Probabilistic Uncertainty Bounding in Output Error Models with Unmodelled Dynamics

Sippe G. Douma and Paul M.J. Van den Hof

Abstract— **In prediction error identification probabilistic model uncertainty bounds are generally derived from the statistical properties of the parameter estimator. The probabilistic bounds are then based on an (asymptotic) normal distribution of the parameter estimator, accompanied by a covariance matrix, which generally has to be estimated from data too. When the primal interest of the identification is in quantifying the parameter uncertainty on the basis of one single experiment, alternative methods exist that do no require the specification of the full pdf of the parameter estimator. The objective then is to have simpler computations and less dependency on (asymptotic) assumptions. While in earlier publications the situation of ARX models has been studied, here we consider the situation of nonlinearly parametrized (Output Error) models. It is shown that for this class relatively simple probabilistic uncertainty bounds can be constructed, that are applicable also to the situation where there is unmodelled dynamics ($\mathcal{S} \notin \mathcal{M}$).**

Index Terms— **system identification, undermodelling, model uncertainty, parameter confidence bounds.**

I. Introduction

In the mainstream approach of system identification, i.e. prediction error identification, model uncertainty quantification is based on the covariance matrix of the parameter estimator, in conjunction with a presumed (or asymptotically achieved if the number of data tends to infinity) Gaussian probability density function, see e.g., [10]. This description leads to probabilistic confidence bounds on estimated parameters, from which also probabilistic confidence bounds on estimated frequency responses can be constructed, with any prechosen level of probability.

In classical prediction error identification explicit and exact expressions for the parameter covariance matrix are available for model structures that are linear-in-the-parameters in the situation that the model structures are correct, i.e. the data generating system is part of the model set, $\mathcal{S} \in \mathcal{M}$. For linear regression models with deterministic regressors (such as FIR and generalized FIR [8]) this holds for finite data length, for ARX models this holds asymptotically. For general model structures, and under the assumption $\mathcal{S} \in \mathcal{M}$, the parameter covariance matrix can be approximated by using first order Taylor expansions. However, in this situation exact system knowledge is also required to compute these approximate expressions for the covariance matrix.

Only in case of linear parametrizations results are available for model uncertainty bounding when the model structures are not correct ($\mathcal{S} \notin \mathcal{M}$), see e.g. [7], [9] and [8], Chapter 7.

For Gaussian distributions parameter confidence bounds that are constructed on the basis of the (exact) covariance matrix of the parameter estimator, lead to the smallest possible parameter uncertainty regions for a given probability level. However, usually the exact covariance matrix is not available, and a replacement has to be made with an estimated covariance.

In this paper we continue the line of research that is presented in [4] where it is shown that alternative probabilistic parameter bounds can be derived for ARX and OE models, that do not require full specification of the estimator probability density function (pdf). This approach is based on the analysis of a data-dependent mapping of parameter estimates. In [5] particularly the case of ARX models is analyzed, and it is shown that the alternative approach shows potentials for analyzing finite-time results also. In the current paper we will focus on the situation of nonlinearly parametrized (output error) models. It will be shown that even in the situation of unmodelled dynamics $\mathcal{S} \notin \mathcal{M}$, probabilistic model uncertainty bounds can be derived in a reliable way. Additionally it will be shown that the presented uncertainty intervals have a very good performance for finite time data records also. Whenever necessary for distinguishing random variables from their realizations, boldface symbols are used to denote random variables.

II. Output Error modelling - the classical approach

In an Output Error (OE) model structure we consider the one-step ahead predictor

$$\hat{y}(t|t-1;\theta) = \frac{B(q,\theta)}{F(q,\theta)}u(t)$$

and we denote the predictor derivative:

$$\psi(t,\theta) = \frac{\partial}{\partial\theta}\hat{y}(t|t-1;\theta),$$

where u and y are input and output signal, and B and F are polynomials in the delay operator q^{-1} [10].

For quantifying parameter uncertainty bounds in the standard approach, the starting point is a first order Taylor expansion:

$$(\hat{\theta}_N - \theta_0) \sim -[\bar{V}''(\theta_0)]^{-1}[V_N'(\theta_0)] \qquad (1)$$

where $V_N'(\theta_0) = \partial V_N(\theta)/\partial\theta|_{\theta=\theta_0}$,
$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta)^2$, $\varepsilon(t,\theta) = y(t) - \hat{y}(t|t-1;\theta)$ and $\bar{V}''$ is the second derivative of $\bar{V} = \bar{\mathbb{E}}\varepsilon(t,\theta)^2$ and

The authors are with the Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands. Tel. +31-15-2784509; Fax: +31-15-2786679; E-mail: p.m.j.vandenhof@tudelft.nl.

Sippe Douma is now with Shell International Exploration and Production, Rijswijk, The Netherlands.

