Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 53-2 (2020) 46–51

Consistent identification of dynamic networks subject to white noise using Weighted Null-Space Fitting ⋆

Stefanie Fonken * **Mina Ferizbegovic** ** **Håkan Hjalmarsson** **

* *Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands (e-mail: s.j.m.fonken@student.tue.nl).*
** *Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {minafe, hjalmars}@kth.se)*

Abstract: Identification of dynamic networks has been a flourishing area in recent years. However, there are few contributions addressing the problem of simultaneously identifying all modules in a network of given structure. In principle the prediction error method can handle such problems but this methods suffers from well known issues with local minima and how to find initial parameter values. Weighted Null-Space Fitting is a multi-step least-squares method and in this contribution we extend this method to rational linear dynamic networks of arbitrary topology with modules subject to white noise disturbances. We show that WNSF reaches the performance of PEM initialized at the true parameter values for a fairly complex network, suggesting consistency and asymptotic efficiency of the proposed method.

Keywords: System identification, dynamic networks, least-squares.

1. INTRODUCTION

A dynamic network consists of spatially interconnected systems, where these interconnections cause the systems to be dynamically correlated with each other. Dynamic networks are generally subjected to external variables, such as excitation and noise signals. Measured signals are represented by nodes in the network while the systems causing the interactions between the nodes are called modules.

With a rising interest in data-driven modeling for identification in dynamic networks, several challenges should be addressed. These can be roughly categorized as follows: detection of network topology (Materassi and Innocenti, 2010; Materassi and Salapaka, 2012), conditions for network identifiability (Gonçalves and Warnick, 2008; Weerts et al., 2015, 2016, 2018a; Gevers et al., 2017), and identification of modules in the dynamic network. Estimating all modules in a dynamic network can be accomplished with the prediction error method (PEM), but often requires non-convex optimization. For large and complex networks the computational effort required for non-convex problems will increase, in addition, the number of local minima tend to grow as the number of modules to estimate increases.

In terms of parameter estimation, the literature focuses on obtaining consistent estimates of separate modules or a collection of modules in a dynamic network. Estimating a particular module in a dynamic network can be achieved with closed-loop identification methods. Methods such as the direct method (Ljung, 1999) and joint-IO, obtain consistent estimates when the noise model has been exactly modeled. Methods that do not depend on exactly estimating the noise model are the two-stage method (Van den Hof and Schrama, 1993) and the instrumental variable (IV) method (Söderström and Stoica, 1983). Both methods require the external excitation signals to be (partially) measurable to obtain consistent estimates. The IV method is the only aforementioned method that does not employ a cost function. A framework for identifying a single module or a collection of modules in dynamic networks has been established in Van den Hof et al. (2013), generalizing the aforementioned closed-loop identification methods. These methods can be applied to one or several multi-input-single-output (MISO) closed-loop problems. Related studies on identifying separate modules or sub-modules in a dynamic network are given in (Dankers et al., 2015; Galrinho et al., 2017; Everitt et al., 2018a,b).

Estimating the modules or collection of modules separately comes at a cost in estimation accuracy and asymptotic efficiency because certain correlations due to the interconnections are not considered.

The subject of identifying all modules simultaneously in dynamic networks is barely covered in system identification literature. The joint-direct method (Weerts et al., 2018b) is an identification approach that does predict all nodes jointly. The identification criterion used here is still in general a non-convex optimization problem.

⋆ This work was supported by the research environment NewLEADS—New Directions in Learning Dynamical Systems, contract 2016-06079; and Wallenberg AI, Autonomous Systems and Software Program (WASP), funded by Knut and Alice Wallenberg Foundation.
This project has received funding from the European Research Council (ERC), Advanced Research Grant SYSDYNET, under the European Unions Horizon 2020 research and innovation programme (Grant Agreement No. 694504).

10.1016/j.ifacol.2020.12.047

Assuming the topology of the network is known, the aim of this paper is to propose a method to simultaneously estimate all modules in a dynamic network providing consistent and asymptotically efficient estimates, without the need to solve non-convex optimization problems. The Weighted Null-Space Fitting (WNSF) method (Galrinho et al., 2019) has been shown to provide consistent and asymptotically efficient estimates for single-input-single-output (SISO) systems. This method avoids non-convex optimization, and extensions to multivariate systems and cascaded systems are available (Galrinho et al., 2018). For networks that can be written as multivariate autoregressive moving average exogenous (ARMAX) models, a multi-step least-squares method is proposed in (Weerts et al., 2018). It is shown that this method can be interpreted as WNSF applied to multivariate ARMAX-models. Here we continue these developments and extend WNSF to linear dynamic networks of known topology, where the nodes are subject to white disturbances rather than the type of noise models considered in the ARMAX case.

The paper proceeds with a problem statement in Section 2 and describes the dynamic network setup that is considered for the extension of the WNSF method. Section 3 provides relevant background regarding the WNSF method, followed by the proposed extension on the WNSF method in Section 4. Section 5 provides a simulation to verify the theoretical results. Here WNSF for dynamic networks is compared to PEM initialized at the true parameters.

