Local module identification in dynamic networks using regularized kernel-based methods

Local module identification in dynamic networks using regularized kernel-based methods

Karthik R. Ramaswamy, Giulio Bottegal and Paul M.J. Van den Hof

***Abstract*— In order to identify a specific system (module) of interest embedded in a dynamic network, one typically has to formulate a multi-input single-output (MISO) identification problem which requires to identify all modules in the MISO structure, and determine their model order. While the former task poses the problem of estimating a large number of parameters that are of no interest to the experimenter, the latter task may result computationally challenging in large-size networks. To avoid these issues and increase the accuracy of the identified module of interest, we use regularized kernel-based methods. Keeping a parametric model for the module of interest, we model the impulse response of the remaining modules in the MISO structure as zero mean Gaussian vectors with covariance matrix (kernel) given by the first-order stable spline kernel, accounting also for the noise model affecting the output of the target model. Using an Empirical Bayes (EB) approach, the target-module parameters are estimated by maximizing the marginal likelihood of the module output. The related optimization problem is solved using the Expectation-Maximization (EM) algorithm. Numerical experiments illustrate the potentials of the introduced method in comparison with the state-of-the-art techniques for local identification.**

I. INTRODUCTION

In recent years increasing attention has been devoted by the system identification community to the identification of dynamic networks. Within a large body of literature in this topic, we can distinguish three main research trends. The first one concerns detection of the network topology [16], [5], [12]. The second research trend focuses on the identification of the full network dynamics [11], [24], [23], [25], while the third studies the identification of a specific component (usually referred to as a module) of the network; under the assumption of known topology, the identification of a target module in a dynamic network is discussed in [21], [7], [13]. In these contributions, classic methods for closed-loop identification such as the *direct method* [14] and the *two-stage method* [22], are generalized to a dynamic network framework. The setting of the aforementioned contributions is generalized in [6], where also sensor noise affecting the measurements is considered. In [10], a method known as *simultaneous minimization of the prediction error* is introduced to identify a specific module in a dynamic network with only sensor noise. The method is extended to a Bayesian setting in [9], where regularized kernel-based methods are used to reduce the mean-square error of the identified target module.

Using a reasoning similar to [9], in this paper we aim at improving the performance of the direct method for dynamic networks introduced in [21]. In that work it was shown that, in order to identify a given module of interest, we have to formulate a multi-input single-output (MISO) identification problem where the modules constituting the MISO structure correspond to the modules of the network sharing the same output with the module of interest (see Sec. III-A for details). This implies that, to avoid possible bias in the parameter estimates, one has to identify all the modules constituting the MISO structure, bringing in the problem a possibly very high number of parameters to be estimated that are of no interest to the experimenter. In addition to it, it may be required to select the number of parameters of each of these additional modules using complexity criteria such as AIC, BIC, or cross validation [14]. If the number of modules is high, one may have to test a huge combination of candidate model orders, making model order selection a computationally infeasible step (e.g., for 5 modules with FIR model structure and orders from 1 to 5, one has to test 5^5 possible combinations).

To avoid model order selection issues and reduce the number of nuisance parameters in local module identification, we introduce a novel identification method based on regularized nonparametric kernel-based methods (see [19] for a survey on this subject). Considering known topology, we keep a parametric model for the module of interest in order to have an accurate description of its dynamics. As for the remaining modules of the MISO structure, we model their impulse responses as zero mean Gaussian processes. The covariance matrix (usually called a *kernel*) is given by the first-order stable spline kernel [4], [19], which encodes stability and smoothness of the processes. We also incorporate the process noise model indirectly through the impulse response modeling. The setup is different from [5], where all the modules are modeled as Gaussian processes.

Using this approach, we obtain a Gaussian probabilistic description that depends on a vector of parameters η containing the parameters of the module of interest, the variance of the noise, and the hyperparamaters characterizing the stable spline kernel. Therefore, by estimating η we can obtain the parameters of interest. To accomplish this task, we use an Empirical Bayes (EB) approach [15], where η is estimated by maximizing the marginal likelihood of the data. The solution of the problem is obtained by using an iterative scheme based on the Expectation-Maximization (EM) method [8], which turns out computationally attractive. Numerical experiments

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).

The authors are with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands {k.r.ramaswamy, g.bottegal, p.m.j.vandenhof}@tue.nl

on a simple dynamic network show the potentials of the developed method in comparison with the direct method and the two-stage method.

