



# Consistent Parameter Bounding Identification for Linearly Parametrized Model Sets\*

RICHARD G. HAKVOORT† and PAUL M. J. VAN DEN HOF†

*The introduction of cross-covariance noise constraints in parameter bounding identification yields consistency properties that are not present in the standard bounded error identification approach with time-domain amplitude noise constraints.*

**Key Words**—Parameter bounding identification; unknown-but-bounded noise; discrete-time systems; linear programming; convergence analysis.

**Abstract**—In standard parameter bounding identification a time-domain bound on the noise signal is used to construct parameter bounds. One of the properties of this standard approach is that there is no consistency if a conservative noise bound has been chosen. It is shown that alternative bounds on the noise signal give rise to consistent parameter bounding identification methods; i.e., asymptotically in the number of data samples, the feasible parameter set converges to the true parameter vector. The first alternative noise bound that is introduced is a bound on the cross-covariance between the noise and some instrumental (input) signal. The noise bound is represented by a small number of linear inequalities, which can be used in parameter bounding by linear programming. It is shown that there is consistency under fairly general conditions, even when a conservative bound has been chosen. Additionally, a procedure is presented to estimate the cross-covariance bound from data. Similar consistency results are shown for two other types of noise bounds: a bound on the discrete Fourier transform of the noise in combination with sinusoidal excitation, and a time-domain bound on the noise after measurement averaging with periodic excitation.

## 1. INTRODUCTION

The literature on set membership, bounded error or parameter bounding identification is now quite extensive. See Walter and Piet-Lahanier (1990), Milanese and Vicino (1991) and Norton and Veres (1991) for overviews of this topic. The idea is to calculate a parameter set of minimal size using measurement data and certain deterministic bounds on the noise. To clarify the discussion, consider the discrete-time linear regression model

$$y(t) = \phi^T(t)\theta + e(t), \quad t = 1, \dots, N, \quad (1)$$

\* Received 25 October 1993; revised 29 June 1994; received in final form 24 November 1994. This paper was not presented at any IFAC meeting. 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 M. J. 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.

where  $y(t)$  is the measured output at time  $t$ ,  $\phi(t)$  the  $n \times 1$  regression vector,  $\theta$  the  $n \times 1$  parameter vector,  $e(t)$  the equation error or residual, and  $N$  the number of samples. This linear regression model can describe a large class of systems, including multi-input single-output and nonlinear systems. The observations are assumed to be generated by

$$y(t) = \phi^T(t)\theta_0 + e_0(t), \quad t = 1, \dots, N, \quad (2)$$

with  $\{e_0(t)\}$  an unknown stochastic noise process. It is emphasized that  $\{e_0(t)\}$  is not assumed to be white noise or uncorrelated with the regression vector  $\{\phi(t)\}$ . Hence the data may be generated in closed loop, and the regression vector may contain samples of the output signal  $\{y(t)\}$ . Basically,  $\{e_0(t)\}$  may also account for undermodelling; however, undermodelling will not be considered in this paper.

In parameter bounding identification the parameter vector  $\theta$  is bounded on the basis of certain bounds on  $\{e(t)\}$ . The most common procedure is to bound the amplitude of the residuals in the time domain,

$$e_l(t) \leq e(t) \leq e_u(t), \quad t = 1, \dots, N, \quad (3)$$

(see e.g. Walter and Piet-Lahanier, 1990). The feasible parameter set is then defined as

$$\Theta_N = \{\theta \mid e_l(t) \leq y(t) - \phi^T(t)\theta \leq e_u(t), \quad t = 1, \dots, N\}.$$

Next an orthotopic outer bounding parameter set can be constructed by calculating

$$\theta_k^{(l)} = \min_{\theta \in \Theta_N} \theta_k, \quad \theta_k^{(u)} = \max_{\theta \in \Theta_N} \theta_k \quad (4)$$

for  $k = 1, \dots, n$ , which requires solving  $2n$  linear programming problems with  $n$  unknowns subject to  $2N$  linear inequality constraints (for details

see Milanese and Belforte, 1982). If  $\{e_0(t)\}$  also satisfies the residual bounds (3) then the parameter vector  $\theta_0$  is guaranteed to be in the identified set.

Note that the feasible parameter set  $\Theta_N$  may be unbounded. This can happen if the data are insufficiently informative with respect to the chosen model parameterization. Basically, this would imply that one or more parameter bounds as identified in (4) are infinite. If this is undesired, prior parameter bounds can be included in the definition of  $\Theta_N$ .

Fogel and Huang (1982) and Veres and Norton (1991) have shown that under certain conditions the outer bounding parameter set converges to  $\theta_0$  if  $N \rightarrow \infty$ , provided the noise  $\{e_0(t)\}$  is at sufficiently many time instants arbitrarily close to the specified noise bounds, without exceeding them. However, it seems impossible to meet this requirement in many practical situations. In practice, the noise bounds have to be chosen as conservative in order to guarantee their correctness. Therefore in general there is no consistency in parameter bounding identification. A similar situation is encountered in the field of robust identification in  $H_\infty$  (see e.g. Helmicki *et al.*, 1991; Gu and Khargonekar, 1992). There, deterministic bounds on the noise are assumed, and basically only consistency is established under the condition that the noise level tends to zero. This highlights the demerit of only using the bounds (3). The explanation for this lack of consistency in the presence of noise is that the noise is assumed to be able to take a worst-case realization within the noise bounds. In particular, the residuals  $\{e(t)\}$  can be heavily correlated with the input signal.

For the prediction-error type of identification procedures it is known that if the number of data samples tends to infinity, the parameter estimate converges to the true parameter vector  $\theta_0$  under fairly general conditions, also for nonzero noise (see e.g. Ljung, 1987). This is due to the fact that stochastic or averaging properties are present in this identification setting. The objective of this paper is to adjust the parameter bounding identification such that a similar consistency property is obtained in the presence of noise. This is achieved by introducing alternative noise bounds, which have a stochastic interpretation. A basic motivation for using the noise bounds (3) is that for small data sets, with  $N$  small, stochastic assumptions on the noise may not be justifiable. However, if large data sets are available, stochastic noise assumptions often can be justified, in which case the noise characterization (3) is overly pessimistic.

Stochastics have already been introduced into

parameter bounding identification by Fogel and Huang (1982), where basically a stochastic interpretation is given for the noise bounds (3). In Veres and Norton (1989) the situation is considered where the noise has a bounded autocovariance, or cross-covariance with a specified signal, but only for the purpose of model structure selection. Fogel (1979) considers bounds on the  $\ell_2$ -norm or energy of the noise, but this does not lead to consistency either. In Gustafsson and Mäkilä (1993) a bound on the  $\ell_1$ -norm of the noise is proposed. But the use of this noise bound leads to huge linear programming problems, and a consistency property is also lacking.

In this paper a cross-covariance bound on the noise is introduced into parameter bounding identification as an alternative to the standard time-domain bound on the noise (3). The sample covariance between the residuals and some given signal is bounded. Typically, this signal is chosen such that it is correlated to the input signal, but uncorrelated to the noise process. In an open-loop experimental situation the (filtered) input signal meets the specifications. In closed-loop operation some external reference signal can be taken. The cross-covariance noise bounds can be represented by a small number of linear inequalities, which can be used to calculate an outer bounding parameter set by means of linear programming, similar to (4). Sufficient conditions are derived that make the feasible parameter region converge to the parameter vector  $\theta_0$ . These conditions are very general: basically, only persistence of excitation is necessary, and it is not required that the specified bounds be tight. A stochastic interpretation is given of the cross-covariance bounds on the noise. Also, a procedure is presented to estimate correct bounds from measurement data.

Similar consistency results appear obtainable with other types of noise bounds, which also give rise to linear constraints usable in parameter bounding identification with linear programming techniques. A frequency-domain bound on the noise is introduced into parameter bounding identification. More specifically, the amplitude of the discrete Fourier transform of the residuals, the square-root of its periodogram, is bounded for a set of specified frequencies. This bound is also utilized in frequency-domain identification procedures (see e.g. Lataire *et al.*, 1991; De Vries and Van den Hof, 1992, 1994). As parameter bounding identification adopts a time-domain setting, the bound requires a translation into linear constraints in the time domain, which is presented in this paper. A stochastic analysis of the noise bound is

presented, as well as sufficient conditions for consistency.

Finally, similar results are shown to be obtainable for a time-domain bound on the noise in combination with measurement averaging and periodic excitation. The use of repeated experiments has also been suggested by Mäkilä (1991) in an  $\ell_1$ -identification setting. And succesful use of repeated experiments has been reported in De Vries and Van den Hof (1992, 1994) in an  $H_\infty$ -identification setting.

The outline of the paper is as follows. In Section 2 the cross-covariance bound on the noise is elaborated and consistency results are derived. In Section 3 it is discussed how to estimate noise bounds from data in a statistically reliable way. In Section 4 the frequency-domain bound on the noise is considered. Next, in Section 5 the time-domain noise constraints are combined with repeating experiments. In Section 6 a simulation example is shown. Finally, in Section 7 conclusions are drawn. This paper is an extended version of Hakvoort *et al.* (1993).

