STOCHASTIC PROCESSES, DETECTION AND ESTIMATION
6.432 Course Notes

Alan S. Willsky, Gregory W. Wornell, and Jeffrey H. Shapiro

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

Fall 2003

6

Detection and Estimation from Waveform Observations: Addendum

6.1 NONRANDOM PARAMETER ESTIMATION FOR GAUSSIAN PROCESSES

In this section, we develop some very useful results for parameter estimation involving stationary Gaussian processes observed over long time intervals, corresponding to the SPLOT scenario of Chapter 5. We focus on the discrete-time results, but comment in advance that analogous results can be developed for the continuous-time case.

To begin, suppose we have observations of the form $y[0], y[1], \ldots, y[N-1]$ where $y[n]$ is a zero-mean stationary Gaussian random process with power spectrum $S_{yy}(e^{j\omega}; \mathbf{x})$, where $\mathbf{x}$ is a vector of unknown parameters. Then provided N is sufficiently large (so that variations in the spectrum are on scales significantly larger than $2\pi/N$), we can exploit our result from the last chapter that the (normalized) discrete Fourier transform (DFT) coefficients

$$Y[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y[n]\, e^{-j2\pi kn/N}, \qquad k = 0, 1, \ldots, N-1 \tag{6.1}$$

are effectively independent Gaussian random variables[1] with variance

$$\text{var}\, Y[k] = E\left[|Y[k]|^2\right] \triangleq \lambda_k(\mathbf{x}) \approx S_{yy}(e^{j2\pi k/N}; \mathbf{x}). \tag{6.2}$$

[1]In fact, for any N, the $Y[k]$ are *circular Gaussians*: their real and imaginary parts are independent and identically-distributed.

Moreover, we recall from Chapter 3 that these normalized DFT coefficients (6.1) can be rewritten as

$$Y[k] = Y_N(e^{j\omega})\Big|_{\omega=2\pi k/N} \tag{6.3}$$

where $Y_N(e^{j\omega})$ is the Fourier transform of a windowed version of $y[n]$, i.e.,

$$Y_N(e^{j\omega}) = \mathcal{F}\left\{w_N[n]\, y[n]\right\} \tag{6.4}$$

with $w_N[n]$ denoting the unit-energy window

$$w_N[n] = \begin{cases} 1/\sqrt{N} & n = 0, 1, \dots, N-1 \\ 0 & \text{otherwise} \end{cases}. \tag{6.5}$$

In terms of parameter estimation, the observations

$$\mathbf{y} = \begin{bmatrix} y[0] & y[1] & \cdots & y[N-1] \end{bmatrix}^{\mathrm{T}} \tag{6.6}$$

and

$$\mathbf{Y} = \begin{bmatrix} Y[0] & Y[1] & \cdots & Y[N-1] \end{bmatrix}^{\mathrm{T}} \tag{6.7}$$

are equivalent. Accordingly, we can write the likelihood function for the observations in the form

$$p_{\mathbf{Y}}(\mathbf{Y};\mathbf{x}) \approx \prod_{k=0}^{N-1} \frac{1}{\sqrt{2\pi\lambda_k(\mathbf{x})}} \exp\left[-\frac{|Y[k]|^2}{2\lambda_k(\mathbf{x})}\right]. \tag{6.8}$$

From (6.8) we obtain, in turn,

$$\begin{aligned} \ell(\mathbf{y};\mathbf{x}) &= \ln p_{\mathbf{Y}}(\mathbf{Y};\mathbf{x}) \\ &\approx -\frac{N}{2}\ln 2\pi - \frac{1}{2}\frac{N}{2\pi}\sum_{k=0}^{N-1}\left[\ln \lambda_k(\mathbf{x}) + \frac{|Y[k]|^2}{\lambda_k(\mathbf{x})}\right]\frac{2\pi}{N} \\ &\approx -\frac{N}{2}\ln 2\pi - \frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\ln S_{yy}(e^{j\omega}) + \frac{|Y_N(e^{j\omega})|^2}{S_{yy}(e^{j\omega})}\right] d\omega \end{aligned} \tag{6.9}$$

where to obtain the last expression in (6.9) we have used the integral approximation

$$\sum_{k=0}^{N-1} f(\lambda_k) \approx \frac{N}{2\pi}\int_{-\pi}^{\pi} f(S_{yy}(e^{j\omega}))\, d\omega \tag{6.10}$$

valid for sufficiently large N.

