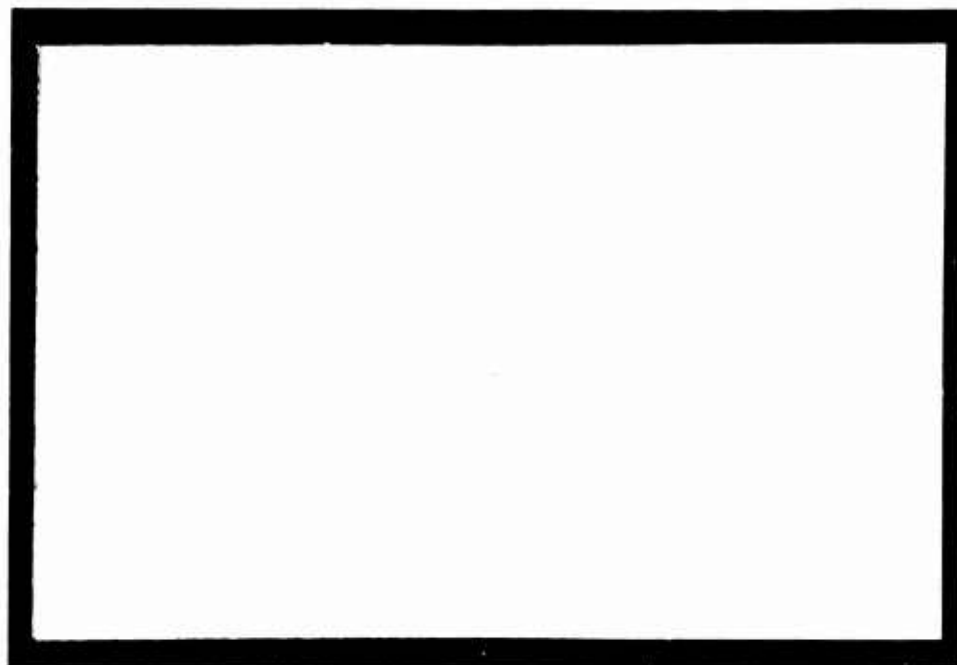


AD630756



MATHEMATICS RESEARCH CENTER

UNITED STATES ARMY

CLEARANCE USE			
FOR FEDERAL, SCIENTIFIC AND			
TECHNICAL INFORMATION			
Library	Microfilm		
3.00	0.75	58	as

Code 1



MATHEMATICS RESEARCH CENTER, UNITED STATES ARMY
THE UNIVERSITY OF WISCONSIN

Contract No.: DA-11-022-ORD-2059

RECENT TRENDS IN MULTIVARIATE
PREDICTION THEORY

P. Masani

MRC Technical Summary Report #637
January 1966

Madison, Wisconsin

ABSTRACT

The paper surveys the present state of the theory of linear, least squares prediction of q -variate weakly stationary stochastic processes with discrete time. The emphasis is on logical order. Hence recent developments are described within the context of a general theory rather than chronologically. Methods for computing the predictor are briefly discussed, but purely statistical questions such as the estimation of covariances are omitted.

Recent Trends in Multivariate Prediction Theory

by

P. Masani

1. Introduction
2. The Gram matricial structure for $\mathbb{H}^q$
3. Elementary solution of the Prediction Problem
4. The shift operator and Wold-Zasuhin decomposition
5. Spectral analysis
6. The space $L_{2, \underline{M}}$ for a non-negative hermitian matrix-valued measure
7. Isomorphism between the temporal and spectral spaces
8. Cross covariance and spectra. Subordination
9. Spectral analysis of a purely non-deterministic SP
10. Spectral analysis of a full-rank SP
11. Concordance of Wold-Zasuhin and Lebesgue-Cramer decompositions
for degenerate ranks
12. Degenerate rank factorization
13. Spectral and autoregressive representations for the predictor of a
purely non-deterministic SP
14. Determination of the generating function from the spectral density
15. The factorization of matricial rational spectral densities

RECENT TRENDS IN MULTIVARIATE PREDICTION THEORY

P. Masani

1. Introduction

From among the many facets of multivariate prediction we will consider only the theory of linear, least squares prediction of q -variate, weakly stationary stochastic processes with discrete time. Our purpose is to give a coherent account of the present state of this theory. We shall therefore refer to recent developments not in isolation but within the context of the general theoretical framework. Our emphasis will be on generality and logical order, but the practical side will also be discussed though somewhat briefly (cf. §§2, 15). Statistical questions of estimation, etc. will be omitted.

To recall the problem involved in such prediction suppose that $\underline{x}$ is a q -dimensional vector quantity associated with some long enduring mechanism in nature, and that $\underline{x}_n$ denotes its value at time $t = n$. Suppose that we have been measuring $\underline{x}$ every second from the remote past up to the present moment $t = 0$, and have so obtained a sequence of readings

$$(1.1) \quad \underline{x}_k = \underline{a}_k, \quad k = 0, -1, -2, \dots$$

Is there some way to forecast the future value $\underline{x}_\nu$, $\nu \geq 1$, on the basis of the information contained in (1.1)? Without further knowledge of the mechanism our answer to this question must be in the negative.

If, however, we assume that our mechanism is such that the sequence

Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No.: DA-11-022-ORD-2059.

$(\underline{x}_k)_{k=-\infty}^0$ is part of a time-sequence (sample-function) of a q-variate stationary stochastic process (SP) $(\underline{f}_n)_{n=-\infty}^{\infty}$ over a probability space $(\Omega, \mathcal{B}, P)$, so that

$$(1.2) \quad \underline{x}_n = \underline{f}_n(\omega_0), \quad \omega_0 \in \Omega, \quad -\infty < n < \infty$$

and that we know the probabilistic structure of this SP, then the answer is in the affirmative as we proceed to indicate.

Denote the forecast value of $\underline{x}_v$ by $\hat{\underline{x}}_v$. As $\hat{\underline{x}}_v$ is to depend on the past and present values $\underline{x}_k$, $k \leq 0$ alone, we must expect that $\hat{\underline{x}}_v \neq \underline{x}_v$ except when our mechanism is purely deterministic. Such mechanisms are of course very important, but they are only of peripheral interest in the theory of probabilistic or statistical prediction, which concerns us here. In this theory the problem is to find the $\hat{\underline{x}}_v$ which comes closest to $\underline{x}_v$ under some preassigned statistical error criterion. In least squares, linear prediction we adopt the root-mean-square (RMS) error criterion, and confine attention to $\hat{\underline{x}}_v$ which are mean limits of linear combinations $\sum_{k=0}^n \underline{A}_k^{(n)} \underline{x}_{-k}$, where $\underline{A}_k^{(n)}$ are $q \times q$ matrices $\begin{pmatrix} 1 \\ \end{pmatrix}$. It can be shown that the $\underline{A}_k^{(n)}$ are determinable and the problem solvable when the covariance structure of the stationary SP $(\underline{f}_n)_{n=-\infty}^{\infty}$ is known.

To state the problem in greater detail, we are given a bisequence $(\underline{f}_n)_{n=-\infty}^{\infty}$ of q-variates

$$(1.3) \quad \underline{f}_n = (f_n^1, \dots, f_n^q), \text{ where } f_n^i \in L_2(\Omega, \mathcal{B}, P),$$

such that the $q \times q$ covariance matrix

$$(1.4) \quad [E(f_m^i \cdot \bar{f}_n^j)] = [\gamma_{m-n}^{(i,j)}] = \underline{\Gamma}_{m-n}$$

depends only on the difference $m-n$ ⁽²⁾. This is the hypothesis of weak stationarity. Now let $\mathfrak{m}_0$ be the (closed, linear) subspace of $L_2(\Omega, \mathcal{B}, P)$ spanned by the f_n^i , with $n \leq 0$ and $1 \leq i \leq q$; in symbols

$$(1.5) \quad \mathfrak{m}_0 = \mathfrak{G}(f_n^i, n < 0, 1 \leq i \leq q).$$

Then our problem may be stated as follows:

1.6 Prediction Problem. Assuming as known the covariance

bisequence $(\underline{\Gamma}_k)_{k=-\infty}^{\infty}$ and given $v \geq 1$, find variates $\hat{f}_v^1, \dots, \hat{f}_v^q \in \mathfrak{m}_0$ such that

$$E(|f_v^i - \hat{f}_v^i|^2) \leq E(|f_v^i - g|^2), \quad g \in \mathfrak{m}_0, \quad 1 \leq i \leq q.$$

Also find the prediction-error covariance matrix

$$\underline{G}_v = [E\{(f_v^i - \hat{f}_v^i)(\bar{f}_v^j - \bar{\hat{f}}_v^j)\}].$$

Now $L_2(\Omega, \mathcal{B}, P)$ is a Hilbert space with the inner product product $(f, g) = E(f\bar{g})$. Since our problem involves only second-order moments, we can restate it as one for a Hilbert space $\mathfrak{H}$, as Kolmogorov first emphasized in 1940, cf. [12]. To get the usual

probabilistic version of the theory we must of course think of this $\mathfrak{H}$ as being $L_2(\Omega, \mathfrak{B}, P)$. But for theoretical purposes it is best to leave $\mathfrak{H}$ unspecified. Adopting this point of view, what we have is a bisequence of vectors $(\underline{f}_n)_{n=-\infty}^{\infty}$ such that

$$\underline{f}_n = (f_n^1, \dots, f_n^q) \text{ where } f_n^i \in \mathfrak{H};$$

i.e. each $\underline{f}_n$ is in the Cartesian product $\mathfrak{H}^q$ of $\mathfrak{H}$ with itself q times. For q -variate prediction the structure of this hyperspace is crucial and must first engage our attention (§2).

2. The Gram matricial structure of $\mathfrak{H}^q$

Let $\mathfrak{H}$ be any (complex) Hilbert space, $q \geq 1$, and $\mathfrak{H}^q$ be the Cartesian product of $\mathfrak{H}$ with itself q times, i.e. the set of all vectors $\underline{f} = (f^1, \dots, f^q)$ such that each $f^i \in \mathfrak{H}$. To make $\mathfrak{H}^q$ serviceable in prediction theory we must endow it with a Gram matricial structure, as Doob [4, p. 594] noticed. For $\underline{f}, \underline{g} \in \mathfrak{H}^q$, the $q \times q$ matrix

$$(2.1) \quad (\underline{f}, \underline{g}) = [(f^i, g^j)]$$

is called the Gramian of the ordered pair $\underline{f}, \underline{g}$ ⁽³⁾. It is reasonable to think of it as a matricial inner-product in view of its properties ⁽⁴⁾:

$$(2.2) \quad (\underline{f}, \underline{f}) \geq 0; \quad (\underline{f}, \underline{f}) = 0 \implies \underline{f} = \underline{0};$$

$$(2.3) \quad \left(\sum_{j \in J} \underline{A}_j \underline{f}_j, \sum_{k \in K} \underline{B}_k \underline{g}_k \right) = \sum_{j \in J} \sum_{k \in K} \underline{A}_j (\underline{f}_j, \underline{g}_k) \underline{B}_k^*,$$

where J, K are finite sets and $\underline{A}_j, \underline{B}_k$ are $q \times q$ matrices.

This suggests defining orthogonality in $\mathfrak{H}^q$ by the relation:

$$(2.4) \quad \underline{f} \perp \underline{g} \iff (\underline{f}, \underline{g}) = 0 \quad (\text{i.e.} \iff f^i \perp g^j, \quad 1 \leq i, j \leq q.)$$

It also suggests taking linear combinations of $\underline{f}_j \in \mathfrak{H}^q$ with $q \times q$ matrix rather than complex coefficients, and calling a subset $\mathfrak{M}$ of $\mathfrak{H}^q$ a linear manifold if and only if

$$\underline{f}, \underline{g} \in \mathfrak{M} \implies \text{for all } q \times q \text{ matrices } \underline{A}, \underline{B}, \quad \underline{A}\underline{f} + \underline{B}\underline{g} \in \mathfrak{M}.$$

The appropriate topology for $\mathfrak{H}^q$ turns out, however, to be the familiar one induced by the (scalar) inner product in $\mathfrak{H}^q$:

$$(2.5) \quad ((\underline{f}, \underline{g})) = \text{trace}(\underline{f}, \underline{g}) = \sum_{i=1}^q f^i \overline{g^i},$$

or rather by the corresponding norm

$$(2.6) \quad |\underline{f}| = \sqrt{((f, f))} = \sqrt{\sum_{i=1}^q |f^i|^2}.$$

It is well known that $\mathfrak{H}^q$ is Hilbert space under the inner product (2.5).

We call $\mathfrak{M}$ a subspace of $\mathfrak{H}^q$, if and only if $\mathfrak{M}$ is both a linear manifold and a closed set. It is easy to check, cf. [36, I, 5.8], that

$$(2.7) \quad \mathfrak{M} \text{ is a subspace of } \mathfrak{H}^q \iff \mathfrak{M} = \mathfrak{M}^q, \text{ where } \mathfrak{M} \text{ is a subspace of } \mathfrak{H}.$$

With these concepts of orthogonality, distance and subspace we can extend to $\mathfrak{H}^q$ the well-known theory of orthogonal projections for Hilbert spaces. Thus we have, cf. [36, I, 5.8; II, 1.17],

2.8 Lemma. If $\underline{f} \in \mathfrak{H}^q$ and $\mathfrak{M}$ is a subspace of $\mathfrak{H}^q$, then there exists a unique $\hat{\underline{f}} \in \mathfrak{H}^q$ satisfying any one (and therefore both) of the

following equivalent conditions

$$(1) \quad \hat{f} \in \underline{m} \quad \& \quad \underline{f} - \hat{f} \perp \underline{m}$$

$$(2) \quad \hat{f} \in \underline{m} \quad \& \quad (\underline{f} - \hat{f}, \underline{f} - \hat{f}) \leq (\underline{f} - \underline{g}, \underline{f} - \underline{g}), \quad \underline{g} \in \underline{m} \quad .$$

Let $\underline{m} = \mathfrak{M}^q$ (cf. 2.7). Then the i^{th} component $\hat{f}^i$ of $\hat{f}$ is the (ordinary) orthogonal projection of the i^{th} component f^i of $\underline{f}$ on $\underline{m}$.

