Pergamon

PII: S0005-1098(98)00052-1

Automatica, Vol. 34, No. 9, pp. 1101–1117, 1998

Printed in Great Britain
0005-1098/98 $—see front matter

Rational Basis Functions for Robust Identification from Frequency and Time-Domain Measurements*

HÜSEYIN AKÇAY† and BRETT NINNESS‡

This paper presents general methods for robust identification from time and frequency domain measurements. Central to the work is the use of a class of rational bases that are orthonormal in H_2 and are natural extensions of the well-known Laguerre and two-parameter Kautz bases.

Key Words—Identification; estimation; worst-case analysis; error analysis; robustness.

Abstract—This paper investigates the use of general bases with fixed poles for the purposes of robust estimation. These bases, which generalise the common FIR, Laguerre and two-parameter Kautz ones, are shown to be fundamental in the disc algebra provided a very mild condition on the choice of poles is satisfied. It is also shown that, by using a min–max criterion, these bases lead to robust estimators for which error bounds in different norms can be explicitly quantified. The key idea facilitating this analysis is to re-parameterise the chosen model structures into a new one with equivalent fixed poles, but for which the basis functions are orthonormal in $H_2(\mathbf{D})$. © 1998 Elsevier Science Ltd. All rights reserved.

1. INTRODUCTION

In connection with the estimation of dynamic models on the basis of observed input–output measurements, many approaches have arisen that are predicated on a stochastic model for disturbances and a viewpoint that errors are best represented as averages over ensembles of possible noise realisations (Ljung, 1987; Söderström and Stoica, 1989; Caines, 1988). Complementary to this is a more recent school of thought that disturbances, and also estimation errors, may be characterised according to a deterministic model under which the worst-case amplitude is quantified (Mäkilä *et al.*, 1995; Ninness and Goodwin, 1995).

There are significant advantages to this latter approach, which is colloquially known as "robust identification". For example, errors due to nonlinearities are easily accommodated, and the resultant models and error bounds are of a form suitable for subsequent robust control design. The formalism of the robust identification school of thought is that discrete time linear models with impulse response sequences $g(k)$ are represented via an associated power series

$$G(z) = \sum_{k=0}^{\infty} g(k) z^k$$

which is the more common discrete time transfer function evaluated at $1/z$. The stability of the system may then be characterised according to $G(z)$ being analytic on the open unit disc $\mathbf{D} = \{z \in \mathbf{C}\colon |z| < 1\}$ or, if the degree of stability is at issue, analytic on an open disc $\mathbf{D}_R = \{z \in \mathbf{C}\colon |z| < R\}$ of radius R. When deriving estimation error bounds, the space in which the true system lies must be characterised, and with analyticity on $\mathbf{D}$ in hand, this leaves the behaviour of $G(z)$ on the boundary $\mathbf{T} = \{z \in \mathbf{C}\colon |z| = 1\}$ to be specified. If $G(z)$ is deemed to be continuous on $\mathbf{T}$, then it is more compactly described as being an element of the disc algebra $A(\mathbf{D})$ while if $G(z)$ is not necessarily continuous on $\mathbf{T}$, but if $|G(z)|^p$ is integrable on $\mathbf{T}$ then $G(z)$ may be succinctly described as being an element of the Hardy space $H_p(\mathbf{D})$. Finally, there is the possibility that one may wish to avoid frequency domain characterisations of $G(z)$ altogether, and instead characterise the system to be identified according to the space ℓ_p that the impulse response $\{g(k)\}$ lives in. Common choices here are ℓ_1 in which $\sum_k |g(k)| < \infty$ and ℓ_∞ in which $\sup_k |g(k)| < \infty$.

Typically, in these robust identification contexts, FIR model structures are employed. Recently,

* Received 24 June 1997; revised 15 December 1997; received in final form 4 March 1998. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor P. Van den Hof under the direction of Editor Torsten Söderström. *Corresponding author* Dr. H. Akçay Tel. +90 216 308 9432; Fax +90 216 308 9427.

†Feza Gürsey Institute, P.O. Box 6, Çengelköy 81220, İstanbul, Turkey. This author gratefully acknowledges support for this work from TÜBİTAK and CIDAC.

‡Centre for Integrated Dynamics and Control (CIDAC) and Department of Electrical and Computer Engineering, University of Newcastle, Callaghan, NSW 2308, Australia. This author gratefully acknowledges the support of CIDAC and the Australian Research Council.

however, in an effort to decrease the undermodelling-induced component of the estimation error, model structures allowing for the encoding of prior knowledge of pole positions have been introduced. For example, in Mäkilä (190) and Wählberg and Ljung (1992), it has been proposed that in trying to robustly estimate the dynamics $G(z)$, a model structure of the form

$$\hat{G}(z,\theta) = \sum_{k=0}^{n-1} \theta_k \mathscr{B}_k(z) \tag{1}$$

be employed where the functions $\{\mathscr{B}_k(z)\}$ are the so-called "Laguerre" basis functions specified as

$$\mathscr{B}_k(z) \triangleq \frac{\sqrt{1-a^2}}{1-az}\left(\frac{z-a}{1-az}\right)^k, \quad k=0,1,\ldots \tag{2}$$

for some fixed a with $-1 < a < 1$. By choosing a according to prior knowledge of the relative stability of $G(z)$, the undermodelling error can be reduced in comparison to the use of an FIR model structure (Mäkilä, 1990; Wahlberg, 1991), which is a special case of the Laguerre structure when $a = 0$.

In the case of systems $G(z)$ for which prior knowledge of a resonant mode exists, then it is more appropriate to employ the so-called two-parameter Kautz basis defined as follows (Wahlberg, 1994). Let

$$\zeta(z) = \frac{cz^2 + bz + 1}{z^2 + bz + c},$$

where b and c are fixed real numbers satisfying $b^2 - 4ac < 0$ and ζ has no poles in the closed unit disk. Let $\psi_0(z) = 1$, $\psi_1(z) = 1/(z^2 + bz + c)$ and $\psi_2(z) = z/(z^2 + bz + c)$. Let $\psi_k(z) = \zeta(z)\psi_{k-2}(z)$ for $k > 2$. Then Kautz models are obtained by orthonormalizing ψ_0, ψ_1 and ψ_2. Laguerre and Kautz models are special cases of general orthonormal bases (Heuberger *et al.*, 1995) where the poles are again restricted to a finite set. More recently, in Ward and Partington (1995, 1996), the "rational wavelet" basis

$$\mathscr{B}_w(z) \triangleq \frac{1}{1-\bar{w}z}, \quad w \in W, \tag{3}$$

has been suggested, where W is a set of discrete points in $\mathbf{D}$ and $\bar{\cdot}$ denotes complex conjugation. Denoting the linear span of the set $\{\mathscr{B}_0, \mathscr{B}_1, \ldots, \mathscr{B}_{n-1}\}$ as $X_n = \mathrm{sp}\{\mathscr{B}_k\}$, the wavelet basis enjoys the advantage of generalising the FIR, Laguerre, two-parameter Kautz, and general orthonormal bases in the sense that the points in W may trivially be chosen so that $\mathrm{sp}\{\mathscr{B}_w; w \in W\} = X_n$. Furthermore, by exploiting the great freedom in the choice of points in W, the wavelet basis would seem to have much greater utility in that it allows the injection of much more prior knowledge of the system $G(z)$. Intuitively, this should lead to smaller undermodelling-induced error when employed for the purposes of system identification (Ward and Partington, 1996).

In the context of robust estimation, perhaps a more important question is that of whether $\mathrm{sp}\{\mathscr{B}_w; w \in W\}$ can arbitrarily well approximate any given element $G(z)$ in the space in question, be it $A(\mathbf{D})$, $H_p(\mathbf{D})$ or ℓ_p. In the sequel, we will refer to this property by the formal definition of a set A being *fundamental* in a space X if the closure of the linear span of A under the norm on X is equal to X.

In Ward and Partington (1996), it was shown that a sufficient condition for $\{\mathscr{B}_w; w \in W\}$ to be fundamental in $A(\mathbf{D})$ was that W be a dyadically spaced lattice of the form:

$$W = \{\xi_{p,k}: \xi_{p,k} = (1 - 2^{-p})e^{j2\pi k/2^p}, \quad k = 0, \ldots, 2^p - 1; p = 0, 1, \ldots\}. \tag{4}$$

The dyadically spaced lattice above satisfies the so-called "Hayman–Lyons condition" considered in Hayman and Lyons (1990).. Many other lattices also satisfy this condition and in the construction of a wavelet basis, kernels different to the Cauchy kernel Equation (3) yet still parameterised by a lattice satisfying the Hayman–Lyons condition can also be employed. The Cauchy kernel was used in Ward and Partington (1996) due to its simplicity. As well, in Ward and Partington (1996), the reasoning behind choosing a dyadically spaced lattice was to provide approximation of systems with poles near the circle more efficiently than (for example) by polynomials.

One of the main results of this paper is to show that in fact a necessary and sufficient condition for $\{\mathscr{B}_w; w \in W\}$ to be fundamental in $A(\mathbf{D})$ and in $H_p(\mathbf{D})$ for all $1 \le p < \infty$ is that with W written as $W = \{\xi_0, \xi_1, \xi_2, \ldots\}$

$$\sum_{k=0}^{\infty} (1 - |\xi_k|) = \infty. \tag{5}$$

Condition (5) is clearly much milder than equation (4). In Section 5, we derive several sufficient conditions for $\{\mathscr{B}_w; w \in W\}$ to be fundamental in ℓ_1.

The key tool in deriving these results is to reparameterise the linear space $\mathrm{sp}\{\mathscr{B}_w; w \in W\}$ as

$$X_n = \mathrm{sp}\left\{\frac{1}{1-\bar{\xi}_k z};\ k = 0, 1, \ldots, n-1\right\} = \mathrm{sp}\{\mathscr{B}_k(z);\ k = 0, 1, 2, \ldots, n-1\},$$

where the functions $\{\mathscr{B}_k(z)\}$, which have been considered in detail in Ninness and Gustafsson (1997) and Ninness and Hjalmarsson (1997), are defined by $\mathscr{B}_0(z) \triangleq \sqrt{1 - |\xi_0|^2}/(1 - \bar{\xi}_0 z)$ and

$$\mathscr{B}_k(z) \triangleq \frac{\sqrt{1-|\xi_k|^2}}{1-\bar{\xi}_k z} \prod_{m=0}^{k-1} \frac{z - \xi_m}{1 - \bar{\xi}_m z}, \quad k = 1, 2, \ldots \tag{6}$$

They are orthonormal in $H_2(\mathbf{D})$ with respect to the inner product

$$\langle \mathscr{B}_n, \mathscr{B}_m \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathscr{B}_n(e^{j\omega}) \overline{\mathscr{B}_m(e^{j\omega})} \, d\omega = \begin{cases} 1, & m = n, \\ 0, & m \neq n. \end{cases} \quad (7)$$

The functions in equation (6) are in fact obtained by applying the Gram–Schmidt procedure to the rational wavelet functions equation (3) with respect to the inner product equation (1). The afore-mentioned Laguerre and two-parameter Kautz bases are special cases of equation (6) where all the $\{\xi_k\}$ are chosen to be the same and real (Laguerre) or complex (Kautz). Formulations of orthonormal bases with general pole locations other than equation (6) are possible; see for example, Heuberger *et al.* (1995) and van den Hof *et al.* (1995) where a state-space approach is taken and Bokor *et al.* (1995a, b), Schipp and Bokor (1997), Schipp *et al.* (1996) and Szabo *et al.* (1997).

Since the linear spaces X_n spanned by the two sets are identical, the approximation properties of X_n with respect to $A(\mathbf{D})$, $H_p(\mathbf{D})$, ℓ_p are identical and any robust estimates obtained will be identical. However, by exploiting the orthonormality property (7) the provision of analytical expressions for approximation error is greatly facilitated. This resonates with the earlier work in a stochastic setting, where it has also been argued that the main utility of orthonormal model structures for system identification is not as an implementational tool (since simpler structures span the same space X_n and hence provide identical estimates), but as an analysis tool (Ninness and Gustafsson, 1997; Ninness and Hjalmarsson, 1997).

Having studied these basic approximation properties, robust estimation using the minimax scheme proposed by Mäkilä (1991), Mäkilä and Partington (1992) and Partington (1994a, 1996) is investigated. Conditions for robust convergence, and explicit quantification of estimation error are derived for each of the spaces $A(\mathbf{D})$, $H_p(\mathbf{D})$ and ℓ_1 and for both frequency-domain and time-domain measurements. As well, analysis of estimation using mixed parametric/non-parametric model structures is provided and implications for model reduction are discussed together with a brief study of how the results may be extended to the multi-variable setting. Finally, an example is given to illustrate the application of the minimax algorithm.