Note that (6.9) implies that for the Gaussian SPLOT scenario the *periodogram* $|Y_N(e^{j\omega})|^2$ is effectively a sufficient statistic: it contains all the features of the data necessary for detection and estimation problems involving $y[n]$. We'll focus on estimation problems in the sequel; detection problems are treated in a similar manner and lead to equally useful algorithms in practice.

6.1.1 Cramér-Rao Bounds

Using (6.9), we can obtain useful asymptotic approximations to the associated Cramér-Rao bounds. In particular, we first obtain

$$\begin{aligned}\frac{\partial}{\partial x_k}\ell(\mathbf{y};\mathbf{x}) &\approx -\frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\partial}{\partial x_k}\ln S_{yy}(e^{j\omega};\mathbf{x}) - \frac{|Y_N(e^{j\omega})|^2}{S_{yy}^2(e^{j\omega};\mathbf{x})}\frac{\partial}{\partial x_k}S_{yy}(e^{j\omega};\mathbf{x})\right]d\omega \\ &\approx -\frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(1-\frac{|Y_N(e^{j\omega})|^2}{S_{yy}(e^{j\omega};\mathbf{x})}\right)\left(\frac{\partial}{\partial x_k}\ln S_{yy}(e^{j\omega};\mathbf{x})\right)d\omega.\end{aligned} \tag{6.11}$$

Differentiating (6.11) with respect to x_l we obtain

$$\begin{aligned}\frac{\partial^2}{\partial x_k \partial x_l}\ell(\mathbf{y};\mathbf{x}) = -\frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigg[&\left(\frac{|Y_N(e^{j\omega})|^2}{S_{yy}^2(e^{j\omega};\mathbf{x})}\frac{\partial}{\partial x_l}S_{yy}(e^{j\omega};\mathbf{x})\right)\left(\frac{\partial}{\partial x_k}\ln S_{yy}(e^{j\omega};\mathbf{x})\right) \\ &+\left(1-\frac{|Y_N(e^{j\omega})|^2}{S_{yy}(e^{j\omega};\mathbf{x})}\right)\left(\frac{\partial^2}{\partial x_k \partial x_l}\ln S_{yy}(e^{j\omega};\mathbf{x})\right)\Bigg],\end{aligned} \tag{6.12}$$

which using [cf. (6.2)]

$$E\left[\left|Y_N(e^{j\omega})\right|^2\right] \approx S_{yy}(e^{j\omega};\mathbf{x}) \tag{6.13}$$

yields

$$\begin{aligned}[\mathbf{I}_{\mathbf{y}}(\mathbf{x})]_{kl} &= -E\left[\frac{\partial^2}{\partial x_k \partial x_l}\ell(\mathbf{y};\mathbf{x})\right] \\ &\approx \frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial}{\partial x_k}\ln S_{yy}(e^{j\omega};\mathbf{x})\right)\left(\frac{\partial}{\partial x_l}\ln S_{yy}(e^{j\omega};\mathbf{x})\right)d\omega.\end{aligned} \tag{6.14}$$

6.1.2 Maximum Likelihood Parameter Estimates

Similar approximations allow us to determine the corresponding maximum likelihood (ML) estimates for this Gaussian SPLOT scenario. To see this, first note that via (6.9) we obtain immediately that these estimates are given by

