Quantification of the Variance of Estimated Transfer Functions in the Presence of Undermodeling

Roland Hildebrand and Michel Gevers

***Abstract*—We study the effect of undermodeling on the parameter variance for prediction error time-domain identification in open loop. We consider linear time-invariant discrete time single-input–single-output systems with known noise model. We examine asymptotic expressions for the variance for large number of data. This quantity depends in general on the fourth order statistical properties of the applied input. However, we establish a sufficient condition under which the asymptotic variance depends on the input power spectrum only. For this case, we deliver exact expressions. For a stochastic input the undermodeling contributes to the parameter variance due to the correlation between the prediction errors and its gradients, while for a deterministic input it has no influence. As an additional contribution, we investigate the parameter variance under the assumptions of the stochastic embedding procedure.**

***Index Terms*—Asymptotic variance, bias, parameter estimation, stochastic embedding.**

I. Introduction

Identification experiments should deliver along with an identified model also an uncertainty region, which specifies the quality of the model. Without this additional information, the model is virtually useless for practical purposes. Within the framework of parametric model structures the uncertainty is usually expressed in terms of the covariance of the identified parameter vector. Often, it is sufficient to consider the asymptotic variance as the number of data tends to infinity, since for common data record lengths these expressions are of satisfying accuracy.

In this note, we consider asymptotic variance expressions for discrete-time, linear, and time-invariant systems. Under mild restrictions on input and noise and under the assumption that the true system dynamics can be exactly reproduced within the model structure, the asymptotic variance expressions are tractable functions of the input power spectrum [7]. In the presence of undermodeling, however, the situation is considerably more complicated. Although closed-form expressions for the asymptotic variance are well-known [7], they are in general intractable. Moreover, they depend on higher order statistical properties of the noise and the input [8]. Basically, both the time domain and the frequency domain approach face the same difficulties when computing exact expressions for the asymptotic parameter covariance. However, in the past decade much advance was made to overcome these problems.

Several results on the asymptotic variance for parametric frequency domain identification were obtained. In [2], the contribution of the noise to the parameter variance was computed, while the contribution of the undermodeling was neglected. In [1], linear model structures and a deterministic input were assumed. In this case, the expressions for the asymptotic parameter variance somewhat simplify. In [10], the prediction errors were assumed to be uncorrelated with their gradients, which also facilitates computations.

For time-domain identification estimators of the asymptotic variance based on input–output data of the experiment were proposed. In [5] and [13], different techniques were presented to obtain a sample estimate of the parameter variance from data gathered during the experiment. One method consisted in introducing an exponential forgetting factor in the expression for the parameter estimate. This led to a windowing effect, which in turn yielded certain ergodicity properties of the so-obtained sequence of parameter estimates. For long data records the sample covariance of this sequence was a good approximation of the true parameter covariance. Another method consisted in estimating the undermodeling with a high-order ARX model and using bootstrap techniques to obtain artificial noise realizations for a Monte Carlo simulation.

These techniques thus assume the availability of input–output data. However, sometimes it is necessary to estimate the parameter variance prior to the experiment, e.g., for purposes of optimal input design. In this framework, one is going to perform an identification experiment and wishes to choose the input sequence for this experiment in order to let the parameter variance have some desired properties. Therefore, one has to know how the parameter variance depends on that input. In most cases, it is sufficient to describe the asymptotic parameter variance as a function of the input power spectrum.

In [6], these variance expressions were derived for closed-loop time domain identification asymptotically as the model order tends to infinity. In [4], expressions for the asymptotic variance were computed that were particularly suited for a posteriori error estimation.

In this note, we address the question of computing asymptotic variance expressions in the presence of undermodeling for prediction error time domain identification in open loop and for finite-model order. We further elaborate the expressions derived in [4] and focus on the dependence of the variance on the input power spectrum. For ease of treatment we restrict our considerations to the single-input–single-output case with known noise properties.

If a stochastic input is used, the asymptotic variance depends in general on the undermodeling as well as on higher order properties of the input. It is known that for parametric frequency domain identification these dependencies do not hold, if a deterministic input is applied [8]. We establish a similar result for time domain identification. A qualitative investigation of the asymptotic parameter variance for time-domain identification was conducted in [9]. It was shown that for deterministic input the variance vanishes as the noise level tends to zero, which is confirmed by our results.

For the case of a stochastic input, we formulate a condition under which the asymptotic parameter variance does not depend on higher order properties of the input. This condition covers a wide class of input sequences and is satisfied, e.g., for filtered Gaussian white noise. Under this condition, we establish explicit expressions for the asymptotic parameter variance as a function of the input power spectrum. We show that the contribution of the undermodeling to the parameter covariance has its origin in the correlation between the prediction errors and its gradients. While this correlation vanishes at lag zero by the nature of the prediction error identification procedure, it is in general nonzero at the other lags if undermodeling is present.