2. PROBLEM STATEMENT

A dynamic network can be expressed by L internal signals or nodes denoted as $w(t) = [w_1(t), \dots, w_L(t)]^\top$. The modules considered are linear time invariant (LTI) systems. Following Van den Hof et al. (2013), the network is expressed as

$$\begin{aligned} w(t) &= G(q,\theta)w(t) + R(q,\theta)r(t) + e(t) \\ w(t) &= \big(I - G(q,\theta)\big)^{-1}\big(R(q,\theta)r(t) + e(t)\big), \end{aligned} \tag{1}$$

where

- $G(q,\theta)$ is an off-diagonal matrix with its elements $G_{ij}(q,\theta)$ either stable proper rational transfer functions containing at least one delay, or zero.
- $r(t) = [r_1(t), \dots, r_M(t)]^\top$ is a vector of external excitation signals, and $R(q,\theta)$ is an $L \times M$ transfer function matrix with proper rational elements. In addition, $R(q,\theta)$ is assumed full rank.
- $e(t) = [e_1(t), \dots, e_L(t)]^\top$ is a vector of unmeasured white noise sequences with zero mean and $\mathbb{E}[e(t)e^\top(s)] = \Lambda\delta_{t-s}$ where δ_τ is the Kronecker delta function.

It is assumed that the nodes $w(t)$ and the external excitations $r(t)$ are available to the user and are sufficiently informative so that the network is identifiable. In addition, it is assumed that the network is well-posed, implying that all principle minors of $\big(I - G(q,\infty)\big)^{-1}$ are nonzero (Dankers, 2014). Thus $I - G(q,\theta)$ is invertible and the inverse consists of causal transfer functions. In addition, the topology of the network is assumed to be known. Furthermore, we assume the data is generated according to (1) with $\theta = \theta_0$, where θ consists of numerator and denominator coefficients of the transfer functions G_{ij}, $i,j = 1,...,L$, and R_{ij}, $i = 1,...L, j = 1,...,M$.

The objective of the paper is to obtain a method that can estimate θ without non-convex optimization.

3. WEIGHTED NULL-SPACE FITTING

The Weighted Null-Space Fitting (WNSF) method from Galrinho et al. (2019) and Galrinho (2018) is a multi-step least-squares method drawing on the seminal work of Ljung and Wahlberg (1992) where the properties of the least-squares method applied to ARX-models with increasing model order is analyzed. WNSF was originally developed for SISO systems, the origins of this method can be traced back to the work of Durbin (Durbin, 1959), see Galrinho (2018). Below we give a brief review which in the next section is followed by our extension to the network case.

The first step is an intermediate step, estimating a non-parametric (high-order) finite impulse response (FIR) or autoregressive exogenous (ARX) model. The second step aims to reduce this estimate to a parametric model using least squares. In the final step, the parametric model is re-estimated with weighted least squares. This leads to a consistent and asymptotically efficient estimate.

Step 1: Non-parametric model
For ease of notation superscript η is used to indicate the non-parametric model while θ is used for the parametric model. Consider a SISO output error (OE) model where the plant is defined by $G(q,\theta)$, assumed to have one delay. The stable rational function $G(q,\theta)$ can be well approximated using a non-parametric FIR model $y(t) = B^\eta u(t) + e(t)$ with $B^\eta = \sum_{k=1}^{n} b_k q^{-k}$, by choosing a sufficiently large order n. This model can be written in regressor form, resulting in the least-squares estimate

$$\hat{b}_N^n = \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^\top(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t). \tag{2}$$

where

$$\varphi(t) = [u(t-1)\ \cdots\ u(t-n)] \tag{3}$$

The estimation error can be expressed as $\epsilon_b^n = \hat{b}_N^n - g_0^n$, with g_0^n being the vector of the n first impulse response coefficients of the true system. This error has approximately a normal distribution with zero mean and covariance

$$P = \sigma_e^2\left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^\top(t)\right]^{-1}, \tag{4}$$

where σ_e^2 is the noise variance of $e(t)$.

Step 2: Reduction to parametric model
While the estimate in Step 1 is attractive from a computational point of view, its variance will typically be high since the number of parameters, n, has to be large in order for the approximation to be valid. Reducing the non-parametric model from the previous step to a parametric model allows us to reduce the variance. Neglecting the truncation error, the relation between the models is given by

$$B^\eta = \sum_{k=1}^{n} b_k q^{-k} = \frac{L^\theta}{F^\theta} = G^\theta \tag{5}$$

By rewriting (5) to

$$F^\theta \sum_{k=1}^{n} b_k q^{-k} - L^\theta = 0, \tag{6}$$

the relation can be rewritten to be linear in the parameters θ according to

$$b^n - Q(b^n)\theta = 0, \text{ with } Q(b^n) = \begin{bmatrix} -Q^b(b^n) & I_{m\times m} \\ & 0_{n-m\times m} \end{bmatrix}, \tag{7}$$

where $Q^b(b^n) = \mathcal{T}_{n\times n}\{[0 \;\; b_1 \;\; \cdots \;\; b_{n-1}]^\top\}$ is a lower triangular Toeplitz matrix with $[0\; b_1 \;\cdots\; b_{n-1}]$ in the first column. By substituting the estimate of b^n in $Q(b^n)$ an estimate of θ can be obtained from (7) via weighted least squares

$$\hat{\theta}_N = [Q^\top(\hat{b}_N^n)WQ(\hat{b}_N^n)]^{-1}Q^\top(\hat{b}_N^n)W\hat{b}_N^n \tag{8}$$

where W is either set to $W = I$ or $W = P^{-1}$ for obtaining the initial estimate $\hat{\theta}_N^{(0)}$. Notice that the unknown noise variance σ_e^2 appears in (4), but since it is a scalar quantity it can be discarded in the weighting matrix without affecting the estimate.