2.9 Def. The $\hat{f}$ mentioned in 2.8 is called the orthogonal projection of $\underline{f}$ on $\underline{m}$ and written $(\underline{f}|\underline{m})$.

We thus obtain a structure for $\mathfrak{H}^q$ which differs from but also closely resembles that of a Hilbert space, and which we shall therefore call "Hilbertian". In terms of this structure we can give a definition of a q -ple, weakly stationary SP, in which all side issues are purged and the essential idea brought to the forefront, cf. [37, §5]:

2.10 Def. A q -ple, weakly stationary SP is a bisequence $(\underline{f}_n)_{n=-\infty}^{\infty}$ such that each $\underline{f}_n \in \mathfrak{H}^q$ and the Gram matrix

$$(\underline{f}_m, \underline{f}_n) = \Gamma_{m-n}$$

depends only on $m-n$. $(\Gamma_k)_{k=-\infty}^{\infty}$ is called the covariance bisequence of the SP .

Associated with a q -ple weakly stationary SP $(\underline{f}_n)_{n=-\infty}^{\infty}$ are the present and past subspaces $\underline{m}_n, \mathfrak{m}_n$ ⁽⁵⁾:

$$(2.11) \quad \begin{cases} \underline{m}_n = \mathcal{G}(\underline{f}_k, k \leq n) \subseteq \mathbb{H}^q \\ \underline{m}_n = \mathcal{G}(\underline{f}_k^i, k \leq n, 1 \leq i \leq q) \subseteq \mathbb{H}, \end{cases}$$

and the terminal subspaces

$$(2.12) \quad \begin{cases} \underline{m}_\infty = \mathcal{G}(\underline{f}_k, \text{all } k), & \underline{m}_\infty = \mathcal{G}(\underline{f}_k^i, \text{all } k, 1 \leq i \leq q) \\ \underline{m}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \underline{m}_n, & \underline{m}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \underline{m}_n. \end{cases}$$

We easily find, cf. [36, I, 6.5], that

$$(2.13) \quad \underline{m}_n = \underline{m}_n^q \quad -\infty \leq n \leq \infty,$$

and obviously

$$(2.14) \quad \underline{m}_{-\infty} \subseteq \underline{m}_n \subseteq \underline{m}_{n+1} \subseteq \underline{m}_\infty.$$

In terms of these subspaces we can easily formulate the concept of determinism and tersely restate the Prediction Problem 1.6:

2.15 Def. We call the SP deterministic, non-deterministic, purely non-deterministic, according as

$$\underline{m}_{-\infty} = \underline{m}_\infty, \quad \underline{m}_{-\infty} \subset \underline{m}_\infty, \quad \underline{m}_{-\infty} = \{0\}.$$

2.16 Prediction Problem. Let $(\underline{f}_n)_{-\infty}^{\infty}$ be a q -ple, weakly stationary SP with covariance bisequence $(\underline{\Gamma}_k)_{-\infty}^{\infty}$ and let $\nu \geq 1$. Find

(i) the matrices $\underline{A}_k^{(n)}$ such that

$$\hat{\underline{f}}_\nu = (\underline{f}_\nu | \underline{m}_0) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \underline{A}_k^{(n)} \underline{f}_{-\nu-k},$$

$$(11) \quad \underline{G}_\nu = (\underline{f}_\nu - \hat{\underline{f}}_\nu, \underline{f} - \hat{\underline{f}}_\nu) .$$

$\underline{G}_\nu$ is called the prediction error matrix for lag ν . Following Zasuhiu [43] we call $\rho = \text{rank } G_1$ the rank of the SP $(\underline{f}_n)_{-\infty}^\infty$.

Obviously

$$(2.17) \quad \text{the SP is deterministic} \iff \rho = 0, \text{ i.e. } \underline{G}_1 = \underline{0} .$$

The deterministic case is only of peripheral interest to us, cf. §1.

Of much theoretical interest, though somewhat pathological, are the non-deterministic cases of degenerate rank $1 \leq \rho < q$. The really interesting case from a practical standpoint is that of full rank $\rho = q$, for which $\det G_1 > 0$. We note that since $\underline{G}_\nu \geq G_1$ for $\nu \geq 1$, we have

$$(2.18) \quad \rho = q \implies \det G_\nu > 0, \quad \nu \geq 1 .$$

3. Elementary solution of the Prediction Problem

Seemingly the easiest way to solve the Problem 2.16 is by an extension of the method of undetermined coefficients. This has been explained in [36, II, §2] , and it will suffice to indicate only a couple of steps. We may choose the $\underline{A}_k^{(n)}$ so that

$$(3.1) \quad \underline{f}_\nu - \sum_{k=0}^n \underline{A}_k^{(n)} \underline{f}_{-\nu-k} \perp \underline{f}_0, \underline{f}_{-1}, \dots, \underline{f}_{-n} ,$$

whence, in block matrix notation,

$$(3.2) \quad [\underline{A}_0^{(n)}, \dots, \underline{A}_n^{(n)}] \begin{bmatrix} \underline{\Gamma}_0 \cdots \underline{\Gamma}_n \\ \vdots \\ \underline{\Gamma}_{-n} \cdots \underline{\Gamma}_0 \end{bmatrix} = [\underline{\Gamma}_\nu, \dots, \underline{\Gamma}_{\nu+n}] .$$

It may be shown that the second $(n+1)q \times (n+1)q$ matrix on the left is invertible if and only if $\det \underline{G}_1 \neq 0$. Thus in the full rank case $p = q$, the coefficients $\underline{A}_k^{(n)}$ can be uniquely determined.

This method involves solving a system of linear equations. It would be feasible for so-called weakly (or wide sense) N-Markovian processes, i. e., cf. [4, pp. 90, 506], weakly stationary ones for which

$$(3.3) \quad \hat{\underline{f}}_{\nu} = (\underline{f}_{\nu} | \underline{G}(\underline{f}_{-k}, k \geq 0)) = (\underline{f}_{\nu} | \underline{G}(\underline{f}_{-k}, 0 \leq k \leq N)), \quad \nu \geq 1$$

and where, consequently, for a given $\nu \geq 1$ there is a fixed set of $N+1$ matrices $\underline{A}_0, \dots, \underline{A}_N$ such that

$$\hat{\underline{f}}_{\nu} = \underline{A}_0 \underline{f}_0 + \underline{A}_1 \underline{f}_{-1} + \dots + \underline{A}_N \underline{f}_{-N}.$$

One might even be able to shorten the computation by adapting for $q > 1$ the interesting devices suggested by Levinson [15, §3] for $q = 1$. But for other types of processes, as time flows and our data accumulate, we would like to let the n in (3.1) increase, and thereby utilize our additional data. This would mean solving larger and larger linear systems de novo, a procedure of questionable efficiency.

It was Wiener's belief that an efficacious computational procedure would emerge from a deeper analysis of the problem. We now turn to such analysis.

4. The shift operator and Wold-Zasuhin decomposition

Let $(\underline{f}_n)_{n=-\infty}^{\infty}$ be a q -ple, weakly stationary SP. Then as Kolmogorov [12] showed, there is a unique unitary operator U on $\mathfrak{M}_{\infty} \subseteq \mathfrak{H}$ onto $\mathfrak{M}_{\infty}$ such that

$$(4.1) \quad U(f_n^i) = f_{n+1}^i, \quad -\infty < n < \infty, \quad 1 \leq i \leq q. \quad (6)$$

U or rather its inflation $\underline{U}$, defined by

$$(4.2) \quad \underline{U}(\underline{f}) = (Uf^1, \dots, Uf^q), \quad \underline{f} = (f^i)_{i=1}^q \in \mathfrak{H}^q$$

is called the shift operator of the SP. Obviously $\underline{U}$ is an operator on $\mathfrak{M}_{\infty}$ onto $\mathfrak{M}_{\infty}$ such that

$$(4.3) \quad \underline{U}(\underline{f}_n) = \underline{f}_{n+1} \quad -\infty < n < \infty.$$

Now

$$U^*(\mathfrak{M}_n) = \mathfrak{M}_{n-1} \subseteq \mathfrak{M}_n.$$

Hence (7)

$$(4.4) \quad V \stackrel{\text{d}}{=} \text{Rstr.}_{\mathfrak{M}_0} U^* = \text{an isometry on } \mathfrak{M}_0 \text{ onto } \mathfrak{M}_{-1}.$$

The theory of this isometry V subsumes the time-domain analysis of our SP, as we shall now indicate.

Since the appearance of von Neumann and Murray's work on operators it has been known in some implicit form that if V is an isometry on a Hilbert space $\mathfrak{H}$ onto $R \subseteq \mathfrak{H}$, then

$$\mathfrak{H} = \bigcap_{k=0}^{\infty} V^k(\mathfrak{H}) + \sum_{k=0}^{\infty} V^k(R^{\perp}), \quad V^j(R^{\perp}) \perp V^k(R^{\perp})$$

where the two subspaces on the right side of the equality are themselves orthogonal. But the great importance of this result has emerged only recently, cf. Halmos [6]. As indicated in [22, 2.8] it extends to $\mathfrak{H}^q$: if $\underline{V}$ is the inflation to $\mathfrak{H}^q$ of an isometry on $\mathfrak{H}$ and $\underline{R} \subseteq \mathfrak{H}^q$ is the range of $\underline{V}$, then

$$(4.5) \quad \mathfrak{H}^q = \bigcap_{k=0}^{\infty} \underline{V}^k(\mathfrak{H}^q) + \sum_{k=0}^{\infty} \underline{V}^k(\underline{R}^\perp), \quad \underline{V}^i(\underline{R}^\perp) \perp \underline{V}^k(\underline{R}^\perp)$$

where the subspaces on the right side of the equality are again orthogonal but in the sense of (2.4).

Turning to our SP, and applying (4.5) with $\mathfrak{H} = \mathfrak{m}_0$ and V as in (4.4), we get at once

$$\mathfrak{m}_0 = \mathfrak{m}_{-\infty} + \sum_{k=0}^{\infty} \underline{U}^{-k}(\mathfrak{m}_{-1}^\perp \cap \mathfrak{m}_0) .$$

Now, and this is crucial, we can show that for a non-deterministic SP

$$\mathfrak{m}_{-1}^\perp \cap \mathfrak{m}_0 = \mathfrak{G}(\underline{g}_0), \quad \text{where } \underline{g}_0 = \underline{f}_0 - (\underline{f}_0 | \mathfrak{m}_{-1}) \neq 0 ,$$

which means roughly that $\mathfrak{R}^\perp = \mathfrak{m}_{-1}^\perp \cap \mathfrak{m}_0$ is "one-dimensional". Letting

$g_k = U^k g_0$, we readily obtain for any n

$$(4.6) \quad \mathfrak{m}_n = \mathfrak{m}_{-\infty} + \sum_{k=0}^{\infty} \mathfrak{G}(g_{n-k}), \quad \mathfrak{m}_{-\infty} \perp \sum_{k=-\infty}^{\infty} \mathfrak{G}(g_k) .$$

This is the Wold-Zasuhin decomposition of the subspace $\mathfrak{m}_n$ ⁽⁸⁾.

In this decomposition the vector

$$(4.7) \quad \underline{g}_n = \underline{U}_d^n \underline{g}_0, \quad \text{where } \underline{g}_0 = \underline{f}_0 - (\underline{f}_0 | \mathfrak{m}_{-1})$$

is called the n^{th} innovation vector of our SP, and $(\underline{g}_n)_{-\infty}^{\infty}$ is called its innovation SP. Obviously

$$(4.8) \quad (\underline{g}_m, \underline{g}_n) = \delta_{mn} \underline{G}, \quad \text{where} \quad \underline{G} = (\underline{g}_0, \underline{g}_0),$$

and since $\underline{U} \underline{g}_0 = \underline{f}_1 - (\underline{f}_1 | \underline{m}_0) \underline{m}_0 = \underline{f}_1 - \hat{\underline{f}}_1$, (cf. 2.16), we see that

$$(4.9) \quad \underline{G} = \underline{G}_1 = \text{the prediction error matrix for lag 1}.$$

It is convenient to "normalize" the innovation vectors. For this we think of the matrix $\underline{G}$ as a linear operator on $\mathbb{C}^q$ to $\mathbb{C}^q$, $\mathbb{C}$ being the complex number field. Let the matrix $\underline{J}$ represent the projection on $\mathbb{C}^q$ onto the range of $\underline{G}$. It is easy to show that there is a unique $q \times q$ matrix $\underline{H}$ such that

$$(4.10) \quad \underline{H} \underline{J}^\perp = \underline{J}^\perp = \underline{J}^\perp \underline{H}, \quad \underline{H} \sqrt{\underline{G}} = \underline{J} = \sqrt{\underline{G}} \cdot \underline{H}.$$

Indeed,

$$(4.11) \quad \underline{H} = (\sqrt{\underline{G}} + \underline{J}^\perp)^{-1}$$

which shows that

$$(4.12) \quad \underline{H} \text{ is invertible, hermitian and positive-definite.}$$

Now let $\underline{h}_n = \underline{H} \underline{g}_n$. Then we find, cf. [22, (3.4) et seq.],

$$(4.13) \quad \underline{g}_n = \underline{J} \underline{g}_n = \sqrt{\underline{G}} \underline{h}_n, \quad (\underline{h}_m, \underline{h}_n) = \delta_{mn} \underline{J}, \quad \underline{J}^\perp \underline{h}_n = \underline{0}.$$

We call $\underline{h}_n$ the n^{th} normalized innovation vector of our SP, and

$(\underline{h}_n)_{-\infty}^{\infty}$ its normalized innovation SP. In the full rank case $p = q$, we have $\det \underline{G} \neq 0$, and so we can define the $\underline{h}_n$ by the simple equation

$\underline{h}_n = (\sqrt{\underline{G}})^{-1} \underline{g}_n$. Since in this case $\underline{J} = \underline{I}$, we have $(\underline{h}_m, \underline{h}_n) = \delta_{mn} \underline{I}$.

