Efficient Pole Optimization of Nonlinear Laguerre Filter Models

Tim Oliver Heinz and Oliver Nelles
Automatic Control, Mechatronics
Department of Mechanical Engineering
University of Siegen
D-57068 Siegen, Germany
Email: tim.heinz@uni-siegen.de

Abstract—The problem of identifying nonlinear dynamic systems with one or more inputs and one output (SISO/MISO) is considered. The properties of orthonormal basis functions (OBF) as inputs for a nonlinear approximator are compared to the common ARX structure which possesses both, delayed input and output values, entering the nonlinearity. The disadvantage of optimizing the pole in the OBF context is overcompensated by the advantages which arise especially in the nonlinear MISO case. The key idea of this paper is the optimization of the OBF pole while fixing the nonlinear model structure, here called Fixed Structure Optimization (FSO). In the SISO case the computational effort can be decreased dramatically. For the MISO case, the increasing complexity of the pole search is tackled with a staggered optimization approach. The reliability is underlined by demonstrating the low correlation between the OBF poles for uncorrelated inputs. The combination of the proposed repeated fixed structure optimization (FSO) with a staggered optimization for the multivariate extension yields many benefits compared to conventional optimization procedures.

I. Introduction

A orthonormal basis function approach is utilized in this contribution, whereby the responses of the orthonormal filters are used as inputs to the nonlinear approximator. It has been shown in the linear case in [1] that the order of filter chain can be significantly lower than the required order of the corresponding FIR or ARX filter chain. The major drawback of this approach is the nonlinear nonconvex optimization problem, which has to be solved to obtain an estimate for the filter chain pole. An thorough analysis of this problem for the linear case can be found in [2]. Our contribution is focused on that problem for the nonlinear case, we investigate on a common pole search and propose a novel method called fixed structure optimization (FSO)

While most of the dynamic model structures methods such as NARX, nonlinear output error (NOE) depend on output feedback, we propose orthonormal basis functions (NOBF) to avoid model feeddback. For a general overview of existing system identification methods, the reader is referred to [3] and [4].

The paper is structured as follows: First, the advantages and disadvantages of the ARX and OBF structure are compared. While both approaches have often complementary but similarly important properties in the case of one input and one output (SISO), this changes for the multiple input case (MISO). By dealing with multiple inputs, several additional disadvantageous arise for the ARX structure. This motivates the focus on OBF whose only main drawback is the required pole determination/optimization.

In the following section, the pole optimization problem is thoroughly discussed. By fixing the nonlinear model structure during the optimization, we obtain a huge computational benefit in contrast to a conventional optimization. With a nonlinear example process, the superiority of the repeated fixed structure optimization (FSO) is outlined.

Usually real world processes have multiple inputs (MISO). While the NARX structure is not well suited for the MISO case, the NOBF approach can be extended to the MISO case easily. But for each input an individual pole has to be optimized. A grid based search fails, if the number of inputs and/or the resolution is high. For the linear case, the correlation between the poles is calculated. We show that for uncorrelated input signals, the loss function does not depend on the interaction between the poles. For the nonlinear case, this does no longer hold, but the resulting correlation be expected to be small. This justifies the introduction of the staggered optimization approach inside the repeated FSO.

In the last part an example is given, which demonstrates the applicability of the two novel methods.

II. A General Comparison of OBF and ARX Model Structures

Dynamic model structures can generally be divided in two different categories: With and without output feedback. One representative of a linear dynamic model structure with output feedback is the well known AutoRegressive with eXogenous input (ARX) structure. The dynamic behavior depends on the delayed inputs and outputs of the system. When the sum of orthonormal basis functions (OBF) in contrast are used for the representation of the system, no output feedback is necessary to approximate the actual process output. The dynamics is represented by filtering the input. The general structure of an NARX and NOBF model is depicted in Fig. 1. Here the Laguerre filter chain is taken as a representative for OBFs. The original input is filtered by the n Laguerre filters [1]

