MATH. SCAND. 5 (1957), 261–266

ON AN EXPLICIT FORMULA IN LINEAR LEAST SQUARES PREDICTION

EDWIN J. AKUTOWICZ

1. Introduction. The theory of linear least squares prediction of a stationary time series, as it has been given by Kolmogoroff, Wiener and others, must be regarded, at least in the one-dimensional case, as essentially complete both in its general scheme and in several more or less specific cases such as, for example, those in which the spectral density is rational. But in spite of the active development of this subject there seem to be one or two open questions which pertain to the general theory and which to this date apparently have not received a satisfactory solution in terms of necessary and sufficient conditions.

The objective of this note is to establish the validity of a linear "predictor" which expresses the prediction of a stationary random sequence into the future with maximum directness in terms of the data of the past, assuming that two numerical series, (5) and either (9) or (10) (v. inf.), converge. These are conditions bearing only upon the spectrum of the original random sequence; however, there still remains a gap between the necessary and the sufficient spectral conditions for the validity of the explicit predictor. Theorem 1 contains an expression for the innovation vector of the random sequence in terms of the past. Reference is made to Chapter XII of Doob's treatise [1] and to a forthcoming book [2] for details about facts that are only briefly recalled here.

Let us suppose that a probability space Ω with points ω and probability element $d\omega$ is given, and that the sequence of random variables with finite second moment,

$$\ldots, \boldsymbol{x}_{-1}(\omega),\ \boldsymbol{x}_0(\omega),\ \boldsymbol{x}_1(\omega),\ \ldots ,$$

constitutes a wide sense stationary random sequence in Ω, cf. [1, Ch. X].

Let $L^2(\Omega, d\omega)$ denote the Hilbert space realized by complex-valued random variables on Ω with finite second moment, the inner product $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ being given by

$$(\boldsymbol{x}_1, \boldsymbol{x}_2) = \int_\Omega \boldsymbol{x}_1(\omega)\overline{\boldsymbol{x}_2(\omega)}\,d\omega .$$

Received October 10, 1957.

In this Hilbert space let $\mathfrak{M}(x_{-n}, x_{-n-1}, \ldots)$ denote the closed linear subspace spanned by $x_{-n}, x_{-n-1}, \ldots$. We shall assume that

$$\bigcap_{n>0} \mathfrak{M}(x_{-n}, x_{-n-1}, \ldots) = (0)\,,$$

which implies that only the absolutely continuous component occurs in the spectrum of $\{x_n\}_{n=0,\pm1,\ldots}$.

The best linear least squares prediction of x_n, where n is a fixed non-negative integer, in terms of the complete past $\{x_{-\nu}\}$, $\nu \geqq 1$, is by definition that random variable $x_n{}^*$ lying in the closed subspace $\mathfrak{M}(x_{-1}, x_{-2}, \ldots)$ for which

$$\int_{\Omega} |x_n(\omega) - x_n{}^*(\omega)|^2 d\omega = \text{minimum}\,.$$

Therefore the task of finding $x_n{}^*$ is exactly that of determining the projection of x_n into $\mathfrak{M}(x_{-1}, x_{-2}, \ldots)$.

The first step in the solution of this problem is to express x_0 in the canonical form due to Wold,

$$x_0(\omega) = \sum_{\nu=0}^{\infty} b_\nu y_{-\nu}(\omega)\,, \tag{1}$$

where the infinite series is convergent in the mean ($d\omega$ integration), and the b_ν are the Fourier coefficients of x_0 with respect to the orthonormal sequence of innovations $\{y_{-\nu}\}$; that is $b_\nu = (x_0, y_{-\nu})$, $\nu \geqq 0$. Here y_0 is the unit vector in $L^2(\Omega, d\omega)$ defined by

$$\sigma \cdot y_0 = x_0 - Px_0\,,$$

where P denotes projection into $\mathfrak{M}(x_{-1}, x_{-2}, \ldots)$, and σ is a positive real number. In terms of these quantities the prediction $x_n{}^*$ is given as

