Weighted Null-Space Fitting for Cascade Networks with Arbitrary Location of Sensors and Excitation Signals*

Mina Ferizbegovic[1], Miguel Galrinho[1] and Håkan Hjalmarsson[1]

Abstract— Identification of a complete dynamic network affected by sensor noise using the prediction error method is often too complex. One of the reasons for this complexity is the requirement to minimize a non-convex cost function, which becomes more difficult with more complex networks. In this paper, we consider serial cascade networks affected by sensor noise. Recently, the Weighted Null-Space Fitting method has been shown to be appropriate for this setting, providing asymptotically efficient estimates without suffering from non-convexity; however, applicability of the method was subject to some conditions on the locations of sensors and excitation signals. In this paper, we drop such conditions, proposing an extension of the method that is applicable to general serial cascade networks. We formulate an algorithm that describes application of the method in a general setting, and perform a simulation study to illustrate the performance of the method, which suggests that this extension is still asymptotically efficient.

I. INTRODUCTION

Dynamic networks have applications in fields such as robotics, transportation, social networks, and others. Identification of such systems gives rise to different challenges. One of the challenges is the detection of network topology [1], [2]. Another interesting area consists of determining conditions for network identifiability studied in [3]–[5]. Moreover, there is the problem of developing appropriate methods for identification of the systems in the network, which is the problem considered in this paper.

The problem of identifying a module of interest embedded in a dynamic network where all nodes are observed with process noise but the measurements are noise free has been studied in [6], [7]. Here, because of the presence of feedback, the main problem is to determine an appropriate set of predictor inputs that guarantee consistent estimates are obtained. When also sensor noise is present, using the prediction error method (PEM) with the noisy internal network variables as inputs to particular modules provides inconsistent estimates. Here, one possibility is to use instrumental variable (IV) methods, extending the errors-in-variables framework to dynamic networks [8]. Alternatively, the methods in [9], [10] can also handle this case, using external excitations and

intermediate non-parametric models, and can provide more accurate estimates than IV.

In this paper, we consider serial cascade networks. Between each module in the network, there is either a known external excitation or a measurement affected by sensor noise. This case differs from the aforementioned dynamic network formulations, as there is no process noise, and not all modules have necessarily a measured output. However, in the presence of sensor noise, one challenge is that obtaining asymptotically efficient estimates requires identifying the full network simultaneously. Although PEM can be applied to a tailor-made parametrization of this type of network, the non-convexity of the cost function is a concern. In [11], it was described how indirect PEM [12] can be useful to provide asymptotically efficient estimates, but only with models for which the predictor is linear in the parameters. In some cases, subspace methods [13] can be applied [14], [15], but they in general do not provide asymptotically efficient estimates.

In [16], it is proposed to use a method that provides asymptotically efficient estimates of all the cascade network modules without solving non-convex optimizations. The proposed method is based on Weighted Null-Space Fitting (WNSF), a multi-step weighted least-squares method proposed in [17], which has been shown to be consistent and asymptotically efficient for single-input single-output (SISO) systems [18]. However, the WNSF-based application to cascade networks in [16] has some limitations regarding the location of excitation signals and sensors. In this paper, we propose an extension based on WNSF for cascade networks that will drop these conditions.

Our contributions are the following. First, we propose a method based on an extension of WNSF for cascade networks that do not impose conditions on the location of sensors and excitation signals. Second, we perform a numerical simulation study, which suggests that the proposed method is asymptotically efficient.

Notation. Let x be a p-dimensional column vector. Then, $\mathcal{T}_{n \times m}\{x\}$ $(n \geq p, n \geq m)$ is the $n \times m$ Toeplitz matrix whose first column is $[x^\top \ 0_{1 \times n-p}]^\top$ and the first row is $[x_1 \ 0_{1 \times n-1}]^\top$, where x_1 is the first entry of x.

II. PROBLEM STATEMENT

We consider a serial cascade network with M modules. Each module is a stable transfer function $G_k(q)$, which is parametrized by a rational transfer function in the parameter vector θ_k:

$$G_k(q, \theta_k) = \frac{B_k(q, \theta_k)}{F_k(q, \theta_k)},$$

[1] Automatic Control Lab and ACCESS Linnaeus Center, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. (e-mail: {minafe, galrinho, hjalmars}@kth.se.)

where $B_k(q, \theta_k)$ and $F_k(q, \theta_k)$ are polynomials

$$
\begin{aligned}
F_k(q, \theta_k) &= 1 + f_1^{(k)} q^{-1} + \cdots + f_m^{(k)} q^{-m}, \\
B_k(q, \theta_j) &= b_0^{(k)} + b_1^{(k)} q^{-1} + \cdots + b_{m-1}^{(k)} q^{-(m-1)},
\end{aligned} \quad (1)
$$

with q being the time-shift operator, and

$$
\theta_k = \begin{bmatrix} f_1^{(k)} & \cdots & f_m^{(k)} & b_0^{(k)} & \cdots & b_{m-1}^{(k)} \end{bmatrix}^\top, \quad (2)
$$

where $k \in \{1, 2, \cdots, M\}$. Although each polynomial in every transfer function may have a different number of parameters or number of delays in the numerator, we assume that the structure is given by (1) for simplicity of notation.

