The Annals of Applied Probability
1993, Vol. 3, No. 2, 467–496

PARAMETRIC SIGNAL MODELLING USING LAGUERRE FILTERS

BY B. WAHLBERG AND E. J. HANNAN

Royal Institute of Technology and Australian National University

Autoregressive (AR) modelling is generalized by replacing the delay operator by discrete Laguerre filters. The motivation is to reduce the number of parameters needed to obtain useful approximate models of stochastic processes, without increasing the computational complexity. Asymptotic statistical properties are investigated. Several AR model estimation results are extended to Laguerre models. In particular, it is shown how the choice of Laguerre time constant affects the resulting estimates. A Levinson-type algorithm for computing the Laguerre model estimates in an efficient way is also given. The Laguerre technique is illustrated by two simple examples.

1. Introduction. The concept of representing complex systems by simple models is fundamental in science. The aim is to reduce a complicated process to a simpler one involving a small number of parameters. The quality of the approximation is determined by its usefulness, for example, its predictive ability. Autoregressive (AR) and autoregressive moving-average (ARMA) models are the dominating parametric models in time series analysis, since they give useful approximations of many processes of interest. The ARMA model leads to nonlinear optimization problems to be solved for best approximation, while the special case of AR modelling only involves a quadratic least squares optimization problem. Hence, AR models are of great importance in applications where fast and reliable computations are necessary. The literature on AR estimation is extensive; see, for example, Hannan (1970), Priestley (1982), Ljung (1987), Hannan and Deistler (1988), Marple (1987) and Söderström and Stoica (1989).

The fact that the true system is bound to be more complex than a fixed order AR model has motivated the analysis of high-order AR approximations, where the model order is allowed to tend to infinity as the number of observations tends to infinity. See Berk (1974) and Hannan and Deistler (1988) for details. However, aspects such as computational limitations and numerical sensitivity set bounds on how high an AR order can be tolerated in practice.

Herein, we shall study discrete Laguerre filter model structures, which reduce the number of parameters to be estimated without increasing the numerical complexity of the estimation algorithm. An early reference on the use of Laguerre networks in estimation theory is Wiener (1956). More refer-

Received May 1990; revised 1992.

AMS 1991 *subject classifications*. Primary 62M10; secondary 62F12, 60G12, 60F05, 60G10.

Key words and phrases. Autoregression, network modeling, time series, spectral analysis, Laguerre functions.

ences on Laguerre series and parameter estimation can be found in Eykhoff (1974). However, in most of these contributions Laguerre functions are used in the time domain to approximate the signals, and then in a second step the dynamics of the system are estimated. See, for example, Hwang and Shih (1982) and Clement (1982) for some ideas.

Suppose that $\{y(t),\ t = \cdots -1, 0, 1, \dots\}$ is a stationary linear regular random process with Wold representation

$$(1) \qquad y(t) = \sum_{k=0}^{\infty} h_k^0 e(t-k), \qquad h_k^0 \in \mathbb{R},\ h_0^0 = 1.$$

Here $\{e(t)\}$ is a sequence of random variables with the properties

$$(2) \qquad E\{e(t)|\mathscr{F}_{t-1}\} = 0, \qquad E\{e(t)^2|\mathscr{F}_{t-1}\} = \sigma_0^2, \qquad E\{e(t)^4\} < \infty,$$

where $\mathscr{F}_{t-1}$ is the σ-algebra generated by $\{e(s),\ s \le t-1\}$. The transfer function

$$(3) \qquad H^0(q) = \sum_{k=0}^{\infty} h_k^0 q^{-k}, \qquad H^0(\infty) = 1$$

is a function of the shift operator q, $qe(t) = e(t+1)$. By q^{-1} we mean the corresponding delay operator $q^{-1}e(t) = e(t-1)$. The power spectral density of $\{y(t)\}$ then equals (we have chosen to not normalize the spectral density with $1/2\pi$)

$$(4) \qquad \Phi_y(e^{i\omega}) = \sigma_0^2 |H^0(e^{i\omega})|^2.$$

We shall assume that the complex function $[H^0(z)]^{-1}$, $z \in \mathbb{C}$, is analytic in $|z| > 1$ and continuous in $|z| \ge 1$. Then

$$(5) \qquad \left[H^0(z)\right]^{-1} = \sum_{k=0}^{\infty} a_k^0 z^{-k}, \qquad |z| \ge 1.$$

We shall impose a further smoothness condition on $[H^0(z)]^{-1}$, namely,

$$(6) \qquad \sum_{k=0}^{\infty} k^{1/2+\delta} |a_k^0| < \infty, \qquad \delta > 0.$$

This condition implies that $[H^0(z)]^{-1} \in \mathrm{Lip}(1/2 + \delta)$, that is, satisfies a Lipschitz condition of order $1/2 + \delta$. Notice that [cf. Zygmund (1968)]

$$(7) \qquad \left[H^0(z)\right]^{-1} \in \mathrm{Lip}(1/2 + \delta) \quad \Rightarrow \quad \sum_{k=0}^{\infty} k^{\delta} |a_k^0| < \infty, \qquad \delta > 0.$$

Hence, the condition (6) is more restrictive than the corresponding Lipschitz condition in (7).

By truncating the expansion (5) at $k = n$, we obtain a nth order autoregressive (AR) approximation of (1),

$$(8) \qquad A_n^0(q) y(t) = e(t), \qquad A_n^0(q) = 1 + \sum_{k=1}^{n} a_k^0 q^{-k}.$$

A crucial question is how large an order n must be chosen to obtain a useful AR approximation. From (5) and (6) we know that the process (1) can be arbitrarily well approximated by an AR model (in the mean square sense) by taking the order n large enough. However, nothing is said about the rate of convergence.

Assume that $H^0(z)$ is a rational function with zeros $\{z_i\}$, $|z_i| < 1$. The error in the AR approximation (8) is then of order δ^n, where $\delta = \max_j |z_j|$. Hence, zeros close to the unit circle imply a slow rate of convergence and consequently a high model order n. What sort of processes have zeros close to the unit circle? An example is a discrete time process obtained by rapid sampling of a continuous time stochastic process. Assume that the zeros of the continuous time process equal $\{\bar{z}_i\}$. The corresponding discrete time zeros will then approximately (for small sampling interval Δ) be equal to $\exp(\bar{z}_i \Delta) \approx 1 + \bar{z}_i \Delta$; see Wahlberg (1990). Hence, for rapidly sampled continuous time stochastic processes the rate of convergence of AR approximations will be very slow. In the limit as the sampling interval $\Delta \to 0$ the discrete time zeros converge to 1, and consequently the AR approximations fail to converge.

This motivates the investigation of alternative approximations which are less sensitive to the location of the zeros (the choice of sampling rate). The problem is that the memory of the delay operator is too short—only one sampling step. In Section 2 we shall generalize AR models by replacing the delay operator with discrete Laguerre filters. Properties of Laguerre approximations are presented in Section 3. In Section 4, Toeplitz forms related to Laguerre models are studied. The statistical properties of the proposed estimation algorithm are analyzed in Section 5 and high-order Laguerre model estimates are considered in Section 6. In Section 7 the estimation procedure is illustrated by two examples. Finally, Section 8 concludes the paper.

2. Laguerre model estimates.

2.1. *Model definition.* The idea of Laguerre models is to replace the delay operator q^{-k} in an AR model by the so-called discrete Laguerre filters

$$L_k(q,a) = \frac{\sqrt{1-a^2}}{q-a}\left(\frac{1-aq}{q-a}\right)^{k-1}, \qquad k \geq 1. \tag{9}$$

A thorough justification for this choice of filters is given in Section 3. The motivation is that the introduction of the Laguerre parameter a reduces the number of parameters needed to obtain useful approximate models.

DEFINITION 2.1. The *Laguerre model structure* is defined by the equation

$$\mathscr{M}\colon\ y(t) + \alpha_1 L_1(q,a)y(t) + \cdots + \alpha_n L_n(q,a)y(t) = e(t), \tag{10}$$

where the Laguerre filters $L_k(q, a)$ are defined by (9), $\{e(t)\}$ is a sequence of random variables with the properties (2) and $\alpha_k \in \mathbb{R}$.

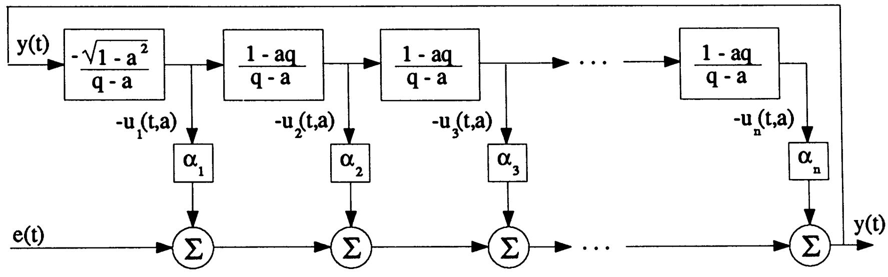

FIG. 1. *Laguerre network.*

REMARK 1. Observe that a Laguerre model simplifies to an AR model for $a = 0$.

The Laguerre model (10) can be rewritten in the linear regression form

$$y(t) = \varphi(t,a)'\theta + e(t), \tag{11}$$

$$\begin{aligned} \varphi(t,a) &= [-u_1(t,a) \cdots -u_n(t,a)]', \\ u_k(t,a) &= L_k(q,a)y(t), \qquad k \geq 1, \end{aligned} \tag{12}$$

$$\theta = (\alpha_1 \dots \alpha_n)' \in \mathbb{R}^n. \tag{13}$$

Figure 1 shows the corresponding Laguerre network configuration. Notice that the regression vector $\varphi(t,a)$ is treated as directly measurable in the network.

2.2. *Computing the estimate.* Given observations of the process (1), $\{y(1) \cdots y(N)\}$, we will discuss how to estimate a Laguerre model in the model set $\mathscr{M}$. In this section a is assumed to be fixed, while in Section 5 we shall discuss how to also estimate a as well as the order n.

In many parameter estimation problems a systematic approach to the estimation procedure is the maximum likelihood (ML) method. One problem is, however, that in most cases one has to solve a complicated optimization problem to find the ML estimate. For AR models and on Gaussian assumptions it is possible to approximate the ML cost function by a quadratic one (the least squares cost function) and still obtain an asymptotically efficient estimate. The same holds for the Laguerre model structure $\mathscr{M}$. Initial conditions have to be specified when forming $\varphi(t,a)$ from the observations. Herein, we will use pre- and postwindowing of $y(t)$, that is, assuming $y(t) = 0$, $t \leq 0$ or $t > N$. Using the linear regression structure (11), the least squares estimate of θ is then given by

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{\infty} [y(t) - \varphi(t,a)'\theta]^2 \quad \Rightarrow \tag{14}$$

$$\hat{\theta} = \left[\frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a)\varphi(t,a)'\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{\infty}\varphi(t,a)y(t)\right]. \tag{15}$$

The normal equations (15) can be solved in an efficient way if the sampled covariance matrix

$$\frac{1}{N}\sum_{t=1}^{\infty}\varphi(t,a)\varphi(t,a)' \tag{16}$$

has a Toeplitz structure. For the AR case ($a = 0$), pre- and postwindowing of $y(t)$ leads to a Toeplitz structure. In Section 4 we will show that such an assumption also gives a Toeplitz structure of (16) for the general ($a \neq 0$) Laguerre case. Observe that pre- and postwindowing does not affect the asymptotic behaviour (large N) of the estimate $\hat{\theta}$. Also notice that the Laguerre model estimation is not restricted to this assumption, and can be applied using only prewindowing or a forward–backward approach. See for instance Wahlberg (1988) for details.