Step 3: Re-estimation of the parametric model
When substituting the estimates of b^n into the left-hand side of (7), the expression no longer equals zero but instead, neglecting the truncation error, it holds that

$$\hat{b}_N^n - Q(\hat{b}_N^n)\theta = T(\theta)\epsilon_b^n, \tag{9}$$

with $T(\theta) = \mathcal{T}_{n\times n}\{[1\; f_1 \;\cdots\; f_m]^\top\}$, depending on only the denominator coefficients of the parametric model. With ϵ_b^n being approximately zero mean normal with covariance matrix P, it follows that the right-hand side of (9) is approximately zero mean normal with covariance matrix $T(\theta)PT^\top(\theta)$. This correlation structure should be accounted for when estimating θ and can be handled using weighted least-squares in the following way. The aforementioned covariance matrix can be approximated using the estimate of θ obtained in Step 2 as $T(\hat{\theta}_N^{(0)})PT^\top(\hat{\theta}_N^{(0)})$. A new estimate of θ is formed by setting the weighting matrix W to the inverse of this matrix, i.e.

$$W = W(\hat{\theta}_N^{(0)}), \text{ where } W(\theta) := T^{-\top}(\theta)P^{-1}T^{-1}(\theta). \tag{10}$$

It is proven in (Galrinho et al., 2019) that both $\hat{\theta}_N^{(0)}$ and $\hat{\theta}_N^{(1)}$ are consistent estimates and in addition, $\hat{\theta}_N^{(1)}$ is proven to be asymptotic efficient. Since finite sample size N is used in practice it could be beneficial to iterate the last step, i.e. to use $\hat{\theta}_N^{(k)}$ in (10) and then computing $\hat{\theta}_N^{(k+1)}$ using (8), iterating until a final value or stopping criterion is reached, see for an example Section 5.

4. WEIGHTED NULL-SPACE FITTING FOR DYNAMIC NETWORKS

4.1 Algorithm

In this section WNSF is extended such that it is suitable for identifying all the modules in a dynamic network simultaneously. The steps described in the previous section can be followed.

Step 1: Non-parametric model
Consider the system defined in (1), where the rational functions of the separate modules in matrices $G(q,\theta)$ and $R(q,\theta)$ are defined as

$$G_{ij}(q,\theta) = \frac{L_{ij}(q,\theta)}{F_{ij}(q,\theta)}, \quad R_{ij}(q,\theta) = \frac{C_{ij}(q,\theta)}{D_{ij}(q,\theta)}. \tag{11}$$

with

$$\begin{aligned} L_{ij}(q,\theta) &= \quad\; l_1^{ij}q^{-1} \;+\cdots+ l_{m_l}^{ij}q^{-m_l}, \\ F_{ij}(q,\theta) &= 1 + f_1^{ij}q^{-1} \;+\cdots+ f_{m_f}^{ij}q^{-m_f}, \\ C_{ij}(q,\theta) &= c_1^{ij} + c_2^{ij}q^{-1} \;+\cdots+ c_{m_c}^{ij}q^{-m_c-1}, \\ D_{ij}(q,\theta) &= 1 + d_1^{ij}q^{-1} \;+\cdots+ d_{m_d}^{ij}q^{-m_d}. \end{aligned} \tag{12}$$

For ease of notation the orders m_f, m_l, m_d, m_c are assumed to be equal for the respective polynomials.

The parameter vector to estimate contains the rational functions of both $G(q,\theta)$ and $R(q,\theta)$ and has structure

$$\theta = [\theta^1 \;\cdots\; \theta^L]^\top \in \mathbb{R}^{L(m_f+m_l+m_d+m_c)}, \tag{13}$$

where

$$\theta^i = [f^{i1} \;\cdots\; f^{iL}\; l^{i1} \;\cdots\; l^{iL}\; d^{i1} \;\cdots\; d^{iL}\; c^{i1} \;\cdots\; c^{iL}], \tag{14}$$

$$\begin{aligned} f^{ij} &= [f_1^{ij} \;\cdots\; f_{m_f}^{ij}], \quad l^{ij} = [l_1^{ij} \;\cdots\; l_{m_l}^{ij}], \\ d^{ij} &= [d_1^{ij} \;\cdots\; d_{m_d}^{ij}], \quad c^{ij} = [c_1^{ij} \;\cdots\; c_{m_c}^{ij}]. \end{aligned} \tag{15}$$

Moreover, to satisfy the structure of $G(q,\theta)$ given in (1) the elements f^{ii} and l^{ii} corresponding to the diagonal elements are omitted for $i = 1,\ldots,L$ since these elements are zero.