There are a noisy measurements available, denoted $\{y_1(t), \ldots, y_a(t)\}$, and b known excitations signals, denoted $\{u_1(t), \ldots, u_b(t)\}$. Between each module there is either a known excitation signal (which is summed to the output of the previous module), a measured signal, or both. An excitation signal is available before the first module, and a measured output is available after the last module. The measured outputs are modeled as

$$
y(t) = G(q, \theta) u(t) + e(t),
$$

where

$$
y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_a(t) \end{bmatrix}, \quad u(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_b(t) \end{bmatrix}, \quad e(t) = \begin{bmatrix} e_1(t) \\ \vdots \\ e_a(t) \end{bmatrix},
$$

where $e(t)$ contains mutually uncorrelated white noise sequences with variances For example, for the network in Fig. 1, the transfer matrix is given by

$$
G(q) = \begin{bmatrix} G_1 G_2 & G_2 \\ G_1 G_2 G_3 & G_2 G_3 \end{bmatrix},
$$

where, inside the matrix, we have dropped the argument q for notational simplicity, which we will often do in this paper.

To guarantee identifiability, we assume that there exists a unique θ^o such that $G(q, \theta^\mathrm{o}) = G(q)$. For a two-module cascade network, specific identifiability conditions have been studied in [11]; for a more general treatment on identifiability in dynamic networks, see for example [3]–[5]. A deeper discussion on identifiability is outside the scope of this work, and we simply assume that the network is identifiable.

A key observation is that each measured output contains information about all transfer functions appearing before it. Then, the complete networks needs to be estimated simultaneously to obtain asymptotically efficient estimates. In this paper, we consider the problem of obtaining asymptotically efficient estimates without the need of minimizing a non-convex cost function.

III. THE PREDICTION ERROR METHOD

The prediction error method is a benchmark for identification of parametric models [19]. For this reason, we summarize this approach and consider its application to identify the type of networks we consider.
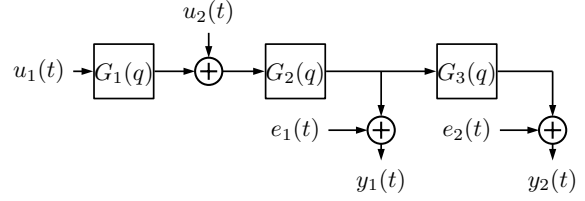

Fig. 1. Serial cascade network example.

The idea of PEM is to minimize a cost function of the prediction errors

$$
\varepsilon(t, \theta) = y(t) - G(q, \theta) u(t).
$$

Then, PEM with a quadratic cost consists of minimizing

$$
V(\theta) = \det \left[\frac{1}{N} \sum_{t=1}^{N} \varepsilon(t, \theta) \varepsilon^\top(t, \theta) \right],
$$

where N is the sample size.

The global minimizer $\hat\theta$ of this cost function is an asymptotically efficient estimate of θ when the noise is Gaussian [19]. The problem is that $V(\theta)$ is, in general, a non-convex cost function, for which optimization algorithms may converge to non-global minimum. To reduce the risk of converging to one such minimum, accurate initialization methods are required. Although this is the case also for SISO systems, the cascade case is of more concern. First, the problem complexity increases as the number of systems in the network increases. Second, methods that provide consistent estimates (and can be used to initialize PEM) in the SISO case may not be directly applicable to cascade networks. This motivates the development of methods that are appropriate to identify cascade network structures.

IV. WEIGHTED NULL-SPACE FITTING

Weighted Null-Space Fitting (WNSF), introduced in [17], is a method for identification of parametric models. In [18], the method has been proved to be consistent and asymptotically efficient for single-input single-output (SISO) systems. In [16], an approach based on WNSF has been proposed for a specific type of serial cascade networks, which we generalize in this paper. For this reason, we first review WNSF for SISO output-error (OE) models, which will be instrumental to our extension for identification of cascade networks. We introduce main concepts and notation, but detailed explanation and motivation for each step can be found in [17], [18].

Consider that the output is given by

$$
y(t) = \frac{B(q, \theta)}{F(q, \theta)} u(t) + e(t),
$$

where $F(q, \theta)$ and $B(q, \theta)$ are polynomials as (1) and θ a parameter vector as (2), but for a particular instance k. For this model structure, WNSF consists of the following three steps. In the first step, we estimate a non-parametric finite impulse response (FIR) model. Second, this estimate is reduced to a parametric model with least squares. Third, the parametric model is re-estimated with weighted least squares.

