A Unifying Construction of Orthonormal Bases for System Identification

Brett Ninness[1] and Fredrik Gustafsson[2]

October 19, 1994

ABSTRACT

In this paper we develop a general and very simple construction for complete orthonormal bases for system identification. This construction provides a unifying formulation of all known orthonormal bases since the common FIR and recently popular Laguerre and Kautz model structures are restrictive special cases of our construction as is another construction method based on balanced realisations of all pass functions. However, in contrast to these special cases, the basis vectors in our unifying construction can have nearly arbitrary magnitude frequency response according to the prior information the user wishes to inject into the problem. We provide results characterising the completeness of our bases, and the accuracy properties of models estimated using our bases.

Technical Report EE9433

Department of Electrical and Computer Engineering, University of Newcastle, Callaghan 2308, AUSTRALIA

---

[1]This work was supported by the Centre for Industrial Control Science, University of Newcastle, Callaghan 2308, Australia. This author can be contacted at email:`brett@tesla.newcastle.edu.au` or FAX: +61 49 601712

[2]Work completed while on leave from the Department of Electrical Engineering, Linköping University, S-581 83, Linköping, Sweden. This author can be contacted at email:`fredrik@isy.liu.se` or FAX: +46 013 282622

1 Introduction

This paper is concerned with the problem of estimating the dynamics of single input, single output linear time invariant systems on the basis of noisy sampled observations of their input-output response. One of the most popular existing methods for dealing with this problem involves modelling the system dynamics via a so-called ARX structure [12, 20]. A problem then is that it is difficult to evaluate the variance of the estimated model except in an asymptotic sense. Additionally, the model parameters generally do not appear linearly and so estimation of them involves numerical solution of a non-linear optimisation problem.

This latter difficulty can be overcome by recasting the problem in a linear regression form, but in this case the parameters to be estimated affect both the dynamic model and the noise model. This can cause estimates of them to be biased [27].

An approach overcoming all these difficulties involves choosing a model structure which is a-priori linear in the parameters. For example

$$y_k = \left(\sum_{n=0}^{p-1} \theta_n \mathcal{B}_n(q) \right) u_k. \tag{1}$$

Here $\{\theta_n\}$ are the parameters to be estimated, $\{u_k\}$ is an observed input, $\{y_k\}$ is an observed output, $\{\mathcal{B}_n(q)\}$ is a set of transfer functions rational in the forward shift operator q, and p is the model order.

Now since $\theta \triangleq [\theta_0, \theta_1, \cdots, \theta_{p-1}]^T$ appears linearly, its least squares estimate $\widehat{\theta}$ can be found in closed form and is linear in $\{y_k\}$ so that if $\{u_k\}$ is not noise corrupted then finite data variances for $\widehat{\theta}$ can be calculated. Furthermore, θ parameterises only the model for the dynamics and so $\widehat{\theta}$ is not biased by measurement noise.

The remaining difficulty is that the estimate could be poor if the $\{\mathcal{B}_n(q)\}$ have been chosen inappropriately with respect to the true underlying dynamics that have generated the observed data. For example, the simplest choice for $\{\mathcal{B}_n(q)\}$ is

$$\mathcal{B}_n(q) = q^{-n}$$

so that (1) represents an FIR model structure. However, if the true dynamics have a slow mode, then the model order p will need to be very large for the model structure (1) to provide an accurate approximation to the true dynamics. The obvious strategy to overcome this is to instead choose

$$\mathcal{B}_n(q) = \frac{1}{q - \xi_n} \tag{2}$$

where the poles $\{\xi_n\}$ are chosen according to a-priori knowledge of the underlying dynamics; in the previous case of a known slow mode at least one of the $\{\xi_n\}$ would be chosen near 1.

This idea of incorporating prior information into the model structure (1) has led to the popular use of the so-called Laguerre model where instead of (2) the choice

$$\mathcal{B}_n(q) = \left(\frac{\sqrt{1-\xi^2}}{q - \xi} \right) \left(\frac{1 - \xi q}{q - \xi} \right)^n \qquad ; |\xi| < 1 \tag{3}$$

is made. Estimation using these models was first proposed in [9, 1] and has been studied in detail in [23, 25]. Where prior knowledge indicates resonant modes the so called Kautz model

$$\mathcal{B}_n(q) = \begin{cases} \dfrac{\sqrt{(1-\alpha^2)(1-\gamma^2)}}{q^2-\alpha(\gamma+1)q+\gamma}\left(\dfrac{\gamma q^2-\alpha(\gamma+1)q+1}{q^2-\alpha(\gamma+1)q+\gamma}\right)^{(n-1)/2} & ; n \text{ odd} \\ \\ \dfrac{\sqrt{(1-\gamma^2)}(q-\alpha)}{q^2-\alpha(\gamma+1)q+\gamma}\left(\dfrac{\gamma q^2-\alpha(\gamma+1)q+1}{q^2-\alpha(\gamma+1)q+\gamma}\right)^{n/2} & ; n \text{ even} \end{cases} \tag{4}$$

where $|\alpha| < 1$ and $|\gamma| < 1$ has been suggested and analysed in [22, 26, 24].

However, with both the Laguerre and Kautz structures, the reader will notice that a restriction over the general $\{\mathcal{B}_n\}$ in (2) has been made in that knowledge of only one mode and not a variety of modes can be incorporated; the magnitude frequency response of all the $\{\mathcal{B}_n(q)\}$ are the same. This restriction is imposed to obtain the $\{\mathcal{B}_n(q)\}$ as an orthonormal system. Orthonormality is important since it leads to

- Improved numerics in solving the least squares normal equations [23, 2].
- Parsimony in representation [27].
- Independence of parameters for broad band excitation signals [12].

The contribution of this paper is to show that the simple construction

$$\mathcal{B}_n(q) = \left(\frac{\sqrt{1-|\xi_n|^2}}{q-\xi_n}\right)\prod_{k=0}^{n-1}\left(\frac{1-\overline{\xi_k}q}{q-\xi_k}\right) \tag{5}$$

preserves orthonormality, but also allows prior knowledge about a **variety** of modes at $\{\xi_0, \xi_1, \cdots, \xi_{p-1}\}$ to be incorporated. Furthermore, the construction (5) provides a unifying formulation of all known system identification orthonormal systems since the well known FIR, Laguerre and Kautz models and a new method using balanced realisations of user chosen dynamics [4] are all special cases of (5) when only one and not a range of modes are incorporated.

Since these known systems are only special cases, our unifying construction (5) allows richer and more useful systems to be developed. More specifically, the flexibility in our construction of allowing prior knowledge about more than one mode to be incorporated, while still preserving orthonormality, leads to more accurate estimation. Finally, our unifying construction has a very simple and physically motivated development which we believe clarifies some earlier derivations for orthonormal systems.

A key tool in developing the model structure (5) is to consider the $\{\mathcal{B}_n\}$ as basis functions for the system identification problem. Using this vector space idea we show that our construction (5) is in essence the Gram-Schmidt procedure and also show that under certain mild necessary and sufficient conditions the basis functions in (5) are H_2 complete.

We then go on to explore links between our orthonormal construction and bases formed from the classical orthogonal polynomials. This leads to a new basis, the 'Legendre' basis, which like the already known Laguerre basis is a special case of our general construction. We find that it is not possible to form any more bases from other classical orthogonal polynomials such as Chebychev and Hermite since non-rational in q discrete time transfer functions are necessary to implement the bases.

Having done this we show how using the model structure (1) our basis functions can be used for system identification. We continue by providing a result which quantifies the accuracy of the resultant estimate. This result generalises the variance expressions presented in [15] for FIR models, in [27] for Laguerre models and in [21] for balanced realization models with the important feature that in contrast to these known results, our result is not asymptotic in the model order.

We finish the paper with some simulation examples that illustrate the utility of using our generalised orthonormal basis functions instead of the more popular Laguerre and Kautz models.

2 Orthonormal Basis Construction

The idea with system identification is to estimate the impulse response $\{g_k\}$ of a system. Most successful methods do not do this directly, but instead estimate some constrained representation of $\{g_k\}$. For example, an ARX model of some fixed order can be found, and this places some smoothness constraints on the estimated $\{g_k\}$. Alternatively, as has been suggested in (1), a linear combination of known fixed impulse responses can be estimated. The fixed responses can then be considered as 'basis functions' for the approximation of the true $\{g_k\}$. This latter idea is the one we wish to pursue in this paper.

As mentioned in the introduction, the paradigm has already been analysed in a number of special cases by a number of authors. If prior knowledge of a first order mode is available, then estimating $\{g_k\}$ as a linear combination of the impulse responses of the Laguerre transfer functions (3) seems appropriate while if knowledge of a resonant mode exists then the Kautz system (4) is preferable.

The deficiency with these existing ideas is that prior knowledge of only one mode can be incorporated. Furthermore, the basis function choices (3) and (4) seem strange; their derivations from the classical Laguerre and Kautz orthogonal polynomials are non-obvious and so the use of Laguerre and Kautz bases may seem obscure or not physically motivated.

In a series of ingenious works Heuberger, Van den Hof and others [4, 5, 6, 21] have addressed this issue by showing how the same and richer classes of orthogonal basis functions may be derived on 'physical' grounds via balanced realisations of user chosen dynamics. In this paper we will refer to this construction method as the 'balanced realisation construction'. Unfortunately, it retains the limitation of only allowing the incorporation of one set of modes.

The purpose of this section is to show how a unifying construction (5) may be trivially derived as an orthonormal basis, allows the incorporation of any number of modes, and gives the Laguerre, Kautz and balanced realisation bases as special cases.

Up to this point we have been discussing ideas rather loosely. To explain our construction clearly we need to become a little more precise by introducing some mathematical formalism. To begin with, we will assume that any impulse response $\{g_k\}$ we are interested in is causal so that $g_k = 0$ for $k < 0$ and is in ℓ_2 so that $\sum_{k=0}^{\infty} |g_k|^2 < \infty$. Although such time domain considerations are physically intuitive, it is often easier to work in the frequency domain by instead considering the Fourier Series (or DFT) $G(e^{j\omega})$ given by

$$G(e^{j\omega}) = \sum_{k=0}^{\infty} g_k e^{-j\omega k}.$$

The essential link between these time and frequency domain characterisations is provided by Plancherel's relation which gives us that for $\{g_k\}, \{h_k\} \in \ell_2$

$$\sum_{k=0}^{\infty} g_k \overline{h_k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(e^{j\omega}) \overline{H(e^{j\omega})} \, \mathrm{d}\omega. \tag{6}$$

Therefore, if $\{g_k\} \in \ell_2$ then the Fourier series $G(e^{j\omega}) \in L_2([-\pi, \pi])$ and in fact the equality in (6) tells us that there is an isometric isomorphism between the representations $\{g_k\}$ and $G(e^{j\omega})$ so that we can work with either equally well.

Additionally, it is well known that (6) defines an inner product on ℓ_2 and L_2. Therefore, if we want to examine the orthogonality properties of 2 basis functions $\mathcal{B}_n(q)$ and $\mathcal{B}_m(q)$ we can do so by calculating

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

Let's see how we can use this formalism to design as set of orthonormal basis functions $\{\mathcal{B}_n\}$ for system identification.

Suppose that we suspect a pole in the true dynamics near ξ_0. Then it makes sense to have a basis function $\mathcal{B}_0$ in the model structure (1) of the form

$$\mathcal{B}_0(q) = A \frac{q^d}{q - \xi_0} \qquad ; d = 0 \text{ or } 1. \tag{8}$$

The choice of d corresponds to a simple time shift on the impulse response of $\mathcal{B}_0$ and depends on whether the user feels that a causal or strictly causal model is most appropriate. It remains to choose the constant A to achieve the normalisation $\|\mathcal{B}_0\| = 1$.

