Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-5 (2016) 061–066

Nonlinear System Identification with Regularized Local FIR Model Networks

Tobias Münker * **Oliver Nelles** *

* *Measurement and Control Engineering - Mechatronics, University of Siegen, Germany (e-mail: tobias.muenker@uni-siegen.de).*

Abstract: An algorithm for the identification of nonlinear black-box systems is introduced using novel techniques for the regularized estimation of impulse responses for linear systems. Based on a comparison of the advantages and disadvantages of (N)FIR and (N)ARX model structures for the linear and nonlinear case it is outlined that the novel regularized FIR model estimation removes the major drawback of high parameter variance from the FIR model and makes it thus usable as a local model structure in local model networks. The estimation of the local FIR models is performed with a special regularization matrix, which is derived from the concept of reproducing kernel Hilbert spaces incorparating the knowledge of the exponential decay of the impulse response of a stable system. The algorithm is applied to a test system and is, in contrast to local ARX models, always able to achieve stability and a fairly good prediction accuracy.

Keywords: System Identification, Local Linear Model Tree (LOLIMOT), Regularization, Model complexity selection, Bias-variance trade-off, FIR

1. INTRODUCTION

In this contribution the problem of nonlinear system identification is considered. Therefore N measurements $Z = \{(u_i, y_i)\}_{i=1,...,N}$ are given, which have been generated by a nonlinear dynamic system described by the input/output relationship

$$y_k = f(y_{k-1}, ..., y_{k-n}, u_k, ..., u_{k-n})) + d_k. \quad (1)$$

Here d_k is a white noise sequence with variance σ^2, y_k is the model output and u_k is the input at the discrete time step k. The problem of identifcation is now to find a model $\hat{f}$ that is able to describe the behaviour of the true process f based on the measurements Z.

When f is a linear function the problem is, compared to the nonlinear case, in a mature terrain which is usually treated by prediction error methods (PEM), for an overview see e.g. Ljung (1999). Nevertheless recently for the identification of linear systems a novel method has been proposed by Pillonetto and De Nicolao (2010) for the identification of FIR systems which is related to classical regularization approaches by Chen et al. (2012). A good overview of these novel methods is provided by Pillonetto et al. (2014). These contributions propose a novel regularization approach, which considers the stability property of the system and penalizes the integral of the squared first or second order derivative of the impulse response. Since compared to other problems in the field of machine learning the number of datapoints strongly exceeds the number of parameters efficient algorithms have been developed by Chen and Ljung (2013), which consider this property. For a general overview of kernel methods in machine learning the reader is referred to the books Schölkopf and Smola (2002) or Rasmussen and Williams (2005).

Unfortunatly the behaviour of many technical and also other systems strongly depends on their operation point. Thus the problem of system identification is nonlinear for these systems. Local model networks identified with the LOLIMOT algorithm, see Nelles (1997) or for a comprehensive treatment Nelles (2001), offer a good way to extend many approaches used for linear system identification to the nonlinear case, since the subsequent splitting of the models provides a good way to deal with the *curse of dimensionality*. In this contribution it will be shown that this is possible for the regularized FIR identification approach as well. Therefore regularized local models will be identified with the LOLIMOT construction algorithm.

The sequel of the contribution is as follows: First ARX and FIR as basic model structures for linear and nonlinear systems are reviewed and compared. In Sect. 3 identification of locally regularized FIR models with LOLIMOT will be described and in Sect. 4 it will be shown how the novel regularization approaches can be incorparated in the local model identification. Section 5 reviews the implementation of the regularized identification suited for system identification and describes how to calculate the number of effective parameters in that framework. In the last section the performance of the proposed algorithm and the shortcomings of local ARX models will be demonstrated with an example.

2. COMPARISON OF ARX AND FIR MODEL STRUCTURES

The most commonly applied models for the identification task are the ARX model and the FIR model. In the ARX case the model $\hat{f}$ depends on the previous n outputs $\mathbf{y}_k = (y_{k-1}, \ldots, y_{k-n})$ and delayed inputs $\mathbf{u}_k = (u_{k-1}, \ldots, u_{k-n})$ of the system, while in the the FIR case only delayed versions of the inputs are utlized in the model. For FIR

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.07.090

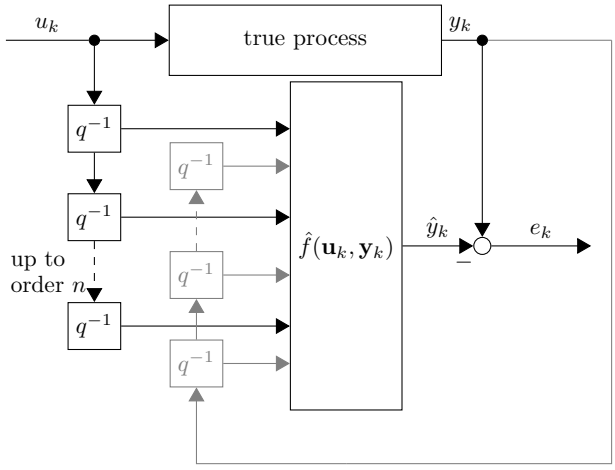

Fig. 1. General NARX and NFIR model structure

models n has to be chosen significantly higher than for ARX models. The general structure is depicted in Fig. 1, where the gray elements have to be omitted in the NFIR case.

