Brett Ninness [*,1] **Håkan Hjalmarsson** [**]

[*] *Dept. of Elec. & Comp. Eng, Uni. Newcastle, Australia. email:*`brett@ee.newcastle.edu.au`*, FAX: +61 2 49 21 69 93*
[**] *Dept. Sensors, Signals & Systems, Royal Inst. Technology, S-100 44 Stockholm, Sweden. email:*`hakan.hjalmarsson@s3.e.kth.se`*, FAX: +46 8 790 7329*

Abstract: There has been recent interest in using orthonormalised forms of fixed denominator model structures for system identification. A key motivating factor in the employment of these forms is that of improved numerical properties. Namely, for white input perfect conditioning of the least-squares normal equations is achieved by design. However, for the more usual case of coloured input spectrum, it is not clear what the numerical conditioning properties should be in relation to simpler and perhaps more natural model structures. This paper provides theoretical and empirical evidence to argue that in fact, even though the orthonormal structures are only designed to provide perfect numerical conditioning for white input, they still provide improved conditioning for a wide variety of coloured inputs.

Keywords: Estimation Algorithms, Estimation Theory, Identification Algorithms.

1. INTRODUCTION

The problems studied in this paper are ones in which N point data records of an input sequence $\{u_t\}$ and output sequence $\{y_t\}$ of a linear time invariant system are available. It is assumed that this data is generated as follows

$$y_t = G(q)u_t + \nu_t.$$

Here $G(q)$ is a stable (unknown) transfer function describing the system dynamics that are to be identified by means of the observations $\{u_t\}$, $\{y_t\}$, and the sequence $\{\nu_t\}$ is some sort of possible noise corruption. The input sequence $\{u_t\}$ is assumed to be quasi–stationary in the sense used by Ljung (Ljung (1987)) and also such that the associated spectral density satisfies $\Phi_u(\omega) > 0$.

The method of estimating the dynamics $G(q)$ which is of interest here is one wherein the following 'fixed denominator' model structure is used

$$G(q,\beta) = \sum_{k=0}^{p-1} \beta_k \mathcal{F}_k(q) \tag{1}$$

where the $\{\beta_k\}$ are real valued co-efficients and the transfer functions $\{\mathcal{F}_k(q)\}$ may be chosen in various ways, but in every case the poles of the transfer functions $\{\mathcal{F}_k(q)\}$ are selected from the set $\{\xi_0, \xi_1, \cdots, \xi_{p-1}\} \subset \mathbf{D}$ where $\mathbf{D} \triangleq \{z \in \mathbf{C} : |z| < 1\}$ with $\mathbf{C}$ being the field of complex numbers. These fixed poles $\{\xi_k\}$ are chosen by the user to reflect prior knowledge of the nature of $G(q)$. That is, in the interests of improved estimation accuracy, they are chosen as close as possible to where it is believed the true poles lie (Wahlberg (1991); Heuberger et al. (1995)).

An advantage of this simple model structure is that it is linearly parameterised in $\{\beta_k\}$, so that with $\beta \triangleq [\beta_0, \beta_1, \cdots, \beta_{p-1}]^T$ then the least-squares estimate

$$\widehat{\beta} = \arg\min_{\beta\in\mathbf{R}^p} \left\{ \frac{1}{N} \sum_{t-0}^{N-1} \left(y_t - G(q,\beta)u_t\right)^2 \right\} \tag{2}$$

is easily computed. Specifically, the solution $\widehat{\beta}$ to (2) can be written in closed form once the model structure (1) is cast in familiar linear regressor form notation as $G(q,\beta)u_t = \psi_t^T \beta$ where

$$\psi_t = \Lambda_p(q)\, u_t, \quad \Lambda_p(q) \triangleq [\mathcal{F}_0(q), \cdots, \mathcal{F}_{p-1}(q)]^T \tag{3}$$

so that (2) is solved as

$$\widehat{\beta} = \left(\sum_{t=0}^{N-1} \psi_t \psi_t^T \right)^{-1} \sum_{t=0}^{N-1} \psi_t y_t \tag{4}$$

provided that the input is persistently exciting enough for the indicated inverse to exist.

