 for node j.

Next we elaborate the vertex-disjoint path conditions by means of an example where a network is subject to reduced rank noise.

5.3.2. Reduced rank noise example

We consider a 6-node network that satisfies Assumption 1 and is subject to reduced rank noise of rank $p = 4$ shown in Fig. 1. This 6-node example is additionally used in the simulations in Section 6, and is further defined in Appendix E. The nodes are ordered such that the first p nodes are subject to full rank noise. Moreover, we assume the disturbance topology is correctly estimated.

The goal of this example is to elaborate on the path-based data informativity conditions given in Propositions 3 and 4. To be more specific, we show which external excitation signals are sufficient in order to satisfy the spectral Condition (2) in Proposition 1 and Condition (1) in Proposition 2. In the example we have external noise signals $e(t) = \begin{bmatrix} e_1(t) & \dots & e_4(t) \end{bmatrix}^\top$ and external excitation signals $r_k(t)$, for simplicity we assume R^0 contains elements that are either 0 or 1.

In order to satisfy Proposition 3, we require $L = 6$ vertex-disjoint paths from $\begin{bmatrix} r_b(t)^\top & e(t)^\top \end{bmatrix}^\top$ to $w(t)$. The first $p = 4$ nodes, denoted by $w_a(t)$, are excited by the noise $e(t)$; we therefore require at least $L - p = 2$ external excitation signals $r_k(t)$ on the last 2 nodes $w_b(t) = \begin{bmatrix} w_5(t) & w_6(t) \end{bmatrix}^\top$, i.e. $r_b(t) = \begin{bmatrix} r_5(t) & r_6(t) \end{bmatrix}^\top$ with $R_b = I \in \mathbb{R}^{2\times 2}$. Therefore we satisfy Proposition 3 since we have 6 vertex-disjoint paths from $\begin{bmatrix} e(t)^\top & r_b(t)^\top \end{bmatrix}^\top$ to $\begin{bmatrix} w_a(t)^\top & w_b(t)^\top \end{bmatrix}^\top$.

To show how Proposition 4 is satisfied, we first consider output node $w_3(t) = G_{31}(\eta)w_1(t) + G_{35}(\eta)w_5(t) + H_{32}(\eta)e_2(t) + H_{33}(\eta)e_3(t)$, that has $w_{\{\mathcal{N}_3\}}(t) = \begin{bmatrix} w_1(t) & w_5(t) \end{bmatrix}^\top$ and $e_{\{\mathcal{V}_3\}}(t) = \begin{bmatrix} e_2(t) & e_3(t) \end{bmatrix}^\top$. We need Cardinal$\{\mathcal{N}_3\} = 2$ vertex-disjoint paths from $\begin{bmatrix} r(t)^\top & e_{\{\mathcal{X}_j\}}(t)^\top \end{bmatrix}^\top$ to $w_{\{\mathcal{N}_3\}}(t)$. There already exist 2 vertex disjoint paths from $e_{\{\mathcal{X}_j\}}(t) = \begin{bmatrix} e_1(t) & e_4(t) \end{bmatrix}^\top$ to $w_{\{\mathcal{N}_3\}}(t)$. This shows that Proposition 4 is satisfied by the two vertex disjoint paths from $e_1(t) \to w_1(t)$ and from $e_4(t) \to w_6(t) \to w_5(t)$ as indicated in red in Fig. 1. If we apply the same reasoning to the other nodes we see that for node

- $w_1(t)$ with $w_{\{\mathcal{N}_1\}}(t) = w_4(t)$, there exists a vertex-disjoint path from $e_2(t) \to w_4(t)$.
- $w_2(t)$ with $w_{\{\mathcal{N}_2\}}(t) = w_5(t)$, there exists a vertex-disjoint path from $e_3(t) \to w_5(t)$.
- $w_4(t)$ with $w_{\{\mathcal{N}_4\}}(t) = w_2(t)$, there exists a vertex-disjoint path from $e_3(t) \to w_5(t) \to w_2(t)$
- $w_5(t)$ with $w_{\{\mathcal{N}_5\}}(t) = \begin{bmatrix} w_1(t) & w_6(t) \end{bmatrix}^\top$, there exist 2 vertex-disjoint paths from $e_1(t) \to w_1(t)$ and from $e_4(t) \to w_6(t)$.
- $w_6(t)$ with $w_{\{\mathcal{N}_3\}}(t) = w_3(t)$, there exists a vertex-disjoint path from $e_3(t) \to w_3(t)$.

In order to satisfy Proposition 4 we therefore do not require additional external excitation signals $r_k(t)$.

Consequently, in order to identify the full network for the given example, it is sufficient to add external signals $r_b(t) = \begin{bmatrix} r_5(t) & r_6(t) \end{bmatrix}^\top$ with $R_b = I \in \mathbb{R}^{2\times 2}$ that satisfies Proposition 3.

6. Numerical simulations

In this section we show the results of different steps in Algorithm 1. We assume $R^0 = I$, and consider the system given in Fig. 1 and Appendix E.

For the simulation study we use normally distributed zero mean white external signals, where $\{r(t)\}$ has a variance of 5 and the vector of e-signals has variances $\{0.1, 0.2, 0.3, 0.4\}$. We simulate the nodes according to $w(t) = (I - G^0)^{-1}(R^0 r(t) + H^0 e(t))$ and perform $M = 100$ Monte Carlo runs over five data lengths logarithmically spaced between 300 and 50000. For each of the data lengths N a specific value of the model order n is chosen according to $n = 10, 20, 30, 40, 40$, for increasing values of N. The actual model orders $m_i, i \in \{l, f, c, d\}$ can be derived from Appendix E.

Next we describe the noise rank estimation results of step 1 of Algorithm 1.