Step 1: Non-parametric model: In the first step, we consider the finite impulse response (FIR) model structure $y(t) = \varphi(t)^\top g + e(t)$, where $\varphi(t) = \begin{bmatrix} u(t) & u(t-1) & \cdots & u(t-n-1) \end{bmatrix}^\top$, and $g = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-1} \end{bmatrix}^\top$ is the parameter vector. An estimate of g is then obtained with least squares, computing

$$\hat{g} = \left[\frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varphi^\top(t) \right]^{-1} \left[\frac{1}{N} \sum_{t=1}^{N} \varphi(t)y(t) \right]. \quad (3)$$

Here, we assume that n has been chosen sufficiently large to capture the system dynamics with certain accuracy, meaning that the error made by truncation is negligible (see [18] for a formal treatment of this), the estimate g is distributed as

$$\sqrt{N}(\hat{g} - g^\mathrm{o}) \sim As\mathcal{N}(0, P),$$

where $\mathcal{N}$ stands for the normal distribution, g^o is a vector containing the first n true impulse response coefficients of the system, and

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

with λ being the noise variance.

Step 2: Estimation of OE Model: In the second step, we use the relation

$$\frac{B(q,\theta)}{F(q,\theta)} = \sum_{k=0}^{n-1} g_k q^{-k} \quad (5)$$

to obtain an estimate of θ from the estimate of $\hat{g}$ obtained in Step 1. This is done by re-writing (5) as

$$F(q,\theta) \sum_{k=0}^{n-1} g_k q^{-k} - B(q,\theta) = 0.$$

In turn, this can be re-written in matrix form as

$$g - Q(g)\theta = 0, \quad (6)$$

where

$$Q(g) = \begin{bmatrix} -\mathcal{T}_{n \times m}\{\Gamma_n g\} & I_{m \times m} \\ & 0_{n-m \times m} \end{bmatrix},$$

with

$$\Gamma_n = \begin{bmatrix} 0_{1 \times n-1} & 0 \\ I_{n-1} & 0_{n-1 \times 1} \end{bmatrix}.$$

We may replace g by the estimate $\hat{g}$ from Step 1, and solve for θ with least squares:

$$\hat{\theta}_{\mathrm{LS}} = \left[Q^\top(\hat{g})Q(\hat{g}) \right]^{-1} Q^\top(\hat{g})\hat{g}. \quad (7)$$

Step 3: Re-estimation of the OE Model: The idea of the third step is to use weighted least squares to re-estimate θ in a statistically sound way. The weighting matrix is chosen in the following way. By replacing the noisy estimate $\hat{g}$ in (6), the residuals we want to minimize can be written as

$$\hat{g} - Q(\hat{g})\theta = T(\theta)\Delta_g, \quad (8)$$

where $T(\theta) = \mathcal{T}_{n \times n}\{\begin{bmatrix} 1 & f_1 & \cdots & f_m \end{bmatrix}^\top\}$ and Δ_g is the noise in the non-parametric estimate $\hat{g}$. Then, the optimal weighting is the inverse of the covariance of the residuals (8). This covariance (i.e., the weighting inverse) is given by

$$W^{-1}(\theta) = T(\theta)PT^\top(\theta). \quad (9)$$

Because the true value of θ is not available to compute (9), we replace it with the estimate $\hat{\theta}_{\mathrm{LS}}$ obtained in Step 2. The noise variance λ in P (4) is also typically unknown, but because it is a scalar, it can be disregarded in the weighting. Then, we re-estimate θ using

$$\hat{\theta}_{\mathrm{WLS}} = \left[Q^\top(\hat{g})W(\hat{\theta}_{\mathrm{LS}})Q(\hat{g}) \right]^{-1} Q^\top(\hat{g})W(\hat{\theta}_{\mathrm{LS}})\hat{g}. \quad (10)$$

The estimate $\hat{\theta}_{\mathrm{WLS}}$ is an asymptotically efficient estimate of θ [18], but for finite sample size iterations may improve the estimation accuracy.

Remark 1. The following three properties are used to apply WNSF:
  1) the parametric model of interest can be approximated by a non-parametric model estimated with least squares;
  2) the non-parametric and parametric models can be related using the form (6);
  3) the residuals (8) are linear in the error of the non-parametric estimate, so that a closed-form expression for the covariance can be obtained.

V. WEIGHTED NULL-SPACE FITTING FOR SPECIFIC SERIAL CASCADE NETWORKS

In this section, we review the WNSF-based approach for cascade networks proposed in [16]. This review will be useful to introduce notation and to identify the main limitations of the approach, which will be used as motivation for the approach proposed in this paper. For illustration purposes, we use the network in Fig. 1 as an example, and follow the three SISO WNSF steps as baseline.