However, a large literature (Wahlberg (1991, 1994); Heuberger et al. (1995); P.M.J. Van den Hof et al. (1995); Bodin et al. (1996); Ninness and Gustafsson (1997)) has developed suggesting that instead of using the model structure (1), one should instead use its so-

[1] This work was supported by the Australian Research Council and the Centre for Integrated Dynamics and Control (CIDAC). Part of this work was completed while visiting S3-Automatic Control, The Royal Institute of Technology, Stockholm, Sweden.

$$G(q,\theta) = \sum_{k=0} \theta_k \mathcal{B}_k(q) \quad (5)$$

where now the $\{\mathcal{B}_k(q)\}$ are transfer functions such that

$$\text{Span}\{\mathcal{F}_0, \cdots, \mathcal{F}_{p-1}\} = \text{Span}\{\mathcal{B}_0, \cdots, \mathcal{B}_{p-1}\} \quad (6)$$

but also such that the $\{\mathcal{B}_k(q)\}$ are orthonormal with respect to the inner product

$$\langle \mathcal{B}_n, \mathcal{B}_m \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{B}_n(e^{j\omega}) \overline{\mathcal{B}_m(e^{j\omega})} \, \mathrm{d}\omega \quad (7)$$

where $\mathbf{T} \triangleq \{z \in \mathbf{C} : |z| = 1\}$ is the complex unit circle. There have been several orthonormal basis function formulations proposed in the literature (Heuberger et al. (1995); Wahlberg (1991, 1994); Bodin et al. (1996)) but this paper focuses on the particular choice discussed in (Ninness and Gustafsson (1997)) of

$$\mathcal{B}_n(q) = \frac{\sqrt{1-|\xi_n|^2}}{q-\xi_n} \prod_{k=1}^{n-1} \left(\frac{1-\overline{\xi_k} q}{q-\xi_k} \right). \quad (8)$$

In this case, defining in a manner analogous to the previous case

$$\phi_t = \Gamma_p(q) u_t, \;\; \Gamma_p(q) \triangleq [\mathcal{B}_0(q), \cdots, \mathcal{B}_{p-1}(q)]^T \quad (9)$$

then the least squares estimate with respect to the model structure (5) is given as

$$\widehat{\theta} = \left(\sum_{t=0}^{N-1} \phi_t \phi_t^T \right)^{-1} \sum_{t=0}^{N-1} \phi_t y_t \quad (10)$$

A key point is that since there is a linear relationship $\phi_t = J\psi_t$ for some non-singular J, then $\widehat{\beta} = J^T \widehat{\theta}$ and hence modulo numerical issues the least-squares frequency response estimate is invariant to the change in model structure between (1) and (5). Specifically:

$$\begin{aligned} G(e^{j\omega}, \widehat{\beta}) &= \Lambda_p^T(e^{j\omega}) \widehat{\beta} \\ &= \Lambda_p^T(e^{j\omega}) \left(\sum_{t=0}^{N-1} \psi_t \psi_t^T \right)^{-1} \sum_{t=0}^{N-1} \psi_t y_t \\ &= [J\Lambda_p(e^{j\omega})]^T \left(\sum_{t=0}^{N-1} \phi_t \phi_t^T \right)^{-1} \sum_{t=0}^{N-1} \phi_t y_t \\ &= \Gamma_p^T(e^{j\omega}) \widehat{\theta} = G(e^{j\omega}, \widehat{\theta}). \end{aligned}$$

Given this exact equivalence of frequency response estimates, it is important to question the motivation for using the structure (8) (which is complicated by the precise definition of the orthonormal bases (8) or whichever other one is used (Heuberger et al. (1995); Bodin et al. (1996))) in place of some other one equivalent orthonormalised version (5).

