Identification of modules in dynamic networks: An empirical Bayes approach

Niklas Everitt, Giulio Bottegal, Cristian R. Rojas, and Håkan Hjalmarsson

***Abstract*— We address the problem of identifying a specific module in a dynamic network, assuming known topology. We express the dynamics by an acyclic network composed of two blocks where the first block accounts for the relation between the known reference signals and the input to the target module, while the second block contains the target module. Using an empirical Bayes approach, we model the first block as a Gaussian vector with covariance matrix (kernel) given by the recently introduced stable spline kernel. The parameters of the target module are estimated by solving a marginal likelihood problem with a novel iterative scheme based on the Expectation-Maximization algorithm. Numerical experiments illustrate the effectiveness of the proposed method.**

I. Introduction

Recently, large interconnected dynamic systems have gained popularity in the system identification community [1], [2], [3]. These systems, usually referred to as *dynamic networks*, find application in different branches of science, such as econometrics, systems biology, social science, and power systems.

In this paper, similarly to [2], we define a dynamic network as the result of the interconnection of *modules*, where each module is a linear time-invariant (LTI) system. The interconnecting signals are the outputs of these modules. In a graph interpretation, the interconnecting signals represent nodes and the modules represent edges. Moreover, we assume that exogenous measurable signals may affect the dynamics of the network.

Two problems arise in dynamic network identification. The first is unraveling the network topology (i.e., identify the edges of the graph), and can be seen as a model structure selection problem. The second is the identification of one or more specific modules in the network.

Some of the recent papers deal with both the aforementioned problems [4], [5] [1], [6], whereas others are mainly focused on the identification of a single module, see [7] [8], [9], [10], [11]. In particular, [7], [2] study the problem of understanding which of the available output measurements should be used to obtain a consistent estimate of a target module. In [9] instead, errors-in-variables dynamic networks are considered, and methods that lead to consistent module estimates are proposed. As observed in [2], dynamic networks with known topology can be seen as a generalization of simple compositions, such as cascaded systems or closed-loop systems. Therefore, identification techniques for dynamic networks may be derived by extending methods already developed for simple structures. This is the idea underlying the method presented in [2], which generalizes the two-stage method, originally developed for closed-loop systems, to dynamic networks [12]. Instrumental variable methods for closed-loop systems [13] are adapted to networks in [9]. Similarly, the methodology proposed in [14] for the identification of cascaded systems is generalized to the context of dynamic networks in [8]. Here, the underlying idea is that a dynamic network can be transformed into an acyclic structure, where any reference signal of the network becomes the input to a cascaded system consisting of two LTI blocks. The first block captures the relation between the reference and the noisy input of the target module, the second one contains the target module. The two LTI blocks are identified simultaneously using the prediction error method (PEM) [15]. In this setup, determining the model structure of the first block of the cascaded structure may be complicated, due to the possibly large number of interconnections in the dynamic network. Therefore, in [8] an unstructured FIR model is utilized for the first block. The major drawback of this approach is that, if the number of available measurements is small, the estimated FIR model may suffer of high variance, affecting the accuracy of the estimated target module.

The objective of this paper is to propose a method for the identification of a module in dynamic networks that circumvents the high variance issue. Following a recent trend in system identification, we use regularization to control the variance [16]. In particular, exploiting the equivalence between regularization and Gaussian process regression [17], we model the impulse response of the first block as a zero-mean stochastic process. The covariance matrix is given by the recently introduced *first-order stable spline kernel* [18], whose structure is parametrized by two *hyperparameters*. An estimate of the target module is then obtained by empirical Bayes (EB) arguments, that is, by maximization of the marginal likelihood of the available measurements [17]. This likelihood depends not only on the parameter of the target module, but also on the kernel hyperparameters and the variance of the measurement noise. Therefore, it is required to estimate all these quantities. This is done by designing a novel iterative solution scheme based on an EM-type algorithm [19], known as the Expectation/Conditional-Maximization (ECM) algorithm [20]. Combining the ECM algorithm with gradient descent strategies, we derive a computationally efficient scheme for solving the marginal likeli-

The authors are with the ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden (neveritt;bottegal;hjalmars@kth.se). This work was supported by the European Research Council under the advanced grant LEARN, contract 267381, and by the Swedish Research Council under contract 621-2009-4017.

hood problem, which provides the estimated target module. The effectiveness of the proposed method is demonstrated through numerical experiment. The method proposed in this paper is close in spirit to some recently proposed kernel-based techniques for blind system identification [21] and Hammerstein system identification [22].

