

# Quantifying Model Errors Caused by Nonlinear Undermodeling in Linear System Identification

A. E. Nordström\*, B. M. Ninness† and T. Wigren‡

Submitted to 13th World Congress of IFAC, June 30–July 5, 1996, San Francisco,  
California, U.S.A.

## Abstract

The problem of quantifying errors due to nonlinear undermodeling is addressed. It is assumed that the system consists of a linear dynamic block in cascade with a static nonlinearity and that the objective is to identify the linear part using a purely linear model. The stochastic embedding approach is applied to capture the on-average properties of the undermodeling. As compared to previous methods, the priors on the covariance matrix of the embedding parameters are reduced. As a result an expression for the amplitude error bounds of the estimated transfer function, that does not require knowledge of the true system parameters, is obtained. The quality of this measure depends on the degree of accuracy with which the unknown nonlinearity can be represented using a set of known basis functions. The proposed method simultaneously delivers error bounds on the estimated transfer function and an indirect estimate of the size of the nonlinearity. The importance of obtaining error bounds on transfer functions and estimates of static nonlinearities for controller design is well established in the literature.

*Keywords:* Nonlinear undermodeling, Stochastic embedding, Error bounds.

---

\*Department of Signals, Sensors and Systems, Royal Institute of Technology, S-100 44 Stockholm, Sweden. Phone: +46 8 764 13 11, Fax: +46 8 790 7329 E-mail: anders.nordstrom@era-t.ericsson.se

†Department of Electrical Engineering, Centre for Control Science, University of Newcastle, N.S.W., 2308, Australia.

‡Systems and Control Group, Uppsala University, P.O. Box 27, S-751 03 Uppsala, Sweden. Phone: +46 18 183074, Fax: +46 18 503611.

# 1 INTRODUCTION

The stochastic embedding principle [1],[2],[3] has been introduced as a method to quantify errors on estimated transfer functions. This method assumes that the noise and the unmodeled dynamics can be described using stochastic models where only the structure of the associated probability density functions need to be known a priori, the parameters of these pdf's are considered unknown.

In this paper the stochastic embedding method is applied to the case where the true system is composed of a linear part followed by a static nonlinearity. Admittedly, this is a simplification in many cases. However, other types of nonlinear effects may also be captured by the methods that are derived here. Some applications where such cascade systems appear include sensor saturation compensation [4], control valves [5], biological systems [6] and sensor evaluation for fault detection functionality, cf. the simulation example in section 5. A more theoretical advantage exists in connection to the work of [4]. Some of the algorithms of that paper can be modified to utilize a priori knowledge of the system in the regressor filtering. Global parametric convergence of these schemes require positive realness of certain transfer functions. One suitable initialization method is then to estimate a linear model to get initial values and to use error bounds on this estimate in order to secure the stipulated positive real conditions. The importance of error bounds on transfer functions and of estimates on the sizes of nonlinearities, is well established in the literature on robust control.

In this paper it is assumed that the objective is to identify the transfer function of the linear part of the system, using a model that is purely linear. To obtain a measure of the quality of the result, an expression for the ensemble mean square error of the transfer function estimate is derived that does not require knowledge of the actual system parameters. This quality-measure only gives information on the error of the amplitude. Furthermore the accuracy of this estimated mean square error depends on the assumption that the unknown nonlinearity can be approximated with a set of known basis functions.

The problem under consideration is described in the next section. An expression for the covariance of the embedded parameters of the unknown nonlinearity is derived in section 3. This result is then used in order to estimate the covariance of the parameter errors in section 4. Error bounds on the estimate of the transfer function are obtained. Computer simulation results are presented in section 5. Finally, a summary is given in the last section.

