Probab. Th. Rel. Fields 79, 145–164 (1988)

Probability Theory and Related Fields

Wold Decomposition, Prediction and Parameterization of Stationary Processes with Infinite Variance

A. G. Miamee* and M. Pourahmadi

Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115, USA

Summary. A discrete time stochastic process $\{X_t\}$ is said to be a p-stationary process $(1<p\leqq 2)$ if $E\left|\sum_{k=1}^{n} b_k X_{t_k+h}\right|^p = E\left|\sum_{k=1}^{n} b_k X_{t_k}\right|^p$, for all integers $n\geqq 1$, $t_1, \ldots, t_n, h$ and scalars $b_1, \ldots, b_n$. The class of p-stationary processes includes the class of second-order weakly stationary stochastic processes, harmonizable stable processes of order α $(1<\alpha\leqq 2)$, and p^{th} order strictly stationary processes. For any nondeterministic process in this class a finite Wold decomposition (moving average representation) and a finite predictive decomposition (autoregressive representation) are given without alluding to any notion of "covariance" or "spectrum". These decompositions produce two unique (interrelated) sequences of scalar which are used as parameters of the process $\{X_t\}$. It is shown that the finite Wold and predictive decomposition are all that one needs in developing a Kolmogorov-Wiener type prediction theory for such processes.

1. Introduction

In recent years there has been considerable interest in developing a prediction theory for and analyzing data from stochastic processes with infinite variance, cf. [2–7, 14, 19] and references therein. The notion of covariance and spectral distribution functions can not be defined properly for such processes. This makes it difficult to construct a spectral-domain and even harder to establish a correspondence between time and spectral domains. Despite this, there has been attempts to mimick the procedure of Kolmogorov-Wiener theory in developing a prediction theory for processes with infinite variance. This is usually done by introducing (pseudo-) covariance and (pseudo-) spectral distribution function for the process under study, cf. [3–5, 9, 19].

The success of the Kolmogorov-Wiener theory of prediction for weakly stationary processes, looking at it in retrospect, can be attributed to two important properties which seem to be valid only for such processes and are intrinsic to the

* *Current address:* Department of Mathematics, Hampton University, Hampton, VA 23668 USA

Hilbertian structure of the time-domain. The first property is that the innovation process $\{\varepsilon_t\}$ is a white noise or orthogonal process, and this is essential in establishing the Wold decomposition Theorem, cf. Sects. 2, 4 and 5. The second property is that for such processes the notions of covariance and spectral distribution function F are defined. Thus, it is possible to define the spectral domain $L^2(F)$ and establish the Kolmogorov isomorphism between time and spectral domains, and use the spectral technique (harmonic analysis) in the development of prediction theory.

The main purpose of this paper is to develop a time-domain theory of prediction for stochastic processes without alluding to any notion of covariance or spectral distribution function even when a genuine covariance or spectral distribution function exists. We rely merely on the linearity of the time-domain and a basic consequence of the notion of linear independence for vectors in a linear space. The core of our approach is the observation that for each t, the closed linear span of the past values of the process, i.e. $H_t = \overline{sp}\{X_s : s \leqq t\}$ can be generated simultaneously by

$$H_t = \overline{sp}\{X_t, H_{t-1}\} = \overline{sp}\{\varepsilon_t, H_{t-1}\},$$

for notation see Sects. 2 and 3. This observation combined with a regression-typed lemma (proved in Sect. 4) give all the results of this paper and they render our approach a regression analytic outlook; this is in accord with the seminal work of Wold [22, Sects. 18, 19, 20, 26] where the idea of multiple linear regression were used to obtain structural representations for stationary processes.

The outline of the paper is as follows. In Sect. 2 a summary of known results concerning prediction of a weakly stationary process is given. The notion of p-stationary process is defined in Sect. 3. Note that such a process may have infinite variance when $p<2$. We show that a p-stationary process has a shift operator U which is an isometric isomorphism from $H(X)$ onto $H(X)$. Also, the predictor process $\{\hat{X}_{t,v}\}$, $v \geqq 1$ fixed, and the innovation process $\{\varepsilon_t\}$ have the same shift U as $\{X_t\}$ itself.

In Sect. 4 by using our Main Lemma 4.3 we find (finite) Wold decomposition (moving average representation) and (finite) predictive decomposition (autoregressive representation) for a nondeterministic p-stationary process, cf. Theorem 4.4.

The second-order properties of a weakly stationary process are determined by its covariance function $\{\gamma_k\}$ or spectral distribution function F and as such either of them provides a natural parametrization for such processes. Of course, for stationary processes with infinite variance there is not such a parametrization. Due to the importance of parameterization in statistical analysis of data it is desirable to have an alternative parameterization for stationary processes with infinite variance. As a by product of our approach, we associate to each stationary process with infinite variance two unique (interrelated) sequences of scalars $\{c_k\}_{k=1}^{\infty}$ and $\{a_k\}_{k=1}^{\infty}$ so that each of them can provide a parameterization for the process under study. The function $\phi(z) = 1 + \sum_{k=1}^{\infty} c_k z^k$ defined, in the open disc with radius 1/2 in the complex-plane, in terms of the c_k's plays roles very similar to the roles of the optimal factor of the spectral density for weakly stationary processes. Since $\phi(0) = 1 \neq 0$, the function $1/\phi(z)$ is analytic in a neighborhood of 0 and has a Taylor expansion with coefficients $d_0=1, d_1, d_2, \ldots$ In Sect. 4, it is shown that $a_k = -d_k, k=1,2,\ldots$ and that the coefficients of the v-step ahead predictor of $\{X_t\}$ can be expressed in terms of the c_k's and d_k's, cf. Theorem 4.18 and Corollary 4.22. These results confirm the distinguished role of the sequence $\{c_k\}$ as the parameter of such stochastic processes. At this point it is important to note that we have obtained all these results for a p-stationary process and its predictor under the mere assumption that $\{X_t\}$ is nondeterministic. In this, the Wold decomposition did not play any role, but its place is taken by finite Wold and predictive decompositions of $\{X_t\}$.

For weakly stationary processes Wold decomposition plays a crucial role in characterizing purely nondeterministic processes as a one-sided moving average of their innovation processes. To obtain such characterization for a purely nondeterministic p-stationary process, the possibility of establishing Wold and predictive decomposition for such processes is studied in Sect. 5. In Theorem 5.1, a simple sufficient condition is given for the convergence of $\sum_{k=1}^{\infty} c_k \varepsilon_{t-k}$ and $\sum_{k=1}^{\infty} a_k X_{t-k}$ in the mean of order p. Theorem 5.2 provides Wold and predictive decompositions for a nondeterministic p-stationary process under the assumption of convergence of these series. As a corollary of this theorem, it is shown that a purely nondeterministic p-stationary process has a one-sided moving average representation in terms of its innovation process, and an autoregressive representation.

It is certainly of interest to know whether a p-stationary process which has a one-sided moving average representation in terms of its innovation process is purely nondeterministic. Although this question has a positive answer, unlike the case of weakly stationary process, its proof requires a careful analysis of every element of $H_t(X)$ and more powerful tools than Lemma 4.3. Our work on this and some other related problems will appear elsewhere.

The paper is self-contained and Lemma 4.3 is the only prerequisite for solving the prediction problem via our approach. When applied to weakly stationary processes one obtains the most general results about the structure and predictor of such processes rather quickly without using any standard analytical tools. It should be noted that the structural results given here are useful for actual prediction in certain circumstances, for example when the process under study has an autoregressive representation, cf. Sect. 5.

2. Summary of Some Results on Prediction of Stationary Processes

In this section for ease of reference and comparison we summarize some of the known results concerning the prediction of a weakly stationary stochastic process and explicit computation of the best linear predictors. For more information and proofs see [12, 21].