N is the number of data points. In the first order Taylor approximation, the asymptotic expressions

$$V_N'(\theta_0) \to \mathcal{N}(0, Q), \quad Q = \sigma_e^2 \mathbb{E}[\tfrac{1}{N}\Psi(\theta_0)^T\Psi(\theta_0)] \quad (2)$$

$$\bar{V}''(\theta_0) = \mathbb{E}[\tfrac{1}{N}\Psi(\theta_0)^T\Psi(\theta_0)] \quad (3)$$

with

$$\Psi(\theta) = \begin{pmatrix} \psi^T(1,\theta) \\ \vdots \\ \psi^T(N,\theta) \end{pmatrix}$$

are then substituted to arrive at the asymptotic covariance matrix of $(\hat{\theta}_N - \theta_0)$ given by

$$P_{oe} = \sigma_e^2[\mathbb{E}\frac{1}{N}\Psi(\theta_0)^T\Psi(\theta_0)]^{-1}. \quad (4)$$

The following asymptotic model uncertainty region can then be formulated based on asymptotic normality of the parameter estimator:

$$\{\theta_0 \in \mathcal{D}_{oe}(\alpha, \hat{\theta}_N)\} \text{ w.p. } \alpha, \text{ with} \quad (5)$$

$$\mathcal{D}_{oe}(\alpha, \hat{\theta}_N) := \{\theta \mid (\theta - \hat{\theta}_N)^T P_{oe}^{-1}(\theta - \hat{\theta}_N) \leq \frac{c_\chi(\alpha, n)}{N}\},$$

where $c_\chi(\alpha, n)$ refers to the number indicating the α-probability level in a χ-squared distributed random variable with n degrees of freedom.

Note that the approximations and assumptions that are involved in this analysis include:

- An approximation in the first order Taylor approximation (1);
- Convergence to (asymptotic) normality of the product term on the right hand side of (1), and
- Convergence of the covariance of this term to the product of the separate terms, as reflected in (2) and (3).

For arriving at a computable expression for the parameter uncertainty, the unknown terms σ_e^2 and $[\mathbb{E}\frac{1}{N}\Psi(\theta_0)^T\Psi(\theta_0)]$ that appear in the covariance matrix (4) are replaced by estimates, to arrive at

$$\hat{P}_{oe} = \hat{\sigma}_e^2[\frac{1}{N}\Psi(\hat{\theta}_N)^T\Psi(\hat{\theta}_N)]^{-1}. \quad (6)$$

III. Alternative Approach for Parameter Uncertainty Bounding

In our new paradigm, the starting point for the analysis of the parameter estimate is the derivative of the identification criterion: $V_N'(\hat{\theta}_N) = 0$ or equivalently

$$\frac{1}{N}\sum_{t=1}^{N}[y(t) - \frac{B(q,\hat{\theta}_N)}{F(q,\hat{\theta}_N)}u(t)] \cdot \psi(t, \hat{\theta}_N) = 0. \quad (7)$$

The data generating system is assumed to be specified by

$$y(t) = G_0(q)u(t) + e(t) \quad (8)$$

with e a white noise process.

To analyse the model error $G_0(q) - G(q, \hat{\theta}_N)$ we introduce the following decomposition

$$G_0(q) - G(q,\hat{\theta}_N) = \Big(G(q,\theta^*) - G(q,\hat{\theta}_N)\Big) + \bar{G}_0(q,\theta^*) \quad (9)$$

where $G(q,\theta^*) \in \mathcal{M}$ with θ^* defined as the minimizing argument: $\theta^* = \arg\min_\theta \bar{\mathbb{E}}\varepsilon(t,\theta)^2$, for an identification experiment with sufficiently exciting input signal. It is the best representative of $G_0(q)$ within the model set; $\bar{G}_0(q,\theta^*)$ denotes the part of $G_0(q)$ which cannot be contained in the model class $\mathcal{M}$. If the data generating system is within the model set ($\mathcal{S} \in \mathcal{M}$) then $G(q,\theta^*) = G_0(q)$ and $\bar{G}_0(q,\theta^*) = 0$.

Whereas in [4] an alternative implicit expression for $\hat{\theta}_N$ is derived to specify probabilistic uncertainty bounds, here we utilize a first order Taylor expansion of the model error. This is to facilitate the handling of unmodelled dynamics as well. It is slightly different from the classical situation where a first order Taylor approximation is made of the cost function.

