IMPROVED AND QUANTIFIED ACCURACY FOR LINEAR SPECTRAL ESTIMATES

*Brett Ninness**

Electrical and Computer Engineering
University of Newcastle,
Callaghan 2308, Australia
brett@ee.newcastle.edu.au

Håkan Hjalmarsson

S3-Automatic Control
Royal Institute of Technology
SE-100 44 Stockholm Sweden
hakan.hjalmarsson@s3.e.kth.se

ABSTRACT

In the context of spectral estimation, this paper examines the use of an 'orthonormal basis' wherein prior knowledge (in the form of fixed poles) may be incorporated in the solution and FIR structures are then seen as a special case of implicitly involving prior knowledge of all poles at the origin. The main technical results are ones that quantify the accuracy of the resulting spectral estimates in a way that clearly exposes how it is affected by the fixed pole choice.

1. INTRODUCTION

Spectral estimation on the basis of observed output data is a problem of great importance [11, 14]; for example, in the context of channel estimation under white-noise excitation it is proving to be central to the implementation of modern telecommunications algorithms [10].

In the interests of low computational complexity, solutions which are linear in parameters and data are preferred. The latter considerations have led to the widespread popularity of solutions that involve FIR type models and 'least–squares' criterions [10, 14].

As will be illustrated by example in the next section, a potential difficulty with this solution (especially for short data sets; ie. fast adaptation) is that unacceptably high noise induced 'variance' errors may occur for FIR tap lengths sufficiently long to accurately model the full impulse response involved.

The solution investigated here is one in which the 'tap length' is decreased to reduce the variance error but, by virtue of allowing the inclusion of fixed pole locations, prior knowledge of system response may be incorporated in the hope of still achieving accurate impulse response modelling for low model orders. The details of this strategy involve the use of an 'orthonormal basis' model [6] which contains the FIR structure as a special case of all poles at the origin.

The ideas of using fixed pole structures in signal processing applications have been examined before [17, 16], as has the idea of using ortho-normalised versions of them [9, 2]. However, this previous work has not examined the particular general basis (see (5)) used here and has also not progressed to using the basis as an analytical tool which allows the quantification of estimation accuracy.

This latter comment refers to the main technical results of this paper (theorems 6.1 and 6.2) which provide quantification of the variance error in the ensuing spectral estimates. The derivations of these expressions rely on rather long arguments based on the recent results in [8] and are therefore not presented here, but instead are available in [7].

*This work was supported by the Australian Research Council and the Centre for Integrated Dynamics and Control (CIDAC)

2. MOTIVATIONAL EXAMPLE

Consider the problem of estimating, on the basis of an observed output data set $\{y_1, y_2, \cdots, y_N\}$, the response of a channel $H(z)$ under the assumption that the input that drove the channel to produce $\{y_k\}$ was white. This can be viewed as a spectrum estimation problem, and a very well known solution involves fitting an FIR model to the data using a Least-Squares criterion [11, 14].

As an example of this solution, suppose that

$$H(z) \triangleq \frac{z^3 - 1.9235z^2 + 1.5910z - 0.5203}{z^3 - 1.9464z^2 + 1.5155z - 0.5368}$$

(so that the underlying spectral density of $\{y_k\}$ is $\Phi_y(\omega) = \left|H(e^{j\omega})\right|^2$) and that $N = 500$ data samples are available. The result of fitting a 4th order FIR model is shown as a dashed line in figure 1. Compared to the true spectrum shown as the solid line, there are

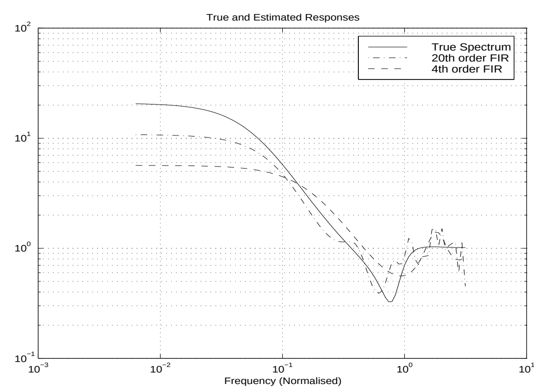

Figure 1: *Comparison of true and estimated Spectral Densities. Solid Line is true Spectral Density, dashed line is the 4th order FIR estimate, and dash–dot line is the 20th order FIR estimate.*

significant estimation errors caused by the order of the model being too low to be able to adequately capture the full impulse response of $H(z)$.

