Estimation Variance is not Model Structure Independent

Brett Ninness *
Dept. of Elec. & Comp. Eng,
Uni. Newcastle, Australia.

Håkan Hjalmarsson †
Dept. S3,
Royal Inst. Tech,
Stockholm, Sweden.

Fredrik Gustafsson ‡
Dept. Elec. Eng,
Linköping Uni,
Linköping,Sweden

Abstract

This paper establishes that when using a least squares criterion to estimate an output error type model structure, then the measurement noise induced variability of the frequency response estimate depends on the estimated (and hence also on the true) pole positions. This dependence on pole position is perhaps counter to prevailing wisdom that for any 'shift invariant' model structure, the variability depends only on model order, data length, and input and noise spectral densities. That is, it is counter to the belief that variance error is model-structure independent.

1 Introduction

When considering the performance of prediction error identification methods using quadratic criteria, a seminal result is that the measurement noise induced variability of the ensuing frequency response estimate may be approximated as (L.Ljung, 1985; L.Ljung and Z.D.Yuan, 1985; Ljung, 1987)

$$\text{Var}\{G_{\hat{\theta}_N}(e^{j\omega})\} \approx \frac{m}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}. \tag{1}$$

Here Φ_ν and Φ_u are respectively the measurement noise and input spectral densities, and $\widehat{\theta}_N$ is the prediction error estimate based on N observed data points of a vector θ parameterising a model structure $G_\theta(q)$ for which (essentially) $m = \dim\theta/2d$ where d is the number of denominator polynomials to be estimated in the model structure.

Apart from its simplicity, a key reason underlying the importance and popularity of the approximation (1) is that, according to its derivation (L.Ljung, 1985; L.Ljung and Z.D.Yuan, 1985; Ljung, 1987), it applies for a very wide class of so-called 'shift invariant' model structures. For example, all the well known FIR, ARX, ARMAX, output error and Box–Jenkins structures are shift invariant (L.Ljung, 1985).

In contrast, in a series of recent works (Wahlberg, 1991, 1994; P.M.J. Van den Hof *et al.*, 1995; Ninness *et al.*, 1997) it has been established that for model structures which can be considered as shift invariant generalisations of FIR (in that the fixed poles are not necessarily at the origin), then in the interests of improved approximation (1) should be modified to become

$$\text{Var}\{G_{\hat{\theta}_N}(e^{j\omega})\} \approx \frac{1}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \tag{2}$$

* Email: *brett@ee.newcastle.edu.au.*
† Email: *hakan.hjalmarsson@s3.e.kth.se*
‡ Email: *fredrik@isy.liu.se*

where the ξ_k are the fixed poles in the model structure, so that (2) reverts to (1) for the FIR case of $\xi_k = 0$.

In (Ninness *et al.*, 1997) it has also been shown that for ARX model structures with fixed noise model zeros, again not necessarily at points $\xi_0, \cdots, \xi_{n-1}$ which are at the origin, then again (2) rather then (1) should be used in the interests of providing the most accurate approximation of $\mathrm{Var}\{G_{\hat{\theta}_N}(e^{j\omega})\}$.

For these afore-mentioned cases, for the purposes of actually calculating $\widehat{\theta}_N$, the process of incorporating the fixed poles or zeros may be achieved by first pre-filtering the data with an all-pole filter, and then fitting a conventional FIR or ARX structure. As such, the more general principle of the original approximation (1) being invariant to the particular model structure (but not to the nature of the input and noise spectral densities) is preserved.

The purpose of this paper is to establish that, in fact, an accurate approximation for the variance $\mathrm{Var}\{G_{\hat{\theta}_N}(e^{j\omega})\}$ *is not* model structure invariant. To establish this, the strategy employed here is to use the output error model structure to illustrate the point. The lack of invariance to model structure stems from the approximation, which outperforms (1), being again of the form (2), but this time with points $\{\xi_0, \cdots, \xi_{n-1}\}$ being the estimated poles of $G_{\hat{\theta}_N}(q)$.