6.1. Rank p and ordering of the nodes

In order to obtain the noise rank p we perform a rank test (singular value decomposition) on covariance matrix $\hat{\Lambda}$ (13). For data length $N = 300$, the singular values averaged over the 100 Monte Carlo runs are
$svd(\hat{\Lambda}_N) = \begin{bmatrix} 0.37 & 0.26 & 0.21 & 0.06 & 2.13 \cdot 10^{-8} & 1.96 \cdot 10^{-9} \end{bmatrix}$,
where we see that the last two singular values are close to zero. As data length increases the last two values converge even closer to zero. For $N = 50000$ we obtain the following averaged singular values
$svd(\hat{\Lambda}_N) = \begin{bmatrix} 0.59 & 0.40 & 0.39 & 0.10 & 4.04 \cdot 10^{-13} & 1.24 \cdot 10^{-13} \end{bmatrix}$,
showing that a clear gap between the fourth and fifth singular value points to a correct rank estimate of 4.

Finally with the noise rank p available we can reorder the nodes such that $\begin{bmatrix} I_p & 0 \end{bmatrix} \Pi^\top \hat{\Lambda} \Pi \begin{bmatrix} I_p & 0 \end{bmatrix}^\top$ has rank p.

Next we show the disturbance topology detection results of step 2 of Algorithm 1.

6.2. Topology estimation of the disturbance model

For the topology detection we are interested in which indices belong in set $\mathcal{V}_j$ for all j, where the indices indicate where the edges are located in the disturbance model. We evaluate the performance of the topology detection by evaluating the trade-off between overestimating and underestimating the number of edges, that is typically used in receiver operating characteristic (ROC) curves (Hajian-Tilaki, 2013).

If an edge is present in both the data generating disturbance and the estimated disturbance topology, we count this edge as a true positive (TP). If an edge is present in the estimated disturbance topology but does not exist in the data generating system, we count this edge as a false positive (FP). Additionally we let *Pos* indicate the total number of existing edges and *Neg* indicates the total number of non-existing edges in the disturbance model. The ROC curve plots the true positive rate (*TPR*) versus the false positive rate (*FPR*), with

$$TPR = \frac{TP}{Pos}, \quad FPR = \frac{FP}{Neg}, \tag{39}$$

where $FPR = 0$ and $TPR = 1$ represented by the point $(0, 1)$, indicates the topology is perfectly reconstructed. We evaluate the closeness to the point $(0, 1)$ by utilizing the distance function

$$dis = \sqrt{FPR^2 + (1 - TPR)^2}, \tag{40}$$

For the structure selection procedure we test all possible combinations in set $\mathcal{V}_j$ and employ AIC, BIC and CV. For AIC we use

$$\frac{1}{2} \log\left(V_{j_N}(\hat{\eta}^n_{j_N})\right) + \frac{n_{p_j}}{N}, \tag{41}$$

with n_{p_j} the number of estimated parameters for node j and

$$V_{j_N}(\hat{\eta}^n_{j_N}) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon_j(t, \hat{\eta}^n_{j_N})^2. \tag{42}$$

For BIC we use

$$N * \log\left(V_{j_N}(\hat{\eta}^n_{j_N})\right) + N(\log(2\pi) + 1) + n_{p_j} \log(N). \tag{43}$$

From these simulations we select set $\mathcal{V}_j$ that gives the smallest AIC or BIC value. For the CV we split the data $Z^N = Z^{(1)}Z^{(2)}$ in a training set $Z^{(1)}$ of length $\frac{2}{3}(N + 1)$ and obtain the estimates for the different combinations in set $\mathcal{V}_j$ according to

$$\hat{\eta}^{(1)}_{j_N} = \underset{\eta}{\operatorname{argmin}} V_{j_N}(\eta_j, Z^{(1)}), \tag{44}$$

With the validation set $Z^{(2)}$, that contains the remaining data of length $N^{(2)} = \frac{1}{3}(N + 1)$, we minimize objective function

$$V_{j_N}(\hat{\eta}^{(1)}_{j_N}, Z^{(2)}) = \frac{1}{N^{(2)}} \sum_{t=1}^{N^{(2)}} \varepsilon_j(t, \hat{\eta}^{(1)}_{j_N})^2, \tag{45}$$

and select the set $\mathcal{V}_j$ that gives the smallest root mean squared error (RMSE)

$$RMSE_j = \sqrt{V_{j_N}(\hat{\eta}^{(1)}_{j_N}, Z^{(2)})}. \tag{46}$$

For Glasso we fully parametrize the disturbance model, using the known topology of G^0 and fixed $R^0 = I$. We inspect all elements of the disturbance model matrix that is parametrized with the Glasso estimates (21). If element $H_{ji}(\hat{\eta}_N)$ of the disturbance model matrix contains nonzero Glasso estimates we say this element contains dynamics, and therefore an edge is present and $i \in \mathcal{V}_j$. To prevent arbitrary small Glasso estimates are seen as dynamics we define a tolerance, where the Glasso estimates are nonzero if the l_2 norm of these estimates is larger than 10^{-3}. The choice to include the estimates of $G_{jl}(\eta)$ in the penalization is due to the implementation of Glasso (Boyd, Parikh, & Chu, 2011). For good estimates on the disturbance topology, we utilize the known topology of G^0 and deal with known $R^0 r(t)$ signals appropriately.

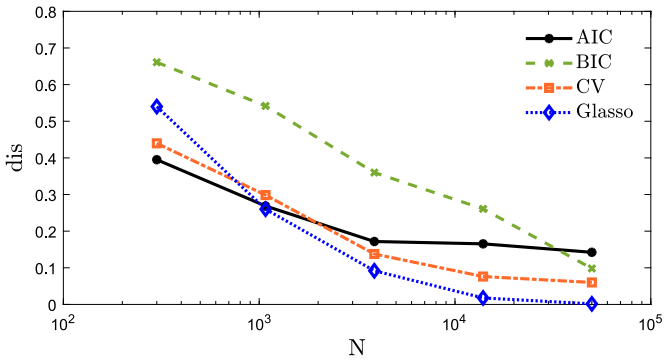

Fig. 2. *dis* as a function of N, averaged over the Monte Carlo runs.

