PRACTICAL ASPECTS OF USING ORTHONORMAL SYSTEM PARAMETERISATIONS IN ESTIMATION PROBLEMS

Brett Ninness [*,1] **Stuart Gibson** [**] **Steve Weller** [***]

[*] *Dept. of Elec. & Comp. Eng, Uni. Newcastle, Australia.*
email:`brett@ee.newcastle.edu.au`, *FAX: +61 49 21 69 93*
[**] *Dept. of Elec. & Comp. Eng, Uni. Newcastle, Australia.*
email:`shgibson@ee.newcastle.edu.au`, *FAX: +6149216993*
[***] *Dept. of Elec. & Comp. Eng, Uni. Newcastle, Australia.*
email:`steve@ee.newcastle.edu.au`, *FAX: +61 49 21 69 93*

Abstract: The last decade has seen a resurgence of interest in the use of ortho-normalised parameterisations for system estimation applications. This has produced a large body of theoretical work that has sought to provide sound scientific underpinnings. The paper here has a different focus in that emphasis is placed on practical dividends associated with orthonormal parameterisations.

Keywords: System Identification, Parameter Estimation, Orthonormal Basis.

1. INTRODUCTION

The idea of using orthonormal parameterisations of linear system representations has a long history (Takenaka 1925, Malmquist 1925, Lee 1933, Lee 1960, Ross 1964, Kautz 1952), but the last decade has provided a particularly strong revitalisation of the area by virtue of work that has strengthened the theoretical foundations on which the ideas rest (Wahlberg 1991*b*, Wahlberg 1994, P.M.J. Van den Hof *et al.* 1995, Heuberger *et al.* 1995, Wahlberg and Mäkilä 1996, Dudley and Partington 1996, Bokor and Schipp 1998, Oliveira e Silva 1997, Ninness *et al.* 1999*c*).

With these new insights and understandings in place, it is now appropriate to focus on practical applications, and the aim of this paper is to make some contribution towards that end.

In the authors opinion, the key aspects that imbue the use of orthonormal representations with a practical utility are two-fold:

(1) The orthonormality provides improved numerical conditioning;
(2) An orthonormal parameterisation is a more convenient basis to use when analysing the performance of an algorithm (even if this basis is not used in the implementation).

The implications of the first aspect are illustrated on two application examples:

(1) Frequency domain estimation from data measured from a flexible structure. In this case, the improved numerical conditioning leads to computable solutions to the associated 'curve-fitting' problem;
(2) Adaptive Tracking using the common LMS update scheme. Here the improved numerical conditioning delivers a faster convergence rate and, as opposed to the previous example, is independent of the numerical precision associated with the particular computing hardware employed.

With regard to the second aspect of practical relevance, we illustrate this by profiling how new quantifications of estimation accuracy may be derived using an orthonormal parameterisation in the theoretical analysis, even if such a parameterisation is not used (in the computer code) for the estimation methods being analysed.

The authors wish to stress that virtually all of the technical material in this paper has been presented elsewhere across a range of principally theoretical papers (P.M.J. Van den Hof *et al.* 1995, Wahlberg 1991*b*, Wahlberg 1991*a*, Ninness *et al.* 1999*b*, Akçay and Ninness 1999, Ninness and Gómez 1998). As such, the contribution of this paper is not meant to be along lines of novelty.

Rather, in consideration of the tutorial session in which this paper is meant to contribute, the effort here uses focussed examples to try to clarify where the use of a rational orthonormal parameterisation can have a practical dividend.

2. ORTHONORMAL BASES

There are two specific orthonormal basis formulations that will be considered here. The first is relevant for modelling sampled data systems using discrete-time models and is formulated as

$$G(q,\theta) = \sum_{k=1}^{n} \theta_k \mathcal{B}_k(q) \qquad (1)$$

[1] This work was supported by the Australian Research Council and the Centre for Integrated Dynamics and Control (CIDAC).

where

$$\mathcal{B}_k(q) = \left(\frac{\sqrt{1-|\xi_k|^2}}{q-\xi_k}\right)\prod_{\ell=1}^{k-1}\left(\frac{1-\overline{\xi_\ell}q}{q-\xi_\ell}\right). \quad (2)$$

Here $G(q,\theta)$ is a model for discrete time dynamics that are parameterised by a vector $\theta \in \mathbf{R}^n$.

