VECTOR-VALUED PROCESSES

HENRY HELSON[1]
UNIVERSITY OF CALIFORNIA, BERKELEY
AND
DAVID LOWDENSLAGER[2]
PRINCETON UNIVERSITY

1. Introduction

The theory of prediction for second order stationary stochastic processes has taken an established place in probability theory, as a result of the complete and elegant treatment achieved by successive authors [2], [3], [4], [6], [12], [13]. More recently the generalization to *multivariate processes* (consisting of several separate but correlated processes with correlations stationary in time) has been taken up by several authors [5], [8], [9], [10], [11], following the beginning made by Wold in this direction. In this paper we survey this problem of multivariate prediction, adopting our own rather eccentric point of view but drawing on the work of others (especially of Masani and Wiener) without hesitation.

2. The prediction problem

Let $y(t)$, with $t = 0, \pm 1, \pm 2, \cdots$, be a vector-valued stochastic process; that is, a random sequence of column vectors, whose components $y^j(t), j = 1, 2, \cdots, N$ are complex random variables. We assume the process is *stationary* in the sense that the inner products

$$(1) \qquad (y^j(t), y^k(s)) = E[y^j(t)\bar{y}^k(s)] = R^{jk}(t-s); \qquad j, k = 1, \cdots, N$$

depend only on the difference $t - s$ of the time arguments. The sequence of matrices $R(t)$, t integral, whose elements are $R^{jk}(t)$ is called the *covariance sequence* of the process. There is a Borel measure, called the *spectral measure* of the process, whose values are positive semidefinite matrices of order N, and whose Fourier-Stieltjes coefficients are the matrices $R(t)$,

$$(2) \qquad R(t) = \int_0^{2\pi} e^{-it\theta}\, dM(e^{i\theta}).$$

(Such matrix or vector equations and integrals can be easily interpreted in terms of the components involved.) For example, if the process is *orthonormal*, which means that each $y^j(t)$ has norm one and is orthogonal to $y^k(s)$ unless $j = k$ and

[1] Supported in part by the National Science Foundation.
[2] Supported in part by the Research and Development Command of the U.S. Air Force.

$s = t$, we find that $R(0) = I$, the identity matrix of order N, and $R(t) = 0$ for other values of t; and $dM\ [\exp(i\theta)] = (1/2\pi)I\ d\theta$.

The basic problem of prediction theory is to determine the linear combination of $\{y^k(t) | k = 1, \cdots, N, t < 0\}$ which is nearest to $y^j(0)$ in the mean square or L^2 metric. The distance to be minimized is then

$$||y^j(0) - \sum_{n>0} A_n^{jk} y^k(-n)||, \tag{3}$$

where A_n^{jk} are arbitrary complex numbers.

From (1) it follows that this problem is as general as the problem of predicting $y^j(s)$ from the past up to s, for any moment s.

We have described a process whose elements are random variables on a probability space with the usual inner product

$$(x, y) = E[x\bar{y}]. \tag{4}$$

However, the prediction question involves only the constructions of a Hilbert space (the formation of linear combinations, inner products, and distances), and we can dispense henceforth with all reference to the underlying probability space. Instead we consider $y^j(t)$ as an element of an abstractly given Hilbert space, whose inner product gives meaning to the statements just made.

Now suppose that $A = A\ [\exp(i\theta)]$ and $B = B\ [\exp(i\theta)]$ are two functions whose values are matrices with complex components of K rows and N columns. Let M be a measure with positive semidefinite matrices of order N as values. Then we can define an inner product by the formula

$$(A, B) = \int_0^{2\pi} \operatorname{tr}\ [A(e^{i\theta})\ dM\ (e^{i\theta}) B^*(e^{i\theta})]. \tag{5}$$

Denote by $L_K^2(M)$ the Hilbert space of measurable K by N matrix functions A for which the norm $||A|| = (A, A)^{1/2}$ is finite, with the inner product just defined. [Here $\operatorname{tr} A = \sum A_{jj}$ and $(A^*)_{jk} = \bar{A}_{kj}$. The integer N is fixed once for all and need not appear in our notation. When $dM = (1/2\pi)I\ d\theta$. so the process is orthonormal, we shall write L_K^2 for $L_K^2(M)$.]