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{ML}}(\mathbf{y}) &\approx \arg\max_{\mathbf{x}} \ell(\mathbf{y};\mathbf{x}) \\ &= \arg\min_{\mathbf{x}} J(\mathbf{x}),\end{aligned} \tag{6.15a}$$

where

$$J(\mathbf{x}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\ln S_{yy}(e^{j\omega};\mathbf{x}) + \frac{|Y_N(e^{j\omega})|^2}{S_{yy}(e^{j\omega};\mathbf{x})}\right]d\omega. \tag{6.15b}$$

From (6.15b) we obtain the following necessary condition satisfied by the ML parameter estimates:

$$\left.\frac{\partial J(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{\mathrm{ML}}} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[1-\frac{|Y_N(e^{j\omega})|^2}{S_{yy}(e^{j\omega};\mathbf{x})}\right]\left[\frac{\partial}{\partial x_i}\ln S_{yy}(e^{j\omega};\mathbf{x})\right]d\omega = 0. \tag{6.16}$$

6.2 AUTOREGRESSIVE SIGNAL MODELING

The results just developed have been used to develop and evaluate efficient estimation algorithms for a host of applications. As an important and illustrative example, in this section we undertake a case study involving autoregressive signal modeling.

The scenario is as follows. Suppose we observe N samples of a discrete-time random signal $y[n]$ that can be modeled as output of a stable LTI filter with the all-pole system function

$$H(z) = \frac{1}{A(z)}, \qquad A(z) = 1 - \sum_{n=1}^{M} a[n]z^{-n} \tag{6.17}$$

driven by zero-mean, wide-sense stationary, white noise $w[n]$ with variance σ^2.

Such a process is referred to as an Mth order *autoregressive* process [AR(M)], since it can be equivalently obtained from the forward recursion defined by the stochastic difference equation

$$y[n] = a[1]\,y[n-1] + a[2]\,y[n-2] + \cdots + a[M]\,y[n-M] + w[n]. \tag{6.18}$$

When, in addition, $w[n]$ is a Gaussian process, then (6.18) implies that $y[n]$ is both Gaussian and a Markov process of Mth order, and in this case we say $y[n]$ is an Mth-order *Gauss-Markov* process [GM(M)]. We'll focus primarily on this case, but will comment on the nonGaussian scenario later in the section.

Autoregressive models have, in general, proven extraordinarily useful in a wide range of applications—from speech modeling for voice compression systems to seismic data modeling for oil exploration systems.

The central estimation problem is as follows: given Gaussian observations

$$\mathbf{y} = \begin{bmatrix} y[0] & y[1] & \cdots & y[N-1] \end{bmatrix}^{\mathrm{T}} \tag{6.19}$$

we seek useful estimates of the unknown, nonrandom parameters

$$\mathbf{x} = \begin{bmatrix} a[1] & a[2] & \cdots & a[M] & \sigma^2 \end{bmatrix}. \tag{6.20}$$

6.2.1 Cramér-Rao Bounds

Let us begin by determining the associated Cramér-Rao bounds for the problem.

First, since

$$\ln S_{yy}(e^{j\omega}) = \ln \sigma^2 - \ln \left|A(e^{j\omega})\right|^2,$$

we obtain

$$\frac{\partial}{\partial \sigma^2} \ln S_{yy}(e^{j\omega}) = \frac{1}{\sigma^2} \tag{6.21a}$$

$$\frac{\partial}{\partial a[k]} \ln S_{yy}(e^{j\omega}) = \frac{A(e^{j\omega})e^{j\omega k} + A^*(e^{j\omega})e^{-j\omega k}}{|A(e^{j\omega})|^2} \tag{6.21b}$$

where

$$A(e^{j\omega}) = \mathcal{F}\{a[n]\} = 1 - \sum_{n=1}^{M} a[n]\, e^{-j\omega n}. \tag{6.22}$$

Using these expressions the Fisher information matrix entries $[\mathbf{I_y}(\mathbf{x})]_{kl}$ are obtained as follows.

