A multi-dimensional prediction problem

By MURRAY ROSENBLATT[*]

1. Introduction

The problem of linear prediction for a weakly stationary stochastic process has been discussed in considerable detail by Kolmogorov [3], Wiener [6] and others. Recently there has been increasing interest in the linear prediction problem for a vector-valued weakly stationary process. Aspects of this problem have been treated in a heuristic manner by Whittle [5] and analytically by Wiener [7]. The discussion in this paper is more probabilistic in orientation and some attention is devoted to the problem of computing the prediction error covariance matrix in a one-step prediction when the process is a two-vector.

2. Preliminary discussion

Let

$$x_t = \begin{pmatrix} {}_1x_t \\ \vdots \\ {}_mx_t \end{pmatrix}, \quad t = \cdots, -1, 0, 1, \ldots, E\,x_t \equiv 0,$$

be an m-vector weakly stationary stochastic process. By this we mean that the sequence of covariance matrices $(m \times m)$

$$r_{t,\tau} = E\,x_t\,x_\tau' = r_{t-\tau}[1]$$

depends only on the difference $t - \tau$. It is then well known that the sequence of covariance matrices r_t can be represented as the Fourier–Stieltjes coefficients

$$r_t = \int_{-\pi}^{\pi} e^{it\lambda}\,d\,F(\lambda) \tag{1}$$

of a matrix-valued $(m \times m)$ non-decreasing function $F(\lambda)$. The function $F(\lambda)$ is said to be non-decreasing since for any given m-vector v

$$v'\,\Delta\,F(\lambda)\,v = v'\,[F(\lambda_2) - F(\lambda_1)]\,v \geq 0,$$

* Indiana University and New York University. Written for the David Taylor Model Basin under ONR contract Nonr 285 (17).

[1] Given the matrix A, A' denotes the conjugated transpose of the matrix A.

$\lambda_2 \geq \lambda_1$. The case of interest to us is that in which $F(\lambda)$ is absolutely continuous, i.e.

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\mu) \, d\mu.$$

The matrix-valued function $(m \times m)$ $f(\lambda)$ is called the spectral density of the process x_t. Note that $f(\lambda)$ is non-negative, i.e. for any given m-vector v

$$v' f(\lambda) v \geq 0.$$

Since $f(\lambda)$ is an $m \times m$ matrix-valued function we shall write $f(\lambda)$ in the form

$$f(\lambda) = \{f_{jk}(\lambda); \ j, k = 1, \ldots, m\}.$$

The function $f_{jj}(\lambda)$ is the spectral density of the jth component $_jx_t$ of the process while $f_{jk}(\lambda)$, $j \neq k$, is the cross-spectral density of $_jx_t$ and $_kx_t$. Since $f(\lambda) \geq 0$ it follows that $f_{jj}(\lambda) \geq 0$. In general $f_{jk}(\lambda)$, $j \neq k$, is complex-valued. The real part of $f_{jk}(\lambda)$, $\mathrm{Re} \, f_{jk}(\lambda)$, is called the cospectrum of $_jx_t$, $_kx_t$ while the imaginary part of $f_{jk}(\lambda)$, $\mathrm{Im} \, f_{jk}(\lambda)$, is called the quadrature spectrum of $_jx_t$, $_kx_t$.

There is a random representation of the process x_t itself analogous to the representation (1)

$$x_t = \int_{-\pi}^{\pi} e^{it\lambda} \, dZ(\lambda),$$

where $Z(\lambda)$ is an m-vector valued process with orthogonal increments

$$E \, dZ(\lambda) \, dZ(\mu)' = \delta_{\lambda\mu} \, dF(\lambda) = \delta_{\lambda\mu} f(\lambda) \, d\lambda.$$

3. The problem

Assume that the vector-valued process $\{x_t\}$ has been observed at times $t = n, \ n - 1, \ldots$ and that we are interested in predicting $\nu > 0$ steps ahead. In particular the case of greatest interest to us will be that in which we predict one step ahead. We limit ourselves to predictors linear in the observations $x_n, \ x_{n-1}, \ldots$ that is, to expressions

$$\sum_{k=0}^{\infty} a_k \, x_{n-k},$$

where the a_k's are $m \times m$ matrices or to limits (in mean square) of such expressions. The best linear predictor $x_{n+\nu}^*$ is the one whose error

$$x_{n+\nu} - x_{n+\nu}^*$$

is smallest. In this context the error is smallest if the covariance matrix

$$E \, (x_{n+\nu} - x_{n+\nu}^*) \, (x_{n+\nu} - x_{n+\nu}^*)'$$

of the error is smaller than that of the error obtained if any other linear predictor in terms of the observations x_n, x_{n-1}, ... were used. The case of greatest interest, as in the case of prediction for a one-dimensional process, is that in which the prediction error is positive, that is, the covariance matrix of the prediction error is a positive definite matrix. A necessary condition for the error to be positive will be obtained in section 5. In section 6 it will be shown that the prediction error is positive if the spectral distribution function is absolutely continuous and the spectral density function is continuous and nonsingular. The question of obtaining simple and explicit formulae for the predictor and the covariance matrix of the prediction error in terms of the spectrum is considered in section 7 when the process is two-dimensional and the diagonal elements of the spectral density function are equal. This question is one of practical interest and in general is difficult.

