

Estimates of an Upper Limit of the Number of Parameters in Nonlinear Model Structures [★]

Jonas Sjöberg* Johan Schoukens**

* *Department of Signals & Systems, Chalmers University of Technology, SE412 96 Gothenburg, Sweden (e-mail: jonas.sjoberg@chalmers.se)*

** *Department of Fundamental Electricity and Instrumentation, Faculty of Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium (e-mail: Johan.Schoukens@vub.ac.be)*

Abstract: An approach to estimate the plausible number of parameters in a nonlinear model in an identification problem is considered. By suggesting a series of experiments using periodic input signal, estimates of the disturbance variance and the nonlinear distortion can be formulated. With these estimates an expression for a reasonable number of parameters is obtained. This is useful help when a user has to choose between different types of nonlinear model structure which differ largely in their number of parameters.

Keywords: Nonlinear systems, System Identification, Parameterization, Dynamic modelling, Input signals

1. INTRODUCTION

There are many possible approaches to nonlinear system identification, each approach with its advantages and disadvantages. For many of the possible alternatives there is a range of the number of parameters in the associated model structure for them to be applicable. The typical approach would be to test the different types of model structures and by trial and error select the most appropriate one.

The goal in this contribution is to obtain an estimate of how many parameters can be estimated to identify the nonlinear part of the system *prior* to any testing of nonlinear model structures. Given that estimate it is possible to suggest type of model structures which are applicable. This can make the choice of model structure easier and faster for the user.

The idea requires that a series of experiments with periodic input signal is conducted on the plant. With the recorded outputs from these experiments, separate estimates of the disturbance, the linear contribution, and the nonlinear distortion can be made. With these estimates, and asymptotic results on the variance contribution to the misfit, the applicable parameter number can be obtained. These ideas build on insights of using periodic inputs covered in, eg, Pintelon and Schoukens (2001); D'haene et al. (2005) and linear approximations of nonlinear systems, see eg, Schoukens et al. (2005); Enqvist (2005); Ljung (2001); Mäkilä and Partington (2004). The general asymptotic theory used is covered in standard textbooks like Ljung (1999).

With an indication of the applicable number of parameters before the actual nonlinear system identification starts, the user can direct their interest to the model structures which typically are used with that number of parameters.

There are, of course no absolute truth which model structure applies given a certain number of parameters, but a rough classification can be made. With *system order* of η , a *memory length* of κ , which roughly corresponds to five times the time constant for a linear system, and, a somewhat fuzzy measure, *nonlinear degree* δ which for polynomial nonlinearities corresponds to the degree, the following model structures with estimates of the order of the number of parameters can be formulated

Block structured models, $O(\kappa^0) + O(\eta)$,
Neural network based models, $O(\eta\delta)$,
Polynomial nonlinear state space models, $O(\eta^\delta)$,
Support vector machines (SVM), $O(\eta^\delta)$,
Non-parametric Volterra series, $O(\kappa^\delta)$.

Block structures are described in Billings and Fakhouri (1982); Westwick and Kearney (IEEE Press). Theory of Volterra series for nonlinear system identification is covered in Schetzen (2006). A general overview of identification of nonlinear systems can be found in Sjöberg et al. (1995), where also neural networks and similar model structures are explained.

The rest of the paper is outlined as follows. Section 2 gives the assumption on the system to be identified and Section 3 explains how the number of possible parameter is estimated given measures of the nonlinear distortion and the noise variance. Then, Section 4 outlines the necessary experiments with a periodic input signal and Section 5 then gives the estimates of the nonlinear distortion and the

* Sponsor and financial support acknowledgment goes here.

noise variance. The method is illustrated with an example in Section 6 and the paper is concluded in Section 7.