2.3. *ARMA models with fixed MA part.* Laguerre models are theoretically equivalent to ARMA modelling with a fixed MA part. By adding up the terms in the model structure (10) to give a rational function, the following model structure is obtained:

$$\bar{\mathscr{M}}:\ \frac{A(q)}{C(q,a)}y(t) = y_c(t,a)e(t), \quad \Rightarrow y_C(t,a) = \bar{\varphi}(t,a)'\bar{\theta} + e(t), \tag{17}$$

$$A(q) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}, C(q,a) = \left(1 - aq^{-1}\right)^n, \tag{18}$$

$$\bar{\varphi}(t,a) = \left[-y_C(t-1,a) \cdots -y_C(t-n,a)\right]', \tag{19}$$

$$y_C(t,a) = \frac{1}{C(q,a)}y(t), \tag{20}$$

$$\bar{\theta} = (a_1 \cdots a_n)'. \tag{21}$$

This model structure corresponds to an ordinary AR model with a fixed prefilter $1/C(q,a)$.

The two model structures $\mathscr{M}$ and $\bar{\mathscr{M}}$ are theoretically equivalent, since $\theta = T\bar{\theta}$, where T is a nonsingular $n \times n$ matrix. However, there are several practical problems associated with the model structure $\bar{\mathscr{M}}$.

First, since we use a direct delay operator representation we will have numerical problems for small sampling intervals. By instead estimating the coefficients of the corresponding δ-operator polynomial, $\delta = (q-1)/\Delta$, this problem is removed. See Middleton and Goodwin (1990) for details.

Second, the model structure $\bar{\mathscr{M}}$ is not well scaled for large n. The problem is related to the condition number of the covariance matrix of the regressor vector. Let

$$\Gamma_n(a) = E\{\varphi(t,a)\varphi(t,a)'\}, \qquad \bar{\Gamma}_n(a) = E\{\bar{\varphi}(t,a)\bar{\varphi}(t,a)'\}. \tag{22}$$

The condition number of a Toeplitz covariance matrix is bounded from below by the ratio between the maximum value and the minumum value of the corresponding spectral density; see Grenander and Szegö (1959). For the model structure $\bar{\mathscr{M}}$ the process $\{y_C(t,a)\}$ generates the Toeplitz matrix $\bar{\Gamma}_n(a)$. The corresponding spectral density equals $|C(e^{i\omega},a)|^{-2}\Phi_y(e^{i\omega})$. Now,

$$\frac{\max_\omega\left[|C(e^{i\omega},a)|^{-2}\Phi_y(e^{i\omega})\right]}{\min_\omega\left[|C(e^{i\omega},a)|^{-2}\Phi_y(e^{i\omega})\right]} \to \infty, \qquad a \neq 0, \tag{23}$$

as $n \to \infty$. This follows from the fact that the ratio of smallest to largest value of $|C(e^{i\omega},a)|^2$ is $[(1-|a|)/(1+|a|)]^{2n}$. Consequently, the condition number of $\bar{\Gamma}_n(a)$ tends to infinity as $n \to \infty$. The reason is that the last components in $\bar{\varphi}(t,a)$ contain very little new information. The condition number of $\bar{\Gamma}_n(a)$ determines the numerical sensitivity in solving the normal equations to find the least squares (LS) estimate of $\bar{\theta}$. From a finite-precision arithmetic point of view it is important to have as well scaled a matrix as possible. As shown below, it is possible to give bounds on the condition number of $\Gamma_n(a)$ in terms of the spectral density of $\{y(t)\}$, while using the model structure $\bar{\mathscr{M}}$ results in an ill-conditioned LS problem for large n.

2.4. *Generalized AR results*. As shown in the preceding section there is a one-to-one mapping between Laguerre models and ARMA models with fixed MA part. By instead considering the filtered process $\{y_C(t,a)\}$, defined by (20), the ARMA problem is reduced to finding an AR model of the process $\{y_C(t,a)\}$. This observation can be used to generalize AR estimation results for a fixed model order n to Laguerre models. For example, it is well known that AR estimates converge to stable polynomials as the number of observations tends to infinity. See Söderström and Stoica (1989) for recent results. Hence, the Laguerre spectral factor estimate

$$\hat{H}(q) = \left[1 + \sum_{k=1}^{n} \hat{\alpha}_k L_k(q,a)\right]^{-1}, \tag{24}$$

where $\hat{\theta} = (\hat{\alpha}_1 \cdots \hat{\alpha}_n)'$ is defined by (15), will be asymptotically stable (for large N), since it will have the same poles as the corresponding AR estimate of $\{y_C(t,a)\}$.

The fundamental difference between the two model structures $\mathscr{M}$ and $\bar{\mathscr{M}}$ becomes clear when one tries to formulate asymptotic results for high-order models. As mentioned above, one gets into serious problems when trying to derive such results for $\bar{\mathscr{M}}$, while, as we shall show in Section 6, this can be done for the model structure $\mathscr{M}$.

3. The discrete Laguerre expansion. In this section we will give a detailed motivation for using Laguerre models. The following lemma provides an alternative series expansion of $[H^0(z)]^{-1}$.

LEMMA 3.1 (Discrete Laguerre expansion). *Assume the function $[H^0(z)]^{-1}$ to be analytic in $|z|>1$, continuous in $|z|\geq 1$ and normalized so that*

$[H^0(\infty)]^{-1} = 1$. *Let* $-1 < a < 1$. *Then there exists a sequence* $\{\alpha_k^0\}$ *such that*

$$(25) \qquad [H^0(z)]^{-1} = 1 + \sum_{k=1}^{\infty} \alpha_k^0 \frac{\sqrt{1-a^2}}{z-a}\left(\frac{1-az}{z-a}\right)^{k-1}, \qquad |z| \geq 1.$$

PROOF. The bilinear transformation

$$(26) \qquad w = \frac{z-a}{1-az} \quad \Leftrightarrow \quad z = \frac{w+a}{1+aw},$$

($|a| < 1$) maps the unit disc onto the unit disc. The discrete time system $[H^0(w+a)/(1-aw))]^{-1} - 1$ is thus analytic in $|w| > 1$, continuous in $|w| \geq 1$, and has at least one zero for $w = -1/a$ (since $[H^0(\infty)]^{-1} = 1$). Thus

$$(27) \qquad \left[H^0\left(\frac{w+a}{1+aw}\right)\right]^{-1} = 1 + \frac{a+w^{-1}}{\sqrt{1-a^2}} \sum_{k=1}^{\infty} \alpha_k^0 w^{-(k-1)}, \qquad |w| \geq 1.$$

Substituting back w using (26) now proves the lemma. □

The filters

$$(28) \qquad L_k(q,a) = \frac{\sqrt{1-a^2}}{q-a}\left(\frac{1-aq}{q-a}\right)^{k-1}, \qquad k \geq 1,$$

which consist of a first order low-pass term and $k-1$ all pass factors are called the discrete Laguerre filters. All pass filters are favorable in terms of numerical sensitivity, and are thus often recommended in filter design. See, for example, Section 8.6 in Lim and Oppenheim (1988) or Dewilde (1982) for details. The functions $\{L_k(z,a)\}$ are the Z-transforms of the discrete Laguerre functions $\{l_k(j,a)\}$,

$$(29) \quad l_k(j,a) = \sqrt{1-a^2} \sum_{l=0}^{k-1} \binom{k-1}{l}\binom{l+j-1}{k-1}(-1)^l a^{2l+j-k}, \qquad j,k \geq 1.$$

This set of functions is orthonormal, that is,

$$(30) \qquad \sum_{j=1}^{\infty} l_n(j,a) l_m(j,a) = \delta_{nm} \quad \Rightarrow$$

$$(31) \qquad \frac{1}{2\pi}\int_{-\pi}^{\pi} L_n(e^{i\omega},a) L_m(e^{-i\omega},a)\, d\omega = \delta_{nm},$$

where δ_{nm} is the Kronecker delta. For basic results on discrete Laguerre functions see Gottlieb (1938) or King and Paraskevopoulos (1977, 1979). General results on the classical continuous Laguerre functions can for example be found in Szegö (1939). More recent results related to approximation of linear systems are given in Mäkilä (1990a, b). A more control-oriented discussion of Laguerre models in system identification is given in Wahlberg (1991).

Since we work with normalized time, $t = 1, 2, 4 \dots$ ($t = t_{\text{true}}/\Delta$) and normalized frequency, $\omega \in [-\pi, \pi]$ ($\omega = \omega_{\text{true}}\Delta$), where Δ is the actual sampling interval, care has to be taken when interpreting the results to follow for a Laguerre parameter a close to 1. By relating the discrete time quantity to its corresponding continuous time counterpart more insight is obtained. A continuous time pole p is mapped to $a = \exp(\Delta p)$ when sampling a stochastic process. Hence, $p \cong (1-a)/\Delta$ for small Δ. In the limit as $\Delta \to 0$, $a \to 1$, but $(1-a)/\Delta$ converges to the corresponding continuous time pole p, which is the interesting quantity here.

Let $H^0(z)$ be a rational transfer function with zeros $\{z_i\}$. The rate of convergence for the Laguerre expansion is determined by the magnitude of the corresponding w-plane zeros:

$$\frac{z_i - a}{1 - a z_i}. \tag{32}$$

For w-plane zeros close to the unit circle the rate of convergence will be slow. To obtain a fast rate of convergence one must choose a close to the dominating zeros of $H^0(z)$. If a is too small compared to z_i, $(z_i - a)/(1 - az_i) \approx z_i$, which causes slow convergence if z_i is close to the unit circle. Taking a too large, that is, too close to 1 compared to z_i, $(z_i - a)/(1 - az_i) \approx -1$, also causes slow convergence. In case of scattered zeros of $H^0(z)$, the rate of convergence will thus be slow.

To improve the rate of convergence one can use more general orthonormal sets of base functions (the Kautz functions), corresponding to multiple time constants. Introducing operators with complex poles will remove slow convergence due to complex (resonant) zeros. This corresponds to replacing the Laguerre filters $L_k(q, a)$ by the Kautz filters

$$\Psi_k(q,b,c) = \begin{cases} \dfrac{\sqrt{1-c^2}\,(q-b)}{q^2 + b(c-1)q - c}\left[\dfrac{-cq^2 + b(c-1)q + 1}{q^2 + b(c-1)q - c}\right]^{(k-1)/2}, & k \text{ odd}, \\[2ex] \dfrac{\sqrt{(1-c^2)(1-b^2)}}{q^2 + b(c-1)q - c}\left[\dfrac{-cq^2 + b(c-1)q + 1}{q^2 + b(c-1)q - c}\right]^{(k-2)/2}, & k \text{ even}, \quad -1 < b < 1,\ -1 < c < 1. \end{cases} \tag{33}$$

Here, the coefficients b and c should be chosen so that the roots of $z^2 + b(c-1)z - c$ are close to the dominating complex zeros of the process to be modelled. The Kautz base functions $\{\Psi_k(z, b, c)\}$ are also orthonormal in L^2. This choice of base functions is useful for modelling narrow-band signals in noise. We refer to Wahlberg and Hannan (1991) for more details.

4. Toeplitz forms. It will be shown that the covariance matrix $\Gamma_n(a)$ of the regression vector, defined by (22), has a Toeplitz structure (the entries of the matrix are constant along each diagonal). This observation will be used to derive a Levinson algorithm to estimate θ. The following lemma explains the Toeplitz structure of $\Gamma_n(a)$.