Let (Ω, F, P) be a probability space and $L^2(\Omega)=L^2(\Omega, F, P)$. A zero-mean, stochastic process $\{X_t\}$ is said to be a weakly stationary stochastic process if

$$X_t \in L^2(\Omega), \quad \text{for all} \quad t,$$

and, for all s, t

$$\text{cov}(X_s, X_t) = (X_s, X_t) = EX_s\bar{X}_t \tag{2.1}$$

depends only on $s-t$. In this case the function $\{\gamma_k\}$ defined by

$$\gamma_k = (X_{t+k}, X_t), \ k=0, \ \pm 1, \ +2, \ldots$$

is referred to as the *covariance function* of $\{X_t\}$. It is well-known that $\{\gamma_k\}$ has the spectral representation

$$\gamma_k = \int_{-\pi}^{\pi} e^{-ik\lambda} F(d\lambda), \tag{2.2}$$

and F is referred to as the *spectral distribution function* of the processes $\{X_t\}$. To each stochastic process $\{X_t\}$ with $X_t \in L^2(\Omega)$ we associate the following important subspaces:

$$\begin{aligned} &H(X) = \overline{\text{sp}}\,\{X_s\colon \text{all integers}\}, \ \textit{time-domain} \text{ of } \{X_t\}, \\ &H_t = H_t(X) = \overline{\text{sp}}\,\{X_s;\ s \leqq t\}, \ \textit{past and present} \text{ (up to } t\text{) of } \{X_t\}, \\ &H_{-\infty} = H_{-\infty}(X) = \bigcap_{t \leqq 0} H_t(X), \ \textit{remote past} \text{ of } \{X_t\}, \end{aligned} \tag{2.3}$$

where $\overline{\text{sp}}\,\{\ldots\}$ stands for the closed linear span of elements of $\{\ldots\}$ in the norm of $L^2(\Omega)$. The non-decreasing chain of subspaces $H_t(X)$, $-\infty < t < \infty$, contains considerable information about the linear structure of $\{X_t\}$. It is said that $\{X_t\}$ is *deterministic* if

$$H_{t-1}(X) = H_t(X), \quad \text{for one and hence all } t.$$

Otherwise, $\{X_t\}$ is said to be *nondeterministic*. Thus, for a nondeterministic process we have

$$X_t \notin H_{t-1}(X) \quad \text{or} \quad H_{t-1}(X) \subsetneqq H_t(X). \tag{2.4}$$

A nondeterministic process is said to be *purely nondeterministic (regular)* if

$$H_{-\infty}(X) = \{0\}. \tag{2.5}$$

For a nondeterministic process the *best linear predictor* of X_{t+v}, $v \geqq 1$, based on its infinite past $X_t, X_{t-1}, \ldots$ is denoted by $\hat{X}_{t,v}$ and is given as the orthogonal projection of X_{t+v} onto the subspace $H_t(X)$. Thus, $\hat{X}_{t,v} \in H_t(X)$ is such that

$$E|X_{t+v} - \hat{X}_{t,v}|^2 \leqq E|X_{t+v} - Y|^2, \quad \text{for all} \quad Y \in H_t(X). \tag{2.6}$$

When $\{X_t\}$ is nondeterministic, it is immediate from (2.4) that the error in predicting X_t by $\hat{X}_{t-1,1}$ is non-zero, i.e.

$$\varepsilon_t = X_t - \hat{X}_{t-1,1} \neq 0, \quad \text{for all} \quad t, \tag{2.7}$$

or

$$\sigma^2 = E|\varepsilon_t|^2 \neq 0, \quad \text{for all} \quad t.$$

The process $\{\varepsilon_t\}$ is referred to as the *innovation process* of $\{X_t\}$ and it plays an important role in the prediction of $\{X_t\}$ as the ε_t's are uncorrelated (orthogonal). We note that even though $\{X_t\}$ or a segment of it is directly observable, this is not the case for the innovation process $\{\varepsilon_t\}$.

By the well-known Wold decomposition Theorem, every nondeterministic process can be written as the sum of two unique uncorrelated stationary processes $\{U_t\}$ and $\{V_t\}$, i.e.

$$X_t = U_t + V_t, \quad \text{for all} \quad t, \tag{2.8}$$

where $\{V_t\}$ is deterministic and $\{U_t\}$, is purely nondeterministic with a one-sided moving average representation

$$U_t = \sum_{k=0}^{\infty} c_k \varepsilon_{t-k}, \quad c_0 = 1, \quad \sum_{k=0}^{\infty} |c_k|^2 < \infty, \quad \sigma^2 = E|\varepsilon_t|^2 < \infty. \tag{2.9}$$

This time-domain decomposition of a nondeterministic $\{X_t\}$ corresponds to a unique (Cramer-Lebesgue) decomposition of its spectral distribution function F into its absolutely continuous and singular parts:

$$F(d\lambda) = f(\lambda)d\lambda + F_s(d\lambda). \tag{2.10}$$

Thus, if $\{X_t\}$ is purely nondeterministic or equivalently F is absolutely continuous with density $0 \leqq f \in L^1$ and $\log f \in L^1$, it follows from (2.8) and (2.9) that

$$X_t = \sum_{k=0}^{\infty} c_k \varepsilon_{t-k}, \quad c_0 - 1, \quad \sum_{k=0}^{\infty} |c_k|^2 < \infty, \quad \sigma^2 = E|\varepsilon_t|^2 < \infty, \tag{2.11}$$

and from (2.2), (2.10) and (2.11) that

$$f(\lambda) = \sigma^2 \left| 1 + \sum_{k=1}^{\infty} c_k e^{ik\lambda} \right|^2 = \sigma^2 |\phi|^2, \tag{2.12}$$

with

$$\phi(\lambda) = 1 + \sum_{k=1}^{\infty} c_k e^{ik\lambda}, \tag{2.13}$$

The function ϕ in (2.13) is referred to as the *transfer function* of $\{X_t\}$. It is well-known that the extension of this function ϕ (also denoted by ϕ) in the open unit disc $D = \{z ; |z| < 1\}$ in the complex plane does not have any zeros, i.e.

$$\phi(z) = 1 + \sum_{k=1}^{\infty} c_k z^k \neq 0, \quad z \in D. \tag{2.14}$$

Thus, its reciprocal $1|\phi(z)$ has a Taylor expansion in D with Taylor coefficients d_k's:

$$1|\phi(z) = \sum_{k=0}^{\infty} d_k z^k, \quad d_0 = 1. \tag{2.15}$$

The coefficients $\{c_k\}$ and $\{d_k\}$ of ϕ and ϕ^{-1}, respectively, play important roles in finding explicit formulae for the predictor of future values of $\{X_t\}$.

From the definition of $\hat{X}_{t,v}$, $v \geqq 1$, cf. (2.6), and the moving average representation of $\{X_t\}$ in (2.11) it follows that

$$\hat{X}_{t,v} = \sum_{k=v}^{\infty} c_k \varepsilon_{t+v-k} \tag{2.16}$$

with the variance of prediction error

$$E|X_{t+v}-\hat{X}_{t,v}|^2=\sigma^2\sum_{k=0}^{v-1}|c_k|^2, \tag{2.17}$$

But, since $\{\varepsilon_t\}$ is the innovation process of $\{X_t\}$ and therefore not directly observable, (2.16) *can not be used to compute* $\hat{X}_{t,v}$ *unless one expresses* ε_t's *in terms of the past values* $X_t, X_{t-1}, \dots$ *of the process* $\{X_t\}$ *or equivalently finds an autoregressive series representation for the predictor* $\hat{X}_{t,v}$ *in the time-domain*, cf. [1, 11, 15, 16, 21].