To date, a major part of addressing this question has been to motivate the use of the orthonormal form (5) along numerical conditioning lines (Wahlberg (1991, 1994); Heuberger et al. (1995); Ninness and Gustafsson (1997)). To elaborate further on this point, it is well known (Golub and Loan (1989)) that the numerical properties of the solution of the normal equations arising in least squares estimation using the model structures (1) and (5) are governed by the condition numbers $\kappa(R_\psi(N))$ and $\kappa(R_\phi(N))$ of the matrices

$$R_\psi(N) \triangleq \frac{1}{N} \sum_{t=1}^{N-1} \psi_t \psi_t^T, \;\; R_\phi(N) \triangleq \frac{1}{N} \sum_{t=1}^{N-1} \phi_t \phi_t^T$$

where the vectors ψ_t and ϕ_t are defined in (3) and (9) respectively. However, by the quasi–stationarity assumption and by Parseval's Theorem, the following limits exist

$$R_\psi \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N-1} \psi_t \psi_t^T = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Lambda_p \Lambda_p^\star \Phi_u \, \mathrm{d}\omega \quad (11)$$

$$R_\phi \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N-1} \phi_t \phi_t^T = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_p \Gamma_p^\star \Phi_u \, \mathrm{d}\omega, \quad (12)$$

(here $\cdot^\star$ denotes 'conjugate transpose') so that the numerical properties of least squares estimation using the model structures (1) and (5) should be closely related to the condition numbers $\kappa(R_\psi)$ and $\kappa(R_\phi)$. These condition number quantities, are defined for a matrix R as (Golub and Loan (1989))

$$\kappa(R) \triangleq \|R\| \, \|R^{-1}\|$$

which is clearly dependant on the matrix norm used. Most commonly, the matrix 2-norm is used (Golub and Loan (1989)), which for positive definite symmetric R is the largest positive eigenvalue. In this case $\kappa(R)$ is the ratio of largest to smallest eigenvalue of R, and is a measure of the Euclidean norm sensitivity of the the solution vector x of the equation $Rx = b$ to errors in the vector b. If not specified otherwise, it will be understood in this paper that this 2-norm defined condition number is being considered.

Now, for white input $\{u_t\}$, by definition its spectrum $\Phi_u(\omega)$ is a constant (say α) so that by orthonormality $R_\phi = \alpha I$ and hence the normal equations are perfectly numerically conditioned. However, an obvious question concerns how the condition numbers of R_ψ and R_ϕ compare for the more commonly encountered coloured input case. A key result in this context is that purely by virtue of the orthonormality in the structure (5), an upper bound on the conditioning of R_ϕ may be guaranteed for any Φ_u by virtue of the fact that (P.M.J. Van den Hof et al. (1995); Ninness et al. (1998)) ($\lambda(R)$ denotes the set of eigenvalues of the matrix R.)

responding to the general (non-orthonormal) structure (1). This suggests that the numerical conditioning associated with (5) might be superior to that of (1) across a range of coloured Φ_u, and not just the white Φ_u that the structure (5) is designed to be perfectly conditioned for.

However, in consideration of this prospect, it would seem natural to also suspect that even though $R_\phi = I$ is designed to occur for unit variance white input, that $R_\psi = I$ might equally well occur for some particular coloured input. If so, then in this scenario the structure (5) would actually be inferior to (1) in numerical conditioning terms. Therefore, in spite of the guarantee (13), it is not clear when and why the structure (5) should be preferred over the often-times simpler one (1) on numerical conditioning grounds.

This paper is devoted to examining these questions.

2. EXISTENCE OF SPECTRA

This section addresses the issue of the existence of a particular coloured Φ_u for which the non-orthonormal model structure (1) leads to perfect conditioning ($R_\psi = I$) and would thus make it a superior choice on numerical grounds than the 'orthonormal' structure (5). This issue is subsumed by that of designing a $\Phi_u(\omega)$ parameterised via real valued co-efficients $\{c_k\}$ as

$$\Phi_u(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{j\omega k} \tag{14}$$

and so as to achieve an arbitrary symmetric, positive definite R_ψ. In turn, this question may be formulated as the search for the solution set $\{\cdots, c_{-1}, c_0, c_1, \cdots\}$ such that

$$\sum_{k=-\infty}^{\infty} c_k \left(\frac{1}{2\pi j} \oint_{\mathbf{T}} \Lambda_p(z)\Lambda_p^\star(z) z^k \, \frac{\mathrm{d}z}{z} \right) = R_\psi$$

which (on recognising that since Φ_u is necessarily real valued then $c_k = c_{-k}$) may be more conveniently expressed as the linear algebra problem

$$\Pi \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \end{bmatrix} = \mathrm{vec}\{R_\psi\} \tag{15}$$