To deal with this, a 20'th order model can also be fitted, and the results of this are shown as the dash–dot line. Again, there are significant estimation errors, this time due to the model order being so high that there is very high variability in the estimates.

These sorts of estimation difficulties are the motivation for this paper. The solution proposed here is one of smoothly extending the

FIR structure to an orthonormal structure which includes the FIR one as a special case, but allows the injection of prior knowledge so as to allow a decreased model order (to fight variance error) while still allowing long-tailed impulse responses to be accurately modelled (to fight undermodelling error).

The main technical result associated with this idea is one that allows the variability of the ensuing estimates to be quantified in the frequency domain; see theorems 6.1 and 6.2.

3. PROBLEM FORMULATION

Suppose that $\{y_k\}$ with $k = \cdots, -1, 0, 1, \cdots$ is a wide sense stationary and zero mean process with corresponding covariance function

$$R_y(\tau) = \mathbf{E}\{y_k y_{k+\tau}\}.$$

Provided $R_y(\tau)$ decreases sufficiently quickly with increasing τ, then the process also possesses a spectral density $\Phi_y(\omega)$ given by

$$\Phi_y(\omega) = \sum_{\tau=-\infty}^{\infty} R_y(\tau) e^{j\omega\tau}$$

and then if $\Phi_y(\omega)$ is continuous and bounded away from zero so that the Paley-Wiener condition

$$\int_{-\pi}^{\pi} \log \Phi_y(\omega)\,\mathrm{d}\omega > -\infty$$

is satisfied, then $\{y_k\}$ possesses a Wold decomposition devoid of deterministic component as follows

$$y_k = e_k + \sum_{n=1}^{\infty} h_n e_{k-n}.$$

Here $\{e_k\}$ is a zero mean i.i.d. process of variance $\mathbf{E}\{e_k^2\} = \sigma^2$ and the transfer function

$$H(z) = 1 + \sum_{n=1}^{\infty} h_n z^{-n} = \exp\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{z+e^{-j\omega}}{z-e^{-j\omega}} \log \Phi_y(\omega)\,\mathrm{d}\omega\right\} \tag{1}$$

and its inverse $H^{-1}(z)$ are both analytic in $|z| > 1$. This permits an alternative expression for the power spectral density $\Phi_y(\omega)$ of $\{y_k\}$ in terms of the so-called spectral factor $H(z)$ as

$$\Phi_y(\omega) = \sigma^2 |H(e^{j\omega})|^2.$$

It is often of interest to estimate this spectral density from observations of realisations of $\{y_k\}$; channel response estimation for telecommunications applications is one example, but many others exist [14].

A common existing strategy for achieving this estimation involves forming a truncated approximation to $H^{-1}(z)$ as

$$H^{-1}(z) \approx 1 - G(z,\theta)$$

$$G(z,\theta) \triangleq \sum_{n=1}^{p-1} \theta_n z^{-n} \qquad \theta^T \triangleq [\theta_0, \cdots, \theta_{p-1}] \tag{2}$$

and then find an estimate $\widehat{\theta}_N$ of θ from the available data so that the spectrum is estimated as

$$\Phi_y(\omega, \widehat{\theta}_N) = \frac{\widehat{\sigma}^2}{|1 - G(e^{j\omega}, \widehat{\theta})|^2}$$

where $\widehat{\sigma}^2$ is an estimate of σ^2.