$$x_n{}^* = \lim_{N\to\infty} \sum_{\nu=n+1}^{N} b_\nu y_{n-\nu}\,, \tag{2}$$

where

$$\sum_{\nu=0}^{\infty} |b_\nu|^2 < \infty\,. \tag{3}$$

It is well known that the numerical coefficents b_ν are uniquely determined by the covariance sequence of $\{x_n\}$ and are such that the holomorphic function

$$\Phi(z) = \sum_{\nu=0}^{\infty} b_\nu z^\nu, \qquad |z| < 1\,, \tag{4}$$

does not vanish. Also the boundary function $\Phi(e^{i\theta})$ of $\Phi(z)$ satisfies a.e. the equation $|\Phi(e^{i\theta})|^2 = F'(\theta)$, where $F'(\theta)$ is the spectral density of

$\{x_n\}$. However, it will be observed that the prediction (2) is by no means given explicitly in terms of the data of the past, $\{x_{-\nu}(\omega)\}$, $\nu \geqq 1$, since the vectors $y_{n-\nu}$ occurring are defined in terms of the original time series through the projection P. We shall now show how this situation can be remedied.

2. An expression for the innovation vector. An obvious property of the function $\Phi(z)$ in (4) is that it belongs to the Hardy class H^2. (For the definition and basic properties of the class H^2, see [3, Ch. VII]).

Let it be assumed that

$$(\Phi(z))^{-1} = \sum_0^\infty a_\mu z^\mu$$

also belongs to H^2, so that

$$\sum_0^\infty |a_\mu|^2 < \infty , \tag{5}$$

and that the spectral density is bounded:

$$F'(\theta) \leqq C . \tag{6}$$

Then the innovation vector y_0 can be written in the form

$$y_0 = \lim_{M\to\infty} \sum_0^M a_\mu x_{-\mu} . \tag{7}$$

To prove (7) observe that for almost all θ (Lebesgue measure), under our hypotheses,

$$\begin{aligned} 1 &= \lim_{r\to 1-} \Phi(re^{i\theta}) \cdot \lim_{r\to 1-} \left(\Phi(re^{i\theta})\right)^{-1} \\ &= \underset{N\to\infty}{\text{l.i.m.}} \sum_0^N b_\nu e^{i\nu\theta} \cdot \underset{M\to\infty}{\text{l.i.m.}} \sum_0^M a_\mu e^{i\mu\theta} \\ &= \underset{M\to\infty}{\text{l.i.m.}} \left(\sum_0^M a_\mu \cdot \underset{N\to\infty}{\text{l.i.m.}} \sum_0^N b_\nu e^{i(\mu+\nu)\theta} \right), \end{aligned}$$

where l.i.m. refers to quadratic mean with respect to Lebesgue measure. Under the boundedness condition (6) mean convergence with respect to Lebesgue measure implies mean convergence with respect to spectral measure. Since log $F'(\theta)$ is Lebesgue integrable, the null sets of Lebesgue and of spectral measure are the same. Hence, by a familiar isometric correspondence,

$$y_0 = \lim_{M\to\infty} \left(\sum_0^M a_\mu \cdot \lim_{N\to\infty} \sum_0^N b_\nu y_{-\nu-\mu} \right) = \lim_{M\to\infty} \sum_0^M a_\mu x_{-\mu} ,$$

where we have used (1).

Conversely, suppose (6) and (7) are valid for some numerical sequence a_μ satisfying (5). Then, using (1),

$$y_0 = \lim_{M\to\infty} \left(\sum_0^M a_\mu \cdot \lim_{N\to\infty} \sum_0^N b_\nu y_{-\nu-\mu} \right).$$