2. CROSS-COVARIANCE CONSTRAINTS ON THE NOISE

First the cross-covariance constraints on the noise are introduced into parameter bounding identification. Consider the linear constraints

$$c_l(p) \leq \frac{1}{\sqrt{N}} \sum_{t=1}^N r_p(t)e(t) \leq c_u(p), \quad p = 1, \dots, s, \tag{5}$$

yielding the feasible parameter set

$$\Theta_N = \left\{ \theta \mid c_l(p) \leq \frac{1}{\sqrt{N}} \sum_{t=1}^N r_p(t)[y(t) - \phi^T(t)\theta] \leq c_u(p), p = 1, \dots, s \right\}, \tag{6}$$

where  $c_l(p)$  and  $c_u(p)$  are specified bounds, and  $\{r_p(t)\}$  is some specified signal, typically equal to a delayed and/or filtered signal that is correlated with the regression vector  $\{\phi(t)\}$  but uncorrelated with the noise process  $\{e_0(t)\}$ , as explained later. The constraints (5) restrict the set of accepted residuals, and therefore the feasible parameter set. As the constraints in (6) are linear in the parameter vector  $\theta$ , they can easily be included in the linear programming problems (4).

Of course, it is desirable to specify the bounds  $c_l(p)$  and  $c_u(p)$  on the sample covariance of the residuals with the signals  $\{r_p(t)\}$  such that they are satisfied by the true noise process  $\{e_0(t)\}$ . Analogously to the noise bounds (3), it is possible to consider the bounds (5) as being

entirely deterministic, and this then does not require the noise to be looked upon as a stochastic process. However, it appears that a nice probabilistic interpretation of the bounds (5) exists if the noise process  $\{e_0(t)\}$  has some stochastic properties, which can often be justified in practical situations. In fact, (5) then boils down to the assumption that the noise process  $\{e_0(t)\}$  is uncorrelated with the signals  $\{r_p(t)\}$ . The underlying noise assumption is that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N r_p(t)e_0(t) = 0,$$

which is a common assumption in identification (see e.g. Ljung, 1987).

For the analysis, some technical assumptions on the signals  $\{e_0(t)\}$ ,  $\{r_p(t)\}$  and  $\{\phi(t)\}$  are needed. The assumptions about  $\{e_0(t)\}$  are as follows.

*Assumption 2.1.* The noise process  $\{e_0(t)\}$  is stationary, satisfying  $e_0(t) = H_0(q)w_0(t)$  for some stable  $H_0(q)$ , and where  $\{w_0(t)\}$  is a sequence of independent random variables with zero mean values, variances  $\lambda_0$  and bounded fourth moments.

The assumptions about  $\{r_p(t)\}$  are as follows.

*Assumption 2.2.* Each signal  $\{r_p(t)\}$  is quasi-stationary, i.e. its autocovariance function

$$R_{r_p}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E r_p(t + \tau)r_p(t)$$

exists  $\forall \tau$ . Moreover,  $\{r_p(t)\}$  satisfies

$$r_p(t) = \bar{r}_p(t) + R_p(t, q)\bar{r}_p(t),$$

where, for each  $p$ ,  $\{\bar{r}_p(t)\}$  is a bounded deterministic signal,  $\{R_p(t, q), t = 1, 2, \dots\}$  is a uniformly stable family of filters, and  $\{\bar{r}_p(t)\}$  is a sequence of independent random variables with zero mean values, variances  $\lambda_{p,i}$  and bounded fourth moments.

The assumptions about  $\{\phi(t)\}$  are as follows.

*Assumption 2.3.* Each signal  $\{\phi_k(t)\}$ , i.e. the  $k$ th element of the vector  $\phi(t)$ , is quasi-stationary, and it satisfies

$$\phi_k(t) = \bar{\phi}_k(t) + S_k(t, q)\bar{\phi}_k(t),$$

where  $\{\bar{\phi}_k(t)\}$  is a bounded deterministic signal,  $\{S_k(t, q), t = 1, 2, \dots\}$  is a uniformly stable family of filters, and  $\{\bar{\phi}_k(t)\}$  is a sequence of

independent random variables with zero mean values, variances  $\mu_{k,t}$  and bounded fourth moments.

The assumptions about joint properties of  $\{r_p(t)\}$  and  $\{\phi_k(t)\}$  are as follows.

*Assumption 2.4.* For each  $p$  and  $k$ , the signals  $\{r_p(t)\}$  and  $\{\phi_k(t)\}$  are jointly quasi-stationary, i.e. they are both quasi-stationary and the cross-covariance function

$$R_{r_p\phi_k}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E r_p(t + \tau) \phi_k(t)$$

exists. Moreover, the signal  $\begin{bmatrix} \tilde{r}_p(t) \\ \tilde{\phi}_k(t) \end{bmatrix}$  has covariances  $v_{p,k,t}$  and bounded fourth moments.

Using these assumptions, a stochastic interpretation of the constraints (5) can be established. In this interpretation the notion of uncorrelation is strengthened to independence. In the following proposition, which gives the stochastic interpretation, fruitfully use is made of results established in Ljung (1987, Theorem 9.1) and Hjalmarsson (1993, Chapter 2).

*Proposition 2.5.* Suppose that  $\{e_0(t)\}$  and  $\{r_p(t)\}$  are independent and that they satisfy Assumptions 2.1 and 2.2 respectively. Denote

$$\Lambda_p^N := E \left[ \frac{1}{\sqrt{N}} \sum_{t=1}^N r_p(t) e_0(t) \right]^2,$$

$$\Lambda_p := \lim_{N \rightarrow \infty} \Lambda_p^N,$$

$$R_p^N(\tau) := \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} E r_p(t) r_p(t + |\tau|),$$

$$\tau = -N + 1, \dots, N - 1.$$

Then

(i) 
$$\Lambda_p^N = \sum_{\tau=-N+1}^{N-1} \frac{N-|\tau|}{N} R_p^N(\tau) R_{e_0}(\tau),$$

(ii) 
$$\Lambda_p = \sum_{\tau=-\infty}^{\infty} R_{r_p}(\tau) R_{e_0}(\tau),$$

(iii) 
$$\frac{1}{\sqrt{N}} \sum_{t=1}^N r_p(t) e_0(t) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, \Lambda_p),$$

where  $\mathcal{N}(0, \Lambda_p)$  denotes the normal distribution with mean 0 and variance  $\Lambda_p$ .

*Proof.* See the Appendix.

On the one hand, the bound (5) is a hard or deterministic bound; on the other, a probabilistic

interpretation has been given in the above proposition.

*Remark 2.6.* In part (iii) of Proposition 2.5 a convergence factor  $f_N = 1/\sqrt{N}$  has been used. The consequence of this choice is that an asymptotic distribution could be derived, which is independent of  $N$ . For example, asymptotically in  $N$ , a 0.9995 probability region is obtained by choosing  $c_u(p) = -c_l(p) = 3.5\sqrt{\Lambda_p}$ . However, other convergence factors can also be of interest. For example, the choice  $f_N = 1/\sqrt{N \log \log N}$  gives, by the law of the iterated logarithm (Heunis, 1988), that w.p. 1,

$$\lim_{N \rightarrow \infty} \left| \frac{1}{\sqrt{N \log \log N}} \sum_{t=1}^N r_p(t) e_0(t) \right| < \sqrt{\Lambda_p}.$$

See Ljung and Wahlberg (1992) for an extensive treatment of convergence factors. Note that from a practical viewpoint there is no difference, since for finite  $N$  a certain choice of  $c_l(p)$  and  $c_u(p)$  in (5) always corresponds to a certain probability, calculable with Proposition 2.5. Moreover,  $c_l(p)$  and  $c_u(p)$  may be chosen functions of  $N$ , although this has not been made explicit in the notation.

Proposition 2.5 also states that  $\Lambda_p$  can be evaluated by considering the second-order statistics of  $\{r_p(t)\}$  and  $\{e_0(t)\}$  separately. The second-order statistics of the signals  $\{r_p(t)\}$  will generally be known exactly; the statistics of the noise process  $\{e_0(t)\}$  have to be estimated from data. How to estimate these statistics is the subject of Section 3.

The covariance bound (5) is especially useful if  $\{r_p(t)\}$  is a signal that is correlated with the regression vector sequence  $\{\phi(t)\}$ , but uncorrelated with the noise  $\{e_0(t)\}$ . This follows from the following consistency result.

*Theorem 2.7.* Suppose that the signal  $\{e_0(t)\}$  satisfies the constraints (5) for given signals  $\{r_p(t)\}$ ,  $p = 1, \dots, s$ , and given and finite  $c_l(p)$  and  $c_u(p)$ . Suppose that  $\{r_p(t)\}$  and  $\{\phi_k(t)\}$  satisfy Assumptions 2.2–2.4. If the matrix

$$R = \begin{bmatrix} R_{r_1\phi_1}(0) & \dots & R_{r_1\phi_s}(0) \\ \vdots & & \vdots \\ R_{r_s\phi_1}(0) & \dots & R_{r_s\phi_s}(0) \end{bmatrix}$$

has full column rank then the feasible parameter region (6) converges to the true parameter vector  $\theta_0$ ,

$$\lim_{N \rightarrow \infty} \max_{\theta \in \Theta_N} |\theta - \theta_0| = 0,$$

with probability 1.