In the first step a non-parametric ARX model is used

$$\bar{A}^\eta y(t) = B^\eta r(t) + e(t). \tag{16}$$

where

$$\bar{A}^\eta = I_L - \begin{bmatrix} 0 & A_{12}^\eta & \ldots & A_{1L}^\eta \\ A_{21}^\eta & 0 & \ddots & A_{2L}^\eta \\ \vdots & \ddots & \ddots & \vdots \\ A_{L1}^\eta & A_{L2}^\eta & \ldots & 0 \end{bmatrix}, \text{ with } A_{ij}^\eta = \sum_{k=1}^{n} a_k^{ij} q^{-k},$$

$$B^\eta \text{ with elements } B_{ij}^\eta = \sum_{k=0}^{n} b_k^{ij} q^{-k}, \tag{17}$$

where the parameter vector η^n is defined as

$$\begin{aligned} &\eta^n = [\eta^1 \;\cdots\; \eta^L]^\top, \\ &\text{with } \eta^i = [a^{i1} \;\cdots\; a^{iL}\; b^{i1} \;\cdots\; b^{iL}], \\ &\text{and } a^{ij} = [a_1^{ij} \;\cdots\; a_n^{ij}], \quad b^{ij} = [b_1^{ij} \;\cdots\; b_n^{ij}]. \end{aligned} \tag{18}$$

The dynamic network, described by the non-parametric model (16), can be expressed in regressor form

$$y(t) = \varphi^\top(t)\eta^n + e(t), \tag{19}$$

with

$$\varphi^\top(t) = \operatorname{diag}\big(\varphi_1^\top(t), \varphi_2^\top(t), \ldots, \varphi_L^\top(t)\big), \tag{20}$$

where the elements of $\varphi(t)$ are functions of delayed outputs and excitation signals

$$\varphi_i^\top(t) = [\varphi_y^\top(t) \quad \varphi_r^\top(t)], \tag{21}$$

with

$$\begin{aligned} \varphi_y^\top(t) &= [\varphi_{y_1}^\top(t), \ldots, \varphi_{y_L}^\top(t)], \\ \varphi_r^\top(t) &= [\varphi_{r_1}^\top(t), \ldots, \varphi_{r_M}^\top(t)] \end{aligned} \tag{22}$$

and

$$\begin{aligned} \varphi_{y_k}^\top(t) &= [y_k(t-1), \ldots, y_k(t-n)], \\ \varphi_{r_k}^\top(t) &= [r_k(t), \ldots, r_k(t-n-1)]. \end{aligned} \tag{23}$$

The least squares estimate of η^n is obtained similar to (2), replacing $\hat{b}_N^n$ with $\hat{\eta}_N^n$ and using $\varphi(t)$ defined in (20).

Furthermore, the estimation error covariance can be derived as was done in Section 3, see (4), but the noise

is no longer scalar. An appropriate approximation of the covariance matrix is given by

$$P(\hat{\Lambda}) = \Big[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\hat{\Lambda}^{-1}\varphi^{\top}(t)\Big]^{-1},$$
$$\hat{\Lambda} = \begin{bmatrix}\hat{\sigma}^2_{e_1} & & 0\\ & \ddots & \\ 0 & & \hat{\sigma}^2_{e_L}\end{bmatrix} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\hat{\eta}^n_N)\varepsilon^{\top}(t,\hat{\eta}^n_N), \tag{24}$$

with prediction error $\varepsilon(t,\hat{\eta}^n_N) = y(t) - \varphi^{\top}(t)\hat{\eta}^n_N$.

Step 2: Reduction to parametric model
The non-parametric model (16) is related to the parametric model (1) by $\bar{A}^{\eta} = 1 - G^{\theta}$ and $B^{\eta} = R^{\theta}$. Since the WNSF algorithm allows element-wise parameterization, the relation using (17) can be rewritten as $A^{\eta}_{ij} = G^{\theta}_{ij}$ and $B^{\eta}_{ij} = R^{\theta}_{ij}$ giving

$$A^{\eta}_{ij} = \sum_{k=1}^{n} a^{ij}_k q^{-k} = \frac{L^{\theta}_{ij}}{F^{\theta}_{ij}} = G^{\theta}_{ij},$$
$$B^{\eta}_{ij} = \sum_{k=0}^{n} b^{ij}_k q^{-k} = \frac{C^{\theta}_{ij}}{D^{\theta}_{ij}} = R^{\theta}_{ij}. \tag{25}$$

Rewriting this gives

$$\sum_{k=1}^{n} a^{ij}_k q^{-k} F^{\theta}_{ij} - L^{\theta}_{ij} = 0, \quad \sum_{k=0}^{n} b^{ij}_k q^{-k} D^{\theta}_{ij} - C^{\theta}_{ij} = 0. \tag{26}$$