2 Motivational Example

In the interests of motivating the analysis to follow, it is preluded by an illustrative simulation example in which the following continuous time system

$$G(s) = \frac{1}{(s+0.9163)^2(s+0.3567)^2(s+0.2231)^3}$$

is considered, and for which input-output samples are obtained at 1 second intervals with zero-order-held inputs. This implies a discrete time representation with poles at $z = 0.8, 0.7, 0.4$ which is estimated using a 7'th order output error model structure and on the basis of observing a length $N = 5000$ sample input-output record for which the output is corrupted by white Gaussian noise of variance $\sigma^2 = 0.001$, and with input which is a realisation of a stationary Gaussian process with spectral density

$$\Phi_u(\omega) = \frac{1}{1.25 - \cos\omega}.$$

The sample mean square error over 500 estimation experiments with different input and noise realisations is used as an estimate of $\mathrm{Var}\{G_{\hat{\theta}_N}(e^{j\omega})\}$ and plotted as a solid line in figure 1. The 'classical' approximation (1) is shown as a dash-dot line in that same figure, and is clearly a poor approximation to the estimated variability. By way of contrast, the modified approximation (2) (with the $\{\xi_k\}$ being the true poles in $G(q)$) shown as the dashed line in figure 1 appears to be quite an accurate approximation to the estimated variability.

This provides clear evidence that the true variability is in fact not model structure invariant, and hence accurate approximation of it may need to take that phenomenon into account. The remainder of the paper is devoted to supplying theoretical analysis of this issue. A key tool in this is the employment of a class of rational orthonormal basis functions that are adaptable to the system being estimated, and which have been discussed in some detail in (Ninness and Gustafsson, 1997; Ninness *et al.*, 1998).

3 Problem Setting

This paper addresses model structures which describe the relationship between an input data record $\{u_t\}$ and an output data record $\{y_t\}$ according to

$$y_t = G_\theta(q)u_t + H_\theta(q)e_t$$

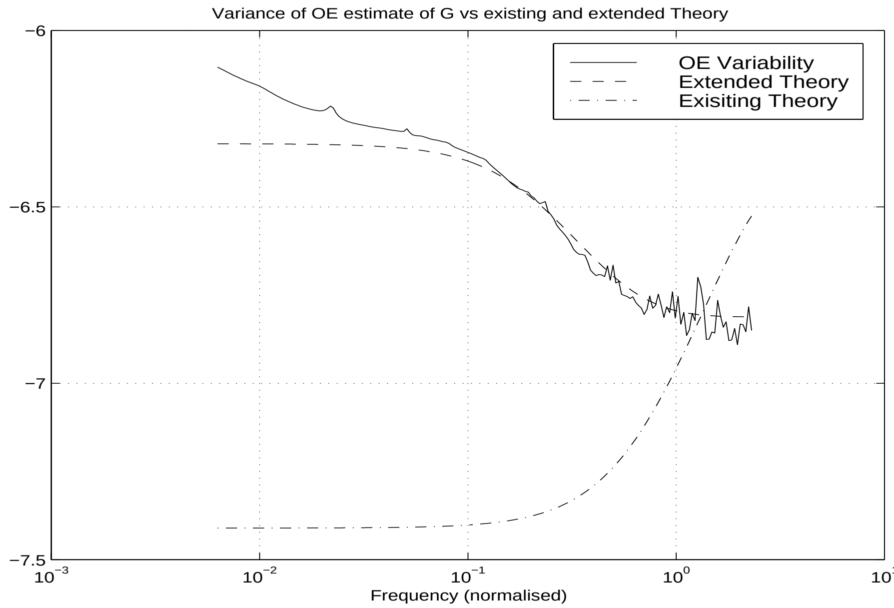

Figure 1: *Variability of Output Error Estimate - True variability vs. new and existing theoretically derived approximations.*

where $\{e_t\}$ is a zero-mean white noise sequence such that $\mathbf{E}\{e_t^2\} = \sigma^2, \mathbf{E}\{|e_t|^{4+\epsilon}\} < \infty$ for some $\epsilon > 0$ and $G_\theta(q), H_\theta(q)$ are transfer functions, rational in the forward shift operator q, and parameterised by a vector $\theta \in \mathbf{R}^n$. The mean-square optimal one-step ahead prediction $\widehat{y}_t(\theta)$ based on this model structure is (Ljung, 1987)

$$\widehat{y}_t(\theta) = [1 - H_\theta^{-1}(q)]y_t + H_\theta^{-1}(q)G_\theta(q)u_t$$

with associated prediction error

$$\varepsilon_t(\theta) = y_t - \widehat{y}_t = H^{-1}(q)\left[y_t - G_\theta(q)u_t\right]$$

