GRAY-BOX IDENTIFICATION USING POLYNOMIAL NARMAX MODELS

Allyne M. dos Santos[*,1], Argimiro R. Secchi[2], Maurício B. de Souza Jr[2], Sigurd Skogestad[1] and Dinesh Krishnamoorthy[1]
[1] Department of Chemical Engineering, Norwegian University of Science and Technology - Trondheim
7491 Trondheim, Norway
[2] PEQ/COPPE/UFRJ - Chemical Engineering Program - Rio de Janeiro
RJ 21.941-914, Brazil

Abstract Overview

It is common to approximate nonlinear models with a collection of linear models, but sometimes it can cause problems, such as lack of representativeness of the nonlinear behavior and generation of nonexistent peaks. Nonlinear identification has been improved to overcome these problems. The type of nonlinear model used in this work is the polynomial Nonlinear AutoRegressive Moving Average models with eXogenous inputs (NARMAX). In the present work, a gray-box identification is compared with the black-box one for optimization and control purposes. The identification was performed using an orthogonal least square algorithm and validation was made using k-step-ahead cross-validation method. Dynamic real-time optimization was set based on both first-principle models and identified models, and compared, to evaluate the performance of identified nonlinear models. The gray-box model was more representative in relation to the nonlinearity of the system and generated closer solutions in the dynamic real-time optimization, when comparing with the solutions based on first-principle model.

Keywords

NARMAX models, Gray-box, System Identification.

Introduction

Nonlinear models have received more attention during the past decades as the computational capacity of solving complex problems has increased. The Nonlinear AutoRegressive Moving Average models with eXogenous inputs (NARMAX) is a parametric model, which uses measured discrete data set from industrial or simulated plant, typically as a black-box model. However, it can also be used with some system information, such as static gain, number of stationary states on the output variables, mass balance, energy balance or qualitative characteristics with respect to the dynamic behavior of the system (as a gray-box model).

The improvement of using some type of system information motivated many researchers to apply system identification using NARMAX models with different types of estimation algorithm (Johansen, 1996). The parameter estimation algorithm used in this work is the Orthogonal Least Square algorithm developed by Chen et al. (1989).

In this work, qualitative characteristics of the case study are available. This type is not much explored because it depends on the quality of the available qualitative characteristics, which usually comes from system manipulation by the user. When high quality information is acquired, gray-box identification improves the capacity of

[*] To whom all correspondence should be addressed

mathematical models to represent the dynamic behavior of nonlinear systems.

The main objective of this work is to develop a methodology for identifying nonlinear systems with prior knowledge using polynomial NARMAX models. The proposed methodology is then compared with a typical black-box approach and applied on dynamic real-time optimization.

Proposed Methodology

In order to acquire nonlinear data, each input variable was perturbed in a way that each excitation lasts enough time to reach a stable response in the its end and with the range of the operating interval (as widely as possible). Measurement uncertainty was added in the form of white noise to all input and output data. Normalization was performed taking the minimum of each variable as reference.

The general structure of NARMAX models is given by,

$$y(k) = P^{\ell}[y(k-1), \dots, y(k-n_y), u_i(k-d), \dots, u_i(k-d-n_{u_i}), e(k-1), \dots, e(k-n_e)] + \varepsilon(k) \quad (1)$$

where P is a nonlinear function with nonlinearity degree ℓ in relation to all variables, with $k = 1, \dots, N$. $y(k)$, $u_i(k)$, $e(k)$ and $\varepsilon(k)$ are the output variable, input variables, noise and prediction error at instant k, respectively. n_y, n_{u_i} and n_e are the maximum lags of system output, inputs, noise, respectively, d is the time delay of the model. The polynomial type is given by,

$$y(k) = \sum_{m_1=1}^{n} \theta_{m_1}\psi_{km_1} + \sum_{m_1=1}^{n}\sum_{m_2=m_1}^{n} \theta_{m_1 m_2}\psi_{km_1}\psi_{km_2} + \sum_{m_1=1}^{n} \dots \sum_{m_\ell=m_{\ell-1}}^{n} \theta_{m_1 \dots m_\ell}\psi_{km_1}\dots\psi_{km_\ell} + \varepsilon(k) \quad (2)$$

$$n = n_y + n_e + \sum_{j=1}^{i} n_{u_j} \quad (3)$$

where i is the number of input variables, ψ_{km_ℓ} is a regressor with nonlinearity degree ℓ at instant k.