The poles ξ_k in (2) are all chosen, according to prior knowledge of the poles of the system being modelled, within the open unit disc $\mathbf{D}$. Indeed, a necessary and sufficient condition for the expansion (1) being able to arbitrary well model a system (with respect to H_p norm, $1 < p < \infty$) is that (Achieser 1992, Hüseyin Açkay and Brett Ninness 1998)

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

The second class of orthonormal basis models studied here are appropriate for modelling continuous time systems and are formulated as

$$G(s,\theta) = \sum_{k=1}^{n}\theta_k B_k(s) \quad (3)$$

where

$$B_k(s) \triangleq \frac{\sqrt{2\text{Re}\{\xi_k\}}}{s+\xi_k}\prod_{\ell=1}^{k-1}\frac{s-\overline{\xi_\ell}}{s+\xi_\ell} \quad (4)$$

and now the co-efficients ξ_k are chosen, again according to prior system knowledge, but this time in the open right half plane $\mathbf{C}^+$. In this case, a necessary and sufficient condition for arbitrary H_p norm modelling accuracy is (Akçay and Ninness 1999, Walsh 1935)

$$\sum_{k=1}^{\infty}\frac{\text{Re}\{\xi_k\}}{1+|\xi_k|^2} = \infty. \quad (5)$$

3. NUMERICAL CONDITIONING ADVANTAGES

A principle practical advantage involved with using an orthonormal model structure is that the numerical conditioning of any required matrix/vector operations is greatly improved.

In some applications, such as the frequency response estimation one profiled next, this leads to more reliable computation with concomitant higher quality estimation results.

In other application, such as the LMS filtering one presented shortly, the advantages of improved numerical conditioning have nothing to do with the finite precision aspects of any computational machine involved. Instead, they are intrinsic to the particular search algorithm employed.

3.1 *Frequency Response Estimation*

The utilisation of rational orthonormal bases for identification from frequency domain data has been considered by a number of authors (Wahlberg and Mäkilä 1996, Mäkilä 1990, Mäkilä 1992, Dudley Ward and Partington 1998), but the focus there has been on estimating only a set of numerator co-efficients.

Here we show how, via trivial manipulations, a denominator may also be estimated as well, but in a manner with numerical conditioning that is much improved over pre-existing methods that parameterise using polynomial bases.

For the sake of concreteness, we focus on a specific problem which involve measurements of the frequency response of a 58cm long, 5mm wide cantilevered piezo-electric laminate beam (for further details, see (Moheimani *et al.* 1998)).

These measurements are shown as dots in Figure 1. For the purposes of control (stiffness compensation using the piezo-electric actuators) a transfer function model that explains this frequency response is required. There are many ways in which this may be achieved (Pintelon *et al.* 1994), but for the purposes of illustrating the efficacy of the basis (4), the simple least-squares method of fitting a model

$$G_n(s) = \frac{N(s)}{D(s)} = \frac{b_n s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}{s^n + d_{n-1}s^{n-1} + \cdots + d_1 s + d_0}$$

to the measured frequency response $\{G(j\omega_1),\cdots,G(j\omega_N)\}$ by means of minimising the cost

$$V_N = \sum_{k=1}^{N}|D(j\omega_k)G(\omega_k) - N(\omega_k)|^2$$

will be studied. As is well known (Pintelon *et al.* 1994), finding this estimate involves solving the so-called 'Normal Equations'

$$\begin{bmatrix} G(j\omega_1)(j\omega_1)^n \\ \vdots \\ G(j\omega_N)(j\omega_N)^n \end{bmatrix} = \Phi \begin{bmatrix} -d_{n-1} \\ \vdots \\ -d_0 \\ b_n \\ \vdots \\ b_0 \end{bmatrix}$$

where

$$\Phi \triangleq \begin{bmatrix} G(j\omega_1)(j\omega_1)^{n-1}, & \cdots & G(j\omega_1), & (j\omega_1)^n, & \cdots & 1 \\ \vdots & & & & & \vdots \\ G(j\omega_N)(j\omega_N)^{n-1}, & \cdots & G(j\omega_N), & (j\omega_N)^n, & \cdots & 1 \end{bmatrix}$$

The numerical stability of the associated solution is highly dependent (Golub and Loan 1989), on the conditioning of the matrix $\Phi^T\Phi$.

However this can be altered via re-parameterisations of the model $G(s)$.