## 2 SYSTEM, MODEL AND MODELING ERRORS

It is assumed that the true system is composed of a dynamic FIR block in cascade with a static nonlinearity, the latter part given by

$$f(x) = x + \sum_{i \geq 2} b_i x^i. \quad (1)$$

Note that this means that the gain of the system has been fixed in the nonlinear block, cf. [7] and [8]. The lack of a  $b_1$  for the linear term also stresses the fact that the focus is on nonlinear undermodeling. The parameters  $b_i$  will be modelled as stochastic variables in the analysis that follows. To avoid notational complications it will generally be assumed in the paper that *the parameters  $b_i$  and the input signal,  $u(n)$ , have zero mean*. Furthermore the output of the nonlinearity is disturbed by additive zero mean noise,  $v(n)$ , with variance  $\sigma_v^2$ . This means that  $y(n)$  will also have zero mean. The resulting output of the true system at time instant  $n$  can now be written,

$$y(n) = \underline{\theta}^T \underline{u}(n) + \sum_{i \geq 2} b_i [\underline{\theta}^T \underline{u}(n)]^i + v(n) \quad (2)$$

where  $\underline{u}(n)$  is the input vector to the system

$$\underline{u}^T(n) = [u(n) \ u(n-1) \ \dots \ u(n-L+1)] \quad (3)$$

and where  $\underline{\theta}$  is the parameter vector of the linear part of the system.

$$\underline{\theta}^T = [\theta_1 \ \theta_2 \ \dots \ \theta_L]. \quad (4)$$

A linear model is then estimated, typically with the least squares method. Since the system is nonlinear, modeling errors arise. Assuming that the estimated parameter vector  $\hat{\underline{\theta}}$  has the same dimension as the true vector  $\underline{\theta}$ , the LS-estimate becomes

$$\hat{\underline{\theta}} = \left[ \sum_{n=1}^N \underline{u}(n) \underline{u}^T(n) \right]^{-1} \sum_{n=1}^N \underline{u}(n) y(n) \quad (5)$$

and the output of the linear model is

$$\hat{y}(n) = \hat{\underline{\theta}}^T \underline{u}(n). \quad (6)$$

Then the error  $w(n)$  due to the undermodeling and the noise becomes

$$w(n) = y(n) - \hat{y}(n)$$

$$\begin{aligned}
&= [\underline{\theta} - \hat{\underline{\theta}}]^T \underline{u}(n) + \sum_{i \geq 2} b_i [\underline{\theta}^T \underline{u}(n)]^i + v(n) \\
&= \underline{\theta}_d^T \underline{u}(n) + \underline{c}^T \underline{g}(n) + v(n)
\end{aligned} \tag{7}$$

where  $\underline{\theta}_d = \underline{\theta} - \hat{\underline{\theta}}$  is the parameter error vector and the elements of  $\underline{g}(n)$  are powers and products of powers of the input signal  $u(n)$ . The elements of  $\underline{c}$  depend on the linear system parameters,  $\underline{\theta}$  and the coefficients  $b_i$  of the nonlinearity. An expression for the associated unknown covariance matrix of  $\underline{c}$ , denoted  $C_c$ , will be derived in the next section. This quantity will be central when error bounds are derived later on. If for a given trial  $N$  measurements are collected, the corresponding  $N \times 1$  error vector can be written, using upper case letters to denote matrices,

$$\underline{w}(n) = \Phi(n)\underline{\theta}_d + A(n)\underline{c} + \underline{v}(n) \tag{8}$$

with

$$\Phi^T(n) = [\underline{u}(n) \ \underline{u}(n-1) \ \dots \ \underline{u}(n-N+1)] \tag{9}$$

$$A^T(n) = [\underline{g}(n) \ \underline{g}(n-1) \ \dots \ \underline{g}(n-N+1)]. \tag{10}$$

As the elements of  $\underline{g}(n)$  are powers and products of powers of the input signal  $u(n)$ ,  $A(n)$  has the form

$$A(n) = \begin{pmatrix} u^2(n) & 2u(n)u(n-1) \dots \\ u^2(n-1) & \dots \\ \vdots & \vdots \\ u^2(n-N+1) & 2u(n-N+1)u(n-N) \dots \end{pmatrix} \tag{11}$$

The columns of  $A(n)$  correspond to the basis functions which are assumed to be known.

