ON ALGORITHMS FOR NONLINEAR PREDICTION

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

SHASHI PHOHA

1976

This is to certify that the

thesis entitled

ON ALGORITHMS FOR NONLINEAR PREDICTION

presented by

Shashi Phoha

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

Major professor

Date 6/4/76

O-7639

ABSTRACT

ON ALGORITHMS FOR NONLINEAR PREDICTION

By

Shashi Phoha

Any claim to prediction, i.e. to foretell the future in whatever limited sense, of a dynamic process, is based on a quantitative understanding of the model. Most processes in nature are such that their future behaviour is not completely determined by their past. For such random phenomena the idea of perfect prediction is replaced by that of a conditional distribution given the past.

In 1959 in their paper in the Harald Cramér volume, N. Wiener and P. Masani formally defined the non-linear prediction problem for a univariate strictly stationary stochastic process and obtained an algorithm for determining the nonlinear predictor in terms of the moments of the process. The corresponding non-linear prediction problem for multivariate processes takes into consideration the information contained in the correlative behaviour of these variables. In this thesis, by introducing certain operations on vectors to define a matrix-algebraic structure on the past, a determinate mathematical problem is established and its solution demonstrated by a corresponding algorithm.

In an attempt to obtain a more efficient algorithm for the problem of non-linear prediction for multivariate processes we proceed

along N. Wiener's idea of linearizing statistics of time-series to obtain the non-linear predictor. The latter half of the thesis deals with transferring the non-linear prediction problem of a univariate stationary process to linear prediction of a related infinite-variate process.

ON ALGORITHMS FOR NONLINEAR PREDICTION

By

Shashi Phoha

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1976

TO MY PARENTS

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Professor H. Salehi whose constant encouragement and guidance made this work possible. Sincere appreciation is also extended to Professor V. Mandrekar whose valuable criticisms gave a shape and form to unformulated ideas. I wish to thank Professors J. Hannan, H.L. Koul and J. Shapiro for serving on my committee. And for the very understanding attitude and personal interest one almost takes for granted here, I wish to thank all members of the faculty. Let me also thank Mrs. Noralee Burkhardt for her excellent typing of this thesis.

TABLE OF CONTENTS

INTRODUCTION

The ultimate objective of modeling a dynamic system is, of course, to predict or control the output of the system by observing and manipulating the inputs. As opposed to the hitherto acceptable idealization that the laws of nature could be formulated linearly, technology in its trends towards both refinements and magnifications of scope and complexity, has been pointing with increasing insistence to the fact that in the formulation of natural laws, modern requirements of precision forbid the suppression of nonlinear elements, for the suitable formulations are almost without exception, nonlinear. Fundamentals of nonlinear prediction theory as an outstanding contribution of World War II, came out of N. Wiener's work on predicting aircraft paths for purposes of fire control. The nonlinear theory is still in a juvenile state. This is a difficult terrain in which the going has been rough. But beyond the challenge to master difficulties, the demand for clearing of paths is arising for more mundane considerations. Wiener's work is geared towards applications. In particular, the theory modified and extended in various ways, constitutes a tool of considerable potential in forecasting and regulation.

In prediction theory of a stationary stochastic process, there is first the problem of defining the prediction, i.e., of showing that we have a determinate mathematical problem, and of demonstrating its theoretical solution. There is then a theoretical analysis of time and spectral domains. Finally, there is an attempt to use the time and spectral analysis to determine the predictor at a more efficient level. The program for linear prediction has largely been carried out in the multivariate case by N. Wiener and P. Masani [20], [21]. Linear predictors for weakly stationary stochastic processes are obtained and a factorization of the spectral density under the boundedness condition yields an algorithm for computing the generating function. An algorithm for computing the linear predictor is then obtained from the generating function. This work was generalized to Hilbert space valued random variables by A.G. Miamee [7]. A needed generalization to random variables defined on a Banach space, is the content of Chapter III.

For nonlinear prediction of a univariate, strictly stationary, discrete parameter stochastic process, work has been done at the first level by N. Wiener and P. Masani [19]; and the ultimate objective to determine the relevant conditional expectation has been established. The predictor, which is the conditional expectation, is obtained in terms of the moments of the process by reducing the problem to a projection on the L_2-closure of the algebra generated by the present and past of the process. In Chapter II of this thesis the corresponding nonlinear problem is defined for multivariate random variables. By introducing certain operators on matrices to define a matrix-algebraic structure on the past, a determinate mathematical problem is established and its solution demonstrated.

But a solution to the problem -- a solution as an algorithm which produces a numerical estimate from numerical observations, perhaps with the intercession of a digital computer, has not yet been satisfactorily obtained in the literature. Chapters III and IV of this thesis deal with transferring the univariate nonlinear prediction to linear prediction of an infinite-variate stationary stochastic process.

In his paper [18] in the Berkeley Symposium, N. Wiener suggested a closely related method of linearizing statistics of time-series to obtain the non-linear predictor. Chapter IV of the thesis deals with this method. For real-valued, measurable, univariate random variables $\{f_n: -\infty < n < \infty\}$, with all powers of f_n Lebesgue integrable, and an ergodic measure preserving transformation T, he suggested developing a linear theory of multiple prediction for random variables X_n whose components are products of all combinations of $f_1(T^{n_1}),\ldots,f_\nu(T^{n_\nu})$. The autocorrelations demanded are expressions of the form

$$E\{f_1(T^{n_1}) \times f_2(T^{n_2}) \times\ldots\times f_\nu(T^{n_\nu})\}.$$

A dense subspace of L^2 generated by $\{f_n: -\infty < n < \infty\}$ and consisting of finite products of $\{f_n\}_{-\infty}^{\infty}$, can then be arranged as an infinite-variate stationary stochastic process $\{X_n: -\infty < n < \infty\}$ whose linear past at any stage is exactly the nonlinear past (polynomials and their limits) of the univariate process $\{f_n: -\infty < n < \infty\}$. Due to duplication of subfactors in the components, however, the vectors $X_n(\omega)$ do not possess a Hilbert norm. X_n's may then possess a Banach norm. The linear predictor for $\{X_n: -\infty < n < \infty\}$ may then be obtained by methods established in Chapter III, which further gives the nonlinear predictor for the univariate process $\{f_n: -\infty < n < \infty\}$. Conditions on the moments of $\{f_n\}$ that would guarantee the boundedness condition for $\{X_n: -\infty < n < \infty\}$ have not been obtained explicitly. An example of a two state Markov chain sheds light on the procedures and some of the intricacies involved.

Following the work of Wiener and Masani much interest has been aroused in the areas of nonlinear prediction. Some of the more recent work related to this problem appears in [4], [8], [9], [10], [11], [14], [16].

CHAPTER I

PRELIMINARIES

While observing a random quantity associated with some long enduring mechanism in nature, from the remote past to the present moment, we obtain a sequence of readings $\{x_k: k = 0,-1,-2,\ldots\}$. For the problems of predicting a future value based on this information we shall be considering $\{x_k: k = 0,-1,-2,\ldots\}$ as part of a stationary stochastic process $\{f_n: n = 0, \pm 1, \pm 2,\ldots\}$ defined on a probability space $(\Omega,\mathcal{F},P)$ such that

$$x_n = f_n(\omega_0), \quad \omega_0 \text{ in } \Omega, \ -\infty < n < \infty .$$

Now f_n may be square integrable and hence lie in some Hilbert space $\mathcal{H}$. Following Masani in [5] we have

1.1. <u>The Gram-matricial structure of</u> $\mathcal{H}^N$ <u>for a fixed positive integer N.</u>

Let

$$\mathcal{H}^N = \{f = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix} : f_i \in \mathcal{H}, \ 1 \le i \le N\}.$$

For f,g in $\mathcal{H}^N$, the matricial inner-product of the ordered pair f,g is defined to be the $N \times N$ matrix

$$(f,g) = [(f_i, g_j)]_{N\times N} .$$

Orthogonality in $\mathcal{H}^N$ is defined as

$$f \perp g \Leftrightarrow (f,g) = 0$$

where 0 denotes the $N \times N$ matrix all whose entries are 0. Also define $\|f\|^2 = \text{trace } (f,f)$. Linear combinations in $\mathcal{H}^N$ are formed with $N \times N$ matrix coefficients.

A subset $\mathcal{M}$ of $\mathcal{H}^N$ is a linear manifold if and only if

$$f,g \in \mathcal{M} \Rightarrow \text{for any } N \times N \text{ matrices } A,B,\ Af + Bg \in \mathcal{M}.$$

1.2. Lemma. Ref. [20] Lemma 5.8 p. 131.

(i) $\mathcal{M}$ is a closed linear manifold of $\mathcal{H}^N$ if and only if $\mathcal{M} = M^N$ where M is a closed linear manifold of $\mathcal{H}$. In fact M consists of all coordinates of elements of $\mathcal{M}$.

(ii) If f is an element of $\mathcal{H}^N$ and $\mathcal{M}$ is a closed linear manifold of $\mathcal{H}^N$ then there exists a unique $\hat{f}$ in $\mathcal{H}^N$ satisfying one of the following equivalent conditions

(1) $\hat{f} \in \mathcal{M}$ and $f - \hat{f} \perp \mathcal{M}$

(2) $\hat{f} \in \mathcal{M}$ and $(f - \hat{f}, f - \hat{f}) \leq (f - g, f - g)$ for any $g \in \mathcal{M}$.

Corollary.

In particular for $\mathcal{H} = L_2(\Omega,\mathcal{F},P)$, let $\mathcal{L}_2$ denote the space $\mathcal{H}^N$. Then $\mathcal{M}$ is a closed linear manifold of $\mathcal{L}_2$ iff $\mathcal{M} = M^N$, where M is the closed linear manifold of L_2 consisting of all co-ordinates of elements of $\mathcal{M}$.

Notation.

$\hat{f}$ of Lemma 1.2(ii) is called the projection of f onto $\mathcal{M}$ and is written as $(f|\mathcal{M})$.

In this thesis n will always denote an integer and $\{-\infty < n < \infty\}$ denotes the set $\{n: n = 0, \pm 1, \pm 2, \ldots\}$.

1.2.1. Definition.

A N-variate weakly stationary stochastic process is a bisequence $\{f_n: -\infty < n < \infty\}$ such that $f_n \in \mathcal{H}^N$ and the Gram matrix

$$R_{m-n} = (f_m, f_n)$$

depends only on $m-n$.

Associated with a N-variate weakly stationary stochastic process $\{f_n: -\infty < n < \infty\}$ are the linear present and past subspaces $\mathcal{M}_n$ for $-\infty < n < \infty$. For a fixed integer n, $\mathcal{M}_n$ is defined to be the closed linear manifold of $\mathcal{H}^N$ spanned by $\{f_k: k \leq n\}$. Also

$$\mathcal{M}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{M}_n .$$

2. The Linear Prediction Problem.

Let $\{f_n: -\infty < n < \infty\}$ be a N-variate weakly stationary stochastic process with covariance bisequence $\{R_k: -\infty < k < \infty\}$ and let $\nu \geq 1$. Determine

(i) matrices $A_k^{(n)}$ such that

$$\hat{f}_\nu = (f_\nu|\mathcal{M}_0) = \lim_{n\to\infty} \sum_{k=0}^{n} A_k^{(n)} f_{-k}$$

(ii) $G_\nu = (f_\nu - \hat{f}_\nu, f_\nu - \hat{f}_\nu)$.

3. The Linear Prediction Problem for a Banach Space Valued Stationary Stochastic Process.

3.1. Notation.

For a Banach space $\mathcal{X}$ let $\mathcal{X}^*$ denote the Banach space of all conjugate bounded linear functionals on $\mathcal{X}$.

For Banach spaces $\mathcal{X}$, $\mathcal{Y}$, let $\mathcal{B}(\mathcal{X},\mathcal{Y})$ denote the Banach space of all bounded linear operators on $\mathcal{X}$ to $\mathcal{Y}$.

An operator f in $\mathcal{B}(\mathcal{X},\mathcal{X}^*)$ is defined to be nonnegative if for each $x \in \mathcal{X}$, $(f(x))\ (x) \geq 0$.

In case $\mathcal{X}$ is a real Banach space we further assume for nonnegative operators that

$$(f(x))\ (y) = (f(y))\ (x) \quad \text{for all} \quad x,y \quad \text{in} \quad \mathcal{X}.$$

We will alternately use the notation $(f(x),\ y)$ for $(f(x))\ (y)$.

Let $\mathcal{B}^+(\mathcal{X},\mathcal{X}^*)$ denote the set of all elements of $\mathcal{B}(\mathcal{X},\mathcal{X}^*)$ which are nonnegative.

3.2. Definitions.

For a Banach space $\mathcal{X}$ and a Hilbert space $\mathcal{K}$ and A, B elements of $\mathcal{B}(\mathcal{X},\mathcal{K})$, (A,B) is the unique bounded operator B^*A defined from $\mathcal{X}$ to $\mathcal{X}^*$. It follows that

$$((A,B)\ x,y) = (Ax,\ By), \quad \forall\ x \in \mathcal{X}, \forall\ y \in \mathcal{Y}\ .$$

Further let $A \perp B$ if $(A,B) = 0$.

3.3. Lemma.

Let $\{f_n : -\infty < n < \infty\}$ be a stochastic process such that for each n, f_n is defined on the probability space $(\Omega, \mathcal{F}, P)$ and takes values in a Banach space $\mathcal{Y}$. If for each x^* in $\mathcal{Y}^*$, $x^*(f_n) \in L_2(\Omega,\mathcal{F},P)$ for $-\infty < n < \infty$, we may identify the process $\{f_n: -\infty < n < \infty\}$ with a process $\{\xi_n: -\infty < n < \infty\}$ where ξ_n is in $\mathcal{B}(\mathcal{X},\mathcal{K})$ for $\mathcal{X} = \mathcal{Y}^*$ and $\mathcal{K} = L_2(\Omega,\mathcal{F},P)$.

Proof.

Define $\xi_n: \mathcal{X} \to L_2(\Omega,\mathcal{F},P)$ as follows

$$\xi_n(x^*)(\omega) = x^*(f_n(\omega)), \quad \omega \in \Omega,\ x^* \in \mathcal{X} .$$

3.4 Definition.

A bisequence $\{\xi_n: -\infty < n < \infty\}$ of elements of $\mathcal{B}(\mathcal{X},\mathcal{K})$ where $\mathcal{X}$ is a Banach space and $\mathcal{K}$ is Hilbert space, is called a $\mathcal{B}(\mathcal{X},\mathcal{K})$ valued weakly stationary stochastic process if the operator $\xi_m^*\xi_n$ in $\mathcal{B}(\mathcal{X},\mathcal{X}^*)$ depends only on m-n. And then the operator sequence $R(n) = \xi_0^*\xi_n$ defined for $-\infty < n < \infty$ is called the covariance bisequence of the process.