This implies

$$1 = \underset{M\to\infty}{\text{l.i.m.}} \sum_0^M a_\mu e^{i\mu\theta} \cdot \underset{N\to\infty}{\text{l.i.m.}} \sum_0^N b_\nu e^{i\nu\theta} ,$$

where l.i.m. refers to spectral measure. However l.i.m. with respect to Lebesgue measure also exists in view of (3) and (5). But if

$$\lim_{N\to\infty} \int_0^{2\pi} \left| \sum_0^N b_\nu e^{i\nu\theta} - \Phi(e^{i\theta}) \right|^2 d\theta = 0$$

and

$$\lim_{N\to\infty} \int_0^{2\pi} \left| \sum_0^N b_\nu e^{ir\theta} - \Psi(e^{i\theta}) \right|^2 F'(\theta)\, d\theta = 0 ,$$

it follows from (6) that $\Phi = \Psi$ a.e. with respect to spectral measure. Therefore $\Phi = \Psi$ a.e. with respect to Lebesgue measure. Therefore, for almost all θ (Lebesgue measure),

$$1 = \Phi_0(e^{i\theta}) \cdot \Phi(e^{i\theta}) ,$$

where $\Phi(e^{i\theta})$ is the boundary function of $(\Phi(z))^{-1}$ and $\Phi(e^{i\theta})$ that of $\Phi(z)$. As $\Phi_0 \Phi$ belongs to L^1 it is determined by its Fourier coefficients,

$$\sum_{\mu+\nu=k} a_\mu b_\nu, \qquad k = 0, 1, 2, \ldots .$$

Hence $\Phi_0 \Phi = 1$ a.e. implies

$$\text{(8)} \qquad \sum_{\mu+\nu=k} a_\mu b_\nu = 0, \qquad k = 1, 2, \ldots , \qquad a_0 b_0 = 1 .$$

Therefore

$$\sum_{\mu=0}^{\infty} a_\mu z^\mu = (\Phi(z))^{-1}, \qquad |z| < 1 ,$$

which identifies the a_μ as the Taylor coefficients of $(\Phi(z))^{-1}$ and proves:

THEOREM 1. *Under the condition of boundedness of the spectral density, the innovation vector y_0 can be expressed in the form*

$$y_0 = \lim_{M\to\infty} \sum_0^M a_\mu x_{-\mu} , \qquad \sum_0^\infty |a_\mu|^2 < \infty ,$$

if and only if the function

$$\Phi(z) = \sum_{0}^{\infty} b_\nu z^\nu, \qquad |z| < 1, \qquad b_\nu = (x_0, y_{-\nu}),$$

is such that its reciprocal belongs to H^2, *and then the* a_μ *are the Taylor coefficients of the reciprocal of* Φ.

3. An explicit predictor. We shall now assume concerning the coefficient sequences $\{b_\nu\}_{\nu=0,1,\dots}$ and $\{a_\mu\}_{\mu=0,1,\dots}$ occurring in the power series expansions about $z=0$ of $\Phi(z)$ and $(\Phi(z))^{-1}$ respectively, that in addition to condition (5) we also have either

$$\sum_{\mu=0}^{\infty} |a_\mu| < \infty, \tag{9}$$

or

$$\sum_{\nu=0}^{\infty} |b_\nu| < \infty. \tag{10}$$

Since Φ determines and is determined by the spectral density of $\{x_n\}$, these assumptions are conditions on the spectrum of the original random sequence. They imply the following statement:

THEOREM 2. *The optimal linear least squares prediction* $x_n{}^*$ *of a stationary time series* $\{x_m\}_{m=0,\pm1,\dots}$ *is given by*

$$x_n{}^* = \lim_{J\to\infty} \sum_{j=1}^{J} \left[\sum_{s=0}^{j} b_{n+s} a_{j-s} \right] x_{-j}. \tag{11}$$

PROOF. By (1),

$$\sum_{j=1}^{J} \left[\sum_{s=0}^{j} b_{n+s} a_{j-s} \right] x_{-j} = \sum_{j=1}^{J} \left[\sum_{s=0}^{j} b_{n+s} a_{j-s} \right] \lim_{N\to\infty} \sum_{\nu=0}^{N} b_\nu y_{-j-\nu}$$
$$= \lim_{N\to\infty} \sum_{j=1}^{J} \sum_{s=0}^{j} \sum_{\nu=0}^{N} b_{n+s} a_{j-s} b_\nu y_{-j-\nu}.$$