It is shown by Wiener and Masani [21] that $\hat{X}_{t,v}$ has a *formal* autoregressive series representation:

$$\hat{X}_{t,v}-\sum_{k=0}^{\infty} e_{k,v}X_{t-k}, \tag{2.18}$$

where,

$$e_{k,v}=\sum_{j=0}^{k} c_{v+j}d_{k-j}, \quad k=0,1,2,\dots\ (c_0=d_0=1). \tag{2.19}$$

The problem of mean-convergence (or any other reasonable mode of convergence) of the infinite series $\sum_{k=1}^{\infty} e_{v,k}X_{t-k}$ is a formidable analytical problem and it has been studied by several authors including [1, 11–13, 15, 16, 21]. It is shown in [16] that the problem of convergence of the series (2.18) is equivalent to the convergence of the (apparently simpler) series

$$\varepsilon_t \sim X_t+\sum_{k=1}^{\infty} d_kX_{t-k}. \tag{2.20}$$

In the spectral-domain the latter convergence problem is the same as the convergence of the Fourier (Taylor)-series of the function ϕ^{-1} in $L^2(f)$.

Rearranging (2.20) one gets

$$X_t \sim \varepsilon_t+\sum_{k=1}^{\infty} a_kX_{t-k}, \quad a_k=-d_k, \quad k\geqq 1. \tag{2.21}$$

Thus, the problem of finding mean-convergent autoregressive series representation for $\hat{X}_{t,v}$, $v\geqq 1$, is equivalent to finding mean-convergent autoregressive representation for the process $\{X_t\}$ itself.

3. *p*-Stationary Stochastic Processes

In this section, for $0<p<\infty$, a definition of stationarity for p-th order stochastic processes is given. It is shown that this is the most natural and useful generalization of the notion of second-order weakly stationary stochastic processes to the $L^p(\Omega)$-setting. Also, it is shown that such stationary processes have shift operators which are isometric isomoprhism from $H(X)$ onto $H(X)$.

(3.1) **Definition.** For $0<p<\infty$, a stochastic process $\{X_t\}$ is said to be a *p-stationary process* if

$$X_t \in L^p(\Omega), \quad \text{for all integers} \quad t,$$

and

$$E\left|\sum_{k=1}^{n} a_k X_{t_k+h}\right|^p = E\left|\sum_{k=1}^{n} a_k X_{t_k}\right|^p, \tag{3.2}$$

for all integers $n \geqq 1$, $t_1, t_2, \ldots, t_n, h$ and scalars $a_1, a_2, \ldots, a_n$.

(3.3) *Remarks.* (a) For $p=2$, it is easy to check that every weakly stationary process $\{X_t\}$ satisfies (3.2) and therefore is a 2-stationary process. Conversely, it follows from the parallelogram law for the Hilbert space $L^2(\Omega)$ that every 2-stationarity process is a weakly stationary process. Thus our definition of p-stationary for $p=2$ coincides with that of weak stationarity.

(b) It is rather easy to show that any p-th order strictly stationary process is a p-stationary process. This shows that the ARMA processes with infinite veriance studied by Brockwell and Cline [6] is a subclass of p-stationary processes with $p<2$.

(c) For $1<\alpha\leqq 2$, by using the spectral representation of a harmonizable symmetric α-stable process, cf. [3, 4], one can show that such processes are p-stationary for any $p<\alpha$.

In the following we produce more examples of p-stationary processes which are important as far as prediction of these processes are concerned.

(3.4) **Theorem.** *Let $\{\varepsilon_j\}_{j=-\infty}^{+\infty}$ be a p-stationary process and $\{c_j\}_{j=-\infty}^{+\infty}$ a sequence of scalars such that the infinite series $\sum_{j=-\infty}^{\infty} c_j\varepsilon_j$ is convergent in the metric of $L^p(\Omega)$. Then, the process $\{X_t\}$ defined by*

$$X_t = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j}, \quad \text{for all integers} \quad t,$$

is a p-stationary process.

Proof. For all integers $n\geqq 1$, $t_1, \ldots, t_n$, h and scalars $a_1, \ldots, a_n$ we have from the definition of $\{X_t\}$ that

$$\sum_{k=1}^{n} a_k X_{t_k} = \sum_{j=-\infty}^{\infty} c_j \sum_{k=1}^{n} a_k \varepsilon_{t_k-j} = \lim_{m\to\infty} \sum_{j=-m}^{m} c_j \left(\sum_{k=1}^{n} a_k \varepsilon_{t_k-j}\right),$$

and

$$\sum_{k=1}^{n} a_k X_{t_k+h} = \sum_{j=-\infty}^{\infty} c_j \sum_{k=1}^{n} a_k \varepsilon_{t_k+h-j} = \lim_{m\to\infty} \sum_{j=-m}^{m} c_j \left(\sum_{k=1}^{n} a_k \varepsilon_{t_k+h-j}\right),$$

where the lim stands for the limit in the metric of $L^p(\Omega)$. Therefore,

$$E\left|\sum_{k=1}^{n} a_k X_{t_k}\right|^p = \lim_{m\to\infty} E\left|\sum_{k=-m}^{m} c_j\left(\sum_{k=1}^{n} a_k \varepsilon_{t_k-j}\right)\right|^p. \tag{1}$$

and

$$E\left|\sum_{k=1}^{n} a_k X_{t_k+h}\right|^p = \lim_{m\to\infty} E\left|\sum_{k=m}^{m} c_j\left(\sum_{k=1}^{n} a_k \varepsilon_{t_k+h-j}\right)\right|^p . \tag{2}$$

Since $\{\varepsilon_j\}$ is a p-stationary process, it follows that the right hand sides of (1) and (2) are equal and so are the left hand sides, i.e. $\{X_t\}$ is a p-stationary process. Q.E.D.

(3.5) **Corollary.** *Let $\{\varepsilon_j\}_{j=-\infty}^{\infty}$ be an independent identically distributed (i.i.d.) sequence of random variables with $\varepsilon_j \in L^p(\Omega)$, for all integers j, and $\{c_j\}_{j=-\infty}^{+\infty}$ be a sequence of scalars such that the series $\sum_{j=-\infty}^{+\infty} c_j \varepsilon_j$ is convergent in the metric of $L^p(\Omega)$. Then, the process $\{X_t\}$ defined by*

$$X_t = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j}, \quad \textit{for all integers} \quad t,$$

is a p-stationary process.

For the prediction of a p-stationary process $\{X_t\}$ we need to, and do restrict p to the range $1<p\leqq 2$. Since for $p\leqq 1$ an element $X\in L^p(\Omega)$ does not have a *unique* projection on a closed subspace M of $L^p(\Omega)$. For $2<p<\infty$, the prediction problem of $\{X_t\}$ is similar to the case $p=2$.

Let $X\in L^p(\Omega)$ be an element, and M a closed subspace of this Banach space such that $X\notin M$. Then, there exists a unique element $\hat{X}$ in M which is the closest to X, i.e.

$$E|X-\hat{X}|^p \leqq E|X-Y|^p, \quad \text{for all} \quad Y\in M, \tag{3.6}$$

this closest element $\hat{X}$ is called the metric projection (or simply the projection) of X on M. An alternative characterization of $\hat{X}$ is given by, cf. [18, p. 56],

$$EY(X-\hat{X})^{\langle p-1\rangle}=0, \quad \text{for all} \quad Y\in M, \tag{3.7}$$

where for a complex number z, $(z)^{\langle p-1\rangle}=|z|^{p-2}\bar{z}$.