Tuning of λ_j is done via a grid based search similar to the CV structure selection. First we select a grid $\lambda_j^{grid} = \{0, 25, 50, \ldots, 2000\}$ containing λ_j values to test. For each grid point we estimate $\hat{\eta}_j^{grid}$ using Glasso, from where the topology is derived by inspecting the disturbance model for dynamics as mentioned before, and fix the topology H_j^{grid} per node. Next we apply CV using topology H_j^{grid} and estimate the RMSE_j. The grid point with the lowest RMSE_j is selected as the λ_j value. Repeating the tuning procedure over a number of runs gives the minimally required value for λ_j. The tuning procedure is applied to all nodes for the different data lengths N.

Fig. 2 shows the topology detection results, with the distance averaged over 100 Monte Carlo runs. The BIC is a consistent information criterion (Kass & Raftery, 1995; Schwarz, 1978), meaning that the estimated disturbance topology will converge to the actual topology if $N \to \infty$. However, as can be seen in the results in Fig. 2, the full convergence of the BIC procedure is not reached for the given data lengths. Until the BIC procedure converges to the actual disturbance topology, it tends to underestimate the number of edges that are actually present, therefore the mismatch in the distance function is caused by not detecting all the TP's. The AIC is not a consistent information criterion, but has a faster convergence rate compared to the BIC (Zhang, 1993). The AIC tends to overestimate the number of edges, meaning the mismatch is caused by detecting the FP's. The CV is comparable to AIC but has a slower convergence rate. Finally the Glasso seems to have the best of both AIC and BIC. However, these results heavily depend on the selected tuning parameter λ, where it is not guaranteed that a suitable λ exists.

Next we show the parametric estimation results of step 3 of Algorithm 1, where we fix the estimated disturbance topology. Based on the results in Fig. 2 we have fixed the correctly estimated disturbance topology obtained with Glasso for $N = 50000$, where $TPR = 1$ and $FPR = 0$.

6.3. Estimating the parametric model

Next we present the results of the estimation of the parametric model. Because Algorithm 1 is consistent we have a negligible

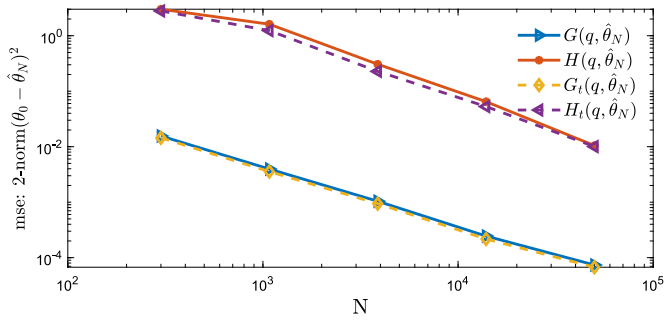

Fig. 3. MSE between $\hat{\theta}_N$ and θ_0 as function of sample size, averaged over the Monte Carlo runs, obtained with Algorithm 1 with $R^0 = I$, where subscript $\{t\}$ indicates the use of the true (unknown) white noise as a predictor input instead of the reconstructed innovation.

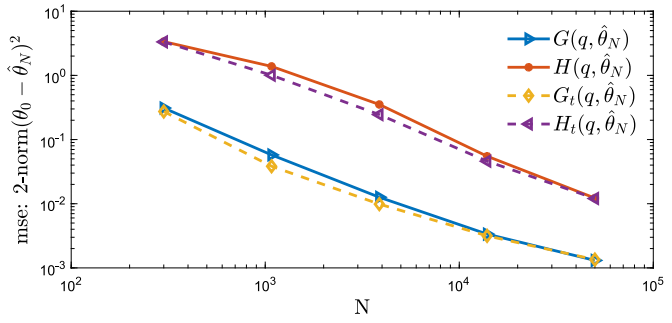

Fig. 4. MSE between $\hat{\theta}_N$ and θ_0 as function of sample size, averaged over the Monte Carlo runs, obtained with Algorithm 1 with $R^0 = \begin{bmatrix} 0 & R_b^{0\top} \end{bmatrix}^\top$ and $R_b = I \in \mathbb{R}^{2\times 2}$, where subscript $\{t\}$ indicates the use of the true (unknown) white noise as a predictor input instead of the reconstructed innovation.

bias and the mean squared error (MSE) represents the variance. For the simulations we use the correct estimated disturbance topology from the previous step. Additionally, for Step 3.2 of Algorithm 1, we compute the $\hat{\theta}^{[0]}_{j_N}$ in (29) using the covariance matrix of the nonparametric model as weighting. Figs. 3 and 4 present the sample MSE that is computed according to $\text{MSE}(N) = \frac{1}{M}\sum_{c=1}^{M}\left\|\hat{\theta}_{N,c}-\theta_0\right\|^2$, where c indicates the Monte Carlo run and $\hat{\theta}_{N,c}$ the final estimate (32). In Fig. 3 we use $R^0 = I$ in the data generating network, and in Fig. 4 we use $R^0 = \begin{bmatrix} 0 & R_b^{0\top} \end{bmatrix}^\top$ with $R_b = I \in \mathbb{R}^{2\times 2}$ according to Section 5.3.2. The solid lines represent Algorithm 1 where the estimates are obtained using the reconstructed innovation as input. The dotted lines represent Algorithm 1 where we use the realization of the actual noise $e(t)$ as input, indicated by subscript $\{t\}$. The results for the whole network are shown, while using L linear regressions. Both simulations shown in Figs. 3 and 4, typically perform $k = 6$ iterations for data length $N = 300$ in (32). As the data length N increases the number of iterations performed decreases, where for $N = 50000$ the simulations typically perform $k = 2$ iterations. The MSE(N) improvement after the iterations is shown in Table 1. From Table 1 we can derive that we benefit most from iterating k in the final step of Algorithm 1 if we do not have full excitation on the network with $R^0 = I$.

