Z. Wahrscheinlichkeitstheorie verw. Geb. 12, 255–266 (1969)

A Factorization Problem and the Problem of Predicting Non-Stationary Vector-Valued Stochastic Processes

J. RISSANEN and L. BARBOSA

Summary. In this paper a prediction theory is developed under the general idea that the infinite dimensional covariance matrix is a self-adjoint element in a symmetric Banach algebra. The usual Wiener's spectral factorization method for solving stationary Wiener-Hopf equations has been extended to this algebra. Finally, a theorem for factoring a positive definite covariance matrix into upper and lower triangular factors with similar inverses has been proved.

1. Introduction

The prediction theory of stationary processes, first scalar-valued and then more recently vector-valued ones, has reached certain degree of completeness, [1–7]. Behind this success has been the rich and well established machinery of harmonic analysis.

The central problem turned out to be a certain factorization of the spectral density function into two factors, one consisting of positive and the other of negative powers of $e^{i\theta}$. This problem was solved first for scalar valued functions by Szego, [8], and then successively for finite and infinite matrix valued functions by a series of authors, [5, 9–11]. The factorization permits a one-sided moving average representation of the process with a simple solution to the prediction problem as the result.

In this paper we show that an analogous approach can be developed for solving the prediction problem of both stationary and non-stationary processes alike. We consider the infinite-dimensional covariance matrix associated with a process as a self-adjoint element of $B(\mathscr{H})$, the symmetric Banach algebra of bounded linear operators on an appropriate Hilbert space into itself. Then, we prove by construction a factorization of self-adjoint operators into upper and lower triangular factors, after which the solution to the prediction problem is obtained without difficulties.

The class of processes with covariance matrices belonging to $B(\mathscr{H})$ has special importance in engineering applications. In this class, the predictor, defined by an infinite dimensional matrix, itself belongs to $B(\mathscr{H})$. Therefore, it can be compound with other elements of $B(\mathscr{H})$ by the product rule. This in applications corresponds to the cascade connection of linear systems with preservation of stability – to use an engineering language.

Because of the importance of the above type of processes, we in this paper investigate them only. Yet, the problem of how to remove the requirement that the covariance function be a bounded operator is an important one. It is conceivable that the powerful factorization results in [9–11] can be applied to perform even the type of factorization needed here, which would permit a harmonic analysis treatment of non-stationary processes.

For other approaches to prediction problems of non-stationary processes, we refer to [12] and [13]; special cases have also been discussed in [14].

2. Preliminaries

To make this paper more self-contained, we give a brief summary of vector-valued stochastic processes. A more thorough treatment can be found in [5].

Let Ω be a space having a Borel field of subsets over which is defined a probability measure P. Let L_2 be the set of all real valued measurable functions f on Ω such that $\int_\Omega f^2(\omega)\,dP(\omega)<\infty$. Then L_2 is a Hilbert space with the inner product

$$(f,g)=\int_\Omega f(\omega)\,g(\omega)\,dP(\omega)=E(f\,g). \tag{2.1}$$

As usual, we do not distinguish functions which differ only on a set of measure zero.

Let $\mathscr{L}_2$ denote the set of all q-dimensional vector-valued functions $f=(f^{(1)},\ldots,f^{(q)})$ on Ω such that $f^{(i)}\in L_2$. $\mathscr{L}_2$ is made into a Hilbert space by setting

$$(f,g)=\sum_{i=1}^{q} E(f^{(i)}\,g^{(i)}). \tag{2.2}$$

The norm induced by (2.2) is given by:

$$\|f\|=(f,f)^{\frac{1}{2}}. \tag{2.3}$$

Definition. By a vector (q-)valued random process f is meant a sequence $\{f_n\}_{-\infty}^{\infty}$ of functions $f_n\in\mathscr{L}_2$. (2.4)

The covariance matrix of f_n and f_m is written as

$$r_{nm}=E\,f_n\,f_m'. \tag{2.5}$$

We call the infinite matrix $r=\{r_{nm}\}$ the covariance function of f.

If two processes f and g have all their covariance matrices identical, the processes are equivalent up to second order moments, [15].

Let M_n denote the subspace spanned (with respect to all possible real $q\times q$-matrix coefficients) by the set $\{f_j\}_{j\leqq n}$, $j\in N=$ set of integers, $f_j\in\mathscr{L}_2$. Then the orthogonal projection of f_{n+m} on M_n, denoted (f_{n+m}/M_n), gives the optimal l.s. prediction of f_{n+m} given $\{f_j\}_{j\leqq n}$, [5].

