ASYMPTOTIC ANALYSIS OF MIMO SYSTEM ESTIMATES BY THE USE OF ORTHONORMAL BASES

Brett Ninness [*,1] **Juan Carlos Gómez** [*,2]

[*] *Centre for Industrial Control Science and*
Department of Electrical and Computer Engineering
University of Newcastle
Callaghan 2308, AUSTRALIA

Abstract. This paper provides asymptotic bias and variance analysis for MIMO system estimates obtained by using generalizations of FIR model structures and least squares techniques. The generalizations are such that prior approximate knowledge of the system poles may be incorporated. The obtained variance expressions provide extensions to well known results that have previously been derived only for FIR model structures. Namely, the asymptotic covariance of the transfer matrix estimate is shown to be proportional not only to the noise-to-signal ratio, but also to a frequency dependent term that depends on the basis functions chosen. By examining a similar expression for the bias error it is shown that it is not possible to minimise the bias error at a particular frequency without increasing the variance error, and vice-versa.

Keywords. Estimation Theory, Identification, Frequency Response.

1. INTRODUCTION

In the last years there has been significant interest in developing schemes for identification of linear time-invariant SISO systems from input-output data, using orthonormal basis function model structures. In these methods, the transfer function of the system $G(z)$ is represented as a series expansion in the basis functions $\{\mathcal{B}_k(z)\}$, and then the identification is performed by estimating a finite number of expansion coefficients using least squares techniques.

The most common model structure is the well known FIR one, that corresponds to the choice $\mathcal{B}_k(z) = z^{-k}$ for the basis functions. However, as pointed out in (Wahlberg, 1991; Wahlberg and Hannan, 1993), the use of such a model structure implicitly involves expanding the true system transfer function around poles located at the origin. If the true poles are slow (i.e. far from the origin) the number of terms in the series expansion necessary to adequately represent the transfer function is high, and this may lead to poor accuracy in the estimated model.

To ameliorate this effect, one approach is to expand $G(z)$ around a pole $z = \xi \in \mathbf{R}$ near the dominating slow mode. This can be achieved using Laguerre models (Wahlberg, 1991; Wahlberg and Hannan, 1993) and amounts to the inclusion of a priori information about a single mode. The FIR model is a special case of the Laguerre structure corresponding to $\xi = 0$. When prior knowledge of a second order mode (ξ is complex) is to be included then a Kautz orthonormal model structure can be used (Wahlberg, 1994; Wahlberg, 1991). The Laguerre structure is a special case of the Kautz one when ξ is real.

When prior information about several modes is to be included in the identification process, then more general orthonormal bases such as those of Heuberger, and coworkers (Heuberger *et al.*, 1995), or those of Ninness and Gustafsson (Ninness and Gustafsson, 1994; Ninness and Gustafsson, 1996), may be employed. The Kautz, Laguerre and FIR model structures are all special cases of these methods.

In this paper the idea of using orthonormal bases in SISO system identification problems is extended to the MIMO setting. Except for the well known special case of FIR model structure studied in (Yuan and Ljung, 1984; Zhu, 1989), the use of more general bases in MIMO system identification has not been studied to date except for the work in (Heuberger *et al.*, 1995) and work by the current authors in (Ninness *et al.*, 1995).

This paper builds on the initial work reported in (Nin-

[1] This author gratefully acknowledges the support of the Australian Research Council
[2] This author gratefully acknowledges the financial support of the CICS

ness *et al.*, 1995) by extending the FIR model structure results of (Yuan and Ljung, 1984) to the case of using general orthonormal bases. More specifically, in (Yuan and Ljung, 1984) the variance of the FIR transfer function matrix estimate was shown to be approximately (for large data length and large model order) equal to

$$\frac{p}{N}\ \Phi_u^{-1}(\omega) \otimes \Phi_\nu(\omega)$$

where Φ_u is the input spectral density, Φ_ν is the output measurement noise spectral density, and $\otimes$ is the Kronecker matrix product. In this paper we show that for general bases functions $\{\mathcal{B}_k(z)\}$ the above expression should be changed to

$$\frac{1}{N}\Phi_u^{-1}(\omega) \otimes \Phi_\nu(\omega) \times \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2.$$

Note that for FIR models $\sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2 = p$ so that the new expression contains the previously known FIR model structure result as a special case.