The paper is organized as follows. In the next section, we introduce the dynamic network model and we give the problem statement. In Section III we present the identification strategy. In Section IV, we describe the solution scheme based on the ECM algorithm. Section V reports the results of a Monte Carlo experiment. Some conclusions end the paper.

II. Problem Statement

A. Dynamic networks

We consider dynamic networks that consist of L scalar *internal variables* $w_j(t)$, $j = 1, \dots, L$ and L scalar external *reference signals* $r_l(t)$, $l = 1, \dots, L$, that can be manipulated by the user. Some of the reference signal may not be present, i.e., they may be identically zero. Define $\mathcal{R}$ as the set of indices of reference signals that are present. In the dynamic network, the internal variables are considered nodes and transfer functions are the edges. Introducing the vector notation $w(t) := [w_1(t) \dots w_L(t)]^T$, $r(t) := [r_1(t) \dots r_L(t)]^T$, the dynamics of the network are defined by the equation

$$w(t) = \mathcal{G}(q)w(t) + r(t)\,, \tag{1}$$

where

$$\mathcal{G}(q) = \begin{bmatrix} 0 & G_{12}(q) & \cdots & G_{1L}(q) \\ G_{21}(q) & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & G_{(L-1)L}(q) \\ G_{L1}(q) & \cdots & G_{L(L-1)}(q) & 0 \end{bmatrix},$$

where $G_{ji}(q)$ is a proper rational transfer function for $j = 1, \dots, L$, $i = 1, \dots, L$. The internal variables $w(t)$ are measured with additive white noise, that is

$$\tilde{w}(t) = w(t) + e(t)\,,$$

where $e(t) \in \mathbb{R}^L$ is a stationary zero-mean Gaussian white-noise process with diagonal noise covariance matrix $\Sigma_e = \operatorname{diag}\{\sigma_1^2, \dots, \sigma_L^2\}$. We assume that the σ_i^2 are unknown. To ensure stability and causality of the network the following assumptions hold for all networks considered in this paper.

Assumption 2.1: The network is well posed in the sense that all principal minors of $\lim_{z\to\infty}(I - \mathcal{G}(z))$ are non-zero [2].

Assumption 2.2: The sensitivity path $S(q)$ is stable

$$S(q) := (I - \mathcal{G}(q))^{-1}\,.$$

Assumption 2.3: The reference variables $\{r_l(t)\}$ are mutually uncorrelated and uncorrelated with the measurement noise $e(t)$.

Thus, we can write

$$\tilde{w}(t) = S(q)r(t) + e(t)\,. \tag{2}$$

We define a $\mathcal{N}_j$ as the set of indices of internal variables that have a direct causal connection to w_j, i.e., $i \in \mathcal{N}_j$ if and only if $G_{ji}(q) \neq 0$. Without loss of generality, we assume that $\mathcal{N}_j = 1, 2, \dots, p$, where p is the number of direct causal connections to w_j (we may always rename the nodes so that this holds). The goal is to identify module $G_{j1}(q)$ given N measurements of the reference $r(t)$, the "output" $\tilde{w}_j(t)$ and the set of p neighbor signals in $\mathcal{N}_j$. To this end, we express $\tilde{w}_j$, the measured output of module $G_{j1}(q)$ as

$$\tilde{w}_j(t) = \sum_{i\in\mathcal{N}_j} G_{ji}(q)w_i(t) + r_j(t) + e_j(t)\,. \tag{3}$$

The above equation depends on the internal variables $w_i(t)$, $i \in \mathcal{N}_j$, which we we only have noisy measurement of; these can be expressed as

$$\tilde{w}_i(t) = w_i(t) + e_i(t) = \sum_{l\in\mathcal{R}} S_{il}(q)r_l(t) + e_i(t)\,. \tag{4}$$

where $S_{il}(q)$ is the transfer function path from reference $r_l(t)$ to output $\tilde{w}_i(t)$. Together, (3) and (4) allows us to express the relevant part of the network as a direct acyclic graph with two blocks connected in cascade.

Example 2.1: As an example consider the network depicted in Figure 1, where, using (3) and (4), the acyclic graph of Figure 2 can describe the relevant dynamics, when $w_j = w_3$ is the output and we wish to identify $G_{31}(q)$.