If $K = 1$ we have row vector-valued functions and the trace is superfluous. In this case let $Y^j(t)$ denote the row vector-valued function with $\exp(-it\theta)$ in the jth position and 0 elsewhere. It is clear that $\{Y^j(t) | j = 1, \cdots, N, t = 0, \pm 1, \cdots\}$ spans $L_1^2(M)$. What is more interesting is that the linear transformation which sends $y^j(t)$ into $Y^j(t)$ is a unitary isomorphism between the closed linear span of $\{y^j(t)|, j = 1, \cdots, N, t = 0, \pm 1, \cdots\}$ and $L_1^2(M)$. This makes it possible to transform Hilbert space problems concerning the process y into isomorphic problems for a canonical process in a function space.

The matrix function space $L_N^2(M)$ is in a similar way a model for the whole vector process $y(t)$, since the matrix is a column vector whose components are row vectors which in turn are isomorphic to the components of y: $y(s)$ corresponds to $[\exp(-is\theta)]I$.

The minimization problem (3) is then the same as the problem of minimizing the square root of

$$\int [Y^j(0) - \sum A_n^{jk} Y^k(-n)]\, dM\, [Y^j(0) - \sum A_n^{jk} Y^k(-n)]^*, \tag{6}$$

where n is summed over 1, 2, $\cdots$, and k over 1, $\cdots$, N. If we sum the integrals in (6) on the index j, we can minimize the integral

$$\int \operatorname{tr} [(I - \sum_{n>0} A_n e^{in\theta})\, dM\, (I - \sum_{n>0} A_n e^{in\theta})^*] \tag{7}$$

so as to solve the minimal problem for $j = 1, 2, \cdots, N$ simultaneously.

The transformation T which sends $y^j(t)$ into $y^j(t+1)$ is described in the space $L_1^2(M)$ by scalar multiplication with $\exp(-i\theta)$. Its effect is to shift all times by 1, and by (1) it is a unitary isomorphism.

The purpose of the work described in this report is to determine how the process, in particular the spectral measure, determines the nature of the prediction problem, and conversely.

3. Moving averages and outer functions

Let $\mathfrak{M}_s$ be the linear span of $\{y^j(t) | j = 1, \cdots, N, t < s\}$. This is the past up to time s. Let $\mathfrak{Z}_s$ be the orthogonal complement of $\mathfrak{M}_s$ in $\mathfrak{M}_{s+1}$ and let $\mathfrak{M}_{-\infty}$ be the intersection of all the $\mathfrak{M}_s$. These are the *innovation manifold* at time s and the *remote past*, respectively. The shift operator T clearly satisfies

$$\begin{aligned} T\mathfrak{M}_s &= \mathfrak{M}_{s+1} \\ T\mathfrak{Z}_s &= \mathfrak{Z}_{s+1} \\ T\mathfrak{M}_{-\infty} &= \mathfrak{M}_{-\infty}. \end{aligned} \tag{8}$$

The dimension r of $\mathfrak{Z}_s$ does not depend on s and satisfies $0 \leqq r \leqq N$; r is called the *rank* of the process. If $r = 0$, the process is called *purely deterministic;* then $\mathfrak{M}_s = \mathfrak{M}_{-\infty}$ for all s.

