SHIFT INVARIANT SPACES AND PREDICTION THEORY

BY

P. MASANI

Indiana University, U.S.A.(1)

1. Introduction

It is apparent that the theory of shift-invariant subspaces of the Hardy class H_2 developed by Beurling [1] in 1949 is related to the prediction theory of weakly stationary stochastic processes as systematized by Kolmogorov [6] in 1941. This has been pointed out by Helson and Lowdenslager [4] and Lax [7], but the extremely close relationship between the two subjects becomes much clearer when seen in the light of a result ((2.1) below) on the decomposition wrought in a Hilbert space by an isometric operator acting on it. This result, known apparently for some time, has been stated in a clear form and put to significant use in a recent paper by Halmos ([3], Lemma 1).

Our purpose is to show that the central theorems of both subjects, viz. the existence of a single isometric generator for shift-invariant subspaces of H_2 (Beurling [1], Theorem IV), and the decomposition of a non-deterministic process into a one-sided moving average and a deterministic process (Wold's Theorem [2], p. 576) are derivable by the same techniques from this result on decomposition (§ 3)(2). We shall also show that the Riesz-Nevanlinna factorization of a function ψ in the Hardy class H_2 into optimal and residual factors(3) is just a variant of the Wold decomposition, these factors being precisely the generating function and normalized innovation of the stochastic process generated by ψ under shifts by $e^{i\theta}$ (§ 4). Beurling's Theorems I–III, which extend to H_2 Wiener's Closure Theorem for L_2, will emerge as corollaries of

(1) Part of the work on this paper was done under a grant from the National Science Foundation and under the sponsorship of the Office of Naval Research.

(2) The writer would like to thank Professor K. Hoffman for calling his attention to the lemma contained in Halmos's paper [3] (then unpublished). In his mimeographed publication [5] Hoffman has also proceeded from this lemma. But as will become clear from the sequel and is commented upon after 3.8, there are differences between our approaches.

(3) Called *outer* and *inner* factors by Beurling.

Wold's Theorem (§ 4). (Beurling's Theorem IV, however, goes beyond the ambit of simple prediction theory, and to prove it recourse has to be taken to the decomposition lemma or to the alternative devices due to Beurling, Lax and others.) Thus we shall reveal the precise and close relationship between the theories of prediction and of functions in the Hardy class H_2.

We shall show all this for *vector-valued processes* and *matrix-valued functions* on the unit circle of the complex plane. For this we shall first extend the decomposition lemma from the Hilbert space $\mathfrak{H}$ to its Cartesian product $\mathfrak{H}^q$, (2.8), which we shall not treat as simply a Hilbert space but endow with a Gram-matricial structure (§ 2). The reasons for dealing with $\mathfrak{H}^q$ under this more complicated structure have been stated in [15, § 5].

A few bibliographical remarks are now in order. Prediction theory was extended to vector-valued processes in 1957–58 independently by Wiener and the writer [18, I, II], Helson and Lowdenslager [4], and Rosanov [17]. Vectorial extensions of Beurling's theorems (especially of Theorem IV) were given by Lax [7] in 1959. The optimal-residual factorization of matrix-valued functions in H_2 was given by the writer [11] in 1959. The further factorization of the residual factor, when it is of full rank, into a matricial Blaschke product and a purely residual factor follows, as the writer showed in [12, 13], from the remarkable work of Potapov [16] in 1955. In this paper all these results will emerge in what seems to us to be the most natural and simple setting.

2. The Wold Decomposition of a Hilbert Space

Halmos's Lemma 1, [3], may be stated as follows. Let S be an isometry on a (complex) Hilbert space $\mathfrak{H}$, and let $\mathfrak{R} = \mathfrak{S}(\mathfrak{H})$ be its range, then

$$\left.\begin{aligned} &S^j(\mathfrak{R}^\perp) \perp S^k(\mathfrak{R}^\perp) \perp \bigcap_{l=0}^{\infty} S^l(\mathfrak{H}) \quad (j > k \geqslant 0) \\ &\mathfrak{H} = \sum_{k=0}^{\infty} S^k(\mathfrak{R}^\perp) + \bigcap_{l=0}^{\infty} S^l(\mathfrak{H}). \end{aligned}\right\} \tag{2.1}$$

In this formulation the result resembles the Wold decomposition of the "past and present" subspace $\mathfrak{M}_0$ of a non-deterministic, weakly stationary stochastic process (S.P.) into its "innovation subspaces" and "remote past" (cf. e.g. [18], Part I, 6.10). We shall therefore speak of (2.1) as the *Wold decomposition of* $\mathfrak{H}$ *wrought by* S. It can be established by refining the arguments used in proving Wold's Theorem for a S.P. (in which $\mathfrak{R}^\perp$ is one-dimensional) after replacing $\mathfrak{M}_0$ by $\mathfrak{H}$ and U^* by S, U being the shift operator of the process.

We wish to extend (2.1) to the Cartesian product $\mathfrak{H}^q$ of q-dimensional (column-) vectors with components in $\mathfrak{H}$, where q is a positive integer. We do not, however, consider $\mathfrak{H}^q$ as simply an inner product space in the usual way, but endow it with a Gram-matricial structure [1]:

$$(\mathbf{f}, \mathbf{g}) = [(f^i, g^j)], \quad \mathbf{f} = (f^i)_{i=1}^q, \quad \mathbf{g} = (g^j)_{j=1}^q, \ \mathbf{f}, \mathbf{g} \in \mathfrak{H}^q. \tag{2.2}$$

We say that
$$\mathbf{f} \perp \mathbf{g}, \quad \textit{if and only if} \quad (\mathbf{f}, \mathbf{g}) = \mathbf{0}. \tag{2.3}$$

The Gram matrix $(\mathbf{f}, \mathbf{g})$ also yields the usual inner product trace $(\mathbf{f}, \mathbf{g})$, which is of no importance to us, and the usual *absolute value*

$$|\mathbf{f}| = \sqrt{\text{trace } (\mathbf{f}, \mathbf{f})} = \sum_{i=1}^{q} |f^i|^2, \tag{2.4}$$

which provides the appropriate topology for $\mathfrak{H}^q$. We take linear combinations of vectors $\mathbf{f}_k$ in $\mathfrak{H}^q$ with $q \times q$ matrix (and not just complex) coefficients, and so define linear manifolds and (closed) subspaces of $\mathfrak{H}^q$. We easily get the following lemma:

2.5 LEMMA. *$\mathcal{M}$ is a subspace of $\mathfrak{H}^q$, if and only if $\mathcal{M} = \mathfrak{M}^q$, where $\mathfrak{M}$ is a subspace of $\mathfrak{H}$; moreover $\mathfrak{M}$ is the set of all components of vectors in $\mathcal{M}$.*

We denote by $\mathfrak{S}(\mathbf{f})$, $\mathfrak{S}(\mathbf{f}_j)_{j \in J}$, the subspaces spanned by $\mathbf{f}$ singly, and by the family $\mathbf{f}_j$, $j \in J$. We have a projection theorem akin to that for $\mathfrak{H}$: every $\mathbf{f} \in \mathfrak{H}^q$ has a unique projection $(\mathbf{f}|\mathcal{M})$ on a subspace $\mathcal{M}$ such that

$$(\mathbf{f}|\mathcal{M}) \in \mathcal{M}, \quad \mathbf{f} - (\mathbf{f}|\mathcal{M}) \perp \mathcal{M}.$$

This yields the following:

2.6 LEMMA. *The subspace $\mathcal{M}$ of $\mathfrak{H}^q$ equals $\mathfrak{S}(\mathbf{f})$, if and only if $\mathbf{f} \in \mathcal{M}$, and $\mathbf{0} \neq \mathbf{g} \in \mathcal{M}$ implies that $\mathbf{g}$ is not orthogonal to $\mathbf{f}$.*