The parameter estimation was performed using an orthogonal least square algorithm named Golub-Householder with Error Reduction Rate (ERR) (Aguirre, 2000). The problem was divided into two parts: NARX part; and MA part (where noise modeling is made). Each part used the OLS algorithm.

The identification algorithm used in this work can be summarized as shown in Figure 1, where J is the objective function of the OLS algorithm. n_{pNARX} and n_{pMA} are the number of terms of the NARX and MA part of the model, respectively.

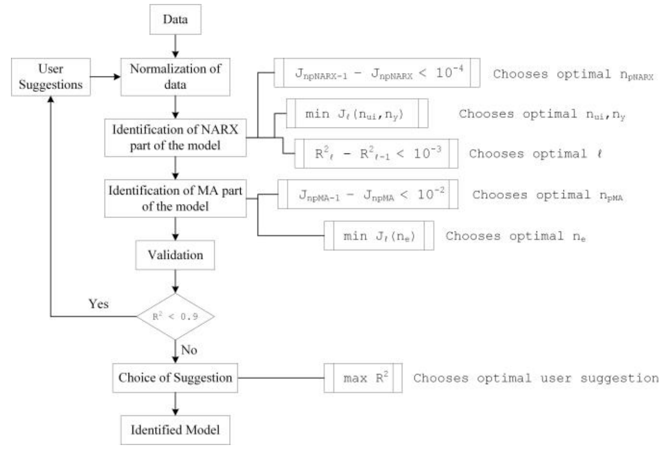

Figure 1. Block diagram of the gray-box identification algorithm

The gray-box identification can be divided into seven main steps:

1. The user starts by suggesting some change on the coordinates or not.
2. The user specifies a range of values for the model orders (n_y, n_{u_i}, n_e and d).
3. Synthetic data is acquired from first-principle models.
4. Normalization is done.
5. Off-line identification algorithm chooses regressors, estimates parameters and optimizes the number of features for each type of model (varying the nonlinearity degree, ℓ) and for each suggestion on coordinate change.
6. Validation step compares the R-squared value of all types of models and chooses the model that has the highest one.
7. The suggestion on coordinate change is also chosen comparing the R-squared values with the one of the black-box identification.

Application Example

The identification was performed on an oil production system with two gas-lift wells (Krishnamoorthy et al., 2018). First, a black-box model was developed according to Eq. (4) and (5).

$$\boldsymbol{y}_{p_{bh_2}} = \boldsymbol{\Psi}^{*}_{p_{bh_2}} \boldsymbol{\theta}_{p_{bh_2}} \quad (4)$$

$$
\Psi^{*T}_{p_{bh_2}} = \begin{bmatrix} y_{p_{bh_2}}(k-1) \\ y_{p_{bh_2}}(k-2) \\ y_{p_{bh_2}}(k-1)u_2(k-1) \\ y_{p_{bh_2}}(k-1)^2 \\ y_{p_{bh_2}}(k-1)u_2(k-2) \\ y_{p_{bh_2}}(k-2)u_2(k-1) \\ u_2(k-4)^2 \\ y_{p_{bh_2}}(k-1)u_2(k-4) \\ y_{p_{bh_2}}(k-2)u_2(k-3) \\ y_{p_{bh_2}}(k-2)^2 \\ u_1(k-1)u_2(k-4) \\ y_{p_{bh_2}}(k-1)u_1(k-3) \\ y_{p_{bh_2}}(k-1)u_1(k-1) \\ y_{p_{bh_2}}(k-2)u_2(k-4) \\ e(k-1)u_2(k-2) \end{bmatrix} \quad \theta_{p_{bh_2}} = \begin{bmatrix} 1.8734 \\ -0.6051 \\ -0.0395 \\ -0.8908 \\ -0.2835 \\ 0.4207 \\ 0.0399 \\ -0.3133 \\ -0.1858 \\ 0.5294 \\ 0.0177 \\ -0.0498 \\ 0.0276 \\ 0.1388 \\ 0.026872 \end{bmatrix} \tag{5}
$$