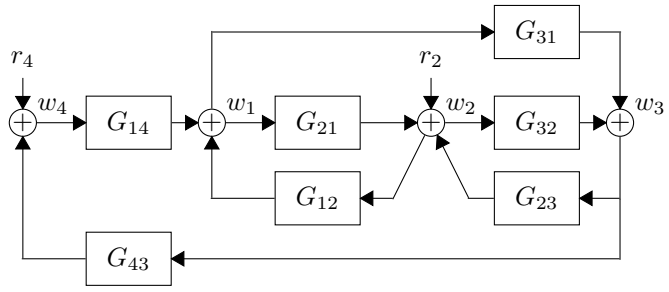

Fig. 1: Network example of 4 internal variables and 2 reference signals.

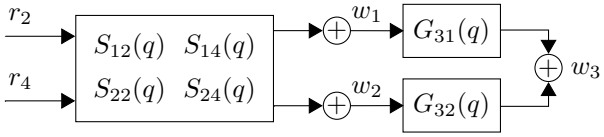

Fig. 2: Direct acyclic graph of part of the network in Figure 1.

B. A two stage method

The first stage of the two-stage method [2], proceeds by finding a consistent estimate $\hat{w}_i(t)$ of all nodes $w_i(t)$ in $\mathcal{N}_j$. This is done by high-order modeling of $\{S_{il}\}$ and estimating it from (4) using the prediction error method. The prediction errors are constructed as

$$\varepsilon_i(t, \alpha) = \tilde{w}_i(t) - \sum_{l\in\mathcal{R}} S_{il}(q, \alpha)r_l(t)\,, \tag{5}$$

where α is a parameter vector. The resulting estimate $S_{il}(q,\hat{\alpha})$ is then used to obtain the node estimate as

$$\hat{w}_i(t) = \sum_{l\in\mathcal{R}} S_{il}(q,\hat{\alpha}) r_l(t)\,. \tag{6}$$

In a second stage, the module of interest $G_{j1}(q)$ (and the other modules in $\mathcal{N}_j$) is parameterized by θ and estimated from (3), again using the prediction error method. The prediction errors are now constructed as

$$\varepsilon_j(t,\theta) = \tilde{w}_j(t) - r_j(t) - \sum_{i\in\mathcal{N}_j} G_{ji}(q,\theta)\hat{w}_i(t)\,. \tag{7}$$

C. Simultaneous minimization of prediction errors

In this section, we briefly introduce the simultaneous minimization of prediction error method (SMPE) [8]. The main idea underlying SMPE is that if, the two prediction errors are simultaneously minimized, the variance will be decreased [14]. In the SMPE method, the prediction error of the measurement $\tilde{w}_j$ depends explicitly on α and is given by

$$\varepsilon_j(t,\theta,\alpha) = \tilde{w}_j(t) - \sum_{i\in\mathcal{N}_j} G_{ji}(q,\theta)\sum_{l\in\mathcal{R}} S_{il}(q,\alpha) r_l(t)\,. \tag{8}$$

The method proceeds to minimize

$$V_N(\theta,\alpha) = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{\varepsilon_j^2(t,\theta,\alpha)}{\sigma_j^2} + \sum_{i\in\mathcal{N}_j}\frac{\varepsilon_i^2(t,\alpha)}{\sigma_i^2}\right]. \tag{9}$$

As an initial estimate of the parameters θ and α, the minimizers of the two-stage method can be taken.

The main drawback is that the least-squares estimation of S may still induce high variance in the estimates. Additionally, if each of the n_s estimated transfer functions in S is estimated by the first n impulse response coefficients, the number of estimated parameters in S alone is $n_s \cdot n$. Already for relatively small dimensions of S the SMPE method is prohibitively expensive. To handle this, a frequency domain approach is taken in [23]. In this paper, we will instead use regularization to reduce the variance and the complexity.

D. Notation

Given a sequence of scalars $\{a(t)\}_{t=1}^{m}$, we denote by a its vector representation $a = [a(1)\ \cdots\ a(m)]^T \in \Re^m$. Given a vector $a \in \Re^m$, we define by $\mathcal{T}_n(a)$ the $m \times n$ Toeplitz matrix whose elements are the entries of a. Lower case letters indicate, in general, column vectors and, when there is no confusion, capital letters indicate their Toeplitz form, so given $a \in \Re^m$, we have that $A = T_n(a)$, where the number n of columns is consistent with the rest of the formula. The symbol "$\otimes$" denotes the standard Kronecker product of two matrices.

III. Empirical Bayes estimation of the module

In this section we derive our approach to the identification of a specific module based on EB. For ease of exposition, we give a detailed derivation in the one-reference-one-module case.