In the sequel, the notation $y_k = O(x_k)$ as $k \to \infty$ will mean y_k/x_k remains bounded. Also, the notation $\|\cdot\|_X$ will denote the norm on the space X, with the understanding that $\|\cdot\|_p$ means the usual $L_p(\mathbf{T})$ or ℓ_p norm as appropriate.

2. PROBLEM FORMULATION

This paper considers the problem of identifying an underlying linear-time-invariant, single-input/single-output, discrete-time system with impulse response $\{g(k)\}$. It is assumed that this system is ℓ_2 bounded-input/bounded-output stable and real so that the associated power-series representation $G(z) = \sum_{k=0}^{\infty} g(k)z^k \in H_\infty(\mathbf{D})$. In particular, if the system is ℓ_∞ bounded-input/bounded-output stable, then $g \in \ell_1$ and $G(z)$ is continuous on $\mathbf{T}$ so that in fact $G(z) \in A(\mathbf{D})$.

The identification of $G(z)$ is performed on the basis of the observed and possibly noise corrupted input–output behaviour of the system, which hereafter is referred to simply as $G(z)$. If the observed behaviour of $G(z)$ is in the frequency domain, then it is assumed that the measurement set-up is as follows:

$$E_k = G(e^{j\omega_k}) + \eta_k, \quad k = 0, \ldots, N \quad (8)$$

where E_k is the observed frequency response at the kth, not necessarily uniformly spaced frequency ω_k and η_k is a corruption to the true frequency response $G(e^{j\omega_k})$. This corruption $\eta = \{\eta_0, \eta_1, \ldots\}$ is assumed to be bounded as $\|\eta\|_\infty \leq \varepsilon$.

The robust identification objective is to produce, on the basis of the observed response $\{E_k\}$, an approximate model $\hat{G}_N \in A(\mathbf{D})$ for G in such a way that the following condition is satisfied:

$$\lim_{\substack{N \to \infty \\ \varepsilon \to 0}} \sup_{\|\eta\|_\infty \leq \varepsilon} \|\hat{G}_N - G\|_\infty = 0 \quad \text{for all } G \in A(\mathbf{D}). \quad (9)$$

In the time-domain problem formulation, the given input–output data $\{u(t), y(t)\}_{t=0}^{N-1}$ of the system is assumed to satisfy the measurement set-up

$$y(t) = (g \circledast u + \eta)(t) = \sum_{k=0}^{\infty} g(k)u(t-k) + \eta(t), \quad (10)$$

where the input signal $u(t)$ is bounded as $\|u\|_\infty \leq 1$ (with $u(t) = 0$ for $t < 0$) and $y(t)$ is the measured output corrupted by a bounded disturbance $\|\eta\|_\infty \leq \varepsilon$. In this case, the robust estimation objective is to again satisfy equation (8) or the following condition under the constraint that $\{\hat{g}_N(k)\} \in \ell_1$

$$\lim_{\substack{N \to \infty \\ \varepsilon \to 0}} \sup_{\|\eta\|_\infty \leq \varepsilon} \|\hat{g}_N - g\|_1 = 0 \quad \text{for all } g \in \ell_1. \quad (11)$$

An identification algorithm that satisfies either one of the above properties is called *convergent* and *robustly convergent* if it does not rely on *a priori* information about the unknown system and noise. We will call η noise although it may be present due to non-linearities, time variations etc.

3. IDENTIFICATION ALGORITHMS

The identification algorithms studied in this paper are results of the works of Partington (1994a,

1996) and Mäkilä (1991) (see also Mäkilä and Partington, 1992) who have derived a general framework to solve robust estimation problems of the form just posed. In their framework, given the linear model structure (1) and linear subspaces $X_k = \text{sp}\{\mathscr{B}_0, \dots, \mathscr{B}_{k-1}\}$, then for frequency domain measurements the robust estimate $G_N(z)$ is found as the solution of the minimax problem

$$\hat{G}_N(z) \triangleq \arg\min_{H \in X_n} \max_{0 \le k \le N} |H(e^{j\omega_k}) - E_k|. \quad (12)$$

A sufficient condition on the model structure X_n and the denseness of the frequency evaluation points $\{\omega_k\}$ such that equation (12) results in a robust estimator satisfying equation (9) is that there exists a fixed $0 < \delta < 1$ such that for each n

$$\max_{0 \le k \le N} |G(e^{j\omega_k})| \ge \delta \|G\|_\infty \quad \text{for all } G \in X_n. \quad (13)$$

In the case of time-domain data, the algorithm (12) takes the following (similar) form

$$\hat{g}_N \triangleq \arg\min_{g \in X_n} \max_{0 \le t \le N-1} |(g \circledast u)(t) - y(t)|, \quad (14)$$

where $\hat{g}_N$ denotes the impulse response of the identified model and a sufficient condition on the model structure X_n and the input u such that equation (14) results in a robust estimator satisfying equation (11) is that also there exists a fixed $0 < \delta < 1$ such that for each n

$$\max_{0 \le t \le N-1} |(g \circledast u)(t)| \ge \delta \|G\|_\infty \ (\text{or } \|g\|_1)$$
$$\text{for all } G \text{ (or } g) \in X_n. \quad (15)$$

Furthermore, for the algorithms defined by equations (12) and (13) provided that the conditions (13) and (15) hold, respectively, then it is possible to specify explicit bounds on the estimation error as (Partington, 1994a; Ward and Partington, 1996)

$$\|G - \hat{G}_N\|_\infty \le \left(\frac{2}{\delta} + 1\right) d(G, X_n; A(\mathbf{D})) + \frac{2}{\delta}\varepsilon, \quad (16)$$

$$\|g - \hat{g}_N\|_1 \le \left(\frac{2}{\delta} + 1\right) d(g, X_n; \ell_1) + \frac{2}{\delta}\varepsilon, \quad (17)$$

where $d(f, X_n; X)$ defined as

$$d(f, X_n; X) \triangleq \inf_{h \in X_n} \|h - f\|_X \quad (18)$$

represents the error in approximating f by some function from the model set X_n. Conditions (13) and (14) are also necessary for robust recovery of systems in the spaces $A(\mathbf{D})$ and ℓ_1 (Partington, 1996).

It should be noted that given the richness of the robust estimation literature, many other estimation approaches are possible other than equation (12) or (14). For example, concentrating on frequency-domain data, if the evaluation points $\{\omega_k\}$ are uniformly spaced, then a class of two-stage non-linear methods are available (Helmicki *et al.*, 1991; Gu and Khargonekar, 1992a, b) for which worst-case error bounds are comparable to Equation (16). However, these sub-optimality properties are crucially dependent on the uniform frequency spacing. If this uniformity requirement is dropped, the error due to undermodelling will decrease polynomially in model order n (Akçay *et al.*, 1994; Partington, 1993) even if G is extremely smooth, whereas using equation (12), the bound (16) shows that the undermodelling error will decrease according to $d(G, X_n, A(\mathbf{D}))$ which, as we shall show, decreases exponentially in n for exponentially stable discrete-time systems, and hence at a rate much faster than the two-stage schemes. (A different implementation of the two-stage algorithms in Akçay (1997) yields undermodelling errors decreasing faster than polynomially but strictly slower than exponentially in n for the same class of systems.)

Given these motivations, the formulation (12) or (14) reduces the worst-case identification problem to a choice of complete model sets X_n. Since the error depends upon $d(G, X_n; X)$, it is desirable to choose (via prior knowledge of $G(z)$) basis functions $\{\mathscr{B}_k(z)\}$ such that the distance $d(G, X_n; X)$ from $G(z)$ to $X_n = \text{sp}\{\mathscr{B}_0, \dots, \mathscr{B}_{n-1}\}$ is as small as possible.

For example, if frequency response data indicate several lightly damped modes, then to speed up the convergence rate of $d(G, X_n; \mathscr{A})$ to zero, some poles in the basis functions $\{\mathscr{B}_k(z)\}$ could be moved toward the boundary of $\mathbf{D}$ with approximately the same arguments as the resonant frequency of the modes.

However, once one has chosen a complete model set $\{X_n\}$ for X, it is necessary to check that it is compatible with the measurement set-up in that the sufficient conditions (13) or (15) for robust convergence are satisfied. Typically (for example, for frequency domain measurements) this will result in the conclusion that for a given number N of data, depending on the choice of model structure X_n, there is a maximum model order $n(N)$ such that equation (13) is satisfied. This is reminiscent of the well-known bias/variance constraint on model order that exists in a stochastic setting for system identification (Ljung, 1987).

The link between n and N for FIR models is provided by Bernstein's inequality (Zygmund, 1959) and for rational models, it can be derived by means of a sharp inequality due to Borwein and Erdelyi (1996). In the time domain, the dependence between n and N is referred to as the *sample complexity* (Poolla and Tikku, 1994). In Sections 6 and 7 we will address the question of how n and N are related for various scenarios, but to begin with we address the more rudimentary question of whether the model structures X_n are complete in the various spaces $A(\mathbf{D})$, $H_p(\mathbf{D})$ and ℓ_1 of interest.

4. FUNDAMENTAL MODEL SETS IN $A(\mathbf{D})$

In this section, we establish that the orthonormal basis (6) is a fundamental set for $A(\mathbf{D})$ if and only if it satisfies the condition

$$\sum_{n=0}^{\infty} (1 - |\xi_n|) = \infty. \tag{19}$$

It is known (Dewilde and Dym, 1981; Ninness and Gustafsson, 1997) that the basis (6) is fundamental in $H_2(\mathbf{D})$ under the same necessary and sufficient condition, and in this section we show that this may be extended to all $H_p(\mathbf{D})$ spaces for $1 \le p < \infty$. In deriving these results, a key tool is the following lemma which is presented in terms of finite Blaschke products defined by

$$\varphi_n(z) \triangleq \prod_{m=0}^{n-1} \frac{z - \xi_m}{1 - \overline{\xi}_m z}. \tag{20}$$

Lemma 1 (Christoffel–Darboux Identity). Let $\{\mathscr{B}_n\}_{n \ge 0}$ be as in equation (6). Then for all $z, \zeta \in D$

$$\sum_{k=0}^{n-1} \overline{\mathscr{B}_k(\zeta)}\mathscr{B}_k(z) = \frac{1 - \overline{\varphi_n(\zeta)}\varphi_n(z)}{1 - \overline{\zeta} z}. \tag{21}$$

Proof. The proof is by induction (see Ninness and Hjalmarsson, 1997). □

A key consequence of this result is that it facilitates a simple integral formulation of $d(G, X_n; A(\mathbf{D})$ as follows.

Lemma 2. Let $G \in A(\mathbf{D})$ and $\mathscr{B}_k(z)$ be as in equation (6). Let $\hat{G}_n$ be the projection

$$\hat{G}_n(\zeta) \triangleq \sum_{k=0}^{n-1} \langle G, \mathscr{B}_k \rangle \mathscr{B}_k(\zeta), \quad \zeta \in \mathbf{D}. \tag{22}$$

Then

$$G(\zeta) - \hat{G}_n(\zeta) = \frac{\varphi_n(\zeta)}{2\pi j} \oint_T \frac{G(z)}{z - \zeta} \overline{\varphi_n(z)} \, dz. \tag{23}$$

Proof. Since $G \in H_1(\mathbf{D})$, it can be represented by the Cauchy integral of its boundary function $G(e^{j\theta})$ [Rudin (1987), Theorem 17.11] as

$$G(\zeta) = \frac{1}{2\pi j} \oint_T \frac{G(z)}{z - \zeta} \, dz, \quad \zeta \in \mathbf{D}.$$

Then using Lemma 1

$$\begin{aligned}\hat{G}_n(\zeta) &= \sum_{k=0}^{n-1} \mathscr{B}_k(\zeta) \frac{1}{2\pi j} \oint_T G(z) \overline{\mathscr{B}_k(z)} \frac{dz}{z} \\ &= \frac{1}{2\pi j} \oint_T \frac{G(z)}{z - \zeta} [1 - \varphi_n(\zeta) \overline{\varphi_n(z)}] \, dz \\ &= G(\zeta) - \frac{\varphi_n(\zeta)}{2\pi j} \oint_T \frac{G(z)}{z - \zeta} \overline{\varphi_n(z)} \, dz.\end{aligned}$$ □

In order to use this result to provide $L_\infty(\mathbf{T})$ error bounds, it is necessary to derive an upper bound on $|\varphi_n(z)|$.