LEMMA 4.1. *Let* $L_k(q, a)$ *be defined by* (9), *with* $-1 < a < 1$. *Then*

$$E\{L_j(q,a)y(t)L_k(q,a)y(t)\} = E\{\bar{y}(t-\tau)\bar{y}(t)\}, \qquad \tau = k-j, \tag{34}$$

where the process $\{\bar{y}(t)\}$ *has power spectral density*

$$\Phi_y\left(\frac{e^{i\omega}+a}{1+ae^{i\omega}}\right). \tag{35}$$

PROOF.

$$\begin{aligned} &E\{L_j(q,a)y(t)L_k(q,a)y(t)\} \\ &\quad = \frac{1}{2\pi}\int_{-\pi}^{\pi} L_j(e^{i\omega},a)L_k(e^{-i\omega},a)\Phi_y(e^{i\omega})\,d\omega \\ &\quad = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\bar{\omega}(j-k)}\Phi_y\left(\frac{e^{i\bar{\omega}}+a}{1+ae^{i\bar{\omega}}}\right)d\bar{\omega} \\ &\quad = E\{\bar{y}(t+(j-k))\bar{y}(t)\}, \end{aligned} \tag{36}$$

where the integration variable has been changed according to

$$e^{i\bar{\omega}} = \frac{e^{i\omega}-a}{1-ae^{i\omega}}. \tag{37}$$

□

REMARK 1. The mapping

$$e^{i\omega} \mapsto \frac{e^{i\omega}+a}{1+ae^{i\omega}} \tag{38}$$

is a standard tool in digital filter design to modify the bandwidth of low-pass filters. For $0 < a < 1$, it maps high frequencies onto lower ones. Consequently, this transformation makes the frequency content of the signal $\{y(t)\}$ appear "wider." We shall show that the transformation (38) improves the condition number of the $\Gamma_n(a)$ matrix.

By using post- and prewindowing of $y(t)$, that is, assuming $y(t) = 0$, $t \le 0$ or $t > N$, the sampled covariance matrix

$$\hat{\Gamma}_n(a) = \frac{1}{N}\sum_{t=1}^{\infty}\varphi(t,a)\varphi(t,a)' \tag{39}$$

will have a Toeplitz structure. This can be seen as follows. If

$$I(e^{i\omega}) = \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{i\omega t}\right|^2, \tag{40}$$

then

$$\left[\hat{\Gamma}_n\right]_{j,k} = \frac{1}{2\pi}\int_{-\pi}^{\pi} L_j(e^{i\omega}, a) L_k(e^{-i\omega}, a) I(e^{i\omega})\, d\omega \tag{41}$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\bar{\omega}(j-k)} I\left(\frac{e^{i\bar{\omega}} + a}{1 + ae^{i\bar{\omega}}}\right) d\bar{\omega},$$

which shows that the above assumptions are appropriate. Notice that the elements of $\varphi(t, a)$ are known, simple analytic functions of $y(t)$, $t = 1 \cdots N$ and a, so the infinite sums above can be computed exactly. Since $y(t)$ is assumed to equal 0 for $t > N$ and the Laguerre filters are asymptotically stable, the filtered variables $u_k(t, a) \to 0$ exponentially fast as $t \to \infty$. Hence, by taking the upper limit in (39) large but finite the errors due to truncation would be negligible. For large N the effects from truncating the sums will be small, since the truncation errors are normalized by N.

Let us introduce the following notation:

$$u_0(t, a) = y(t), \qquad y(t) = 0,\ t \leq 0 \text{ or } t > N, \tag{42}$$

$$u_k(t, a) = L_k(q, a) y(t), \qquad k \geq 1,\ y(t) = 0,\ t \leq 0 \text{ or } t > N, \tag{43}$$

$$\hat{c}_{j-k}(a) = \frac{1}{N}\sum_{t=1}^{\infty} u_j(t, a) u_k(t, a), \qquad j \geq 1,\ k \geq 1, \tag{44}$$

$$\hat{d}_j(a) = \frac{1}{N}\sum_{t=1}^{N} u_0(t, a) u_j(t, a), \qquad j \geq 1, \tag{45}$$

$$\hat{\gamma}_n(a) = \left(\hat{d}_1(a) \cdots \hat{d}_n(a)\right)'. \tag{46}$$

Notice that $\hat{c}_{j-k}(a)$ is the (j, k) element of $\hat{\Gamma}_n(a)$. We shall write

$$c_{j-k}(a) = E\{u_j(t, a) u_k(t, a)\}, \tag{47}$$
$$d_j = E\{u_0(t, a) u_j(t, a)\} = E\{y(t) u_j(t, a)\}.$$

Since a is assumed to be fixed in the algorithm to follow, the argument a will be omitted in the remainder of this section. The post- and prewindowed LS estimate is given by the normal equation

$$\hat{\Gamma}_n \hat{\theta}_n = -\hat{\gamma}_n. \tag{48}$$

Using the Toeplitz structure of $\hat{\Gamma}$ and the fact that $\hat{\gamma}_n = (\hat{\gamma}_{n-1}' \ \hat{d}_n)'$ it is possible to solve the normal equation (48) in a computationally efficient way.

A Levinson type algorithm. The Levinson type recursion for estimation of $\theta_n = (\alpha_{n,1} \cdots \alpha_{n,n})'$ (here we have indicated the order n by an extra subscript) is

$$(49)\qquad \hat{\alpha}_{n,j} = \hat{\alpha}_{n-1,j} + \hat{\alpha}_{n,n}\hat{\beta}_{n-1,n-j}, \qquad \hat{\alpha}_{n,0} = 1,$$

$$(50)\qquad \hat{\beta}_{n,j} = \hat{\beta}_{n-1,j} + \hat{\beta}_{n,n}\hat{\beta}_{n-1,n-j}, \qquad \hat{\beta}_{n,0} = 1,\ \hat{\beta}_{n,j} = 0,\ j > n,$$

$$(51)\qquad \hat{\beta}_{n,n} = \sum_{j=0}^{n-1} \hat{\beta}_{n-1,j}\hat{c}_{n-j}/\hat{\sigma}_{n-1}^2,$$

$$(52)\qquad \hat{\alpha}_{n,n} = \sum_{j=0}^{n-1} \hat{\beta}_{n-1,j}\hat{d}_{n-j}/\hat{\sigma}_{n-1}^2,$$

$$(53)\qquad \hat{\sigma}_n^2 = \left(1 - \hat{\beta}_{n,n}^2\right)\hat{\sigma}_{n-1}^2, \qquad \hat{\sigma}_0^2 = \hat{c}_0.$$

The proof is given in the Appendix.

REMARK 2. The algorithm can be viewed as first solving the standard Yule–Walker problem

$$(54)\qquad \hat{\Gamma}_n\left(\hat{\beta}_{n,1}\dots\hat{\beta}_{n,n}\right)' = -(\hat{c}_1\dots\hat{c}_n)'.$$

The solution to $\hat{\Gamma}_n\hat{\theta}_n = -\hat{\gamma}_n$, that is, a general right-hand side, is then obtained by modifying the Yule–Walker solution. We refer to Chapter 4.7 of Golub and Van Loan (1989) for details.

REMARK 3. It seems difficult to find simple lattice procedures. However, there are connections between the theory given herein and realization theory based on the Schur algorithm. See, for example, Dewilde and Dym (1981).

5. Statistical properties. Here we shall analyze the special case where $\{y(t)\}$ is generated by a true system (11) with $n = n_0 < \infty$, $a = a_0$ with $|a_0| \le 1 - \delta$, $\delta > 0$ and $\theta = \theta^0$. We assume that δ is known, that is, we can estimate a by optimizing over $a \in [-1 + \delta, 1 - \delta]$ for some $\delta > 0$. Although $n_0 < \infty$ will not in general hold, a value $\hat{n}$ will have to be chosen in practice; for such a suitably chosen $\hat{n}$, a system of that order will provide an excellent approximation to the truth. We use $\hat{a}_n$, $\hat{\theta}_n$ and $\hat{n}$, for the estimated values.

The Laguerre time constant a can be estimated by minimizing $\hat{\sigma}_n^2(a)$ given by (53) with respect to a, or equally,

$$(55)\qquad \hat{a}_n = \arg\min_a \left[\frac{1}{N}\sum_{t=1}^{N} y(t)^2 - \hat{\gamma}_n(a)'\hat{\Gamma}_n(a)^{-1}\hat{\gamma}_n(a)\right] \quad \Rightarrow$$

$$(56)\qquad \hat{a}_n = \arg\max_a \left[\hat{\gamma}_n(a)'\hat{\Gamma}_n(a)^{-1}\hat{\gamma}_n(a)\right].$$

Given $\hat{a}_n$, the estimate $\hat{\theta}_n$ is obtained as described in the preceding section, with $a = \hat{a}_n$. We choose $\hat{n}$ to minimize the criterion

$$\text{(57)} \qquad \mathrm{BIC}(n) = \log \hat{\sigma}_n^2(\hat{a}_n) + \frac{n \log N}{N}, \qquad n \leq M.$$

M may be arbitrarily large. For a discussion of (57) see Hannan and Deistler (1988).

THEOREM 5.1. *Under the above conditions, $\hat{n} \to n_0$ a.s. so that $\hat{n} = n_0$ for $N \geq N_0$, where $P(N_0 < \infty) = 1$. Also $(\hat{a}_n, \hat{\theta}_n) \to (a_0, \theta^0)$ a.s. and, indeed, $|\hat{a}_n - a_0| = \mathscr{O}(Q_N)$ a.s., $\|\hat{\theta}_n - \theta^0\|_1 = \mathscr{O}(Q_N)$, a.s., where $Q_N = [\log\log N/N]^{1/2}$ and $n = \hat{n}$.*

The proof is given in the Appendix.

REMARK 1. N_0 is not, of course, a stopping time, that is, one cannot tell whether $N \geq N_0$ by examining the history of $t \leq N$. By $\mathscr{O}(Q_N)$ we mean, for example, that $|\hat{a}_n - a_0|/Q_N$ is a.s. a bounded sequence ($N \geq 3$).

Let $\tau_n = [\theta_n^*(a)', a]'$, where

$$\text{(58)} \quad \theta_n^*(a) = \arg\min_{\theta} E\left\{[y(t) - \varphi_n(t,a)'\theta]^2\right\}, \quad \text{for given } a \text{ and order } n.$$

Let

$$\text{(59)} \quad \eta_n(a) = \gamma_n(a)'\Gamma_n(a)^{-1}\gamma_n(a), \qquad \hat{\eta}_n(a) = \hat{\gamma}_n(a)'\hat{\Gamma}_n(a)^{-1}\hat{\gamma}_n(a),$$

where

$$\text{(60)} \quad \Gamma_n(a) = E\{\varphi(t,a)\varphi(t,a)'\}, \qquad \gamma_n(a) = -E\{\varphi(t,a)y(t)\}.$$

We shall write $\hat{\tau}_n = [\hat{\theta}_n', \hat{a}_n]'$. For brevity write $\hat{\tau}_0 = \hat{\tau}_{n0}$ and so also for $\hat{\theta}_0$, $\hat{a}_0$, $\eta_0(a)$, $\hat{\eta}_0(a)$ and so on. It will be sufficient to establish a central limit theorem at $\hat{n} = n_0$ for it can be shown that there is, for every $\varepsilon > 0$, a set Ω_1 with $P(\Omega_1) \geq 1 - \varepsilon$, in the sample space of all realizations of $\{y(t)\}$, on which $\hat{n} = n_0$ for $N \geq N_0 < \infty$. The same central limit theorem holds therefore for $\hat{a}_n$ as well as $\hat{\theta}_n$.