The value for A is most easily calculated by writing the inner product (7) as a contour integral around the unit circle $\mathbf{T}$ by using the change of variable $z = e^{j\omega}$,

$$\langle \mathcal{B}_n, \mathcal{B}_m \rangle = \frac{1}{2\pi j} \oint_{\mathbf{T}} \mathcal{B}_n(z) \overline{\mathcal{B}_m(z)} \, \frac{\mathrm{d}z}{z} \tag{9}$$

and then using Cauchy's Residue Theorem and the fact that on $\mathbf{T}$ we have $\bar{z} = z^{-1}$ to give

$$\|\mathcal{B}_0\|^2 = \frac{A^2}{2\pi j} \oint_{\mathbf{T}} \frac{\mathrm{d}z}{(z - \xi_0)(1 - \overline{\xi_0} z)} = \frac{A^2}{1 - |\xi_0|^2} \tag{10}$$

so that if we want a unit norm basis vector we should put $A = \sqrt{1 - |\xi_0|^2}$.

Now, suppose we also suspect another mode in the true dynamics which we can represent as a pole in the near ξ_1. Then it makes sense to include a second basis function $\mathcal{B}_1(q)$ given by

$$\mathcal{B}_1(q) = A' \frac{q^d(1 - \overline{\xi}_0 q)}{(q - \xi_0)(q - \xi_1)}. \tag{11}$$

This choice of the structure of $\mathcal{B}_1$ may seem unusual, but is explained by noting that the non-minimum phase zero at $q = 1/\overline{\xi_0}$ is included to ensure that the orthogonality condition $< \mathcal{B}_0, \mathcal{B}_1 >= 0$ is satisfied.

To see this, notice that $\mathcal{B}_0(z)$ has a pole within the unit disk $\mathbf{D}$ at ξ_0, so to ensure orthogonality of $\mathcal{B}_0$ and $\mathcal{B}_1$ we must include a zero in $\overline{\mathcal{B}_1(z)}$ at ξ_0 to cancel this pole. This follows since now $\mathcal{B}_0(z)\overline{\mathcal{B}_1(z)}$ is analytic in $\mathbf{D}$ which ensures $< \mathcal{B}_0, \mathcal{B}_1 >= 0$ by Cauchy's Integral Theorem:

$$\langle \mathcal{B}_0, \mathcal{B}_1 \rangle = \frac{\sqrt{1 - |\xi_0|^2}}{2\pi j} \oint_{\mathbf{T}} \frac{A'(z - \xi_0)}{(z - \xi_0)(1 - \overline{\xi_0} z)(1 - \overline{\xi_1} z)} \, \mathrm{d}z = 0. \tag{12}$$

Again we need to choose A' to achieve the normalisation $\|\mathcal{B}_1\|^2 = 1$:

$$\|\mathcal{B}_1\|^2 = \frac{A'^2}{2\pi j} \oint_{\mathbf{T}} \frac{(\xi_0 - z)(\overline{\xi_0} z - 1)}{(z - \xi_0)(z - \xi_1)(1 - \overline{\xi_0} z)(1 - \overline{\xi_1} z)} \, \mathrm{d}z = \frac{A'^2}{1 - |\xi_1|^2}. \tag{13}$$

Note that in the choice of $\mathcal{B}_1$ the pole at ξ_0 could in fact be chosen anywhere in the unit disk and orthogonality would be preserved. However, we have already decided in the choice of $\mathcal{B}_0$ that dynamics at ξ_0 are important. Continuing in this fashion for user chosen poles at $\{\xi_0, \xi_1, \xi_2, \cdots, \xi_{p-1}\}$ then provides our unifying construction for orthonormal basis functions:

$$\mathcal{B}_n(z) = z^d \left(\frac{\sqrt{1 - |\xi_n|^2}}{z - \xi_n} \right) \prod_{k=0}^{n-1} \left(\frac{1 - \overline{\xi}_k z}{z - \xi_k} \right) \qquad ; d = 0 \text{ or } 1. \tag{14}$$

Now, if we restrict ourselves to knowledge of only one real mode by choosing $\xi_k = \xi$ for every pole then the general construction (14) gives the Laguerre basis (3) as a special case when we also choose $d = 0$. If we go further and set $\xi = 0$ then we have the FIR basis when we choose $d = 1$.

When we wish to incorporate a complex mode we need to rethink our strategy a little. This is so since as soon as we choose one pole, say ξ_n, complex then the impulse responses for the $\{\mathcal{B}_k\}$ for $k \geq n$ become complex, and this is physically unreasonable in our system identification setting. The solution is to only include complex modes in conjugate pairs. This is achieved as follows.

Suppose n modes $\{\xi_0, \cdots, \xi_{n-1}\}$ have been included in $\{\mathcal{B}_0, \cdots, \mathcal{B}_{n-1}\}$ and we now wish to include a complex mode at ξ_n. Then two new basis functions $\mathcal{B}'_n, \mathcal{B}''_n$ with real impulse responses should be formed as linear combinations of $\mathcal{B}_n, \mathcal{B}_{n+1}$ generated by (14). These new functions then replace $\mathcal{B}_n$ and $\mathcal{B}_{n+1}$ in the model structure (1). The linear combination we are suggesting can be expressed as

$$\begin{pmatrix} \mathcal{B}'_n \\ \mathcal{B}''_n \end{pmatrix} = \begin{pmatrix} c_0 & c_1 \\ c'_0 & c'_1 \end{pmatrix} \begin{pmatrix} \mathcal{B}_n \\ \mathcal{B}_{n+1} \end{pmatrix}, \qquad c_0, c'_0, c_1, c'_1 \in \mathbf{C}. \tag{15}$$

Now it is obvious that we have preserved orthogonality in our construction in that $<\mathcal{B}'_n, \mathcal{B}_k>=0$ for $k<n$. Furthermore, since $<\mathcal{B}_n, \mathcal{B}_{n+1}>=0$ we have $\|\mathcal{B}'_n\|^2 = |c_0|^2 + |c_1|^2$. Requiring $\mathcal{B}'_n$ and $\mathcal{B}''_n$ to be unit norm then uses up two degrees of freedom in our choice of the real and imaginary parts of c_0, c'_0, c_1, c'_1. If $\mathcal{B}'_n$ and $\mathcal{B}''_n$ are to be orthogonal to one another we need $c_0\overline{c'_0} + c_1\overline{c'_1} = 0$ and this uses up one more degree of freedom. Finally, if we want the co-efficients of $\mathcal{B}'_n$ and $\mathcal{B}''_n$ to be purely real then this uses up four more degrees of freedom. But there are eight variables in the real and imaginary parts of c_0, c'_0, c_1, c'_1, so one degree of freedom remains. We should therefore be able to form an infinite number of orthonormal real co-efficient orthonormal basis vectors $\mathcal{B}'_n$ and $\mathcal{B}''_n$ by exploiting this last degree of freedom.

This in fact is the case, and is most clearly seen by first writing the unifying construction (14) in a recursive form (assume in what follows that we have selected $d=0$)

$$\mathcal{B}_n = \sqrt{\frac{1-|\xi_n|^2}{1-|\xi_{n-1}|^2}} \left(\frac{1-\overline{\xi_{n-1}}z}{z-\xi_n}\right) \mathcal{B}_{n-1}(z).$$

Using this in (15) where we choose complex poles in conjugate pairs as $\xi_{n+1} = \overline{\xi_n}$ then gives

$$\mathcal{B}'_n(z) = \sqrt{\frac{1-|\xi_n|^2}{1-|\xi_{n-1}|^2}} \mathcal{B}_{n-1}(z) \left\{ \frac{(1-\overline{\xi_{n-1}}z)(\beta z + \mu)}{z^2 - (\xi_n + \overline{\xi_n})z + |\xi_n|^2} \right\} \tag{16}$$

where the co-efficients β, μ are related to the choice of c_0, c_1 by

$$c_0 = \frac{\beta + \overline{\xi_n}\mu}{1-\overline{\xi_n}^2}, \qquad c_1 = \frac{\overline{\xi_n}\beta + \mu}{1-\overline{\xi_n}^2}.$$

Therefore, to ensure normality we must choose β and μ according to the constraint that $|c_0|^2 + |c_1|^2 = 1$ which (assuming β and μ are real) becomes

$$(1+|\xi_n|^2)(\beta^2+\mu^2) + 2(\overline{\xi_n}+\xi_n)\beta\mu = 1 + |\xi_n|^4 - (\xi_n^2 + \overline{\xi_n}^2)$$

or more compactly

$$x^T M x = |1-\xi_n^2|^2 \tag{17}$$

where

$$\begin{aligned} x &\triangleq (\beta, \mu)^T \\ M &\triangleq \begin{pmatrix} 1+|\xi_n|^2 & 2\mathrm{Re}\{\xi_n\} \\ 2\mathrm{Re}\{\xi_n\} & 1+|\xi_n|^2 \end{pmatrix}. \end{aligned}$$

The eigenvalues of M are $\lambda_1, \lambda_2 = |1 \pm \xi_n|^2 > 0$ so that choices of β and μ giving unit norm vectors lie on an ellipse. Now, suppose we make two pairs of choices $x \triangleq (\beta, \mu)$ giving a basis function $\mathcal{B}'_n$ and $y \triangleq (\beta', \mu')$ giving a basis function $\mathcal{B}''_n$. These two choices correspond to two pairs of choices $\{c_0, c_1\}$ and $\{c'_0, c'_1\}$. The requirement $c_0\overline{c'_0} + c_1\overline{c'_1} = 0$ ensuring orthogonality of $\mathcal{B}'_n$ and $\mathcal{B}''_n$ can be expressed as needing

$$x^T M y = 0 \tag{18}$$

to hold. As predicted, there is one degree of freedom remaining and hence many solutions x and y to (17) and (18) exist. To formulate them, suppose we begin by choosing any x satisfying (17). Then a y that also satisfies (17) but also satisfies (18) may be found by rotating x by ninety degrees in the normalised eigenspace of M:

$$y = M^{-1/2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} M^{1/2} x$$

or, to be more explicit, after some arithmetic we have

$$\begin{pmatrix} \beta' \\ \mu' \end{pmatrix} = \frac{1}{\sqrt{1-\alpha^2}} \begin{pmatrix} \alpha & 1 \\ -1 & -\alpha \end{pmatrix} \begin{pmatrix} \beta \\ \mu \end{pmatrix} \qquad ; \alpha \triangleq \frac{\xi_n + \overline{\xi_n}}{1 + |\xi_n|^2}. \tag{19}$$

To summarise this discussion, if we want to include complex modes in the model structure (1) then we obtain two basis vectors $\mathcal{B}'_n$ and $\mathcal{B}''_n$ from two linear combinations of $\mathcal{B}_n$ and $\mathcal{B}_{n+1}$ that come from the unifying construction (14). The first basis function $\mathcal{B}'_n$ is found as

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

where $x^T = (\beta, \mu)$ is chosen to lie anywhere on the ellipse (17). A vector $y^T = (\beta', \mu')$ is then found that also lies on the ellipse (17) by using the formula (19). The second basis function $\mathcal{B}''_n$ is then obtained as

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

These real valued impulse response basis vectors $\mathcal{B}'_n$ and $\mathcal{B}''_n$ are then used in the model structure (1) instead of $\mathcal{B}_n$ and $\mathcal{B}_{n+1}$. If we require further basis functions with complex modes then we repeat the process in (15) by forming $\mathcal{B}'_{n+1}$ and $\mathcal{B}''_{n+1}$ from linear combinations of $\mathcal{B}_{n+2}$ and $\mathcal{B}_{n+3}$ and so on.