Lemma 3. Let φ_n be as in equation (20). Then for each $z \in \mathbf{D}$,

$$|\varphi_n(z)| \le \exp\left(-\frac{1}{2}(1 - |z|) \sum_{k=0}^{n-1} (1 - |\xi_k|) \right). \tag{24}$$

Proof. Let $z = re^{j\theta}$ and $\xi_n = Re^{j\omega_n}$ denote polar decompositions of z and ξ_n. Put $\psi = \theta - \omega_n$. Then a simple algebraic manipulation yields

$$\begin{aligned}\left| \frac{z - \xi_n}{1 - \overline{\xi}_n z} \right|^2 &= \left| \frac{re^{j\psi} - R}{1 - rRe^{j\psi}} \right|^2, \\ &= 1 - \frac{(1 - r^2)(1 - R^2)}{1 + r^2R^2 - 2rR\cos\psi}, \\ &\le 1 - (1 - r)(1 - R)\frac{(1 + r)(1 + R)}{(1 + rR)^2}, \\ &\le 1 - (1 - r)(1 - R), \\ &\le \exp(-(1 - r)(1 - R)),\end{aligned} \tag{25}$$

where the last inequality follows from the fact that $e^{-x} \ge 1 - x$ for all x. Consideration of equations (25) and (20) completes the proof. □

Use of these results allows the calculation of the $L_\infty(\mathbf{T})$-norm distance $d(G, X_n; A(\mathbf{D}))$ from the space $X_n = \mathrm{sp}\{\mathscr{B}_0, \ldots, \mathscr{B}_{n-1}\}$ to an arbitrary $G(z)$ with relative stability such that the power series $G(z)$ is convergent on the disc $\mathbf{D}_R$ for some $R > 1$ in which case we write $G \in A(\mathbf{D}_R, K)$ if $|G(z)|$ is bounded by $K < \infty$ in $\mathbf{D}_R$.

Lemma 4. Let $\{X_n\}_{n \ge 0}$ be the set spanned by equation (6). Let $G \in A(\mathbf{D}_R, K)$ and $d(G, X_n; X)$ be as in equation (16). Then

$$d(G, X_n; A(\mathbf{D})) \le \frac{KR}{R - 1} \exp\left(-\frac{R - 1}{2R} \sum_{k=0}^{n-1} (1 - |\xi_k|) \right). \tag{26}$$

Proof. Since $d(G, X_n; A(\mathbf{D})) \le \|G - H\|_\infty$ for all $H \in \mathrm{sp}\,\{\mathscr{B}_k\}_{k=0}^{n-1}$, we can use G_n in equation (20) to over-bound $d(G, X_n; A(\mathbf{D}))$. Lemma 2 provides the following expression for $G - \hat{G}_n$:

$$G(\zeta) - \hat{G}_n(\zeta) = \frac{\varphi_n(\zeta)}{2\pi j} \oint_T \frac{G(z)}{(z - \zeta)\varphi_n(z)} \, dz, \quad \zeta \in \mathbf{D}.$$

The integrand above is meromorphic in $\mathbf{D}_R$. Thus, the contour integrals on $\mathbf{T}$ and $(R - \varepsilon)\mathbf{T}$ are equal by the residue theorem (Rudin, 1987, Theorem 10.42] if $R - \varepsilon \ge 1$ since each pole is encircled once by $\mathbf{T}$ and $(R-\varepsilon)\mathbf{T}$. Therefore

$$|G(\zeta)-\hat{G}_n(\zeta)| = \left|\frac{\varphi_n(\zeta)}{2\pi j}\oint_{(R-\varepsilon)T}\frac{G(z)}{(z-\zeta)\varphi_n(z)}\,dz\right|,$$
$$\leq \frac{KR}{R-1-\varepsilon}\sup_{|z|=R-\varepsilon}|1/\varphi_n(z)|$$
$$= \frac{KR}{R-1-\varepsilon}\sup_{|z|=1/(R-\varepsilon)}|\varphi_n(z)|| \qquad (27)$$

Let $\varepsilon \to 0$. Then

$$\|G-\hat{G}_n\|_\infty \leq \frac{KR}{R-1}\sup_{|z|=R^{-1}}|\varphi_n(z)|.$$

Use of Lemma 3 now completes the proof. □

The remainder of this section is devoted to examining the question of whether or not $\{\mathscr{B}_k\}$ is fundamental in certain spaces. Key to this analysis is that via a well-known application of the Hahn–Banach theorem, it is possible to establish (see, for example, Achieser (1992), Section 30) that a set $A \subset X$ is fundamental in X if and only if any bounded linear functional vanishing on A also vanishes on X. This will be used firstly to examine the case of $X = A(\mathbf{D})$, and in this case it is also useful to know that a special case of the set $\{\mathscr{B}_k\}$ considered here, namely the Laguerre basis, is fundamental in $A(\mathbf{D})$. A proof of this fact is as follows.

Lemma 5. The Laguerre basis (2) (which is equation (6) with $\xi_k = a$, $\forall k$) is fundamental in $A(\mathbf{D})$ for all $-1 < a < 1$.

Proof. Let

$$w = \psi_a(z) = \frac{z-a}{1-az}.$$

Since ψ_a is a bilinear map, it suffices to show that the closed linear span of the set $\{\mathscr{B}_k \circ \psi_a^{-1}\}$ (here $\circ$ denotes composition of functions $(f\circ g)(z) = f(g(z))$ denoted by L equals to $A(\mathbf{D})$. A simple calculation yields

$$\mathscr{B}_k \circ \psi_a^{-1}(w) = \frac{1+aw}{\sqrt{1-a^2}}\,w^k, \quad k = 0, 1, \ldots. \qquad (28)$$

Thus $L \subset A(\mathbf{D})$. The reverse inclusion follows from the well-known fact that polynomials are dense in $A(\mathbf{D})$ (Rudin, 1987, Theorem 20.5) and any $F \in A(\mathbf{D})$ can be written as $F = (1+aw)H$ where $H(w) = F(w)/(1+aw)$ since $(1+aw)^{-1} \in A(\mathbf{D})$ whenever $|a| < 1$. □

Combining the previous lemmata with these observations, we have the main result of this section characterising the fundamentalness of the orthonormal bases (6) *in* $A(\mathbf{D})$ in terms of the chosen pole positions $\{1/\bar{\xi}_k\}$

Theorem 6. The orthonormal set in equation (6) is fundamental in $A(\mathbf{D})$ if and only if $\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$.

Proof. Suppose first that $\sum(1-|\xi_k|) = \infty$. Let B be the set of all Laguerre functions defined in equation (2) for the case of $a = 1/2$. Since it has just been established that the Laguerre basis is fundamental in $A(\mathbf{D})$, the set sp B is dense in $A(\mathbf{D})$. Notice (using the notation defined just before Lemma 4) that $B \subset A(\mathbf{D}_2, \sqrt{3})$. Therefore, the set sp $A(\mathbf{D}_2, \sqrt{3})$ is dense in $A(\mathbf{D})$. But by Lemma 4, it also holds that sp $A(\mathbf{D}_2, \sqrt{3}) \subset \overline{\mathrm{sp}\{\mathscr{B}_k\}_{k\geq 0}}$. Therefore $\{\mathscr{B}_k\}_{k\geq 0}$ is fundamental in $A(\mathbf{D})$.

Conversely assume that $\sum(1-|\xi_k|) < \infty$. Then the finite Blaschke products in equation (20) modulated by $\prod^{n-1} - \bar{\xi}_m/|\xi_m|$ converge uniformly on $\mathbf{D}$ to a function $\varphi(z) \in H_\infty$ which has zeros precisely at the points ξ_k (Rudin, 1987, Theorem 15.21). In this case, the linear functional F defined on $A(\mathbf{D})$ by

$$F(h) = \frac{1}{2\pi j}\oint_T h(z)\overline{\varphi(z)}\frac{dz}{z} \qquad (29)$$

is clearly nontrivial and also bounded. However, by Cauchy's Integral Theorem it also vanishes for any $\mathscr{B}_k$ of the form equation (6). Therefore, the linear span of the set $\{\mathscr{B}_k\}$ defined by equation (6) is not dense in $A(\mathbf{D})$. □

The lemmata may also be combined to characterise the fundamentalness of the orthonormal bases (6) in $H_p(\mathbf{D})$ for $1 \leq p < \infty$ in terms of the chosen pole positions $\{1/\bar{\xi}_k\}$.

Corollary 7. The orthonormal set in equation (6) is fundamental in $H_p(\mathbf{D})$ for all $1 \leq p < \infty$ if and only if $\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$.

Proof. If $\sum(1-|\xi_k|) < \infty$, then the non-trivial bounded linear functional F defined by equation (29) extends to a bounded linear functional on $H_p(\mathbf{D})$ for all $p \geq 1$ and hence equation (6) is not fundamental in $H_p(\mathbf{D})$. The sufficiency follows from the facts that the set of polynomials in z is dense in $H_p(\mathbf{D})$ for all $0 < p < \infty$ (Duren, 1970, Theorem 3.3) and $H_\infty(\mathbf{D})$ norm dominates all $H_p(\mathbf{D})$ norms for $p > 0$. □

The set $\{\mathscr{B}_k\}$ defined via equation (6) is a minimal spanning set in $H_2(\mathbf{D})$ since its elements are orthonormal and removal of any element from the set diminishes the span.

Theorem 6 and Corollary 7 can also be obtained directly from the results in (Achieser (1992, Section A.2). Nevertheless, our results are self-contained and further results in the subsequent sections will be based on the explicit error bounds that were derived in this section. Theorem 6 also has the following corollary.

Corollary 8. Consider the rational wavelet basis $\mathscr{B}_w$, $w \in W$ in equation (3) where W is an arbitrary subset of $\mathbf{D}$. Then $\{\mathscr{B}_w, w \in W\}$ is a fundamental set in $A(\mathbf{D})$ if and only if $\sum_{w \in W}(1-|w|) = \infty$.

Proof. Without loss of generality, assume that W is countable and let $\{\xi_n\}$ denote an enumeration of W. Now construct an orthonormal base from W as in equation (6). Then clearly $\mathrm{sp}\{\mathscr{B}_w\}_{w \in W} = \mathrm{sp}\{\mathscr{B}_n\}_{n \geq 0}$. Direct application of Theorem 6 then provides the result. □

The lattice W may be modified so as to contain an element w a finite or infinite number of times if each repeated w can be associated uniquely with a basis function in the form $(1-\bar{w}z)^{-k}$ for some integer $k > 1$. The base constructed in this manner does not contain polynomials in its linear span. This deficiency can be remedied by adjoining polynomials into the base. In contrast to the basis construction presented here, *hybrid* models, i.e. models containing both rationals and polynomials, are generated in Ward and Partington (1995) by the Hardy–Sobolev norm on *smooth* subsets of $A(\mathbf{D})$.

5. FUNDAMENTAL MODEL SETS IN ℓ_1

Having considered fundamental model sets applicable for robust estimation from frequency domain data, we now turn to the problem of estimation from time domain data. In this scenario, it is more natural to specify fundamental model sets according to the time domain properties of their elements, in which case the most common choice is to consider the space of systems whose impulse responses lie in ℓ_1; see for example Mäkilä *et al.* (1995) for a review, results and further references. One reason for the choice of space being ℓ_1 is the appeal of providing models suitable for subsequent ℓ_1 controller design (Tse *et al.*, 1993; Dahleh and Khammash, 1993).

In light of these motivations, we turn to the issue of formulating rational model sets that are fundamental in ℓ_1. In the sequel we will investigate their use in robust estimation from time-domain data. As in the previous section, a key tool will be to employ the orthonormal set $\{\mathscr{B}_k\}$ defined via equation (6).