In [3], a method called *stochastic embedding* was introduced. Within this framework, the undermodeling is treated as being stochastic with zero mean. Hence, the undermodeling error can be treated as a variance

Manuscript received October 10, 2002; revised December 29, 2003. Recommended by Associate Editor E. Bai. This work was supported in part by The European Commission. The support is provided via the Program Training and Mobility of Researchers (TMR) and Project System Identification (ERB FMRX CT98 0206) to the European Research Network System Identification (ERNSI). This work presents research results of the Belgian Programme on Interuniversity Poles of Attraction, Phase V, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors.

R. Hildebrand is with the Laboratoire de Modélization et Calcul, Université Joseph Fourier, 38400 St Martin d'Hères, France (e-mail: roland.hildebrand@imag.fr).

M. Gevers is with the Center for Systems Engineering and Applied Mechanics (CESAME), Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium (e-mail: gevers@csam.ucl.ac.be).

Digital Object Identifier 10.1109/TAC.2004.832651

error along with the error introduced by the noise. Frequency domain identification by means of frequency response function measurements within the stochastic embedding framework was investigated in [12].

In this note, we also consider the parameter variance in the framework of stochastic embedding. For ease of treatment we restrict our consideration to linear model structures. We derive asymptotic expressions for the total variance and show by an example that the total parameter variance does not necessarily decrease when the input power is increased and can even increase. Thus, the usual property of monotonicity of the variance with respect to the input power spectrum is not satisfied.

The remainder of this note is structured as follows. In Section II, we give formal definitions and remind the expressions for the asymptotic variance as given in [7]. We prove that the undermodeling has no effect on the asymptotic parameter variance if a zero mean periodic input is used. In Section III, we consider the case of a stochastic input. We establish a condition under which the parameter variance is independent of the higher order properties of the input. Assuming this condition, we deduce expressions for the parameter variance as a function of the input power spectrum. Section IV, is devoted to the investigation of the variance expressions in the framework of stochastic embedding. In Section V, we give an example, which shows that under adoption of the stochastic embedding paradigms the information matrix need not be monotonic with respect to the input power spectrum. Finally, we give some conclusions in Section VI.

II. General Asymptotic Variance Expressions

In this section, we restate general asymptotic variance expressions in the presence of undermodeling [7]. We show that undermodeling may have a worsening effect on the variance only in the case of a stochastic input.

Let the true system be given by

$$y = G_0 u + He$$

and let $\theta \in \mathbf{R}^n$ be the parameter vector in the model structure $G(\theta)$. Here, u is the scalar input, y the scalar output, G_0 is the transfer function of the system, and e is white noise with variance λ_0, which is filtered through the monic stable and inversely stable noise filter H. We assume H to be known. The input u is assumed to be a quasi-stationary sequence with zero time average and with power spectrum Φ_u. The filters G_0 and H are functions of the forward shift operator q. They are represented as analytic functions of the complex parameter z.

Identification of θ is performed in open loop by minimizing the squared deviation of the output y from the 1-step ahead predictor $\hat{y}(\theta) = (1 - H^{-1})y + H^{-1}G(\theta)u$. The prediction error is given by $\varepsilon(\theta) = y - \hat{y}(\theta) = H^{-1}y - H^{-1}G(\theta)u$. The identified parameter vector $\hat{\theta}_N$ minimizes the cost function $V_N(\theta) = (1/2N)\sum_{t=1}^{N} \varepsilon_t^2$, where t indexes the time instants and N is the number of data samples, $\hat{\theta}_N = \arg\min_\theta V_N(\theta)$. Under mild assumptions (see [7] for details), the time average $\bar{V}(\theta) = \lim_{N\to\infty} V_N(\theta)$ is defined and $\hat{\theta}_N$ tends to the minimizer θ^* of $\bar{V}(\theta)$ as the number of data N tends to ∞

$$\theta^* = \arg\min_\theta \bar{V}(\theta) \qquad \lim_{N\to\infty} \hat{\theta}_N = \theta^* \text{ with prob. 1.}$$

It is well known that the vector θ^* is not completely determined by the properties of the system, but depends on the input u, specifically on its power spectrum Φ_u (see [7] and [8]).

The vector θ^* minimizes the variance of $\varepsilon(\theta) - e = H^{-1}(G_0 - G(\theta))u$. Hence, θ^* admits the following frequency domain interpretation. Define a pseudoscalar product on the space $\mathcal{H}_\infty$ of stable transfer functions by

$$\langle A, B\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_u(\omega)}{|H(e^{j\omega})|^2} A(e^{j\omega})B^*(e^{j\omega})\, d\omega. \tag{1}$$

Then, θ^* corresponds to that transfer function $G(\theta^*)$ within the model structure that realizes the minimal distance to the true transfer function G_0 with respect to the pseudoscalar product (1). In other words, the mismatch $G_0 - G(\theta^*)$ is orthogonal to the model structure at $G(\theta^*)$ with respect to (1).