Now consider a context in which such a prediction problem arises naturally. Even though the stationary processes here are continuous parameter stationary processes, the prediction problem that arises is quite analogous to that discussed above. W. J. Pierson Jr. has constructed the following model of a storm generated ocean surface. The ocean surface is given by

$$\eta(u, v, t) = \int_0^\infty \int_{-\pi}^\pi \left[\cos\left(\frac{\mu^2}{g}(u \cos \theta + v \sin \theta) - \mu t\right) dZ_1(\mu, \theta) \right.$$

$$\left. + \sin\left(\frac{\mu^2}{g}(u \cos \theta + v \sin \theta) - \mu t\right) dZ_2(\mu, \theta) \right],$$

where (u, v) is the position on the ocean surface and t is the time. Here μ is the frequency and g is the gravitational constant. The random processes $Z_1(\mu, \theta)$, $Z_2(\mu, \theta)$ are orthogonal to each other and are processes with orthogonal increments, that is

$$E Z_1(\mu, \theta) = E Z_2(\mu, \theta) \equiv 0$$

$$E Z_1(\mu, \theta) Z_2(\mu', \theta') \equiv 0$$

and

$$E dZ_i(\mu, \theta) dZ_i(\mu', \theta') = \delta_{\mu\mu'} \delta_{\theta\theta'} f(\mu, \theta) d\mu d\theta, \quad i = 1, 2.$$

Thus $dZ_1(\mu, \theta)$, $dZ_2(\mu, \theta)$ are the random amplitudes of long crested waves of frequency μ with direction of propagation at an angle θ to the u-axis. The ocean surface is a superposition of such waves. This process is stationary in (u, v) and t. It can, of course, only be considered as a realistic model well within the storm area and over time intervals short compared to the time length of the storm. Assume that the sea surface is observed at the two points (u, v) and $(u + \Delta, v)$ up to time $t = 0$ and that one wishes to predict the sea surface at $(u + \Delta, v)$ at time $\tau > 0$. The two-dimensional process

$$x_t = \begin{pmatrix} \eta(u + \Delta, v, t) \\ \eta(u, v, t) \end{pmatrix}$$

has the matrix (2×2) spectral density

$$f(\lambda) = \frac{1}{2} \left\{ \begin{array}{cc} \displaystyle\int_{-\pi}^{\pi} f(\lambda, \theta) \, d\theta & \displaystyle\int_{-\pi}^{\pi} e^{-i\frac{\lambda^2 \Delta}{\sigma} \cos\theta} f(\lambda, \theta) \, d\theta \\[4mm] \displaystyle\int_{-\pi}^{\pi} e^{i\frac{\lambda^2 \Delta}{\sigma} \cos\theta} f(\lambda, \theta) \, d\theta & \displaystyle\int_{-\pi}^{\pi} f(\lambda, \theta) \, d\theta \end{array} \right\}$$

if $\lambda \geq 0$. We define $f(-\lambda) = \overline{f(\lambda)}$. Note that the problem of predicting

$$\eta(u + \Delta, v, \tau)$$

is the same as that of predicting $\eta(u, v, \tau)$ from the past of both processes. In other words, it really is equivalent to that of predicting the two-vector x_t from its past.

4. Remarks on orthogonalization

Assume that x_t is a process with an absolutely continuous spectral distribution function and a non-singular spectral density function $f(\lambda)$. Suppose that $x_n, x_{n-1}, \ldots, x_{n-p}$ have been observed and we wish to predict x_{n+1} by the best linear predictor in terms of the least squares criterion we have adopted. The predictor has the form

$$x^*_{n+1} = \sum_{k=0}^{n} a_k x_{n-k}$$

and is such that

$$E\left(x_{n+1} - x^*_{n-1}\right)\left(x_{n+1} - x^*_{n+1}\right)'$$

$$= \min_{b_i} E\left(x_{n+1} - \sum_{k=0}^{p} b_k x_{n-k}\right)\left(x_{n+1} - \sum_{k=0}^{p} b_k x_{n-k}\right)'.$$

In terms of the spectral representation we can look at the prediction problem as a minimization problem

$$E\left(x_{n+1} - x^*_{n+1}\right)\left(x_{n+1} - x^*_{n+1}\right)'$$

$$= \int_{-\pi}^{\pi} \left(I - \sum_{k=0}^{p} a_k e^{-i(k+1)\lambda}\right) f(\lambda) \left(I - \sum_{k=0}^{p} a_k e^{-i(k+1)\lambda}\right)' d\lambda$$

$$= \min_{b_i} \int_{-\pi}^{\pi} \left(I - \sum_{k=0}^{p} b_k e^{-i(k+1)\lambda}\right) f(\lambda) \left(I - \sum_{k=0}^{p} b_k e^{-i(k+1)\lambda}\right)' d\lambda.$$

Now we will briefly discuss matrix-valued $(m \times m)$ orthogonal polynomials in $e^{i\lambda}$ and then interpret aspects of our minimization problem in terms of these polynomials. Consider generating the system $\varphi_0(\lambda)$, $\varphi_1(\lambda)$, ... of orthogonal polynomials in $e^{i\lambda}$ with weight function $f(\lambda)$ recursively by the Gram–Schmidt orthogonalization procedure from I, $e^{i\lambda}I$, ... Then

410

$$\int_{-\pi}^{\pi} \varphi_j(\lambda) f(\lambda) \varphi_k(\lambda)' d\lambda = \delta_{jk} I.$$

Assume that $\varphi_0(\lambda), \ldots, \varphi_p(\lambda)$ have been generated and that we wish to approximate a given function $g(\lambda)$ in the mean square by a linear combination of $\varphi_0(\lambda), \ldots, \varphi_p(\lambda)$. Now

$$\int_{-\pi}^{\pi} \left[g(\lambda) - \sum_{j=0}^{p} c_j \varphi_j(\lambda) \right] f(\lambda) \left[g(\lambda) - \sum_{j=0}^{p} c_j \varphi_j(\lambda) \right]' d\lambda$$

$$= (g, g) + \sum_{j=0}^{p} [c_j - (g, \varphi_j)][c_j - (g, \varphi_j)]' - \sum_{j=0}^{p} (g, \varphi_j)(g, \varphi_j)', \qquad (2)$$

where $(g, h) = \int_{-\pi}^{\pi} g(\lambda) f(\lambda) h(\lambda)' d\lambda$. The error (2) is minimized if we set