For a p-stationary process $\{X_t\}$ the time-domain $H(X)$, the past and present subspaces $H_t(X)$, $-\infty<t<\infty$, and the remote past $H_{-\infty}(X)$ are defined as in (2.3) with the exception that the closure is taken in the norm of $L^p(\Omega)$. Similarly, the notions of deterministic, nondeterministic, purely nondeterministic and innovation process can be defined.

To each p-stationary process $\{X_t\}$ we associate an operator U which is an *isometric isomorphism* from $H(X)$ onto $H(X)$. To define this operator, consider the linear map U defined on the linear span of $\{X_t\}$, i.e. on $L(X)=\text{sp}\,\{X_s\,;\text{all integers } s\}$, by

$$U\left(\sum_{k=1}^{n} a_k X_{t_k}\right) = \sum_{k=1}^{n} a_k X_{t_k+1},$$

for all integers $n\geqq 1$, $t_1,\dots,t_n$ and scalars $a_1, a_2,\dots,a_n$. Then, from (3.2) it is immediate that U is an isometry from $L(X)$ onto $L(X)$, and therefore by continuity it can be extended to a unique isometric isomorphism from $H(X)$ onto $H(X)$. We refer to this extension of U (also denoted by U) as the *shift operator* of the process. We note that for a 2-stationary process the shift operator U is a unitary operator.

For any integer $v \geqq 1$, $\hat{X}_{t,v}$ denotes the best linear predictor of the future value X_{t+v} based on $X_t, X_{t-1}, \ldots$, i.e. $\hat{X}_{t,v}$ is such that $\|X_{t+v} - \hat{X}_{t,v}\|_p \leqq \|X_{t+v} - Y\|_p$ for all $Y \in H_t(X)$. For fixed $v \geqq 1$, $\{\hat{X}_{t,v}\}$ can be viewed as a stochastic process. The next theorem is essential for our study of the prediction problems of a p-stationary process.

(3.8) **Theorem.** *Let $\{X_t\}$ be a nondeterministic p-stationary process with the shift operator U, innovation process $\{\varepsilon_t\}$ and predictor process $\{\hat{X}_{t,v}\}$. Then,* (a) *the process $\{\hat{X}_{t,1}\}$ is a p-stationary process with shift U.* (b) *for any $v \geqq 1$, the process $\{\hat{X}_{t,v}\}$ is a p-stationary process with shift operator U.* (c) *the innovation process $\{\varepsilon_t\}$ is a p-stationary process with shift operator U.*

Proof. (a) Since $\hat{X}_{t,1} \in H_t(X)$, it can be expressed as

$$\hat{X}_{t,1} = \lim_{n \to \infty} \sum_{k=0}^{n} a_{n,k} X_{t-k}, \tag{1}$$

for some scalars $a_{n,k}$; $k = 0, 1, 2, \ldots, n$; $n = 1, 2, \ldots$. Thus, by definition of $\hat{X}_{t,1}$, cf. (3.6),

$$\lim_{n \to \infty} E\left|X_{t+1} - \sum_{k=0}^{n} a_{n,k} X_{t-k}\right|^p = E|X_{t+1} - \hat{X}_{t,1}|^p \tag{2}$$

Since $\{X_t\}$ is a p-stationary process cf. (3.2), it follows from (2) that

$$\lim_{n \to \infty} E\left|X_{t+2} - \sum_{k=0}^{n} a_{n,k} X_{t+1-k}\right|^p \leqq E|X_{t+2} - Y|^p, \text{ for all } Y \in H_{t+1}(X). \tag{3}$$

Now, since $\sum_{k=0}^{n} a_{n,k} X_{t-k}$ is convergent, as $n \to \infty$, and hence Cauchy in $H(X)$, it follows that $\sum_{k=0}^{n} a_{n,k} X_{t+1-k}$ is also Cauchy and hence convergent (because $\{X_t\}$ is a p-stationary process). Therefore, we get from (3) that

$$E\left|X_{t+2} - \lim_{n \to \infty} \sum_{k=0}^{n} a_{n,k} X_{t+1-k}\right|^p \leqq E|X_{t+2} - Y|^p,$$

for all $Y \in H_{t+1}(X)$, i.e.

$$\hat{X}_{t+1,1} = \lim_{n \to \infty} \sum_{k=0}^{n} a_{n,k} X_{t+1-k} = \lim_{n \to \infty} \sum_{k=0}^{n} a_{n,k} U X_{t-k}$$

$$= \lim_{n \to \infty} U \sum_{k=0}^{n} a_{n,k} X_{t-k} = U\left(\lim_{n \to \infty} \sum_{k=0}^{n} a_{n,k} X_{t-k}\right) = U\hat{X}_{t,1}.$$

The fourth equality holds because U is an isometry and hence bounded. This completes proof of (a). Proof of (b) is similar to (a), and (c) follows from (a), since $\varepsilon_{t+1} = X_{t+1} - \hat{X}_{t,1} = UX_t - U\hat{X}_{t-1,1} = U\varepsilon_t$ Q.E.D.

4. Finite Wold and Predictive Decompositions

Let $\{X_t\}$ be a nondeterministic p-stationary process, $1<p<2$, with innovation process $\{\varepsilon_t\}$. Since $\{\varepsilon_t\}$ is not a white noise process in general the Wold decomposition of $\{X_t\}$, even when it exists [3, 4], is not as useful in finding the predictors of the future values of $\{X_t\}$ as it is in the case of 2-stationary processes, cf. (2.16). In this section we provide finite Wold and predictive decompositions for a nondeterministic p-stationary process and show that these simultaneous decompositions of X_t do provide valuable information about the predictor and structure of such processes. These decompositions are obtained by noting that for any nondeterministic p-stationary process with innovation $\{\varepsilon_t\}$ we have

(4.1) $$H_t=\overline{\text{sp}}\,\{X_t, H_{t-1}\}, \quad X_t\notin H_{t-1}, \quad \text{for all} \quad t.$$

and also

(4.2) $$H_t=\overline{\text{sp}}\,\{\varepsilon_t, H_{t-1}\}, \quad \varepsilon_t\notin H_{t-1}, \quad \text{for all} \quad t.$$

The following basic and important lemma plays a crucial role in bringing out the geometrical meaning of (4.1), (4.2) and its implications as far as the problems of prediction of $\{X_t\}$ are concerned. This lemma constitutes the cornerstone of our approach to the prediction problem of processes with infinite variance.

(4.3) **Main Lemma.** *Let M be a subspace of a Banach space B and let X be an element of B so that $X\notin M$. Then, a fixed $Y\in\overline{sp}\,\{X, M\}$ can be written as*

$$Y=aX+e$$

for a unique (scalar) a and vector $e\in M$.

Proof. Since $Y\in\overline{\text{sp}}\,\{X, M\}$, there exists a convergent sequence $\{Y_n\}$ such that

(1) $$Y_n=a_nX+e_n\in\overline{\text{sp}}\,\{X, M\} \quad \text{and} \quad Y=\lim_{n\to\infty} Y_n.$$

To establish the Lemma we need to show that the sequence of constants $\{a_n\}$ is bounded. If $\{a_n\}$ is not bounded, then there exists a subsequence $\{a_{n_k}\}$ such that $0\neq a_{n_k}\to\infty$ as $k\to\infty$. But, since from (1)

(2) $$Y_{n_k}=a_{n_k}X+e_{n_k},$$

it follows from the boundedness of $\{Y_{n_k}\}$ that

$$X=\lim_{n\to\infty}\frac{Y_{n_k}}{a_{n_k}}-\lim_{n\to\infty}\frac{e_{n_k}}{a_{n_k}}=-\lim_{n\to\infty}\frac{e_{n_k}}{a_{n_k}}\in M,$$

which is a contradiction. Thus, $\{a_n\}$ is bounded. Since $\{a_n\}$ is bounded, it has (by Bolzano-Weierstrass theorem) a convergent subsequence $\{a_{n_k}\}$. Let $\lim_{n\to\infty} a_{n_k}=a$. From (2) it follows that $\{e_{n_k}\}$ is convergent, and by letting $e=\lim_{n\to\infty} e_{n_k}$ we get that $e\in M$ and

$$Y=aX+e.$$

To prove the uniqueness of a and e, note that if Y has two different representations;

$$Y = a_1 X + e_1 ,$$

$$Y = a_2 X + e_2 ,$$

then by subtraction we get that

$$(a_1 - a_2) X = e_2 - e_1 \in M ,$$

which is impossible unless $a_1 = a_2$ and $e_1 = e_2$. Q.E.D.