Denote the predictor gradient at θ^* by $\psi, \psi = -(\partial\varepsilon/\partial\theta) = H^{-1}(\partial G/\partial\theta)u = H^{-1}G'(\theta^*)u$, and the difference between the residual signal ε at $\theta = \theta^*$ and the noise e by $\tilde{\varepsilon}, \tilde{\varepsilon} = \varepsilon(\theta^*) - e = H^{-1}(G_0 - G(\theta^*))u$. Then, we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_u}{|H|^2} G'(\theta^*)(G_0 - G(\theta^*))^*\, d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{\psi\tilde{\varepsilon}}\, d\omega = 0. \tag{2}$$

Here, $\Phi_{\psi\tilde{\varepsilon}}$ is the cross-spectrum between the signals ψ and $\tilde{\varepsilon}$.

Moreover, the quantity $\sqrt{N}(\hat{\theta}_N - \theta^*)$ is asymptotically normally distributed and its asymptotic covariance is given by (see [7])

$$P_\theta = (\bar{V}'')^{-1}\left(\lim_{N\to\infty} NE(V_N'(\theta^*)V_N'^T(\theta^*)\right)(\bar{V}'')^{-1}$$

with $\bar{V}'' = \bar{E}(\psi\psi^T)$. The central term on the right-hand side is given by (see also [4])

$$\begin{aligned}
&\lim_{N\to\infty} N^{-1}E\sum_{t=1}^{N}\sum_{s=1}^{N} \varepsilon_t(\theta^*)\varepsilon_s(\theta^*)\psi_t\psi_s^T \\
&= \lim_{N\to\infty} N^{-1}\sum_{t,s=1}^{N} E(e_t + \tilde{\varepsilon}_t)(e_s + \tilde{\varepsilon}_s)\psi_t\psi_s^T \\
&= \lambda_0\bar{E}(\psi\psi^T) + \lim_{N\to\infty} N^{-1}\sum_{t,s=1}^{N} E\left\{\tilde{\varepsilon}_t\tilde{\varepsilon}_s\psi_t\psi_s^T\right\}.
\end{aligned}$$

Here, t, s index time instants. Let us denote the second term in the last expression by Ξ. Then P_θ can be written as

$$P_\theta = \lambda_0(\bar{E}(\psi\psi^T))^{-1} + (\bar{E}(\psi\psi^T))^{-1}\Xi(\bar{E}(\psi\psi^T))^{-1}. \tag{3}$$

Thus the asymptotic covariance of the parameter estimate is the sum of two terms. While the first term in (3) is induced by the noise e, the second term is due to the undermodeling. A similar situation holds for parametric frequency domain identification. It is known (see [2], [10], and [8, Sec. 7.11.4]) that in this case the second contribution in fact is due to the variability of the input u. Hence, for a stochastic input the variance of the parameter estimate in general does not vanish even in the absence of noise. However, it vanishes if the input is deterministic. A similar result was established in [9] for time-domain identification. Let us formalize this fact.

The term Ξ can be written as

$$\begin{aligned}
\Xi &= \lim_{N\to\infty} N^{-1}E\left\{\left(\sum_{t=1}^{N}\tilde{\varepsilon}_t\psi_t\right)\left(\sum_{s=1}^{N}\tilde{\varepsilon}_s\psi_s^T\right)\right\} \\
&= \lim_{N\to\infty} N^{-1}E\left\{\left(\sum_{t=1}^{N}\tilde{\varepsilon}_t\psi_t\right)\left(\sum_{t=1}^{N}\tilde{\varepsilon}_t\psi_t\right)^T\right\}.
\end{aligned} \tag{4}$$

Proposition 1: If the input signal u is a multisine, then $\Xi = 0$. We hereby assume that the period of the multisine remains constant when the number of data tends to infinity, i.e., the number of periods tends to infinity. Observe that this proposition also covers the case where u is a square wave signal.

Proof: Suppose u is a multisine. Then the signals $\tilde{\varepsilon}$ and ψ are also multisines. Since products of trigonometric functions can be expressed as sums and differences of trigonometric functions, the signal $\tilde{\varepsilon} \cdot \psi$ is also a multisine and, hence periodic, with a spectrum consisting of sums and differences of the frequencies present in the spectrum of u. By (2), the sum of $\tilde{\varepsilon} \cdot \psi$ over a complete period is zero. But then the cumulative sum of $\tilde{\varepsilon} \cdot \psi$ is periodic, specifically bounded. The proposition now follows from (4). □

Proposition 1 states that in the case of a deterministic input the undermodeling has an impact only on the value of the asymptotic estimate θ^*, but not on the variance of $\hat{\theta}_N - \theta^*$. The latter is entirely due to the noise.