We consider a dynamic network with one reference signal $r_1(t)$. Without loss of generality, we assume that the module of interest is $G_{21}(q)$, and hence $G_{22}(q),\dots,G_{2L}(q)$ are assumed zero. We parametrize the target module by means of a parameter vector $\theta \in \mathbb{R}^{n_\theta}$. Using the vector notation introduced in the previous section, we denote by $\tilde{w}_1$ the stacked measurements $\tilde{w}_1(t)$ before the module of interest $G_{21}(q,\theta)$, and by $\tilde{w}_2$ the stacked output of this module $\tilde{w}_2(t)$. We define the impulse response coefficients of $G_{21}(q,\theta)$ by the inverse discrete-time Fourier transform

$$g_\theta(t) \coloneqq \int_{-\pi}^{\pi} G_{21}(e^{j\omega},\theta)e^{j\omega t}\,\mathrm{d}\omega\,. \tag{10}$$

Similarly we define s_{11} as the impulse response coefficients of $S_{11}(q)$, where $S_{11}(q)$ is as before the sensitivity path from $r_1(t)$ to $w_1(t)$, and $e_1(t)$ and $e_2(t)$ are the measurement noise sources (which we have assumed white and Gaussian). Their variance is denoted by σ_1^2 and σ_2^2, respectively. We rewrite the dynamics as

$$\begin{aligned}\tilde{w}_1 &= \quad S_{11}r_1 + e_1\,,\\ \tilde{w}_2 &= G_\theta S_{11} r_1 + e_2\,,\end{aligned} \tag{11}$$

where G_θ is the $N \times N$ Toeplitz matrix (with null initial conditions) of the N first impulse response g_θ. The same notation holds for the impulse response s_{11} and its Toeplitz-matrix version $S_{11} = \mathcal{T}_N(s_{11})$. We further rewrite (11) as

$$\begin{aligned}\tilde{w}_1 &= \quad R_1 s_{11} + e_1\,,\\ \tilde{w}_2 &= G_\theta R_1 s_{11} + e_2\,.\end{aligned} \tag{12}$$

where $R_1 = \mathcal{T}_N(r_1)$. For computational purposes, we only consider the first n samples of s_{11}, where n is large enough to capture the relevant dynamics of the sensitivity $S_{11}(q)$. Let $z \coloneqq [\tilde{w}_1^T\ \tilde{w}_2^T]^T$; we rewrite (12) as

$$z = W_\theta s_{11} + e\,, \qquad W_\theta = \begin{bmatrix} R_1 \\ G_\theta R_1\end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ e_2\end{bmatrix} \tag{13}$$

Note that e is a random vector such that

$$\Sigma_e \coloneqq \mathrm{E}\left[ee^T\right] = \begin{bmatrix}\sigma_1^2 I & 0\\ 0 & \sigma_2^2 I\end{bmatrix}. \tag{14}$$

A. Bayesian model of the sensitivity path

To reduce the variance in the sensitivity estimate (and also reduce the number of estimated parameters), we cast our problem in a Bayesian framework and model the sensitivity function as a zero-mean Gaussian stochastic vector, i.e.,

$$p(s_{11};\lambda,K_\beta) \sim N(0,\lambda K_\beta)\,. \tag{15}$$

The structure of the covariance matrix is given by the *first-order stable spline kernel* [18]:

$$\{K_\beta\}_{i,j} = \beta^{\max(i,j)}, \qquad \beta \in [0,\,1)\,. \tag{16}$$

The parameter β regulates the decay velocity of the realizations from (15), whereas, λ tunes their amplitude.

B. The marginal likelihood estimator

Since s_{11} is assumed stochastic, it admits a probabilistic description jointly with the vector of observation z, parametrized by the vector

$$\eta = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \lambda & \beta & \theta \end{bmatrix} . \tag{17}$$

In particular, having assumed a Gaussian distribution of the noise, the joint description is also Gaussian, that is,

$$p\left(\begin{bmatrix} z \\ s_{11} \end{bmatrix} ; \eta \right) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_z & \Sigma_{zs} \\ \Sigma_{sz} & \lambda K_\beta \end{bmatrix} \right) , \tag{18}$$

where $\Sigma_z = W_\theta \lambda K_b W_\theta^T + \Sigma_e$, and $\Sigma_{zs} = \Sigma_{sz}^T = W_\theta \lambda K_\beta$. It is instrumental to derive the posterior distribution of s_{11} given the measurement vector z. It is given by

$$p(s_{11}|z;\, \eta) \sim N(P W_\theta^T \Sigma_e^{-1} z, P) , \tag{19}$$

$$P = (W_\theta^T \Sigma^{-1} W_\theta + (\lambda K_\beta)^{-1})^{-1} , \tag{20}$$

and it is also parametrized by the vector η.