THEOREM 5.2. *Under the assumptions above, $\sqrt{N}(\hat{\tau}_0 - \tau_0)$ has a distribution converging to a multivariate Gaussian law with zero mean and covariance matrix*

$$\text{(61)} \quad P = \sigma_0^2 \begin{pmatrix} \Gamma_0(a_0) & E\{\varphi_0(t,a_0)\psi_0(t,a_0)\} \\ E\{\varphi_0(t,a_0)'\psi_0(t,a_0)\} & E\{\psi_0(t,a_0)^2\} \end{pmatrix}^{-1},$$

where

$$\text{(62)} \qquad \psi_0(t,a_0) = \frac{\partial}{\partial a}\varphi_0(t,a)'\theta, \quad \text{at } a = a_0,\ \theta = \theta^0,$$

$$\text{(63)} \qquad \sigma_0^2 = E\left\{e(t)^2|\mathscr{F}_{t-1}\right\}.$$

PROOF. Theorem 5.2 follows directly from Theorem 9.1 and Expression (9.17) in Ljung (1987). □

Finally in this section we point out the following. Let us consider the determination of $\hat{a}_n$ by an iteration. Since eventually $\hat{n} = n_0$ and since we do not need to actually determine $\hat{a}_n$ – but only (sufficiently accurately) the maximum of $\hat{\eta}_n(a)$ – to evaluate $\hat{\sigma}_n^2$ it is, in principle, sufficient to discuss $n = n_0$. A natural iteration is

$$\hat{a}_0^{(k+1)} = \hat{a}_0^{(k)} - \frac{\hat{\eta}_0'\big(\hat{a}_0^{(k)}\big)}{\hat{\eta}_0''\big(\hat{a}_0^{(k)}\big)}, \tag{64}$$

where now we use $\hat{\eta}_0'(a)$ for the derivative, for notational simplicity.

THEOREM 5.3. *There is an open interval,* **I**, *about* a_0 *and a random time* N_0, $P(T_0 < \infty) = 1$, *such that if* $\hat{a}_0^{(1)} \in \mathbf{I}$ *and* $N \geq N_0$ *then* $\hat{a}_o^{(k)} \to \hat{a}$ *a.s.*

The proof is given in the Appendix.

REMARK 2. Of course (64) is used at $n = \hat{n}$ but again for N_0 sufficiently large we have $\hat{n} = n_0$, so the theorem applies.

6. High-order properties. The assumption that $H^0(z)$ can be exactly represented by a finite order Laguerre model is, of course, not realistic in practice. To improve the flexibility of the model, the order can be increased. Hence, it is of importance to analyze the properties of high model orders. In this section the order is allowed to increase with the number of observations, that is, we will generalize the high-order AR convergence results of Berk (1974) to cover Laguerre models.

6.1. *Convergence of generalized covariance functions.* A fundamental difference between the AR case ($a = 0$) and the general Laguerre case is the behaviour of $\sum_{j=1}^{\infty}|l_k(j,a)|$. For $a = 0$ this sum equals 1, while for $a \neq 0$ we have to rely on the following bound:

$$\sum_{j=1}^{\infty} |l_k(j,a)| \leq 2k \sup_{\omega} |L_k(e^{i\omega}, a)|. \tag{65}$$

This bound follows from the facts that the L^1-norm of a kth order stable system is less than two times the nuclear norm, and that the nuclear norm is less than k times the L^∞ norm. The corresponding continuous time result is proved in Glover, Curtain and Partington (1988). The proof of the discrete time result is similar. For completeness, a brief proof is included in the Appendix.

Using the above bound the following result is immediate.

LEMMA 6.1. *Let $c_j(a)$, $\hat{c}_j(a)$, $d_j(a)$ and $\hat{d}_j(a)$ be defined as in Section* 2.2. *Under the conditions* (2) *and* (6) *we have*

$$\sup_{0 \le j \le (n-1)} |\hat{c}_j(a) - c_j(a)| = \mathscr{O}\big(n(\log N/N)^{1/2}\big), \quad a.s., \tag{66}$$

$$\sup_{1 \le j \le n} |\hat{d}_j(a) - d_j(a)| = \mathscr{O}\big(n(\log N/N)^{1/2}\big), \quad a.s. \tag{67}$$

The proof is given in the Appendix.

6.2. *Grenander–Szegö bounds.* Using Lemma 4.1 and the well known bound for the eigenvalues of Toeplitz matrices given in Grenander and Szegö (1959), it is easy to establish that

$$\min_{-\pi \le \overline{\omega} \le \pi} \Phi_y\left(\frac{e^{i\overline{\omega}} + a}{1 - ae^{i\overline{\omega}}}\right) \le \lambda_j(\Gamma_n(a)) \le \max_{-\pi/ \le \overline{\omega} \le \pi} \Phi_y\left(\frac{e^{i\overline{\omega}} + a}{1 + ae^{i\overline{\omega}}}\right) \tag{68}$$

$$\Leftrightarrow$$

$$\min_{-\pi \le \omega \le \pi} \Phi_y(e^{i\omega}) \le \lambda_j(\Gamma_n(a)) \le \max_{-\pi/ \le \omega \le \pi} \Phi_y(e^{i\omega}), \tag{69}$$

where $\lambda_j(\Gamma_n(a))$ denotes the jth eigenvalue of $\Gamma_n(a)$. This shows that $\Gamma_n(a)$ is well conditioned independently of the order n, provided that the spectral density satisfies $0 < C_l \le \Phi_y(e^{i\omega}) \le C_u$, $\forall\ \omega$.

The following theorem gives a n-dependent lower bound that may be better than the one given above, especially if the power spectral density is small for high frequencies.

THEOREM 6.1. *Let $\Gamma_n(a)$ be defined by* (22). *For ε such that $\varepsilon n/\pi < 1$, we have*

$$(1 - \varepsilon n/\pi) \min_{-(\pi-\varepsilon) \le \omega \le (\pi - \varepsilon)} \Phi_y\left(\frac{e^{i\overline{\omega}} + a}{1 + ae^{i\overline{\omega}}}\right) \le \lambda_j(\Gamma_n(a)). \tag{70}$$

The proof is given in the Appendix.

REMARK 1. By restricting the frequency interval to $-(\pi - \varepsilon) \le \overline{\omega} \le (\pi - \varepsilon)$, we can take advantage of the mapping (38). The minimum over this frequency range may be much larger than the global minimum, since in fact [due to the mapping (38)] we only search over low and medium frequencies. The extra price we have to pay is the n-dependent factor in the lower bound, which decreases as n increases.

6.3. *Error bounds.* In this section we shall derive a bound on the error in the approximations obtained by fitting an nth order Laguerre model to a process that may be of infinite order. Since a here is assumed to be fixed, we shall drop the a argument, wherever it causes no confusion.

Let $H_n(e^{i\omega})$ denote the spectral factor associated with θ_n. For $\theta_n = \theta_n^*$ [given by (58)] we write $H_n^*(e^{i\omega})$ for the corresponding spectral factor. By θ_n^0, we mean $(\alpha_1^0 \cdots \alpha_n^0)'$, that is, the first n Laguerre parameters of $[H^0(z)]^{-1}$.

First, let us give a frequency domain interpretation of $H_n^*(e^{i\omega})$. Using $H_n(\infty) = H^0(\infty) = 1$ and Lemma 4.1 it is easy to establish that

$$\begin{aligned}
&E\{[y(t) - \varphi(t,a)'\theta_n]^2\} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} |H_n(e^{i\omega})|^{-2}\Phi_y(e^{i\omega})\,d\omega \\
&= \sigma_0^2\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|H^0(e^{i\omega}) - H_n(e^{i\omega})|^2}{|H_n(e^{i\omega})|^2}\,d\omega + 1\right) \qquad (71)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} |H^0(e^{i\omega})^{-1} - H_n(e^{i\omega})^{-1}|^2\Phi_y(e^{i\omega})\,d\omega + \sigma_0^2 \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{k=1}^{\infty}\alpha_k^0 e^{-i\bar{\omega}k} - \sum_{k=1}^{n}\alpha_k e^{-i\bar{\omega}k}\right|^2 \Phi_y\left(\frac{e^{i\bar{\omega}} + a}{1 + ae^{i\bar{\omega}}}\right)d\bar{\omega} + \sigma_0^2.
\end{aligned}$$

Recall that θ_n^* minimizes (71). Hence, the elements of θ_n^* equal the Fourier coefficients of the optimal nth order weighted L^2 approximation of the function $\sum_{k=1}^{\infty}\alpha_k^0 e^{-i\bar{\omega}k}$. The weighting function is the transformed spectral density (35).

Based on the work of Baxter (1963), we shall derive a bound on $\|\theta_n^0 - \theta_n^*\|_1$.

LEMMA 6.2. *Assume that*

$$\left[H^0\left(\frac{w+a}{1+aw}\right)\right]^{-1} = \sum_{k=0}^{\infty}\tilde{a}_k^0 w^{-k}, \qquad \sum_{k=0}^{\infty}|\tilde{a}_k^0| < \infty. \tag{72}$$

Define $\theta_n^0 = (\alpha_1^0 \dots \alpha_n^0)'$, *and let* θ_n^* *be given by* (58). *Then*

$$\|\theta_n^0 - \theta_n^*\|_1 \le M\sum_{j=n+1}^{\infty}|\alpha_j^0|, \quad \textit{for } n > \tilde{n} \tag{73}$$

and

$$\begin{aligned}
&\sup_{\omega}\left|H^0(e^{i\omega})^{-1} - H_n^*(e^{i\omega})^{-1}\right| \\
&\le (M+1)\left(\frac{1+|a|}{1-|a|}\right)^{1/2}\sum_{j=n+1}^{\infty}|\alpha_j^0|, \quad \textit{for } n > \tilde{n},
\end{aligned} \tag{74}$$

with $\tilde{n}$ *such that*

$$\left\|H^0\left(\frac{w+a}{1+aw}\right)\sum_{k=\tilde{n}+1}^{\infty}\tilde{a}_k^0 w^{-k}\right\|_1 \le \varepsilon < 1 \tag{75}$$

and with

$$M = \frac{3-\varepsilon}{1-\varepsilon}\left\|\left[\sigma_0 H^0\left(\frac{w+a}{1+aw}\right)\right]^{-1}\right\|_1^2 \left\|\Phi_y\left(\frac{w+a}{1+aw}\right)\right\|_1. \tag{76}$$

For functions $\|\cdot\|_1$ *denotes the* L^1 *norm, for example,* $\|[H^0(w+a)/(1+aw)]^{-1}\|_1 = \sum_{k=0}^{\infty}|\tilde{a}_k^0|$.

The proof is given in the Appendix.

REMARK 2. We have chosen to state the conditions in Lemma 6.2 in terms of $\tilde{a}_k^0$. However, it is easy to verify that

$$\tilde{a}_0^0 = 1 + \frac{a\alpha_1^0}{\sqrt{1-a^2}}, \qquad \tilde{a}_k^0 = \frac{\alpha_{k-1}^0 + a\alpha_k^0}{\sqrt{1-a^2}}, \qquad k \geq 1. \tag{77}$$

Hence, Lemma 6.2 can equally well be formulated in the Laguerre parameters. Observe that

$$\sum_{k=1}^{\infty}|\alpha_k^0| < \infty \quad \Leftrightarrow \quad \sum_{k=0}^{\infty}|\tilde{a}_k^0| < \infty. \tag{78}$$

REMARK 3. Condition (72) is crucial for the above result. From Kahane (1956) it is known that in general $\|[H^0(z)]^{-1}\|_1 < \infty$ does not imply that $\|[H^0((w+a)/(1+aw))]^{-1}\|_1 < \infty$ (only affine mappings guarantee absolute convergence). However, the assumption (6) implies that $[H^0((w+a)/(1+aw))]^{-1} \in \mathrm{Lip}(1/2+\delta)$ and thus that [see Zygmund (1968)] $\|[H^0((w+a)/(1+aw))]^{-1}\|_1 < \infty$.