In Figs. 3 and 4 we see convergence between the solid and dotted lines as the data length N increases. This indicates that as data length N increases the reconstructed innovation converges to the actual noise. Furthermore all MSE results continue to converge towards zero which is in line with the consistency proof.

Table 1
MSE improvement:
$\frac{1}{M}\sum_{c=1}^{M}\left\|\hat{\theta}_{N,c}-\theta_0\right\|^2 - \frac{1}{M}\sum_{c=1}^{M}\left\|\hat{\theta}^{(1)}_{N,c}-\theta_0\right\|^2$ over k iterations.

N	300	1078	3873	13916	50000
$R^0 = I$	$1.6\cdot10^{-3}$	$5.1\cdot10^{-5}$	$-1.2\cdot10^{-6}$	$-1.9\cdot10^{-7}$	$3.7\cdot10^{-8}$
$R_b^0 = I$	0.43	0.26	0.15	0.07	0.01

The results of this simulation study support the consistency proof and we consistently estimate the BJ model structure, while employing a row-wise optimization.

7. Conclusions

In this paper we present a multi-step least squares method for network identification, that can handle reduced rank noise with low computational burden. We follow a step wise procedure where we first extend the SLR identification method to detect the disturbance topology, and thereafter extend the WNSF method to consistently identify networks of general model structure, including a BJ model structure. For a BJ network, usually a non-convex MIMO identification method is needed. In this paper, we show that we identify the BJ network using analytical solutions. Simulation results indicate that we can identify the disturbance topology of the given network with low error if the data length N is sufficiently large. We show that the presented method is consistent, and provide path based data informativity conditions, that guides where to allocate external excitation signals for the experimental design. Considering large networks subject to correlated and/or reduced rank noise, the presented method is promising due to its scalability and low variance results.

Appendix A. Proof of Proposition 1

Consider the prediction error for the predictor $\hat{w}(t|t-1,\zeta)$ from (35):

$$\begin{aligned}\varepsilon(t,\zeta) &= w(t) - \hat{w}(t|t-1,\zeta) = \breve{A}(\zeta)w(t) - \breve{B}(\zeta)r(t),\\ &= \breve{A}(\zeta)w(t) - B(\zeta)r_a(t) - \begin{bmatrix} 0 & R_b^\top \end{bmatrix}^\top r_b(t).\end{aligned} \tag{A.1}$$

With the data generating system (1) given as

$$\begin{aligned}w(t) &= (\breve{A}^0)^{-1}\breve{B}^0 r(t) + (\breve{A}^0)^{-1}\breve{e}(t),\\ &\text{with } \breve{A}^0 = (\breve{H}^0)^{-1}(I-G^0),\quad \breve{B}^0 = (\breve{H}^0)^{-1}R^0\end{aligned} \tag{A.2}$$

we can rewrite the prediction error as

$$\varepsilon(t,\zeta) = \big(\breve{A}^0 - \Delta\breve{A}(\zeta)\big)w - \big(\breve{B}^0 - \Delta\breve{B}(\zeta)\big)r \tag{A.3}$$

with $\Delta\breve{A}(\zeta) = \breve{A}^0 - \breve{A}(\zeta)$ and $\Delta\breve{B}(\zeta) = \breve{B}^0 - \breve{B}(\zeta)$. Then with (A.2) it follows that

$$\varepsilon(t,\zeta) = \Delta\breve{B}(\zeta)r - \Delta\breve{A}(\zeta)w + \breve{e}, \tag{A.4}$$

and since the second block column of $\breve{B}(\zeta)$ is fixed and known, it follows that $\Delta\breve{B}(\zeta)r = \Delta B(\zeta)r_a$. We now proceed by evaluating the jth component

$$\varepsilon_j(t,\zeta) = \quad \Delta B_j(\zeta)r_a - \Delta\breve{A}_j(\zeta)w + \breve{e}_j, \tag{A.5}$$

where $\Delta\breve{A}_j(\zeta)$ and $\Delta B_j(\zeta)$ are the rows of matrices $\Delta\breve{A}(\zeta)$ and $\Delta B(\zeta)$ belonging to node j.

The consistency proof consists of two steps:

(1) Show that the objective function is bounded from below by the noise variance[5] $\bar{V}_j(\zeta) := \bar{\mathbb{E}}\varepsilon_j^2(t,\zeta) \geq \sigma^2_{\breve{e}_j}$, where the minimum is achieved for $\Delta\breve{A}_j(\zeta) = 0$ and $\Delta\breve{B}_j(\zeta) = 0$.
(2) Show that the global minimum is unique.

[5] $\bar{\mathbb{E}}$ refers to the generalized expectation operator $\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}$.

A.1. Consistency proof step (1)

With (A.2) substituted into (A.5), the expression for $\varepsilon_j(t,\zeta)$ becomes

$$\Delta B_j(\zeta)r_a - \Delta\breve{A}_j(\zeta)\Big((\breve{A}^0)^{-1}\breve{B}^0 r + (\breve{A}^0)^{-1}\breve{e}\Big) + \breve{e}_j \tag{A.6}$$

from which, due to the fact that $\Delta\breve{A}_j(\zeta)$ is strictly proper and r and e are uncorrelated, it follows that $\breve{e}_j$ is uncorrelated with the remaining terms in the expression. As a result, the objective function is given by

$$\bar{V}_j(\zeta) = \bar{\mathbb{E}}\Big[\Big(\Delta B_j(\zeta)r_a - \Delta\breve{A}_j(\zeta)w\Big)^2\Big] + \sigma^2_{\breve{e}_j}, \tag{A.7}$$

from which we can infer that $\bar{V}_j(\zeta) \geq \sigma^2_{\breve{e}_j}$ with equality for $\Delta\breve{A}_j(\zeta) = 0$ and $\Delta B_j(\zeta) = 0$.