*Proof.* See the Appendix.

This theorem thus provides a consistency result for bounded error identification without requiring tight error bounds. If the values  $c_u(p)$  and  $c_l(p)$  in (5) are chosen too large, convergence will still take place. The cross-covariance noise bounds possess a certain averaging property that is not present in standard parameter bounding identification with time-domain noise bounds. Also note that this consistency result holds *without* assuming stationary noise: Assumption 2.1 has not been used in Theorem 2.7.

Though the bounds may be chosen as conservative, they are required to be correct for the convergence result to hold. Proposition 2.5 shows that the bounds in (5) can be chosen such that they will be correct with any prespecified probability. Moreover, in Remark 2.6 a probability-1 result is given. These results are applicable if the signals  $\{r_p(t)\}$  are independent of the noise  $\{e_0(t)\}$ , and the noise is a stationary stochastic process. No assumptions on the distribution or colour of the noise  $\{e_0(t)\}$  are needed.

The other conditions of Theorem 2.7 are not very restrictive. The matrix  $R$  will have full column rank if the identification experiment is sufficiently informative and the signals  $\{r_p(t)\}$  have been chosen suitably. These signals should be chosen such that they are correlated with the regression vector  $\{\phi(t)\}$ . If identification takes place in open loop, the (filtered and/or delayed) input signal is a suitable choice. If identification takes place in closed loop, a (filtered and/or delayed) external reference signal is appropriate.

Note that the assumption of quasi-stationarity of the signals  $\{r_p(t)\}$  and  $\{\phi_k(t)\}$  can be relaxed. Basically, it is not necessary that the matrix  $R$  in Theorem 2.7 exist. It is merely essential that the matrix

$$\frac{1}{N} \sum_{t=1}^N \begin{bmatrix} r_1(t)\phi_1(t) & \dots & r_1(t)\phi_n(t) \\ \vdots & & \vdots \\ r_s(t)\phi_1(t) & \dots & r_s(t)\phi_n(t) \end{bmatrix}$$

be bounded and have full column rank for large enough  $N$ . But convergence of this matrix for  $N \rightarrow \infty$  is not needed. Moreover, the assumption that  $c_l(p)$  and  $c_u(p)$  are finite can be relaxed. These bounds are allowed to tend to infinity, provided the divergence is slower than  $\sqrt{N}$ .

*Remark 2.8.* Note that the use of cross-covariance bounds is closely related to the instrumental variable identification method (see Söderström and Stoica, 1989, Chapter 8). In fact, the signals  $\{r_p(t)\}$  can be regarded as

instrumental variables. Also note the close connection of the consistency result of Theorem 2.7 with consistency for instrumental variable identification techniques (see Ljung, 1987, Chapter 8). In the present parameter bounding identification setting, as well as the instrumental variable identification setting, consistency has been shown under fairly general conditions. For example, in both cases there is still consistency if the input signal is correlated with the noise process (closed loop identification).

### 3. ESTIMATING THE CROSS-COVARIANCE BOUNDS FROM DATA

In bounded error identification an a priori specification of noise bounds is required. When time-domain noise constraints (3) are used, the bounds  $e_l(t)$  and  $e_u(t)$  need to be specified. When cross-covariance noise constraints (5) are used, the bounds  $c_l(p)$  and  $c_u(p)$  need to be specified. It may be that these bounds are known a priori, e.g. from physical laws. However, it is not at all impossible that this is not the case, and that measurement data have to be used to establish the noise bounds. Unfortunately, very little attention is paid in the parameter bounding literature to the problem of estimating the noise bounds from data.

In this section the problem is considered of estimating appropriate cross-covariance bounds from data. Proposition 2.5 shows that this boils down to estimating  $\Lambda_p^N$ , which is related to the second-order noise statistics. If knowledge of the noise statistics is not available from physical considerations of the process, measurements have to be used to estimate these. This is a valid procedure if the noise is stationary, i.e. the statistical properties do not change in time. In that case any measurement sequence may be used to estimate the noise statistics. The estimated statistics will then still hold for the measurement sequence used in the parameter bounding identification procedure.

Note that if an exact value for  $\Lambda_p^N$  is not obtainable, an upper bound is still of use. If  $\Lambda_p^N$  is overestimated, the resulting noise bounds  $c_l(p)$  and  $c_u(p)$  are conservative but correct (with a certain specified probability), and the resulting feasible parameter region will be correct, i.e. it will contain the true parameter vector. Even if the noise bounds are chosen as conservative, the consistency result given in Theorem 2.7 remains valid.

Suppose that there is available a measurement sequence generated by (2). As mentioned above, this need not be the same measurement sequence as that used in the parameter bounding

identification procedure: it may be an independent data set. In this section it is assumed that the regression process  $\{\phi(t)\}$  is uncorrelated with the noise process  $\{e_0(t)\}$ . This implies that measurements have to take place in open loop and that the regression vector may only contain filtered and/or delayed samples of the input signal, since the output signal is disturbed by noise. This does not necessarily imply a restriction to FIR identification. Also, identification with Laguerre polynomials (Wahlberg, 1991) and identification with generalized orthonormal polynomials (Heuberger and Bosgra, 1990; Heuberger *et al.*, 1995) fit into this setting. Let there be available a nominal model  $\hat{\theta}$ , which has been obtained independently of the given data set, but for example by physical modelling or identification based on another dataset. The prediction or output error  $\hat{e}(t)$  for this nominal model  $\hat{\theta}$  is given by

$$\hat{e}(t) := y(t) - \phi^T(t)\hat{\theta} = \psi(t) + e_0(t), \quad (7)$$

with

$$\psi(t) := \phi^T(t)(\theta_0 - \hat{\theta}). \quad (8)$$

The idea is to use this prediction error in order to estimate the second-order statistics of the noise process. First some technical assumptions are made with respect to the signal  $\psi(t)$ .

*Assumption 3.1.* The signal  $\{\psi(t)\}$  is quasi-stationary and it satisfies

$$\psi(t) = \bar{\psi}(t) + P(t, q)\tilde{\psi}(t),$$

where  $\{\bar{\psi}(t)\}$  is a bounded deterministic signal,  $\{P(t, q), t = 1, 2, \dots\}$  is a uniformly stable family of filters, and  $\{\tilde{\psi}(t)\}$  is a sequence of independent random variables with zero mean values, variances  $\mu_t$  and bounded fourth moments. Moreover, the autocovariance function of the signal  $\{\psi(t)\}$  is exponentially decaying, i.e.  $R_\psi(\tau) \leq M\rho^\tau \forall \tau$  for certain finite  $M$  and  $\rho < 1$ .

Next denote

$$\hat{R}_\varepsilon^N(\tau) := \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} \hat{e}(t)\hat{e}(t + |\tau|),$$

and consider the following estimate for  $\Lambda_p^N$ :

$$\hat{\Lambda}_p^N = \sum_{\tau=-w(N)}^{w(N)} c_w(\tau) \frac{N - |\tau|}{N} R_{r_p}^N(\tau) \hat{R}_\varepsilon^N(\tau), \quad (9)$$

where  $c_w(\tau)$  is a window function, similar to those used in spectral analysis (see Ljung, 1987, Chapter 6). Notice that the quantities  $R_{r_p}^N(\tau)$  appearing in the estimate  $\hat{\Lambda}_p^N$  are assumed to be known precisely. This is a realistic situation, since the signals  $\{r_p(t)\}$  are generally user-

determined. These signals may be deterministic, and completely known, or stochastic, with known second-order statistics, or mixed deterministic stochastic, with given autocovariance function.

In order to be able to establish a useful convergence result, some technical assumptions are made with respect to the window used in the estimate (9). These assumptions, which are taken from Hjalmarsson (1993, Theorem 3.1), in fact state that the window should converge to 1 as  $N \rightarrow \infty$ , but slow enough in comparison with the number of data  $N$ .

*Assumption 3.2.* The sequence of integers  $\{w(N)\}$  is a positive, monotonically increasing sequence such that, for some  $k > \frac{5}{2}$ ,

$$\lim_{N \rightarrow \infty} \frac{w(N)}{\sqrt{N}/\log^k N} = C$$

for some finite  $C$ . The real-valued window function  $c_w(\tau)$  is such that

$$|c_w(\tau)| < C \quad \forall w, \tau,$$

and

$$\lim_{w \rightarrow \infty} c_w(\tau) = 1 \quad \forall \tau.$$

The following theorem states that, under the given assumptions, the estimate  $\hat{\Lambda}_p^N$  asymptotically overbounds  $\Lambda_p^N$ .