involved with the quadratic estimation criterion

$$V_N(\theta) = \frac{1}{2N}\sum_{t=1}^{N}\varepsilon_t^2(\theta)$$

used to define the prediction error estimate $\widehat{\theta}_N$ of θ as

$$\widehat{\theta}_N \triangleq \arg\min_{\theta\in\mathbf{R}^n} V_N(\theta).$$

As has been established in (L.Ljung, 1978; Ljung, 1987), under the assumption that the input $\{u_t\}$ is quasi-stationary (Ljung, 1987), then $\widehat{\theta}_N$ converges with increasing N and with probability one according to

$$\lim_{N\to\infty}\widehat{\theta}_N = \theta_\circ \triangleq \arg\min_{\theta\in\mathbf{R}^n}\lim_{N\to\infty}\mathbf{E}\left\{V_N(\theta)\right\},$$

As well, it also holds that as N increases the estimate $\widehat{\theta}_N$ tends to be Normally distributed about $\theta_\circ$ according to (L.Ljung and P.E.Caines, 1979; Caines, 1988; Ljung, 1987)

$$\sqrt{N}(\widehat{\theta}_N - \theta_\circ) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_n), \tag{3}$$

$$P_n \triangleq R_n^{-1} Q_n R_n^{-1}$$

where with the definition of prediction error gradient $\psi_t(\theta)$ as ($\cdot'$ denotes differentiation with respect to θ)

$$\psi_t(\theta) \triangleq \widehat{y}_t'(\theta) = H_\theta^{-1}(q)\left[G_\theta'(q)u_t + H_\theta'(q)\varepsilon_t(\theta)\right] \tag{4}$$

so that more succinctly

$$\psi_t(\theta) = H_\theta^{-1}(q)\Pi_\theta'(q)\zeta_t(\theta)$$

where

$$\Pi_\theta(q) \triangleq [G_\theta(q), H_\theta(q)], \qquad \Pi_\theta'(q) \triangleq \left[G_\theta'(q), H_\theta'(q)\right],$$

$$\zeta_t(\theta) \triangleq \begin{bmatrix} u_t \\ \varepsilon_t(\theta) \end{bmatrix}.$$

Therefore,

$$R_n \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\left[\mathbf{E}\left\{\psi_t(\theta_\circ)\psi_t^T(\theta_\circ)\right\} - \mathbf{E}\left\{\varepsilon_t(\theta_\circ)\left(\psi_t'(\theta)\right)^T\right\}\right]$$

and

$$Q_n \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t,\ell=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ)\psi_\ell^T(\theta_\circ)\varepsilon_t(\theta_\circ)\varepsilon_\ell(\theta_\circ)\right\}.$$

While an asymptotic distributional result like (3) is very satisfying theoretically, for practical applications it is rather less appealing, mainly due to the (just presented) intricate definition of P_n via Q_n, R_n and $\psi_t(\theta)$.

This was recognised in (L.Ljung, 1985; L.Ljung and Z.D.Yuan, 1985; Ljung, 1987) which proposed the solution of investigating how (3) manifested itself in the variability of the frequency responses $G_{\hat{\theta}_N}(e^{j\omega}), H_{\hat{\theta}_N}(e^{j\omega})$, the result being approximations such as (1).

4 Synopsis of Output Error Case

To see how this strategy of evaluating the implications of (3) in the frequency domain, and in such a way as to expose the genesis of the inaccuracy illustrated in figure 1 of the approximation (1), this section focuses on the case of the model structure being of the output error form

$$y_t = G_\theta(q)u_t + e_t = \frac{B_\theta(q)}{A_\theta(q)}u_t + e_t$$

where the numerator and denominator polynomials are of the form

$$A_\theta(q) = q^m + a_{m-1}q^{m-1} + \cdots + a_1 q + a_0,$$

$$B_\theta(q) = b_{m-1}q^{m-1} + \cdots + b_1 q + b_0,$$

and $\theta \in \mathbf{R}^n$ (with $n = 2m$) is defined as

$$\theta^T = [a_0, b_0, a_1, b_1, \cdots, a_{m-1}, b_{m-1}].$$

In this case

$$\begin{aligned} \frac{\mathrm{d}G_\theta(q)}{\mathrm{d}\theta} &= \left[-\frac{B_\theta(q)}{A_\theta^2(q)}, \frac{1}{A_\theta(q)}, -\frac{qB_\theta(q)}{A_\theta^2(q)}, \frac{q}{A_\theta(q)}, \cdots, -\frac{q^{m-1}B_\theta(q)}{A_\theta^2(q)}, \frac{q^{m-1}}{A_\theta(q)}\right]^T \\ &= [\Lambda_m(q) \otimes I_2] Z_\theta(q)\frac{1}{A_\theta(q)} \end{aligned}$$