The most interesting case is when $\mathfrak{M}_{-\infty} = \{0\}$. Then the components of $y(t)$ are the sums of their projections on $\mathfrak{Z}_s$ for $s \leqq t$. Choose once and for all r orthonormal elements $x^1(0), \cdots, x^r(0)$ spanning $\mathfrak{Z}_0$, and let $x^i(s) = T^s x^i(0)$, so that the process $x(s)$ is orthonormal and $\{x^j(s) | j = 1, \cdots, r; s < t\}$ spans $\mathfrak{M}_t$. There exist N by r matrices A_n such that

$$y^j(t) = \sum_{n \geqq 0, k} A_n^{jk} x^k(t - n). \tag{9}$$

Taking into account the stationarity condition (1), we see that (9) holds for all t and that the A do not depend on t. Equation (9) may be condensed into the convolution equation

$$y = A * x. \tag{10}$$

The process y is said to be a *one-sided moving average* of x.

Conversely, suppose that y is a one-sided moving average of an orthonormal process (or of any process whose remote past is $\{0\}$). Then the manifold $\mathfrak{M}_s$ is contained in the corresponding manifold $\mathfrak{N}_s$ for x. Therefore

$$\mathfrak{M}_{-\infty} = \bigcap_s \mathfrak{M}_s \subset \bigcap_s \mathfrak{N}_s = \{0\}. \tag{11}$$

In case $y = A * x$ a direct calculation shows that the spectral measure is given by

$$dM(e^{i\theta}) = \frac{1}{2\pi} A(e^{i\theta})A^*(e^{i\theta})\, d\theta,$$
$$A(e^{i\theta}) \sim \sum_{n=0}^{\infty} A_n e^{in\theta}. \tag{12}$$

Again the converse is true: if $dM = (1/2\pi)AA^*\, d\theta$ with A [exp $(i\theta)$] as in (12) the y process is a one-sided moving average of an orthonormal process. The function A above satisfies

$$\|A\|^2 = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{tr}\,[A(e^{i\theta})A^*(e^{i\theta})]\, d\theta \tag{13}$$
$$= \sum_0^{\infty} \operatorname{tr}\,[A_n A_n^*] < \infty.$$

The class of all such functions is a Hilbert space which will be denoted by H^2. (The functions in H^2 have N rows, instead of N columns like the elements of L_k^2.)

We have seen that if $\mathfrak{M}_{-\infty} = \{0\}$, the process is a moving average of a special orthonormal process, its innovation process. On the other hand, if $y = A * x$ with x orthonormal, we cannot say x is the innovation process of y. For instance if $A_1 = I$, A_0 and all the other A are 0, then the components of $x(s-1)$ and not $x(s)$ span the innovation manifold $\mathfrak{I}_s$ of y. Moreover, if some of the columns of A [exp $(i\theta)$] are zero, or more generally A [exp $(i\theta)$] has some fixed null space, the vectors in this null space do not affect y at all and thus are irrelevant. They are even orthogonal to all the y.

Beurling [1] has studied the approximate inversion of such one-sided convolutions in the scalar case ($N = 1$), using the isomorphic function theoretic problem. In L_N^2, the backward shift T^{-1} is multiplication by exp $(i\theta)$. Linear combinations of the past will have the form $B_n \exp(in\theta)A$ [exp $(i\theta)$], where $n > 0$, so that the prediction problem becomes the problem of minimizing

$$\|(I - \sum_{n>0} B_n e^{in\theta})A(e^{i\theta})\| \tag{14}$$

in L^2, or actually in $H^2 \subset L^2$. Using the Parseval identity we see that the minimum distance is at least $[\operatorname{tr}\,(A_0 A_0^*)]^{1/2}$. The function A is called *outer* if this is exactly the minimum, that is if the closure of the set of functions $[I - \sum_{n>0} B_n \exp(in\theta)]A$ [exp $(i\theta)$] in H^2 contains the constant A_0. The definition has meaning if A is any N by r matrix, even with $r > N$.

It is possible to prove that if A is outer, the null space of A [exp $(i\theta)$] is almost everywhere the null space of A_0, and that the span of the elements (14) includes every element of H^2 whose null space includes that of A_0 almost every-

where. We shall assume henceforth, without loss of generality, that the null space of outer functions is $\{0\}$. Then the innovation manifold of $y = A * x$ is identical with the innovation manifold of x.