II. Problem statement

Following the setting of [21], we consider a dynamic network that is built up of L measurable internal variables or nodes $w_j(t)$, $j = 1, \dots, L$. The dynamic network can be defined by the equation (time and frequency dependence is omitted below),

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G_{12}^0 & \dots & G_{1L}^0 \\ G_{21}^0 & 0 & \dots & G_{2L}^0 \\ \vdots & \ddots & \ddots & \vdots \\ G_{L1}^0 & G_{L2}^0 & \dots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_L \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_L \end{bmatrix} = G^0(q)w(t) + r(t) + v(t). \tag{1}$$

In this equation,

- q^{-1} is the delay operator i.e. $q^{-1}u(t) = u(t-1)$;
- $G_{ji}^0(q)$ is a strictly proper rational transfer function for $j = 1, \dots, L$ and $i = 1, \dots, L$;
- $v_j(t)$ is unmeasured process noise entering node $w_j(t)$. It is a realization of a stationary stochastic process represented by $v_j(t) = H_j^0(q)e_j(t)$, with $e_j(t)$ a Gaussian white noise process with unknown variance σ_j^2 and $H_j^0(q)$ a monic, stable and minimum phase filter;
- $r_j(t)$ is a measured external reference signal entering node $w_j(t)$. In some nodes, it may be absent.

The following assumptions are made on the considered dynamic network.

Assumption 1: The dynamic network is stable, i.e. $(I - G^0(q))^{-1}$ is stable, and well posed (see [21] for details).

Assumption 2: The process noise $v_j(t)$ entering the node $w_j(t)$ is uncorrelated with the process noise entering any other node of the network.

We assume that we have collected N measurements of the internal variables $\{w_k(t)\}_{t=1}^N$, $k = 1, \dots, L$, and that we are interested in building a model of the module directly linking node i to node j, that is $G_{ji}^0(q)$, using these measurements. To this end, we choose a parameterization of $G_{ji}^0(q)$, denoted as $G_{ji}(q, \theta)$, that describes the dynamics of the module of interest for a certain parameter vector $\theta_0 \in \mathbb{R}^{n_\theta}$.

III. The developed identification technique

In this section we describe the developed method for identification of module $G_{ji}^0(q)$.

A. The standard direct method

Following the above definition of a dynamic network, the node signal $w_j(t)$ follows the equation

$$w_j(t) = \sum_{k \in \mathcal{N}_j} G_{jk}^0(q)w_k(t) + r_j(t) + v_j(t), \tag{2}$$

where $\mathcal{N}_j$ is the set of indices of internal variables $w_k(t)$ ($k \neq j$ and $G_{jk}^0(q) \neq 0$) that have direct causal connections with $w_j(t)$. The above equation represents a MISO structure and is the starting point of the methodology presented in this paper, which is based on extending the direct method [21]. We construct the one-step ahead predictor [14] of $w_j(t)$:

$$\hat{w}_j(t|t-1;\theta) = \big(1 - (H_j^0)^{-1}(q)\big)w_j(t) + (H_j^0)^{-1}(q)G_{ji}(q,\theta)w_i(t) + (H_j^0)^{-1}(q)\big(\sum_{k \in \mathcal{N}_j \backslash \{i\}} G_{jk}^0(q)w_k(t) + r_j(t)\big),$$

which is a function of the parameter vector θ entering the target module $G_{ji}^0(q)$. In the standard direct method for dynamic networks [21], all the modules $G_{jk}^0(q)$, $k \in \mathcal{N}_j$, and the noise model $H_j^0(q)$, are suitably parameterized with additional parameters. The parameter vector of interest θ is identified by minimizing the prediction error $\varepsilon_j(t) = w_j(t) - \hat{w}_j(t|t-1;\theta)$. We note that in this formulation, the prediction error depends also on the additional parameters entering the remaining modules and the noise model, which need to be identified to guarantee consistent estimates of θ. Therefore, the total number of parameters may grow large if the cardinality of $\mathcal{N}_j$ is large, with a detrimental effect on the variance of the estimate of θ in the case where N is not very large.

B. The direct method meets the Empirical Bayes approach

We now discuss how to use regularized kernel-based methods to avoid parameterization of the additional modules in the MISO structure. We define the following quantities:

$$S_j^0(q) := 1 - (H_j^0)^{-1}(q) \quad , \quad S_{jk}^0(q) := (H_j^0)^{-1}G_{jk}(q)\,,$$

where $k \in \{\mathcal{N}_j\}\backslash i$. Considering the parameterization of $G_{ji}^0(q)$, Eq. (2) can be re-written as

$$w_j(t) = S_j^0(q)w_j(t) + (1 - S_j^0(q))(G_{ji}(q,\theta)w_i(t) + r_j(t)) + \sum_{k \in \mathcal{N}_j \backslash \{i\}} S_{jk}^0(q)w_k(t) + e_j(t). \tag{3}$$

Next, we consider a time-domain description of the module dynamics. Our goal is to obtain a vector description of the network dynamics for the available N measurements. Since $H_j^0(q)$ is monic, proper and has a stable inverse, then $(H_j^0)^{-1}(q)$ is also monic, proper and stable. The impulse response of $(H_j^0)^{-1}(q)$ is represented as

$$(H_j^0)^{-1}(q) = \tilde{h}_j(0) + \sum_{d=1}^{\infty} \tilde{h}_j(d)q^{-d}, \tag{4}$$

where $\tilde{h}_j(0) = 1$. Similarly, the impulse response of $S_j^0(q)$ is represented as $S_j^0(q) = \sum_{d=1}^{\infty} s_j(d)q^{-d}$, and the impulse response of the target module is written as $G_{ji}(q,\theta) = \sum_{d=1}^{\infty} g_{ji}(d,\theta)q^{-d}$. For notation purposes, we consider the first N samples of $g_{ji}(d,\theta)$ and collect them in the N-dimensional vector g_{ji} (which will also depend on θ, although we will keep this dependence tacit). Similarly, we define the vector s_k, $k \in \{\mathcal{N}_j\}\backslash i$, and s_j as the vectors containing the first l coefficients of the impulse responses of $S_{jk}^0(q)$, $k \in \{\mathcal{N}_j\}\backslash i$, and $S_j^0(q)$, respectively. The integer l is chosen large enough to ensure $s_k(l+1), s_j(l+1) \simeq 0$.