The first order Taylor coefficient of the model error term $(G(q,\theta^*) - G(q,\hat{\theta}_N))$ around $(\theta^* - \hat{\theta}_N)$ is given by $\partial G(q,\hat{\theta}_N)/\partial\hat{\theta}_N$. By writing

$$G(q,\hat{\theta}_N) = \frac{\Omega_1^T(q)\hat{\theta}_b}{\Omega_2^T(q)\hat{\theta}_f}$$

with

$$\Omega_1(q) = \begin{bmatrix} 1 & q^{-1} & \cdots & q^{-n_b-1} \end{bmatrix}^T q^{-n_k} \quad (10)$$

$$\Omega_2(q) = \begin{bmatrix} q^{-1} & \cdots & q^{-n_f} \end{bmatrix}^T, \quad (11)$$

and $\hat{\theta}_N = [\hat{\theta}_f^T \ \hat{\theta}_b^T]^T$ it follows that $\partial G(q,\hat{\theta}_N)/\partial\hat{\theta}_N$ is given by

$$K(q,\hat{\theta}_N) := \begin{bmatrix} -G(q,\hat{\theta}_N)\frac{1}{F(q,\hat{\theta}_N)}\Omega_2^T(q) & \frac{1}{F(q,\hat{\theta}_N)}\Omega_1^T(q) \end{bmatrix}^T \quad (12)$$

As a result expression (9) can be rewritten as

$$G_0(q) - G(q,\hat{\theta}_N) = K^T(q,\hat{\theta}_N)\left(\theta^* - \hat{\theta}_N\right) + \tilde{G}_0(q,\theta^*,\hat{\theta}_N) \quad (13)$$

where $\tilde{G}_0(q,\theta^*,\hat{\theta}_N) = \bar{G}_0(q,\theta^*) + \breve{G}(q,\theta^*,\hat{\theta}_N)$, the latter term $\breve{G}$ being the linearization error that is induced by the first order Taylor approximation in (12).

Now substituting expressions (8) and (13) into (7) it follows that

$$\frac{1}{N}\sum_{t=1}^{N}[K^T(q,\hat{\theta}_N)u(t)(\theta^* - \hat{\theta}_N) + \\ + e(t) + w(t)] \cdot \psi(t,\hat{\theta}_N) = 0, \quad (14)$$

with

$$w(t) = \tilde{G}_0(q,\theta^*,\hat{\theta}_N)u(t). \quad (15)$$

Utilizing the fact that

$$K^T(q,\hat{\theta}_N)u(t) = \psi^T(t,\hat{\theta}_N)$$

the parameter error $(\theta^* - \hat{\theta}_N)$ in (14) can be written in a linear regression-type equation as:

$$\theta^* - \hat{\theta}_N = -(\Psi(\hat{\theta}_N)^T\Psi(\hat{\theta}_N))^{-1}\Psi(\hat{\theta}_N)^T(\mathbf{e} + \mathbf{w}) \quad (16)$$

where $\mathbf{e}$ and $\mathbf{w}$ are column vectors with stacked values of $e(t)_{t=1,\cdots N}$ and $w(t)_{t=1,\cdots N}$ respectively.

IV. Situation $\mathcal{S} \in \mathcal{M}$ and neglecting linearization error

If the system is in the model set, and we neglect the error due to first order Taylor approximation, the expression for the parameter error becomes:

$$\hat{\theta}_N - \theta_0 = (\Psi^T\Psi)^{-1}\Psi^T\mathbf{e} \tag{17}$$

where we have used the short-hand notation $\Psi = \Psi(\hat{\theta}_N)$. For specifying the parameter uncertainty bounds, we now consider the expression

$$\frac{1}{\sqrt{N}}\Psi^T\Psi(\hat{\theta}_N - \theta_0) = \frac{1}{\sqrt{N}}\Psi^T\mathbf{e}$$

The random variable $\boldsymbol{\beta} = \frac{1}{\sqrt{N}}\boldsymbol{\Psi}^T\mathbf{e}$ is asymptotically Gaussian distributed with zero mean and covariance matrix

$$Q = \sigma_{\mathbf{e}^2}\mathbb{E}[\frac{1}{N}\boldsymbol{\Psi}^T\boldsymbol{\Psi}]$$

Mapping the uncertainty region of $\boldsymbol{\beta}$ to $\hat{\theta}_N - \theta_0$, using the similar reasoning as presented in [4] then leads to the asymptotic uncertainty region

$$\{\theta_0 \in \mathcal{D}_{oe,n}(\alpha, \hat{\theta}_N)\} \text{ w.p. } \alpha, \text{ with}$$

$$\mathcal{D}_{oe,n}(\alpha, \hat{\theta}_N) := \{\theta \mid (\theta - \hat{\theta}_N)^T P_{oe,n}^{-1}(\theta - \hat{\theta}_N) \leq \frac{c_\chi(\alpha, n)}{N}\}$$

with

$$P_{oe,n} = (\frac{1}{N}\Psi^T\Psi)^{-1} \cdot Q \cdot (\frac{1}{N}\Psi^T\Psi)^{-1}. \tag{18}$$

In order to arrive at a computable expression for the parameter uncertainty, $\sigma_{\mathbf{e}}^2$ is replaced by an estimate $\hat{\sigma}_e^2$ and $\mathbb{E}[\frac{1}{N}\boldsymbol{\Psi}^T\boldsymbol{\Psi}]$ is replaced by $[\frac{1}{N}\Psi^T\Psi]$ leading to the estimate

$$\hat{P}_{oe,n} = (\frac{1}{N}\Psi^T\Psi)^{-1}\hat{\sigma}_e^2. \tag{19}$$

Note that $P_{oe,n}$ does not have the interpretation of the covariance matrix of $\hat{\theta}_N$. It only serves to specify the parameter uncertainty region. Note also that the computational expression for the parameter uncertainty region has become exactly the same as the expressions (5) and (6) in the classical case. However whereas the classical analysis relies on asymptotic normality of the right hand side of (1), the alternative analysis relies on asymptotic normality of $\frac{1}{\sqrt{N}}\boldsymbol{\Psi}^T\mathbf{e}$.