The same procedure for determining $Q(\eta^n)$ from Section 3 can be followed, giving

$$Q(\eta^n) = \mathrm{diag}(Q^1,\ldots,Q^L),$$
$$Q^i(\eta^n) = \begin{bmatrix}Q^{f,l} & 0\\ 0 & Q^{d,c}\end{bmatrix}, \tag{27}$$

where $Q^{f,l}$ and $Q^{d,c}$ share the same structure, defined by

$$Q^{f,l} = \mathrm{diag}\big([-Q^{f^{i1}}Q^{l^{i1}}],\ldots,[-Q^{f^{iL}}Q^{l^{iL}}]\big)$$
$$Q^{d,c} = \mathrm{diag}\big([-Q^{d^{i1}}Q^{c^{i1}}],\ldots,[-Q^{d^{iL}}Q^{c^{iL}}]\big) \tag{28}$$

Moreover,

$$Q^{f^{ij}}(\eta^n) = \mathcal{T}_{n,m_f}\{[a^{ij}_1,\ldots,a^{ij}_L]^{\top}\} \quad Q^{l^{ij}}(\eta^n) = \bar{I}_{n\times m_l},$$
$$Q^{d^{ij}}(\eta^n) = \mathcal{T}_{n,m_d}\{[b^{ij}_1,\ldots,b^{ij}_L]^{\top}\} \quad Q^{c^{ij}}(\eta^n) = \bar{I}_{n\times m_c}. \tag{29}$$

where the Toeplitz matrices are lower triangular and $\bar{I}_{n\times m}$ has $I_{m\times m}$ at the top with zeros elsewhere.

The initial estimate $\hat{\theta}^{(0)}_N$ can now be obtained analogously to (8), using $Q(\hat{\eta}^n_N)$, and with $W = P^{-1}(\hat{\Lambda})$ as weighting.

Step 3: Re-estimation of the parametric model
With the initial estimate of θ obtained in the previous step, the noise variance estimate is updated according

$$\hat{\Lambda} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\hat{\theta}^{(0)}_N)\varepsilon^{\top}(t,\hat{\theta}^{(0)}_N), \tag{30}$$

obtaining the prediction error with

$$\varepsilon(t,\hat{\theta}^{(0)}_N) = y(t) - \begin{bmatrix}G(q,\hat{\theta}^{(0)}_N) & R(q,\hat{\theta}^{(0)}_N)\end{bmatrix}\begin{bmatrix}y(t)\\ r(t)\end{bmatrix}. \tag{31}$$

Again the same procedure for determining $T(\theta)$ from Section 3 can be followed. $T(\theta)$ is now defined by

$$T(\theta) = \mathrm{diag}\big(T^1(\theta),\ \ldots,\ T^L(\theta)\big)$$
$$T^i(\theta) = \mathrm{diag}\big(T^{f^{i1}}(\theta),\ldots,T^{f^{iL}}(\theta),T^{\,d^{i1}}(\theta),\ \ldots,T^{\,d^{iL}}(\theta)\big), \tag{32}$$

$$\text{where } T^{\,f^{ij}}(\theta) = \mathcal{T}_{n,n}\{[1,f^{ij}_1,\ldots,f^{ij}_L]^{\top}\},$$
$$T^{\,d^{ij}}(\theta) = \mathcal{T}_{n,n}\{[1,d^{ij}_1,\ldots,d^{ij}_L]^{\top}\}, \tag{33}$$

are lower triangular Toeplitz matrices. The weighting matrix can now be defined as $W = W(\hat{\theta}^{(0)}_N,\hat{\Lambda})$, where

$$W(\theta,\Lambda) := T^{-\top}(\theta)P^{-1}(\Lambda)T^{-1}(\theta). \tag{34}$$

Again analogously to (8) the new estimate $\hat{\theta}^{(1)}_N$ is obtained by using the newly defined W in (34) and $Q(\hat{\eta}^n_N)$ in (27). Similar to the SISO case, one can make the choice to continue iterating to improve accuracy.

Algorithm 1 The algorithm for WNSF suitable for general dynamic networks is constructed as

(1) estimate non-parametric ARX model (16), using least squares (2) to obtain $\hat{\eta}^n_N$,
(2) reduce the non-parametric model to a parametric model (8), using $Q(\hat{\eta}^n_N)$ defined in (27) and $W = P^{-1}(\hat{\Lambda})$ from (24), obtaining $\hat{\theta}^{(0)}_N$,
(3) improve the estimates by updating weighting matrix in (8) to $W(\hat{\theta}^{(0)}_N,\hat{\Lambda})$ according to (34) with $\hat{\Lambda}$ from (30), resulting in $\hat{\theta}^{(1)}_N$. Continuing to iterate gives $\hat{\theta}^{(k+1)}_N$.