Next, we introduce a vector notation for the node $w_j(t)$:

$$w_j := \begin{bmatrix} w_j(1) & \dots & w_j(N) \end{bmatrix}^T, \tag{5}$$

and we denote by $W_k \in \mathbb{R}^{N\times l}$ the Toeplitz matrix of the vector $\mathbf{w}_k := \begin{bmatrix}0 & w_k(1) & \dots & w_k(N-1)\end{bmatrix}^T$, $k \in \{\mathcal{N}_j \cup j\}\backslash\{i\}$ and $W_i \in \mathbb{R}^{N\times N}$ the Toeplitz matrix of the vector $w_i := \begin{bmatrix}0 & w_i(1) & \dots & w_i(N-1)\end{bmatrix}^T$. Similarly, we denote by $\bar{W}_i \in \mathbb{R}^{N\times l}$ the Toeplitz matrix of the vector $\bar{w}_i := \begin{bmatrix}0 & 0 & -w_i(1) & \dots & -w_i(N-2)\end{bmatrix}^T$, and by G_θ the Toeplitz of g_{ji}. Without loss of generality, we can assume $r_j(t) = 0$ for the sake of brevity. With the above notations in place, we can rewrite the network dynamics in the following vector form:

$$w_j = \tilde{W}s_j + W_i g_{ji} + \sum_{k\in\mathcal{N}_j\backslash\{i\}} W_k s_k + e_j, \tag{6}$$

where $\tilde{W} := W_j + G_\theta \bar{W}_i$ and e_j is the vectorized noise.

Having set a convenient notation for the network dynamics, we now discuss in details our modeling strategy. Our goal is to limit the number of parameters necessary to describe w_j in (6), in order to increase the accuracy of the estimated parameter vector of interest θ. Therefore, while we keep a parametric model for g_{ij}, for the remaining impulse responses in (6) we use nonparametric models induced by Gaussian processes [20]. The choice of Gaussian processes is motivated by the fact that, with a suitable choice of the prior covariance matrix, we can get a significant reduction in the variance of the estimated impulse responses [19]. Therefore, we model s_j and s_k, $k \in \mathcal{N}_j\backslash\{i\}$, as independent[1] Gaussian processes (vectors in this case) with zero-mean. The covariance matrix of these vectors, usually referred to as a kernel in this context, is chosen to be corresponding to the so-called *First-order Stable Spline kernel*. The general structure of this kernel is given by

$$[K_\beta]_{x,y} = \beta^{\max(x,y)}, \tag{7}$$

where $\beta_j \in [0, 1)$ is a *hyperparameter* that regulates the decay velocity of the realizations of the corresponding Gaussian vector, while $\lambda \geq 0$ tunes their amplitude. The choice of this kernel is motivated by the fact that it enforces enjoyable properties such as stability and smoothness in the estimated impulse responses [17], [18]. Therefore, we have that

$$s_j \sim \mathcal{N}(0, \lambda_j K_{\beta_j}), \tag{8}$$

$$s_k \sim \mathcal{N}(0, \lambda_k K_{\beta_k}) \quad , k \in \mathcal{N}_j\backslash\{i\}, \tag{9}$$

where we have assigned different hyperparameters to the impulse response priors to guarantee flexible enough models.

We define

$$s := \begin{bmatrix}s_j^\top & s_{k_1}^\top & s_{k_2}^\top & \dots & s_{k_p}^\top\end{bmatrix}^\top, \tag{10}$$

where $k_1, \dots, k_p$ are the elements of the set $\mathcal{N}_j\backslash\{i\}$, and

$$\mathbf{W} := \begin{bmatrix}\tilde{W} & W_{k1} & W_{k2} & \dots & W_{kp}\end{bmatrix}, \tag{11}$$

[1]It is clear that these impulse responses share some common dynamics given by the pre-multiplication with the inverse of the noise model $H_j(q)$. However, for computational purposes it is convenient to treat the impulse responses as independent. Furthermore, incorporating the mutual dependence through a suitable choice of prior distribution seems a non-trivial problem that deserves a thorough analysis that is outside the scope of this paper.

$$\mathbf{K} := \operatorname{diag}\{\lambda_j K_{\beta_j}, \lambda_{k1} K_{\beta_{k1}}, \dots, \lambda_{kp} K_{\beta_{kp}}\}. \tag{12}$$