$$u_{L,i}(k) = \frac{\sqrt{1-p_L^2}}{q-p_L}\left(\frac{1-p_L q}{q-p_L}\right)^{i-1} \cdot u(k) \quad . \quad (1)$$
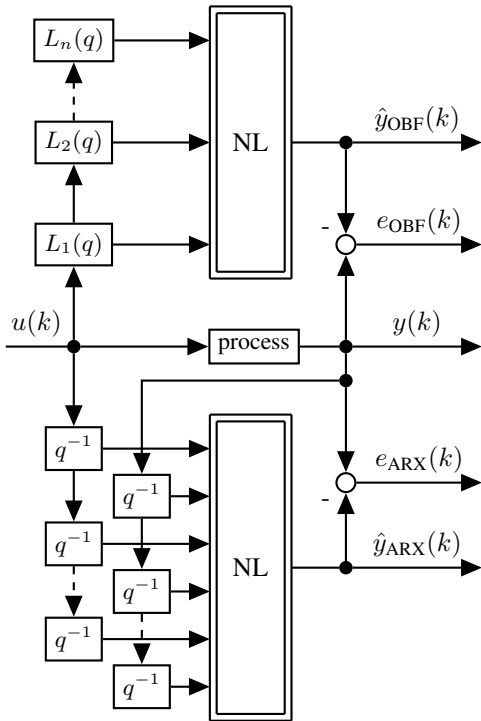
Fig. 1. NOBF (top) and NARX (bottom) models.

In contrast to (N)ARX, a Laguerre filter chain restricts the properties of the estimated dynamic system. The user implies, that the process is non.- or only weakly oscillatory (the pole of Laguerre Filter has just a real part). For an oscillatory process, Kautz Filter (with conjugate-complex poles) can be used [5]. Additionally, the process is assumed to be stable. This constrains the Laguerre pole to $p_L \in [0 \dots 1]$. The very theoretic case of $p_L < 0$ will be omitted here.

A. Linear Models

If the model $\hat{y}$ is linear, the general NARX and NOBF structure simplifies to:

$$\hat{y}_{\text{ARX}}(k) = b_1 u(k-1) + b_2 u(k-2) + \dots + b_m u(k-m)$$
$$- a_1 y(k-1) - a_2 y(k-2) - \dots - a_m y(k-m) \tag{2}$$

$$\hat{y}_{\text{OBF}}(k) = b_1 u_{L,1}(k) + b_2 u_{L,2}(k) + \dots + b_n u_{L,n}(k) \tag{3}$$

After selecting the model order n, the parameters a_i, b_i can be estimated using least squares. One drawback of the OBFs is the determination of the pole in (1). Proposals to determine p_L will be given in Sect. III and Sect. V.

In the linear case, there are different advantages and drawbacks for ARX and OBF.

1) ARX:

+ easy to estimate
+ possible link between parameters and first principles
- suboptimal for multiple inputs (MISO/MIMO), because a common denominator polynomial is required

- ARX optimizes the equation (one-step prediction) error. Usually the simulation error is the true objective and the equation error is just used as a proxy to keep the model linear in its parameters.
- Parameter estimation for ARX is biased for white output disturbances favoring high frequencies
- Typically very sensitive to the correct choice of model order m

2) OBF:

+ easy to estimate
+ optimizes simulation error (output error) which is the true objective in nearly every case
+ easily extendible to MISO systems
+ insensitive to the filter order n
o For badly chosen pole p_L the model order n has to be chosen very large.
- no link between parameters and first principles
- pole/poles of the OBF have to be determined

A more detailed comparison of OBF and ARX properties can be found in [2].

B. Nonlinear Models

In the nonlinear case, all of the above considerations 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.