Step 1: Non-parametric model: For the cascade network in Fig. 1, we use the non-parametric MIMO FIR model

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \tilde{G}_{1,1}(q,g_{11}) & \tilde{G}_{1,2}(q,g_{12}) \\ \tilde{G}_{2,1}(q,g_{21}) & \tilde{G}_{2,2}(q,g_{22}) \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix},$$

where, for $i,j = \{1,2\}$, $\tilde{G}_{i,j}(q,g_{ij}) = \sum_{k=0}^{n-1} g_k^{(ij)} q^{-k}$, and $g_{ij} = \begin{bmatrix} g_0^{(ij)} & g_1^{(ij)} & \cdots & g_{n-1}^{(ij)} \end{bmatrix}^\top$. In the remainder of the paper, $\tilde{G}$ is used to emphasize that the transfer function is non-parametric, and a subscript consisting of two values separated by a comma indicate the row and column indexes in the transfer matrix. Then, the PEM estimate of $g = \begin{bmatrix} g_{11}^\top & g_{12}^\top & g_{21}^\top & g_{22}^\top \end{bmatrix}^\top$ is obtained with least squares using (3), but in this case with $y(t) = \begin{bmatrix} y_1(t) & y_2(t) \end{bmatrix}^\top$ and

$$\varphi(t) = \begin{bmatrix} \varphi_{11}(t) & 0 \\ \varphi_{12}(t) & 0 \\ 0 & \varphi_{21}(t) \\ 0 & \varphi_{22}(t) \end{bmatrix},$$

where

$$\varphi_{11}(t) = \begin{bmatrix} u_1(t) & u_1(t-1) & \cdots & u_1(t-n-1) \end{bmatrix}^\top = \varphi_{21}(t),$$

$$\varphi_{12}(t) = \begin{bmatrix} u_2(t) & u_2(t-1) & \cdots & u_2(t-n-1) \end{bmatrix}^\top = \varphi_{22}(t).$$

Now, the covariance of $\hat{g}$ is given by, similarly to (4) but for the MIMO case,

$$P = \left[\sum_{t=1}^{N} \varphi(t) \Lambda^{-1} \varphi^{\top}(t) \right]^{-1}, \qquad (11)$$

where Λ is a diagonal matrix with diagonal elements $\{\lambda_1, \lambda_2\}$.

Step 2: Estimation of Parametric Model: The second step consists of estimating θ from the non-parametric estimate. In this case, we have

$$\begin{bmatrix} G_1^{\theta} G_2^{\theta} & G_2^{\theta} \\ G_1^{\theta} G_2^{\theta} G_3^{\theta} & G_2^{\theta} G_3^{\theta} \end{bmatrix} = \begin{bmatrix} \tilde{G}_{1,1}^{g} & \tilde{G}_{1,2}^{g} \\ \tilde{G}_{2,1}^{g} & \tilde{G}_{2,2}^{g} \end{bmatrix}, \qquad (12)$$

where we used the simplified notation $G_k^{\theta} := G_k(q, \theta_k)$ and $\tilde{G}_{i,j}^{g} = \tilde{G}(q, g_{ij})$. In order to apply WNSF for estimation of (12), we need to satisfy all points in Remark 1. However, the second point is not satisfied (we have products of θ polynomials), so we cannot apply WNSF directly. To circumvent this issue, we re-write (12) as

$$\begin{bmatrix} \tilde{G}_{1,2}^{g} G_1^{\theta} & G_2^{\theta} \\ \tilde{G}_{2,2}^{g} G_1^{\theta} & \tilde{G}_{1,2}^{g} G_3^{\theta} \end{bmatrix} = \begin{bmatrix} \tilde{G}_{1,1}^{g} & \tilde{G}_{1,2}^{g} \\ \tilde{G}_{2,1}^{g} & \tilde{G}_{2,2}^{g} \end{bmatrix}, \qquad (13)$$

where we replaced some of the products of θ by non-parametric models. Replacing G_k^{θ} with the respective numerator and denominator, we can re-write (13) as

$$\begin{bmatrix} \tilde{G}_{1,1}^{g} F_1^{\theta} - \tilde{G}_{1,2}^{g} B_1^{\theta} & \tilde{G}_{1,2}^{g} F_2^{\theta} - B_2^{\theta} \\ \tilde{G}_{2,1}^{g} F_1^{\theta} - \tilde{G}_{2,2}^{g} B_1^{\theta} & \tilde{G}_{2,2}^{g} F_3^{\theta} - \tilde{G}_{1,2}^{g} B_3^{\theta} \end{bmatrix} = 0, \qquad (14)$$

which is linear in θ, and can be written as (6) with

$$Q(g) = \begin{bmatrix} Q_1(g) & 0 & 0 \\ 0 & Q_2(g) & 0 \\ Q_3(g) & 0 & 0 \\ 0 & 0 & Q_4(g) \end{bmatrix},$$

where

$$Q_1(g) = \begin{bmatrix} -\mathcal{T}_{n \times m}\{\Gamma_n g_{11}\} & \mathcal{T}_{n \times m}\{g_{12}\} \end{bmatrix},$$
$$Q_2(g) = \begin{bmatrix} -\mathcal{T}_{n \times m}\{\Gamma_n g_{12}\} & I_{m \times m} \\ & 0_{n-m \times m} \end{bmatrix},$$
$$Q_3(g) = \begin{bmatrix} -\mathcal{T}_{n \times m}\{\Gamma_n g_{21}\} & \mathcal{T}_{n \times m}\{g_{22}\} \end{bmatrix},$$
$$Q_4(g) = \begin{bmatrix} -\mathcal{T}_{n \times m}\{\Gamma_n g_{22}\} & \mathcal{T}_{n \times m}\{g_{12}\} \end{bmatrix}.$$

Then, the estimate $\hat{g}$ from Step 1 can be used to obtain an estimate of $\hat{\theta}_{\mathrm{LS}}$ with least squares, using (7).

