Pergamon

0005-1098(94)00099-9

Automatica, Vol. 31, No. 4, pp. 543–557, 1995
Elsevier Science Ltd
Printed in Great Britain
0005-1098/95 $9.50 + 0.00

Quantification of Uncertainty in Transfer Function Estimation: a Mixed Probabilistic–Worst-case Approach*

DOUWE K. DE VRIES† and PAUL M. J. VAN DEN HOF†

The identification of a linear nominal model and an additive model error bound is presented, where the error bound has both a deterministic component due to unmodelled dynamics, and a probabilistic component due to noise.

Key Words—System identification; transfer function estimation; model uncertainty; linear systems; error analysis; robust control.

Abstract—In this paper an identification problem is solved which is directed towards the use of the identified model as a basis for robust control design. A procedure is presented to identify, on the basis of time domain measurement data, a reduced order finite impulse response (FIR) model together with an upper bound on the model error of the corresponding transfer function, using only minor prior information. We assume the measurement data to be contaminated with a stochastic noise disturbance with unknown spectral properties. By applying a procedure similar to Bartlett's procedure of periodogram averaging, in conjunction with a periodic input signal, the statistics of the model error asymptotically can be obtained from the data. The model error consists of two parts: a probabilistic part, due to the stochastic noise disturbance, and a worst-case part, due to the unmodelled dynamics. The latter is explicitly bounded with a hard error bound, while for the former a confidence interval can be specified asymptotically. This enables an explicit trade-off between undermodelling (bias) and variance terms. The resulting error bound appears to be tight.

1. INTRODUCTION

This paper addresses the problem of identifying, on the basis of experimental data, an accurate transfer function model, together with a bound on the uncertainty that is present in this estimate.

*Received 20 December 1993; received in final form 20 June 1994. The original version of this paper was presented at the 12th IFAC World Congress which was held in Sydney, Australia, during 18–23 July 1993. The Published Proceedings of this IFAC Meeting may be ordered from: Elsevier Science Limited, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, U.K. This paper was recommended for publication in revised form by Associate Editor Bo Wahlberg under the direction of Editor Torsten Söderström. Corresponding author Dr Paul Van den Hof. Tel. +31 15 784509; Fax +31 15 784717; E-mail vdhof@tudw03.tudelft.nl.

† Mechanical Engineering Systems and Control Group, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

The identification of linear models together with uncertainty bounds is an important issue that has recently attracted a lot of attention. In general terms, an uncertainty bound on an estimated model is indispensible for assessing the quality of any model describing the dynamical system that underlies the measured data. However, recent interest in identifying models specifically directed towards the consecutive design of robust control systems, in particular, has been a stimulus for developing identification methods that provide model error bounds.

Early references addressing the problem of describing the uncertainty in a transfer function estimate include e.g. Jenkins and Watts (1968), Schweppe (1973) and Ljung and Caines (1979), whereas comprehensive accounts of either spectral analysis and parametric identification methods are given in Brillinger (1981) and Ljung (1987). The recent approaches to this estimation problem can be divided into two areas, characterized by the type of uncertainty bounds that arise: soft (probabilistic) or hard (worst-case) error bounds. The two different approaches are simply shown to be the result of different assumptions (priors) on the underlying data generating system; see e.g. Ljung *et al.* (1991) and Hjalmarsson (1993).

In view of robust control design, one would like to have available a (nominal) model together with a hard bound on the model error, measued in e.g. the $\mathcal{H}_\infty$-norm of the additive model error. Employing system identification methods, provision of such an uncertainty bound is achieved by using the prior that the available experimental data are contaminated by 'hard-bounded' disturbance signals, as e.g. an l_∞-bounded time or frequency domain signal. This type of prior implies that the identification procedure has to allow a worst-case disturbance on the data at any time instant. As a result, the uncertainty bounds that are obtained will be highly conservative if this worst-case disturbance on the data is not actually present. Examples of approaches to this problem set-up can be found e.g. in Helmicki *et al.* (1990), LaMaire *et al.* (1991), Gu and Khargonekar (1992), and De Vries and Van den Hof (1994), using the prior that the noise is bounded in the frequency domain, and Wahlberg and Ljung (1992), Hakvoort (1993) and many works in the area of bounded error and set-membership identification, such as e.g. Fogel and Huang (1982), Norton (1987), and Milanese and Vicino (1991), using the prior that the noise is bounded in the time domain.

Experiences with real-life measurements show that quite often it can very well be motivated to describe disturbance signals on experimental data as realizations of stochastic processes; see also Goodwin *et al.* (1992), Ninness and Goodwin (1994). One of the typical advantages of this description is that it allows the disturbance signal to be characterized as a random (indifferent) signal, which is contaminating the data but which is not worst-case at all time instants, i.e. it is not 'playing against' the experimenter. The assumption of stochastic noise disturbance on the data leads to identification procedures that provide soft (probabilistic) bounds on the model uncertainty; see e.g. Zhu (1989), Zhu and Backx (1991), Goodwin *et al.* (1992) and Bayard (1992).

The presence of noise disturbance on the experimental data is not the only cause of model uncertainty, as uncertainty is also implied by undermodelling and the effect of initial conditions on the data.

Our problem formulation will be the following. Let the data generating system be linear and time-invariant, generating data sequences according to

$$y(t) = G_0(q)u(t) + v(t), \tag{1}$$

with $v(t)$ a realization of a stationary zero-mean stochastic process with rational spectral density $\Phi_v(\omega)$, G_0 an exponentially stable proper transfer function and q the forward shift operator. We identify, on the basis of an experimental data sequence of input $u(t)$ and output $y(t)$, a finite-dimensional linear time-invariant model, together with a bound on the additive model uncertainty, where this bound distinguishes between the three sources of uncertainty: (1) undermodelling, (2) noise disturbance and (3) unknown initial conditions affecting the data.

The distinction between the different sources of uncertainty is necessary in order to obtain information *a posteriori* about how to improve the identification result and the error bound in order to meet required performance specifications.

Our approach to provide a solution to the problem mentioned above contains two characteristic features. The first one is that we will consider the model errors due to unmodelled dynamics and unknown past input signals as deterministic unknown-but-bounded (worst-case) quantities, whereas the noise disturbance is considered to be stochastic (averaging). As a result, the error bounds that we will derive will have both soft and hard components. That is, the error bounds will hold with a probability level that is bounded from below. The second feature is that we will employ a periodic input signal, which allows us to clearly distinguish between disturbance effects which are supposed to average out, and structural effects as unmodelled dynamics which remain unchanged when experiments are repeated.

The combination of hard and soft components in the model error bounds is the main deviation from existing methods, where either a worst-case approach is taken, leading to the previously mentioned conservativeness, or a stochastic approach is taken, which makes it complex to handle bias due to undermodelling.

The approach that we take consists of the following two steps. Employing a periodic input signal, we will first estimate an averaged (non-parametric) empirical transfer function estimate (ETFE), which is only defined in a finite number of frequency points, together with an asymptotic error bound that reflects the effects of unknown initial conditions and noise disturbance. The procedure followed is similar to Bartlett's procedure of periodogram averaging; see e.g. Oppenheim and Schafer (1975). The construction of this error bound requires only minor *a priori* information concerning the data generating process: an upper bound on past values of the input signal and on the system's pulse response. In the second step of the procedure we will fit a finite impulse response (FIR) model to the frequency domain estimates obtained in the first step. In this step a bias contribution due to undermodelling is introduced, but we can reduce the variance contribution (due to the noise) by a reduction in the number of parameters that has to be estimated from the data. An asymptotic bound on the error in the transfer function of the estimated FIR model will be established. This error bound can be separated into components which account for the different sources of uncertainty.

Comparing the above results to closely related previous work, we have the following remarks. The classical results, see e.g. Ljung (1985a, b, 1987), Ljung and Yuan (1985) and Brillinger (1981), specify the asymptotic distribution of the error for parametric identification methods and spectral analysis. Although these results are excellent for analysis, they are less fit to actually bound the error. It is difficult to provide a good quantification of the bias due to undermodelling c.q. windowing, while the bias can be considerable. Only implicit expressions for the bias are available, and the variance contributions contain the unknown noise spectrum.

The results given in Zhu (1989) and Zhu and Backx (1991) are in line with the classical results as stated above. The error bounds are obtained by neglecting the bias contribution and the fact that the variance is estimated.

In Goodwin *et al.* (1992), stochastic embedding is used for both the influence of the noise as well as for the error due to undermodelling. Both the distribution of the noise and the distribution of the undermodelling error are assumed to be known, up to a number of free parameters which are estimated from the data. This leads to non-asymptotic error bounds.

In Bayard (1992), the noise is assumed to be normally distributed, the noise filter is assumed to be known and the deviation from a steady-state situation is not taken into account. However, the results of Bayard (1992) are non-asymptotic.

The paradigm that we choose concerning the data generating system (1) definitely has its limitations. Robustness of our identification method with respect to slight deviations of the assumptions on linearity and time-invariance of the data generating system is examined elsewhere; see De Vries (1994).

The remainder of this paper is organized as follows. Section 2 contains the assumptions and *a priori* information. In Section 3 an error bound for the empirical transfer function estimate (ETFE) is formulated. Subsequently, in Section 4, a finite impulse response (FIR) model is fitted to the ETFE, and an error bound is established for the transfer function of this model using the results of Section 3. This error bound constitutes the main result of this paper. A simulation example is presented in Section 5, and Section 6 contains the conclusions.

2. PRELIMINARIES

The transfer function of the system G_0, see (1), can be written as

$$G_0(e^{j\omega}) = \sum_{k=0}^{\infty} g_0(k)e^{-j\omega k}, \tag{2}$$

with $g_0(k)$ the impulse response of the plant. We will consider scalar (single input, single output) systems. The output disturbance $v(t)$ is represented as

$$v(t) = H_0(q)e(t), \tag{3}$$

where $e(t)$ is a sequence of independent identically distributed random variables with zero-mean, variance σ_e^2 and all moments finite, and where H_0 is a stable proper transfer function.

