Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-15 (2018) 646–651

Nonlinear FIR Identification with Model Order Reduction Steiglitz-McBride ⋆

Mina Ferizbegovic * **Miguel Galrinho** * **Håkan Hjalmarsson** *

* *Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {minafe,galrinho,hjalmars}@ kth.se).*

Abstract: In system identification, many structures and approaches have been proposed to deal with systems with non-linear behavior. When applicable, the prediction error method, analogously to the linear case, requires minimizing a cost function that is non-convex in general. The issue with non-convexity is more problematic for non-linear models, not only due to the increased complexity of the model, but also because methods to provide consistent initialization points may not be available for many model structures. In this paper, we consider a non-linear rational finite impulse response model. We observe how the prediction error method requires minimizing a non-convex cost function, and propose a three-step least-squares algorithm as an alternative procedure. This procedure is an extension of the Model Order Reduction Steiglitz-McBride method, which is asymptotically efficient in open loop for linear models. We perform a simulation study to illustrate the applicability and performance of the method, which suggests that it is asymptotically efficient.

Keywords: System identification, non-linear, least-squares identification

1. INTRODUCTION

Non-linear system identification is a vast field, as many different structures can be considered to model non-linear behaviors. Early work described non-linear models with Volterra and Wiener series (Schetzen, 1980; Rugh, 1981). Although this approach is impractical for many identification purposes because of the typically large number of parameters, it still serves as basis for some identification algorithms (Kibangou and Favier, 2006, 2009; Birpoutsoukis et al., 2017; Stoddard et al., 2017).

However, the most popular approaches to non-linear identification can be seen as generalizations of structures that are commonly used in linear identification. One category consists of non-linear state-space models, for which different non-linearities have been considered (Suykens et al., 1995; Paduart et al., 2010; Schön et al., 2011). Another category consists of block-structured models, where blocks with linear dynamic transfer functions and blocks with static non-linearities are connected in a network structure. The papers by Billings (1980); Giri and Bai (2010); Schoukens and Tiels (2016) reflect the evolution of structures and methods that have been used. Finally, a third category consists of non-linear generalizations of standard parametric linear models, of which the non-linear ARMAX (NARMAX) is the most general (Leontaritis and Billings, 1985; Chen and Billings, 1989; Billings, 2013). Also here, for each model structure different assumptions can be made on the non-linear function.

With linear models, it is often possible to assume, as consequence of the central limit theorem, that the prediction errors have a Gaussian distribution, and maximum likelihood can be applied. However, for many structures in the latter category of non-linear models, it is not possible to obtain a closed-form expression for the noise source, and the prediction errors will not have a known distribution in general, which hinders the application of maximum likelihood. Nevertheless, prediction error methods can still be applied (Billings and Voon, 1986), although they may not always be the most appropriate for complex NARMAX models (Chen et al., 1989; Billings and Mao, 1998).

However, for some non-linear model structures where it is possible to solve for the noise source, the prediction error cost function has a form identical to the linear case, and corresponds to maximum likelihood. Then, the prediction error method (PEM) may be applied, providing asymptotically efficient estimates under mild assumptions if the model structure is correct (Caines and Ljung, 1976). One limitation with this approach is that the cost function, whose non-convexity can already be problematic for some linear models, becomes yet more complicated to minimize.

This limitation is present even for relatively simple non-linear models. In this paper, we consider a non-linear finite impulse response (NLFIR) model, where the non-linearity is a rational function whose numerator and denominator are non-linear in the input. This problem has been inspired by the models used in reaction networks (Hagrot et al., 2016). For linear FIR models—as well as non-linear with a polynomial nonlinearity in the input—the predictor is linear in the model parameters, and the minimizer of a quadratic prediction error cost function can be found by least squares. However, for rational functions, the problem becomes non-convex, and appropriate consistent methods for initialization are scarcer than for the linear case: in

⋆ This work was supported by the Swedish Research Council under contracts 2015-05285 and 2016-06079.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.09.218

particular, instrumental variable and least-squares methods have been proposed, but only for polynomial non-linearities, and recursive implementations of the algorithms are often required (Billings and Voon, 1984).