$$c_j = (g, \varphi_j).$$

If in particular $g(\lambda) = e^{i(p+1)\lambda} I$, the best linear approximation in terms of $\varphi_0(\lambda), \ldots, \varphi_p(\lambda)$ is

$$\sum_{j=0}^{p} (e^{i(p+1)\lambda} I, \varphi_j(\lambda)) \varphi_j(\lambda).$$

Note that

$$e^{i(p+1)\lambda} I - \sum_{j=0}^{p} (e^{i(p+1)\lambda} I, \varphi_j(\lambda)) \varphi_j(\lambda) \qquad (3)$$

is orthogonal to $\varphi_0(\lambda), \ldots, \varphi_p(\lambda)$. On normalizing (3) we obtain $\varphi_{p+1}(\lambda)$. Note that

$$x_{n+1}^* = \int_{-\pi}^{\pi} e^{i(n+p)\lambda} \sum_{j=0}^{p} (e^{i(p+1)\lambda} I, \varphi_j) \varphi_j(\lambda) dZ(\lambda)$$

in the prediction problem spoken of at the beginning of this section. The prediction error

$$E(x_{n+1} - x_{n+1}^*)(x_{n+1} - x_{n+1}^*)'$$

Is the square [1] of the inverse of the coefficient of $e^{i(p+1)\lambda}$ in $\varphi_{p+1}(\lambda)$. We have in effect given here procedures which can be used to obtain the predictor and predictor error in a computing program.

5. A necessary condition for positive prediction error

Let x_t, $E x_t \equiv 0$, $t = \cdots, -1, 0, 1, \ldots$ be a weakly stationary m-vector stochastic process, i.e.

$$r_{t,\tau} = r_{t-\tau} = \int_{-\pi}^{\pi} e^{i(t-\tau)\lambda} dF(\lambda),$$

---

[1] The square of a matrix M is understood to be MM'.

where $F(\lambda)$ is a non-decreasing $(m \times m)$ matrix-valued function. It is clear that one ought to assume r_0 non-singular, for if this were not true one could in any case reduce x_t to a weakly stationary process y_t of lower dimension but equal rank $m' < m$ with a non-singular covariance matrix of lag zero. Consider the projection of x_t on the closed linear manifold $\mathfrak{M}_{t-1}$([1]) generated by $x_{t-1}, x_{t-2}, \ldots$. This projection x_t^* is an m-vector whose components are the projections of the components of x_t on $\mathfrak{M}_{t-1}$ respectively. Note that the one-step prediction error

$$E(x_t - x_t^*)(x_t - x_t^*)' = \min E(x_t - y)(x_t - y)', \qquad {}_1y, \ldots, {}_my \in \mathfrak{M}_{t-1},$$

where

$$y = \begin{pmatrix} {}_1y \\ \vdots \\ {}_my \end{pmatrix}$$

Let $\eta_t = x_t - x_t^*$. The η_t's are orthogonal to each other, i.e.

$$E\,\eta_t\,\eta_\tau' = 0 \quad \text{if} \quad t \neq \tau.$$

Consider $E\,\eta_t\,\eta_t'$. The case of interest to us is that in which the *prediction error is positive*, that is, $E\,\eta_t\,\eta_t'$ is a non-singular matrix (a matrix of rank m). Normalize the random vector η_t. Call the normalized vector ξ_t

$$\xi_t = (E\,\eta_t\,\eta_t')^{-\frac{1}{2}}\,\eta_t.$$

Let

$$c_j = E\,x_t\,\xi_{t-j}'.$$

Then

$$x_t = \sum_{j=0}^{\infty} c_j\,\xi_{t-j} + v_t,$$

where $v_t \in \bigcap_{t=-\infty}^{\infty} \mathfrak{M}_t$. Note that the process v_t is orthogonal to the weakly stationary process ξ_t. The process ξ_t is a process of orthonormal random vectors

---

[1] This closed linear manifold $\mathfrak{m}_{t-1}$ is simply the space of random variables which can be obtained either as finite linear combinations of the components of $x_{t-1}, x_{t-2}, \ldots$ or as limits in the mean square of such random variables. $\mathfrak{m}_{t-1}$ is a linear space. The random variables of $\mathfrak{m}_{t-1}$ can be considered points in $\mathfrak{m}_{t-1}$. The distance between two points (random variables) in $\mathfrak{m}_{t-1}$ x, y is

$$\| x - y \| = E^{\frac{1}{2}} | x - y |^2. \tag{4}$$

Most of the usual techniques appropriate in the case of a finite dimensional space can be carried over to this infinite dimensional space. In particular the notion of a projection on a closed linear manifold is valid just as in the finite dimensional space, i.e. the projection $p\,z$ of a random variable z on $\mathfrak{m}_{t-1}$ is the random variable in $\mathfrak{m}_{t-1}$ which is closest to z using the distance (4). The difference between z and its projection on $\mathfrak{m}_{t-1}$ is orthogonal to $\mathfrak{m}_{t-1}$ (perpendicular to $\mathfrak{m}_{t-1}$), i.e. the inner product of $z - p\,z$ and any element x of $\mathfrak{m}_{t-1}$

$$(z - p\,z,\ x) = E\,(z - p\,z)\,x' = 0.]$$

See Halmos [1] for a detailed discussion of these points. An m-vector v is loosely spoken of as being in the closed linear manifold $\mathfrak{m}$ if all its components are in $\mathfrak{m}$.

412

$$E \, \xi_t \, \xi_\tau' = \delta_{t\tau} I.$$

The process x_t is called a purely non-deterministic process if $v_t \equiv 0$.