- In the nonlinear case, where the dynamic of the process varies strongly over the input range, it is hard to determine the best pole for an NOBF.
- The above point is one reason why NARX is much more widely used than NOBF. The dynamics of a NARX model is determined by the estimated parameters. So NARX has the potential to fit the dynamics in every operating point.
- NARX models can tend to run into stability problems. 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. Although significant research results are available for local model networks[6]. This issue is a constant concern in industry and calls for better solutions.
- Most nonlinear structures with output feedback (ARX, OE,) approximate the one-step prediction surface. Therefore, they interpolate transfer functions. In contrast, all nonlinear orthonormal basis function approaches (NFIR, NOBF) interpolate signals. Interpolating transfer functions can lead to very strange effects for models of order higher than 1. This is due to the fact that interpolating polynomials can yield unexpected roots and thus dynamics [4].
- The advantages of linear regression structures like ARX and OBF in the linear world are lost or at least diminish in the nonlinear world. Only for certain model architectures (like local model networks or Gaussian process models) that heavily rely on least squares, big benefits can be drawn from ARX or OBF. Generally, i.e., for nonlinearly
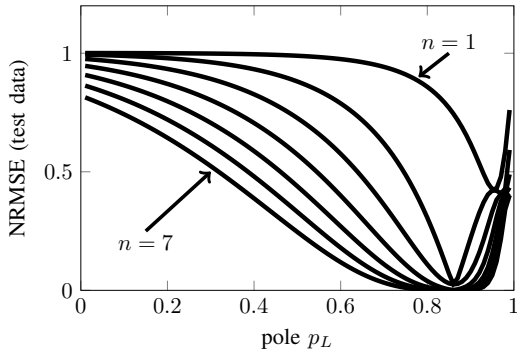
Fig. 2. Normalized root mean squared error (NRMSE) of the training data for different choices of the Laguerre pole.

Fig. 3. Structure of a nonlinear Wiener-Hammerstein system.

Fig. 4. NRMSE of the training data for different choices of the Laguerre pole.

parameterized approximators, the computational advantages compared to OE structures fade.

In summary, ARX and OBF 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 contrast, NARX possesses many disadvantages especially in the MISO case. In comparison, NOBF retains their benefits from the SISO case.

III. Pole Optimization - SISO

In comparison to ARX/NARX, the OBF/NOBF structure has an additional hyperparameter (the pole p_L of the Laguerre filter chain). There is no closed formula, to determine p_L, even in the linear case. Thus a fast and robust pole optimization is required.

In the following, linear and nonlinear second order processes shall be considered.

A. Linear Case

In the linear case (poles at $p = [0.8, 0.9]$), the normalized root mean squared error (NRMSE) for different OBF models is shown in Fig. 2 for different model orders $n = 1, 2, \ldots, 7$ and poles between 0 and 1. It is obvious, that the optimum pole is sensitive to the filter order. For increasing values of n, the NRMSE becomes more insensitive w.r.t. p_L.

For estimating p_L of a linear model several estimation procedures exist (see [7], [8]). An adaptive pole optimization approach is proposed in [9]. This paper focuses on the pole optimization for nonlinear models.

B. Conventional Nonlinear Pole Search

For nonlinear models, the order n affects the number of inputs/ dimensions of the nonlinear approximator directly. To escape the curse of dimensionality, n has to be chosen small. As can be seen from Fig. 2 low orders yield sharp optima and therefore require an quite accurate pole search for a high model quality.

For the nonlinear approximation, the well known Local Linear Model Tree (LOLIMOT) [10] is used to partition the input space with orthogonal splits subsequently in half.
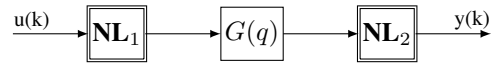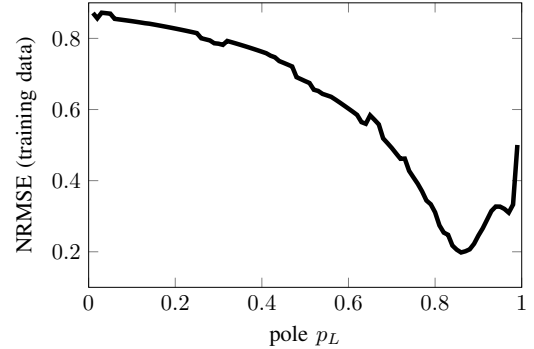
For each separated area, an individual local model (LM) is estimated. To avoid overfitting, the number of local models is determined with the Bayesian information criterion (BIC). The BIC was chosen over the Akakike information criterion (AIC) [11] because its penalty term is bigger and thus simpler models are selected. This is reasonable due to the computational demand and significant dynamic errors due to still suboptimal pole choices.