In this paper, we propose an alternative procedure to avoid the non-convexity of PEM, based on the Model Order Reduction Steiglitz-McBride (MORSM) method (Everitt et al., 2017), which has been applied to linear model structures with the same purpose. For the linear case, it has been observed that the method is robust against convergence to local minima, and shown to be consistent and asymptotically efficient (Everitt et al., 2017).

Our contributions are the following. First, we propose an algorithm based on an extension of MORSM to obtain estimates for the considered non-linear model structure. Second, we perform numerical simulations illustrating the applicability of the method as a standalone method or to provide initial conditions for PEM. Third, we argue that the method is asymptotically efficient and support this claim with simulations.

2. PRELIMINARIES

In this section, we introduce the model and the identification problem, as well as the main limitation of the standard existing approach, which motivates the proposed method.

2.1 Model and Identification Problem

A parametric non-linear FIR model can be written as

$$y(t) = F(u(t), u(t-1), \dots, u(t-p); \theta) + e(t), \tag{1}$$

where $u(t)$ is the input at time t, p is the number of delayed inputs that influence the output $y(t)$, F is a non-linear function, $e(t)$ is Gaussian white noise with finite variance and uncorrelated with $u(t)$, and θ is a column vector containing the model parameters.

We consider non-linear FIR models where F is given by

$$F(u_{t-p}^t, \theta) = \frac{f(u_{t-p}^t)\theta}{1 + g(u_{t-p}^t)\theta}, \tag{2}$$

where $f(u_{t-p}^t)$ and $g(u_{t-p}^t)$ are row vectors whose entries are known smooth non-linear functions of the last p inputs $u_{t-p}^t = \{u(t), u(t-1), ..., u(t-p)\}$. We assume that the data has been generated by the considered model structure evaluated at some true parameters θ^o. Also, for simplicity of presentation of the method we will propose, we assume that $|g(u_{t-p}^t)\theta^o| < 1$; in Section 6, we address how this assumption could be relaxed.

The problem is to estimate $\theta = [\theta_1 \ \theta_2 \ \cdots \ \theta_m]^\top$ using the known input sequence $\{u(t)\}$ and the measured output sequence $\{y(t)\}$, for $t = 1, 2, ..., N$. This can be solved with PEM, which we now proceed to review.

2.2 Prediction Error Method for FIR models

PEM is a benchmark for identification of parametric linear models. The essential idea of PEM is to obtain an estimate of the model parameters θ by minimizing a cost function of the prediction errors, denoted $\varepsilon(t, \theta)$, which are chosen such that at the true parameter values θ^o, $\varepsilon(t, \theta^o) = e(t)$. For Gaussian errors and a correct model structure, PEM with the quadratic cost function $V_N(\theta) = \frac{1}{N}\sum_{t=1}^N \varepsilon^2(t, \theta)$ is equivalent to maximum likelihood, and provides asymptotically efficient estimates (Ljung, 1999).

In the case of some non-linear model structures, it is not possible to solve for $e(t)$ to define the prediction errors $\varepsilon(t, \theta)$. In this case, a prediction error method may still be applied, but it does not correspond to maximum likelihood, which is not applicable because the prediction errors have an unknown distribution (Billings and Voon, 1986). However, this is not the case of the problem we consider, where the prediction errors defined by $\varepsilon(t, \theta) = y(t) - F(u_{t-p}^t, \theta)$ correspond to $e(t)$ at the true parameters θ^o. Thus, the global minimizer of the PEM cost function is an asymptotically efficient estimate for our problem (Caines and Ljung, 1976). PEM is then an appropriate method to apply, and we now proceed to review what this algorithm consists of for different choices of $F(u_{t-p}^t, \theta)$.

Linear FIR model Let us consider a linear FIR model:

$$y(t) = \sum_{k=0}^{p} \theta_k u(t-k) + e(t). \tag{3}$$

This corresponds to, in the considered model (2), having $f(u_{t-p}^t) = [u(t) \ u(t-1) \ \cdots \ u(t-p)]$ and $g(u_{t-p}^t) = 0_{1\times p+1}$. The cost function for this model, given by

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[y(t) - \sum_{k=0}^{p}\theta_k u(t-k)\right]^2,$$

is quadratic in the model parameters θ, so the global minimizer can be obtained with linear least squares:

$$\hat{\theta}_N = \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^\top(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t), \tag{4}$$

with $\varphi(t) = f^\top(u_{t-p}^t)$. In fact, this is the case also if the non-linearity is only in the input and not in the parameters (i.e., $f(u_{t-p}^t)$ may contain smooth non-linear functions, as long as $g(u_{t-p}^t) = 0$).