The function
$$c(z) = \sum_{j=0}^{\infty} c_j \, z^j$$

is a matrix-valued $(m \times m)$ function of a complex argument z. The function $c(z)$ is analytic in the unit circle $|z| < 1$ since

$$\sum_{j=0}^{\infty} c_j c_j' < \infty. \tag{5}$$

Moreover (5) implies that
$$c(e^{-i\lambda}) = \lim_{r \to 1-} c(r e^{-i\lambda})$$

exists almost everywhere. We shall show that *the determinant* $|c(z)|$ *is not zero at any point in the unit circle* $|z| < 1$. Now $|c(0)| = |c_0|$ is not zero since we have assumed positive prediction error, that is, c_0 is assumed non-singular. If $c(z)$ were singular at some point $z_0 \neq 0$, $|z_0| < 1$, there would be some vector α such that

$$c(z_0) \, \alpha = \sum_{j=0}^{\infty} c_j z_0^j \, \alpha = 0.$$

Consider
$$x = \sum_{j=0}^{\infty} z_0^j \, \xi_{t-j}' \, \alpha.$$

Now $x \in \mathfrak{M}_t$ but not $\mathfrak{M}_{t-1}$. However

$$E \, \beta' \, x_t \, x = \beta' \sum_{j=0}^{\infty} c_j z_0^j \, \alpha = 0$$

for all vectors β so that $x \in \mathfrak{M}_{t-1}$ and we are thus led to a contradiction.

The processes x_t, ξ_t, v_t are weakly stationary. But ξ_t, v_t have the reprensentation

$$\xi_t = \int_{-\pi}^{\pi} e^{it\lambda} \, dZ_\xi(\lambda), \qquad E \, dZ_\xi(\lambda) \, dZ_\xi(\mu)' = \delta_{\lambda\mu} I \, d\mu,$$

$$v_t = \int_{-\pi}^{\pi} e^{it\lambda} \, dZ_v(\lambda), \qquad E \, dZ_v(\lambda) \, dZ_v(\mu)' = \delta_{\lambda\mu} \, dF_v(\lambda).$$

We know that

$$x_t = \sum_{j=0}^{\infty} c_j \, \xi_{t-j} + v_t$$

$$= \int_{-\pi}^{\pi} e^{it\lambda} c(e^{-i\lambda}) \, dZ_\xi(\lambda) + \int_{-\pi}^{\pi} e^{it\lambda} \, dZ_v(\lambda).$$

Because of the orthogonality of the ξ_t and v_t processes, it is clear that

$$F(\lambda) = \int\limits_{-\pi}^{\lambda} \frac{1}{2\pi} c(e^{-i\mu}) c(e^{-i\mu})' \, d\mu + F_v(\lambda).$$

Since the components of ξ_t are in $\mathfrak{M}_t$, it follows that

$$\xi_t = \int\limits_{-\pi}^{\pi} e^{it\lambda} \Phi(\lambda) \, dZ(\lambda)$$

$$= \int\limits_{-\pi}^{\pi} e^{it\lambda} \Phi(\lambda) c(e^{-i\lambda}) \, dZ_\xi(\lambda) + \int\limits_{-\pi}^{\pi} e^{it\lambda} \Phi(\lambda) \, dZ_v(\lambda)$$

$$= \int\limits_{-\pi}^{\pi} e^{it\lambda} \, dZ_\xi(\lambda).$$

The function $|c(z)|$ is a power series in z that belongs to H_2 and hence it follows that $|c(e^{-i\lambda})|$ is not zero for almost all λ (see G. Szego [4], p. 267). But then $c(e^{-i\lambda})$ is non-singular for almost all λ (Lebesgue measure). Thus

$$\Phi(\lambda) c(e^{-i\lambda}) = I$$

for almost all λ (Lebesgue measure) and $\Phi(\lambda) = 0$ for almost all λ (F_v measure). Thus

$$F'(\lambda) = \frac{1}{2\pi} c(e^{-i\lambda}) c(e^{-i\lambda})'$$

and

$$|c(e^{-i\lambda})|^2 = (2\pi)^m |F'(\lambda)|.$$

Since $|c(z)|$ has no zeros in the unit circle $|z| < 1$, it follows that

$$\log |F'(\lambda)|$$

is integrable and that the determinant of the one-step prediction error covariance matrix

$$|c(0)|^2 = |c_0|^2 = (2\pi)^m \exp\left\{ \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \log |F'(\lambda)| \, d\lambda \right\} \tag{6}$$

(see J. L. Doob [1], p. 577). We have thus shown that *if the prediction error is positive,* $\log |F'(\lambda)|$ *is integrable and that the determinant of the prediction error covariance is given by* (6). Note that though the determinant of the prediction error covariance matrix is easily obtained, this is in general not true of the elements of the covariance matrix. It is the latter problem that is of greatest interest.

414

6. A sufficient condition for positive prediction error

We shall restrict ourselves in this section to an m-vector weakly stationary process x_t, $E x_t \equiv 0$, with *an absolutely continuous spectral distribution function* $F(\lambda)$. Moreover *the spectral density* $f(\lambda)$ *is assumed to be continuous and non-singular for all* λ. Under these conditions, the covariance matrix of the one-step prediction will be shown to be positive definite.