Theorem 9. The orthonormal set in equation (6) is fundamental in ℓ_1 if

$$\lim_{m\to\infty} \exp\left(-\frac{1}{2}\sum_{k=0}^{m-1}(1-|\xi_k|)\right)\sum_{k=0}^{m-1}\frac{1}{1-|\xi_k|} = 0. \quad (30)$$

Proof. Let $g = \{g_k\} \in \ell_1$ and $\varepsilon > 0$. Truncate $\{g_k\}$ at the $k = n$th term to provide $g_n = \{g_0, \ldots, g_{n-1}\}$ where n is chosen such that $\|g - g_n\|_1 = \sum_{k=n}^{\infty}|g(k)| < \varepsilon$. Let $\hat{G}_m$ be the estimate of $G_1(z) = \sum_{k=0}^{n-1} g(k)z^k$ as in equation (22) and let ψ denote the impulse response of $\hat{G}_m$. By Hardy's inequality (Duren, 1970, Theorem 3.15), we bound $\|g_n - \psi\|_1$ as follows:

$$\|g_n - \psi\|_1 \leq |g_n(0) - \psi(0)| + \frac{1}{2\pi}\int_0^{2\pi} |G_1'(e^{j\theta}) - \hat{G}_m'(e^{j\theta})|\,d\theta. \quad (31)$$

Since $G_1 \in A(\mathbf{D}_R, \sup_{|z|\leq R}|G_1(z)|)$ for all $R > 1$, we have from the proof of Lemma 4,

$$|g_n(0) - \psi(0)| \leq \|G_1 - \hat{G}_m\|_\infty \leq \sup_{|z|\leq R}|G_1(z)|\frac{R}{R-1} \exp\left(-\frac{R-1}{2R}\sum_{k=0}^{n-1}(1-|\xi|_k)\right). \quad (32)$$

As in the proof of Lemma 4, we have the following expression for $G_1 - \hat{G}_m$ for all $R > 1$:

$$G_1(e^{j\theta}) - \hat{G}_m(e^{j\theta}) = \frac{\varphi_m(e^{j\theta})}{2\pi j}\oint_{RT}\frac{G_1(z)}{(z-e^{j\theta})\varphi_m(z)}dz. \quad (33)$$

Hence, the integrand in equation (31) is calculated from equation (33) as

$$\frac{d}{d\theta}[G_1(e^{j\theta}) - \hat{G}_m(e^{j\theta})] = e^{j\theta}\frac{\varphi_m'(e^{j\theta})}{2\pi}\oint_{RT}\frac{G_1(z)}{(z-e^{j\theta})\varphi_m(z)}dz + e^{j\theta}\frac{\varphi_m(e^{j\theta})}{2\pi}\oint_{RT}\frac{G_1(z)}{(z-e^{j\theta})^2\varphi_m(z)}dz. \quad (34)$$

We first bound $\varphi_m'(z) \triangleq d\varphi_m/dz$. Write $\varphi_m(z)$ as $\varphi_m(z) = z^\alpha\varphi_\beta(z)$ where α denotes the multiplicity of a zero at 0. Without loss of generality, assume $\xi_k = 0$ for $\beta \leq k < m$. Then

$$\varphi_m'(z) = \alpha z^{\alpha-1}\varphi_\beta(z) + z^\alpha\sum_{k=0}^{\beta-1}\frac{(1-|\xi_k|^2)\varphi_\beta(z)}{(1-\bar{\xi}_k z)(z-\xi_k)}.$$

Hence

$$\|\varphi_m'\|_\infty \leq \alpha + \sum_{k=0}^{\beta-1}\frac{1+|\xi_k|}{1-|\xi_k|} = \sum_{k=0}^{m-1}\frac{1+|\xi_k|}{1-|\xi_k|}. \quad (35)$$

We bound the first integral in equation (31) from Lemma 3 as

$$\left|\frac{1}{2\pi}\oint_{RT}\frac{G_1(z)}{(z-e^{j\theta})\varphi_m(z)}dz\right| \leq \frac{R}{R-1}\sup_{|z=R}|G_1(z)|\sup_{|z|R^{-1}}|\varphi_m(z)| \leq \frac{R}{R-1}\sup_{|z=R}|G_1(z)|\exp\left(-\frac{R-1}{2R}\sum_{k=0}^{m-1}(1-|\xi_k|)\right).$$

The second integral in equation (34) can be bounded in a similar fashion. Therefore

$$\|g-\psi\|_1 \le \varepsilon + \frac{R}{R-1}\sup_{|z|=R}|G_1(z)| \times \exp\left(-\frac{R-1}{2R}\sum_{k=0}^{m-1}(1-|\xi_k|)\right) \times \left[\sum_{k=0}^{m-1}\frac{1+|\xi_k|}{1-|\xi_k|}+\frac{R}{R-1}\right]. \quad (36)$$

The right-hand side of equation (36) tends to ε as $m \to \infty$ for fixed n and $R = 2$ under the hypothesis of the theorem. □

Corollary 10. The orthonormal set in equation (6) is fundamental in ℓ_1 if $k^{-\alpha} = O(1-|\xi_k|)$ for some $0 < \alpha < 1$.

Proof. For some constants C_1, C_2, the left-hand side of equation (30) is bounded as

$$\log \sum_{k=0}^{m-1}\frac{1}{1-|\xi_k|} \le \log \sum_{k=0}^{m-1} C_1 k^{\alpha} \le \log C_1 \frac{m(m-1)}{2} \le C_2 \log m.$$

The term on the right-hand side of equation (30) is calculated as

$$\sum_{k=0}^{m-1}(1-|\xi_k|) \ge \sum_{k=0}^{m-1} C_3 k^{-\alpha} = C_4 m^{1-\alpha}$$

for some constants C_3, C_4. Hence the result. □

Thus, a base $\{\mathscr{B}_k\}$ can be fundamental in ℓ_1 without requiring the set of points $\{\xi_k\}$ have an accumulation point in **D**. When the set $\{\xi_k\}$ has an accumulation point in **D**, the conclusion of Theorem 9 easily follows from (Ward and Partington, 1995, Theorem 2), since in this case the set of basis functions $\{\mathscr{B}_k\}$ will be a fundamental set in the (Hardy–Sobolev) $H^{2,1}$ norm, which dominates the ℓ_1 norm.

Corollary 10 provides a rather tight criterion. For example, if $1-|\xi_k| = O(1/k(\log k)^{\beta})$ for some $\beta > 1$, then from Theorem 6 the base $\{\mathscr{B}_k\}$ will not even be fundamental in $A(\mathbf{D})$.

The following result will be required in a later section.

Corollary 11. Let $\{X_n\}_{n\ge 0}$ be the model set spanned by the orthonormal set in equation (6), where the chosen poles lie in the complement of $\mathbf{D}_r$ for some fixed $r > 1$. Let g denote the impulse response of a transfer function $G \in A(\mathbf{D}_R, K)$ and $d(g, X_n; \ell_1)$ be as in equation (7). Then

$$d(g, X_n; \ell_1) \le \frac{KR}{R-1}\left(\frac{r+1}{r-1}n + \frac{R}{R-1}\right)e^{-(R-1)(r-1)n/2Rr}.$$

Proof. In the proof of Theorem 9, set $G_1 = G$ and replace $R\mathbf{T}$ by $(R-\delta)\mathbf{T}$ for sufficiently small $\delta > 0$ and r by r^{-1}. Then after these changes and letting $\delta \to 0$, we obtain

$$\|g-\psi\|_1 \le \frac{KR}{R-1}\left(\frac{r+1}{r-1}n + \frac{R}{R-1}\right)e^{-(R-1)(r-1)n/2Rr}, \quad (37)$$

where ψ is the impulse response of $\hat{G}_n$ in equation (22). The inequality above is an upper bound for $d(g, X_n; \ell_1)$. □

6. ROBUST IDENTIFICATION IN $A(\mathbf{D})$ FROM FREQUENCY RESPONSE MEASUREMENTS

In this section, we present a solution to the first problem formulated in Section 2; namely that of identification from frequency domain data. The model sets are assumed to be complete in $A(\mathbf{D})$ but arbitrary. The frequency response data is not required to be uniformly spaced. This problem for the equally spaced case has been well studied and a range of robustly convergent algorithms are available in the literature (Gu and Khargonekar, 1992a, b; Helmicki *et al.*, 1991; Partington, 1991). The non-uniformly spaced case is more difficult to handle and several robustly convergent (Akçay, 1997; Akçay *et al.*, 1994; Partington, 1993) and some tuned convergent algorithms (Chen *et al.*, 1995; Gu *et al.*, 1993) have been proposed in the literature. The common feature of the algorithms (Akçay, 1997; Akçay *et al.*, 1994; Chen *et al.*, 1995; Gu, 1994; Gu and Khargonekar, 1992a, b; Gu *et al.*, 1993; Helmicki *et al.*, 1991; Partington, 1991, 1993) is that the identified model is chosen from the set of finite-impulse response systems.

As we pointed out in Section 3, given fundamental model sets $\{X_n\}$ it is necessary to derive the relationship $n(N)$ (known as a sampling theorem for $A(\mathbf{D})$) such that the sufficient condition (12) holds for robust convergence of the scheme (11) to exist. In order to derive this, assume first (without loss of generality) that $\omega_0 = 0$ and $\omega_1 = \pi$. We define the maximum angular gap between the first $N+1$ points by

$$\delta_N \triangleq \max_{0\le k\le N}\ \min_{\substack{\ell \ne k \\ 0\le \ell \le N}} |\omega_k - \omega_\ell| \quad (38)$$

and via this we can derive the relationship $n(N)$ by the following lemma.

Lemma 12. Let $\{\mathscr{B}_k\}_{k=0}^{n-1}$ be a set of rational functions analytic in $\mathbf{D}_r$ for some $r > 1$ and let p denote the number of poles of $\{\mathscr{B}_k\}_{k=0}^{n-1}$ (including poles at ∞). Let δ_N be as in equation (38). Then equation (13) holds for some $0 < \delta < 1$ provided that

$$\delta_N \le 2\frac{(1-\delta)(r-1)}{p(r+1)}. \quad (39)$$

To prove Lemma 12, we need the following lemma which is a corollary of a result in Borwein and Erdelyi (1996).

Lemma 13. Suppose $a_1, \ldots, a_n$ are in the complement of $\mathbf{D}_r$ for some $r > 1$. Then

$$|g'(z)| \leq \frac{r+1}{r-1} m \|g\|_\infty, \quad z \in \mathbf{D}, \tag{40}$$

where

$$g(z) = \frac{\prod_{k=1}^{m}(z - b_k)}{\prod_{k=1}^{n}(z - a_k)}$$

and $m > n$.

Proof. Split g as

$$g(z) = \sum_{k=0}^{m-n} c_k z^k + \frac{P(z)}{\prod_{k=1}^{n}(z - a_k)} = g_1(z) + g_2(z),$$

where $P(z)$ is a polynomial of degree at most $n - 1$. Let

$$h_1(z) = g_1(z) - (1 - \alpha z)^{n-m} g_1(z),$$

where $\alpha \in \mathbf{D}$ and $h_2(z) = g(z) - h_1(z)$. Then h_2 is a proper rational function with m poles and is analytic in $\mathbf{D}_r$ provided $|\alpha| < 1/r$. The derivative of h_2 is bounded from Borwein and Erdelyi (1996) as

$$|h_2'(z)| \leq \frac{r+1}{r-1} m \|h_2\|_\infty, \quad z \in \mathbf{D}. \tag{41}$$

The derivative $h_1'(z)$ is calculated as

$$h_1'(z) = [1 - (1 - \alpha z)^{n-m}] g_1'(z) + \alpha(n - m)(1 - \alpha z)^{n-m-1} g_1(z).$$

Therefore, as α tends to zero, h_1 and h_1' tend to zero uniformly on $\mathbf{D} \cup \mathbf{T}$. Letting α tend to zero also in equation (41) then provides equation (40). □

Proof of Lemma 12. Suppose $\|g\|_\infty$ is attained at the point $e^{j\alpha}$. Now $|\alpha - \omega_k| \leq \delta_N/2$ for some k and since

$$g(e^{j\omega_k}) = \int_\alpha^{\omega_k} g'(e^{j\theta}) j e^{j\theta} \, d\theta + \|g\|_\infty,$$

we have

$$|g(e^{j\omega_k})| \geq \|g\|_\infty - \int_\alpha^{\omega_k} |g'(e^{j\theta})| \, d\theta \geq \|g\|_\infty - \|g'\|_\infty \delta_N/2. \tag{42}$$

Thus, from equations (42) and (40), we have $|g(e^{j\omega_k})| \geq \delta \|g\|_\infty$ provided that equation (39) is satisfied. □

Corollary 14. Consider the orthonormal set in equation (6). Let $r_n = \max_{k<n} |\xi_k|$. Then equation (13) holds for some $0 < \delta < 1$ provided that

$$\delta_N \leq 2 \frac{1-\delta}{n} \left(\frac{1 - r_n}{1 + r_n} \right). \tag{43}$$

Thus, if frequencies are uniformly spaced and the chosen poles lie in the complement of $\mathbf{D}_r$ for some fixed $r > 1$, condition (13) is satisfied provided $N \geq (r+1)\pi n / 2(1 - \delta)(r - 1)$. This condition is weaker than the requirement for the rational wavelets developed in Ward and Partington (1996). In Theorem 5 and Corollary 6 of Ward and Partington (1996), n and N satisfy the relations $n = 2^{p+1} - 1$ and $N \geq \pi 2^{2(p+1)}/(1 - \delta)$, where p is the lattice parameter in equation (4). Thus, the sampling theorem for the radial basis functions and uniformly spaced data can be expressed as $N \geq \pi(n+1)^2/(1 - \delta)$. Furthermore, n model poles lie in the complement of the open disk $\mathbf{D}_{n+1/(n-1)}$ since p in equation (4) is related to r by the expression $1/r = 1 - 2^{-p}$.