3.4.1. Assumption.

Assume from now on that $\mathcal{X}$ is separable.

3.4.2. The concepts and theorems from here on up to the end of Section 5 are outlined by A.G. Miamee [7].

With the stationary stochastic process $\{\xi_n: -\infty < n < \infty\}$ are associated the following subspaces

M_∞ the subspace of $\mathcal{K}$ spanned by $\{\xi_k(x): -\infty < k < \infty,\ x \in \mathcal{X}\}$,

M_n the subspace of $\mathcal{K}$ spanned by $\{\xi_k(x): -\infty < k \le n,\ x \in \mathcal{X}\}$,

and $M_{-\infty} = \bigcap_{-\infty<n<\infty} M_n$.

3.4.3. The process $\{\xi_n: -\infty < n < \infty\}$ is said to be

(i) singular if $M_{-\infty} = M_n$ for $-\infty < n < \infty$

(ii) nondeterministic if $M_{-\infty} \subsetneq M_n$ for some finite n,

(iii) regular if $M_{-\infty} = \{0\}$.

4. Time Domain Analysis.

For a $\mathcal{B}(X,K)$ valued regular weakly stationary stochastic process $\{\xi_n: -\infty < n < \infty\}$ there exist mutually orthogonal isometries S_k and $A_k \in \mathcal{B}(X,K)$ such that

$$\xi_n = \sum_{k=0}^{\infty} S_{n-k} A_k,$$

convergence being in the weak sense. Ref. [7] Theorem 3.3.7, p. 18.

5. Spectral Analysis.

5.1. Definitions: Ref. [22], p. 130-132.

A function f defined on a measure space $(S,\mathcal{B},\mu)$, with values in a Banach space $\mathcal{Y}$, is said to be finitely-valued if it is of the form $\sum_{i=1}^{n} I_{B_i} y_i$, $y_i \in \mathcal{Y}$ and $n < \infty$, where B_i's are disjoint measurable sets of finite μ-measure.

The function f is said to be strongly measurable if there exists a sequence of finitely-valued functions strongly convergent to f a.e. on S.

Further, f is said to be Bochner integrable if there exists a sequence of finitely-valued functions f_n strongly convergent to f a.e. in such a way that

$$\lim_{n\to\infty} \int_S \|f(s) - f_n(s)\| \mu(ds) = 0.$$

For any $B \in \mathcal{B}$, the Bochner integral of f over B is then defined by

$$\int_B f(s)\mu(ds) = \text{s-}\lim_{n\to\infty} \int_S I_B(s) f_n(s)\mu(ds).$$

Remark: Note that f is Bochner integrable if and only if f is strongly measurable and $\int \|f\| d\mu < \infty$.

5.1.1. For the weakly stationary stochastic process $\{\xi_n: -\infty < n < \infty\}$ which is $\mathcal{B}(\mathcal{X},\mathcal{K})$ valued, M_∞ denotes the closed linear subspace of the Hilbert space $\mathcal{K}$, generated by $\{\xi_n x: x \in \mathcal{X}, -\infty < n < \infty\}$. The shift operator $\mathcal{U}$ defined on M_∞ as follows

$$\mathcal{U}\xi_n x = \xi_{n+1} x, \quad x \in \mathcal{X}, -\infty < n < \infty$$

has a spectral resolution $\mathcal{U} = \frac{1}{2\pi}\int_0^{2\pi} e^{-i\theta} E(d\theta)$, where E is a projection valued measure over $([0,2\pi],\mathcal{B})$, $\mathcal{B}$ being the σ-algebra of Borel sets. Ref. [5], p. 359-360. Now define for $B \in \mathcal{B}$,

$$F(B) = \xi_0^* E(B)\xi_0.$$

Now $F(\theta) = \xi_0^* E(0,\theta]\xi_0$ is $\mathcal{B}(\mathcal{X},\mathcal{X}^*)$ valued such that for $x,y \in \mathcal{X}$

$$(\xi_n x, \xi_0 y) = (\frac{1}{2\pi}\int_0^{2\pi} e^{-in\theta} E(d\theta)\xi_0 x, \xi_0 y)$$

$$= \frac{1}{2\pi}\int_0^{2\pi} e^{-in\theta} (E(d\theta)\xi_0 x, \xi_0 y)$$

$$= \frac{1}{2\pi}\int_0^{2\pi} e^{-in\theta} (\xi_0^* E(d\theta)\xi_0 x, y)$$

i.e.

$$(R(n)x,y) = \frac{1}{2\pi}\int_0^{2\pi} e^{-in\theta} (F(d\theta)x,y).$$

<u>Definition.</u>

F defined in 5.1.1 is called the spectral distribution function for the process $\{\xi_n: -\infty < n < \infty\}$.

<u>Assumption.</u>

Assume from now onwards that there exists a $\mathcal{B}^+(X,X^*)$ valued function $f(\theta)$ defined on $[0,2\pi)$ such that

(i) $f(\theta)$ is strongly measurable

(ii) $f(\theta)$ is Bochner integrable

and (iii) for each Borel measurable $A \subset [0,2\pi)$

$$F(A) = \int_A f(\theta)d\theta.$$

This $f(\theta)$ will then be called the spectral density of the process $\{\xi_n: -\infty < n < \infty\}$. In general none of the above assumptions may hold.

5.2. Let K be a separable Hilbert space and let $L_2(K)$ denote the Hilbert space of all K-valued functions on the unit circle which have a square summable norm with the inner product of g_1,g_2 in $L_2(K)$ defined as

$$(g_1,g_2) = 1/2\pi \int_0^{2\pi} (g_1(e^{i\theta}),\ g_2(e^{i\theta}))d\theta.$$

Define

$$L_2^{0+}(K) = \{g \in L_2(K):\ 1/2\pi \int_0^{2\pi} e^{-in\theta}g(e^{i\theta})d\theta = 0 \quad \text{for} \quad n < 0\}$$

$$L_2^{0-}(K) = \{g \in I_2(K):\ 1/2\pi \int_0^{2\pi} e^{-in\theta}g(e^{i\theta})d\theta = 0 \quad \text{for} \quad n > 0\}.$$

5.3. Further, define a weakly measurable $\mathcal{B}(X,K)$ valued function $A(e^{i\theta})$ on the unit circle to be analytic, if for each $x \in X$, $A(e^{i\theta})(x)$ is in $L_2^{0+}(K)$,

conjugate analytic, if for each $x \in X$, $A(e^{i\theta})(x) \in L_2^{0-}(K)$.

5.2.1. Factorization of the Spectral Density.

A nonnegative weakly summable $B(X,X^*)$ valued function defined on the unit circle is said to be factorable if there exists a Hilbert space K and a conjugate analytic $B(X,K)$ valued function $A(e^{i\theta})$ defined on the unit circle such that $f(e^{i\theta}) = A^*(e^{i\theta})A(e^{i\theta})$, in the sense that $(f(e^{i\theta})\ x)(y) = (A(e^{i\theta})x,\ A(e^{i\theta})y)$ for all $x \in X$, $y \in X$.

Regarding factorization of the spectral density of a $B(X,K)$ valued regular stationary stochastic process, the following has been established in [7] Theorem 3.3.12, p. 23.

5.3.2. Theorem.

The spectral distribution F of a regular $B(X,K)$ valued stationary stochastic process is absolutely continuous and

$$\frac{d}{d\theta}\,(F(e^{i\theta})x)(x) = \|\phi(e^{i\theta})x\|^2$$

where

$$\phi(e^{i\theta})(x) = \sum_{k=0}^{\infty} e^{-ik\theta}A_k(x), \quad A_k \in B(X,K).$$

5.3.3. Definition.

ϕ as defined in the above theorem is called a generating function of the stationary stochastic $B(X,K)$ valued process $\{\xi_n: -\infty < n < \infty\}$.

6. Transferring the Linear Prediction Problem from the Time Domain to the Spectral Domain.

For the N-variate weakly stationary stochastic processes $\{f_n: -\infty < n < \infty\}$ an elementary solution to the linear prediction problem is summarized in [5] by choosing $A_k^{(n)}$ such that

$$f_\nu - \sum_{k=0}^{n} A_k^{(n)} f_{-k} \perp f_0, f_{-1}, \ldots, f_{-n},$$

whence

$$\left[A_0^{(n)}, \ldots A_n^{(n)}\right] \begin{bmatrix} (f_0, f_0) & (f_0, f_{-1}) \ldots (f_0, f_{-n}) \\ \ldots\ldots\ldots & \ldots\ldots\ldots \\ (f_{-n}, f_0) & (f_{-n}, f_{-1}) \ldots (f_{-n}, f_{-n}) \end{bmatrix}$$

$$= \left[(f_\nu, f_0) \quad (f_\nu, f_{-1}) \ldots (f_\nu, f_{-n})\right].$$

This method is not efficient when n is increasing as our observed data accumulates. Wiener and Masani transferred the problem to the spectral domain instead, establishing an isomorphism between the time domain and the spectral domain and thus obtaining the linear predictor at a more efficient level. This was done in [21] for a multivariate real-valued stationary stochastic process. A.G. Miamee [7] has extended their results to the case of a $\mathcal{B}(K,K)$ valued stochastic process for some Hilbert space K. He established in [7] Chapter VII that under regularity conditions on the spectral density, the matricial frequency-response or transfer function Y_ν that corresponds to the linear predictor $\hat{f}_\nu$ in the time domain, is given as

$$Y_\nu = [\, e^{-i\nu\theta} \phi(e^{i\theta})]_{0+} \; \phi^{-1} (e^{i\theta})$$

where $\phi(e^{i\theta})$ is a generating function of the stochastic process.

In chapter III we obtain a factorization of the spectral density for $\mathcal{B}(X, K)$ valued regular stochastic process and thus obtain the linear predictor and the prediction error matrix G_ν, for any $\nu \geq 1$, for such a process. Chapter IV proceeds to suggest the possibility of utilizing this analysis for nonlinear prediction problem. Results in chapter IV are not yet complete.

CHAPTER II

MULTIVARIATE NONLINEAR PREDICTION

In genuine applications one almost always works with systems whose states must be described by several random variables. Weather forecasts depend on a set of interdependent variables, for example. Although it is reasonable that ideas and methods should be developed first for the univariate case, ultimately one must be able to cope with the multivariate situation.

The price $\{p_n\}$ and demand $\{d_n\}$ of a certain commodity, tabulated by days, in themselves form univariate time-series, which are by no means independent. Nonlinear predictors for each of the processes $\{p_n: -\infty < n < \infty\}$ and $\{d_n: -\infty < n < \infty\}$ may be obtained under regularity conditions by methods of Wiener and Masani in [19]. However the predictors $\hat{p}_\nu$ and $\hat{d}_\nu$ for $\nu \geq 1$, thus obtained would ignore the information contained in the correlative behavior of p_n and d_n for $n \leq 0$. It would therefore be natural to consider $\{\binom{p_n}{d_n}: -\infty < n < \infty\}$ as a multivariate process and to define a corresponding structure on the past which formalizes the notions of nonlinear predictor and prediction error for a multivariate process. In fact with the appropriate formalism, all of the work of Wiener and Masani in [19] generalizes to this case. We proceed to introduce the same in this chapter and to obtain the corresponding generalization.

1. Multivariate Stationarity.

Let $f_{n1}, f_{n2}, \ldots, f_{nN}$, $-\infty < n < \infty$, be real random variables defined on a probability space (X, A, μ). Then the multivariate stochastic process $\{f_n: -\infty < n < \infty\}$ where $f_n = \begin{pmatrix} f_{n1} \\ f_{n2} \\ \vdots \\ f_{nN} \end{pmatrix}$ is defined to be strictly stationary if and only if for each integer $\nu > 0$,

$\mu[f_{n_1+\nu} \in B_1, f_{n_2+\nu} \in B_2, \ldots, f_{n_k+\nu} \in B_k] =$

$\mu[f_{n_1} \in B_1, f_{n_2} \in B_2, \ldots, f_{n_k} \in B_k]$, whatever $B_1, \ldots, B_k$ Borel measurable in R^N and $-\infty < n_1 < n_2 < \ldots < n_k < \infty$.

1.1. Note.

(i) The multivariate stochastic process $\{f_n: -\infty < n < \infty\}$ is strictly stationary if and only if for every integer $\nu > 0$

$$\mu[\omega \in S: f_{i+\nu j}(\omega) \in A_{ij}, (i,j) \in I] =$$

$$\mu[\omega \in S: f_{ij}(\omega) \in A_{ij}, (i,j) \in I]$$

whatever finite $I \subset \{-\infty < n < \infty\} \times \{1, 2, \ldots, N\}$ and Borel measurable $A_{ij} \subset R$.

(ii) If $\{f_n: -\infty < n < \infty\}$ is a multivariate strictly stationary process then the univariate stochastic processes $\{f_{nj}: -\infty < n < \infty\}$ for $j = 1, 2, \ldots, N$ are all strictly stationary.

1.2. A Characterization of Multivariate Strictly Stationary Stochastic Processes.

Notation.

For $s, t \in R^N$ let

$$st = \sum_{i=1}^{N} s_i t_i .$$

Theorem.

The process $\{f_n: -\infty < n < \infty\}$ is strictly stationary iff for each $t_1,\ldots,t_k \in R^N$ and $-\infty < n_1 < n_2 <\ldots< n_k < \infty$

$$t_1 f_{n_1} + t_2 f_{n_2} +\ldots+ t_k f_{n_k}$$

and

$$t_1 f_{n_1+1} + t_2 f_{n_2+1} +\ldots+ t_k f_{n_k+1}$$

are identically distributed.

Proof.

If $\{f_n: -\infty < n < \infty\}$ is strictly stationary then $t_1 f_{n_1+1} + t_2 f_{n_2+1} +\ldots+ t_k f_{n_k+1}$ is a finite linear combination of univariate random variables and the result then follows from Note 1.1(ii).

Conversely for every $\nu > 0$ and $t_j = (t_{j_1},\ldots,t_{j_N})^T \in R^N$, for $1 \le j \le k$ and any $u \in R$

$$E[\exp iu(t_1 f_{n_1} +\ldots+ t_k f_{n_k})] = E[\exp\{i \sum_{j=1}^{N} \sum_{\ell=1}^{k} ut_{j\ell} f_{n_j \ell}\}]$$

$$= E[\exp\{i \sum_{j=1}^{k} \sum_{\ell=1}^{N} ut_{j\ell} f_{n_j+1\ell}\}]$$

$$= E[\exp\{iu(t_1 f_{n_1+1} + t_2 f_{n_2+1} +\ldots+ t_k f_{n_k+1})\}].$$

Since u and t_j's were chosen arbitrarily, the characteristic function of the joint distributions of $\{f_{n_j \ell}: 1 \le j \le k, 1 \le \ell \le N\}$ and $\{f_{n_j+1\ell}: 1 \le j \le k, 1 \le \ell \le N\}$ agree.