Using the above, we can rewrite (6) in compact form as

$$w_j = \mathbf{W}s + W_i g_{ji} + e_j. \tag{13}$$

Having assumed a Gaussian distribution of the noise, we can write the joint probabilistic description of s and w_j, which is jointly Gaussian, as:

$$p\left(\begin{bmatrix}s\\ w_j\end{bmatrix};\eta\right) \sim \mathcal{N}\left(\begin{bmatrix}0\\ W_i g_{ji}\end{bmatrix}, \begin{bmatrix}\mathbf{K} & \mathbf{K}\mathbf{W}^\top\\ \mathbf{W}\mathbf{K} & \mathbf{P}\end{bmatrix}\right), \tag{14}$$

where

$$\mathbf{P} := \sigma_j^2 I_N + \tilde{W}\lambda_j K_{\beta_j}\tilde{W} + \sum_{k\in\mathcal{N}_j\backslash\{i\}} W_k \lambda_k K_{\beta_k} W_k{}^\top. \tag{15}$$

We note that this pdf depends upon the vector of parameters

$$\eta := \begin{bmatrix}\theta & \lambda_j & \lambda_{k1} & \dots & \lambda_{kp} & \beta_j & \beta_{k1} & \dots & \beta_{kp} & \sigma_j^2\end{bmatrix},$$

which contains the parameter vector of the target module, the hyperparameters of the kernels of the impulse response models of the other modules, and the variance of the noise corrupting $w_j(t)$. Therefore, we focus on the estimation of η, since it contains the parameter of interest θ. To this end, we apply an Empirical Bayes (EB) approach. We consider the marginal pdf of w_j, which is obtained by integrating out the dependence on s and which corresponds to

$$p(w_j;\eta) \sim \mathcal{N}(W_i g_{ji}, \mathbf{P}). \tag{16}$$

Then, the estimate of η is obtained by maximizing the marginal likelihood of w_j, namely

$$\begin{aligned}\hat{\eta} &= \arg\max_\eta p(w_j;\eta)\\ &= \arg\min_\eta \log\det\mathbf{P} + \big(w_j - W_i g_{ji}\big)^\top \mathbf{P}^{-1}\big(w_j - W_i g_{ji}\big).\end{aligned} \tag{17}$$

Solving this optimization problem can be a cumbersome task, because it is nonlinear. In the next section, we study how to solve the marginal likelihood problem through a dedicated iterative scheme.

IV. Solving the marginal likelihood estimation

In this section we focus on solving the problem in (17). We use an iterative solution scheme through the EM algorithm. To do so, we first have to define a *latent variable* whose estimation simplifies the computation of the marginal likelihood. In our case, a natural choice is s. Then, the EM algorithm iterates among the following two steps:

- *E-Step:* Given an estimate $\hat{\eta}^{(n)}$ computed at the n^{th} iteration, compute

$$Q^{(n)}(\eta) = \mathbb{E}[\log p(w_j, s;\eta)], \tag{18}$$

  which is the expected value of the joint log-likelihood of w_j and s with respect to the posterior $p(s|w_j;\hat{\eta}^{(n)})$;
- *M-Step:* Update $\hat{\eta}$ by solving

$$\hat{\eta}^{(n+1)} = \arg\max_\eta Q^{(n)}(\eta). \tag{19}$$

Iterating among the above steps, convergence to a stationary point of the marginal likelihood is ensured [3]. Clearly, we get an advantage from using the EM algorithm if repetitively solving (19) is significantly easier than solving the original marginal likelihood problem (17). In the following subsections, we show that this is indeed the case.

A. Computation of E-step

First we focus on the E-step. The posterior distribution of s given w_j and an estimate of η is Gaussian and corresponds to (see also [1]),

$$p(s|w_j;\eta) \sim \mathcal{N}\big(C(w_j - W_i g_{ji}), \mathbf{P}_s\big), \tag{20}$$

where

$$\mathbf{P}_s = \left(\frac{\mathbf{W}^\top \mathbf{W}}{\sigma_j^2} + \mathbf{K}^{-1}\right)^{-1} \quad C = \frac{\mathbf{P}_s \mathbf{W}^\top}{\sigma_j^2}.$$

Let $\hat{s}^{(n)}$ and $\hat{\mathbf{P}}_s^{(n)}$ be the posterior mean and covariance of s obtained from (20) using $\hat{\eta}^{(n)}$. We define $\hat{\mathbf{S}}^{(n)} := \hat{\mathbf{P}}_s^{(n)} + \hat{s}^{(n)}\hat{s}^{(n)\top}$, and consider its $l \times l$ diagonal blocks, which we denote by $\hat{\mathbf{S}}_j^{(n)}$, $\hat{\mathbf{S}}_{k_1}^{(n)}$, $\ldots$, $\hat{\mathbf{S}}_{k_p}^{(n)}$, respectively. These submatrices correspond to the posterior second moments of the estimated impulse responses $\hat{s}_j^{(n)}$,$\hat{s}_{k_1}^{(n)}$, $\ldots$, $\hat{s}_{k_p}^{(n)}$.