As shown in [22, 3.2, 3.5] the decomposition (4.6) of the subspace $\underline{m}_n$ yields a decomposition the process $(\underline{f}_n)_{-\infty}^{\infty}$ itself:

$$(4.14) \quad \underline{f}_n = \underline{u}_n + \underline{v}_n, \quad \underline{u}_m \perp \underline{v}_n, \quad -\infty < m, n < \infty$$

where $(\underline{u}_n)_{-\infty}^{\infty}$ is a (purely non-deterministic) one-sided moving average of the normalized innovations and has the same rank as $(\underline{f}_n)_{-\infty}^{\infty}$, and $(\underline{v}_n)_{-\infty}^{\infty}$ is purely deterministic. More fully,

$$(4.15) \quad \underline{u}_n = \sum_{k=0}^{\infty} \underline{A}_k \sqrt{\underline{G}} \underline{h}_{-k}, \quad \sum_{k=0}^{\infty} |\underline{A}_k \sqrt{\underline{G}}|_E^2 < \infty, \quad (9)$$

where

$$(4.16) \quad \underline{A}_k \sqrt{\underline{G}} = (\underline{f}_0, \underline{h}_{-k}), \quad \underline{A}_0 \sqrt{\underline{G}} = \sqrt{\underline{G}}, \quad \underline{A}_0 \underline{g}_0 = \underline{g}_0, \quad \underline{A}_k \underline{G} = (\underline{f}_0, \underline{g}_{-k})$$

are unique, although the $\underline{A}_k$ are not unique. Also

$$(4.17) \quad \underline{v}_n = (\underline{f}_n | \underline{m}_{-\infty}).$$

The relations (4.14)-(4.17) constitute an alternative form of the Wold-Zasuhin decomposition.

It is well known that the conditions (4.14) and

$$(4.18) \quad \begin{cases} (\underline{u}_n)_{-\infty}^{\infty} \text{ is purely non-deterministic} \\ (\underline{v}_n)_{-\infty}^{\infty} \text{ is deterministic} \end{cases}$$

do not together characterize the Wold-Zasuhin decomposition. An extra condition is needed, which is usually stated (with an obvious notation) in the form

$$(4.19) \quad \underline{m}_n^{(u)} \subseteq \underline{m}_n^{(f)} \text{ for some integer } n \text{ (10)}.$$

Does (4.19) work with $n = \infty$? (11). Recently Robertson [26, App. B] has shown that the answer is in the negative for $q > 1$, but that the stronger condition

$$(4.20) \quad \underline{m}_\infty^{(u)} \subseteq \underline{m}_\infty^{(f)} \quad \& \quad \text{rank}(\underline{u}_n)_{-\infty}^\infty = \text{rank}(\underline{f}_n)_{-\infty}^\infty$$

does work for any q ; i.e. (4.14), (4.18) and (4.20) together characterize the Wold-Zasuhin decomposition. A recent result of Robertson [27] on the wandering subspaces of unitary operators yields a nice, spectral free, proof.

5. Spectral analysis

The shift operator U of our q -ple weakly stationary $SP(\underline{f}_n)_{-\infty}^\infty$ has a spectral resolution:

$$(5.1) \quad U = \int_0^{2\pi} e^{-i\theta} E(d\theta)$$

where E is a projection-valued measure over $([0, 2\pi], \mathcal{B})$, $\mathcal{B}$ being the family of Borel subsets of $[0, 2\pi]$. By taking the inflation $\underline{E}$ of E we associate two new measures with our SP :

(i) a $\mathbb{H}^q$ -valued countably-additive, orthogonally-scattered (c.a.o.s.) measure $\underline{\xi}$ defined by

$$(5.2) \quad \underline{\xi}(B) = \underline{E}(B) \underline{f}_0, \quad B \in \mathcal{B};$$

so-called, because of its decisive property

$$B, C \in \mathfrak{B} \text{ \& } B, C \text{ disjoint} \implies \underline{\xi}(B) \perp \underline{\xi}(C) ; \quad (12)$$

(ii) a $q \times q$ non-negative, hermitian matrix-valued measure $\underline{M}$

defined by

$$(5.3) \quad \underline{M}(B) = \left(\int_d \underline{E}(B) \underline{f}_0, \int_d \underline{E}(B) \underline{f}_0 \right) = (\underline{\xi}(B), \underline{\xi}(B)) \quad B \in \mathfrak{B}.$$

We then introduce the well known $q \times q$ spectral distribution $\underline{F}$ of our SP by the definition

$$(5.4) \quad \underline{F}(\theta) = \int_d 2\pi \underline{M}(0, \theta] \quad 0 \leq \theta \leq 2\pi.$$

Likewise one could define the q -ple process of orthogonal increments

$(\eta_\theta, \theta \in (0, 2\pi])$ associated with our process by

$$\eta_\theta = \int_d 2\pi \underline{\xi}(0, \theta] \quad 0 \leq \theta \leq 2\pi.$$

Next, for a complex-valued function ϕ on $[0, 2\pi]$ we define the integrals

$$\int_0^{2\pi} \phi(\theta) \underline{\xi}(d\theta), \int_0^{2\pi} \phi(\theta) \underline{M}(d\theta), \int_0^{2\pi} \phi(\theta) d\underline{F}(\theta)$$

to be

$$\left(\int_0^{2\pi} \phi(\theta) \xi^i(d\theta) \right)_{i=1}^q, \left[\int_0^{2\pi} \phi(\theta) M_{ij}(d\theta) \right], \left[\int_0^{2\pi} \phi(\theta) dF_{ij}(\theta) \right].$$

These definitions make sense, since the components ξ^i of $\underline{\xi}$ are $\mathbb{H}$ -valued c.a.o.s. measures for which a theory of integration akin to that given in Doob's book [4, Ch. IX, §2] is available, and the entries M_{ij}, F_{ij} of $\underline{M}, \underline{F}$ are

complex-valued measures, and complex-valued functions of bounded variation. With these definitions we easily get the spectral representation of our SP and of its covariance (cf. [36, I, 7.1]):

$$(5.5) \quad \underline{f}_n = \int_0^{2\pi} e^{-ni\theta} \underline{E}(d\theta) \underline{f}_0 = \int_0^{2\pi} e^{-ni\theta} \underline{\xi}(d\theta)$$

$$(5.6) \quad \underline{\Gamma}_n = \int_0^{2\pi} e^{-ni\theta} \underline{M}(d\theta) = \frac{1}{2\pi} \int_0^{2\pi} e^{-ni\theta} d\underline{F}(\theta) .$$

Finally, we define matricial Riemann Stieltjes integrals of the form

$\int_0^{2\pi} \underline{\Phi}(\theta) d\underline{F}(\theta) \underline{\Psi}(\theta)$, where $\underline{\Phi}, \underline{\Psi}$ are continuous matrix-valued functions, and $\underline{F}$ a matrix-valued function of bounded variation, by adopting the classical pattern, cf. [36, I, §4]. From (5.6) we then get

$$(5.7) \quad \left(\sum_{j \in J} \underline{A}_j \underline{f}_{-j}, \sum_{k \in K} \underline{B}_k \underline{f}_{-k} \right) = \frac{1}{2\pi} \int_0^{2\pi} \left(\sum_{j \in J} \underline{A}_j e^{ji\theta} \right) d\underline{F}(\theta) \left(\sum_{k \in K} \underline{B}_k e^{ki\theta} \right)^*$$

where J, K are finite sets of integers and $\underline{A}_j, \underline{B}_k$ are $q \times q$ matrices, cf. [36, I, 7.9(a)].

It is natural to ask if the equality (5.7) continues to hold when limits of linear combinations and of trigonometrical polynomials are taken on the two sides.

This raises the preliminary question as to how $\int_0^{2\pi} \underline{\Phi}(\theta) d\underline{F}(\theta) \underline{\Psi}(\theta)$ or equivalently $\int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Psi}(\theta)$ is to be defined when $\underline{\Phi}, \underline{\Psi}$ are any (discontinuous) matrix-valued functions on $[0, 2\pi]$ with Borel measurable entries. We can pose this question for any non-negative, hermitian matrix-valued measure $\underline{M}$, not just

the one defined in (5.3). The further development of the spectral theory of q -ple processes hinges on the answer (§6).

6. The space $L_{2,M}$ for a non-negative hermitian matrix-valued measure.

Let $\underline{M}$ be any $q \times q$ non-negative, hermitian matrix-valued measure over $([0, 2\pi], \mathfrak{B})$ and suppose that we have in some way defined the integrals

$$(1) \quad \int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Psi}(\theta)$$

for $q \times q$ matrix-valued functions $\underline{\Phi}, \underline{\Psi}$ with Borel measurable entries.

It would then be natural to define the L_2 class with respect to the measure $\underline{M}$ by

$$(6.1) \quad L_{2,M} = L_2([0, 2\pi], \mathfrak{B}, \underline{M}) \stackrel{d}{=} \{ \underline{\Phi} : \int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Phi}^*(\theta) \text{ exists} \}.$$

Now one of the fundamental properties of the class $L_{2,M}$ when $q = 1$, i.e. when Φ, M are complex-valued, is its completeness under the norm

$$|\Phi|_M = \sqrt{\int_0^{2\pi} |\Phi(\theta)|^2 M(d\theta)}.$$

This is the core of the celebrated Riesz-Fischer Theorem. For $q > 1$, the corresponding norm would appear to be

$$(6.2) \quad |\underline{\Phi}|_M = \sqrt{\text{trace} \int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Phi}^*(\theta)}.$$

Our definition of the integral (1) would be useless, were the space $\underline{L}_{2,\underline{M}}$ defined in (6.1) to be incomplete under the norm (6.2). We are thus faced with the following problem:

Problem. Define the integrals (1) in such a way that the space $\underline{L}_{2,\underline{M}}$ defined in (6.1) is complete under the norm (6.2).

This problem was settled by M. Rosenberg [32, §3] for rectangular matrices $\underline{\Phi}, \underline{\Psi}$ and by Yu. A. Rosanov [31, Ch. I, §7] for vectorial $\underline{\Phi}, \underline{\Psi}$, independently around 1963. We shall follow Rosenberg's more inclusive treatment. He observed that a $q \times q$ non-negative, hermitian matrix-valued measure $\underline{M}$ is invariably absolutely continuous with respect to the non-negative, real measure trace $\underline{M}$. Writing $\tau \underline{M}$ for trace $\underline{M}$, it follows that each entry of $\underline{M}$ has a Radon-Nikodym derivative with respect to $\tau \underline{M}$. The $q \times q$ matrix $d\underline{M}/d\tau \underline{M}$ of these derivatives has nice properties, and this suggests adoption of the definition

$$(6.3) \quad \int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Psi}(\theta) = \int_0^{2\pi} \underline{\Phi}(\theta) \frac{d\underline{M}}{d\tau \underline{M}}(\theta) \underline{\Psi}(\theta) \cdot \tau \underline{M}(d\theta),$$

the last integral being defined (earlier) as the matrix of Lebesgue integrals of the entries of $\underline{\Phi}(d\underline{M}/d\tau \underline{M}) \underline{\Psi}$ with respect to the ordinary measure $\tau \underline{M}$. Rosenberg showed that this definition solves the problem. Thus

6.4. Thm. (Rosenberg-Rosanov) With the definitions 6.3 and 6.1 the space $\underline{L}_{2,\underline{M}}$ is complete under the norm (6.2). ⁽¹³⁾

In case the measure $\underline{M}$ is absolutely continuous with respect to Lebesgue measure L , it follows at once from simple properties of Radon-Nikodym derivatives that (6.3) is equivalent to the simpler definition

$$(6.5) \quad \int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Psi}(\theta) = \frac{1}{2\pi} \int_0^{2\pi} \underline{\Phi}(\theta) \underline{F}'(\theta) \underline{\Psi}(\theta) d\theta, \quad$$

where $\underline{F}$ is as in (5.4). The work of Rosenberg and Rosanov thus subsumes the partial results obtained previously on the basis of (6.5), e.g. those in [36, II, §4].

Having defined the integrals (1), we can introduce in $\underline{L}_{2,\underline{M}}$ matrix- and complex-valued inner products by the definitions:

$$(6.6) \quad (\underline{\Phi}, \underline{\Psi})_{\underline{M}} = \int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Psi}^*(\theta), \quad \underline{\Phi}, \underline{\Psi} \in \underline{L}_{2,\underline{M}}$$

$$(6.7) \quad ((\underline{\Phi}, \underline{\Psi}))_{\underline{M}} = \text{trace } (\underline{\Phi}, \underline{\Psi})_{\underline{M}}.$$

The norm introduced in (6.2) can then be written

$$(6.8) \quad |\underline{\Phi}|_{\underline{M}} = \sqrt{((\underline{\Phi}, \underline{\Phi}))_{\underline{M}}}.$$

The equations (6.6)-(6.8) are comparable to (2.1), (2.5), (2.6).

The fact that $\underline{L}_{2,\underline{M}}$ is complete under the norm (6.8) thus shows that $\underline{L}_{2,\underline{M}}$ is Hilbertian in the sense of §2. Thus every non-negative, hermitian matrix-valued measure $\underline{M}$ generates a Hilbertian space $\underline{L}_{2,\underline{M}}$.

This result holds in particular for the measure $\underline{M}$, defined in (5.3), which is associated with the SP $(\underline{f}_n)_{-\infty}^{\infty}$. Thus every q -ple, weakly

stationary SP possesses two Hilbertian spaces $\mathbb{H}_\infty \subseteq \mathbb{H}^q$ and $\underline{L}_{2, \underline{M}}$.

We shall refer to them as the temporal and spectral spaces of the SP.

For $q = 1$ we know that they are isomorphic Hilbert spaces under a natural correspondence, cf. [12, (2.7)]. This isomorphism survives when $q > 1$ (§7).