In the next theorem by using Lemma 4.3 we obtain finite Wold and predictive decompositions for nondeterministic p-stationary processes.

(4.4) **Main Theorem.** *Let $\{X_t\}$ be a nondeterministic p-stationary process with innovation process $\{\varepsilon_t\}$. Then, for any integer $n \geqq 1$,*

(a) *there exist unique constants $a_1, \dots, a_n$ and a unique p-stationary Process $\{e_{t,n}\}$ with $e_{t,n} \in H_{t-n-1}(X)$ such that*

(4.5)
$$\begin{cases} \hat{X}_{t-1,1} = \sum_{k=1}^{n} a_k X_{t-k} + e_{t,n} , \\ X_t = \varepsilon_t + \sum_{k=1}^{n} a_k X_{t-k} + e_{t,n} . \end{cases}$$

(b) *there exist unique constants $c_1, \dots, c_n$ and a unique p-stationary-process $\{V_{t,n}\}$ with $V_{t,n} \in H_{t-n-1}$ such that*

(4.6)
$$\begin{cases} \hat{X}_{t-1,1} = \sum_{k=1}^{n} c_k \varepsilon_{t-k} + V_{t,n} , \\ X_t = \varepsilon_t + \sum_{k=1}^{n} c_k \varepsilon_{t-k} + V_{t,n} . \end{cases}$$

Proof. (a) Since $\hat{X}_{0,1} \in H_0 = \overline{\text{sp}}\,\{X_0, H_{-1}\}$ by applying Lemma 4.3 with $X = X_0$, $M = H_{-1}$, it follows that there exist a unique constant a_1 and $e_1 \in H_{-1}$ such that

(1)
$$\hat{X}_{0,1} = a_1 X_0 + e_1 .$$

By applying the same argument to $e_1 \in \overline{\text{sp}}\,\{X_{-1}, H_{-2}\}$ and repeating it we get that

(2)
$$\hat{X}_{0,1} = a_1 X_0 + a_2 X_{-1} + \dots + a_n X_{1-n} + e_n ,$$

for unique constants $a_1, \dots, a_n$ and $e_n \in H_{-n}$. Let U be the shift operator of the process $\{X_t\}$, by applying the operator U^{t-1} to both sides of (2) it follows from Theorem 3.8 that

$$\hat{X}_{t-1,1} = \sum_{k=1}^{n} a_k X_{t-k} + e_{t,n} ,$$

where $e_{t,n} = U^{t-1} e_n \in H_{t-n-1}$. Furthermore, $\{e_{t,n}\}$ is a p-stationary process.

The second equation in (4.5) follows from the first by observing that $X_t - \hat{X}_{t-1,1} = \varepsilon_t$. Proof of part (b) is similar to (a) by noting that $H_t = \overline{\text{sp}}\,\{\varepsilon_t, H_{t-1}\}$, cf. (4.2). Q.E.D.

Since the integer n in Theorem 4.4 is arbitrary and the set of coefficients a_k's and c_k's do not depend on n, and furthermore they are unique, it follows that the two infinite sequences $\{a_k\}_{k=1}^{\infty}$ and $\{c_k\}_{k=1}^{\infty}$ must carry considerable information about a nondeterministic p-stationary process. It is even more important to note that the two sequences are closely related, i.e. knowing one it is possible to find the other uniquely, cf. Corollary 4.7. This suggests the possibility that either of these sequences be used to *parameterize* a general nondeterministic p-stationary process.

(4.7) **Corollary.** *The sequences $\{a_k\}$ and $\{c_k\}$ of Theorem 4.4 are related by the recursion*

$$(4.8)\qquad \begin{cases} c_0=1\,, \\ \sum_{k=0}^{l-1} c_k a_{l-k}=c_l,\ l=1,2,\dots . \end{cases}$$

Proof. To establish (4.8), we note that for any $n\geqq 1$, $a_1,\dots,a_n$ satisfy, cf. (4.5),

$$\hat{X}_{t-1,1}-\sum_{k=1}^{n} a_k X_{t-k}\in H_{t-n-1}\,,$$

and from (4.6) we have

$$(1)\qquad \hat{X}_{t-1,1}-\sum_{k=1}^{n} a_k X_{t-k}=\sum_{l=1}^{n}\left(c_l-\sum_{k=0}^{l-1} c_k a_{l-k}\right)\varepsilon_{t-l}+R_{n,t}\,,$$

where $R_{n,t}\in H_{t-n-1}$. It follows from (1) that $\hat{X}_{t-1,1}-\sum_{k=1}^{n} a_k X_{t-k}\in H_{t-n-1}$, if and only if (4.8) is satisfied. Q.E.D.

When the sequence $\{c_k\}$ is known, one can write (4.8) as the following (Toeplitz) system of equations with $\{a_k\}$ as the vector of unknowns;

$$(4.9)\qquad \begin{bmatrix} 1 & 0 & 0 & 0 \\ c_1 & 1 & 0 & 0 \\ c_2 & c_1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \vdots \end{bmatrix}=\begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \vdots \end{bmatrix}.$$

In the following we study two examples where the $\{a_k\}$ sequence is found in terms of a given sequence $\{c_k\}$. As usual we assume that $\{\varepsilon_t\}$ is the innovation process of the p-stationary process $\{X_t\}$ and our starting point in these examples are the finite Wold decomposition of $\{X_t\}$, cf. Theorem 4.4 (b).

(4.10) *Example.* Let $\{X_t\}$ be given by

$$X_t=\varepsilon_t-\varepsilon_{t-1}\,,\quad \text{for all}\quad t\,,$$

i.e. $c_0=1$, $c_1=1$, $c_k=0$, $k\geqq 2$. Then, it follows from (4.8) that

$$a_k=-1\,,\quad \text{for all}\quad k\geqq 1\,,$$

and

$$e_{t,n}=\hat{X}_{t-1,1}-\sum_{k=1}^{n} a_k X_{t-k}=-\varepsilon_{t-n-1}\,.$$

(4.11) *Example.* Let $\{X_t\}$ be given by

$$X_t = \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}, \quad \text{for all} \quad t,$$

i.e. $c_0 = 1$, $c_1 = -2$, $c_2 = 1$, $c_k = 0$, for all $k \geqq 3$. Then, it follows from (4.8) that

$$a_k = -(k+1), \quad \text{for} \quad k \geqq 1,$$

and

$$e_{t,n} = \hat{X}_{t-1,1} - \sum_{k=1}^{n} a_k X_{t-k} = -(n+2)\varepsilon_{t-n-1} + (n+1)\varepsilon_{t-n-2}.$$

In the previous two examples, it is interesting to note that although the $\{c_k\}$ sequences are very similar in nature, this is not the case for their corresponding $\{a_k\}$ sequences. The first one is bounded while the other is unbounded. This difference in the boundedness property of $\{a_k\}$ plays an important role in studying the convergence of $e_{t,n}$, as $n \to \infty$.