First consider the case of a positive definite matrix-valued $(m \times m)$ polynomial of finite order in $e^{i\lambda}$

$$f(\lambda) = \sum_{k=-p}^{p} c_k e^{ik\lambda} > 0.$$

Now one can construct a weakly stationary m-vector x_t, $E x_t \equiv 0$, with spectral density $f(\lambda)$. Consider the random vector x_t^* whose components are the projections of the components of x_t on $\mathfrak{M}_{t-1}$. Let

$$\eta_t = x_t - x_t^*.$$

The random vectors η_t are orthogonal, i.e.,

$$E \eta_t \eta_\tau' = 0 \quad \text{if } t \neq \tau.$$

The random vectors η_t all have the same covariance matrix, let us say R. From the form of $f(\lambda)$, it follows that the components of x_t are orthogonal to $\mathfrak{M}_{t-p-1}$. This means that

$$x_t = \sum_{k=0}^{p} a_k \eta_{t-k}.$$

Now

$$x_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda), \qquad E \, dZ(\lambda) \, dZ(\mu)' = \delta_{\lambda\mu} f(\lambda) \, d\lambda,$$

and

$$\eta_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ_\eta(\lambda), \qquad E \, dZ_\eta(\lambda) \, dZ_\eta(\mu)' = \frac{\delta_{\lambda\mu}}{2\pi} R \, d\lambda.$$

Thus

$$x_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda) = \sum_{k=0}^{p} a_k \eta_{t-k}$$

$$= \int_{-\pi}^{\pi} e^{it\lambda} \sum_{k=0}^{p} a_k e^{-ik\lambda} dZ_\eta(\lambda).$$

It then follows that

$$Z(\lambda) = \int_{-\pi}^{\lambda} \sum_{k=0}^{p} a_k e^{-ik\lambda} dZ_\eta(\lambda). \tag{7}$$

415

On taking the covariance of both sides of equation (7) we have

$$\int\limits_{-\pi}^{\lambda} f(\mu)\, d\mu = \int\limits_{-\pi}^{\lambda} \left(\sum_{k=0}^{p} a_k\, e^{-ik\mu} \right) \frac{1}{2\pi} R \left(\sum_{k=0}^{p} a_k\, e^{-ik\mu} \right)' d\mu$$

so that

$$f(\lambda) = \frac{1}{2\pi} \left(\sum_{k=0}^{p} a_k\, e^{-k\lambda} \right) R \left(\sum_{k=0}^{p} a_k\, e^{-ik\lambda} \right)'.$$

Since $f(\lambda)$ is positive definite, R must be non-singular. But then the η_t's can be normalized so as to obtain the ξ_t's

$$\xi_t = R^{-\frac{1}{2}}\, \eta_t$$

$$E\, \xi_t\, \xi_\tau' = \delta_{t,\tau}\, I.$$

Using the argument given above one can see that there are coefficients b_k such that

$$x_t = \sum_{k=0}^{p} b_k\, \xi_{t-k}.$$

Thus

$$f(\lambda) = \frac{1}{2\pi} \left(\sum_{k=0}^{p} b_k\, e^{-ik\lambda} \right) \left(\sum_{k=0}^{p} b_k\, e^{-ik\lambda} \right)'.$$

Note that

$$x_t - x_t^* = b_0\, \xi_t$$

and the ξ_t's are constructed so that $b_0 = (b_0\, b_0')^{\frac{1}{2}}$ where $b_0\, b_0'$ is the prediction error. It is clear that the optimal one-step predictor is

$$x_t^* = \sum_{k=1}^{p} b_k\, \xi_{t-k}.$$

Note that b_0 is non-singular so that the prediction error covariance $b_0\, b_0'$ is positive definite. Given the mode of construction of the coefficients b_k, it is clear that

$$|b(z)| \neq 0 \quad \text{for all } |z| < 1,$$

where

$$b(z) = \sum_{k=0}^{p} b_k\, z^k.$$

We have in effect given a generalization of Fejér's theorem, for we have shown that *given a matrix-valued $(m \times m)$ positive definite polynomial $f(\lambda)$ of order p in $e^{i\lambda}$*

$$f(\lambda) = \sum_{k=-p}^{p} c_k\, e^{ik\lambda} > 0$$

one can write $f(\lambda)$ *in the form*

$$f(\lambda) = \frac{1}{2\pi} \left(\sum_{k=0}^{p} b_k e^{-ik\lambda} \right) \left(\sum_{k=0}^{p} b_k e^{-ik\lambda} \right)'$$

with b_0 *positive definite and* $|b(z)| \neq 0$ *when* $|z| < 1$. Here $b(z) = \sum_{k=0}^{p} b_k z^k$. This representation is uniquue.

Now we use the generalization of Fejér's theorem to obtain the desired result for any positive definite continuous matrix-valued $(m \times m)$ function $f(\lambda)$. Given any such function $f(\lambda)$, one can find a positive definite polynomial $g(\lambda)$ in $e^{i\lambda}$ of finite order such that $f(\lambda) \geq g(\lambda)$. The prediction error for a weakly stationary process with spectral density $g(\lambda)$ has positive definite covariance matrix and hence this also must be true of a weakly stationary process with spectral density $f(\lambda)$. Using the argument of the previous section we can then show that $f(\lambda)$ has the representation

$$f(\lambda) = \frac{1}{2\pi} \left(\sum_{k=0}^{\infty} a_k e^{-ik\lambda} \right) \left(\sum_{k=0}^{\infty} a_k e^{-ik\lambda} \right)'$$

with a_0 positive definite and $|a(z)| \neq 0$ when $|z| < 1$. Here $a(z) = \sum_{k=0}^{\infty} a_k z^k$. Moreover, this representation is unique.

It is worthwhile looking briefly at a process whose spectral density is the inverse of a matrix-valued $(m \times m)$ polynomial in $e^{i\lambda}$ of finite order

$$f(\lambda) = \frac{1}{2\pi} \left(\sum_{k=-p}^{p} c_k e^{-ik\lambda} \right)^{-1} > 0.$$

From the result cited above it is clear that we can write

$$f(\lambda) = \frac{1}{2\pi} \left(\sum_{k=0}^{p} a_k e^{-ik\lambda} \right)^{-1} \left(\sum_{k=0}^{p} a_k e^{-ik\lambda} \right)'^{-1},$$

where a_0 is positive definite and $|a(z)| \neq 0$ when $|z| < 1$. Here $a(z) = \sum_{k=0}^{p} a_k z^k$. Now the process x_t has the representation

$$x_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda).$$

Let

$$\xi_t = \int_{-\pi}^{\pi} e^{it\lambda} \left(\sum_{k=0}^{p} a_k e^{-ik\lambda} \right) dZ(\lambda).$$