It would be incorrect, however, to speak of $\mathfrak{S}(\mathbf{f})$ as a "one-dimensional" subspace; for invariably there will be non-zero vectors $\mathbf{g}$, $\mathbf{h}$ in $\mathfrak{S}(\mathbf{f})$ such that $\mathbf{g} \perp \mathbf{h}$. Only those linear operators $\mathbf{T}$ on $\mathfrak{H}^q$ to $\mathfrak{H}^q$ are relevant to us, which stem from a linear operator T on $\mathfrak{H}$ by "inflation":

$$\mathbf{T}(\mathbf{f}) = (Tf^i)_{i=1}^q, \quad \mathbf{f} = (f^i)_{i=1}^q \in \mathfrak{H}^q. \tag{2.7}$$

For a detailed discussion of $\mathfrak{H}^q$ see [18, I, § 5], where unfortunately we have taken $\mathfrak{H} = L_2(\Omega, B, P)$, but our arguments are not dependent on this specific realization. The result (2.1) extends to $\mathfrak{H}^q$ as follows:

---

[1] Our usage of bold face and script letters is as follows: $\mathbf{f}$, $\mathbf{g}$ etc. denote members (i.e. vectors) of $\mathfrak{H}^q$, and $\mathcal{M}$, $\mathcal{N}$, etc. denote subspaces of $\mathfrak{H}^q$. $\mathbf{A}$, $\mathbf{B}$, etc. denote $q \times q$ matrices with complex-entries, and $\mathbf{\Phi}$, $\mathbf{\Psi}$, $\mathbf{X}$ etc. denote $q \times q$ matrix-valued functions.

2.8 THEOREM. (Wold decomposition of $\mathfrak{H}^q$). *Let S be an isometry on the (complex) Hilbert space $\mathfrak{H}$, $\mathbf{S}$ be the operator on $\mathfrak{H}^q$ to $\mathfrak{H}^q$ induced by S, cf. (2.7), and $\mathcal{R} = \mathbf{S}(\mathfrak{H}^q)$. Then*

$$\text{(a)} \quad \mathbf{S}^j(\mathcal{R}^\perp) \perp \mathbf{S}^k(\mathcal{R}^\perp) \perp \bigcap_{l=0}^{\infty} \mathbf{S}^l(\mathfrak{H}^q) \quad (j > k \geqslant 0),$$

$$\text{(b)} \quad \mathfrak{H}^q = \sum_{k=0}^{\infty} \mathbf{S}^k(\mathcal{R}^\perp) + \bigcap_{l=0}^{\infty} \mathbf{S}^l(\mathfrak{H}^q).$$

This theorem is deducible from (2.1) in a straightforward way. For instance, (a) can be had from the relations

$$\mathcal{R} = \mathfrak{R}^q, \quad \mathcal{R}^\perp = (\mathfrak{R}^\perp)^q, \quad \mathbf{S}^k(\mathcal{R}^\perp) = \{S^k(\mathfrak{R}^\perp)\}^q,$$

and

$$\mathfrak{M} \perp \mathfrak{N} \textit{ in } \mathfrak{H} \quad \textit{implies} \quad \mathfrak{M}^q \perp \mathfrak{N}^q \textit{ in } \mathfrak{H}^q.$$

We leave the details to the reader.

A concrete realization of $\mathfrak{H}^q$, which shall concern us, is obtained by taking as $\mathfrak{H}$ the space of q-dimensional (row-)vector-valued functions on the unit circle $C = [|z| = 1]$ of the complex plane, whose components are in L_2. It is easy to check that $\mathfrak{H}^q$ is then the space $\mathbf{L}_2$ of $q \times q$ matrix-valued functions on C whose entries are in L_2, and that

$$(\mathbf{\Phi}, \mathbf{\Psi}) = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{\Phi}(e^{i\theta}) \mathbf{\Psi}^*(e^{i\theta})\, d\theta \quad (\mathbf{\Phi}, \mathbf{\Psi} \in L_2). \tag{2.9}$$

The nth Fourier coefficient of $\mathbf{\Phi}$ is defined by

$$\mathbf{A}_n = \frac{1}{2\pi} \int_0^{2\pi} e^{-ni\theta} \mathbf{\Phi}(e^{i\theta})\, d\theta. \tag{2.10}$$

We shall denote by $\mathbf{L}_2^{0+}$, $\mathbf{L}_2^{+}$ the subspaces of $\mathbf{L}_2$ comprising functions $\mathbf{\Phi}$ for which $\mathbf{A}_n$ vanishes for $n < 0$, $n \leqslant 0$, respectively. Similarly we define $\mathbf{L}_2^{0-}$, $\mathbf{L}_2^{-}$ and also $\mathbf{L}_\infty$, $\mathbf{L}_\infty^{0+}$, etc.

3. The Wold and Beurling Theorems

As our first corollary of 2.8 we shall get Wold's Theorem for a non-deterministic process, cf. [18, I, 6.9, 6.10]:

3.1 WOLD'S THEOREM (Form I). *Let* (i) $(\mathbf{f}_n)_{-\infty}^{\infty}$ *be a q-variate, weakly stationary S.P., i.e. a bisequence in $\mathfrak{H}^q$ for which $(\mathbf{f}_m, \mathbf{f}_n) = \mathbf{\Gamma}_{m-n}$ depends only on the difference $m - n$;*

(ii) $$\mathfrak{M}_n = \mathfrak{S}(\mathfrak{f}_k)_{k \leq n}, \quad \mathfrak{M}_{-\infty} = \bigcap_{n \to -\infty}^{\infty} \mathfrak{M}_n;$$

(iii) $(\mathfrak{f}_n)_{-\infty}^{\infty}$ *be non-deterministic, i.e.*

$$\mathfrak{g}_n = \mathfrak{f}_n - (\mathfrak{f}_n \mid \mathfrak{M}_{n-1}) \neq 0 \quad (-\infty < n < \infty).$$

Then (a) $(\mathfrak{g}_n)_{-\infty}^{\infty}$ *is orthogonal,* i.e. $(\mathfrak{g}_m, \mathfrak{g}_n) = \delta_{mn}(\mathfrak{g}_0, \mathfrak{g}_0)$;

(b) $$\mathfrak{M}_n = \sum_{k=0}^{\infty} \mathfrak{S}(\mathfrak{g}_{n-k}) + \mathfrak{M}_{-\infty}, \quad \sum_{j=-\infty}^{\infty} \mathfrak{S}(\mathfrak{g}_j) \perp \mathfrak{M}_{-\infty}.$$

Proof. We know, cf. [18, I, § 6], that there exists a unitary operator U on $\mathfrak{H}$ such that

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

where $\mathfrak{f}_n = (f_n^i)_{i=1}^q$. Now by 2.5, $\mathfrak{M}_n = \mathfrak{M}_n^q$, where $\mathfrak{M}_n$ is a subspace of $\mathfrak{H}$. Since $U^*(\mathfrak{M}_n) = \mathfrak{M}_{n-1} \subseteq \mathfrak{M}_n$, therefore the restriction S of U^* to $\mathfrak{M}_n$ is an isometry on $\mathfrak{M}_n$. Hence by 2.8

$$S^j(\mathcal{R}^\perp) \perp S^k(\mathcal{R}^\perp) \perp \bigcap_{l=0}^{\infty} S^l(\mathfrak{M}_n) \quad (j > k \geq 0), \tag{1}$$

$$\mathfrak{M}_n = \sum_{k=0}^{\infty} S^k(\mathcal{R}^\perp) + \bigcap_{l=0}^{\infty} S^l(\mathfrak{M}_n), \tag{2}$$

where $$\mathcal{R} = S(\mathfrak{M}_n) = \mathfrak{M}_{n-1}, \quad \mathcal{R}^\perp = \mathfrak{M}_n \cap \mathfrak{M}_{n-1}^\perp.$$

Obviously $\bigcap_{l=0}^{\infty} S^l(\mathfrak{M}_n) = \mathfrak{M}_\infty$. Hence (a), (b) will follow from (1), (2) if we can show that

$$S^k(\mathfrak{M}_n \cap \mathfrak{M}_{n-1}^\perp) = \mathfrak{M}_{n-k} \cap \mathfrak{M}_{n-k-1}^\perp = \mathfrak{S}(\mathfrak{g}_{n-k}).$$

But this is obvious, since from (ii), (iii)

$$\mathfrak{M}_{n-k} = \mathfrak{M}_{n-k-1} + \mathfrak{S}(\mathfrak{f}_{n-k}) = \mathfrak{M}_{n-k-1} + \mathfrak{S}(\mathfrak{g}_{n-k}),$$

and $\mathfrak{M}_{n-k-1} \perp \mathfrak{g}_{n-k}$ (Q.E.D.).