A special case of this construction is when we have only one fixed complex mode $\xi_k = \xi$ to consider and where we make the following special choice satisfying (17)

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

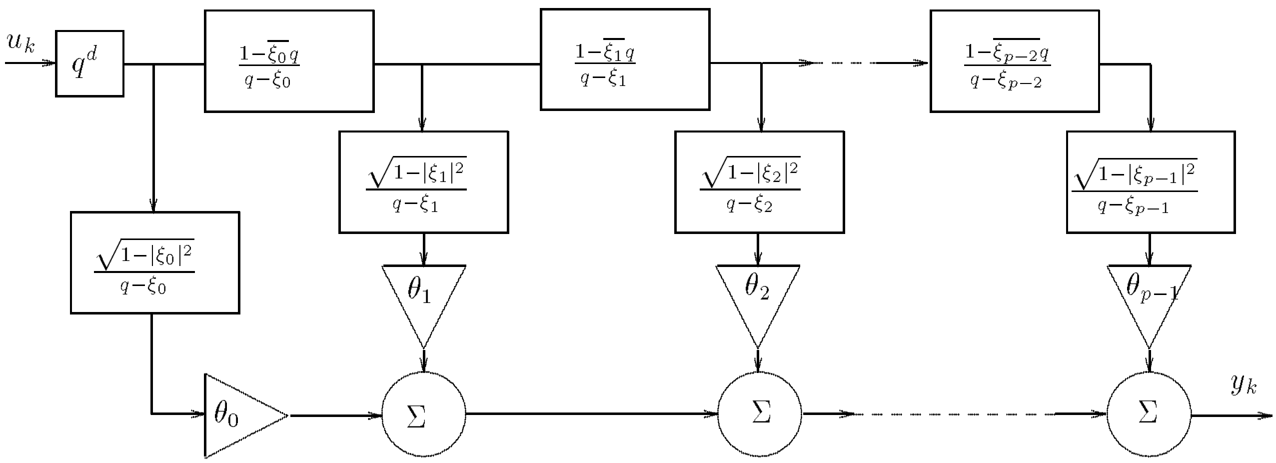

Figure 1: *Diagrammatic illustration of the use of our unifying basis.*

in which case (19) gives

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

With the initialisation $\mathcal{B}_{-1} = 1, \xi_{-1} = 0$ in (16) this gives the Kautz basis (4) if we associate $\gamma = |\xi|^2$. Different initial choices for (β, μ) satisfying (17) give an infinite number of second order bases other than the Kautz one. This is interesting, since despite this plentiful nature of second order bases, the particular case of the Kautz basis is the only instance we can find in the literature save for the balanced realisation construction [4, 5, 6, 21].

In fact, as we range through all possible choices of β and μ that satisfy (17) we range through all possible second order orthonormal bases given by the balanced realisation construction. With this latter system this corresponds to ranging through all possible balanced realisations of the all pass function

$$\frac{1 - 2\text{Re}\{\xi\}z + |\xi|^2 z^2}{z^2 - 2\text{Re}\{\xi\}z + |\xi|^2}.$$

The details of this equivalence between the balanced realisation construction and our construction are presented in section 5 and appendix B following.

Therefore, all the known orthonormal bases for system identification are special cases of our unifying construction in (14) under the restriction of fixed modes. However, we suggest that once the simple construction (8)-(13) is recognised it is unnecessary to make fixed mode restrictions, and as much prior information as possible should be incorporated into the estimation problem by using the basis choice (14) and a range of poles $\{\xi_0, \cdots, \xi_{p-1}\}$.

To finish this derivation, our unifying construction, together with the model structure (1) is illustrated diagrammatically in figure 1. The reader can compare this illustration to the corresponding figures in [23, 26, 21] in order to gain a perhaps more intuitive impression of how our unifying construction embodies previously known ones, and also extends them.

3 Gram-Schmidt Interpretation

The derivation (8)-(13) of our general basis (14) has the flavour of a Gram-Schmidt construction. With this latter scheme a set of basis vectors say $\{\mathcal{V}_n\}$ are chosen and an orthonormal set $\{\mathcal{B}_n\}$ that span the same space as $\{\mathcal{V}_n\}$ are constructed via the well known recursion

$$\mathcal{W}_n = \mathcal{V}_n - \sum_{k=0}^{n-1} \langle \mathcal{V}_n, \mathcal{B}_k \rangle \mathcal{B}_k \tag{20}$$

$$\mathcal{B}_n = \|\mathcal{W}_n\|^{-1/2} \mathcal{W}_n. \tag{21}$$

Indeed, any suspicion that a Gram-Schmidt construction has been performed in (8)-(13) is correct.

Lemma 1. *The general basis functions we arrived at in (14) can also be derived using the Gram-Schmidt procedure (20)-(21) with the initial basis vector choice $\{\mathcal{V}_n\}$ as the ones we suggested in (2):*

$$\mathcal{V}_n(q) = \frac{1}{q - \xi_n}$$

Proof. See Appendix A. □□□

4 Completeness

The question now arises as to whether the unifying basis functions suggested in (14) are 'complete'. That is, do they span a comprehensive space of physically reasonable frequency responses? Intuitively one would expect Span$\{\mathcal{B}_0, \cdots, \mathcal{B}_n, \cdots\}$ to be 'rich' since each $\mathcal{B}_k(e^{j\omega})$ can be identified in a 1-1 fashion with the standard basis $e^{-j\omega k}$ via a Fourier decomposition.

Unfortunately, intuition is often false in infinite dimensional spaces. Specifically, a 1-1 map from an infinite dimensional space to itself is not necessarily onto so that the completeness question is not trivial.

Nevertheless, in this section we answer the completeness questions as follows. If there are infinitely many poles $\{\xi_k\}$ selected strictly inside the stability boundary $|z| = 1$ then our basis functions (14) span all causal systems with square summable impulse responses. Furthermore, this is a necessary and sufficient condition for completeness.

In order to state and prove this precisely it is necessary to associate such causal ℓ_2 impulse responses with a Hardy Space of frequency responses since this lets us draw on known results on Blaschke products. The only difficulty with this is that the definition of a Hardy space is tied to the classical definition of a Fourier series which is an expansion in terms of $\{e^{j\omega n}\}$ instead of the basis $\{e^{-j\omega n}\}$ which spans the frequency responses of causal systems. With this difficulty in mind we begin by noting that by definition a function f is in $L_2([-\pi, \pi])$ if

$$\int_{-\pi}^{\pi} |f(\omega)|^2 \, d\omega < \infty$$

in which case it can be written as a Fourier series

$$f(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{j\omega k} \tag{22}$$

where equality is in the sense of norm convergence. Since f can be written in terms of $e^{j\omega}$, then f can also be considered as defined on the circle $\mathbf{T}$ so that $f \in L_2(\mathbf{T})$. Now, if the $\{c_k\}$ happen to be the impulse response co-efficients of a discrete time system, then $f(-\omega)$ is the frequency response of the system and furthermore the Fourier co-efficients $\{c_k\}$ can be calculated from $f(\omega)$ as

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)e^{-j\omega k}\, \mathrm{d}\omega. \tag{23}$$

The Hardy space $H_2(\mathbf{T})$ is formally defined [3] as the closed subspace of $L_2(\mathbf{T})$ on which the elements have Fourier co-efficients which are zero for negative k. That is, $c_k = 0$ for $k < 0$.

Given this pre-amble, if we are only interested in stable and causal systems, then we are only interested in systems whose frequency responses $f(\omega)$ evaluated at $-\omega$ are in $H_2(\mathbf{T})$, and with a slight abuse of terminology we will refer to this as a search for a basis for $H_2(\mathbf{T})$.

The question of the richness of the basis functions in (14) can then be phrased abstractly as a question of their completeness in $H_2(\mathbf{T})$. This is answered with a necessary and sufficient condition on the choice of the poles $\{\xi_k\}$ as follows.

Theorem 1. *Consider the basis functions $\mathcal{B}_k(e^{j\omega})$ defined in (14). Then* $\mathrm{Span}\{\mathcal{B}_k(e^{-j\omega})\}$ *is dense in $H_2(\mathbf{T})$ if and only if*

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

Proof. See Appendix C. □□□

Of course, what we really require is completeness of the impulse responses of the $\{\mathcal{B}_n\}$ in the signal space of causal ℓ_2 sequences which are also real valued. But Theorem 1 implies this since it proves completeness in a larger class.

To be specific, by Theorem 1 any sequence in the signal space can be made up as a possibly complex linear combination of the impulse responses of the $\{\mathcal{B}_n\}$. If any of these impulse responses are complex valued, then they may be written as a complex linear combination of the real valued impulse responses of $\mathcal{B}'_n$ and $\mathcal{B}''_n$ introduced in the previous section. Therefore, by Theorem 1 equating real and imaginary parts shows that any signal space impulse response can be written as a real linear combination of the real impulse responses of our basis functions.

5 Relationship to Construction Methods using Balanced Realisations

The first authors to generalise the construction of orthonormal bases were Heuberger, Van den Hof and co-workers [4, 5, 6, 21]. Here we provide our interpretation of their method in order to show how it is subsumed by the unifying construction (14).

Suppose a basis with modes at $\{\xi_0, \cdots, \xi_{n-1}\}$ is desired. Then form an all pass prototype function $G(z)$ which has poles at $\{\xi_0, \cdots, \xi_{n-1}\}$ and has state space description

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k. \end{aligned}$$

In this case, the $\mathcal{Z}$ transform $X(z)$ of the state sequence $\{x_k\}$ is related to the $\mathcal{Z}$ transform $U(z)$ of the input sequence $\{u_k\}$ as

$$X(z) = (zI - A)^{-1}BU(z) = V(z)U(z) \tag{24}$$

where we write the vector $V(z)$ as

$$V(z) = \frac{N(z)}{d(z)}$$

with $N(z)$ being a vector of polynomials in z and $d(z)$ being

$$d(z) = \prod_{k=0}^{n-1}(z - \xi_k).$$

Now suppose that $\{u_k\}$ is a zero mean unit variance white noise sequence. Then by Parseval's Theorem

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} V(e^{j\omega})V^\star(e^{j\omega})\, \mathrm{d}\omega = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}\left\{x_k x_k^T\right\} \triangleq P.$$

But, assuming x_0 is zero we have

$$x_k = \sum_{m=0}^{k-1} A^{n-m-1}Bu_m$$

so that

$$\begin{aligned} \frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}\left\{x_k x_k^T\right\} &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{m=0}^{k-1}\sum_{r=0}^{k-1} A^{n-m-2}B\mathbf{E}\left\{u_m u_r\right\}B^T(A^T)^{n-r-2} \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{m=0}^{k-1} A^{n-m-2}BB^T(A^T)^{n-m-2} \\ &= \sum_{k=0}^{n-1}\left(1 - \frac{k}{n}\right)A^kBB^T(A^T)^k. \end{aligned}$$

Therefore, by the properties of Césaro means [10], provided all the poles $\{\xi_k\}$ satisfy $|\xi_k| < 1$ then P can be written as

$$P = \sum_{k=0}^{\infty} A^k BB^T (A^T)^k$$

which is better known as the controllability Grammian satisfying the Lyapanov equation

$$APA^T + BB^T = P. \tag{25}$$

The observability Grammian Q may be similarly defined as the solution of

$$A^T QA + C^T C = Q. \tag{26}$$

The key to the balanced realisation scheme is to notice that if (A, B, C, D) is a balanced realisation of G then by definition $P = Q =$ diagonal. But since $G(z)$ is all-pass, then it is also true that $PQ = I$ so that in fact $P = Q = I$. In this case

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} V(e^{j\omega})V^\star(e^{j\omega})\,\mathrm{d}\omega = I$$

which means that the n transfer functions in the vector $V(z)$ are orthonormal to one another. Furthermore, if we consider

$$V_m(z) = V(z)G^m(z) = \frac{N(z)}{d(z)}\left(\frac{d(1/z)}{d(z)}\right)^m$$

then since $|G(e^{j\omega})|^m = 1$ we have $V_m V_m^\star = VV^\star$ so that the n transfer functions in $V_m(z)$ are also orthonormal to one another. Finally

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} V_m(e^{j\omega})V_n^\star(e^{j\omega})\,\mathrm{d}\omega = \frac{1}{2\pi j}\oint_{\mathbf{T}} \frac{N(z)N(1/z)^T}{d(z)d(1/z)}\left(\frac{d(z)}{d(1/z)}\right)^{m-n}\frac{\mathrm{d}z}{z}$$

so that if $m > n$ then $V_m - V_n$ since the integrand is analytic inside $\mathbf{T}$ and hence the integral is zero.

In summary then, one takes an all pass prototype function $G(z)$ with poles $\{\xi_0, \cdots, \xi_{n-1}\}$ as desired, finds a balanced realisation (A, B, C, D) of $G(z)$, forms the vector $V(z)$ as in (24) of n orthonormal basis functions and then extends this to an infinite set of orthonormal basis functions by forming $V(z)G^m(z)$.