Suppose $y = B * z$ is a one-sided convolution of an orthonormal process. Then $\mathfrak{M}_{-\infty} = \{0\}$ and y is a one-sided convolution of its innovation process. Therefore we have $2\pi\, dM\, [\exp(i\theta)] = B\, [\exp(i\theta)] B^*\, [\exp(i\theta)]\, d\theta = A\, [\exp(i\theta)] A^*\, [\exp(i\theta)]\, d\theta$, that is, if B is arbitrary in H^2 there is an outer A with $BB^* = AA^*$.

We can also consider the uniqueness of the factorization $2\pi\, dM = AA^*\, d\theta$. Each such outer factorization corresponds to a choice of vectors $x^i(0)$ to span the innovation manifold $\mathfrak{I}_0$. Two such sets can differ only by a unitary map. Therefore if $AA^* = BB^*$ with both A and B outer, there is a constant unitary matrix U with $A = BU$.

The formal solution of the prediction problem is within view, in the case where $\mathfrak{M}_{-\infty} = \{0\}$. The representation $y = A * x$ where A is outer, so that the components of $x(s)$ span $\mathfrak{I}_s$, shows that the vector in $\mathfrak{M}_t$ nearest to y_s is $\sum_{n>0} A_n x_{s-n}$. Therefore, formally, the problem is resolved into two related parts, that of finding the coefficients of A and that of finding x. We have already indicated how the problem of finding A is the problem of factoring the spectral measure: $2\pi\, dM = AA^*\, d\theta$, A outer. This factorization we have seen to be unique up to a right multiplication by a constant unitary matrix. In the next section we shall outline criteria for the existence of such a factorization.

The problem of finding the x from the y is the problem of inverting the one-sided convolution (10) by another one-sided convolution, that is finding B with $B * y = B * A * x = x$. The definition of outer function implies that approximating finite sequences exist, but it is not completely understood, even in the scalar case, when an element B actually exists or how to find it. Several special cases have been discussed by Wiener and Masani.

In case $\mathfrak{M}_{-\infty} \neq \{0\}$ at least a portion of the process is deterministic. Let $\mathfrak{N}$ be the manifold spanned by all the $\mathfrak{I}_s$, so that $\mathfrak{N}$ is the orthogonal complement of $\mathfrak{M}_{-\infty}$ in the space spanned by the process. Let $y(x) = u(s) + r(s)$ where $u^i(s) \in \mathfrak{N}$ and $v^i \in \mathfrak{M}_{-\infty}$. One can prove that the processes u and v are *pure* in the sense that u is a moving average and v is a purely deterministic process. This decomposition is due to H. Wold. Thus every process is an "orthogonal direct sum" of the two pure types. This type of orthogonal direct sum offers some paradoxes. For instance, even in the scalar case the sum of a moving average process and a deterministic process may be a moving average; this will happen when the spectral measure of v is absolutely continuous.

4. The Szegö theorem

One of the key theorems in our treatment is a generalization of a classical result of Szegö:

$$\inf \frac{1}{2\pi} \int_0^{2\pi} \operatorname{tr} \{[B_0 - P(e^{i\theta})]\, dM\, (e^{i\theta})[B_0 - P(e^{i\theta})]^*\} = \exp\left\{\frac{1}{2\pi} \int_0^{2\pi} \log \det M'(e^{i\theta})\, d\theta\right\}. \tag{15}$$

The infimum is taken over all matrices B_0 with $\det B_0 = 1$ and all trigonometric polynomials $P[\exp(i\theta)] = \sum_{n>0} B_n \exp(in\theta)$. M' is the Radon-Nikodym derivative of dM with respect to Lebesgue measure. The minimizing function $B_0 - H[\exp(i\theta)]$ satisfies

$$(B_0 - H)\, dM\, (B_0 - H)^* = C\, d\theta, \tag{16}$$

where C is a constant positive semidefinite matrix, so that $B_0 - H$ vanishes on the set carrying the part of dM singular with respect to Lebesgue measure. The term in the braces on the right of (15) can be $-\infty$ but not $+\infty$. If it is $-\infty$, both sides of (15) are 0.