4.2 Theoretical analysis

According to Remark 5 (Chapter 3) in Ljung and Wahlberg (1992) their asymptotic results for SISO ARX models can be extended to multivariate systems. The SISO case is therefore used to prove the asymptotic properties of the dynamic network by considering the element-wise notation. The data generating system, denoted in the element-wise ARX structure is given by

$$A^0_{ij}(q) = \sum_{k=1}^{\infty} a^{0,ij}_k q^{-k}, \qquad B^0_{ij}(q) = \sum_{k=0}^{\infty} b^{0,ij}_k q^{-k} \tag{35}$$

It should be observed that the term $b^{0,ij}_0$ is no longer necessarily assumed zero, since $R(q,\theta)$ does not have to contain a delay. Although this case is not covered in Assumption S1 of Ljung and Wahlberg (1992), Remark 4 (Chapter 2) is still true, meaning the impulse responses still decrease at a certain lowest rate and Assumption S1 still holds. In addition, the input is obtained in open loop, where the sequence $\{r(t)\}$ is assumed to be sufficiently exciting and independent of $\{e(t)\}$, where $\{e(t)\}$ is considered to be Gaussian white noise.

The assumptions and conditions from Galrinho (2018) and Ljung and Wahlberg (1992) therefore still hold. This indicates that Theorem 7.1 from Galrinho (2018), for applying WNSF to multivariate systems, also is valid for dynamic networks. Hence, for LTI dynamic networks subjected to white noise disturbances, WNSF obeys the following properties under assumptions of identifiability and sufficient richness of the excitation:

(i) $\hat{\theta}^{(0)}_N \to \theta_0$, as $N \to \infty$ w.p.1.,
(ii) $\hat{\theta}^{(1)}_N \to \theta_0$, as $N \to \infty$ w.p.1.,
(iii) $\sqrt{N}(\hat{\theta}^{(1)}_N - \theta_0) \sim As\mathcal{N}(0,M)$

with

$$M = \lim_{n\to\infty}[Q^\top(\eta_0^n)T^{-\top}(\theta_0)\bar{P}^{-1}(\Lambda)T^{-1}(\theta_0)Q(\eta_0^n)]^{-1}, \tag{36}$$

$$\text{where } \bar{P}^{-1}(\Lambda) = \bar{\mathbb{E}}\big[\varphi(t)\Lambda^{-1}\varphi^\top(t)\big]. \tag{37}$$

5. NUMERICAL ILLUSTRATION

In this section we compare the performance of Algorithm 1 with PEM initialized at the true parameters by means of a simulation study. PEM provides asymptotically efficient estimates when converging to a global minimum. The simulation results suggest that Algorithm 1 has the same large sample properties as PEM, supporting the theoretical considerations in the preceding section.

In the simulation we consider the following system

$$G(q,\theta) = \begin{bmatrix} 0 & 0 & 0 & G_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & G_{26} \\ G_{31} & 0 & 0 & 0 & G_{35} & 0 \\ 0 & G_{42} & 0 & 0 & 0 & 0 \\ G_{51} & 0 & 0 & 0 & 0 & G_{56} \\ 0 & 0 & G_{63} & 0 & 0 & 0 \end{bmatrix}, \tag{38}$$

$$R(q,\theta) = \text{diag}(R_{11}, R_{22}, R_{33}, R_{44}, R_{55}, R_{66}),$$

with the elements $G_{ij}(q,\theta) = \frac{L_{ij}}{F_{ij}}$ given by

$$\begin{bmatrix} L_{14} \\ L_{26} \\ L_{31} \\ L_{35} \\ L_{42} \\ L_{51} \\ L_{56} \\ L_{63} \end{bmatrix} = \begin{bmatrix} 0.13 & 0 \\ 0.11 & 0 \\ 0.13 & 0 \\ -0.67 & 0.44 \\ -0.11 & 0 \\ -0.30 & 0.28 \\ -0.32 & 0.19 \\ -0.11 & 0 \end{bmatrix} \begin{bmatrix} q^{-1} \\ q^{-2} \end{bmatrix}, \tag{39}$$

$$\begin{bmatrix} F_{14} \\ F_{26} \\ F_{31} \\ F_{35} \\ F_{42} \\ F_{51} \\ F_{56} \\ F_{63} \end{bmatrix} = 1 + \begin{bmatrix} -1.20 & 0.50 & 0 & 0 & 0 & 0 \\ 1.11 & 1.85 & -1.12 & 0.89 & -0.26 & 0.11 \\ -2.14 & 2.30 & -1.57 & 0.70 & -0.16 & 0 \\ -0.98 & 0.25 & 0 & 0 & 0 & 0 \\ -2.06 & 2.06 & -1.07 & 0.28 & 0 & 0 \\ -1.93 & 1.46 & -0.51 & 0.07 & 0 & 0 \\ -1.31 & 0.51 & 0 & 0 & 0 & 0 \\ -1.79 & 1.67 & -0.98 & 0.26 & 0 & 0 \end{bmatrix} \begin{bmatrix} q^{-1} \\ q^{-2} \\ q^{-3} \\ q^{-4} \\ q^{-5} \\ q^{-6} \end{bmatrix} \tag{40}$$

and the elements $R_{ij}(q,\theta)$ given by

$$R_{11} = \frac{1}{1+0.25q^{-1}}, \qquad R_{22} = \frac{0.20-0.40q^{-1}}{1-1.30q^{-1}+0.36q^{-2}},$$
$$R_{33} = \frac{1}{1+0.45q^{-1}}, \qquad R_{44} = \frac{0.50}{1+0.58q^{-1}},$$
$$R_{55} = \frac{1.40+0.45q^{-1}}{1+0.95q^{-1}}, \quad R_{66} = \frac{0.21}{1+0.73q^{-1}}, \tag{41}$$

where the parameters are rounded to two decimal places.