The module identification strategy we propose in this paper relies on an EB approach. We introduce the marginal probability density function (pdf) of the measurements

$$p(z;\, \eta) = \int p(z,\, s_{11})\, ds_{11} \sim \mathcal{N}(0,\, \Sigma_z) , \tag{21}$$

that is, the pdf of the measurements after having integrating out the dependence on the sensitivity path s_{11}. Then, we can define the (log) marginal likelihood (ML) criterion as the maximum of the marginal pdf defined above

$$\hat{\eta} = \arg\min_{\eta} \left(\log\det \Sigma_z + z^T \Sigma_z^{-1} z \right) , \tag{22}$$

whose solution provides also an estimate of θ and thus of the module of interest.

IV. Computation of the solution of the marginal likelihood criterion

Problem (22) is nonlinear and may involve a large number of decision variables, if n_θ is large. In this section, we derive an iterative solution scheme based on the Expectation/Conditional-Maximization (ECM) algorithm [20], which is a generalization of the standard Expectation-Maximization (EM) algorithm. In order to employ EM-type algorithms, one has to define a *latent variable*; in our problem, a natural choice is s_{11}. Then, a (local) solution to (22) is achieved by iterating over the following steps:

(E-step) Given an estimate $\hat{\eta}^{(k)}$ (computed at the k-th iteration of the algorithm), compute

$$Q^{(k)}(\eta) := \mathbb{E}\left[\log p(z,\, s_{11};\, \eta^{(k)}) \right] , \tag{23}$$

where the expectation is taken with respect to the posterior of s_{11} when the estimate $\eta^{(k)}$ is used, i.e., $p(s_{11}|z,\, \hat{\eta}^{(k)})$;

(M-step) Solve the problem

$$\hat{\eta}^{(k+1)} = \arg\max_{\eta} Q^{(k)}(\eta) . \tag{24}$$

First, we turn our attention on the computation of the E-step, i.e., the derivation of (23). Let $\hat{s}_{11}^{(k)}$ and $\hat{P}^{(k)}$ be the posterior mean and covariance matrix of s_{11}, computed from (19) using $\hat{\eta}^{(k)}$. Define $\hat{S}_{11}^{(k)} := \hat{P}^{(k)} + \hat{s}_{11}^{(k)} \hat{s}_{11}^{(k)T}$. The following proposition provides an expression for the function $Q^{(k)}(\eta)$.

Lemma 4.1: Let $\hat{\eta}^{(k)} = [\sigma_1^{2(k)}\, \sigma_2^{2(k)}\, \lambda^{(k)}\, \beta^{(k)}\, \theta^{(k)}]$ be an estimate of η after the k-th iteration of the EM method. Then

$$Q^{(k)}(\eta) = -\frac{1}{2} Q_0^{(k)}(\sigma_1^2,\, \sigma_2^2,\, \theta) - \frac{1}{2} Q_s^{(k)}(\lambda,\, \beta) , \tag{25}$$

where

$$Q_0^{(k)}(\sigma_1^2,\, \sigma_2^2,\, \theta) = \Bigg(\log\det \Sigma_e + z^T \Sigma_e^{-1} z \\ -2 z^T W_\theta \hat{s}_{11}^{(k)} + \mathrm{Tr}\Big\{ W_\theta^T \Sigma_e^{-1} W_\theta \hat{S}_{11}^{(k)} \Big\} \Bigg) , \tag{26}$$

$$Q_s^{(k)}(\lambda,\, \beta) = \log\det \lambda K_\beta + \mathrm{Tr}\Big\{ (\lambda K_\beta)^{-1}\, \hat{S}_{11}^{(k)} \Big\} . \tag{27}$$

Having computed the function $Q^{(k)}(\eta)$, we now focus on its maximization. We first note that the decomposition (25) shows that the kernel hyperparameters can be updated independently of the rest of the parameters:

Proposition 4.1: Define

$$Q_\beta(\beta) = \log\det K_\beta + n \log \mathrm{Tr}\, \{ K_\beta^{-1} \hat{S}_{11}^{(k)} \}. \tag{28}$$

Then

$$\hat{\beta}^{(k+1)} = \arg\min_{\beta \in [0,1)} Q_\beta(\beta) , \tag{29}$$

$$\hat{\lambda}^{(k+1)} = \frac{1}{n} \mathrm{Tr}\, \{ K_{\hat{\beta}^{(k+1)}}^{-1} \hat{S}_{11}^{(k)} \} . \tag{30}$$

Therefore, the update of the scaling hyperparameter is available in closed-form, while the update of β requires the solution of a scalar optimization problem in the domain $[0,\, 1]$, an operation that requires little computational effort, see [24] for computational details.