V. Extension of the OE results

The probabilistic expressions in the previous section, are based on the statistical properties of the expression $\boldsymbol{\beta} := \frac{1}{\sqrt{N}}\boldsymbol{\Psi}^T\mathbf{e}$. In order to further isolate the role of the noise distribution, the following Lemma will be instrumental. For completeness the proof is also added in the appendix.

Lemma 1 ([5]): Consider random vectors $\mathbf{z}, \mathbf{e} \in \mathbb{R}^{N\times 1}$ and a random matrix $\mathbf{V} \in \mathbb{R}^{N\times N}$ related through

$$\mathbf{z} = \mathbf{V}^T\mathbf{e}.$$

If the following properties are satisfied:

1) $\mathbf{e}$ has independent identical Gaussian distributed entries, $\mathcal{N}(0, \sigma^2)$, and
2) $\mathbf{e}$ and $\mathbf{V}$ are independent, and
3) $\mathbf{V}$ is unitary, i.e. $\mathbf{V}^T\mathbf{V} = I$

then the vector elements of $\mathbf{z}$ are independent identically distributed with Gaussian distribution $\mathcal{N}(0, \sigma^2)$. □

The result of this Lemma is quite remarkable. Irrespective of the *pdf* of the elements of matrix $\mathbf{V}$, the resulting random variable $\mathbf{z}$ has a Gaussian distribution. We can now formulate the following result that is relevant for OE models.

Proposition 1: Given a random matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N\times n}$ and a random vector $\mathbf{e} \in \mathbb{R}^{N\times 1}$. The vector $\mathbf{e}$ contains independent identically distributed Gaussian random variables with distribution $\mathcal{N}(0, \sigma^2)$.
If $\boldsymbol{\Psi}$ and $\mathbf{e}$ are statistically independent, then for every realization (z, Ψ) of the random variables $(\mathbf{z}, \boldsymbol{\Psi})$ with

$$z := (\boldsymbol{\Psi}^T\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^T\mathbf{e}$$

it holds that

$$z^T(\Psi^T\Psi)z \leq c_\chi(\alpha, n) \quad \text{is true w.p. } \alpha.$$

This result is exact for all values of N.

Proof Utilizing the singular value decomposition of $\boldsymbol{\Psi}^T$: $\boldsymbol{\Psi}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = I$, it follows that $\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{z} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{e}$.
We now denote

$$\boldsymbol{\beta} := \boldsymbol{\Sigma}^{-1}\mathbf{U}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{z} = \mathbf{V}^T\mathbf{e}. \tag{20}$$

If $\mathbf{V}$ and $\mathbf{e}$ are independent (upon assumption) then according to Lemma 1 $\boldsymbol{\beta} \in \mathcal{N}(0, \sigma^2 I_n)$ and consequently for every realization β corresponding to (z, Ψ) with $\Psi = U\Sigma V^T$ it follows that

$$\beta^T\beta \leq c_\chi(\alpha, n) \quad \text{w.p. } \alpha.$$

Substituting β in the left hand side then delivers

$$z^T\Psi\Psi^T U\Sigma^{-1}\Sigma^{-1}U^T\Psi^T\Psi z$$

which by using $\Psi^T\Psi = U\Sigma^2U^T$, simplifies to $z^T(\Psi^T\Psi)z$ which proves the result. □

When applying Proposition 1 to estimated OE models, the suggested consequence is that for all N with probability α:

$$(\hat{\theta}_N - \theta_0)^T\tfrac{1}{N}\Psi^T\Psi(\hat{\theta}_N - \theta_0) \leq \quad \frac{c_\chi(\alpha, n)}{N}$$

and consequently $\theta_0 \in \mathcal{D}(\alpha, \hat{\theta}_N)$, w.p. α,

$$\text{with } \mathcal{D}(\alpha, \hat{\theta}_N) := \tag{21}$$
$$\{\theta \mid (\theta - \hat{\theta}_N)^T\frac{1}{N}\Psi^T\Psi(\theta - \hat{\theta}_N) \leq \frac{c_\chi(\alpha, n)}{N}\}.$$

This - very strong looking - result however is not exactly true, due to the fact that Proposition 1 requires statistical independence of $\mathbf{V}$ and $\mathbf{e}$, a condition that is not satisfied for OE models. However the statistical relation between $\boldsymbol{\Psi}$ and $\mathbf{e}$ is only through the fact that they both depend on $\hat{\theta}_N$. $\boldsymbol{\Psi}$ does not contain signals (as e.g. output signals) that are correlated directly to the noise signal. Therefore the

correlation between $\mathbf{\Psi}$ and $\mathbf{e}$ is expected to be rather small. Here we conjecture that failure to meet this independence condition hardly affects the Gaussian character of the test statistic $\boldsymbol{\beta}$. Support for this conjecture is illustrated in the following example.