The question now arises of how this neat construction method relates to our unified formulation (14). In fact it is a special case of (14) where the modes are restricted to a finite set $\{\xi_0, \cdots, \xi_{n-1}\}$ instead of being allowed to be extended indefinitely.

To see this, note that for the case of the all-pass prototype $G(z)$ being first order the balanced realisation method generates the Laguerre basis [4] which is a special case of our unifying construction (14) when one fixed real mode $\xi_k = \xi$ is chosen. When $G(z)$ is second order, then as detailed in appendix B the balanced realisation method is again a special case of our unifying construction method, this time when one fixed complex mode is chosen.

For higher order $G(z)$ the equivalence is in the sense of linear combinations. Take the third order case as an example. The balanced realisation method provides the first three basis functions as all third order, whereas our unifying basis provides either a first order and then two third order basis functions, or two second order functions and then a third order one. Either way, by a simple partial fraction argument, the first three balanced realisation bases are a linear combination of the first three basis vectors of the unifying construction (14). Continuing in this fashion for higher order $G(z)$ shows that the balanced realisation bases are subsumed by the unifying formulation (14).

6 Links to Classical Orthogonal Polynomials

The basis vectors $\{\mathcal{B}_k\}$ given in (3), as we've already mentioned, have been extensively studied under the title of 'Laguerre Models'. They inherit this name from the classical Laguerre orthonormal polynomials from which they are derived. This raises an obvious question. Can orthonormal bases suitable for system identification be generated from the other classical orthogonal polynomials such as Legendre, Chebychev and Hermite which are more commonly used for the solution of certain partial differential equations?

In order to answer this question we need to chart the course of the Laguerre polynomial to discrete basis development in order to ascertain if the same template can be successfully applied to other polynomial systems.

This development seems to be shrouded in secrecy in the literature. Here we give our interpretation which we have not been able to find elsewhere. To begin with, the classical Laguerre polynomials [19] denoted $L_n(x)$ which most commonly arise in the solution of Schrödingers wave equation are defined by

$$L_n(x) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \begin{pmatrix} n \\ k \end{pmatrix} x^k \tag{27}$$

and satisfy the orthogonality condition

$$\int_0^\infty e^{-x} L_n(x) L_m(x)\, \mathrm{d}x = \begin{cases} 1 & ; n = m \\ 0 & ; n \neq m \end{cases} \tag{28}$$

so that the so-called 'Laguerre functions' given by $\psi_n(x) \triangleq e^{-x/2} L_n(x)$ are $L_2([0,\infty))$ orthonormal. Using a change of variable $x = 2\alpha t, \alpha > 0$ in (28) and Parseval's Theorem then gives

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\psi}_n\left(\frac{\omega}{2\alpha}\right) \overline{\widehat{\psi}_m\left(\frac{\omega}{2\alpha}\right)}\, \mathrm{d}\omega = \begin{cases} 2\alpha & ; n = m \\ 0 & ; n \neq m \end{cases} . \tag{29}$$

where we have used the $\widehat{\cdot}$ notation to denote Fourier Transform. Now, using (27), the Binomial Theorem and the fact that the Laplace Transform of $x^n e^{-ax}$ is given by $n!(s +$

$a)^{-(n+1)}$ we are led to

$$\widehat{\psi}_n\left(\frac{\omega}{2\alpha}\right) = \sum_{k=0}^{n}\left(\begin{array}{c} n \\ k \end{array}\right)\frac{(-1)^k(2\alpha)^{k+1}}{(j\omega+\alpha)^{k+1}} = \frac{2\alpha}{(j\omega+\alpha)}\left[1-\frac{2\alpha}{(j\omega+\alpha)}\right]^n = 2\alpha\frac{(j\omega-\alpha)^n}{(j\omega+\alpha)^{n+1}}. \tag{30}$$

Therefore, by (29) the 'continuous time' basis functions

$$\mathcal{B}'(j\omega) = \frac{\sqrt{2\alpha}}{(j\omega+\alpha)}\left(\frac{j\omega-\alpha}{j\omega+\alpha}\right)^n \tag{31}$$

are $H_2(\mathbf{R})$ orthonormal. Finally, if we use the Bilinear transformation which is defined by $j\omega = (e^{j\varphi}-1)(e^{j\varphi}+1)^{-1} \triangleq M(\varphi)$ and which maps the imaginary axis to $\mathbf{T}$ we get an $H_2(\mathbf{T})$ orthogonal set of functions:

$$\delta(m-n) = \frac{1}{2\pi j}\int_{-\pi}^{\pi}\mathcal{B}'_n(M(\varphi))\overline{\mathcal{B}'_m(M(\varphi))}\frac{\mathrm{d}M(\varphi)}{\mathrm{d}\varphi}\,\mathrm{d}\varphi = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{B}_n(e^{j\varphi})\overline{\mathcal{B}_m(e^{j\varphi})}\,\mathrm{d}\varphi \tag{32}$$

where

$$\mathcal{B}_n(e^{j\varphi}) = \frac{\sqrt{1-\xi^2}}{(e^{j\varphi}-\xi)}\left(\frac{\xi e^{j\varphi}-1}{e^{j\varphi}-\xi}\right)^n \qquad ;\xi \triangleq \frac{1-\alpha}{1+\alpha} \tag{33}$$

which are the basis functions we arrived at much more simply in (14) with the special choice $\xi_k = \xi$ for every k.

Actually, to be completely precise, the careful reader will notice that (33) does not come from (14) since the sign of the d.c. gain in (33) is negative for n odd, and this is not the case in (14). Nevertheless, (14) with $\xi_k = \xi$ is the construction which has become known in the literature as the 'Laguerre Basis'. The fact that it differs slightly from (33) merely highlights that this accepted terminology is a trivial misnomer.

Having recognised this construction of discrete time bases from classical orthogonal polynomials we can use the same analysis as for the Laguerre case but begin with the use of Legendre polynomials which are useful in describing the solution of Laplace's equation in the sphere and are defined as satisfying the Rodrigues formula [18]:

$$P_n(x) \triangleq \frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(x^2-1)^n \tag{34}$$

and which satisfy the orthogonality property

$$\int_{-1}^{1} P_n(x)P_m(x)\,\mathrm{d}x = \left\{\begin{array}{cl} 2(2n+1)^{-1} & ;n=m \\ 0 & ;n\neq m \end{array}\right. .$$

Therefore, using the substitution $x = 2e^{-\alpha t}-1, \alpha > 0$ we obtain

$$\int_{0}^{\infty} P_n(2e^{-\alpha t}-1)P_m(2e^{-\alpha t}-1)e^{-\alpha t}\,\mathrm{d}t = \left\{\begin{array}{cl} [\alpha(2n+1)]^{-1} & ;m=n \\ 0 & ;m\neq n \end{array}\right.$$

so that the 'Legendre functions' $p_n(t)$ defined by

$$p_n(t) \triangleq \sqrt{\alpha(2n+1)}e^{-\alpha t/2}P_n(2e^{-\alpha t}-1)$$

are $L_2([0,\infty))$ orthonormal. But from (34)

$$\begin{aligned} p_n(t) &= \frac{\sqrt{\alpha(2n+1)}e^{-\alpha t/2}}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x(t)^n}\left(x^2(t)-1\right)^n \quad ; x(t) \triangleq 2e^{-\alpha t}-1 \\ &= \frac{\sqrt{\alpha(2n+1)}e^{-\alpha t/2}}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x(t)^n}\sum_{k=0}^{n}\left(\begin{array}{c} n \\ k \end{array}\right) x^{2k}(t)(-1)^{n-k} \\ &= \frac{\sqrt{\alpha(2n+1)}e^{-\alpha t/2}}{2^n}\sum_{m=0}^{[n/2]}\left(\begin{array}{c} n \\ m \end{array}\right)\left(\begin{array}{c} 2n-2m \\ n \end{array}\right)(-1)^m x^{n-2m} \\ &= \sqrt{\alpha(2n+1)}\sum_{m=0}^{[n/2]}\sum_{r=0}^{n-2m}\left(\begin{array}{c} n \\ m \end{array}\right)\left(\begin{array}{c} 2n-2m \\ n \end{array}\right)\left(\begin{array}{c} n-2m \\ r \end{array}\right)(-1)^{n-m-r}2^{r-n}e^{-\alpha t(r+1/2)} \end{aligned}$$

so that

$$\widehat{p}_n(\omega) = \sqrt{\alpha(2n+1)}\sum_{m=0}^{[n/2]}\sum_{r=0}^{n-2m}\left(\begin{array}{c} n \\ m \end{array}\right)\left(\begin{array}{c} 2n-2m \\ n \end{array}\right)\left(\begin{array}{c} n-2m \\ r \end{array}\right)\frac{(-1)^{n-m-r}2^{r-n}}{j\omega+\alpha(r+1/2)}.$$

This is a very complicated expression, but in fact it is the partial fraction expansion of the following formula which is given in [11]

$$\widehat{p}_n(\omega) = \frac{\sqrt{\alpha(2n+1)}}{j\omega+(n+1/2)\alpha}\prod_{k=0}^{n-1}\left(\frac{j\omega-(k+1/2)\alpha}{j\omega+(k+1/2)\alpha}\right). \tag{35}$$

So, as for the Laguerre case, using Parseval's Theorem we have that $\widehat{p}_n(\omega)$ given by (35) is an orthonormal basis for $H_2(\mathbf{R})$, and again by using the Bilinear transform as for the Laguerre basis we get the Legendre orthonormal basis for $H_2(\mathbf{T})$:

$$\mathcal{B}_{n-1}(e^{j\omega}) = \frac{\sqrt{1-\xi_n^2}}{e^{j\omega}-\xi_n}\prod_{k=0}^{n-1}\left(\frac{1-\xi_k e^{j\omega}}{e^{j\omega}-\xi_k}\right) \tag{36}$$

$$\xi_k \triangleq \frac{2-\alpha(2k+1)}{2+\alpha(2k+1)} \tag{37}$$

which we could have obtained much more simply from the general construction (14) with the substitution for ξ_k given by (37).

We can find no reference to these Legendre functions being used for system identification even though they would intuitively seem more useful than the popular Laguerre basis functions since they have a progression of the pole position and hence allow approximation of a range of modes. They are used for continuous time system synthesis in the seminal work of Wiener and Lee [11].

Having examined the method of passing from orthogonal polynomials to $H_2(\mathbf{T})$ basis functions, we now show that bases generated from other classical orthogonal polynomials will not be feasible since the orthonormal bases generated will not by rational in $e^{j\omega}$. This is due to the nature of the kernel in the orthogonality representation.

To be more explicit, the Laguerre and Legendre polynomials are special cases of the most general class of orthogonal polynomials [19] which may be found by orthonormalising $\{1, t, t^2, t^3, \cdots\}$ using Gram-Schmidt and the inner product definition

$$\langle p(t), m(t) \rangle = \int_{-\infty}^{\infty} p(t)m(t)K(t)\,\mathrm{d}t$$

where $p(t)$ and $m(t)$ are polynomials and $K(t)$ is a positive definite function we call the 'kernel function'. Different choices of kernel give different sets of polynomials, some of which occur so commonly as to be given special names such as 'Laguerre' and 'Legendre'. These latter two correspond to kernel choices of

$$K(t) = \begin{cases} e^{-t} & ; t \geq 0 \\ 0 & ; \text{Otherwise} \end{cases}$$

for Laguerre and

$$K(t) = \begin{cases} 1 & ; |t| \leq 1 \\ 0 & ; \text{Otherwise} \end{cases}$$

for Legendre. The crucial point is that both these kernels have Laplace transforms which are rational in s so that the procedure we have just illustrated of turning orthogonal polynomials into orthonormal discrete time bases yields functions $\{\mathcal{B}_k(q)\}$ that are rational in q.

In contrast, if we take the Hermite and Chebychev polynomials as examples, they have orthogonality kernels $K(t)$ of e^{-t^2} and $(1-t^2)^{-1/2}$ respectively, whose Laplace transforms are not rational in s. This will lead to bases not rational in q and hence not implementable with finite dimensional filters. Such basis functions are therefore, for practical intents, ruled out.