Our results in this section are summarised in the following theorem.

Theorem 15. Consider the orthonormal set in equation (6). Suppose $\{e^{j\omega_k}\}_{k \geq 0}$ is dense in $\mathbf{T}$. Let δ_N be as in equation (38).Then the algorithm given in equation (12) is robustly convergent if $\sum_{n=0}^{\infty}(1 - |\xi_n|) = \infty$ and δ_N satisfies equation (43) with $r_n = \max_{k<n} |\xi_k|$. In particular, for each fixed $r > 1$, an orthonormal set of rational functions with poles in the complement of $\mathbf{D}_r$ can be chosen such that the algorithm given in equation (12) is robustly convergent if

$$\delta_N \leq 2 \frac{1-\delta}{n} \left(\frac{r-1}{r+1} \right) \tag{44}$$

or when the frequencies are uniformly spaced

$$N \geq \frac{\pi n}{2(1-\delta)} \left(\frac{r+1}{r-1} \right). \tag{45}$$

The inequality (16) provides a simple upper bound for the identification error. However, the distance $d(G, X_n; A(\mathbf{D}))$ is difficult to evaluate in most cases because it depends not only on the system but also the chosen model sets. In this section, we will simplify the analysis and assume that $G \in A(\mathbf{D_R}, K)$ where the meaning of the latter notation was defined just before Lemma 4. Even with this simplification, it is still hard to calculate exact values of $d(G, X_n; A(\mathbf{D}))$ for arbitrary model sets. However, Lemma 4 provides a useful bound on $d(G, X_n; A(\mathbf{D}))$ which when combined with equation (16) provides the following result.

Theorem 16. Consider the orthonormal set in equation (6). Let $r_n = \max_{k<n} |\xi_k|$. Suppose $\{e^{j\omega_k}\}_{k \geq 0}$ is dense in $\mathbf{T}$. Let δ_N be as in equation (38). For each N, choose an n such that equation (43) is satisfied. Let $\hat{G}_n$ be the estimate of $g \in A(\mathbf{D_R}, K)$, $R > 1$ by the algorithm given in equation (12).

Then

$$\|G - \hat{G}_n\|_\infty \le \left(\frac{2}{\delta} + 1\right)\frac{KR}{R-1} \exp\left(-\frac{R-1}{2R}\sum_{k=0}^{n-1}(1-|\xi_k|)\right) + \frac{2}{\delta}\varepsilon. \tag{46}$$

In particular for each fixed $r > 1$, an orthonormal set of rational functions with poles in the complement of $\mathbf{D}_r$ can be chosen such that if δ_N satisfies Equation (44) or N satisfies equation (44) when the frequencies are uniformly spaced, then

$$\|G - \hat{G}_n\|_\infty \le \left(\frac{2}{\delta} + 1\right)\frac{KR}{R-1} e^{-(R-1)(r-1)n/2Rr} + \frac{2}{\delta}\varepsilon. \tag{47}$$

Theorem 16 extends the Laguerre and Kautz results in Ward and Partington (1996) to arbitrary orthonormal bases. Notice that the upper bound in equation (47) is minimized for $r = \infty$. This conforms with the *n-width* result (Pinkus, 1985) that polynomial models are optimal for the class $A(\mathbf{D}_R, K)$.

7. ROBUST IDENTIFICATION IN $A(\mathbf{D})$ FROM TIME DOMAIN MEASUREMENTS

In this section, we present solutions to the second problem formulated in Section 2. Condition (15) places severe restrictions on the choice of inputs. Following the terminology introduced in Harrison *et al.* (1997), we will call an input signal u that satisfies equation (15) a δ-cover of X_n. The length of the shortest δ-cover of X_n is said to be the *sampling size* for u. The sampling sizes and δ-covers are known for polynomials and certain compact subsets of $A(\mathbf{D})$ and ℓ_1.

In Dehleh *et al.* (1993), Harrison and Ward (1996), Poolla and Tikku (1994), the sampling size for the set of nth-order polynomials denoted by $\mathscr{P}_n$ in the ℓ_1-norm was shown to be $O(\beta^n)$ for some $\beta \in (1, 2]$. On the other hand, the sampling size for the same set of polynomials in the $H_\infty(\mathbf{D})$ -norm is $O(n^2)$ (Harrison *et al.*, 1996). In Harrison *et al.* (1996), Kacewicz and Milanese (1995) and Mäkilä (1991), examples of the δ-covers of polynomial models for the ℓ_1 and $H_\infty(\mathbf{D})$ norms are presented. The δ-covers for polynomial models can be used in the construction of δ-covers for compact rational model sets (with the same norm). An example is the set of nth order, strictly proper transfer functions which are analytic on $\mathbf{D}_r$ for some fixed $r > 1$, denoted by $\mathscr{V}(n, r^{-1})$. In Harrison *et al.* (1997), it is shown that each $0.2 + 0.8\delta$-cover of $\mathscr{P}_m$ is also a δ-cover of $\mathscr{V}(n, r^{-1})$, where m can be chosen to be any integer satisfying

$$m \ge \frac{4nr}{r-1}\ln\left(\frac{20r}{(1-\delta)(r-1)}\right). \tag{48}$$

Example 17. Let $G \in \mathscr{V}(n,0.9)$. Set $\delta = 0.5$. Then equation (48) reads $m \ge 240n$. Thus each 0.6-cover of $\mathscr{P}_{240 \cdot n}$ is a 0.5-cover of $\mathscr{V}(n,0.9)$. Suppose $X = \ell_1$ and let u be a sequence that contains only all possible m-tuples of ± 1. Set $m = 240n$. Thus $N = 2^m + m - 1$ and u is a 1-cover (and hence 0.6-cover) of $\mathscr{P}_m$ (Dahleh *et al.*, 1993; Kacewicz and Milanese, 1995; Mäkilä, 1991; Poolla and Tikku, 1994). Then u yields $\|g \circledast u\|_\infty \ge 0.5\|g\|_1$ for all $G \in \mathscr{V}(n, 0.9)$, where g denotes the impulse response of G.

The following lemma is immediate.

Lemma 18. Let $\{X_n\}_{n \ge 0}$ be the model set spanned by the orthonormal set in equation (6), where the chosen poles lie in the complement of $\mathbf{D}_r$ for some fixed $r > 1$. For each n choose an integer m satisfying equation (48). Let u be the $0.2 + 0.8\delta$-cover of $\mathscr{P}_m$ in X, where X denotes either $A(\mathbf{D})$ or ℓ_1, and let N be the length of u. Then

$$\max_{0 \le t \le N-1} |(g \circledast u)(t)| \ge \delta \|g\|_X \quad \text{for all } g \in X_n. \tag{49}$$

Use of Lemma 18 together with Theorem 9 provides the following robust convergence result for $A(\mathbf{D})$ and ℓ_1.

Theorem 19. Consider the orthonormal set in Equation (6), where the chosen poles lie in the complement of $\mathbf{D}_r$ for some fixed $r > 1$. Let X denote either ℓ_1 or $A(\mathbf{D})$ and let the inputs for the system in equation (10) be chosen as in Lemma 18. Then the algorithm given in equation (14) robustly converges in X.

As in Section 6, from equations (16), (17), Lemma 4, Corollary 11, and Lemma 18, we obtain the following worst-case identification error bounds in the $H_\infty(\mathbf{D})$ and ℓ_1 norms.

Theorem 20. Consider the orthonormal set in equation (6), where the chosen poles lie in the complement of $\mathbf{D}_r$ for some fixed $r > 1$. Let the inputs for the system in equation (10) be chosen as in Lemma 18. Let $\hat{G}_n$ be the estimate of $G \in A(\mathbf{D}_R, K)$, by the algorithm given in equation (12). Then

$$\|G - \hat{G}_n\|_\infty \le \left(\frac{2}{\delta} + 1\right)\frac{KR}{R-1} e^{-(R-1)(r-1)n/2Rr} + \frac{2}{\delta}\varepsilon. \tag{50}$$

Theorem 21. Consider the orthonormal set in equation (6), where the chosen poles lie in the complement of $\mathbf{D}_r$ for some fixed $r > 1$. Let the inputs for the system in equation (10) be chosen as in Lemma 18. Let $\hat{g}_n$ be the estimate of g, the impulse response of $G \in A(\mathbf{D}_R, K)$, by the algorithm given in equation (14). Then

$$\|g - \hat{g}_n\|_1 \le \left(\frac{2}{\delta} + 1\right)\frac{KR}{R-1}\left(\frac{r+1}{r-1}n + \frac{R}{R-1}\right) \times e^{-(R-1)(r-1)n/2Rr} + \frac{2}{\delta}\varepsilon. \qquad (51)$$

When $X = H_2(\mathbf{D})$ or $X = \ell_2$, a necessary and sufficient condition for the existence of robustly convergent algorithms in X is that there exists a fixed $0 < \delta < 1$ such that for each n

$$\max_{0 \le t \le N-1} |(g \circledast u)(t)| \ge \delta \|g\|_2 = \delta \|G\|_2 \quad \text{for all } g \text{ (or } G) \in X_n. \qquad (52)$$

Furthermore, for the algorithms defined by equations (12) and (14) provided that equation (52) holds, an explicit bound on the ℓ_2 norm of the estimation error is obtained from Partington (1994a, 1996)

$$\|g - \hat{g}_N\|_2 \le \left(\frac{2}{\delta} + 1\right) d(g, X_n, \ell_2) + \frac{2}{\delta}\varepsilon. \qquad (53)$$

Condition (52) places mild restrictions on the choice of input signal despite the fact that it appears to be stronger than the usual persistence of excitation condition in Ljung (1987), Söderström and Stoica (1989). In particular, the sampling size of δ-covers for polynomial models is $O(n)$ (Partington, 1994b). As well, the least-squares algorithm has robust convergence property in ℓ_2 (or $H_2(\mathbf{D})$) identification. Moreover, input signal can be chosen such that to identify a system $G \in \mathscr{P}_n$ with an error of $O(\varepsilon)$ one requires only $O(n)$ measurements (Partington, 1994a). Extension of this result to power-bounded noise case is discussed in Partington and Mäkilä (1995).

8. MIXED PARAMETRIC/NON-PARAMETRIC MODELS

The upper bounds on the worst-case identification errors in the $H_\infty(\mathbf{D})$ and ℓ_1 norms given in Theorems 16, 20, and 21 are minimised if polynomial models are used when the unknown system is in the class $A(\mathbf{D}_R, K)$. Indeed, the linear span of polynomials form optimal model sets in the Kolmogorov's n-width sense (see Pinkus (1985) for a comprehensive treatment of n-widths). However, in practice, it is more common that prior knowledge about $G(z)$ is significantly richer than the simple statement $G \in A(\mathbf{D}_R, K)$. In this section, we show that large improvements can be obtained over the non-parametric approach of using polynomial models by instead employing more general classes of models constituted of mixed parametric and non-parametric models.

To illustrate this, consider the example of the following perturbation model:

$$G(z) = \sum_{k=0}^{m-1} \frac{\alpha_k}{1 - \bar{\beta}_k z} + h(z) = H(z) + h(z), \quad (54)$$

where $\beta_k \ne 0$ for all k and $h \in A(\mathbf{D})$ is arbitrary but its norm satisfies $\|h\|_\infty \le C_h$. Let $r = \max_k |\beta_k|$ and $\sum_{k=0}^{m-1} |\alpha_k| = 1 - r$ for normalisation. We assume that the points β_k lie in the circles

$$D(\gamma_k) = \{z \in \mathbf{C}: |z - \gamma_k| < (1 - |\gamma_k|)\mu\}, \quad k = 0, \ldots, m-1$$

for some $\mu < 1$ so that the uncertainty radius is modulated according to the pole position such as to preclude crossing of the stability boundary. This choice of uncertainty structure is predicated on the assumption that in practice, while one may be unsure of system time constants, one is normally confident of whether or not the system is stable. With this choice of uncertainty structure, observe that $D(\gamma_k) \subset \mathbf{D}$ for all $\gamma_k \in \mathbf{D}$. Although this section will concentrate on the model structure (44), note that other examples of mixed parametric and non-parametric models have also appeared in the literature (Elia and Milanese, 1993; Giarre *et al.*, 1997; Kosut *et al.*, 1992).