*Theorem 3.3.* Consider  $\Lambda_p^N$  as given in part (i) of Proposition 2.5, and the estimate  $\hat{\Lambda}_p^N$  defined in (9) with a window that satisfies Assumption 3.2. Suppose that  $\{e_0(t)\}$  satisfies Assumption 2.1,  $\{r_p(t)\}$  Assumption 2.2 and  $\{\psi(t)\}$  Assumption 3.1. Moreover, suppose that the signal  $\{\psi(t)\}$  is uncorrelated with  $\{e_0(t)\}$ , and additionally satisfies the weak conditions (2.11), (3.13) and (3.14) of Hjalmarsson (1993). Finally, suppose that the estimate  $\hat{\theta}$  used in (7) has been established independently of the noise process  $\{e_0(t)\}$ . Then

$$(i) \quad \lim_{N \rightarrow \infty} \hat{\Lambda}_p^N = \Lambda_p + \sum_{\tau=-\infty}^{\infty} R_{r_p}(\tau) R_\psi(\tau), \quad \text{w.p. 1,}$$

$$(ii) \quad \sum_{\tau=-\infty}^{\infty} R_{r_p}(\tau) R_\psi(\tau) \geq 0.$$

*Proof.* See the Appendix.

Hence asymptotically correct cross-covariance bounds  $c_l(p)$  and  $c_u(p)$  can be established as the estimated variance  $\hat{\Lambda}_p^N$  is overbiased. The conservatism decreases if the nominal model  $\hat{\theta}$

becomes a more accurate description of the true system  $\theta_0$ .

*Remark 3.4.* For finite  $N$  the estimate  $\hat{\Lambda}_p^N$  has a nonzero variance as well. In fact, it is a kind of spectral estimate, for which it has been shown by Ljung (1987, p. 160) that the variance is asymptotically linearly proportional to  $w(N)/N$ , which tends to zero. Hence the variance is negligible if  $N$  is large enough.

*Remark 3.5.* The most severe restriction in Assumption 3.1 is that  $R_\psi(\tau)$  is exponentially decaying. This means that  $\psi(t)$ , and hence  $\hat{e}(t)$ , is not allowed to contain undecaying deterministic components such as sinusoids. If they are present in the prediction error, they can be detected and removed from the signal, e.g. by taking the DFT of the prediction error, removing the peak values, and taking the inverse DFT of the remaining part. Note that asymptotically this will not influence the contribution of  $\{e_0(t)\}$  to  $\{\hat{e}(t)\}$ .

*Remark 3.6.* It is desirable to introduce as little conservatism as possible when establishing the noise bounds  $c_l(p)$  and  $c_u(p)$ , in order to avoid unnecessarily large parameter sets being identified in the parameter bounding identification procedure. Part (ii) of Theorem 3.3 shows that asymptotically the estimate  $\hat{\Lambda}_p^N$  upper-bounds  $\Lambda_p^N$ . Only in the special case where  $\hat{e}(t)$  equals  $e_0(t)$  is the estimate unbiased, yielding minimal conservatism. However, there is another special situation where it is possible to derive a non-conservative estimate of the noise statistics, with a procedure different from that described above. This is the case if a repeated experiment has been performed, i.e. if

$$\phi(t + T) = \phi(t), \quad t = 1, \dots, T,$$

for some period time  $T$ . If the regression vector only contains (filtered) samples of the input, this is realized by applying a periodic input signal. Now consider the signal

$$\epsilon(t) := [y(t + T) - y(t)]/\sqrt{2}, \quad t = 1, \dots, T,$$

which, with (2), can be written as

$$\begin{aligned} \epsilon(t) &= [\phi(t + T) - \phi(t)]\theta_0/\sqrt{2} \\ &\quad + [e_0(t + T) - e_0(t)]/\sqrt{2} \\ &= [e_0(t + T) - e_0(t)]/\sqrt{2}. \end{aligned}$$

The signal  $\epsilon(t)$  actually appears to have second-order statistics identical to those of the noise process  $e_0(t)$  (in the asymptotic case  $T \rightarrow \infty$ ). Also, if the regression process is not perfectly periodic, e.g. because of different initial

conditions, the given signal  $\epsilon(t)$  can still be used to estimate the second-order noise statistics of  $e_0(t)$ . Using an argument similar to that in Theorem 3.3, it can be shown that the estimate  $\hat{\Lambda}_p^N$  will then be overbiased, yielding correct bounds  $c_l(p)$  and  $c_u(p)$ .

#### 4. FREQUENCY-DOMAIN CONSTRAINTS ON THE NOISE

It appears that consistency results in parameter bounding identification, similar to those of Theorem 2.7, can be obtained with another type of noise constraint. This concerns a frequency-domain bound on the noise. Consider the function  $E(\omega_j)$  defined by

$$E(\omega_j) := \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t)e^{-i\omega_j t}, \quad (10)$$

which for  $\omega_j = 2\pi j/N$ ,  $j = 1, \dots, N$ , is the discrete Fourier transform of the sequence  $\{e(1), \dots, e(N)\}$ . With some abuse of terminology,  $E(\omega_j)$  will be called the discrete Fourier transform of the signal  $\{e(t)\}$ , no matter what  $\omega_j$  is, and keeping in mind that it is dependent on  $N$ . By bounding its amplitude, the square root of the so-called periodogram, the residuals are bounded in the frequency domain. Consider the constraints

$$|E(\omega_j)| \leq f(\omega_j), \quad \omega_j \in [0, \pi], \quad j = 1, \dots, l,$$

for some specified bounds  $f(\omega_j)$ . This type of noise constraint is also used in Lamaire *et al.* (1991) and De Vries and Van den Hof (1992, 1994) in a frequency-domain identification setting. Substituting  $e(t) = y(t) - \phi^T(t)\theta$  gives the feasible parameter set

$$\Theta_N = \left\{ \theta \mid \left| \frac{1}{\sqrt{N}} \sum_{t=1}^N [y(t) - \phi^T(t)\theta]e^{-i\omega_j t} \right| \leq f(\omega_j), \omega_j \in [0, \pi], j = 1, \dots, l, \right\}.$$

If parameter outer-bounding by linear programming is carried out, linear constraints are required. However, the constraints given above are nonlinear owing to the fact that the DFT is a complex-valued quantity. Fortunately it appears possible to approximate each nonlinear constraint by a number of linear constraints. Consider the linear constraints

$$\begin{aligned} f_1(\omega_j, \alpha_{k'}) &\leq \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t) \cos(\omega_j t - \alpha_{k'}) \\ &\leq f_u(\omega_j, \alpha_{k'}), \\ \omega_j &\in [0, \pi], \quad j = 1, \dots, l, \\ \alpha_{k'} &\in [0, \pi], \quad k' = 1, \dots, m, \end{aligned} \quad (11)$$

yielding the feasible parameter set

$$\Theta_N = \left\{ \theta \left| \begin{aligned} & f_l(\omega_j, \alpha_{k'}) \\ & \leq \frac{1}{\sqrt{N}} \sum_{t=1}^N \cos(\omega_j t - \alpha_{k'}) [y(t) - \phi^T(t)\theta] \\ & \leq f_u(\omega_j, \alpha_{k'}), \quad \omega_j \in [0, \pi], j = 1, \dots, l, \\ & \alpha_{k'} \in [0, \pi], k' = 1, \dots, m \end{aligned} \right. \right\}, \quad (12)$$

where  $f_l(\omega_j, \alpha_{k'})$  and  $f_u(\omega_j, \alpha_{k'})$  are specified bounds,  $\omega_j$  specified frequencies and  $\alpha_{k'}$  specified phase shifts. As the constraints in (12) are linear in  $\theta$ , they can easily be included in linear programming problems like those in (4).

The relation of the constraints (11) to the amplitude of the DFT of  $\{e(t)\}$  is investigated. This amplitude can be written as

$$\begin{aligned} |E(\omega_j)| &= \left| \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t) e^{-i\omega_j t} \right| \\ &= \max_{\alpha \in [0, 2\pi]} \operatorname{Re} \left[ e^{i\alpha} \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t) e^{-i\omega_j t} \right] \\ &= \max_{\alpha \in [0, 2\pi]} \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t) \operatorname{Re} (e^{-i\omega_j t + \alpha}) \\ &= \max_{\alpha \in [0, 2\pi]} \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t) \cos(\omega_j t - \alpha), \quad (13) \end{aligned}$$

which shows that the bound (11) approximates the bound on  $|E(\omega_j)|$ . The approximation improves if more phase shifts  $\alpha_{k'}$  are used. This is formalized in the next proposition.

*Proposition 4.1.* Let  $E(\omega_j)$  be defined by (10) and  $\alpha_{k'}$  be given by  $\alpha_{k'} = \pi k' / m$ ,  $k' = 1, \dots, m$ ; then

$$\begin{aligned} \max_{k'} \left| \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t) \cos(\omega_j t - \alpha_{k'}) \right| \\ \leq |E(\omega_j)| \leq \frac{\max_{k'} \left| \frac{1}{\sqrt{N}} \sum_{t=1}^N e(t) \cos(\omega_j t - \alpha_{k'}) \right|}{\cos(\pi/2m)}. \end{aligned}$$

*Proof.* See the Appendix.