Case: $1 \leq k, l \leq M$

From (6.14) we get,

$$\begin{aligned}
[\mathbf{I_y}(\mathbf{x})]_{kl} &\approx \frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\partial}{\partial a[k]} \ln S_{yy}(e^{j\omega};\mathbf{x})\right)\left(\frac{\partial}{\partial a[l]} \ln S_{yy}(e^{j\omega};\mathbf{x})\right) d\omega \\
&= \frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|A(e^{j\omega})|^4}\left[A(e^{j\omega})e^{j\omega k} + A^*(e^{j\omega})e^{-j\omega k}\right]\left[A(e^{j\omega})e^{j\omega l} + A^*(e^{j\omega})e^{-j\omega l}\right] d\omega \\
&= \frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left[\frac{1}{A(e^{j\omega})^2}e^{-j\omega(k+l)} + \frac{1}{A^*(e^{j\omega})^2}e^{j\omega(k+l)}\right. \\
&\qquad\qquad \left. + \frac{1}{|A(e^{j\omega})|^2}e^{j\omega(k-l)} + \frac{1}{|A(e^{j\omega})|^2}e^{-j\omega(k-l)}\right] d\omega \\
&= N\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{A(e^{j\omega})^2}e^{-j\omega(k+l)}\, d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|A(e^{j\omega})|^2}e^{j\omega(k-l)}\, d\omega\right\},
\end{aligned} \tag{6.23}$$

where to obtain the last equality in (6.23) we have used that $A^*(e^{j\omega}) = A(e^{-j\omega})$ since the sequence $a[n]$ is real-valued.

Next we let $b[n]$ denote the sequence whose Fourier transform is

$$B(e^{j\omega}) = \frac{1}{A^2(e^{j\omega})} \tag{6.24}$$

and $c[n]$ the sequence whose Fourier transform is

$$C(e^{j\omega}) = \frac{1}{|A(e^{j\omega})|^2}. \tag{6.25}$$

Then with this notation (6.23) becomes

$$[\mathbf{I}_{\mathsf{y}}(\mathbf{x})]_{kl} \approx Nb[-k-l] + Nc[k-l], \qquad 1 \le k, l \le M \tag{6.26}$$

However, since we can write

$$b[n] = h[n] * h[n] \tag{6.27}$$

where $h[n]$ is the causal sequence with Fourier transform

$$H(e^{j\omega}) = \frac{1}{A(e^{j\omega})}, \tag{6.28}$$

we see that $b[n]$ is also causal, so

$$b[-k-l] = 0, \qquad 1 \le k, l \le M. \tag{6.29}$$

Hence, using (6.29) and the fact that

$$c[n] = \frac{1}{\sigma^2} K_{yy}[n] \tag{6.30}$$

since

$$S_{yy}(e^{j\omega}) = \frac{\sigma^2}{|A(e^{j\omega})|^2},$$

we obtain, finally,

$$[\mathbf{I}_{\mathsf{y}}(\mathbf{x})]_{kl} \approx \frac{N}{\sigma^2} K_{yy}[k-l], \qquad 1 \le k, l \le M. \tag{6.31}$$

Case: $1 \le k \le M$, $l = M+1$

For this case, we have

$$\begin{aligned}
[\mathbf{I}_{\mathsf{y}}(\mathbf{x})]_{kl} &\approx \frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\partial}{\partial a[k]} \ln S_{yy}(e^{j\omega};\mathbf{x})\right)\left(\frac{\partial}{\partial \sigma^2} \ln S_{yy}(e^{j\omega};\mathbf{x})\right) d\omega \\
&= \frac{N}{2\sigma^2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|A(e^{j\omega})|^2}\left[A(e^{j\omega})e^{j\omega k} + A^*(e^{j\omega})e^{-j\omega k}\right] d\omega \\
&= \frac{N}{\sigma^2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{A(e^{j\omega})} e^{-j\omega k}\, d\omega \\
&= h[-k] = 0
\end{aligned} \tag{6.32}$$

where we have again used that $h[n]$ is causal.