We will calculate an upper bound on $d(G, X_n; A(\mathbf{D}))$ for a suitably chosen base using the prior information on β_k. We pick the orthonormal set in equation (6) with

$$\xi_{k+(m+1)p} = \begin{cases} \gamma_k, & 0 \le k < m;\ 0 \le p < M, \\ 0, & k = m;\ 0 \le p < M. \end{cases} \quad (55)$$

Set $n = mM + m$ and let $\hat{H}_n$ denote the least-squares estimate of H as in equation (22). Then from Lemma 2, we obtain by an application of Cauchy formula

$$\begin{aligned} H(\zeta) - \hat{H}_n(\zeta) &= \frac{\varphi_n(\zeta)}{2\pi j}\sum_{k=0}^{m-1}\alpha_k \oint_T \frac{1}{(1 - \bar{\beta}_k z)(z - \zeta)\varphi_n(z)}\,dz \\ &= -\frac{\varphi_n(\zeta)}{2\pi j}\sum_{k=0}^{m-1}\alpha_k \oint_{-T} \frac{\overline{\varphi_n(\bar{z})}}{(z - \bar{\beta}_k)(1 - \zeta z)}\,dz \\ &= \varphi_n(\zeta)\sum_{k=0}^{m-1} \frac{\overline{\varphi_n(\beta_k)}}{(1 - \zeta\bar{\beta}_k)} \end{aligned} \quad (56)$$

for all $\zeta \in \mathbf{D}$. But from our uncertainty description, we have for all $k \neq m$

$$|\varphi_n(\beta_k)| = \prod_{l=0}^{n-1} \left| \frac{\beta_k - \xi_l}{1 - \bar{\xi}_l \beta_k} \right|$$
$$= |\beta_k|^M \prod_{p=0}^{M-1} \left| \frac{\beta_k - \xi_{k+(m+1)p}}{1 - \bar{\xi}_{k+(m+1)p}\beta_k} \right|$$
$$\times \prod_{\substack{l=0 \\ l \neq k}}^{m-1} \prod_{p=0}^{M-1} \left| \frac{\beta_k - \xi_{l+(m+1)p}}{1 - \bar{\xi}_{l+(m+1)p}\beta_l} \right|$$
$$< (r\mu)^M.$$

Thus $\|H - \hat{H}_n\|_\infty \leq (r\mu)^M$. Hence for this choice of basis functions, it follows that

$$d(H, X_n; A(\mathbf{D})) \leq (r\mu)^M.$$

This result applies for any rational function of the form $H + P_s$ where $P_s \in \mathscr{P}_s$ and $M \geq s$ since

$$d(P_s, X_n; A(\mathbf{D})) = 0 \quad \text{for all } M \geq s.$$

Thus, the model sets $\{X_n\}_{n \geq 0}$ spanned by the orthonormal set (6) whose parameters are chosen as in equation (55) satisfy

$$d(G, X_n; A(\mathbf{D})) \leq (r\mu)^M + d(h, \mathscr{P}_{M-1}; A(\mathbf{D})).$$

Let $\hat{G}_N$ be the estimate of G by the algorithm given in equation (12) where the basis functions are chosen as above, the frequencies are dense in $\mathbf{T}$, and the maximum angular gaps satisfy

$$\delta_N \leq 2\frac{(1-\delta)(1-r)}{n(1+r)}. \tag{57}$$

Then from equation (16), we have

$$\|G - \hat{G}_N\|_\infty \leq \left(\frac{2}{\delta} + 1\right)(r\mu)^M + \left(\frac{2}{\delta} + 1\right) d(h, \mathscr{P}_{M-1}; A(\mathbf{D}) + \frac{2}{\delta}\,\varepsilon. \tag{58}$$

The first term tends to zero geometrically and the second term, which asymptotically tends to zero, is bounded by $(2/\delta + 1)C_h$. This inequality shows that by using mixed parametric–nonparametric models, one can obtain large improvements in estimation error. In particular, the number of measurements needed to estimate a transfer function, within a given level of accuracy, can be reduced dramatically in comparison to impulse response models. If only one model pole is chosen in $D(\gamma_{k_0})$, then the first term above must be replaced by $r\mu O(|\beta_{k_0}|^M)$. Thus in order to improve approximation error converge rate, one has to choose multiple poles around the dominant poles of the system. Indeed, the orthonormal base with infinitely repeated poles at $1/\bar{\gamma}_k$, $k = 0, \ldots, m-1$ has been shown to be optimal in the Kolmogorov's n-width sense (Oliveira & Silva, 1996).

In this paper, the distance $d(G, X_n; X)$ has been estimated by means of the orthogonal projection (20) which has proven to be an effective tool in the computation of approximation errors. In spite of this, the projection (20) has not been used for robust estimation. This is due to a result presented by Partington (1992) which states that if $\hat{G}_N$ is the least-squares estimate of G, then the $L_\infty(\mathbf{T})$ norm of the worst-case identification error diverges as $O(\log N)\varepsilon$. It is also interesting to observe that the least-squares estimate diverges as the number of data tends to infinity even if $\varepsilon = 0$ provided that $h \in A(\mathbf{D})$ is arbitrary subject to the constraint $\|h\|_\infty \leq C_h$ (Somorjai, 1980). The two-stage algorithms in Akçay *et al.* (1994), Gu and Khargenekar (1992a, b) can be used in the estimation of equation (54) but under more restrictive sampling conditions than equation (57) (Akçay, 1997).

8.1. *Model reduction*

When both r and μ are close to one, M and consequently N by equation (57) must be chosen large enough to keep the $O(r\mu^M)$ term in equation (58) within acceptable limits. Then a model reduction procedure is necessary to extract a nominal model from the identified model $\hat{G}_N$. The optimal Hankel norm model reduction and balanced truncations are the most frequently used techniques. In practice, both methods often work well. They both require a balanced state-space realisation of $\hat{G}_N$ or $\hat{g}_N$. The subspace-based system identification algorithms in McKelvey *et al.* (1996) and Kung (1978) can be used to effectively compute state-space parameters for a particular realisation.

Assume $h(z)$ in equation (54) is constant so that model sets contain only proper rational functions. (If $h(z)$ is not constant and modeled by polynomials, then the polynomial and the rational parts of $\hat{G}_N$ can be reduced separately). The input to the algorithm in McKelvey *et al.* (1996) are the computed frequency response of $\hat{G}_N$ at an arbitrary set of equally spaced frequencies. This method exactly retrieves an nth-order transfer function when the frequency response measurements are noise-free and the number of measurements is at least $n + 2$. Moreover, the returned state-space realisation is close to being in balanced form McKelvey *et al.* (1996). However, in order to obtain accurate results, the number of computed frequency response samples must be kept rather large in comparison to the identified model order.

This technique was used in the identification of a power transformer, where it was not possible to obtain a balanced realisation directly from the identified model for a subsequent model reduction

(Akçay *et al.*, 1997).. This problem is related to the fact that if the system order is high, poles and zeros of the system are sensitive to polynomial factoring. The algorithm in Kung (1978) is even simpler. An nth-order system can exactly be retrieved from its first $2n + 1$ noise-free impulse response coefficients. Then again a balancing transformation on the computed state-space parameters is performed.

8.2. *Choice of pole locations*

The choice of pole positions plays an important role in the quality of the approximation of a given system by a truncated series. Although optimal solution in the Kolmogorov's n-width sense is conceptually simple:

$$X_n \triangleq \arg\inf_{X_n} \sup_{G \in S} \inf_{H \in X_n} \|G - H\|_X,$$

where $S \subset X$ captures prior information on G, except few isolated cases it is hardly computable. Several methods to determine optimal pole locations for simple uncertainty descriptions and model structures have been proposed in the literature. The discussion of these methods is beyond the scope of the current paper. We refer the interested reader to Bodin and Wahlberg (1994), den Brinker (1996), Oliveira e Silva (1995), Fu and Dumont (1993), and Zimmermann and Williamson (1994) and the references therein.

9. MULTI-INPUT/MULTI-OUTPUT SYSTEMS

There is no difficulty extending the results of this paper to multi-variable systems. We show this for the frequency-domain formulation. The time-domain extension is similar.

Let $G(z)$ be $p \times q$ matrix-valued transfer function of the unknown system with entries in $A(\mathbf{D})$. Let $\|G\|_\infty$ denote the $H_\infty(\mathbf{D})^{p\times q}$ norm of G defined by $\sup_\omega \sigma_1(G(e^{j\omega}))$ where σ_1 denotes the largest singular value. Assume that noise in equation (8) is bounded as $\sigma_1(\eta_k) \le \varepsilon$ for all k. Given a sequence of scalar-valued functions $\{\mathscr{B}_k\}_{k\ge 0}$ in $A(\mathbf{D})$, model sets are defined by

$$X_n \triangleq \left\{\psi \in A(\mathbf{D})^{p\times q}:\ \psi = \sum_{k=0}^{n-1} \lambda_k \mathscr{B}_k;\ \lambda_k \in \mathbf{R}^{p\times q},\right.$$
$$\left. k = 0, \ldots, n-1\right\}.$$

Since G is real, the linear span of basis functions must be defined with respect to the real field when the basis functions are real. We choose basis poles in complex conjugate pairs so that each basis function is real—see Ninness and Gustafsson (1997) for more detail on this point. Obviously, $\{X_n\}_{n\ge 0}$ is a complete model set for $A(\mathbf{D})^{p\times q}$ if and only if $\{\mathscr{B}_k\}_{k\ge 0}$ is fundamental in $A(\mathbf{D})$. It only remains to derive an upper bound on the worst-case identification error of the minimax algorithm introduced next. For this purpose, we define the identified model $\hat{G}_N \in X_n$ to be a solution of the minimax problem

$$\hat{G}_N \triangleq \arg\min_{H\in X_n} \max_{\substack{0\le k\le n \\ 1\le l\le p;\ 1\le m\le q}} |H^{lm}(e^{j\omega_k}) - E_k^{lm}|. \tag{59}$$

The proof of the following result is adapted from Partington (1994a).

Theorem 22. Consider the orthonormal set equation (6). Let $r_n = \max_{k<n}|\xi_k|$. Suppose $\{e^{j\omega_k}\}_{k\ge 0}$ is dense in $\mathbf{T}$. Let δ_N be as in equation (38). For each N, choose an n such that equation (43) is satisfied for some $0 < \delta < 1$. Let $\hat{G}_N$ be the estimate of $G \in A(\mathbf{D})^{p\times q}$ given in equation (59). Then

$$\|G - \hat{G}_N\|_\infty \le \left(\frac{2}{\tilde{\delta}} + 1\right) d(G, X_n; A(\mathbf{D})^{p\times q}) + \frac{2}{\tilde{\delta}}\varepsilon \tag{60}$$

where $\tilde{\delta} = \delta/\sqrt{pq}$.