Following Wiener, [5], we call the process $\{f_n\}_{-\infty}^{\infty}$

(a) *non-deterministic*, if for all n, $f_{n+1}\notin M_n$

(b) *regular*, if $(f_n/M_{-m})\to 0$, as $m\to\infty$ (2.6)

(c) *full rank*, if for all n

$$k_1\,I<E\,g_n\,g_n'<k_2\,I,$$

where k_1 and k_2 are positive real numbers, and

$$g_n=f_n-(f_n/M_{n-1}). \tag{2.7}$$

The process $\{g_n\}$ is called the innovation process associated with f. It is clear that a full rank process is non-deterministic.

If $S(g_{m+1}, g_{m+2}, \dots, g_n)$ denotes the closed linear manifold spanned by the orthogonal vectors $g_{m+1}, \dots, g_n$, the following holds:

$$M_n = M_m + S(g_{m+1}, \dots, g_n); \qquad S(g_{m+1}, \dots, g_n) \perp M_m. \tag{2.8}$$

This and regularity imply that f_n can be represented as a one-sided moving average:

$$f_n = \sum_{i \leqq n} b_{ni}\, g_i, \tag{2.9}$$

where b_{ni} are $q \times q$-matrices such that

$$\sum_{i \leqq n} \|b_{ni}\|^2 < \infty. \tag{2.10}$$

The notation $\|b_{ni}\|$ is used for the matrix norm:

$$\|b_{ni}\| = \big(t(b_{ni}\, b'_{ni})\big)^{\frac{1}{2}} \tag{2.11}$$

the prime indicating transpose of a matrix and $t(\ )$ the trace.

Clearly, for a full rank process,

$$E\, g_i\, g'_j = \delta_{ij}\, G_i; \qquad k_1\, I < G_i < k_2\, I \tag{2.12}$$

where $\delta_{ij} = 1$ for $i = j$ and $= 0$ for $i \neq j$.

Conversely, we may consider Eq. (2.9) for a given process $g = \{g_i\}$ in (2.12) and a set of coefficients $\{b_{ni}\}$ satisfying (2.10) to define a class of random processes. Clearly, this class contains all regular, full rank processes; and this is why we may confine our study to one-sided moving average representation of processes.

Because of the condition in (2.12) the orthogonal process g may be normalized by setting

$$w_i = G_i^{-\frac{1}{2}}\, g_i,$$

whereupon we may rewrite the class of interest as:

$$\begin{aligned} & f_n = \sum_{i \leqq n} a_{ni}\, w_i, \qquad n \in N \\ & \sum_{i \leqq n} \|a_{ni}\|^2 < \infty \\ & E\, w_i\, w'_j = \delta_{ij}\, I. \end{aligned} \tag{2.13}$$

By direct computation the covariance function of the process (2.13) is seen to be given by:

$$r_{nm} = \sum_{i \leqq m} a_{ni}\, a'_{mi} \quad \text{for} \quad m \leqq n; \qquad r_{nm} = r'_{mn} \qquad \text{for all } m, n \in N. \tag{2.14}$$

Defining the infinite matrices $r=\{r_{nm}\}$, $a=\{a_{ij}\}$, and a^* by $(a^*)_{ij}=a'_{ji}$, we may write Eq. (2.14) as the single equation:

$$r=a\,a^*. \tag{2.15}$$

Notice that r has been factored into upper and lower triangular factors. The converse problem: given r, find the factorization (2.15), is of central interest; the main part of the paper is concerned with that problem.

The basic prediction problem is to determine for each n the matrix coefficients p_{ni} such that the infinite series $\sum_{i\leqq n} p_{ni} f_i$ converges and satisfies the equality:

$$(f_{n+m}/M_n)=\sum_{i\leqq n} p_{ni} f_i, \qquad n\in N. \tag{2.16}$$

But in many applications this is not enough. Often, the prediction itself is not the main goal; and the data obtained from the predictor, considered as a linear system defined by the p_{ni}'s must be processed by other linear systems for regulation purposes, say. This brings up the question whether the resulting process is well defined, or as often called, stable.

It is clear that to ensure this additional property on the predictor, the class of processes (2.13) must be appropriately reduced, which is done in the next sections.