There is a huge literature going back at least as far as Wiener [15] that analyses this scheme. Much of it proceeds by using the ideas of Szegö [12] wherein $G(z,\theta)$ is considered as an expansion of $H^{-1}(z)$ in the polynomial (or trigonometric) basis $\{z^{-k}\}$ which is orthogonal on the unit circle $\mathbf{T}$ with respect to the inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{j\omega})\overline{g(e^{j\omega})}\,\mathrm{d}\omega = \frac{1}{2\pi j}\oint_{\mathbf{T}} f(z)\overline{g(z)}\,\frac{\mathrm{d}z}{z}. \tag{3}$$

The idea in this paper is to use these ideas of expansion in orthonormal bases, but to extend them by replacing the $\{z^{-k}\}$ basis functions by more general ones $\{\mathcal{B}_k(z)\}$ such that the approximation in (2) becomes

$$H^{-1}(z) \approx 1 - G(z,\theta), \qquad G(z,\theta) = \sum_{n=0}^{p-1} \theta_n \mathcal{B}_n(z). \tag{4}$$

As will be shown, the utility of this idea will be that the approximation in (4) can be significantly improved if even a small amount of prior knowledge about the nature of the spectrum $\Phi_y(\omega)$ is available.

4. ORTHONORMAL MODEL STRUCTURES

The construction of a scalar orthonormal basis $\{\mathcal{B}_k(z)\}$ that allows the incorporation of prior knowledge of the zeros $\{\xi_0, \cdots, \xi_{p-1}\}$ of $H(z)$ has been developed elsewhere [6] where the following basis functions are presented

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

so that the model structure (4) can be represented in signal form as shown in figure 2.

Notice that for the choice $\xi_k = 0$ for all k this basis reduces to that corresponding to the conventional FIR expansion basis (or trigonometric basis) model structure as given in (1). For the choice of $\xi_k = \xi \in \mathbf{R} \cap \mathbf{D}$ of all poles being the same and real, the Laguerre basis [13] is also obtained as a special case. However, it would seem more appropriate to not use this basis formulation in (5) in such a restricted setting, but rather choose the poles $\{\xi_k\}$ in a distributed fashion that most accurately reflects prior knowledge of the zeros of $H(z)$.

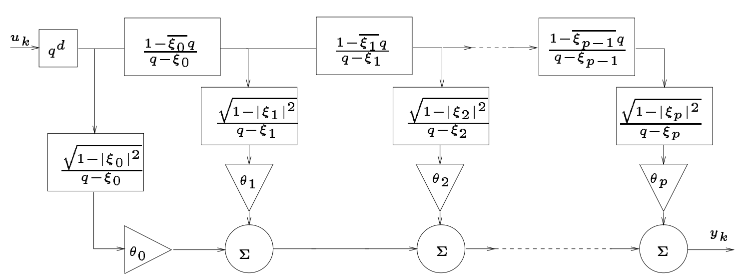

Figure 2: *Diagrammatic representation of Orthonormal Model structure employed in this paper.*

5. LINEAR ESTIMATION METHOD

A key reason for the large literature on using FIR expansions such as (2) for modelling the process $\{y_k\}$ is that it leads to straight-forward and computationally cheap methods for estimating $\Phi_y(\omega)$ from realisations of $\{y_k\}$. This is so because (2) is linear in the unknown parameters $\{\theta_0, \cdots, \theta_{p-1}\}$.

Since the basis function expansion (4) has the FIR expansion (2) as a special case (when $\xi_k = 0, \forall k$) then (2) will lead to equally simple estimation schemes, again because of the linearity in the unknown parameters.

More specifically, using q to denote the forward in time shift operator, (4) implies a signal model of

$$y_k = \phi_k^T \theta + e_k$$

where

$$\phi_k^T = [\mathcal{B}_0(q)y_k, \mathcal{B}_1(q)y_k, \cdots, \mathcal{B}_{p-1}(q)y_k].$$