7. Isomorphism between the temporal and spectral spaces

Let $\underline{\xi}$ be any $\mathbb{H}^q$ -valued, c.a.o.s. measure over $([0, 2\pi], \mathcal{B})$, cf. §5, and let

$$\underline{M}_{\underline{\xi}}(B) = (\underline{\xi}(B), \underline{\xi}(B)), \quad B \in \mathcal{B}.$$

Then $\underline{M}_{\underline{\xi}}$ is clearly a $q \times q$ non-negative, hermitian matrix-valued measure.

The crucial fact that the associated space $\underline{L}_{2, \underline{M}_{\underline{\xi}}}$ has a (complete) Hilbertian structure enables us to define integrals of the type $\int_0^{2\pi} \underline{\Phi}(\theta) \underline{\xi}(d\theta)$, where $\underline{\Phi} \in \underline{L}_{2, \underline{M}_{\underline{\xi}}}$, by following essentially the procedure adopted for stochastic integration in Doob's book [4, Ch. IX, §2], and to prove the following theorem (for details, cf. Rosenberg [32, §4]):

7.1 Thm. Let (i) $\underline{\xi}$ be any $\mathbb{H}^q$ -valued, c.a.o.s. measure over $([0, 2\pi], \mathcal{B})$, (ii) $\underline{M}_{\underline{\xi}}(B) \stackrel{d}{=} (\underline{\xi}(B), \underline{\xi}(B))$, $B \in \mathcal{B}$, (iii) $\underline{g}^{(\underline{\xi})} \stackrel{d}{=} \bigcup \{\underline{\xi}(B) : B \in \mathcal{B}\} \subseteq \mathbb{H}^q$ (¹⁴). Then

$$(a) \quad \underline{g} \in \underline{g}^{(\underline{\xi})} \iff \exists \underline{\Phi} \in \underline{L}_{2, \underline{M}_{\underline{\xi}}} \text{ such that } \underline{g} = \int_0^{2\pi} \underline{\Phi}(\theta) \underline{\xi}(d\theta)$$

(b) the $\underline{\Phi}$ in (a) is uniquely defined up to a set of zero $\underline{M}_{\underline{\xi}}$ measure

(c) the correspondence $\underline{\Phi} \rightarrow \int_0^{2\pi} \underline{\Phi}(\theta) \underline{\xi}(d\theta)$ is an isomorphism on $\underline{L}_{2, \underline{M}_{\underline{\xi}}}$ onto the subspace $\underline{g}^{(\underline{\xi})}$ of $\mathbb{H}^q$; i.e. it is one-one on $\underline{L}_{2, \underline{M}_{\underline{\xi}}}$ onto $\underline{g}^{(\underline{\xi})}$ and, cf. (6.6),

$$(\underline{\Phi}, \underline{\Psi})_{\underline{M}_{\underline{\xi}}} = \left(\int_0^{2\pi} \underline{\Phi}(\theta) \underline{\xi}(d\theta), \int_0^{2\pi} \underline{\Psi}(\theta) \underline{\xi}(d\theta) \right), \quad \underline{\Phi}, \underline{\Psi} \in \underline{L}_{2, \underline{M}_{\underline{\xi}}}.$$

Each $\mathfrak{H}^q$ -valued, c.a.o.s. measure thus carries with it two isomorphic Hilbertian spaces $\underline{g}^{(\xi)}$ and $\underline{L}_{2, \underline{M}_\xi}$. This applies in particular to the c.a.o.s. measure, defined in (5.2), which is associated with the SP $(\underline{f}_n)_{-\infty}^\infty$. But now $\underline{\xi}(B) = \underline{E}(B) \underline{f}_0$, and so we have

$$(7.2) \quad \underline{g}^{(\xi)} = \mathfrak{G}\{\underline{E}(B) \underline{f}_0 : B \in \mathfrak{B}\} = \mathfrak{G}\{\underline{U}^n \underline{f}_0 : -\infty < n < \infty\} = \underline{m}_\infty.$$

Here the first and third equalities are obvious, and the second stems from the basic connection between $\underline{U}$ and $\underline{E}$ given in (5.1), as is well known from the theory of cyclic subspaces of Hilbert spaces. We thus get as a corollary of 7.1 the result:

7.3 Thm. For a q -ple, weakly stationary SP $(\underline{f}_n)_{-\infty}^\infty$, the correspondence $\underline{\Phi} \rightarrow \int_0^{2\pi} \underline{\Phi}(\theta) \underline{E}(d\theta) \underline{f}_0$ is an isomorphism on the spectral space $\underline{L}_{2, \underline{M}}$ onto the temporal space $\underline{m}_\infty \subseteq \mathfrak{H}^q$.

This theorem shows of course that the equality (5.7) holds when limits are taken in the $\mathfrak{H}^q$ and $\underline{L}_{2, \underline{M}}$ topologies on the two sides.

8. Cross-covariance and spectra. Subordination

To treat simultaneously two or more q -ple, weakly stationary SP's $(\underline{f}_n)_{-\infty}^\infty, (\underline{g}_n)_{-\infty}^\infty, \dots$, it is convenient to use subscripts or superscripts $f, g, \dots$ to distinguish their possessions, e.g. to denote their temporal spaces by $\underline{m}_\infty^{(f)}, \underline{m}_\infty^{(g)}, \dots$. The processes $(\underline{f}_n)_{-\infty}^\infty, (\underline{g}_n)_{-\infty}^\infty$ are said

to be stationarily cross-correlated, if and only if the Gram matrix

$$(8.1) \quad (\underline{f}_m, \underline{g}_n) = \Gamma_{m-n}^{(f,g)}$$

depends on $m-n$ alone. The bisequence $(\Gamma_k^{(f,g)})_{k=-\infty}^{\infty}$ is then called the cross-covariance of $(\underline{f}_n)_{-\infty}^{\infty}$ with $(\underline{g}_n)_{-\infty}^{\infty}$. Obviously $\Gamma_k^{(f,f)}$ is the Γ_k introduced in (2.10).

By a slight extension of an argument of Kolmogorov [12, Thm. 1] it follows that if $(\underline{f}_n)_{-\infty}^{\infty}$, $(\underline{g}_n)_{-\infty}^{\infty}$ are stationarily cross-correlated, then there is a unique unitary operator U on the subspace $\text{clos. } \{m_{\infty}^{(f)} + m_{\infty}^{(g)}\}$ onto itself such that $U f_n^i = f_{n+1}^i$ and $U g_n^i = g_{n+1}^i$, $1 \leq i \leq q$. In dealing simultaneously with several such processes it is therefore legitimate to start out with a single shift operator U on $\mathfrak{H}$ to $\mathfrak{H}$:

$$(8.2) \quad U = \int_0^{2\pi} e^{-i\theta} E(d\theta)$$

and to suppose that our SP's are of the form $(\underline{U}^n \underline{f})_{-\infty}^{\infty}$, $(\underline{U}^n \underline{g})_{-\infty}^{\infty}$, $\dots$, where $\underline{f}, \underline{g}, \dots \in \mathfrak{H}^q$, and $\underline{U}$ is the inflation of U .

With each ordered pair of $\underline{f}, \underline{g} \in \mathfrak{H}^q$ we associate the $q \times q$ matrix-valued cross-measure $\underline{M}_{fg}$, no longer hermitian-valued, and the $q \times q$ cross-spectral distribution $\underline{F}_{fg}$ defined by

$$(8.3) \quad \underline{M}_{fg}(B) = (\underline{E}(B) \underline{f}, \underline{E}(B) \underline{g}) \quad B \in \mathfrak{B}$$

$$(8.4) \quad \underline{F}_{fg}(\theta) = \underline{M}_{fg}(0, \theta], \quad 0 \leq \theta \leq 2\pi.$$

Obviously

$$\underline{M}_{gf}(B) = \underline{M}_{fg}^*(B), \quad B \in \mathcal{B}$$

and

$$(8.5) \quad \Gamma_n^{(f,g)} = \int_0^{2\pi} e^{-ni\theta} \underline{M}_{fg}(d\theta) = \frac{1}{2\pi} \int_0^{2\pi} e^{-ni\theta} d\underline{F}_{fg}(\theta) .$$

With $\underline{g} = \underline{f}$, the $\underline{M}_{ff}$, $\underline{F}_{ff}$, $\Gamma_n^{(f,f)}$ we get are of course the $\underline{M}$, $\underline{F}$, Γ_n introduced for the SP $(\underline{f}_n)_{-\infty}^{\infty}$ in (5.3), (5.4), (5.6).

We can define integrals of the form $\int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Psi}(\theta)$, where $\underline{M}$ is any (non-hermitian) matrix-valued measure, by a slight extension of (6.3):

$$(8.6) \quad \int_0^{2\pi} \underline{\Phi}(\theta) \underline{M}(d\theta) \underline{\Psi}(\theta) = \int_0^{2\pi} \underline{\Phi}(\theta) \frac{d\underline{M}}{d\sigma}(\theta) \underline{\Psi}(\theta) \sigma(d\theta)$$

where σ is not necessarily $\tau \underline{M}$ but some non-negative real measure with respect to which $\underline{M}$ is absolutely continuous. We can show that the definition does not depend on the choice of σ . We can then prove the following basic result by using a slightly extended version of the operational calculus: ⁽¹⁵⁾

8.7 Thm. Let (i) $\underline{f}, \underline{g} \in \mathcal{H}^q$, (ii) $\hat{\underline{g}} = (\underline{g}|_{\mathcal{M}_{\infty}}^{(f)})$, (iii) $\underline{\Phi}_{\hat{\underline{g}}} \in \underline{L}_{2, \underline{M}_{ff}}$ be the isomorph of $\hat{\underline{g}}$, i.e.

$$\hat{\underline{g}} = (\underline{g}|_{\mathcal{M}_{\infty}}^{(f)}) = \int_0^{2\pi} \underline{\Phi}_{\hat{\underline{g}}}(\theta) \underline{E}(d\theta) \underline{f}_0 .$$

Then

$$(a) \quad \underline{M}_{gf}(B) = \underline{M}_{\hat{\underline{g}}f}(B) = \int_B \underline{\Phi}_{\hat{\underline{g}}}(\theta) \underline{M}_{ff}(d\theta), \quad B \in \mathcal{B}$$

$$(b) \quad \underline{M}_{\hat{\underline{g}}\hat{\underline{g}}}(B) = \int_B \underline{\Phi}_{\hat{\underline{g}}}(\theta) \underline{M}_{ff}(d\theta) \underline{\Phi}_{\hat{\underline{g}}}^*(\theta) = \int_B \underline{\Phi}_{\hat{\underline{g}}}(\theta) \underline{M}_{f\hat{\underline{g}}}(d\theta), \quad B \in \mathcal{B} .$$

Remark. 8.7(a) suggests that $\Phi_{\hat{g}}$ is in some sense the Radon-Nikodym derivative of $\underline{M}_{gf}$ with respect to $\underline{M}_{ff}$. But a theory of such derivatives for matricial measures has not been developed so far, and it would be premature to write

$$\hat{g} = (g | \underline{m}_{\infty}^{(f)}) = \int_0^{2\pi} \frac{d\underline{M}_{gf}}{d\underline{M}_{ff}}(\theta) \underline{E}(d\theta) \underline{f},$$

when $q > 1$, even though this equality prevails for $q = 1$.

Following Kolmogorov [12, §4] we say that the $SP(\underline{U}^n \underline{g})_{-\infty}^{\infty}$ is subordinate to the $SP(\underline{U}^n \underline{f})_{-\infty}^{\infty}$, briefly, $\underline{g}$ is subordinate to $\underline{f}$, if and only if $\underline{m}_{\infty}^{(g)} \subseteq \underline{m}_{\infty}^{(f)}$. The last theorem then yields the following extension of Kolmogorov's Thms. 8, 9 in [12]:

8.8 Cor. The following conditions are equivalent:

- (i) $\underline{g}$ is subordinate to $\underline{f}$, i.e. $\underline{m}_{\infty}^{(g)} \subseteq \underline{m}_{\infty}^{(f)}$
- (ii) $\exists \Phi \in L_{2, \underline{M}_{ff}}$ such that $\underline{g} = \int_0^{2\pi} \Phi(\theta) \underline{E}(d\theta) \underline{f}$
- (iii) $\exists \Phi \in L_{2, \underline{M}_{ff}}$ such that for any $B \in \mathcal{B}$

$$\underline{M}_{gf}(B) = \int_B \Phi(\theta) \underline{M}_{ff}(d\theta), \quad \underline{M}_{gg}(B) = \int_B \Phi(\theta) \underline{M}_{ff}(d\theta) \Phi^*(\theta).$$

In case $\underline{g}$ is subordinate to $\underline{f}$, the functions Φ in (ii) and (iii) are equal a.e. ($\underline{M}_{ff}$).