We will denote the discrete time intervals for the measured input and output signals by the integer intervals $T^N = [0, N-1]$, $T^N_{N_s} = [N_s, N + N_s - 1]$ with N and N_s appropriate integers. We will partition the time interval $T^N_{N_s}$ with $N = rN_0$ in r time intervals of length N_0, denoting $T_i = T^{N_0}_{(i-1)N_0+N_s}$, $i = 1, \ldots, r$. The subscript i will indicate a variable that originates from the ith time interval T_i, e.g.

$$x_i(t) := x(t + (i-1)N_0 + N_s), \quad \text{where} \quad t \in T^{N_0}. \tag{4}$$

For a signal $x_i(t)$ we will denote the discrete Fourier transform (DFT) by

$$X_i(e^{j(2\pi k/N_0)}) := \frac{1}{\sqrt{N_0}} \sum_{t=0}^{N_0-1} x_i(t)e^{-j(2\pi k/N_0)t} \quad \text{for} \quad k \in T^{N_0}. \tag{5}$$

For a signal $x(t)$ being defined on $T^N_{N_s}$, we will denote

$$X^s(e^{j(2\pi k/N)}) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t + N_s)e^{-j(2\pi k/N)t} \quad \text{for} \quad k \in T^N. \tag{6}$$

Some specific sets of frequencies that arise in the DFT are denoted as

$$\Omega_{N_0} := \left\{\omega_k = \frac{2\pi k}{N_0},\ k = 0, 1, \ldots, N_0 - 1\right\}, \tag{7}$$

$$\Omega^{u_i}_{N_0} := \{\omega_k \in \Omega_{N_0} \mid |U_i(e^{j\omega_k})| \neq 0\}. \tag{8}$$

Finally, we will denote

$$\bar{u} = \max_{t \in T^{N+N_s}} |u(t)|.$$

Throughout this paper we will assume to have available the following *a priori* information on the past input signal and the system.

Assumption 1. We have as *a priori* information that

(i) there exists a finite and known $\bar{u}^p \in \mathbb{R}$, such that $|u(t)| \le \bar{u}^p$ for $t < 0$;
(ii) there exist finite and known M, $\rho \in \mathbb{R}$, $\rho > 1$, such that $|g_0(k)| \le M\rho^{-k}$, for $k \in \mathbb{N}$.

The *a priori* information on M and ρ need not be tight in the first instance. It can be improved using the measurement data, as will be discussed in Section 4.2. The *a priori* information on $\bar{u}^p$ is given by the actuator constraints.

3. ERROR BOUND FOR AVERAGED ETFE

3.1. *Introduction*

In this section we will construct a non-parametric estimate $\hat{G}$ of the system's transfer function G_0, by averaging over a set of empirical transfer function estimates (ETFEs). For this estimate, which is only defined at a finite number of frequency points, we will establish a confidence interval $\alpha(\omega_k)$ such that

$$|G_0(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \le \alpha(\omega_k)$$

with a prespecified probability. In order to achieve this, we will pursue the following strategy. Experimental data are available over a time set of length $N_s + N$. This time set is decomposed into a first interval of length N_s, not used for identification, and consecutively r intervals of length N_0. We consider an input signal that is periodic with period N_0, so that in each of the r intervals the same input signal is applied. Subsequently, an ETFE is made over each interval, and the final estimate $\hat{G}$ is obtained by averaging over these ETFEs. This procedure is similar to Bartlett's procedure of periodogram averaging, see Brillinger (1981) and Oppenheim and Schafer (1975), and is also proposed by Ljung (1985a). The choice of using a periodic input signal and averaging, and the significance of N_s, will be clarified at the end of Section 3.2. Because of the periodicity of the input signal $u(t)$, it follows that

$$\Omega_{N_0}^{u_i} = \Omega_{N_0}^{u} \quad \text{for all} \quad i = 1, \ldots, r.$$

3.2. *A non-parametric transfer function estimate*

We denote the following estimates:

$$\hat{G}_i(e^{j\omega_k}) := \frac{Y_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} \quad \text{for} \quad i = 1, 2, \ldots, r, \qquad \omega_k \in \Omega_{N_0}^u, \tag{9}$$

$$\hat{G}(e^{j\omega_k}) := \frac{1}{r}\sum_{i=1}^{r} \hat{G}_i(e^{j\omega_k}). \tag{10}$$

Employing the system's equations, similar to Ljung (1987), results in

$$\hat{G}_i(e^{j\omega_k}) = G_0(e^{j\omega_k}) + S_i(e^{j\omega_k}) + \frac{V_i(e^{j\omega_k})}{U_i(e^{j\omega_k})}, \tag{11}$$

where

$$S_i(e^{j\omega_k}) = \frac{R_i(e^{j\omega_k})}{U_i(e^{j\omega_k})}, \tag{12}$$

with $R_i(e^{j\omega_k})$ a component which is due to unknown past inputs, i.e. input samples preceding the time interval that is considered. We will split the analysis into two parts, separating bias and variance issues, by defining the auxiliary variables $\tilde{G}_i(e^{j\omega_k})$ and their average $\tilde{G}(e^{j\omega_k})$ as

$$\tilde{G}_i(e^{j\omega_k}) := \hat{G}_i(e^{j\omega_k}) - S_i(e^{j\omega_k}) = G_0(e^{j\omega_k}) + \frac{V_i(e^{j\omega_k})}{U_i(e^{j\omega_k})}, \tag{13}$$

$$\tilde{G}(e^{j\omega_k}) := \frac{1}{r}\sum_{i=1}^{r} \tilde{G}_i(e^{j\omega_k}) \quad \text{for} \quad \omega_k \in \Omega_{N_0}^u. \tag{14}$$

Using the triangle inequality we find

$$|G_0(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \le |G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})| + |S(e^{j\omega_k})| \tag{15}$$

with

$$S(e^{j\omega_k}) := \hat{G}(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k}) = \frac{1}{r}\sum_{i=1}^{r} S_i(e^{j\omega_k}). \tag{16}$$

Considering the inequality (15), the first term on the right hand side is the variance contribution to the error, which is due to the noise disturbance on the data. The second term is the bias contribution, which is due to unknown past inputs. A confidence interval for the first term has to be determined on the basis of its distribution, whereas a hard bound will be derived for the second term.

For the bias contributions $S(e^{j\omega_k})$ and $S_i(e^{j\omega_k})$, which will be of interest in the sequel of the paper, we have the following results.

Lemma 1. Consider the experimental set-up as presented above. Let the input signal be periodic with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$. Then, for all $\omega_k \in \Omega_{N_0}^u$ and $i = 1, \ldots, r$,

$$|S_i(e^{j\omega_k})| \le \frac{1}{\sqrt{N_0}} \frac{\bar{u}^P + \bar{u}}{|U_i(e^{j\omega_k})|} \times \frac{M\rho(1-\rho^{-N_0})}{(\rho-1)^2} \rho^{-(i-1)N_0 - N_s}, \tag{17}$$

$$|S(e^{j\omega_k})| \le \frac{1}{r\sqrt{N_0}} \frac{\bar{u}^P + \bar{u}}{|U_i(e^{j\omega_k})|} \times \frac{M\rho(1-\rho^{-N})}{(\rho-1)^2} \rho^{-N_s}. \tag{18}$$

□

Proof. See Appendix A1. □

Upper bounds on the bias contribution $S(e^{j\omega_k})$ in situations of general non-periodic input signals can be found in Ljung (1987) and LaMaire *et al.* (1991).

Establishing the distribution of the variance contribution $|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|$ is more involved, and only results are available that are asymptotic in N_0. Because the difference $G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})$ is entirely due to the noise, see (14) and (13), a first essential step is to find the asymptotic distribution of the DFT of the noise. This has been achieved by Brillinger (1981).

Theorem 1 (Brillinger, 1981). Consider $v_i(t)$ as defined in (3) and (4), and let $V_i(e^{j\omega_k})$ be the DFT of $v_i(t)$ as defined in (5) with $\omega_k \in \Omega_{N_0}$. Let

$$\check{V}_{N_0} := \begin{bmatrix} \mathrm{Re}\{V_i(e^{j\omega_k})\} \\ \mathrm{Im}\{V_i(e^{j\omega_k})\} \\ \mathrm{Re}\{V_i(e^{j\omega_l})\} \\ \mathrm{Im}\{V_i(e^{j\omega_l})\} \\ \mathrm{Re}\{V_m(e^{j\omega_l})\} \\ \mathrm{Im}\{V_m(e^{j\omega_l})\} \end{bmatrix},$$

where $\omega_k, \omega_l \in \Omega_{N_0}$, $\omega_k \neq \omega_l$, and $i, m = 1, \ldots, r$, $i \neq m$. Then as $N_0 \to \infty$

$$\check{V}_{N_0} \xrightarrow{d} \mathcal{N}(0, \Lambda),$$

meaning that $\check{V}_{N_0}$ converges in distribution to the normal distribution with zero mean and covariance matrix Λ, where Λ is a diagonal matrix with diagonal elements given by

$$\begin{aligned} \mathrm{var}[\mathrm{Re}\{V_q(e^{j\omega_p})\}] &= \mathrm{var}[\mathrm{Im}\{V_q(e^{j\omega_p})\}] \\ &= \tfrac{1}{2}\sigma_e^2 |H_0(e^{j\omega_p})|^2 \qquad \text{for} \quad \omega_p \neq 0, \pi \\ \mathrm{var}[\mathrm{Re}\{V_q(e^{j\omega_p})\}] &= \sigma_e^2 |H_0(e^{j\omega_p})|^2 \qquad \text{for} \quad \omega_p = 0, \pi \\ \mathrm{var}[\mathrm{Im}\{V_q(e^{j\omega_p})\}] &= 0 \qquad \text{for} \quad \omega_p = 0, \pi \end{aligned}$$

for $\omega_p \in \Omega_{N_0}$ and $q = 1, \ldots, r$. □

Note that Theorem 1 states that the elements of the vector $\check{V}_{N_0}$ are asymptotically uncorrelated and jointly normally distributed, which implies that they are asymptotically independent; see e.g. Priestley (1981). Because U_i is independent of i due to the periodic input, it follows from Theorem 1 and (13) that asymptotically the auxiliary variables $\tilde{G}_i(e^{j\omega_k})$ are independent and identically normally distributed. Considering (13) it now follows that $\tilde{G}(e^{j\omega_k})$ is asymptotically normally distributed with expectation $G_0(e^{j\omega_k})$. However, the variance of $\tilde{G}(e^{j\omega_k})$ is still unknown. We will quantify this variance on the basis of measurement data. To this end we consider the estimate

$$\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k})) := \frac{1}{r(r-1)} \sum_{i=1}^{r} |\tilde{G}(e^{j\omega_k}) - \tilde{G}_i(e^{j\omega_k})|^2. \quad (19)$$

Although this estimate is not available from data—as the auxiliary $\tilde{G}(e^{j\omega_k})$ and $\tilde{G}_i(e^{j\omega_k})$ are unknown—we can bound its difference with the estimate

$$\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k})) := \frac{1}{r(r-1)} \sum_{i=1}^{r} |\hat{G}(e^{j\omega_k}) - \hat{G}_i(e^{j\omega_k})|^2 \quad (20)$$

as formulated in the following lemma.