The rest of the paper is organized as follows. In Section 2, the identification problem is stated. A description of the MIMO system is presented, where the transfer matrix is represented as a series expansion in scalar bases with matrix coefficients. With this parameterization, the system can be put in linear regressor form, so that standard least squares techniques can be used for the parameter estimation. In Section 3. we analize the so-called 'undermodelling' induced error that results from the parsimony of the model structure due to a finite number of terms in the series expansion representing the transfer matrix of the system. In Section 4., the main contribution of the paper is derived. Namely, an expression for the asymptotic (in the number of data and in the model order) covariance of the transfer matrix estimate is obtained. The derived result indicates that the asymptotic covariance of the transfer matrix estimate is proportional to the (generalized) noise-to-signal ratio. This result is consistent with the results presented in (Yuan and Ljung, 1984) for the particular case of FIR model structures.

2. PROBLEM FORMULATION

2.1 *MIMO Model Structure*

We address the problem of identification of linear time-invariant MIMO systems from observed input-output data. To be more specific, the paper assumes that the system has n inputs and m outputs and that an N point data record of n input sequences $\{u_k^1\}, \{u_k^2\}, \ldots, \{u_k^n\}$, as well as m output sequences $\{y_k^1\}, \{y_k^2\}, \ldots, \{y_k^m\}$ is available. We assume also that the data are related according to

$$y_k^i = \sum_{j=1}^{n} G_{ij}(q) u_k^j + \sum_{s=1}^{m} H_{is}(q) e_k^s \qquad (1)$$

with $i = 1, 2, \cdots, m$.

The scalar transfer functions $\{G_{ij}(q)\}$ describe the (assumed stable) unknown system dynamics that are to be identified. A notational simplification is possible by vectorising:

$$y_k \stackrel{\triangle}{=} (y_k^1, y_k^2, \ldots, y_k^m)^T, \quad u_k \stackrel{\triangle}{=} (u_k^1, u_k^2, \ldots, u_k^n)^T,$$

$$e_k \stackrel{\triangle}{=} (e_k^1, e_k^2, \ldots, e_k^m)^T$$

$$G(q) = \begin{bmatrix} G_{11}(q) & G_{12}(q) & \ldots & G_{1n}(q) \\ G_{21}(q) & G_{22}(q) & \ldots & G_{2n}(q) \\ \vdots & \vdots & \ldots & \vdots \\ G_{m1}(q) & G_{m2}(q) & \ldots & G_{mn}(q) \end{bmatrix},$$

$$H(q) = \begin{bmatrix} H_{11}(q) & H_{12}(q) & \ldots & H_{1m}(q) \\ H_{21}(q) & H_{22}(q) & \ldots & H_{2m}(q) \\ \vdots & \vdots & \ldots & \vdots \\ H_{m1}(q) & H_{m2}(q) & \ldots & H_{mm}(q) \end{bmatrix},$$

so that (1) can be rewritten in matrix form

$$y_k = G(q)\ u_k + H(q)\ e_k = G(q)\ u_k + \nu_k. \qquad (2)$$

Here $\{e_k\}$ is assumed to be a sequence of independent random vectors, with zero mean values and covariance matrix Λ. Hence, the disturbance term $\nu_k = H(q)\ e_k$ will be a stationary process with spectral density

$$\Phi_\nu(\omega) = H(e^{j\omega})\Lambda H^\star(e^{j\omega})$$

The model is parameterized by representing the transfer matrix $G(q)$ as a truncated series expansion in the general scalar orthonormal basis $\{\mathcal{B}_\ell(q)\}$ introduced in (Ninness and Gustafsson, 1994; Ninness and Gustafsson, 1996), as follows

$$G(q, \Theta) = \sum_{\ell=0}^{p-1} \theta_\ell^T \mathcal{B}_\ell(q) \qquad (3)$$

where θ_ℓ^T are m by n matrices of expansion coefficients. The expression for the elements $\mathcal{B}_\ell(z)$ of the bases is given by (Ninness and Gustafsson, 1994; Ninness and Gustafsson, 1996)

$$\mathcal{B}_\ell(z) = \left(\frac{\sqrt{1-|\xi_\ell|^2}}{z-\xi_\ell}\right) \prod_{k=0}^{\ell-1} \left(\frac{1-\overline{\xi_k} z}{z-\xi_k}\right) \qquad (4)$$

and allows for prior information about several modes to be incorporated in the estimation process.