In Section III, we quantify the impact of undermodeling on the parameter variance if a stochastic input signal is used.

III. Parameter Variance in the Case of Stochastic Input

In this section, we examine the asymptotic covariance matrix (3) for zero mean quasistationary stochastic inputs. We establish a condition under which the asymptotic covariance depends only on the second order properties of the involved signals. Assuming this condition, we derive an explicit frequency domain expression for the asymptotic parameter variance as a function of the input power spectrum Φ_u.

If u is filtered white noise, then the signal $\tilde{\varepsilon} \cdot \psi$ has by (2) zero mean, but the standard deviation of its cumulative sum grows proportionally to the square root of the number of summands. Therefore, the term Ξ in (3) might be nonzero. By (4) the matrix Ξ is positive semidefinite and, as expected, undermodeling can only increase the asymptotic parameter variance.

It is known that in the presence of undermodeling the asymptotic covariance depends on the fourth-order properties of the input and the noise (see [8, p. 198]). Indeed, the definition of Ξ involves fourth-order products and powers of the input. Therefore, the asymptotic variance cannot be described as a function of the second order properties of u alone. In general, it will depend also on the fourth-order cumulant spectrum [11]. This poses serious difficulties, e.g., for input design. However, if we restrict the fourth-order cumulants of u to be zero, then the asymptotic variance is a function of the input power spectrum Φ_u only. Denote the autocorrelation function $\bar{E}(u_t u_{t-\tau})$ of u by $R_u(\tau)$. Then, the vanishing of the fourth-order cumulants of u can be equivalently rewritten as the condition

$$\bar{E}(u_{p+t}u_{q+t}u_{r+t}u_{s+t}) = R_u(p-r)R_u(q-s)+R_u(p-s)R_u(q-r) + R_u(p-q)R_u(r-s) \qquad \forall p,q,r,s. \quad (5)$$

Here, the time average is taken with respect to t and the numbers p, q, r, s are assumed to be fixed. Condition (5) is in fact not very restrictive. It is satisfied for instance for filtered zero mean white noise, where the probability density function of the white noise has zero *kurtosis* "peakedness," see, e.g., [11]. This is equivalent to the condition that the second and fourth moments m_2, m_4 of the probability density function satisfy the relation $m_4 = 3m_2^2$. This relation holds, e.g., for a Gaussian distribution.

We now proceed to compute an expression of Ξ in terms of signal spectra for the case where condition (5) holds. For this, we need some technical lemmas.

Lemma 1: Let A, B, C, D be stable transfer functions and $A_k, B_k, C_k, D_k, k = 0, 1, \ldots$ the coefficients of their Laurent expansions around $z = 0$. Let u be a quasi-stationary stochastic zero mean process with power spectrum Φ_u and autocorrelation coefficients $R_u(\tau)$ satisfying condition (5). Let $a = Au, b = Bu, c = Cu, d = Du$ be signals obtained by filtering u through A, B, C, D and let $\alpha, \beta, \gamma, \delta$ be delays. Then

$$\bar{E}(a_{t-\alpha}b_{t-\beta}c_{t-\gamma}d_{t-\delta}) = R_{ab}(\beta-\alpha)R_{cd}(\delta-\gamma) + R_{ac}(\gamma-\alpha)R_{bd}(\delta-\beta) + R_{ad}(\delta-\alpha)R_{bc}(\gamma-\beta)$$

where the time average is taken with respect to t and $R_{gh}(\tau)$ denotes the cross-correlation of signals g, h at lag τ.

The result is easily verified by direct calculation. □

We now compute the asymptotic covariance of the parameter estimate for quasistationary zero mean stochastic input satisfying condition (5). By Lemma 1 and (2), we obtain from (4)

$$\begin{aligned}\Xi &= \sum_{\tau=-\infty}^{+\infty} \bar{E}\left\{\tilde{\varepsilon}_{t+\tau}\tilde{\varepsilon}_t \psi_{t+\tau}\psi_t^T\right\} \\ &= \sum_{\tau=-\infty}^{+\infty} \{R_{\tilde{\varepsilon}\tilde{\varepsilon}}(\tau)R_{\psi\psi^T}(\tau) + R_{\tilde{\varepsilon}\psi}(0)R_{\tilde{\varepsilon}\psi^T}(0) \\ &\quad + R_{\tilde{\varepsilon}\psi}(-\tau)R_{\tilde{\varepsilon}\psi^T}(\tau)\} \\ &= \sum_{\tau=-\infty}^{+\infty} \left\{R_{\tilde{\varepsilon}\tilde{\varepsilon}}(\tau)R_{\psi\psi^T}(\tau) + R_{\psi\tilde{\varepsilon}}(\tau)R_{\tilde{\varepsilon}\psi^T}(\tau)\right\}. \end{aligned} \quad (6)$$