If we set $B_0 = I$ and let P vary, the positive semidefinite matrix

$$\frac{1}{2\pi} \int_0^{2\pi} (I - P)\, dM\, (1 - P)^* \tag{17}$$

has a minimum D and $\det D$ is the right side of (15). This is a version due to Masani and Wiener and the relation between it and the one above is given by the following equation which holds for any positive semidefinite matrix D:

$$\inf \exp\{\operatorname{tr}(B_0 D B_0^*)\} = \det D, \qquad (\det B_0 = 1). \tag{18}$$

The problem of determining D from dM is unsolved; its importance lies in the fact that $A_0 A_0^* = D$ if $2\pi\, dM = AA^*\, d\theta$ is an outer factorization of M.

Using the linear and norm-preserving isomorphism between the Hilbert space spanned by the components of $y(s)$ and $L_1^2(M)$, our theorem has the following interpretation: If we replace the y process by all possible processes of the form $z = B_0 y$ with $\det B_0 = 1$, and compute for each of these processes the sum of the squares of the prediction errors for each component, the minimum is the right side of (15).

If this right side is positive, the innovation manifolds contain N linearly independent elements, so that the process is of rank N, and conversely. In fact, if a linear combination of the components of $y(0)$ lies in the past, this linear combination can be used as one component of $B_0 y$ to make the sum of the errors in predicting the other components arbitrarily small.

In case $r = N$ so that the process is of full rank, the Wold decomposition is easily described. The interval from 0 to 2π can be split into two sets U and V, one of Lebesgue measure 1 and the other 0. Then M is absolutely continuous on U and singular on V with respect to Lebesgue measure. The Wold decomposition $y = u + v$ is accomplished by multiplying the functions representing y by the characteristic function of U and V respectively. No such simple characterization is known if $0 < r < N$.

Again if the left side of (15) is not 0, we find from (16) that

$$M' = (B_0 + H)^{-1}C^{-1/2}((B_0 + H)^{-1}C^{-1/2})^* = AA^* \tag{19}$$

is an outer factorization of M' with $\det A_0A_0^* \neq 0$. The result implies that a necessary and sufficient condition that $M' = AA^*$ with A outer and $\det A_0A_0^* \neq 0$ is that

$$\frac{1}{2\pi}\int_0^{2\pi} \log \det M'(e^{i\theta})\, d\theta > -\infty. \tag{20}$$

It also follows that for any A in H^2,

$$\frac{1}{2\pi}\int_0^{2\pi} \log \det AA^*\, d\theta \geqq \log \det A_0A_0^*, \tag{21}$$

and that if $\det A_0A_0^* \neq 0$, equality is necessary and sufficient for A to be outer. This is a generalization of a study by Beurling [1] of such problems in the scalar case.

5. The singular case and analytic subspaces

In this section we assume for convenience that M is absolutely continuous, with derivative $M' = W$. The case when $r < N$ so that the left side of (15) is 0 is called the *singular case*. To see what might happen, consider the case when the measure W is diagonal

$$W = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & w_N \end{pmatrix}, \tag{22}$$

when each component of the process is orthogonal to every other component. It is clear that we can factor W if we can factor each w_j, with $j = 1, \cdots, N$. The converse is also true. Each of these can be factored if $\int_0^{2\pi} \log \det w_j\, d\theta > -\infty$ for each j for which $w_j \neq 0$. Therefore, if $\mathfrak{M}$ is the range of W, which reduces W, and DW is the determinant of the restriction to $\mathfrak{M}$ of W,

$$\int_0^{2\pi} \log DW\, d\theta > -\infty \tag{23}$$

is sufficient for a factorization to exist. This inequality is also necessary for factorization. An analogous statement holds whenever the range of W is constant, and a proof proceeds easily from consideration of the nonsingular case.