For example, in (Bayard 1992) the parameterisation

$$N(s) = b_0 + \sum_{k=1}^{n} b_k p_k(s),$$

$$D(s) = s^n + \sum_{k=0}^{n-1} d_k p_k(s) \qquad (6)$$

where each $p_k(s)$ is an order k 'modified Tchebychev' polynomial ($p_0 = 1$) is suggested as a means of improving numerical conditioning.

In Figure 1, the dash-dot line shows the results of using the above Tchebychev parameterisation to fit an $n = 18$'th order model to the observed frequency response. Note that the second resonance peak is completely missed, and that there are three close pole-zero cancellation 'spikes' from the 4'th resonant mode onwards (this is still much improved from the results obtained by using the standard basis $p_k(s) = s^k$).

However, if the model is parameterised using the orthonormal basis (4) as

$$N(s) = b_0 + \sum_{k=1}^{n} b_k B_k(s)$$
$$D(s) = s^n + \sum_{k=1}^{n} d_k B_k(s) \qquad (7)$$

with the (Laguerre) pole choice $\xi_k = \xi = 2\omega_N$, then the ensuing 18'th order least squares estimate is the solid line shown in Figure 1. It now captures the second resonance peak, and does not have high frequency near pole-zero cancellations.

Since the model structures (6) and (7) both span the same manifold of rational models, the only explanation for the difference in results is that of differences in numerical conditioning. Figure 2 shows the singular values of Φ for three model parameterisation choices.

Considering the log-scale employed, the parameterisation using the basis (7) enjoys a two order of magnitude better conditioning (ratio of largest to smallest singular value) than either a Tchebychev polynomial, or conventional polynomial parameterisation.

As a consequence, for such applications of modelling resonant structures over large bandwidths, we suggest that the basis (4) should be employed in the interests of the resultant frequency response having no artifacts due to poor numerical conditioning.

3.2 *Adaptive Filtering*

Techniques of recursively updating an estimate of underlying (possibly time varying) linear dynamics are fundamental to many areas of signal processing, particularly ones aimed to support modern wireless telecommunications methods (Solo and Kong 1995, Verdu 1998).

The problem may be represented abstractly as one in which an observed input sequence $\{u_t\}$ is related to an observed output sequence $\{y_t\}$ according to

$$y_t = G_t(q)u_t + \nu_t \qquad (8)$$

where $\{\nu_t\}$ is a noise process (that may model a range of practical effects) and

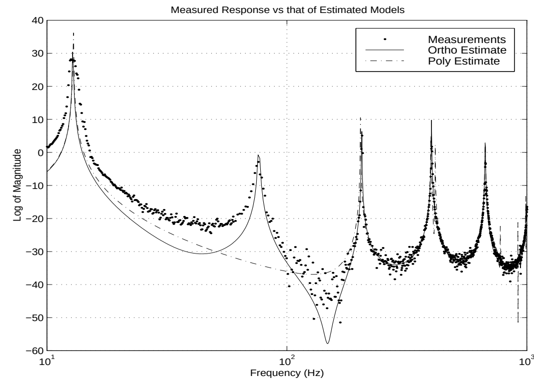

Fig. 1. *Estimation using polynomial and orthonormal basis. The dots are the measurements, the solid line is the estimate using basis (4) to parameterise the model, the dash-dot line is the estimate using (modified) Tchebychev polynomials to parameterise the model.*

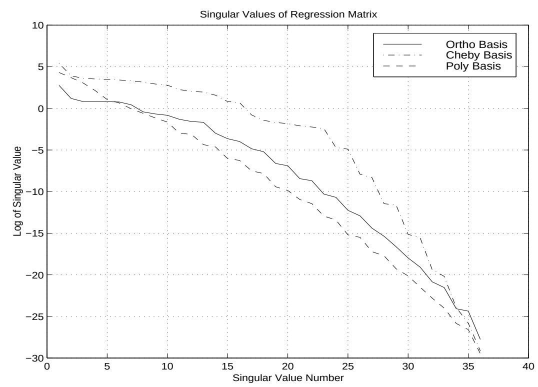

Fig. 2. *Singular Values of Φ using polynomial (natural and Tchebychev) and the rational orthonormal basis (4).*

$$G_t(q) = \sum_{n=1}^{\infty} g_t(n) q^{-n}$$

is a possibly time varying linear system with impulse response $\{g_t(n)\} \in \ell_2$.