Stationarity of $\{f_n\}_{-\infty}^{\infty}$ follows by Note 1.1(i).

1.3. Theorem.

With the multivariate strictly stationary process $\{f_n\}_{-\infty}^{\infty}$ defined on a complete probability space $(X,\mathcal{A},\mu)$, can be associated a probability space $(\Omega,\mathcal{F},P)$, a measure preserving transformation T on Ω onto itself and a random variable $\tilde{f}$ on Ω such that the stochastic process $\{\tilde{f}_n\}_{-\infty}^{\infty}$ defined by

$$\tilde{f}_n(\omega) = \tilde{f}(T^n\omega), \quad \omega \in \Omega$$

has the same joint distribution functions as $\{f_n\}_{-\infty}^{\infty}$, i.e. (in obvious notation)

$$\tilde{F}_{n_1,\ldots,n_q} = F_{n_1,\ldots,n_q}, \ -\infty < n_1 <\ldots< n_q < \infty.$$

Proof.

Let Ω be the set of all bisequences of elements of R^N, $\mathcal{A}$ the σ-algebra generated by cylinder sets and μ the Kolmogorov measure obtained by using the well-known Kolmogorov construction. Ref. [1] Theorem 2.2, p. 605. Letting T be the forward shift of co-ordinates, $\tilde{f}$ is obtained as the projection on the 0^{th} co-ordinate, $\mathcal{F}$ as the completion of the σ-algebra generated by $\mathcal{A}$ and P as the extension of μ to $\mathcal{F}$.

2. Preliminaries and Notation.

Let H_N denote the family of all $N \times N$ real matrices.

For $1 \leq p < \infty$, $L_p(\Omega,\mathcal{F},P) = \{f = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix}$: f is an equivalence class of $(\Omega,\mathcal{F})$-measurable functions and $\int_\Omega \{\sqrt{f^*f}\}^p dP < \infty\}$. For $f_1,f_2 \in L_2(\Omega,\mathcal{F},P)$ the gramian is defined as

$$(f_1,f_2) = \int f_1 f_2^* dP = [(f_{1i},f_{2j})]_{N\times N}$$

where $f_k = \begin{pmatrix} f_{k1} \\ \vdots \\ f_{kN} \end{pmatrix}$ for $k = 1,2$. $L_\infty(\Omega,F,P) = \{f: f$ is an equivalence class of (Ω,F) measurable functions such that there exists an $a \in R^N$ with $|f_i| \leq a_i$, $i = 1,\ldots,N\}$.

2.1. For a strictly stationary N-dimensional stochastic process $\{f_n: -\infty < n < \infty\}$ with $f_n \in L_\infty$, define for $-\infty < n < \infty$ and $q \geq 0$

$$B_n = \sigma\{f_{kj}^{-1}(A): A \subset R, \text{ A Borel measurable}, -\infty < k \leq n,$$

$$1 \leq j \leq N\} \cup \{\Lambda \subseteq \Omega : P(\Lambda) = 0\}.$$

$$B_{-\infty} = \bigcap_{-\infty<n<\infty} B_n$$

$$B_{0,q} = \sigma\{f_{kj}^{-1}(A): \text{A Borel measurable in } R, -q \leq k \leq 0,$$

$$1 \leq j \leq N\} \cup \{\Lambda \in \Omega: P(\Lambda) = 0\}$$

where σ denotes the generated σ-algebra.

Further define

N_n as the set of all R^N-valued B_n-measurable functions on Ω,

$N_{0,q}$ as the set of all R^N-valued $B_{0,q}$ measurable functions on Ω,

$M_n = N_n \cap L_2$

$M_{0,q} = N_{0,q} \cap L_2$.

2.2. <u>A Matricial Algebra</u>.

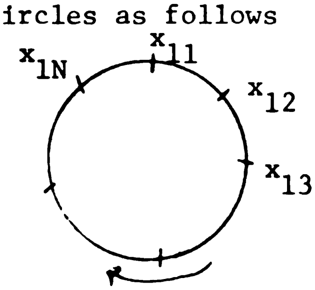
<u>Definitions</u>.

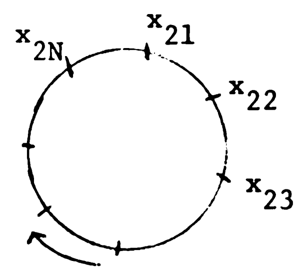
2.2.1. For $x_1,x_2 \in R^N$ let

$$x_1 * x_2 = \begin{pmatrix} x_{11}\ x_{21} + x_{12}\ x_{22} + \ldots + x_{1N}\ x_{2N} \\ x_{11}\ x_{22} + x_{12}\ x_{23} + \ldots + x_{1N}\ x_{21} \\ \cdots \\ x_{11}\ x_{2N} + x_{12}\ x_{21} + \ldots + x_{1N}\ x_{2N-1} \end{pmatrix} .$$

<u>Note</u>.

(i) If co-ordinates of x_1 and x_2 were arranged along two different circles as follows





then $x_1 * x_2$ is a $N \times 1$ vector whose k^{th} co-ordinate is obtained as the sum of products of corresponding entries on the two circles after the second circle is rotated anti-clockwise through $k-1$ units.

(ii) The aim of introducing this special type of multiplication is to be able to obtain all finite products of co-ordinates of x_1 and x_2 through the operations of $*$ and premultiplication by suitable $N \times N$ matrices on x_1 and x_2. More precisely, for any product $\prod_{\substack{i \in I \\ j \in J}} x_{1i} x_{2j}$ for $I, J \subset \{1,2,\ldots,N\}$, we have

$$\begin{pmatrix} \prod_{\substack{i \in I \\ j \in J}} x_{1i} x_{2j} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \{ \prod_{\substack{i \in I \\ j \in J}} (I_{1i} x_1) * (I_{1j} x_2) \}$$

where I_{1i} denotes the $N \times N$ matrix having 1 in the (1,i)th place and all other entries 0, and Π on the right hand side denotes $*$ multiplication from right to left.

Note that by premultiplication with a suitable elementary matrix, the position of the finite product on the left hand side of the equality may be suitably changed.

(iii) (a) Note that $*$ multiplication is not commutative.

(b) Also, $*$ multiplication is not associative since

$$\begin{pmatrix}1\\1\\0\end{pmatrix} * \left\{\begin{pmatrix}0\\1\\1\end{pmatrix} * \begin{pmatrix}1\\0\\1\end{pmatrix}\right\} = \begin{pmatrix}3\\3\\2\end{pmatrix}$$

and

$$\left\{\begin{pmatrix}1\\1\\0\end{pmatrix} * \begin{pmatrix}0\\1\\1\end{pmatrix}\right\} * \begin{pmatrix}1\\0\\1\end{pmatrix} = \begin{pmatrix}2\\3\\3\end{pmatrix} .$$

(c) For $A \subset H_N$, $A(x_1 * x_2) \neq Ax_1 * x_2$ in general. For example

$$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\left\{\begin{pmatrix}x_{11}\\0\end{pmatrix} * \begin{pmatrix}0\\x_{22}\end{pmatrix}\right\} = \begin{pmatrix}0\\x_{11} \quad x_{22}\end{pmatrix}$$

while

$$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} \begin{pmatrix}x_{11}\\0\end{pmatrix} * \begin{pmatrix}0\\x_{22}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} .$$

2.2.2. In the absence of these properties we make the following definition.

Any expression of the form

2.2.3. $\alpha_0 + \Sigma\, A_{n_1,\ldots,n_\ell}(A_{n_1}x_{n_1} * A_{n_2}x_{n_2} *\ldots* A_{n_\ell}x_{n_\ell})$

for $1 \leq n_1,n_2,\ldots,n_\ell \leq r$, $A_{n_i} \neq 0$ in H_N, α_0 is $N \times 1$ vector of scalars and Σ a sum over finitely many terms, is called a polynomial in $x_1,\ldots,x_r$. Note that the $*$ multiplication operation is always performed from right to left.

Denote a polynomial in $x_1,\ldots,x_r$ as $P(x_1,\ldots,x_r)$.

Note.

(i) $P(x_1,\ldots,x_r)$ is a $N \times 1$ vector each of whose co-ordinates is obtained by finite operations of addition and (ordinary) multiplication on the co-ordinates of $x_1,\ldots,x_r$. Each co-ordinate of $P(x_1,\ldots,x_r)$ is therefore some polynomial in $\{x_{ij}: 1 \le i \le r,\ 1 \le j \le N\}$ with scalar coefficients. We may thus write

$$P(x_1,\ldots,x_r) = \begin{pmatrix} p_1(x_{ij}) \\ p_2(x_{ij}) \\ \vdots \\ p_N(x_{ij}) \end{pmatrix} : \begin{matrix} 1 \le i \le r \\ 1 \le j \le N \end{matrix}$$

where $p_1,\ldots,p_N$ are polynomials in scalar coefficients.

(ii) Note that sum of two polynomials is also a polynomial and so is the pre-multiplication by an element of H_N. Furthermore the $*$ multiplication of two polynomials is also a polynomial as may be deduced from the following

$$(A_1x_1 + A_2x_2) * (B_1y_1 + B_2y_2)$$

$$= (A_1x_1 * B_1y_1) + (A_1x_1 * B_2y_2) + (A_2x_2 * B_1y_1) + (A_2x_2 * B_2y_2).$$

(iii) The Degree of a Polynomial.

Degree of a summand $A_{n_1,\ldots,n_\ell}(A_{n_1}x_{n_1} * \ldots * A_{n_\ell}x_{n_\ell})$ is defined to be ℓ. The degree of a polynomial $P(X_1,\ldots,x_r)$ is the maximum of the degrees of its summands.

Lemma.

The degree of a polynomial

$$P(x_1,\ldots,x_r) = \begin{pmatrix} p_1(x_{ij}) & \\ \vdots & : \begin{array}{l} 1 \le i \le r \\ 1 \le j \le N \end{array} \\ p_N(x_{ij}) & \end{pmatrix}$$

is the maximum of the degrees of the ordinary polynomials $p_1,\ldots,p_N$.

Proof.

Let the degree of $P(x_1,\ldots,x_r)$ be m. Then there exists a summand of this polynomial of the form

$$A_{n_1,\ldots,n_m}(A_{n_1}x_{n_1} * \ldots * A_{n_m}x_{n_m})$$

for $1 \le n_1,\ldots,n_m \le r$. Since $A_{n_i} \neq 0$ for $i = 1,2,\ldots,m$, each $N \times 1$ vector $A_{n_i}x_{n_i}$ has at least one nonzero entry. The co-ordinate polynomials of the $N \times 1$ vector $A_{n_{m-1}}x_{n_{m-1}} * A_{n_m}x_{n_m}$ therefore have degree at most 2 and at least one of these has degree exactly 2. Continuing to perform the $*$ multiplication operation from right to left it is easy to see that the maximum of the degrees of the polynomials is exactly m. Since $P(x_1,\ldots,x_r)$ consists of terms of the above form, the result follows.

2.2.4. A Matricial Algebra of N-variate Measurable Functions.

A subset $\mathcal{A}$ of the family of all measurable functions on $(\Omega,\mathcal{F})$ into R^N is called a matricial algebra if and only if it is closed for the following binary operations

(i) $f_1,f_2 \in \mathcal{A}$ and $A_1,A_2 \in H_N \Rightarrow A_1f_1 + A_2f_2 \in \mathcal{A}$

(ii) $f_1,f_2 \in \mathcal{A} \Rightarrow f_1 * f_2 \in \mathcal{A}$.

Note.

(i) If co-ordinates of f_1 and f_2 are bounded in absolute value so are the co-ordinates of $f_1 * f_2$.

(ii) Intersection of a family of matricial algebras is again a matricial algebra. Hence given a subfamily $\mathcal{E}$ of $\mathcal{A}$, the intersection of all matricial algebras containing $\mathcal{E}$ is the smallest matrcial algebra containing $\mathcal{E}$.

2.3. Notation.

For any set A let A^N denote the set of all $N \times 1$ vectors of elements of A, i.e. $A^N = \left\{ \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} : a_i \in A \right\}$.

For any positive integer n let P_r^n denote the family of n-variate polynomials in r variables $x_1, \ldots, x_r$.

Theorem.

$$P_r^N = (P_{Nr}^1)^N .$$

Proof.

$$P_r^N \subseteq (P_{Nr}^1)^N \quad \text{by Note 2.2.2(i).}$$

Conversely for $\begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix} \in (P_{Nr}^1)^N$,

2.3.1 $$\begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix} = \sum_{\ell=1}^{N} I_\ell \begin{pmatrix} p_\ell \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$ where I_ℓ is the elementary matrix obtained by interchanging the 1st and ℓth rows of the identity matrix.

Also each p_ℓ itself is a finite linear combination of finite products of $\{x_{ij}: 1 \le i \le r, 1 \le j \le N\}$. So if

$$p_\ell = \sum_{m=1}^{k} a_{I(m)} \prod_{(i,j)\in I(m)} x_{ij}$$

where $k < \infty$ and $I(m)$ contains a finite number of elements, repetitions allowed, of $\{1 \le i \le r\} \times \{1,2,\ldots,N\}$, then

$$\begin{pmatrix} p_\ell \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{m=1}^{k} A_{I(m)} \{ \prod_{(i,j)\in I(m)} I_{1j} x_i \}$$

where $A_{I(m)}$ is the $N \times N$ matrix with $a_{I(m)}$ in the (1,1)th place and 0 elsewhere and I_{1j} is the matrix with 1 in the (1,j)th place and 0 elsewhere. Note that Π here denotes * multiplication. Thus by 2.3.1 above it follows that $(P^1_{Nr})^N \subseteq P^N_r$. Hence the result.

<u>Corollary</u>.

Consider the set $\{x_r: r$ in a countable set $I\}$ of $N \times 1$ vectors. P^N, the family of polynomials over this set, is defined as the union of the families of polynomials over any finite subset of $\{x_r: r \in I\}$. It is immediate from the theorem then that

$$P^N = (P^1)^N.$$

Here P^1 is the family of all polynomials in $\{x_{rj}: 1 \le r < \infty, 1 \le j \le N\}$.

2.4. Let A_n denote the smallest matricial algebra containing the function $1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ and $\{f_k: k \le n\}$.

Let $A_{0,q}$ denote the smallest matricial algebra containing the function 1 and $\{f_k: -q \leq k \leq 0\}$.

Let A_n denote the linear algebra generated by the function 1 and $\{f_{ij}: -\infty < i \leq n, 1 \leq j \leq N\}$.