But the range of W need not be constant. For instance if

$$A(e^{i\theta}) = \begin{pmatrix} 1 & e^{i\theta} \\ e^{i\theta} & e^{2i\theta} \end{pmatrix} \tag{24}$$

and $W = AA^*$ then W has an outer factorization, by an earlier remark. Here the range of A is the range of W, and it is not constant.

We shall introduce a type of function from the unit circle into subspaces of compex N-space which will characterize the possible ranges of factorable functions W. A subspace-valued function $\mathfrak{M}[\exp(i\theta)]$ is *analytic* provided there exists an H^2 function A so that the range of $A[\exp(i\theta)]$ is $\mathfrak{M}[\exp(i\theta)]$ almost everywhere.

The characterization of factorable functions is this: $W = AA^*$ with A in H^2 if and only if the range $\mathfrak{M}[\exp(i\theta)]$ is an analytic subspace and

$$\int \log DW \, d\theta > -\infty. \tag{25}$$

Here DW denotes the determinant of the restriction of W to $\mathfrak{M}$.

With each analytic subspace $\mathfrak{M}$ we shall associate a special outer element U of H^2. U is a partial isometry: UU^* is a projection onto the analytic subspace $\mathfrak{M}$ and U^*U is a projection onto the orthogonal complement of the null space of U, which is constant since U is outer. One finds U in the following way. Take A to be an outer N by p function with $\mathfrak{M}$ as range. Then A^*A is a nonsingular p by p function, which is equal to B^*B where B is p by p and nonsingular, using the existence theorem in the full rank case for vector processes of dimension p. From the equation $A^*A = B^*B$ we deduce $(AB^{-1})^*(AB^{-1}) = I$, and we may choose $U = AB^{-1}$. It follows from the uniqueness theorem for outer factorings stated earlier that U is unique up to right multiplication by a constant unitary matrix.

There is a curious statement of this theorem in terms of analytic vectors: if $v_1[\exp(i\theta)], \cdots, v_k[\exp(i\theta)]$ are analytic vector-valued functions with N components, we can find vector functions $u_1[\exp(i\theta)], \cdots, u_p[\exp(i\theta)]$ analytic, orthonormal, and spanning almost everywhere the same subspace.

We can indicate an outline of a proof of the factoring theorem using this construction. Take a partial isometry U corresponding to the analytic subspace $\mathfrak{M}$ and replace W by U^*WU, which turns out to be a p by p spectral function corresponding to a p by p process of full rank. Then

$$DW = \det U^*WU, \tag{26}$$

where the determinant is the p by p determinant. This reduces the problem to the case of full rank processes, which has been solved already.

Another way of looking at this analysis is to consider the general one-sided moving average $y = A * x$ as an iterated moving average $y = U * B * x$, where B is a p by p element of full rank and U is the special kind of outer partial isometry analyzed above. If A is not outer, we may achieve a similar factorization $A = UB$, but B will no longer be outer. We may then factor B so that $B = CD$ where C and D are square and $C^*C = B^*B$ and C is outer. It is not clear what implication this factoring has for the prediction problem.

If $\mathfrak{M}$ is an analytic subspace we take an outer isometry U with $\mathfrak{M}$ as range. Let $U[\exp(i\theta)] = \sum_{n \geq 0} U_n \exp(i\theta)$ and define

$$d(\mathfrak{M}) = -\log \det (U_0^* U_0). \tag{27}$$

Then $d \geqq 0$ *and* $d = 0$ *if and only if* $\mathfrak{M}$ *is constant.* The number d measures the deviation of $\mathfrak{M}$ from constancy. Suppose now A in H^2 has range $\mathfrak{M}$. Then

$$\frac{1}{2\pi}\int_0^{2\pi} \log D(A^*A) \geqq \log D(A^*A) + d, \tag{28}$$

and equality is necessary and sufficient in order that A be outer. This result generalizes equality in (21) which characterizes the outer functions of full rank. However, if A is N by p, both sides are $-\infty$ unless $N \geqq p$ and the rank of A is p almost everywhere.