Hence this proposition states that (11) bounds the amplitude of the residuals  $\{e(t)\}$  in the frequency domain. This apparently boils down to excluding the residuals to contain sinusoids with the specified frequencies  $\omega_j$ ,  $j = 1, \dots, l$ . For  $\omega_j = 0$ , (11) is a bound on the mean value of the residuals.

The frequency-domain bounds on the noise are especially useful if the system is excited by sinusoids. This follows from the following consistency result.

*Theorem 4.2.* Suppose that the signal  $\{e_0(t)\}$  satisfies the constraints (11) for given and finite  $f_l(\omega_j, \alpha_{k'})$  and  $f_u(\omega_j, \alpha_{k'})$ ,  $j = 1, \dots, l$ ,  $\alpha_{k'} = \pi k' / m$ ,  $k' = 1, \dots, m \geq 2$ . Suppose that  $\{\phi_k(t)\}$  satisfies Assumption 2.3. For each  $j$  and  $k$ , denote

$$\Psi_k(\omega_j) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E \phi_k(t) e^{-i\omega_j t}.$$

If the matrix

$$\Psi = \begin{bmatrix} \operatorname{Re} [\Psi_1(\omega_1)] & \dots & \operatorname{Re} [\Psi_n(\omega_1)] \\ \vdots & & \vdots \\ \operatorname{Re} [\Psi_1(\omega_l)] & \dots & \operatorname{Re} [\Psi_n(\omega_l)] \\ \operatorname{Im} [\Psi_1(\omega_1)] & \dots & \operatorname{Im} [\Psi_n(\omega_1)] \\ \vdots & & \vdots \\ \operatorname{Im} [\Psi_1(\omega_l)] & \dots & \operatorname{Im} [\Psi_n(\omega_l)] \end{bmatrix}$$

has full column rank then the feasible parameter region (12) converges to the true parameter vector  $\theta_0$ ,

$$\lim_{N \rightarrow \infty} \max_{\theta \in \Theta_N} |\theta - \theta_0| = 0,$$

with probability 1.

*Proof.* This is similar to that of Theorem 2.7.

Again this is a consistency result for bounded error identification without requiring tight error bounds. If the values  $f_l(\omega_j, \alpha_{k'})$  and  $f_u(\omega_j, \alpha_{k'})$  are chosen too large, convergence will still take place under the given conditions.

*Remark 4.3.* In general  $\Psi_k(\omega_j)$  is nonzero if  $\{\phi_k(t)\}$  contains a sinusoid with frequency  $\omega_j$ . A sinusoid  $s(t) = a \sin(\omega_j t + \phi)$  has the well-known property that its periodogram is unbounded, since the following relation holds:

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N s(t) e^{-i\omega_j t} \\ = \begin{cases} \frac{1}{2} a (\sin \phi - i \cos \phi), & \omega_j \in (0, \pi), \\ a \sin \phi, & \omega_j = 0, \pi. \end{cases} \end{aligned}$$

Hence the consistency result of Theorem 4.2 will only hold if the system has been excited by a sum of at least  $\frac{1}{2}n$  sinusoids. This type of excitation has also been exploited by De Vries and Van den Hof (1992, 1994) and Bayard (1992).

It is desirable to specify the bounds  $f_l(\omega_j, \alpha_{k'})$  and  $f_u(\omega_j, \alpha_{k'})$  such that they are satisfied by the true noise process  $\{e_0(t)\}$ . The bounds (11) may be regarded as entirely deterministic, which does not require any stochastic assumptions about the noise. However, analogously to the bounds (5), a nice probabilistic interpretation exists when the noise has some stochastic properties. In the following proposition probabilistic properties of the periodogram of the noise are evaluated.

*Proposition 4.4.* Suppose that  $\{e_0(t)\}$  satisfies Assumption 2.1. Let  $E_0(\omega_j)$  be the DFT of  $\{e_0(t)\}$ , defined analogously to (10); then

$$|E_0(\omega_j)|^2 \xrightarrow{N \rightarrow \infty} \begin{cases} \frac{1}{2} \Phi_{e_0}(\omega_j) \chi^2(2), & \omega_j \in (0, \pi), \\ \Phi_{e_0}(\omega_j) \chi^2(1), & \omega_j = 0, \pi, \end{cases}$$

where  $\Phi_{e_0}(\omega_j)$  denotes the autospectrum of the process  $\{e_0(t)\}$ , defined by

$$\Phi_{e_0}(\omega_j) = \sum_{\tau=-\infty}^{\infty} R_{e_0}(\tau) e^{-i\tau\omega_j},$$

and  $\chi^2(n)$  denotes the chi-squared distribution with  $n$  degrees of freedom.

*Proof.* See Brillinger (1981, Theorem 5.2.6); the result can also be established along the lines of the proof of Proposition 2.5.

Again note that the bound (11) is a hard or deterministic bound, which, however, enables a probabilistic interpretation. For example, asymptotically in  $N$ , a 0.999 probability region is obtained by choosing  $f_u(\omega_j, \alpha_{k'}) = -f_l(\omega_j, \alpha_{k'}) = c\sqrt{\Phi_{e_0}(\omega_j)}$ ,  $k' = 1, \dots, m$ , with

$$c = \begin{cases} 2.63, & \omega_j \neq 0, \pi, \\ 3.29, & \omega_j = 0, \pi. \end{cases}$$

*Remark 4.5.* The information required to establish the frequency-domain bounds on the noise consists of  $\Phi_{e_0}(\omega_j)$ ; see Proposition 4.4. Analogously to the procedure of Section 3, it is possible to use the prediction error to (conservatively) estimate the noise spectrum. With the definitions and assumptions of Section 3, it follows that

$$\Phi_\varepsilon(\omega_j) = \Phi_\psi(\omega_j) + \Phi_{e_0}(\omega_j) \geq \Phi_{e_0}(\omega_j),$$

where the latter inequality follows from the fact that an autospectrum is nonnegative. The spectrum of the prediction error can be estimated with standard techniques for spectral estimation (see Brillinger, 1981).

#### 5. TIME-DOMAIN CONSTRAINTS ON THE NOISE REVISITED

It appears that a consistency result similar to those presented above can be achieved with the standard time-domain noise constraints in combination with periodic excitation and measurement averaging. Consider one measurement sequence, split up into  $m+1$  parts, each of length  $T$ :

$$y(t+k'T) = \phi^T(t+k'T)\theta + e(t+k'T),$$

$$t = 1, \dots, T, \quad k' = 0, \dots, m,$$

$$\frac{1}{\sqrt{m+1}} \sum_{k'=0}^m y(t+k'T) = \frac{1}{\sqrt{m+1}} \sum_{k'=0}^m \phi^T(t+k'T)\theta + \frac{1}{\sqrt{m+1}} \sum_{k'=0}^m e(t+k'T).$$

Now, instead of bounding  $e(t)$ , the averaged residuals are bounded by requiring that

$$m_l(t) \leq \frac{1}{\sqrt{m+1}} \sum_{k'=0}^m e(t+k'T) \leq m_u(t), \quad t = 1, \dots, T. \quad (14)$$

These constraints yield the feasible parameter set

$$\Theta_N = \left\{ \theta \mid m_l(t) \leq \frac{1}{\sqrt{m+1}} \sum_{k'=0}^m [y(t+k'T) - \phi^T(t+k'T)\theta] \leq m_u(t), \quad t = 1, \dots, T \right\}, \quad (15)$$

which can be used to set up linear programming problems for parameter bounding like those in (4).

As already mentioned, the idea is to combine the averaging in combination with repeated experiments. The following consistency result shows the benefits of this approach.

*Theorem 5.1.* Suppose that the signal  $\{e_0(t)\}$  satisfies the constraints (14) for given and finite  $m_l(t)$  and  $m_u(t)$ ,  $t = 1, \dots, T$ . Also suppose that  $\{\phi_k(t)\}$  satisfies Assumption 2.3. For each  $k$ , denote

$$M_{\phi_k}(t) := \lim_{m \rightarrow \infty} \frac{1}{m+1} \sum_{k'=0}^m E\phi_k(t+k'T).$$

If the matrix

$$M = \begin{bmatrix} M_{\phi_1}(1) & \dots & M_{\phi_n}(1) \\ \vdots & & \vdots \\ M_{\phi_1}(T) & \dots & M_{\phi_n}(T) \end{bmatrix}$$

has full column rank then the feasible parameter region (15) converges to the true parameter vector  $\theta_0$ ,

$$\lim_{m \rightarrow \infty} \max_{\theta \in \Theta_N} |\theta - \theta_0| = 0,$$

with probability 1.

*Proof.* This is similar to that of Theorem 2.7.