Case: $k = l = M + 1$

Finally, for this case we have

$$[\mathbf{I}_{\mathsf{y}}(\mathbf{x})]_{kl} \approx \frac{N}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial}{\partial\sigma^2}\ln S_{yy}(e^{j\omega};\mathbf{x})\right)^2 d\omega = \frac{N}{2\sigma^4}. \tag{6.33}$$

Combining (6.31), (6.32), and (6.33), we obtain the Fisher information matrix

$$\mathbf{I}_{\mathsf{y}}(\mathbf{x}) = \begin{bmatrix} (N/\sigma^2)\mathbf{\Lambda}_{\mathsf{y}}(\mathbf{x}) & \mathbf{0} \\ \mathbf{0} & N/(2\sigma^4) \end{bmatrix} \tag{6.34a}$$

where

$$[\mathbf{\Lambda}_{\mathsf{y}}(\mathbf{x})]_{ij} = K_{yy}[i-j], \qquad i,j = 1,2,\ldots,M. \tag{6.34b}$$

From (6.34) we obtain the following Cramér-Rao bounds on unbiased estimates of the parameters:

$$\operatorname{var}\hat{a}[k] \geq \frac{\sigma^2}{N}\left[\mathbf{\Lambda}_{\mathsf{y}}^{-1}(\mathbf{x})\right]_{kk}, \qquad k = 1,2,\ldots,M \tag{6.35}$$

$$\operatorname{var}\frac{\hat{\sigma}^2}{\sigma^2} \geq \frac{2}{N} \tag{6.36}$$

Useful insight can be obtained from examining the special case corresponding to $M = 1$. In this instance, recall from Example 4.15 that the associated first-order spectrum

$$S_{yy}(z) = \frac{\sigma^2}{(1-a[1]z^{-1})(1-a[1]z)} \tag{6.37}$$

corresponds to the autocovariance

$$K_{yy}[n] = \frac{\sigma^2}{1-a^2[1]}\,(a[1])^{|n|} \tag{6.38}$$

so that (6.35) specializes to the condition

$$\operatorname{var}\hat{a}[1] \geq \frac{\sigma^2}{NK_{yy}[0]} = \frac{1}{N}\left(1-a^2[1]\right). \tag{6.39}$$

Hence, the closer the pole of the shaping filter $H(z)$ is to the unit-circle, the better we might expect to be able to estimate the pole location. This behavior holds more generally for other values of M—the closer the poles are to the unit-circle, the more peaky the spectrum, and the more effective we can expect an estimation algorithm to be.

6.2.2 Maximum Likelihood Parameter Estimates

Let's determine the form of the ML estimates for the Gaussian autoregressive modeling problem. First, since

$$S_{yy}(e^{j\omega}) = \frac{\sigma^2}{|A(e^{j\omega})|^2} \tag{6.40}$$

the function $J(\mathbf{x})$ in (6.15b) specializes to

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\ln\sigma^2 - \ln\left|A(e^{j\omega})\right|^2 + \frac{1}{\sigma^2}\left|Y_N(e^{j\omega})\right|^2\left|A(e^{j\omega})\right|^2\right]\,d\omega \\ &= \ln\sigma^2 - \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left|A(e^{j\omega})\right|^2\,d\omega + \frac{1}{\sigma^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|Y_N(e^{j\omega})\right|^2\,\left|A(e^{j\omega})\right|^2\,d\omega. \end{aligned} \tag{6.41}$$

The second term in (6.41) is zero. To see this, first note that $A(e^{j\omega})$ is the frequency response of a causal, BIBO-stable system, so $A(z)$ converges in some region of the z-plane including the unit-circle, and, hence so does $A(1/z)$ and

$$B(z) = \tilde{B}(z) + \tilde{B}(1/z) \tag{6.42}$$

where

$$\tilde{B}(z) = \ln A(z) = \ln\left[1 - \sum_{n=1}^{M} a[n]z^{-n}\right]. \tag{6.43}$$

Since the second term in (6.41) is $b[0]$, the sequence whose z-transform is (6.42) evaluated at $n = 0$, it suffices to show that $b[0] = 0$.