Non-linear Rational FIR model We now consider the general case of the model structure (1), where in addition $g(u_{t-p}^t)$ is non-zero and potentially non-linear. Then, the cost function to minimize is given by

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[y(t) - \frac{f(u_{t-p}^t)\theta}{1 + g(u_{t-p}^t)\theta}\right]^2.$$

Unlike the previous case, this cost function is now non-convex in general, and may converge to non-global minima. To reduce the risk of converging to one such minimum, accurate initialization points are required.

This problem with PEM is not specific for non-linear models: for many linear model structures, PEM also requires minimizing a non-convex cost function. For this reason, we now review PEM for a linear output-error model, as well as some methods that do not suffer from non-convexity and can be used to provide initialization points for the PEM optimization.

2.3 Prediction Error Method for Linear OE Models

Consider the output-error (OE) model, given by

$$y(t) = \frac{B(q,\theta)}{F(q,\theta)}u(t) + e(t), \tag{5}$$

where $F(q)$ and $B(q)$ are polynomials in the delay operator q^{-1}, according to

$$F(q) = 1 \ + f_1 q^{-1} + ... + f_{m_f} q^{-m_f}$$
$$B(q) = b_0 + b_1 q^{-1} + ... + b_{m_b} q^{-m_b},$$

and $\theta = \begin{bmatrix} f_1 & \cdots & f_{m_f} & b_0 & b_1 & \cdots & b_{m_b}\end{bmatrix}^\top$. For this model, the PEM cost function is

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[y(t) - \frac{B(q,\theta)}{F(q,\theta)}u(t)\right]^2. \tag{6}$$

In general, this cost function is non-convex in θ, and may converge to non-global minima.

To initialize this cost function, instrumental variable methods (Söderström and Stoica, 2002) and subspace methods (van Overschee and de Moor, 1996), which provide consistent estimates of the model in (5) and do not suffer from non-convexity, are typically used. However, they are in general not asymptotically efficient, and their accuracy is not always possible to establish. Other methods have been proposed as an alternative to PEM that are asymptotically efficient under quite general scenarios (Young, 2008; Galrinho et al., 2014; Zhu and Hjalmarsson, 2016; Everitt et al., 2017).

2.4 Problem Statement and Motivation

As discussed, the non-linear identification problem considered in this paper can be solved with PEM; however, it requires minimizing a non-convex cost function. Although this is the case also for some linear model structures, many methods are available to provide initialization points for PEM, of which some even have the same asymptotic properties. On the other hand, in the case of non-linear model structures, such methods are scarcer: approaches based on recursive instrumental variables and least-squares methods have been proposed by Billings and Voon (1984) for non-linear model structures more general than FIR, but they are not applicable when the non-linearity is a rational function, even in the simpler FIR case.

Motivated by this, the problem considered in this paper is how to solve the identification problem in Section 2.1 without minimizing a non-convex cost function. The method we propose is based on the Model Order Reduction Steiglitz-McBride (MORSM) method when applied to linear OE models, which we proceed to review.

3. MORSM FOR LINEAR OE MODELS

MORSM is a method for estimation of parametric models. It provides asymptotically efficient estimates of the plant in open loop without requiring a parametric noise model and without solving a non-convex optimization problem. Because MORSM uses the Steiglitz-McBride method (Steiglitz and McBride, 1965), we start by reviewing this method.

3.1 The Steiglitz-McBride Method

Let us consider the OE model given by (5). As previously mentioned, the PEM cost function (6) associated with this model is non-convex in general, and requires local non-linear optimization routines. The Steiglitz-McBride method was proposed for output-error models to avoid this limitation of PEM, using iterative least squares.