We are left with the maximization of the function $Q_0^{(k)}(\sigma_1^2,\, \sigma_2^2,\, \theta)$. In order to simplify this step, we split the optimization problem into constrained subproblems that involve fewer decision variables. This operation is justified by the ECM paradigm, which, under mild conditions [20], guarantees the same convergence properties of the EM algorithm even when the optimization of $Q^{(k)}(\eta)$ is split into a series of constrained subproblems. In our case, we decouple the update of the noise variances from the update of θ. By means of the ECM paradigm, we split the maximization of $Q_0^{(k)}(\sigma_1^2,\, \sigma_2^2,\, \theta)$ in a sequence of two constrained optimization subproblems:

$$\hat{\theta}^{(k+1)} = \arg\max_{\theta} Q_0^{(k)}(\sigma_1^2,\, \sigma_2^2,\, \theta) \\ \text{s.t. } \sigma_1^2 = \sigma_1^{2(k)},\, \sigma_2^2 = \sigma_2^{2(k)} , \tag{31}$$

$$\sigma_1^{2(k+1)},\, \sigma_2^{2(k+1)} = \arg\max_{\sigma_1^2,\, \sigma_2^2} Q_0^{(k)}(\sigma_1^2,\, \sigma_1^2,\, \theta) \\ \text{s.t. } \theta = \hat{\theta}^{(k+1)} . \tag{32}$$

The following result provides the solution of the above problems.

Proposition 4.2: Introduce the matrix $D \in \mathbb{R}^{N^2 \times N}$ such that $D\tilde{w}_1 = \mathrm{vec}(\mathcal{T}_N(a))$, for any $a \in \mathbb{R}^N$. Define

$$\hat{A}^{(k)} = D^T(R_1 \hat{S}_{11}^{(k)} R_1^T \otimes I_N)D\,,\ \hat{b}^{(k)} = \mathcal{T}_N(R_1 \hat{s}_{11}^{(k)})^T y\,. \quad (33)$$

Then

$$\hat{\theta}^{(k+1)} = \arg\min_{\theta} g_\theta^T \hat{A}^{(k)} g_\theta - 2\hat{b}^{(k)T} g_\theta\,. \quad (34)$$

The closed form updates of the noise variances are as follows

$$\hat{\sigma}_1^{2(k+1)} = \frac{1}{N}\Big(\|\tilde{w}_1 - R_1 \hat{s}_{11}^{(k)}\|_2^2 + \mathrm{Tr}\left\{R_1 \hat{P}^{(k)} R_1^T\right\}\Big)\,,$$
$$\hat{\sigma}_2^{2(k+1)} = \frac{1}{N}\Big(\|\tilde{w}_2 - G_{\hat{\theta}^{(k+1)}} R_1 \hat{s}_{11}^{(k)}\|_2^2 + \mathrm{Tr}\left\{G_{\hat{\theta}^{(k+1)}} R_1 \hat{P}^{(k)} R_1^T G_{\hat{\theta}^{(k+1)}}^T\right\}\Big)\,. \quad (35)$$

Each variance is the result of the sum of one term that measures the adherence of the identified systems to the data and one term that compensates for the bias in the estimates introduced by the Bayesian approach. The update of the parameter θ involves a (generally) nonlinear least-squares problem, which can be solved using gradient descent strategies. Note that, in case the impulse response g_θ is linearly parametrized (e.g., it is an FIR system or orthonormal basis functions are used [25]), then the update of θ is also available in closed-form.

Example 4.1: Assume that the linear parametrization $g_\theta = L\theta$, $L \in \mathbb{R}^{N \times n_\theta}$ is used, then

$$\hat{\theta}^{(k+1)} = \left(L^T \hat{A}^{(k)} L\right)^{-1} L^T \hat{b}^{(k)}\,. \quad (36)$$

A. Identification algorithm

The proposed method for module identification can be summarized in the following steps.

1) Find an initial estimate of $\hat{\eta}^{(0)}$, set $k = 0$.
2) Compute $\hat{s}_{11}^{(k)}$ and $\hat{P}^{(k)}$ from (19).
3) Update the kernel hyperparameters using (30), (29).
4) Update the vector θ solving (34).
5) Update the noise variances from (35).
6) Check if the algorithm has converged. If not, set $k = k + 1$ and go back to step 2.