Table 1. Successful computation of estimates with Algorithm 1 given in % over sample sizes N

N	300	1212	4899	19797	80000
Algorithm 1	28%	61%	91%	100%	100%

The excitation signals $\{r(t)\}$ and noise sequences $\{e(t)\}$ are normally distributed with zero mean, where the variances of $\{r(t)\}$ are set to 1 and the variances of $\{e(t)\}$ are set to $\{3,\ 1,\ 2,\ 1,\ 3,\ 2.5\}$. The output data is gathered by defining the system as $(I - G(q,\theta))^{-1}$ with the input data given by $R(q,\theta)r(t) + e(t)$.

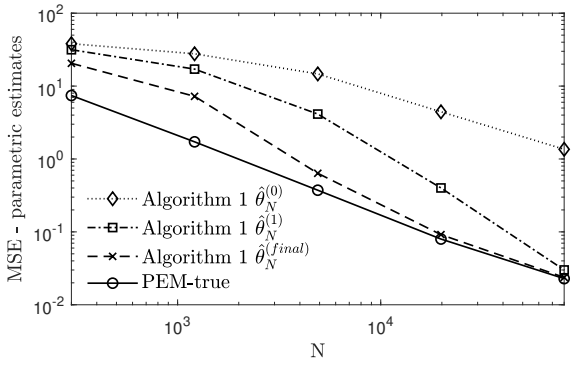

Fig. 1. MSE between $\hat{\theta}_N$ and θ_0 as function of sample size, averaged over the Monte Carlo runs

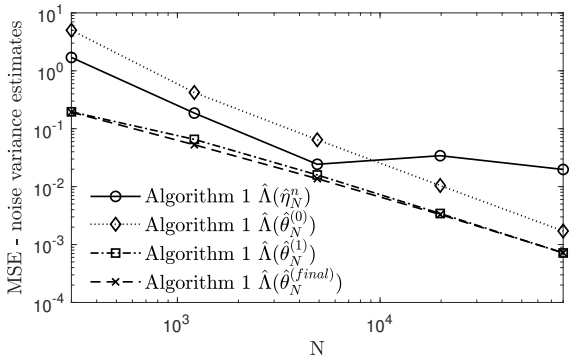

Fig. 2. MSE between $\hat{\Lambda}$ and Λ_0 as function of sample size, averaged over the Monte Carlo runs

In Algorithm 1 different model orders $n = \{20,\ 30,\ 40\}$ for the ARX model are tested and the n that results in the smallest prediction error is selected. Also implemented is the option to continue to iterate a 1000 times, the iteration stops earlier when $||\hat{\theta}_N^{(final)} - \hat{\theta}_N^{(final-1)}||/||\hat{\theta}_N^{(final-1)}|| < 0.0001$ is reached, where the final estimate for the iteration process $\hat{\theta}_N^{(k+1)}$ is indicated as $\hat{\theta}_N^{(final)}$. PEM is implemented using `idgrey` in the Matlab System Identification Toolbox, Matlab version 2018b and is initialized at the true parameters.

We perform $M = 100$ Monte Carlo runs for five integer sample sizes logarithmically spaced between 300 and 80000. Fig. 1 presents the sample mean-square error (MSE) as a function of sample size N, i.e. $\text{MSE}(N,k) = \frac{1}{M}\sum_{l=1}^{M}||\hat{\theta}_{N,l}^k - \theta_0||^2$, where $\hat{\theta}_{N,l}^k$ is the estimate in the k^{th} iteration of Monte Carlo run l. The figure shows the different steps of Algorithm 1 and PEM initialized at the true parameters. The results show that steps 2 and 3 of Algorithm 1 converge to PEM as the sample size N grows. Continuing to iterate is most beneficial for smaller N, where it improves accuracy the most. It should be

Table 2. Average computation time in seconds over sample sizes N

N	300	1212	4899	19797	80000
Algorithm 1	15.75	7.60	7.52	22.84	99.26
PEM-true	104.84	55.67	44.67	48.20	93.58

noted that Algorithm 1 does not handle small N well when the orders of the module transfer functions are increased, meaning it does not always successfully compute an estimate of θ. This success rate, for the simulation shown in Fig. 1, is presented in Table 1. Therefore, the MSE is only taken over the successful Monte Carlo runs.

Table 2 presents the average computation time for both methods. For Algorithm 1 the computation time is shown using $\hat{\theta}_N^{(final)}$. As seen from the table, the computation time for Algorithm 1 increases with the sample size. However, the code has not been optimized and we have seen that tailoring the code to the network topology can significantly reduce the computation time.