Let $A_{0,q}$ denote the linear algebra generated by the function 1 and $\{f_{ij}: -q \leq i \leq 0, 1 \leq j \leq N\}$.

Lemma.

$$A_n = \{\alpha_0 + \sum_{n_1,\ldots,n_\ell \in I} A_{n_1,\ldots,n_\ell} (\prod_{i=1}^{\ell} A_{n_i} f_{n_i}):$$

$$I \text{ is a finite subset of } \{k: -\infty < k \leq n\}, A_{n_i} \in H_N\}\}.$$

Proof.

Since A_n is closed for * multiplication and addition and premultiplication by elements of H_N,

$$A_n \supseteq \{\alpha_0 + \sum_{n_1,\ldots,n_\ell \in I} A_{n_1,\ldots,n_\ell} (\prod_{i=1}^{\ell} A_{n_i} f_{n_i}):$$

$$I \text{ a finite subset of } \{k: -\infty < k \leq n\}, A_{n_i} \in H_N\} .$$

Conversely the family on the right is a matricial algebra since it is closed for the required operations. It also contains the function 1 and $\{f_k: k \leq n\}$. Hence it contains A_n. Hence the result

Theorem.

For the strictly stationary process $\{f_n: -\infty < n < \infty\}$

$$\bar{A}_n = \overline{A_n^N} = (\bar{A}_n)^N$$

$$\bar{A}_{0,q} = \overline{A_{0,q}^N} = (\bar{A}_{0,q})^N .$$

A bar on top denotes closure.

Pro

Als

1 <

nor

2.5

fir

for

Bi

the

whe

T

Cor

ces

is

Proof.

$\mathcal{A}_n$ is the family $\mathcal{P}^N$ of polynomials over $\{f_i: -\infty < i \leq n\}$. Also A_n is the family $\mathcal{P}^1$ of polynomials over $\{f_{ij}: -\infty < i \leq n,$ $1 \leq j \leq N\}$. So by Theorem 2.3, $\mathcal{A}_n = A_n^N$. Taking closures in the trace norm gives the result.

Similarly for $\mathcal{A}_{0,q}$.

2.5. The moments of the stochastic process $\{f_n: -\infty < n < \infty\}$ are defined as

$$\alpha_{(n_1,j_1)(n_2,j_2),\ldots,(n_k,j_k)} = \int_\Omega f_{n_1j_1} \times f_{n_2j_2} \times\ldots\times f_{n_kj_k} dP(\omega)$$

for $-\infty < n_1 \leq n_2 \leq\ldots\leq n_k$ and $1 \leq j_1,j_2,\ldots,j_k \leq N$.

Birkhoff's Ergodic Theorem. Ref. [17] §2.2.

If T is a measure-preserving transformation on $(\Omega,\mathcal{F},P)$ then for any random variable f defined on Ω such that $E|f| < \infty$

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\nu=0}^{n} f(T^\nu\omega) = E[f|\mathcal{T}](\omega) \quad \text{a.e.},$$

where $\mathcal{T}$ is the family of T-invariant measurable sets. Further if T is ergodic, then

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\nu=0}^{n} f(T^\nu\omega) = E[f] \quad \text{a.s.}$$

Corollary 1. For a multivariate strictly stationary stochastic process $\{f_n: -\infty < n < \infty\}$, if the unique shift operator defined in 1.3 is ergodic then letting T_N denote the inflation of T to N-dimensions

$$\lim_{\nu\to\infty} \frac{1}{\nu+1} \sum_{\mu=0}^{\nu} T_N^{\mu}\{I_{1j_1} f_{n_1} * I_{1j_2} f_{n_2} * \ldots * I_{1j_r} f_{n_r}(\omega)\}$$

$$= \alpha_{(n_1,j_1)(n_2,j_2),\ldots,(n_r,j_r)} \quad \text{a.e.}$$

whatever $-\infty < n_1 \leq n_2 \leq \ldots \leq n_r < \infty$ and $1 \leq j_1, j_2, \ldots, j_r \leq N$.

<u>Proof</u>. The shift operator T_N is measure-preserving. Also

$$T_N^{\mu}\{I_{j_1} f_{n_1} * I_{j_2} f_{n_2} * \ldots * I_{j_r} f_{n_r}\}$$

$$= T_N^{\mu} \begin{pmatrix} 0 \\ \vdots \\ \prod_{k=1}^{r} f_{n_k j_k} \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ \vdots \\ T^{\mu} \prod_{k=1}^{r} f_{n_k j_k} \\ \vdots \\ 0 \end{pmatrix}.$$

Using Birkhoff's ergodic theorem the result now follows.

<u>Corollary 2</u>. For a multivariate stationary stochastic process $\{f_n : -\infty < n < \infty\}$, if the unique shift operator T is ergodic then for any polynomial P in N-dimensional variables $f_{n_1}, \ldots, f_{n_r}$ with matricial coefficients in H_N

$$\lim_{\nu\to\infty} \frac{1}{\nu+1} \sum_{\mu=0}^{\infty} T_N^{\mu}\{P(f_{n_1}, \ldots, f_{n_r})\} = E\{P(f_{n_1}, f_{n_2}, \ldots, f_{n_r})\}.$$

Proof. Note that

$$P(f_{n_1},\dots,f_{n_r}) = \begin{pmatrix} p_1(f_{n_i j}): & 1 \le i \le r,\ 1 \le j \le N \\ p_2(f_{n_i j}): & 1 \le i \le r,\ 1 \le j \le N \\ \vdots & \\ p_N(f_{n_i j}): & 1 \le i \le r,\ 1 \le j \le N \end{pmatrix}$$

where $p_1, p_2, \dots, p_N$ are some polynomials with real coefficients. The ergodic theorem holds for each coordinate now. This gives the desired result.

Remark. The assumption of ergodicity of the stationary process is not necessary in this thesis but with that assumption Birkhoff's ergodic theorem justifies the possibility of determining moments of the process from data collected from time-series observations. Ref. [19] §3, pp. 195-196. The assumption of stationarity reduces significantly the number of moments to be determined.

3. Statement of the Problem

3.1. Definition. For a strictly stationary stochastic process $\{f_n: -\infty < n < \infty\}$ with zero expectations and for any $\nu > 0$, define the predictor $\hat{f}_\nu$ with lead ν of the function f_ν as

$$\hat{f}_\nu = \begin{pmatrix} E(f_{\nu 1} | \mathcal{B}_0) \\ E(f_{\nu 2} | \mathcal{B}_0) \\ \vdots \\ E(f_{\nu N} | \mathcal{B}_0) \end{pmatrix} .$$

Let the right hand side of the equation be denoted by $E(f_\nu | \mathcal{B}_0)$.

3.2. The Prediction Problem.

Given the moments $\alpha_{(n_1,j_1)(n_2,j_2),\ldots,(n_k,j_k)}$ for $-\infty < n_1 \leq n_2 \leq \ldots \leq n_k < \infty$ and $1 \leq j_1,j_2,\ldots,j_k \leq N$ of the multivariate strictly stationary stochastic process $\{f_n: -\infty < n < \infty\}$ with $f_n \in L_\infty$ and with zero expectations, to determine polynomials Φ_q of $q + 1$ N-dimensional variables with matricial coefficients such that

$$\hat{f}_\nu = E[f_\nu | B_0] = \lim_{q\to\infty} \Phi_q(f_0,f_1,\ldots,f_{-q}).$$

Note.

1. Since the joint distributions $F_{n_1,\ldots,n_q}$ of $f_{n_1},f_{n_2},\ldots,f_{n_q}$ have compact support, their moments characterize them completely [12] Cor. 1.1, p. 11.

2. The assumption $f_n \in L_\infty$ is made to reduce the prediction problem to a projection on the closure of the algebra generated by the past. The weaker condition of existence of all moments is not sufficient. Ref. [19] Theorem 6.5.

4. Main Theorems.

4.1. The following are based on [5].

L_2 with the inner-product $(f_1,f_2) = \{\int f_1^* f_2 dP\}^{1/2}$ which yields the trace norm, is complete (almost everywhere equal functions are identified).

Theorem.

For $\nu > 0$, $\hat{f}_\nu$ is the orthogonal projection in L_2 of f_ν on M_0. $E[f_\nu | B_{0,q}]$ is the orthogonal projection in L_2 of f_ν on $M_{0,q}$.

Proof.

$M_0 = N_0 \cap L_2$ is a subspace of $L_2 = L_2^N$. In fact $M_0 = L_2^N(\mathcal{B}_0)$. Hence M_0, the space of all co-ordinates of elements of M_0, is $L_2(\mathcal{B}_0)$. Also as in the proof of Lemma 2.8 of [5], p. 354.

$$(f_\nu | M_0) = \begin{pmatrix} E(f_{\nu 1}|M_0) \\ E(f_{\nu 2}|M_0) \\ \vdots \\ E(f_{\nu N}|M_0) \end{pmatrix} = \begin{pmatrix} E(f_{\nu 1}|\mathcal{B}_0) \\ E(f_{\nu 2}|\mathcal{B}_0) \\ \vdots \\ E(f_{\nu N}|\mathcal{B}_0) \end{pmatrix}$$
$$= E(f_\nu | \mathcal{B}_0).$$

Similarly for projections on $M_{0,q}$.

4.2. Theorem.

A function $f \in N_{0,q}$ iff there exists a Baire function $\phi: (R^N)^{q+1} \to R^N$ such that $f = \phi(f_0, f_{-1}, \ldots, f_{-q})$.

4.3. Theorem

For F a distribution function on $(R^N)^q$ with a compact carrier J^q for some compact set $J \subset R^N$ and for $1 \leq p < \infty$, let $L_{p,F}$ be the space of all R^N-valued functions ϕ on $(R^N)^q$ which are measurable w.r.t. the Lebesgue-Stieltjes measure generated by F and are such that $\int |\phi|^p dF < \infty$ where $|\phi| = \sqrt{\phi^* \phi}$. Then under the norm $\|\phi\|_{F,p} = \{ \int_{J^q} |\phi|^p dF\}^{1/p}$, the polynomials in q-variables with values in R^N and coefficients in H_N form an everywhere dense subset of $L_{p,F}$.

Proof.

$$L_{p,F} = \{\phi: (R^N)^q \to R^N : \int\{\phi^* \phi\}^p dF < \infty\}.$$

As in 2.3 $\mathcal{P}_q^N$ denotes the family of polynomials in q N-variate vari-

ables $\{x_i\}_{i=1}^q$. $(P_{Nq}^1)^N$ is the family of polynomials in N_q variables $\{x_{ij}: 1 \leq i \leq q,\ 1 \leq j \leq N\}$. By Theorem 2.3 we have $P_q^N = (P_{Nq}^1)^N$.

Now P_{Nq}^1 is a dense subset of the space $L_{p,F}$ of all Lebesgue-Stieltjes measurable functions ψ defined on R^{Nq} into R^N with $\int(\psi^*\psi)^p dF < \infty$, Ref. [19] §6.2, p. 203. Hence for each coordinate ϕ_i of $\phi \in L_{p,F}$ there corresponds for given $\varepsilon > 0$, $p_i \in P_{Nq}^1$ such that $\|\phi_i - p_i\|_p^p < \varepsilon/N^p$. And then $p = \begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix} \in (P_{Nq}^1)^N = P_q^N$ is such that

$$\|\phi - F\|_{F,p}^p = \int\{\Sigma(\phi_i - p_i)^2\}^{p/2} dF \leq N^{p-1}\Sigma\|\phi_i - p_i\|_p^p < \varepsilon.$$

Hence P_q^N is everywhere dense in $L_{p,F}$.

4.4. Theorem.

(i) M_0 is the closure in L_2 of the linear manifold $\bigcup_{q=0}^{\infty} M_{0,q}$.

(ii) $M_{0,q} = \bar{A}_{0,q}$

(iii) $M_0 = \bar{A}_0$.

Proof.

Let M_0 denote the subspace of L_2 consisting of all co-ordinates of elements of M_0. Similarly define $M_{0,q}$. Then by Lemma 1.2 of Chapter I

$$M_{0,q} = M_{0,q}^N, \qquad M_0 = M_0^N.$$

Now let $f \in M_0$. Since the truncations $\{f\chi_{|f| \leq n}\}$ of f approach f in the trace norm, each co-ordinate of f is approximable by simple B_0 measurable functions. Since B_0 is the Borel algebra generated by $\bigcup_{q=0}^{\infty} \mathcal{B}_{0,q}$, simple Borel measurable functions are approximable by simple functions of the form $\sum_{i=1}^{k} \alpha_i I_{F_i}$ with $F_i \in \bigcup_{q=0}^{\infty} \mathcal{B}_{0,q}$ which obviously belong to $\bigcup_{q=0}^{\infty} N_{0,q}$. Such functions being bounded are in L_2 and hence in $\bigcup_{q=0}^{\infty} (N_{0,q} \cap L_2) = \bigcup_{q=0}^{\infty} M_{0,q}$. Hence the result.

(ii) Let $f \in M_{0,q}$. Then by Theorem 4.2, $f = \phi(f_0, f_{-1}, \ldots, f_{-q})$ where $\phi\colon (R^N)^{q+1} \to R^N$ is a Baire function.

Now $f \in L_2$, hence $\phi \in L_{2,F}$ where F denotes the joint distribution of $\{f_i\colon -q \le i \le 0\}$, because

$$\|f\|^2_{L_2} = \|\phi\|^2_{L_{2,F}} .$$

Also F has a compact carrier. Hence by Theorem 4.3 there exists a sequence of real polynomials Q_n of $q+1$ N-variate variables such that

4.4.1 $$\|\phi - Q_n\|_F \to 0 \quad \text{as} \quad n \to \infty.$$

Now define $\psi_n = Q_n(f_0, f_{-1}, \ldots, f_{-q})$. Then $\psi_n \in A_{0,q}$. Since $f \in L_2$, $\psi_n \in L_2$ and $\|f - \psi_n\|_{L_2} = \|\phi - Q_n\|_F$. By 4.4.1 then

$$f = \lim_{n\to\infty} \psi_n \quad \text{in } L_2\text{-norm}.$$

(iii) From Definition 2.4 it follows immediately that $A_0 = \bigcup_{q=0}^{\infty} A_{0,q}$. Also from (ii) above $\bigcup_{q=0}^{\infty} A_{0,q}$ is everywhere dense in $\bigcup_{q=0}^{\infty} M_{0,q}$. By (i) therefore, A_0 is everywhere dense in M_0 i.e. $\overline{A}_0 = M_0$.

Corollaries.