Lemma 2: Let A, B, C, D be stable transfer functions. Let u be a quasi-stationary process with power spectrum Φ_u and let $a = Au, b = Bu, c = Cu, d = Du$ be signals obtained by filtering u through A, B, C, D. Then, the following relation holds:

$$\begin{aligned}&\sum_{\tau=-\infty}^{+\infty} R_{ab}(\tau)R_{cd}(\tau) \\ &\quad = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{ab}\bar{\Phi}_{cd}\, d\omega \\ &\quad = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{ac}\bar{\Phi}_{bd}\, d\omega = \sum_{\tau=-\infty}^{+\infty} R_{ac}(\tau)R_{bd}(\tau).\end{aligned}$$

This result is a consequence of Parseval's theorem. □

Applying Lemma 2 componentwise to the products $R_{\tilde{\varepsilon}\tilde{\varepsilon}}(\tau)R_{\psi\psi^T}(\tau)$ in (6) and taking into consideration (2) leads to

$$\begin{aligned}\Xi &= \sum_{\tau=-\infty}^{+\infty} (R_{\tilde{\varepsilon}\psi}(\tau) + R_{\psi\tilde{\varepsilon}}(\tau))R_{\tilde{\varepsilon}\psi^T}(\tau) \\ &= \sum_{\tau=1}^{\infty}(R_{\psi\tilde{\varepsilon}}(\tau) + R_{\tilde{\varepsilon}\psi}(\tau))(R_{\psi\tilde{\varepsilon}}(\tau) + R_{\tilde{\varepsilon}\psi}(\tau))^T.\end{aligned} \quad (7)$$

Thus, we have represented Ξ as a sum of squares for the case where the input satisfies condition (5). By Lemma 2, representation (7) yields the following frequency domain expression for Ξ:

$$\Xi = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{Re}\Phi_{\psi\tilde{\varepsilon}} \cdot \mathrm{Re}\Phi_{\psi\tilde{\varepsilon}}^T\, d\omega. \quad (8)$$

We obtain the following result.

Proposition 2: Assume an experimental setting as described in Section II. Let the input u be a quasistationary stochastic process satisfying condition (5). Then undermodeling does not have an impact on the asymptotic parameter covariance if and only if the cross-spectrum between the signals $\psi = H^{-1}G'u$ and $\tilde{\varepsilon} = H^{-1}(G_0 - G(\theta^*))u$ is a purely imaginary function of the frequency. If the cross-spectrum has a nonzero real part, then the undermodeling increases the asymptotic parameter covariance. □

Thus, the condition $\Xi = 0$ admits the following interpretation. Suppose for some frequency ω we have $\Phi_u(\omega) \neq 0$. Then $\mathrm{Re}\Phi_{\psi\tilde{\varepsilon}}(\omega) = 0$

implies $\arg(G_0(e^{j\omega})-G(\theta^*)(e^{j\omega}))-\arg(\partial G/\partial\theta_k)(e^{j\omega}) = \pm(\pi/2)$ for all $k = 1,\ldots,n$. This means that the bias $G_0 - G(\theta^*)$ is orthogonal to the model structure not only in the sense of the pseudoscalar product (1), but frequency-wise in the Nyquist plane at the excited frequency ranges.

Inserting (8) into (3) yields the following frequency domain expression for the asymptotic parameter covariance:

$$P_\theta = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_{\psi\psi^T}\,d\omega\right)^{-1}\left(\frac{\lambda_0}{2\pi}\int_{-\pi}^{\pi}\Phi_{\psi\psi^T}\,d\omega + \frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{Re}\Phi_{\psi\tilde{\varepsilon}}\cdot Re\Phi_{\psi\tilde{\varepsilon}}^T\,d\omega\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_{\psi\psi^T}\,d\omega\right)^{-1}. \quad (9)$$

It is valid if (5) holds.

In this section, we considered the case of a stochastic input and derived expressions for the asymptotic parameter covariance as a function of the input power spectrum under assumption (5). As (7) shows, the contribution of the undermodeling to the parameter covariance is caused by the correlation between the prediction errors $\tilde{\varepsilon}_t + e_t$ and its negative gradients ψ_t at time lags other than zero.

IV. Variance in the Framework of Stochastic Embedding

In this section, we compute the asymptotic parameter variance in the framework of stochastic embedding. In this framework, the parameter error is entirely described as a variance error. It is then interesting to compute that part of this variance error that is due to the undermodeling.