The method can be initialized in several ways. One option is to first estimate $\hat{S}_{11}(q)$ by an empirical Bayes method using only r_1 and $\tilde{w}_1$. Then, $\hat{w}_1$ is constructed from (6), using the obtained $\hat{S}_{11}(q)$. Finally, G is estimated using the prediction error method, using $\hat{w}_1$ as input and $\tilde{w}_2$ as output.

V. Numerical experiments

In this section, we present results from a Monte Carlo simulation of $MC = 300$ iterations to illustrate the performance of the proposed method, which we abbreviate as *Network Empirical Bayes* (NEB). We consider a SISO system operating in closed loop, where the proposed method is compared with both the SMPE method and the prediction error method (*PEM*). The reference signals used are zero-mean unit-variance Gaussian white noise. The noise signals e_k are zero-mean Gaussian white noise with variances such that $\mathrm{Var}\, w_k / \mathrm{Var}\, e_k = 10$. The setting of the compared methods are provided in some more detail below, where the model order of the plant $G(q)$ is known for both the SMPE method and the proposed NEB method.

TABLE I: Mean of parameter estimates plus/minus one standard deviation, dynamic network setup.

	NEB	SMPE	θ^0
$\hat{b}_1$	$.051 \pm 0.005$	$.052 \pm 0.008$	.05
$\hat{a}_1$	$.56 \pm 0.09$	$.57 \pm 0.13$	.6
$\hat{b}_2$	$.089 \pm 0.004$	$.089 \pm 0.006$	.09
$\hat{a}_2$	$.51 \pm 0.03$	$.50 \pm 0.05$	.5

NEB: The method is initialized by first estimating $\hat{S}(q)$ by an empirical Bayes method. Using $\tilde{w}_j$ and $\hat{w}$ as input obtained from (6), G is estimated using the *pem* command in *MATLAB*. Then, the iterative method outlined in Section IV-A is employed with the stopping criteria $\|\eta^{k+1} - \eta^k\|/\|\eta^k\| < 10^{-10}$.

SMPE: The method is initialized by the two-stage method. First, $\hat{S}(q)$ is estimated by least-squares. Second, G is estimated using the *pem* command in *MATLAB* from $\hat{w}$ obtained from (6) and $\tilde{w}_j$. Then, a greybox-model is used to minimize the cost function (9) using the *pem* command with the tolerance set to 10^{-2} and unlimited number of iterations.

The second Monte Carlo simulation compares the NEB method with the SMPE method on data from the network of Example 2.1, illustrated in Figure 1, where

$$G_{14} = G_{21} = G_{43} = \frac{0.4q^{-1} - 0.5q^{-2}}{1 + 0.3q^{-1}}$$
$$G_{12} = G_{23} = \frac{0.4q^{-1} + 0.5q^{-2}}{1 + 0.3q^{-1}}$$
$$G_{31} = \frac{b_1 q^{-1}}{1 + a_1 q^{-1}}, \quad G_{32} = \frac{b_2 q^{-1}}{1 + a_2 q^{-1}}$$

Two reference signals, $r_2(t)$ and $r_4(t)$ are available and $N = 200$ data samples are used with the goal to estimate $G_{31}(q)$. In total 6 transfer functions are estimated, $\{S_{12}(q), S_{24}(q), S_{22}(q), S_{24}(q), G_{31}(q), G_{32}(q)\}$, where $\{S_{12}(q), S_{24}(q), S_{22}(q), S_{24}(q)\}$ are each parameterized by $n = 50$ impulse response coefficients in both methods. The modules $G_{31}(q)$ and $G_{32}(q)\}$ are parametrized with 2 parameters each where the true parameters are $\theta^0 = [b_1^0, a_1^0, b_2^0, a_2^0] = [0.05, 0.6, 0.09, 0.5]$. The results of the experiment seem to confirm that the proposed NEB method performs better than the SMPE method, with more consistent performance, cf. Figures 6-7 and Table II. The mean execution time τ is one order of magnitude slower for the SMPE method (2557 s) compared to NEB (171 s).

VI. Conclusion

In this paper, we have addressed the identification of a module in dynamic networks with known topology. The problem is cast as the identification of two systems in series. The second system corresponds to the target module, while the first represents the dynamic relation between exogenous signals and the input to the target module. This system is modeled following a Bayesian kernel-based approach, which enables the identification of the target module using

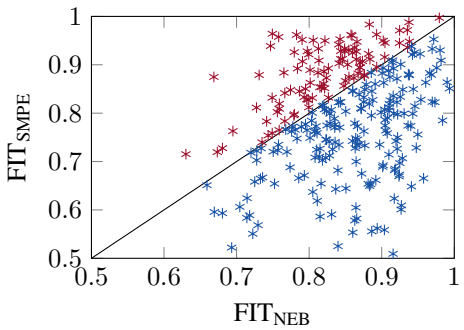

Fig. 3: Scatter plots of the fits of the impulse response of G_{31} obtained by the methods NEB and SMPE at each Monte Carlo, dynamic network setup.