A.2. Consistency proof step (2)

For the second step we show that the minimum is unique, by showing that $\bar{V}_j(\zeta) = \sigma^2_{\breve{e}_j}$ implies $\Delta\breve{A}_j(\zeta) = 0$ and $\Delta B_j(\zeta) = 0$. With (A.7) and by applying Parseval's theorem, $\bar{V}_j(\zeta) = \sigma^2_{\breve{e}_j}$ implies

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Delta x^\top(e^{j\omega},\zeta)^\top \Phi_\kappa(\omega)\Delta x(e^{-j\omega},\zeta)\mathrm{d}\omega = 0, \tag{A.8}$$

with $\Delta x^\top = \begin{bmatrix}\Delta B_j(\zeta) & -\Delta\breve{A}_j(\zeta)\end{bmatrix}$ and $\kappa = \begin{bmatrix}r_a^\top & w^\top\end{bmatrix}^\top$.

By Condition (2) the spectral density $\Phi_\kappa(\omega)$ is positive definite. Therefore Eq. (A.8) holds only for $\Delta x^\top = 0$ which is satisfied by Condition (3). The global minimum of $\bar{V}_j(\zeta)$ is thus unique for $\breve{A}_j(\zeta) = \breve{A}^0_j$ and $\begin{bmatrix}B_j(\zeta) & \bar{R}_j\end{bmatrix} = \breve{B}^0_j$, with $\bar{R}_j = 0$ for $j = 1,\dots,p$ and $\bar{R}_j$ is a row of R_b for $j = p+1,\dots,L$. □

Appendix B. Proof of Proposition 2

For ease of notation we start with the MIMO notation of the one-step-ahead predictor (20)

$$\hat{w}(t|t-1,\eta) = G(\eta)w + Rr + \bar{H}(\eta)\varepsilon_a(\hat{\zeta}^n_N), \tag{B.1}$$

From Proposition 1 we know $\hat{\zeta}^n_N$ is consistent, therefore

$$\varepsilon(\hat{\zeta}^n_N) \to \breve{e} \quad \text{w.p. 1 as } N \to \infty\, \forall t, \tag{B.2}$$

and we can rewrite the one-step-ahead predictor as

$$\hat{w}(t|t-1,\eta) = G(\eta)w + Rr + \bar{H}(\eta)e \tag{B.3}$$

Considering the data generating system in (1) the residual becomes

$$\begin{aligned}\varepsilon(t,\eta) &= w(t) - \hat{w}(t|t-1,\eta)\\ &= \Delta G(\eta)w + H^0 e - \bar{H}(\eta)e\\ &= \Delta G(\eta)w + \Delta\bar{H}(\eta)e + \begin{bmatrix}I\\ \Gamma^0\end{bmatrix}e,\end{aligned} \tag{B.4}$$

where $\Delta G(\eta) = G^0 - G(\eta)$, and $\Delta\bar{H}(\eta) = \begin{bmatrix}\Delta\bar{H}_a(\eta)\\ \Delta\bar{H}_b(\eta)\end{bmatrix}$, with $\Delta\bar{H}_a(\eta) = \bar{H}^0_a - \bar{H}_a(\eta)$, with $\bar{H}_a = H_a - I$ and $\Delta\bar{H}_b(\eta) = \bar{H}^0_b - \bar{H}_b(\eta)$, with $\bar{H}_b = H_b - \Gamma$.

The residual per node is written as

$$\varepsilon_j(t,\eta) = \sum_{l\in\mathcal{N}_j}\Delta G_{jl}(\eta)w_l + \sum_{s\in\mathcal{V}_j}\Delta\bar{H}_{js}(\eta)e_s + \breve{e}_j, \tag{B.5}$$

where $\Delta G_{jl}(\eta) = G^0_{jl} - G_{jl}(\eta)$ is an element of matrix $\Delta G(\eta)$, and $\Delta\bar{H}_{js}(\eta)$ is an element of matrix $\Delta\bar{H}(\eta)$.

The consistency proof consists of two steps

(1) Show that the objective function is bounded from below by the noise variance $\bar{V}_j(\theta) := \bar{\mathbb{E}}\varepsilon^2_j(t,\theta) \geq \sigma^2_{\breve{e}_j}$, where the minimum is achieved for $\Delta G_{jl} = 0$ and $\Delta\bar{H}_{js} = 0$.
(2) Show that the global minimum is unique.

Step 1 By using the property that all ΔG- and $\Delta\bar{H}$-terms are strictly proper, it follows from (B.5) that

$$\bar{V}_j(\eta) = \bar{\mathbb{E}}\Big[\Big(\sum_{l\in\mathcal{N}_j}\Delta G_{jl}(\eta)w_l + \sum_{s\in\mathcal{V}_j}\Delta\bar{H}_{js}(\eta)e_s\Big)^2\Big] + \sigma^2_{\breve{e}_j} \tag{B.6}$$

and $\bar{V}_j(\eta) \geq \sigma^2_{\breve{e}_j}$ with equality for $\Delta G_{jl} = 0$ and $\Delta\bar{H}_{js} = 0$ for all $l \in \mathcal{N}_j$ and $s \in \mathcal{V}_j$.

Step 2 Showing that the minimum is unique is done by showing that $\bar{V}_j(\eta) = \sigma^2_{\breve{e}_j}$ implies $\Delta G_{jl} = 0$ and $\Delta\bar{H}_{js} = 0$ for all $l \in \mathcal{N}_j$ and $s \in \mathcal{V}_j$. With (B.6) and by applying Parseval's theorem, $\bar{V}_j(\zeta) = \sigma^2_{\breve{e}_j}$ implies

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Delta x^\top(e^{j\omega},\eta)^\top\Phi_{\bar{\kappa}}(\omega)\Delta x(e^{-j\omega},\eta)\mathrm{d}\omega = 0, \tag{B.7}$$

with $\Delta x^\top = \begin{bmatrix}\Delta G_{jl\in\mathcal{N}_j} & \Delta\bar{H}_{js\in\mathcal{V}_j}\end{bmatrix}$ and $\bar{\kappa} = \begin{bmatrix}w^\top_{\{\mathcal{N}_j\}} & e^\top_{\{\mathcal{V}_j\}}\end{bmatrix}^\top$.