As shown in [18, I, 6.11] this result on the subspaces of the process can be translated into one on the *moving-average representation* of the process itself:

3.2 WOLD'S THEOREM (Form II). *Let* $\mathfrak{f}_n$, $\mathfrak{g}_n$, *etc. be as in* 3.1 *and let* $\mathfrak{G} = (\mathfrak{g}_0, \mathfrak{g}_0)$. *Then*

$$\mathfrak{f}_n = \sum_{k=0}^{\infty} A_k \mathfrak{g}_{n-k} + (\mathfrak{f}_n \mid \mathfrak{M}_{-\infty}), \quad \mathfrak{g}_j \perp (\mathfrak{f}_n \mid \mathfrak{M}_{-\infty}),$$

where
$$\mathbf{A}_k \mathbf{G} = (\mathbf{f}_0, \mathbf{g}_{-k}), \quad \mathbf{A}_0 \mathbf{g}_0 = \mathbf{g}_0,$$

$$\sum_{k=0}^{\infty} | \mathbf{A}_k \sqrt{\mathbf{G}} |_E^2 < \infty, \text{ (1)} \quad \mathbf{A}_0 \sqrt{\mathbf{G}} = \sqrt{\mathbf{G}}.$$

The matrices $\mathbf{A}_k$ *are not necessarily unique, but* $\mathbf{A}_k \mathbf{G}$ *and* $\mathbf{A}_k \sqrt{\mathbf{G}}$ *are unique* (*cf.* [9]).

The bisequence $(\mathbf{g}_n)_{-\infty}^{\infty}$, given in 3.1 (iii) is called the *innovation process* of $(\mathbf{f}_n)_{-\infty}^{\infty}$, and $\mathbf{G} = (\mathbf{g}_0, \mathbf{g}_0)$ is called its *lag* 1 *prediction error matrix.* The innovation process can be "normalized". For this let $\mathfrak{C}^q$ be the space of q-dimensional (column-) vectors with complex components, and $\mathbf{R}$ be the range of the linear operator on $\mathfrak{C}^q$ represented by the matrix $\mathbf{G}$ in the privileged basis of $\mathfrak{C}^q$, and let the matrix $\mathbf{J}$ represent the orthogonal projection from $\mathfrak{C}^q$ onto $\mathbf{R}$ (same basis). Then

$$\mathbf{J} = \mathbf{J}^* = \mathbf{J}^2 \quad \sqrt{\mathbf{G}}\mathbf{J} = \sqrt{\mathbf{G}} = \mathbf{J}\sqrt{\mathbf{G}}, \tag{3.3}$$

and from 3.1 (a) it follows that $\mathbf{J}\mathbf{g}_n = \mathbf{g}_n$. Elementary considerations show that there exists a positive definite, *invertible* matrix $\mathbf{H}$ such that

$$\mathbf{H}\sqrt{\mathbf{G}} = \mathbf{J} = \sqrt{\mathbf{G}}\mathbf{H}. \tag{3.4}$$

Let $\mathbf{h}_n = \mathbf{H}\mathbf{g}_n$; then we easily find that

$$(\mathbf{h}_m, \mathbf{h}_n) = \delta_{mn} \mathbf{H}\mathbf{G}\mathbf{H} = \delta_{mn} \mathbf{J},$$

$$\mathbf{g}_n = \mathbf{J}\mathbf{g}_n = \sqrt{\mathbf{G}}\mathbf{h}_n.$$

$(\mathbf{h}_n)_{-\infty}^{\infty}$ is called the *normalized innovation process* of $(\mathbf{f}_n)_{-\infty}^{\infty}$. We thus get

3.5 WOLD'S THEOREM (Form III).(2) *Let* $\mathbf{f}_n$, $\mathbf{g}_n$, *etc. be as in* 3.1, *and* $\mathbf{J}$ *represent the projection onto the range of the linear transformation given by* $\mathbf{G} = (\mathbf{g}_0, \mathbf{g}_0)$. *Then*

$$\mathbf{f}_n = \sum_{k=0}^{\infty} \mathbf{A}_k \sqrt{\mathbf{G}} \mathbf{h}_{n-k} + (\mathbf{f}_n | \mathfrak{M}_{-\infty}), \quad \mathbf{h}_k \perp (\mathbf{f}_n | \mathfrak{M}_{-\infty}),$$

where $(\mathbf{h}_n)_{-\infty}^{\infty}$ *is the normalized innovation process:*

$$(\mathbf{h}_m, \mathbf{h}_n) = \delta_{mn} \mathbf{J}, \quad \mathbf{g}_n = \sqrt{\mathbf{G}}\mathbf{h}_n, \quad \mathbf{h}_n = \mathbf{H}\mathbf{g}_n,$$

---

(1) The *Euclidean norm* $|\mathbf{A}|_E$ of a matrix $\mathbf{A} = [a_{ij}]$ is defined by

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

(2) This result is an extension of [18, I, 6.12], in which we assumed that rank $\mathbf{G} = q$.

H *is as in* (3.4),

$$\mathbf{A}_k\sqrt{\mathbf{G}} = (\mathbf{f}_0, \mathbf{h}_{-k}), \quad \mathbf{A}_0\sqrt{\mathbf{G}} = \sqrt{\mathbf{G}}, \quad \sum_{k=0}^{\infty}|\mathbf{A}_k\sqrt{\mathbf{G}}|_E^2 < \infty.$$

The function $\mathbf{\Phi} = \sum_{k=0}^{\infty}\mathbf{A}_k\sqrt{\mathbf{G}}\,e^{ki\theta}$, which is clearly in $\mathbf{L}_2^{0+}$, is called the *generating function* of the process $(\mathbf{f}_n)_{-\infty}^{\infty}$. The relations (3.3), (3.4) yield a canonical expression for $\mathbf{\Phi}$. Thus, let $\mathbf{\Omega} = (\mathbf{I}-\mathbf{J}) + \mathbf{\Phi}\mathbf{H}$. Then clearly $\mathbf{\Omega}\in\mathbf{L}_2^{0+}$, $\mathbf{\Omega}_+(0) = \mathbf{I}$,(1) and $\mathbf{\Omega}\sqrt{\mathbf{G}} = \mathbf{\Phi}$. Since $\det\mathbf{\Omega}_+$ is in the Hardy class $H_{2/q}$ on the disk $[|z|<1]$, and does not vanish identically, it follows that $\det\mathbf{\Omega}$ vanishes almost nowhere on C. Thus

3.6 Corollary. *The generating function* $\mathbf{\Phi}$ *of a non-deterministic S.P. is always expressible in the form* $\mathbf{\Omega}\sqrt{\mathbf{G}}$, *where* $\mathbf{G}$ *is the prediction error matrix, and* $\mathbf{\Omega}\in\mathbf{L}_2^{0+}$ *is invertible a.e. on* C, *and* $\mathbf{\Omega}_+(0) = \mathbf{I}$.

A crucial property of the generating function is given by the following lemma proved by us in [8, 2.9]:

3.7 Basic Lemma. *If* $\mathbf{\Phi}$ *is the generating function of a non-deterministic S.P., then* (2) $\mathbf{\Phi}_+(0) \succ \mathbf{0}$, *and*

$$\mathbf{0} \neq \mathbf{\Psi}\in\mathbf{L}_2^{0+} \quad \textit{and} \quad \mathbf{\Psi}\mathbf{\Psi}^* = \mathbf{\Phi}\mathbf{\Phi}^* \ \textit{a.e.} \quad \textit{implies} \quad \mathbf{\Psi}_+(0)\mathbf{\Psi}_+(0)^* \prec \{\mathbf{\Phi}_+(0)\}^2.$$

We turn next to the application of 2.8 to the theory of shift invariant subspaces of L_2^{0+}. This yields the following matricial extension of Beurling's Theorem IV [1]:

3.8 Beurling–Lax Theorem. *Let* $\mathbf{M}$ *be a proper subspace of* $\mathbf{L}_2^{0+}$ *such that* $e^{i\theta}\mathfrak{M}\subseteq\mathfrak{M}$.