Then

$$E\, \xi_t \xi_\tau' = \delta_{t\tau} I.$$

Note that

$$\sum_{k=0}^{p} a_k x_{t-k} = \int_{-\pi}^{\pi} e^{it\lambda} \left(\sum_{k=0}^{p} a_k e^{-ik\lambda} \right) dZ(\lambda) = \xi_t$$

and

$$E\,\xi_t x_{t-j}' = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{it\lambda} \left(\sum_{k=0}^{p} a_k e^{-ik\lambda} \right) f(\lambda)\, e^{-i(t-j)\lambda}\, d\lambda$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ij\lambda} \left(\sum_{k=0}^{p} a_k e^{-ik\lambda} \right)'^{-1} d\lambda = 0$$

if $j > 0$. Thus the best linear predictor of x_t given $x_{t-1}, x_{t-2}, \ldots$ is

$$x_t^* = \sum_{k=1}^{p} a_0^{-1} a_k x_{t-k}.$$

One can obtain the function $\sum_{k=0}^{p} a_k e^{-ik\lambda}$ in the case of such a spectral density $f(\lambda)$ by the following procedure. By the Gram–Schmidt orthogonalization procedure we obtain the polynomial in $e^{i\lambda}$

$$\sum_{k=0}^{p} b_k e^{ik\lambda}$$

which is normalized and orthogonal to $I, e^{i\lambda}I, \ldots, e^{i(p-1)\lambda}I$ with respect to the weight function $f(\lambda)$. On multiplying this polynomial by $e^{-ip\lambda}$ one obtains the desired expression.

7. Computational aspects of the problem

It is interesting to examine the computational aspect of the prediction problem in the case of a two-dimensional process. We shall see that there are various difficulties that arise in the attempt to get an explicit and simple representation of $c(e^{-i\lambda})$ (see section 5) and the prediction error covariance in terms of $f(\lambda)$.

Before going on let us review a few aspects of the one-dimensional case. Let x_t be a one-dimensional weakly stationary process, $t = \cdots, -1, 0, 1, \ldots, E\,x_t \equiv 0$, with spectral density $f(\lambda)$. Let $\log f(\lambda)$ be integrable.

Set

$$D(f; z) = \sqrt{2\pi} \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1 + z\,e^{i\theta}}{1 - z\,e^{i\theta}} \log f(\theta)\, d\theta \right\}.$$

Then $|D(f; z)| \neq 0$ when $|z| < 1$ and $D(f; 0) > 0$.

Now on setting

$$c(e^{-i\lambda}) = \lim_{r \to 1-} D(f; r\,e^{-i\lambda}) = D(f; e^{-i\lambda})$$

one finds that $\qquad\qquad f(\lambda) = |c(e^{-i\lambda})|^2$.

The prediction error variance for a one-step prediction is

$$2\pi \exp\left\{\frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \log f(\lambda)\, d\lambda\right\}.$$

Thus there is a neat and elegant way of obtaining $c(e^{-i\lambda})$ and the prediction error variance in the one-dimensional case.

Consider a two-vector weakly stationary process x_t, $t = \cdots, -1, 0, 1, \ldots, E\, x_t \equiv 0$, with spectral density

$$f(\lambda) = \begin{pmatrix} f_1(\lambda) & f_2(\lambda) \\ \overline{f_2(\lambda)} & f_1(\lambda) \end{pmatrix}$$

where $f_1(\lambda) \geq 0$, $f_1^2(\lambda) - |f_2(\lambda)|^2 = |f(\lambda)| \geq 0$, and $\log |f(\lambda)|$ is integrable. If we try to follow the procedure used in the one-dimensional case uniformly, we run into difficulties very soon as the logarithm of a matrix does not satisfy the equation

$$\log MN = \log M + \log N$$

unless $MN = NM$. We shall therefore first limit ourselves to those spectral densities $f(\lambda)$ which are such that $f(\lambda)f(\mu) = f(\mu)f(\lambda)$ for all λ, μ.

Given two matrices

$$A = \begin{pmatrix} a_1 & a_2 \\ \overline{a_2} & a_1 \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 & b_2 \\ \overline{b_2} & b_1 \end{pmatrix}$$

where $A, B \geq 0$, the matrices A, B commute if and only if $a_2 \overline{b_2} = \overline{a_2}\, b_2$ or $a_2 \overline{b_2}$ is real. Let us now add the additional condition that

$$f_1(\lambda) = f_1(-\lambda), \quad f_2(\lambda) = \overline{f_2(-\lambda)}$$

as this is satisfied by the spectral density of a process with real-valued components and we are particularly interested in such processes. Now $f_2(\lambda)\overline{f_2(\mu)}$ is real for all λ, μ if and only if either $f_2(\lambda)$ is real-valued for all λ or $f_2(\lambda)$ is pure imaginary for all λ. Thus the only cases in which commutativity holds are those where there is either zero cospectrum ($f_2(\lambda)$ pure imaginary) or zero quadrature spectrum ($f_2(\lambda)$ real).