A necessary condition for this consistency property to hold is that there is some deterministic, repetitive component in the regression vector  $\{\phi(t)\}$ . If, for example,  $\phi(t+k'T) = \phi(t) \forall t, k'$  then  $M_{\phi_k}(t) = \phi_k(t) \forall t$ .

*Remark 5.2.* In fact, by applying a periodic input

signal in combination with averaging, the signal-to-noise ratio improves, whereas the number of data samples to be used in the identification procedure remains small. This data reduction is an attractive property, since the complexity of the bounded error identification problem increases with the number of constraints used in the linear programming problems (4). Also, when  $m$  is small, the effect of the procedure described will be that the signal-to-noise ratio improves. Note that this method of improving the signal-to-noise ratio, by applying repeated experiments and data reduction, is not restricted to the set membership identification setting, but can be applied in combination with any identification method.

Again the constraint (14) can be considered entirely deterministic without giving a stochastic interpretation. But stochastic theory can fruitfully be applied when evaluating the meaning of these bounds. The following result is then obtained.

*Proposition 5.3.* Suppose that  $\{e_0(t)\}$  satisfies Assumption 2.1. Then

$$\frac{1}{\sqrt{m+1}} \sum_{k'=0}^m e_0(t+k'T) \xrightarrow{m, T \rightarrow \infty} \mathcal{N}(0, \Lambda), \quad \Lambda = R_{e_0}(0), \quad \forall t.$$

*Proof.* See the Appendix.

This result states that if the number of repetitions increases, the distribution of the averaged noise converges to the normal distribution.

*Remark 5.4.* The information required to establish the time-domain bounds on the noise in combination with repeated experiments consists of  $R_{e_0}(0)$ . Analogously to the results of Section 3, the estimate  $R_{\hat{e}}(0)$ , which is based on the prediction error, can be used as a conservative estimate of this quantity as

$$R_{\hat{e}}(0) = R_{\psi}(0) + R_{e_0}(0) \geq R_{e_0}(0).$$

6. EXAMPLE

Consider the data-generating system

$$y(t) = \theta_0^{(1)}u(t) + \theta_0^{(2)}u(t-1) + \theta_0^{(3)}u(t-2) + e_0(t),$$

$$\theta_0^{(1)} = 2, \quad \theta_0^{(2)} = 1, \quad \theta_0^{(3)} = 0.6,$$

with the noise process given by

$$e_0(t) = n(t) + 0.8n(t-1) + 0.2n(t-2) + 0.1n(t-3),$$

where  $\{n(t)\}$  is a white-noise process uniformly distributed between  $-0.25$  and  $0.25$ . The input signal  $\{u(t)\}$  is chosen to be

$$u(t) = \begin{cases} 0, & t \leq 0, \\ \sin \frac{1}{4}\pi t + \sin \frac{1}{2}\pi t, & t = 1, \dots, N; N = 800. \end{cases}$$

The following parameter bounding identification procedures have been carried out.

1. *Parameter bounding with bounds on the amplitude of the noise as in (3).* Since  $|e_0(t)| \leq 0.25(1 + 0.8 + 0.2 + 0.1) = 0.525$ , the bounds have been chosen as  $e_u(t) = -e_l(t) = 0.525 \forall t$ . Altogether, 6 linear programming problems have to be solved, each with 3 unknowns subject to 1600 inequality constraints.
2. *Parameter bounding with bounds on the cross-covariance of the noise as in (5).* The signals  $\{r_p(t)\}$  have been chosen as  $r_p(t) = u(t + \tau_p)$ ,  $\tau_p = 1 - p$ ,  $p = 1, \dots, 8$ , which, with Proposition 2.5, yields  $\Lambda_p = 0.0388$ ,  $p = 1, \dots, 8$ . The bounds  $c_u(p) = -c_l(p) = 3\sqrt{\Lambda_p}$ ,  $p = 1, \dots, 8$ , have been chosen; hence each bound is satisfied with probability 0.997. Altogether, 6 linear programming problems have to be solved, each with 3 unknowns and 16 constraints. According to the rule of Bonferroni (Manoukian, 1986), the resulting parameter region is correct with a probability larger than or equal to  $1 - 8(1 - 0.997) = 0.98$ .
3. *Parameter bounding with bounds on the periodogram of the noise as in (11).* The frequencies  $\omega_1 = \frac{1}{4}\pi$  and  $\omega_2 = \frac{1}{2}\pi$  were selected, and  $\alpha_{k'} = \pi k' / m$ ,  $k' = 1, \dots, 4$ . Straightforward calculations show that  $\Phi_{e_0}(\omega_1) = 0.2473$  and  $\Phi_{e_0}(\omega_2) = 0.1534$ . The bounds

$$f_u(\omega_j, \alpha_{k'}) = -f_l(\omega_j, \alpha_{k'}) = 2.146\sqrt{\Phi_{e_0}(\omega_j)},$$

$$j = 1, 2, \quad k' = 1, \dots, 4,$$

have been chosen, which corresponds to a 0.99 confidence interval for each frequency separately. Hence 6 linear programming problems have to be solved, each with 3 unknowns subject to 16 constraints. The resulting parameter region is correct with probability larger than or equal to  $1 - 2(1 - 0.99) = 0.98$ .

4. *Parameter bounding with bounds on the averaged noise as in (14).* The input sequence is periodic, with a period of 8 samples. The

Table 1. Results of parameter bounding identification

|              | $\theta_1^{(l)}$ | $\theta_1^{(u)}$ | $\theta_2^{(l)}$ | $\theta_2^{(u)}$ | $\theta_3^{(l)}$ | $\theta_3^{(u)}$ |
|--------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Procedure 1  | 1.890            | 2.117            | 0.862            | 1.127            | 0.468            | 0.707            |
| Procedure 2  | 1.964            | 2.044            | 0.963            | 1.046            | 0.557            | 0.637            |
| Procedure 3  | 1.954            | 2.048            | 0.981            | 1.028            | 0.554            | 0.648            |
| Procedure 4  | 1.959            | 2.062            | 0.955            | 1.056            | 0.542            | 0.644            |
| Exact values | 2.0              |                  | 1.0              |                  | 0.6              |                  |

last 99 periods were averaged; the first period has been used to settle the initial conditions of the system. The bounds have been set to

$$m_u(t) = -m_l(t) = 3\sqrt{R_{e_0}(0)} = 0.563, \\ t = 1, \dots, 8.$$

This means that each linear programming problem has 3 unknowns and 16 inequality constraints. The probability that  $\theta_0$  is an element of the resulting parameter region is at least  $1 - 8(1 - 0.997) = 0.98$ .

The resulting upper and lower bounds on the parameters are shown in Table 1. It is concluded that for the first and third parameter the tightest bounds are obtained with the cross-covariance bounds on the noise; for the second parameter this is the case with frequency-domain bounds on the noise. The parameter uncertainty intervals estimated with cross-covariance noise constraints (Procedure 2) are a factor 3 smaller than those estimated with time-domain noise constraints (Procedure 1). The uncertainty interval for the second parameter estimated with the frequency-domain noise constraints (Procedure 3) is even a factor 5.7 smaller than the corresponding interval for Procedure 1. The new parameter bounding identification methods 2, 3 and 4 appear to outperform the standard parameter bounding identification Procedure 1. Much smaller parameter uncertainty intervals are estimated with these new methods compared with the result of identification with bounds on the amplitude of the noise in the time domain.

*Remark 6.1.* For simplicity, exact knowledge of the second-order noise statistics has been used to establish the bounds on the noise. The procedure of Section 3 may be used to estimate these statistics. In particular, the method indicated in Remark 3.6 is applicable, since the input is periodic.

## 7. CONCLUSIONS

Several valuable alternatives have been presented for time-domain bounds on the noise in parameter bounding identification by linear programming. In particular, cross-covariance bounds on the noise are powerful, since consistency has been proved for an arbitrary

persistently exciting input signal. Frequency-domain bounds on the noise are powerful if the input signal contains sinusoids, in which case consistency has been proved as well. Consistency has also been proved in the case where time-domain bounds on the noise are combined with periodic excitation and averaging. In all these cases the number of constraints does not increase with increasing number of measurements; hence the identification problem remains tractable for large  $N$ . The example showed a considerable reduction of parameter uncertainty when the new types of noise constraints are utilized.

A stochastic analysis has been presented of the new types of noise bounds. The basic assumptions needed to justify this analysis are that the number of samples is large enough and that the noise is stationary, i.e. its stochastic properties do not change in time. The analysis showed that the bounds for the noise can be specified such that they are correct with a certain probability. This means that a parameter set estimate is calculated that contains the true parameter vector with a certain probability. It is emphasized that the probability density function of the noise is arbitrary and need not be known. Only knowledge of the second-order statistics of the noise process is required. A procedure has been presented to estimate these statistics from data.

In Hakvoort (1992, 1993, 1994a) bounded error identification with linear programming techniques is applied to identification in  $\ell_1$  and  $H_\infty$ . The alternative noise bounds presented here can be fruitfully applied in those settings. In those papers the problem of undermodelling is also considered. More details can also be found in Hakvoort (1994b).