Lemma 2. Consider the estimates $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))$, $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$ as defined in (20), (19). Let $u(t)$ be a periodic input signal with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$. Then

$$|\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k})) - \hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))| \leq v(\omega_k)$$

with

$$v(\omega_k) = \frac{1}{r(r-1)} \sum_{i=1}^{r} (2\,|A_i(e^{j\omega_k})|\,|B_i(e^{j\omega_k})| + |B_i(e^{j\omega_k})|^2) \quad (21)$$

and

$$|A_i(e^{j\omega_k})| = |\hat{G}(e^{j\omega_k}) - \hat{G}_i(e^{j\omega_k})|$$

$$|B_i(e^{j\omega_k})| = \frac{r-2}{r}|S_i(e^{j\omega_k})| + \frac{1}{r}\sum_{m=1}^{r} |S_m(e^{j\omega_k})|,$$

where $S_i(e^{j\omega_k})$, $S_m(e^{j\omega_k})$ are bounded by (17). □

Proof. See Appendix A2. □

Note that the difference between the random variables (20) and (19) contains known realizations of random variables only, i.e. realizations of $\hat{G}_i$ and $\hat{G}$, and therefore can be bounded with a hard error bound.

Using Theorem 1 and the auxiliary estimate $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$, the distribution of the difference between the auxiliary transfer function $\tilde{G}(e^{j\omega_k})$ and the system's transfer function $G_0(e^{j\omega_k})$ can now be completely specified asymptotically in N_0.

Lemma 3. Consider the auxiliary transfer function $\tilde{G}(e^{j\omega_k})$, (14), (13) and the estimate of its variance (19). Let the input signal be periodic with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$, and let $r > 1$. Then as $N_0 \to \infty$

$$\frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} \xrightarrow{d} \begin{cases} F(2, 2(r-1)) & \text{for } \omega_k \neq 0, \pi \\ F(1, r-1) & \text{for } \omega_k = 0, \pi, \end{cases}$$

where $F(n, d)$ denotes the F distribution with n degrees of freedom in the numerator and d degrees of freedom in the denominator. □

Proof. See Appendix A3. □

Combining Lemmas 1, 2 and 3 and (15) leads to an error bound in terms of a confidence interval that can be calculated on the basis of measurement data. In formulating this confidence interval, we adopt the following notation:

$$F_\alpha(n, d) = \{P[x \leq \alpha], x \in F(n, d)\},$$

meaning that a random variable x which is distributed as $F(n, d)$ is smaller than α with probability (w.p.) given by $F_\alpha(n, d)$.

Theorem 2. Consider the transfer function $\hat{G}(e^{j\omega_k})$, (10), (9) and the estimate of its variance (20). Let the input signal be periodic with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$, and let $r > 1$. Then asymptotically in N_0,

$$|G_0(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq |S(e^{j\omega_k})| + \gamma_\alpha(\omega_k)$$
$$\text{w.p.} \geq \begin{cases} F_\alpha(2, 2(r-1)) & \omega_k \neq 0, \pi \\ F_\alpha(1, r-1) & \omega_k = 0, \pi, \end{cases}$$

where

$$\gamma_\alpha(\omega_k) = \sqrt{\alpha}\,(\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k})) + v(\omega_k))^{1/2}$$

while $v(\omega_k)$ is given by (21), and $S(e^{j\omega_k})$ is bounded by (18). □

Proof. See Appendix A4. □

Further analysis of the properties of the estimates being used gives rise to the following remarks.

- The estimate $\hat{G}(e^{j\omega_k})$ is a consistent estimate of $G_0(e^{j\omega_k})$; its bias decays as $\rho^{-N_s}/\sqrt{rN}$, see (18), and its variance decays as $1/r$, see (23). Asymptotically in N_0 the estimate $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))$ is a consistent estimate in r of $\mathrm{var}\,[\hat{G}(e^{j\omega_k})]$; asymptotically in N_0 the estimate is unbiased and its variance decays as $1/(r-1)$, basically following from (A.6) and (A.5) in Appendix A3. Thus, for the components of the error bound given in Theorem 2 there holds

$$\begin{aligned} |S(e^{j\omega_k})| &\to 0 \quad \text{as} \quad \rho^{-N_s}/\sqrt{rN}, \\ v(\omega_k) &\to 0 \quad \text{as} \quad \rho^{-N_s}/(r\sqrt{rN}), \\ \hat{\sigma}_r^2(\hat{G}(e^{j\omega_k})) &\to \mathrm{var}\,[\hat{G}(e^{j\omega_k})] \\ &\text{as } 1/(r-1) \text{ asymptotically in } N_0, \\ \mathrm{var}\,[\hat{G}(e^{j\omega_k})] &\to 0 \quad \text{as} \quad /r. \end{aligned} \tag{22}$$

- The variance of $\hat{G}(e^{j\omega_k})$ is given as the noise to signal ratio in the frequency domain

$$\mathrm{var}\,[\hat{G}(e^{j\omega_k})] = \frac{\mathrm{var}\,[V^s(e^{j\omega_k})]}{|U^s(e^{j\omega_k})|^2} = \frac{1}{r}\frac{\mathrm{var}\,[V^s(e^{j\omega_k})]}{|U_i(e^{j\omega_k})|^2}, \tag{23}$$

see (36), (10), (9) and (6), where

$$\mathrm{var}\,[V^s(e^{j\omega_k})] \to \Phi_v(\omega_k) = \sigma_e^2\,|H_0(e^{j\omega_k})|^2 \quad \text{as} \quad 1/N;$$

see Ljung (1985a).
- The estimate $\hat{G}(e^{j\omega_k})$ is only defined at the finite number of frequency points in $\Omega^u_{N_0}$.

Summarizing, the periodic input yields an estimate having a variance that quickly decays with the number of averages, and enables a considerable reduction of the bias due to the unknown past input signals by using N_s. Furthermore, the periodic input offers the opportunity to completely characterize the asymptotic distribution of the estimate, i.e. including the fact that an estimate of the variance is used, without relying heavily on *a priori* knowledge of the noise; see Lemma 3.

The advantage of averaging (10), in conjunction with a periodic input signal, over windowing techniques which average over neighbouring frequency points is that an estimate of the true transfer function with decaying variance is obtained without introducing bias due to windowing. However, averaging reduces the frequency resolution: an estimate is obtained at only N_0 instead of N frequency points.

Finally, note that Theorem 2 can very well be used to provide an estimate of the information that is assumed to be available as *a priori* information in many algorithms for identification in $\mathcal{H}_\infty$ based on frequency domain data; see e.g. Helmicki *et al.*, (1990). In these algorithms a model and a hard (worst-case) error bound are obtained in the $\mathcal{H}_\infty$-space. Note, however, that the error bound formulated in Theorem 2 is a soft (probabilistic) one, while the bound needed in Helmicki *et al.* (1990) is a hard (worst-case) one.

3.3. *Input design*

In view of experiment design, the error bound of Theorem 2 has the following implications. The 'bias' contributions $|S(e^{j\omega_k})|$ and $v(\omega_k)$ can be reduced most effectively by increasing N_s; see (18). Considering the estimated variance $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))$ it follows from (22) and (23) that $|U_i(e^{j\omega_k})|$ should not be too small at any frequency $\omega_k \in \Omega^u_{N_0}$. To arrive at an error bound which is less than some prespecified frequency domain bound $K(\omega_k)$, resulting e.g. from the robustness properties of a controller that has been designed for the system, the input $|U_i(e^{j\omega_k})|$ should be chosen to be proportional to $\sqrt{\mathrm{var}\,[V^s(e^{j\omega_k})]}/K(\omega_k)$ for all $\omega_k \in \Omega^u_{N_0}$. Such an input, also having a small time domain magnitude $\bar{u}$, can be designed using the results of e.g. Schroeder (1970) and Schoukens and Pintelon (1991). Filtered white noise or pseudo-random binary sequences (PRBS) are not adequate input signals: $|U_i(e^{j\omega_k})|^2$ will be an erratic function of frequency, resulting in an erratic estimate and error bound.

4. ERROR BOUND FOR FIR ESTIMATE

4.1. *Introduction*

In the previous section we have obtained a sequence of non-parametric estimates $\hat{G}_i(e^{j\omega_k})$, $\omega_k \in \Omega^u_{N_0}$, $i = 1, \ldots, r$, together with a bound for the model error in $\hat{G}(e^{j\omega_k})$. This implies that we only have model information in a finite number of points on the frequency axis. In this section we will use the frequency domain data, as obtained in the previous section, to provide a reduced order parametric model, together with a model error bound that is continuous on the frequency interval $\omega \in [0, 2\pi)$.

The parametric identification introduces bias due to undermodelling. On the other hand, the variance can be reduced due to the restricted number of estimated parameters. Hence, by choosing the model order, bias can be traded against variance, allowing a lower error bound. We will derive explicit bounds for the bias and variance errors, enabling a clear trade-off between these terms.

We focus on parametric methods instead of spectral analysis techniques (windowing) to improve our results, since Ljung (1985a) has shown that parametric methods allow for a faster convergence to zero of the mean square model error.

For the parametric identification we will use the FIR model structure. First, we will estimate the parameters of an FIR model in the frequency domain, by using the set of non-parametric frequency domain estimates $\hat{G}_i(e^{j\omega_k})$, and an error bound for the estimated prameters will be derived. Based on these results, an error bound will be established for the corresponding transfer function of the model.

The presentation will be short; the methods used are similar to those of Section 3.

4.2. *Parameter estimate*

We will consider an FIR model structure, determined by

$$G(e^{j\omega}, \theta) = \sum_{l=0}^{n_p-1} g(l)e^{-j\omega l} = \phi(e^{j\omega})\theta,$$

where

$$\theta = [g(0) \quad g(1) \quad \ldots \quad g(n_p - 1)]^T,$$
$$\phi(e^{j\omega}) = [1 \quad e^{-j\omega} \quad \ldots \quad e^{-j(n_p-1)\omega}]. \tag{24}$$

Now for each time interval $T_i, i = 1, \ldots, r$, a corresponding FIR parameter estimate is obtained by the following weighted least squares criterion:

$$\hat{\theta}_i := \arg\min_{\theta} \sum_{k=1}^{N_p} |w_k[\hat{G}_i(e^{j\zeta_k}) - \phi(e^{j\zeta_k})\theta]|^2, \tag{25}$$

where θ ranges over an appropriate parameter space $\Theta \subset \mathbb{R}^{n_p}$, and $w_k \in \mathbb{C}$ are weighting factors. The frequency points $\{\zeta_k\}_{k=1,\ldots,N_p}$ constitute a set $\Omega^p \subseteq \Omega^u_{N_0}$, satisfying that $e^{j\zeta_k}$ comes in complex conjugate pairs.