The following lemma provides the structure of the function $Q^{(n)}(\eta)$ for the particular situation of our setup in (17).

Lemma 1: Let $\hat{\eta}^{(n)}$ be the estimate of η at n^{th} iteration of the EM algorithm. Then

$$Q^{(n)}(\eta) = Q_0^{(n)}(\sigma_j^2, \theta) + \sum_{k \in \{\mathcal{N}_j \cup j\} \backslash \{i\}} {Q_{s_k}}^{(n)}(\lambda_k, \beta_k) \tag{21}$$

where

$$\begin{aligned} Q_o^{(n)}(\sigma_j^2, \theta) = &- N \log(\sigma_j^2) - \frac{1}{\sigma_j^2}\Big[w_j^\top w_j - 2 w_j^\top W_i g_{ji} + \\ &g_{ji}^\top W_i^\top W_i g_{ji} - 2 w_j^\top \mathbf{W} \hat{s}^{(n)} + 2 g_{ji}^\top W_i^\top \mathbf{W} \hat{s}^{(n)} \\ &+ \operatorname{tr}\big(\mathbf{W}^\top \mathbf{W} \hat{\mathbf{S}}^{(n)}\big)\Big], \end{aligned} \tag{22}$$

$${Q_{s_k}}^{(n)}(\lambda_k, \beta_k) = - \log\det(\lambda_k K_{\beta_k}) - \operatorname{tr}\big((\lambda_k K_{\beta_k})^{-1} \hat{\mathbf{S}}_k^{(n)}\big). \tag{23}$$

■

It is seen that the function $Q^{(n)}(\eta)$ is the summation of several terms that depend on different components of the vector η. In particular, we have a term of the type ${Q_{s_k}}^{(n)}(\lambda_k, \beta_k)$ for each module in the MISO structure, and a term $Q_0^{(n)}(\sigma_j^2, \theta)$ for the module of interest and the noise variance. Therefore, the update of η splits into a number of independent and smaller optimization problems.

B. Computation of M-step

We now focus on the M-step. From (21), it is evident that each kernel hyperparameters can be updated independently of the rest of the parameters. The following theorem, inspired from [2] and [9], shows how to update the kernel hyperparameters.

Theorem 1: Define

$$Q_{\beta_k}^{(n)}(\beta_k) = \log\det(K_{\beta_k}) + l \log\left(\operatorname{tr}\big((K_{\beta_k})^{-1} \hat{\mathbf{S}}_k^{(n)}\big)\right) \tag{24}$$

for $k \in \{\mathcal{N}_j \cup j\} \backslash i$. Then,

$$\hat{\beta_k}^{(n+1)} = \underset{\beta_k \in [0,1)}{\arg\min}\, Q_{\beta_k}^{(n)}(\beta_k); \tag{25}$$

$$\hat{\lambda_k}^{(n+1)} = \frac{1}{l} \operatorname{tr}\big((K_{\hat{\beta}_k^{(n+1)}})^{-1} \hat{\mathbf{S}}_k^{(n)}\big). \tag{26}$$

■

We note that from (26) we get closed form solutions for all λ_k, $k \in \{\mathcal{N}_j \cup j\} \backslash i$, while the β_k, $k \in \{\mathcal{N}_j \cup j\} \backslash i$, can be updated by solving scalar optimization problems in domain $[0, 1)$, as detailed in (25). Therefore, the hyperparameters update turns out to be a computationally fast operation.

We now turn our attention to θ and σ_j^2. We notice that the optimum with respect to θ does not depend on the optimal value of σ_j^2. Then, we can first update θ and then use its optimal value to update σ_j^2. How to update of θ is explained in the following theorem.

Theorem 2: The parameter vector θ are updated by solving the nonlinear least-squares problem

$$\hat{\theta}^{(n+1)} = \underset{\theta}{\arg\min} \left[g_{ji}^\top \hat{A}^{(n)} g_{ji} - 2 \hat{b}^{(n)\top} g_{ji}\right], \tag{27}$$

where $\hat{A}^{(n)}$ and $\hat{b}^{(n)}$ are computed using the current estimates $\hat{s}^{(n)}$ and $\hat{\eta}^{(n)}$. ■

The expression of $\hat{A}^{(n)}$ and $\hat{b}^{(n)}$ is given in the appendix. Therefore, the parameter vector of the target module is updated by solving a (generally) nonlinear least-squares problem given by (27). If g_{ji} is linearly parameterized with θ (e.g. in case of FIR models), the above problem becomes quadratic and a closed-form solution is achieved. That is, if $g_{ji} = M\theta$ where $M \in \mathbb{R}^{N \times n_\theta}$, then

$$\hat{\theta}^{(n+1)} = \big(M^\top \hat{A}^{(n)} M\big)^{-1} M^\top \hat{b}^{(n)}. \tag{28}$$

We are left with updating σ_j^2. This is done in the following statement.