## REFERENCES

- Bayard, D. S. (1992). Statistical plant set estimation using Schroeder-phased multi-sinusoidal input design. In *Proc. American Control Conf.*, Chicago, IL, pp. 2988–2995.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.
- De Vries, D. K. and P. M. J. Van den Hof (1992). Quantification of model uncertainty from data: input design, interpolation and connection with robust control design specifications. *Proc. American Control Conf.*, Chicago, IL, pp. 3170–3175.
- De Vries, D. K. and P. M. J. Van de Hof (1994). Quantification of model uncertainty from data. *Int. J. Robust and Nonlinear Control*, **4**, 301–319.
- Fogel, E. (1979). System identification via membership set constraints with energy constrained noise. *IEEE Trans. Autom. Control*, **AC-24**, 752–758.
- Fogel, E. and Y. F. Huang (1982). On the value of information in system identification—bounded noise case. *Automatica*, **18**, 229–238.
- Gu, G. and P. P. Khargonekar (1992). Linear and nonlinear algorithms for identification in  $H_\infty$  with error bounds. *IEEE Trans. Autom. Control*, **AC-37**, 953–963.

- Gustafsson, T. K. and P. M. Mäkilä (1993). Modelling of uncertain systems via linear programming. In *Preprints of 12th IFAC World Congress*, Sydney, Vol. 5, pp. 293–298.
- Hakvoort, R. G. (1992). Worst-case system identification in  $\xi_1$ : error bounds, optimal models and model reduction. In *Proc. 31st IEEE Conf. on Decision and Control*, Tucson, AZ, pp. 499–504.
- Hakvoort, R. G. (1993). Worst-case system identification in  $H_\infty$ : error bounds and optimal models. In *Preprints 12th IFAC World Congress*, Sydney, Vol. 8, pp. 161–164.
- Hakvoort, R. G. (1994a). A linear programming approach to the identification of frequency domain error bounds. In *Preprints of SYSID'94 10th IFAC Symp. on System Identification*, Copenhagen, Vol. 2, pp. 195–200.
- Hakvoort, R. G. (1994b). System identification for robust process control: nominal models and error bounds. PhD thesis, Mechanical Engineering Systems and Control Group, Delft University of Technology.
- Hakvoort, R. G., P. M. J. Van den Hof and O. H. Bosgra (1993). Consistent parameter bounding identification using cross-covariance constraints on the noise. In *Proc. 32nd IEEE Conf. on Decision and Control*, San Antonio, TX, pp. 2601–2606.
- Helmicki, A. J., C. A. Jacobson and C. N. Nett (1991). Control-oriented system identification: a worst-case/deterministic approach in  $H_\infty$ . *IEEE Trans. Autom. Control*, **AC-36**, 1163–1176.
- Heuberger, P. S. C. and O. H. Bosgra (1990). Approximate system identification using system based orthonormal functions. In *Proc. 29th IEEE Conf. on Decision and Control*, Honolulu, HI, pp. 1086–1092.
- Heuberger, P. S. C., P. M. J. Van den Hof and O. H. Bosgra (1995). A generalized orthonormal basis for linear dynamical systems. *IEEE Trans. Autom. Control*, **AC-40**, No. 3. An abbreviated version appeared in *Proc. 32nd IEEE Conf. on Decision and Control*, San Antonio, TX, 1993, pp. 2850–2855.
- Heunis, A. J. (1988). Asymptotic properties of prediction error methods in approximate system identification. *Stochastics*, **24**, 1–43.
- Hjalmarsson, H. (1993). Aspects on incomplete modelling in system identification, PhD thesis, Department of Electrical Engineering, Linköping University.
- 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. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
- Ljung, L. and B. Wahlberg (1992). Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Adv. Appl. Prob.*, **24**, 412–440.
- Mäkilä, P. M. (1991). Robust identification and Galois-sequences. *Int. J. Control*, **54**, 1189–1200.
- Manoukian, E. B. (1986). *Modern Concepts and Theorems of Mathematical Statistics*. Springer-Verlag, New York.
- Milanese, M. and G. Belforte (1982). Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: linear families of models and estimators. *IEEE Trans. Autom. Control*, **AC-27**, 408–414.
- Milanese, M. and A. Vicino (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, **27**, 997–1009.
- Norton, J. P. and S. M. Veres (1991). Developments in parameter bounding. In L. Gerencser and P. E. Caines (Eds), *Topics in Stochastic Systems: Modelling, Estimation and Adaptive Control*, pp. 137–158. Lecture Notes in Control and Information Sciences, Vol. 161, Springer-Verlag, London.
- Söderström, T. and P. Stoica (1989). *System Identification*. Prentice-Hall, Hemel Hempstead, U.K.
- Veres, S. M. and J. P. Norton (1989). Structure identification of parameter bounding models by use of noise structure bounds. *Int. J. Control*, **50**, 639–649.
- Veres, S. M. and J. P. Norton (1991). Structure selection for bounded parameter models: consistency conditions and selection criterion. *IEEE Trans. Autom. Control*, **AC-36**, 474–481.

- Walter, E. and H. Piet-Lahanier (1990). Estimation of parameter bounds from bounded error data: a survey. *Math. Comput. Simul.*, **32**, 449–468.
- Wahlberg, B. (1991). System identification using Laguerre models. *IEEE Trans. Autom. Control*, **AC-36**, 551–562.

## APPENDIX—PROOFS

*Proof of Proposition 2.5*

Since  $\{e_0(t)\}$  is a stationary stochastic process and is independent of the quasi-stationary process  $\{r_p(t)\}$ , the product  $\{e_0(t)r_p(t)\}$  is also quasi-stationary, both in the sense of Ljung (1987) and in the sense of Hjalmarsson (1993), since the condition (2.11) in the latter reference disappears because  $Ee_0(t)r_p(t) = 0$ . For the same reason, the definitions of the covariance function coincide. The autocovariance function of  $\{e_0(t)r_p(t)\}$  is equal to  $R_{e_0}(\tau)R_{r_p}(\tau)$ .

As the filters  $H_0(q)$  and  $R_p(t, q)$ ,  $t = 1, 2, \dots$ , are exponentially stable, and the stochastic terms of  $\{e_0(t)\}$  and  $\{r_p(t)\}$  have bounded fourth moments, these stochastic terms are exponentially forgetting processes of order 4 (Hjalmarsson, 1993, p. 48). Furthermore, since these stochastic parts are independent, their product is again exponentially forgetting of order 4 (Hjalmarsson, 1993, Lemma 2.4). Hence  $\{e_0(t)r_p(t)\}$  is exponentially forgetting of order 4, since multiplication with bounded deterministic signals does not change this property.

As  $\{e_0(t)r_p(t)\}$  is quasi-stationary and exponentially forgetting of order 4, part (ii) of Lemma 2.17 in Hjalmarsson (1993) can be applied, which immediately yields result (ii) of this proposition. Corollary 2.8.1 in Hjalmarsson (1993) gives (iii).

Again using the fact that  $\{r_p(t)\}$  and  $\{e_0(t)\}$  are independent, and using the substitution  $\tau = t' - t$ , it is possible to write for  $\Lambda_p^N$ ,

$$\begin{aligned} \Lambda_p^N &= E \frac{1}{N} \sum_{t=1}^N \sum_{t'=1}^N r_p(t) e_0(t) r_p(t') e_0(t') \\ &= \frac{1}{N} \sum_{t=1}^N \sum_{\tau=1-t}^{N-t} E r_p(t) e_0(t) r_p(t+\tau) e_0(t+\tau) \\ &= \frac{1}{N} \sum_{\tau=-N+1}^0 \sum_{t=-\tau+1}^N E r_p(t) r_p(t+\tau) R_{e_0}(\tau) \\ &\quad + \frac{1}{N} \sum_{\tau=1}^{N-1} \sum_{t=1}^{N-\tau} E r_p(t) r_p(t+\tau) R_{e_0}(\tau) \\ &= \frac{1}{N} \sum_{\tau=-N+1}^{N-1} R_{e_0}(\tau) \sum_{t=1}^{N-|\tau|} E r_p(t) r_p(t+|\tau|) \\ &= \frac{1}{N} \sum_{\tau=-N+1}^{N-1} R_{e_0}(\tau) (N-|\tau|) R_{r_p}^N(\tau), \end{aligned}$$

which proves part (i) of this proposition.  $\square$

It is also possible to give a proof of this proposition along the lines of that of Theorem 9.1 in Ljung (1987). This has been carried out in Hakvoort (1994b, Chapter 3).