As a demonstration example, a Wiener-Hammerstein process is considered, see Fig. 3. The nonlinearities are formed by an inverse tangent function and the dynamic system $G(z)$ is a process of second order. The system is excited by an amplitude modulated pseudo random binary signal (APRBS) [4], to cover the whole input space with data. With the resulting input/output signals, the system is identified using Laguerre filter chains with $n = 2$. A nonlinear model is identified for different poles $p_L \in [0, \ldots, 1]$ of the Laguerre filter chain. Figure 4 demonstrates the resulting error on training data. Obviously the global maximum is close to $p_L = 0.86$. For each pole, a separate local model network was trained with LOLIMOT.

While the training of each local model network was terminated after the BIC increased two times. The local model network with the lowest BIC was chosen. Since the convergence curve of the BIC depends on the data distribution, poles which are close together can create networks with a different number of local models. So the loss function in Fig. 4 has a non-continuous behavior, which makes it unsuitable for gradient-based optimization techniques. Evolution- or global approaches (e.g. Simulated Annealing, swarm-based optimizations, etc.) may find the true optimum. The main shortcomings of theses approaches are in this context:

1) Many evaluations of the loss functions, to circumvent to get stuck at local optima. This leads to a high computational demand.
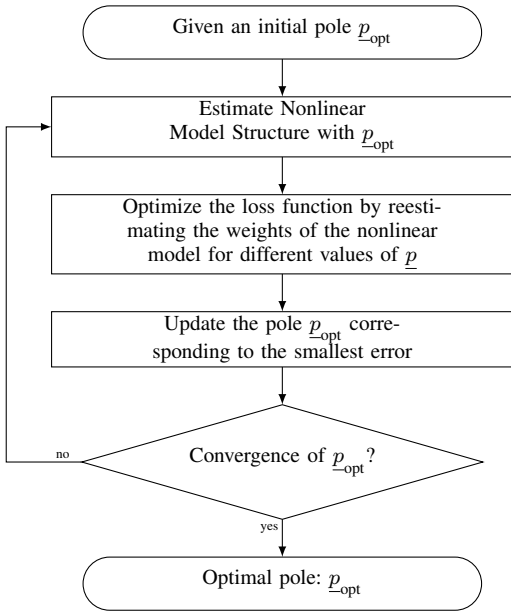
Fig. 5. Flow chart of the repeated fixed structure optimization (FSO)

2) Each evaluation of the loss function requires the time consuming training of a nonlinear model.

C. Fixed Structure Optimization in the Nonlinear Case

To accelerate the optimization procedure in Sect.III-B, the training of the nonlinear model can be approximated following the flow chart in Fig. 5. The FSO algorithm alternates between two steps:

1) Create a model structure (e.g. by LOLIMOT) depending on an initial pole.
2) Find the optimal pole p_L for the fixed structure from 1). Update the pole and repeat these two steps, until p_L converges.

This algorithm is independent of the choice of the nonlinear approximator.

The precondition, to apply the FSO is a pole indepedent gain of the Laguerre Filters. Otherwise the limits of the data distribution in the input space would change with the pole. It is obvious, that the gain of the filter outputs depend on the chosen pole p_L:

$$\lim_{z \to 1}(z-1)\frac{\sqrt{1-p_L^2}}{z-p_L}\left(\frac{1-p_L z}{z-p_L}\right)^{i-1}\frac{z}{z-1} = \frac{\sqrt{1-p_L^2}}{1-p_L} \tag{4}$$

The pole independent gain can be achieved, by normalizing the transfer functions from Eq. 1.

$$K = \frac{p_L}{\sqrt{1-p_L^2}} \tag{5}$$

$$\tilde{U}_{L,i}(z) = KG_{L,i}(z) \cdot U(z) \tag{6}$$

$$\Rightarrow \quad \lim_{z \to 1}(z-1)KG_{L,i}(z)\frac{z}{z-1} = 1 \tag{7}$$

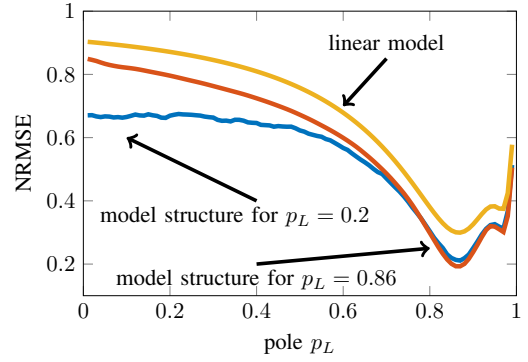Now the gain of each transfer function $KG_{L,i}(z)$ is independent of p_L.