where $\otimes$ is the Kronecker tensor product of matrices, I_2 is a 2×2 identity matrix, and

$$\Lambda_m(q) \triangleq [1, q, \cdots, q^{m-1}]^T, \;\; Z_\theta(q) \triangleq \begin{bmatrix} -G_\theta(q) \\ 1 \end{bmatrix} \tag{5}$$

Therefore, using (4) and noting that there is no parameterised noise model $H_\theta(q)$, leads to an expression for the gradient of the prediction error as a filtered version of the input $\{u_t\}$

$$\psi_t(\theta) = [\Lambda_m(q) \otimes I_2] Z_\theta(q) \frac{1}{A_\theta(q)} u_t.$$

Now, assuming that the true system is in the model structure so that a true parameter vector $\theta_\circ$ exists, then $\varepsilon_t(\theta_\circ) = e_t$ so that use of Parseval's Formula leads to (the $e^{j\omega}$ dependence will not be made explicit in what follows in order to improve readability, and $\cdot^\star$ will be used to denote'conjugate transpose')

$$\begin{aligned} R_n &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbf{E}\left\{\psi_t(\theta_\circ)\psi_t^T(\theta_\circ)\right\} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [\Lambda_m(e^{j\omega}) \otimes I_2] \frac{Z_\circ \Phi_u Z_\circ^\star}{|A_\theta(e^{j\omega})|^2} [\Lambda_m^\star \otimes I_2] \, \mathrm{d}\omega. \end{aligned}$$

In this case, with a $2m \times 2m$ block-Toeplitz matrix being defined by a 2×2 positive definite matrix valued function $F(\omega)$ as

$$T_n(F) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} [\Lambda_m \otimes I_2] F(\omega) [\Lambda_m^\star \otimes I_2] \, \mathrm{d}\omega \tag{6}$$

then

$$R_n = T_n \left(\frac{Z_\circ(e^{j\omega})\Phi_u(\omega) Z_\circ^\star(e^{j\omega})}{|A_\theta(e^{j\omega})|^2} \right).$$

Also, again under the assumption of $\varepsilon_t(\theta_\circ) = e_t$

$$Q_n \triangleq \lim_{N\to\infty} \frac{\sigma^2}{N} \sum_{t=1}^{N} \mathbf{E}\left\{\psi_t(\theta_\circ)\psi_t^T(\theta_\circ)\right\} = \sigma^2 R_n$$

so that the matrix P_n quantifying the parameter space variability of $\widehat{\theta}_N$ via (3) is in fact expressible as a block Toeplitz matrix associated with a particular spectral density as follows

$$P_n = \sigma^2 R_n^{-1} = \sigma^2 T_n^{-1} \left(\frac{Z_\circ \Phi_u(\omega) Z_\circ^\star}{|A_\theta|^2} \right). \tag{7}$$

This formulation of P_n is a key ingredient underlying the methods of (L.Ljung, 1985; L.Ljung and Z.D.Yuan, 1985; Ljung, 1987) that arrive at the approximation (1). A second fundamental idea is to relate this parameter space variability to frequency domain variability of $G_{\hat{\theta}_N}(e^{j\omega})$ via Taylor expansion according to

$$\begin{aligned} \widetilde{G}_N(\omega) &\triangleq G_{\hat{\theta}_N}(e^{j\omega}) - G_\circ(e^{j\omega}) \\ &= \left[\left.\frac{\mathrm{d}G_\theta}{\mathrm{d}\theta}\right|_{\theta=\theta_\circ} \right]^T (\widehat{\theta}_N - \theta_\circ) + o(\|\widehat{\theta}_N - \theta_\circ\|) \\ &= A_{\theta_\circ}^{-1} Z_\circ^T [\Lambda_m^T \otimes I_2](\widehat{\theta}_N - \theta_\circ) + o(\|\widehat{\theta}_N - \theta_\circ\|) \end{aligned} \tag{8}$$