6.4. *Consistency.* We are now able to prove the following result on consistency of the high-order LS Laguerre model estimate.

THEOREM 6.2. *Let* $y(t)$ *be generated by* (1) *with* (2) *and* (6) *holding. Then for* $n = n(N)$ *such that* $n^2(\log N/N)^{1/2} \to 0$ *as* $N \to \infty$,

$$\|\hat{\theta}_n - \theta_n^*\|_2 = \mathscr{O}\left(n^{3/2}(\log N/N)^{1/2}\right) \quad a.s., \tag{79}$$

$$\|\hat{\theta}_n - \theta_n^*\|_1 = \mathscr{O}\left(n^2(\log N/N)^{1/2}\right) \quad a.s., \tag{80}$$

and

$$\|\hat{\theta}_n - \theta_n^0\|_1 = \mathscr{O}\left(n^{3/2}(\log N/N)^{1/2}\right) + \mathscr{O}\left(\sum_{j=n+1}^{\infty}|\alpha_j^0|\right) \quad a.s. \tag{81}$$

The proof is given in the Appendix.

The convergence of the variance estimate $\hat{\sigma}_n^2$ is easy to establish from (81). It is also straightforward to prove the consistency (uniformly in ω) of the spectral density estimate

$$\left[\hat{H}_n(e^{i\omega})\right]^{-1} = 1 + \sum_{k=1}^{b} \hat{\alpha}_k L_k(e^{i\omega}, a), \qquad \hat{\theta}_n = (\hat{\alpha}_1 \cdots \hat{\alpha}_n)', \tag{82}$$

and thus also the corresponding spectral density estimate

$$\hat{\Phi}_y(e^{i\omega}) = \hat{\sigma}_n^2 |\hat{H}_n(e^{i\omega})|^2. \tag{83}$$

Notice that

$$\begin{aligned} &\left|\left[\hat{H}_n(e^{i\omega})\right]^{-1} - \left[H^0(e^{i\omega})\right]^{-1}\right| \\ &\quad \le \max_k \left|L_k(e^{i\omega}, a)\right| \left(\|\hat{\theta}_n - \theta_n^0\|_1 + \sum_{j=n+1}^{\infty} |\alpha_j^0| \right) \end{aligned} \tag{84}$$

and

$$|\hat{H}_n(e^{i\omega}) - H^0(e^{i\omega})| \le |\hat{H}_n(e^{i\omega})|\, |H^0(e^{i\omega})| \left|\left[\hat{H}_n(e^{i\omega})\right]^{-1} - \left[H^0(e^{i\omega})\right]^{-1}\right|. \tag{85}$$

6.5. Asymptotic distribution and variance. Having investigated the almost sure convergence issues, we now turn to the asymptotic distribution of the estimates.

Introduce the row vector

$$W_n(\omega) = \left[L_1(e^{i\omega}, a) \cdots L_n(e^{i\omega}, a)\right], \tag{86}$$

so that the estimate of the inverse spectral factor can be written as

$$\hat{H}_n(e^{i\omega})^{-1} = 1 + W_n(\omega)\hat{\theta}_n. \tag{87}$$

Under the assumptions of Theorem 5.2 with $a = a_0$, $n \ge n_0$, we have

$$\sqrt{\frac{N}{n}} \begin{pmatrix} \hat{H}(e^{i\omega_1})^{-1} - H^0(e^{i\omega_1})^{-1} \\ \hat{H}(e^{i\omega_2})^{-1} - H^0(e^{i\omega_2})^{-1} \end{pmatrix} \sim \mathrm{AsN_c}(0, P_n(\omega_1, \omega_2)), \tag{88}$$

with

$$[P_n(\omega_1, \omega_2)]_{ij} = \frac{\sigma_0^2}{n} W(\omega_i)\Gamma_n(a)^{-1} W(-\omega_j)', \qquad i, j = 1, 2. \tag{89}$$

Here $\mathrm{AsN_c}$ means a complex normal asymptotic distribution. In the preceding results we have assumed that $y(t)$ is generated by a process with fixed order n_0, and known $a = a_0$. These assumptions are made to simplify the analysis, and are of course unrealistic but adequate. The correct choice of a is not crucial since $[H^0(z)]^{-1}$ can be arbitrary well approximated by an Laguerre model for any choice of a ($|a| < 1$) by just increasing the order n. By fixing a and taking n large, we can assume that the model set is flexible enough to include an good approximation of the true process. These heuristic arguments can be justified more rigorously by the following observations: Assume that

$y(t)$ is generated by (1). Theorem 9.1 in Ljung (1987) then shows that the covariance matrix in the asymptotic distribution of the error $N^{1/2}(\hat{\theta}_n - \theta_n^*)$ (for fixed n) should be modified to

$$\Gamma_n(a)^{-1} Q_n(a) \Gamma_n(a)^{-1}, \tag{90}$$

where

$$\begin{aligned} Q_n(a) = \lim_{N\to\infty} NE\Bigg\{&\left[\frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a)[H_n^*(q,a)]^{-1} y(t)\right] \\ &\times \left[\frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a)[H_n^*(q,a)]^{-1} y(t)\right]'\Bigg\}. \end{aligned} \tag{91}$$

Compare (139). Now $[H^0(q)]^{-1} y(t) = e(t)$ and thus

$$\frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a)[H_n^*(q,a)]^{-1} y(t) = \frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a) e(t) + \Delta_n, \tag{92}$$

$$\Delta_n = \frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a)\left([H_n^*(q,a)]^{-1} - [H^0(q)]^{-1}\right) y(t). \tag{93}$$

The matrix $\Gamma_n(a)^{-1} Q_n(a) \Gamma_n(a)^{-1}$ can now be rewritten as

$$\begin{aligned} \Gamma_n(a)^{-1} Q_n(a) \Gamma_n(a)^{-1} &= \sigma_0^2 \Gamma_n(a)^{-1} \\ &\quad + \Gamma_n(a)^{-1} \lim_{N\to\infty} E\{N\Delta_n \Delta_n'\} \Gamma_n(a)^{-1}. \end{aligned} \tag{94}$$

Notice that Δ_n is proportional to the error $[H_n^*(q,a)]^{-1} - [H^0(q)]$, which tends to 0 as $n \to \infty$. The L^2 norm of the second term in (94) can be shown to tend to 0 as $n \to \infty$ under more restrictive conditions on the system and the moments of the noise $e(t)$ than have been assumed so far. Since this analysis is quite tedious we shall not pursue it here.

By letting $n \to \infty$ the frequency domain covariance expression (88) simplifies considerably, as is shown by the following result.

THEOREM 6.3. *Assume the spectral density of* $\{y(t)\}$ *to be continuous and bounded away from* 0:

$$\Phi_y(e^{i\omega}) \geq \delta > 0, \qquad \forall\, \omega. \tag{95}$$

Then $(i, j = 1, 2)$

$$\lim_{n\to\infty} [P_n(\omega_1, \omega_2)]_{ij} = \begin{cases} 0, & \text{if } i \neq j, \\ \dfrac{(1-a^2)}{|e^{i\omega_1} - a|^2} \sigma_0^2 \left[\Phi_y^0(e^{i\omega_1})\right]^{-1}, & \text{if } i = j. \end{cases} \tag{96}$$

The proof is given in the Appendix.

Theorem 6.3 directly implies the following result for the spectral density estimate.

COROLLARY 6.1. *Under the assumptions of Theorem* 5.2 *with* $a = a_0$, $n \geq n_0$, $\hat{\Phi}_y(e^{i\omega}) = \hat{\sigma}_n^2 |\hat{H}_n(e^{i\omega})|^2$, *we have*

$$\sqrt{\frac{N}{n}}\left(\left[\hat{\Phi}_y(e^{i\omega_1}) - \Phi_y(e^{i\omega_1})\right], \left[\hat{\Phi}_y(e^{i\omega_2}) - \Phi_y(e^{i\omega_2})\right]\right)' \sim \mathrm{AsN_c}(0, R_n(\omega_1, \omega_2)), \tag{97}$$

where

$$[R_n(\omega_1, \omega_2)]_{ij} \to \begin{cases} 0, & \text{if } i \neq j, \\ \dfrac{(1-a^2)}{|e^{i\omega_i} - a|^2} 2\left[\Phi_y(e^{i\omega_i})\right]^2, & \text{if } i = j,\ \omega_i \neq 0, \pi, \\ \dfrac{(1-a^2)}{|e^{i\omega_i} - a|^2} 4\left[\Phi_y(e^{i\omega_i})\right]^2, & \text{if } i = j,\ \omega_i = 0, \pi, \end{cases} \tag{98}$$

as $n \to \infty$.

The result of Corollary 6.1 is unexpected. The factor

$$\frac{(1-a^2)}{|e^{i\omega_1} - a|^2} \tag{99}$$

does not appear in the corresponding AR result, given in Berk (1974). This factor is the square amplitude of a first order low-pass filter with a pole at a and the gain $[(1+a)/(1-a)]^{1/2}$. This means that the variance for higher frequencies will be reduced compared to the variance for lower frequencies. The factor makes sense since we have indirectly assumed $[H^0(e^{i\omega})]^{-1} - 1$ to be small (by choosing $a > 0$) for high frequencies. Thus the absolute variance for high frequencies should decrease. It is also important to remember that we have normalized the variance by the "order" n, which is closely related to the choice of a. Taking a too large or too small means that we have to increase the order n to reduce the bias. Consequently, a cannot be seen as a direct design variable at this stage.

7. Examples. In this section we shall give two simple examples that illustrate the advantage of choosing $a > 0$, that is, using a general Laguerre network instead of an AR model.

EXAMPLE 7.1. Consider the continuous time process

$$v(t) = \frac{1}{p+1} e_c(t), \tag{100}$$

where $\{e_c(t)\}$ is a continuous time white noise process, with incremental variance $0.5\,dt$. Here p denotes the differential operator. Assume that $\{v(t)\}$ is measured together with additive white measurement noise with variance

σ_m^2, that is,

(101) $$y(t) = v(t) + e_m(t), \qquad t = k\Delta, k = 1, \dots, N.$$

The spectral representation of the discrete time process $\{y(t)\}$ then equals

(102) $$y(t) = \frac{q + c_1^0}{q + a_1^0} e(t), \qquad E\{e^2(t)\} = \sigma_0^2.$$

Choosing $\sigma_m^2 = 3.57$ and $\Delta = 0.1$ gives

(103) $$c_1^0 = -0.82, \qquad a_1^0 = -0.90, \qquad \sigma_0^2 = 1.83.$$

To emphasize the bias effects, the limiting estimates θ_n^* [given by (71)] and the corresponding spectral factor approximation $H_n^*(e^{i\omega})$ are calculated. Figure 2 shows the true spectral density together with the approximations using

(104) $$[a = 0, n = 8], \qquad [a = 0.5, n = 4], \qquad [a = 0.75, n = 2].$$

The orders have been chosen via the BIC criterion using a single realization with $N = 10^4$. By choosing $a = 0.82$, the system can be described within the chosen model set. Hence, this choice is not shown in the figure.