Notice that if any of the poles $\{\xi_k\}$ are chosen as complex, then the formulation (4) has a complex valued impulse response which is inappropriate. In (Ninness and Gustafsson, 1994; Ninness and Gustafsson, 1996) it is shown how this may easily be circumvented by still using the construction (4), but if ξ_n is chosen as complex, then another pole ξ_{n+1} must also be chosen as the complex conjugate $\xi_{n+1} = \overline{\xi}_n$. This leads now to two basis function $\mathcal{B}_n$ and $\mathcal{B}_{n+1}$ with complex valued impulse responses. The idea now is that these may be linearly combined in an infinite variety of ways to yield two new basis function $\mathcal{B}'_n, \mathcal{B}''_n$ which have the same complex valued poles, are orthonormal to one another, and have **real** valued impulse responses. These latter two basis function are the ones used in the identification procedure. The details of this linear combination are unimportant here, but lead to $\mathcal{B}'_n$ having the form

$$\mathcal{B}'_n(z) = \frac{\sqrt{1-|\xi_n|^2}(\beta z + \mu)}{z^2 + (\xi_n + \overline{\xi_n})z + |\xi_n|^2} \prod_{k=0}^{n-1} \left(\frac{1 - \overline{\xi_k} z}{z - \xi_k} \right)$$

where $x = (\beta, \mu)$ is any choice lying on the ellipse

$$x^T M x = |1 - \xi_n^2|^2 \tag{5}$$

with

$$M \triangleq \begin{pmatrix} 1 + |\xi_n|^2 & 2\mathrm{Re}\{\xi_n\} \\ 2\mathrm{Re}\{\xi_n\} & 1 + |\xi_n|^2 \end{pmatrix}.$$

The next basis function $\mathcal{B}''_n$ is then found as

$$\mathcal{B}''_n(z) = \frac{\sqrt{1-|\xi_n|^2}(\beta' z + \mu')}{z^2 + (\xi_n + \overline{\xi_n})z + |\xi_n|^2} \prod_{k=0}^{n-1} \left(\frac{1 - \overline{\xi_k} z}{z - \xi_k} \right)$$

with (β', μ') related to the initial choice of (β, μ)

$$\begin{pmatrix} \beta' \\ \mu' \end{pmatrix} = \frac{1}{\sqrt{1-\alpha^2}} \begin{pmatrix} \alpha & 1 \\ -1 & -\alpha \end{pmatrix} \begin{pmatrix} \beta \\ \mu \end{pmatrix} \; ; \alpha \triangleq \frac{\xi_n + \overline{\xi_n}}{1 + |\xi_n|^2}. \tag{6}$$

A special case of this construction is when only one fixed complex mode $\xi_k = \xi$ is considered and where the following special choice satisfying (5) is made

$$(\beta, \mu) = \left(0, \sqrt{(1-\alpha^2)(1+|\xi_n|^2)}\right)$$

in which case (6) gives

$$(\beta', \mu') = \sqrt{(1+|\xi_n|^2)}(1, -\alpha).$$

which is the so-called Kautz basis employed in (Wahlberg, 1994). Different initial choices for (β, μ) satisfying (5) give an infinite number of second order bases other than the Kautz one.

The representation in (3) leads to the convenient linear regressor form

$$\begin{aligned} y_k &= G(q,\Theta)u_k + \nu_k \\ &= \left(\theta_0^T, \theta_1^T, \cdots, \theta_{p-1}^T\right) \begin{pmatrix} I_n\mathcal{B}_0(q) \\ I_n\mathcal{B}_1(q) \\ \vdots \\ I_n\mathcal{B}_{p-1}(q) \end{pmatrix} u_k + \nu_k \\ &= \Theta^T \Gamma_p(q) u_k + \nu_k \\ &= \Theta^T \phi_k + \nu_k \end{aligned} \tag{7}$$

where we have defined

$$\begin{aligned} \Theta^T &= \left(\theta_0^T, \theta_1^T, \cdots, \theta_{p-1}^T\right) \\ \Gamma_p^T(q) &= (I_n\mathcal{B}_0(q), I_n\mathcal{B}_1(q), \cdots, I_n\mathcal{B}_{p-1}(q)) \\ \phi_k &= \Gamma_p(q) u_k \end{aligned}$$