2. PLANT ASSUMPTIONS

The process to be identified is assumed to be a discrete time deterministic nonlinear dynamic system, described by

$$y(t) = f(u^t) + e(t) \quad (1)$$

where f is the nonlinear dynamic function, u^t consists lagged input signals $u(t)$ up to time t , and $e(t)$ is the disturbance with variance σ^2 . The input is assumed to be known exactly and the noise variance is unknown. Often it is more convenient to work with data in the frequency domain, especially since periodic inputs are used. All signals can as well be described in the frequency domain, and capital letters will then be used. The system dynamic, described by f , is however more conveniently described in the time domain since it is nonlinear.

3. ESTIMATING THE PLAUSIBLE NUMBER OF PARAMETERS IN NONLINEAR MODELS

The contributions to the misfit of a model can be described as a sum of a bias part B , or, equivalently, as systematic part, and a variance part. The bias part depends on the chosen model structure and describes the distance between the best possible model and the true function, f given statistical properties on the input signal. The variance part is stochastic and depends on the disturbance in the measurements. Hence, if the identification is repeated with the same model structure but with new data with the same statistical properties as the first data set, the bias part remains the same but realization of the variance part of the misfit will be different. However, with a quadratic measure, in mean the variance contribution will be

$$\lambda \frac{d}{N} \quad (2)$$

where d is the number of parameters in the model, λ is the variance of the disturbance, and N is the number of data used in the identification. See, eg, Ljung (1999).

As will be described in the next section, if linear models are considered to model the unknown system f , the nonlinear distortion will act as an additional stochastic disturbance with variance σ_{NL}^2 . Hence, if linear models are considered, $\lambda = \sigma^2 + \sigma_{NL}^2$ in (2). For a linear model structure, which can describe the Best possible Linear Approximation (BLA), one obtains the following expression for the expected misfit

$$E[(y(t) - \hat{y}(t))^2] = B + d \frac{\sigma^2 + \sigma_{NL}^2}{N}. \quad (3)$$

where σ_{NL}^2 is the bias which is equivalent to the nonlinear distortion.

Consider now the case that a nonlinear model which is able to describe f for some parameter values, ie, if there is no bias error. Then the expected variance of an estimated nonlinear model becomes contribution becomes

$$E[(y(t) - \hat{y}_{NL}(t))^2] = d_{NL} \frac{\sigma^2}{N} \quad (4)$$

where d_{NL} are the number of parameters in the nonlinear model.

Now, the prediction with the nonlinear model will be better than with the linear model if (4) is smaller than (3), ie, if

$$d_{NL} < N \frac{B}{\sigma^2} + d + d \frac{\sigma_{NL}^2}{\sigma^2}. \quad (5)$$

Typically, the number of parameters in a linear model is much smaller than in a nonlinear model so in many cases, for reasonable values on N , the second term in (5) can be neglected. So, to use (5) to obtain an estimate of the number of parameters one needs estimates of, B , σ_{NL}^2 and σ^2 . This is described in the following.

4. EXPERIMENT OUTLINE

A series of experiments with periodic excitation need to be performed following the setup described in, eg, Pintelon and Schoukens (2001). The description is included here for consistency.

The input signal is chosen as a multi-sine signal where the phase is chosen randomly uniformly distributed at each of the experiments. Let P periods of data recorded in each experiment, assuming measurements starting after that transients have settled, and let M indicate the number of experiments. Let $u(t)^{[m,p]}$ indicate the input signal at experiment m and period p and let each period have N_p data sample. Let $y(t)^{[m,p]}$ be the corresponding notation for the output signal.

Now, according to Pintelon and Schoukens (2001), the output can be described as

$$y(t)^{[m,p]} = y_{BLA}(t)^{[m,p]} + y_{NL}(t)^{[m,p]} + e(t)^{[m,p]} \quad (6)$$

where $y_{BLA}(t)^{[m,p]}$ is the output of the best linear approximation of the unknown plant, $y_{NL}(t)^{[m,p]}$ is the nonlinear distortion of the plant, and $e(t)^{[m,p]}$ is the disturbance at experiment m and period p . Notice that the first two terms in (6) equals the nonlinear plant, ie,

$$f(u^t) = y_{BLA}(t) + y_{NL}(t) \quad (7)$$

In Pintelon and Schoukens (2001) it is, further explained that $y_{BLA}(t)$ is described by the transfer function of the best linear approximation G_{BLA} ,