It should be remarked that Masani and Wiener [10] have found a different characterization of those 2 by 2 matrix measures which can be factored. Moreover Masani has a structure theory [8] for processes of less than full rank which is different from the one which has been described. The whole problem does not seem to us to be closed.

6. Invariant subspaces

Let $\mathfrak{N}$ denote the subspace of L_1^2 spanned by $x^j(s)$, with $j = 1, \cdots, N$ and $s \leqq 0$, where x is the canonical orthonormal process. A subspace $\mathfrak{M}$ of $\mathfrak{N}$ such that $T^{-1}\mathfrak{M} \subset \mathfrak{M}$ is called an *invariant* subspace. An element B of H^2 generates an invariant subspace as follows. Take a one-sided moving average $y = B * x$ of the orthonormal process, and consider $\mathfrak{M}_s$, as defined in section 3. We have $\mathfrak{M}_s \subset \mathfrak{N}$ for $s \leqq 1$, and since $T^{-1}\mathfrak{M}_s = \mathfrak{M}_{s-1} \subset \mathfrak{M}_s$, $\mathfrak{M}_s$ is an invariant subspace. Lax [7], extending results of Beurling [1], has proved that, conversely, every invariant subspace is of this form. In particular, he has proved that the functions $A\,[\exp(i\theta)] = \sum_{n \geqq 0} A_n \exp(in\theta)$ which are partial isometries are themselves enough to generate the invariant subspaces. This gives an alternate way to approach the factoring of elements of H^2. Since $B * x$ generates an invariant subspace of $\mathfrak{N}$, there exists an isometry A which generates the same subspace, and A is a factor of B in the sense that $B = AC$, with C in H^2 also.

REFERENCES

[1] A. Beurling, "On two problems concerning linear transformations in Hilbert space," *Acta Math.*, Vol. 81 (1948), pp. 239–255.

[2] J. L. Doob, *Stochastic Processes*, New York, Wiley, 1953.

[3] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, New York, Wiley, 1957.

[4] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, California Monographs in Mathematical Sciences, Berkeley and Los Angeles, University of California Press, 1958.

[5] H. Helson and D. Lowdenslager, "Prediction theory and Fourier series in several variables," *Acta Math.*, Vol. 99 (1958), pp. 165–202.

[6] A. Kolmogorov, "Stationary sequences in Hilbert space," *Bolletin Moskovskogo Gosudarstvenogo Universiteta*, Vol. 2 (1941), 40 pp. (In Russian.)

[7] P. Lax, "Translation invariant subspaces," *Acta Math.*, Vol. 101 (1959), pp. 163–178.

[8] P. MASANI, "Sur la fonction génératrice d'un processus stochastique vectoriel," *C. R. Acad. Sci. Paris*, Vol. 249 (1959), pp. 360–362; "Isormorphie entre les domaines temporel et spectral d'un processus vectoriel, régulier," pp. 496–498; "Sur les fonctions matricielles de la classe de Hardy H_2," I and II, pp. 873–875, pp. 906–907.

[9] P. MASANI and N. WIENER, "The prediction theory of multivariate stochastic processes," I and II, *Acta Math.*, Vol. 98 (1957), pp. 111–150, Vol. 99 (1958), pp. 93–137.

[10] ———, "On bivariate stationary processes and the factorization of matrix-valued functions," *Teor. Veroyatnost. i Primenen;* Vol. 4 (1959), pp. 322–331.

[11] M. ROSENBLATT, "A multidimensional prediction problem," *Ark. Mat.*, Vol. 3 (1958), pp. 407–424.

[12] N. WIENER, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, New York, Wiley, and Cambridge, Mass., Technology Press, 1949.

[13] H. WOLD, *A Study in the Analysis of Stationary Time Series*, Stockholm, Almqvist & Wiksell, 1938.