so that using (3) and the previous Toeplitz matrix formulation of P_n

$$\sqrt{N}\widetilde{G}_N(\omega) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Delta_n(\omega)) \tag{9}$$

where (with obvious compactification of notation involving $\theta_\circ$) using (7)

$$\begin{aligned}\Delta_n &= \frac{1}{|A_\circ|^2} Z_\circ^\star[\Lambda_m^\star \otimes I_2] P_n [\Lambda_m \otimes I_2] Z_\circ \\ &= \frac{\sigma^2}{|A_\circ|^2} Z_\circ^\star[\Lambda_m^\star \otimes I_2] T_n^{-1}\left(\frac{Z_\circ \Phi_u Z_\circ^\star}{|A_\circ|^2}\right)[\Lambda_m \otimes I_2] Z_\circ.\end{aligned}$$

A potential difficulty in continuing the analysis of this expression is that although the Toeplitz matrix involved is non-singular, the matrix valued function $Z_\circ \Phi_u Z_\circ^\star$ is itself, by definition, singular when evaluated at any one frequency. To circumvent this, define a perturbed matrix

$$\Delta_n(\omega, \delta) \triangleq \frac{\sigma^2}{|A_\circ|^2} Z_\circ^\star[\Lambda_m^\star \otimes I_2] T_n^{-1}\left(\frac{Z_\circ \Phi_u(\omega) Z_\circ^\star}{|A_\circ|^2} + \delta I_2\right)[\Lambda_m \otimes I_2] Z_\circ$$

so that since matrix inversion is continuous

$$\Delta_n(\omega) = \lim_{\delta \to 0} \Delta_n(\omega, \delta).$$

The final principle underlying the analysis of (L.Ljung, 1985) is that that the quadratic form defining $\Delta_n(\omega)$, by virtue of being formulated in terms of inverses of Toeplitz matrices parameterised by spectral densities, it can be viewed as an n'th order Fourier reconstruction of the inverse of the spectral density. Specifically, by Lemma A.1

$$\begin{aligned}&\lim_{m \to \infty} \frac{1}{m} \Delta_{2m}(\omega, \delta) \triangleq \Delta(\omega, \delta) \\ &= \frac{\sigma^2}{|A_\circ|^2} Z_\circ^\star\left[\frac{Z_\circ \Phi_u(\omega) Z_\circ^\star}{|A_\circ|^2} + \delta I_2\right]^{-1} Z_\circ.\end{aligned} \tag{10}$$

Therefore, using the matrix inversion lemma provides

$$\Delta(\omega, \delta) = \frac{\sigma^2}{|A_\circ|^2} \frac{Z_\circ^\star Z_\circ |A_\circ|^2}{\Phi_u Z_\circ^\star Z_\circ + \delta |A_\circ|^2}.$$

Therefore

$$\lim_{\delta \to 0} \Delta(\omega, \delta) = \frac{\sigma^2}{\Phi_u(\omega)}$$

and hence

$$\lim_{\delta \to 0} \lim_{m \to \infty} \frac{1}{m} \Delta_{2m}(\omega, \delta) = \frac{\sigma^2}{\Phi_u(\omega)}. \tag{11}$$

Consequently, assuming that the convergence has approximately occurred for finite m leads to the approximation

$$\Delta_{2m}(\omega) \approx m \frac{\sigma^2}{\Phi_u(\omega)} \tag{12}$$

and in a similar vein, assuming that convergence of (9) has approximately occurred for finite N provides the approximation

$$\mathbf{E}\left\{|\widetilde{G}_N(\omega)|^2\right\} \approx \frac{1}{N} \Delta_{2m}(\omega).$$

Combining these expressions then furnishes the overall approximation

$$\mathbf{E}\left\{|G_{\hat{\theta}_N}(e^{j\omega}) - G_\circ(e^{j\omega})|^2\right\} \approx \frac{m}{N} \frac{\sigma^2}{\Phi_u(\omega)} \tag{13}$$

which is, as mentioned earlier, a special case of (1) for the output-error model structure case of $\Phi_\nu = \sigma^2$.

4.1 Genesis of Impaired Approximation

The point of the preceding synopsis of the methods originating in (L.Ljung, 1985) is to isolate why, as illustrated in figure 1, the phenomenon of (1) providing poor approximation occurs.

Put simply, it arises since (1) is predicated on the convergence in (11) having approximately occurred for finite model order m so that (12) can be concluded and in fact, it is problematic as to whether the Fourier series convergence underlying this has in fact approximately converged.