Remark 2. Although not all equations defined in (14) need to be used to obtain consistent estimates of all transfer functions, discarding "redundant" equations has a cost in terms of efficiency.

Remark 3. The replacement from (12) to (13) to achieve linearity in θ is not unique. For example,

$$\begin{bmatrix} \tilde{G}_{1,2}^{g} G_1^{\theta} & G_2^{\theta} \\ \tilde{G}_{1,1}^{g} G_3^{\theta} & \tilde{G}_{1,2}^{g} G_3^{\theta} \end{bmatrix} = \begin{bmatrix} \tilde{G}_{1,1}^{g} & \tilde{G}_{1,2}^{g} \\ \tilde{G}_{2,1}^{g} & \tilde{G}_{2,2}^{g} \end{bmatrix} \qquad (15)$$

is also possible. As seen in paper [16], simulations support the chosen replacement—in this case, from (12) to (13) or from (12) to (15)—does not affect the asymptotic properties of the estimate as long as we use all equations.

Step 3: Re-estimation of Parametric Model: In the third step, we re-estimate θ with weighted least squares. The residuals of (14) can be written linearly in the non-parametric estimate noise, as in (8). In this case, $T(\theta)$ is given by

$$T(\theta) = \begin{bmatrix} T_{11}(\theta) & 0_{2n \times 2n} \\ T_{21}(\theta) & T_{22}(\theta) \end{bmatrix},$$

where

$$T_{11}(\theta) = \begin{bmatrix} \mathcal{T}_{n \times n}\{f^{(1)}\} & -\mathcal{T}_{n \times n}\{b^{(1)}\} \\ 0_{n \times n} & \mathcal{T}_{n \times n}\{f^{(2)}\} \end{bmatrix},$$
$$T_{21}(\theta) = \begin{bmatrix} 0_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & -\mathcal{T}_{n \times n}\{b^{(3)}\}\} \end{bmatrix},$$
$$T_{22}(\theta) = \begin{bmatrix} \mathcal{T}_{n \times n}\{f^{(1)}\} & -\mathcal{T}_{n \times n}\{b^{(1)}\} \\ 0_{n \times n} & \mathcal{T}_{n \times n}\{f^{(3)}\} \end{bmatrix},$$

with $f^{(k)} = [1 \; f_1^{(k)} \; \cdots \; f_m^{(k)}]^{\top}$ and $b^{(k)} = [b_0^{(k)} \; \cdots \; b_{m-1}^{(k)}]^{\top}$, while the covariance of Δ_g is given by (11). Then, we denote by $\hat{P}$ the matrix (11) with Λ replaced by $\hat{\Lambda}$ that is estimated from data using $\hat{\Lambda} = \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t, \hat{\theta}_{\mathrm{LS}}) \varepsilon^{\top}(t, \hat{\theta}_{\mathrm{LS}})$. We can re-estimate θ with weighted least squares using (10), where

$$W^{-1}(\theta) = T(\theta) \hat{P} T^{\top}(\theta).$$

This approach, suggested in [16], can be seen as a direct extension of WNSF, in the sense that all properties in Remark 1 are satisfied after the adequate replacements of parametric with non-parametric transfer functions. This suggests, along with simulation results in [16], that this approach is consistent and asymptotically efficient. Moreover, a possible proof for this asymptotic properties would follow the same steps of the proof in [18] for the SISO case.

However, this approach covers only serial cascade network where all excitation signals appear prior to all measurements. Consider the cascade network in Fig. 2. In this case, the relation between the parametric and non-parametric models would be given by

$$\begin{bmatrix} G_1^{\theta} & 0 \\ G_1^{\theta} G_2^{\theta} G_3^{\theta} & G_3^{\theta} \end{bmatrix} - \begin{bmatrix} \tilde{G}_{1,1}^{g} & 0 \\ \tilde{G}_{2,1}^{g} & \tilde{G}_{2,2}^{g} \end{bmatrix} = 0. \qquad (16)$$

Although (16) can be made linear in θ by replacing $G_1^{\theta} G_2^{\theta} G_3^{\theta} = \tilde{G}_{2,2}^{g} G_2^{\theta} \tilde{G}_{1,1}^{g}$, a closed-form expression for the weighting is not available, because it will not be possible to write the residuals of (16) linearly in the non-parametric estimate noise (the third point in Remark 2 is not satisfied). This is a consequence of the product $\tilde{G}_{1,1}^{g} \tilde{G}_{2,2}^{g}$.