Fig. 6. Loss functions for different FSO and a linear model. The nonlinearity of the process is formed by inverse tangent functions.
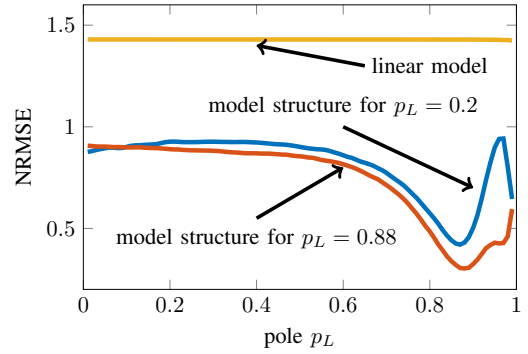

Fig. 7. Loss functions for different FSO and a linear model. The nonlinearity $\mathbf{NL}_2(u)$ is a quadratic function.

IV. EXAMPLES FOR SISO SYSTEMS

For the presented Hammerstein-Wiener system, the optimal pole shall be found using the repeated FSO. Without any knowledge about the process, the initial pole is set to $p_L = 0.2$ to demonstrate a bad initial condition. With the chosen order of the filter chain $n = 2$ a nonlinear model is trained using LOLIMOT. While keeping this partitioning fixed, the loss function can be evaluated for several poles p_L requiring just several least squares estimations to update the local model parameters. The result is the loss function shown in Fig. 6. In the first step, the optimal pole is $p_L = 0.86$. With this pole, the structure of the nonlinear model is updated by a new LOLIMOT training. The corresponding loss function in Fig. 6 demonstrates the convergence of the pole. After reaching the user given accuracy $\Delta p_L \leq 0.01$ the repeated FSO is terminated in the global optimum.

With lacking knowledge about the process a linear model can be used to estimate the dominant pole. For the example in Fig. 6 the loss function of the linear model shows a similar behavior to the loss functions obtained with the FSO.

However, a Hammerstein-Wiener process with a non-monotonic nonlinearity $\mathbf{NL}_2(u) = u^2$, the linear model is not able to represent the process behavior (see Fig. 7). So the initialization of p_L using a linear model might not be a good choice in general.

Since the true loss function in Fig. 4 is not smooth, the approximative loss functions in Fig. 6 are better suited for an gradient-based optimization. Beside this, the evaluation of the approximative loss function is significant faster than the true loss function.

To compare the computational effort of the repeated FSO, to a conventional pole search (CPS), the optimization is done with a grid-based search. With the pole resolution Δp_L, the number of FSO repetitions R, the model training time (MT) and the time for all local models estimation (LS), the computational time $t_c(\cdot)$, the overall computational time can be summarized:

$$t_c(\text{CPS}) = \frac{1}{\Delta p_L} \cdot t_c(\text{MT}) \tag{8}$$

$$t_c(\text{repeated FSO}) = R \cdot t_c(\text{MT}) + \frac{R}{\Delta p_L} \cdot t_c(\text{LS}) \tag{9}$$

The computational effort for the model training is typically much larger than the least squares estimation of the weights. For a LOLIMOT training, the ratio $t_c(\text{MT})/t_c(\text{LS})$typically is around 10100. For less efficient model structures it can be orders of magnitude larger. Here in Fig. 6 we need 2 repetitions of the FSO ($R = 2$) for a resolution of $\Delta p_L = 0.01$. By assuming, that the average time for a model training is 100 times bigger than for the least squares estimation, the relative time saving can be calculated to

$$\frac{t_c(\text{CPS})}{t_c(\text{repeated FSO})} = \frac{\frac{1}{\Delta p_L} \cdot t_c(MT)}{\frac{R}{\Delta p_L} \cdot t_c(LS) + R \cdot t_c(MT)} \tag{10}$$

i.e., for this example, the computational effort of a conventional pole search is 25 times higher than for the repeated FSO. The computation time of one model training depends strongly on the chosen nonlinear model structure and the number of samples N. So the benefit of the repeated FSO is even bigger, if the model training is very time consuming.