where the vec$\{\cdot\}$ operator is one which turns a matrix into a vector by stacking its columns on top of one another in a left-to-right sequence and the matrix Π, which will be referred to frequently in the sequel, is defined as

$$\Pi \triangleq \frac{1}{2\pi j} \oint_{\mathbf{T}} [\Lambda_p(z) \otimes I_p]\overline{\Lambda_p(z)}[1, z + z^{-1}, \cdots] \, \frac{\mathrm{d}z}{z}. \tag{16}$$

$$\begin{bmatrix} c_0 & c_1 & c_2 & \cdots \\ c_1 & c_0 & c_1 & \\ c_2 & & \ddots & \\ \vdots & & & \ddots \end{bmatrix}$$

is positive definite, which is a necessary and sufficient condition (S.Pillai and T.Shim (1993)) for $\Phi_u(\omega) > 0$.

Now it might be supposed that since (15) is an equation involving $p(p+1)/2$ constraints, but with an infinite number of degrees of freedom in the choice $c_0, c_1, \cdots$ then it should be possible to solve for an arbitrary symmetric positive definite R_ψ.

Perhaps surprisingly, this turns out not to be the case, the reason being that (as established in Theorem 4.1 following) the rank of Π in (16) is always only p. In fact therefore, the achievable R_ψ live only in a sub-manifold of the $p(p+1)/2$ dimensional manifold of $p \times p$ symmetric matrices, and this sub-manifold *may not contain a perfectly conditioned matrix*. Furthermore, as can be seen by (16), this sub-manifold that the possible R_ψ lie in will be completely determined by the choice of the functions $\mathcal{F}_k(z)$ in the model structure (1) and hence also in the definition for $\Lambda_p(z)$ in (3). These principles are most clearly exposed by considering a simple two dimensional example.

3. TWO DIMENSIONAL EXAMPLE

Consider the simplest case of $p = 2$ wherein there are only 3 constraints inherent in (15), and one may as well neglect the third row of $[\Lambda_p(z) \otimes I_p]\overline{\Lambda_p(z)}$ (since it is equal, by symmetry, to the second row) and instead consider

$$\begin{bmatrix} \mathcal{F}_0(z)\mathcal{F}_0(\frac{1}{z}) \\ \mathcal{F}_0(z)\mathcal{F}_1(\frac{1}{z}) \\ \mathcal{F}_1(z)\mathcal{F}_1(\frac{1}{z}) \end{bmatrix} = \begin{bmatrix} \mathcal{F}_0(\frac{1}{\xi_0})\mathcal{F}_0(z) + \frac{1}{z\xi_0}\mathcal{F}_0(\frac{1}{\xi_0})\mathcal{F}_0(\frac{1}{z}) \\ \mathcal{F}_1(\frac{1}{\xi_0})\mathcal{F}_0(z) + \frac{1}{z\xi_1}\mathcal{F}_0(\frac{1}{\xi_1})\mathcal{F}_1(\frac{1}{z}) \\ \mathcal{F}_1(\frac{1}{\xi_1})\mathcal{F}_1(z) + \frac{1}{z\xi_1}\mathcal{F}_1(\frac{1}{\xi_1})\mathcal{F}_1(\frac{1}{z}) \end{bmatrix}$$

where in forming the right hand side of the above equation it has been assumed that $\mathcal{F}_0(z)$ has a pole at $z = \xi_0$, $\mathcal{F}_1(z)$ has a pole at $z = \xi_1$, that $\mathcal{F}_0(0) \neq 0$, $\mathcal{F}_1(0) \neq 0$ and that $\xi_0, \xi_1 \in \mathbf{R}$. That is $\mathcal{F}_0(z)$ and $\mathcal{F}_1(z)$ are of the simple form ($\xi_0, \xi_1 \in \mathbf{R}$)

$$\mathcal{F}_0(z) \triangleq \frac{1}{z - \xi_0}, \quad \mathcal{F}_1(z) \triangleq \frac{1}{z - \xi_1}, \tag{17}$$