In fact, there seems to be no other classical orthogonal polynomials other than Laguerre and Legendre with appropriate kernels. Other non-classical polynomials such as Kautz can, of course, be generated with appropriate choice of $K(t)$, but they will lead to basis functions that can be more directly derived by just using the unifying formulation (14).

7 System Identification using the General Basis

The original ambit of this paper was to attack a system identification problem. We are now finally ready to address this after having examined the choice of basis functions $\{\mathcal{B}_k\}$ to be used in the model structure (1) that we intend to employ.

The problems we are interested in are ones in which N point data records of an input sequence $\{u_k\}$ and output sequence $\{y_k\}$ of a linear time invariant system are available.

We assume this data is generated as follows

$$y_k = G_T(q)u_k + H(q)\nu_k.$$

Here $G_T(q)$ is a stable (unknown) transfer function describing the system dynamics that we wish to identify by using $\{u_k\}$ and $\{y_k\}$. Unfortunately $\{y_k\}$ is possibly noise corrupted by a zero mean stationary white noise process $\{\nu_k\}$ which has finite variance $\sigma_\nu^2 = \mathbf{E}\{\nu_k^2\}$. In fact, this noise corruption may be coloured by the stable filter $H(q)$. Finally we assume that $\{u_k\}$ is a quasi-stationary sequence [12] with spectral density $\Phi_u(\omega)$.

The method of estimating the dynamics $G_T(q)$ that we wish to examine is the least squares method that we mentioned in the introduction. To be specific, we examine the use of the model structure $G(q,\theta)$ given in (1)

$$G(q,\theta) = \sum_{k=0}^{p-1} \theta_k \mathcal{B}_k(q). \tag{38}$$

where $\theta \triangleq [\theta_0, \cdots, \theta_{p-1}]$ is a vector of parameters. We take $G(q,\widehat{\theta})$ as out estimate of the dynamics $G_T(q)$ where the vector $\widehat{\theta}$ is found by minimising the squared error

$$\widehat{\theta} = \arg\min_{\theta} \left\{ \frac{1}{N} \sum_{k-0}^{N-1} \left(y_k^f - G(q,\theta)u_k^f \right)^2 \right\} \tag{39}$$

and $\{y_k^f\}$, $\{u_k^f\}$ are filtered versions of the observed data $\{y_k\}$ and $\{u_k\}$

$$y_k^f = F(q)y_k, \qquad u_k^f = F(q)u_k.$$

The filter $F(q)$ is user chosen and stable. The solution $\widehat{\theta}$ to (39) is well known once the model structure (38) is cast in more familiar linear regressor form notation as

$$G(q,\theta)u_k^f = \phi_k^T\theta.$$

Here

$$\phi_k^T = [\mathcal{B}_0(q)u_k^f, \cdots, \mathcal{B}_{p-1}u_k^f] \tag{40}$$

and (39) then has the closed form solution

$$\widehat{\theta} = \left(\sum_{k=0}^{N-1} \phi_k \phi_k^T \right)^{-1} \sum_{k=0}^{N-1} \phi_k y_k \tag{41}$$

provided that the indicated inverse exists. The estimated frequency response is taken as $G(e^{j\omega},\widehat{\theta})$. The purpose of the rest of the paper is to examine how the choice of the basis functions $\{\mathcal{B}_k\}$ affects the accuracy of the estimate $G(e^{j\omega},\widehat{\theta})$.

8 Accuracy of Estimate

Given the experimental conditions we have assumed, if $|F(e^{j\omega})|^2\Phi_u(\omega)$ is never zero and Lipschitz continuous then we can use the results of [13] to conclude that as the data length N grows the estimate $\widehat{\theta}$ converges as

$$\widehat{\theta} \xrightarrow{\text{a.s.}} \theta_\star \quad \text{as } N \to \infty$$

where $\theta_\star$ minimises a weighted quadratic norm error

$$\theta_\star = \arg\min_\theta \left\{ \int_{-\pi}^{\pi} |G_T(e^{j\omega}) - G(e^{j\omega},\theta)|^2 |F(e^{j\omega})|^2 \Phi_u(\omega)\, \mathrm{d}\omega \right\}. \tag{42}$$

Therefore, if the basis function set $\{\mathcal{B}_k\}$ is H_2 complete we can make the asymptotic in N estimation error as small as we like by making the model order p sufficiently large.

Of course, in practice we never have the luxury of infinite data. Our finite data estimate $\widehat{\theta}$ will therefore have a 'variance' error component affecting the accuracy of our estimate. It is due to the noise corrupting process $\{\nu_k\}$ and prevents us from making p arbitrarily large.

This phenomenon is part of the folk wisdom of system identification that the variance error component is inversely proportional to the data length N and proportional to the model order p. The net result is that for a given N there is an optimal p for which the variance error due to noise is balanced against the so called 'bias error' that results from a parsimonious model structure. At this point the estimation error, by some measure, is a minimum. This is called the bias-variance tradeoff.

Rigorous results have been obtained in this area for FIR models by Ljung and Yuan [15], for ARX models by Ljung [14] for Laguerre models by Wahlberg [23] and for the balanced realisation construction [21]. These results are all of the nature that asymptotically in both N and p the estimation error due to noise is proportional to p/N times a signal to noise ratio term.

In the following Theorem we provide the same results for our unifying basis. This encapsulates the previously known variance error results [15, 23, 21] as special cases. Furthermore, the result is **not** asymptotic in the model order p, although this is obtained at a cost of not providing strict convergence results except in special cases.

Theorem 2. *Using the model structure (1) with $H_2(\mathbf{T})$ orthonormal basis functions $\{\mathcal{B}_k\}$ and least squares estimation the asymptotic variance in the frequency response estimate satisfies*

$$\lim_{N\to\infty} N\mathsf{E}\left\{ |G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2 \right\} \geq \underline{\gamma} \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2. \qquad \textit{w.p.1.} \tag{43}$$

Furthermore, in the special case where we have white measurement noise which corresponds to $|H(e^{j\omega})|^2 = 1$ then we have an upper bound as well

$$\underline{\gamma} \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2 \leq \lim_{N\to\infty} N\mathsf{E}\left\{ |G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2 \right\} \leq \overline{\gamma} \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2 \qquad \textit{w.p.1.} \tag{44}$$

In both cases we have

$$\underline{\gamma} \triangleq \sigma_\nu^2 \min_{\varphi\in[-\pi,\pi]} \left\{ \frac{|H(e^{j\varphi})|^2}{|F(e^{j\varphi})|^2 \Phi_u(\varphi)} \right\}, \qquad \overline{\gamma} \triangleq \sigma_\nu^2 \max_{\varphi\in[-\pi,\pi]} \left\{ \frac{|H(e^{j\varphi})|^2}{|F(e^{j\varphi})|^2 \Phi_u(\varphi)} \right\}.$$

Proof. See Appendix E. □□□

The bounds in (43) and (44) are tight since for $\{\mathcal{B}_n(e^{j\omega})\}$ an FIR basis (43) leads to

$$\lim_{N\to\infty} \frac{N}{p} \mathbf{E}\left\{ |G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2 \right\} \geq \min_{\varphi\in[-\pi,\pi]} \left\{ \frac{|H(e^{j\varphi})|^2\sigma_\nu^2}{|F(e^{j\varphi})|^2 \Phi_u(\varphi)} \right\} \quad \text{w.p.1.} \tag{45}$$

which can be compared to the convergence result of Ljung [15]

$$\frac{N}{p} \mathbf{E}\left\{ |G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2 \right\} \xrightarrow{\text{a.s.}} \frac{|H(e^{j\omega})|^2\sigma_\nu^2}{|F(e^{j\omega})|^2 \Phi_u(\omega)} \quad \text{as } N, p \to \infty. \tag{46}$$

Also, for $\{\mathcal{B}_n(e^{j\omega})\}$ a Laguerre basis (43) leads to

$$\lim_{N\to\infty} \frac{N}{p} \mathbf{E}\left\{ |G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2 \right\} \geq \min_{\varphi\in[-\pi,\pi]} \left\{ \frac{|H(e^{j\varphi})|^2\sigma_\nu^2}{|F(e^{j\varphi})|^2 \Phi_u(\varphi)} \right\} \left(\frac{1-\xi^2}{|1-\xi e^{-j\omega}|^2} \right) \quad \text{w.p.1.} \tag{47}$$

which can be compared to Wahlberg's result [23]

$$\frac{N}{p} \mathbf{E}\left\{ |G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2 \right\} \xrightarrow{\text{a.s.}} \frac{|H(e^{j\omega})|^2\sigma_\nu^2}{|F(e^{j\omega})|^2 \Phi_u(\omega)} \left(\frac{1-\xi^2}{|1-\xi e^{j\omega}|^2} \right) \quad \text{as } N, p \to \infty \tag{48}$$

Notice the important point that the convergence results we are comparing our bounds to are asymptotic in both the data length N **and** the model order p whereas our bounds hold for a fixed and finite model order p.

Furthermore in the special case case of white input as well as white measurement noise, that is $|F|^2\Phi_u = \sigma_u^2 =$ constant and $|H|^2 = 1$, Theorem 2 does provide a convergence result that is **not** asymptotic in p and gives (46) and (48) as special cases:

$$N\mathbf{E}\left\{ |G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2 \right\} \xrightarrow{\text{a.s.}} \frac{\sigma_\nu^2}{\sigma_u^2} \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2 \quad \text{as } N \to \infty.$$

In summary, Theorem 2 indicates that the variance error in the estimate at a particular frequency may be minimised by including as few basis functions as possible with significant response at that frequency. Of course, one should also keep in mind that this strategy could result in increased bias error. This tradeoff will be illustrated in the next section.

9 Simulation Examples

We believe that our general construction (14) is useful for two reasons. Firstly, it provides a unifying construction embodying all known bases in a physically intuitive fashion. Secondly, the general construction (14) can provide bases that are richer than the known ones.

They are richer in the sense that more prior information can be injected into the system identification problem by the incorporation of a variety of modes in the orthonormal basis functions. This is in contrast to the already known bases which only allow the inclusion of prior information about one mode.

In this section we will illustrate the utility of this richness by presenting some simulation examples. To begin with, suppose we have an underlying continuous time system

$$G_T(s) = \frac{e^{-2s}}{(s+1)(10s+1)} \tag{49}$$

which we sample with period one second. The zero order hold equivalent discrete time model relating output samples to input samples then is

$$G_T(q) = \frac{q^{-2}(0.0355q + 0.0247)}{(q-0.9048)(q-0.3679)}. \tag{50}$$

Suppose also that our mission is to estimate the dynamics (50) on the basis of observing $N = 500$ samples of the output $\{y_k\}$ of G_T when the input $\{u_k\}$ is a unit amplitude square wave of fundamental frequency 0.02Hz. Finally, suppose we have prior knowledge about the time constants in (49) that is accurate to within 30 per cent. That is, out prior knowledge is that the time constants are 7 seconds and 1.3 seconds.

Now there are a range of options as to how we might complete our system identification mission, but let us investigate the use of least squares estimation (40),(41) using the model structure (1) and the unifying basis (14) so that we can take advantage of our prior information about the modes of the system.

To begin with we choose all the poles $\{\xi_k\}$ the same and reflecting the prior information about a 7 second time constant. That is we put $\xi_k = \xi = e^{-1/7}$ so that a Laguerre model structure is employed that incorporates prior information about the slowest mode. The results of these choices are shown in the left hand plot of figure 2 where we have chosen the model order $p = 3$. If we use our general basis (14) which allows the incorporation of the prior knowledge of **both** modes with the choice $\xi_0 = \xi_1 = e^{-1/7}, \xi_2 = e^{-1/1.3}$ then the result is shown in the right hand plot of figure 2. As can be seen, using the general basis with a range of modes can result in much more accurate modelling when compared to the more common Laguerre basis.