$$y_{BLA}(t)^{[m,p]} = G_{BLA} u(t)^{[m,p]} \quad (8)$$

and since the input is periodic, $u(t)^{[m,p_1]} = u(t)^{[m,p_2]}$, the output will be p -independent (with exception for the disturbance $e(t)$) and we can write $y_{BLA}(t)^{[m]}$ and $y_{NL}(t)^{[m]}$. Due to the random phase of the input signal, the nonlinear distortion will act as an additional stochastic disturbance term. Hence

$$E[Y_{NL}(\omega)^{[m_1]} Y_{NL}(\omega)^{[m_2]}] = \begin{cases} \sigma_{NL}^2(\omega) & \text{if } m_1 = m_2 \\ 0 & \text{else.} \end{cases} \quad (9)$$

So, if a linear model is used to model the data, the disturbance will be $y_{NL}(t)^{[m,p]} + e(t)^{[m,p]}$ in (6), with the variance $\sigma^2 + \sigma_{NL}^2$.

We can then formulate an expression for the bias as

$$\begin{aligned} B &= E[(f(u^t) - y_{BLA}(t))^2] = \\ &= E[(y_{BLA}(t) + y_{NL}(t) - y_{BLA}(t))^2] \\ &= E[y_{NL}^2(t)] = \sigma_{NL}^2. \end{aligned} \quad (10)$$

The expectation is under the condition that the stochastic properties of the input signal are the same as that defining

$y_{BLA}(t)$. If the stochastic properties are allowed to change, then the bias will in general be larger.

5. ESTIMATION OF NOISE AND NONLINEAR DISTORTION

With the experimental setting and the notations from previous section, it easy to formulate estimates of σ^2 and σ_{NL}^2 , they also follow the insights given in Pintelon and Schoukens (2001).

An estimate of σ^2 is obtained in the following way. First, $y_{BLA}(t)^{[m,p]}$ and $y_{NL}(t)^{[m,p]}$ can be removed from the output by using one period in each experiment. That is,

$$\begin{aligned} y(t)^{[m,p]} - y(t)^{[m,P]} \\ = e(t)^{[m,p]} - e(t)^{[m,P]} \quad p \neq P. \end{aligned} \quad (11)$$

Taking the square and expectation gives

$$\begin{aligned} \mathbb{E}[(y(t)^{[m,p]} - y(t)^{[m,P]})^2] \\ = \mathbb{E}[(e(t)^{[m,p]} - e(t)^{[m,P]})^2] = 2\sigma^2(t). \end{aligned} \quad (12)$$

An estimate can then be formulated as

$$\hat{\sigma}^2(t) = \frac{1}{M(P-1)} \sum_{m=1}^M \sum_{p=1}^{P-1} (y(t)^{[m,p]} - y(t)^{[m,P]})^2. \quad (13)$$

A common assumption would be that the disturbance is time-invariant and then one can include a mean value over the length of one period in the experiment, N_p , in (13). The estimate (13) can also be expressed in the frequency domain and it then becomes

$$\hat{\sigma}^2(\omega) = \frac{1}{M(P-1)} \sum_{m=1}^M \sum_{p=1}^{P-1} (Y(\omega)^{[m,p]} - Y(\omega)^{[m,P]})^2. \quad (14)$$

An estimate of σ_{NL}^2 is obtained by proceeding similar as above. To remove the BLA part between different experiments one must first compensate for the phase in the different experiments. This is most easily done by multiplying the correspondence of (6) in the frequency domain with the phase opposite to the one of the input signal, ie,

$$\begin{aligned} Y(\omega)^{[m,p]} \frac{\bar{U}(\omega)^{[m]}}{|U(\omega)|} = G_{BLA} |U(\omega)^{[m]}| \\ + Y_{NL}(\omega)^{[m,p]} \frac{\bar{U}(\omega)^{[m]}}{|U(\omega)|} + E(\omega)^{[m,p]} \frac{\bar{U}(\omega)^{[m]}}{|U(\omega)|}. \end{aligned} \quad (15)$$