The following is M. Rosenberg's unpublished generalization of Kolmogorov's Thm. 10 in [12]:

8.9. Cor. Let $\underline{g}$ be subordinate to $\underline{f}$, and Φ be as in 8.8(ii). Then $\underline{f}$ is subordinate to $\underline{g}$ (i.e. $\underline{f}, \underline{g}$ are mutually subordinate), if and only if

$$\text{rank} \left\{ \Phi(\theta) \frac{d\underline{M}_{ff}}{d\tau \underline{M}_{ff}}(\theta) \Phi^*(\theta) \right\} = \text{rank} \frac{d\underline{M}_{ff}}{d\tau \underline{M}_{ff}}(\theta), \quad \text{a.e. } (\tau \underline{M}_{ff}).$$

Thm. 8.7 has many applications. For instance, we can derive from it the spectral distribution of the projection of a component of a q -ple SP on the orthogonal complement of the space spanned by the rest of its components. First by taking Besicovitch derivatives [1] of our matricial measures in 8.7 with respect to Lebesgue measure L we can show that ⁽¹⁶⁾

$$(8.10) \quad (\det \underline{F}'_{ff}) \underline{F}'_{\hat{g}\hat{g}} = \underline{F}'_{\hat{g}f} (\text{adj } \underline{F}'_{ff}) \underline{F}'_{f\hat{g}}, \quad \text{a.e. } (L)$$

whether or not the functions $\underline{F}_{ff}$, $\underline{F}_{\hat{g}\hat{g}}$, $\underline{F}_{f\hat{g}}$ are absolutely continuous on $[0, 2\pi]$. From (8.9) we can in turn deduce the following

8.11 Lemma. Let (i) $(\underline{f}_n)_{n=-\infty}^{\infty}$ be a q -ple, weakly stationary SP, and Δ_q be the determinant of the derivative of its spectral distribution, (ii) Δ_{q-1} be defined similarly for the $(q-1)$ -ple SP formed by the first $q-1$ components of $(\underline{f}_n)_{n=-\infty}^{\infty}$, (iii) $\tilde{f}_n^{(q)}$ be the projection of the last component of $\underline{f}_n$ on $\{\mathfrak{E}(f_m^i, -\infty < m < \infty, 1 \leq i \leq q-1)\}^{\perp}$. Then the (real-valued) spectral distribution F_q of the (1) -ple SP $(\tilde{f}_n^{(q)})_{n=-\infty}^{\infty}$ satisfies the equation

$$\Delta_{q-1}(\theta) \cdot F'_q(\theta) = \Delta_q(\theta), \quad \text{a.e. (Leb.)}.$$

In case $\Delta_{q-1} > 0$ a.e. (Leb), we have of course

$$F'_q(\theta) = \Delta_q(\theta) / \Delta_{q-1}(\theta), \quad \text{a.e. (Leb.)}.$$

This result was obtained by Matveev in 1960, cf [23, p. 35, (5)], in the course of deriving spectral conditions for the determinism of a q -ple SP. With $q = 2$ it reappears in 1964 in a paper of L. H. Koopmans [13, Thm. 1], who seems to have been unaware of Matveev's work. Indeed, many of Koopmans' results on coherence of processes [13, 14] are deducible from those given above and standard theorems on Besicovitch derivatives [1].

9. Spectral analysis of a purely non-deterministic SP

It is easy to show, cf. [36, I, 7.7(a)], that if $(f_n)_{n=-\infty}^{\infty}$ is a moving average SP, i.e.

$$(9.1) \quad f_n = \sum_{k=-\infty}^{\infty} \underline{C}_k h_{n-k}, \quad (h_m, h_n) = \delta_{mn} I, \quad \sum_{k=-\infty}^{\infty} |\underline{C}_k J|_E^2 < \infty$$

where J is a projection matrix, then its spectral distribution $\underline{F}$ is absolutely continuous and

$$(9.2) \quad \underline{F}'(\theta) = \underline{\Psi}(e^{i\theta}) \underline{\Psi}(e^{i\theta})^*, \text{ a.e. (Leb.)}, \text{ where } \underline{\Psi}(e^{i\theta}) = \sum_{k=-\infty}^{\infty} \underline{C}_k J e^{ki\theta}.$$

The inequality in (9.1) shows that each entry of $\underline{\Psi}$ is in L_2 on the unit circle $C = \{|z| = 1\}$ of the complex plane, a fact we shall express by writing $\underline{\Psi} \in \underline{L}_2(C)$.

Now let $(f_n)_{n=-\infty}^{\infty}$ be any non-deterministic process. Then, as emphasized in §4, the coefficients $\underline{A}_k \sqrt{\underline{G}}$, which occur in its Wold-Zasuhin decomposition:

$$(9.3) \quad \begin{cases} f_n = u_n + v_n = \sum_{k=0}^{\infty} \underline{A}_k \sqrt{\underline{G}} h_{n-k} + (f_n | m_{-\infty}) , \\ \sum_{k=0}^{\infty} |\underline{A}_k \sqrt{\underline{G}}|_E^2 < \infty , \end{cases}$$

are uniquely determined by the SP. This fact and (9.2) suggest associating with our SP the function $\underline{\Phi}$ defined by

$$(9.4) \quad \underline{\Phi}(e^{i\theta}) = \sum_{k=0}^{\infty} \underline{A}_k \sqrt{\underline{G}} e^{ki\theta}.$$

We call $\underline{\Phi}$ the generating function of $(\underline{f}_n)_{-\infty}^{\infty}$. It plays a vital role in the theory. The inequality in (9.3) shows that $\underline{\Phi} \in \underline{L}_2(C)$. But the Fourier series of $\underline{\Phi}$ is devoid of negative frequency terms, and actually $\underline{\Phi} \in \underline{L}_2^{0+}(C)$, where

$$(9.5) \quad \underline{L}_2^{0+}(C) = \{ \underline{\Psi} : \underline{\Psi} \in \underline{L}_2(C) \text{ \& } \int_0^{2\pi} e^{-ki\theta} \underline{\Psi}(e^{i\theta}) d\theta = 0 \text{ for } k \leq 0 \}.$$

From (9.2) we immediately get:

9.6 Thm. The purely non-deterministic part $(\underline{u}_n)_{-\infty}^{\infty}$ in the Wold-Zasuhin decomposition of $(\underline{f}_n)_{-\infty}^{\infty}$ has an absolutely continuous spectral distribution $\underline{F}_{-u}$ such that

$$\underline{F}'_u(\theta) = \underline{\Phi}(e^{i\theta}) \underline{\Phi}^*(e^{i\theta}) \quad \text{a.e. (Leb.)}$$

where $\underline{\Phi} \in \underline{L}_2^{0+}(C)$ is the generating function of $(\underline{f}_n)_{-\infty}^{\infty}$, and $\underline{\Phi}^*(\cdot) = \{\underline{\Phi}(\cdot)\}^*$.

In case $(\underline{f}_n)_{-\infty}^{\infty}$ is itself purely non-deterministic, we have $\underline{v}_n = 0$, $\underline{u}_n = \underline{f}_n$, $\underline{F}_u = \underline{F}$, and it follows from 9.6 that $\underline{F}$ itself is absolutely continuous and $\underline{F}' = \underline{\Phi} \underline{\Phi}^*$ a.e., where $\underline{\Phi} \in \underline{L}_2^{0+}(C)$. The converse also holds as Rosanov [29] has shown, cf. also [37, 2.3]. We thus get the following spectral characterization of purely non-deterministic SP's :

9.7 Thm. (Rosanov). A q-ple SP is purely non-deterministic, if and only if its spectral distribution $\underline{F}$ is absolutely continuous and $\underline{F}'$ admits a factorization

$$\underline{F}'(\theta) = \underline{\Psi}(e^{i\theta}) \underline{\Psi}^*(e^{i\theta}) \quad \text{a.e. (Leb.)}, \text{ where } \underline{\Psi} \in \underline{L}_2^{0+}(C).$$

A function $\underline{\Psi}$ in $\underline{L}_2^{0+}(C)$ admits a holomorphic extension to the inner disk $D_+ = \{ |z| < 1 \}$ and its adjoint $\underline{\Psi}^*$ a holomorphic extension to the outer

disk $D_- = \{1 < |z| \leq \infty\}$. Thm. 9.7 thus reveals an interesting connection between q -ple prediction and the inner-outer factorization of $q \times q$ matrix-valued functions in $L_1(C)$.

10. Spectral analysis of a full-rank SP

Let $\underline{F}$ be the spectral distribution of a q -ple, weakly stationary SP $(\underline{f}_n)_{n=-\infty}^{\infty}$, and $\underline{G}$ be its prediction error matrix for lag 1, cf. (4.8), (4.9). Then, cf. [36, I, 7.10; or 8, I, Thm. 8],

$$(10.1) \quad \det \underline{G} = \exp \left\{ \frac{1}{2\pi} \int_0^{2\pi} \log \det \underline{F}'(\theta) d\theta \right\}.$$

This fundamental equality, first stated by Whittle [34] in 1953, is a determinantal extension of the Szego-Kolmogorov identity [12, (8.44)] for $q = 1$, and shows at once that

$$(10.2) \quad \rho = \underset{d}{\text{rank of SP}} = q \iff \log \det \underline{F}(\cdot) \in L_1[0, 2\pi].$$

We thus have a perfect spectral characterization of the full-rank case. A less obvious consequence of (10.1) is the following important result

[36, I, 7.11]:

10.3. Thm. Let (i) $(\underline{f}_n)_{n=-\infty}^{\infty}$ be a q -ple, weakly stationary SP of full rank q , (ii) $\underline{f}_n = \underline{u}_n + \underline{v}_n$ be its Wold decomposition (4.14 et seq), (iii) $\underline{F}, \underline{F}_u, \underline{F}_v$ be the spectral distributions of the processes $(\underline{f}_n)_{n=-\infty}^{\infty}, (\underline{u}_n)_{n=-\infty}^{\infty}, (\underline{v}_n)_{n=-\infty}^{\infty}$, (iv) $\underline{F}_a, \underline{F}_b$ be the absolutely continuous, and non-absolutely continuous parts of $\underline{F}$ ⁽¹⁷⁾. Then

$$\underline{F}_u = \underline{F}_a \quad \text{and} \quad \underline{F}_v = \underline{F}_b \quad .$$

We may paraphrase this result by saying that in the full rank case there is concordance between the Wold-Zasuhin decomposition in the time-domain and the Lebesgue-Cramer decomposition in the spectral domain.

On combining 10.3 and 9.6 we get another important result on full-rank processes. Since $\underline{F}_u = \underline{F}_a$ and $\underline{F}'_b = 0$, we have $\underline{F}' = \underline{F}'_u$, whence:

10.4 Thm. The derivative $\underline{F}'$ of the spectral distribution of any full-rank SP admits the factorization

$$\underline{F}'(\theta) = \underline{\Phi}(e^{i\theta}) \underline{\Phi}^*(e^{i\theta}), \quad \text{a.e. (Leb.)} \quad ,$$

where $\underline{\Phi} \in \underline{L}_2^{0+}(C)$ is the generating function (cf. 9.4).

The results (10.2), 10.3, 10.4 shed much light on the full-rank case $p = q$, and it is natural to ask whether corresponding results are available for $0 \leq p < q$. We have three questions:

- Q. 1. Given $0 \leq p < q$, what is the spectral n. & s.c. that a SP have the rank p ?
- Q. 2. For a SP of rank p such that $0 \leq p < q$, are the Wold-Zasuhin and Lebesgue-Cramer decompositions in concordance? If not, what extra condition would restore this concordance?
- Q. 3. For a SP of rank p such that $0 < p < q$, does $\underline{F}'$ admit a factorization $\underline{F}' = \underline{\Psi} \underline{\Psi}^*$ a.e.(Leb.), where $\underline{\Psi} \in \underline{L}_2^{0+}(C)$? If not, what additional condition would ensure such factorization?

Of these the most basic question Q.1 is still unanswered, despite some important work by Matveev which we shall discuss in §12. Q.2 and Q.3 have, however, been answered satisfactorily, cf. §§11, 12.

11. Concordance of Wold-Zasuhin and Lebesgue-Cramer decompositions for degenerate ranks

In 1959 the writer showed that the answer to the first part of question Q.2 (§10) is in the negative. He gave an example of a 2-ple process of rank 1 for which

$$(11.1) \quad \{0\} \neq \underline{M}_{-\infty} \subset \underline{M}_{\infty} \quad \& \quad \underline{F} \text{ is absolutely continuous,}$$

[19, §3]. For this SP concordance between the W.Z. and L.C. decompositions fails since $\underline{F}_v \neq 0 = \underline{F}_b$, cf. 10.3 et seq.

For 2-ple processes of rank 1, the writer also gave the extra condition needed for concordance, viz. $\det \underline{F}'(\theta) = 0$, a.e. (Leb.), cf. [19, 4.5]. Now it is easy to show that $\text{rank } \underline{F}'(\theta) \geq \text{rank } \underline{F}'_u(\theta) = p$, a.e. (Leb.). (Just combine 9.6 with 13.3 below). Hence for $q = 2$, $p = 1$ our result may be written:

$$\text{concordance} \iff \text{rank } \underline{F}'(\theta) = 1, \text{ a.e. (Leb.)}.$$

A complete generalization of this result was obtained by Robertson [26, 10.2] in 1963:

11.2 Thm. (Robertson) For any q -ple, weakly stationary SP of (any) rank p , there is concordance between the W.Z. and L.C. decompositions,

if and only if $\text{rank}' \underline{F}'(\theta) = \rho$ a.e. (Leb.), where $\underline{F}$ is the spectral distribution of the SP.

To prove this theorem Robertson used a result on the ranges of the matrices $\underline{F}'(\theta)$, viewed as linear transformations on $\mathbb{C}^q$ to $\mathbb{C}^q$, $\mathbb{C}$ being the complex-number field. This result is itself interesting [26, 9.11]:

11.3 Lemma (Robertson). Let $(\underline{x}_n)_{-\infty}^{\infty}$, $(\underline{y}_n)_{-\infty}^{\infty}$, $(\underline{z}_n)_{-\infty}^{\infty}$ be q -ple weakly stationary SP's with the same shift operator, and let

$$\underline{x}_n = \underline{y}_n + \underline{z}_n$$

$$\underline{m}_{-\infty}^{(x)} = \underline{m}_{-\infty}^{(y)} + \underline{m}_{-\infty}^{(z)}, \quad \underline{m}_{-\infty}^{(y)} \perp \underline{m}_{-\infty}^{(z)}.$$

Then, with an obvious notation,

- (a) $\text{Range } \underline{F}'_x(\theta) = \text{Range } \underline{F}'_y(\theta) + \text{Range } \underline{F}'_z(\theta), \quad \text{a.e. (Leb.)}$
- (b) $\text{Range } \underline{F}'_y(\theta) \cap \text{Range } \underline{F}'_z(\theta) = \{0\}, \quad \text{a.e. (Leb.)}$
- (c) $\text{rank } \underline{F}'_x(\theta) = \text{rank } \underline{F}'_y(\theta) + \text{rank } \underline{F}'_z(\theta) \quad \text{a.e. (Leb.)}.$

Some of these results were duplicated independently by Jang Ze-pei [10].