3. A Banach Algebra and a Class of Stochastic Processes

We begin by defining the real Hilbert space $\mathscr{H}$ as the direct sum of countable number of the real Euclidean q-dimensional spaces R^q: $\mathscr{H}=\sum_{i\in N}\oplus H_i$, $H_i=R^q$ for all $i\in N$. The elements of $\mathscr{H}$ are thus of the form

$$x=\{x_i\}_{-\infty}^{\infty}, \qquad x_i=(x_i^{(1)},\ldots,x_i^{(q)})\in R^q,$$

having finite norm:

$$\|x\|=\Big(\sum_{i\in N}\|x_i\|^2\Big)^{\frac{1}{2}}<\infty, \tag{3.1}$$

where $\|x_i\|$ is the usual Euclidean norm $\left(\sum_{j=1}^{q}(x_i^{(j)})^2\right)^{\frac{1}{2}}$.

Consider now the symmetric Banach algebra $B(\mathscr{H})$ of all bounded linear operators on $\mathscr{H}$ into itself. Each operator in $B=B(\mathscr{H})$, [16], can be described by an infinite dimensional matrix of the type:

$$a=\{a_{nm}\colon\ n,m\in N,\ a_{nm}\in B(\mathscr{H}_i)\}=\begin{pmatrix} & \ddots & & \\ \cdots\ a_{n,n+1} & a_{n,n} & a_{n,n-1}\ \cdots & \\ & \cdots\ a_{n-1,n} & a_{n-1,n-1} & a_{n-1,n-2}\ \cdots \\ & & \ddots & \end{pmatrix} \tag{3.2}$$

such that the norm

$$\|a\|=\operatorname*{Sup}_{\substack{x\in\mathscr{H}\\ \|x\|=1}}\|a\,x\|<\infty. \tag{3.3}$$

In (3.2) the a_{nm}'s can be considered as $q \times q$-real matrices. The product in B, defined as the composite map $a \circ b$, becomes a matrix multiplication:

$$(a\,b)_{nm} = \sum_{i \in N} a_{ni}\, b_{im}. \tag{3.4}$$

The element $e = \{e_{ij}\colon e_{ij} = \delta_{ij} I,\ I = q \times q\text{-identity matrix}\}$ is the identity element in B. The map $a \to a^*$, defined as $(a^*)_{nm} = a'_{mn}$, is called the involution; and the equalities

$$\|a\,a^*\| = \|a\|^2, \qquad \|a\| = \|a^*\| \tag{3.5}$$

follow from the definitions. Also, $(a^*)^{-1} = (a^{-1})^*$.

We already met the upper and lower triangular matrices in (2.15). Their role in the present discussion is central, and we define the following subspaces of B:

$$\begin{aligned} B^+ &= \{a\colon\ a \in B,\ a_{nm} = 0,\ m > n\} \\ B^- &= \{a\colon\ a \in B,\ a_{nm} = 0,\ m < n\}. \end{aligned} \tag{3.6}$$

That they also are subalgebras follows from the proposition:

Proposition. $a, b \in B^+(B^-) => a\,b \in B^+(B^-)$. (3.7)

Proof. $a, b \in B^+ =>$

$$(a\,b)_{nm} = \begin{cases} \sum\limits_{i=m}^{n} a_{ni}\, b_{im} & \text{for } m \leqq n \\ 0 & \text{for } m > n. \end{cases}$$

Similar proof holds for B^-.

We next define a product between an element $a \in B$ and the orthonormal process $w = \{w_i\}_{-\infty}^{\infty}$ as follows:

$$(a\,w)_n = \sum_{i \in N} a_{ni}\, w_i, \qquad n \in N. \tag{3.8}$$

Since $a \in B$ implies $\sum_{i \in N} \|a_{ni}\|^2 < \infty$, where $\|a_{ni}\|$ is defined in (2.11), Eq. (3.8) defines an element in $\mathscr{L}_2$ for each n by Riesz-Fischer theorem.

We conclude this section by defining a class of processes of the type:

$$C = \{f\colon f = a\,w,\ a \in B^+,\ w \text{ orthonormal}\}. \tag{3.9}$$

Contrary to the entire class of processes of the type (2.13), we may now consider products $g = b\,f$, $b \in B^+$, $f \in C$, to mean: $g = (b\,a)\,w$, where $f = a\,w$. Clearly $g \in C$.