To study the variance aspects, the simplified variance expression (96), which is also asymptotic in the model order, is compared with the more complicated finite model order variance expression (89). To avoid effects from under modelling, $a = 0.82$. The results for $n = 1$ and $n = 5$ are shown in

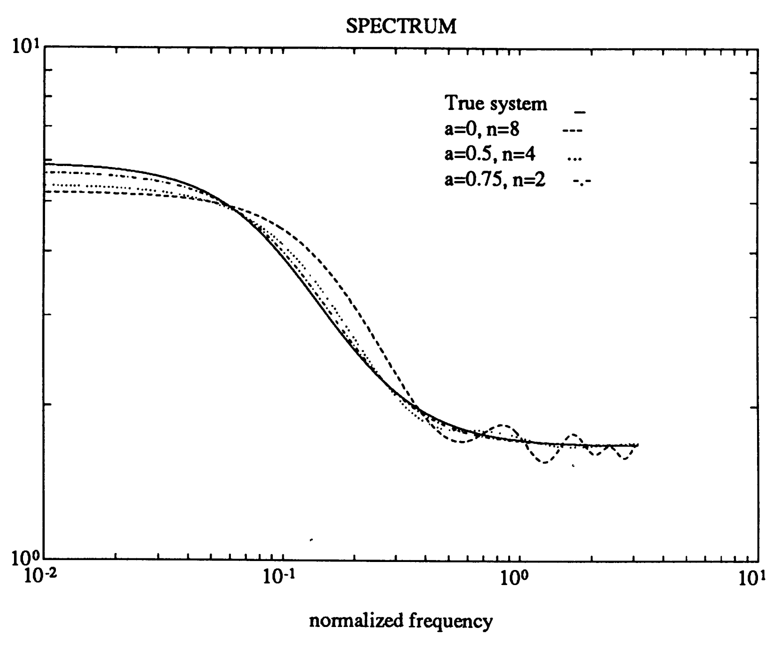

FIG. 2. *Example* 5.1: *Spectral density approximations.*

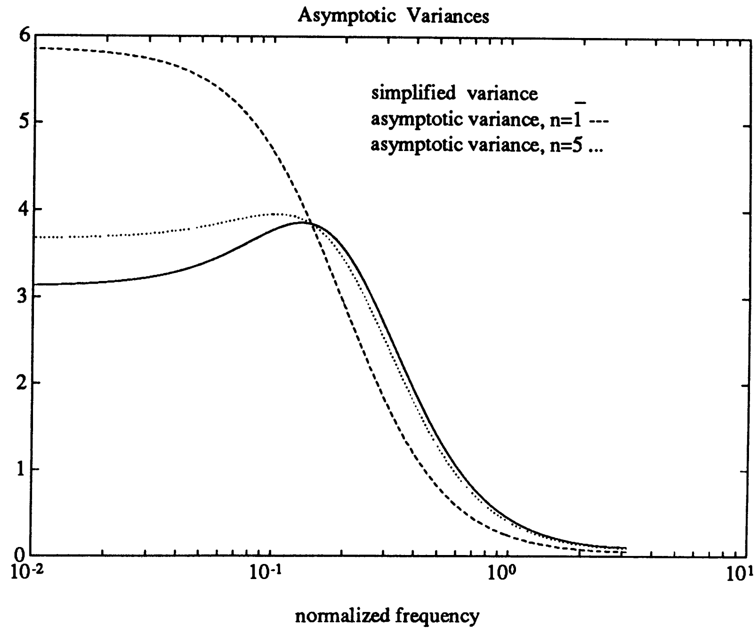

FIG. 3. *Comparison of the asymptotic variance expressions* (90) *and* (97), *for* $a = 0.82$.

Figure 3. As can be seen, the simplified high-order variance expression is a reasonable approximation even for moderate large values of n ($n = 5$). Notice the influence of the low-pass filter (99), which gives a lower absolute variance for higher frequencies.

EXAMPLE 7.2. Replace the low-frequency process in Example 5.1 with the more narrow band process

$$v(t) = \frac{p+1}{(p+5)^2} e_c(t), \tag{105}$$

where $\{e_c(t)\}$ is a continuous time white noise process, with incremental variance $\sigma_c^2\, dt$. The corresponding sampled process can be well approximated (for small sampling periods Δ) with the discrete time process

$$v(t) = \frac{q(q - e^{-\Delta})}{(q - e^{-5\Delta})^2} e_\Delta(t), \qquad E\{e_\Delta^2(t)\} = \frac{\sigma_c^2}{\Delta}. \tag{106}$$

See Wahlberg (1988) for details. For $\sigma_c^2 = 10$, $\sigma_m^2 = 1$ and $\Delta = 0.1$ the following discrete time process will be observed:

$$y(t) = v(t) + e_m(t) = \frac{(q + c_1^0)(q + c_2^0)}{(q + a_1^0)^2} e(t), \tag{107}$$

$$c_1^0 = -0.84, \qquad c_2^0 = -0.21, \qquad a_1^0 = -0.61, \qquad \sigma_0^2 = 2.07. \tag{108}$$

In Figure 4 the optimal (w.r.t. a) second order Laguerre spectral density approximation ($a_{\mathrm{opt}} = 0.55$) is compared to a second order AR ($a = 0$) estimate. As we can see the AR estimate is unable to account for the low frequency behavior of the process, while the Laguerre approximation makes a compromise between good fit for high and low frequencies. Notice that the true process has two rather scattered zeros (0.84 and 0.21). Hence it is difficult to describe the process using one single Laguerre time constant. By using two Laguerre time constants we can of course expect perfect fit. The results for fifth order ($a_{\mathrm{opt}} = 0.58$) models are shown in Figure 5. Here we obtain very good fit for the optimal Laguerre model, while we still have a considerable model mismatch for low frequencies using an AR model. This illustrates the advantage of adapting the Laguerre time constant, instead of just fixing it to zero as in the AR case.

8. Conclusions. Properties of discrete Laguerre filter representations of time series have been investigated. By appropriate choice of Laguerre time constant the number of parameters needed to obtain useful approximations can be considerably reduced compared to AR modelling. Several results on AR parameter estimation have been generalized to Laguerre models, such as asymptotic statistical properties and the Levinson algorithm. The key observation is that the general Laguerre case can be transformed to the AR case using Lemma 4.1. This makes it possible to generalize AR results to Laguerre modelling.

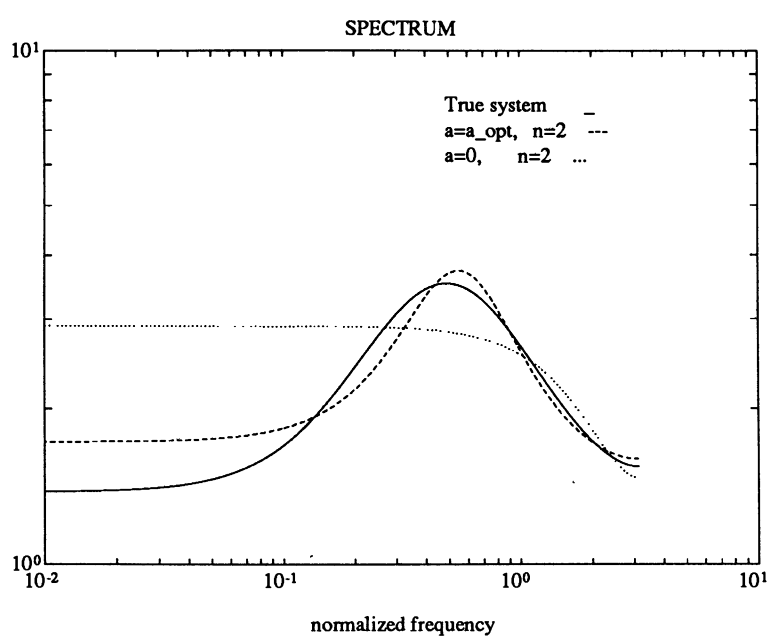

FIG. 4. *Example* 5.2: *Second order spectral density approximations.*

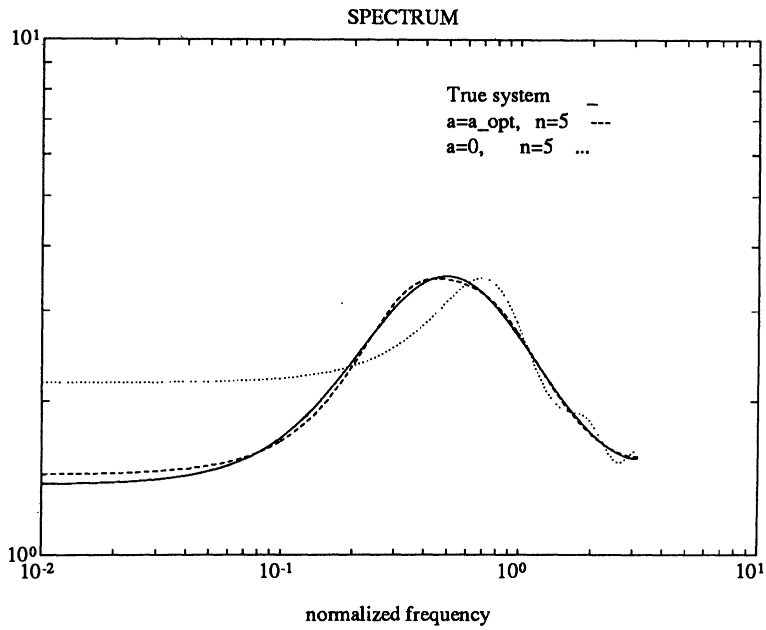

FIG. 5. *Example* 5.2: *Fifth order spectral density approximations.*

More practical experience with Laguerre models is needed before one can fully evaluate the potential in practical applications. Several promising results in adaptive control have been reported by Dumont and co-workers; see, for example, Zervos and Dumont (1988).

APPENDIX

A.1. Derivation of the Levinson type algorithm (49–53). Write

$$\sum_{j=0}^{n} \hat{\beta}_{n,j} u_{j+1}(t) = e(t), \qquad \hat{\beta}_{n,0} = 1, \tag{109}$$

where this is obtained by regressing $u_1(t)$ on the $-u_{j+1}(t)$, $j = 1, \ldots, n$. Then (50), (51) and (53) follow from the Levinson recursion via the Toeplitz nature of $\hat{\Gamma}_n$, and its symmetry. Next consider the regression

$$\begin{aligned} \sum_{j=0}^{n} \hat{\alpha}_{n,j} u_j(t) &= \hat{\varepsilon}(t), \qquad \hat{\alpha}_{n,0} = 1, \\ &= \sum_{j=0}^{n-1} \tilde{\alpha}_{n,j} u_j(t) \\ &\quad + \tilde{\alpha}_{n,n} \left(\sum_{k=0}^{n-1} \hat{\beta}_{n-1,k} u_{n-k}(t) \right), \qquad \tilde{\alpha}_{n,0} = 1. \end{aligned} \tag{110}$$