The method starts by estimating an ARX model

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

by minimizing the cost function

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[F(q,\theta)y(t) - B(q,\theta)u(t)\right]^2 \tag{7}$$

Unlike (5), this model has a cost function quadratic in the model parameters, and is estimated by least squares, providing an estimate $\hat{\theta}_N^{(0)}$ of θ. If the true system is of OE structure, this is a biased estimate of θ. Therefore, the method then filters the input and output according to

$$y_0^f(t) = \frac{1}{F(q,\hat{\theta}_N^{(0)})}y(t), \qquad u_0^f(t) = \frac{1}{F(q,\hat{\theta}_N^{(0)})}u(t), \tag{8}$$

and re-estimates an ARX model, now using $\{y_0^f(t), u_0^f(t)\}$ as data set. This provides a new estimate $\hat{\theta}_N^{(1)}$ of θ. Finally, the method continues to iterate until convergence, creating a new filtered data set using the previous estimate.

The reason for applying this procedure is that, at iteration k, we minimize the cost function

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{F(q,\theta)}{F(q,\hat{\theta}_N^{(k)})}y(t) - \frac{B(q,\theta)}{F(q,\hat{\theta}_N^{(k)})}u(t)\right]^2. \tag{9}$$

Then, upon convergence to some true parameters θ^{o}, the cost function (9) corresponds to (6). Convergence of the Steiglitz-McBride method has been studied by Stoica and Söderström (1981). The method is locally convergent when the additive output noise is white. Additionally, it is globally convergent if the signal-to-noise ratio is sufficiently large. However, this method is not asymptotically efficient. In the following, we review MORSM, which uses the Steiglitz-McBride algorithm applied to another data set, and is asymptotically efficient in one iteration.

3.2 MORSM

Although MORSM is applicable when the noise is colored, we now cover only the OE case, which is sufficient for the purpose of our extension to non-linear FIR models.

Consider that we want to estimate the model (5). MORSM does so in the following three steps. First, estimate a non-parametric FIR model. Second, use the non-parametric model estimate to generate a simulated data set. Third, apply the Steiglitz-McBride method.

Step 1: non-parametric FIR model Consider a linear FIR model as in (3), but where we now use a different notation to emphasize that it is a non-parametric model:

$$y(t) = B(q, b^n)u(t) + e(t),$$

where $B(q, b^n) = \sum_{k=0}^{n} b_k q^{-k}$, and $b^n = \begin{bmatrix} b_0 & b_1 & \cdots & b_n\end{bmatrix}^\top$ is the parameter vector. By non-parametric, we mean that the order n should be chosen arbitrarily large to model (5) with good accuracy. The PEM estimate $\hat{b}^n$ of a linear FIR model is obtained with least squares, as in (4).

Step 2: generate a simulated data set Motivated by an asymptotic maximum likelihood cost function (Wahlberg, 1989), it is possible to use the non-parametric FIR estimate to obtain estimates of the OE model of interest by minimizing the cost function (Everitt et al., 2017)

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[B(q,\hat{b}^n)u(t) - \frac{B(q,\theta)}{F(q,\theta)}u(t)\right]^2. \quad (10)$$

We notice that (10) has the same form as the OE cost function (6), if we interpret the term $B(q,\hat{b}^n)u_t$ as an output. Then, we may define $\hat{y}(t) := B(q,\hat{b}^n)u_t$ and use the data set $\{\hat{y}(t), u(t)\}$ to estimate the model of interest.

Step 3: Steiglitz-McBride The cost function (10) can be minimized using non-linear optimization. This is the idea of the ASYM method, which has some numerical advantages with respect to PEM, even if the cost function is still non-convex (Zhu, 2001). However, with MORSM we use instead the Steiglitz-McBride method with the data set $\{\hat{y}(t), u(t)\}$. This provides asymptotically efficient estimates in open loop with one Steiglitz-McBride iteration.

4. MORSM FOR NON-LINEAR RATIONAL FIR MODELS

In this section, we propose the algorithm of this paper: an extension of MORSM for linear OE models to estimate non-linear FIR modules with structure (2). The idea is the following. First, estimate a non-parametric FIR model based on a function expansion of the non-linear function. Second, simulate the output signal using the non-parametric estimate. Third, similarly to Steiglitz-McBride, apply an iterative least-squares algorithm to estimate the parametric model coefficients.