The advantage of the re–parameterisation into causal and anti–causal components is that it then straightforward to calculate Π from (16) as

$$\Pi = \begin{bmatrix} \mathcal{F}_0(\frac{1}{\xi_0})(\frac{1}{\xi_0}) & 2\mathcal{F}_0(\frac{1}{\xi_0}) & \cdots \\ \mathcal{F}_0(\frac{1}{\xi_1})(\frac{1}{\xi_1}) & \mathcal{F}_1(\frac{1}{\xi_0}) + \mathcal{F}_0(\frac{1}{\xi_1}) & \cdots \\ \mathcal{F}_1(\frac{1}{\xi_1})(\frac{1}{\xi_1}) & 2\mathcal{F}_1(\frac{1}{\xi_1}) & \cdots \end{bmatrix}. \tag{18}$$

provided that

$$\mathcal{F}_0(1/\xi_1)\xi_0 = \mathcal{F}_1(1/\xi_0)\xi_1 \qquad (20)$$

which is certainly true for the first order $\mathcal{F}_0(z), \mathcal{F}_1(z)$ in (17). Therefore, Π is of row (and hence column) rank no more than two. Therefore, regardless of the choice of Φ_u, it is only possible to manipulate (via change of Φ_u) the corresponding R_ψ in a two dimensional sub-manifold of the full three dimensional manifold of symmetric two-by-two matrices.

Furthermore, the identity matrix is not part of the two-dimensional sub-manifold, since if it were to lie in the subspace spanned by the columns of Π, it would have to be orthogonal to the normal vector specifying the orientation of this subspace (the left hand row vector in (19)). But it isn't, since

$$\left[\frac{\mathcal{F}_1(1/\xi_0)}{2\mathcal{F}_0(1/\xi_0)}, -1, \frac{\mathcal{F}_0(1/\xi_1)}{2\mathcal{F}_1(1/\xi_1)},\right] \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \neq 0$$

provided $\mathcal{F}_0, \mathcal{F}_1$ are of the form shown in (17). Therefore, even though Φ_u can be viewed as an infinite dimensional quantity, its effect on R_ψ is not powerful enough to achieve an arbitrary positive definite symmetric matrix. In particular, there is no Φ_u for which the simple and natural fixed denominator basis (17) is perfectly conditioned.

4. KEY RESULT

Given these motivating arguments specific to a two-dimensional case, it is of interest to consider the case of arbitrary dimension. As the arithmetic considered in the previous section illustrated, such a study will become very tedious as the dimension is increased. To circumvent this difficulty, the key idea of this section is to in fact replace the study of the rank of Π associated with an arbitrary basis $\{\mathcal{F}_n(q)\}$ by its rank with respect to the orthonormal basis $\{\mathcal{B}_n(q)\}$ specified in (8). Fundamental to this strategy is that via the span equivalence condition (6) the rank is invariant to the change of basis, so the most tractable one may as well be employed.

Using these ideas leads to the following key results exposing the limited flexibility available in the assignment of R_ϕ, R_ψ by manipulation of the spectral density Φ_u.

Theorem 4.1. With Π defined as in (16), and for all bases that maintain the same span as in condition (6) then

$$\operatorname{Rank}\Pi = p.$$

Proof: The main idea of the proof is to recognise that the rank of Π defined in (16) is invariant to a change of

This theorem exposes the key feature imbuing orthonormal parameterisations with numerical robustness beyond the white input case. Specifically, for white input, $R_\phi = I$ is perfectly numerically conditioned, while for this same white input $R_\psi \triangleq \Sigma \neq I$ which has inferior conditioning. As Φ_u is changed from the white case, both R_ϕ and R_ψ will change, but *but only in p-dimensional sub-manifolds*.

This feature of highly restricted mobility raises the possibility that since (by construction) I is in the manifold of possible R_ϕ, but may not (as the previous section illustrated) be in the manifold of possible R_ψ, then the orthonormal model structure (8) may imbue a numerical robustness to the associated normal equations across a range of coloured Φ_u.