Now, as mentioned in [13], it is common to wish to solve estimation problems via the ideas of Maximum Likelihood (ML). However, this can lead to difficult non-convex optimisation problems, so that a compromise is often made with FIR models in the Gaussian case by approximating the ML cost function with a quadratic one. This same approach of using the 'least-squares' estimate will be employed here. More specifically, the estimate $\widehat{\theta}_N$ is chosen as

$$\widehat{\theta}_N = \arg\min_{\theta \in \mathbf{R}^p} V_N(\theta), \qquad V_N(\theta) \triangleq \frac{1}{N}\sum_{k=0}^{N-1}(y_k - \phi_k^T\theta)^2$$

which is well known to have closed form solution

$$\widehat{\theta}_N = R_p(N)^{-1}\frac{1}{N}\sum_{k=0}^{N-1}\phi_k y_k, \qquad (6)$$

$$R_p(N) \triangleq \frac{1}{N}\sum_{k=0}^{N-1}\phi_k\phi_k^T. \qquad (7)$$

This in turn provides an estimate of the inverse spectral factor $H^{-1}(z)$ as

$$\begin{aligned} H^{-1}(z, \widehat{\theta}_N) &= 1 - \widehat{\theta}_N^T \Gamma_p(z), \\ \Gamma_p(z) &\triangleq [\mathcal{B}_0(z), \cdots, \mathcal{B}_{p-1}(z)]^T, \end{aligned}$$

and a corresponding spectral density estimate

$$\Phi_y(\omega, \widehat{\theta}_N) = \frac{\widehat{\sigma}^2}{|H^{-1}(e^{j\omega}, \widehat{\theta}_N)|^2}.$$

6. MAIN TECHNICAL RESULTS

The statistical nature of these estimates may be investigated in an asymptotic sense by using Theorem 8.3 of [4] to conclude that under the given assumptions on $\{y_k\}$ then

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

where

$$\theta_\star \triangleq \arg\min_{\theta \in \mathbf{R}^p} \mathbf{E}\left\{(y_k - \phi_k^T\theta)^2\right\}$$

and then the accuracy of an estimate of $\widehat{\theta}_N$ may be assessed in terms of its fluctuations about $\theta_\star$ by quantifying its distributional properties.

In particular, by using Theorem 9.1 of [4] it may be concluded

$$\sqrt{N}(\widehat{\theta}_N - \theta_\star) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_p) \quad \text{as } N \to \infty$$

so that since

$$G(e^{j\omega}, \widehat{\theta}_N) = \widehat{\theta}_N^T \Gamma_p(e^{j\omega})$$

it is possible to further conclude that as $N \to \infty$

$$\sqrt{N}\begin{bmatrix} \gamma_p(\omega_1) & 0 \\ 0 & \gamma_p(\omega_2) \end{bmatrix}^{-1/2} \begin{bmatrix} H^{-1}(e^{j\omega_1}, \widehat{\theta}_N) - H^{-1}(e^{j\omega_1}, \theta_\star) \\ H^{-1}(e^{j\omega_2}, \widehat{\theta}_N) - H^{-1}(e^{j\omega_2}, \theta_\star) \end{bmatrix}$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(0, \Lambda_p(\omega_1, \omega_2))$$

where

$$\begin{aligned} \Lambda_p(\omega_1, \omega_2) &= \frac{\Gamma_p^\star(e^{j\omega_1})P_p\Gamma_p(e^{j\omega_2})}{\gamma_p(\omega_1)}, \\ \gamma_p(\omega) &\triangleq \sum_{k=0}^{p-1}|\mathcal{B}_k(e^{j\omega})|^2, \\ P_p &\triangleq R_p^{-1}Q_pR_p^{-1}, \quad R_p \triangleq \mathbf{E}\left\{V_N''(\theta_\star)\right\}, \\ Q_p &\triangleq \lim_{N\to\infty} N\mathbf{E}\left\{V_N'(\theta_\star)\left(V_N'(\theta_\star)\right)^T\right\} \end{aligned}$$

and where $\cdot^\star$ denotes conjugate transpose. Simple algebra then provides

$$R_p = \mathbf{E}\left\{\phi_k\phi_k^T\right\},$$

and using Parseval's Theorem this can be expressed in the frequency domain as

$$R_p = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Gamma_p(e^{j\omega})\Gamma_p^\star(e^{j\omega})\Phi_y(\omega)\,\mathrm{d}\omega.$$

The matrix Q_p is somewhat more difficult to handle. It is possible to show that it may be expressed as [7]

$$Q_p = \sigma^2 R_p + \Delta_p$$

where

$$\lim_{p\to\infty}\|\Delta_p\|_2 = 0.$$

This leads to the idea (as per the analysis of [1, 5, 13]) of examining the statistical properties of the estimation scheme by looking at the asymptotic properties when both the model order p as well as the number of allowed data N are allowed to tend to infinity.