Similar to the previous section, the final estimate is obtained by averaging over the different sections of the data:

$$\hat{\theta} := \frac{1}{r}\sum_{i=1}^{r} \hat{\theta}_i. \tag{26}$$

Before analysing these parameter estimates, we have to introduce some additional notation. As an alternative expression for (25) we can write

$$\hat{\theta}_i = \Psi\hat{\mathbf{G}}_i, \tag{27}$$

with

$$\Psi = (\Phi^*W^*W\Phi)^{-1}\Phi^*W^*W,$$
$$\Phi = [\phi^T(e^{j\zeta_1}) \quad \phi^T(e^{j\zeta_2}) \quad \ldots \quad \phi^T(e^{j\zeta_{N_p}})]^T,$$
$$W = \mathrm{diag}\,(w_1, \ldots, w_{N_p}),$$
$$\hat{\mathbf{G}}_i = [\hat{G}_i(e^{j\zeta_1}) \quad \hat{G}_i(e^{j\zeta_2}) \quad \ldots \quad \hat{G}_i(e^{j\zeta_{N_p}})]^T,$$

where an asterisk denotes the complex conjugate transpose. Furthermore, let $x\langle m\rangle$ denote the mth element of the vector x and $X\langle m, i\rangle$ the (m, i)th element of the matrix X. We will use the following estimate for the variance of $\hat{\theta}\langle k\rangle$:

$$\hat{\sigma}_r^2(\hat{\theta}\langle k\rangle) := \frac{1}{r(r-1)}\sum_{i=1}^{r} |\hat{\theta}\langle k\rangle - \hat{\theta}_i\langle k\rangle|^2. \tag{28}$$

To separate bias and variance issues, similar to the technique as presented in the previous section, see (13), we introduce the auxiliary variables $\tilde{\theta}_i$ and $\tilde{\theta}$,

$$\tilde{\theta}_i := \theta_0 + \Psi \mathbf{F}_i,$$

$$\tilde{\theta} := \frac{1}{r}\sum_{i=1}^{r} \tilde{\theta}_i,$$

with

$$\theta_0 = [g_0(0) \quad g_0(1) \quad \ldots \quad g_0(n_p - 1)]^{\mathrm{T}}, \tag{29}$$

$$\mathbf{F}_i = \left[\frac{V_i(\mathrm{e}^{\mathrm{j}\zeta_1})}{U_i(\mathrm{e}^{\mathrm{j}\zeta_1})} \quad \frac{V_i(\mathrm{e}^{\mathrm{j}\zeta_2})}{U_i(\mathrm{e}^{\mathrm{j}\zeta_2})} \quad \ldots \quad \frac{V_i(\mathrm{e}^{\mathrm{j}\zeta_{N_p}})}{U_i(\mathrm{e}^{\mathrm{j}\zeta_{N_p}})}\right]^{\mathrm{T}}.$$

In line with Lemma 2, we can now bound the difference between the estimated variances of the parameter estimate $\hat{\theta}$ and the unknown auxiliary variable $\tilde{\theta}$, as formulated in the following lemma.

Lemma 4. Consider the estimates $\hat{\sigma}_r^2(\hat{\theta}\langle k\rangle)$, $\hat{\sigma}_r^2(\tilde{\theta}\langle k\rangle)$ as defined in (28). Let the input signal be periodic with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$. Then

$$|\hat{\sigma}_r^2(\tilde{\theta}\langle k\rangle) - \hat{\sigma}_r^2(\hat{\theta}\langle k\rangle)| \le v^p\langle k\rangle$$

with

$$v^p\langle k\rangle = \frac{1}{r(r-1)}\sum_{i=1}^{r}(2\,|a_i^p\langle k\rangle|\,|b_i^p\langle k\rangle| + |b_i^p\langle k\rangle|^2) \tag{30}$$

and

$$|a_i^p\langle k\rangle| = |\hat{\theta}\langle k\rangle - \hat{\theta}_i\langle k\rangle|$$

$$|b_i^p\langle k\rangle| \le \frac{1}{r}\sum_{m=1}^{r}|s_m^p\langle k\rangle + \frac{r-2}{r}|s_i^p\langle k\rangle|$$

$$|s_i^p\langle k\rangle| \le \sum_{m=1}^{N_p}|\Psi\langle k, m\rangle|\,|S_i(\mathrm{e}^{\mathrm{j}\zeta_m})|,$$

where $S_i(\mathrm{e}^{\mathrm{j}\zeta_m})$ is bounded by (17). □

Proof. The proof is similar to the proof of Lemma 2, and is skipped for brevity. □

We are now ready to formulate an error bound on the parameter estimate.

Theorem 3. Consider the estimated parameters $\hat{\theta}$, (26), (27) and the estimated variance (28). Let the input signal be periodic with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$, let $r > 1$ and $n_p \le N_p \le N_0$. Then, asymptotically in N_0,

$$|\theta_0\langle k\rangle - \hat{\theta}\langle k\rangle| \le |s^p\langle k\rangle| + |z^p\langle k\rangle| + \gamma_\alpha^p\langle k\rangle$$
$$\text{w.p.} \ge F_\alpha(1, r-1) \tag{31}$$

where

$$|s^p\langle k\rangle| \le \sum_{l=1}^{N_p}|\Psi\langle k, l\rangle|\,|S(\mathrm{e}^{\mathrm{j}\zeta_l})|$$

$$|z^p\langle k\rangle| \le M\sum_{m=n_p}^{n_h-1}\rho^{-m}\left|\sum_{l=1}^{N_p}\Psi\langle k, l\rangle \mathrm{e}^{-\mathrm{j}\zeta_l m}\right| + \frac{M\rho}{\rho-1}\rho^{-n_h}\sum_{l=1}^{N_p}|\Psi\langle k, l\rangle|$$

$$\gamma_\alpha^p\langle k\rangle \le \sqrt{\alpha}\,(\hat{\sigma}_r^2(\hat{\theta}\langle k\rangle) + v^p\langle k\rangle)^{1/2}$$

for any $n_h \ge n_p$, while $v^p\langle k\rangle$ is given by (30), and $S(\mathrm{e}^{\mathrm{j}\zeta_l})$ is bounded by (18). □

Proof. See Appendix A5. □

The bound on the parameter error formulated in (31) consists of three separate terms. The first term reflects the error due to unknown initial conditions, the second term refers to the effect of unmodelled dynamics, whereas the third (probabilistic) term represents the fact that the measurements are contaminated by noise.

Comparing the above result with the results for parametric identification as obtained by Ljung (1985a, 1987, Appendix II.2), we have the following three remarks. Firstly, we have obtained an expression for the distribution of the error that can be calculated from the data, also if the noise is coloured. Secondly, the bias due to undermodelling has been explicitly bounded. Finally, a consistent estimate of the covariance matrix of the non-parametric estimate $\hat{G}(\mathrm{e}^{\mathrm{j}\omega_k})$ is available, and this estimate is independent of the parametric model. A good approximation of the best linear unbiased estimate (BLUE) can therefore be obtained. This latter statement can be understood by realizing that a standard result from statistics is that the variance of the estimated parameters $\hat{\theta}$ is minimized by choosing the matrix W^*W to be proportional to the inverse covariance matrix of the estimate $\hat{G}(\mathrm{e}^{\mathrm{j}\omega_k})$. Asymptotically in N this covariance matrix is diagonal; see (23) and Theorem 1. Therefore, a good choice for W is

$$W = \mathrm{diag}\,(1/\hat{\sigma}_r(\hat{G}(\mathrm{e}^{\mathrm{j}\zeta_1})), \ldots, 1/\hat{\sigma}_r(\hat{G}(\mathrm{e}^{\mathrm{j}\zeta_{N_p}}))).$$

The weighting matrix W of course also can be used to affect the bias distribution over frequency of the estimated FIR model; see Ljung (1985a).

Because the *a priori* information M and ρ on the impulse response of the system largely determines the bounds on the errors due to unknown past inputs $s^p\langle k\rangle$ and undermodelling $z^p\langle k\rangle$, a correct and not overly conservative choice for M and ρ is of major importance. This also holds for the frequency domain error bound on the transfer function of the estimated FIR model as given in Theorem 4 in the next section. However, obtaining the *a priori* information M and ρ is largely an open issue in model uncertainty estimation. Helmicki *et al.* (1993) state that in practical applications such *a priori* information typically is obtained through some 'engineering leap of faith'. An important contribution of Theorem 3 is that it provides a simple and effective way to validate the prior information M and ρ: the prior information should be consistent with the resulting error bound on the estimated parameters. This implies that the prior information M and ρ can be improved iteratively by adapting M and ρ to match the resulting error bounds on the estimated parameters. Clearly, such a procedure gives no guarantee that the prior information given by M and ρ is correct, but it significantly reduces the 'engineering leap of faith' involved.

Remark 1. A similar procedure is possible when the parameters of the FIR model are estimated directly from the time domain data, since the estimated parameters are again asymptotically normally distributed; see Ljung and Caines (1979) and Ljung (1987). The time domain approach, however, leads to algorithms for the computation of the error bounds that are considerably more demanding.

4.3. *Transfer function estimate*

Using the estimated FIR model, an estimate of the transfer function with an error bound can be obtained for any frequency $\omega \in [0, 2\pi)$. The estimate of the transfer function at ω is, for any $\omega \in [0, 2\pi)$, defined as

$$\hat{G}_i^f(e^{j\omega}) := P(e^{j\omega})[\mathcal{W}\ 0]\hat{\theta}_i, \tag{32}$$

where

$$P(e^{j\omega}) = [1\ e^{-j\omega}\ \ldots\ e^{-j(n_f-1)\omega}]$$

and $\mathcal{W}$ is an $n_f \times n_f$ weighting matrix. This weighting matrix, which will be discussed in more detail later on, can be seen as a regularization of the estimated parameters towards agreement with a prior knowledge as e.g. an exponentially decaying impulse response.

Let $\hat{G}^f(e^{j\omega})$ and $\hat{\sigma}_r^2(\hat{G}^f(e^{j\omega}))$ denote the averaged estimate and the estimated variance, respectively, with the usual definitions. Define the auxiliary variables

$$\tilde{G}_i^f(e^{j\omega}) := G_0(e^{j\omega}) + \Upsilon\mathbf{F}_i,$$

where

$$\Upsilon(e^{j\omega}) = P(e^{j\omega})[\mathcal{W}\ 0]\Psi. \tag{33}$$

Because the real and imaginary parts of the averaged auxiliary variable $\tilde{G}^f(e^{j\omega})$ are not independent (as was the case for $\tilde{G}(e^{j\omega_k})$ in Section 3.2), we will give separate confidence intervals for the real and imaginary part of the error.