2.2 *Least Squares Estimation*

Now, the most obvious scheme for estimating the parameter matrix Θ is the least squares method. The least squares estimate $\widehat{\Theta}$ of Θ is given by

$$\widehat{\Theta} = \arg\min_{\Theta} \frac{1}{N} \mathrm{Tr} \left\{ \sum_{k=0}^{N-1} \varepsilon_k(\Theta) \varepsilon_k^T(\Theta) \right\}, \tag{8}$$

$$\varepsilon_k(\Theta) = y_k - G(q,\Theta)u_k. \tag{9}$$

Adopting the vectorized notation

$$\begin{aligned} Y^T &= (y_0, y_1, \cdots, y_{N-1}) \\ \Phi^T &= (\phi_0, \phi_1, \cdots, \phi_{N-1}) \\ V^T &= [\nu_0, \nu_1, \cdots, \nu_{N-1}] \end{aligned}$$

the model for the N point observed data record can be written as:

$$Y = \Phi\Theta + V. \tag{10}$$

It is then well known that the estimate $\widehat{\Theta}$ defined in equation (8) can be written in closed form as

$$\widehat{\Theta} = \Phi^\dagger Y \tag{11}$$

where $\Phi^\dagger$ is the left pseudoinverse of Φ. If there is sufficient input excitation for the indicated inverse to exist this will be given by

$$\Phi^\dagger = (\Phi^T\Phi)^{-1}\Phi^T. \tag{12}$$

The transfer matrix estimate is then given by

$$G(e^{j\omega}, \widehat{\Theta}) = \widehat{\Theta}^T \Gamma_p(e^{j\omega}) \tag{13}$$

The rest of this paper is concerned with quatifying the error $G(e^{\omega}) - G(e^{j\omega}, \widehat{\Theta})$ in the estimated frequency response.

3. UNDERMODELLING ERROR

The section addresses the question of analysing the performance of the proposed system identification scheme. In our limited context of linear time invariant systems, estimation errors can arise from two sources. One of them is a so-called 'undermodelling' induced error that results from the parsimony of the model structure (3) which cannot completely describe the true dynamics $G(q)$. The second source of errors arises from the noise corruption $H(q)e_k$ of the measurements.

For the SISO case and for particular orthonormal bases the undermodelling error quantification has been performed in (Wahlberg, 1991; Wahlberg, 1994; Heuberger *et al.*, 1995). The methods employed in these works involve establishing an isomorphism between the bases used in those works to the common FIR basis $\{e^{j\omega n}\}$. Standard techniques from Fourier analysis are then employed.

In comparison with the bases studied in (Wahlberg, 1991; Wahlberg, 1994; Heuberger *et al.*, 1995), the bases (4) studied in this paper enjoy greater flexibility in the possible choice of pole location. Unfortunately, there is a price to be paid for this increased flexibility in that the isomorphic (with respect to FIR) structure is lost.

Nevertheless, it is still possible to develop results equivalent to those in (Wahlberg, 1991; Wahlberg, 1994; Heuberger *et al.*, 1995), but with considerably more effort. The main idea is to draw inspiration from the Szegö theory of orthonormal polynomials (Szegö, 1959) and develop a so-called 'Christoffel-Darboux' formula for the Reproducing Kernel associated with the basis. This result and the following ones were originally presented in (Ninness *et al.*, 1995) but are re-presented here for ease of comparison to following new results quantifying the noise induced estimation error.

Theorem 1. Christoffel-Darboux formula:

Define the Blaschke product

$$\varphi_p(z) = \prod_{k=0}^{p-1} \frac{1 - \overline{\xi_k} z}{z - \xi_k}.$$

Then for $|\mu| < 1, |z| < 1$ the Reproducing Kernel of the orthonormal system $\{\mathcal{B}_k\}$ can be expressed as

$$\sum_{k=0}^{p-1} \overline{\mathcal{B}_k(\mu^{-1})} \mathcal{B}_k(z^{-1}) = \left\{ \frac{1 - \overline{\varphi_p(\mu^{-1})} \varphi_p(z^{-1})}{1 - \overline{\mu} z} \right\} \overline{\mu} z.$$

PROOF: This is too lengthy to be presented here. It does not use the methods of Szegö which rely heavily on the properties of polynomials. Rather the proof is constructive.