Since $a[n]$ is causal, the region of convergence of $A(z)$ extends to $|z| \to \infty$, so we can exploit that for suitably large z, we have[2]

$$\tilde{B}(z) = -\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\left[\sum_{n=1}^{M} a[n]z^{-n}\right]^k \tag{6.44}$$

which we see consists of strictly negative powers of z. Hence,

$$\tilde{b}[n] = 0, \qquad \text{for } n \leq 0. \tag{6.45}$$

Finally, from (6.42) we see

$$b[n] = \tilde{b}[n] + \tilde{b}[-n] \tag{6.46}$$

so combining (6.45) with (6.46) we obtain $b[0] = 0$ as claimed.

---

[2]Here we have used the familiar power series

$$\ln(1 - x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots$$

valid for $|x| < 1$.

Thus, (6.41) simplifies to

$$J(\mathbf{x}) = \ln \sigma^2 + \frac{1}{\sigma^2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|Y_N(e^{j\omega})\right|^2 \left|A(e^{j\omega})\right|^2 \, d\omega. \tag{6.47}$$

The objective function (6.47) has a unique local minimum corresponding to the ML parameter estimates, which are the solutions to the stationary point equations

$$\frac{\partial J}{\partial \sigma^2} = \frac{1}{\sigma^2} - \frac{1}{\sigma^4}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|Y_N(e^{j\omega})\right|^2 \left|A(e^{j\omega})\right|^2 \, d\omega = 0 \tag{6.48a}$$

$$\frac{\partial J}{\partial a[k]} = \frac{1}{\sigma^2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left[A(e^{j\omega})e^{j\omega k} + A^*(e^{j\omega})\, e^{-j\omega k}\right] \left|Y_N(e^{j\omega})\right|^2 \, d\omega = 0. \tag{6.48b}$$

From (6.48a), we see that the ML estimate of σ^2 is specified in terms of the ML estimates of the filter parameters:

$$\hat{\sigma}^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|Y_N(e^{j\omega})\right|^2 \left|\hat{A}(e^{j\omega})\right|^2 \, d\omega \tag{6.49}$$

where

$$\hat{A}(e^{j\omega}) = 1 - \sum_{n=1}^{M} \hat{a}[n] e^{-j\omega n}. \tag{6.50}$$

To obtain the filter parameter estimates, we note that since the sequence $a[n]$ is real, (6.48b) simplifies to the condition

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{A}(e^{j\omega}) \left|Y_N(e^{j\omega})\right|^2 e^{j\omega k} \, d\omega = 0, \tag{6.51}$$

which using (6.50) can be expressed directly in terms of the coefficients $\hat{a}[n]$, i.e.,

$$\sum_{l=1}^{M} \hat{a}[l] \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|Y_N(e^{j\omega})\right|^2 e^{j\omega(k-l)} \, d\omega\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|Y_N(e^{j\omega})\right|^2 e^{j\omega k} \, d\omega. \tag{6.52}$$

To put (6.52) in its final form, note that as a consequence of (6.4) we have

$$\left|Y_N(e^{j\omega})\right|^2 = \mathcal{F}\left\{\hat{K}_{yy}[n]\right\} \tag{6.53}$$

where $\hat{K}_{yy}[n]$ is the sample covariance function

$$\hat{K}_{yy}[n] = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{k=0}^{N-1-|n|} y[k]\, y[k+|n|] & |n| \le N-1 \\ 0 & \text{otherwise} \end{cases}. \tag{6.54}$$