There are many approaches to how $G_t(q)$ might be estimated from this data in an on-line fashion, but a common theme (Goodwin and Sin 1984) is to express the dependence (8) in a linear regression form

$$y_t = \phi_t^T \theta_t + \nu_t \qquad (9)$$

where the 'regression vector' ϕ_t depends on measurements of $\{u_t\}$ and $\{y_t\}$ up until $t = k$ and $\theta_t \in \mathbf{R}^n$ is a vector of n parameters in a model structure $G(q, \theta_t)$ that attempts to describe the true dynamics $G_t(q)$.

An estimate of $G_t(q)$ is then obtained as $G(q, \widehat{\theta}_t)$ where the estimate $\widehat{\theta}_t$ is obtained recursively via

$$\widehat{\theta}_{t+1} = \widehat{\theta}_t + L_t(y_t - \phi_t^T \widehat{\theta}_t), \quad \mu \in (0, 1) \qquad (10)$$

where L_t is a gain vector that may be computed in various ways.

A common choice for this gain vector is $L_t = \mu\phi_t,\ \mu \in (0,1)$ in which case (10) is known as the 'gradient' or 'least mean square' (LMS) algorithm.

3.2.1. *Model Structures*

The regression vector ϕ_t in (9), (10) is determined by the model structure chosen for $G_t(q)$ and, of course, this paper is interested in considering the orthonormal structure (1) so that ϕ_t becomes

$$\phi_t^T \triangleq [\mathcal{B}_1(q)u_t, \mathcal{B}_1(q)u_t, \cdots, \mathcal{B}_n(q)u_t]\,.$$

If all the poles ξ_k are chosen at the origin, then this is nothing more than the very well known FIR model structure.

However, empirical evidence supports the fact (Williamson and Zimmermann 1996) that in an adaptive filtering context, a significant improvement in estimation accuracy is possible by avoiding poles $\{\xi_k\}$ all fixed at the origin, and instead distributing them in the unit disk so as to be as close as possible to the true poles of $G_t(q)$.

Indeed, motivated by this (Williamson and Zimmermann 1996) has proposed the so-called 'fixed pole adaptive filters' that are formulated as

$$G(q,\theta_t') = \left[\prod_{k=1}^{n}(q-\xi_k)\right]^{-1}\sum_{k=1}^{n}\theta_t'(k)q^k. \quad (11)$$

Clearly, there is a one–one correspondence between the model structures (11) and (1) in that for any θ there exists a θ' such that $G(q,\theta)$ in (1) is exactly the same transfer function as $G(q,\theta')$ in (11).

However, as the remainder of this section will highlight, despite this theoretical equivalence, there can be a great practical difference between the use of these two model structures in terms of LMS convergence rate.

3.2.2. *Convergence Rates and Numerical Conditioning*

As is well known (Guo and Ljung 1995) the convergence properties of the general learning algorithm (10) is determined by the (time varying) linear update relation

$$\widetilde{\theta}_{t+1} = (I - L_t\phi_t^T)\widetilde{\theta}_t.$$

where $\widetilde{\theta}_t \triangleq \widehat{\theta}_t - \theta_t$ is the error in the parameter estimation at time t.

For the LMS algorithm where $L_t = \mu\phi_t$, this suggests (Solo and Kong 1995) that the convergence (averaged over an ensemble of input realisations) is determined by the choice of μ combined with the eigenvalues of $R = \mathbf{E}\{\phi_t\phi_t^T\}$. Specifically, in order to guarantee stability it is necessary that the step–size μ be limited as

$$\frac{1}{\mu} > \lambda_{\max}(R)$$

in which case the parameter space convergence in the direction of the eigenvector associated with $\lambda_{\min}(R)$ is limited to be no faster than the rate of decay of

$$\left[1 - \frac{\lambda_{\min}(R)}{\lambda_{\max}(R)}\right]^t.$$

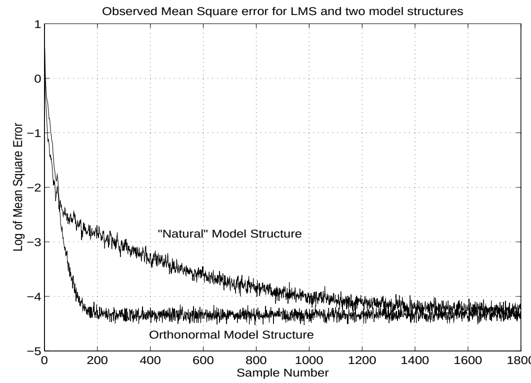

Fig. 3. *Sample mean square error rates (averaged over 500 realisations) for LMS algorithm with (top plot) 'natural' fixed denominator model structure (11) and (bottom plot) orthonormal model structure (1).*

Therefore, in the interests of LMS convergence speed, it is desirable to choose a model structure $G(q,\theta)$ such that the ensuing regressors ϕ_k implies the minimum possible condition number for R.