For simplicity, we assume a linear model structure. Let $G_0 = \theta_0^T\Lambda + \eta^T Z$, where θ_0 is a fixed parameter vector, $\eta \in \mathbf{R}^L, \Lambda$ is a column vector of n stable transfer functions, and $Z = (Z_1,\ldots,Z_L)^T$ is a vector of L stable transfer functions. Within the stochastic embedding framework, η is assumed to be a random variable with zero mean and covariance matrix C_η. For any given identification experiment the vector η assumes a fixed value, which is drawn according to its probability distribution. For details and a justification of the procedure see [3]. Hence, under the assumptions of stochastic embedding the vector θ_0 reflects intrinsic properties of the system and can be considered as the "true" parameter vector.

Thus, the asymptotic value θ^* of the parameter estimate as well as its asymptotic covariance P_θ become a function of $\eta, \theta^* = \theta^*(\eta), P_\theta = P_\theta(\eta)$. By (2), we have

$$\theta^* - \theta_0 = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{|H|^2}\Lambda\Lambda^*\,d\omega\right)^{-1}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{|H|^2}\Lambda Z^*\,d\omega\right)\eta$$

$$\tilde{\varepsilon} = (\theta_0 - \theta^*)^T\psi + \eta^T H^{-1}Zu. \quad (10)$$

Note that both $\tilde{\varepsilon}$ and $\theta^* - \theta_0$ are linear in η. Observe that by $E\,\eta = 0$ we have $E\theta^* = \theta_0$, where the expectation is taken over η. Similarly $E\tilde{\varepsilon} = 0$. Further, Ξ becomes a matrix-valued positive semidefinite quadratic form in η. Let us write this as $\tilde{\varepsilon} = \sum_i \eta_i\tilde{\varepsilon}^i$, $\theta^* - \theta_0 = \sum_i \eta_i\Delta\theta^i$, and $\Xi = \sum_{i,j}\eta_i\eta_j\Xi^{ij}$. Here

$$\tilde{\varepsilon}^i = \left(\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{|H|^2}Z_i\Lambda^*\,d\omega\right) \times \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{|H|^2}\Lambda\Lambda^*\,d\omega\right)^{-1} H^{-1}\Lambda + H^{-1}Z_i\right)u$$

$$\Delta\theta^i = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{|H|^2}\Lambda\Lambda^*\,d\omega\right)^{-1}\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{|H|^2}\Lambda Z_i^*\,d\omega$$

and Ξ^{ij} are matrices, which under (5), are given by

$$\Xi^{ij} = \sum_{\tau=1}^{\infty}(R_{\psi\tilde{\varepsilon}^i}(\tau) + R_{\tilde{\varepsilon}^i\psi}(\tau))(R_{\psi\tilde{\varepsilon}^j}(\tau) + R_{\tilde{\varepsilon}^j\psi}(\tau))^T = \frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{Re}\Phi_{\psi\tilde{\varepsilon}^i}\cdot\mathrm{Re}\Phi_{\psi\tilde{\varepsilon}^j}^T\,d\omega.$$

Let us now compute the variance of the parameter estimate $\hat{\theta}_N$ after averaging over η. Since η has zero mean, the expectation of $\hat{\theta}_N$ is equal to θ_0. Assuming a normal distribution for η, the asymptotic covariance $\mathbf{P}_N$ of $\hat{\theta}_N$ is given by

$$\begin{aligned}
&E(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T \\
&= \int_{\mathbf{R}^{n+L}}\frac{1}{\sqrt{(2\pi)^{n+L}|C_\eta||N^{-1}P_\theta(\eta)|}} \\
&\quad\times\exp\left\{-\frac{(\theta-\theta^*)^T NP_\theta^{-1}(\eta)(\theta-\theta^*) + \eta^T C_\eta\eta}{2}\right\} \\
&\quad\cdot(\theta-\theta_0)(\theta-\theta_0)^T\,d\theta d\eta \\
&= \int_{\mathbf{R}^L}\frac{1}{\sqrt{(2\pi)^L|C_\eta|}}e^{-\frac{\eta^T C_\eta\eta}{2}}(N^{-1}P_\theta(\eta) \\
&\quad + (\theta^*-\theta_0)(\theta^*-\theta_0)^T)\,d\eta \\
&= \lambda_0(N\bar{E}(\psi\psi^T))^{-1} + \sum_{i,j}(C_\eta)_{ij}[N^{-1}(\bar{E}(\psi\psi^T))^{-1}\Xi^{ij} \\
&\quad\times(\bar{E}(\psi\psi^T))^{-1} + \Delta\theta^i(\Delta\theta^j)^T]. \qquad (11)
\end{aligned}$$

Here $(C_\eta)_{ij}$ are the entries of the covariance matrix C_η. Besides the familiar variance term $\lambda_0(N\bar{E}(\psi\psi^T))^{-1}$ caused by the noise we have two different contributions from the undermodeling to the total variance. The term $\sum_{i,j}(C_\eta)_{ij}\Delta\theta^i(\Delta\theta^j)^T$ is due to the bias, i.e., the shift away from θ_0 of the asymptotic value θ^* of the estimate. It is included into the variance only by the stochastic embedding procedure.