Hence, the ML equations can be written in the form

$$\sum_{l=1}^{M} \hat{a}[l]\, \hat{K}_{yy}[k-l] = \hat{K}_{yy}[k], \qquad k = 1, 2, \ldots, M, \tag{6.55}$$

i.e.,

$$\begin{bmatrix} \hat{K}_{yy}[0] & \hat{K}_{yy}[1] & \cdots & \hat{K}_{yy}[M-1] \\ \hat{K}_{yy}[1] & \hat{K}_{yy}[0] & \cdots & \hat{K}_{yy}[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{K}_{yy}[M-1] & \hat{K}_{yy}[M-2] & \cdots & \hat{K}_{yy}[0] \end{bmatrix} \begin{bmatrix} \hat{a}[1] \\ \hat{a}[2] \\ \vdots \\ \hat{a}[M] \end{bmatrix} = \begin{bmatrix} \hat{K}_{yy}[1] \\ \hat{K}_{yy}[2] \\ \vdots \\ \hat{K}_{yy}[M] \end{bmatrix}. \tag{6.56}$$

The equations (6.56) are referred to as the *estimated Yule-Walker equations*, and the resulting algorithm as the *autocorrelation method of linear prediction* for reasons that will become apparent.[3] It is also possible to show that the solution of these equations have a variety of attractive characteristics, perhaps the most important of which is that for any N the resulting $\hat{a}[n]$'s always correspond to an $\hat{A}(z)$ with its zeros strictly inside the unit circle, so that the resulting all-pole modeling filter $1/\hat{A}(z)$ is always stable. Furthermore, it can be shown that the resulting parameter estimates are consistent.

Equally importantly, the equations (6.56) can be efficiently solved using a fast algorithm referred to as *Levinson's recursion*. The problem of solving a set of M linear equations (e.g., by Gaussian elimination) has, in general, a computational complexity of $\mathcal{O}(M^3)$. However, with Levinson's recursion (6.56) can be solved with $\mathcal{O}(M^2)$ complexity. While a detailed development of Levinson's algorithm is beyond the scope of these notes, it is worth pointing out that the algorithm exploits the special structure of estimated Yule-Walker equations, recursively computing the Mth-order model parameters from those of the $(M-1)$st-order model, etc. This means, in addition, that as a byproduct the Levinson algorithm yields all lower-order solutions to the modeling problem, which can be useful in scenarios when the appropriate model order is not known *a priori*.

Finally, note that $\hat{\sigma}^2$ can be readily obtained from the filter parameter estimates. In particular, using (6.50) and (6.53) in (6.49) we obtain

$$\hat{\sigma}^2 = \hat{K}_{yy}[0] - \sum_{k=1}^{M} \hat{a}[k]\, \hat{K}_{yy}[k]. \tag{6.57}$$

It is also worth pointing out that for nonGaussian problems—and even purely deterministic ones—the same parameter estimates are obtained as the solution corresponding to a different performance criterion that does not require knowledge of process statistics. In particular, let $\hat{w}[n]$ denote the output of an FIR filter with system function $\hat{A}(z)$ driven by the zero-padded observations

$$\tilde{y}[n] = \begin{cases} y[n] & n = 0, 1, \dots, N-1 \\ 0 & \text{otherwise} \end{cases}. \tag{6.58}$$

[3]By contrast, the exact ML estimates for the closely related Gauss-Markov parameter estimation problem in the homework correspond to the *autocovariance* method of linear prediction.