V. POLE OPTIMIZATION - MISO

The computational disadvantages, which arise due to the pole optimization in the SISO case, are extended, by increasing the number of dynamic inputs. The computational effort of the grid-based method would increase exponentially with the number of inputs.

To allow many inputs in practice, this problem requires a new method to determine the poles $\underline{p}_L$ of each Laguerre filter chain. Basically, the repeated FSO can be used to speed up the optimization procedure. But the curse of dimensionality will remain, if a grid-based search is conducted for all poles simultaneously.

A. Loss Function for a Linear MISO System

It can be shown that the poles are slightly correlated even in the linear case. This is analyzed for a system with 2 inputs and 1 output. Each input u_i passes through a Laguerre filter chain
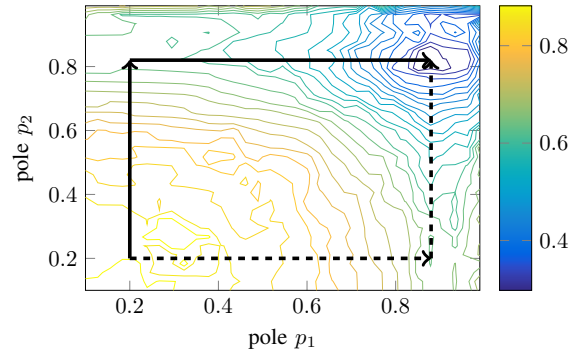

Fig. 8. Contour lines of the true loss function. The arrows show the paths of staggered optimizations.

with its individual pole $p_{L,i}$ generating its output contribution $\hat{y}_i$.

$$\hat{y}(k, p_{L,1}, p_{L,1}) = \hat{y}_1(k, p_{L,1}) + \hat{y}_2(k, p_{L,1}) \tag{11}$$

$$MSE = \frac{1}{N} \sum_{k=1}^{N} (y(k) - \hat{y}(k, p_{L,1}, p_{L,1}))^2 \tag{12}$$

$$MSE = \frac{1}{N} \sum_{k=1}^{N} \Big[y^2(k) + \hat{y}_1^2(k, p_{L,1}) + \hat{y}_2^2(k, p_{L,1}) \\ - 2y(k)\hat{y}_1(k, p_{L,1}) - 2y(k)\hat{y}_2(k, p_{L,1}) \\ + 2\underbrace{\hat{y}_1(k, p_{L,1})\hat{y}_2(k, p_{L,1})}_{\text{interaction between poles}} \Big] \ . \tag{13}$$

The mean squared error (MSE), which should be minimized with respect to the poles $p_{L,1}$ and $p_{L,2}$ depends on the outputs $\underline{\hat{y}}_1$, $\underline{\hat{y}}_2$ and $\underline{y}$. In the case of 2 inputs, only one summand of the loss function contains an interaction between the two poles. By dropping the arguments k and $p_{L,i}$, the MSE can be rewritten as:

$$MSE = \frac{1}{N} \Big(\underline{y}^T \underline{y} + \underline{\hat{y}}_1^T \underline{\hat{y}}_1 + \underline{\hat{y}}_2^T \underline{\hat{y}}_2 \\ - 2\underline{y}^T \underline{\hat{y}}_1 - 2\underline{y}^T \underline{\hat{y}}_2 + 2\underline{\hat{y}}_1^T \underline{\hat{y}}_2 \Big) \tag{14}$$

The value of $\underline{\hat{y}}_1^T \underline{\hat{y}}_2$ depends on the chosen poles, the estimated parameters and of course on the choice of the input signals. It can be shown that the outputs of linear transfer functions are uncorrelated by applying uncorrelated inputs. So the pole optimization for each pole can be done separately.