12. Degenerate rank factorization

The writer's example mentioned in connection with the question Q. 2 in §11 also shows that the answer to the first part of Q. 3 (§10) is in the negative. For this consider the 2-ple process of rank 1 satisfying (11.1).

Were $\underline{F}' = \underline{\Psi} \underline{\Psi}^*$ a.e. with $\underline{\Psi} \in \underline{L}_2^{0+}(C)$, then since $\underline{F}$ is absolutely continuous, it would follow from Rosanov's Thm. 9.7 that the SP is purely non-deterministic, in contradiction to the assertion $\underline{m}_{-\infty} \neq \{0\}$ in (11.1).

As in the case of Q.2, the extra conditions needed to secure a positive answer to Q.3 were first given for the case $q = 2$, this time by Wiener and the writer [37, 4.1] in 1959; and the result was then fully generalized by Matveev [24] in 1960, but for processes with continuous time. Whereas in the full rank case $p = q$ we encounter (holomorphic) functions of the Hardy class H_2 on the disk D_+ , or rather their radial limits in $\underline{L}_2^{0+}(C)$, for the degenerate rank cases $1 \leq p < q$ we encounter quotients of H_∞ functions on D_+ , i.e. the (meromorphic) beschränktartige functions introduced by R. Nevanlinna to round off the Hardy class theory. (For a brief, relevant account see [37, §3 & Note on p. 308].) The final result, cf. [24, Thm. 1], is as follows:

12.1 Thm. (Matveev) Let $\underline{F}$ be the spectral distribution of a q -ple SP. Then $\underline{F}'$ admits a factorization

$$\underline{F}'(\theta) = \underline{\Psi}(e^{i\theta}) \underline{\Psi}^*(e^{i\theta}) \text{ a.e. (Leb.)}, \text{ where } \underline{\Psi} \in \underline{L}_2^{0+}(C),$$

if and only if

(1) $\text{rank } \underline{F}'(\theta) = \text{const. } p, \text{ a.e. (Leb.)}$

(2) there is a principal $p \times p$ minor ⁽¹⁸⁾ $\underline{M} = \underline{M}_{i, \dots, i}^{i, \dots, i}_p$ of $\underline{F}'$ such that

$$\log \det \underline{M} \in L_1[0, 2\pi],$$

and for each $i \in \{i_1, \dots, i_p\}$, and each $k \in \{1, \dots, p\}$

$$\frac{\det \underline{M}_{i_k}^i(\theta)}{\det \underline{M}(\theta)} = \lim_{r \rightarrow 1-} \Psi_{i, i_k}^{(r)}(e^{i\theta}), \text{ a.e. (Leb.)}, \Psi_{i, i_k} \text{ beschränktartige,}$$

where $\underline{M}_j^i$ denotes the minor obtained from $\underline{M}$ by replacing the i_k^{th} row of $\underline{M}$ by the appropriate entries of the i^{th} row of $\underline{F}'$.

Matveev's proof is based on the fact:

$$(12.2) \quad \underline{\Psi} \in \underline{L}_2^{0+}(C) \implies \text{rank } \underline{\Psi}(e^{i\theta}) = \text{const. a.e. (Leb.)},$$

which emerges on applying theorems on Hardy class functions to the sums of the determinants of principal $r \times r$ minors of $\underline{\Psi}$, $1 \leq r \leq q$, cf. [16, 2.5].

Matveev showed that if the constant in (12.2) is p , then

$$\underline{\Psi}(e^{i\theta}) \underline{\Psi}^*(e^{i\theta}) = \underline{X}(e^{i\theta}) \underline{X}^*(e^{i\theta}),$$

where
$$\underline{X} = \begin{bmatrix} \dots & 0 \\ q \times p & q \times q-p \end{bmatrix} \in \underline{L}_2^{0+}(C).$$

The example mentioned above and Thm. 12.1 together answer the question Q.3 completely. Thm. 12.1 also answers completely the following question related to Q.1:

Q.1'. Given $0 \leq p < q$, what is the spectral n. & s.c. that a SP be purely non-deterministic and have rank p ?

The answer is immediate from the Theorems 9.7 and 10.1 of Rosanov and Matveev:

12.3 Cor. A q -ple SP is purely non-deterministic and of rank ρ , if and only if its spectral distribution $\underline{F}$ is absolutely continuous and $\underline{F}'$ satisfies the conditions (1), (2) of Thm. 12.1.

Unfortunately, this still leaves us in the dark as regards Q.1. For instance, the answer:

$\underline{F}'$ satisfies the conditions (1), (2) of Thm. 12.1

won't do. Indeed, by Robertson's Thm. 11.2 the condition (1) of 12.1 ensures concordance, whereas we know that there are non-deterministic processes for which concordance fails. A proper answer to Q.1 would be a major contribution.

13. Spectral and autoregressive representations for the predictor of a purely non-deterministic SP

We shall now turn to prediction proper. Let $(\underline{f}_n)_{-\infty}^{\infty}$ be a q -ple, non-deterministic SP, $\underline{h}_n$ be its n^{th} normalized innovation, and $\hat{\underline{f}}_v = (\underline{f}_v | \underline{m}_0)$ be the prediction of $\underline{f}_v$ with lag $v \geq 1$. Since $\underline{h}_0, \hat{\underline{f}}_v \in \underline{m}_{\infty}$, they have, cf. Thm. 7.3, isomorphs $\underline{W}, \underline{Y}_v \in \underline{L}_{2, \underline{M}}$ such that ⁽¹⁹⁾

$$(13.1) \quad \underline{h}_n = \int_0^{2\pi} e^{-ni\theta} \underline{W}(e^{i\theta}) \underline{E}(d\theta) \underline{f}_0, \quad \hat{\underline{f}}_v = \int_0^{2\pi} \underline{Y}_v(e^{i\theta}) \underline{E}(d\theta) \underline{f}_0.$$

Our first problem is to find $\underline{Y}_v$ for a purely non-deterministic SP. For such a process the Wold-Zasuhin decomposition (4.14) - (4.17) shows that

$$\underline{f}_n = \sum_{k=0}^{\infty} \underline{A}_k \sqrt{\underline{G}} \underline{h}_{n-k}, \quad \hat{\underline{f}}_v = \sum_{k=v}^{\infty} \underline{A}_k \sqrt{\underline{G}} \underline{h}_{v-k}.$$

Letting $n = 0$, taking isomorphs and proceeding heuristically, we get

$$\underline{I} = \sum_{k=0}^{\infty} \underline{A}_k \sqrt{\underline{G}} e^{ki\theta} \underline{W}(e^{i\theta}) = \underline{\Phi}(e^{i\theta}) \underline{W}(e^{i\theta})$$

$$(13.2) \quad \underline{Y}_\nu(e^{i\theta}) = \sum_{k=\nu}^{\infty} \underline{A}_k \sqrt{\underline{G}} e^{(k-\nu)i\theta} \underline{W}(e^{i\theta}) = [e^{-\nu i\theta} \underline{\Phi}(e^{i\theta})]_{0+} \underline{W}(e^{i\theta})$$

where $\underline{\Phi}$ is the generating function of the SP, and $[\dots]_{0+}$ denotes the function obtained from $\dots$ by cutting off the negative frequency terms from its Fourier series. The first equation yields $\underline{W}(\cdot) = \{\underline{\Phi}(\cdot)\}^{-1}$, which is wrong since $\underline{\Phi}$ need not be invertible. Our heuristic procedure is thus untenable, but it reveals that the determination of $\underline{Y}_\nu$ is tied up with the possibility of inverting the generating function $\underline{\Phi}$.

To investigate the invertibility of $\underline{\Phi}$, we first note that its degeneracies stem from a constant matrix, as the following canonical factorization given in [16, 3.1] and also [22, 3.6] makes clear:

13.3 Thm. The generating function $\underline{\Phi}$ of any q -ple, non-deterministic SP is expressible in the form $\underline{\Omega}(\cdot)\sqrt{\underline{G}}$, where $\underline{G}$ is the prediction error matrix with lag 1, and $\underline{\Omega} \in \underline{L}_2^{0+}(\mathbb{C})$ is invertible a.e. (Leb.), and $\binom{20}{}$ $\underline{\Omega}_+(0) = \underline{I}$. In fact

$$\underline{\Omega}(e^{i\theta}) = \underline{J}^\perp + \underline{\Phi}(e^{i\theta}) \underline{H} = \underline{I} + \sum_{k=1}^{\infty} \underline{A}_k \underline{J} e^{ki\theta},$$

where $\underline{J}$ and $\underline{H}$ are as in (4.10)-(4.12).

Since, $\underline{J}$, $\underline{H}$, $\underline{\Phi}$ are uniquely determined by the SP (cf. (4.9), (9.3), et seq.), so is $\underline{\Omega}$. In fact, as the writer showed in [17, 2.2], its inverse

$\underline{\Omega}^{-1}(\cdot) \stackrel{\text{d}}{=} \{\underline{\Omega}(\cdot)\}^{-1}$ is the isomorph in $\underline{L}_{2,\underline{M}}$ of the (non-normalized) innovation $\underline{g}_0$ in the purely non-deterministic case, i. e.

$$(13.4) \quad \underline{g}_n = \int_0^{2\pi} e^{-ni\theta} \underline{\Omega}^{-1}(e^{i\theta}) \underline{E}(d\theta) \underline{f}_0, \quad \underline{\Omega}^{-1} \in \underline{L}_{2,\underline{M}}.$$

Since $\underline{h}_0 = \underline{H}\underline{g}_0$, its isomorph $\underline{W}$ is of course $\underline{H}\underline{\Omega}^{-1} = (\underline{\Omega}\underline{H}^{-1})^{-1}$.

From 13.3 it easily follows that $\underline{\Omega}\underline{H}^{-1} = \underline{J}^\perp + \underline{\Phi}$. Thus, for a purely non-deterministic SP we find that

$$(13.5) \quad \underline{W}(e^{i\theta}) = \{\underline{J}^\perp + \underline{\Phi}(e^{i\theta})\}^{-1} \quad \text{in} \quad \underline{L}_{2,\underline{M}}.$$

For $q = 1$ we know the corresponding result for any (purely or impurely) non-deterministic SP, viz.

$$(13.6) \quad W(e^{i\theta}) = \chi_A(\theta)/\Phi(e^{i\theta}) \quad \text{in} \quad L_{2,M}$$

where A is any subset of $[0, 2\pi]$ such that A, A' are carriers of the (mutually singular) measures $|E(\cdot)u_0|^2, |E(\cdot)v_0|^2$, u_0, v_0 being as in the Wold decomposition. But for $q > 1$ the difficulties caused by rank deficiencies and the failure of concordance (§11) have prevented so far the discovery of a full-fledged generalization of (13.6).

Inserting the value of $\underline{W}$ given by (13.5) into the heuristically obtained equation (13.2), we get

$$(13.7) \quad \underline{Y}_\nu(e^{i\theta}) = [e^{-\nu i\theta} \underline{\Phi}(e^{i\theta})]_{0+} \{\underline{J}^\perp + \underline{\Phi}(e^{i\theta})\}^{-1} \quad \text{in} \quad \underline{L}_{2,\underline{M}}.$$

This equality was proved for purely non-deterministic SP's of full rank q in [36, II, 4.11]; a slight variation of the arguments used therein shows

its validity for $1 \leq p < q$. The equations (13.1), (13.5), (13.7) thus yield spectral expressions for the predictor and for the innovations of a purely non-deterministic SP.

We must next investigate the expressibility of the predictor $\hat{f}_v$ directly in terms of the f_{-k} , $k \geq 0$. For this we must appeal to a most basic property of the generating function, established by the writer [19, 2.9], viz. its optimality ⁽²¹⁾ :

13.8 Basic Lemma (a) The generating function Φ of a q-ple, non-deterministic SP of any rank p is an optimal function in $L_2^{0+}(C)$, i.e.

$$(1) \quad \Phi_+(0) \geq 0$$

$$(2) \quad \Psi \in L_2^{0+}(C) \quad \& \quad \Psi \Psi^* = \Phi \Phi^*, \quad \text{a.e. (Leb.) on } C$$

$$\implies \sqrt{\{\Psi_+(0) \Psi_+^*(0)\}} \leq \Phi_+(0) \quad .$$

(b) In case $p = q$, we have

$$\det \Phi_+(z) = \exp \left\{ \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \log \det \sqrt{F'(\theta)} d\theta \right\}, \quad |z| < 1 \quad .$$

Now let us confine attention to purely non-deterministic processes of full rank. By 13.8(b), the holomorphic (matrix-valued) function Φ_+ on the inner disk D_+ :

$$\Phi_+(z) = \sum_{k=0}^{\infty} C_k z^k, \quad z \in D_+, \quad \text{where} \quad C_k = A_k \sqrt{G} \quad ,$$

is invertible at each $z \in D_+$, and hence

$$\{\Phi_+(z)\}^{-1} = \sum_{k=0}^{\infty} D_k z^k, \quad z \in D_+, \quad ,$$

where the D_k are matrix coefficients, satisfying

$$\underline{C}_0 \underline{D}_0 = \underline{I}, \underline{C}_0 \underline{D}_1 + \underline{C}_1 \underline{D}_0 = \underline{0}, \underline{C}_0 \underline{D}_2 + \underline{C}_1 \underline{D}_1 + \underline{C}_2 \underline{D}_0 = \underline{0}, \dots$$

It follows from (13.7) that $\underline{Y}_\nu$ has a holomorphic extension $(\underline{Y}_\nu)_+$ to D_+ given by

$$(\underline{Y}_\nu)_+(z) = \sum_{k=\nu}^{\infty} \underline{C}_k z^k \cdot \sum_{k=0}^{\infty} \underline{D}_k z^k = \sum_{k=0}^{\infty} \underline{E}_{\nu k} z^k, \quad z \in D_+,$$

where

$$(13.9) \quad \underline{E}_{\nu k} = \sum_{j=0}^k \underline{C}_{\nu+j} \underline{D}_{k-j}.$$

This suggests that in some sense we should have

$$\underline{Y}_\nu(e^{i\theta}) \sim \sum_{k=0}^{\infty} \underline{E}_{\nu k} e^{ki\theta}.$$

But in general $\underline{Y}_\nu \notin L_1(C)$, and the $\underline{E}_{\nu k}$ will not be the ordinary Fourier coefficients of $\underline{Y}_\nu$. There will, however, be circumstances under which

$$(13.10) \quad \sum_{k=0}^n \underline{E}_{\nu k} e^{ki\theta} \rightarrow \underline{Y}_\nu(e^{i\theta}) \text{ in } L_{2,M} \text{ topology } (^{22}), \text{ as } n \rightarrow \infty,$$

and hence by our Isomorphism Thm. 7.3,

$$(13.11) \quad \sum_{k=0}^n \underline{E}_{\nu k} f_{-k} \rightarrow \hat{f}_\nu \text{ in } \underline{m}_\infty, \text{ as } n \rightarrow \infty.$$

In short, there are processes for which the coefficients $\underline{E}_{\nu k}$ given in

(13.9) provide an autoregressive representation for the predictor, to wit

$$(13.12) \quad \hat{f}_\nu = \sum_{k=0}^{\infty} \underline{E}_{\nu k} f_{-k}.$$

For such processes we call $(\underline{E}_{\nu k})_{k=0}^{\infty}$ the time-domain weighting sequence for the predictor $\hat{\underline{f}}_{\nu}$.