*Proof of Theorem 2.7*

Under the given assumptions on  $\{r_p(t)\}$  and  $\{\phi_k(t)\}$ , Theorem 2B.1 in Ljung (1987) is applicable, and consequently, with probability 1,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N r_p(t) \phi_k(t) = R_{r_p \phi_k}(0) \quad \forall p, k.$$

Combination of (1) and (2) gives  $\phi^T(t)(\theta_0 - \theta) = e(t) - e_0(t)$ , and hence

$$\begin{aligned} \frac{1}{N} \sum_{t=1}^N r_p(t) \phi^T(t) (\theta_0 - \theta) \\ = \frac{1}{N} \sum_{t=1}^N r_p(t) e(t) - \frac{1}{N} \sum_{t=1}^N r_p(t) e_0(t), \quad p = 1, \dots, s. \end{aligned}$$

By assumption,  $\theta = \theta_0$  is a feasible solution of this set of constraints, since  $e(t) = e_0(t) \forall t$  is feasible. Taking the limit

as  $N \rightarrow \infty$  gives

$$R(\theta_0 - \theta) = \lim_{N \rightarrow \infty} \frac{c}{\sqrt{N}},$$

$$c_p = \frac{1}{\sqrt{N}} \sum_{t=1}^N r_p(t)e(t) - \frac{1}{\sqrt{N}} \sum_{t=1}^N r_p(t)e_0(t), \quad p = 1, \dots, s.$$

Here  $c$  is an  $s$ -dimensional vector with elements  $c_p$ ,  $p = 1, \dots, s$ . The first contribution to  $c_p$  is bounded by the parameter bounding identification constraint (5). The second contribution to  $c_p$  is bounded by assumption, and hence  $c_l(p) - c_u(p) \leq c_p \leq c_u(p) - c_l(p)$ . Finally, the matrix  $R$  has full column rank. Hence there exists a nonsingular square matrix  $R_s$ , consisting of  $n$  rows of  $R$ , such that for any feasible  $\theta$  as  $N \rightarrow \infty$ ,

$$R_s(\theta_0 - \theta) = \lim_{N \rightarrow \infty} \frac{c}{\sqrt{N}} \Leftrightarrow \theta_0 - \theta = \lim_{N \rightarrow \infty} \frac{R_s^{-1}c}{\sqrt{N}} = 0,$$

since both  $R_s^{-1}$  and  $c$  are bounded. □

*Proof of Theorem 3.3*

The facts that  $\{\psi(t)\}$ , given in (8), is uncorrelated with  $\{e_0(t)\}$  and that the latter process is stationary give

$$E\hat{R}_\varepsilon^N(\tau) = R_\varepsilon^N(\tau) = R_\psi^N(\tau) + R_{e_0}(\tau), \quad |\tau| \leq N - 1.$$

Hence, with (9),

$$E\hat{\Lambda}_p^N = \sum_{\tau=-w}^w c_w(\tau) \frac{N - |\tau|}{N} R_p^N(\tau) [R_\psi^N(\tau) + R_{e_0}(\tau)].$$

As  $\{r_p(t)\}$  is quasi-stationary,  $R_p^N(\tau)$  is bounded. Moreover the autocovariance function of the prediction error,

$$R_\varepsilon(\tau) = \lim_{N \rightarrow \infty} R_\varepsilon^N(\tau) = R_\psi(\tau) + R_{e_0}(\tau),$$

shows an exponential decay rate in  $\tau$ , since both  $R_\psi(\tau)$  and  $R_{e_0}(\tau)$  show exponential decay rates. Also, the following asymptotic relation holds for  $R_p^N(\tau)$  defined in Proposition 2.5:

$$\lim_{N \rightarrow \infty} R_p^N(\tau) = R_p(\tau) \quad \forall |\tau| \leq w,$$

where  $w$  is allowed to tend to infinity as long as  $w/N$  tends to 0, which is assumed to be the case. In combination with the facts that

$$\lim_{N \rightarrow \infty} \frac{N - |\tau|}{N} = 1 \quad \forall |\tau| \leq w,$$

that  $R_\varepsilon(\tau)$  shows exponential decay rate, and that the window  $c_w(\tau)$  satisfies Assumption 3.2, this gives

$$\lim_{N \rightarrow \infty} E\hat{\Lambda}_p^N = \sum_{\tau=-\infty}^{\infty} R_p(\tau)R_{e_0}(\tau) + \sum_{\tau=-\infty}^{\infty} R_p(\tau)R_\psi(\tau)$$

$$= \Lambda_p + \sum_{\tau=-\infty}^{\infty} R_p(\tau)R_\psi(\tau).$$

Under the given assumptions on  $\{e_0(t)\}$ ,  $\{\psi(t)\}$  and the window  $c_w(\tau)$ , Theorem 3.1 in Hjalmarrsson (1993) can be applied, which gives

$$\lim_{N \rightarrow \infty} \hat{\Lambda}_p^N = \lim_{N \rightarrow \infty} E\hat{\Lambda}_p^N, \quad \text{w.p. 1.}$$

Finally, because the autocovariance function of  $\psi(t)$  is exponentially decaying, there exists a stationary stochastic process  $\{s(t)\}$  with exponentially decaying autocovariance function for which  $R_s(\tau) = R_\psi(\tau) \quad \forall \tau$ . Without loss of generality, it may be assumed that  $\{s(t)\}$  is independent of  $\{r_p(t)\}$ . Define the signal  $p(t) = r_p(t)s(t)$ ; then

$$\sum_{\tau=-\infty}^{\infty} R_p(\tau)R_\psi(\tau)$$

$$= \sum_{\tau=-\infty}^{\infty} R_p(\tau)R_s(\tau)$$

$$= \sum_{\tau=-\infty}^{\infty} E s(t)s(t + \tau) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E r_p(t)r_p(t + \tau)$$

$$= \sum_{\tau=-\infty}^{\infty} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E s(t)r_p(t)s(t + \tau)r_p(t + \tau)$$

$$= \sum_{\tau=-\infty}^{\infty} R_p(\tau) = \Phi_p(0) \geq 0,$$

where  $\Phi_p(\omega)$  denotes the autospectrum of  $\{p(t)\}$ . This completes the proof. □

*Proof of Proposition 4.1*

It follows from (13) that it is sufficient to prove that, for any complex number  $x$ ,

$$\max_{k'} |\operatorname{Re}(e^{i\alpha k' x})| \leq |x| \leq \frac{\max_{k'} |\operatorname{Re}(e^{i\alpha k' x})|}{\cos(\pi/2m)}.$$

The left-hand inequality follows from the fact that, for any  $k'$ ,

$$|\operatorname{Re}(e^{i\alpha k' x})| \leq |e^{i\alpha k' x}| = |x|.$$

The right-hand inequality is equivalent to the statement

$$|x| \leq \frac{\max_{k'=1, \dots, 2m} \operatorname{Re}(e^{i\alpha k'/m x})}{\cos(\pi/2m)}.$$

Now for any  $x$  there exists an integer  $l$  and  $\delta \in [-\pi/2m, \pi/2m]$  such that  $x = |x| e^{i(\pi l/m + \delta)}$ , yielding

$$\operatorname{Re}(e^{i\alpha k'/m x}) = \operatorname{Re}(e^{i\alpha k'/m |x|} e^{i\alpha k' l/m + \delta})$$

$$= |x| \operatorname{Re}(e^{i\alpha(k' + l)/m + \delta})$$

$$= |x| \cos\left[\frac{\pi(k' + l)}{m} + \delta\right].$$

If  $k' = k^*$  is now chosen such that  $k^* = l = 2nm$  for some integer  $n$ , this gives

$$\operatorname{Re}(e^{i\alpha k^*/m x}) = |x| \cos(2\pi n + \delta) = |x| \cos \delta \geq |x| \cos\left(\frac{\pi}{2m}\right)$$

$$\Leftrightarrow |x| \leq \frac{\operatorname{Re}(e^{i\alpha k^*/m x})}{\cos(\pi/2m)},$$

which proves the statement. □

*Proof of Proposition 5.3*

Substitution of  $r_p(t) = 1 \quad \forall t$  into Proposition 2.5 immediately gives that,  $\forall t, T$ ,

$$\frac{1}{\sqrt{m+1}} \sum_{k'=0}^m e_0(t + k'T) \xrightarrow{m \rightarrow \infty} \mathcal{N}(0, \Lambda_T),$$

with

$$\Lambda_T = \lim_{m \rightarrow \infty} E \left[ \frac{1}{\sqrt{m+1}} \sum_{k'=0}^m e_0(t + k'T) \right]^2$$

$$= \lim_{m \rightarrow \infty} \frac{1}{m+1} \sum_{k'=0}^m \sum_{k''=0}^m E e_0(t + k'T) e_0(t + k''T)$$

$$= \lim_{m \rightarrow \infty} \frac{1}{m+1} \sum_{k'=0}^m \sum_{k''=0}^m R_{e_0}((k' - k'')T),$$

and, owing to the exponential decay rate of  $R_{e_0}(\tau)$ ,

$$\Lambda = \lim_{T \rightarrow \infty} \Lambda_T = \lim_{m \rightarrow \infty} \frac{1}{m+1} \sum_{k'=0}^m \sum_{k''=0}^m \lim_{T \rightarrow \infty} R_{e_0}((k' - k'')T)$$

$$= \lim_{m \rightarrow \infty} \frac{1}{m+1} \sum_{k'=0}^m R_{e_0}(0) = R_{e_0}(0). \quad \square$$