By Condition (1) the spectral density $\Phi_{\bar{\kappa}}$ is positive definite. Therefore Eq. (B.7) holds only for $\Delta x^\top = 0$. The Parseval's theorem shows the global minimum of $\bar{V}_j(\eta)$ is unique for $G_{jl}(\eta) = G^0_{jl}$ and $\bar{H}_{js}(\eta) = \breve{H}^0_{js} - I_{js}$ by Condition (2). □

Appendix C. Proof of Proposition 3

The vector signal κ is written as

$$\kappa = \begin{bmatrix}r_a\\ w\end{bmatrix} = \underbrace{\begin{bmatrix}I & 0 & 0\\ J_{wa} & J_{wb} & J_{we}\end{bmatrix}}_{J}\begin{bmatrix}r_a\\ r_b\\ e\end{bmatrix} \tag{C.1}$$

with J_{wa}, J_{wb}, J_{we} appropriate transfer function matrices. Since $\rho = \begin{bmatrix}r_a^\top & r_b^\top & e^\top\end{bmatrix}^\top$ is persistently exciting, i.e. $\Phi_\rho(\omega) \geq 0$ for all ω, it follows from Lemma 1 in Van den Hof and Ramaswamy (2020) that κ is persistently exciting if and only if matrix J has full row rank. Since full row rank of J is equivalent to a full row rank of $[J_{wb}\ J_{we}]$, the result of Proposition 1 in Van den Hof and Ramaswamy (2020) then shows the equivalence with the condition that there are L vertex disjoint paths from the inputs of $[J_{wb}\ J_{we}]$, i.e. r_b and e, to its outputs, i.e. w. □

Appendix D. Proof of Proposition 4

Similar to the line of reasoning in the proof of Proposition 3, the vector signal $\bar{\kappa}$ is written as

$$\bar{\kappa} = \begin{bmatrix}w_{\{\mathcal{N}_j\}}\\ e_{\{\mathcal{V}_j\}}\end{bmatrix} = \underbrace{\begin{bmatrix}J_{wr} & J_{wx} & J_{wv}\\ 0 & 0 & I\end{bmatrix}}_{\bar{J}}\begin{bmatrix}r\\ e_{\{\mathcal{X}_j\}}\\ e_{\{\mathcal{V}_j\}}\end{bmatrix} \tag{D.1}$$

with J_{wr}, J_{wx}, J_{wv} appropriate transfer function matrices. Since $\bar{\rho} = \begin{bmatrix}r^\top & e^\top_{\{\mathcal{X}_j\}} & e^\top_{\{\mathcal{V}_j\}}\end{bmatrix}^\top$ is persistently exciting, i.e. $\Phi_{\bar{\rho}}(\omega) \geq 0$ for all ω, it follows from Lemma 1 in Van den Hof and Ramaswamy (2020) that $\bar{\kappa}$ is persistently exciting if and only if matrix $\bar{J}$ has full row rank. Since full row rank of $\bar{J}$ is equivalent to a full row rank of $[J_{wr}\ J_{wx}]$, the result of Proposition 1 in Van den Hof and Ramaswamy (2020) then shows the equivalence with the condition that there are $Cardinal\{\mathcal{N}_j\}$ vertex disjoint paths from the inputs of $[J_{wr}\ J_{wx}]$, i.e. r and $e_{\{\mathcal{X}_j\}}$, to its outputs, i.e. $w_{\{\mathcal{N}_j\}}$. □

Appendix E. System used in simulations

In the simulation results in Section 6 we use the data generating network of which the graph is represented in Fig. 1. The data generating transfer functions G and H are given by

$$G = \begin{bmatrix} 0 & 0 & 0 & G_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & G_{25} & 0 \\ G_{31} & 0 & 0 & 0 & G_{35} & 0 \\ 0 & G_{42} & 0 & 0 & 0 & 0 \\ G_{51} & 0 & 0 & 0 & 0 & G_{56} \\ 0 & 0 & G_{63} & 0 & 0 & 0 \end{bmatrix}, \tag{E.1}$$

with the elements of G_{jl}

$$\begin{aligned} G_{14} &= \frac{0.38q^{-1}+0.24q^{-2}}{1-1.35q^{-1}+0.54q^{-2}}, & G_{25} &= \frac{0.20q^{-1}}{1-1.30q^{-1}+0.60q^{-2}}, \\ G_{31} &= \frac{0.39q^{-1}}{1-0.80q^{-1}+0.20q^{-2}}, & G_{35} &= \frac{0.16q^{-1}}{1-1.23q^{-1}+0.51q^{-2}}, \\ G_{42} &= \frac{-0.30q^{-1}}{1-0.60q^{-1}+0.20q^{-2}}, & G_{51} &= \frac{-0.60q^{-1}}{1+0.45q^{-1}+0.12q^{-2}}, \\ G_{56} &= \frac{-0.22q^{-1}}{1-1.22q^{-1}+0.46q^{-2}}, & G_{63} &= \frac{-0.11q^{-1}}{1-1.49q^{-1}+0.62q^{-2}}, \end{aligned} \tag{E.2}$$

and

$$H = \begin{bmatrix} H_{11} & 0 & 0 & H_{14} \\ 0 & H_{22} & 0 & 0 \\ 0 & H_{32} & H_{33} & 0 \\ 0 & H_{42} & 0 & H_{44} \\ 0 & H_{52} & H_{53} & 0 \\ 0 & H_{62} & 0 & H_{64} \end{bmatrix}, \tag{E.3}$$