Now the BLA part can be removed by subtracting (15) with a particular choice of m from the other. That is,

$$\begin{aligned} (Y(\omega)^{[m,p]} - Y(\omega)^{[M,p]}) \frac{\bar{U}(\omega)^{[m]}}{|U(\omega)|} = \\ (Y_{NL}(\omega)^{[m,p]} - Y_{NL}(\omega)^{[M,p]}) \frac{\bar{U}(\omega)^{[m]}}{|U(\omega)|} \\ + (E(\omega)^{[m,p]} - E(\omega)^{[M,p]}) \frac{\bar{U}(\omega)^{[m]}}{|U(\omega)|} \quad m \neq M. \end{aligned} \quad (16)$$

Taking the square and expectation gives

$$\begin{aligned} \mathbb{E}[(Y(\omega)^{[m,p]} - Y(\omega)^{[M,p]})^2] \\ = \mathbb{E}[(Y_{NL}(\omega)^{[m,p]} - Y_{NL}(\omega)^{[M,p]} + E(\omega)^{[m,p]} - E(\omega)^{[M,p]})^2] \\ = 2\sigma_{NL}^2(\omega) + 2\sigma^2(\omega). \end{aligned} \quad (17)$$

The estimate of the nonlinear distortion can now be formulated as

$$\begin{aligned} \hat{\sigma}_{NL}^2(\omega) \\ = \frac{1}{2(M-1)P} \sum_{m=1}^{M-1} \sum_{p=1}^P (Y(\omega)^{[m,p]} - Y(\omega)^{[M,p]})^2 - \hat{\sigma}^2(\omega) \end{aligned} \quad (18)$$

in the frequency domain.

The estimates (14) and (18) are given in the frequency domain. These expressions can, for example, reveal if the disturbances and the nonlinearities are dominating different frequency regions, which indicates that filtering might be interesting to apply.

The estimates of σ^2 and σ_{NL}^2 , to be used in (5), are obtained by taking the mean value over all frequencies in (18) and (14). These estimates can also be obtained by using time domain data and assuming time-invariance as

$$\hat{\sigma}^2 = \frac{1}{N_p M (P-1)} \sum_{t=1}^{N_p} \sum_{p=1}^{M, P-1} (y(t)^{[m,p]} - y(t)^{[m,P]})^2 \quad (19)$$

and

$$\begin{aligned} \hat{\sigma}_{NL}^2 = \frac{1}{2N_p(M-1)P} \cdot \\ \sum_{t=1}^{N_p} \sum_{p=1}^{M-1, P} (y(t)^{[m,p]} - y(t)^{[M,p]})^2 - \hat{\sigma}^2. \end{aligned} \quad (20)$$

In the next section the proposed method will be used on some small examples.

6. EXAMPLES

Consider a simple example where the nonlinear system is of Wiener type

$$y(t) = f(Gu(t)) + e(t) \quad (21)$$

where f is a static nonlinear function, $e(t)$ is white noise with variance 1 and G is transfer function

$$G(z) = \frac{z^2}{z^2 - 1.2z + 0.52}.$$

Two different choices of f will be illustrated, namely

$$\begin{aligned} f_1(x) &= x + \left(\frac{x}{150}\right)^3 \\ f_2(x) &= x + \left(\frac{x}{90}\right)^3 \end{aligned}$$

The experiment parameters were chosen as follows: $M = 5$, $P = 2$, $N = 128$. The input signal was chosen to contain 20 frequencies. An example of the input signal in the time domain from one experiment is shown in Figure 1 a). In Figure 1 b) the Bode plot of the linear plant is shown together with the frequency excitation of the input signal.

f_1 : The nonlinearity has very little influence since the output from the linear part is divided by 150 before the cubic nonlinearity get into action. In Figure 2 a) and