Example 1: Monte Carlo simulations are generated of estimating the two parameters θ_f and θ_b of the transfer function

$$G_0(q) = \frac{\theta_b}{1 + \theta_f q^{-1}}$$

on the basis of a data-generating system

$$y(t) = G_0(q)u(t) + e(t)$$

with $u(t)$ and $e(t)$ independent Gaussian distributed white noise sequences with variance $\sigma_u^2 = \sigma_e^2 = 1$. The parameters θ_f and θ_b are estimated in the OE model structure with a least-squares identification criterion.

The top of Figure 1 depicts the second element of $(\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{\Psi}^T\mathbf{e}$ corresponding with $\hat{\theta}_b$ as a function of data length N. The bottom depicts the distribution of the second element of $\mathbf{V}^T\mathbf{e}$ corresponding with $\hat{\theta}_b$ as a function of data length N. Clearly, the bottom row is nearly indistinguishable from the Gaussian distribution, while the top, even for $N = 50$, only approaches the Gaussian. The figure is based on 2000 Monte Carlo simulations.

The top of Figure 2 depicts the first element of $(\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{\Psi}^T\mathbf{e}$ corresponding with $\hat{\theta}_f$ as a function of data length N. The bottom depicts the distribution of the first element of $\mathbf{V}^T\mathbf{e}$ corresponding with $\hat{\theta}_f$ as a function of data length N. Clearly, the bottom row is nearly indistinguishable from the Gaussian distribution, while the top, even for $N = 50$, does not even come close to the Gaussian distribution. The figure is based on 2000 Monte Carlo simulations.

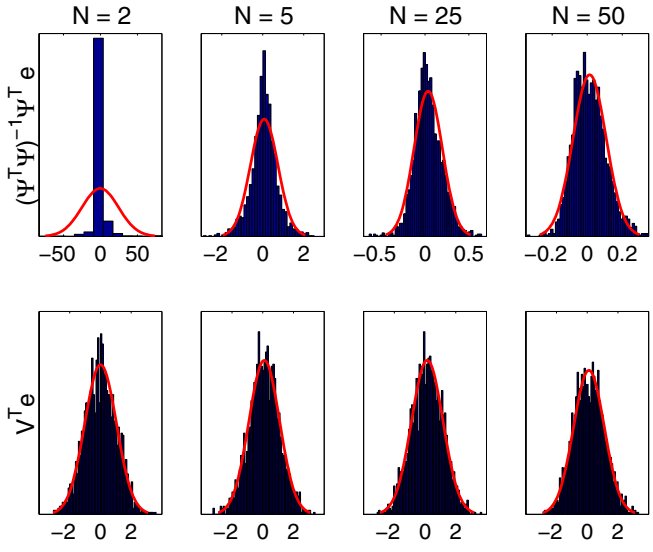

Fig. 1. Distribution of parameters in OE structure. Top: second element of $(\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{\Psi}^T\mathbf{e}$ corresponding to $\hat{\theta}_b$ for data length $N = 2, 5, 25, 50$. Bottom: the distribution of the second element of $\mathbf{V}^T\mathbf{e}$ corresponding to $\hat{\theta}_b$.

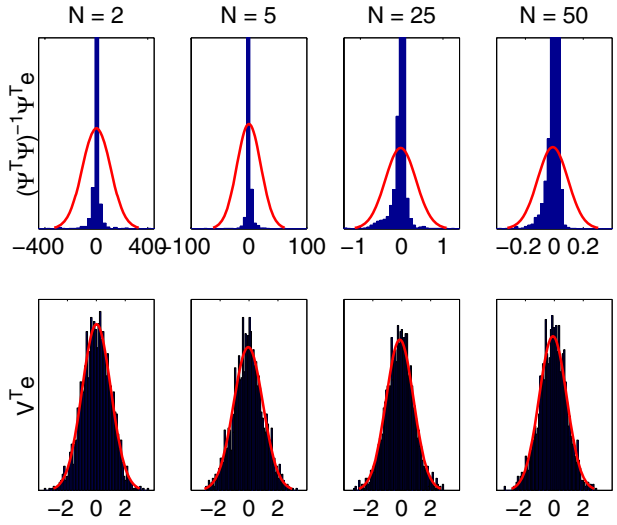

Fig. 2. Similar simulation results as in Figure 1 but then for the test statistic related to the denominator parameter $\hat{\theta}_f$.

The results of the example suggest that the existing correlation between $\mathbf{V}$ and $\mathbf{e}$ hardly affects the normality of the test statistic. This would allow to apply the Gaussian distribution to finite time signals also.