For the case of white input spectrum $\Phi_u(\omega) =$ constant, the orthonormal model structure (1) achieves this, and is therefore optimal (in the white input convergence sense) among the class of all fixed denominator model structures, of which the more 'natural' or obvious choice (11) is a member.

This optimality is illustrated via simulation in figure 3. The top (slowly converging) plot is the observed mean square error (averaged over 500 simulations with different input and measurement noise realisations) obtained when using the LMS algorithm with the 'natural' fixed–denominator model structure (11). The bottom (quickly converging) plot is the observed mean square error using the orthonormal model structure (1) and the same LMS algorithm, but with μ changed to keep the steady state error invariant to the change in model structure; the convergence rate improvement obtained by using the orthonormal structure is clear. In both cases the model structure was third order with poles chosen at $\xi_0 = 0.4$, $\xi_1 = 0.85$, $\xi_2 = 0.6$ and the true system $G_k(q)$ was time invariant with true poles at $\gamma_0 = 0.9$ and $\gamma_1 = 0.37$. The output was corrupted by white Gaussian distributed noise of variance $\sigma_\nu^2 = 0.01$, and the input was white Gaussian distributed noise of variance $\Phi_u(\omega) = 10$.

For the case of non–white $\Phi_u(\omega)$, this convergence–rate optimality of the orthonormal model structure (1) is lost, but a feature retained by the orthonormal structure (1) is that a bound on the numerical conditioning of R in terms of the properties of $\Phi_u(\omega)$ may be derived(Ninness and Gómez 1998).

$$\min_{\omega\in[-\pi,\pi]}\Phi_u(\omega) \le \lambda(R) \le \max_{\omega\in[-\pi,\pi]}\Phi_u(\omega). \quad (12)$$

This implies an upper bound on the on–average convergence rate of

$$\left[1 - \frac{\min_{\omega\in[-\pi,\pi]}\Phi_u(\omega)}{\max_{\omega\in[-\pi,\pi]}\Phi_u(\omega)}\right]^t. \quad (13)$$

4. PERFORMANCE ANALYSIS

Consider the general method of prediction error estimation wherein, for a given input data record $\{u_t\}$ and output data record $\{y_t\}$, a model is postulated as

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

Here $\{e_t\}$ is a zero-mean white noise sequence such that $\mathbf{E}\{e_t^2\} = \sigma^2$, and $G_\theta(q), H_\theta(q)$ are transfer functions, rational in the forward shift operator q, and parameterised by a vector $\theta \in \mathbf{R}^n$.

The mean-square optimal one-step ahead prediction $\widehat{y}_t(\theta)$ based on this model structure is ((Ljung 1987))

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

with associated prediction error

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

involved with the quadratic estimation criterion

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

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

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

This class of estimation methods is in very common use, and it is of key practical importance to understand the factors that influence the accuracy of any estimate obtained by them.

For this purpose, it is expedient to consider the effect on the estimated frequency response accuracy, and a by-now classical approximate expression for the noise-induced variability of this quantity is (Ljung 1987)

$$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta}_N)\} \approx \frac{n}{N}\frac{\sigma^2|H(e^{j\omega},\widehat{\theta}_N)|^2}{\Phi_u(\omega)}. \quad (14)$$

Here, Φ_u is the spectral density of the input, and the factor n may, according to model structure choice, be somewhat less that the dimension of the whole parameter vector θ.

Given the simplicity of (14), it is of great utility in the design of estimation schemes and in understanding their importance. As a simple example, (14) illustrates that if the model $G(q,\widehat{\theta}_N)$ is to be used for a control system design in which high modelling accuracy is required near a cross-over frequency ω_c, then the input spectrum should be designed so the $\Phi_u(\omega_c)$ is as large as possible since this minimises the variability of the estimate at that frequency.