Step 1: non-parametric FIR model Consider the denominator of (2). For simplicity, we assume $|g(u_{t-p}^t)\theta| < 1$ and take the polynomial expansion

$$\frac{1}{1+g(u_{t-p}^t)\theta} = \sum_{k=0}^{\infty} a_k \tilde{g}^k(u_{t-p}^t)$$

for some a_k parameters, and where the scalars $\tilde{g}(u_{t-p}^t)$ are determined by the polynomial expansion and the input signal values. Multiplying with the numerator, we can write

$$f(u_{t-p}^t)\theta \sum_{k=0}^{\infty} a_k \tilde{g}^k(u_{t-p}^t) = \sum_{k=0}^{\infty} b_k h^k(u_{t-p}^t)$$

for some b_k parameters, where the values $h(u_{t-p}^t)$ are derived from the product of the entries in $f(u_{t-p}^t)$ with $\tilde{g}(u_{t-p}^t)$. Truncating this expansion to n coefficients, we have the model

$$y(t) = \sum_{k=0}^{n} b_k^n h^k(u_{t-p}^t) + e(t),$$

with $b^n = [b_0\ b_1\ \cdots\ b_n]^\top$. This is an FIR model linear in the model parameters, which can be estimated with least squares, as in (4), with b^n instead of θ and $\varphi(t) = \left[h^0(u_{t-p}^t)\ h^1(u_{t-p}^t)\ \cdots\ h^n(u_{t-p}^t)\right]^\top$.

Step 2: generate a simulated data set Similarly to MORSM for linear models, we now use the non-parametric model estimate to generate simulated data according to

$$\hat{y}(t) = \sum_{k=0}^{n} \hat{b}_k^n h^k(u_{t-p}^t),$$

where $\hat{b}^n$ is the estimate of b^n.

Step 3: Iterative Least Squares Analogously to (10), we could now obtain estimates of θ by minimizing

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\hat{y}(t) - \frac{f(u_{t-p}^t)\theta}{1+g(u_{t-p}^t)\theta}\right]^2. \quad (11)$$

The problem with minimizing (11) is that, in general, it is non-convex in θ. Nevertheless, the similarities between (10) and (11) suggest that an iterative least-squares algorithm like Steiglitz-McBride could be used instead. However, the Steiglitz-McBride algorithm cannot be used as in the linear case, because the denominator $1 + g(u_{t-p}^t)\theta$ is not a filter, and (8) cannot be applied. We now consider how to apply an iterative least squares algorithm to this scenario.

First, in order to make the problem solvable with least squares, we consider instead the cost function

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[(1+g(u_{t-p}^t)\theta)\hat{y}(t) - f(u_{t-p}^t)\theta\right]^2.$$

This is equivalent to the first step in the Steiglitz-McBride method for linear OE models, where the cost function (7) is minimized. An initial least squares estimate $\hat{\theta}_N^{(0)}$ is then given by (4), with $y(t)$ replaced by our simulated output $\hat{y}(t)$ and $\varphi^\top(t) = f(u_{t-p}^t) - g(u_{t-p}^t)\hat{y}(t)$.

Second, to mimic the cost function (9), we proceed with re-estimating θ by minimizing the cost function

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{1+g(u_{t-p}^t)\theta}{1+g(u_{t-p}^t)\hat{\theta}_N^{(1)}}\hat{y}(t) - \frac{f(u_{t-p}^t)\theta}{1+g(u_{t-p}^t)\hat{\theta}_N^{(1)}}\right]^2.$$

This consists in computing the least-squares solution (4) with $y(t)$ replaced by the corrected simulated output

$$\hat{y}_0(t) = \frac{\hat{y}(t)}{1+g(u_{t-p}^t)\hat{\theta}_N^{(0)}}$$

and with $\varphi^\top(t)$ replaced by

$$\varphi_0^\top(t) = \frac{f(u_{t-p}^t) - g(u_{t-p}^t)\hat{y}(t)}{1+g(u_{t-p}^t)\hat{\theta}_N^{(0)}}.$$

Then, the estimate $\hat{\theta}_N^{(0)}$ can be used to initialize an iterative procedure that consists in using least-squares to minimize, at iteration k, the cost function

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{1+g(u_{t-p}^t)\theta}{1+g(u_{t-p}^t)\hat{\theta}_N^{(k)}}\hat{y}(t) - \frac{f(u_{t-p}^t)\theta}{1+g(u_{t-p}^t)\hat{\theta}_N^{(k)}}\right]^2.$$

Because this method is an extension of MORSM for non-linear models, we will refer to it as non-linear MORSM (NMORSM). As we will see in Section 5, a simulation study suggests that, analogously to MORSM for linear models, the estimate $\hat{\theta}_N^{(1)}$ of NMORSM is an asymptotically efficient estimate of θ for the considered problem.