### 3 COVARIANCE OF THE EMBEDDED PARAMETERS

The problem of estimating the covariance matrix of the embedded parameters  $\underline{c}$  in (8) will now be addressed<sup>1</sup>. Returning to (2), vectorize the true system output,  $y(n)$  corresponding to a collection of  $N$  measurements for a particular trial as

$$\underline{y}(n) = \Phi(n)\underline{\theta} + A(n)\underline{c} + \underline{v}(n) \tag{12}$$

with

$$\underline{y}^T(n) = [y(n) \ y(n-1) \ \dots \ y(n-N+1)]. \tag{13}$$

---

<sup>1</sup>Note that this means that no direct estimate of the size of  $b_i$  is obtained

To obtain an expression that is independent of  $\Phi(n)$  project (12) on the subspace orthogonal to that spanned by the columns of  $\Phi(n)$ . With the orthogonal projection matrix

$$P_{\Phi}^{\perp}(n) = I - \Phi(n)[\Phi^T(n)\Phi(n)]^{-1}\Phi^T(n) \quad (14)$$

where  $I$  is the identity matrix, then

$$P_{\Phi}^{\perp}(n)\Phi(n) = 0 \quad (15)$$

holds and the following expression is obtained from (12)

$$P_{\Phi}^{\perp}(n)\underline{y}(n) = P_{\Phi}^{\perp}(n)A(n)\underline{c} + P_{\Phi}^{\perp}(n)\underline{v}(n). \quad (16)$$

The covariance matrix of the true system output  $\underline{y}(n)$  can be written

$$\Omega(n) = E[\underline{y}(n)\underline{y}^T(n)] = \lim_{NT \rightarrow \infty} \frac{1}{NT} \sum_{k=1}^{NT} \underline{y}_k \underline{y}_k^T \quad (17)$$

where the expectation is taken w.r.t. the embedded parameters  $\underline{c}$  and the noise  $\underline{v}(n)$ .  $NT$  equals the number of independent trials.

Now form the covariance of (16), assume that  $\underline{c}$  and  $\underline{v}(n)$  are uncorrelated and use the estimate  $\hat{\Omega}(n)$ , corresponding to a large but finite  $NT$  in (17). Then

$$\begin{aligned} & P_{\Phi}^{\perp}(n)\hat{\Omega}(n)P_{\Phi}^{\perp}(n) \\ &= P_{\Phi}^{\perp}(n)A(n)E[\underline{c}\underline{c}^T]A^T(n)P_{\Phi}^{\perp}(n) + \hat{\sigma}_v^2 P_{\Phi}^{\perp}(n) \\ &= Q(n)E[\underline{c}\underline{c}^T]Q^T(n) + \hat{\sigma}_v^2 P_{\Phi}^{\perp}(n) \end{aligned} \quad (18)$$

where the estimate of the measurement noise covariance,  $\hat{\sigma}_v^2$  has been used. Furthermore  $Q(n) = P_{\Phi}^{\perp}(n)A(n)$ .

With  $Q^{\dagger}(n) = [Q^T(n)Q(n)]^{-1}Q^T(n)$  being the pseudoinverse of  $Q(n)$ , the estimated covariance of the embedded parameters,  $\underline{c}$  might be obtained from (18) as

$$\hat{C}_c = \hat{E}[\underline{c}\underline{c}^T] = Q^{\dagger}(n)\hat{\Omega}(n)Q^{\dagger T}(n) - \hat{\sigma}_v^2[Q^T(n)Q(n)]^{-1} \quad (19)$$

where  $\hat{E}$  denotes the estimated expectation. However, there is no guarantee that  $\hat{C}_c$ , as obtained from (19), is positive semidefinite, unless  $\hat{\sigma}_v^2$  is zero. In order to circumvent this problem use the assumption that the embedded parameters,  $\underline{c}$  are orthogonal to the measurement noise,  $\underline{v}$  and introduce the eigendecomposition