The term $\sum_{i,j}(C_\eta)_{ij}N^{-1}(\bar{E}(\psi\psi^T))^{-1}\Xi^{ij}(\bar{E}(\psi\psi^T))^{-1}$ is due to the increase of the asymptotic parameter variance by the undermodeling for any fixed η.

Note that the covariance matrix $\mathbf{P}_N$ does not completely describe the distribution of $\hat{\theta}_N - \theta_0$, even if η is normally distributed. The distribution of $\hat{\theta}_N$ will not be Gaussian if Ξ is not identically zero, because its asymptotic covariance $P_\theta(\eta)$ is a function of the random vector η, as stated before. The same holds for the distribution of the transfer function in the frequency domain. The definition of uncertainty regions at certain probability levels will therefore face considerable difficulties. We stress that this property does in no way contradict the familiar theorems on asymptotic normality of the parameter estimate. The nonnormality of $\hat{\theta}_N - \theta_0$ is an artifact introduced by randomizing the undermodeling, i.e., by averaging with respect to the probability distribution of the actually constant undermodeling parameter vector η. However, if Ξ is zero, i.e., when using multisines as input, then averaging over η yields a Gaussian probability distribution of $\hat{\theta}_N$, given η is normally distributed.

V. Nonmonotonicity of the Information Matrix

In this section, we investigate the monotonicity properties of the total parameter variance with respect to the input power spectrum under the assumptions of stochastic embedding. We will work with the information matrix M, the inverse of the covariance matrix, $M = [E(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T]^{-1}$. This information matrix depends on the input power spectrum Φ_u, and one would expect that the usual monotonicity property holds, namely $M(\Phi_u + \Phi'_u) \succeq M(\Phi_u)$ for any power spectra Φ_u, Φ'_u. This is in general not true, as demonstrated by the following example.

Example: Consider prediction error identification in a stochastic embedding framework for the following system:

$$y = \theta_0 z^{-1} u + \eta z^{-2} u + e$$

with model structure $G(\theta) = \theta\Lambda = \theta z^{-1}$. Thus, $n = 1, L = 1, H \equiv 1, \Lambda = z^{-1}, Z = z^{-2}$. Let $x_k(\Phi_u), k = 0, 1$, be the first trigonometric moments of the power spectrum Φ_u, i.e.,

$$x_k(\Phi_u) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u(\omega) e^{-jk\omega}\, d\omega, \qquad k = 0, 1.$$

Suppose further that the inputs are multisines, so that $\Xi = 0$. For this example we have $\psi = z^{-1}u, \bar{E}(\psi\psi^T) = x_0(\Phi_u), \theta^* - \theta_0 = \Delta\theta = x_0^{-1}(\Phi_u)\, x_1(\Phi_u)\, \eta$ using (10). Consequently, by (11) we have $E(\hat{\theta}_N - \theta_0)^2 = (\lambda_0/N x_0) + C_\eta(x_1^2)/(x_0^2)$, and hence

$$M(\Phi_u) = \left[\frac{\lambda_0}{N x_0} + C_\eta \frac{x_1^2}{x_0^2}\right]^{-1} = \frac{x_0^2(\Phi_u)}{\lambda_0 N^{-1} x_0(\Phi_u) + C_\eta x_1^2(\Phi_u)}.$$

Let us now consider the three multisine sequences $u_t = (\sqrt{3}/3)\cos(\pi t) + (2\sqrt{3}/3)\sin((\pi/3)t)$, $u'_t = (\sqrt{2}/2)\cos(\pi t) + \sin((\pi/2)t), u''_t = (\sqrt{30}/6)\cos(\pi t) + \sin((\pi/2)t) + (2\sqrt{3}/3)\sin((\pi/3(t)$. Their respective power spectra Φ_u, Φ'_u and Φ''_u are related by the equality $\Phi''_u = \Phi_u + \Phi'_u$ and have moments $x_0(\Phi_u) = x_0(\Phi'_u) = 1, x_1(\Phi'_u) = x_1(\Phi''_u) = -(1/2), x_1(\Phi_u) = 0, x_0(\Phi''_u) = 2$. Hence, the information matrices of experiments performed with inputs u and u'' are given by

$$M(\Phi_u) = \frac{N}{\lambda_0} \quad M(\Phi_u + \Phi'_u) = \frac{4}{2\lambda_0 N^{-1} + \frac{1}{4}C_\eta}.$$