In (Ninness and Gustafsson, 1994) the basis (4) was shown to be H_2 complete if and only if $\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$. The Christoffel-Darboux like result in the previous theorem shows that under the same conditions for H_2 completeness, (4) is also complete in the disk algebra; this latter space is commonly employed in the so-called 'Estimation in H_∞' literature.

Perhaps more importantly though, the previous Theorem allows the undermodelling error to be bounded as follows.

Theorem 2. Suppose $G_{ij}(z)$ has partial fraction expansion

$$G_{ij}(z) = \sum_{\ell=0}^{r-1} \frac{\alpha_\ell^{ij}}{z - \gamma_\ell^{ij}}$$

where all the poles satisfy $|\gamma_\ell^{ij}| < 1$. Put $\widehat{G}_{ij}(z)$ as the best H_2 approximation to $G_{ij}(z)$ with respect to the p basis functions $\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{p-1}\}$

$$\widehat{G}_{ij}(z) = \sum_{k=0}^{p-1} \langle G_{ij}, \mathcal{B}_k \rangle \, \mathcal{B}_k(z).$$

Then

$$|G_{ij}(e^{j\omega}) - \widehat{G}_{ij}(e^{j\omega})| < \sum_{\ell=0}^{r-1} \left| \frac{\alpha_\ell^{ij}}{e^{j\omega} - \gamma_\ell^{ij}} \right| \left| \prod_{k=0}^{p-1} \left| \frac{\gamma_\ell^{ij} - \xi_k}{1 - \overline{\xi_k} \gamma_\ell^{ij}} \right| \right. \tag{14}$$

PROOF: Use Cauchy's Integral Theorem to write $G_{ij}(\mu^{-1})$ with $|\mu| < 1$ as

$$G_{ij}(\mu^{-1}) = \frac{1}{2\pi j} \oint_{|z|=1} \frac{G_{ij}(z^{-1})}{z - \mu} \, \mathrm{d}z$$

and then use the previous Theorem together with Cauchy's Residue Theorem and the Triangle inequality.

Note that this bound is tight in the sense that if perfect knowledge of the poles of $G(z)$ is available (so that $\xi_n = \gamma_k^{ij}$ for some n) then the upper bound is zero. The interpretation of the result is that the convergence of the expansion (3) to the true $G(z)$ can be very much

faster than that of the special case of FIR expansion ($\xi_k = 0, \forall k$) if reasonable guesses $\{\xi_k\}$ of the poles of $G(z)$ can be made. This is so since

$$\frac{z - \xi_k}{1 - \overline{\xi_k} z} \tag{15}$$

is analytic on $\mathbf{D}$ and of modulus 1 on the boundary $\mathbf{T}$. Therefore, by the maximum modulus theorem, the factor (15) with $z = \gamma_\ell^{ij}$ that appears p times in the bound (14) is of modulus less than one, and hence the bound (14) decreases **geometrically** with model order p; the geometric rate will depend on the error term $|\gamma_\ell^{ij} - \xi_k|$ in the guess ξ_k in the true pole position γ_ℓ^{ij}. For example, with $r = 1$ and $\gamma_\ell^{ij} = 0.8$ then if all the guesses ξ_k satisfy $|\xi_k| > 0.4$ the approximation error is at least $(0.8/0.4 \times 0.68)^{10} \approx 22$ times smaller for a $p = 10$th order expansion in the general basis (4) than it is for the more common FIR expansion.

4. ASYMPTOTIC COVARIANCE OF THE TRANSFER MATRIX ESTIMATE

In this section we analyze the asymptotic properties of the measurement noise induced variability in the estimated frequency response of the system. Since $G(e^{j\omega}, \widehat{\Theta})$ given by (13) is a matrix of estimated frequency responses and we want to quantify the covariance in the individual entries in this matrix, it will be more convenient to introduce a vectorized notation as follows

$$\begin{aligned} y_k &= \text{vec}\,\{\Theta^T \phi_k\} + \nu_k \\ &= \underbrace{\left(\phi_k^T \otimes I_m\right)}_{=\psi_k^T} \underbrace{\text{vec}\,\Theta^T}_{=\eta} + \nu_k \\ &= \psi_k^T \eta + \nu_k \end{aligned} \tag{16}$$