In [16], it was suggested that this problem could be solved by using an intermediate estimation step. The idea is to, in the first step, use WNSF to obtain estimates of $G_1^{\bar{\theta}}$, $G_{12}^{\bar{\theta}}$ and $G_3^{\bar{\theta}}$, according to the relations

$$\begin{bmatrix} G_1^{\bar{\theta}} & 0 \\ \tilde{G}_{2,2}^{g} G_{12}^{\bar{\theta}} & G_3^{\bar{\theta}} \end{bmatrix} - \begin{bmatrix} \tilde{G}_{1,1}^{g} & 0 \\ \tilde{G}_{2,1}^{g} & \tilde{G}_{2,2}^{g} \end{bmatrix} = 0.$$

where $G_{12}^{\bar{\theta}}$ is a parametric over-parametrization of the product $G_1 G_2$. In the second step, $G_{12}^{\bar{\theta}}$ can be used to re-apply WNSF and obtain estimates G_1^{θ}, G_2^{θ}, and G_3^{θ}.

The main limitation of this approach is that it is requires an additional WNSF step, which makes it computationally

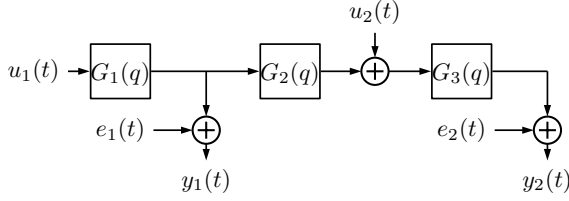Fig. 2. Another serial cascade network example.

Fig. 3. Serial cascade network used for simulation.

heavier. Here, we will propose an alternative approach that does not require an additional step.

VI. EXTENSION OF WNSF FOR GENERAL SERIAL CASCADE NETWORKS

Recall that application of the WNSF-based method in [16] to the network in Fig. 2 was hindered by the third property in Remark 1 not being satisfied: the residuals cannot be made linear in the noise, and there is no closed-form expression for the optimal weighting matrix. In this section, we propose a method that approximates these residuals, avoiding the use of an intermediate step. Despite the approximation, simulation results in the next section support that the method is still asymptotically efficient.

A. An Example

Consider the network in Fig. 2. Then, we propose the following approach to identify the network transfer functions.

Step 1: Non-parametric model: In the first step, we obtain a non-parametric model estimate, identically to Step 1 in Section V.

Step 2: Estimation of Parametric Model: In this step, we estimate the parametric model from the non-parametric estimate in Step 1. In order to make (16) linear in θ, we can rewrite it as

$$\begin{bmatrix} \tilde{G}_{1,1}^g F_1^\theta - B_1^\theta & 0 \\ \tilde{G}_{2,1}^g F_2^\theta - \tilde{G}_{2,2}^g \tilde{G}_{1,1}^g B_2^\theta & \tilde{G}_{2,2}^g F_3^\theta - B_3^\theta \end{bmatrix} = 0. \quad (17)$$

We denote $\tilde{G}_{2,2}^g \tilde{G}_{1,1}^g = \sum_{k=1}^\infty g_{13}^{(k)} q^{-k}$, and we let $g_{13} = [g_{13}^{(1)} \dots g_{13}^{(n)}]^\top$. Then, we can write (17) as (6) with

$$Q(g) = \begin{bmatrix} Q_1(g) & 0 & 0 \\ 0 & Q_2(g) & 0 \\ 0 & 0 & Q_3(g) \end{bmatrix},$$

where

$$Q_1(g) = \begin{bmatrix} -\mathcal{T}_{n\times m}\{\Gamma_n g_{11}\} & I_{m\times m} \\ & 0_{n-m\times m} \end{bmatrix},$$
$$Q_2(g) = \begin{bmatrix} -\mathcal{T}_{n\times m}\{\Gamma_n g_{21}\} & \mathcal{T}_{n\times m}\{\Gamma_n g_{13}\} \end{bmatrix},$$
$$Q_3(g) = \begin{bmatrix} -\mathcal{T}_{n\times m}\{\Gamma_n g_{22}\} & I_{m\times m} \\ & 0_{n-m\times m} \end{bmatrix}.$$

The estimate $\hat{g}$ from Step 1 can now be used to obtain an estimate $\hat{\theta}_{\text{LS}}$ with least squares, using (7).