Now let us increase the difficulty of the problem by corrupting the observed output data $\{y_k\}$ with stationary and white Gaussian distributed noise of variance $\sigma_\nu^2 = 0.005$. Let us also reduce the amount of observed data to 50 samples. In this case, the results are presented in figure 3, where the information presented differs from figure 2 in that

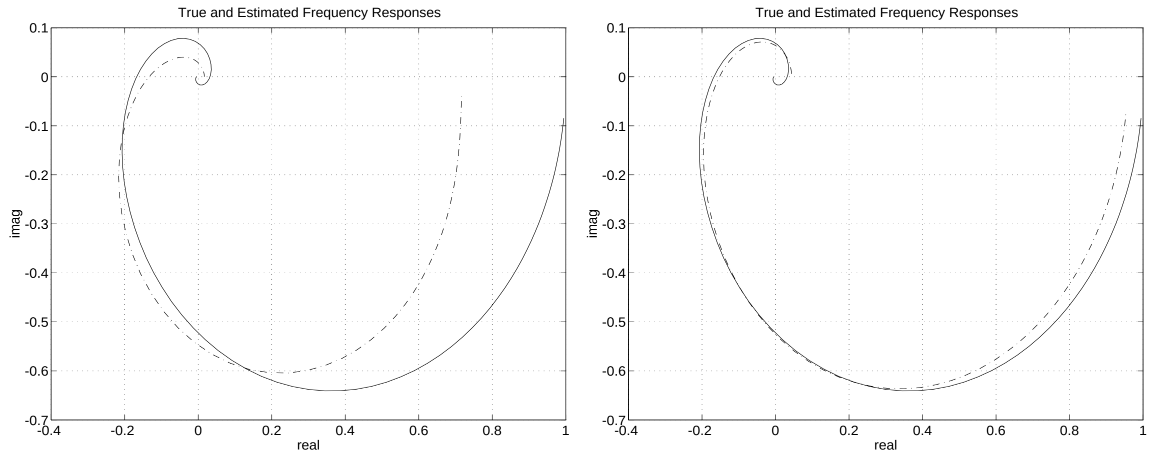

Figure 2: *Comparison of true and estimated frequency responses. Solid Line is true Nyquist, dash-dot line is the estimate. On the left a Laguerre model with all modes at 7 seconds has been used. On the right a generalised basis model with modes at 7 and* 1.3 *seconds has been used.*

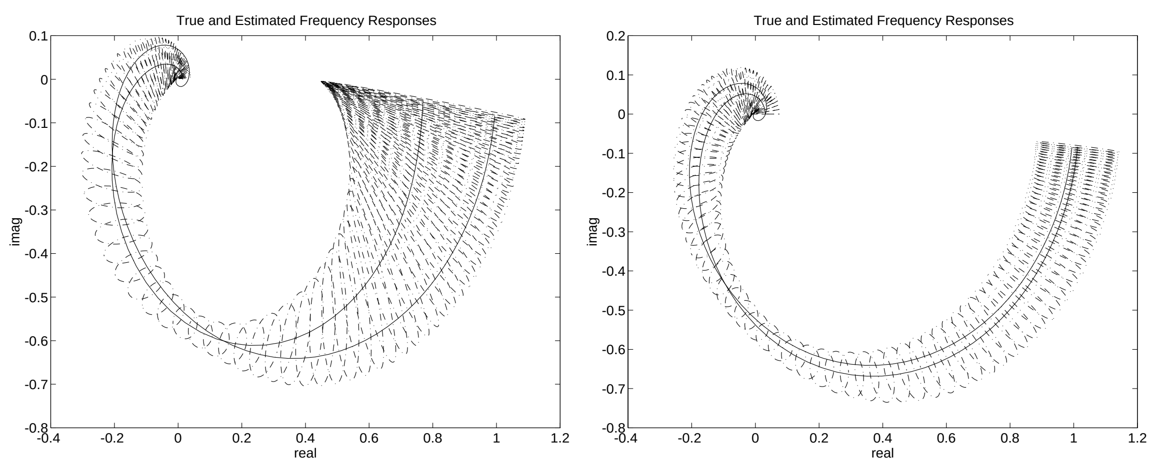

Figure 3: *This shows the same information as in the previous figure save that measurement noise has been introduced into the simulation and hence* 99.5% *confidence ellipses are shown on these diagrams*

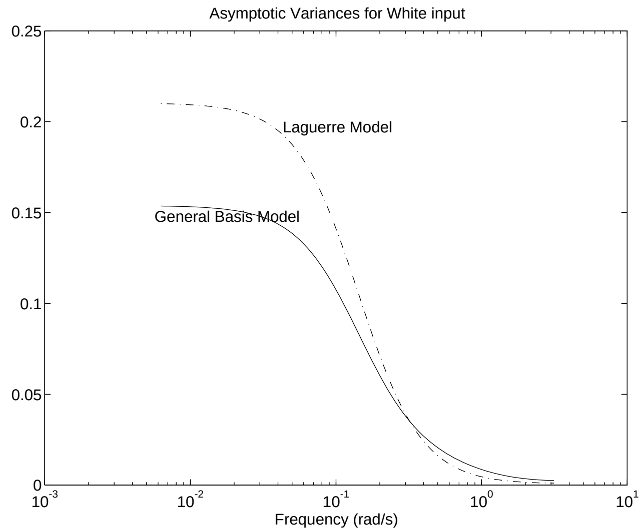

Figure 4: *Plot of* $\sigma_\nu^2 \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2$ *for basis functions used in previous two figures. The solid line is for the general basis with distributed modes. The dash-dot line is for the Laguerre basis.*

---

ellipsoidal 99.5% confidence regions on the estimated frequency response have been drawn. For the model structure (1) we are using these may be calculated as

$$?^T(\omega) P^{-1} ?(\omega) \leq \lambda$$

where

$$\begin{aligned} ?^T(\omega) &\triangleq \begin{pmatrix} \mathrm{Re}\{\mathcal{B}_0(e^{j\omega})\}, & \cdots & ,\mathrm{Re}\{\mathcal{B}_{p-1}(e^{j\omega})\} \\ \mathrm{Im}\{\mathcal{B}_0(e^{j\omega})\}, & \cdots & ,\mathrm{Im}\{\mathcal{B}_{p-1}(e^{j\omega})\} \end{pmatrix} \\ P &\triangleq \mathbf{E}\left\{\widehat{\theta}\widehat{\theta}^T\right\} = \sigma_\nu^2 \left(\sum_{k=0}^{N-1} \phi_k \phi_k^T\right)^{-1}. \end{aligned}$$

The constant λ is chosen as 12.85 to give 99.5% confidence regions by recognising that $?^T P^{-1} ?$ is χ^2 distributed with $p = 3$ degrees of freedom.

Notice that these confidence regions are significantly tighter for the right hand plot where our generalised basis with distributed modes is used than they are for the Laguerre model in the left hand plot.

This improved accuracy is commensurate with the results of Theorem 2 which tells us that the variability in the estimated frequency response is proportional to $\sigma_\nu^2 \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2$ which we have plotted in in figure 4 for both the Laguerre basis and our multi-mode basis. From this plot Theorem 2 predicts the variance for our general basis model to be smaller than for the Laguerre model at all except high frequencies. Indeed, careful inspection of figure 3 shows negligibly larger confidence regions for the generalised basis estimate at high frequencies.

Together, figures 1 and 2 show that the two error components in system identification, namely 'bias error' due to parsimonious model structure and 'variance error' due to noise corruption of measurements can both be significantly reduced by taking advantage of the multi-modal nature of our unifying basis construction.

To complete this illustration, we close the simulation study by replacing the system under study in (49) with a resonant system

$$G_T(s) = \frac{\omega_c^2 e^{-2s}}{(3s+1)(s^2+2\zeta\omega_c s+\omega_c^2)} \qquad ;\omega_c = 0.8, \zeta = 0.2. \tag{51}$$

If we drive this system with the unit amplitude, 0.02Hz fundamental frequency, 500 sample square wave that we used previously and then estimate the system dynamics using the model structure (1) and a least squares criterion, then the results for different basis function choices are shown in figures 5 and 6. No measurement noise was included in the simulation.

In the left hand plot of figure 5 we show the use of a 4th order Laguerre model with all modes at 3 seconds. Here the resonant and time delay nature of G_T has been completely missed in the estimated model. In the right hand plot, a 4th order Kautz model with poles commensurate with the resonant mode $\omega_c = 0.8, \zeta = 0.2$ in G_T is used. Again, the error is very large, this time due to the inability to capture the 3 second mode and the time delay.

To remedy this, we use a 4th order unifying basis model with mixed modes; two real ones with 3 second time constant as in the Laguerre model and 2 resonant ones as in the Kautz model. The result is shown in the left hand plot of figure 6. The results are much more accurate than for the previous cases with the only error component here resulting from approximation of the time delay nature of G_T.

This illustrates that bias error can be significantly reduced in trying to estimate (51) by using our unifying basis with distributed modes. To see the effect of our basis on the variance error, we appeal to Theorem 2 which indicates that this will be proportional to $\sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2$, which we plot for the Laguerre, Kautz, and unifying basis choices of $\{\mathcal{B}_k\}$ in the right hand diagram of figure 6.

The results are not so conclusive as they were for the previous simulation example. Depending on the frequency range of interest, figure 6 shows that any one of the three model structures could be expected to give an estimate least sensitive to measurement noise. However, the unifying basis model appears to be the least sensitive of the three models on average over all frequencies. Combined with the much improved bias error for the unifying basis model, this would appear to be the best choice for minimising the total error consisting of both a bias and variance component.

10 Conclusions

This paper has attempted to unify an area of recent interest in the system identification literature. Namely, the study of using orthogonal basis functions to estimate system dynamics. Towards this aim we showed how a very general orthonormal basis formulation may be trivially derived by a process that is essentially a Gram-Schmidt construction. This

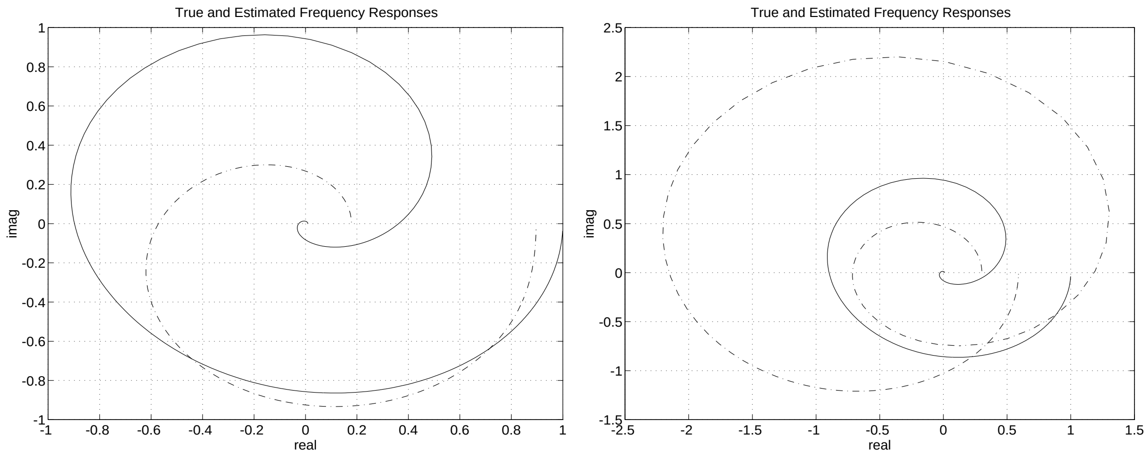

Figure 5: *Comparison of true and estimated frequency responses for the case of G_T having resonant modes. Solid Line is true Nyquist, dash-dot line is the estimate. On the left a 4th order Laguerre model with all modes at 3 seconds has been used, on the right a 4th order Kautz Basis model with all modes at a resonant frequency of 0.8 rad/s and damping constant $\zeta = 0.2$ has been used.*

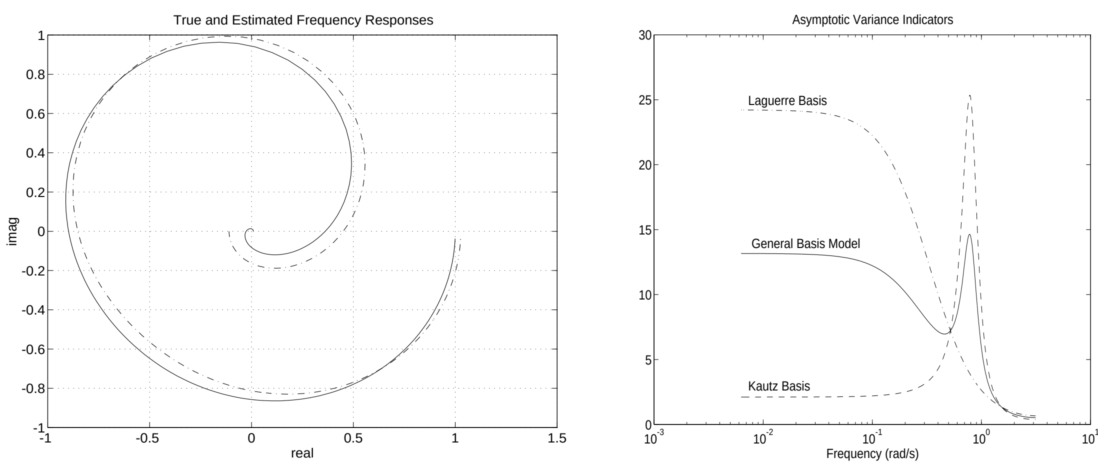

Figure 6: *Results of estimation on resonant plant when using our unifying basis with 2 real modes at 3 seconds and one resonant mode at 0.8 rad/s, damping factor $\zeta = 0.2$. On the left is a comparison of the true and approximate Nyquist plots. On the right is a plot of $\sum_{k=0}^{3} |\mathcal{B}_k(e^{j\omega})|^2$ for all three model structures. This latter plot allows us to judge the size of the random error component.*

general formulation achieves our unification aim since, as we showed, all other currently known orthonormal bases are simple special cases of our general construction.