Since equation (16) is still in linear regressor form, the least squares estimate $\widehat{\eta}$ of η is given by the well known expression

$$\widehat{\eta} = \left(\sum_{k=0}^{N-1} \psi_k \psi_k^T \right)^{-1} \sum_{k=0}^{N-1} \psi_k y_k$$

provided that the indicated inverse exists. It is not difficult to show that $\widehat{\eta} = \text{vec}\ \widehat{\Theta}^T$. However, $\widehat{\eta}$ is more convenient to work with than $\widehat{\Theta}$ since it linearly parameterizes $\text{vec}\left\{G(e^{j\omega}, \widehat{\Theta})\right\}$ as

$$\begin{aligned} g(e^{j\omega}, \widehat{\eta}) &= \text{vec}\,\{\widehat{\Theta}^T \Gamma_p(e^{j\omega})\} \\ &= \left(\Gamma_p^T(e^{j\omega}) \otimes I_m\right) \text{vec}\ \widehat{\Theta}^T \\ &= \Psi(e^{j\omega}) \widehat{\eta} \end{aligned} \tag{17}$$

where we have defined $\Psi(e^{j\omega}) \triangleq \Gamma_p^T(e^{j\omega}) \otimes I_m$.

We are now able to compute the asymptotic covariance of the transfer function estimate $g(e^{j\omega}, \widehat{\eta})$ given by (17) when the number of data N together with the model order p increase to infinity. The main result is summarized in the following theorem.

Theorem 3. Let $\Phi_u(\omega)$ and $\Phi_\nu(\omega)$ be the input and noise spectral densities, and let $\gamma_p(\omega)$ be defined as

$$\gamma_p(\omega) = \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2$$

Then, for $H(e^{j\omega}) = I, \Phi_u > 0$ and Φ_u having finite-dimensional spectral factorization,

$$\lim_{p\to\infty} \lim_{N\to\infty} \frac{N}{\gamma_p} \text{Cov}\left\{g(e^{j\omega}, \widehat{\eta})\right\} = \Phi_u^{-1}(\omega) \otimes \Phi_\nu(\omega)$$

PROOF: Due to space limitations only a sketch of the proof is presented here.

By Theorem 9.1 in (Ljung, 1987)

$$\lim_{N\to\infty} N\ \text{Cov}\ \widehat{\eta} = R_p^{-1} Q_p R_p^{-1}$$

where it is possible to show that R_p and Q_p have spectral representations

$$\begin{aligned} R_p &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_p(\omega) \Phi_u(\omega) \Gamma_p(\omega)^\star\ \mathrm{d}\omega \otimes I_m, \\ Q_p &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_p(\omega) \Phi_u(\omega) \Gamma_p(\omega)^\star \otimes \Phi_\nu(\omega)\ \mathrm{d}\omega + \Delta_p, \end{aligned}$$

where $\| \Delta_p \| \to 0$ as $p \to \infty$.

Therefore, for $\text{E}\left\{e_k e_k^T\right\} = \Lambda$ and $H(e^{j\omega}) = I$ we have that since

$$g(e^{j\omega}, \widehat{\eta}) = \left(\Gamma_p(\omega)^T \otimes I_m\right)\ \widehat{\eta}$$

then

$$\lim_{N\to\infty} N\text{Cov}\left\{g(e^{j\omega}, \widehat{\eta})\right\} = \Gamma_p(\omega)^\star M_p^{-1}(\Phi_u) \Gamma_p(\omega) \otimes \Lambda$$

where $M_p(\cdot)$ is defined as

$$M_p(W) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_p(\omega) W(\omega) \Gamma_p(\omega)^\star\ \mathrm{d}\omega$$

for an aribitrary spectral density $W(\omega)$. In this case

$$\lim_{\substack{N\to\infty\\ p\to\infty}} \frac{N}{\gamma_p}\mathrm{Cov}\left\{g(e^{j\omega},\widehat{\eta})\right\} = \lim_{p\to\infty} \frac{\Gamma_p^\star M_p^{-1}(\Phi_u)\Gamma_p}{\gamma_p}\otimes\Lambda$$

$$= \lim_{p\to\infty}\left\{\frac{\Gamma_p^\star M_p(\Phi_u^{-1})\Gamma_p}{\gamma_p} + \right.$$

$$\left. + \frac{\Gamma_p^\star M_p^{-1}(\Phi_u)\left[I - M_p(\Phi_u)M_p(\Phi_u^{-1})\right]\Gamma_p}{\gamma_p}\right\}\otimes\Lambda \quad (18)$$

But then it is possible to show that

$$\lim_{p\to\infty} \frac{\Gamma_p(\omega)^\star M_p(\Phi_u^{-1})\Gamma_p(\omega)}{\gamma_p} = \Phi_u^{-1}$$

and that for $W, X > 0$ with finite dimensional spectral factorizations

$$M_p(W)M_p(X) \overset{p\to\infty}{\longrightarrow} M_p(WX)$$

Applying this last result to the second term in (18) and using the fact that it can be shown that if $\Phi_u > 0$ then $\| M_p^{-1}(\Phi_u) \| < \infty$, yields the result.