Where orthonormal parameterisations enter is that, by a strategy of *pretending* that they parameterise $G(q,\theta)$ (even when any underlying computer code in fact does not) then an improved accuracy variance approximation can be derived.

This new approximation is

$$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta}_N)\} \approx \frac{1}{N}\frac{\sigma^2|H(e^{j\omega},\widehat{\theta}_N)|^2}{\Phi_u(\omega)} \times \sum_{k=1}^{n}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \quad (15)$$

and the definition of the ξ_k depend on the model structure being used. If it is one of fixed denominator, then the ξ_k are the poles of that fixed denominator (P.M.J. Van den Hof *et al.* 1995, Ninness *et al.* 1999*b*), while if it is ARX, then the ξ_k are the fixed zeros of $H(q,\theta)$ (Ninness *et al.* 1999*b*).

In both these cases, if these poles or zeros are all at the origin (FIR, conventional ARX), then (15) is identical to (14), and hence provides a smooth extension that can offer improved accuracy (Ninness *et al.* 1999*b*).

Of more recent discovery (Ninness *et al.* 1999*a*) is the fact that in the case of output-error modelling ($H(q,\theta) = 1$), then (15) with the ξ_k being the *estimated* poles can provide an approximation with better accuracy than (14).

This is illustrated in figure 4, where for a particular experimental setup (Ninness *et al.* 1999*a*), (14) is shown as the dash-dot line, (15) as the dashed line, and the true variability (estimated via Monte-Carlo analysis) is shown as the solid line.

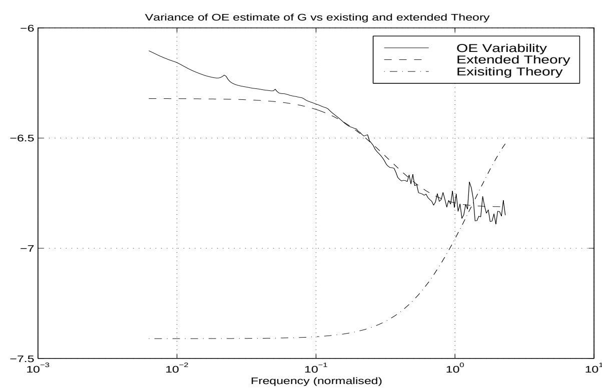

Fig. 4. *Variability of Output Error Estimate - True variability vs. new and existing theoretically derived approximations.*

In this case, the practical utility of the orthonormal parameterisation is not realised so directly as when it is used in actually coding an estimation algorithm. Rather, the benefit is achieved by enabling a theoretical analysis that is most closely matched to experimental conditions.

5. CONCLUSION

This paper has attempted to provided a brief précis illustrating the practical advantage of using an orthonormal parameterisation *for certain* problems of system estimation. A key theme was to delineate these advantages according to whether they are accrued do to improved numerical conditioning, or due to greater insight into performance. Within this framework, there are several further practical aspects that, due to space limitations, are not discussed here, but are treated in (Bokor *et al.* 1999).

6. REFERENCES

Achieser, N. (1992). *Theory of Approximation*. Dover Publications.

Akçay, H. and Brett Ninness (1999). ‘Orthonormal basis functions for continuous–time systems and l_p–convergence’. *Mathematics of Control, Signals and Systems (MCSS)* **12**(3), 295–305.

Bayard, D. (1992). Multivariable frequency domain identification via 2-norm minimization. In ‘Proc. American Control Conference’. pp. 1253–1257.

Bokor, J. and F. Schipp (1998). ‘Approximate identification in Laguerre and Kautz bases’. *Automatica J. IFAC* **34**(4), 463–468.

Bokor, J., Peter Heuberger, Brett Ninness, Tomás Oliveira e Silva, Paul Van den Hof and Bo Wahlberg (1999). *Modelling and Identification with Orthogonal Basis Functions*. Short Course Notes, 14th IFAC World Congress, 1999. Available on request to Prof. Paul Van den Hof.

Dudley, N. and J.R. Partington (1996). ‘Robust identification in the disc algebra using rational wavelets and orthonormal basis functions’. *Int. J. Control* **64**(3), 409–423.

Dudley Ward, N. F. and Jonathan R. Partington (1998). ‘A construction of rational wavelets and frames in Hardy-Sobolev spaces with applications to system modeling’. *SIAM J. Control Optim.* **36**(2), 654–679 (electronic).