In more detail, the convergence of (11) was highlighted to be one of Fourier series convergence in (10) of a matrix valued function

$$\frac{Z_\circ(e^{j\omega})\Phi_u(\omega)Z_\circ^\star(e^{j\omega})}{|A_\circ(e^{j\omega})|^2} + \delta I_2. \tag{14}$$

As is well known (Edwards, 1979), the rate of Fourier series convergence is governed by the smoothness of the function being reconstructed. Furthermore, the variation (and hence smoothness) of (14) is degraded by the division by $|A_\circ(e^{j\omega})|^2$ term since, supposing for simplicity that all the zeroes of $A_\circ(z)$ are in the left half plane, then (see figure 2) $|A_\circ(e^{j0})| \leq \eta^m$ for some $\eta < 1$ and $|A_\circ(e^{j\pi})| \geq \gamma^m$ for some $\gamma > 1$ so that division of a function by $|A_\circ(e^{j\omega})|^2$ can magnify the maximum and minimum values of that function by factors of $1/\eta^{2m}$ and $1/\gamma^{2m}$ respectively. Therefore, as the model order m

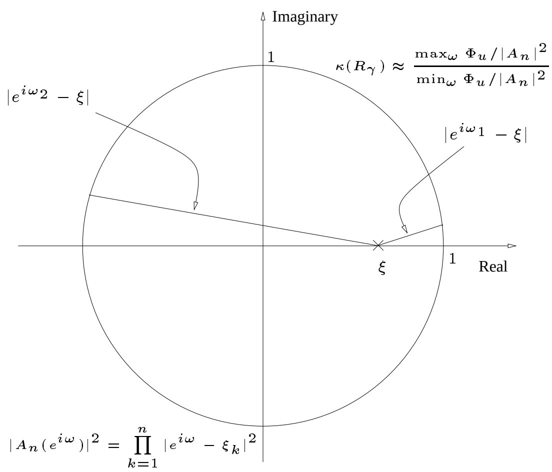

Figure 2: *Graphical illustration of how magnitude of $A_\circ(e^{j\omega})$ depends on ω.*

grows, the function (14) being implicitly Fourier reconstructed in (10) develops greater variation, which necessitates more terms in it's Fourier expansion before approximate convergence can be assumed in the step (12) leading to (1). But the number of terms in the implicit Fourier reconstruction (10) is also given by the quantity m.

The net result is that it is problematic as to whether Fourier convergence can be assumed to hold in such a way that the approximation (13) can be concluded. This appears, at a theoretical level, to be the genesis of the approximation discrepancy shown in figure 1.

5 Resolution

As pre-cursed via the plot of the modified expression (2) in figure 1, these Fourier series convergence difficulties are not insurmountable, the solution being to change the orthonormal basis involved in the Fourier series to one that is adapted to the function (14) being reconstructed. That is, the problem may be re-parameterised in terms of a new orthonormal basis that is 'adapted' to $A_{\theta_\circ}(z)$ in the sense that this polynomial is incorporated into the basis. A basis suitable for this has been discussed in (Ninness and Gustafsson, 1997), analysed in (Ninness *et al.*, 1998), applied to FIR and ARX model structures in (Ninness *et al.*, 1997), and is formulated as a sequence of rational functions $\{\mathcal{B}_0(z), \cdots, \mathcal{B}_{m-1}(z)\}$ defined by a choice of poles $\{\xi_0, \cdots, \xi_{m-1}\}$ all contained in the open unit disk $\mathbf{D} = \{z \in \mathbf{C} : |z| < 1\}$ by

$$\mathcal{B}_n(z) \triangleq \frac{\sqrt{1-|\xi_n|^2}}{z-\xi_n} \prod_{k=0}^{n-1} \left(\frac{1-\overline{\xi}_k z}{z-\xi_k} \right). \tag{15}$$

These functions are orthonormal with respect to the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega)\overline{g(\omega)}\, \mathrm{d}\omega$$

and the usual trigonometric basis $\{e^{-j\omega}, \cdots, e^{-jm\omega}\}$ (which is the one used in the previous analysis of expressions like (10)) is obtained as a special case of $\xi_0 = \xi_1 = \cdots = \xi_{m-1} = 0$. In the sequel the function $K_m(\omega, \mu)$ defined as

$$K_m(\omega, \mu) = \sum_{k=0}^{m-1} \mathcal{B}_k(e^{j\omega})\overline{\mathcal{B}_k(e^{j\mu})},$$

$$K_m(\omega, \omega) \triangleq \sum_{k=0}^{m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}$$

will be important (it happens to be the reproducing kernel associated with the space Span$\{\mathcal{B}_k\}_{k=0}^{m-1}$, hence the notation).