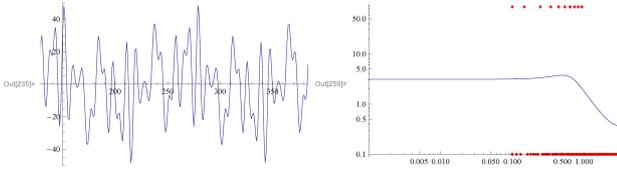


Fig. 1. a) The input signal from one of the 5 experiments. b) Bode plot of the linear part of the plant together with the excited frequencies.

b) the output from the linear part of the system and the plant output from one of the experiments are shown, respectively. No trace of any nonlinear distortion of the signal can be seen. No nonlinearity can be detected in this

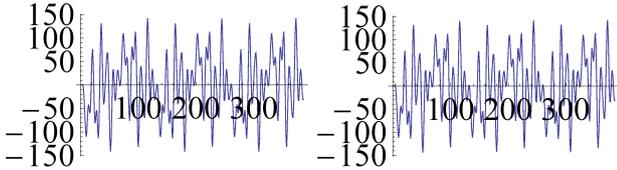


Fig. 2. From one of the experiments with f_1 : a) Output from the linear part of the system. b) System output.

case and the estimate of the nonlinear distortion becomes negative, $\hat{\sigma}_{NL}^2 = -0.32$. The noise estimate is however reasonable, $\hat{\sigma}^2 = 0.93$.

$$f(x) = x + \left(\frac{x}{90}\right)^3$$

f_1 : In the second case the nonlinearity is slightly stronger. However, any nonlinear distortion is hardly visible if one compare the output from the linear part and the system output, see Figure 3 a) and b). However, the estimate of

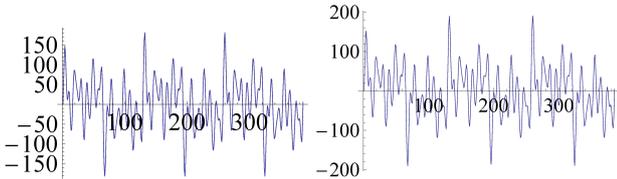


Fig. 3. From one of the experiments with f_2 : a) Output from the linear part of the system. b) System output.

the nonlinear distortion becomes $\hat{\sigma}_{NL}^2 = 0.36$ and the estimate of the noise $\hat{\sigma}^2 = 0.97$. Then an estimate of a plausible number of parameters can be calculated as $d_{NL} = 478$.

7. CONCLUSIONS

An algorithm to estimate the plausible number of parameters in a nonlinear model has been presented. It requires data from a series of experiments so that estimates of the noise level and the nonlinear distortion can be made, on which the estimate relies.

REFERENCES

Billings, S. and Fakhouri, S. (1982). Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1), 15–26.

D'haene, T., Pintelon, R., Schoukens, J., and Van Gheem, E. (2005). Variance analysis of frequency response function measurements using periodic excitations. *IEEE Transactions on Instrumentation and Measurement*, 54(4), 1452–1456.

Enqvist, M. (2005). *Linear Models of Nonlinear systems*. Ph.D. thesis, Linköping University.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition.

Ljung, L. (2001). Estimating linear time-invariant models of nonlinear time-varying systems. *European Journal of Control*, 7(2-3), 203–219.

Mäkilä, P. and Partington, J. (2004). Least-squares lti approximation of nonlinear systems and quasistationarity analysis. *Automatica*, 40(7), 1157–1169.

Pintelon, R. and Schoukens, J. (2001). *System Identification: A Frequency Domain Approach*. IEEE-press, Piscataway.

Schetzen, M. (2006). *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley and Sons.

Schoukens, J., Pintelon, R., Dobrowiecki, T., and Rolain, Y. (2005). Identification of linear systems with nonlinear distortions. *Automatica*, 41, 451–504.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B., Glorennec, P.Y., Hjalmarsson, H., and Juditsky, A. (1995). Non-linear black-box modeling in system identification: a unified overview. *Automatica*, 31(12), 1691–1724.

Westwick, D. and Kearney, R. (IEEE Press). *Identification of nonlinear physiological systems*. IEEE Press.