$$P_{\Phi}^{\perp}(n)\hat{\Omega}(n)P_{\Phi}^{\perp}(n)$$

$$\begin{aligned}
&= Q(n)E[\underline{c}\underline{c}^T]Q^T(n) + \hat{\sigma}_v^2 P_{\Phi}^{\perp}(n) \\
&= \hat{E}_s(n)\hat{\Lambda}_s(n)\hat{E}_s^T(n) + \hat{E}_n(n)\hat{\Lambda}_n(n)\hat{E}_n^T(n)
\end{aligned} \tag{20}$$

where the columns of  $\hat{E}_s(n)$  are the estimated eigenvectors which span the signal subspace. Its orthogonal complement, the noise subspace, is spanned by the estimated eigenvectors that are columns of  $\hat{E}_n(n)$ . Noting that the matrix  $P_{\Phi}^{\perp}(n)\hat{\Omega}(n)P_{\Phi}^{\perp}(n)$  is real and positive semidefinite, its associated eigenvalues are real and greater than or equal to zero. The  $d$  largest eigenvalues, that are associated with the "nonlinear term" in (20), are elements of the diagonal matrix  $\hat{\Lambda}_s(n)$ . The smaller eigenvalues, that are elements of  $\hat{\Lambda}_n(n)$ , are equal to the estimated noise covariance,  $\hat{\sigma}_v^2$ . It is assumed that the number  $d$  can be determined using an appropriate threshold. Furthermore since  $\hat{E}_n(n)\hat{E}_n^T(n) = I - \hat{E}_s(n)\hat{E}_s^T(n)$  it follows from (20) that

$$\begin{aligned}
&Q(n)E[\underline{c}\underline{c}^T]Q^T(n) + \hat{\sigma}_v^2 P_{\Phi}^{\perp}(n) \\
&= \hat{E}_s(n)\hat{\Lambda}_s(n)\hat{E}_s^T(n) + \hat{\sigma}_v^2 I - \hat{\sigma}_v^2 \hat{E}_s(n)\hat{E}_s^T(n).
\end{aligned} \tag{21}$$

Now postmultiplying (21) with  $P_{\Phi}^{\perp}(n)$ , noting that  $P_{\Phi}^{\perp}(n) = P_{\Phi}^{\perp 2}(n)$  and  $Q^T(n)P_{\Phi}^{\perp}(n) = Q^T(n)$ , it follows that

$$\begin{aligned}
&Q(n)E[\underline{c}\underline{c}^T]Q^T(n) \\
&= \hat{E}_s(n)\hat{\Lambda}_s(n)\hat{E}_s^T(n)P_{\Phi}^{\perp}(n) - \hat{\sigma}_v^2 \hat{E}_s(n)\hat{E}_s^T(n)P_{\Phi}^{\perp}(n) \\
&= \hat{E}_s(n)\hat{\Lambda}_t(n)\hat{E}_s^T(n)P_{\Phi}^{\perp}(n)
\end{aligned} \tag{22}$$

where  $\hat{\Lambda}_t(n) = \hat{\Lambda}_s(n) - \hat{\sigma}_v^2 I$  is positive.

Consequently the estimated covariance of the embedded parameters,  $\underline{c}$  can be obtained from (22) as

$$\hat{C}_c = \hat{E}[\underline{c}\underline{c}^T] = Q^{\dagger}(n)\hat{E}_s(n)\hat{\Lambda}_t(n)\hat{E}_s^T(n)Q^{\dagger T}(n). \tag{23}$$

which is the main result of this section. Now  $\hat{C}_c$  is positive semidefinite by construction.