The use of this basis to obtain the approximation (2) appearing as the improved dashed line approximation in figure 1 now involves the analysis used in (L.Ljung, 1985) up until (8) being retained. However, if the poles $\{\xi_k\}$ of the bases (15) are chosen the same as the zeros of $A_{\theta_\circ}(z)$, then with the definition

$$\Gamma_m \triangleq [\mathcal{B}_0(z), \mathcal{B}_1(z), \cdots, \mathcal{B}_{m-1}(z)]^T$$

it holds that for some non-singular $n \times n$ matrix J, the matrix $\Lambda_m(z)$ appearing in (8), and defined in (5) in terms of a trigonometric basis, is expressible as $A_{\theta_\circ}^{-1}\Lambda_m = J\Gamma_m$. Therefore

$$(\Lambda_m \otimes I_2)A_\theta^{-1} = J\Gamma_m \otimes I_2 = (J \otimes I_2)(\Gamma_m \otimes I_2)$$

so that with the further (generalised block Toeplitz matrix) definition

$$M_n(F) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} [\Gamma_m \otimes I_2] F(\omega) [\Gamma_m^\star \otimes I_2]\, \mathrm{d}\omega \tag{16}$$

then the block Toeplitz matrix formulation (7) is expressible as

$$T_n(F/|A_\circ^2|) = (J \otimes I_2) M_n(F) (J^T \otimes I_2)$$

and hence the quantity $\Delta_n(\omega, \delta)$ previously analysed via Fourier theory with respect to the trigonometric basis becomes

$$\Delta_n(\omega, \delta) \quad = \quad \frac{\sigma^2}{|A_\circ|^2} Z_\circ^\star [\Gamma_m^\star \otimes I_2] M_n^{-1} \left(Z_\circ \Phi_u Z_\circ^\star + \delta |A_\circ|^2 I_2 \right) [\Gamma_m \otimes I_2] Z_\circ.$$

In (Ninness *et al.*, 1998, 1997) a generalised Fourier theory involving the generalised basis (15) is developed, for which the most pertinent result in the current context is the convergence one reproduced in the appendix as Lemma A.2. Applying it provides the conclusion

$$\begin{aligned} \lim_{m\to\infty} \frac{\Delta_{2m}}{K_m} &= \frac{\sigma^2}{|A_\circ|^2} Z_\circ^\star \left(Z_\circ \Phi_u Z_\circ^\star + \delta |A_\circ|^2 I_2 \right)^{-1} Z_\circ \\ &= \frac{\sigma^2 Z_\circ^\star Z_\circ}{\Phi_u(\omega) Z_\circ^\star Z_\circ + \delta |A_\circ|^2 I_2.} \end{aligned}$$

so that

$$\lim_{\delta\to 0} \lim_{m\to\infty} \frac{\Delta_{2m}}{K_m(\omega,\delta)} = \frac{\sigma^2}{\Phi_u(\omega)}.$$

A vital point here is that the implicit Fourier reconstruction operating here, by virtue of the use of the basis (15), is of a function $Z_\circ(e^{j\omega})\Phi_u(\omega)Z_\circ^\star(e^{j\omega})$ whose smoothness is constant with respect to Fourier reconstruction length m.

Therefore, following the same line of argument established in (Ljung, 1987) that led to (13) provides

$$\begin{aligned} \mathbf{E}\left\{ |G_{\hat\theta_N}(e^{j\omega}) - G_\circ(e^{j\omega})|^2 \right\} &\approx \frac{\sigma^2}{N} \frac{K_m(\omega,\omega)}{\Phi_u(\omega)} \\ &= \frac{1}{N} \frac{\Phi_\nu(\omega)}{\Phi_u(\omega)} \sum_{k=0}^{m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \end{aligned} \tag{17}$$

which is the new approximation (2).

6 Conclusion

The main theme of this paper was to highlight that the variability of quadratic cost prediction error estimates is not invariant to the choice of model structure, nor is it necessarily invariant to the dynamics of the actual system being estimated.