First let us look at the case in which $f_2(\lambda)$ is real so that

$$f(\lambda) = \begin{pmatrix} f_1(\lambda) & f_2(\lambda) \\ f_2(\lambda) & f_1(\lambda) \end{pmatrix}$$

Then

$$U f(\lambda) U = \begin{pmatrix} f_1(\lambda) + f_2(\lambda) & 0 \\ 0 & f_1(\lambda) - f_2(\lambda) \end{pmatrix}$$

where

$$U = \begin{pmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\[2mm] \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \end{pmatrix}.$$

Now

$$U f(\lambda) U = \begin{pmatrix} D(f_1 + f_2; e^{-i\lambda}) & 0 \\ 0 & D(f_1 - f_2; e^{-i\lambda}) \end{pmatrix} \begin{pmatrix} D(f_1 + f_2; e^{i\lambda}) & 0 \\ 0 & D(f_1 - f_2; e^{i\lambda}) \end{pmatrix}.$$

Thus

$$c(e^{-i\lambda}) = U \begin{pmatrix} D(f_1 + f_2; e^{-i\lambda}) & 0 \\ 0 & D(f_1 - f_2; e^{-i\lambda}) \end{pmatrix} U$$

$$= \frac{1}{2} \begin{pmatrix} D(f_1 + f_2; e^{-i\lambda}) + D(f_1 - f_2; e^{-i\lambda}) & D(f_1 + f_2; e^{-i\lambda}) - D(f_1 - f_2; e^{-i\lambda}) \\ D(f_1 + f_2; e^{-i\lambda}) - D(f_1 - f_2; e^{-i\lambda}) & D(f_1 + f_2; e^{-i\lambda}) + D(f_1 - f_2; e^{-i\lambda}) \end{pmatrix}.$$

The covariance matrix of the error in a one-step prediction is

$$\pi \left\{ \begin{vmatrix} \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 + f_2)\, d\lambda\right\} & \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 + f_2)\, d\lambda\right\} \\[4mm] + \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 - f_2)\, d\lambda\right\} & - \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 - f_2)\, d\lambda\right\} \\[4mm] \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 + f_2)\, d\lambda\right\} & \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 + f_2)\, d\lambda\right\} \\[4mm] - \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 - f_2)\, d\lambda\right\} & + \exp\left\{\dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \log (f_1 - f_2)\, d\lambda\right\} \end{vmatrix} \right\}$$

The prediction error variance that would have been obtained for the first component if only the past of the first component of the process were used is

$$2\pi \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_1\, d\lambda\right\} \tag{8}$$

The improvement in prediction accuracy can therefore be represented by the ratio of the prediction variance using the past of both components to the prediction variance (8), that is,

420

$$\frac{1}{2}\left[\exp\left\{\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1+f_2)\,d\lambda\right\}+\exp\left\{\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1-f_2)\,d\lambda\right\}\right]\Bigg/$$

$$\exp\left\{\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ f_1\,d\lambda\right\}\leq 1.$$

Now let us look at the case in which there is no cospectrum. Then we can write

$$f(\lambda)=\begin{pmatrix} f_1(\lambda) & i\,f_2(\lambda) \\ -i\,f_2(\lambda) & f_1(\lambda) \end{pmatrix}$$

where $f_2(\lambda)$ is real and $f_2(\lambda)=-f_2(-\lambda)$. Now

$$U f(\lambda)\,U'=\begin{pmatrix} f_1(\lambda)+f_2(\lambda) & 0 \\ 0 & f_1(\lambda)-f_2(\lambda) \end{pmatrix}$$

where

$$U=\begin{pmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{i}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} & \dfrac{-i}{\sqrt{2}} \end{pmatrix}.$$

Thus

$$c\,(e^{-i\lambda})$$

$$=U'\begin{pmatrix} D\,(f_1+f_2;\,e^{-i\lambda}) & 0 \\ 0 & D\,(f_1-f_2;\,e^{-i\lambda}) \end{pmatrix}U$$

$$=\frac{1}{2}\begin{pmatrix} D\,(f_1+f_2;\,e^{-i\lambda})+D\,(f_1-f_2;\,e^{-i\lambda}) & i\,D\,(f_1+f_2;\,e^{-i\lambda})-i\,D\,(f_1-f_2;\,e^{-i\lambda}) \\ -i\,D\,(f_1+f_2;\,e^{-i\lambda})+i\,D\,(f_1-f_2;\,e^{-i\lambda}) & D\,(f_1+f_2;\,e^{-i\lambda})+D\,(f_1-f_2;\,e^{-i\lambda}) \end{pmatrix}.$$

Note that if we expand $c\,(e^{-i\lambda})$ as a series in $e^{-i\lambda}$, all the coefficients are matrices with real elements since $f_2(\lambda)=-f_2(-\lambda)$. The covariance matrix of a one-step prediction is

$$\pi\left\{\begin{vmatrix} \exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1+f_2)\,d\lambda\right\} & i\,\exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1+f_2)\,d\lambda\right\} \\[2ex] +\;\exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1-f_2)\,d\lambda\right\} & -i\,\exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1-f_2)\,d\lambda\right\} \\[2ex] -i\,\exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1+f_2)\,d\lambda\right\} & \exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1+f_2)\,d\lambda\right\} \\[2ex] +i\,\exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1-f_2)\,d\lambda\right\} & +\;\exp\left\{\dfrac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log\ (f_1-f_2)\,d\lambda\right\} \end{vmatrix}\right\}$$

$$= 2\pi \left\{ \begin{array}{cc} \exp\left\{\dfrac{1}{2\pi} \int\limits_0^\pi \log (f_1^2 - f_2^2)\, d\lambda\right\} & 0 \\[2em] 0 & \exp\left\{\dfrac{1}{2\pi} \int\limits_0^\pi \log (f_1^2 - f_2^2)\, d\lambda\right\} \end{array} \right\}.$$

The cases discussed thus far are the commutative cases in which precise analogues of one-dimensional techniques can be used and where, for example, the covariance matrix for the one-step prediction error is given by

$$2\pi \exp\left\{\dfrac{1}{2\pi} \int\limits_{-\pi}^\pi \log f(\lambda)\, d\lambda\right\}. \tag{9}$$

The cases of greatest interest are those in which one has non-commutativity of $f(\lambda)$ with $f(\mu)$, $\lambda \neq \mu$, that is, when both the cospectrum and quadrature spectrum are not zero. We consider a simple example of this type to show that one can no longer expect results like (9) to hold generally in this context.