(4.12) *Remarks.* (a) The representation $\hat{X}_{t-1,1} = \sum_{k=1}^{n} a_k X_{t-k} + e_{t,n}$ in (4.5) provides a formula for approximating the predictor $\hat{X}_{t-1,1}$ when only n observations $X_{t-1}, \ldots, X_{t-n}$ from the past are available. The interesting feature of this approximation is that the coefficients a_k do not depend on n, and therefore they do not change as n increases or as more observations become available. This is in sharp contrast to the best linear predictor of X_t based on n observations $X_{t-1}, \ldots, X_{t-n}$. Denoting this by $X^*_{t,n}$, we have to find $a_{n,k}$'s such that

$$E|X_t - X^*_{t,n}|^p = E|X_t - \sum_{k=1}^{n} a_{n,k} X_{t-k}|^p,$$

$1 < p \leqq 2$, is minimized. Finding these $a_{n,k}$'s is not easy. However, when $p = 2$, i.e. $\{X_t\}$ is a 2-stationary process with covariance function $\{\gamma_\kappa\}$ the $a_{n,k}$'s can be found by solving the following (Toeplitz) system of linear equations

$$\text{(4.13)} \qquad \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \vdots \\ \vdots & \vdots & \cdots \gamma_0 & \gamma_1 \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} a_{n,1} \\ a_{n,2} \\ \vdots \\ a_{n,n} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}.$$

Despite some differences between $\hat{X}_{t-1,1}$ and $X^*_{t,n}$, it is interesting to note that in computing both $\hat{X}_{t-1,1}$, and $X^*_{t,n}$ (when $p = 2$) one needs to deal with Toeplitz matrices. For $1 < p < 2$, there is no simple and clear way to find $X^*_{t,n}$, but even in this case $\hat{X}_{t-1,1}$ can be approximated by solving a finite section of (4.9). This shows the importance of the sequence $\{c_k\}$ in determining $\hat{X}_{t-1,1}$ and other important characteristics of a p-stationary process.

In view of the important role of $\{c_k\}$ as the parameter of a nondeterministic p-stationary process it is desirable to have more information about this sequence. For nondeterministic weakly stationary processes it is well-known that the sequence $\{c_k\}$ is square-summable and therefore it is bounded. For a general p-stationary process all we know about the sequence $\{c_k\}$ is that $|c_k| < c^{2^k}$, $k = 1, 2, \ldots$ for a constant c and this is proved in the next theorem. Also, this theorem sheds some light on the meaning of the process $\{V_{t,n}\}$ appearing in Theorem 4.4(b).

(4.14) **Theorem.** *Let* $\{X_t\}$ *be a nondeterministic p-stationary process with innovation* $\{\varepsilon_t\}$, *and* $\{c_k\}$, $\{V_{t,n}\}$, $n \geqq 1$, *be as in Theorem* 4.4(b). *Then, for any* $n \geqq 1$,

(a) $$V_{t,n} = \underset{H_{t-n-1}}{P} \left(X_t - \sum_{k=0}^{n-1} c_k \varepsilon_{t-k} \right),$$

where $P_M X$ *denotes the projection of* X *onto the subspace* M *of* $H(X)$, *cf.* (3.6).

(b) $|c_k| \leqq c 2^k, 1 = 1, 2, \ldots$ for some constant c.

Proof. (a) From Theorem 4.4(b), since $V_{t,n} \in H_{t-n-1}$ and

(1) $$X_t - \sum_{k=0}^{n-1} c_k \varepsilon_{t-n} - V_{t,n} = c_n \varepsilon_{t-n},$$

it is enough to show that

$$EY(c_n \varepsilon_{t-n})^{\langle p-1 \rangle} = 0, \quad \text{for all} \quad Y \in H_{t-n-1},$$

and this follows from the definition of the innovation process and (3.7). (b) It follows from (a) and (1) above that

(2) $$\|c_n \varepsilon_{t-n}\|_p \leqq \left\| X_t - \sum_{k=0}^{n-1} c_k \varepsilon_{t-k} \right\|_p \leqq \|X_t\|_p + \sum_{k=0}^{n-1} \|c_k \varepsilon_{t-k}\|_p.$$

Since

$$\|\varepsilon_t\|_p = \|X_t - \hat{X}_{t-1,1}\|_p \leqq \|X_t\|_p,$$

it follows from (2) and stationarity of $\{X_t\}$ and $\{\varepsilon_t\}$, cf. Theorem 3.8, that

$$|c_n| \leqq \frac{\|X_0\|_p}{\|\varepsilon_0\|_p} 2^n, \quad n = 1, 2, \ldots.$$

Q.E.D.

Now, by using the unique sequence $\{c_k\}$ associated to each nondeterministic p-stationary process $\{X_t\}$, we define a function ϕ

(4.15) $$\phi(z) = 1 + \sum_{k=1}^{\infty} c_k z^k, \quad |z| < 1/2,$$

which is analytic in the open disc with radius 1/2, cf. Theorem 4.14(b), and refer to it as the *transfer function of* $\{X_t\}$. It is immediate from (4.8) or (4.9) that the sequence $\{a_k\}$ of Theorem 4.4(a) can be obtained from

(4.16) $$a_k = -d_k, \quad k = 1, 2, \ldots$$

where the sequence $\{d_k\}$ is the Taylor coefficients of the reciprocal of ϕ, i.e.

(4.17) $$\phi^{-1}(z) = 1 + \sum_{k=1}^{\infty} d_k z^k.$$

This argument shows that by having the transfer function ϕ of a nondeterministic p-stationary process $\{X_t\}$ one can find the two parameter sequences $\{c_k\}$ and $\{a_k\}$ at once. Next, we explore the possibility of finding the coefficients of the v-

step ahead predictor of $\{X_t\}$, i.e. the coefficients of $\hat{X}_{t,v}$, in terms of the coefficients of the functions ϕ and ϕ^{-1}. The following theorem which is more general than Theorem 4.4 and Corollary 4.7 combined, provides the basic ingredients for finding these coefficients.

(4.18) **Theorem.** *Let $\{X_t\}$ be nondeterministic p-stationary process with innovation $\{\varepsilon_t\}$. Then, for a fixed $v \geqq 1$ and any integer $n \geqq 1$,*

(a) *there exist unique constants $a_{1,v}, \ldots, a_{n,v}$ and a unique p-stationary process $\{e_{t,n,v}\}_{t=-\infty}^{\infty}$ with $e_{t,n,v} \in H_{t-n-1}(X)$ such that*

$$\hat{X}_{t-1,v} = \sum_{k=1}^{n} a_{k,v} X_{t-k} + e_{t,n,v}. \tag{4.19}$$

(Note: For $v=1$, $a_{k,v}=a_k$, for $k \geqq 1$.)

(b) *there exists unique constants $c_{1,v}, \ldots, c_{n,v}$ and a unique p-stationary process $\{V_{t,n,v}\}_{t=-\infty}^{\infty}$ with $V_{t,n,v} \in H_{t-n-1}(X)$ such that*

$$\hat{X}_{t-1,v} = \sum_{k=1}^{n} c_{k,v} \varepsilon_{t-k} + V_{t,n,v}. \tag{4.20}$$

(Note: For $v=1$, $c_{k,v}=c_k$, for $k \geqq 1$.)

(c) *for a fixed $v \geqq 1$, the sequences $\{a_{k,v}\}$ and $\{c_{k,v}\}$ of parts* (a) *and* (b) *are related by the recursion*

$$\begin{cases} c_0 = 1, \\ \sum_{k=0}^{l-1} c_k a_{l-k,v} = c_{l-v}, \quad l = 1, 2, \ldots, \end{cases}$$

where $\{c_k\}$ is the sequence of coefficients of the transfer function of X_t, cf. (4.15).