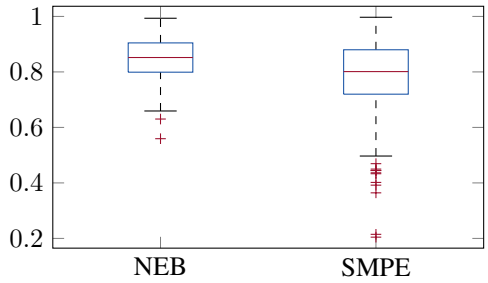

Fig. 4: Box plots of the fits of the impulse response of G_{31}, dynamic network setup.

empirical Bayes arguments. In particular, the target module is estimated using a marginal likelihood criterion, whose solution is obtained by a novel iterative scheme designed through the ECM algorithm. The performance of the method is evaluated through a numerical experiment. In particular, as compared to SMPE, the computation time is significantly reduced and the accuracy is improved.

REFERENCES

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

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

[3] H. Hjalmarsson, "System identification of complex and structured systems," *European J. of Control*, vol. 15, no. 3-4, pp. 275–310, 2009.

[4] 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.

[5] A. Chiuso and G. Pillonetto, "A Bayesian approach to sparse dynamic network identification," *Automatica*, vol. 48, no. 8, pp. 1553–1565, 2012.

[6] D. Hayden, Y. Yuan, and J. Gonçalves, "Network reconstruction from intrinsic noise: Minimum-phase systems," in *Proceedings of the 2014 American Control Conference*, pp. 4391–4396, 2014.

[7] A. Dankers, P. M. J. Van den Hof, and P. S. C. Heuberger, "Predictor input selection for direct identification in dynamic networks," in *Proceeding of the 52nd IEEE Annual Conference on Decision and Control*, pp. 4541–4546, IEEE, 2013.

[8] B. Gunes, A. Dankers, and P. M. J. Van den Hof, "A variance reduction technique for identification in dynamic networks," in *Proceedings of the 19th IFAC World Congress*, 2014.

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

[10] A. Haber and M. Verhaegen, "Subspace identification of large-scale interconnected systems," *IEEE Trans. on Automatic Control*, vol. 59, no. 10, pp. 2754–2759, 2014.

[11] P. Torres, J. W. van Wingerden, and M. Verhaegen, "Output-error identification of large scale 1D-spatially varying interconnected systems," *IEEE Trans. on Automatic Control*, vol. 60, no. 1, pp. 130–142, 2014.

[12] U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, pp. 1215–1241, 1999.

[13] M. Gilson and P. M. J. Van den Hof, "Instrumental variable methods for closed-loop system identification," *Automatica*, vol. 41, no. 2, pp. 241 – 249, 2005.

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

[15] L. Ljung, *System identification*. Springer, 1998.

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

[17] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[18] G. Pillonetto and G. De Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, no. 1, pp. 81–93, 2010.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[20] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

[21] G. Bottegal, R. S. Risuleo, and H. Hjalmarsson, "Blind system identification using kernel-based methods," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 466–471, 2015.

[22] R. S. Risuleo, G. Bottegal, and H. Hjalmarsson, "A kernel-based approach to hammerstein system identification," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 1011–1016, 2015.

[23] A. Dankers and P. M. J. Van den Hof, "Non-parametric identification in dynamic networks," in *Proceeding of the 54th IEEE Conference on Decision and Control)*, pp. 3487–3492, 2015.

[24] G. Bottegal, A. Y. Aravkin, H. Hjalmarsson, and G. Pillonetto, "Robust EM kernel-based methods for linear system identification," *Automatica*, vol. 67, pp. 114–126, 2016.

[25] B. Wahlberg, "System identification using Laguerre models," *IEEE Trans. on Automatic Control*, vol. 36, pp. 551–562, 1991.

[26] G. Prando, A. Chiuso, and G. Pillonetto, "Bayesian and regularization approaches to multivariable linear system identification: the role of rank penalties," in *Proceeding of the 53rd IEEE Annual Conference on Decision and Control*, pp. 1482–1487, 2014.