with noise rank $p = 4$ and elements

$$\begin{aligned} H_{11} &= \frac{1+0.52q^{-1}}{1+0.41q^{-1}}, & H_{14} &= \frac{0.41q^{-1}}{1-0.56q^{-1}}, \\ H_{22} &= \frac{1+0.44q^{-1}}{1+0.35q^{-1}}, & H_{32} &= \frac{-0.56q^{-1}}{1-0.40q^{-1}}, \\ H_{33} &= \frac{1-0.20q^{-1}}{1+0.43q^{-1}}, & H_{42} &= \frac{0.26q^{-1}}{1-0.62q^{-1}}, \\ H_{44} &= \frac{1+0.52q^{-1}}{1+0.45q^{-1}}, & H_{52} &= \frac{0.49q^{-1}}{1-0.49q^{-1}}, \\ H_{53} &= \frac{1+0.66q^{-1}}{1+0.51q^{-1}}, & H_{62} &= \frac{1+0.24q^{-1}}{1+0.53q^{-1}}, \\ H_{64} &= \frac{-0.56q^{-1}}{1-0.56q^{-1}+0.21q^{-2}}, \end{aligned} \tag{E.4}$$

where $\Gamma^0 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

References

Araki, M., & Saeki, M. (1983). A quantitative condition for the well-posedness of interconnected dynamical systems. *IEEE Transactions on Automatic Control*, *28*(5), 569–577.

Bach, F., Jenatton, R., Mairal, J., Obozinski, G., et al. (2011). Convex optimization with sparsity-inducing norms. *Optimization for Machine Learning*, *5*, 19–53.

Bolstad, A., Van Veen, B. D., & Nowak, R. (2011). Causal network inference via group sparse regularization. *IEEE Transactions on Signal Processing*, *59*(6), 2628–2641.

Bombois, X., & Hjalmarsson, H. (2021). Network topology detection via uncertainty analysis of an identified static model. *IFAC-PapersOnLine*, *54*(7), 595–600, 19th IFAC Symposium on System Identification SYSID 2021.

Boyd, S., Parikh, N., & Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

Camba-Méndez, G., & Kapetanios, G. (2009). Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. *Econometrics Reviews*, *28*(6), 581–611.

Cao, W., Picci, G., & Lindquist, A. (2021). Identification of low rank vector processes. arXiv:2111.10899.

Chiuso, A., & Pillonetto, G. (2012). A Bayesian approach to sparse dynamic network identification. *Automatica*, *48*(8), 1553–1565.

Dankers, A. G. (2019). *Optimization method for obtaining estimates in a dynamic network*. Technical note.

Dankers, A. G., Van den Hof, P. M. J., Heuberger, P. S. C., & Bombois, X. (2016). Identification of dynamic models in complex networks with prediction error methods: Predictor input selection. *IEEE Transactions on Automatic Control*, *61*(4), 937–952.

Dankers, A. G., Van den Hof, P. M. J., Bombois, X., & Heuberger, P. S. C. (2015). Errors-in-variables identification in dynamic networks — Consistency results for an instrumental variable approach. *Automatica*, *62*, 39–50.

Dankers, A. G., Van den Hof, P. M. J., Heuberger, P. S. C., & Bombois, X. (2012). Dynamic network structure identification with prediction error methods - Basic examples. *IFAC Proceedings Volumes*, *45*(16), 876–881, 16th IFAC Symposium on System Identification.

Deistler, M., Anderson, B. D. O., Filler, A., Zinner, C., & Chen, W. (2010). Generalized linear dynamic factor models: An approach via singular autoregressions. *European Journal of Control*, *16*(3), 211–224.

Deistler, M., Scherrer, W., & Anderson, B. D. O. (2015). The structure of generalized linear dynamic factor models. In J. Beran, Y. Feng, & H. Hebbel (Eds.), *Empirical economic and financial research: Theory, methods and practice* (pp. 379–400). Cham: Springer International Publishing.

Dimovska, M., & Materassi, D. (2017). Granger-causality meets causal inference in graphical models: Learning networks via non-invasive observations. In *2017 IEEE 56th annual conference on decision and control* (pp. 5268–5273).

Everitt, N., Galrinho, M., & Hjalmarsson, H. (2018). Open-loop asymptotically efficient model reduction with the Steiglitz–McBride method. *Automatica*, *89*, 221–234.

Fonken, S. J. M., Ferizbegovic, M., & Hjalmarsson, H. (2020). Consistent identification of dynamic networks subject to white noise using weighted null-space fitting. *IFAC-PapersOnLine*, *53–2*, 46–51, Proc. 21st IFAC World Congress, Berlin, Germany.

Galrinho, M., Rojas, C., & Hjalmarsson, H. (2019). Parametric identification using weighted null-space fitting. *IEEE Transactions on Automatic Control*, *64*(7), 2798–2813.

Gevers, M., Bazanella, A. S., & Vian da Silva, G. (2018). A practical method for the consistent identification of a module in a dynamical network. *IFAC-PapersOnLine*, *51–15*, 862–867, Proc. 18th IFAC Symp. System Identification (SYSID2018).

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*(7), 1–15.

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627.

Hickman, R., Van Verk, M. C., Van Dijken, A. J. H., Mendes, M. P., Vroegop-Vos, I. A., Caarls, L., et al. (2017). Architecture and dynamics of the jasmonic acid gene regulatory network. *The Plant Cell*, *29*(9), 2086–2105.

Van den Hof, P. M. J., Dankers, A. G., Heuberger, P. S. C., & Bombois, X. (2013). Identification of dynamic models in complex networks with prediction error methods: Basic methods for consistent module estimates. *Automatica*, *49*(10), 2994–3006.