Proof. If δ_N is chosen as in Equation (43), then from Lemma 12 we have that for all l, m and $H^{lm} \in \mathrm{sp}\{\mathscr{B}_k\}_{k=0}^{n-1}$

$$\max_{0\le k\le N} |H^{lm}(e^{j\omega_k})| \ge \delta\, \|H^{lm}\|_\infty.$$

Since for all ω

$$\sigma_1(H(e^{j\omega}) \le \sqrt{pq} \max_{1\le l\le p;\ 1\le m\le q} |H^{lm}(e^{j\omega})|,$$

it follows that whenever $H^{lm} \in \mathrm{sp}\{\mathscr{B}_k\}_{k=0}^{n-1}$

$$\max_{1\le l\le p;\ 1\le m\le q} \max_{0\le k\le N} |H^{lm}(e^{j\omega_k})| \ge \frac{\delta}{\sqrt{pq}} \|H\|_\infty. \tag{61}$$

Let Ψ be the closest element of X_n to G in $A(\mathbf{D})^{p\times q}$. Since $\hat{G}_N \in X_n$, from equation (61) we obtain for some s, t and i

$$|\Psi^{st}(e^{j\omega_i}) - \hat{G}_N^{st}(e^{j\omega_i})| \ge \tilde{\delta}\, \|\Psi - \hat{G}_N\|_\infty. \tag{62}$$

The left-hand side of equation (62) is bounded as

$$\begin{aligned} &|\Psi^{st}(e^{j\omega_i}) - \hat{G}_N^{st}(e^{j\omega_i})| \\ &\le |\Psi^{st}(e^{j\omega_i}) - E_i^{st}| + |E_i^{st} - \hat{G}_N^{st}(e^{j\omega_i})| \\ &\le \|\Psi - G\|_\infty + \varepsilon + |E_i^{st} - \Psi^{st}(e^{j\omega_i})| \\ &\le 2\|\Psi - G\|_\infty + 2\varepsilon, \end{aligned} \tag{63}$$

where the second inequality is due to the fact that Ψ is a candidate minimiser of equation (59). Last, we have the following triangle inequality:

$$\|G - \hat{G}_N\|_\infty \le \|G - \Psi\|_\infty + \|\Psi - \hat{G}_N\|_\infty. \tag{64}$$

Thus from equations (62)–(64), we obtain equation (60). □

The minimax problem in equation (59) is complex and hence difficult to implement. However, it can be re-cast as a real-parameter minimax problem at the expense of slightly increased error bounds as follows. To simplify the notation, we assume G is single-input/single-output and define the matrices

$$\mathscr{B} \triangleq \begin{bmatrix} \mathscr{B}_0(e^{j\omega_0}) & \cdots & \mathscr{B}_{n-1}(e^{j\omega_0}) \\ \vdots & \ddots & \vdots \\ \mathscr{B}_0(e^{j\omega_N}) & \cdots & \mathscr{B}_{n-1}(e^{j\omega_N}) \end{bmatrix}, \quad (65)$$

$$E \triangleq \begin{bmatrix} E_0 & \cdots & E_N \end{bmatrix}^{\mathrm{T}}, \quad (66)$$

$$\Lambda \triangleq \begin{bmatrix} \lambda_0 & \cdots & \lambda_{n-1} \end{bmatrix}^{\mathrm{T}}, \quad (67)$$

and let Φ_R and Φ_I denote, respectively, the real and imaginary parts of Φ. Let E_R and E_I be the real and imaginary parts of E. Define the identified model as $\tilde{G}_N = \sum_{k=0}^{n-1} \tilde{\lambda}_k \mathscr{B}_k$ where $\tilde{\Lambda} = (\tilde{\lambda}_1 \cdots \tilde{\lambda}_{n-1})^{\mathrm{T}}$ is a solution of the minimax problem

$$\tilde{\Lambda} \triangleq \arg\min_{\Lambda \in \mathbf{R}^n} \left\| \begin{bmatrix} \Phi_R \\ \Phi_I \end{bmatrix} \Lambda - \begin{bmatrix} E_R \\ E_I \end{bmatrix} \right\|_\infty. \quad (68)$$

The multi-variable form of equation (68) is obtained by concatenation. This minimax problem is a linear programming problem involving real matrices and vectors and can be solved efficiently by the algorithm of Barrodale and Phillips (1975). A similar real-parameter minimax algorithm was used in Mäkilä and Partington (1992) with the same purpose of obtaining an easier numerical solution. The algorithm in Mäkilä and Partington (1992) uses non-uniformly spaced data and works well with Laguerre models. The proof of the following result is modified from Theorem (22).

Theorem 23. Consider the orthonormal set equation (6). Let $r_n = \max_{k<n}|\xi_k|$. Suppose $\{e^{j\omega_k}\}_{k\geq 0}$ is dense in $\mathbf{T}$. Let δ_N be as in equation (38). For each N, choose an n such that equation (43) is satisfied for some $0 < \delta < 1$. Let $\tilde{G}_N$ be the estimate of $G \in A(\mathbf{D})$ given in equation (68). Then

$$\|G - \tilde{G}\|_\infty \leq \left(\frac{2}{\tilde{\delta}} + 1\right) d(G, X_n; A(\mathbf{D})) + \frac{2}{\tilde{\delta}}\varepsilon \quad (69)$$

where $\tilde{\delta} = \delta/(1 + \sqrt{2})$.

Proof. Let $\hat{G}_N = \sum_{k=0}^{n-1} \hat{\lambda}_k \mathscr{B}_k$ be a solution in equation (12) and let $\hat{\Lambda} = (\hat{\lambda}_0 \cdots \hat{\lambda}_{n-1})^{\mathrm{T}}$. Since $\tilde{G}_N \in X_n$, by the same argument as in the proof of Theorem 22, we derive the following inequality similar to equation (62):

$$\max_{0 \leq k \leq N} |\Psi(e^{j\omega_k}) - \tilde{G}_N(e^{j\omega_k})| \geq \delta \|\Psi - \tilde{G}_N\|_\infty, \quad (70)$$

where Ψ is the closest element of X_n in $A(\mathbf{D})$ and the inequalities in equation (63) are replaced by

$$|\Psi(e^{j\omega_k}) - \tilde{G}_N(e^{j\omega_k})| \leq \|\Psi - G\|_\infty + \varepsilon + \|E - \Phi\tilde{\Lambda}\|_\infty. \quad (71)$$

Next

$$\begin{aligned} \|E - \Phi\tilde{\Lambda}\|_\infty &\leq \sqrt{2} \left\| \begin{bmatrix} \Phi_R \\ \Phi_I \end{bmatrix} \tilde{\Lambda} - \begin{bmatrix} E_R \\ E_I \end{bmatrix} \right\|_\infty \\ &\leq \sqrt{2} \left\| \begin{bmatrix} \Phi_R \\ \Phi_I \end{bmatrix} \hat{\Lambda} - \begin{bmatrix} E_R \\ E_I \end{bmatrix} \right\|_\infty \\ &\leq \sqrt{2}\|\Phi\hat{\Lambda} - E\|_\infty \\ &\leq \sqrt{2}(\|G - \Psi\|_\infty + \varepsilon), \end{aligned} \quad (72)$$

where the second inequality is due to the fact $\hat{\Lambda}$ is a candidate minimiser of equation (68). Lastly, we have the following inequality

$$\|G - \tilde{G}_N\|_\infty \leq \|G - \Psi\|_\infty + \|\Psi - \tilde{G}_N\|_\infty. \quad (73)$$

Thus from equations (70)–(73), we obtain equation (69). □

10. EXAMPLE

In this section, we use a simulation example to illustrate the minimax algorithm (12). Work is in progress to use the methods developed and analysed here on real data. Consider the approximation of the infinite-dimensional system

$$G(s) = \frac{e^{-s}}{\sqrt{s^2 + \sqrt{3s + 1}}}$$

by a rational transfer function. (The square root of $s = e^{j\theta}$ is defined by $e^{j\theta/2}$ for $\theta \neq \pm\pi$.) The identification of this system for $N = 256$ equally spaced frequencies in $[0, \pi)$ as obtained through by the bilinear map

$$s = \psi(z) = \frac{1 - z}{1 + z}$$

was investigated in Akçay *et al.* (1993). We use the same transformation so that $G(\psi) \in A(\mathbf{D})$. Then the continuous-time identified system is obtained by the back transformation $z = \psi^{-1}(s)$ from the discrete-time identified system. This map preserves the sup-norm.

We will compare the minimax algorithm of this paper with a Fourier series based algorithm. The real-parameter minimax algorithm in equation (68) is implemented with the basis functions $\mathscr{B}_0(z) = 1$

and

$$\mathscr{B}_k(z) = \frac{(1 + 0.1k)z + 1}{z + 1 + 0.1k}, \quad k = 1, \ldots, 50.$$

In the Fourier-based algorithm, first the frequency response data are extended into the interval $(\pi, 2\pi)$ using complex conjugate symmetry of G. Then the impulse-response coefficients of G are estimated from the extended frequency response by 512-point inverse discrete-Fourier transform. In the third step, a linear model is calculated as $\sum_{k=0}^{200} \hat{g}(k)z^k$. The eighth-order nominal model is obtained by the balanced truncation of $\hat{g}$ as implemented by the Kung's realization algorithm (Kung, 1978). The eighth-order nominal model for the minimax algorithm is extracted from the real-parameter minimax solution by first calculating 51 impulse-response coefficients in $\sum_{k=0}^{50} c_k \mathscr{B}_k(z)$, where c_k are the coefficients sought in the minimax problem, and then applying the model reduction technique described above.

In Figs. 1 and 2, the frequency responses of G, nominal models, and the identification errors

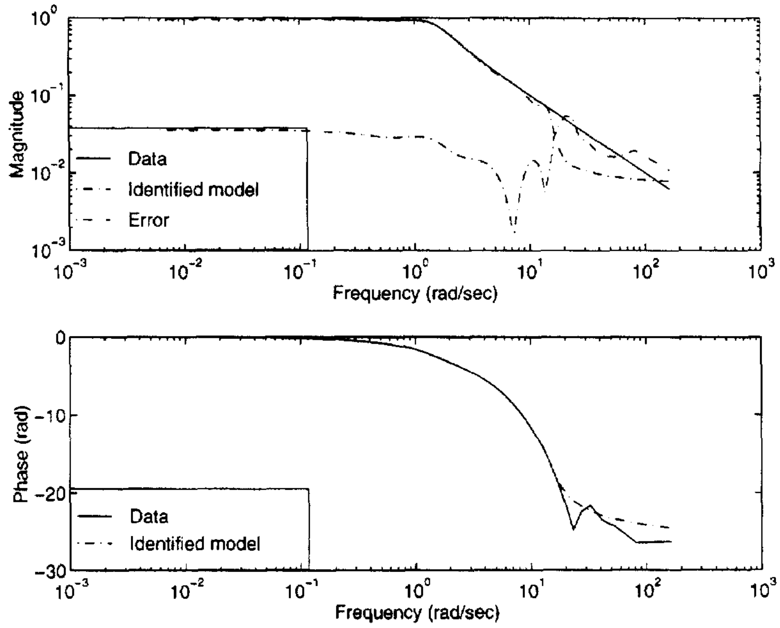

Fig. 1. Plot of G and the eighth-order model frequency responses and the approximation error magnitude using the Fourier-series-based algorithm

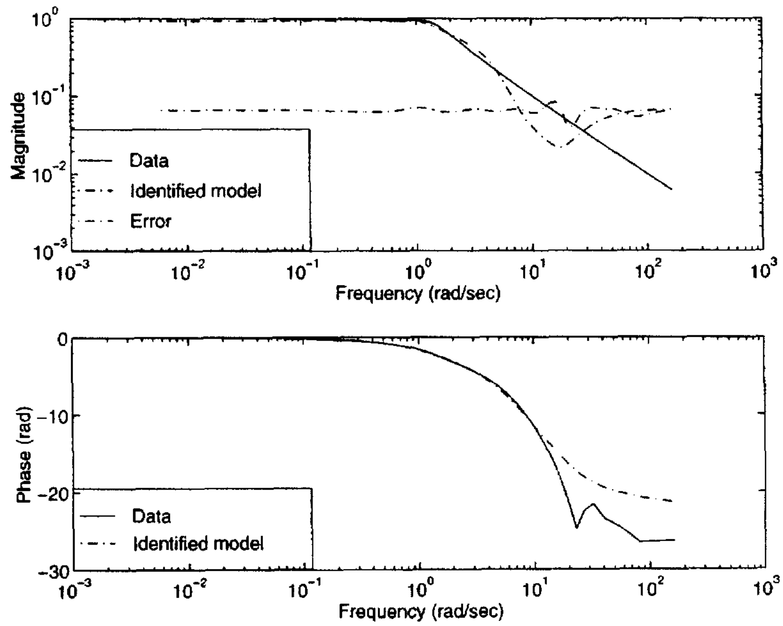

Fig. 2. Plot of G and the eighth-order model frequency responses and the approximation error magnitude using the real-parameter minimax algorithm.

produced by the algorithms on the same noise-free data set are plotted, respectively. The $L_\infty(\mathbf{T})$ errors of the nominal models were computed 0.0542 for the Fourier series based and 0.0831 for the real-parameter minimax algorithm. Both algorithms are successful in capturing the low-frequency dynamics of G and outside the bandwidth, the approximation errors are relatively small in comparison to $\|G\|_\infty$.

11. CONCLUSIONS

This paper has provided an analysis of the use of rational model structures in a robust estimation context. A key result of this analysis is the provision of necessary and sufficient conditions on the poles of the rational model structures for them to form a fundamental set in $A(\mathbf{D})$ and $H_p(\mathbf{D})$ for $1 \leq p < \infty$. For $A(\mathbf{D})$ this condition, which restricts how quickly the poles may approach the stability boundary, is much milder than sufficient conditions that have been put forward by other authors.