VI. Model error bounding for $\mathcal{S} \notin \mathcal{M}$

A. Model error expression

We are now going to extend the results for parameter uncertainty intervals in the previous sections to the situation of uncertainty bounds on the frequency response of the identified models for the more general situation that undermodelling can be present ($\mathcal{S} \notin \mathcal{M}$). To this end we start again from the parameter error relation (16), that together with (13) can be rewritten into the model error form:

$$G_0(q) - G(q, \hat{\theta}_N) = K^T(q, \hat{\theta}_N)(\Psi^T\Psi)^{-1}\Psi^T(\mathbf{e} + \mathbf{w}) + \\ + \tilde{G}_0(q, \theta^*, \hat{\theta}_N). \qquad (22)$$

We model the error term $\tilde{G}_0(q, \theta^*, \hat{\theta}_N)$ into a series expansion:

$$\tilde{G}_0(q, \theta^*, \hat{\theta}_N) = \Omega_\Delta^T(q)\theta_\Delta$$

with

$$\Omega_\Delta^T(q) = \begin{bmatrix} 1 & q^{-1} & q^{-2} & \cdots \end{bmatrix}^T.$$

θ_Δ then represents the (infinite) pulse response vector of $\tilde{G}_0(q, \theta^*, \hat{\theta}_N) = \left[\bar{G}_0(q, \theta^*) + \breve{G}(q, \theta^*, \hat{\theta}_N)\right]$.

With $w(t)$ given by (15), the vector $\mathbf{w}$ can be written as

$$\mathbf{w} = \Phi_\Delta\theta_\Delta$$

where Φ_Δ is the matrix corresponding to stacking $\left\{\Omega_\Delta^T(q)u(t)\right\}_{t=1,\cdots,N}$.

As a result the model error expression can be written as

$$G_0(q) - G(q, \hat{\theta}_N) = \\ K^T(q, \hat{\theta}_N)(\Psi^T\Psi)^{-1}\Psi^T\mathbf{e} + B_\Delta^T(q)\theta_\Delta \qquad (23)$$

with

$$K_\Delta^T(q) = \left(K^T(q,\hat{\theta}_N)(\Psi^T\Psi)^{-1}\Psi^T\Phi_\Delta + \Omega_\Delta^T(q)\right).$$

B. Bounding the frequency response error

To evaluate the frequency responses of $G_0(q)$ and $G(q,\hat{\theta}_N)$ in terms of their real and imaginary parts, we define the function f_G as

$$f_G(\omega) := \left[\ \Re\left(G(e^{i\omega})\right) \quad \Im\left(G(e^{i\omega})\right)\ \right]^T$$

and note that at a particular frequency ω it holds with expression (23) that

$$f_{G_0}(\omega) - f_{\hat{G}}(\omega) = f_{K^T}(\omega)(\Psi^T\Psi)^{-1}\Psi^T\mathbf{e} + f_{K_\Delta^T}(\omega)\theta_\Delta.$$

For the construction of probabilistic uncertainty bounds we can now follow the same reasoning as was applied in sections IV and V, but now taking account of the undermodelling and linearization error.

To this end we define the random variable

$$\begin{aligned}\beta(\omega) &:= \frac{1}{\sqrt{N}}\Sigma_1^{-1}U^T\left[(f_{G_0}(\omega) - f_{\hat{G}}(\omega)) - f_{K_\Delta^T}(\omega)\theta_\Delta\right] \\ &= \frac{1}{\sqrt{N}}V_1^T\mathbf{e}, \end{aligned} \tag{24}$$

where $U^T = U^{-1} \in \mathbb{R}^{2\times 2}$, $\Sigma_1 \in \mathbb{R}^{2\times 2}$, $V_2 \in \mathbb{R}^{N\times(N-2)}$, $V_1 \in \mathbb{R}^{N\times 2}$ satisfying $V_1^TV_1 = I^{2\times 2}$, such that

$$f_{K^T}(\omega)(\Psi^T\Psi)^{-1}\Psi^T = U\left[\ \Sigma_1 \quad 0\ \right]\left[\begin{array}{c} V_1^T \\ V_2^T \end{array}\right].$$

Conditioned on the reasoning as explained in Section V, Proposition 1,the unknown term on the right hand side of equation (24) is known to satisfy

$$\boldsymbol{\beta} = \frac{1}{\sqrt{N}}V_1^T\mathbf{e} \in \mathcal{N}(0, \frac{\sigma_e^2}{N}\cdot I^{2\times 2}).$$

When taking this result for granted, an uncertainty bound for realizations of the random variable $\boldsymbol{\beta}$ can be specified:

$$\beta \in \mathcal{D}_\beta(\alpha, 0) \text{ w.p.}\alpha, \text{ with}$$

$$\mathcal{D}_\beta(\alpha,0) := \left\{\beta \mid \|\beta\|_2^2 \le \frac{c_\chi(\alpha,2)\sigma_e^2}{N}\right\}.$$