The papers [36, II; 21] are devoted largely to the demarcation of processes for which the equivalent conditions (13.10) - (13.12) prevail. No complete characterization has been obtained so far — only sufficient conditions, the best being perhaps the one in [21, 5.2]:

$$(13.13) \quad \begin{cases} \underline{F} \text{ is absolutely continuous on } [0, 2\pi] , \\ \underline{F}' \in \underline{L}_{\infty}[0, 2\pi] \quad \& \quad (\underline{F}')^{-1} \in \underline{L}_1[0, 2\pi] . \end{cases}$$

Also sufficient, of course, is the stronger boundedness condition, of greater practical interest, cf. [36, II, 5.1 & 7.3]:

$$(13.14) \quad \begin{cases} \underline{F} \text{ is absolutely continuous} \\ \lambda \underline{I} \leq \underline{F}'(e^{i\theta}) \leq \lambda' \underline{I} \text{ a.e. (Leb.)}, \text{ where } 0 < \lambda < \lambda' < \infty . \end{cases}$$

It is easy to show that the autoregressive relation (13.12) is equivalent to the discrete matricial Hopf-Wiener equation:

$$(13.15) \quad \underline{\Gamma}_{n+\nu} = \sum_{k=0}^{\infty} \underline{E}_{\nu k} \underline{\Gamma}_{n-k}, \quad n \geq 0 .$$

The continuous parameter analogues of (13.12), (13.15) are, for a given real lag $h > 0$,

$$(13.12') \quad \underline{f}_h = \int_0^{\infty} d\underline{E}_h(\tau) \cdot \underline{f}_{-\tau} ,$$

$$3.15') \quad \underline{\Gamma}(t+h) = \int_0^{\infty} d\underline{E}_h(\tau) \cdot \underline{\Gamma}(t-\tau), \quad t \geq 0 ,$$

where the weighing $\underline{E}_h(\cdot)$ is a $q \times q$ matrix-valued function of bounded variation on $[0, \infty)$. These are the equations with which Wiener began the subject of multivariate prediction [35, Ch. IV]. He showed that in simple cases of practical interest the weighting $\underline{E}_h(\cdot)$ can be found by solving the matricial Hopf-Wiener equation (13.15'). We now see that his pioneering work belongs to a rather late chapter of the general theory.

14. Determination of the generating function from the spectral density.

Given the covariance bisequence $(\underline{\Gamma}_k)_{-\infty}^{\infty}$ or equivalently the spectral density $\underline{F}'$ of a q -ple, purely non-deterministic SP of full rank, the determination of its generating function $\underline{\Phi}$ is of great importance for prediction. This is because once $\underline{\Phi}$ or its Fourier coefficients $\underline{C}_k$ are found, we can get $\underline{\Phi}^{-1}$ and its Taylor coefficients $\underline{D}_k$, and thereafter the crucial function $\underline{Y}_v$ and its coefficients $\underline{E}_{vk}$ required to determine the predictor $\hat{\underline{f}}_v$; cf. (13.7), (13.9), (13.1), (13.12).

In the case $q = 1$ Φ can be found in principle from the equation

$$(14.1) \quad \Phi_+(z) = \exp \left\{ \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \log \sqrt{F'(\theta)} d\theta \right\}, \quad |z| < 1,$$

and its coefficients C_k found from

$$(14.2) \quad \sum_{k=0}^{\infty} C_k e^{ki\theta} = \exp \left\{ \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} \alpha_k e^{ki\theta} \right\},$$

where α_k is the k^{th} Fourier coefficient of $\log F'$. These are canonical

expressions for optimal Hardy class functions in terms of the norms of their boundary values. But for $q > 1$ analogous expressions are not available because the exponential law, $\exp(\underline{A} + \underline{B}) = \exp \underline{A} \cdot \exp \underline{B}$, fails for matrices. In fact, no general, closed-form expression for Φ or for its leading coefficient $\sqrt{G}$ in terms of $\underline{F}'$ is known for the cases $q \geq 2$.

Its discovery would be a major contribution.

Fortunately, we do have an infinite series expansion for Φ and G in terms of $\underline{F}'$ in case the conditions (13.13) are met, i.e. for the only known case in which the predictor has an autoregressive representation (13.12), cf. [21, 4.7]. Since an explanation of this result and its proof would require a digression, we shall only describe how with its aid the crucial weighting coefficients $\underline{E}_{\nu k}$ may be computed from the $\underline{\Gamma}_k$. For simplicity we shall assume that our SP satisfies the stronger boundedness condition (13.14) rather than (13.13). For details the reader should see the papers [36, II, §6; 21, §§4, 5].

Knowing the covariances $\underline{\Gamma}_k$ and the bounds λ, λ' in (13.14), we first obtain the slightly modified coefficients (²³)

$$\underline{\Gamma}'_0 = \frac{2}{\lambda + \lambda'} \underline{\Gamma}_0 - \underline{I}, \quad \underline{\Gamma}'_n = \frac{2}{\lambda + \lambda'} \underline{\Gamma}_n, \quad n \neq 0.$$

We then compute $\underline{A}_0, \underline{A}_1, \underline{A}_2, \dots$, where $\underline{A}_0 = \underline{I}$, and for $m > 0$

$$(14.3) \quad \underline{A}_{-m} = -\underline{\Gamma}'_{-m} + \sum_{n=1}^{\infty} \underline{\Gamma}'_{-n} \underline{\Gamma}'_{-m-n} - \sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \underline{\Gamma}'_{-p} \underline{\Gamma}'_{-n-p} \underline{\Gamma}'_{-m-n} + \dots$$

We next find $\underline{B}_0, \underline{B}_1, \underline{B}_2, \dots$ from the recurrence relations

$$\underline{A}_0 \underline{B}_0 = \underline{I}, \quad \underline{A}_0 \underline{B}_1 + \underline{A}_1 \underline{B}_0 = 0, \quad \underline{A}_0 \underline{B}_2 + \underline{A}_1 \underline{B}_1 + \underline{A}_2 \underline{B}_0 = 0, \dots$$

Since $\underline{A}_0 = \underline{I}$, this computation does not involve matrix inversion. Finally,

for any given $\nu \geq 1$, we compute the coefficients $\underline{E}_{\nu 0}, \underline{E}_{\nu 1}, \underline{E}_{\nu 2}, \dots$

from

$$(14.4) \quad \underline{E}_{\nu k} = \sum_{j=0}^k \underline{B}_{\nu+j} \underline{A}_{k-j}, \quad k \geq 0.$$

There are the weighting coefficients required in the autoregressive series

(13.12) for the predictor $\hat{\underline{f}}_{\nu}$. To complete the solution of the Prediction

Problem 2.16, we must find the prediction error matrix $\underline{G}_{\nu}$ for lag ν .

For this, we first compute the crucial matrix $\underline{G}$ from

$$(14.5) \quad \underline{G} = \frac{2}{\lambda + \lambda^*} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \underline{A}_{-m} \underline{\Gamma}_{-n-m} \underline{A}_{-n}^*,$$

and then $\underline{G}_{\nu}$ from

$$(14.6) \quad \underline{G}_{\nu} = \sum_{k=0}^{\nu-1} \underline{B}_{-k} \underline{G} \underline{B}_{-k}^*.$$

The practical utility of this scheme of computation will be discussed in §15. In it the generating function Φ has been by-passed. But if Φ is wanted, its Fourier coefficients $\underline{C}_k$ can be found from the equations

$$(14.7) \quad \underline{C}_k = \underline{B}_{-k} \sqrt{\underline{G}}.$$

15. The factorization of matricial rational spectral densities

In most practical problems of prediction the q -ple SP $(\underline{f}_n)_{n=-\infty}^{\infty}$ has a spectral density $\underline{F}'$ on $\mathbb{C}$, ⁽²⁴⁾ which can be analytically continued to the entire complex plane $\mathbb{C}$ to yield a matrix rational function, i.e. one whose entries are complex-valued rational functions. Such a spectral density is said to be rational. Retaining the symbol $\underline{F}'$ for the extension to $\mathbb{C}$, we have

$$(15.1) \quad \underline{F}'(z) = \frac{1}{p(z)} \underline{P}(z), \quad z \in \mathbb{C}$$

where $\underline{P}$ is a $q \times q$ matrix polynomial and $p(\cdot)$ a complex polynomial.

We infer easily that there is an integer r , $1 \leq r \leq q$, such that $\text{rank } \underline{F}'(z) \leq r$, for any z , and $\text{rank } \underline{F}'(z) = r$ except for a finite number of z . We call r the rank of $\underline{F}'$.

A basic result, known generally but properly enunciated and proved by Polyak and Rosanov, cf. [30, Lemma 3; 31, Ch.I, 10.2], is that a (matricial) rational spectral density $\underline{F}'$ admits a factorization

$$(15.2) \quad \underline{F}'(e^{i\theta}) = \underline{\Psi}(e^{i\theta}) \underline{\Psi}^*(e^{i\theta}), \quad 0 \leq \theta \leq 2\pi,$$

where $\underline{\Psi} \in \underline{L}_2^{0+}(\mathbb{C})$ and $\underline{\Psi}$ is rational, i.e. its analytic continuation to $\mathbb{C}$ is rational. It follows from Rosanov's Thm. 9.7 that a q -ple process with a rational spectral density of rank $r > 0$ is purely non-deterministic and of rank r . We owe to Rosanov [30, Thm. 7] the proof that its generating

function $\underline{\Phi}$ itself has a rational extension $\underline{R}$ to $\mathbb{C}$, and moreover

$$(15.3) \quad \text{rank } \underline{R}(z) = r, \quad z \in D_+ . \quad (25)$$

Thus by Thm. 9.6

$$(15.4) \quad \underline{F}'(z) = \underline{R}(z) \underline{R}^*(z), \quad z \in \mathbb{C} .$$

To carry out the prediction for lag ν for our process when it has full rank q , we must find $\underline{R}$ and

$$(15.5) \quad \underline{Y}_\nu(e^{i\theta}) = [e^{-\nu i\theta} \underline{R}(e^{i\theta})]_{0+} \{ \underline{R}(e^{i\theta}) \}^{-1} ,$$

cf. §14 and (13.7). The methods proposed for this fall into two broad categories: (i) algebraic, (ii) analytic.

(i) An algebraic method has been proposed by Yaglom [41, §1] for continuous time processes, in which $\underline{R}$ is by-passed, and $\underline{Y}_\nu$ found directly. In the discrete version, it is assumed that $\text{rank } \underline{F}'(e^{i\theta}) = q$ for all θ , so that condition (13.14) is fulfilled. Since the $\underline{Y}_\nu$ in (15.5) is rational, Yaglom starts out with a rational function $\underline{Y}_\nu$ with undetermined coefficients, and shows that these can be found from the conditions

$$\underline{Y}_\nu(z) \text{ is holomorphic for } z \in D_+$$

$$\{z^{-\nu} \underline{I} - \underline{Y}_\nu(z)\} \underline{F}'(z) \text{ is holomorphic for } z \in D_- .$$

The first of these is obvious from (15.5), and the second is just a spectral paraphrasing of the condition $\underline{f}_\nu - \hat{\underline{f}}_\nu \perp \underline{f}_k, \quad k \leq 0$. Yaglom attacks $\underline{Y}_\nu$ row by row, and obtains a system of linear equations for the unknown

coefficients of $\underline{Y}_v$. He also adapts this method to prediction on the basis of a bounded time-interval in the past, [41, §2].

Youla [42] has given an algebraic technique for carrying out the factorization (15.4) even for $r < q$, again for continuous time, based on the diagonalization of polynomial matrices by elementary transformations, cf., [5, p.139]. In the discrete case we get from (15.1)

$$\underline{F}'(z) = \frac{1}{p(z)} \underline{C}_1(z) \cdot \underline{D}(z) \cdot \underline{C}_2(z)$$

where $\underline{D}(\cdot)$ is a diagonal matrix polynomial and $\underline{C}_1(\cdot)$, $\underline{C}_2(\cdot)$ are matrix polynomials with constant-valued determinants. By exploiting the hitherto unused fact that $\underline{F}'$ is a spectral density, Youla shows that the last factorization can be brought into the form (15.4), where $\underline{R}$ is a $q \times r$ rational matrix, holomorphic and left-invertible on D_+ . (A slight variation of his technique would yield a $q \times q$ rational $\underline{R}$ of rank r .) In effect, Youla proves a factorization theorem, but his proof is constructive and provides linear algebraic equations for the determination of $\underline{R}$.