5. ROBUSTNESS

Having indicated via theorem 4.1 that the orthonormal parameterisation provides a numerical conditioning advantage that is robust to the nature of the input spectral density Φ_u, this section delves deeper on this issue, and in order to do so it is expedient to split Φ_u into 'causal' and 'anti–causal' components as

$$\Phi_u(\omega) = \varphi(e^{j\omega}) + \varphi(e^{-j\omega}) \qquad (21)$$

where $\varphi(z)$ is known as the 'positive real' part of Φ_u and is given by the so-called Hergloz-Riesz transform (S.Pillai and T.Shim (1993)) as

$$\varphi(z) = \frac{c_0}{2} + \sum_{k=1}^{\infty} c_k z^k = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{1+ze^{j\omega}}{1-ze^{j\omega}}\right)\Phi_u(\omega)\,\mathrm{d}\omega. \qquad (22)$$

With this definition in hand, the following lemma will prove to be useful.

Lemma 5.1. The matrix R_ϕ defined via (12), (8) and (9) has entries given by

$$[R_\phi]_{m,n} = \begin{cases} \varphi(\xi_n) + \varphi(\overline{\xi_n}) & ; n = m, \\ \sum_{i=m}^{n} A^i_{m,n}\varphi(\overline{\xi_i}) & ; n > m \end{cases}$$

where

$$A^i_{m,n} \triangleq \frac{\sqrt{(1-|\xi_m|^2)(1-|\xi_n|^2)}(1-|\xi_i|^2)}{(1-\xi_m\overline{\xi}_i)(1-\xi_n\overline{\xi}_i)} \times \prod_{\substack{k=m \\ k\neq i}}^{n}\left(\frac{1-\xi_k\overline{\xi}_i}{\overline{\xi}_i - \xi_k}\right)$$

and it is understood that the array indexing of R_ϕ begins at $m, n = 0$.

Proof: See (Ninness et al. (1998)). ■

R_ϕ are contained in regions $\Delta_0, \Delta_1, \cdots, \Delta_{p-1}$ defined by

$$\Delta_m \triangleq \{x \in \mathbf{R} : |x - 2\mathrm{Re}\,\varphi(\xi_m)| \le \alpha_m\}$$

where

$$\alpha_m^2 \triangleq \sum_{\substack{n=0\\n\neq m}}^{p-1} \left(\sum_{i=m}^{n-1} |A^i_{m,n}||\varphi(\overline{\xi}_i) - \varphi(\overline{\xi}_{i+1})| \right)^2.$$

Proof: See (Ninness et al. (1998)). ■

Note that this theorem provides a tight characterisation in the sense that for white input, $\varphi(\xi_k) = c_0/2$ a constant, in which case the theorem provides the eigenvalues as being all at $\lambda_k = c_0$ with tolerance $\alpha_k = 0$.

However, more generally the theorem provides further indication of the general robustness of the condition number of R_ϕ. Specifically, if $\varphi(z)$ is smooth, then Theorem 5.1 indicates that since in this case the terms $|\varphi(\overline{\xi}_i) - \varphi(\overline{\xi}_{i+1})|$ will then be small, then the bounds α_m on the eigenvalue locations $\{2\mathrm{Re}\,\varphi(\xi_m)\}$ will be tight, and so the true eigenvalues should be very near to the locations $\{2\mathrm{Re}\varphi(\xi_m)\}$ which again if $\varphi(z)$ is smooth, will be relatively tightly constrained.

6. ASYMPTOTIC ANALYSIS

As mentioned in the introduction, a key feature of the orthonormal parameterisation (5) is that associated with it is a covariance matrix with numerical conditioning guaranteed by the bounds

$$\min_{\omega\in[-\pi,\pi]} \Phi_u(\omega) \le \lambda(R_\phi) \le \max_{\omega\in[-\pi,\pi]} \Phi_u(\omega). \quad (23)$$

A natural question to consider is how tight these bounds are. In (Ninness et al. (1998)), this was addressed by a strategy of analysis that is asymptotic in p. Specifically, define $M_\phi \triangleq \lim_{p\to\infty} R_\phi$. In this case, M_ϕ is an operator $\ell_2 \to \ell_2$, so that the eigenvalues of the finite dimensional matrix R, generalize to the continuous spectrum $\lambda(R_\infty)$ of the operator M_ϕ defined as (Böttcher and Silbermann (1983))

$$\lambda(M_\phi) = \{\lambda \in \mathbf{R} : \lambda I - M_\infty \ \text{ is not invertible}\}.$$

This spectrum can be characterized as follows.