2.1 The linear case

If the model $\hat{f}$ is chosen linear the general NARX model simplifies to

$$\begin{aligned}\hat{y}_k = b_1 u_{k-1} + \ldots + b_n u_{k-n} \\ + a_1 y_{k-1} + \ldots + a_n y_{k-n}\end{aligned} \tag{2}$$

and the FIR model can be written as

$$\hat{y}_k = b_1 u_{k-1} + \ldots + b_n u_{k-n}. \tag{3}$$

This is illustrated in Fig. 2 where also the effect of the noise d_k is displayed.

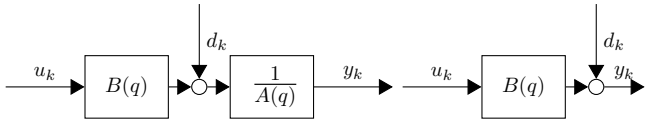

Fig. 2. Model structure of linear ARX and FIR model

ARX and FIR are both linear regression structures and thus can be estimated by least squares. Beside the fact that both are simple and relatively easy to handle this is most certainly the reason why both are very popular. In the assessment of the authors, the number of ARX compared to FIR model applications might be around a ratio of 2:1. One difference between the ARX and the FIR model is the error that is optimized. While ARX optimizes the equation (one-step prediction) error, FIR optimizes the output (simulation) error. Usually the simulation error is the true objective and the equation error is just required as an approximation to keep the model linear in its parameters. Furthermore the parameter estimation for the ARX is biased for white output disturbances favoring high frequencies, while the FIR structure is not. Comparing the selected model order ARX typically is very sensitive with respect to the 'correct' model structure, that is order and dead time. Also here FIR is superior since the order corresponds just to the degree of accuracy. Dead time and order of the process are kind of automatically 'detected' by FIR due to the estimation of the impulse response coefficients. An advantage of the ARX model is, that it can be utilized for the description of unstable processes. Unfortunatly this also means that ARX models can become unstable for stable processes because of errors in the estimated coefficients. The greatest advantage of the ARX model structure compared to FIR models is that the number of parameters is significantly lower, which influences the bias-variance tradeoff in a considerably positive way. Recently this disadvantage has been removed with the application of regularization techniques by Pillonetto and De Nicolao (2010) and this makes FIR structures now very advantageous for the linear case. Before continuing the discussion of how to apply these results in the nonlinear case, it will be discussed how ARX and FIR models in general behave in that case.

2.2 The nonlinear case

For models with a nonlinear structure all of the considerations for linear systems hold as well. Additionally, some new aspects need to be taken into account and some of the properties above have to be weighted very differently. While in the linear case FIR requires much more *parameters* than ARX, in the nonlinear case it requires much more *dimensions*. Estimation of paramters is problematic because the compelxity of least squares solution is $O(n^3)$ and FIR models often exhibit too high variance errors. However, more dimensions are often catastrophic due to the curse of dimensionality. This is the main reason why NARX is much more widely used than NFIR. In the assessment of the authors, the number of NARX vs. NFIR model applications might be around a ratio of at least 20:1. Unfortunately NARX models tend to run into stability problems (see the example in Sect. 6). While in the linear case, model stability can at least be readily checked by calculating the poles, this is not so easy in the nonlinear case. This issue is a constant concern in industry and calls for better solutions.

Most of the above arguments for (N)FIR models generalize to any orthonormal basis functions model (like Laguerre or Kautz). However, important new and complex issues arise with respect to the determination of the filter pole(s) that are beyond the scope of this paper.

In summary, ARX and FIR approaches have their own, often opposite, advantages and drawbacks. In the linear case the overall assessment depends on the circumstances. Both approaches are realistic for applications. In the nonlinear world, the high dimensionality of NFIR practically ruins this approach.

Local linear model networks offer a loophole to this dilemma. The input space for the local linear models may be chosen high-dimensional therefore yielding many parameters to be estimated for the local FIR models. But the input space for the validity functions that control which local linear model is how much active, can often be chosen very low-dimensional comparable (or identical) to NARX. This brings the tradeoffs between NARX and NFIR back to (or at least closer to) the tradeoffs in the linear scenario.