Proof. Since $\hat{X}_{t-1,v} \in H_{t-1}(X) = \overline{\text{sp}}\,\{X_{t-1}, H_{t-2}\} = \overline{\text{sp}}\,\{\varepsilon_{t-1}, H_{t-2}\}$, proofs of parts (a) and (b) are exactly the same as those in Theorem 4.4. Proof of (c) is the same as that of Corollary 4.7. Q.E.D.

This theorem shows that for a fixed $v \geqq 1$, the two sequences $\{a_{k,v}\}$ and $\{c_{k,v}\}$ are related via the recursion (4.21). Thus to find the sequence $\{a_{k,v}\}$, i.e. the coefficients of $\hat{X}_{t-1,v}$ one needs to know the sequence $\{c_{k,v}\}$. However, in general it is not that easy to find the sequence $\{c_{k,v}\}$, cf. [3, p. 610]. In the following two important cases we show that $\{c_{k,v}\}$ can be found. In fact, for these two cases we show that

$$c_{k,v} = c_{k+v-1}, \quad k = 1, 2, \ldots.$$

(4.22) **Corollary.** *Let $\{X_t\}$ be a nondeterministic p-stationary process with innovation process $\{\varepsilon_t\}$. If (i) $p=2$ or (ii) the $\{\varepsilon_t\}$ is an i.i.d. sequence of random variables, then for a fixed $v \geqq 1$, and any integer $n \geqq 1$,*

(a)
$$\hat{X}_{t-1,v} = \sum_{k=1}^{n} c_{v+k-1} \varepsilon_{t-k} + V_{t-1+v,n+v-1},$$

where
$$V_{t-1+v,n+v-1} \in H_{t-n-1}.$$

(b)
$$\hat{X}_{t-1,v} = \sum_{k=1}^{n} a_{k,v} X_{t-k} + e_{t,n,v}, \quad e_{t,n,v} \in H_{t-n-1}$$

and

$$a_{k,v}=\sum_{j=0}^{k-1} c_{v+j}d_{k-j}, \quad k=1,2,\dots \ (c_0=d_0=1).$$

Proof. (a) For any integer $m\geqq 1$ and fixed v, $m\geqq v\geqq 1$, we have from the second relation in (4.5) that

$$(1) \qquad X_{t-1+v}=\varepsilon_{t-1+v}+\sum_{k=1}^{v-1} c_k\varepsilon_{t-1+v-k}+\sum_{k=v}^{m} c_k\varepsilon_{t-1+v-k}+V_{t-1+v,m},$$

with $V_{t-1+v,m}\in H_{t-1+v-m-1}$. Since for $p=2$, $\varepsilon_{t-1+v}+\sum_{k=1}^{v-1} c_k\varepsilon_{t-1+v-k}$ is orthogonal to H_{t-1}, it follows from (1) that

$$\hat{X}_{t-1,v}=\sum_{k=v}^{m} c_k\varepsilon_{t-1+v-k}+V_{t-1+v,m},$$

or equivalently

$$(2) \qquad \hat{X}_{t-1,v}=\sum_{k=1}^{m-v+1} c_{k+v-1}\varepsilon_{t-k}+V_{t-1+v,m}$$

$$=\sum_{k=1}^{n} c_{k+v-1}\varepsilon_{t-k}+V_{t-1+v,n+v-1},$$

where $n=m-v+1$ and $V_{t-1+v,n+v-1}\in H_{t-n-1}$. Comparing (2) with (4.20) it follows, from the uniqueness of $\{c_{k,v}\}$, that

$$c_{k,v}=c_{k+v-1}, \quad \text{for all} \quad k\geqq 1,\ v\geqq 1.$$

(b) follows from Theorem 4.18 (a), (c) and the fact that $c_{k,v}=c_{k+v-1}$. Q.E.D.

This shows that even for v-step ahead prediction of some p-stationary processes one actually does not need the full power of Wold decomposition in finding the coefficients of the v-step ahead predictor. Only the finite Wold decomposition suffices and in this the orthogonality (independence) of the random variables of the innovation process $\{\varepsilon_t\}$ plays a crucial role. The latter case is important in the statistical analysis of data from processes with infinite variance, cf. [2, 7].

5. Wold and Predictive Decompositions

The role and importance of Wold decomposition in prediction and characterization of purely nondeterministic weakly stationary processes are well-known. An important factor in establishing Wold decomposition for such processes is the orthogonality of the random variables of the innovation process. Since for p-stationary processes, $1<p<2$, the innovation process is not an orthogonal process, it is natural to see if there is any kind of Wold decomposition for such processes. In this section we study nondeterministic p-stationary processes for which one can find Wold and predictive decompositions. Our starting point is Theorem 4.4 where finite Wold and predictive decompositions for such processes are given.

In view of Theorem 4.4, existence of a Wold or predictive decomposition for $\{X_t\}$ depends very much on convergence properties of the processes $\{V_{t,n}\}$ and $\{e_{t,n}\}$ as $n\to\infty$, and in turn is related to the covvergence properties of the series $\sum_{k=1}^{\infty} c_k\varepsilon_{t-k}$ and $\sum_{k=1}^{\infty} a_k X_{t-k}$ in the norm of $L^p(\Omega)$ or in the mean of order p. Thus, *the key problem in establishing such decomposition is the problem of norm-convergence of these series, and as such the degrees of difficulty of establishing Wold decomposition, and predictive decomposition for stochastic process are the same in general.* There is the simple and most important special case of weakly stationary processes for which the problem of norm-convergence of $\sum_{k=1}^{\infty} c_k\varepsilon_{t-k}$ is a simple consequence of the orthogonality of $\{\varepsilon_t\}$. In this case since $\{\varepsilon_t\}$ is a white noise process, it follows from (4.6) that

$$\sigma^2 \sum_{k=1}^{n} |c_k|^2 \leqq E|X_0|^2 < \infty\,, \quad \text{for all} \quad n\,,$$

and therefore $\sum_{k=1}^{\infty} |c_k|^2 < \infty$, which implies that the series $\sum_{k=1}^{\infty} c_k\dot{\varepsilon}_{t-k}$ is convergent in the mean of order 2. Note that for such processes the problem of norm-convergence of the series $\sum_{k=1}^{\infty} a_k X_{t-k}$ is quite complicated, since the X_t's are not uncorrelated, cf. [11–13, 15, 16, 21].

For other p-stationary processes it is difficult to find conditions for the mean-convergence of the series $\sum_{k=1}^{\infty} c_k\varepsilon_{t-k}$, let alone $\sum_{k=1}^{\infty} a_k X_{t-k}$. The next theorem provides a simple sufficient condition for the mean-convergence of these series.

(5.1) **Theorem.** *Let $\{X_t\}$ be a nondeterministic p-stationary process with innovation $\{\varepsilon_t\}$. Then,*

(a) $\sum_{k=1}^{\infty} a_k X_{t-k}$ *is convergent in the mean of order p, if* $\sum_{k=1}^{\infty} |a_k| < \infty$.

(b) $\sum_{k=1}^{\infty} c_k\varepsilon_{t-k}$ *is convergent in the mean of order p, if*

$$\sum_{k=1}^{\infty} |c_k| < \infty\,.$$

Proof. It is immediate from the fact that for all $m \leqq n$,

$$\left\| \sum_{k=m}^{n} a_k X_{t-k} \right\|_p \leqq \|X_0\|_p \sum_{k=m}^{n} |a_k|\,.$$

Q.E.D.

The next theorem provides Wold and predictive decompositions for such processes under the assumption that $\sum_{k=1}^{\infty} a_k X_{t-k}$ and $\sum_{k=1}^{\infty} c_k\varepsilon_{t-k}$ are convergent in the mean of order p.

