PREDICTION THEORY AND FOURIER SERIES IN SEVERAL VARIABLES. II

BY

HENRY HELSON

Berkeley, California

AND

DAVID LOWDENSLAGER

Princeton, New Jersey (1)

1. Introduction

This second paper on Prediction Theory, like our first one [10], is divided into two parts: the first, consisting of the first eight sections, treats complex-valued functions defined on rather general groups, and the second part deals with matrix-valued functions defined on the unit circle. In both parts we are concerned with degeneracies which were excluded by our hypotheses before, but which turn out to be interesting from both the function-theoretic and the prediction-theoretic points of view.

Unlike the first paper, this one has to do with difficulties which do not exist at all for the classical case of scalar functions defined on the circle group. Both parts of the paper leave interesting problems unsolved. In this introduction we shall try to present the questions of this paper and our contribution to their solution in broad terms.

Let $f = f(e^{ix})$ be a summable function defined on the circle whose Fourier series has the form

$$f(e^{ix}) \sim \sum_{n \geq 0} a_n e^{nix}.$$

It is important and well-known that

$$\int \log |f| \, d\sigma \text{ (2)}$$

(1) The authors acknowledge the support of the National Science Foundation, and in the case of the second-named author, of the United States Air Force.

(2) $d\sigma$ always denotes normalized invariant measure on the compact group being considered. Here it is $dx/2\pi$ on $(0, 2\pi)$.

is finite, unless f is the null-function. In particular, f cannot vanish on a set of positive measure.

In our first paper we obtained a generalization of this theorem to the class of compact abelian groups whose duals are linearly ordered, but under the hypothesis that a_0, the mean value of f, is different from 0. For the circle group this is no restriction at all, because we can consider $e^{-pix}f$ (where p is a positive integer chosen appropriately) in place of f, but this device fails if the ordered group which generalizes the integer group has no least positive element. Indeed, trivial examples show that if $a_0 = 0$ the function may vanish on a non-empty open set without vanishing identically.

Under the hypothesis that *the order relation is archimedean* the counter-examples just referred to are defeated, and Arens [3] has proved that the finiteness of the integral follows from the *continuity* of f. We shall prove in the same direction that f cannot vanish on a set of positive measure unless it vanishes identically, without assuming that $a_0 \neq 0$ or that f is continuous. It may be unexpected therefore that even *bounded* functions exist for which the logarithmic integral diverges. (This negative result is a corollary of a prediction theorem of different character.) We have additional information, but no definitive result, on the question which non-negative functions are the modulus of some function f of the class considered.

In the second part of the paper we reconsider the problems of [10] for positive semi-definite matrix functions $W = W(e^{ix})$ of less than full rank. In prediction-theoretic terms, we study a process of less than full rank whose covariance matrix is absolutely continuous. For a process with covariance matrix W (of full rank or not) these conditions are known to be equivalent: the process has no remote past; the process is a moving average; W has the form AA^*, where A is an analytic matrix function. For processes of full rank, these properties are equivalent to this analytic condition on W:

$$\int \log \det W \, d\sigma > -\infty .$$

But this integral always diverges if W is singular, for example if W is AA^* for a singular analytic function A; in this form the integral is thus too crude to give information about processes of less than full rank.

At almost every point $W(e^{ix})$ is a positive semi-definite matrix which operates as a non-singular transformation on its range $\mathfrak{H}(e^{ix})$. Denote the determinant of this transformation by $\Delta W(e^{ix})$. If we replace the determinant function in the integral by Δ, then the finiteness of the integral is the first condition for the process to be a moving average.

There is an obvious second necessary condition. If $W = AA^*$ for some analytic function A, then $\mathfrak{H}(e^{ix})$ coincides at almost every point with the range of $A(e^{ix})$. In order for W to have this form it is evidently necessary for $\mathfrak{H}(e^{ix})$ to be the range of *some* analytic function.

We prove that these conditions together are sufficient as well as necessary for the process to be a moving average, or for W to have the form AA^* for some analytic function A. Of course in application it may be difficult to decide whether the range of W coincides with the range of an analytic function, and the criterion found by Masani and Wiener [17] for W to be a square when it is of order two therefore has independent interest.

The theorem on the factoring of W depends on structure theorems for analytic matrix functions which are proved first. We follow Lax [14] in taking Beurling's notions of inner function and outer function as fundamental, but we offer our own definition for matrix functions which leads to a new description of singular analytic functions.