As this probabilistic expression is also valid for the particular estimate

$$\beta = \frac{1}{\sqrt{N}}\Sigma_1^{-1}U^T\left[(f_{G_0}(\omega) - f_{\hat{G}}(\omega)) - f_{K_\Delta^T}(\omega)\theta_\Delta\right],$$

it follows that

$$\left\|\Gamma(\omega)\left[(f_{G_0}(\omega) - f_{\hat{G}}(\omega)) - f_{K_\Delta^T}(\omega)\theta_\Delta\right]\right\|_2^2 \le \frac{c_\chi(\alpha,2)\sigma_e^2}{N}$$

w.p. α, where $\Gamma(\omega)\Gamma^T(\omega) = P^{-1}(\omega)$ and

$$P(\omega) = f_{K^T}(\omega)(\frac{1}{N}\Psi^T\Psi)^{-1}f_{K^T}^T(\omega).$$

With a norm bounded undermodelling term θ_Δ it follows that with probability α:

$$\left\|\Gamma(\omega)\left(f_{G_0}(\omega) - f_{\hat{G}}(\omega)\right) - f_{K_\Delta^T}(\omega)\theta_\Delta\right\|_2 \le \sqrt{\frac{c_\chi(\alpha,2)\sigma_e^2}{N}}$$

leading to

$$\left\|\Gamma(\omega)\left(f_{\hat{G}}(\omega) - f_{G_0}(\omega)\right)\right\|_2 \le \sqrt{\frac{c_\chi(\alpha,2)\sigma_e^2}{N}} + \left\|\Gamma_\omega(\omega)f_{K_\Delta^T}(\omega)\theta_\Delta\right\|_2 \quad \text{w.p. } \ge \alpha$$

The result refers to an ellipsoidal region in the frequency domain that specifies - for each prechosen value of ω - an α probability uncertainty region for the system's frequency response $f_{G_0}(\omega)$. Effects of undermodelling and linearization lead to an increase of the size of the uncertainty ellipsoids.

C. Estimating a bound on the undermodelling

In a practical situation (a bound on) the undermodelling and linearization term θ_Δ has to be estimated from data to determine the model error of expression (25). From expression (16) it follows that

$$\varepsilon(\hat{\theta}_N) = \left(I - \Psi(\Psi^T\Psi)^{-1}\Psi^T\right)(\Phi_\Delta\theta_\Delta + \mathbf{e}), \tag{25}$$

with $\varepsilon(\hat{\theta}_N)$ a stacked column vector with the residual signal for $t = 1, \cdots N$. Note that in this expression now all terms are known except for θ_Δ and $\mathbf{e}$. The expression can be used to estimate the (bound on the) undermodelling term θ_Δ. This can be done in several ways: In model error modelling, [9], the term θ_Δ is directly estimated from (25). In stochastic embedding, [6], the term θ_Δ is parametrized and estimated together with the nominal model. Alternatively, in the approach of [2] and [7] the pulse response represented in θ_Δ is bounded on the basis of the extrapolation of the estimated parameters $\hat{\theta}_N$ itself.

D. Remarks

When $\mathbf{e}$ is not white but has a covariance $\Lambda\sigma_e^2$ then a similar derivation will lead to $P(\omega) = f_{K^T}(\omega)(\frac{1}{N}\Psi^T\Lambda\Psi)^{-1}f_{K^T}^T(\omega)$ by defining

$$U\left[\ \Sigma_1 \quad 0\ \right]\left[\begin{array}{c} V_1^T \\ V_2^T \end{array}\right] := f_{K^T}(\omega)(\Psi^T\Psi)^{-1}\Psi^T\Lambda^{\frac{1}{2}}.$$

Note that the result is very straightforward and exact, except perhaps for the fact that V_1^T is not exactly uncorrelated with $\mathbf{e}$.

Further as more frequency points are considered simultaneously (say n_ω), it follows that more rows of V are to be considered as V_1, resulting in the fact that

$$V_1\mathbf{e} \in \mathcal{N}(0, \frac{\sigma_e^2}{N}\cdot I^{\max(n,2n_\omega)\times\max(n,2n_\omega)}).$$

Consequently, for the total frequency response error to be bounded over all frequencies the bound $\frac{c_\chi(\alpha,2)\sigma_e^2}{N}$ is to be replaced by $\frac{c_\chi(\alpha,n)\sigma_e^2}{N}$, see also e.g. [1].

VII. CONCLUSIONS

Using an alternative paradigm for probabilistic uncertainty bounding, it has been shown that uncertainty bounds can be formulated for nonlinearly parametrized (Output Error) models, using tools and expressions that usually are only fit for linear regresson models. The presented approach appears to have high potentials for use also in situations of

finite data, while - in contrast with existing methods - the situation of undermodelling ($\mathcal{S} \notin \mathcal{M}$) is covered as well. A further generalization of the results to the Box-Jenkins model structure is developed in [3].