Theorem 3: Let $\hat{g}_{ji}^{(n+1)}$, $\hat{\mathbf{W}}^{(n+1)}$ be constructed by inserting $\hat{\theta}^{(n+1)}$ in the general expression of g_{ji} and $\mathbf{W}$. Then

$$\begin{aligned} (\hat{\sigma}_j^2)^{(n+1)} = \frac{1}{N}\Big[&\|w_j - W_i \hat{g}_{ji}^{(n+1)}\|_2^2 - 2 w_j^\top \hat{\mathbf{W}}^{(n+1)} \hat{s}^{(n)} + \\ &2 \hat{g}_{ji}^{(n+1)\top} W_i^\top \hat{\mathbf{W}}^{(n+1)} \hat{s}^{(n)} + \operatorname{tr}\big(\hat{\mathbf{W}}^{(n+1)\top} \hat{\mathbf{W}}^{(n+1)} \hat{\mathbf{S}}^{(n)}\big)\Big]. \end{aligned}$$

■

Thus, a closed form solution for the noise variance is also obtained.

All-in-all, we have obtained a fast iterative procedure that provides a local solution to the marginal likelihood problem (17). Except for θ that requires solving a nonlinear optimization problem at each iteration, all the updates follow simple rules that allow for fast iterative computation. Algorithm 1 summarizes the steps to follow to obtain $\hat{\eta}$ and therefore $\hat{\theta}$. The initialization can be done by randomly choosing η

Algorithm 1 Algorithm for local identification in dynamic networks

Input: $\{w_k(t)\}_{t=1}^{N}$, $k = 1, \ldots, p$
Output: $\hat{\theta}$

1) Set $n = 0$, Initialize $\hat{\eta}^{(0)}$.
2) Compute $\hat{\mathbf{P}}_s^{(n)}$, $\hat{C}^{(n)}$, $\hat{\mathbf{S}}^{(n)}$ and $\hat{s}^{(n)}$.
3) Update hyperparameters $\hat{\beta}_k^{(n+1)}$ and $\hat{\lambda}_k^{(n+1)}$ using (25) and (26) respectively for all $k \in \{\mathcal{N}_j \cup j\}\backslash\{i\}$.
4) Update $\hat{\theta}^{(n+1)}$ by solving (27).
5) Update $\hat{\sigma}_j^{2(n+1)}$ as in Theorem 3.
6) Set $\hat{\eta}^{(n+1)}$
$= \big[\, \hat{\theta}^{(n+1)} \;\; \hat{\lambda}_j^{(n+1)} \;\; \hat{\lambda}_{k_1}^{(n+1)} \;\; \ldots \;\; \hat{\lambda}_{k_p}^{(n+1)}$
$\hat{\beta}_j^{(n+1)} \;\; \hat{\beta}_{k_1}^{(n+1)} \;\; \ldots \;\; \hat{\beta}_{k_p}^{(n+1)} \;\; \hat{\sigma}_j^{2(n+1)} \,\big]^\top$
7) Set $n = n + 1$.
8) Repeat steps (2) to (7) until convergence.

considering the constraints of hyperparameters. The convergence criterion for the algorithm depend on the value of $\frac{\|\hat{\eta}^{(n)} - \hat{\eta}^{(n-1)}\|}{\|\hat{\eta}^{(n-1)}\|}$. This value should be small for convergence so that the algorithm can be terminated. A value of 10^{-9} is considered for the numerical experiment in Sec. V.

V. Numerical Experiments

Numerical experiments are performed to evaluate the performance of the developed method, which we abbreviate as Empirical Bayes Direct Method (EBDM). To this end, the EBDM is compared with the standard direct method and the two-stage method (see [21] for details). The comparison is made on the dynamic network depicted in Fig. 1. The goal is to identify $G_{31}^0(q)$. The network modules are given by

$$G_{31}^0 = \frac{q^{-1} + 0.05q^{-2}}{1 + q^{-1} + 0.6q^{-2}} = \frac{b_1^0 q^{-1} + b_2^0 q^{-2}}{1 + a_1^0 q^{-1} + a_2^0 q^{-2}};$$

$$G_{32}^0 = \frac{-0.367q^{-1} - 0.063q^{-2} + 0.02q^{-3} + 0.005q^{-4}}{1 - 0.895q^{-1} - 0.104q^{-2} + 0.052q^{-3} + 0.011q^{-4}};$$

$$G_{34}^0 = \frac{1.184q^{-1} - 0.647q^{-2} + 0.151q^{-3} - 0.082q^{-4}}{1 - 0.8q^{-1} + 0.279q^{-2} - 0.048q^{-3} + 0.01q^{-4}};$$

$$G_{14}^0 = G_{21}^0 = \frac{0.4q^{-1} - 0.5q^{-2}}{1 + 0.3q^{-1}}; H_1^0 = \frac{1}{1 + 0.2q^{-1}};$$

$$G_{12}^0 = G_{23}^0 = \frac{0.4q^{-1} + 0.5q^{-2}}{1 + 0.3q^{-1}}; H_2^0 = \frac{1}{1 + 0.3q^{-1}};$$

$$H_3^0 = \frac{1 - 0.505q^{-1} + 0.155q^{-2} - 0.01q^{-3}}{1 - 0.729q^{-1} + 0.236q^{-2} - 0.019q^{-3}}; H_4^0 = 1.$$