In connection with the first part of the paper, we draw attention to the related work of Arens, Hoffman, and Singer [2, 3, 4, 5, 12]. At a crucial stage of our research we had the good fortune to talk at length with Professor P. Malliavin, and we are grateful for his permission to incorporate his ideas in this paper. As far as possible we shall identify his contributions in context.

As with our first paper, this one overlaps the work of Masani and Wiener [16] to some extent. Following several notes, a third paper has been published recently by Masani [15]. These authors refer in their various papers to Russian work of which we have taken no account.

2. Mise en scène

In this section we lay down notation and recount known results which will be used throughout the first part of the paper.

R denotes the real line, and R_d the same group in the discrete topology. The elements of R_d will always be called λ or τ. The character group of R_d is a compact abelian group B also obtained as the Bohr compactification of R, with elements $x, y, \ldots$, and Haar measure $d\sigma$ (normalized to have unit total mass). For each λ in R_d let χ_λ be the character of B defined by

$$\chi_\lambda(x) = x(\lambda) \quad (\text{all } x \in B). \tag{1}$$

The correspondence of λ to χ_λ is an isomorphism of R_d with the character group of B, and we have for example

$$\chi_0(x) \equiv 1, \quad \chi_\lambda(x) \cdot \chi_\tau(x) \equiv \chi_{\lambda+\tau}(x). \tag{2}$$

We shall nevertheless preserve the notational distinction between elements of R_d and characters on B, for convenience rather than for purity.

The Fourier series of a function f which is defined and summable on B has the form

$$f(x) \sim \sum_{\lambda} a(\lambda)\, \chi_\lambda(x), \tag{3}$$

with coefficients defined by

$$a(\lambda) = \hat{f}(\lambda) = \int \overline{\chi_\lambda(x)}\, f(x)\, d\sigma(x). \tag{4}$$

The Fourier-Stieltjes series of a finite complex measure μ is given by similar formulas.

It is known that every discrete abelian group with an archimedean order relation is isomorphic (with preservation of order) to a subgroup of R_d. All the results of the first part of the paper could be stated and proved for arbitrary archimedean-ordered discrete groups and their duals, which means for subgroups of R_d and the dual quotient groups of B. Indeed the archimedean hypothesis is not used until section five, so that our first theorems are true in the more general conditions of [10]. We prefer, however, to write explicitly about R_d and B in order to avoid notational difficulty.

Among the characters of R_d are some having the form

$$e_u(\lambda) = e^{iu\lambda} \tag{5}$$

for some real number u. The mapping from u to e_u carries real numbers into B. It is known that this mapping is a one-one continuous isomorphism of R onto a dense subgroup B_0 of B. Sometimes it is convenient to identify u with e_u and think of R as a subset of B. We shall be concerned later with the ergodic properties of B_0 in B.

As generalizations of the Hardy spaces of analytic functions defined in the circle, we consider the spaces H^p $(1 \leqslant p \leqslant \infty)$ of functions defined on B, belonging to L^p, whose Fourier coefficients (4) vanish for all $\lambda < 0$. When the value of p is unimportant we may call such a function *analytic*. In particular an *analytic trigonometric polynomial* is a finite sum of the form

$$\sum_{\lambda \geqslant 0} a(\lambda)\, \chi_\lambda(x). \tag{6}$$

For each p, H_0^p is the subspace of H^p consisting of those functions whose mean value $a(0)$ vanishes.

A function f in some H^p is called *outer* [6] if

$$\int \log|f|\, d\sigma = \log\left|\int f\, d\sigma\right| > -\infty. \tag{7}$$

Beurling proved the fundamental result that *f in H^2 is outer if and only if the set of functions Pf, where P ranges over all analytic trigonometric polynomials, is dense in H^2*. (Beurling's Theorem, originally proved for the circle group, was extended in [10] to compact groups with ordered duals.)

From (7) it follows that every function w of the form $|f|^2$ for some outer function f in H^2 must satisfy

$$\int \log w \, d\sigma > -\infty. \tag{8}$$

Conversely, *every non-negative summable function w satisfying* (8) *is such a square, and moreover f is unique up to a constant factor of modulus* 1. This theorem was also proved (not quite explicitly) in [10]; the unicity of f will be reconsidered in section seven, where the notion of outer function is further developed.

If f is in H^1 and *has mean value different from* 0, then [10] there is an outer function g such that $|f| = |g|$ almost everywhere. If we write $f = g \cdot h$, then the function h is analytic also. An analytic function with modulus 1 almost everywhere is called an *inner function* by Beurling.

This notation and these results will be used in the first part of the paper without further reference.