More interestingly, the known bases are seen to be quite restrictive special cases in which prior knowledge about only one mode is incorporated. Our unifying construction therefore leads to an infinite number of new orthonormal basis systems that allow incorporation of prior knowledge of any number and type of mode. Taking advantage of this flexibility results in more accurate estimation which we illustrated both by theoretical analysis of estimate variance and by simulation example.

Appendix A Proof of Gram-Schmidt Lemma

Proof. To begin with, we have $\mathcal{W}_0 = \mathcal{V}_0$ so normalisation gives

$$\mathcal{B}_0(z) = \frac{\sqrt{1-|\xi_0|^2}}{z-\xi_0}.$$

Continuing, $\mathcal{B}_1$ is given by Gram-Schmidt as

$$\begin{aligned}\mathcal{W}_1(z) &= \frac{1}{z-\xi_1} - \left(\frac{1}{2\pi j}\oint_{\mathbf{T}}\left(\frac{z}{z-\xi_1}\right)\left(\frac{1-|\xi_0|^2}{1-\overline{\xi_0}z}\right)\frac{\mathrm{d}z}{z}\right)\frac{1}{(z-\xi_0)} \\ &= \frac{1}{z-\xi_1} - \left(\frac{1-|\xi_0|^2}{1-\xi_1\overline{\xi_0}}\right)\frac{1}{(z-\xi_0)}.\end{aligned}$$

It is easy to verify that $\mathcal{W}_1(1/\overline{\xi_0}) = 0$ so that after the normalisation step (21) we have that $\mathcal{B}_1(z)$ is

$$\mathcal{B}_1(z) = \frac{\sqrt{1-\xi_1^2}}{(z-\xi_1)}\left(\frac{1-\overline{\xi_0}z}{z-\xi_0}\right)$$

so that we have proved the Lemma for $n = 0, 1$. Now suppose that the Lemma is true for some arbitrary $n > 1$. Then

$$\mathcal{W}_{n+1}(z) = \frac{1}{z-\xi_{n+1}} - \sum_{k=0}^{n}\left\langle\frac{1}{z-\xi_n}, \mathcal{B}_k\right\rangle\mathcal{B}_k(z) = \frac{N(z)}{\prod_{k=0}^{n+1}(z-\xi_k)}$$

for some numerator polynomial $N(z)$. Now, by the orthogonality implied by the Gram-Schmidt procedure

$$0 = \langle\mathcal{W}_{n+1}, \mathcal{B}_0\rangle = \frac{1}{2\pi j}\oint_{\mathbf{T}}\mathcal{B}_0(z)\overline{\mathcal{W}_{n+1}(z)}\frac{\mathrm{d}z}{z} = \sqrt{1-|\xi_0|^2}\,\overline{\mathcal{W}_{n+1}(\xi_0)}$$

so that $\overline{\mathcal{W}_{n+1}(\xi_0)} = 0$. Using this and the fact that Gram-Schmidt implies $< \mathcal{W}_{n+1}, \mathcal{B}_1 >= 0$ then gives that $\overline{\mathcal{W}_{n+1}(\xi_1)} = 0$ and so on so that $\mathcal{W}_{n+1}$ has zeros at $1/\bar{\xi}_0, \cdots, 1/\bar{\xi}_n$. Therefore, $N(z)$ must be of the form $N(z) = \prod_{k=0}^{n}(1-\bar{\xi}_k z)$. After normalising $\mathcal{W}_{n+1}$ to produce $\mathcal{B}_{n+1}$ we find $\mathcal{B}_{n+1}$ must be of the form

$$\mathcal{B}_{n+1}(z) = \frac{\sqrt{1-|\xi_{n+1}|^2}}{z-\xi_{n+1}}\prod_{k=0}^{n}\left(\frac{1-\overline{\xi}_k z}{z-\xi_k}\right).$$

The Lemma then follows by induction. □□□

Appendix B Correspondence with the Balanced Realisation Construction

Here we show that the second order basis vectors generated by the scheme of Heuberger, Van den Hof and others [4, 5, 6, 21] are identical to our construction under the restrictive special assumption of fixed modes. That is $\xi_k = \xi$ for all k. As explained in section 5 in the balanced realisation scheme one begins with an all pass version $G(z)$ of the second order dynamics of interest:

$$G(z) = \frac{1 - 2\mathrm{Re}\{\xi\}z + |\xi|^2 z^2}{z^2 - 2\mathrm{Re}\{\xi\}z + |\xi|^2}.$$

This has observer form state space description $G = (A, B, C, D)$ given as

$$A = \begin{pmatrix} 2\mathrm{Re}\{\xi\} & 1 \\ -|\xi|^2 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} (|\xi|^2 - 1)2\mathrm{Re}\{\xi\} \\ 1 - |\xi|^4 \end{pmatrix}, \quad C = (1, 0), \quad D = |\xi|^2.$$

One then finds a balanced realisation of these dynamics. Here we use the algorithm of Moore [16] to find this realisation. We begin by solving the appropriate Lyapanov equations (25) and (26) for the controllability and observability Grammians P and Q as

$$\begin{aligned} P &= (1 - |\xi|^2) \begin{pmatrix} 1 + |\xi|^2 & -2\mathrm{Re}\{\xi\} \\ -2\mathrm{Re}\{\xi\} & 1 + |\xi|^2 \end{pmatrix} \\ Q &= \frac{1}{(1 - |\xi|^2)|1 - \xi^2|^2} \begin{pmatrix} 1 + |\xi|^2 & 2\mathrm{Re}\{\xi\} \\ 2\mathrm{Re}\{\xi\} & 1 + |\xi|^2 \end{pmatrix}. \end{aligned}$$

We then Cholesky factor Q as $Q = R^T R$ where

$$R = \frac{1}{\sqrt{2(1 - |\xi|^2)}} \begin{pmatrix} |1 - \xi|^{-1} & |1 - \xi|^{-1} \\ -|1 + \xi|^{-1} & |1 + \xi|^{-1} \end{pmatrix}$$

and go on to form $S = RPR^T$ and then spectrally factor S as $S = U\Sigma U^T$. However, in the special case we are considering of starting with an all pass function the Grammians satisfy $PQ = I$ so that S as defined is the identity matrix which is invariant under a similarity transformation consisting of rotation by any angle φ. Therefore, we can perform the spectral factorisation of S with $\Sigma = I$ and

$$U = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

The final step of forming a balanced realisation of G is to form the transformation matrix T as $T = U^T R$. The balanced realisation (A_B, B_B, C_B, D_B) is then formed as $(TAT^{-1}, TB, CT^{-1}, D)$. The balanced realisation orthonormal construction scheme then finds basis functions $\{\mathcal{B}'_n, \mathcal{B}''_n\}$ from the balanced realisation as

$$\begin{pmatrix} \mathcal{B}'_n \\ \mathcal{B}''_n \end{pmatrix} = z(zI - A_B)^{-1} B_B G(z)^n$$

which in our case evaluates to

$$\begin{pmatrix} \mathcal{B}'_n \\ \mathcal{B}''_n \end{pmatrix} = \begin{pmatrix} \beta z + \mu \\ \beta' z + \mu' \end{pmatrix} \frac{z\sqrt{1-|\xi|^2}}{z^2 - 2\mathrm{Re}\{\xi\}z + |\xi^2|} G(z)^n$$

where

$$\begin{aligned} \sqrt{2}\beta &= |1-\xi|\cos\varphi + |1+\xi|\sin\varphi && \text{(B.52)} \\ \sqrt{2}\mu &= |1-\xi|\cos\varphi - |1+\xi|\sin\varphi && \text{(B.53)} \\ \sqrt{2}\beta' &= |1+\xi|\cos\varphi - |1-\xi|\sin\varphi && \text{(B.54)} \\ -\sqrt{2}\mu' &= |1+\xi|\cos\varphi + |1-\xi|\sin\varphi. && \text{(B.55)} \end{aligned}$$

Adding and subtracting (B.52) and (B.53) in order to express $\cos\varphi$ and $\sin\varphi$ in terms of $(\beta+\mu)$ and $(\beta-\mu)$ then gives that in the balanced realisation construction β and μ are constrained to satisfy

$$(1+|\xi|^2)(\beta^2+\mu^2) + 2(\xi+\overline{\xi})\beta\mu = |1-\xi^2|^2$$

after we use the identity $\cos^2\varphi + \sin^2\varphi = 1$. This is identical to the constraint (17) on β and μ in our construction. Finally, if we equate expressions for $\cos\varphi$ and $\sin\varphi$ obtained from (B.52) and (B.53) and also (B.54) and (B.55) we obtain

$$\begin{aligned} \frac{\mu-\beta}{|1+\xi|} &= \frac{\mu'+\beta'}{|1-\xi|} && \text{(B.56)} \\ \frac{\beta+\mu}{|1-\xi|} &= \frac{\beta'-\mu'}{|1+\xi|}. && \text{(B.57)} \end{aligned}$$

Solving (B.56) and (B.57) for β' and μ' in terms of β and μ then gives

$$\begin{pmatrix} \beta' \\ \mu' \end{pmatrix} = \frac{1}{\sqrt{1-\alpha^2}} \begin{pmatrix} \alpha & 1 \\ -1 & -\alpha \end{pmatrix} \begin{pmatrix} \beta \\ \mu \end{pmatrix}, \qquad ;\alpha = \frac{\xi+\overline{\xi}}{1+|\xi|^2}.$$

which is identical to the rotation (19) used in our unifying construction. The balanced realisation construction with second order $G(z)$ prototype is therefore identical to the unifying construction (14) when the modes ξ_k are restricted to be fixed and in complex conjugate pairs.

Appendix C Proof of Completeness

Proof. Throughout we use the idea that a set is dense in a Hilbert Space if and only if there does not exist a non-zero element of the space that is orthogonal to all the elements in the set [17].