Specifically, the first main theoretical result of this paper is one that provides an engineering relevant simplification of asymptotic (in N) covariance $\Lambda_p(\omega_1, \omega_2)$ in $G(e^{j\omega}, \widehat{\theta}_N)$:

Theorem 6.1. *If $\Phi_y(\omega)$ is Lipschitz continuous of some order $\alpha > 0$ and if all the poles $\{\xi_k\}$ are such that $|\xi_k| \le \delta < 1$, then*

$$\lim_{p\to\infty}\Lambda_p(\omega_1, \omega_2) = \sigma^2\begin{bmatrix} \Phi_y^{-1}(\omega_1) & 0 \\ 0 & \Phi_y^{-1}(\omega_2) \end{bmatrix}$$

Proof. See [7]. □

Using the Gaussian approximation formula [3], this result then provides a further one providing insight into the noise induced variability of the estimate of the spectral density $\Phi_y(\omega, \widehat{\theta}_N) = \sigma^2|H(e^{j\omega}, \widehat{\theta}_N)|^2$ itself:

Theorem 6.2. *Under the assumptions of Theorem 6.1*

$$\sqrt{N}\begin{bmatrix} \gamma_p(\omega_1) & 0 \\ 0 & \gamma_p(\omega_2) \end{bmatrix}^{-1/2} \begin{bmatrix} \Phi_y(e^{j\omega_1},\widehat{\theta}_N) - \Phi_y(e^{j\omega_1},\theta_\star) \\ \Phi_y(e^{j\omega_2},\widehat{\theta}_N) - \Phi_y(e^{j\omega_2},\theta_\star) \end{bmatrix}$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_p(\omega_1,\omega_2)) \quad \text{as } N\to\infty$$

where for $\omega_1,\omega_2 \neq 0,\pi$

$$\lim_{p\to\infty} \Sigma_p(\omega_1,\omega_2) = 2\begin{bmatrix} \Phi_y^2(\omega_1) & 0 \\ 0 & \Phi_y^2(\omega_2) \end{bmatrix}.$$

while for $\omega_1,\omega_2 = 0,\pi$

$$\lim_{p\to\infty} \Sigma_p(\omega_1,\omega_2) = 4\begin{bmatrix} \Phi_y^2(\omega_1) & 0 \\ 0 & \Phi_y^2(\omega_2) \end{bmatrix}.$$

Proof. See [7]. □

The interpretation of this last theorem is that we can approximate the variance in our estimate of the spectral density as

$$\text{Var}\left\{\Phi_y(e^{j\omega},\widehat{\theta}_N)\right\} \approx \frac{2}{N}\Phi_y^2(\omega)\sum_{k=0}^{p-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \qquad (8)$$

This explicitly shows how the choice of fixed zero location $\{\xi_k\}$ in the estimation of $H(z)$ affects the sensitivity of resulting spectral density estimates - the last term in (8) provides a 'shaping' of the variance errors as they are distributed over frequency. Importantly, note that although these errors can be shaped, they are also in some sense 'conserved' since (by orthonormality of the $\{\mathcal{B}_k(z)\}$) the area under the last term in (8) is equal to p regardless of the choice of the $\{\xi_k\}$.

7. SIMULATION EXAMPLE

We continue the example presented in §2. Recall that, for the underlying dynamics $H(z)$ involved, the use of an FIR model led to poor estimates due to either undermodelling error for low model orders, or variance error for high model orders.

The solution proposed in this paper has been one of reducing the variance error by reducing the model order, while at the same time injecting prior knowledge of zero locations so as to not suffer a penalty of high undermodelling error.