We run 50 independent Monte Carlo experiments where the data is generated using known reference signals $r_2(t)$ and $r_4(t)$ that are realizations of white noise with unit variance. The number of data samples is N = 500. The noise sources $e_1(t)$, $e_2(t)$, $e_3(t)$ and $e_4(t)$ have variance 0.05, 0.08, 1, 0.1, respectively. We assume that we know the model order of $G_{31}^0(q)$. In the case of direct method, we solve a 3-input/1-output MISO identification problem with $w_1(t)$, $w_2(t)$ and $w_4(t)$ as inputs. In the two-stage method, the projection of the three inputs on external signals $r_1(t)$ and $r_2(t)$ are used as inputs to the MISO identification problem. For both

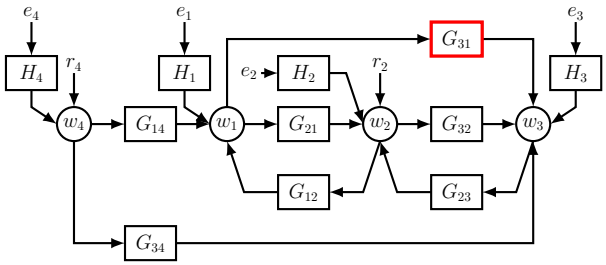

Fig. 1. Network example with 4 internal nodes, 2 reference signals and a noise sources at each node.

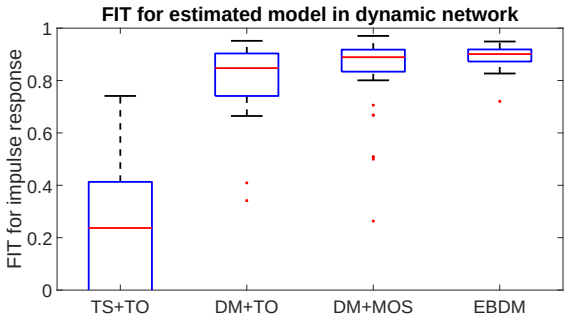

Fig. 2. Box plot of the fit of the impulse response of $\hat{G}_{31}$ obtained by the Two-stage method, Direct method and proposed method. Number of data samples used for estimation is N = 500.

these methods, we consider the case where the model order selection of all the modules in the MISO structure (except for the target module) is required, and the case where the model orders are known. Moreover, in order to improve the accuracy of the identified module in the two-stage method, we identify the noise model even though it is not necessary.

The box plots of the fits of the impulse response of $G_{31}(q)$ are shown in Fig. 2, where we have compared the two-stage method with true model orders ('TS+TO'), the direct method with true model orders and model orders selected via BIC ('DM+TO' and 'DM+MOS', respectively), and the Empirical Bayes Direct Method ('EBDM'). As for the latter, we choose l = 100. It can be noted that in this setup the EBDM achieves better fit than the classic methods. Fig. 3 shows the mean and standard deviation of the parameter estimates

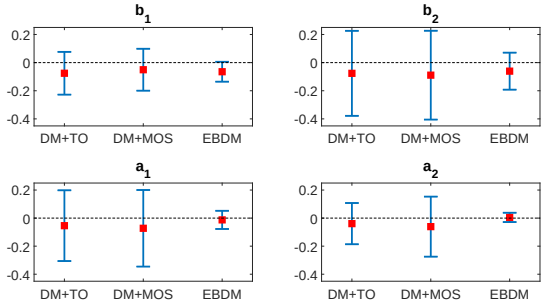

Fig. 3. Bias and standard deviation of each parameter obtained from 50 MC simulations using different identification methods.

of G_{31}. It is evident that the EBDM gives a smaller bias and a greatly reduced variance compared to the other considered identification methods. The reduction in variance is attributed to the regularization approach used in this method. Considering a relatively small sized network with 3 modules in the MISO structure, the developed method proves effective. When the size of the network grows, the results of the direct method may deteriorate further due to increase in variance; furthermore, it is expected that in large networks the model order selection step contributes to inaccurate results. Thus the EBDM, by offering reduced variance and circumventing the problem of model order selection, can stand out as an effective local module identification method in large dynamic networks.

VI. Conclusions

An effective regularized kernel-based approach for local module identification in large dynamic networks has been introduced in this paper. The introduced method (EBDM) circumvents the model order selection step for all the modules that are not of primary interest to the experimenter, but still need to be identified in order to get a consistent estimate of the target module. Furthermore, by using regularized nonparametric methods, the number of parameters to be estimated is greatly reduced, with a clear benefit in terms of mean square error of the estimated target module. Numerical experiments performed with a dynamic network example illustrate the potentials of the developed method on comparison with the already available methods. The developed method provides better estimates and a reduced variance is observed in the identified model due to the integration of the regularization approach in the method.