3. The Wold decomposition

Let μ be a non-negative finite measure defined on the Borel subsets of B. We form the Hilbert space L^2_μ with inner product

$$(f, g) = \int f \bar{g} \, d\mu. \tag{9}$$

The functions χ_λ $(-\infty < \lambda < \infty)$ belong to L^2_μ and form a complete set. Moreover they form a *stationary stochastic process* in the sense that the inner products

$$(\chi_\lambda, \chi_\tau) = \int \chi_\lambda \bar{\chi}_\tau \, d\mu = \varrho(\tau - \lambda) \tag{10}$$

depend only on the difference $\tau - \lambda$. We know that any stationary process depending on a real parameter (without any continuity hypothesis) is isomorphic, with respect to all Hilbert space properties, to the process of characters in L^2_μ for a suitable measure μ.

For each λ in R_d we define a unitary operator S_λ in L^2_μ by setting

$$S_\lambda f = \chi_\lambda \cdot f. \tag{11}$$

A closed subspace $\mathfrak{M}$ of L^2_μ is called *invariant* if it is carried onto itself by each S_λ.

LEMMA 1. *Let $\mathfrak{M}$ be a closed invariant subspace of L^2_μ. The orthogonal projection* $\mathbf{P}$ *on* $\mathfrak{M}$ *has the form*

$$\mathbf{P}\, f = e \cdot f \tag{12}$$

for a function e in L^2_μ taking the values 0 *and* 1. *Moreover* $e = \mathbf{P}\,\chi_0$.

The proof, which offers no difficulty, is omitted.

For each real τ, let $\mathfrak{M}_\tau$ be the smallest closed subspace of L^2_μ containing the functions χ_λ with $\lambda > \tau$. The subspaces $\mathfrak{M}_\tau$ are nested, and decrease as τ increases; we set

$$\mathfrak{H}_3 = \bigcap_{\tau < \infty} \mathfrak{M}_\tau. \tag{13}$$

Then $\mathfrak{H}_3$, which is clearly closed and invariant, is called the *remote past* of the process $\{\chi_\lambda\}$. Denote the corresponding projection function of the lemma by e_3.

If χ_0 does not belong to $\mathfrak{M}_0$, let y_0 be the part of χ_0 orthogonal to $\mathfrak{M}_0$, so that

$$\chi_0 = y_0 + z_0 \quad (y_0 \perp \mathfrak{M}_0,\ z_0 \in \mathfrak{M}_0). \tag{14}$$

It is easy to see that

$$y_\tau = S_\tau y_0 \tag{15}$$

is the part of χ_τ orthogonal to $\mathfrak{M}_\tau$. Evidently the y_τ form an orthogonal set in L^2_μ, and their linear combinations span a closed subspace $\mathfrak{H}_1$ which is invariant and orthogonal to $\mathfrak{H}_3$. Let e_1 be the function realizing the orthogonal projection on $\mathfrak{H}_1$. (If $\mathfrak{M}_0$ contains χ_0, then $\mathfrak{H}_1$ is defined to be $\{0\}$ and e_1 the null function.)

The linear sum $\mathfrak{H}_1 \oplus \mathfrak{H}_3$ may not be all of L^2_μ; its orthogonal complement is a third closed invariant subspace $\mathfrak{H}_2$ with projecting function e_2. Then by definition L^2_μ is the orthogonal sum of $\mathfrak{H}_1$, $\mathfrak{H}_2$, and $\mathfrak{H}_3$, and

$$e_1 + e_2 + e_3 = 1, \quad e_j e_k = 0 \text{ a.e.} \quad (j \neq k). \tag{16}$$

If only one summand $\mathfrak{H}_j$ is different from $\{0\}$, the process is said to be *pure*, and of type 1, 2, or 3 depending on whether $\mathfrak{H}_1$, $\mathfrak{H}_2$, or $\mathfrak{H}_3$ is non-trivial. More descriptive names have been invented: a process purely of type 1 is an *innovation process*, with y_τ its innovation at time τ; one of type 3 is *deterministic*; and we suggest the adjective *evanescent* for a process of type 2.

Set
$$d\mu_j = e_j\, d\mu \quad (j = 1, 2, 3); \tag{17}$$

these measures are mutually singular and their sum is μ. We have obviously

$$(e_j \chi_\lambda, e_j \chi_\tau) = \int \chi_\lambda \bar{\chi}_\tau e_j\, d\mu = \int \chi_\lambda \bar{\chi}_\tau\, d\mu_j, \tag{18}$$

from which we see that $\{e_j \chi_\lambda\}$ is a stationary stochastic process in L^2_μ, isomorphic with the canonical process $\{\chi_\lambda\}$ in $L^2_{\mu j}$ for $j = 1, 2, 3$. Each process $\{e_j \chi_\lambda\}$ lies in $\mathfrak{H}_j$. These component processes have the following fundamental property.