Consider a matrix-valued (2×2) spectral density $f(\lambda) > 0$ which can be factored

$$a(\lambda)\, a(\lambda)' = \begin{pmatrix} a_{11}(\lambda) & a_{12}(\lambda) \\ a_{21}(\lambda) & a_{22}(\lambda) \end{pmatrix} \begin{pmatrix} \overline{a_{11}(\lambda)} & \overline{a_{21}(\lambda)} \\ \overline{a_{12}(\lambda)} & \overline{a_{22}(\lambda)} \end{pmatrix}$$

$$= f(\lambda) = \begin{pmatrix} f_1(\lambda) & f_2(\lambda) \\ \overline{f_2(\lambda)} & f_1(\lambda) \end{pmatrix},$$

$a(\lambda) = \sum\limits_{j=0}^\infty a_j e^{-ij\lambda}$, where $a_{12}(\lambda) \equiv 0$. We shall also assume that $\sum\limits_{j=0}^\infty a_j z^j = a(z)$ is non-singular in the unit circle. Then

$$|a_{11}(\lambda)|^2 = f_1(\lambda),$$

$$a_{11}(\lambda)\, \overline{a_{21}(\lambda)} = f_2(\lambda),$$

$$|a_{21}(\lambda)|^2 + |a_{22}(\lambda)|^2 = f_1(\lambda).$$

Then we can take

$$a_{11}(\lambda) = D(f_1(\lambda); e^{-i\lambda})$$

$$\overline{a_{21}(\lambda)} = f_2(\lambda) / D(f_1(\lambda); e^{-i\lambda}).$$

Now

$$|a_{22}(\lambda)|^2 = f_1(\lambda) - |f_2(\lambda)|^2 / f_1(\lambda) = \frac{|f(\lambda)|}{f_1(\lambda)}$$

so that

$$a_{22}(\lambda) = D\left(\frac{|f(\lambda)|}{f_1(\lambda)}; e^{-i\lambda}\right).$$

We can now see that a necessary and sufficient condition that $f(\lambda)$ have such a factorization is that $f_2(\lambda)/f_1(\lambda)$ have the one-sided expansion

422

$$f_2(\lambda)/f_1(\lambda) = \sum_{k=0}^{\infty} \alpha_k e^{-ik\lambda}.$$

The prediction error covariance matrix for a one-step prediction is

$$2\pi \left\{ \begin{matrix} \exp\left\{\dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \log f_1 \, d\lambda\right\} & 0 \\[2em] \exp\left\{\dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \log \dfrac{|f_2|^2}{f_1} \, d\lambda\right\} & \exp\left\{\dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \log \dfrac{f(\lambda)}{f_1(\lambda)} \, d\lambda\right\} \end{matrix} \right\} \times$$

$$\times \left\{ \begin{matrix} \exp\left\{\dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \log f_1 \, d\lambda\right\} & \exp\left\{\dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \log \dfrac{|f_2|^2}{f_1} \, d\lambda\right\} \\[2em] 0 & \exp\left\{\dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \log \dfrac{|f(\lambda)|}{f_1(\lambda)} \, d\lambda\right\} \end{matrix} \right\}$$

$$= 2\pi \left\{ \begin{matrix} \exp\left\{\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \log f_1 \, d\lambda\right\} & \exp\left\{\dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \log |f_2|^2 \, d\lambda\right\} \\[2em] \exp\left\{\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \log |f_2|^2 \, d\lambda\right\} & \exp\left\{\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \log \dfrac{|f_2|^2}{f_1} \, d\lambda\right\} + \\[2em] & + \exp\left\{\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \log \dfrac{|f|}{f_1} \, d\lambda\right\} \end{matrix} \right\}.$$

Note that in the case of such a spectral density (which is not generally commutative), the prediction error covariance is not of the same form as (9). Various other special non-commutative cases can be treated similarly.

8. Some final remarks

Let us discuss the case in which there is no quadrature spectrum in the context of Pierson's model of storm-generated ocean waves. Assume that the sea surface is observed at two points separated by a distance Δ as in section 3 and that the quadrature spectrum of the observed two-vector process is zero. What kind of spectrum $f(\lambda, \theta)$ of the sea surface is it that would lead to such a situation? The quadrature spectrum of the observed process is

$$\int_{-\pi}^{\pi} \sin\left(\frac{\lambda^2 \Delta}{g} \cos \theta\right) f(\lambda, \theta) \, d\theta \equiv 0. \tag{10}$$

An interesting sufficient condition for equation (10) to hold is given by the restraint

$$f(\lambda, \theta) + f(\lambda, -\theta) = f(\lambda, \pi - \theta) + f(\lambda, \theta - \pi) \tag{11}$$

on the sea spectrum $f(\lambda, \theta)$. It should be noted, however, that this is not a necessary condition. It would be interesting to choose interesting spectra satisfying condition (11) and determine the optimal distance Δ between the points of observation on the sea surface so as to minimize the prediction error.

REFERENCES

1. DOOB, J. L., Stochastic Processes. John Wiley (1953).
2. HALMOS, P., Introduction to Hilbert Space and the Theory of Spectral Multiplicity. Chelsea (1951).
3. KOLMOGOROV, A., Sur l'interpolation et extrapolation des suites stationnaires. C. R. Acad. Sci. Paris *208*, 2043–2045 (1939).
4. SZEGÖ, G., Orthogonal Polynomials, vol. 23. Amer. Math. Soc. Collog. Pub. (1939).
5. WHITTLE, P., The analysis of multiple stationary time series. Journ. Roy. Stat. Soc., Ser. B, *15*, 125 (1953).
6. WIENER, N., Extrapolation, Interpolation and Smoothing of Stationary Time Series. Cambridge–New York (1949).
7. ——, On the factorization of matrices. Comment. Math. Helvet. 29, 97–111 (1955).