Step 3 has been developed as an adaptation of Steiglitz-McBride for the considered non-linear model. Then, like Steiglitz-McBride, it could be directly applied to measured data (we will refer to this as non-linear Steiglitz-McBride).

However, unlike Steiglitz-McBride for linear OE models, this will not provide consistent estimates of the non-linear FIR model of interest, as we will illustrate in Section 5.

An Example To illustrate the algorithm of NMORSM, let us consider the rational non-linear FIR model given by

$$y(t) = \frac{\theta_1 u(t-1)}{1+\theta_2 u^2(t)} + e(t).$$

In terms of (2), we have $f(u^t_{t-p}) = [u(t-1)\ 0]$, $g(u^t_{t-p}) = [0\ u^2(t)]$, and $\theta = [\theta_1\ \theta_2]^\top$.

In Step 1, for $|\theta_2 u^2(t)| < 1$, we perform the expansion

$$\frac{\theta_1 u(t-1)}{1+\theta_2 u^2(t)} = \theta_1 u(t-1)\sum_{k=0}^{\infty} a_k u^{2k}(t) = \sum_{k=0}^{\infty} b_k u(t-1) u^{2k}(t).$$

Here, the coefficients a_k and b_k are related to θ by $a_k = \theta_2^k$ and $b_k = \theta_1\theta_2^k$, but algorithm does not use this explicitly. Truncating the expansion to n coefficients, our model is

$$y(t) = \sum_{k=0}^{n} b_k u(t-1)u^{2k}(t) + e(t),$$

for which we estimate the parameter b^n by least squares with $\varphi(t) = \left[u(t-1)\ \ u(t-1)u^2(t)\ \cdots\ u(t-1)u^{2n}(t))\right]^\top$.

In Step 2, we generate simulated data using

$$\hat{y}(t) = \sum_{k=0}^{n} \hat{b}^n_k u(t-1)u^2(t),$$

where $\hat{b}^n$ is the estimate of b^n. Finally, in Step 3, we iterate by minimizing the cost function

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{1+\theta_2 u^2(t)}{1+\hat{\theta}_2^{(k)} u^2(t)}\hat{y}(t) - \frac{\theta_1 u(t-1)}{1+\hat{\theta}_2^{(k)} u^2(t)}\right]^2$$

using least-squares.

5. NUMERICAL SIMULATIONS

Let us consider the rational non-linear FIR model given by

$$y(t) = \frac{\theta_1 u(t-1) + \theta_2 u^2(t-2)}{1+\theta_3 u^2(t) + \theta_4 u(t-1)u(t-2)} + e(t),$$

where $u(t)$ is uniformly distributed noise in the interval $[-1, 1]$, the noise $e(t)$ has standard deviation $\sigma = 0.5$, and $\theta^{\mathrm{o}} = [1\ 2\ 0.6\ 0.3]^\top$. The model has correct structure. A thousand Monte Carlo simulation were performed with ten different sample sizes N between 100 to 100000.

We compare PEM and NMORSM. For PEM, a concern is how to initialize the optimization procedure. We considered two different initialization points: the true parameter values and the estimates provided by NMORSM. For NMORSM, a concern is how to choose the order n of the non-parametric model. The solution is identical to the linear case: we apply NMORSM for a grid of values of n, compute the PEM cost function for each of the parametric models obtained, and choose the model estimate at which this cost function is lowest (Everitt et al., 2017). To illustrate how non-linear Steiglitz-McBride (i.e., Step 3 of NMORSM if applied to measured data) is biased, we include also this procedure.

Thus, the following methods are compared:

- NMORSM with a grid of non-parametric orders $n = \{3, 5, 7\}$ and ten iterations (i.e., the estimate $\hat{\theta}_N^{(10)}$);
- PEM initialized at the true model parameters, using the `fminsearch` function in MATLAB2017b, with a maximum number of 1000 iterations (PEM-true);
- PEM with the same implementation, but initialized at the NMORSM estimate (PEM-NMORSM).
- non-linear Steiglitz-McBride, with a maximum of 1000 iterations (SM-nl).