Then

(a) $$\mathfrak{M} = \mathbf{L}_2^{0+}\mathbf{X},$$

where $\mathbf{X} = \mathbf{J}\mathbf{U}\in\mathbf{L}_\infty^{0+}$, $\mathbf{J}$ *is a (constant) projection matrix and* $\mathbf{U}$ *is an a.e. unitary matrix-valued function on* C;

(b) *In* (a) *the function* $\mathbf{X} = \mathbf{J}\mathbf{U}\in\mathbf{L}_\infty^{0+}$ *is unique up to a constant unitary matrix pre-factor.*

---

(1) Here and throughout the sequel the symbol $\mathbf{\Psi}_+$ will denote the holomorphic extension to the disk $[|z|<1]$ of a function $\mathbf{\Psi}\in\mathbf{L}_2^{0+}$ on $C = [|z| = 1]$.

(2) Here and in the sequel for matrices $\mathbf{A}$, $\mathbf{B}$ we shall write $\mathbf{A}\succ\mathbf{B}$ or $\mathbf{B}\prec\mathbf{A}$ to mean that $\mathbf{A}-\mathbf{B}$ is non-negative hermitian.

Relation between 3.8 and Lax's Theorem. Theorem 3.8 differs in form from Lax's vectorial extension [7, p. 164] of Beurling's Theorem IV. In the first place, Lax deals with functions on $(-\infty, \infty)$ and not on $C = [|z| = 1]$; but we may suppose that this difference is removed by an initial conformal transformation of the upper half-plane onto the disk $[|z| < 1]$. Even so, Lax's functions are q-dimensional vector-valued, whereas ours are $q \times q$ matrix-valued. Thus, Lax deals with a shift invariant subspace $\dot{\mathfrak{M}}$ of $(L_2^{0+})^q$. The generating element of $\mathfrak{M}$ turns out to be a function F whose values are almost everywhere isometries from a fixed subspace $\mathfrak{C}^p$ of $\mathfrak{C}^q$ to $\mathfrak{C}^q$ $(p \leqslant q)$. His result takes the form

$$\mathfrak{M} = F(L_2^{0+})^p.$$

We, on the other hand, never leave the space $\mathbf{L}_2^{0+}$ of $q \times q$ matrix valued functions, and provide for possible degeneracies in rank by introduction of the projection matrix $\mathbf{J}$. Our generator $\mathbf{X}$ occurs as a post-factor and not like Lax's F as a pre-factor. Notwithstanding these differences the two results are equivalent, and hence our title for 3.8. Our proof of 3.8, based on 2.8, is close to K. Hoffman's, who uses the Halmos lemma (2.1) [5, p. 153]. But Hoffman adheres to Lax's enunciation, and accordingly has, like Lax, to get hold of a lower dimensional Hilbert space. This is avoided in our approach, which we feel best reveals the connection between this subject and prediction theory. Indeed, functions of the form $\mathbf{JU} \in \mathbf{L}_\infty^{0+}$ enter quite naturally in the prediction theory proofs of factorization theorems, as we first noticed in [11] and as will be seen in § 4 below.

Proof of 3.8. (a) By 2.5, $\mathcal{M} = \mathfrak{M}^q$, where $\mathfrak{M}$ is the space of all vectors which are rows of matrix functions in $\mathcal{M}$. It follows from our hypothesis that $e^{i\theta}\mathfrak{M} \subseteq \mathfrak{M}$. Hence letting

$$S(\psi) = e^{i\theta}\psi \quad (\psi \in \mathfrak{M}),$$

it follows that S is an isometry on $\mathfrak{M}$ to $\mathfrak{M}$. Now consider the induced operator S on $\mathcal{M}$:

$$\mathrm{S}(\Psi) = e^{i\theta}\Psi \quad (\Psi \in \mathcal{M}).$$

We have
$$\bigcap_{l=0}^{\infty} \mathrm{S}^l(\mathcal{M}) \subseteq \bigcap_{l=0}^{\infty} e^{li\theta}\mathbf{L}_2^{0+} = \{0\}.$$

Hence by Theorem 2.8

$$\mathcal{M} = \sum_{k=0}^{\infty} \mathrm{S}^k(\mathcal{R}^\perp), \quad \mathrm{S}^j(\mathcal{R}^\perp) \perp \mathrm{S}^k(\mathcal{R}^\perp) \quad (j > k \geqslant 0), \tag{1}$$

where
$$\mathcal{R} = e^{i\theta}\mathcal{M}, \quad \mathcal{R}^\perp = \mathcal{M} \cap (e^{i\theta}\mathcal{M})^\perp \subseteq \mathbf{L}_2^{0+}. \tag{2}$$

It only remains to show that $\mathcal{R}^{\perp}=\mathfrak{S}(\mathbf{X}_1)$ where $\mathbf{X}_1=\mathbf{J}_1\mathbf{U}_1\in\mathbf{L}_{\infty}^{0+}$. For then, $\mathbf{S}^k(\mathcal{R}^{\perp})=e^{ki\theta}\mathfrak{S}(\mathbf{X}_1)$ and (1) will imply that for any $\mathbf{\Psi}\in\mathcal{M}$,

$$\mathbf{\Psi}=\sum_{k=0}^{\infty}\mathbf{A}_k e^{ki\theta}\mathbf{X}_1=\mathbf{\Phi}\mathbf{X}_1,\quad \mathbf{\Phi}\in\mathbf{L}_2^{0+},$$

as desired. In the light of Lemma 2.6, our objective reduces to showing that

$$\mathbf{X}_1=\mathbf{J}_1\mathbf{U}_1\in\mathcal{R}^{\perp}\ \textit{such that}\ 0\neq\mathbf{\Psi}\in\mathcal{R}^{\perp}\ \textit{implies}\ \mathbf{\Psi}\ \textit{is not orthogonal to}\ \mathbf{X}_1. \tag{A}$$

Let $\mathbf{\Psi},\mathbf{X}\in\mathcal{R}^{\perp}$. Then from the second relation in (1)

$$\mathbf{\Psi}\perp e^{ki\theta}\mathbf{X},\quad \mathbf{X}\perp e^{ki\theta}\mathbf{\Psi}\quad (k\geqslant 1).$$

It follows from (2.3), (2.9), (2.10) that

$$\mathbf{\Psi}\mathbf{X}^*=\text{const.}=(\mathbf{\Psi},\mathbf{X})\quad \textit{a.e. on } C.$$

From this we draw two inferences:

$$\textit{for all}\ \mathbf{\Psi}\in\mathcal{R}^{\perp},\ \mathbf{\Psi}\mathbf{\Psi}^*=(\mathbf{\Psi},\mathbf{\Psi})\quad \textit{a.e., and so}\ \mathbf{\Psi}\in\mathbf{L}_{\infty}^{0+}; \tag{3}$$

$$\textit{for all}\ \mathbf{\Psi},\mathbf{X}\in\mathcal{R}^{\perp},\ \mathbf{\Psi}\perp\mathbf{X}\ \textit{if and only if}\ \mathbf{\Psi}\mathbf{X}^*=0\quad \textit{a.e.} \tag{4}$$