Here $u_j(t)$, $j = 1, \dots, n$ is replaced by $u_j(t)$, $j = 1, \dots, n-1$ and $\sum_{k=0}^{n-1} \hat{\beta}_{n-1,k} u_{n-k}(t)$, which is the residual from the regression of $u_n(t)$ on $-u_1(t), \dots, -u_{n-1}(t)$. This follows from standard arguments for the Levinson recursion; see Hannan and Deistler [(1988), page 212]. Since $\sum_{k=0}^{n-1} \hat{\beta}_{n-1,k} u_{n-k}(t)$ is Toeplitz orthogonal to $u_1(t), \dots, u_{n-1}(t)$, then $\tilde{\alpha}_{n,j} = \hat{\alpha}_{n-1,j}$ and since $\hat{\beta}_{n-1,0} = 1$, $\tilde{\alpha}_{n,n} = \hat{\alpha}_{n,n}$. Thus $\hat{\alpha}_{n,j} = \hat{\alpha}_{n-1,j} + \hat{\alpha}_{n,n} \hat{\beta}_{n-1,n-j}$, as required and $\hat{\alpha}_{n,n} = \sum_{j=0}^{n-1} \hat{\beta}_{n-1,j} \hat{d}_{n-j} / \hat{\sigma}_{n-1}^2$. This completes the proof. □

A.2. Proof of Theorem 5.1. All quoted results used here are in Hannan and Deistler (1988), especially Chapter 5. We shall carry out the proofs as if the calculations are done in Toeplitz form, that is, with $y(t) = 0$, $t \le 0$ or $t > N$. Let

$$\hat{r}(j) = \frac{1}{N} \sum_{t=1}^{N} y(t) y(t+j), \quad \text{with } \hat{r}(j) = 0, |j| \ge N, \tag{111}$$

$$r(j) = E\{y(t) y(t+j)\}. \tag{112}$$

Then the following hold:

$$\sup_{0 \le j < \infty} |\hat{r}(j) - r(j)| = \mathscr{O}([\log N/N]^{1/2}), \quad \text{a.s.}, \tag{113}$$

$$\sup_{0 \le j < v_N} |\hat{r}(j) - r(j)| = \mathscr{O}(Q_N), \quad \text{a.s.}; \ v_N = \mathscr{O}([\log N]^b), \ b < \infty. \tag{114}$$

We shall establish (114) for $\hat{c}_j(a)$, $\hat{d}_j(a)$, $c_j(a)$, $d_j(a)$, where $c_j(a) = E\{u_{j+1}(t,a) u_1(t,a)\}$, $d_j(a) = E\{u_j(t,a) y(t)\}$, but only for $j \le M < \infty$. However, results for M increasing with N that are weaker (and perhaps necessarily so) are presented in Section 4.1. For example,

$$d_j(a) = \sum_{i=1}^{\infty} l_j(i,a) r(i), \qquad \hat{d}_j(a) = \sum_{i=1}^{\infty} l_j(i,a) \hat{r}(i), \tag{115}$$

where the $\{l_j(i,a)\}$ are the Laguerre functions, that is, the coefficients in the expansion of $L_j(z,a)$ in powers of z^{-1}. For the analogue of (114) we may truncate the sum over i at $c \log N$ for arbitrary large c, so that the remainder is $\mathscr{O}(N^{-\alpha})$, a.s., for any $\alpha > 0$ (and suitable c), and the analogue of (114) holds. Now it follows that $\hat{\eta}_n(a)$ converges a.s. and uniformly in a to $\eta_n(a) = \gamma_n(a)' \Gamma_n(a)^{-1} \gamma_n(a)$ with $\Gamma_n(a) = E\{\varphi(t,a)\varphi(t,a)'\}$, $\gamma_n(a) = -E\{\varphi(t,a) y(t)\}$. For $n \ge n_0$, $\eta_n(a)$ has a unique maximum at $a = a_0$, as is easily checked. For $n < n_0$ then $\sup_a \eta_n(a) < \eta_0(a_0)$. Thus for N sufficiently large and $n < n_0$, $\hat{\sigma}_n^2$ is strictly larger than $\hat{\sigma}_{n_0}^2$ by a fixed amount and BIC(n) cannot be minimized for such N, at $n < n_0$. For $n \ge n_0$, $\|\hat{\tau}_n - \tau_0\|_1 = \mathscr{O}(Q_N)$. To show this we first show that $|\hat{a}_n - a_0| = \mathscr{O}(Q_N)$. The proof for the other component $\hat{\theta}_n$ of $\hat{\tau}_n$ is then relatively straight forward and is of the same nature as that for an autoregression. Now in much the same way as in Hannan and Deistler [(1988), Section 6.6], we may show that $d\hat{\eta}_n(a)/da = \eta_n'(a) + \mathscr{O}(Q_N)$, where we use $\eta_n'(a)$ to denote the derivative. Thus since $\eta_n'(a) = (a - a_0)\eta_n''(\bar{a})$, $|\bar{a} - a_0| \le |a - a_0|$, then

$$(116) \qquad \frac{d}{da}\hat{\eta}_n(a)\bigg|_{a=\hat{a}_n} = (\hat{a}_n - a_0)\eta_n''(\bar{a}) + \mathscr{O}(Q_N), \qquad |\bar{a} - a_0| \le |\hat{a}_n - a_0|.$$

Since the left-hand side of (116) equals 0, this establishes the last line of Theorem 5.1 for $\hat{a}_n$ and thus for $\hat{\theta}_n$.

To complete the proof that $\hat{n} \to n_0$ we need only to show that, for sufficiently large N, BIC(n) cannot be minimized at $n > n_0$. This is established by showing that

$$(117) \qquad \hat{\sigma}_n^2 = \frac{1}{N}\sum_{t=1}^{N} e(t)^2 + \mathscr{O}(Q_N^2).$$

Since $(n \log N)/N$ increases by $\log N/N$ as n increases and $(\log N/N)/Q_N^2 \to \infty$, (117) is sufficient to complete the proof of Theorem 5.1.

However, putting $e(t) = 0$ for $t > N$ we have

$$(118) \qquad \begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{N}\sum_{t=1}^{\infty}\left(y(t) - \varphi(t, \hat{a}_n)'\hat{\theta}_n\right)^2 \\ &= \frac{1}{N}\sum_{t=1}^{\infty}\left(e(t) + \varphi(t, a_0)'\left(\theta^0 - \hat{\theta}_n\right) + \left(\varphi(t, a_0) - \varphi(t, \hat{a}_n)\right)'\hat{\theta}_n\right)^2, \end{aligned}$$

and the last line of Theorem 5.1 may be used to establish (117). This completes the proof of Theorem 5.1. □

A.3. Proof of Theorem 5.3. The proof essentially follows Jennrich (1969). Indeed

$$(119) \qquad \left(\hat{a}_0^{(k+1)} - \hat{a}_0\right) = \left(\hat{a}_0^{(k)} - \hat{a}_0\right)\left[1 - \frac{\hat{\eta}_0''(\bar{a})}{\hat{\eta}_0''\left(\hat{a}_0^{(k)}\right)}\right],$$

where $|\bar{a} - \hat{a}_0| \le |\hat{a}_0^{(k)} - \hat{a}_0|$. We can choose $\mathbf{I}$ and N_0 so that the last factor in (119) is less than unity so long as $\hat{a}_0^{(k)}$ is sufficiently near $\hat{a}_0$. For suitable $\mathbf{I}$ and N_0 this will be true for $\hat{a}_0^{(1)}$ since $\hat{a}_0$ will be sufficiently near to a_0. Thus $|\hat{a}_0^{(2)} - \hat{a}_0| \le |\hat{a}_0^{(1)} - \hat{a}_0|$, and so on, so that $\hat{a}_0^{(k)} \to \hat{a}_0$, a.s.

A.4. Proof of the bound (65). Let $\mathscr{H}$ be the Hankel matrix with (n, m)th element $l_k(n + m - 1, a)$, and the singular value decomposition (SVD)

$$(120) \qquad \mathscr{H} = \sum_{i=1}^{k} \sigma_i u_i v_i', \qquad u_i = (u_i(1)u_i(2)\cdots)', \qquad v_i = (v_i(1)v_i(2)\cdots)'.$$

Then

$$(121) \qquad l_k(2j, a) = \sum_{i=1}^{k} \sigma_i u_i(j) v_i(j+1)$$

and consequently

$$\sum_{j=1}^{\infty}|l_k(2j,a)| = \sum_{j=1}^{\infty}\left|\sum_{i=1}^{k}\sigma_i u_i(j)v_i(j+1)\right|$$
$$\leq \sum_{i=1}^{k}\sigma_i\|u_i\|_2\|v_i^{\uparrow}\|_2 = \sum_{i=1}^{k}\sigma_i\sqrt{1-v_i(1)^2} \leq \sum_{i=1}^{k}\sigma_i, \tag{122}$$

where the first inequality follows from the Schwarz inequality, and the second one from $\|u_i\|_2 = \|v_i\|_2 = 1$. $x^{\uparrow}$ means an upshift in the elements of x. Similarly,

$$\sum_{j=1}^{\infty}|l_k(2j-1,a)| = \sum_{j=1}^{\infty}\left|\sum_{i=1}^{k}\sigma_i u_i(j)v_i(j)\right|$$
$$\leq \sum_{i=1}^{k}\sigma_i\|u_i\|_2\|v_i\|_2 = \sum_{i=1}^{k}\sigma_i. \tag{123}$$

Hence,

$$\sum_{j=1}^{\infty}|l_k(j,a)| \leq 2\sum_{i=1}^{k}\sigma_i = \text{two times the nuclear norm.} \tag{124}$$

Also $\mathscr{H}\mathscr{H}' \leq \mathscr{L}$, where $\mathscr{L}$ is the symmetric Toeplitz matrix with (n,m)th element $\sum_{i=1}^{\infty} l_k(i,a)l_k(i+|n-m|,a)$, so that

$$\sigma_1^2 \leq \sup_{\omega}|L_k(e^{i\omega},a)|^2. \tag{125}$$

Thus

$$\sum_{j=1}^{\infty}\left|l_k(j,a)\right| \leq 2k\sigma_1 \leq 2k\sup_{\omega}\left|L_k(e^{i\omega},a)\right|, \tag{126}$$

which proves (65). □

A.5. Proof of Lemma 6.1. Using

$$c_j(a) = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} l_{j+1}(i,a)l_1(k,a)r(i-k),$$
$$\hat{c}_j(a) = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} l_{j+1}(i,a)l_1(k,a)\hat{r}(i-k), \tag{127}$$

we obtain

$$|c_j(a) - \hat{c}_j(a)| \leq \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}|l_{j+1}(i,a)l_1(k,a)|\,|\hat{r}(i-k) - r(i-k)|$$
$$\leq \sup_i|\hat{r}(i) - r(i)|\sum_{i=1}^{\infty}|l_{j+1}(i,a)|\sum_{k=1}^{\infty}|l_1(k,a)|$$
$$\leq 4(j+1)\frac{1+|a|}{1-|a|}\sup_i|\hat{r}(i) - r(i)|, \tag{128}$$

where the last inequality follows from (65) and

$$(129) \qquad |L_j(e^{i\omega}, a)| \leq \frac{\sqrt{1-a^2}}{1-|a|}.$$

The result (66) now follows from the result (113) in the Appendix. The proof of (67) follows the same lines, and is therefore omitted. □

A.6. Proof of Theorem 6.1. Let $d = (d_1 \cdots d_n)'$, $\|d\|_2 = 1$. Then using Lemma 4.1 we have

$$(130) \qquad d'\Gamma_n(a)d \geq \frac{1}{2\pi}\int_{-(\pi-\varepsilon)}^{\pi-\varepsilon}\left|\sum_{j=1}^{n} d_j e^{i\varpi(j-1)}\right|^2 \Phi_y\left(\frac{1+ae^{i\varpi}}{e^{i\varpi}+a}\right)d\varpi$$

$$(131) \qquad \geq \min_{-(\pi-\varepsilon)\leq\varpi\leq(\pi-\varepsilon)} \Phi_y\left(\frac{1+ae^{i\varpi}}{e^{i\varpi}+a}\right)\left(\|d\|_2^2 - \frac{\varepsilon}{\pi}\left(\sum_{j=1}^{n}|d_j|\right)^2\right).$$