(i) For the strictly stationary stochastic process $\{f_n: -\infty < n < \infty\}$ and for $\nu > 0$

$$\hat{f}_\nu = \lim_{q\to\infty} E[f_\nu | M_{0,q}] \qquad \text{in } L_2 .$$

This is immediate from Theorem 4.4(i) above.

(ii) For the strictly stationary stochastic process $\{f_n: -\infty < n < \infty\}$ there exists a sequence ϕ_q of Baire functions on $(R^N)^{q+1}$ to R^N such that

$$\hat{f}_\nu = \lim_{q\to\infty} \phi_q(f_0, f_{-1}, \ldots, f_{-q}) \quad \text{in } L_2.$$

Proof.

From Corollary (i) $\hat{f}_\nu = \lim_{q\to\infty} E[f_\nu | M_{0,q}]$.

Also by Theorem 4.2 there exists a Baire function ϕ_q such that $E[f_\nu | M_{0,q}] = \phi_q(f_0, f_{-1}, \ldots, f_{-q})$. Hence the result.

(iii) For the strictly stationary stochastic process $\{f_n: -\infty < n < \infty\}$ and for any integer $\nu > 0$, there exists a sequence of polynomials Q_q in $q+1$ N-variate variables with coefficients in H_N and values in R^N such that

$$\hat{f}_\nu = \lim_{q\to\infty} Q_q(f_0, f_{-1}, \ldots, f_{-q}) \quad \text{in } L_2.$$

Proof.

By Corollary (ii) above there exist Baire functions ϕ_q such that $\|\hat{f}_\nu - \phi_q(f_0, f_{-1}, \ldots, f_{-q})\|_2 \to 0$. Furthermore for given $\varepsilon > 0$ there exist polynomials Q_q of the above type such that

$\|\phi_q - Q_q\|_2 < \frac{\varepsilon}{2^q}$. Thus $\|\phi_q(f_0,f_1,\ldots,f_{-q}) - Q_q(f_0,f_{-1},\ldots,f_{-q})\|_2 < \frac{\varepsilon}{2^q}$. Hence $\hat{f}_\nu = \lim_{q\to\infty} Q_q(f_0,f_{-1},\ldots,f_{-q})$ in L_2.

5. Computation of the Predictor.

In this section we shall obtain an orthonormal basis for the space $M_0 = \bar{A}_0$ in L_2 with the trace norm. The expansion of the predictor in this basis is then actually computable from the given data.

Define a subset $H \subset L_2$ to be linearly independent iff for any finite collection $h_1,\ldots,h_n \in H$ and $A_i \in H_N$ with $1 \le i \le n$,

$$\Sigma A_i h_i = 0 \Rightarrow A_i = 0 \;\forall\, i \text{ with } 1 \le i \le n .$$

Define the span $S(H)$ of a subset H of L_2 to be the family

$$\{\sum_{i=1}^{n} A_i h_i \colon A_i \in H_N,\; h_i \in H \text{ and } 0 < n < \infty\} .$$

A subset H of a subspace M of L_2 is said to be a basis for M iff H is linearly independent and spans M.

A basis H of a subspace M of L_2 is said to be orthogonal iff for $h_1 \ne h_2 \in H$

$$(h_1,h_2) = Eh_1^* h_2 = 0 .$$

An orthogonal basis H is said to be orthonormal if in addition $\forall\, h \in H$

$$\|h\|^2 = \text{trace}(h,h) = 1.$$

5.1. Lemma.

A subset $\mathcal{H} \subset L_2$ is linearly independent in L_2 iff the corresponding set $H = \{h_j : 1 \le j \le N,\ h = \begin{pmatrix} h_1 \\ \vdots \\ h_N \end{pmatrix} \in \mathcal{H}\}$ is linearly independent in L_2.

Proof.

Let $h_1,\ldots,h_n$ be a finite set which is linearly independent in $\mathcal{H}$. Then

$$\sum_{i=1}^{n}\sum_{j=1}^{N} a_{ij}h_{ij} = 0 \Leftrightarrow \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}\begin{pmatrix} h_{11} \\ h_{12} \\ \vdots \\ h_{1N} \end{pmatrix} + \cdots$$

$$+ \begin{bmatrix} a_{n1} & \cdots & a_{nN} \\ 0 & \cdots & 0 \\ 0 & \vdots & 0 \end{bmatrix}\begin{pmatrix} h_{n1} \\ \vdots \\ h_{nN} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Leftrightarrow a_{ij} = 0 \text{ for } 1 \le i \le n,\ 1 \le j \le N.$$

Conversely if $\{h_{ij}: 1 \le i \le n, 1 \le j \le N\}$ is linearly independent in L_2 then $h_1,\ldots,h_n$ must be linearly independent in $\mathcal{L}_2$ since for matrices $A_k = [a^k_{ij}]_{N\times N}$, $\sum_{k\in I} A_k h_k = 0 \Leftrightarrow a^k_{ij} = 0$ for $1 \le i \le N$, $1 \le j \le N$ and $I \subset \{1,2,\ldots,n\}$, i.e. $A_k = 0$ for each $k \in I$. Hence $\{h_1,\ldots,h_n\}$ is linearly independent in $\mathcal{L}_2$.

The case when the set H is not finite is immediate now.

5.1.1. A linear arrangement of the set $H = \{f_{ij}: -\infty < i \le 0, 1 \le j \le N\}$ is:

$$f_{01},\ldots,f_{0N},f_{-11},\ldots,f_{-1N},f_{-21},\ldots,f_{-k\ell},\ldots$$

Then $H = \{h_n: \ 0 \leq n < \infty\}$ with $h_0 = 1$ and for each $n > 0$

$$h_n = f_{-[\frac{n-1}{N}],n-[\frac{n-1}{N}]N}$$

where $[\frac{n}{N}]$ denotes the integral part of $\frac{n}{N}$. One method of arranging all finite products of elements of H in a row is given by Wiener and Masani as follows. Ref. [19] §7, p. 205.

Let $p_k = k+1^{th}$ prime, and $\forall\ j > 1$ with the prime factorization

$$j+1 = p_{k_1} p_{k_2} \cdots p_{k_r} \quad , \qquad 0 \leq k_1 \leq \cdots \leq k_r$$

define

$$\phi_j = h_{k_1} h_{k_2} \cdots h_{k_r} .$$

$\{\phi_j: \ 1 \leq j < \infty\}$ is then a linear rearrangement of all finite products of elements of H.

Thus

$M_0 = \text{clos } S\{\phi_j: \ 0 \leq j < \infty\}$ where $\phi_0 = 1$

M_0 is therefore the closure of the subspace of L_2 spanned by the set $M = \left\{ \begin{pmatrix} \phi_j \\ 0 \\ \vdots \\ 0 \end{pmatrix} : \ 0 \leq j < \infty \right\}$.

5.2. A Basis for A_0.

Lemma.

A subset $M \subset A_0$ forms an orthonormal basis for A_0 in L_2 if and only if the corresponding set $\mathcal{M} = \left\{ \begin{pmatrix} h \\ 0 \\ \vdots \\ 0 \end{pmatrix} : h \in M \right\}$ forms an orthonormal basis for $\mathcal{A}_0$ in $\mathcal{L}_2$.

Proof.

By Theorem 2.4 $\overline{\mathcal{A}}_0 = \overline{A}_0^N$. By Lemma 1.2(i) of Chapter I the result follows since coefficients in linear combinations of elements of $\mathcal{A}_0$ in $\mathcal{L}_2$ come from H_N and

$$\left\| \begin{pmatrix} h \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|_2 = \|h\|_{L_2} .$$

Now the set $\{\phi_j : 0 \leq j < \infty\}$ need not be linearly independent in general Ref. [19], §7.3. Although the following procedure of computation of the predictor is possible without having a linearly independent set at this stage, we may not be able to keep track of the number of variables involved in the approximating polynomials. Of course the computations get more complicated also. We therefore make the following

Assumption.

The strictly stationary stochastic process $\{f_n : -\infty < n < \infty\}$ is such that for any finite set $I \subset \{-\infty < n < \infty\} \times \{1,2,\ldots,N\}$, the spectrum of the joint distribution of $\{f_{ij} : (i,j) \in I\}$ has positive Lebesgue measure in the $|I|$ dimensional space, where $|I|$ denotes the cardinality of the set I.

5.2.1. Lemma.

Under the above assumptions the set $\{\phi_j: 0 \leq j < \infty\}$ is linearly independent.

Proof.

Let $\psi = \sum_{i=1}^{n} c_i\phi_i$, $c_i \neq 0$, be any linear combination. Note that ϕ_i's are univariate random variables. Let the factors occurring in the products ϕ_i for $i = 1,2,\ldots,n$ be $\{f_{ij}: (i,j) \in I\}$. Then $\psi = Q(f_{ij}: (i,j) \in I)$ for some non-zero polynomial Q in $|I|$ univariate variables and

$$\|\psi\|^2 = \int |Q(x_{ij}: (i,j) \in I)|^2 dF_I(x_{ij})$$

where F_I denotes the joint distribution of $\{f_{ij}: (i,j) \in I\}$. Note that F_I has compact support. Since Q can vanish only on at most a $|I| - 1$ dimensional algebraic surface which is of zero $|I|$-dimensional Lebesgue measure and the spectrum of F_I has positive measure, it follows that $\|\psi\| > 0$, i.e. $\psi \neq 0$. Hence $\{\phi_j: 0 < j < \infty\}$ is linearly independent.

5.2.2. Corollary.

Under the above assumptions the set $\{\phi_j: 0 \leq j < \infty\}$ can be orthonormalized to obtain an orthonormal basis $\{\psi_j: 0 \leq j < \infty\}$ for A_0 in L_2 as follows:

$\psi_0 = 1$

$\psi_1 = \phi_1/\sqrt{(\phi_1,\phi_1)}$

....

$$\psi_k = \frac{1}{\sqrt{\Delta_{k-1}\Delta_k}} \begin{vmatrix} (\phi_0,\phi_0) & (\phi_0,\phi_1) & \cdots & (\phi_0,\phi_{k-1}) & \phi_0 \\ (\phi_1,\phi_0) & (\phi_1,\phi_1) & \cdots & (\phi_1,\phi_{k-1}) & \phi_1 \\ \vdots & \vdots & & \vdots & \vdots \\ (\phi_k,\phi_0) & (\phi_k,\phi_1) & \cdots & (\phi_k,\phi_{k-1}) & \phi_k \end{vmatrix}$$

where $\Delta_k = \det[(\phi_i,\phi_j)]_{0 \le i,j \le k}$, $k > 1$.

5.2.3. Corollary.

For every integer $\nu > 0$, $\hat{f}_{\nu j} = E[f_{\nu j}|\mathcal{B}_0] = \lim_{n\to\infty} \sum_{k=0}^{n} (f_{\nu j},\psi_k)\psi_k$.

5.3. Computation of the Predictor.

Theorem.

For every $\nu > 0$, $\hat{f}_\nu = \lim_{n\to\infty} Q_n(f_0,f_{-1},\ldots,f_{-m_n})$ where m_n is an increasing sequence of nonnegative integers and Q_n is a polynomial in m_n+1 N-variate variables with co-efficients in H_N, these co-efficients being computable in terms of the moments

$\alpha_{(n_1,j_1)(n_2,j_2)\ldots(n_k,j_k)}$.

Proof.

As in the univariate case (Ref. [19] Theorem 7.9, p. 209), let μ_j be the subscript of the last prime in the factorization of $j+1$ and let, for any $n > 0$

$$m_n = \max\{\mu_1,\ldots,\mu_n\}.$$

Then ψ_k is expressible as a linear polynomial in terms of $\{\phi_j\colon 0 \le j \le k\}$, each one of which is further expressible as a polynomial in $\{h_j : 1 \le j \le k\}$.

Thus from the Corollary 5.2.3 there exist real polynomials q_{nj} such that

$$\hat{f}_{\nu j} = \lim_{n\to\infty} \sum_{k=0}^{n} (f_{\nu j}, \psi_k)\psi_k = \lim_{n\to\infty} q_{nj}(f_{i\ell} : -m_{n_j} \le i \le 0,\ 1 \le \ell \le N).$$

Thus

$$\hat{f} = \lim_{n\to\infty} \begin{pmatrix} q_{n_1}(f_{ij} : -m_n \le i \le 0,\ 1 \le j \le N) \\ q_{n_2}(f_{ij} : -m_n \le i \le 0,\ 1 \le j \le N) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ q_{n_N}(f_{ij} : -m_n \le i \le 0,\ 1 \le j \le N \end{pmatrix} .$$

Since $A_{0,m_n} = A^N_{0,m_n}$, there exist polynomials $Q_n \in A_{0,m_n}$ such that

5.3.1 $$\hat{f}_\nu = \lim_{n\to\infty} Q_n(f_0, f_{-1}, f_{-2}, \ldots, f_{-m_n}).$$

<u>Note</u>: Actually q_{nj} is a linear polynomial in m_n+1 variables and

$$\hat{f}_{\nu j} = \lim_{n\to\infty} q_{nj}(h_\ell : -m_n \le \ell \le 0)$$

$$= \lim_{n\to\infty} q_{nj}(f_{-[\frac{\ell}{N}],\ell-[\frac{\ell}{N}]N} : -m_n \le \ell \le 0).$$

So the coordinates of the predictor are obtained through the procedure of this section and are written in terms of limits of polynomials in N-variate variables in 5.3.1.

CHAPTER III

EXTENSION OF THE ALGORITHM FOR LINEAR PREDICTION TO BANACH SPACE VALUED STATIONARY STOCHASTIC PROCESSES

An algorithm for computation of the linear predictor for Banach space valued random variables is obtained in this chapter. Time domain and spectral analysis for such processes, including Wold Cramer concordance and necessary and sufficient conditions for factorability of the spectral density were obtained by A. G. Miammee Ref. [7], Chapter III. However the algorithm of Wiener and Masani under the boundedness condition, was obtained only for Hilbert space valued random variables using Fourier analysis of infinite matrix valued functions Ref. [7], Chapter VII. Under an extension of the boundedness condition of Wiener and Masani on the spectral density of the Banach space valued stationary stochastic process, a corresponding algorithm for computing the generating function and the linear predictor is obtained.

1.1 The Boundedness condition on the Spectral Density

Let the spectral density f_θ , of the $B(X,K)$ valued stationary stochastic process $\{\xi_n : -\infty < n < \infty\}$ satisfy

$$0 < m(\theta)A^*A \le f_\theta \le M(\theta)A^*A \quad \text{a.e.} \quad \theta \in [0,2\pi)$$

for some $A: X \to K$ with $\|A\| = 1$ and $M(\theta)$, $\frac{1}{m(\theta)}$, $\frac{M(\theta)}{m(\theta)}$ summable.