From (3) and the polar decomposition we get $\mathbf{\Psi}=\sqrt{(\mathbf{\Psi},\mathbf{\Psi})}\,\mathbf{U}$, where $\mathbf{U}(e^{i\theta})$, is unitary a.e. Now put $\mathbf{G}=(\mathbf{\Psi},\mathbf{\Psi})$ and define the matrices $\mathbf{J},\mathbf{H}$ as was done above in (3.3), (3.4). Then it easily follows that

$$\textit{for all}\quad \mathbf{\Psi}\in\mathcal{R}^{\perp},\quad \mathbf{\Psi}=\sqrt{(\mathbf{\Psi},\mathbf{\Psi})}\,\mathbf{X},\quad \mathbf{X}=\mathbf{J}\mathbf{U}=\mathbf{H}\mathbf{\Psi}\in\mathcal{R}^{\perp}\subseteq\mathbf{L}_{\infty}^{0+}. \tag{5}$$

In view of (5) and (4) our goal (A) may be restated:

$$\textit{There exists a}\quad \mathbf{X}_1=\mathbf{J}_1\mathbf{U}_1\in\mathcal{R}^{\perp}\quad \textit{such that}\quad 0\neq\mathbf{X}=\mathbf{J}\mathbf{U}\in\mathcal{R}^{\perp}\quad \textit{implies}\quad \mathbf{X}_1\mathbf{X}^*\neq 0. \tag{A$'$}$$

To prove (A′), we note that by (5) the rank of any $\mathbf{\Psi}$ in $\mathfrak{R}^{\perp}$ is essentially constant. Let

$$p=\max\{r: r=\operatorname{rank}\mathbf{\Psi},\ \text{a.e.},\ \mathbf{\Psi}\in\mathcal{R}^{\perp}\}. \tag{6}$$

Then $\qquad$ *there exists a* $\mathbf{X}_0=\mathbf{J}_0\mathbf{U}_0\in R^{\perp}$ *such that* $\operatorname{rank}\mathbf{J}_0=p$.

In case $p=q$, we have $\mathbf{J}_0=\mathbf{I}$. Hence we can take $\mathbf{X}_1=\mathbf{X}_0=\mathbf{U}_0$ in (A′), which then becomes utterly obvious. In case $1\leqslant p<q$, we can find a unitary matrix $\mathbf{V}_0$ such that

$$\mathbf{J}_1=\mathbf{V}_0\mathbf{J}_0\mathbf{V}_0^*=\text{diag.}\,[1,\ldots,1,0,\ldots,0]\ \textit{of rank}\ p.\text{[1]} \tag{7}$$

[1] diag. $[a_1,\ldots,a_q]$ denotes the $q\times q$ diagonal matrix with entries $a_1,\ldots,a_q$ along the diagonal.

Then, letting $\mathbf{U}_1 = \mathbf{V}_0\mathbf{U}_0$ we get

$$\mathbf{X}_1 = \mathbf{J}_1\mathbf{U}_1 = \mathbf{V}_0\mathbf{X}_0 \in \mathcal{R}^\perp. \tag{8}$$

Now suppose, in contradiction to (A′), that

$$0 \neq \mathbf{J}\mathbf{U} \in \mathcal{R}^\perp \quad \text{and} \quad \mathbf{X}_1\mathbf{U}^*\mathbf{J} = 0. \tag{9}$$

Take any projection matrix $\mathbf{K}$ such that $\mathbf{J}\mathbf{K} = \mathbf{K}$, and $0 < r = \operatorname{rank} \mathbf{K} < q - p$. We can find a unitary matrix $\mathbf{V}$ such that

$$\mathbf{J}_2 = \mathbf{V}\mathbf{K}\mathbf{V}^* = \text{diag. } [0, \ldots, 0, 1, \ldots, 1] \textit{ of rank } r.$$

Then
$$\mathbf{X}_2 = \mathbf{V}\mathbf{K}\mathbf{J}\mathbf{U} = \mathbf{V}\mathbf{K}\mathbf{U} = \mathbf{J}_2\mathbf{V}^*\mathbf{U} = \mathbf{J}_2\mathbf{U}_2,$$

where $\mathbf{U}_2 = \mathbf{V}\mathbf{U}$. From (9)

$$\mathbf{X}_2 = \mathbf{J}_2\mathbf{U}_2 \in \mathcal{R}^\perp \quad \textit{and} \quad \mathbf{X}_1\mathbf{X}_2^* = 0. \tag{10}$$

From (8) and (10), $\mathbf{X}_1 + \mathbf{X}_2 \in \mathcal{R}^\perp$. But since the non-zero rows of $\mathbf{X}_1, \mathbf{X}_2$ have disjoint locations, in view of our selection of $\mathbf{J}_1, \mathbf{J}_2$ and the fact that $0 < r < q - p$, and since $\mathbf{X}_1\mathbf{X}_2^* = \mathbf{0}$, it follows easily that

$$\operatorname{rank} (\mathbf{X}_1 + \mathbf{X}_2) = \operatorname{rank} \mathbf{X}_1 + \operatorname{rank} \mathbf{X}_2 = p + r > p.$$

This contradicts (6). Hence our supposition (9) is untenable, i.e. we have (A′). This completes the proof of (a).

(b) Suppose that $\mathcal{M} = \mathbf{L}_2^{0+}\mathbf{X}_1$, $\mathbf{X}_1 = \mathbf{J}_1\mathbf{U}_1 \in \mathbf{L}_\infty^{0+}$. Then $\mathbf{X}_1\mathbf{X}_1^* = \mathbf{J}_1$, and so $(\mathbf{X}_1, \mathbf{X}_1 e^{ki\theta}) = \mathbf{0}$ for $k > 0$. Thus

$$\mathbf{X}_1 \perp \mathfrak{S}(\mathbf{X}_1 e^{ki\theta})_{k\geq 1} = e^{i\theta}\mathfrak{S}(\mathbf{X}_1 e^{ki\theta})_{k\geq 0} = e^{i\theta}\mathcal{M}.$$

Hence $\mathbf{X}_1 \in \mathcal{M} \cap (e^{i\theta}\mathcal{M})^\perp = \mathcal{R}^\perp$, cf. (2). It follows from (A) and Lemma 2.6 that

$$\mathbf{X}_1 = \mathbf{A}\mathbf{X}, \textit{ where } \mathbf{X} = \mathbf{J}\mathbf{U} \in \mathcal{R}^\perp \textit{ is as in } \text{(a)}. \tag{11}$$

Obviously, $\mathbf{J}_1 = \mathbf{A}\mathbf{J}\mathbf{A}^*$, and hence by the polar decomposition $\mathbf{A}\mathbf{J} = \mathbf{J}_1\mathbf{V}_1$, where $\mathbf{V}_1$ is unitary. It follows that

$$\mathbf{J}_1(\mathbf{V}_1\mathbf{J}\mathbf{V}_1^*) = \mathbf{A}\mathbf{J}\mathbf{J}\mathbf{V}_1^* = \mathbf{A}\mathbf{J}\mathbf{V}_1^* = \mathbf{J}_1.$$

Hence
$$\mathbf{J}_1 \prec \mathbf{V}_1\mathbf{J}\mathbf{V}_1^*. \tag{12}$$

But since $\mathbf{X} \in \mathcal{M}$, i.e. $\mathbf{J}\mathbf{U} \in \mathbf{L}_2^{0+}\mathbf{J}_1\mathbf{U}_1$, therefore $\operatorname{rank} \mathbf{J} \leqslant \operatorname{rank} \mathbf{J}_1$. This together with (12) shows that $\mathbf{J}_1 = \mathbf{V}_1\mathbf{J}\mathbf{V}_1^*$, from which we get $\mathbf{A}\mathbf{J} = \mathbf{J}_1\mathbf{V}_1 = \mathbf{V}_1\mathbf{J}$. (11) now gives

$$\mathbf{X}_1 = \mathbf{A}\mathbf{J}\mathbf{X} = \mathbf{V}_1\mathbf{J}\mathbf{X} = \mathbf{V}_1\mathbf{X},$$

as desired. (Q.E.D.)