REFERENCES

[1] X. Bombois, B.D.O. Anderson and M. Gevers. Mapping parametric confidence ellipsoids to Nyquist plane for linearly parametrized transfer functions. In: G.C. Goodwin (Ed.), *Model Identification and Adaptive Control*, pp. 53-71. Springer-Verlag, London, UK, 2001.

[2] D.K. de Vries and P.M.J. Van den Hof. Quantification of uncertainty in transfer function estimation: A mixed probabilistic-worst-case approach. *Automatica*, 31(4):543–557, 1995.

[3] S.G. Douma. *From Data to Performance - System Identification Uncertainty and Robust Control Design*. PhD thesis, Delft Center for Systems and Control, Delft Univ. Technology, 2006.

[4] S.G. Douma and P.M.J. Van den Hof. An alternative paradigm for probabilistic uncertainty bounding in prediction error identification. *Proc. 44th IEEE Conf. Decision and Control and European Control Conference ECC'05, CDC-ECC'05*, December 12-15, 2005, Sevilla, Spain, pp. 4970-4975.

[5] S.G. Douma and P.M.J. Van den Hof. Probabilistic model uncertainty bounding: an approach with finite-time perspectives. *Proc. 14th IFAC Symp. System Identification*, Newcastle, Australia, March 29-31, 2006, Paper FrA2.2.

[6] G.C. Goodwin, M. Gevers, and B.M. Ninness. Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Autom. Control*, 37:913–928, 1992.

[7] R.G. Hakvoort and P.M.J. Van den Hof. Identification of probabilistic uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Trans. Automatic Control*, 42(11):1516–1528, 1997.

[8] P.S.C. Heuberger, P.M.J. Van den Hof, and B. Wahlberg, editors. *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer Verlag, 2005.

[9] L. Ljung. Model validation and model error modeling. In B. Wittenmark and A. Rantzer, editors, *The Åström Symposium on Control*, pages 15–42, Lund, Sweden, August 1999. Studentlitteratur.

[10] L. Ljung. *System Identification — Theory for the User*. Prentice-Hall, Upper Saddle River, NJ, second edition, 1999.

[11] M.B. Priestley. *Spectral Analysis and Time Series*. Acad. Press, London, UK, 1981.

APPENDIX

Proof of Lemma 1.

Define the vector valued function $g(\cdot) : \mathbb{R}^{\left(n+n^2\right)\times 1} \to \mathbb{R}^{\left(n+n^2\right)\times 1}$, defined by

$$g(\mathbf{z},\mathbf{v}) := \left[\begin{array}{c} \mathbf{e} \\ \mathbf{v} \end{array}\right] = \left[\begin{array}{cc} \mathbf{V} & 0 \\ 0 & I^{n^2\times n^2} \end{array}\right] \left[\begin{array}{c} \mathbf{z} \\ \mathbf{v} \end{array}\right], \qquad \text{(A.26)}$$

with $\mathbf{v} = col(\mathbf{V}^T)$ a vector containing all elements of $\mathbf{V}^T$. When denoting $\mathbf{e}' := [\mathbf{e}^T \ \mathbf{v}^T]^T$ and $\mathbf{z}' := [\mathbf{z}^T \ \mathbf{v}^T]^T$ it follows that

$$p_{\mathbf{z}}(z) = \int_v p_{\mathbf{z}'}(z')dv.$$

Using the mapping from $\mathbf{z}'$ to $\mathbf{e}'$ it follows from standard theory on the transformation of random variables [11] that

$$p_{\mathbf{z}'}(z') = p_{\mathbf{e}'}(g(z')) \cdot \det(J(g(z')))$$

with the Jacobian given by

$$J(g(z')) = \left[\begin{array}{cc} V & Z \\ 0 & I^{n^2\times n^2} \end{array}\right],$$

and Z containing the partial derivatives of $V\mathbf{z}$ to $\mathbf{v}$. Consequently

$$\begin{aligned} p_{\mathbf{z}}(z) &= \int_v p_{\mathbf{e}'}(g(z')) \cdot \det(J)dv \\ &= \int_v p_{\mathbf{e}}(Vz)p_{\mathbf{v}}(v) \cdot \det(J)dv \end{aligned}$$

where the latter equation follows from the fact that $\mathbf{e}$ and $\mathbf{V}$ are independent. Using the Gaussian distribution of $\mathbf{e}$ and the fact that $\det(J) = \det(V) = 1$ it follows that

$$\begin{aligned} p_{\mathbf{z}}(z) &= \int_v \frac{1}{\sigma(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sigma^{-2}z^T V^T V z} \ p_{\mathbf{v}}(v)dv \\ &= \int_v \frac{1}{\sigma(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sigma^{-2}z^T z} \ p_{\mathbf{v}}(v)dv \\ &= p_{\mathbf{e}}(z) \int_v p_{\mathbf{v}}(v)dv = p_{\mathbf{e}}(z). \end{aligned}$$

□