In relation to this, note that $H(z)$ presented in §2 has zeros at $z = 0.7165$ and $z = 0.852e^{\pm j0.784}$, and it would seem feasible that coarse prior knowledge of these zeros being at $\{\xi_0,\cdots,\xi_3\} = \{0.5, 0.5, 0.5+j0.5, 0.5-j0.5\}$ could reasonably be assumed in many applications.

In this case, if this prior knowledge is incorporated in the estimation problem via the formulation (5), then the results are as shown in figure 3. Clearly, the incorporation of the prior knowledge of zeros has led to a very greatly improved estimate.

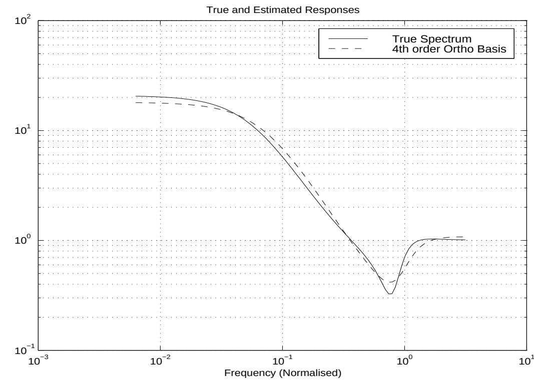

Figure 3: *Comparison of true and estimated Spectral Densities. Solid Line is true spectral density, dashed line is the orthonormal basis model estimate using prior knowledge poles at* $\{0.5, 0.5, 0.5+j0.5, 0.5-j0.5\}$.

REFERENCES

[1] K. BERK, *Consistent autoregressive spectral estimates*, Annals of Statistics, 2 (1974), pp. 489–502.

[2] G. DAVIDSON AND D. FALCONER, *Reduced complexity echo cancellation using orthonormal functions*, IEEE Trans. Cir. & Sys, 38 (1991), pp. 20–28.

[3] E. LEHMANN, *Theory of Point Estimation*, John Wiley & Sons, 1983.

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

[5] L.LJUNG AND Z.D.YUAN, *Asymptotic properties of black-box identification of transfer functions*, IEEE Trans. Auto. Cont, 30 (1985), pp. 514–530.

[6] B. NINNESS AND F. GUSTAFSSON, *A unifying construction of orthonormal bases for system identification*, IEEE Trans. Auto. Cont, 42 (1997), pp. 515–521.

[7] B. NINNESS AND H. HJALMARSSON, *Modelling of random processes using orthonormal bases*, tech. rep., Dept EE&CE, Uni. Newc, 1999.

[8] B. NINNESS, H. HJALMARSSON, AND F. GUSTAFSSON, *Generalised Fourier and Toeplitz results for rational orthonormal bases*, SIAM J. Cont. & Optim, 37 (1999), pp. 429–460.

[9] H. PEREZ AND S. TSUJII, *A system identification algorithm using orthogonal functions*, IEEE Trans. on Sig. Proc, 38 (1991), pp. 752–755.

[10] H. V. POOR AND G. W. (EDITORS), *Wireless communications : signal processing perspectives*, Prentice Hall, 1998.

[11] P. STOICA AND R. MOSES, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

[12] G. SZEGÖ, *Orthogonal Polynomials*, vol. 23, Colloq. pub - AMS, New York, 1939.

[13] B. WAHLBERG AND E. HANNAN, *Parameteric signal modelling using Laguerre filters*, Ann. Appl. Prob, 3 (1993), pp. 467–496.

[14] B. WIDROW, *Adaptive Signal Processing*, Prentice Hall, 1985.

[15] N. WIENER, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, M.I.T. Press, 1949.

[16] G. A. WILLIAMSON, *Tracking random walk systems with vector space adaptive filters*, IEEE Trans. Circ. & Sys-II 42 (1995), pp. 543–547.

[17] G. A. WILLIAMSON AND S. ZIMMERMANN, *Globally convergent adaptive IIR filters based on fixed pole locations*, IEEE Trans. Sig. Proc, 44 (1996), pp. 1418–1427.