This result extends the more preliminary analysis contained in (Ninness *et al.*, 1995) where only bounds on the asymptotic variance (as opposed to the convergence results presented here) were given. However, the result in the previous Theorem only pertains to temporally white measurement noise. Work is in progress to obtaining a complete convergence result that applies for temporally coloured measurement noise ($H(e^{j\omega}) \neq I$).

Together, theorems 2 and 3 illustrate a fundamental limitation of basis function type estimates in terms of a trade-off between bias and variance errors. To be more specific, suppose that $G(z)$ contains a resonant pole at $\gamma \in \mathbf{C}$. Then according to theorem 2, in order to minimise undermodelling (bias) errors, it is necessary to include a pole ξ_k near γ (ideally equal to γ) in the basis functions used. However, this strategy will cause a peak in the functions $\sum_{k=0}^{p-1}|\mathcal{B}_k(e^{j\omega})|^2$ near Imγ, so that by theorem 3 the noise induced (variance) error will be increased. In other words, theorems 2 and 3 show that decrease in undermodelling error is paid for by increase in measurement noise induced error. However, depending on the spectral densities Φ_u and Φ_ν, it is almost always possible to achieve a total error smaller than for an FIR model structure by sensible choice of $\{\xi_k\}$.

5. CONCLUSIONS

This paper has examined the idea of extending the strategy of least squares estimation of MIMO systems using FIR model structures to more general model structures called 'basis function' type structures. A result on the asymptotic variance error of the frequency response estimate was provided which extends earlier FIR results of Yuan and Ljung. It should be pointed out that 'basis function' type model structures are nothing more than model structures with fixed denominator. The utility of re-writing this structure in terms of orthogonal basis functions (aside from a possible numerical advantage) is to provide a mechanism for deriving theoretical results that aid in the practical application of fixed denominator estimation techniques.

6. REFERENCES

Heuberger, P., P. V. den Hof and O. Bosgra (1995). 'A generalized orthonormal basis for linear dynamical systems'. *IEEE Transactions on Automatic Control* **40**(3), 451–465.

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

Ninness, B., J. C. Gómez and S. Weller (1995). MIMO system identification using orthonormal basis functions. In 'Proceedings of the 34th IEEE Conference on Decision and Control, New Orleans'. pp. 703–708.

Ninness, B. and F. Gustafsson (1994). A unifying construction of orthonormal bases for system identification. In 'Proceedings of the 33rd IEEE Conference on Decision and Control'. pp. 3388–3393.

Ninness, B. and F. Gustafsson (1996). 'A unifying construction of orthonormal bases for system identification'. *To appear in IEEE Transactions on Automatic Control*.

Szegö, G. (1959). *Orthogonal Polynomials*. Vol. 23. Colloquium publications - American Mathematical Society. New York.

Wahlberg, B. (1991). 'System identification using Laguerre models'. *IEEE Transactions on Automatic Control* **AC-36**(5), 551–562.

Wahlberg, B. and E. Hannan (1993). 'Parameteric signal modelling using Laguerre filters'. *The Annals of Applied Probabililty* **3**(2), 467–496.

Wahlberg, B. (1991). Identification of resonant systems using Kautz filters. In 'Proceedings of the 30th Conference on Decision and Control'. pp. 2005–2010.

Wahlberg, B. (1994). 'System identification using Kautz models'. *IEEE Transactions on Automatic Control* **AC-39**(6), 1276–1282.

Yuan, Z. and L. Ljung (1984). 'Black box identification of multivariable transfer functions: Asymptotic properties and optimal input design'. *International Journal of Control* **40**, 233–256.

Zhu, Y.-C. (1989). 'Black box identification of MIMO transfer functions: Asymptotic properties of prediction error models'. *International Journal of Adaptive Control and Signal Processing* **3**, 357–373.