4. Prediction in C

The prediction problem (2.16) for processes in class C is very simple. Let $f \in C$ be defined by an $a \in B^+$:

$$f = a\,w. \tag{4.1}$$

If $a^{-1} \in B^+$, then the equality $w = a^{-1} f$ shows that the two sets of elements in $\mathscr{L}_2$: $\{w_i\}_{-\infty}^n$ and $\{f_i\}_{-\infty}^n$ span the same subspace M_n. Further, we have from (4.1):

$$f_{n+m} = a_{n+m,\,n+m}\, w_{n+m} + \cdots + a_{n+m,\,n+1}\, w_{n+1} + \sum_{i \leqq n} a_{n+m,\,i}\, w_i, \tag{4.2}$$

which shows that the optimal predictor is given by:

$$(f_{n+m}/M_n) = \sum_{i \leqq n} a_{n+m,\,i}\, w_i, \qquad n \in N. \tag{4.3}$$

We define the shift operator U^m in B as:

$$(U^m x)_n = x_{n+m}, \qquad x \in \mathscr{H}, \tag{4.4}$$

and the truncation of an $a \in B$ as:

$$(a_+)_{ij} = \begin{cases} a_{ij} & \text{if } j \leqq i \\ 0 & \text{if } j > i. \end{cases} \tag{4.5}$$

With the notation $\hat{f} = \{\hat{f}_i\}_{-\infty}^{\infty}$, $\hat{f}_i = (f_{i+m}/M_i)$ we, then, can write Eq. (4.3) in the concise form:

$$\hat{f} = (U^m a)_+\, a^{-1} f, \tag{4.6}$$

where $p = (U^m a)_+\, a^{-1} \in B^+$ by Proposition (3.7).

5. Factorization of the Covariance Function

In the previous section we saw how neatly the prediction problem is solved in terms of the generating element $a \in B^+$, if we know that $a^{-1} \in B^+$, too. In this section we study the converse problem: given a covariance function $r \in B$, find the factorization (2.15). In addition, we are going to determine conditions on r to insure that $a^{-1} \in B^+$.

The fact that we are dealing with vector valued processes introduces certain trivial but annoying complications. To dispose of them a few notations and preliminary remarks are needed.

We begin by introducing one more Hilbert space. Let V be the space of all sequences $y = \{y_i : y_i$ real valued $q \times q$-matrices, $\sum_{i \in N} t(y_i\, y_i') < \infty\}$. The inner product in V is defined as:

$$(y, z) = \sum_{i \in N} t(y_i\, z_i'), \tag{5.1}$$

and the induced norm is denoted by $\|y\|$. Sometimes it is useful to consider y to consist of q columns or rows $y^{(j)}$, each belonging to $\mathscr{H}$.

Any element of $B(\mathscr{H})$ defines also a linear map $V \to V$ by:

$$(a\, y)_n = \sum_{i \in N} a_{ni}\, y_i, \qquad n \in N, \tag{5.2}$$

which clearly is bounded.

In addition to multiplication of elements of V by scalars, it is convenient to define a product of a $y \in V$ with a $q \times q$-matrix Q as:

$$(Q\,y)_n = Q\,y_n, \qquad n \in N. \tag{5.3}$$

We use the notation $r>0$ for an $r \in B$ to mean $(x, r\,x) > k > 0$ for all $x \in \mathscr{H}$ such that $\|x\| = 1$, and $r \geqq 0$ to mean that $(x, r\,x) \geqq 0$ for all $x \in \mathscr{H}$. In what follows we assume $r = r^*$. Further, if $r>0$, then $r^{-1} \in B$. But, also $r^{-1} > 0$. To see this, observe that $r>0$ implies first the existence of $r^{\frac{1}{2}}$ such that $r^{-\frac{1}{2}} \in B$. Thus $r^{-1} = r^{-\frac{1}{2}}\,r^{-\frac{1}{2}}$, which is enough to establish that $r^{-1} > 0$. We also make frequent use of the fact that for an $r>0$, the expression $(x, r\,x)^{\frac{1}{2}}$ defines a new norm in $\mathscr{H}$ which is equivalent to $(x, x)^{\frac{1}{2}}$.

Consider the $q \times q$-matrix defined as:

$$[y\,r\,z] = \sum_{i,j \in N} y_i\, r_{ij}\, z_j'; \qquad y,\ z \in V,\ r \in B,\ r>0. \tag{5.4}$$

This definition is meaningful, for the $m\,n$'th element of $[y\,r\,z]$ can be written as $(y^{(m)}, r\,z^{(n)})$, where $y^{(m)} \in \mathscr{H}$ denotes the m'th column of y and $z^{(n)}$ the n'th row of z.

In particular, it is easy to see that $[y\,r\,y]$ is positive definite if and only if $[y\,y] = [y\,e\,y]$ is; or, equivalently, if the columns $y^{(j)}$ of y, considered as elements of $\mathscr{H}$, are linearly independent. We use the usual ordering for positive definite $q \times q$-matrices, defined by: $Q > P$ if $Q - P$ is positive definite.