Having established these results, and similar ones for ℓ_1, the paper showed how robust estimation algorithms using both time and frequency domain data could be constructed together with explicit error bounds on the estimation accuracy. These results have implications, as discussed, for mixed parametric/non-parametric estimation, model reduction and may be extended to the multivariable setting.

An important and perhaps surprising point is that although the main spaces of interest are not the Hilbert space $H_2(\mathbf{D})$, the key analytical tool used is to in fact re-formulate the various problems studied into equivalent problems expressed in terms of a basis that is orthonormal in $H_2(\mathbf{D})$, with the orthonormality itself playing an important role.

This is reminiscent of equivalent strategies employed in the classical theory of orthogonal polynomials where, as in this paper, the key use of the orthonormal property is the derivation of a so-called Christoffel–Darboux formula for the reproducing kernel associated with the subspace in which the estimated model is constrained to lie according to the chosen model structure. This same strategy of reformulating estimation problems with respect to an orthonormal basis for the purposes of facilitating analysis has also been employed in a stochastic prediction error setting in Ninness *et al.* (1997).

Acknowledgements—The authors are grateful to K. J. Harrison, J. R. Partington, and J. A. Ward for sending them a preprint.

REFERENCES

Achieser, N. I. (1992). *Theory of Approximation*. Dover Publications, New York, (Transl. C. J. Hyman).

Akçay, H. (1997) Algorithms for robust identification in $\mathscr{H}_\infty$ with nonuniformly spaced frequency response data. In *Proc. 36th IEEE Conf. Dec. Control*, San Diego, CA, *Math. Control Signals Systems*, to appear.

Akçay, H., G. Gu and P. P. Khargonekar (1993). A class of algorithms for identification in $\mathscr{H}_\infty$: continuous-time case. *IEEE Trans. Automat. Control*, **38**, 289–294.

Akçay, H., G. Gu and P. P. Khargonekar (1994). Identification in $\mathscr{H}_\infty$ with nonuniformly spaced frequency response measurements. *Int. J. Robust Nonlinear Control*, **4**, 613–629.

Akçay, H., S. M. Islam and B. Ninness (1997). Identification of power transformer models from frequncy response data: a case study. *Signal Processing*, to appear.

Barrodale, I. and C. Phillips (1975). Algorithm 495, solution of an overdetermined system of linear equations in the Chebyshev norm [f4]. *ACM Trans. Math. Software*, **1**, 264–270.

Bodin, P. and B. Wahlberg (1994). Thresholding in high order transfer function estimation. In *Proc. 33rd IEEE Conf. Dec. Control*, Lake Buena Vista, FL, pp. 3400–3405.

Bokor, J., L. Gianone and Z. Szabo (1995a). Construction of generalized orthonormal basis in $\mathscr{H}_2$. Technical report, Computer and Automation Institute, Hungarian Academy of Sciences.

Bokor, J., F. Schipp and L. Gianone (1995b). Approximate $\mathscr{H}_\infty$ identification using partial sum operators in disc algebra basis. In *Proc. Amer. Control Conf.*, pp. 1981–1985.

Borwein, P. and T. Erdelyi (1996). Sharp extensions of Bernstein's inequality to rational spaces. *Mathematika*, **43**, 413–423.

Caines, P. (1996). *Linear Stochastic Systems*. Wiley, New York.

Chen, J., C. N. Nett and M. K. H. Fan (1995). Worst case system identification in $\mathscr{H}_\infty$: validation of *a priori* information, essentially optimal algorithms and error bounds. *IEEE Trans. Automat. Control*, **40**, 1260–1265.

Dahleh, M. and M. Khammash. Controller design for plants with structured uncertainty. *Automatica*, **29**, 37–56.

Dahleh, M. A., T. Theodosopoulos and J. N. Tsitsiklis (1993). The sample complexity of worst-case identification of FIR linear systems. *Systems Control Lett.* **20**, 157–166.

den Brinker, A. C. (1996). Optimality conditions for a specific class of truncated Kautz series. *IEEE Trans. Circuits Systems—Part II*, **43**, 597–600.

Dewilde, P. and H. Dym (1981). Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences. *IEEE Trans. Information Theory*, **27**, 446–461.

Duren. P. L. (1970). *Theory of H^p Spaces*. Academic Press, New York and London.

Elia, N. and M. Milanese (1993). Worst-case ℓ_1 identification using mixed parametric-nonparametric models. In *Proc. 32nd IEEE Conf. Dec. Control*, San Antonio, Texas, pp. 545–550.

Fu, Y. and G. A. Dumont (1993). On determination of Laguerre filter pole through step or impulse response data. In *Proc. 12th IFAC World Congress*, Sydney, Australia, Vol. 5, pp. 303–307.

Giarre, L., M. Milanese and M. Taragna (1997). H_∞ identification and model quality evaluation. *IEEE Trans. Automat. Control*, **42**, 188–199.

Gu, G. (1994). Suboptimal algorithms for worst-case identification in $\mathscr{H}_\infty$ and model validation. *IEEE Trans. Automat. Control*, **39**, 1657–1661.

Gu, G. and P. P. Khargonekar (1992a). A class of algorithms for identification in $\mathscr{H}_\infty$. *Automatica*, **28**, 229–242.

Gu, G. and P. P. Khargonekar (1992b). Linear and nonlinear algorithms for identification in $\mathscr{H}_\infty$ with error bounds. *IEEE Trans. Automat. Control*, **37**, 953–963.

Gu, G., D. Xiong and K. Zhou (1993). Identification in $\mathscr{H}_\infty$ using Pick's interpolation. *Systems Control Lett.*, **20**, 263–272.

Harrison, K. J., J. R. Partington and J. A. Ward (1997). Complexity of identification of discrete-time linear systems with rational transfer functions. Preprint.

Harrison, K. J. and J. A. Ward (1996). Fractional covers for convolution products. *Result. Math.*, **30**, 67–78.

Harrison, K. J., J. A. Ward and D. K. Gamble (1996). Sample complexity of worst-case identification. *Systems Control Lett.*, **27**, 255–260.

Hayman, W. K. and T. J. Lyons (1990). Bases for positive continuous functions. *J. London Math. Soc.*, **42**, 292–308.

Helmicki, A. J., C. A. Jacobson and C. N. Nett (1991). Control-oriented system identification: A worst-case/deterministic approach in $\mathscr{H}_\infty$. *IEEE Trans. Automat. Control*, **36**, 1163–1176.

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

Kacewicz, B. and M. Milanese (1995). Optimality properties in finite sample ℓ_1 identification with bounded noise. *Int. J. Adaptive Control Signal Processing*, **9**, 87–96.

Kosut, R. L., M. K. Lau and S. P. Boyd (1992). Set membership identification of systems with parametric and nonparametric uncertainty. *IEEE Trans. Automat. Control*, **37**, 929–941.

Kung, S. Y. (1978). A new identification and model reduction algorithm via singular value decomposition. In *Proc. 12th Asilomar Conference on Circuits, Systems and Computers*, Pacific Grove, CA, pp. 705–714.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.

Mäkilä, P. (1990). Laguerre series approximation of infinite dimensional systems. *Automatica*, **26**, 985–995.

Mäkilä, P. M. (1990) Robust identification and Galois sequences. *Int. J. Control*, **54**, 1189–1200.

Mäkilä, P. M. and J. R. Partington (1992). Robust identification of strongly stabilizable systems. *IEEE Trans. Automat. Control*, **37**, 1709–1716.

Mäkilä, P. M., J. R. Partington and T. Gustafsson (1995). Worst-case control-relevant identification. *Automatica*, **31**, 1799–1820.

McKelvey, T., H. Akçay and L. Ljung (1996). Subspace-based multivariable system identification from frequency response data. *IEEE Trans. Automat. Control*, **41**, 960–979.

Ninness, B. and G. Goodwin (1995). Estimation of model quality. *Automatica*, **31**, 32–74.

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

Ninness, B. and H. Hjalmarsson (1997). Generalised Fourier and Toeplitz results for rational orthonormal bases. Technical Report EE9740, Dept. Elec. and Comp. Eng., Univ. Newcastle, Australia. *SIAM J. Control Optim.*, (submitted).

Ninness, B., H. Hjalmarsson and F. Gustafsson (1997). The Fundamental role of orthonormal bases in system identification. Technical Report EE9739, Dept. Elec. and Comp. Eng., Univ. Newcastle, Australia. *IEEE Trans. Automat. Control.* (submitted).

Oliveira e Silva, T. (1995). On the determination of the optimal pole position of Laguerre filters. *IEEE Trans. Signal Process.*, **49**, 2079–2087.

Oliveira e Silva, T. (1996). N-width result for the generalized orthonormal basis function model. In *Proc. 13th IFAC World Congr.*, Vol. I, pp. 375–380, San Fransisco, CA.

Partington, J. R. (1991). Robust identification and interpolation in $\mathscr{H}_\infty$. *Int. J. Control*, **54**, 1281–1290.

Partington, J. R. (1992). Robust identification in $\mathscr{H}_\infty$. *J. Math. Anal. Appl.*, **166**, 428–441.

Partington, J. R. (1993). Algorithms for identification in $\mathscr{H}_\infty$ with unequally spaced function measurements. *Int. J. Control*, **58**, 21–31.

Partington, J. R. (1994). Interpolation in normed spaces from the values of linear functionals. *Bull. London Math. Soc.*, **26**, 165–170.

Partington, J. R. (1994). Worst-case identification in ℓ^2: linear and non-linear algorithms. *Systems Control Lett.*, **22**, 93–98.

Partington, J. R. (1996). Recovery of functions by interpolation and sampling. *J. Math. Anal. Appl.*, **198**, 301–309.

Partington, J. R. and P. M. Mäkilä (1995). Worst-case analysis of the least-squares method and related identification methods. *Systems Control Lett.*, **24**, 193–200.

Pinkus, A. (1985). *n-Widths in Approximation Theory*. Springer, Berlin.

Poolla, K. and A. Tikku (1994). On the time complexity of worst-case system identification. *IEEE Trans. Automat. Control*, **39**, 944–950.

Rudin, W. (1987). *Real and Complex Analysis*. 3rd edn. McGraw-Hill, New York.

Schipp, F. and J. Bokor (1997). Approximation by discrete Laguerre and Kautz functions in $\mathscr{H}_\infty$ norm. Preprint.

Schipp, F., L. Gianone and Z. Szabo (1996). Identification in generalized orthonormal basis—a frequency domain approach. In *Proc. 13th IFAC World Congress*, San Fransisco, pp. 387–392.

Söderström, T. and P. Stoica. (1989). *System Identification*. Prentice Hall Int., Hemel Hempstead, Hertfordshire.

Somorjai, G. (1977). On discrete linear operators in the function space *A*. In *Proc. Constructive Function Theory*, Blagoevgrad, Sofia, 1980. H. Bulgarian Acad. Sci. pp. 489–500.

Szabo, Z., J. Bokor and F. Schipp. (1997). Identification of rational approximate models in $\mathscr{H}_\infty$ using generalized orthonormal basis. *IEEE Trans. Automat. Control*, (to appear).

Tse, T., M. Dahleh and J. Tsitsiklis (1993). Optimal asymptotic identification under bounded disturbances. *IEEE Trans. Automat. Control*, **38**, 1176–1190.

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

Wahlberg, B. (1991). System identification using Laguerre models. *IEEE Trans. Automat. Control*, **36**, 551–562.

Wahlberg, B. (1994). System identification using Kautz models. *IEEE Trans. Automat. Control*, **AC-39**, 1276–1282.

Wahlberg, B. and L. Ljung (1992). Hard frequency-domain model error bounds from least-squares like identification techniques. *IEEE Trans. Automat. Control*, **AC-37**, 900–912.

Ward, N. F. D. and J. R. Partington (1995). Rational wavelet decomposition of transfer functions in Hardy–Sobolev classes. *Math. Control Signals Systems*, **8**, 257–278.

Ward, N. F. D. and J. R. Partington (1996). Robust identification in the disk algebra using rational wavelets and orthonormal basis functions. *Int. J. Control*, **64**, 409–423.

Zimmermann, S. and G. A. Williamson (1994). Selecting pole locations for system identifiers with fixed poles. In *Proc. 10th IFAC Symp. Systems Identif.*, Copenhagen, Denmark, Vol. 3, pp. 519–524.

Zygmund, A. (1959). *Trigonometric Series*. Cambridge University Press, Cambridge.