In the real world, the preconditions of totally uncorrelated inputs can not be fulfilled. By violating this precondition slightly, the separated optimization will lead to wrong results with small errors. As mentioned above, in the linear case, the error is insensitive to p_L for an adequate choice of n. If the quality of the determined pole is not sufficient, the resulting poles are good initial values for a possible subsequent nonlinear optimization.
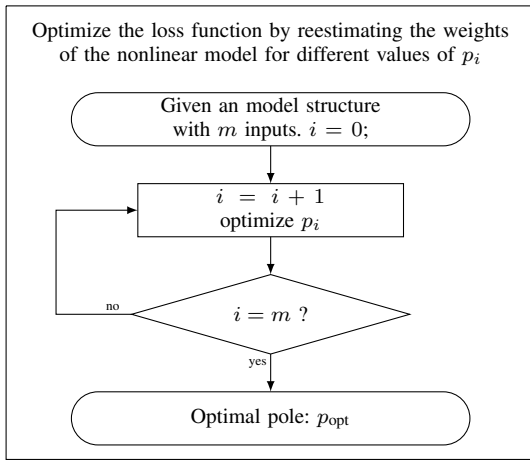
Fig. 9. Staggered optimization of the poles inside the repeated FSO

Fig. 10. Block diagram of Wiener Hammerstein system with two inputs.

Fig. 11. FSO with $\underline{p}_L = [0.2\,0.2]$ and an optimum value at $\underline{p}_L = [0.85\,0.82]$. The arrows show the paths of staggered optimizations.

B. Loss Function for a Nonlinear MISO System

Usually all poles are optimized simultaneously. Then the only alternatives are

1) local (gradient-based) search which runs into problems with the wiggly loss function and multiple local optima
2) global grid-based search which fully underlies the curse of dimensionality

For nearly uncorrelated inputs it is observed, that the interaction between the Laguerre poles are small. With this information, a staggered optimization[1] can be used to overcome it. The idea behind staggered optimization is to only optimize one pole, while all remaining ones are kept fixed. After convergence of the 1-dimensional optimization, it continues with the next pole. This is illustrated by Fig. 8, which shows the influence of the poles on the NRMSE of the training data.

The main directions of the contour lines are orthogonal to the pole-axes. This reveals the weak correlation between the poles. A staggered optimization will find the global optimum within 1 or 2 iterations.

The previously introduced repeated FSO can be extended to the multivariate case. The optimization block in Fig. 5 has to be replaced by the block in Fig. 9.

VI. EXAMPLE FOR A MISO SYSTEM

As an example the nonlinear process is considered as the superposition of two Wiener-Hammerstein systems as shown in Fig. 10. The true loss function is depicted in Fig. 8.

Similar to the SISO example, the initial poles for repeated FSO are set to $p_L = [0.2\,0.2]$. Figure 11 demonstrates the loss function for the fixed nonlinear model structure. By optimizing the pole successively in $p_{L,1}$-direction, and $p_{L,2}$-direction, the result is close to the global optimum after the first iteration.

By updating the nonlinear model structure with $\underline{p}_L = [0.85\,0.82]$, the loss function in Fig. 12 is obtained. The

---

[1]In [12] it is mentioned, that the staggered optimization is a synonym for: successive variation of the variables, relaxation, parallel axis search, univariate or uni-variant search, one-variable-at-a-time method, axial iteration technique, cyclic coordinate ascent method, alternating variable search
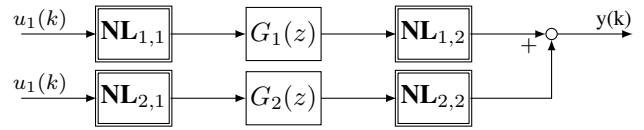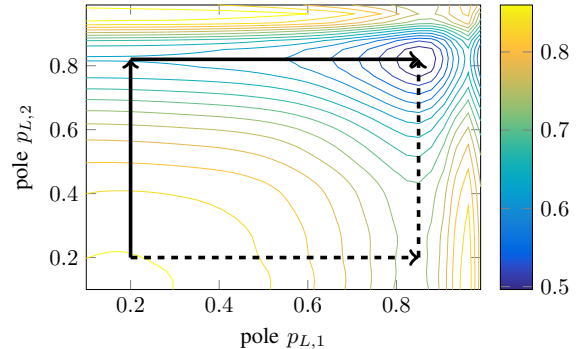
optimization terminates after reaching the chosen accuracy $\Delta p_L \leq 0.01$ at $\underline{p}_L = [0.88\,0.82]$.