1.2. Lemma.

Under boundedness condition 1.1 for $N_\theta = \frac{f_\theta}{a_\theta} - A^*A$ with $a_\theta = \frac{m(\theta) + M(\theta)}{2}$,

$$\|N_\theta\|_B \leq \frac{M(\theta) - m(\theta)}{M(\theta) + m(\theta)} < 1.$$

Proof.

Let $\hat{f}_\theta$ denote the quadratic form of f_θ, i.e.

$$\forall\ x \in X,\ \hat{f}_\theta(x) = (f_\theta x, x).$$

For $x, y \in X$

$$\hat{f}_\theta(x+y) = \hat{f}_\theta(x) + \hat{f}_\theta(y) + (f_\theta x, y) + (f_\theta y, x)$$

$$\hat{f}_\theta(x-y) = \hat{f}_\theta(x) + \hat{f}_\theta(y) - (f_\theta x, y) - (f_\theta y, x).$$

But f_θ is Hermitian, i.e. $(f_\theta x, y) = (f_\theta y, x)$. So

$$(f_\theta x, y) = \frac{1}{4}[\hat{f}_\theta(x+y) - \hat{f}_\theta(x-y)].$$

Similarly

$$(Ax, Ay) = \frac{1}{4}[(A^*A(x+y), x+y) - (A^*A(x-y), x-y)].$$

Therefore

$$\begin{aligned}(N_\theta(x), y) &= \frac{(f_\theta x, y)}{a_\theta} - (A^*Ax, y) \\ &= \frac{1}{4}[\{\frac{\hat{f}_\theta(x+y)}{a_\theta} - (A^*A(x+y), x+y) - \{\frac{\hat{f}_\theta(x-y)}{a_\theta} - (A^*A(x-y), x-y)\}].\end{aligned}$$

Due to boundedness assumption 1.1 then

$$|(N_\theta(x),y)| \leq \frac{1}{4}[\{\frac{M(\theta)}{a_\theta}(A^*A(x+y),x+y) - (A^*A(x+y),x+y)\} -$$

$$\{\frac{m(\theta)}{a_\theta}(A^*A(x-y),x-y) - (A^*A(x-y),x-y)\}] \quad \text{a.e. } \theta$$

$$= \frac{1}{4}[\frac{M(\theta)-m(\theta)}{a_\theta}\{(A^*Ax,x) + (A^*Ay,y)\} + (\frac{M(\theta)+m(\theta)}{a_\theta} - 2) \times$$

$$\{(A^*Ax,y) + (A^*Ay,x)\}] \quad \text{a.e. } \theta .$$

Note that $\frac{M(\theta)+m(\theta)}{a_\theta} - 2 = 0$, which gives

$$|(N_\theta(x),y)| \leq \frac{M(\theta)-m(\theta)}{M(\theta)+m(\theta)}\{\frac{(A^*Ax,x) + (A^*Ay,y)}{2}\} \quad \text{a.e. } \theta$$

Thus

1.2.1. $$\|N_\theta\|_B \leq \frac{M(\theta)-m(\theta)}{M(\theta)+m(\theta)} \sup_{\substack{x:\|x\|=1\\ y:\|y\|=1}} \frac{\|Ax\|^2 + \|Ay\|^2}{2} \quad \text{a.e. } \theta$$

$$\leq \frac{M(\theta)-m(\theta)}{M(\theta)+m(\theta)} \quad \text{a.e. } \theta \quad (\text{since } \|A\|^2 = 1).$$

1.3. Lemma.

Let the spectral density f satisfy boundedness condition 1.1 and the image $A\mathcal{X}$ be dense in $\mathcal{K}$. If A is one-to-one onto $A\mathcal{X}$ then

(a) A^* is one-to-one

(b) $A^{*-1} = (A^{-1})^*$

and

(c) $|A^{*-1}N_\theta A^{-1}(k),\ell| \leq \frac{M(\theta)-m(\theta)}{M(\theta)+m(\theta)}\|k\|\,\|\ell\|$ a.e. θ for $k,\ell \in A\mathcal{X}$.

Proof.

(a) Since $A\mathcal{X}$ is dense in $\mathcal{K}$, A^* is defined on $\mathcal{K}$ to $\mathcal{X}^*$ as follows $\forall\ k \in \mathcal{K}$, $A^*(k) = x^*$ where $x^*(y) = (k,Ay)\ \forall\ y \in \mathcal{X}$.

Further, for each $k, k' \in K$

$$A^*(k) = A^*(k') \Leftrightarrow \forall y \in X,\ (k, Ay) = (k', Ay)$$

$$\Leftrightarrow \forall \ell \in AX,\ (k, \ell) = (k', \ell)$$

$$\Leftrightarrow \forall \ell \in K,\ (k, \ell) = (k', \ell)$$

(since AX is dense in K)

$$\Leftrightarrow k = k' .$$

Hence A^* is one-to-one.

(b) Range of $A^* = \{x^* \in X^* : \exists\ k \in K \text{ with } x^*(y) = (k, Ay)\ \forall y \in X\}$.

Also A^{-1} is defined on a dense subspace of K, and domain of $(A^{-1})^*$ is a subspace of X^*. In fact

$$\mathcal{D}((A^{-1})^*) = \{x^* \in X^* : \exists k \in K \text{ with } x^*(A^{-1}\ell) = (k, \ell)\ \forall \ell \in AX\}$$

$$= \{x^* \in X^* : \exists k \in K \text{ with } x^*(y) = (k, Ay)\ \forall y \in X\}$$

(since A is one-to-one onto AX from X).

Therefore $\mathcal{D}(A^{*-1}) = \mathcal{D}((A^{-1})^*)$.

Also for $x^* \in \mathcal{D}(A^{*-1})$

$$A^{*-1}(x^*) = k \Leftrightarrow x^*(y) = (k, Ay)\ \ \forall y \in X$$

$$\Leftrightarrow x^*(A^{-1}\ell) = (k, AA^{-1}\ell)\ \ \forall \ell \in AX$$

$$\Leftrightarrow x^*(A^{-1}\ell) = (k, \ell)\ \forall \ell \in AX$$

$$\Leftrightarrow (A^{-1})^*(x^*) = k.$$

(c) From 1.2.1 in the proof of the previous lemma 1.2, for $x, y \in X$

$$|N_\theta(x), y| \le \frac{M(\theta) - m(\theta)}{M(\theta) + m(\theta)} \frac{\|Ax\|^2 + \|Ay\|^2}{2} \text{ a.e. } \theta .$$

Thus for $k, \ell \in A(X)$

$$|(A^{*-1}N_\theta A^{-1}(k),\ell)| = |(N_\theta A^{-1}(k),A^{-1}\ell)|$$

$$\leq \frac{M(\theta) - m(\theta)}{M(\theta) + m(\theta)} \frac{\|AA^{-1}(k)\|^2 + \|AA^{-1}(\ell)\|^2}{2} \quad \text{a.e. } \theta$$

$$= \frac{M(\theta) - m(\theta)}{M(\theta) + m(\theta)} \frac{\|k\|^2 + \|\ell\|^2}{2} \quad \text{a.e. } \theta \, .$$

Thus

$$\|A^{*-1}N_\theta A^{-1}\|_B \leq \sup_{\substack{\|k\|=1 \\ \|\ell\|=1}} \frac{M(\theta) - m(\theta)}{M(\theta) + m(\theta)} \frac{\|k\|^2 + \|\ell\|^2}{2}$$

$$= \frac{M(\theta) - m(\theta)}{M(\theta) + m(\theta)} \quad \text{a.e. } \theta \, .$$

Hence the result.

1.4. <u>Relationship with the Case of Hilbert Space Valued Random Variables.</u>

<u>Main Theorem I.</u>

If the spectral density f_θ satisfies the boundedness condition and $A\colon \mathcal{X} \to \mathcal{K}$ is one-to-one and $A\mathcal{X}$ dense in $\mathcal{K}$ then there is a unique stationary stochastic process $\{\eta_n : -\infty < n < \infty\}$ which is $\mathcal{B}(\mathcal{K},\mathcal{K})$ valued and is such that

(i) $R_\eta(n) = A^{*-1}R_\xi(n)A^{-1}$ on $A\mathcal{X}$

(ii) $f_\eta(\theta) = \frac{2}{M(\theta) + m(\theta)} [I_{\mathcal{K}} + A^{*-1}N_\theta A^{-1}] = A^{*-1}f_\xi(\theta)A^{-1}$ on $A\mathcal{X}$

where $I_{\mathcal{K}}$ denotes the identity operator on $\mathcal{K}$ and $f_\xi(\theta)$ is f_θ in our previous notation.

<u>Proof.</u>

By Lemma 1.3 $A^{*-1}f_\xi(\theta)A^{-1} = \frac{2}{M(\theta) + m(\theta)} [I_{\mathcal{K}} + A^{*-1}N_\theta A^{-1}]$ is a bounded operator defined on $A\mathcal{X}$. Let g_θ denote its unique continuous extension to $\mathcal{K}$. g_θ is then a $B^+(\mathcal{K},\mathcal{K})$ valued function defined on the unit circle. Further $g(\theta)$ is strongly measurable since $f(\theta)$ is assumed to be so. Also by 1.3(c) $\|g_\theta\| \in L_1[0,2\pi)$. Hence g_θ is Bochner integrable. Thus for any n, $\xi_n A^{-1}$ defined on $A\mathcal{X}$ is such that $\forall\ k \in A\mathcal{X}$

$$\|\xi_n A^{-1}(k)\|^2 = \frac{1}{2\pi}\int_0^{2\pi} (A^{*-1} f_\theta A^{-1}(k), k)\,d\theta$$

$$= \frac{1}{2\pi}\int_0^{2\pi} (g_\theta(k), k)\,d\theta$$

$$\leq \|g\|_{L_1[0,2\pi)} \|k\|^2 < \infty.$$

Hence $\xi_n A^{-1}$ admits of a unique continuous extension to $\mathcal{K}$, say η_n. $\{\eta_n: -\infty < n < \infty\}$ is then a $\mathcal{B}(\mathcal{K},\mathcal{K})$ valued stochastic process. In fact $\{\eta_n: -\infty < n < \infty\}$ is stationary as shown in the following reasoning.

For $k, \ell \in \mathcal{K}$ we must show that $(\eta_n k, \eta_m \ell)$ depends only on $m-n$. Since $A(\mathcal{X})$ is dense in $\mathcal{K}$, $\exists$ sequences $\{x_p\}$, $\{y_q\}$ in $\mathcal{X}$ such that

$$A(x_p) \to k \text{ in } \mathcal{K}$$

and $$A(y_p) \to \ell \text{ in } \mathcal{K}.$$

Then, since η_n and η_m are bounded

$$(\eta_n k, \eta_m \ell) = \lim_{p\to\infty} (\eta_n(Ax_p), \eta_m(Ay_p))$$

$$= \lim_{p\to\infty} (\xi_n A^{-1}(Ax_p), \xi_m A^{-1}(Ay_p))$$

$$= \lim_{p\to\infty} (\xi_n x_p, \xi_m y_p)$$

$$= \lim_{p\to\infty} \frac{1}{2\pi}\int_0^{2\pi} e^{-i(n-m)\theta} (f_\theta x_p, y_p)\,d\theta$$

which depends on m and n only through n-m.

Furthermore

$$(\eta_n k, \eta_m \ell) = \lim_{p\to\infty} \frac{1}{2\pi} \int_0^{2\pi} e^{-i(n-m)\theta} (f_\theta A^{-1}(Ax_p), A^{-1}(Ay_p))\, d\theta$$

$$= \lim_{p\to\infty} \frac{1}{2\pi} \int_0^{2\pi} e^{-i(n-m)\theta} (A^{*-1} f_\theta A^{-1}(Ax_p), Ay_p)\, d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} e^{-i(n-m)} \{\lim_{p\to\infty} (A^{*-1} f_\theta A^{-1}(Ax_p), Ay_p)\}\, d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} e^{-i(n-m)\theta} (g_\theta(k), (\ell))\, d\theta,$$

the last two steps being true since g_θ is a bounded operator. Thus on $A(X)$

$$R_{\eta_n} = (\eta_n, \eta_0) = (\xi_n A^{-1}, \xi_0 A^{-1}) = A^{*-1}(\xi_n, \xi_0) A^{-1} = A^{*-1} R_{\xi_n} A^{-1}$$

$$f_{\eta_n}(\theta) = A^{*-1} f_{\xi_n}(\theta) A^{-1},$$

and due to continuity of all functions involved, R_{η_n} and $f_{\eta_n}(\theta)$ are the unique continuous extensions of $A^{*-1} R_{\xi_n} A^{-1}$ and $A^{*-1} f_{\xi_n}(\theta) A^{-1}$ respectively to K.

1.5. Factorization of the spectral density.

Corollary.

If $\phi_\theta: K \to K$ is the generating function given in [7] 7.3.5, for the $B(K,K)$ valued stationary stochastic process $\{\eta_n : -\infty < n < \infty\}$ and if f_θ satisfies the boundedness condition and $A: X \to K$ is one-to-one and AX dense in K then $\phi_\theta A: X \to K$ is such that $f_\theta = (\phi_\theta A)^*(\phi_\theta A)$.

Proof.

By Theorem 1.4, g_θ is the unique continuous extension of $A^{*-1}f_\theta A^{-1}$. So that

$$\begin{aligned} f_\theta &= A^* g_\theta A \\ &= A^* \phi_\theta^* \phi_\theta A = (\phi_\theta A)^* (\phi_\theta A). \end{aligned}$$

1.6. The Prediction Error Matrix and the Predictor for a Banach Space Valued Process.

For the $\mathcal{B}(\mathcal{K},\mathcal{K})$ valued stationary stochastic process $\{\eta_n: -\infty < n < \infty\}$ a schematic algorithm to obtain the prediction error matrix G_η and the linear predictor $\hat{\eta}_\nu$ of η_ν for $\nu > 0$ based on the past $\{\eta_n: n \leq 0\}$, is given in [7], Chapter VII. We shall now find the same for the $\mathcal{B}(\mathcal{X},\mathcal{K})$ valued process $\{\xi_n: -\infty < n < \infty\}$.

1.6.1. Notation.

For $\mathcal{S} \subset \mathcal{B}(\mathcal{X},\mathcal{K})$ let

$\bar{\sigma}(\mathcal{S})$ denote the smallest (strongly) closed subspace of $\mathcal{B}(\mathcal{X},\mathcal{K})$ containing the set $\{SB: S \in \mathcal{S}, B \in \mathcal{B}(\mathcal{X},\mathcal{X})\}$

$\sigma(\mathcal{S})$ denote the smallest closed subspace of $\mathcal{K}$ containing the set $\{Sx: S \in \mathcal{S}, x \in \mathcal{X}\}$.