To conclude these preliminaries let $\mathscr{H}_n$ denote the subspace of $\mathscr{H}$ consisting of all the elements satisfying $x_i = 0$ for $i > n$, and let $a_{(n)}$ denote the operator obtained from $a \in B$ by the truncation:

$$(a_{(n)})_{ij} = \begin{cases} a_{ij} & \text{for } i \text{ and } j \leqq n \\ 0 & \text{for } i \text{ or } j > n. \end{cases} \tag{5.5}$$

We also write $B_n = B(\mathscr{H}_n)$, $B_n^+ = B^+ \cap B_n$, and similarly for B_n^-.

If $r>0$, then with $x \in \mathscr{H}_n$, $\|x\| = 1$, the relations

$$0 < k_1 < (x, r\,x) = (x, r_{(n)}\,x) < k_2 \tag{5.6}$$

show that $r_{(n)}: \mathscr{H}_n \to \mathscr{H}_n$ has an inverse $r_{(n)}^{-1}$ which is uniformly bounded in n.

We now proceed to the main theorem:

Theorem. *Let* $r \in B$, *and*

$$r = r^*, \qquad r > 0. \tag{5.7}$$

Then there exists an element $a \in B^+$ *with* $a^{-1} \in B^+$ *such that*

$$r = a\,a^*$$

The factorization is unique up to the equivalence under diagonal unitary transformations.

Proof. We begin by constructing an element $b = b_{(0)} \in B_0^-$ satisfying the equality:

$$b\, r_{(0)}^{-1}\, b^* = \mu, \tag{5.8}$$

where $\mu = \mu_{(0)} \in B_0^+ \cap B_0^-$; i.e. μ is a diagonal. Denote the i'th row of b by b^i:

$$b^i = (\dots 0, b_{i,0}, b_{i,-1}, \dots, b_{ii}, 0, \dots), \qquad i \leqq 0.$$

The rows b^i will be determined by an immediate adaptation of Schmidt's orthogonalization method. Letting $\{e^i\}_{-\infty}^0$ denote the orthogonal set in V: $(e^i)_j = \delta_{ij} I$, the rows b^i are generated by the recurrence relations:

$$\begin{aligned} b^0 &= e^0 \\ b^i &= e^i - \sum_{j=i+1}^{0} [e^i r_{(0)}^{-1} b^j]\,[b^j r_{(0)}^{-1} b^j]^{-1} b^j, \qquad i \leqq -1. \end{aligned} \tag{5.9}$$

Here, we used the definition (5.3) of the product between a $q \times q$-matrix and a vector in V. The inverse $[b^j r_{(0)}^{-1} b^j]^{-1}$ exists since by the construction the q columns of the element $b^i \in V$ are linearly independent in $\mathscr{H}_0$, and $r_{(0)}^{-1} > 0$.

Any two vectors b^i and b^j are seen to satisfy the orthogonality (with respect to $r_{(0)}^{-1}$) relation:

$$[b^i r_{(0)}^{-1} b^j] = \delta_{ij} \mu_{ii} \tag{5.10}$$

where

$$0 < k_1 I < \mu_{ii} < k_2 I, \qquad i \leqq 0. \tag{5.11}$$

Indeed, by writing $\alpha_{ij} = [e^i r_{(0)}^{-1} b^j]$, the equality

$$\mu_{ii} + \sum_{j=i+1}^{0} \alpha_{ij} \mu_{jj}^{-1} \alpha'_{ij} = [e^i r_{(0)}^{-1} e^i] \tag{5.12}$$

proves the upper bound for μ_{ii}. The lower bound is obtained by observing that $[b^i b^i] \geqq I$, which entails $[b^i r_{(0)}^{-1} b^i] > k_1 I$ for some $k_1 > 0$.

Setting $\mu = \{\mu_{ii}\}$ we obtain (5.8). That $b^* \in B^+$ follows from the second inequality below:

$$k_1 \|x\|^2 \leqq (x, \mu x) = (x, b\, r_{(0)}^{-1} b^* x) = (b^* x, r_{(0)}^{-1} b^* x) \leqq k_2 \|x\|^2, \qquad x \in \mathscr{H}_0 \tag{5.13}$$

and the equivalence of the norms $(x, x)^{\frac{1}{2}}$ and $(x, r_{(0)}^{-1} x)^{\frac{1}{2}}$.