Lemma 5. Let the input signal be periodic with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$. Then

$$|\hat{\sigma}_r^2(\mathrm{Re}\{\tilde{G}^f(e^{j\omega})\}) - \hat{\sigma}_r^2(\mathrm{Re}\{\hat{G}^f(e^{j\omega})\})| \le v^f(\omega)$$

with

$$v^f(\omega) = \frac{1}{r(r-1)}\sum_{i=1}^{r}(2\,|A_i^f(e^{j\omega})|\,|B_i^f(e^{j\omega})| + |B_i^f(e^{j\omega})|^2) \tag{34}$$

and

$$|A_i^f(e^{j\omega})| = |\hat{G}^f(e^{j\omega}) - \hat{G}_i^f(e^{j\omega})|$$

$$|B_i^f(e^{j\omega})| \le \frac{1}{r}\sum_{m=1}^{r}|S_m^f(e^{j\omega})| + \frac{r-2}{r}|S_i^f(e^{j\omega})|$$

$$|S_i^f(e^{j\omega})| \le \sum_{m=1}^{N_p}|\Upsilon\langle m\rangle|\,|S_i(e^{j\zeta_m})|$$

where $S_i(e^{j\zeta_m})$ is bounded by (17). □

Proof. Similar to the proof of Lemma 2. □

Theorem 4. Let the input signal be periodic with period N_0 for $t \in T^{N+N_s}$, with $N = rN_0$, let $r > 1$ and $n_f \le n_p \le N_p \le N_0$. Then, asymptotically in N_0,

$$\begin{aligned}&|\mathrm{Re}\{G_0(e^{j\omega}) - \hat{G}^f(e^{j\omega})\}|\\ &\quad\le |S^f(e^{j\omega})| + |\mathrm{Re}\{\Lambda(e^{j\omega}) - Z^f(e^{j\omega})\}|\\ &\qquad + |\mathrm{Re}\{\Gamma(e^{j\omega})\}| + \gamma_\alpha^f(\omega)\quad \text{w.p.} \ge F_\alpha(1, r-1)\end{aligned}$$

where

$$|S^f(e^{j\omega})| \le \sum_{l=1}^{N_p}|\Upsilon\langle l\rangle|\,|S(e^{j\zeta_l})|$$

$$\begin{aligned}&|\mathrm{Re}\{\Lambda(e^{j\omega}) - Z^f(e^{j\omega})\}|\\ &\quad\le M\sum_{m=n_f}^{n_p-1}\rho^{-m}\,|\mathrm{Re}\{e^{-j\omega m}\}|\\ &\qquad + M\sum_{m=n_p}^{n_h-1}\rho^{-m}\left|\mathrm{Re}\left\{e^{-j\omega m} - \sum_{l=1}^{N_p}\Upsilon\langle l\rangle e^{-j\zeta_l m}\right\}\right|\\ &\qquad + \frac{M\rho}{\rho-1}\rho^{-n_h}\left(1 + \sum_{l=1}^{N_p}|\Upsilon\langle l\rangle|\right)\end{aligned}$$

$$\begin{aligned}&|\mathrm{Re}\{\Gamma(e^{j\omega})\}|\\ &\quad\le M\sum_{l=0}^{n_f-1}\Big(\rho^{-l}\,|1 - \mathcal{W}\langle l+1, l+1\rangle|\\ &\qquad + \sum_{m=0}^{l-1}\rho^{-m}\,|\mathcal{W}\langle l+1, m+1\rangle|\\ &\qquad + \sum_{m=l+1}^{n_f-1}\rho^{-m}\,|\mathcal{W}\langle l+1, m+1\rangle|\Big)\,|\mathrm{Re}\{e^{-j\omega l}\}|\end{aligned}$$

$$\gamma_\alpha^f(\omega) \le \sqrt{\alpha}\,(\hat{\sigma}_r^2(\mathrm{Re}\{\hat{G}^f(e^{j\omega})\}) + v^f(\omega))^{1/2}$$

for any $n_h \geq n_p$, while $\nu^f(\omega)$ is given by (34), and $S(e^{j\zeta_i})$ is bounded by (18). □

Proof. See Appendix A6. □

Replacing Re$\{\cdot\}$ in Lemma 5 and Theorem 4 with Im$\{\cdot\}$ gives a confidence interval for the imaginary part of the error. In Theorem 4 S^f and ν^f are the errors due to the unknown past inputs, Z^f and Λ are the errors due to undermodelling, and Γ is the error due to weighting with $\mathcal{W}$. For these error components, hard bounds are provided. A probabilistic bound is given for the error due to the noise γ^f_α.

If $\mathcal{W}$ is diagonal, the last two terms in the upper bound for $\Gamma(e^{j\omega})$ as given in Theorem 4 are zero. A good choice for the weighting matrix $\mathcal{W}$ is a diagonal matrix having its first diagonal elements equal to one, and decreasing thereafter (using e.g. a cosine taper on an exponential with base ρ). By choosing the point at which the elements start to decay appropriately, the variance can be reduced significantly, while relatively little bias is introduced. Usually, a good choice for this point is the point at which the confidence interval for the estimated impulse response (Theorem 3) becomes larger than the interval given by the prior information M and ρ. For such a choice, the influence of $\mathcal{W}$ can be seen as a regularization of the estimated parameters towards zero, thus using the prior knowledge that the impulse response converges to zero, which reduces the sensitivity to the actual choice for the order n_f; see Ljung *et al.* (1993).

Note that we have obtained explicit and separate bounds for the errors due to noise, undermodelling, weighting and unknown past inputs, providing valuable insight into the effective ways to reduce the error bound by input design and by choosing the orders, the weighting functions and N_s. Note also that the procedure allows for non-uniformly spaced frequency points $\zeta_k \in \Omega^p$.

The result given in Theorem 4 does not directly provide a specification of a region in the complex plane which, with a certain probability, contains the true transfer function. A straightforward solution to this problem of specifying the simultaneous probability of two dependent random variables follows from the Bonferroni inequality, which reduces a simultaneous probability statement to individual probability statements:

$$P[x_i < a_i, i = 1, \ldots, m] \geq 1 - \sum_{i=1}^{m} (1 - P[x_i < a_i]). \quad (35)$$

Thus, we can specify a lower bound on the probability that the true transfer function lies in the rectangle in the complex plane specified by using Theorem 4 for the real and imaginary parts of the error.

Remark 2. Alternatively, a joint confidence interval for the real and imaginary part of the error can be derived, using Hotelling's T^2-statistic, see e.g. Johnson and Wichern (1988), as is done in De Vries (1993). This results in an ellipsoidal confidence region in the complex plane, which provides more information as to the direction of the uncertainty. □

The procedure above has some clear resemblances with that of Gu and Khargonekar (1992), where an estimated impulse response is obtained by applying the inverse DFT to a non-parametric estimate of the transfer function at a finite number of frequency points. This estimated impulse response is truncated and weighted to provide the necessary reduction of the influence of the noise, and an estimate of the transfer function with an error bound is obtained. The main difference is that in Gu and Khargonekar (1992), the problem is formulated entirely in the unknown-but-bounded setting, leading to hard (worst-case) error bounds, which in general will be rather conservative.

Comparing the result of Theorem 4 with the frequency domain results given in Ljung (1985a, b, 1987) for the asymptotic bias and variance of parametric transfer function estimates, we have the following remarks. In contrast to the results of Ljung (1985a, b, 1987), Theorem 4 holds for finite model orders n_p and n_f, and we have explicit expressions for the various bias contributions. Also note that we use an estimate of the variance, rather than the unknown true spectrum of the noise.

Finally, we have derived explicit bounds for the bias and variance errors, enabling a clear trade-off between these terms by e.g. selecting the model order. If bias errors are not taken into account, as in Zhu (1989) and Zhu and Backx (1991), or if no explicit error bounds are available, as in Ljung (1985a, 1987) and Brillinger (1981), such a trade-off cannot be made.

Remark 3. The analysis presented in this and the previous sections can be extended to regression schemes other than the FIR-type of regression as discussed here. This includes the use of more general orthogonal basis functions such as Laguerre or Kautz functions, as well as generalizations thereof, see e.g. Heuberger *et al.* (1992, 1993). An extended analysis of this situation is presented in De Vries (1993, 1994).

5. EXAMPLE

To illustrate our results, a simulation was made of a fifth order system:

$$G_0(z) = \frac{0.7027 - 0.8926z^{-1} + 0.2400z^{-2} + 0.5243z^{-3} - 0.9023z^{-4} + 0.4009z^{-5}}{1 - 2.4741z^{-1} + 2.8913z^{-2} - 1.9813z^{-3} + 0.8337z^{-4} - 0.1813z^{-5}},$$

whose impulse response $g_0(k)$ satisfies a bound given by $M_0 = 2$ and $\rho_0 = 1.23$. The input signal was chosen to obey $\bar{u}^p = 2$ and $\bar{u} = 1$. We used 1100 data points with $N = 1024$, $N_0 = 128$ and $N_s = 76$. There was 10% (standard deviation) coloured noise $v(t) = H_0(q)e(t)$ added to the output. For the noise filter we used a third order highpass filter:

$$H_0(z) = \frac{0.7184 - 0.2206z^{-1} + 0.2390z^{-2} - 0.0060z^{-3}}{1 + 0.1177z^{-1} + 0.3208z^{-2} - 0.0182z^{-3}}$$

and $e(t)$ was chosen to be a *uniformly* distributed random sequence. For the input signal we used a random-phased multi-sine, with phases uniformly distributed in the interval $[0, 2\pi)$. The magnitude of the DFT of the input signal over one period is given in Fig. 1. Note that the frequency points where $|U_i(e^{j\omega_k})| > 0$ are not equidistant. The probability level for all confidence regions that will be shown is 99%.