4. The Factorization and Closure Theorems for $\mathbf{L}_2^{0+}$

The results 3.7 and 3.8 suggest the following definition:

4.1. DEFINITION.(1) (a) *We call* $\mathbf{\Phi}$ *an* optimal function *in* L_2^{0+}, *if*

$$\mathbf{\Phi} \in \mathbf{L}_2^{0+}, \quad \mathbf{\Phi}_+(0) > \mathbf{0},$$

and $$\mathbf{\Psi} \in \mathbf{L}_2^{0+} \text{ and } \mathbf{\Psi}\mathbf{\Psi}^* = \mathbf{\Phi}\mathbf{\Phi}^* \text{ a.e. implies } \mathbf{\Psi}_+(0)\mathbf{\Psi}_+(0)^* < \{\mathbf{\Phi}_+(0)\}^2.$$

(b) *We call* $\mathbf{X}$ *a* residual function, *if* $\mathbf{X} = \mathbf{J}\mathbf{U} \in \mathbf{L}_\infty^{0+}$, *where* $\mathbf{J}$ *is a (constant) projection matrix and* $\mathbf{U}(e^{i\theta})$ *is unitary a.e.*

The Parseval Identity [18, I, 3.9(c)] at once yields:

4.2. LEMMA. *Let* $\mathbf{X} = \mathbf{J}\mathbf{U} \in \mathbf{L}_\infty^{0+}$. *Then* $\mathbf{X}_+(0)\mathbf{X}_+(0)^* < \mathbf{J}$; *moreover* $\mathbf{X}_+(0) = \mathbf{J}$ *implies* $\mathbf{X} = \mathbf{J}$ *a.e.*

A culminating point in the work of Riesz, Nevanlinna and Szegö is the result that complex-valued functions in L_δ^{0+}, $0 < \delta \leqslant \infty$, can be factored into unique optimal and residual functions. In [11] we extended this result to the matricial space $\mathbf{L}_2^{0+}$ by using the isomorphism between the time and spectral domains of a S.P., cf. [10]. We shall now indicate how recourse to this isomorphism can be avoided by applying the Wold Theorem 3.5 directly to the weakly stationary matricial S.P. $(\mathbf{\Psi}e^{-ki\theta})_{-\infty}^{\infty}$, for which $\mathbf{\Psi} \in \mathbf{L}_2^{0+}$, and the shift into the future is given by multiplication by $e^{-i\theta}$. This process is a bisequence in the space $H^q = \mathbf{L}_2$, cf. § 2. For this S.P.

$$\mathcal{M}_{-n} = \mathfrak{S}(\mathbf{\Psi}e^{ki\theta})_{k \geqslant n} \subseteq \mathfrak{S}(\mathbf{I}e^{ki\theta})_{k \leqslant n},$$

and hence $\mathcal{M}_{-\infty} = \{0\}$. Wold's Theorem 3.5 is therefore applicable, and shows that the process is a one-sided moving average of its normalized innovation process. This yields the desired theorem:

4.3. FACTORIZATION THEOREM. *Let* $0 \neq \mathbf{\Psi} \in \mathbf{L}_2^{0+}$, *and* $\mathbf{\Phi}, \mathbf{X}$ *be the generating function and normalized innovation of the S.P.* $(\mathbf{\Psi}e^{-ki\theta})_{-\infty}^{\infty}$. *Then*

(a) $$\mathbf{\Psi} = \mathbf{\Phi}\mathbf{X} \text{ a.e.};$$

(b) $\mathbf{X}$ *is residual,* $\mathbf{X}\mathbf{X}^* =$ *projection on range of* $\mathbf{\Phi}_+(0)$, *and* $\mathfrak{S}(\mathbf{X}e^{ki\theta})_{k=0}^{\infty} = \mathfrak{S}(\mathbf{\Psi}e^{ki\theta})_{k=0}^{\infty}$;

(c) $\mathbf{\Phi}$ *is optimal in* $\mathbf{L}_2^{0+}$ *and* $\mathbf{\Phi}\mathbf{\Phi}^* = \mathbf{\Psi}\mathbf{\Psi}^*$.

---

(1) Beurling [1] uses the terms "outer" and "inner" in place of "optimal" and "residual".

Proof. (a) Let $\mathbf{G}$ be the prediction error matrix of the S.P. $(\mathbf{\Psi} e^{-ki\theta})_{-\infty}^{\infty}$, and $\mathbf{J}$ be the projection onto the range of $\mathbf{G}$. Then by 3.5 $(\mathbf{X}, \mathbf{X} e^{ki\theta}) = \delta_{0k}\mathbf{J}$, and therefore

$$\mathbf{X}\mathbf{X}^* = \mathbf{J} \text{ a.e.}, \quad \mathbf{X} \in \mathbf{L}_\infty. \tag{1}$$

Next, since $\mathbf{M}_{-\infty} = \{0\}$, therefore again by 3.5

$$\mathbf{\Psi} = \sum_{k=0}^{\infty} \mathbf{A}_k \sqrt{\mathbf{G}} \mathbf{X} e^{ki\theta} = \mathbf{\Phi}\mathbf{X},$$

the last equality being permissible since $\sum_{k=0}^{\infty} |\mathbf{A}_k \sqrt{\mathbf{G}}|_E^2 < \infty$ and $\mathbf{X} \in L_\infty$.

(b) Since $\sqrt{\mathbf{G}} = \mathbf{A}_0 \sqrt{\mathbf{G}} = \mathbf{\Phi}_+(0)$, it follows cf. (1) that $\mathbf{J}$, i.e. $\mathbf{X}\mathbf{X}^*$, is the projection onto range of $\mathbf{\Phi}_+(0)$. Next, by (1) and the polar decomposition

$$\mathbf{X} = \mathbf{J}\mathbf{U}, \quad \mathbf{U}(e^{i\theta}) = \text{unitary a.e.} \tag{2}$$

Again since $\mathbf{M}_{-\infty} = \{0\}$, and $\sqrt{\mathbf{G}}\mathbf{X}$ is the (non-normalized) innovation, therefore by 3.1(b)

$$\mathfrak{S}(\mathbf{\Psi} e^{ki\theta})_{k=0}^{\infty} = \mathbf{M}_0 = \mathfrak{S}(\sqrt{\mathbf{G}}\mathbf{X} e^{ki\theta})_{k=0}^{\infty}. \tag{3}$$

But $\mathbf{X} = \mathbf{H}(\sqrt{\mathbf{G}}\mathbf{X})$, where $\mathbf{H}$ is invertible; hence (3) reduces to the second equality in (b). This equality shows in particular that $\mathbf{X} \in \mathbf{L}_2^{0+}$. But by (1) $\mathbf{X} \in \mathbf{L}_\infty$. Hence $\mathbf{X} \in \mathbf{L}_\infty^{0+}$, i.e. $\mathbf{X}$ is residual.

(c) The Basic Lemma 3.7 and Definition 4.1(a) show that $\mathbf{\Phi}$ is optimal in $\mathbf{L}_2^{0+}$. Finally, since $\mathbf{A}_k \sqrt{\mathbf{G}} = \mathbf{A}_k \sqrt{\mathbf{G}}\mathbf{J}$, therefore $\mathbf{\Phi} = \mathbf{\Phi}\mathbf{J}$; hence by (a) and (1)

$$\mathbf{\Psi}\mathbf{\Psi}^* = \mathbf{\Phi}\mathbf{X}\mathbf{X}^*\mathbf{\Phi}^* = \mathbf{\Phi}\mathbf{J}\mathbf{\Phi}^* = \mathbf{\Phi}\mathbf{\Phi}^*. \quad \text{(Q.E.D.)}$$

4.4. Uniqueness Theorem. *Let* $0 \neq \mathbf{\Psi} \in \mathbf{L}_2^{0+}$. *Then*

(a) *there exists a unique function* $\mathbf{\Phi}$ *such that* $\mathbf{\Phi}$ *is optimal in* $\mathbf{L}_2^{0+}$ *and* $\mathbf{\Phi}\mathbf{\Phi}^* = \mathbf{\Psi}\mathbf{\Psi}^*$ *a.e.*;

(b) *there exists a unique function* $\mathbf{X}$ *such that* $\mathbf{X}$ *is residual,* $\mathbf{\Psi} = \mathbf{\Phi}\mathbf{X}$, $\mathfrak{S}(\mathbf{X} e^{ki\theta})_{k=0}^{\infty} = \mathfrak{S}(\mathbf{\Psi} e^{ki\theta})_{k=0}^{\infty}$, *where* $\mathbf{\Phi}$ *is as in* (a).