Firstly, suppose that $\sum(1-|\xi_k|)<\infty$. Then by Lemma D.1 the Blaschke product $f(z)$ defined by

$$f(z)=\prod_{k=0}^{\infty}\frac{\bar{\xi}_k}{|\xi_k|}\left(\frac{\xi_k-z}{1-\bar{\xi}_k z}\right)$$

is a well defined inner function on $\mathbf{D}$. Therefore $f(e^{j\omega})$ is a non-zero function in $H_2(\mathbf{T})$ for which

$$\left\langle f(e^{j\omega}),\mathcal{B}_n(e^{-j\omega})\right\rangle=\frac{1}{2\pi j}\oint_{\mathbf{T}}f(z)\overline{\mathcal{B}_n(z^{-1})}\frac{\mathrm{d}z}{z}=0$$

for any n by Cauchy's Integral Theorem. Therefore Span$\{\mathcal{B}_n(e^{-j\omega})\}$ is not complete in $H_2(\mathbf{T})$. Conversely, suppose that Span$\{\mathcal{B}_n(e^{-j\omega})\}$ is not complete in $H_2(\mathbf{T})$. Then there exists a function $f\in H_2(\mathbf{T})$ not equal to zero such that $<f,\mathcal{B}_n>=0$ for every n. Now, by Cauchy's Integral Theorem, if f is to be orthogonal to $\mathcal{B}_0(e^{-j\omega})$, then f must have a zero at ξ_0. By the same argument, if f is to be orthogonal to both $\mathcal{B}_0(e^{-j\omega})$ and $\mathcal{B}_1(e^{-j\omega})$ then since f is already known to have a zero a ξ_0 then f must have zeros at both ξ_0 and ξ_1. Continuing this argument, if f is to be orthogonal to all the $\{\mathcal{B}_k\}$, then $f(z)$ must have zeros at $\{\xi_0,\xi_1,\xi_2,\cdots\}$. If $f(z)$ is also to be non-zero and in $H_2(\mathbf{T})$, then by Lemma D.2 it must be that $\sum(1-|\xi_k|)<\infty$. □□□

Appendix D Results on Blaschke Products

Lemma D.1. *Let $\{\alpha_n\}$ be a sequence of non-zero complex numbers in $\mathbf{D}$. A necessary and sufficient condition that the Blaschke product*

$$B(z) = z^m \prod_{k=0}^{\infty} \frac{\bar{\alpha}_k}{|\alpha_k|} \left(\frac{\alpha_k - z}{1 - \bar{\alpha}_k z} \right)$$

should converge uniformly on compact subsets of $\mathbf{D}$ is that

$$\sum_{k=0}^{\infty} (1 - |\alpha_k|) < \infty$$

in which case $B(z)$ is an inner function and has no other zeros than those at $\alpha_0, \alpha_1, \alpha_2, \cdots$ (and 0 if $m > 0$).

Proof. See [7] page 64 or [17] page 310 (Theorem 15.21). □□□

Lemma D.2. *Suppose $f(z) \in H_p(\mathbf{D})$ for some p and suppose $f \not\equiv 0$. Suppose $\{\alpha_n\}$ are the zeros of $f(z)$. Then*

$$\sum_{k=0}^{\infty} (1 - |\alpha_k|) < \infty$$

Proof. See [8] page 53(Theorem 2.1). □□□

Appendix E Proof of Variance Bounds

Proof. A combination of Slutsky's Theorem and Theorem 9.1 in [12] gives the asymptotic result

$$\lim_{N\to\infty} N\mathbf{E}\left\{|G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2\right\} = ?(\omega)^\star R^{-1}QR^{-1}?(\omega) \qquad \text{w.p.1} \tag{E.58}$$

where

$$R \triangleq \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\phi_k\phi_k^T = \frac{1}{2\pi}\int_{-\pi}^{\pi} ?(\omega)?^\star(\omega)|F(e^{j\omega})|^2\Phi_u(\omega)\,\mathrm{d}\omega \tag{E.59}$$

$$Q \triangleq \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\sum_{m=0}^{N-1}\phi_k\phi_m^T\mathbf{E}\left\{H\nu_k H\nu_m\right\} = \frac{\sigma_\nu^2}{2\pi}\int_{-\pi}^{\pi} ?(\omega)?^\star(\omega)|F(e^{j\omega})|^2|H(e^{j\omega})|^2\Phi_u(\omega)\,\mathrm{d}\omega.$$

Furthermore, by construction the following matrix is positive semi-definite

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_u}{|H|^2\sigma_\nu^2}\left[\begin{array}{c} ?F \\ ?F|H|^2\sigma_\nu^2\end{array}\right]\left[\begin{array}{cc} ?^\star\overline{F} & ?^\star\overline{F}|H|^2\sigma_\nu^2\end{array}\right]\mathrm{d}\omega = \left[\begin{array}{cc} Z & R \\ R & Q\end{array}\right]$$

where

$$Z \triangleq \int_{-\pi}^{\pi} ?(\omega)?^\star(\omega)\frac{|F(e^{j\omega})|^2\Phi_u(\omega)}{|H(e^{j\omega})|^2\sigma_\nu^2}\,\mathrm{d}\omega.$$

Therefore, by a standard result for partitioned positive semi-definite matrices[3] we have that $Z^{-1} \leq R^{-1}QR^{-1}$. Using this, and using Lemma E.1 with the substitution $f = |F|^2\Phi_u|H|^{-2}\sigma_\nu^{-2}$ then gives the general lower bound (43). Finally, when $|H(e^{j\omega})|^2 = 1$ then $\sigma_\nu^2 R = Q$ so that

$$\lim_{N\to\infty} N\mathbf{E}\left\{|G(e^{j\omega},\widehat{\theta}) - G(e^{j\omega},\theta_\star)|^2\right\} = \sigma_\nu^2 ?(\omega)^\star R^{-1}?(\omega) \qquad \text{w.p.1} \tag{E.60}$$

so that we can again use Lemma E.1 but this time with the substitution $f = |F|^2\Phi_u\sigma_\nu^{-2}$ to get the white noise specific upper and lower bounds. □□□

Lemma E.1. *If* $?(\omega)^T = [\mathcal{B}_0(e^{j\omega}),\cdots,\mathcal{B}_{p-1}(e^{j\omega})]$ *and the* $\{\mathcal{B}_k(e^{j\omega})\}$ *are* $L_2(\mathbf{T})$ *orthonormal, then for* $f(\varphi)$ *any positive definite function*

$$\min_{\varphi\in[-\pi,\pi]}\frac{\mathcal{S}(\omega)}{f(\varphi)} \leq ?^\star(\omega)\left(\int_{-\pi}^{\pi} ?(\varphi)?^\star(\varphi)f(\varphi)\,\mathrm{d}\varphi\right)^{-1}?(\omega) \leq \max_{\varphi\in[-\pi,\pi]}\frac{\mathcal{S}(\omega)}{f(\varphi)}$$

where

$$\mathcal{S}(\omega) \triangleq 2\pi\sum_{k=0}^{p-1}\left|\mathcal{B}_k(e^{j\omega})\right|^2$$

[3] See Lemma A.3 in [20]

Proof. Take $x \in \mathbf{C}^p$ such that $x^*x = 1$ and put

$$P \triangleq \int_{-\pi}^{\pi} ?(\varphi)?^*(\varphi) f(\varphi)\,\mathrm{d}\varphi. \tag{E.61}$$

Then

$$\begin{aligned} x^*Px &= \sum_{r=0}^{p-1}\sum_{k=0}^{p-1} x_r\overline{x_k}\int_{-\pi}^{\pi}\mathcal{B}_r(e^{j\varphi})\overline{\mathcal{B}_k(e^{j\varphi})}f(\varphi)\,\mathrm{d}\varphi & \text{(E.62)}\\ &= \int_{-\pi}^{\pi} f(\varphi)\left|\sum_{r=0}^{p-1} x_r\mathcal{B}_r(e^{j\varphi})\right|^2 \mathrm{d}\varphi & \text{(E.63)}\\ &\leq \max_{\varphi\in[-\pi,\pi]} f(\varphi)\sum_{r=0}^{p-1}\sum_{k=0}^{p-1} x_r\overline{x_k}\int_{-\pi}^{\pi}\mathcal{B}_r(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})}\,\mathrm{d}\omega = 2\pi\max_{\varphi\in[-\pi,\pi]} f(\varphi) & \text{(E.64)} \end{aligned}$$

so

$$2\pi \min_{\varphi\in[-\pi,\pi]} f(\varphi) \leq x^*Px \leq 2\pi \max_{\varphi\in[-\pi,\pi]} f(\varphi) \tag{E.65}$$

Now P is by construction non-negative definite and symmetric so it's spectrum $\sigma(P)$ is the range of x^*Px with x in the unit ball. Finally, the spectrum $\sigma(P^{-1})$ is the reciprocal of $\sigma(P)$. □□□

REFERENCES

[1] P. R. Clement, *Laguerre functions in signal analysis and parameter identification*, Journal of the Franklin Institute, 313 (1982), pp. 85–95.

[2] G.C.Goodwin, B.M.Ninness, and H.V.Poor, *On the relationship between continuous and discrete Levinson-type algorithms for estimation*, in Proceedings of the Symposium on Identification,Budapest, IFAC, 1991, pp. 1179–1184.

[3] P. Halmos, *A Hilbert Space Problem Book*, D. Van Nostrand Company, New York, 1967.

[4] P. Heuberger and O. Bosgra, *Approximate system identification using system based orthonormal functions*, in Proceedings of the 29th Conference on Decision and Control, IEEE, December 1990, pp. 1086–1092.

[5] P. Heuberger, P. V. den Hof, and O. Bosgra, *A generalised orthonormal basis for linear dynamical systems*, Tech. Rep. N-404, Department of Mechanical Engineering, Delft University of Technology, The Netherlands, October 1992.

[6] ——, *On orthogonal basis functions that contain system dynamics*, Selected Topics in Identification Modelling and Control, 5 (1992), pp. 37–44.

[7] K. Hoffman, *Banach Spaces of Analytic Functions*, Prentice-Hall, Inc., New Jersey, 1962.

[8] J.B.Garnett, *Bounded Analytic Functions*, Academic Press,New York, 1981.

[9] R. King and P. Paraskevopoulos, *Parametric identification of discrete time SISO systems*, International Journal of Control, 30 (1979), pp. 1023–1029.

[10] T. Körner, *Fourier Analysis*, Cambridge University Press, 1988.

[11] Y. W. Lee, *Statistical Theory of Communication*, Wiley, New York, 1960.

[12] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc., New Jersey, 1987.

[13] L.Ljung, *Convergence analysis of parametric identification methods*, IEEE Transactions on Automatic Control, AC-23 (1978), pp. 770–783.

[14] ——, *Asymptotic variance expressions for identified black-box transfer function models*, IEEE Transactions on Automatic Control, AC-30 (1985), pp. 834–844.

[15] L.Ljung and Z.D.Yuan, *Asymptotic properties of black-box identification of transfer functions*, IEEE Transactions on Automatic Control, AC-30 (1985), pp. 514–530.

[16] B. Moore, *Principal component analysis in linear systems:Controllability, observability, and model reduction*, IEEE Transactions on Automatic Control, AC-26 (1981), pp. 17–31.

[17] W. Rudin, *Real and Complex Analysis*, McGraw Hill, New York, 1966.

[18] G. Sansone, *Orthogonal Functions*, Interscience Publishers, New York, 1959.

[19] G. Szegö, *Orthogonal Polynomials*, vol. 23, Colloquium publications - American Mathematical Society, New York, 1959.

[20] T.Söderström and P.Stoica, *System Identification*, Prentice Hall, New York, 1989.

[21] P. Van den Hof, P. Heuberger, and J. Bokor, *Identification with generalized orthonormal basis functions-statistical analysis and error bounds*, Selected Topics in Identification Modelling and Control, 6 (1993), pp. 39–48.

[22] B. Wahlberg, *Identification of resonant systems using Kautz filters*, in Proceedings of the 30th Conference on Decision and Control, 1991, pp. 2005–2010.

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

[24] B. Wahlberg, *System identification using kautz models*, IEEE Transactions on Automatic Control, AC-39 (1994), pp. 1276–1282.

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

[26] B. Wahlberg and P. Lindskog, *Approximate modeling by means of orthonormal functions*, In Modeling, Estimation and Control of Systems with Uncertainty, G.B. Di Masi, A Gombani, A.B Kurzhansky editors, Birhkäuser Boston', (1991), pp. 449–467.

[27] B. Wahlberg and L.Ljung, *Design variables for bias distribution in transfer function estimation*, IEEE Transactions on Automatic Control, AC-31 (1986), pp. 134–144.