But

$$(132) \qquad \left(\sum_{j=1}^{n}|d_j|\right)^2 \leq n\|d\|_2^2,$$

which proves Theorem 6.1. □

A.7. Proof of Lemma 6.2. Since Γ_n is the Toeplitz matrix associated with the spectral density

$$(133) \qquad \Phi_y\left(\frac{e^{i\varpi}+a}{1+ae^{i\varpi}}\right) = \sigma_0^2\left|H^0\left(\frac{e^{i\varpi}+a}{1+ae^{i\varpi}}\right)\right|^2$$

and

$$(134) \qquad \Gamma_n(\theta_n^0 - \theta_n^*) = E\{\varphi(t,a)\Delta(q)y(t)\},$$

where $\Delta(q) = \sum_{k=n+1}^{\infty}\alpha_k^0 L_k(q,a)$. Theorem 1.1 in Baxter (1963) implies that

$$(135) \qquad \|\theta_n^0 - \theta_n^*\|_1 \leq \overline{M}\|E\{\varphi(t)\Delta(q)y(t)\}\|_1, \quad \text{for } n > \tilde{n},$$

where $\overline{M} = M/(\|\Phi_y((1+aw)/(w+a))\|_1)$, with M and $\tilde{n}$ as above. Observe that condition (72) guarantees that Theorem 1.1 of Baxter is applicable. The right-hand side of (135) can be bounded as

$$(136) \qquad \begin{aligned} &\|E\{\varphi(t,a)\Delta(q)y(t)\}\|_1 \\ &\leq \sum_{j=1}^{n}\sum_{k=n+1}^{\infty}|\alpha_k^0|\left|E\{L_k(q,a)y(t)L_j(q,a)y(t)\}\right| \\ &\leq \sum_{k=n+1}^{\infty}|\alpha_k^0|\left\|\Phi_y\left(\frac{w+a}{1+aw}\right)\right\|_1, \end{aligned}$$

where the last equality follows from

$$(137)\quad \Phi_y\left(\frac{w+a}{1+aw}\right) = \sum_{\tau=-\infty}^{\infty} c_\tau w^{-\tau}, \qquad c_{k-j} = E\{L_k(q,a)y(t)L_j(q,a)y(t)\}.$$

This proves (73). The statement (74) follows directly from (73), using $|L_j(e^{i\omega}, a)| \le [(1+|a|)/(1-|a|)]^{1/2}$. □

A.8. Proof of Theorem 6.2. By definition,

$$(138)\qquad y(t) = \varphi(t,a)'\theta_n^* + [H_n^*(q,a)]^{-1}y(t),$$

which implies

$$(139)\qquad \hat{\Gamma}_n(\hat{\theta}_n - \theta_n^*) = \frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a)[H_n^*(q,a)]^{-1}y(t).$$

Row j of the right-hand side of (139) equals

$$(140)\qquad \hat{d}_j + \sum_{k=1}^{n} \alpha_k^* \hat{c}_{k-j}, \qquad j = 1,\dots,n.$$

Also, from the optimality of θ_n^*,

$$(141)\qquad d_j + \sum_{k=1}^{n} \alpha_k^* c_{k-j} = 0, \qquad j = 1,\dots,n.$$

Hence, row j of the right-hand side of (139) can be bounded by

$$(142)\qquad \begin{aligned} &|\hat{d}_j - d_j| + \sum_{k=1}^{n} |\alpha_k^*|\,|\hat{c}_{k-j} - c_{k-j}| \\ &\le |\hat{d}_j - d_j| + \max_{0\le k\le(n-1)} |\hat{c}_k - c_k| \sum_{k=1}^{n} |\alpha_k^*| \\ &= \left(n(\log N/N)^{1/2}\right), \quad \text{a.s.} \end{aligned}$$

The last equality follows from Lemma 6.1 and

$$(143)\qquad \sum_{k=1}^{n} |\alpha_k^*| \le \|\theta_n^* - \theta_n^0\|_1 + \|\theta_n^0\|_1 \le C\sum_{k=1}^{\infty} |\alpha_k^0| < \infty,$$

using Lemma 6.2. From Lemma 6.1,

$$(144)\qquad \|\Gamma_n - \hat{\Gamma}_n\|_2 \le \sup_{1\le i\le n} \sum_{j=1}^{n} |c_{j-i} - \hat{c}_{j-i}| = \mathscr{O}\left(n^2(\log N/N)^{1/2}\right).$$

Consequently, $\|\hat{\Gamma}^{-1}\|_2$ is bounded a.s. for large enough N, since (69) implies that $\|\Gamma^{-1}\|_2$ is bounded from above uniformly in n. To prove (79), we now use

$$(145)\qquad \begin{aligned} \|\hat{\theta}_n - \theta_n^*\|_2 &\le \|\hat{\Gamma}^{-1}\|_2 \left\| \frac{1}{N}\sum_{t=1}^{\infty} \varphi(t,a)[H_n^*(q,a)]^{-1}y(t)\right\|_2 \\ &= \mathscr{O}\left(n^{3/2}\left(\frac{\log N}{N}\right)^{1/2}\right), \end{aligned}$$

since $\|x\|_2 \le n^{1/2} \max_i |x_i|$ for vectors of dimension n. The second statement in Theorem 6.2 follows from $\|x\|_1 \le n^{1/2} \|x\|_2$ for any $n \times 1$ vector x. Combining (80), Lemma 6.2 and

$$\|\hat{\theta}_n - \theta_n^0\|_1 \le \|\hat{\theta}_n - \theta_n^*\|_1 + \|\theta_n^0 - \theta_n^*\|_1 \tag{146}$$

yields (81). □

A.9. Proof of Theorem 6.3. Introduce $e^{i\bar{\omega}} = (1 - ae^{i\omega})/(e^{i\omega} - a)$; then

$$\begin{aligned}\frac{\sigma_0^2}{n} & W(\omega_1)\Gamma_n^{-1} W(-\omega_2)' \\ &= \frac{(1 - a^2)}{(e^{i\omega_1} - a)(e^{-i\omega_2} - a)} \\ &\quad \times \frac{\sigma_0^2}{n} (1\ e^{i\bar{\omega}_1} \cdots e^{i\bar{\omega}_1(n-1)}) \Gamma_n^{-1} (1\ e^{-i\bar{\omega}_2} \cdots e^{-i\bar{\omega}_2(n-1)}).\end{aligned} \tag{147}$$

From Hannan and Wahlberg (1989),

$$\begin{aligned}\frac{1}{n} & (1\ e^{i\bar{\omega}_1} \cdots e^{i\bar{\omega}_1(n-1)}) \Gamma_n^{-1} (1\ e^{-i\bar{\omega}_2} \cdots e^{-i\bar{\omega}_2(n-1)}) \\ & \to \begin{cases} 0, & \text{if } \bar{\omega}_1 \ne \bar{\omega}_2, \\ \left[\Phi_y\left(\dfrac{e^{i\bar{\omega}_1} + a}{1 + ae^{i\bar{\omega}_1}}\right)\right]^{-1}, & \text{if } \bar{\omega}_1 = \bar{\omega}_2, \end{cases}\end{aligned} \tag{148}$$

as $n \to \infty$. Substituting (148) into (147) now proves the theorem. □

REFERENCES

BAXTER, G. (1963). A norm inequality for a "finite section" Wiener–Hopf equation. *Illinois J. Math.* **7** 97–103.

BERK, K. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502.

CLEMENT, P. R. (1982). Laguerre functions in signal analysis and parameter identification. *J. Franklin Inst.* **313** 85–95.

DEWILDE, P. (1982). Stochastic modelling with orthogonal filters. In *l'Analyse de Systemes et le Traitement du Signal* **2**. Editions du CNRS, Paris.

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

EYKHOFF, P. (1974). *System Identification*. Wiley, New York.

GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications*. Univ. California Press, Berkeley.

GLOVER, K., CURTAIN, R. F. and PARTINGTON, J. R. (1988). Realization and approximation of linear infinite-dimensional systems with error bounds. *SIAM J. Control Optim.* **26** 863–898.

GOLUB, G. H. and VAN LOAN, C. F. (1989). *Matrix Computations*, 2nd ed. Johns Hopkins Univ. Press.

GOTTLIEB, M. J. (1938). Polynomials orthogonal on finite or enumerable set of points. *Amer. J. Math.* **60** 453–458.

HANNAN, E. J. and DEISTLER, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York.

HANNAN, E. J. and WAHLBERG, B. (1989). Convergence rates for inverse Toeplitz matrix forms. *J. Multivariate Anal.* **31** 127–135.

HWANG, C. and SHIH, Y.-P. (1982). Parameter identification via Laguerre polynomials. *Internat. J. Systems Sci.* **13** 209–217.

JENNRICH, R. I. (1969). Asymptotic properties of non linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

KAHANE, J. P. (1956). Sur certaines classes de sèries de Fourier absolument convergentes. *J. Math. Pures Appl.* **35** 249–259.

KING, R. E. and PARASKEVOPOULOS, P. N. (1977). Digital Laguerre filters. *Circuit Theory Appl.* **5** 81–91.

KING, R. E. and PARASKEVOPOULOS, P. N. (1979). Parametric identification of discrete-time SISO systems. *Internat. J. Control* **30** 1023–1029.

LIM, J. S. and OPPENHEIM, A. V. (1988). *Advanced Topics in Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

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

MÄKILÄ, P. (1990a). Approximation of stable systems by Laguerre filters. *Automatica* **26** 333–345.

MÄKILÄ, P. (1990b). Laguerre series approximation of infinite dimensional systems. *Automatica* **26** 985–996.

MARPLE, S. L. (1987). *Digital spectral analysis: with applications*. Prentice-Hall, Englewood Cliffs, NJ.

MIDDLETON, R. H. and GOODWIN, G. C. (1990). *Digital Estimation and Control: A Unified Approach*. Prentice-Hall, Englewood Cliffs, NJ.

SÖDERSTRÖM, T. and STOICA, P. (1989). *System Identification*. Prentice-Hall, Englewood Cliffs, NJ.

SZEGÖ, G. (1939). *Orthogonal Polynomials* **23**. *Amer. Math. Soc. Coll. Pub.*

WAHLBERG, B. (1988). Limit results for sampled systems. *Internat. J. Control* **48** 1267–1284.

WAHLBERG, B. (1990). ARMA spectral estimation of narrowband processes via model reduction. *IEEE Trans. Acoust. Speech Signal Process.* **38** 1144–1154.

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

WAHLBERG, B. and HANNAN, E. J. (1991). Estimation of noise models by means of discrete Laguerre/Kautz filters. In *Preprints 9th IFAC/IFORS Symposium on System Identification and System Parameter Estimation* 1236–1241. Budapest, Hungary.

WIENER, N. (1956). The theory of prediction. In *Modern Mathematics for Engnr. Bechenbach* 183–184. McGraw-Hill, New York.

ZERVOS, C. C. and DUMONT, G. A. (1988). Deterministic adaptive control based on Laguerre series representations. *Internat. J. Control* **48** 2333–2359.

ZYGMUND, A. (1968). *Trigonometric Series*. Cambridge Univ. Press.

DEPARTMENT OF AUTOMATIC CONTROL
ROYAL INSTITUTE OF TECHNOLOGY
S-100 44 STOCKHOLM
SWEDEN

DEPARTMENT OF STATISTICS
FACULTY OF ECONOMICS AND COMMERCE
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4, ACT 2601
AUSTRALIA