Proof. (a) Suppose that $\mathbf{\Phi}_1$ is optimal in $\mathbf{L}_2^{0+}$ and $\mathbf{\Phi}_1\mathbf{\Phi}_1^* = \mathbf{\Psi}\mathbf{\Psi}^*$ a.e. In view of 4.3(c), to prove (a) we need only show that $\mathbf{\Phi}_1 = \mathbf{\Phi}$, where $\mathbf{\Phi}$ is the generating function of the S.P. $(\mathbf{\Psi} e^{-ki\theta})_{-\infty}^{\infty}$. Now since $\mathbf{\Phi}_1\mathbf{\Phi}_1^* = \mathbf{\Psi}\mathbf{\Psi}^*$, the processes $(\mathbf{\Psi} e^{-ki\theta})_{-\infty}^{\infty}$, $(\mathbf{\Phi}_1 e^{-ki\theta})_{-\infty}^{\infty}$ have the same covariance structure and therefore the same generating function, viz. $\mathbf{\Phi}$. Hence by 4.3

$$\mathbf{\Phi}_1 = \mathbf{\Phi}\mathbf{X}_1, \quad \mathbf{X}_1 = \mathbf{J}\mathbf{U}_1 \in \mathbf{L}_\infty^{0+}, \tag{1}$$

where $\mathbf{J}$ is the projection onto the range of $\mathbf{\Phi}_+(0)$. Moreover by 4.3(c) $\mathbf{\Phi\Phi}^* = \mathbf{\Phi}_1\mathbf{\Phi}_1^*$. From this and the optimality of $\mathbf{\Phi}, \mathbf{\Phi}_1$, it follows that

$$\mathbf{\Phi}_+(0) = \mathbf{\Phi}_{1+}(0) = \sqrt{\mathbf{G}}, \text{ say,}$$

i.e. by (1) $\sqrt{\mathbf{G}} = \sqrt{\mathbf{G}}\mathbf{X}_{1+}(0)$. Premultiplying this by the matrix $\mathbf{H}$ of (3.4) we get $\mathbf{J} = \mathbf{J}\mathbf{X}_{1+}(0)$, i.e. $\mathbf{J} = \mathbf{X}_{1+}(0)$. Hence by 4.2, $\mathbf{X}_1 = \mathbf{J}$, a.e. The equation (1) thus reduces to $\mathbf{\Phi}_1 = \mathbf{\Phi J}$, i.e. to $\mathbf{\Phi}_1 = \mathbf{\Phi}$.

(b) Suppose that $\mathbf{X}_1 = \mathbf{J}_1\mathbf{U}_1 \in \mathbf{L}_\infty^{0+}$, $\mathbf{\Psi} = \mathbf{\Phi X}_1$, and

$$\mathfrak{S}(\mathbf{X}_1 e^{ki\theta})_{k=0}^{\infty} = \mathfrak{S}(\mathbf{\Psi} e^{ki\theta})_{k=0}^{\infty}. \tag{2}$$

In view of 4.3(b), it will suffice to show that $\mathbf{X}_1 = \mathbf{X}$, where $\mathbf{X}$ is the normalized innovation of the S.P. $(\mathbf{\Psi} e^{\;ki\theta})_{-\infty}^{\infty}$. Now

$$\mathbf{\Phi X} = \mathbf{\Psi} = \mathbf{\Phi X}_1, \tag{3}$$

By 3.6 $\mathbf{\Phi} = \mathbf{\Omega}\sqrt{\mathbf{G}}$, where $\mathbf{\Omega}$ is invertible a.e. Premultiplying (3) by $\mathbf{\Omega}^{-1}\mathbf{H}$, where $\mathbf{H}$ is as in (3.4), we get

$$\mathbf{JX} = \mathbf{JX}_1, \quad \text{i.e. } \mathbf{X} = \mathbf{JX}_1. \tag{4}$$

It follows that $\mathbf{J} = \mathbf{JJ}_1\mathbf{J}$, i.e. $\mathbf{J} \prec \mathbf{J}_1$. But in addition to this we have rank $\mathbf{J}_1 =$ rank $\mathbf{J}$; for from (2) and Theorem 3.8(b) $\mathbf{J}_1 = \mathbf{VJV}^*$, where $\mathbf{V}$ is unitary. Hence $\mathbf{J} = \mathbf{J}_1$. This reduces (4) to $\mathbf{X} = \mathbf{J}_1\mathbf{X}_1 = \mathbf{X}_1$. (Q.E.D.)

4.5. Definition. *Let* $\mathbf{0} \neq \mathbf{\Psi} \in \mathbf{L}_2^{0+}$. *Then the unique optimal and residual functions* $\mathbf{\Phi}, \mathbf{X}$ *given by* 4.4(a), 4.4(b) *will be called the* optimal *and* residual factors *of* $\mathbf{\Psi}$.

We see from the preceding results that these factors are precisely the generating function and the normalized innovation of the matricial process $(\mathbf{\Psi} e^{-ki\theta})_{-\infty}^{\infty}$.

We know from the classical theory that every complex-valued residual function can be factored into a constant $e^{i\alpha}$, α real, a Blaschke product and the radial limit of

$$\exp\left\{-\frac{1}{2\pi}\int_0^{2\pi}\frac{z+e^{i\theta}}{z-e^{i\theta}}d\mu(\theta)\right\}, \tag{1}$$

where μ is a monotone increasing, right continuous function on $[0, 2\pi]$ such that $\mu'(\theta) = 0$ a.e. In 1955 Potapov [16] gave a remarkable generalization of this result for functions $\mathbf{\Psi} \in \mathbf{L}_\infty^{0+}$ such that $\mathbf{\Psi\Psi}^* \prec \mathbf{I}$ and rank $\mathbf{\Psi} = q$ a.e., in which (1) is replaced by a multiplicative integral with respect to a monotone increasing, right continuous, hermitian matrix-valued function $\mathbf{E}$. In [12, 13] we have shown that in case $\mathbf{\Psi}$ is residual, the weighting $\mathbf{E}$ is jump-singular, i.e. $\mathbf{E}' = \mathbf{0}$, a.e. Thus when $\mathbf{\Psi}$ is in $\mathbf{L}_2^{0+}$ and has full rank q, its residual factor admits a further factorization into a constant unitary matrix, a matricial Blaschke product, and the radial limit of a multiplicative integral taken with respect to a purely jump-singular, monotone increasing hermitian weighting, in full analogy with the classical situation. For details see [12, 13], where, however, the constant unitary factor has been overlooked.

One of Beurling's objectives in [1] was to solve "Wiener's closure problem" for H_2, or equivalently L_2^{0+}, i.e. to find functions in L_2^{0+} the one-sided shifts of which span the whole of L_2^{0+}. The following result extends his solution to $\mathbf{L}_2^{0+}$:

4.6. COROLLARY (Closure Theorem for $\mathbf{L}_2^{0+}$). *Let* $0 \neq \mathbf{\Psi} \in \mathbf{L}_2^{0+}$. *Then*

$$\mathfrak{S}(\mathbf{\Psi} e^{ki\theta})_{k=0}^{\infty} = \mathbf{L}_2^{0+} \cdot \mathbf{K}, \tag{1}$$

where $\mathbf{K}$ *is a projection matrix,*(1) *if and only if* $\mathbf{\Psi} = \mathbf{\Phi V}_0$, *where* $\mathbf{\Phi}$ *is the optimal factor of* $\mathbf{\Psi}$, *and* $\mathbf{V}_0$ *is a unitary matrix such that* $\mathbf{V}_0\mathbf{K}\mathbf{V}_0^*$ *gives the projection onto the range of* $\mathbf{\Phi}_+(0)$.