Van den Hof, P. M. J., & Ramaswamy, K. R. (2020). Path-based data-informativity conditions for single module identification in dynamic networks. In *Proc. 59th IEEE conf. on decision and control* (pp. 4354–4359). Jeju Island, Republic of Korea: IEEE.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Ljung, L. (1999). *System identification. Theory for the user* (2nd ed.). Prentice-Hall.

Ljung, L., & Wahlberg, B. (1992). Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Advances in Applied Probability*, *24*(2), 412–440.

Materassi, D., & Innocenti, G. (2010). Topological identification in networks of dynamical systems. *IEEE Transactions on Automatic Control*, *55*(8), 1860–1871.

Materassi, D., & Salapaka, M. V. (2012). On the problem of reconstructing an unknown topology via locality properties of the Wiener filter. *IEEE Transactions on Automatic Control*, *57*(7), 1765–1777.

Materassi, D., & Salapaka, M. V. (2020). Signal selection for estimation and identification in networks of dynamic systems: A graphical model approach. *IEEE Transactions on Automatic Control*, *65*(10), 4138–4153.

Materassi, D., Salapaka, M. V., & Giarrè, L. (2011). Relations between structure and estimators in networks of dynamical systems. *Proceedings of the IEEE Conference on Decision and Control*, 162–167.

Pagani, G. A., & Aiello, M. (2013). The power grid as a complex network: A survey. *Physica A: Statistical Mechanics and its Applications*, *392*(11), 2688–2700.

Ramaswamy, K. R., Bottegal, G., & Van den Hof, P. M. J. (2021). Learning linear models in a dynamic network using regularized kernel-based methods. *Automatica*, *129*(109591).

Ramaswamy, K. R., & Van den Hof, P. M. J. (2021). A local direct method for module identification in dynamic networks with correlated noise. *IEEE Transactions on Automatic Control*, *66*(11), 3237–3252.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Shi, S., Bottegal, G., & Van den Hof, P. M. J. (2019). Bayesian topology identification of linear dynamic networks. In *2019 18th European control conference* (pp. 2814–2819).

Van den Hof, P. M. J., & Ramaswamy, K. R. (2021). Learning local modules in dynamic networks. In *Proc. machine learning research*: *Vol. 144*, *Proceedings of the 3rd conference on learning for dynamics and control* (pp. 176–188).

Veedu, M. S., & Salapaka, M. V. (2020). Topology identification under spatially correlated noise. arXiv:2012.04175.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*(1), 92–107.
Weerts, H. H. M., Galrinho, M., Bottegal, G., Hjalmarsson, H., & Van den Hof, P. M. J. (2018). A sequential least squares algorithm for ARMAX dynamic network identification. *IFAC-PapersOnLine*, *51*(15), 844–849.
Weerts, H. H. M., Van den Hof, P. M. J., & Dankers, A. G. (2017). Identification of dynamic networks with rank-reduced process noise. *IFAC-PapersOnLine*, *50*(1), 10562–10567.
Weerts, H. H. M., Van den Hof, P. M. J., & Dankers, A. G. (2018a). Identifiability of linear dynamic networks. *Automatica*, *89*, 247–258.
Weerts, H. H. M., Van den Hof, P. M. J., & Dankers, A. G. (2018b). Prediction error identification of linear dynamic networks with rank-reduced noise. *Automatica*, *98*, 256–268.
Van der Woude, J. W. (1991). A graph-theoretic characterization for the rank of the transfer matrix of a structured system. *Mathematics of Control, Signals, and Systems*, *4*(1), 33–40.
Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B.*, *68*, 49–67.
Yuan, Y., Stan, G., Warnick, S., & Gonçalves, J. (2011). Robust dynamical network structure reconstruction. *Automatica*, *47*(6), 1230–1235.
Zhang, P. (1993). On the convergence rate of model selection criteria. *Communications in Statistics. Theory and Methods*, *22*(10), 2765–2775.

Stefanie J.M. Fonken received her B.Sc. degree in Mechatronic Engineering in 2014 from Fontys University of Applied Sciences in Eindhoven.

After her Bachelor she worked as a Machine Engineer at Bosch Transmission Technology for 3 years.

In 2020, she obtained the M.Sc. degree in Systems and Control (with distinction cum laude) from Eindhoven University of Technology (TU/e).

Currently, she is working as a Ph.D. candidate in the project System Identification in Dynamic Networks (SYSDYNET) in the Control Systems group of TU/e.

Karthik Raghavan Ramaswamy was born in 1989.

He received his Bachelor's in Electrical and Electronics Engineering (with Distinction) in 2011 from Anna University and Master's in Systems and Control (with great appreciation) from the Eindhoven University of Technology in 2017.

From 2011 to 2015 he was Control & Automation design engineer at Larsen & Toubro and also holds an M.B.A. Currently, he is a Ph.D. researcher with the Control Systems research group, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands.

His research interests are in the area of machine learning, data-driven modeling of complex systems, and dynamic network identification.

Paul M.J. Van den Hof received the M.Sc. and Ph.D. degrees in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1982 and 1989, respectively.

In 1986 he moved to Delft University of Technology, where he was appointed as Full Professor in 1999. From 2003 to 2011, he was founding co-director of the Delft Center for Systems and Control (DCSC). As of 2011, he is a Full Professor in the Electrical Engineering Department, Eindhoven University of Technology. His research interests include data-driven modeling, identification for control, dynamic network identification, and model-based control and optimization, with applications in industrial process control systems and high-tech systems. He holds an ERC Advanced Research grant for a research project on identification in dynamic networks.

Paul Van den Hof is an IFAC Fellow and IEEE Fellow, and Honorary Member of the Hungarian Academy of Sciences, and IFAC Advisor. He has been a member of the IFAC Council (1999–2005, 2017–2020), the Board of Governors of IEEE Control Systems Society (2003–2005, 2022–2024), and an Associate Editor and Editor of Automatica (1992–2005). In the triennium 2017–2020 he served as Vice-President of IFAC.