(5.2) **Theorem** (Wold and predictive decompositions). *Let* $\{X_t\}$ *be a nondeterministic p-stationary process with innovation* $\{\varepsilon_t\}$ *and* $\{a_k\}$, $\{c_k\}$ *as in Theorem 4.4.*

(a) *If* $\sum_{k=1}^{\infty} a_k X_{t-k}$ *is convergent in the mean of order p, then there exists a unique p-stationary process* $\{e_t\}$ *with* $e_t \in H_{-\infty}(X)$ *such that, for all t,*

$$X_t = \varepsilon_t + \sum_{k=1}^{\infty} a_k X_{t-k} + e_t. \quad \textit{(Predictive decomposition)}.$$

(b) *If* $\sum_{k=1}^{\infty} c_k \varepsilon_{t-k}$ *is convergent in the mean of order p, then there exists a unique p-stationary process* $\{V_t\}$ *with* $V_t \in H_{-\infty}(X)$ *such that, for all t,*

$$X_t = \varepsilon_t + \sum_{k=1}^{\infty} c_k \varepsilon_{t-k} + V_t = U_t + V_t, \quad \textit{(Wold decomposition)}.$$

where the p-stationary process $\{U_t\}$ *is a one-sided moving-average of* $\{\varepsilon_t\}$, *i.e.* $U_t = \varepsilon_t + \sum_{k=1}^{\infty} c_k \varepsilon_{t-k}$.

Proof. (a) We have from (4.5) that, for all $n \geqq 1$,

$$e_{t,n} = X_t - \varepsilon_t - \sum_{k=1}^{n} a_k X_{t-k} \in H_{t-n-1}(X). \tag{1}$$

Since the series $\sum_{k=1}^{\infty} a_k X_{t-k}$ is convergent in the mean of order p, it follows from (1) that, for each t, $e_{t,n}$ converges to an element e_t in the norm of order p, and furthermore $e_t \in \bigcap_{n=1}^{\infty} H_{t-n-1}(X) = H_{-\infty}(X)$. Now, the result follows by letting $n \to \infty$ on both sides of (1). Proof of (b) is similar to that of (a).

An important consequence of Wold decomposition for nondeterministic weakly stationary process $\{X_t\}$ with innovation $\{\varepsilon_t\}$ is the characterization of purely nondeterministic processes: A nondeterministic weakly stationary process $\{X_t\}$ is purely nondeterministic, if and only if

$$X_t = \sum_{k=1}^{\infty} c_k \varepsilon_{t-k}, \quad c_0 = 1, \quad \sum_{k=1}^{\infty} |c_k|^2 < \infty,$$

i.e. X_t *is a one-sided moving average of its innovation process.* In proving this the orthogonality of $\{\varepsilon_t\}$, or $\{U_t\}$ and $\{V_t\}$ plays an important role. Lack of orthogonality of $\{\varepsilon_t\}$, or $\{U_t\}$ and $\{V_t\}$, cf. Theorem 5.2(b), makes it difficult to find such characterizations of pure nondeterminism for p-stationary processes, $1 < p < 2$. Only a necessary condition for this is given in the following Corollary, part (b).

(5.3) **Corollary.** *Let* $\{X_t\}$ *be a purely nondeterministic p-stationary process with innovation* $\{\varepsilon_t\}$ *and* $\{a_k\}$, $\{c_k\}$ *as in Theorem 4.4.*

(a) *If* $\sum_{k=1}^{\infty} a_k X_{t-k}$ *is convergent in the mean of order p, then for all t,*

$$X_t = \varepsilon_t + \sum_{k=1}^{\infty} a_k X_{t-k},$$

i.e. $\{X_t\}$ *has an autoregressive representaiton.*

(b) *If* $\sum_{k=1}^{\infty} c_k \varepsilon_{t-k}$ *is convergent in the mean of order p, then for all t,*

$$X_t = \sum_{k=1}^{\infty} c_k \varepsilon_{t-k}, \quad c_0 = 1,$$

i.e. $\{X_t\}$ *is a one-sided moving average of its innovation process.*

Proof of this corollary is immediate from Theorem 5.2 since $H_{-\infty}(X) = \{0\}$, and therefore $e_t = V_t = 0$, for all t.

Acknowledgement. We wish to express our gratitude to a referee for very helpful comments regarding the presentation of ideas in this paper. The second author was supported by the NSF Grants MCS-8301240 and DMS-8601858.

References

1. Akutowicz, E.J.: On an explicit formula in linear least squares prediction. Math. Scand. **5**, 261–266 (1957)
2. Bhansali, R.J.: Order determination for process with infinite variance. In: Franke, J., Härdle, W., Martin, D. (eds.) Robust and nonlinear time series analysis. (Lect. Not. Statist., vol. 26) Berlin Heidelberg New York Tokyo: Springer 1984
3. Cambanis, S., Soltani, A.R.: Prediction of stable processes: Spectral and moving average representations. Z. Wahrscheinlichkeitstheor. Verw. Geb. **66**, 593–612 (1984)
4. Cambanis, S., Miamee, A.G.: On prediction of harmonizable stable processes. Sakhya (to appear)
5. Cambanis, S., Hardin, C.D., Weron, A.: Innovations and Wold decompositions of stable processes. Probab. Th. Rel. Fields **79**, 1–27 (1988)
6. Cline, D.B.H., Brockwell, P.J.: Linear prediction of ARMA processes with infinite variance. Stochastic Proc. Appl. **19**, 281–296 (1985)
7. Hannan, E.J., Kanter, M.: Autoregressive process with infinite variance. J. Appl. Probab. **14**, 411–415 (1977)
8. Hall, P, Heyde, C.C.: Martingale limit theory and its application. New York: Academic Press (1980)
9. Hosoya, Y.: Harmonizable stable processes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **60**, 517–533 (1982)
10. Kolmogorov, A.N.: Stationary seqzences in Hilbert space. Bull. Math. Univ. Moscow, **2**, 1–40 (1941)
11. Masani, P.R.: The prediction theory of multivariate stochastic processes; III. Acta Math. **104**, 141–162 (1960)
12. Masani, P.R.: Recent trends in multivariate prediction theory. In: Krishnaiah, P.R. (eds) Multivariate analysis, pp. 351–382. New York: Academic Press 1966
13. Miamee, A.G., Salehi, H.: On an explicit representation of the linear predictor of a weakly stationary stochastic sequence. Boletin de la Sociedad Matematica Mexicana, **28**, 81–93 (1983)
14. Pourahmadi, M.: On minimality and interpolation of harmonizable stable processes. SIAM J. Appl. Math. **44**, 1023–1030 (1984)
15. Pourahmadi, M.: A matricial extension of the Helson-Szego theorem and its application in multivariate prediction. J. Multivariate Anal. **16**, 265–275 (1985)

16. Pourahmadi, M.: Autoregressive representations of multivariate stationary stochastic processes. Prob. Th. Rel. Fields (To appear)
17. Rozanov, Yu.A.: Stationary random processes. San Francisco: Holden-Day 1967
18. Singer, I.: Best approximation in normed linear spaces by elements of linear subspaces. Berlin Heidelberg New York: Springer 1970
19. Urbanik, K.: Harmonizable sequences of random variables. Colloques Internationaux du C.N.R.S. **186**, 345–361 (1970)
20. Weron, A.: Harmonizable stable processes on groups: Spectral, ergodic and interpolation properties. Z. Wahrscheinlichkeitstheor. Verw. Geb. **68**, 473–491 (1985)
21. Wiener, N., Masani, P.R.: The prediction theory of multivariate stochastic processes. Part I.: Acta Math. **98**, 111–150: part II: Acta Math. **99**, 93–137 (1958)
22. Wold, H.: A study in the analysis of stationary time series. Stockholm: Almquist & Wiksell 1954

Received March 18, 1987