This is perhaps counter to previous thought that has argued that since only shift invariance is required of the model structure for (1) to hold, then since this depends only on model order, data length and signal-to-noise ratio, then in fact the precise choice of model structure or the dynamics of the estimated system are irrelevant to the variance error.

As illustrated here, this argument can fail since convergence of a certain Fourier series is central to the approximation (1) being accurate, and this can easily be upset in the output error model structure case. A strategy circumventing this was shown to involve re-parameterisation with a certain rational orthonormal basis, which lead to an extension (2) of (1) which can offer improved accuracy by explicitly accounting for factors (such as estimated pole positions) that may otherwise destroy convergence, and hence approximation.

A Technical Lemmata

In the following lemmata, the definitions (6) and (16) need to be recalled.

Lemma A.1 *Provided $F(\omega)$ of dimension $p\times p$ is positive definite and (componentwise) continuous for all $\omega\in[-\pi,\pi]$, then*

$$\lim_{m\to\infty} \frac{1}{m} [\Lambda_m^\star \otimes I_p] T_{mp}^{-1}(F) [\Lambda_m \otimes I_p] = F^{-1}(\omega)$$

componentwise and uniformly on $\omega\in[-\pi,\pi]$.

Proof: See (Hannan and Wahlberg, 1989). ∎

Lemma A.2 *Provided $F(\omega)$ of dimension $p \times p$ is positive definite and (componentwise) Lipschitz continuous of order $\alpha > 0$ for all $\omega \in [-\pi, \pi]$, then*

$$\lim_{m \to \infty} \frac{1}{K_m(\omega, \omega)} [\Gamma_m^\star \otimes I_p] M_{mp}^{-1}(F) [\Gamma_m \otimes I_p] = F^{-1}(\omega)$$

componentwise and uniformly on $\omega \in [-\pi, \pi]$.

Proof: See (Ninness *et al.*, 1997). ∎

References

Caines, P. (1988). *Linear Stochastic Systems*, John Wiley and Sons, New York.

Edwards, R. (1979). *Fourier Series, a Modern Introduction.*, vol. 2 of *Graduate texts in mathematics number 64.*, Springer Verlag.

Hannan, E. and B. Wahlberg (1989). "Convergence rates for inverse Toeplitz matrix forms," *Journal of Multivariate Analysis*, **31**, no. 1, pp. 127–135.

Ljung, L. (1987). *System Identification: Theory for the User*, Prentice-Hall, Inc., New Jersey.

L.Ljung (1978). "Convergence analysis of parametric identification methods," *IEEE Transactions on Automatic Control*, **AC-23**, no. 5, pp. 770–783.

L.Ljung (1985). "Asymptotic variance expressions for identified black-box transfer function models," *IEEE Transactions on Automatic Control*, **AC-30**, no. 9, pp. 834–844.

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

L.Ljung and Z.D.Yuan (1985). "Asymptotic properties of black-box identification of transfer functions," *IEEE Transactions on Automatic Control*, **30**, no. 6, pp. 514–530.

Ninness, B. and F. Gustafsson (1997). "A unifying construction of orthonormal bases for system identification," *IEEE Transactions on Automatic Control*, **42**, no. 4, pp. 515–521.

Ninness, B., H. Hjalmarsson, and F. Gustafsson (1997). "The fundamental role of general orthonormal bases in system identification," *Technical Report EE9739, Department of Electrical and Computer Engineering, University of Newcastle, Australia. Accepted subject to revision, IEEE Transactions on Automatic Control.*

Ninness, B., H. Hjalmarsson, and F. Gustafsson (1998). "Generalised Fourier and Toeplitz results for rational orthonormal bases," *Technical Report EE9740, Department of Electrical and Computer Engineering, University of Newcastle, Australia. To appear in SIAM Journal on Control and Optimization.*

P.M.J. Van den Hof, P.S.C. Heuberger, and J. Bokor (1995). "System identification with generalized orthonormal basis functions," *Automatica*, **31**, no. 12, pp. 1821–1834.

Wahlberg, B. (1991). "System identification using Laguerre models," *IEEE Transactions on Automatic Control*, **AC-36**, no. 5, pp. 551–562.

Wahlberg, B. (1994). "System identification using Kautz models," *IEEE Transactions on Automatic Control*, **AC-39**, no. 6, pp. 1276–1282.