Due to the smoother loss function in Fig. 11 the staggered optimization is much more robust w.r.t. finding the global optimum than in the original loss function case shown in Fig. 8.

VII. CONCLUSION AND OUTLOOK

In this paper we introduced the repeated fixed structure optimization (FSO). This innovative method fixes the structure of the nonlinear approximators and optimizes the pole with this structure. By repeating the FSO, we exemplified the fast convergence for one input. This method is completely independent of the nonlinear approximator.

Because of the low correlation between the poles, the optimization inside the FSO can be replaced by a staggered optimization. With this approach the complexity of the optimization scales roughly only linearly with the number of
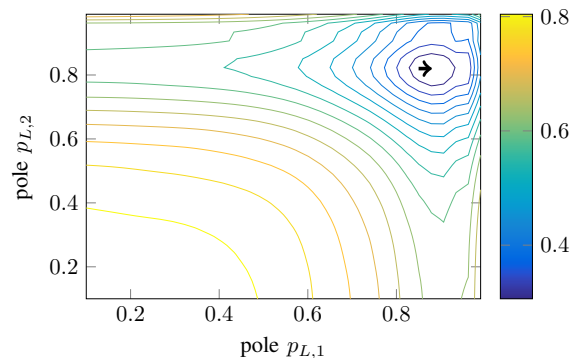

Fig. 12. FSO with $\underline{p}_L = [0.85\,0.82]$ and an optimum value at $\underline{p}_L = [0.88\,0.82]$. The arrow show the only path of the staggered optimization.

inputs. For an example system with two inputs the poles converged after three iteration FSO.

With the two proposed methods, the computational effort decreases dramatically especially for the MISO case.

In the future, the repeated FSO can be accelerated further by the following two procedures:

1) Premature termination:

   In early phases of the pole optimization the model training can be carried out very roughly (e.g. few local models or early stopping) because the model dynamics is only very approximate.

2) Adaptive pole resolution:

   Starting from a coarse discretization, the pole resolution can be increased in latter iterations of the staggered FSO search.

REFERENCES

[1] B. Wahlberg, "System identification using laguerre models," *Automatic Control, IEEE Transactions on*, vol. 36, no. 5, pp. 551–562, 1991.

[2] G. H. Oliveira, A. da Rosa, R. J. Campello, J. B. Machado, and W. C. Amaral, "An introduction to models based on laguerre, kautz and other related orthonormal functions–part i: linear and uncertain models," *International Journal of Modelling, Identification and Control*, vol. 14, no. 1-2, pp. 121–132, 2011.

[3] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.

[4] O. Nelles, *Nonlinear system identification*. Springer Science & Business Media, 2002.

[5] B. Wahlberg, "System identification using kautz models," *Automatic Control, IEEE Transactions on*, vol. 39, no. 6, pp. 1276–1282, 1994.

[6] K. Tanaka, "Stability and stabilizability of fuzzy-neural-linear control systems," *Fuzzy Systems, IEEE Transactions on*, vol. 3, no. 4, pp. 438–447, 1995.

[7] C. Boukis, D. P. Mandic, A. G. Constantinides, and L. C. Polymenakos, "A novel algorithm for the adaptation of the pole of laguerre filters," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 429–432, 2006.

[8] N. Tanguy, P. Vilbé, and L.-C. Calvez, "Optimum choice of free parameter in orthonormal approximations," *IEEE Transactions on Automatic Control*, vol. 40, no. 10, pp. 1811–1813, 1995.

[9] H. Belt and A. den Brinker, "Laguerre filters with adaptive pole optimization," in *Circuits and Systems, 1996. ISCAS'96., Connecting the World., 1996 IEEE International Symposium on*, vol. 2. IEEE, 1996, pp. 37–40.

[10] O. Nelles and R. Isermann, "Basis function networks for interpolation of local linear models," in *Decision and Control, 1996., Proceedings of the 35th IEEE Conference on*, vol. 1. IEEE, 1996, pp. 470–475.

[11] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2002.

[12] H.-P. P. Schwefel, *Evolution and optimum seeking: the sixth generation*. John Wiley & Sons, Inc., 1993.