We shall use the same notation also for subsets $\mathcal{S}$ of $\mathcal{B}(\mathcal{K},\mathcal{K})$. In this notation then

$$\bar{\sigma}(\mathcal{S}) = \mathcal{B}(\mathcal{X},\sigma(\mathcal{S})) \qquad \text{(Ref. [7], 3.2.3, Chapter III, p. 9).}$$

1.6.2. Definition.

For the stationary stochastic $B(\mathcal{X},\mathcal{X})$ valued process $\{\xi_n: -\infty < n < \infty\}$ define

$\mathcal{M}_n = \bar{\sigma}\{\xi_k : k \le n\}$, $M_n = \sigma\{\xi_k x : x \in X;\ k \le n\}$, $-\infty < n < \infty$

$\mathcal{M}_{-\infty} = \bigcap_n \mathcal{M}_n$, $\qquad M_{-\infty} = \bigcap_n M_n$

$\mathcal{M}_\infty = \bar{\sigma}\{\xi_k : -\infty < k < \infty\}$, $M_\infty = \sigma\{\xi_k x : x \in X,\ -\infty < k < \infty\}$.

Furthermore let $\mathcal{B}_n$, B_n for $-\infty \le n \le \infty$ denote the corresponding subspaces for the $\mathcal{B}(K,K)$ valued stationary process $\{\eta_n : -\infty < n < \infty\}$.

1.6.3. Main Theorem II.

The two stationary stochastic processes $\{\xi_n : -\infty < n < \infty\}$ and $\{\eta_n : -\infty < n < \infty\}$ are further related as follows

(i) For each integer $\nu > 0$, if $\hat{\xi}_\nu$ denotes the projection of ξ_ν on $\mathcal{M}_0$ and $\hat{\eta}_\nu$ the projection of η_ν on $\mathcal{B}_0$ then

$$\hat{\xi}_\nu = \hat{\eta}_\nu A.$$

(ii) $G_\xi = A^* G_\eta A$

(iii) $\hat{\xi}_\nu = \lim_{n\to\infty} \sum_{k=0}^{n} E_{\nu k} \xi_{-k}$

where $E_{\nu k}$ is the kth Fourier coefficient of $[e^{-i\nu\theta}\phi(e^{i\theta})]_{0+}\phi^{-1}$, ϕ being the generating function of the process $\{\eta_n : -\infty < n < \infty\}$.

Proof.

(i) Note that $\xi_k = \eta_k A$ for each k. Also AX is dense in K and η_k is bounded. Therefore for each k

$$\sigma\{\xi_k x : x \in X\} = \sigma\{\eta_k \ell : \ell \in K\}.$$

Now for each $x \in X$

$$\begin{aligned}
\hat{\eta}_\nu A(x) &= (\eta_\nu | \mathcal{B}_0)(Ax) \\
&= (\eta_\nu Ax | B_0) \\
&= (\xi_\nu A^{-1} Ax | B_0) \\
&= (\xi_\nu x | B_0) \\
&= (\xi_\nu x | M_0) \\
&= (\xi_\nu | \mathcal{M}_0)(x) \\
&= \hat{\xi}_\nu(x).
\end{aligned}$$

(ii) Now for each $x \in \mathcal{X}$, $y \in \mathcal{X}$

$$\begin{aligned}
(A^* G_\eta A(x))(y) &= ((\eta_1 - \hat{\eta}_1)(Ax), (\eta_1 - \hat{\eta}_1)(Ay)) \quad \text{(by definition of } G_\eta) \\
&= (\eta_1 Ax - \hat{\eta}_1 Ax, \eta_1 Ax - \hat{\eta}_1 Ay) \\
&= (\xi_1 A^{-1}(Ax) - \xi_1(x), \xi_1 A^{-1}(Ax) - \xi_1(y)) \\
&\qquad \text{(using } \eta_n = \xi_n A^{-1} \text{ on } A\mathcal{X} \text{ and (i))} \\
&= (\xi_1 x - \hat{\xi}_1 x, \xi_1 x - \hat{\xi}_1 x) \\
&= G_\xi(x).
\end{aligned}$$

(iii) For each integer $\nu > 0$,

$$\hat{\eta}_\nu(x) = \lim_{n\to\infty} \left(\sum_{k=0}^{n} E_{\nu k} \eta_{-k} \right)(x) \quad \text{Ref. [7], 7.4.11, Chapter VII, p. 108.}$$

Therefore

$$\hat{\xi}_\nu(x) = \hat{\eta}_\nu A(x) = \hat{\eta}_\nu(Ax)$$

$$= \lim_{n\to\infty} \left(\sum_{k=0}^{n} E_{\nu k}\eta_{-k}\right)(Ax)$$

$$= \lim_{n\to\infty} \left(\sum_{k=0}^{n} E_{\nu k}\xi_{-k}A^{-1}\right)(Ax)$$

$$= \lim_{n\to\infty} \sum_{k=0}^{n} E_{\nu k}\xi_{-k}A^{-1}(Ax)$$

$$= \lim_{n\to\infty} \sum_{k=0}^{n} E_{\nu k}\xi_{-k}(x).$$

1.7. Note.

For results in this chapter the boundedness assumption 1.1 was made on the spectral density of the process and it was further assumed that the map $A: X \to K$ be one-to-one with the image of X dense in K. The restriction of AX being dense in K is easily deleted by replacing K by the Hilbert space H generated by AX in defining the process $\{\eta_n: -\infty < n < \infty\}$. Generalization when A is not one-to-one calls for a closer look and may be handled as follows: Let $K(P)$ denote the kernel of any operator P. Then due to boundedness assumption 1.1

1.7.1 $$K(A) = K(\hat{f}_\theta) \quad \text{a.e. } \theta$$

where $\hat{f}_\theta$ denotes the quadratic form of f_θ. Let the quotient space, denoted by $\tilde{X}$, be such that

$$\forall\ x \in X,\ \|\tilde{x}\| = \inf_{\delta \in K(A)} \|x - \delta\| = d(x, K(A))$$

where $\tilde{x}$ is the equivalence class $x + K(A)$ of elements of x. Now $(\tilde{X}, \|\ \|)$ is a Banach space. Ref. [3], p. 140.

The linear map A_Q defined on it as follows

$$A_Q(\tilde{x}) = Ax \quad \text{for} \quad \tilde{x} \in \tilde{X}$$

is continuous in the norm of $\tilde{X}$. This is shown as follows

$$\|A_Q\| = \sup_{\|\tilde{x}\|=1} \|A_Q(\tilde{x})\| = \sup_{x:d(x,K(A))=1} \|Ax\| .$$

Also for each $x \in \tilde{X}$ with $d(x,K(A)) = 1$, $x = x + \delta - \delta$ whatever $\delta \in K(A)$. So

$$\|Ax\| \leq \inf_{\delta \in K(A)} \{\|A(x + \delta)\| + \|A\delta\|\} \leq \inf_{\delta \in K(A)} \{\|x + \delta\|\} \leq 1$$

since $A\delta = 0$ for $\delta \in K(A)$ and $\|A\| = 1$.

Hence A_Q is continuous. Furthermore A_Q is such that

$$(A_Q^* A_Q(\tilde{x}), \tilde{y}) = (Ax, Ay) = (A^* Ax, y) \quad \text{for} \quad x, y \in X.$$

Thus

$$m(\theta) A_Q^* A_Q \leq f_\theta \leq M(\theta) A_Q^* A_Q \quad \text{a.e.} \quad \theta$$

and A_Q: $\tilde{X} \to K$ is one-to-one, and without loss of generality $A_Q(\tilde{X})$ is dense in K.

To make sense of the definition of $\xi_n A_Q^{-1}$ on the image of $\tilde{X}$ under A_Q, we must have ξ_n uniquely defined on $\tilde{X}$. It is here that we would need the assumption of linearity of ξ_n. Let, for x and y in X, $Ax = Ay$. Then

$$\|\xi_n(x) - \xi_n(y)\|^2 = \|\xi_n(x - y)\|^2$$

$$= (\xi_n(x - y), \xi_n(x - y))$$

$$= (\xi_0(x - y), \xi_0(x - y))$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \hat{f}_\theta(x - y) d\theta$$

$$= 0 \qquad \text{(due to 1.7.1).}$$

So if $x = y \bmod K(A)$ then $\xi_n(x) = \xi_n(y)$. Hence $\forall\, x \in X$ we may define $\xi_n(\tilde{x}) = \xi_n(x)$. And the preceding procedures now apply to $\xi_n A_Q^{-1}$ to ultimately yield the prediction error matrix and the predictor for the process $\{\xi_n: -\infty < n < \infty\}$.

CHAPTER IV

ON LINEARIZING STATISTICS OF TIME-SERIES FOR NONLINEAR PREDICTION

The problem of nonlinear prediction of a stationary stochastic process was dealt with, in the second chapter of this thesis at the level of defining the predictor and showing that we have a determinate mathematical problem and then obtaining the predictor by a more or less direct attack. In order to utilize the time and spectral domain analysis to obtain an algorithm for the predictor at a more efficient level, N. Weiner in [18] suggests a method of relating the nonlinear prediction problem to linear prediction of an infinite-variate stationary stochastic process. In this chapter we will explore this relationship further.

1. The Related Infinite-Variate Process

Let $\{f_n : -\infty < n < \infty\}$ be a univariate strictly stationary stochastic process defined on a probability space (Ω, F, P), with zero expectations. Let T denote the shift associated with the process. Further assume that there exists $a > 0$ such that $|f_0| \leq a$.

Let $H = \{ \prod_{i \in I} f_i : I \subset \{0,-1,-2,\dots\}$ and I finite$\}$, where elements of H are written as products of f_i with decreasing indices. Let $\{\phi_{j'} : 1 \leq j' < \infty\}$ be a linear arrangement of H as in 5.1 of Chapter II. Let $\{\phi_j : 1 \leq j < \infty\}$ be the subfamily $\{\phi_{j'} : j'$ odd$\}$. Note that for j odd, $j + 1$ contains a factor of 2 and $p_0 = 2$ being the first prime, the family $\{\phi_j : 1 \leq j < \infty\}$ consists of those finite products of $\{f_i : i \leq 0\}$, written in descending order, which begin with f_0.

Let

$$X_0 = \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_i \\ \vdots \end{pmatrix}$$

Due to duplication of subfactors, for $\omega \varepsilon \Omega$, $X_0(\omega)$ may not be square summable. Now the shift T acting successively on any finite product $\prod_{i \varepsilon I} f_i$ defines a stochastic process in its own right since

$$T^n (\prod_{i \varepsilon I} f_i) = \prod_{i \varepsilon I} f_{i+n}.$$

(Ref. Thm. 1.3, Chapter I).

Define for $-\infty < n < \infty$

$$X_n = \begin{pmatrix} T^n \phi_0 \\ T^n \phi_1 \\ \vdots \\ T^n \phi_i \\ \vdots \end{pmatrix}$$

1.1. Notation

Let X_{nk} denote the k^{th} coordinate of X_n for $0 \leq k < \infty$.

Let $\mathcal{M}_k$ denote the family of square integrable functions which are measurable w.r.t. the σ-algebra generated by $\{f_n : n \leq k\}$, along with all null sets $0 \leq k < \infty$. Also let M_k denote the closed manifold of $L_2(\Omega)$ generated by the co-ordinate functions $\{X_{ni} : 0 \leq i < \infty, n \leq k\}$.

1.2. Main Theorem

$\{X_n : -\infty < n < \infty\}$ is a Banach space ℓ^∞- valued strictly stationary stochastic process with shift which is the inflation of T. Furthermore this process is related to the univariate process $\{f_n : -\infty < n < \infty\}$ as follows

(i) for $-\infty < k < \infty$

$$\mathcal{M}_k = M_k,$$

(ii) for $\nu > 0$

$$(f_\nu | M_0) = \hat{X}_{\nu 0}.$$

where $\hat{X}_\nu$ denotes the linear predictor of X_ν based on the present and past of the process $\{X_n : -\infty < n < \infty\}$.

Proof

(i) The polynomials in $\{f_n : n \leq k\}$ form a dense subset of M_k. M_k, on the other hand, is the closure of finite linear combinations of products from $\{f_n : n \leq k\}$.

Both M_k and M_k therefore, contain a common dense set. Both being closed, it follows that

$$M_k = M_k.$$

(ii) then follows from the definition of predictors.

1.3. MAIN RESULT I .

The nonlinear predictor $(f_\nu | M_0)$ may thus be obtained as the 0^{th} coordinate of the linear predictor $\hat{X}_\nu$ of the ℓ^∞-valued strictly stationary random process $\{X_n : -\infty < n < \infty\}$.

1.4 Let C_0 denote the set of all infinite vector sequences that tend to zero. C_0 is then a (separable) Banach space and its dual space ℓ' is separable.

1.4.1. MAIN RESULT II

Under the assumption that $X_n \in C_0$, the process $\{\xi_n : -\infty < n < \infty\}$ defined from the separable Banach space $C_0^* = \ell^1$ to $L^2(\Omega)$ as follows

$$\forall\ x \in \ell^1,\ \xi_n(x)\ (\omega) = x(X_n(\omega))$$

may be identified with the process $\{X_n : -\infty < n < \infty\}$.

This identification yields the linear predictor by methods used in Chapter III, under the boundedness condition on the spectral density of $\{\xi_n : -\infty < n < \infty\}$. The problem of translating this condition in terms of moments of the process $\{f_n : -\infty < n < \infty\}$ is still open.

An attempt to obtain extension of the work of Robertson Ref., [10], [11], has not been made here. This is to be investigated later.

2. Special Case

We shall, now, consider a particular stationary stochastic process $\{f_n : -\infty < n < \infty\}$ and its related infinite variate process $\{X_n : -\infty < n < \infty\}$ to shed some light on the difficulty in translating the condition of boundedness of the spectral density of $\{X_n : -\infty < n < \infty\}$ in terms of moments of $\{f_n : -\infty < n < \infty\}$. Let $\{f_n : -\infty < n < \infty\}$ be a two state stationary Markov process with state space $\{1, -1\}$ and the transition matrix

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix}$$

such that $P[X_n = 1] = P[X_n = -1] = \frac{1}{2}$ for $-\infty < n < \infty$.