Step 3: Re-estimation of Parametric Model: Although (17) is linear in θ, it is not linear in g, and it will not be possible to write the residuals linearly in the non-parametric estimation noise. So, we will neglect the non-linear terms in the residuals:
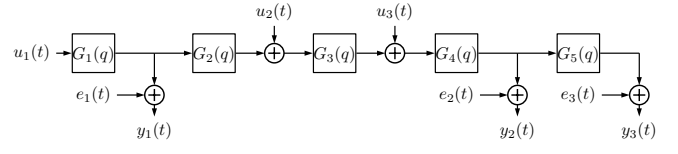
when estimates $\hat{\tilde{G}}_{i,j}^g$ are plugged in (17), we approximate the corresponding residuals for element $(2,1)$ as

$$\begin{aligned} &F_2^\theta \hat{\tilde{G}}_{2,1}^g - B_2^\theta \hat{\tilde{G}}_{1,1}^g \hat{\tilde{G}}_{2,2}^g \\ &= F_2^\theta(\tilde{G}_{2,1}^g + \Delta_{2,1}^g) - B_2^\theta(\tilde{G}_{2,2}^g + \Delta_{2,2}^g)(\tilde{G}_{1,1}^g + \Delta_{1,1}^g) \\ &= F_2^\theta \Delta_{2,1}^g - B_2^\theta \tilde{G}_{2,2}^g \Delta_{1,1}^g - B_2^\theta \tilde{G}_{1,1}^g \Delta_{2,2}^g - B_2^\theta \Delta_{1,1}^g \Delta_{2,2}^g \\ &\approx F_2^\theta \Delta_{2,1}^g - B_2^\theta \tilde{G}_{2,2}^g \Delta_{1,1}^g - B_2^\theta \tilde{G}_{1,1}^g \Delta_{2,2}^g, \end{aligned}$$
$$(18)$$

where $\Delta_{i,j}^g$ are polynomials in q^{-1} whose coefficients contain the estimation noise in $\hat{\tilde{G}}_{i,j}^g$, and whose covariance follows from the least-squares estimate. As with SISO WNSF, (18) can be written in vector form, and the true parameters replaced with an estimate obtained in a previous iteration. The difference here is that not only the parametric coefficients θ appear, but also the non-parametric coefficients g. Although they could be replaced with the estimated g coefficients, these have high variance; in order to obtain a more accurate estimate of the residual covariance, we may replace $\tilde{G}_{1,1}^g$ and $\tilde{G}_{2,2}^g$ by the corresponding parametric descriptions: the impulse response coefficients of G_1^θ and G_3^θ, respectively. We denote the first n coefficients of $B_2^\theta G_1^\theta$ and $B_2^\theta G_3^\theta$ by η_1 and η_3, respectively. Now, we can express $T(\theta)$ as

$$T(\theta) = \begin{bmatrix} \mathcal{T}_{n\times n}\{f^{(1)}\} & 0_{n\times n} & 0_{n\times n} \\ \mathcal{T}_{n\times n}\{\eta_3(\theta)\} & \mathcal{T}_{n\times n}\{f^{(2)}\} & \mathcal{T}_{n\times n}\{\eta_1(\theta)\} \\ 0_{n\times n} & 0_{n\times n} & \mathcal{T}_{n\times n}\{f^{(3)}\} \end{bmatrix}.$$

Then, we re-estimate θ by using equation (10) where W is obtained from (9).

As we will see in Section VII, a simulation study suggests that, despite the approximation of the residuals statistics, the estimate $\hat{\theta}_{\text{WLS}}$ is an asymptotically efficient estimate of θ for the considered problem.

VII. SIMULATIONS

Let us consider the network shown in Fig. 3 with

$$G_1(q) = \frac{0.7q^{-1} + 0.5q^{-2}}{1 - 1.2q^{-1} + 0.5q^{-2}}, \quad G_2(q) = \frac{0.6 - 0.2q^{-1}}{1 - 1.3q^{-1} + 0.6q^{-2}},$$
$$G_3(q) = \frac{0.6 + 0.8q^{-1} - 1.2q^{-2}}{1 - 0.75q^{-1} + 0.56q^{-2}},$$
$$G_4(q) = \frac{0.6 + 0.8q^{-1}}{1 - 1.1q^{-1} + 0.7q^{-2}}, \quad G_5(q) = \frac{0.6 + 0.7q^{-1}}{1 - 1.2q^{-1} + 0.4q^{-2}}.$$

We use colored input given by $u_i(t) = (1 - 0.8q^{-1})^{-1}e_i^u(t)$, where $\{e_1^u(t), e_2^u(t)\}$ are unit-variance mutually-uncorrelated Gaussian white-noise sequences. The noises $\{e_1(t), e_2(t)\}$ have unitary variance. We performed 100 Monte Carlo runs for seven different sample sizes N logarithmically spaced between 500 and 50000.
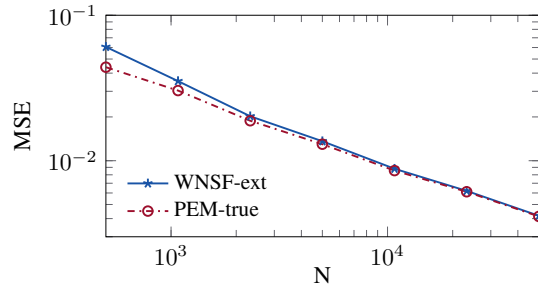
Fig. 4. Average MSE as function of sample size.