Therefore, we have $M(\Phi_u) \preceq M(\Phi_u + \Phi'_u)$ if and only if $C_\eta N \leq 8\lambda_0$. Thus, if the SNR is large enough (i.e., λ_0 is small), if the undermodeling effects begin to dominate the noise effects, or if the number of data becomes large, then the input u'' yields a smaller information matrix than the input u, although its power spectrum Φ''_u is larger or equal to Φ_u frequency-wise. □

This leads to the counterintuitive conclusion that an increase in input power at some frequencies without decrease at the other frequencies does not necessarily imply an increase of information, and may even lead to a decrease. This is an artifact caused by the stochastic embedding procedure, which randomizes the undermodeling by considering it as being of stochastic nature and lumping it together with the actual parameter variance. The increase of variance can therefore be attributed to an increase of the bias.

A weaker monotonicity property does hold, however. The following assertion is a consequence of (11).

Corollary 1: Let Φ_u be a power spectrum and let $\beta > 1$ be a constant. Let $M(\Phi_u)$ denote the information matrix when the input signal has power spectrum Φ_u and let $M(\beta\Phi_u)$ be the corresponding information matrix for inputs with power spectrum $\beta\Phi_u$. Assume further that the inputs satisfy condition (5). Then $M(\beta\Phi_u) \succeq M(\Phi_u)$. □

VI. Conclusion

In this note, we have investigated the asymptotic parameter variance under time-domain prediction error identification in open loop in the presence of undermodeling. It was shown that under identification with a periodic input the undermodeling does not influence the parameter variance. This result is summarized in Proposition 1.

For stochastic input, undermodeling leads to an increase of the variance. In general the amount of this increase is difficult to evaluate and depends on higher order properties of the input signal. Under assumption (5) on the input signal, however, it is proportional to the integral of the squared real part of the cross-spectrum between the prediction error and its gradient. This result is formalized in formula (9). Thus, the undermodeling impacts the parameter variance through the correlation between the prediction error and its gradient.

Further, we investigated the asymptotic parameter covariance within the stochastic embedding framework for linear model structures. An explicit expression is given by formula (11), but note that in general the distribution of the parameter vector is not normal. This can be attributed to the way the bias error is randomized under the assumptions of stochastic embedding, and does not contradict the standard asymptotic normality theorems of prediction error identification. Normality of the distribution is preserved under the conditions of Proposition 1. For the information matrix as a function of the input power spectrum a weak monotonicity property holds, which is formalized in Corollary 1. A counterexample to the usual monotonicity condition has also been given.

References

[1] D. K. de Vries and P. M. J. van den Hof, "Frequency domain identification with generalized orthonormal basis functions," *IEEE Trans. Automat. Contr.*, vol. 43, pp. 656–669, May 1998.

[2] G. C. Goodwin, M. Gevers, and D. Q. Mayne, "Bias and variance distribution in transfer function estimation," presented at the 9th IFAC Symp. Identification, Budapest, Hungary, July 1991.

[3] G. C. Goodwin, M. Gevers, and B. Ninness, "Quantifying the error in estimated transfer functions with application to model order selection," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 913–928, July 1992.

[4] H. Hjalmarsson, "A model variance estimator," in *Preprint 12th IFAC World Congr.*, vol. 9, Sydney, NSW, Australia, 1993, pp. 5–10.

[5] H. Hjalmarsson and L. Ljung, "Estimating model variance in the case of undermodeling," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 1004–1008, July 1992.

[6] H. Hjalmarsson and B. Ninness, "Identification in dosed loop: Asymptotic high order variance for restricted complexity models," in *Proc. 37th IEEE Conf. Decision Control*, Tampa, FL, 1998, pp. 3396–3401.

[7] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[8] R. Pintelon and J. Schoukens, *System Identification. A Frequency Domain Approach*. Piscataway, NJ: IEEE Press, 2001.

[9] ——, "Some peculiarities of identification in the presence of model errors," *Automatica*, vol. 38, pp. 1683–1693, Oct. 2002.

[10] R. Pintelon, J. Schoukens, and Y. Rolain, "Uncertainty of transfer function modeling using prior estimated noise models," Dept. ELEC, Vrije Universiteit Brussels, Brussels, Belgium, Tech. Rep. RP042 002, 2002.

[11] M. Rosenblatt, *Stationary Sequences and Random Fields*. Boston, MA: Birkhäuser, 1985.

[12] J. Schoukens and R. Pintelon, "Quantifying model errors of identified transfer functions," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 1733–1737, Aug. 1994.

[13] F. Tjarnstrom and L. Ljung, "Using the bootstrap to estimate the variance in the case of undermodeling," *IEEE Trans. Automat. Contr.*, vol. 47, pp. 395–398, Feb. 2002.