By changing the order of summation this finite triple sum can be written in the form:

$$\sum_{r=1}^{N+J} \left[\sum_{s=0}^{(r,J)^-} b_{n+s} \sum_{t=((0,r-N)^+,s)^+}^{(r,J)^-} b_{r-t} a_{t-s} \right] y_{-r}$$
$$= \sum_{r=1}^{(N,J)^-} \left[\sum_{s=0}^{r} b_{n+s} \sum_{t=0}^{r-s} b_{r-s-t} a_t \right] y_{-r} + \sum_{r=(N,J)^-+1}^{N+J} \left[\sum_{t=(0,r-N)^+}^{(r,J)^-} b_{r-t} \sum_{s=0}^{t} b_{n+s} a_{t-s} \right] y_{-r},$$

where $(p,q)^- = \min(p,q)$, $(p,q)^+ = \max(p,q)$. Using the equations (8) in the first term we obtain, in view of (2),

$$\lim_{J\to\infty} \left(\lim_{N\to\infty} \sum_{r=1}^{(N,J)^-} \left[\sum_{s=0}^{r} b_{n+s} \sum_{t=0}^{r-s} b_{r-s-t} a_t \right] y_{-r} \right) = \lim_{J\to\infty} \sum_{r=1}^{J} b_{n+r} y_{-r} = x_n{}^*.$$

So it remains to show that applying $\lim_{N\to\infty}$ followed by $\lim_{J\to\infty}$ to the second term yields zero. Using equations (8) in the form

$$\sum_{s=0}^{t} b_{n+s}a_{t-s} = -\sum_{q=1}^{n} b_{n-q}a_{q+t},$$

this amounts to showing that for $q=1, \ldots, n$ the quantity

$$\sum_{r=(N,J)^-+1}^{N+J}\left[\sum_{t=(0,r-N)^+}^{(r,J)^-} b_{r-t}a_{q+t}\right]y_{-r} \tag{12}$$

tends to zero in $L^2(\Omega, dw)$ as $N\to\infty$ and then $J\to\infty$. Thus it is permissible to assume $N>J$ in the rest of the argument, so that $(N, J)^- = J$ and $(r, J)^- = J$ in (12). Using (8) once more we write (12) in the form

$$-\sum_{r=N}^{N+J}\left[\sum_{t=0}^{J} b_{r-t}a_{q+t}\right]y_{-r} - \sum_{t=J+1}^{N+J} a_{q+t}\sum_{r=t}^{N+J} b_{r-t}y_{-r}, \tag{13}$$

and here the first sum clearly tends to zero for every fixed J as $N\to\infty$, while the second is a linear combination of vectors

$$v_{t,N,J} = \sum_{r=t}^{N+J} b_{r-t}y_{-r}$$

of length

$$\|v_{t,N,J}\| \leqq \left(\sum_{\nu=0}^{\infty}|b_\nu|^2\right)^{\frac{1}{2}} = B.$$

Therefore, by Minkowski's inequality,

$$\left\|\sum_{t=J+1}^{N+J} a_{q+t}v_{t,N,J}\right\| \leqq B\cdot\sum_{t=J+1}^{N+J}|a_{q+t}|,$$

a quantity tending to zero as $N\to\infty$ and then $J\to\infty$, if (9) holds. Therefore (11) is proved in this case. (11) will also follow if we reverse the order of summation in the second sum in (13) and use (10) instead of (9).

I wish to take this opportunity to state that I have incurred a considerable scientific debt to Professor Norbert Wiener through numerous discussions and through the probability seminar conducted by him.

REFERENCES

1. J. L. Doob, *Stochastic processes*, New York, 1953.
2. E. J. Akutowicz, W. T. Martin, B. Rankin, A. Siegel, and N. Wiener, *A book on applications of certain stochastic processes*. To appear.
3. A. Zygmund, *Trigonometrical series*, Warszawa–Lwow, 1935.
4. P. Masani and N. Wiener, *The prediction theory of multivariate stochastic processes*, Acta Math. 98 (1957).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASS., U. S. A.