Then for each $n > 0$, the n step transition matrix is

$$P^n = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}(2\alpha-1)^n & \frac{1}{2} - \frac{1}{2}(2\alpha-1)^n \\ \frac{1}{2} - \frac{1}{2}(2\alpha-1)^n & \frac{1}{2} + \frac{1}{2}(2\alpha-1)^n \end{bmatrix}$$

All moments of finite products of $\{f_n\}_{-\infty}^{\infty}$ are then obtainable as follows. Since for any integer $\nu > 0$, f_n^{ν} is 1 or f_n according as ν is even or odd respectively, to obtain moments of finite products of $\{f_n\}_{-\infty}^{\infty}$ we need just evaluate $E(f_{n_1} f_{n_2} \dots f_{n_k})$ for $-\infty < n_1 < n_2 < \dots < n_k < \infty$. Notice that for an integer $\nu > 0$

$$E(f_\nu|f_0=1) - E(f_\nu|f_0=-1) = P^\nu(1,1) - P^\nu(1,-1) -\{P^\nu(-1,1) - P^\nu(-1,-1)\}$$

$$= \frac{1}{2} + \frac{1}{2}(2\alpha-1)^\nu -\{\frac{1}{2} - \frac{1}{2}(2\alpha-1)^\nu\} - \{\frac{1}{2} - \frac{1}{2}(2\alpha-1)^\nu - \frac{1}{2} - \frac{1}{2}(2\alpha-1)^\nu\}$$

$$= 2\ (2\alpha-1)^\nu.$$

Similarly

$$E(f_\nu|f_0=1) + E(f_\nu|f_0 = -1) = P^\nu(1,1) - P^\nu(1,-1) + P^\nu(-1,1) - P^\nu(-1,-1) = 0.$$

Now

$$E(f_0 f_{n_1} f_{n_2} \cdots f_{n_k}) = E(f_{n_1} f_{n_2} \cdots f_{n_k}|f_0 = 1)\ P(f_0 = 1) -$$

$$E(f_{n_1} f_{n_2} \cdots f_{n_k}|f_0 = -1)\ P\ (f_0 = -1)$$

$$= \frac{1}{2}[E(f_{n_1} f_{n_2} \cdots f_{n_k}|f_0 = 1) - E(f_{n_1} f_{n_2} \cdots f_{n_k}|f_0 = -1)]$$

$$= \frac{1}{2}[E(f_{n_2} \ldots f_{n_k}|f_{n_1} = 1,\ f_0 = 1)\ P(f_{n_1} = 1|f_0=1)$$

$$-E(f_{n_2} \ldots f_{n_k}|f_{n_1} =-1, f_0=1)\ P(f_{n_1} =-1|f_0=1)$$

$$-E(f_{n_2} \ldots f_{n_k}|f_{n_1} =1, f_0=-1)\ P(f_{n_1} =1|f_0 = -1)$$

$$+E(f_{n_2} \ldots f_{n_k}|f_{n_1} = -1,\ f_0=-1)\ P(f_{n_1}=-1|f_0=-1)]$$

$$= \frac{1}{2}[E(f_{n_2} \ldots f_{n_k}|f_{n_1} =1)\ P^{n_1}\ (1,1)-E(f_{n_2} \ldots f_{n_k}|f_{n_1}= -1)P^{n_1}(-1,1)$$

$$-E(f_{n_2} \ldots f_{n_k}|f_{n_1} =1)\ P^{n_1}(-1,1)+E(f_{n_2}\ldots f_{n_k}|f_{n_1} =-1)P^{n_1}\ (-1,-1)\ ]$$

$$= \frac{1}{2}[E(f_{n_2}\cdots f_{n_k}|f_{n_1}=1)\ \{P^{n_1}(1,1)-P^{n_1}(-1,1)\} + E(f_{n_2}\cdots f_{n_k}|f_{n_1}=-1)$$

$$\times\{P^{n_1}(-1,-1)-P^{n_1}(-1,1)\}]$$

$$= \frac{1}{2}[E(f_{n_2}\cdots f_{n_k}|f_{n_1}=1)\ (2\alpha-1)^{n_1} + E(f_{n_2}\cdots f_{n_k}|f_{n_1}=-1)(2\alpha-1)^{n_1}]$$

$$= \frac{1}{2}\ (2\alpha-1)^{n_1}\ [E(f_{n_2}\cdots f_{n_k}|f_{n_1}=1) + E(f_{n_2}\cdots f_{n_k}|f_{n_1}=-1)].$$

Again repeating the same process

$$E(f_0 f_{n_1}\cdots f_{n_k}) = \frac{1}{2}(2\alpha-1)^{n_1}\ [E(f_{n_3}\cdots f_{n_k}|f_{n_2}=1)\ \{P^{n_2-n_1}(1,1)+P^{n_2-n_1}(-1,1)\}$$

$$-E(f_{n_3}\cdots f_{n_k}|f_{n_2}=-1)\ \{P^{n_2-n_1}(1,-1)+P^{n_2-n_1}(-1,-1)\ ]$$

$$= \frac{1}{2}(2\alpha-1)^{n_1}\ [E(f_{n_3}\cdots f_{n_k}|f_{n_2}=1)\{\ \frac{1}{2} + (2\alpha-1)^{n_2-n_1} + \frac{1}{2} - \frac{1}{2}(2\alpha-1)^{n_2-n_1}\}$$

$$-E(f_{n_3}\cdots f_{n_k}|f_{n_2}=-1)\ \ \frac{1}{2} - \frac{1}{2}\ (2\alpha-1)^{n_2-n_1} + \frac{1}{2} - \frac{1}{2}\ (2\alpha-1)^{n_2-n_1}\}$$

$$= \frac{1}{2}(2\alpha-1)^{n_1}\ [E(f_{n_3}\cdots f_{n_k}|f_{n_2}=1) - E(f_{n_3}\cdots f_{n_k}|f_{n_2}=-1)]$$

Proceeding thus we obtain

$$E(f_0 f_{n_1}\cdots f_{n_k}) = \frac{1}{2}(2\alpha-1)^{n_1}\ (2\alpha-1)^{n_3-n_2}\cdots[E(f_{n_k}|f_{n_{k-1}}=1)+(-1)^k E(f_{n_k}|f_{n_{k-1}}=-1)]$$

$$= \frac{1}{2}\{(2\alpha-1)^{n_1+(n_3-n_2)+\cdots}\}\times 2\ (2\alpha-1)^{n_k-n_{k-1}}\ \text{if } k \text{ is odd}$$

$$0 \text{ if } k \text{ is even}$$

$$= (2\alpha-1)^{n_1+n_3+\ldots+n_k-(n_2+n_4+\ldots+n_{k-1})} \quad \text{if } k \text{ is odd}$$

$$0 \text{ if } k \text{ is even}$$

We will actually consider the corresponding prediction problem for $\{g_n: -\infty<n<\infty\}$ where $g_n = \delta f_n$ for some fixed δ with $0<\delta<1$. Now if the condition of boundedness of the spectral density of the related infinite variate process $\{X_n: -\infty<n<\infty\}$ was satisfied, we could determine the nonlinear predictor for the process $\{g_n: -\infty<n<\infty\}$, hence for the process $\{f_n: -\infty<n<\infty\}$.

The spectral density for the process $\{X_n: -\infty<n<\infty\}$ is given as follows:

$$f_\theta(i,j) = \sum_{h=-\infty}^{\infty} E(X_{0i}X_{hj})e^{i\theta h} \qquad 0\leq i,j<\infty$$

where $X_{hj} = T^h X_{0j}$ is the h^{th} shift of the j^{th} coordinate of X_0, T being the shift for the stationary stochastic process $\{g_n: -\infty<n<\infty\}$

Let $X_{0i} = g_0 g_{n_1}\cdots g_{n_I}$ and $X_{0j} = g_0 g_{m_1}\cdots g_{m_J}$.

Then

$$f_\theta(i,j) = \sum_{h=-\infty}^{\infty} E(g_0 g_{n_1}\cdots g_{n_I}\, g_n g_{m_{1+h}}\cdots g_{m_J+h})\, e^{i\theta h}$$

$$= \sum_{|h|\leq n_I} E(X_{0i}\, X_{hj})\, e^{i\theta h} + \sum_{|h|>n_I} E(g_0 g_{n_1}\cdots g_{n_I} g_h g_{m_1+h}\cdots g_{m_J+h})e^{i\theta h}$$

$$= \sum_{|h|\le n_I} E(X_{0i}\, X_{hj})\, e^{i\theta h} + 0, \text{ if } I + J \text{ is odd}$$

$$\sum_{|h|\le n_I} E(X_{0i}\, X_{hj})\, e^{i\theta h} + \sum_{|h|>n_I} (2\alpha-1)^{n_1+n_3-n_2+\ \ldots+n_I-n_{I-1}+m_1+m_3-m_2+\ldots+m_J-m_{J-1}}$$

if I and J are both odd

$$\sum_{|h|\le n_I} E(X_{0i}\, X_{hj})\, e^{i\theta h} + \sum_{|h|>n_I} \{(2\alpha-1)^{n_1+n_3-n_2+\ldots+(h-n_I)+(m_2+h-m_1-h)+\ldots+(m_J+h-m_{J-1}-h)}$$

$$\times\ e^{i\theta h}\}$$ if I and J are both even

$$= \sum_{|h|\le n_I} E(X_{0i}\, X_{hj})\, e^{i\theta h} \quad \text{if } I + J \text{ is odd}$$

$$\sum_{|h|\le n_I} E(X_{0i}\, X_{hj})\, e^{i\theta h} + c(i,j) \sum_{|h|>n_I} e^{i\theta h} \quad \text{if } I,\ J \text{ odd}$$

$$\sum_{|h|\le n_I} E(X_{0i}\, X_{hj})\, e^{i\theta h} + c(i,j) \sum_{|h|>n_I} (2\alpha-1)^h\, e^{i\theta h} \quad \text{if } I,\ J \text{ even,}$$

where $c(i,j)$ is the appropriate constant bounded by 1.

The spectral density in all these cases is bounded above by some $M(\theta)$ which is integrable since

$$\text{(i)} \left| \sum_{|h|\le n_I} E(g_0 g_{n_1} \cdots g_{n_I} g_h g_{m_I+h})\, e^{i\theta h} \right|$$

$$\le \sum_{|h|\le n_I} \delta^{n_I}\, |e^{i\theta h}| = 2n_I \delta^{n_I}$$

(ii) $c(i,j) \sum_{|h|>n_I} e^{i\theta h} = 2\, c(i,j) \dfrac{e^{i(n_I+1)\theta}}{1 - e^{i\theta}}$

which is integrable on the unit circle,

(iii) $$|c(i,j) \sum_{|h|>n_I} (2\alpha-1)^h e^{i\theta h}| \leq c(i,j)\, 2(2\alpha-1)^{n_I+1} \sum_{h>0} (2\alpha-1)^h$$

$$= \frac{2c(i,j)\ (2\alpha-1)^{n_I+1}}{1-(2\alpha-1)}$$

$$= \frac{c(i,j)\ (2\alpha-1)^{n_I+1}}{1-\alpha}$$

The spectral density, however, is not bounded away from 0. Consider the diagonal element $f_\theta(i,i)$ where X_{0i} is $g_0 g_1 \ldots g_I$.

Then

$$|\sum_{|h|\leq I} E\, X_{0i}\, X_{hj}\ e^{i\theta h}| \leq \delta^{2I} \times 2I$$

which tends to 0 as I tends to ∞.

Also

$|\sum_{|h|>I} E\, X_{0i}\, X_{hj}\, e^{i\theta h}$ tends to 0 as I tends to ∞.

Thus there exists a subsequence $\{i_n\}$ such that $f_\theta\ (i_n, i_n)$ tends to 0 as n tends to ∞. Hence f_θ cannot be bounded away from zero.

So if we consider the infinite variate process $\{\frac{1}{2|n|}\ g_n: -\infty<n<\infty\}$, its spectral density will also not be bounded away from zero.

Remark.

The above example shows that a non-deterministic stochastic process with bounded spectral density may very well have its spectral density not bounded away from zero. Since the boundedness condition 1.1 of Chapter III is not satisfied, to use Wiener and Masani's algorithm for computing the predictor one must prove a stronger version of our algorithm techniques given in Chapter III which would be valid in the absence of a positive lower bound. Once such an algorithm is available, examples such as above can be more fully investigated.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Doob, J.L., Stochastic Processes, Wiley, 1960.

2. Gangolli, R., Wide sense stationary sequences of distributions on Hilbert space and the factorization of operator valued functions. J. Math. Mech. 12(1963), 893-910.

3. Kato, T., Perturbation Thoery for Linear Operators, Springer-Verlag New York Inc., 1960.

4. Mandrekar, V. and Salehi, H., The square integrability of operator-valued functions with respect to a non-negative operator-valued measure and Kolmogorov isomorphism theorem. Indiana Univ. Math. J. 20(1970), 545-563.

5. Masani, P., Recent trends in Multivariate Prediction Theory, Multivariate Analysis, Academic Press, New York, 1966, 351-382.

6. ________, The Prediction Theory of Multivariate Stochastic Processes, III, Acta Math. 104(1960), 142-162.

7. Miamee, A.G., Banach Space Valued Stationary Stochastic Processes and factorization of nonnegative operator valued functions on a Banach space. Ph.D. thesis, MSU (1973). (To be published in Sankhyā).

8. Nadkarni, M.G., Prediction theory of infinite variate weakly stationary stochastic processes. Sankhyā, Series A, Vol. 32, Part 2 (1970), 145-172.

9. ________, On a paper of Remesh Gangolli, J. Math. Mech. 17(1967), 403-405.

10. Robertson, J.B., An algorithm for computing the non-linear predictor, Ann. Math. Statist. 42(1971), 75-80.

11. ________, On a problem in nonlinear prediction theory. Ann. Math. Statist. 34(1965), 1554-1555.

12. Rosenblatt, M., Stationary processes as shifts of functions of independent random variables, J. Math. Mech., 8(1959), 665-681.

13. Rozanov, Yu, Some approximation problems in the theory of stationary processes, J. Multivariate Anal., 2(1972), 135-144.

14. Salehi, H., On determination of the optimal factor of a non-negative matrix valued function, Proc. Amer. Math. Soc. 29(1971), 383-389.

15. Shohat, J.A. and Tamashin, J.D., The problem of moments, American Mathematical Society Surveys, No. 1, New York (1943).

16. Sobel, E., Prediction of a noise distorted mutlivariate non-stationary signal. J. Appl. Prob. 4(1967), 330-342.

17. Spitzer, F., Random fields and interacting particle systems, Notes on Lectures given at the 1971 MAA summer seminar.

18. Wiener, N., Nonlinear prediction and dynamics. Proc. of Third Berkeley Symposium on Math. Stat. and Prob. III(1956), 247-252.

19. Wiener, N. and Masani, P., Nonlinear Prediction, Probability and Statistics, The Harald Cramer Volume, Ed. by Ulf Grenader, Stockholm (1959), 190-212.

20. ________, The prediction theory of multivariate stochastic processes, I., Acta Math. 98(1957), 111-150.

21. ________, The prediction theory of multivariate stochastic processes, II., Acta Math. 99(1958), 93-137.

22. Yosida, K., Functional Analysis, Springer-Verlag New York Inc., 1971.