Lemma 6.1. Suppose that $\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$. Then $\lambda(M_\phi) = \mathrm{Range}\{\Phi_u(\omega)\}$.

Proof: See (Ninness et al. (1998)). ■

This provides evidence, that at least for large p (when the issue of numerical conditioning is most important), that the bounds (23) are in fact tight, and therefore might be expected to be a reasonable approximation.

Of course, what would also be desirable is a similar approximation for R_ψ, and of course this will depend on the nature of the definition of the $\{\mathcal{F}_k(q)\}$. One particularly natural definition is that of

$$F_k(q) = \frac{z^k}{D_p(z)}, \qquad D_n(q) = \prod_{\ell=0}^{p-1}(z-\xi_\ell) \quad (25)$$

for $k = 0, 1, \cdots, p-1$ and $\{\xi_0, \cdots, \xi_{p-1}\} \in \mathbf{D}$ the fixed pole choices. This case is considered important, since possibly the most straightforward way of realising a fixed-pole estimate $G(q, \widehat{\beta})$ as originally defined in (2) of § 1 would be to simply use pre-existing software for estimating FIR model structures, but after having pre-filtering the input sequence $\{u_t\}$ with the all-pole filter $1/D_p(q)$. This is identical to using the general model structure (1) with the $\{\mathcal{F}_k(q)\}$ choice of (25) above, with estimated FIR co-efficients then simply being the numerator co-efficient estimates $\{\beta_0, \cdots, \beta_{p-1}\}$.

Fortunately, for this common structure, it is also possible to develop an approximation of the condition number $\kappa(R_\psi)$ via the following asymptotic result which is a direct corollary of Theorem 6.1.

Corollary 6.1. Consider the choice for the $\{\mathcal{F}_k(q)\}$ defining R_ψ via (3) and (11) given in (25). Suppose that only a finite number of the poles $\{\xi_k\}$ are chosen away from the origin so that

$$D(\omega) \triangleq \lim_{p\to\infty} \prod_{\ell=0}^{p-1} |e^{j\omega} - \xi_\ell|^2 \quad (26)$$

exits. Define, in a manner analogous to that pertaining to Lemma 6.1, the operator $M_\psi : \ell_2 \to \ell_2$ as

$$M_\psi \triangleq \lim_{p\to\infty} R_\psi.$$

Then

$$\lambda(M_\psi) = \mathrm{Range}\left\{\frac{\Phi_u(\omega)}{D(\omega)}\right\}.$$

Proof: See (Ninness et al. (1998)) ■

In analogy with the previous approximation, it is tempting to apply this asymptotic result for finite p to derive the approximation

$$\kappa(R_\psi) \approx \frac{\max_\omega \Phi_u(\omega)/|D_p(e^{j\omega})|^2}{\min_\omega \Phi_u(\omega)/|D_p(e^{j\omega})|^2}. \quad (27)$$

Now, considering that $|D_p(e^{j\omega})|^2 = \prod_{\ell=0}^{p-1}|e^{j\omega} - \xi_\ell|^2$ can take on both very small values (especially if some of the ξ_ℓ are close to the unit circle) and also very large values (especially if all the $\{\xi_\ell\}$ are chosen in the estimation with respect to the orthonormal form (5) could be expected to be much better conditioned than that with respect to the model structure (3) with the simple choice (25) for a very large class of Φ_u - an obvious exception here would be $\Phi_u = |D_p|^2$ for which $R_\psi = I$.

However, this conclusion depends on the accuracy of applying the asymptotically derived approximations (24) and (27) for finite p. In the absence of theoretical analysis, which appears intractable, simulation study can be pursued. Consider p in the range 2-30 with all the $\{\xi_\ell\}$ chosen at $\xi_\ell = 0.5$, and $\Phi_u(\omega) = 0.36/(1.36 - \cos\omega)$. Then the maximum and minimum eigenvalues for R_ψ and R_ϕ are shown as solid lines in the left and (respectively) right diagrams in figure (1). The dash-dot lines in these figures are the approximations (24) and (27). Clearly, in this case the approximations are quite accurate, even for what might be considered small p. Note that the minimum eigenvalue of R_ψ is shown only up until $p = 18$ since it was numerically impossible to calculate it for higher p. Again, this provides evidence that even