Proof. Let $\mathbf{X} = \mathbf{JU}$ be the residual factor of $\mathbf{\Psi}$. Then from (1), 4.4(b) and 3.8(b)

$$\mathbf{X} = \mathbf{V}_0\mathbf{K}, \quad \mathbf{V}_0 = \text{unitary matrix}. \tag{2}$$

Hence

$$\mathbf{J} = \mathbf{XX}^* = \mathbf{V}_0\mathbf{K}\mathbf{V}_0^*. \tag{3}$$

Hence by 4.3(a), (2) and (3)

$$\mathbf{\Psi} = \mathbf{\Phi X} = \mathbf{\Phi V}_0\mathbf{K} = \mathbf{\Phi J V}_0 = \mathbf{\Phi V}_0. \tag{4}$$

Conversely, let (3) and (4) hold. Then $\mathbf{\Phi} = \mathbf{\Phi V}_0$. Premultiplying by $\mathbf{\Omega}^{-1}\mathbf{H}$, cf. 3.6, (3.4), we get $\mathbf{X} = \mathbf{JV}_0$, i.e. by (3) $\mathbf{X} = \mathbf{V}_0\mathbf{K}$. Since $\mathbf{V}_0$ is invertible, it follows that

$$\mathfrak{S}(\mathbf{X} e^{ki\theta})_{k=0}^{\infty} = \mathfrak{S}(\mathbf{K} e^{ki\theta})_{k=0}^{\infty} = \mathbf{L}_2^{0+} \cdot \mathbf{K}.$$

Since by 4.3(b), the L.H.S. $= \mathfrak{S}(\mathbf{\Psi} e^{ki\theta})_{k=0}^{\infty}$, we have (1). (Q.E.D.)

We also easily get matricial extensions of Beurling's Theorems I, III in [1]:

4.7. COROLLARY. *Let* $\mathbf{\Psi}_1, \mathbf{\Psi}_2 \in \mathbf{L}_2^{0+}$. *Then*

(a) $\mathbf{\Psi}_2 \in \mathfrak{S}(\mathbf{\Psi}_1 e^{ki\theta})_{k=0}^{\infty}$, *if and only if the residual factor* $\mathbf{X}_1$ *of* $\mathbf{\Psi}_1$ *is a right divisor of the residual factor* $\mathbf{X}_2$ *of* $\mathbf{\Psi}_2$, *i.e.*

$$\mathbf{X}_2 = \mathbf{\Psi X}_1, \quad \mathbf{\Psi} \in \mathbf{L}_2^{0+}.$$

(1) In particular, we can take $\mathbf{K} = \mathbf{I}$.

(b) clos. $\{\mathfrak{S}(\mathbf{\Psi}_1 e^{ki\theta})_{k=0}^{\infty} + \mathfrak{S}(\mathbf{\Psi}_2 e^{ki\theta})_{k=0}^{\infty}\} = \mathfrak{S}(\mathbf{X} e^{ki\theta})_{k=0}^{\infty}$, *where* $\mathbf{X}$ *is a greatest common residual right divisor of* $\mathbf{X}_1, \mathbf{X}_2$.

Proof. (a) This follows from the results, cf. 4.4(b), 4.5,

$$\mathfrak{S}(\mathbf{\Psi}_j e^{ki\theta})_{k=0}^{\infty} = \mathfrak{S}(\mathbf{X}_j e^{ki\theta})_{k=0}^{\infty} = \mathbf{L}_2^{0+}\mathbf{X}_j \quad (j=1,2),$$

$$\mathbf{\Psi}_2 \in \mathfrak{S}(\mathbf{\Psi}_1 e^{ki\theta})_{k=0}^{\infty} \quad \textit{if and only if} \quad \mathfrak{S}(\mathbf{\Psi}_2 e^{ki\theta})_{k=0}^{\infty} \subseteq \mathfrak{S}(\mathbf{\Psi}_1 e^{ki\theta})_{k=0}^{\infty},$$

the corresponding equivalence for $\mathbf{X}$.

(b) Let $\mathfrak{M}$ be the subspace on L.H.S. (b). Then $e^{i\theta}\mathfrak{M} \subseteq \mathfrak{M}$, and hence by Theorem 3.8(a), there exists a residual function $\mathbf{X}$ such that

$$\mathfrak{M} = \mathbf{L}_2^{0+}\mathbf{X}. \tag{1}$$

Since $\mathbf{\Psi}_1, \mathbf{\Psi}_2 \in \mathfrak{M}$, therefore by (a)

$$\mathbf{X} \textit{ is a right divisor of } \mathbf{X}_1 \textit{ and } \mathbf{X}_2. \tag{2}$$

Next, let $\mathbf{X}_3$ be any residual right divisor of both $\mathbf{X}_1, \mathbf{X}_2$. Then by (a)

$$\mathbf{\Psi}_1, \mathbf{\Psi}_2 \in \mathfrak{S}(\mathbf{X}_3 e^{ki\theta})_{k=0}^{\infty},$$

and so $\mathfrak{M} \subseteq \mathfrak{S}(\mathbf{X}_3 e^{ki\theta})_{k=0}^{\infty}$. Hence again by (a) and (1), $\mathbf{X}_3$ is a right divisor of $\mathbf{X}$. It follows from (2) that $\mathbf{X}$ is a greatest common, residual, right divisor $\mathbf{X}_1, \mathbf{X}_2$. (Q.E.D.)

References

[1]. Beurling, A., On two problems concerning linear transformations in Hilbert space. *Acta Math.*, 81 (1949), 239–255.

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

[3]. Halmos, P., Shifts of Hilbert spaces. *Crelle's Journal*, 208 (1961), 102–112.

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

[5]. Hoffman, K., *H^p Spaces*. Mass. Inst. of Technology, Cambridge, Mass., 1961 (mimeographed).

[6]. Kolmogorov, A. N., Stationary sequences in Hilbert space (Russian). *Bull. Math. Univ. Moscou*, 2, No. 6 (1951), 40 pp. (English translation by Natasha Artin.)

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

[8]. Masani, P., *Cramer's Theorem on monotone matrix-valued functions and the Wold decomposition, Probability and Statistics, the Harald Cramer Volume*, Ed. Ulf Grenander, Stockholm, 1959, pp. 175–189.

[9]. ——, Sur la fonction génératrice d'un processus stochastique vectoriel. *C. R. Acad. Sci. Paris*, 249 (1959), 360–362.

[10]. ——, Isomorphie entre les domaines temporel et spectral d'un processus vectoriel, régulier. *C. R. Acad. Sci. Paris*, 249 (1959) 496–498.

[11]. MASANI, P., Sur les fonctions matricielles de la classe de Hardy H_2. *C. R. Acad. Sci. Paris*, 249 (1959), 873–875.

[12]. ——, Sur les fonctions matricielles de la classe de Hardy H_2. *C. R. Acad. Sci. Paris*, 249 (1959), 906–907.

[13]. ——, Une généralisation pour les fonctions matricielles de la classe de Hardy H_2 d'un théorème de Nevanlinna, *C. R. Acad. Sci. Paris*, 251 (1960), 318–320.

[14]. ——, The prediction theory of multivariate stochastic processes, III. *Acta Math.*, 104 (1960), 141–162.

[15]. MASANI, P. & WIENER, N., *On bivariate stationary processes and the factorization of matrix-valued functions. Theory of Probability and its Applications*, IV (1959), 300–308.

[16]. POTAPOV, V. P., The multiplicative structure of J-contractive matrix functions. *Trudi Mosk. Matem. Obsch*, 4 (1955), 125–236. (In Russian.) (*Amer Math. Soc. Transl.*, 15 (1960), 131–243.)

[17]. ROZANOV, YU. A., The spectral theory of multi-dimensional stationary stochastic processes with discrete time. *Uspekhi Mat. Nauk*, XXIII, (1957), 989–992. (In Russian.)

[18]. WIENER, N. & MASANI, P., The prediction theory of multivariate stochastic processes, Part I. *Acta Math.*, 98 (1957), 111–150, Part II, *Acta Math.*, 99 (1958) 93–137.

Received October, 11, 1961