We still have to show that $b_{(0)}^{*-1} \in B_0^+$. Let $\mathscr{H}_{0,m}$, $m \leqq 0$, denote the subspace of $\mathscr{H}_0$ consisting of the elements satisfying $x_i = 0$ for $i < m$. $\mathscr{H}_{0,m}$ is invariant for the operators a in B_0^+; denote the induced truncated operators $a_{(0,m)}$: $(a_{(0,m)})_{ij} = a_{ij}$ for $0 \geqq i, j \geqq m$, with all the other components zero.

The first inequality in (5.13) implies that for all $x \in \mathscr{H}_{0,m}$,

$$k'_1 \|x\| \leqq \|b^* x\| = \|b^*_{(0,m)} x\|, \qquad k'_1 > 0. \tag{5.14}$$

But this means that the inverse $b_{(0,m)}^{*-1}$: $\mathscr{H}_{0,m} \to \mathscr{H}_{0,m}$ is uniformly bounded in m. Moreover, the inverse is upper triangular. Since $\overline{\bigcup_{m \leqq 0} \mathscr{H}_{0,m}} = \mathscr{H}_0$, the induced operators $c_{(0,m)} \triangleq b_{(0,m)}^{*-1}$, satisfying the relation $(c_{(0,m-1)})_{(0,m)} = c_{(0,m)}$ define an operator $c_{(0)} \in B_0^+$ which by (3.4) is $b_{(0)}^{*-1}$.

Write now $\lambda_{(0)} = \mu^{-1} = \{\mu_{ii}^{-1}, i \leqq 0\}$. Eq. (5.8), rewritten as

$$c_{(0)} r_{(0)} c_{(0)}^* = \lambda_{(0)} \tag{5.15}$$

shows that the rows of $c_{(0)}$ form an $r_{(0)}$ orthogonal set in V.

This suggests a continuation of the Schmidt's orthogonalization scheme (with respect to r) for increasing indices. Let $c^j = (\dots 0, I, c_{j,j-1}, c_{j,j-2}, \dots) \in V$. For $j \leqq 0$, we let c^j coincide with the rows of $c_{(0)}$. For $j > 0$, we define

$$c^j = e^j - \sum_{i=-\infty}^{j-1} [e^j r c^i][c^i r c^i]^{-1} c^i, \qquad j > 0. \tag{5.16}$$

The sum on the right is bounded uniformly in j, since with the notations $\beta_{ij} = [e^j r c^i]$, $\lambda_{ii} = [c^i r c^i]$, the equality

$$\lambda_{jj} + \sum_{i \leqq j-1} \beta_{ij} \lambda_{ij}^{-1} \beta'_{ij} = [e^j r e^j], \qquad j > 0, \tag{5.17}$$

obtained from (5.16), holds.

Eq. (5.16) implies that

$$[c^k r c^l] = \delta_{kl} \lambda_{kk}, \tag{5.18}$$

where, as in (5.11),

$$0 < k_1'' I < \lambda_{ii} < k_2'' I, \qquad i > 0.$$

Just as above one can show that the operator $c_{(i)}$, defined by the rows $c^j, j \leqq i$, is uniformly bounded in i, and that the same is true about the inverse $c_{(i)}^{-1}$ with domain $\mathscr{H}_i$. Moreover, $c_{(i)} \in B_i^+$, $(c_{(i+1)})_{(i)} = c_{(i)}$, which together with the fact that $c_{(0)}^{-1} \in B_0^+$ entail that $c_{(i)}^{-1} \in B_i^+$, $i > 0$. Also, $(c_{(i+1)}^{-1})_{(i)} = c_{(i)}^{-1}$. By (5.18)

$$c_{(i)} r_{(i)} c_{(i)}^* = \lambda_{(i)} \tag{5.19}$$

where $\lambda_{(i)} = \{\lambda_{ii}\colon \lambda_{ii} = \mu_{ii}^{-1}, i \leqq 0, \lambda_{ii} = [c^i r c^i], i > 0\}$.

The operators $c_{(i)}$, $c_{(i)}^{-1}$, and $\lambda_{(i)}$, $i \in N$, define the three operators $c, c^{-1} \in B^+$, and $\lambda \in B^+ \cap B^-$, respectively, such that

$$c r c^* = \lambda \tag{5.20}$$

as can be verified by (3.4).

Setting $a = c^{-1} \lambda^{\frac{1}{2}}$ we obtain the desired factorization.