As *a priori* information on the impulse response we choose $M = 3$ and $\rho = 1.2$. Figure 2 depicts the prior bound on the true impulse response as given by M and ρ, together with the true impulse response and the confidence interval of Theorem 3, using $n_p = 45$ and $W = \mathrm{diag}\,(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k}))$. Figure 2 shows that the prior information given by M and ρ and the resulting confidence interval for the estimated FIR parameters are consistent, and that the prior information given by M and ρ is reasonably tight. Figure 2 also gives the weighting function $\mathcal{W}$ that will be used for the parametric estimate $\hat{G}^f(e^{j\omega})$. Comparing the prior bound and the confidence interval, it can be seen that $\mathcal{W}$ is chosen as indicated in Section 4.3. Figure 3 shows the confidence interval for the estimate $\hat{G}(e^{j\omega_k})$ of Theorem 2, which is only defined at $\omega_k \in \Omega^u_{N_0}$, and the confidence interval for an estimate $\hat{G}^f(e^{j\omega})$ of Theorem 4 with $n_f = 40$ and $\mathcal{W}$ as given in Fig. 2, together with the true errors. Note that very good estimates are obtained for those frequencies where $|U_i(e^{j\omega_k})|$ was chosen to be large. Finally, Fig. 4 gives the Nyquist plot of the true system, together with the confidence region according to the estimate $\hat{G}^f(e^{j\omega})$. Note that in all cases the error bounds are tight, i.e. at some frequencies the actual error gets close to the upper bound.

6. CONCLUSIONS

By applying a procedure similar to Bartlett's procedure of periodogram averaging in conjunction with a periodic input, a reliable and tight error bound for the transfer function of an estimated FIR model has been established, when the period length of the input signal is large. The error bound is formulated in terms of a confidence region. The result is based on the asymptotic normality of the DFT of a filtered sequence of independent identically distributed random variables, and only minor prior information is required.

In the derivation a probabilistic setting for the

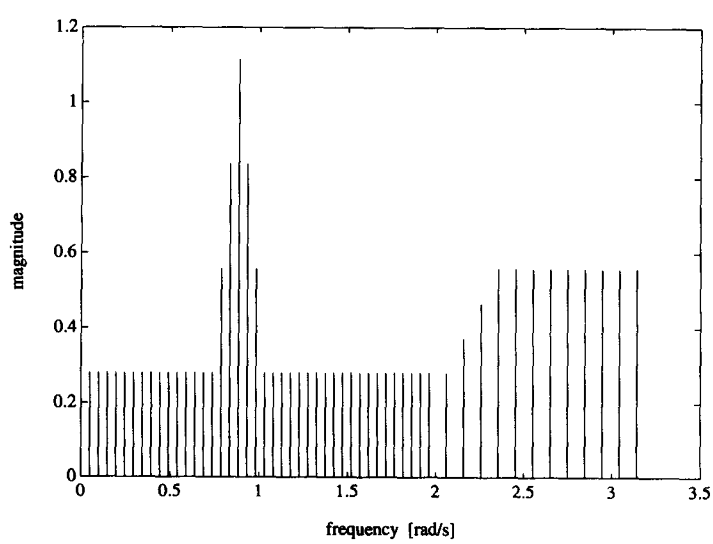

Fig. 1. Magnitude of the DFT over one period of the input signal, $|U_i(e^{j\omega_k})|$.

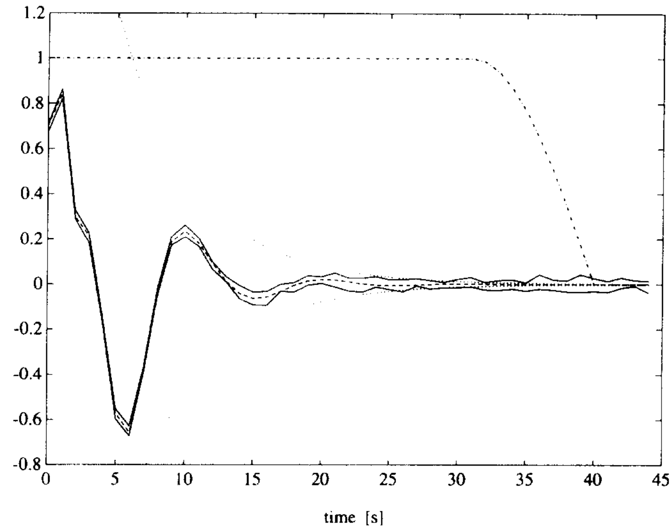

Fig. 2. True impulse response $g_0(k)$ (--), confidence interval of $\hat{\theta}\langle k\rangle$ (—), *a priori* bound $M\rho^{-k}$ ($\cdots$) and weighting function $\mathcal{W}$ (- · · -).

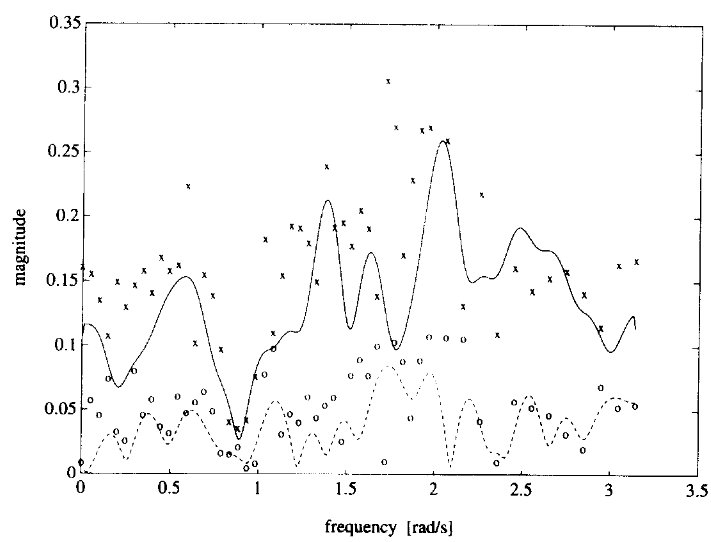

Fig. 3. Error bound for the non-parametric estimate $\hat{G}(e^{j\omega_k})$ (×), defined at $\omega_k \in \Omega^u_{N_0}$, true error $|\hat{G}(e^{j\omega_k}) - G_0(e^{j\omega_k})|$ (O), error bound for the parametric estimate $\hat{G}^f(e^{j\omega})$ (−) and true error $|\hat{G}^f(e^{j\omega}) - G_0(e^{j\omega})|$ (--).

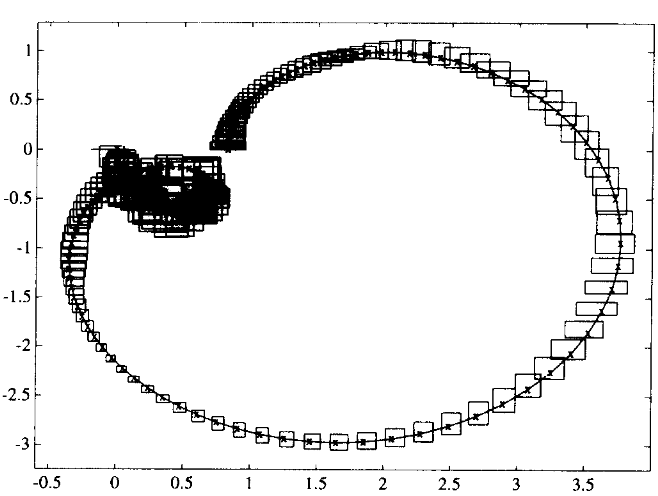

Fig. 4. Nyquist plot of the true system and confidence region of the parametric estimate $\hat{G}^f(e^{j\omega})$.

noise was chosen, whereas the input signal outside the measurement interval and the effects of undermodelling were considered as being unknown-but-bounded. The fact that the variance of the noise is estimated from the data is accounted for in the asymptotic error distribution.

As intermediate results, we have obtained an error bound on the ETFE, and on the parameters of the estimated FIR model. The latter error bound can be used to validate and improve the prior information on the impulse response, thus reducing the 'engineering leap of faith' in choosing this prior information.

In all cases, explicit and separate bounds are obtained for the different sources of uncertainty. This allows for an explicit trade-off between the bias and variance contributions to the error bound by selecting the model order, and provides valuable insight into the effective ways of improving the error bound when the error bound is too large in view of the robust control design requirements.

Acknowledgements—The authors would like to thank the anonymous referees for their careful evaluation and constructive remarks that contributed to the improvement of this paper.

REFERENCES

Bayard, D. S. (1992). Statistical plant set estimation using Schroeder-phased multi-sinusoidal input design. *Proc. Am. Control Conf.*, Chicago, IL, pp. 2988–2995.

Brillinger, D. R. (1981). *Time Series. Data Analysis and Theory*, Expanded Edition. Holden-Day, San Francisco.

De Vries, D. K. (1993). Quantification of model uncertainty from experimental data using system based orthonormal basis functions. Report N-439, Mechanical Engineering Systems and Control Group, Delft University of Technology.

De Vries, D. K. (1994). Identification of model uncertainty for control design. Dr. Dissertation, Delft University of Technology, The Netherlands, September 1994.

De Vries, D. K. and P. M. J. Van den Hof (1994). Quantification of model uncertainty from data. *Int. J. Robust Nonlinear Control*, **4**, 301–319.

Fogel, E. and Y. F. Huang (1982). On the value of information in system identification—bounded noise case. *Automatica*, **18**, 229–238.

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

Gu, G. and P. P. Khargonekar (1992). A class of algorithms for identification in H_∞. *Automatica*, **28**, 299–312.

Hakvoort, R. G. (1993). Worst-case system identification in H_∞: error bounds and optimal models. *Preprints 12th IFAC World Congress*, Sydney, Australia, Vol. 8, pp. 161–164.

Helmicki, A. J., C. A. Jacobson and C. N. Nett (1990). Identification in H_∞: a robustly convergent, nonlinear algorithm. *Proc. Am. Control Conf.*, Chicago, IL, pp. 386–391.

Helmicki, A. J., C. A. Jacobson and C. N. Nett (1993). Least squares methods for H_∞ control-oriented system identification. *IEEE Trans. Autom. Control*, **AC-38**, 819–826.

Heuberger, P. S. C, P. M. J. Van den Hof and O. H. Bosgra (1992). A generalized orthonormal basis for linear dynamical systems. Report N-404, Mechanical Engineering Systems and Control Group, Delft University of Technology, To appear in *IEEE Trans. Autom. Control*, **AC-40**, No. 3 (March 1995).

Heuberger, P. S. C., P. M. J. Van den Hof and O. H. Bosgra (1993). A generalized orthonormal basis for linear dynamical systems. *Proc. 32nd IEEE Conf. Decision and Control*, 15–17 December 1993, San Antonio, TX, pp. 2850–2855.

Hjalmarsson, H. (1993). Aspects of incomplete modelling in system identification. Dr. Dissertation, Linköping Studies in Science and Technology, No. 298, Linköping, Sweden.