Then the coefficients $\hat{a}[n]$ that minimize the total energy at the output of the filter, i.e., $\sum_n \hat{w}^2[n]$, are obtained as the least-squares solution of a (overdetermined) set of $N + M$ linear equations for the M unknown coefficients, which in turn corresponds to the solution of (6.56). Moreover, $\hat{\sigma}^2$ corresponds to the average power in the resulting minimum energy output, i.e.,

$$\hat{\sigma}^2 = \frac{1}{N+M} \sum_n \hat{w}^2[n]. \tag{6.59}$$

6.3 LINEAR PREDICTION

Let us finish with an observation of a key connection between the modeling problems developed above, and a fundamentally different Bayesian estimation problem involving prediction. Suppose $y[n]$ is a zero-mean wide-sense stationary random process with known covariance function $K_{yy}[n]$. In this problem, $K_{yy}[n]$ is arbitrary—we do not assume that $y[n]$ is an autoregressive process of any order. And let us develop a linear-least squares estimate of $y[n]$ based on $y[n-1], y[n-2], \ldots, y[n-M]$. We write this estimate in the form

$$\hat{y}[n] = a[1]\, y[n-1] + a[2]\, y[n-2] + \cdots + a[M]\, y[n-M], \tag{6.60}$$

and recognize that the one-step prediction process $y[n]$ can be viewed as the output of a M-tap FIR filter with system function

$$A(z) = \sum_{n=1}^{M} a[n]\, z^{-n} \tag{6.61}$$

driven by $y[n-1]$. Then by the orthogonality principle developed in Chapter 3, the optimum predictions have the property that the corresponding prediction errors are orthogonal to the data, leading to the normal equations

$$E\left[(\hat{y}[n] - y[n])\, y[n-k]\right] = 0, \qquad k = 1, 2, \ldots, M, \tag{6.62}$$

which can be written in the form

$$E\left[\sum_{l=1}^{M} \hat{a}[l]\, y[n-l]\, y[n-k]\right] = E\left[y[n]\, y[n-k]\right], \qquad k = 1, 2, \ldots, M. \tag{6.63}$$

Hence,

$$\sum_{l=1}^{M} \hat{a}[l]\, K_{yy}[k-l] = K_{yy}[k], \qquad k = 1, 2, \ldots, M \tag{6.64}$$

which, in matrix form, corresponds to

$$\begin{bmatrix} K_{yy}[0] & K_{yy}[1] & \cdots & K_{yy}[M-1] \\ K_{yy}[1] & K_{yy}[0] & \cdots & K_{yy}[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ K_{yy}[M-1] & K_{yy}[M-2] & \cdots & K_{yy}[0] \end{bmatrix} \begin{bmatrix} \hat{a}[1] \\ \hat{a}[2] \\ \vdots \\ \hat{a}[M] \end{bmatrix} = \begin{bmatrix} K_{yy}[1] \\ K_{yy}[2] \\ \vdots \\ K_{yy}[M] \end{bmatrix}. \tag{6.65}$$

The normal equations (6.65) for this problem are referred to as the *Yule-Walker* or *linear prediction* equations. Interestingly they are strikingly similar to those that arose in the very different estimation problem involving autoregressive modeling, viz., (6.56). Note that (6.65) differs from (6.56) in that samples of the true covariance function for $y[n]$ are involved in the former rather than the estimated covariance function. Nevertheless, the similarities in the form of the equations can be exploited in a variety of ways. As one important example, Levinson's recursion can be used to efficiently solve this linear prediction problem.

It is also interesting to note that the resulting $\hat{a}$'s in (6.65) do not depend on n—i.e., the prediction filter is time-invariant. This is a consequence of the fact that the process $y[n]$ is wide-sense stationary. Also, the variance of the prediction error follows from an application of Pythagoras' theorem exploiting the orthogonality characteristics of the error:

$$\lambda_{\text{LLS}} = \text{var}\,[\hat{y}[n] - y[n]] = K_{yy}[0] - \sum_{n=1}^{M} \hat{a}[n]\, K_{yy}[n], \tag{6.66}$$

which we see also has a form directly analogous to that of (6.57).

In the next section of the course, we will more generally explore linear-least squares estimation problems involving random processes.