There is not a straightforward procedure to initialize PEM applied to this network structure. One reasonable way to obtain initialization estimates, which was used in the simulation in [16], is to identify a MIMO OE model with PEM and use model reduction to recover the individual network transfer functions. With this initialization, PEM often failed to attain the global minimum for the network in Fig. 1. Therefore, with the more complex network in Fig. 3 used in this simulation, we will not focus on the robustness of the PEM, but on illustrating that the method we propose, despite the approximation used, is still asymptotically efficient.

Then, we compare the following methods:

- PEM initialized at the true model parameters (PEM-true);
- the proposed extension of weighted null-space fitting using approximation of the non-parametric noise statistics (WNSF-ext).

For PEM, the network structure is implemented with the MATLAB function `greyest`, and then minimized using the `pem` function, with a maximum of 1000 iterations. For WNSF, we apply the algorithm for a grid of non-parametric orders $n = \{30, 50, 70\}$ and a maximum of 100 iterations.

In Fig. 4, we plot the obtained average mean-square error (MSE $= ||\hat{\theta} - \theta_{\mathrm{o}}||^2$) as function of the sample size N, where $\hat{\theta}$ is the estimate obtained for a particular method and sample size. This simulation suggests that the proposed method (WNSF-ext) has the same asymptotic performance as PEM initialized at the true parameters (PEM-true), supporting that the method is asymptotically efficient.

VIII. CONCLUSION

In this paper, we considered a particular case of dynamic networks: serial cascade networks, where at each node either a known external excitation or a measurement affected by sensor noise (i.e., not all nodes are necessarily measured). Although PEM can be applied to solve this kind of the problem, the non-convexity of the cost function is a concern. The Weighted Null-Space Fitting method has been shown to be appropriate for this setting [16]: it does not suffer from non-convexity, and similarities with the SISO case and simulation studies support that the method is consistent and asymptotically efficient under Gaussian noise. However, applicability of the method in [16] was subject to some conditions on the locations of sensors and excitation signals.

In this paper, we proposed an extension based on WNSF for the same type of networks, but which is applicable with arbitrary locations of excitation external signals and sensors.

For the extension, we use the approximation of the noise statistics, which a simulations supports does not affect the asymptotic properties.

For future work, we will perform a theoretical analysis to show consistency and asymptotic efficiency of the method proposed in this paper. Also, we will propose WNSF-based approaches to other networks besides serial cascade, and include colored noise.

REFERENCES

[1] D. Materassi and G. Innocenti, "Topological identification in networks of dynamical systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1860–1871, 2010.

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

[3] H. H. Weerts, A. G. Dankers, and P. M. Van den Hof, "Identifiability in dynamic network identification," in *17th IFAC Symposium on System Identification*, 2015, pp. 1409–1414.

[4] H. H. Weerts, P. M. Van den Hof, and A. G. Dankers, "Identifiability of dynamic networks with part of the nodes noise-free," in *12th IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, 2016, pp. 19–24.

[5] M. Gevers, A. S. Bazanella, and A. Parraga, "On the identifiability of dynamical networks," in *20th IFAC World Congress*, 2017, pp. 10 580–10 585.

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

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

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

[9] N. Everitt, G. Bottegal, C. R. Rojas, and H. Hjalmarsson, "Identification of modules in dynamic networks: An empirical bayes approach," in *55th IEEE Conference on Decision and Control*, 2016, pp. 4612–4617.

[10] M. Galrinho, N. Everitt, and H. Hjalmarsson, "Incorporating noise modeling in dynamic networks using non-parametric models," in *20th IFAC World Congress*, 2017, pp. 4612–4617.

[11] B. Wahlberg, H. Hjalmarsson, and J. Mårtensson, "Variance results for identification of cascade systems," *Automatica*, vol. 45, no. 6, pp. 1443–1448, 2009.

[12] T. Söderström, P. Stoica, and B. Friedlander, "An indirect prediction error method for system identification," *Automatica*, vol. 27, no. 1, pp. 183–188, 1991.

[13] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory, Implementation, Applications*. Springer Science & Business Media, 2012.

[14] B. Wahlberg and H. Sandberg, "Cascade structural model approximation of identified state space models," in *47th IEEE Conf. on Decision and Control*, 2008, pp. 4198–4203.

[15] P. Hägg, B. Wahlberg, and H. Sandberg, "On subspace identification of cascade structured systems," in *49th IEEE Conf. on Decision and Control*, 2010, pp. 2843–2848.

[16] M. Galrinho, R. Prota, M. Ferizbegovic, and H. Hjalmarsson, "Weighted null-space fitting of cascade networks," in *18th IFAC Symposium on System Identification*, 2018.

[17] M. Galrinho, C. R. Rojas, and H. Hjalmarsson, "A weighted least-squares method for parameter estimation in structured models," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 3322–3327.

[18] ——, "Parametric identification with weighted null-space fitting," *submitted to Trans. on Automatic Control (arXiv:1708.03946)*, 2017.

[19] L. Ljung, *System Identification. Theory for the User*. Prentice-Hall, 1999.