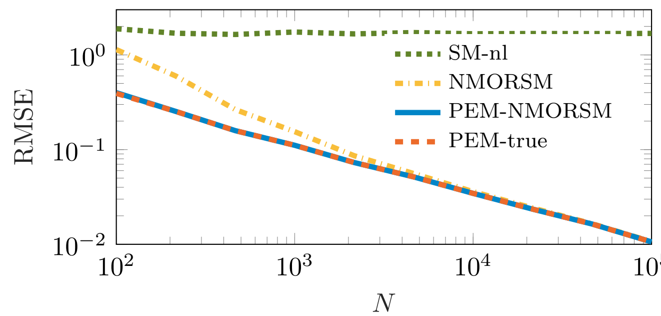

Fig. 1. Average RMSE as function of sample size for different methods.

The accuracy of each estimate is computed by measuring the root mean-square error (RMSE $= ||\hat{\theta}_N - \theta^{\mathrm{o}}||$), where $\hat{\theta}_N$ is a vector with the estimated model parameters for a particular method.

The average RMSE for each sample size and method is presented in Fig. 1. This simulation suggests that NMORSM has the same asymptotic performance as PEM initialized at the true model parameters, which implies that the method may be asymptotically efficient. For lower sample sizes, NMORSM does not have the same performance as PEM, but its estimate can be used to initialize PEM: with this initialization, PEM had the same performance as with initialization at the true parameters. Finally, the RMSE of non-linear Steiglitz McBride (i.e., Step 3 of MORSM if it were applied to measured data) does not decrease as N increases. This implies that these estimates are biased, and it motivates the importance of using the non-parametric estimate and simulated data.

6. DISCUSSION

A system identification method has been proposed for estimating rational non-linear FIR models. This method is an extension of MORSM, which has been proposed to deal with the non-convexity of PEM for linear model structures. The idea of the method is to first estimate a non-parametric model that approximates the true system; this is then used to generate a simulated data set to which the Steiglitz-McBride algorithm is applied. In the linear case, it has been shown by Everitt et al. (2017) that the method is asymptotically efficient in open loop.

In this contribution, we have extended this procedure to be applicable to rational non-linear FIR model structures. First, we estimate a non-parametric model based on a function expansion of the non-linear rational function. Second, we use this estimate to generate a simulated data set. Third, we apply iterative least squares intercalated with data correction between each iteration, analogously to Steiglitz-McBride for linear systems. A numerical simulation suggests that, similarly to linear MORSM, the proposed method is asymptotically efficient for rational non-linear FIR models. For smaller sample sizes, it can also be useful to provide initial conditions for PEM.

For the non-parametric model estimated in the first step, we used a polynomial expansion of $[1 + g(u^t_{t-p})\theta]^{-1}$ in the input. The choice of polynomial expansion is for simplicity, but it has some disadvantages. First, we must have that $|g(u^t_{t-p})\theta^o| < 1$. Second it imposes some excitation conditions: for example, if the input $u(t)$ has repeated values, the least-squares problem becomes ill-conditioned. These limitations require the use of inputs that are within certain bounds and unlikely to have repeated values (such as uniformly distributed noise, as we used in the example).

To deal with these limitations, in future work we will consider alternative function expansions. We will provide a theoretical analysis of the method, perform more simulations to study robustness to model errors, and consider extensions to more general non-linear model structures.

REFERENCES

Billings, S. (1980). Identification of nonlinear systems: a survey. *IEEE Proceedings D—Control Theory and Applications*, 127(6), 272–285.

Billings, S. (2013). *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. John Wiley & Sons.

Billings, S. and Mao, K. (1998). Model identification and assessment based on model predicted output. *U.o.S. ACSE Trans, University of Sheffield*, Research Rep. 714.

Billings, S. and Voon, W. (1984). Least squares parameter estimation algorithms for non-linear systems. *International Journal of Systems Science*, 15.

Billings, S.A. and Voon, W.S.F. (1986). A prediction-error and stepwise-regression estimation algorithm for non-linear systems. *Int. Journal of Control*, 44(3), 803–822.