Wiener's original approach [35, Ch. IV] may be classified as analytic-cum-algebraic. To solve the Hopf-Wiener equation (13.15') a Fourier analytic technique is to be used, which leads to the rational factorization problem (15.4). But to solve this problem an algebraic method is proposed, cf. [35, p.108].

(ii) The only purely analytic method known to us is the one outlined in §14. This will work when $\underline{F}'$ satisfies condition (13.13), i.e. for

rational $\underline{F}'$, only when $\det F'(e^{i\theta}) > 0$ for all θ . This method has been adapted to continuous time processes by Wong and Thomas [39], who also point out some short cuts in case $\underline{F}'$ is rational. But their paper contains several obscurities. This question of the extension of the factorization algorithm to continuous time is also the subject of a recent dissertation of H. Salehi [33]. In many practical situations, we would expect "weak memory", i.e. $\underline{\Gamma}_k = 0$ for $|k| > N$. In such cases $\underline{F}'$ would be a matrix trigonometrical polynomial (i.e. a rational function with poles only at 0 and ∞). Each series on the right-hand side of (14.3) would then terminate as would the series (14.5), and the method would gain in efficiency.

Which of these methods for finding $\underline{R}$ and $\underline{Y}_v$ is best suited to the modern digital computer? A weakness of the analytical technique of §14 is the occurrence of alternating signs in the series (14.3) resulting perhaps in slow convergence. On the other hand, the algebraic techniques that have been proposed involve the solution of large systems of linear equations, and it is not clear to the writer if they are generally more efficient. A comparative study of the effectiveness of all these methods on the computer would be very useful. Some interest has been aroused recently in this question because of its bearing on the discrimination of seismic signals due to different types of subterranean disturbances, cf. [28], [40]. With a slight smile one may say that an answer could even contribute to world peace.

REFERENCES

1. A. S. Besicovitch, A general form of the covering principle and relative differentiation of additive functions, II, Proc. Cambridge Phil. Soc., 42 (1946), 1-10.
2. A. Beurling, On two problems concerning linear transformations in Hilbert space, Acta Math., 81 (1949), 239-255.
3. H. Cramer, On the theory of stationary random processes, Ann. of Math., 41 (1940), 215-230.
4. J. L. Doob, Stochastic Processes, New York, 1953.
5. F. R. Gantmacher, The theory of matrices, vol. I, New York (1959).
6. P. Halmos, Shifts of Hilbert spaces, Crelle's Journal 208 (1961), 102-112.
7. O. Hanner, Deterministic and non-deterministic stationary random processes, Ark. for Mat., 1 (1950), 161-177.
8. H. Helson and D. Lowdenslager, Prediction theory and Fourier series in several variables, Part I, Acta Math. 99, 165 (1959); Part II, *ibid.* 106, (1962), 175-213.
9. _____, Vector-valued processes, Proc. of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, U. of Cal. Press, Berkeley, (1961), 203-212.

10. Jang Ze-Pei, The prediction theory of multivariate stationary processes, I, Chinese Mathematics, 13 (1963), 291-322.
11. K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung, Ann. Acad. Sci. Fennicae, A, I, 37 (1947).
12. A. N. Kolmogorov, Stationary sequences in Hilbert space, Bull. Math. Univ. Moscow, 2, No. 6 (1941), 1-40. (English translation by Natasha Artin).
13. L. H. Koopmans, On the coefficient of coherence for weakly stationary stochastic processes, Ann. Math. Stat., 35 (1964), 532-549.
14. _____, On the multivariate analysis of weakly stationary stochastic processes, Ann. of Math. Stat., 35 (1964), 1765-1780.
15. N. Levinson, The Wiener RMS error criterion in filter design and prediction, J. Math. and Physics, 25 (1947), 261-278. (Reprinted as Appendix B in [34]).
16. P. Masani, Sur la fonction génératrice d'un processus stochastique vectoriel, C. R. Acad. Sci., Paris, 249 (1959), 360-362.
17. _____, Isomorphie entre les domaines temporel et spectral d'un processus vectoriel, régulier, C. R., Acad. Sci., Paris, 249 (1959), 496-498.
18. _____, Sur les fonctions matricielles de la classe de Hardy H_2 , C. R. Acad. Sci., Paris, 249 (1959), 873-875; 906-907.

19. _____, Cramer's Theorem on monotone matrix-valued functions and the Wold Decomposition, (in "Probability and Statistics", The Harald Cramer Volume, Ed. by U. Grenander, Stockholm, 1959, 175-189).
20. _____, Une généralisation pour les fonctions matricielles de la classe de Hardy H_2 d'un théorème de Nevanlinna, C.R. Acad. Sci., Paris, 251 (1960), 318-320.
21. _____, The prediction theory of multivariate stochastic processes, III, Acta Math. 104 (1960), 141-162.
22. _____, Shift invariant spaces and prediction theory, Acta Math. 107 (1962), 275-290.
23. R. F. Matveev, On singular multidimensional stationary processes, Theory of Probability and its Applications, English Edition, 5 (1960), 33-39.
24. _____, On multidimensional regular stationary processes, Theory of Probability and its Applications, English Edition, 6 (1961), 149-165.
25. V. P. Potapov, The multiplicative structure of J-contractive matrix functions, Trudi Mosk. Matem. Obsch, 4 (1955), 125-236 (In Russian). (Amer. Math. Soc. Transl. 15 (1960), 131-243.)
26. J. B. Robertson, Multivariate continuous parameter weakly stationary stochastic processes, Thesis, Indiana University, 1963.

27. _____, "On wandering subspaces for unitary operators,
Proc. Amer. Math. Soc., 16 (1965), 233-236.
28. E. A. Robinson, Mathematical development of discrete filters for
detection of nuclear explosions, J. Geophys. Res., 68 (1963),
5559-5567.
29. Iu. A. Rosanov, The spectral theory of multi-dimensional stationary
random processes with discrete time, Uspekhi Mat. Nauk, XIII,
(1958), 93-142. (In Russian).
30. _____, Spectral properties of multivariate stationary processes
and boundary properties of analytic matrices, Theory of Probability
and its Applications, English Edition, 5(1960), 362-376.
31. _____, Stationary random processes, Moscow (1963) (Russian).
(English translation by M. Ravindranathan, Indian Statistical Institute,
Calcutta, (1965)).
32. M. Rosenberg, The square-integrability of matrix-valued functions
with respect to a non-negative hermitian measure, Duke Math. J.,
31 (1964), 291-298.
33. H. Salehi, The prediction theory of multivariate stochastic processes
with continuous time, Thesis, Indiana University, 1965.
34. P. Whittle, The analysis of multiple stationary time series, J. Roy.
Statist. Soc., Ser. B., 15 (1953), 125-139.
35. N. Wiener, Extrapolation, interpolation and smoothing of stationary
time series with engineering applications, New York, 1949.

36. N. Wiener and P. Masani, The prediction theory of multivariate stochastic processes, Part I, Acta Math., 98 (1957), 111-150; Part II, Acta Math., 99 (1958), 93-137.
37. _____, On bivariate stationary processes and the factorization of matrix-valued functions, Theory of Probability and its Applications (Moscow), English Edition, 4(1959), 300-308.
38. H. Wold, A study in the analysis of stationary time series, 2nd Edition, Stockholm, 1954.
39. E. Wong and J. B. Thomas, On the multidimensional prediction and filtering problem and the factorization of spectral matrices, J. Franklin Inst., 272 (1961), 87-99.
40. C. Wunch, The multiple spectral factorization method of Wiener and Masani, IEEE Trans. on Information Theory, IT-11 (1965), 175-182.
41. A. M. Yaglom, Effective solutions of linear approximation problems for multivariate stationary processes with a rational spectrum, Theory of Probability and Its Application, English edition, 5 (1960), 239-264.
42. D. C. Youla, On the factorization of rational matrices, IRE Trans., IT-7 (1961), 172-189.
43. V. Zasuhin, On the theory of multidimensional stationary random processes, (Russian) C. R. (Doklady) Acad. Sci., U. R. S. S., 33 (1941), 435-437.

Indiana University
Bloomington, Indiana

FOOTNOTES

1. A precise rendition of this statement will follow in a moment.
 2. In calling Γ_k a covariance matrix we are assuming tacitly that each $E(f_n^i) = 0$. This assumption entails no real loss of generality since our SP is stationary and therefore the vector $E(f_n)$ is independent of n . Alternatively, we may allow our SP $(f_n)_{n=-\infty}^{\infty}$ to be non-stationary, assuming only that (1.4) holds for $\tilde{f}_n^i$, where $\tilde{f}_n^i = f_n^i - E(f_n^i)$.
 3. Our usage of bold face letters is as follows: $\underline{f}$, $\underline{g}$, etc. denote members of $\mathfrak{H}^q$ and $\underline{m}$, $\underline{n}$ denote subsets of $\mathfrak{H}^q$. $\underline{A}$, $\underline{B}$, etc. denote $q \times q$ matrices with complex entries, and $\underline{\Phi}$, $\underline{\Psi}$, etc. denote $q \times q$ matrix-valued functions.
 4. We write $\underline{A} \geq \underline{B}$ or $\underline{B} \leq \underline{A}$ to mean that the matrix $\underline{A} - \underline{B}$ is non-negative definite. $\underline{A}^*$ denotes the adjoint of $\underline{A}$.
 5. For $\underline{G} \subseteq \mathfrak{H}^q$, $\mathfrak{G}(\underline{G})$ denotes the smallest subspace of $\mathfrak{H}^q$ containing $\underline{G}$.
N.B. Linear combinations must be taken with matrix coefficients.
- The symbol $\stackrel{d}{=}$ should be read "equals by definition". We shall often use it to introduce previously undefined expressions.
6. U can of course be extended (non-uniquely) to a unitary operator on $\mathfrak{H}$ onto $\mathfrak{H}$.
 7. $\text{Rstr.}_D A$ denotes the restriction of the operator A to the subset D of its domain.
 8. For $q = 1$, it was first proved by Wold [38] in 1938, and extended to

continuous time by Karhunen [11] and Hanner [7]. For $q > 1$, it was conjectured by Zasuhiu [43], and proved in the full rank case by Doob [4] and in general by Wiener and the writer [36, I]. The present method of obtaining it is given in [22, 3.1].

9. The Euclidean norm $|\underline{A}|_E$ of a matrix $\underline{A} = [a_{ij}]$ is defined by

$$|\underline{A}|_E^2 = \text{trace } \underline{A} \underline{A}^* = \sum_{i=1}^q \sum_{j=1}^q |a_{ij}|^2 .$$

10. and hence by stationarity for all integers n (and also $n = \infty$) .

11. With $n = \infty$ (4.19) reads: $(\underline{u}_n)_{-\infty}^{\infty}$ is subordinate to $(\underline{f}_n)_{-\infty}^{\infty}$,
cf. §8 below.

12. With the probabilistic interpretation of $\mathfrak{H}$, viz. $\mathfrak{H} = L_2(\Omega, \mathfrak{F}, P)$, ξ becomes a q-variate random measure over $([0, 2\pi], \mathfrak{B})$, but with the nice property that the q -variates corresponding to disjoint Borel sets are uncorrelated.

13. Actually Rosenberg takes rectangular $\underline{\Phi}, \underline{\Psi}$ in Def. (6.3), of sizes $p \times q$ and $q \times r$, respectively, and his result [32, 3.9] applies to all the $L_{2, \underline{M}}$ spaces obtained with different choices of p .

14. See Footnote 5 for the meaning of this symbol $\mathfrak{G}$.

15. Unfortunately there is no published work which treats the results of this section from our point of view. A treatment from a somewhat different standpoint is available in Rosanov [31, Ch. I, §§7, 8].

16. $\text{adj } \underline{A}$ denotes the adjugate matrix of $\underline{A}$, i.e. the transpose of the matrix formed by the cofactors of $\underline{A}$, so that

$$\underline{A} \cdot (\text{adj } \underline{A}) = (\det \underline{A}) \underline{I} = (\text{adj } \underline{A}) \cdot \underline{A} .$$

17. That every matricial distribution $\underline{F}$ possesses such parts $\underline{F}_a, \underline{F}_b$ is a celebrated result of Cramer [3, §4, Thm. 2]. See [19, 1.1] for a formulation of this result, especially suitable for our purposes.
18. $\underline{M}_{j_1, \dots, j_p}^{i_1, \dots, i_p}$ denotes the $p \times p$ minor of $\underline{F}'$ made up of the rows $i_1, \dots, i_p$, and the columns $j_1, \dots, j_p$.
19. For convenience we have transplanted the functions $\underline{W}, \underline{Y}_\nu$ from $[0, 2\pi]$ to the circle $C = [|z| = 1]$.
20. $\underline{\Psi}_+$ denotes the holomorphic extension to $D_+ = [|z| < 1]$ of a function $\underline{\Psi}$ in $\underline{L}_2^{0+}(C)$.
21. The word "maximal" is used instead of "optimal" in the English translations of the Russian literature, in which a less explicit but related result appears, cf. [30, Thm. 4]. For $q = 1$, the word "outer" has been used by Beurling [2] and his disciples.
22. i.e. since the spectral distribution $\underline{F}$ is absolutely continuous,
- $$\int_0^{2\pi} \left| \sum_{k=0}^n \underline{E}_{\nu k} e^{ki\theta} - \underline{Y}_\nu(e^{i\theta}) \right| \sqrt{\underline{F}'(\theta)} d\theta \rightarrow 0, \text{ as } n \rightarrow \infty.$$
23. Obviously, Γ'_k is the k^{th} Fourier coefficient of $\frac{2}{\lambda + \lambda'} \underline{F}'(\cdot) - \underline{I}$.
24. It is now convenient to transplant $\underline{F}'$ from $[0, 2\pi]$ to C .
25. Rosanov's proof can be simplified by appealing to the generalization of the classical canonical factorization of Hardy class functions to matrix-valued functions of the Hardy class H_2 on D_+ given in [18, 20, 22]. This generalization employed prediction theory as well as Potapov's important work [25], and illustrates how the subject has ramified.