## 4 ERROR BOUNDS

Assuming that the linear part of the system is estimated using a least squares estimate,  $\Phi^{\dagger}(n)\underline{y}$ , then an expression for the parameter error  $\underline{\theta}_d = \hat{\underline{\theta}} - \underline{\theta}$  is found by premultiplying (12) with  $\Phi^{\dagger}(n)$ . The result is

$$\underline{\theta}_d = \hat{\underline{\theta}} - \underline{\theta} = \Phi^{\dagger}(n)A(n)\underline{c} + \Phi^{\dagger}(n)\underline{v} \tag{24}$$

since

$$\hat{\underline{\theta}} = \Phi^\dagger(n)\underline{y} \quad (25)$$

$$\Phi^\dagger(n)\Phi(n) = I \quad (26)$$

An estimate of the covariance of the parameter errors  $\theta_d$  is obtained from (24) as

$$\hat{E}[\underline{\theta}_d\underline{\theta}_d^T] = \Phi^\dagger(n)A(n)\hat{C}_cA^T(n)\Phi^{\dagger T}(n) + \hat{\sigma}_v^2[\Phi^T(n)\Phi(n)]^{-1} \quad (27)$$

where the estimated covariance of the embedded parameters,  $\hat{C}_c$  is obtained from (23) in the previous section.

The square root of the frequency domain equivalent of (27) is used as a measure of the standard deviation of the amplitude of the estimated transfer function. With

$$\Gamma(e^{-j\omega}) = [1 \ e^{-j\omega} \ \dots \ e^{-j(L-1)\omega}] \quad (28)$$

the on-average estimate of the transfer function magnitude error becomes

$$\begin{aligned} & \sqrt{\hat{E}|H(e^{-j\omega}, \theta - \hat{\theta})|^2} = \\ & \sqrt{\hat{E}|\Delta H(e^{-j\omega})|^2} = \sqrt{|\Gamma(e^{-j\omega})\hat{E}[\underline{\theta}_d\underline{\theta}_d^T]\Gamma^T(e^{j\omega})|}. \end{aligned} \quad (29)$$

## 5 NUMERICAL EXAMPLES

In this section the possibility to estimate the stochastic embedding parameters and to quantify their effect on an estimated transfer function is studied with simulation. A practical application where this occurs is sensor evaluation for fault detection functionality in e.g. jet engines [9]. In such a situation samples of sensors, possibly with unknown linearity properties, may be available. The system designer is then typically interested in obtaining linearity measures on the sensors and also in assessing the impact on the associated estimated transfer functions. Performing independent trials with one sensor at a time the proposed methodology can be used for the estimation of all wanted quantities.

*Example 1:* A system and a model, as described in section 2, both of order 12 were simulated. The impulse response of the system was given by

$$\begin{aligned} \underline{\theta}^T = & (0.1578 \ 0.3548 \ 0.3820 \ 0.2872 \ 0.1388 \ -0.0018 \\ & -0.0954 \ -0.1300 \ -0.1147 \ -0.0706 \ -0.0201 \ 0.0197). \end{aligned} \quad (30)$$

In this example a polynomial form, as defined by (1), of order three was chosen to represent the nonlinearity. The input to the system was zero mean white Gaussian noise (WGN) with variance 1.0000 and the parameters  $\underline{b} = [b_2 \ b_3]$  were generated as WGN with zero mean and covariance  $diag[0.5000 \ 0.5000]$ . 800 trials were run, each with  $N=600$  measurements. Furthermore the covariance of the measurement noise was 0.25 and the basis functions were chosen according to (11), with the highest order equal to three. The estimate of the transfer function error,  $\sqrt{\hat{E}|\Delta H(e^{-j\omega})|^2}$ , according to (29) can be compared to the corresponding true error,  $\sqrt{E|\Delta H(e^{-j\omega})|^2}$ , calculated using the true parameters  $\underline{\theta}$  in Fig.1. In this case the two curves almost coincide. So the estimated error bounds give a good indication of the true modeling errors in the frequency domain also for significant output nonlinearities.  $\square$

*Example 2:* In this example a more normal size of the nonlinearity was considered in order to study practical situation a little further. For that purpose amplitude frequency domain errors were computed as described in the paper. In Fig. 2 the true transfer function amplitude is plotted together with the estimate plus and minus two times the standard deviation, as given by (29). As seen in Fig.2 there is a fair agreement between the estimated bounds and the true transfer function.