To complete the proof, let $b \in B^+$ be another operator factoring r:

$$r = b b^*. \tag{5.21}$$

Then there exists a unitary transformation u such that $b = a u$. Indeed, $u \triangleq a^{-1} b \in B^+$, and $u u^* = a^{-1} b b^* a^{*-1} = e = u^* u$. For $x, y \in \mathscr{H}_{0,m}$ $(m \leqq 0)$

$$(x, y) = (x, u^* u y) = (u x, u y) = (u_{(0,m)} x, u_{(0,m)} y) = (u_{(0,m)}^* u_{(0,m)} x, y);$$

which shows that $u_{(0,m)}$ is unitary. But $u_{(0,m)}^* = u_{(0,m)}^{-1}$ is upper triangular, for $u_{(0,m)}$ is. Thus, $u_{(0,m)}$ is diagonal. But this implies that u is diagonal, too.

Remark 1. A strict application of the method to construct the factors in Theorem (5.7) is not very practical, because the inverse $r_{(0)}^{-1}$ is required. However, in the important special case of practice where the process is of interest only for $i \geqq 0$; i.e. the process begins at a fixed time instant: $x = \{x_i\}_0^\infty$, a convenient algorithm is obtained by a slight modification of the technique above. Let $\mathscr{H}'_{(0)}$ denote the elements of $\mathscr{H}$ satisfying $x_i = 0$, $i < 0$; and let $a'_{(0)} \in B(\mathscr{H}'_{(0)})$ denote the associated truncation of $a \in B(\mathscr{H})$. Then by substituting in (5.15) $r'_{(0)}$, $c'_{(0)}$, and $\lambda'_{(0)}$ for $r_{(0)}$, $c_{(0)}$, and $\lambda_{(0)}$, respectively, Eq. (5.16) give directly the inverse $c'_{(0)} \triangleq a^{-1}$ needed in the predictor formula (4.6). The same algorithm has applications also for stationary processes and when the process is described by difference equations, [17].

We conclude this section by showing that the covariance function in Theorem (5.7) defines a regular, full rank process.

Proposition. *If r is the covariance function of a process $f = \{f_n\}_{-\infty}^{\infty}$ such that $r \in B$, $r > 0$, then f is regular and of full rank.* (5.22)

Proof. We show first regularity, i.e., that $(f_n/M_m) \to 0$ as $m \to -\infty$. By Theorem (5.7) $r = a a^*$; $a, a^{-1} \in B^+$. Thus the process aw, w orthonormal process, is equivalent to f up to second order moments.

We compute

$$(f_n/M_m) = \hat{f}_m(n-m) = \sum_{i \leqq m} p_{m,i} f_i.$$

The vector $p_m = (p_{m,m}, p_{m,m-1}, \ldots) \in V$ is obtained as the m'th row of the matrix $p \in B^+$ given by (4.6):

$$p = (U^{(n-m)} a)_+ a^{-1}.$$

Let $c_m \triangleq (a_{n,m}, a_{n,m-1}, \ldots) \in V$ denote the m'th row of $(U^{(n-m)} a)_+$. Then

$$\|p_m\| \leqq \|a^{-1}\| \, \|c_m\| \to 0 \qquad \text{as } m \to -\infty.$$

This entails that $(f_n/M_m) \to 0$ as $m \to -\infty$ proving the regularity of f. That f is of full rank follows from the fact that $r > 0$. Indeed, let

$$g_n = f_n - \sum_{i \leqq n-1} p_{n-1,i} f_i.$$

Writing $y \triangleq (I, -p_{n-1,n-1}, -p_{n-1,n-2}, \ldots) \in V$ we observe that

$$G_n = E g_n g'_n = [y r y] > 0,$$

uniformly in n, since $[y y] \geqq I$.

6. Wiener-Hopf Equations in *B*

As we saw in Section 4 and 5 the basic prediction problem for processes defined by their covariance function is solved by first finding a class C-representation for the process, after which the simple reasoning of Section 4 applies. But since in more complex least squares estimation problems quite the same approach would not apply, we wish to present here another more general one.

The solution of a general class of least squares estimation problems can be reduced to that of solving the so-called generalized Wiener-Hopf equations:

$$\sum_{i \leqq n} p_{ni} r_{ik} - q_{nk} = 0, \qquad k = n, n-1, \ldots, \text{ all } n \in N. \tag{6.1}$$

Here, r_{ik} and q_{nk} are known covariance matrices, and p_{ni} are the matrix coefficients of the optimal estimator to be determined. Let us recall that Eqs. (6.1) are merely a restatement of the fact that the estimation error is to be orthogonal to the manifold spanned by the observed random variables.