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal filtering*. Englewood Cliffs, N.J., USA: Prentice-Hall, 1979.

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

[3] R. A. Boyles. On the convergence of the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(1):47–50, 1983.

[4] T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and gaussian processes - revisited. *Automatica*, 48(8):1525–1535, 2012.

[5] A. Chiuso and G. Pillonetto. A bayesian approach to sparse dynamic network identication. *Automatica*, 48(8):1553–1565, 2012.

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

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

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[9] N. Everitt, G. Bottegal, and H. Hjalmarsson. An empirical bayes approach to identification of modules in dynamic networks. *Automatica*, 91:144–151, 2018.

[10] B. Gunes, A. Dankers, and P. M. J. Van den Hof. A variance reduction technique for identication in dynamic networks. In *IFAC Proceedings Volumes*, volume 47, pages 2842–2847, 2014.

[11] A. Haber and M. Verhaegen. Subspace identication of large-scale interconnected systems. *IEEE Transactions on Automatic Control*, 59(10):2754–2759, 2014.

[12] D. Hayden, Y. H. Chang, J. Goncalves, and C. J. Tomlin. Sparse network identifiability via compressed sensing. *Automatica*, 68:9–17, 2016.

[13] J. Linder and M. Enqvist. Identification of systems with unknown inputs using indirect input measurements. *International Journal of Control*, 90(4):729–745, 2017.

[14] L. Ljung. *System Identification: Theory for the User*. Prentice Hall PTR, 1999.

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

[16] D. Materassi and G. Innocenti. Topological identication in networks of dynamical systems. *IEEE Transactions on Automatic Control*, 55(8):1860–1871, 2010.

[17] G. Pillonetto, A. Chiuso, and G. De Nicolao. Prediction error identification of linear systems: a nonparametric gaussian regression approach. *Automatica*, 47(2):291–305, 2011.

[18] G. Pillonetto and G. De Nicolao. A new kernel-based approach for linear system identification. *Automatica*, 46(1):81–93, 2010.

[19] 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*, 50(3):657–682, 2014.

[20] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. The MIT Press, 2006.

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

[22] P. M. J. Van den Hof and R. J. P. Schrama. An indirect method for transfer function estimation from closed loop data. *Automatica*, 29(6):1523–1527, 1993.

[23] H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Identification of dynamic networks with rank-reduced process noise. *IFAC-PapersOnLine*, 50(1):10562–10567, 2015.

[24] H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Identification of dynamic networks operating in the presence of algebraic loops. In *55nd IEEE Conf. on Decision and Control (CDC)*, pages 4606–4611, 2016.

[25] H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Prediction error identication of linear dynamic networks with rank – reduced noise. *arXiv preprint arXiv:1711.06369*, 2017. to appear in Automatica.

Appendix

Let $D_1 \in \mathbb{R}^{N^2 \times N}$ and $D_2 \in \mathbb{R}^{N^2 \times N}$ are two matrices such that, for any vector $\mathbf{w} \in \mathbb{R}^N$, $D_1\mathbf{w} = \mathrm{vec}(W)$, where W is the Toeplitz matrix of $\mathbf{w}$, and $D_2\mathbf{w} = \mathrm{vec}(W^\top)$. Let us define $\breve{s}^{(n)} \in \mathbb{R}^N$ be a vector such that, if $N \leq l$, $\breve{s}^{(n)}$ is the vector of first N elements of $\hat{s}^{(n)}$ and if $N > l$, $\breve{s}^{(n)}$ is a vector with the first l elements equal to $\hat{s}^{(n)}$ and the remaining ones equal to 0. Let $\breve{S}^{(n)}$, $\breve{W}_i \in \mathbb{R}^{N \times N}$ be the Toeplitz matrix of $\hat{s}^{(n)}$ and $\bar{w}_i$ respectively. Then

$$\mathcal{X} = \begin{bmatrix} W_j & W_{k1} & W_{k2} & \dots & W_{kp} \end{bmatrix}, \quad \hat{\mathcal{Y}}^{(n)} = \breve{S}^{(n)}\breve{W}_i$$

and $\mathcal{Z} = \begin{bmatrix} \bar{W}_i & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{N \times (p+1)l}$. Then we have

$$\hat{A}^{(n)} = [W_i^\top W_i + D_1^\top (\mathcal{Z}\hat{\mathbf{S}}^{(n)}\mathcal{Z}^\top \otimes I_N)D_1 + 2W_i^\top \hat{\mathcal{Y}}^{(n)}]$$

and

$$\hat{b}^{(n)} = \Big[w_j^\top W_i + w_j^\top \hat{\mathcal{Y}}^{(n)} - \hat{s}^{(n)\top}\mathcal{X}^\top W_i - \frac{1}{2}\mathrm{vec}(\mathcal{Z}\hat{\mathbf{S}}^{(n)\top}\mathcal{X}^\top)^\top D_2 - \frac{1}{2}\mathrm{vec}(\mathcal{X}\hat{\mathbf{S}}^{(n)\top}\mathcal{Z}^\top)^\top D_1 \Big]^\top.$$