Jenkins, G. M. and D. G. Watts (1968). *Spectral Analysis and its Applications*. Holden-Day, San Francisco.

Johnson, R. A. and D. W. Wichern (1988). *Applied Multivariate Statistical Analysis*, 2nd edn. Prentice-Hall, Englewood Cliffs, NJ.

LaMaire, R. O., L. Valavani, M. Athans and G. Stein (1991). A frequency-domain estimator for use in adaptive control systems. *Automatica*, **27**, 23–38.

Ljung, L. (1985a). On the estimation of transfer functions. *Automatica*, **21**, 677–696.

Ljung, L. (1985b). Asymptotic variance expressions for identified black-box transfer function models. *IEEE Trans. Autom. Control*, **AC-30**, 834–844.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.

Ljung, L. and P. E. Caines (1979). Asymptotic normality of prediction error estimators for approximate system models. *Stochastics*, **3**, 29–46.

Ljung, L. and Z. D. Yuang (1985). Asymptotic properties of black-box identification of transfer functions. *IEEE Trans. Autom. Control*, **30**, 514–530.

Ljung, L., B. Wahlberg and H. Hjalmarsson (1991). Model quality: the role of prior knowledge and data information. *Proc. 30th IEEE Conf. Decision and Control*, Brighton, U.K. pp. 273–278.

Milanese, M. and A. Vicino (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, **27**, 997–1009.

Ninness, B. M. and G. C. Goodwin (1994). Estimation for robust control. In R. Smith and M. Dahleh (Eds), *The Modelling of Uncertainty in Control Systems. Lecture Notes in Control and Information Sciences*, Vol. 192, pp. 235–259. Springer, London.

Norton, J. P. (1987). Identification and application of bounded-parameter models. *Automatica*, **23**, 497–507.

Oppenheim, A. V. and R. W. Schafer (1975). *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, London.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.

Schoukens, J. and R. Pintelon (1991). *Identification of Linear Systems. A Practical Guideline to Accurate Modeling*. Pergamon Press, Oxford.

Schroeder, M. R. (1970). Synthesis of low peak-factor signals and binary sequences of low auto-correlation. *IEEE Trans. Information Theory*, **IT-16**, 85–89.

Schweppe, F. C. (1973). *Uncertain Dynamic Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Sjöberg, J., L. Ljung and T. McKelvey (1993). On the use of regularization in system identification. *Preprints 12th IFAC World Congress*, Sydney, Australia, Vol. 7, pp. 381–386.

Wahlberg, B. and L. Ljung (1992). Hard frequency-domain model error bounds from least-squares like identification techniques. *IEEE Trans. Autom. Control*, **AC-37**, 900–912.

Zhu, Y. C. (1989). Estimation of transfer functions: asymptotic theory and a bound of model uncertainty. *Int. J. Control*, **49**, 2241–2258.

Zhu, Y. C. and A. C. P. M. Backx (1991). MIMO process identification for controller design: test signals, nominal model and error bounds. In *Preprints of 9th IFAC/IFORS*

Symp. on Identification and System Parameter Estimation, Budapest, pp. 1202–1207.

APPENDIX

*A*1. *Proof of Lemma* 1

Proof of inequality (17): using (12) we have $|S_i| = |R_i|/|U_i|$. The upper bound on $|R_i(e^{j\omega_k})|$ directly follows from the results of De Vries and Van den Hof (1994).

Proof of inequality (18): from (16) and (12) we have

$$|S| = \frac{1}{r\,|U_i|}\left|\sum_{i=1}^{r} R_i\right|,$$

where we used the fact that U_i is independent of i because the input is periodic. Using (6), (4), (5) and the fact that $e^{-j2\pi l} = 1$ for $l \in \mathbb{Z}$, we find for $k \in T^{N_0}$ and $N = rN_0$

$$\begin{aligned}
X^s(e^{j(2\pi k/N_0)}) &= \frac{1}{\sqrt{N}}\sum_{t=0}^{N-1} x(t+N_s)e^{-j(2\pi k/N_0)t} \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{r}\sum_{t=(i-1)N_0}^{iN_0-1} x(t+N_s)e^{-j(2\pi k/N_0)t} \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{r}\sum_{l=0}^{N_0-1} x(l+(i-1)N_0+N_s) \\
&\quad\times e^{-j(2\pi k/N_0)(l+(i-1)N_0)} \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{r}\sum_{l=0}^{N_0-1} x_i(l)e^{-j(2\pi k/N_0)l} \\
&= \frac{1}{\sqrt{r}}\sum_{i=1}^{r} X_i(e^{j(2\pi k/N_0)}).
\end{aligned} \tag{A.1}$$

Note that it is not assumed that $x(t)$ is periodic. The result now follows from the upper bound on $|E^s(e^{j\omega_k})| = |R^s(e^{j\omega_k})|$ that is given in De Vries and Van den Hof (1994).

*A*2. *Proof of Lemma* 2

Using (20) gives

$$\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G}) = \frac{1}{r(r-1)}\sum_{i=1}^{r}(|\tilde{G}-\tilde{G}_i|^2 - |\hat{G}-\hat{G}_i|^2). \tag{A.2}$$

From the fact that $\tilde{G}_i = \hat{G}_i - S_i$, we find

$$\begin{aligned}
\tilde{G}-\tilde{G}_i &= \hat{G} - S - \hat{G}_i + S_i \\
&= (\hat{G}-\hat{G}_i) + (S_i - S) = A_i + B_i
\end{aligned} \tag{A.3}$$

$$|\tilde{G}-\tilde{G}_i|^2 = |A_i+B_i|^2 = |A_i|^2 + A_iB_i^* + B_iA_i^* + |B_i|^2. \tag{A.4}$$

Combining (A.2) and (A.4) gives

$$\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G}) = \frac{1}{r(r-1)}\sum_{i=1}^{r} A_iB_i^* + B_iA_i^* + |B_i|^2,$$

resulting in

$$|\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G})| \le \frac{1}{r(r-1)}\sum_{i=1}^{r} 2\,|A_i|\,|B_i| + |B_i|^2,$$

with

$$\begin{aligned}
|A_i| &= |\hat{G}-\hat{G}_i| \\
|B_i| &= |S_i - S| = \left|S_i - \frac{1}{r}\sum_{k=1}^{r} S_k\right| \\
&= \left|\frac{r-1}{r}S_i - \frac{1}{r}\sum_{k=1}^{i-1} S_k - \frac{1}{r}\sum_{k=i+1}^{r} S_k\right| \\
&\le \frac{r-1}{r}|S_i| + \frac{1}{r}\sum_{k=1}^{i-1}|S_k| + \frac{1}{r}\sum_{k=i+1}^{r}|S_k| \\
&= \frac{r-2}{r}|S_i| + \frac{1}{r}\sum_{k=1}^{r}|S_k|,
\end{aligned}$$

which completes the proof.

*A*3. *Proof of Lemma* 3

We will use the following standard results from probability theory; see e.g. Priestley (1981).

1. If $x_i \in \mathcal{N}(0,1)$, $i = 1,\ldots,k$, are independent random variables and $z = \sum_{i=1}^{k} x_i^2$, then $z \in \chi^2(k)$ with mean k and variance $2k$.
2. If $x \in \chi^2(a)$ and $y \in \chi^2(b)$ are independent random variables and $z = \dfrac{x/a}{y/b}$, then $z \in F(a,b)$.
3. Let x_i, $i = 1,\ldots,k$, be a sequence of independent identically distributed real random variables, with $x_i \in \mathcal{N}(\mu,\sigma^2)$. Let $\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k} x_i$ and let $\hat{\sigma}^2 = \frac{1}{k-1}\times \sum_{i=1}^{k}(\hat{\mu}-x_i)^2$. Then, $\hat{\mu}$ and $\hat{\sigma}^2$ are independent, and $\hat{\mu} \in \mathcal{N}\left(\mu, \frac{1}{r}\sigma^2\right)$, and $\dfrac{(k-1)\hat{\sigma}^2}{\sigma^2} \in \chi^2(k-1)$.
4. Uncorrelated jointly normally distributed random variables are independent.

The convergence in distribution follows from the Continuous Mapping Theorem, see Pollard (1984, p. 46), which states that if the sequence of random vectors x_n converges in distribution to x, then $f(x_n)$ converges in distribution to $f(x)$, for all continuous functions $f: \mathbb{R}^k \to \mathbb{R}^s$. We denote

$$D(e^{j\omega_k}) := \frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))}$$

and

$$\mathbf{V}(e^{j\omega_k}) := \begin{bmatrix} \mathrm{Re}\{V_1(e^{j\omega_k})\} \\ \vdots \\ \mathrm{Re}\{V_r(e^{j\omega_k})\} \\ \mathrm{Im}\{V_1(e^{j\omega_k})\} \\ \vdots \\ \mathrm{Im}\{V_r(e^{j\omega_k})\} \end{bmatrix}$$

Clearly we have that $D(e^{j\omega_k}) = f(\mathbf{V}(e^{j\omega_k}))$, where $f(\cdot)$ denotes some function. For a periodic input signal, i.e. $U_i(e^{j\omega_k})$ independent of i, it follows by direct calculation that

$$D = \frac{L(\mathrm{Re}\{V_l\}\,\mathrm{Re}\{V_m\}, \mathrm{Re}\{V_l\}\,\mathrm{Im}\{V_m\}, \mathrm{Im}\{V_l\}\,\mathrm{Im}\{V_m\})}{L(\mathrm{Re}\{V_l\}\,\mathrm{Re}\{V_m\}, \mathrm{Re}\{V_l\}\,\mathrm{Im}\{V_m\}, \mathrm{Im}\{V_l\}\,\mathrm{Im}\{V_m\})},$$

where $L(\cdot)$ denotes some linear combination of its arguments, with real coefficients. Hence $f(\mathbf{V}): \mathbb{R}^{2r} \to \mathbb{R}$ is a continuous function for all $\hat{\sigma}_r^2(\tilde{G}) > 0$. From the Continuous Mapping Theorem it now follows that if the vector $\mathbf{V}(e^{j\omega_k})$ converges in distribution to $\mathcal{V}(e^{j\omega_k})$ then $D(e^{j\omega_k}) = f(\mathbf{V}(e^{j\omega_k}))$ converges in distribution to $f(\mathcal{V}(e^{j\omega_k}))$, taking into account that the event $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k})) = 0$ has probability zero. The distribution of $\mathcal{V}(e^{j\omega_k})$ is completely specified by the asymptotic distribution of $\check{V}_{N_0}$ as given in Theorem 1. Hence for the derivation of the asymptotic distribution of $D(e^{j\omega_k})$, we can use $\check{V}_{N_0} \in \mathcal{N}(0,\Lambda)$ with Λ as given in Theorem 1.