Birpoutsoukis, G., Marconato, A., Lataire, J., and Schoukens, J. (2017). Regularized nonparametric volterra kernel estimation. *Automatica*, 82, 324–327.

Caines, P. and Ljung, L. (1976). Asymptotic normality and accuracy of prediction error estimators. In *Joint Automatic Control Conference*, 13, 586–592.

Chen, S. and Billings, S. (1989). Representations of non-linear systems: the NARMAX model. *Int. Journal of Control*, 49(3), 1013–1032.

Chen, S., Billings, S., and Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of control*, 50(5), 1873–1896.

Everitt, N., Galrinho, M., and Hjalmarsson, H. (2017). Open-loop asymptotically efficient model reduction with the Steiglitz-Mcbride method. *accepted for publication in Automatica*, arXiv:1610.08534.

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

Giri, F. and Bai, E. (2010). *Block-oriented Nonlinear System Identification, Lecture Notes in Control and Information Sciences*. Berlin Heidelberg: Springer.

Hagrot, E., Oddsdóttir, H.Æ., Hosta, J.G., Jacobsen, E.W., and Chotteau, V. (2016). Poly-pathway model, a novel approach to simulate multiple metabolic states by reaction network-based model–Application to amino acid depletion in CHO cell culture. *Journal of biotechnology*, 228, 37–49.

Kibangou, A. and Favier, G. (2006). Identification of Wiener-Hammerstein systems modeling using diagonal Volterra kernels coefficients. *IEEE Signal Processing Letters*, 13, 381–384.

Kibangou, A. and Favier, G. (2009). Parallel-cascade Wiener systems using joint diagonalization of third-order Volterra kernel slices. *IEEE Signal Proc. Letters*, 16, 188–191.

Leontaritis, I. and Billings, S. (1985). Input-output parametric models for non-linear systems part I: deterministic non-linear systems. *Int. Journal of Control*, 41(2), 303–328.

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

Paduart, J., Lauwers, L., Swevers, J., Smolders, K., Schoukens, J., and Pintelon, R. (2010). Identification of nonlinear systems using polynomial nonlinear state space models. *Automatica*, 46(4), 647–656.

Rugh, W.J. (1981). *Nonlinear System Theory: The Volterra Wiener Approach.* Johns Hopkins University Press.

Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems.* Wiley.

Schön, T.B., Wills, A., and Ninness, B. (2011). System identification of nonlinear state-space models. *Automatica*, 47(1), 39 – 49.

Schoukens, M. and Tiels, K. (2016). Identification of non-linear block-oriented systems starting from linear approximations: A survey. *under revision*, arXiv:1607.01217.

Söderström, T. and Stoica, P. (2002). Instrumental variable methods for system identification. *Circuits, Systems and Signal Processing*, 21(1), 1–9.

Steiglitz, K. and McBride, L. (1965). A technique for the identification of linear systems. *IEEE Trans. on Automatic Control*, 10, 461–464.

Stoddard, J.G., Welsh, J.S., and Hjalmarsson, H. (2017). EM-based hyperparameter optimization for regularized Volterra kernel estimation. *IEEE control systems letters*, 1(2), 388–393.

Stoica, P. and Söderström, T. (1981). The Steiglitz-McBride identification algorithm revisited—convergence analysis and accuracy aspects. *IEEE Transactions on Automatic Control*, 26(3), 712–717.

Suykens, J., Moor, B.D., and Vandewalle, J.R. (1995). Nonlinear system identification using neural state space models, applicable to robust control design. *International Journal of Control*, 62, 129–152.

van Overschee, P. and de Moor, B. (1996). *Subspace identification for linear systems: theory, implementation, applications.* Kluwer Academic, Boston.

Wahlberg, B. (1989). Model reduction of high-order estimated models: the asymptotic ML approach. *International Journal of Control*, 49(1), 169–192.

Young, P. (2008). The refined instrumental variable method: Unified estimation of discrete and continuous-time transfer function models. *Journal Européen des Systèmes Automatisés*, 42, 149–179.

Zhu, Y. (2001). *Multivariable System Identification for Process Control.* Pergamon.

Zhu, Y. and Hjalmarsson, H. (2016). The Box-Jenkins Steiglitz-McBride algorithm. *Automatica*, 65, 170–182.