Fig. 2 shows the MSE for the noise variances $\text{MSE}_\sigma(N,\varepsilon) = \frac{1}{M}\sum_{l=1}^{M} ||\hat{\Lambda}_l(\varepsilon) - \Lambda_0||^2$, where only the diagonal elements $\sigma^2_{e_i}$ and its estimates are evaluated and where the prediction error ε depends on $\hat{\eta}_N^n$ or $\hat{\theta}_N^{(0,1,final)}$ for Monte Carlo run l. The estimated noise variances converge to the true noise variances as N increases, except for the noise variance computed with $\varepsilon(t,\hat{\eta}_N^n)$. A possible explanation for this may be that the truncation error becomes visible. Besides, it should be noted that there is no significant difference between deriving the noise variance using prediction error $\varepsilon(t,\hat{\theta}_N^{(1)})$ or $\varepsilon(t,\hat{\theta}_N^{(final)})$. The latter phenomenon also appears in other simulations we have tested. This suggests that the improvements in accuracy for $\hat{\theta}_N^{(k+1)}$ mainly comes from the improved estimate of $T(\theta_0)$ used in the weighting matrix $W(\hat{\theta}_N^{(k)},\hat{\Lambda})$ (34). Moreover, other simulation tests also seem to suggest the $R(q,\theta)$ is not restricted to be full rank.

The simulation results confirm the consistency of $\hat{\theta}_N^{(0,1,k+1)}$, derived in the preceding section. In addition, the results suggest that for a large sample size Algorithm 1 is asymptotically efficient for $\hat{\theta}_N^{(1,k+1)}$.

6. CONCLUSION

In this contribution we have presented an extension of WNSF that is tailored to identify LTI dynamic networks with known topology, without the need to solve non-convex optimization problems. The main assumption is that the nodes are subject to white noise. Simulations on a fairly challenging network indicate that the method is consistent and competitive with PEM (even in the idealized situation when this method is initialized at the true parameter values). An interesting extension under study is to handle colored noise.

REFERENCES

Dankers, A., Van den Hof, P., Bombois, X., and Heuberger, P. (2015). Errors-in-Variables identification in dynamic networks—consistency results for an instrumental variable approach. *Automatica*, 62, 39–50.

Dankers, A. (2014). *System Identification in Dynamic networks*. Ph.D. thesis, Delft University of Technology.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika*, 46(3/4), 306–316.

Everitt, N., Bottegal, G., and Hjalmarsson, H. (2018a). An empirical Bayes approach to identification of modules in dynamic networks. *Automatica*, 91, 144–151.

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

Galrinho , M. (2018). *System Identification with Multi-Step Least-Squares Methods*. Ph.D. thesis, KTH Royal Institute of Technology.

Galrinho, M., Everitt, N., and Hjalmarsson, H. (2017). Incorporating noise modeling in dynamic networks using non-parametric models. *IFAC-PapersOnLine*, 50(1), 10568 – 10573.

Galrinho, M., Rojas, C., and Hjalmarsson, H. (2019). Parametric identification using Weighted Null-Space Fitting. *IEEE Tran. on Automatic Control*, 64(7), 2798–2813.

Galrinho, M., Prota, R., Ferizbegovic, M., and Hjalmarsson, H. (2018). Weighted Null-Space Fitting for identification of cascade networks. *IFAC-PapersOnLine*, 51(15), 856–861.

Gevers, M., Bazanella, A.S., and Parraga, A. (2017). On the identifiability of dynamical networks. *IFAC PapersOnLine*, 50(1), 10580–10585.

Gonçalves, J. and Warnick, S. (2008). Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks. *IEEE Trans. on Automatic Control*, 53(7), 1670–1674.

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

Ljung, L. and 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. and Innocenti, G. (2010). Topological identification in networks of dynamical systems. *IEEE Trans. on Automatic Control*, 55(8), 1860–1871.

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

Söderström, T. and Stoica, P. (1983). Optimal instrumental variable estimation and approximate implementations. *IEEE Trans. on Automatic Control*, 28(7), 757–772.

Van den Hof, P. and Schrama, R. (1993). An indirect method for transfer function estimation from closed loop data. *Automatica*, 31(6), 1523–1527.

Van den Hof, P., Dankers, A., Heuberger, P., and 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.

Weerts, H., Dankers, A., and Van den Hof, P. (2015). Identifiability in dynamic network identification. *IFAC-PapersOnLine*, 48(28), 1409–1414.

Weerts, H., Van den Hof, P., and Dankers, A. (2016). Identifiability of dynamic networks with part of the nodes noise-free. *IFAC-PapersOnLine*, 49(13), 19–24.

Weerts, H., Galrinho, M., Bottegal, G., Hjalmarsson, H., and Van den Hof, P. (2018). A sequential least squares algorithm for ARMAX dynamic network identification. *IFAC-PapersOnLine*, 51(15), 844–849.

Weerts, H., Van den Hof, P., and Dankers, A. (2018a). Identifiability of linear dynamic networks. *Automatica*, 89, 247–258.

Weerts, H.H., den Hof, P.M.V., and Dankers, A.G. (2018b). Prediction error identification of linear dynamic networks with rank-reduced noise. *Automatica*, 98, 256 – 268.