Note that $\mathrm{Im}\{\tilde{G}(e^{j\omega_k})\} \equiv 0$ for $\omega_k \in \{0,\pi\}$. We will therefore split the remainder of the proof in two parts. First, we will derive the distribution for $\omega_k = 0, \pi$, and next for $\omega_k \ne 0, \pi$. We denote

$$B(e^{j\omega_k}) := \frac{(r-1)\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))}{\mathrm{var}\,[\tilde{G}(e^{j\omega_k})]} \tag{A.5}$$

$$A(e^{j\omega_k}) := \frac{\tilde{G}(e^{j\omega_k}) - G_0(e^{j\omega_k})}{\sqrt{\mathrm{var}\,[\tilde{G}(e^{j\omega_k})]}}.$$

Case $\omega_k = 0, \pi$. From $\check{V}_{N_0} \in \mathcal{N}(0,\Lambda)$ with Λ as given in Theorem 1, it follows that the $V_i(e^{j\omega_k})$, $i = 1,\ldots,r$, are normally distributed and that they are independent and identically distributed. Therefore, $A(e^{j\omega_k}) \in \mathcal{N}(0, 1)$ and $B(e^{j\omega_k}) \in \chi^2(r-1)$. Because $\mathrm{Im}\{A(e^{j\omega_k})\} \equiv 0$ for $\omega_k = 0, \pi$ we have that $A^2 \in \chi^2(1)$. Furthrmore, $\tilde{G}(e^{j\omega_k})$ and $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$ are independent, so that $D(e^{j\omega_k})$ is a quotient of two independent χ^2 distributions, which is an F distribution:

$$\frac{|A|^2/1}{B/(r-1)} = \frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} = F(1, r-1).$$

Case $\omega_k \neq 0, \pi$. From $\check{V}_{N_0} \in \mathcal{N}(0, \Lambda)$ with Λ as given in Theorem 1, it follows that $A(e^{j\omega_k}) \in \mathcal{N}(0, 1)$. By direct calculation it also follows that $\mathrm{Re}\{\tilde{G}(e^{j\omega_k})\}$ and $\mathrm{Im}\{\tilde{G}(e^{j\omega_k})\}$ have the same variance, and hence that $\mathrm{Re}\{A(e^{j\omega_k})\}$ and $\mathrm{Im}\{A(e^{j\omega_k})\}$ have the same variance, for $\omega_k \neq 0, \pi$. Noting that for any complex random variable z there holds $\sigma^2(z) = \mathbb{E}[z^*z] = \sigma^2(\mathrm{Re}\{z\}) + \sigma^2(\mathrm{Im}\{z\})$, we find that $\mathrm{Re}\{A(e^{j\omega_k})\} \in \mathcal{N}(0, \frac{1}{2})$ and $\mathrm{Im}\{A(e^{j\omega_k})\} \in \mathcal{N}(0, \frac{1}{2})$. Hence $\mathrm{Re}^2\{\sqrt{2}A\} \in \chi^2(1)$ and $\mathrm{Im}^2\{\sqrt{2}A\} \in \chi^2(1)$. Finally, from $\check{V}_{N_0} \in \mathcal{N}(0, \Lambda)$ it follows by direct calculation that $\mathrm{Re}\{A(e^{j\omega_k})\}$ and $\mathrm{Im}\{A(e^{j\omega_k})\}$ are uncorrelated and jointly normally distributed. Hence $\mathrm{Re}\{A(e^{j\omega_k})\}$ and $\mathrm{Im}\{A(e^{j\omega_k})\}$ are independent. The above observations result in

$$2|A|^2 = 2\,\mathrm{Re}^2\{A\} + 2\,\mathrm{Im}^2\{A\} \in \chi^2(2).$$

Using a similar argument, we find that

$$2B(e^{j\omega_k}) \in \chi^2(2(r-1)). \tag{A.6}$$

Again, $\tilde{G}(e^{j\omega_k})$ and $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$ are independent, so that

$$\frac{|A|^2/2}{2B/(2(r-1))} = \frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} = F(2, 2(r-1)),$$

which concludes the proof.

A4. Proof of Theorem 2

From Lemma 3 we find

$$\frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} \leq \alpha \quad \text{w.p.} \quad \begin{cases} F_\alpha(2, 2(r-1)) & \omega_k \neq 0, \pi \\ F_\alpha(1, r-1) & \omega_k = 0, \pi. \end{cases}$$

Hence

$$|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})| \leq \sqrt{\alpha \hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))}. \tag{A.7}$$

Combining (A.7) and (15) gives

$$|G_0(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq |S(e^{j\omega_k})| + \sqrt{\alpha \hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))}.$$

Using Lemma 2 and the triangle inequality

$$\hat{\sigma}_r^2(\tilde{G}) \leq \hat{\sigma}_r^2(\hat{G}) + |\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G})|$$

completes the proof.

A5. Proof of new elements in Theorem 3

Using (2), (29) and (24) we rewrite the system's equation (11) as

$$\hat{G}_i(e^{j\omega_k}) = \phi(e^{j\omega_k})\theta_0 + Z(e^{j\omega_k}) + S_i(e^{j\omega_k}) + \frac{V_i(e^{j\omega_k})}{U_i(e^{j\omega_k})}, \tag{A.8}$$

with

$$Z(e^{j\omega_k}) = \sum_{l=n_p}^{\infty} g_0(l)e^{-j\omega_k l}.$$

Note that $Z(e^{j\omega_k})$ is the error due to undermodelling (truncation of the impulse response of the true system), $S_i(e^{j\omega_k})$ is the error due to the unknown past inputs, and the last term in (A.8) is the error due to the noise.

From (27) it now follows that

$$\begin{aligned} \hat{\theta}_i &= \Psi\Phi\theta_0 + \Psi\mathbf{Z} + \Psi\mathbf{S}_i + \Psi\mathbf{F}_i \\ &= \theta_0 + \Psi\mathbf{Z} + \Psi\mathbf{S}_i + \Psi\mathbf{F}_i, \end{aligned} \tag{A.9}$$

with

$$\mathbf{Z} = [Z(e^{j\zeta_1}) \quad Z(e^{j\zeta_2} \quad \ldots \quad Z(e^{j\zeta_{N_p}})]^T$$
$$\mathbf{S}_i = [S_i(e^{j\zeta_1}) \quad S_i(e^{j\zeta_2}) \quad \ldots \quad S_i(e^{j\zeta_{N_p}})]^T.$$

Using (26) and (16) now gives

$$\hat{\theta} = \theta_0 + \Psi\mathbf{Z} + \Psi\mathbf{S} + \Psi\frac{1}{r}\sum_{i=1}^{r}\mathbf{F}_i,$$

with

$$\mathbf{S} = [S(e^{j\zeta_1}) \quad S(e^{j\zeta_2}) \quad \ldots \quad S(e^{j\zeta_{N_p}})]^T.$$

We will now only prove the inequalities for $z^p := \Psi\mathbf{Z}$ and $s^p := \Psi\mathbf{S}$. The asymptotic distribution follows using a similar argument as in Section A3, while noting that the estimates $\hat{\theta}_i$ are real so that we do not have additional degrees of freedom due to independent real and imaginary parts.

The inequality for $z^p := \Psi\mathbf{Z}$ follows as

$$\begin{aligned} z^p\langle k\rangle &= \sum_{l=1}^{N_p}\Psi\langle k, l\rangle\mathbf{Z}\langle l\rangle \\ &= \sum_{l=1}^{N_p}\Psi\langle k, l\rangle\sum_{m=n_p}^{\infty}g_0(m)e^{-j\zeta_l m} \\ &= \sum_{m=n_p}^{\infty}g_0(m)\sum_{l=1}^{N_p}\Psi\langle k, l\rangle e^{-j\zeta_l m} \\ |z^p\langle k\rangle| &\leq \sum_{m=n_p}^{n_h-1}|g_0(m)|\left|\sum_{l=1}^{N_p}\Psi\langle k, l\rangle e^{-j\zeta_l m}\right| \\ &\quad + \left(\sum_{m=n_h}^{\infty}|g_0(m)|\right)\left(\sum_{l=1}^{N_p}|\Psi\langle k, l\rangle|\right). \end{aligned}$$

The result now follows by noting that $|g_0(m)| \leq M\rho^{-m}$ and using the Taylor series expansion of $1/(1-x)$ with $x = 1/\rho$:

$$\sum_{m=n_h}^{\infty}|g_0(m)| \leq \frac{M\rho}{\rho-1}\rho^{-n_h}.$$

The inequality for $s^p := \Psi\mathbf{S}$ follows directly from

$$s^p\langle k\rangle = \sum_{l=1}^{N_p}\Psi\langle k, l\rangle S(e^{j\zeta_l}).$$

A6. Proof of new elements in Theorem 4

From (A.9), (32) and (33) it follows that

$$\hat{G}_i^f(e^{j\omega}) = G_0(e^{j\omega}) + \Upsilon\mathbf{Z} + \Upsilon\mathbf{S}_i - \Lambda(e^{j\omega}) - \Gamma(e^{j\omega}) + \Upsilon\mathbf{F}_i,$$

where $\Lambda(e^{j\omega})$ is an undermodelling error and $\Gamma(e^{j\omega})$ is the error due to weighting with $\mathcal{W}$:

$$\Lambda(e^{j\omega}) = G_0(e^{j\omega}) - P[I \quad 0]\theta_0$$
$$\Gamma(e^{j\omega}) = P[I - \mathcal{W} \quad 0]\theta_0.$$

We will only prove the inequality for Γ. Everything else is similar to the arguments given in Sections A5, A4 and A3. The inequlity for Γ follows as

$$\begin{aligned} \Gamma(e^{j\omega}) &= P(e^{j\omega})[I - \mathcal{W} \quad 0]\theta_0 \\ &= \sum_{l=0}^{n_f-1}\left(g_0(l) - \sum_{m=0}^{n_f-1}\mathcal{W}\langle l+1, m+1\rangle g_0(m)\right)e^{-j\omega l} \\ &= \sum_{l=0}^{n_f-1}\left(g_0(l)(1 - \mathcal{W}\langle l+1, l+1\rangle)\right. \\ &\quad - \sum_{m=0}^{l-1}\mathcal{W}\langle l+1, m+1\rangle g_0(m) \\ &\quad \left. - \sum_{m=l+1}^{n_f-1}\mathcal{W}\langle l+1, m+1\rangle g_0(m)\right)e^{-j\omega l}. \end{aligned}$$