The conditions were the same as in Example 1 with the following exceptions. No measurement noise was present, 100 realizations were considered and  $\underline{b} = [b_2 \ b_3]$  were generated as WGN with zero mean and covariance  $diag[0.0250 \ 0.0250]$ .  $\square$

## 6 CONCLUSIONS

The problem of quantifying errors due to certain nonlinear undermodeling was addressed. The true system was supposed to consist of a FIR dynamic block in cascade with a static nonlinearity. Although the parameters of the system were considered unknown, it was assumed that the nonlinear part could be represented with a set of known basis functions. The linear part of the system was estimated with the least squares method, using a purely linear model. Consequently modeling errors occurred. With data from a large number of trials available, the stochastic embedding approach was applied to capture the on-average properties of the errors.

First an expression for the covariance of the embedded unknown parameters of the non-

linearity was derived. This result was then used to obtain an expression for the covariance of the parameter errors. This expression does not require knowledge of the true system parameters and was further used in order to obtain a measure of the average error in the estimated transfer function. Computer simulations showed that the estimated error bounds give a good indication of the true modeling errors in the frequency domain. As compared to previous work, the necessary prior information on the second order properties of the undermodeling is reduced.

Further analysis of the properties of the proposed estimates, e.g. in terms of conditions for convergence to the true values, seems worthwhile. The choice of basis functions could be studied in more detail, especially for the case when the true nonlinearity function can not be fitted exactly to a polynomial. It might be advantageous to use orthogonal functions, e.g. Hermite polynomials. A generalization to the IIR model case also seems to be practically important. At least for small nonlinearities this should be achievable by linearization around the obtained estimates. The estimation of a more direct estimate of the nonlinearity, preferably covariances of the  $b_i$ , is another interesting topic for research.

## REFERENCES

- [1] G. C. Goodwin and M. Salgado, "A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models", *Int. J. Adaptive Contr. Signal Processing*, vol. 3, no. 4, pp. 333–356, 1989.
- [2] B. M. Ninness, "Stochastic and Deterministic Modelling", Ph.D. Dissertation, University of Newcastle, N.S.W., Australia, 1993.
- [3] G. C. Goodwin, M. Gevers and B. M. Ninness, "Quantifying the error in estimated transfer functions with application to model order selection", *IEEE Trans. Automat. Contr.*, vol. AC-37, no. 7, pp. 913–928, 1992.
- [4] T. Wigren, "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model", *IEEE Trans. Automat. Contr.*, vol. AC-39, no. 11, pp. 2191–2206, 1994.
- [5] M. G. Singh, J. P. Elloy, R. Mezencev and N. Munro, *Applied Industrial Control*, Oxford, UK.: Pergamon Press, 1980.

- [6] I. W. Hunter and M. J. Korenberg, “The identification of nonlinear biological systems: Wiener and Hammerstein cascade models”, *Biol. Cybern.*, vol. 55, no. 3, pp. 135–144, 1986.
- [7] G. A. Pajunen, “Application of a model reference adaptive technique to the identification and control of Wiener type nonlinear processes”, Ph.D. Dissertation, Acta Polytechnica Scandinavica, no. 52, Dept. Elec. Eng., Helsinki University of Technology, Helsinki, Finland, 1984.
- [8] T. Wigren, “Recursive prediction error identification using the nonlinear Wiener model”, *Automatica*, vol. 29, no. 4, pp. 1011–1025, 1993.
- [9] G. Merrington, O-K. Kwon, G. C. Goodwin and B. Carlsson, “Fault detection and diagnosis in gas turbines”, *ASME J. Engineering for Gas Turbines and Power*, vol. 113, pp. 276–282, 1991.