Golub, G. and Charles Van Loan (1989). *Matrix Computations*. Johns Hopkins University Press.

Goodwin, G. and K.W. Sin (1984). *Adaptive Filtering Prediction and Control*. Prentice-Hall, Inc.. New Jersey.

Guo, L. and Lennart Ljung (1995). ‘Performance analysis of general tracking algorithms’. *IEEE Transactions on Automatic Control* **40**(8), 1388–1402.

Heuberger, P., P.M.J. Van den Hof and O.H. Bosgra (1995). ‘A generalized orthonormal basis for linear dynamical systems’. *IEEE Transactions on Automatic Control* **AC-40**(3), 451–465.

Hüseyin Açkay and Brett Ninness (1998). ‘Rational basis functions for robust identification from frequency and time domain measurements’. *Automatica* **34**(9), 1101–1117.

Kautz, W. H. (1952). Network synthesis for specified transient response. Technical Report 209. Massachusetts Institute of Technology. Research Laboratory of Electronics.

Lee, Y. (1933). ‘Synthesis of electric networks by means of the Fourier transforms of Laguerre’s functions.’. *Journal of Mathematics and Physics* **XI**, 83–113.

Lee, Y. W. (1960). *Statistical Theory of Communication*. Wiley. New York.

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

Mäkilä, P. (1990). ‘Laguerre series approx’n of infinite dim. systems’. *Automatica* **26**, 985–995.

Mäkilä, P. (1992). ‘Worst-case input-output identification’. *Int. J. Control* **56**(3), 673–689.

Malmquist, F. (1925). ‘Sur la détermination d’une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de points’. *Comptes Rendus du Sixième Congrès des mathématiciens scandinaves (Kopenhagen)* pp. 253–259.

Moheimani, S., H.R. Pota and I.R. Peterson (1998). Spatial control for active vibration control of piezoelectric laminates. In ‘Proceedings of the IEEE Conference on Decision and Control’.

Ninness, B. and Juan Carlos Gómez (1998). ‘Frequency domain analysis of tracking and noise properties of adaptive algorithms’. *IEEE Transactions on Signal Processing* **46**(5), 1314–1332.

Ninness, B., Håkan Hjalmarsson and Fredrik Gustafsson (1999*a*). Asymptotic variance expressions for output error model structures. In ‘Proceedings of the 14th IFAC World Congress, Beijing’.

Ninness, B., Håkan Hjalmarsson and Fredrik Gustafsson (1999*b*). ‘The fundamental role of general orthonormal bases in system identification’. *IEEE Transactions on Automatic Control* **44**(8), to appear.

Ninness, B., Håkan Hjalmarsson and Fredrik Gustafsson (1999*c*). ‘Generalised Fourier and Toeplitz results for rational orthonormal bases’. *SIAM Journal on Control and Optimization* **37**(2), 429–460.

Oliveira e Silva, T. (1997). ‘Stationary conditions for the L^2 error surface of the generalized orthonormal basis functions lattice filter’. *Signal Processing* **56**(3), 233–253.

Pintelon, R., P. Guillaume, Y. Rolain, J. Schoukens and H. Van hamme (1994). ‘Parametric identification of transfer functions in the frequency domain-a survey’. *IEEE Transactions on Automatic Control* **39**(11), 2245–2260.

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

Ross, D. (1964). ‘Orthonormal exponentials’. *IEEE Transactions on Communication and Electronics* **71**(12), 173–176.

Solo, V. and Xuan Kong (1995). *Adaptive Signal Processing Algorithms*. Prentice Hall.

Takenaka, S. (1925). ‘On the orthogonal functions and a new formula of interpolation’. *Japanese Journal of Mathematics* **II**, 129–145.

Verdu, S. (1998). *Multiuser Detection*. Cambridge University Press.

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

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

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

Wahlberg, B. and P.M. Mäkilä (1996). ‘On approximation of stable linear dynamical systmes using Laguerre and Kautz functions’. *Automatica* **32**(5), 693–708.

Walsh, J. (1935). *Interpolation and Approximation by Rational Functions in the Complex Domain*. Vol. XX of *Colloquium Publications*. American Mathematical Society.

Williamson, G. A. and Sergio Zimmermann (1996). ‘Globally convergent adaptive IIR filters based on fixed pole locations’. *IEEE Trans. Signal Processing* **44**(6), 1418–1427.