The validation of the black-box model gave a R-squared value of 0.5038.

The change in the coordinates that gave higher improvement was substituting u_1 for u_1/u_2 and u_2 for u_2/u_1. Using this information, a gray-box model was developed according to Eq. (4) and (6). The validation of the gray-box model gave an R-squared value of 0.8867, which is higher than the one from black-box model. This is because the algorithm chose a NARX model instead of a NARMAX one.

$$
\Psi^{*T}_{p_{bh_2}} = \begin{bmatrix} y_{p_{bh_2}}(k-1) \\ y_{p_{bh_2}}(k-2) \\ u_1(k-1) \\ u_2(k-4) \\ u_2(k-2) \\ u_1(k-4) \\ u_1(k-2) \\ u_2(k-3) \\ y_{p_{bh_2}}(k-3) \\ u_1(k-3) \end{bmatrix} \quad \theta_{p_{bh_2}} = \begin{bmatrix} 1.4030 \\ -0.4210 \\ 0.0325 \\ 0.0063 \\ -0.0053 \\ -0.0422 \\ 0.0458 \\ 0.0104 \\ -0.0194 \\ -0.0014 \end{bmatrix} \tag{6}
$$

Dynamic Real-time Optimization

The closed loop responses of the NARMAX models to the DRTO actions on the input variables were compared with the one using the first-principle model, the result is shown in Figure 2.

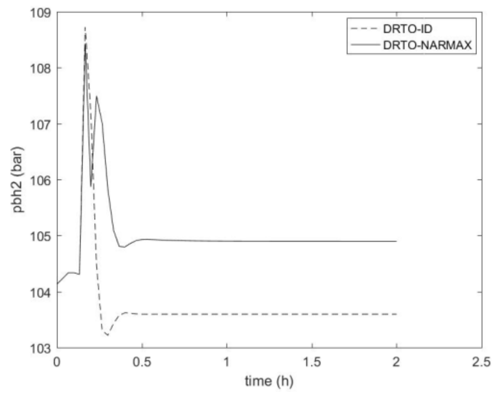

Figure 2. Comparison of DRTO performances (first-principle model and black-box model)

The difference between the solutions was expected because the NARMAX model accounts with noise, which is different (none in this case) than the noise in the original data set. When using gray-box models, the difference between the solutions reduced significantly, reassuring the improvement noted with the system identification results.

Conclusion

The gray-box identification algorithm used in this work has low complexity, although it can take a lot of time depending on the knowledge of the user about the process (if one makes a lot of suggestions to the change in coordinates). The gray-box method improved the modeling of the system, generating higher R-squared value, which was illustrated with an application in dynamic real-time optimization.

Acknowledgments

The main author gratefully acknowledge financial support from the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ); the International Partnerships for Excellence Education, Research and Innovation (INTPART); and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

Aguirre, L. A. (2000). *Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais*, Belo Horizonte, UFMG.

Chen, S., Billings, S. A., Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification, *Int. J. Control*, 50, 5, 1873.

Krishnamoorthy, D., Foss, B., Skogestad, S. (2018). Steady-state real-time optimization using transient measurements. *Comput. Chem. Eng.*, 115, 34.

Johansen, T. A. (1996). Identification of non-linear systems using empirical data and prior knowledge – an optimization approach, *Automatica*, 32, 3, 337-356.