We plan to present a neat way to solve the Eqs. (6.1). In that purpose, define the operators $r = \{r_{ik}\}$, $q = \{q_{nk}\}$, and $p = \{p_{ni}\}$; and assume that $r \in B$ and $q \in B^+$. We want conditions on these matrices such that $p \in B^+$.

With the notation $\hat{B}^- = \{a\colon\ a \in B,\ a_{nm} = 0 \text{ for } m \leqq n\}$ Eqs. (6.1) read:

$$p\,r - q \in \hat{B}^-. \tag{6.2}$$

We prove the following theorem.

Theorem. *If* $r \in B$, $r > 0$, *and* $q \in B^+$, *then* Eq. (6.2) *has the unique solution* (6.3)

$$p = (q\,a^{*-1})_+\,a^{-1} \tag{6.4}$$

in B^+, *where* $a \in B^+$ *is any factor solving* $r = a\,a^*$.

Proof. Theorem (5.7) ensures the existence of an $a \in B^+$ such that $a^{-1} \in B^+$. Then $a^{*-1} \in B^-$. Let $p \in B^+$ solve (6.2). Then by Proposition (3.7) and making use of the fact that $p\,r - q$ belongs to $\hat{B}^-$ rather than merely to B^- we have:

$$(p\,r - q)\,a^{*-1} = p\,a - q\,a^{*-1} \in \hat{B}^-. \tag{6.5}$$

Since $p\,a \in B^+$, the inclusion relation in (6.5) can hold only when

$$p\,a - (q\,a^{*-1})_+ = 0,$$

which is the required formula.

To show the uniqueness of p, let also $r = b\,b^*$, $b \in B^+$. Then by Theorem (5.7), $b = a\,u$, u diagonal and unitary; and

$$p = (q\,b^{*-1})_+\,b^{-1} = (q\,a^{*-1}u)_+\,u^{-1}a^{-1} = (q\,a^{*-1})_+\,a^{-1}.$$

Remark. Formula (6.4), when applied to the prediction problem of Section 4, reduces easily to (4.6); and Theorem (6.3) thus ensures that the optimal predictor is uniquely determined by r.

References

1. Kolmogorov, A.: Stationary sequences in Hilbert space. Bolletin Moskov. Gosud, Univ. **2**, 40 (1941) [in Russian].
2. Wiener, N.: Extrapolation, interpolation, and smoothing of stationary time series. New York: Wiley 1949.
3. Wold, H.: A study in the analysis of stationary time series. Stockholm: Almquist and Wiksell 1938.
4. Grenander, U., and M. Rosenblatt: Statistical analysis of stationary time series. New York: Wiley 1957.
5. Wiener, N., and P. Masani: The prediction theory of multivariate stochastic processes. I and II. Acta math. **98**, 111–150 (1957); **99**, 93–137 (1958).

6. Helson, H., and D. Lowdenslager: Prediction theory and Fourier series in several variables. Acta math. **99**, 165–202 (1958).
7. – Vector valued processes. Proc. Fourth Berkeley Sympos. math. Statist. Probab., vol. II, 203–212, 1961.
8. Szego, G.: Beiträge zur Theorie der Toeplitzschen Formen, I, II. Math. Z. **6**, 167–202 (1920); **9**, 167–190 (1921).
9. Devinatz, A.: The factorization of operator valued functions. Ann. of Math., II. Ser. **73**, 458–495 (1961).
10. Lowdenslager, D.: On factoring matrix valued functions. Ann. of Math., II. Ser. **78**, 450–454 (1963).
11. Helson, H.: Lectures on invariant subspaces. New York: Academic Press 1964.
12. Parzen, E.: Extraction and detection problems and reproducing Hilbert spaces. J. Soc. industr. appl. Math., Ser. A: Control, **1**, 35–62 (1962).
13. Rao, M.: Inference in stochastic processes. III. Z. Wahrscheinlichkeitstheorie verw. Geb. **8**, 49–72 (1967).
14. Kalman, R.: A new approach to linear filtering and prediction problems. Trans. ASME, Ser. D, Journal of Basic Engrg. **82**, 35–45 (1960).
15. Doob, J.: Stochastic processes. New York: Wiley 1962.
16. Naimark, M.: Normed rings. Groningen, The Netherlands: P. Noordhoff N.V. 1959.
17. Rissanen, J., and L. Barbosa: Properties of infinite covariance matrices and stability of optimum predictors. Information Sciences: An International Journal (to appear).

Dr. J. Rissanen and
Dr. L. Barbosa
IBM Research Laboratory
Monterey & Cottle Roads
San Jose, California 95114, USA

(Received September 28, 1967)