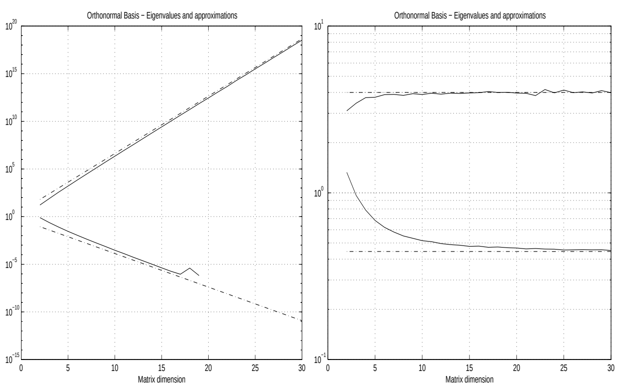

Fig. 1. *Solid lines are maximum and minimum eigenvalues of (left figure) R_ψ and (right figure) R_ϕ for a range of dimensions p. The dash dot lines are the approximations (24) and (27).*

though model structures (5) parameterised in terms of orthonormal $\{\mathcal{B}_k(q)\}$ are only designed to provide superior numerical conditioning properties for white input, they seem to also provide them for a very wide range of coloured inputs as well.

7. CONCLUSIONS

A variety of arguments have been presented to indicate that the condition numbers $\kappa(R_\psi)$ and $\kappa(R_\phi)$, which govern the numerical properties of least squares estimation associated with (respectively) simple 'fixed denominator' model structures and their orthonormalised forms, are such that $\kappa(R_\psi) \geq \kappa(R_\phi)$ for a very wide class of input spectra Φ_u. While this might be considered somewhat surprising, since it is only designed to occur (by the construction of the 'orthonormal' model structure) for white Φ_u, it is also analysis is made in counter-argument to the charge (as illustrated in the introduction), that a change of model structure is not the same as a change of estimation method - equivalent structures provide identical estimates, modulo the numerical issues considered here.

References

P.Bodin, T.Oliveira e Silva, and Bo Wahlberg. On the construction of orthonormal basis functions for system identification. In *Proc. 13'th IFAC World Congress*, pages 291–296, 1996.

J. Bokor, L. Gianone, and Z. Szabo. Construction of generalised orthonormal bases in $\mathcal{H}_2$. Tech. rep, Computer & Auto. Inst, Hungarian Acad. Sci, 1995.

A. Böttcher and B. Silbermann. *Invertibility and Asymptotics of Toeplitz Matrices*. Akadamie–Verlag, 1983.

G.Golub and C.Van Loan. *Matrix Computations*. Johns, 1989.

P.S.C. Heuberger, P.M.J. Van den Hof, and O.H. Bosgra. A generalized orthonormal basis for linear dynamical systems. *IEEE Trans. Auto. Cont.*, AC-40(3):451–465, 1995.

Horn and Johnson. *Matrix Analysis*. Cambridge Uni. Press, 1985.

Lennart Ljung. *System Identification: Theory for the User*. Prentice-Hall, 1987.

B.Ninness and F.Gustafsson. A unifying construction of orthonormal bases for system identification. *IEEE Trans. Auto. Cont.*, 42(4):515–521, 1997.

B.Ninness, H.Hjalmarsson, and F.Gustafsson. Generalised Fourier and Toeplitz results for rational orthonormal bases. *To appear in SIAM J. Cont. & Optim.*, 1999.

B.Ninness and H.Hjalmarsson. Model Structure and Numerical Properties of Normal Equations, *Tech Report EE9801, Dept. EE&CE, Uni. Newcastle, Aust.*, 1998.

P.M.J. Van den Hof, P.S.C. Heuberger, and J. Bokor. System identification with generalized orthonormal basis functions. *Automatica*, 31(12):1821–1834, 1995.

S.Pillai and T.Shim. *Spectrum Estimation and System Identification*. Springer–Verlag, 1993.

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

Bo Wahlberg. System identification using Kautz models. *IEEE Trans. Auto. Cont.*, AC-39(6):1276–1282, 1994.