THEOREM 1. *The process* $\{e_j \chi_\lambda\}$ *in* L^2_μ *is purely of type* j, *for* $j = 1, 2, 3$.

This theorem is a generalization of the well-known result of Wold [21] for the circle group (where only processes of types 1 and 3 occur). A generalization of Wold's Theorem, for processes depending *continuously* on a real parameter, was given by Hanner [8], and his method of proof will be adapted to prove our theorem, but in reality his result is quite different from ours and lies deeper.

The theorem contains three statements, which we prove in order.

Statement 1: $j = 1$.—The functions y_λ are an orthogonal set fundamental in $\mathfrak{H}_1$, and so we have the decomposition

$$\chi_0 = \sum a(\lambda) y_\lambda + w \quad (w \perp \mathfrak{H}_1). \tag{19}$$

Unless $\mathfrak{H}_1$ is trivial, it is clear that $a(\lambda) = 0$ for $\lambda < 0$, but $a(0) \neq 0$. First projecting (19) into $\mathfrak{H}_1$ and then applying S_τ gives

$$e_1 \chi_\tau = \sum_{\lambda \geqslant 0} a(\lambda) y_{\lambda+\tau}. \tag{20}$$

This shows that the linear subspace spanned by $\{e_1 \chi_\lambda\}_{\lambda \geqslant 0}$ is contained in the subspace spanned by $\{y_\lambda\}_{\lambda \geqslant 0}$.

The converse holds also. For y_0 can be approximated by linear combinations of χ_λ $(\lambda \geqslant 0)$, and since $y_0 \in \mathfrak{H}_1$, by the same linear combinations of $e_1 \chi_\lambda$ $(\lambda \geqslant 0)$. The same argument applies to y_τ if $\tau > 0$.

Combining these results, and translating by τ, we see that $\{e_1 \chi_\lambda\}_{\lambda \geqslant \tau}$ and $\{y_\lambda\}_{\lambda \geqslant \tau}$ span the same subspace of L^2_μ, and indeed of $\mathfrak{H}_1$, for each τ. The y_λ are orthogonal and so form an innovation process with innovation y_τ at time τ; therefore $\{e_1 \chi_\lambda\}$ is an innovation process with the same innovation y_τ. This proves the first part of the theorem.

Statement 2: $j = 2$.—We have to prove that the process $\{e_2 \chi_\lambda\}$ has neither innovation nor remote past. As to innovation, we can form finite sums

$$\chi_0 + \sum_{\lambda > 0} b(\lambda) \chi_\lambda \tag{21}$$

approximating y_0, which belongs to $\mathfrak{H}_1$, and so the corresponding sums

$$e_2 (\chi_0 + \sum_{\lambda > 0} b(\lambda) \chi_\lambda) \tag{22}$$

approximate $e_2 y_0 = 0$. In other words, $e_2 \chi_0$ is the limit of sums of the form

$$-\sum_{\lambda > 0} b(\lambda) e_2 \chi_\lambda. \tag{23}$$

This is exactly the statement that $\{e_2\chi_\lambda\}$ has no innovation.

To consider the remote past we need a result which will be referred to again, and which is therefore stated formally.

LEMMA 2. *$e_j\mathfrak{M}_0$ is a closed subspace of $\mathfrak{M}_0$ for $j=1, 2, 3$.*

Let f be an analytic trigonometric polynomial with mean value zero. Writing f as the sum of its projections and using (20) we have

$$f=\sum_{\lambda>0} a(\lambda)\, y_\lambda + w + z \quad (w\in\mathfrak{H}_2, z\in\mathfrak{H}_3). \tag{24}$$

Since y_λ belongs to $\mathfrak{M}_0$ for $\lambda>0$, the first sum, equal to $e_1 f$, lies in $\mathfrak{M}_0$. Now $z=e_3 f$ is in $\mathfrak{H}_3$ which is entirely contained in $\mathfrak{M}_0$. Hence the remaining term $w=e_2 f$ is in $\mathfrak{M}_0$ as well. This set of functions f is dense in $\mathfrak{M}_0$, and so $e_j\mathfrak{M}_0$ is contained in $\mathfrak{M}_0$ for $j=1, 2, 3$.

Now in the decomposition

$$\mathfrak{M}_0 = e_1\mathfrak{M}_0 \oplus e_2\mathfrak{M}_0 \oplus e_3\mathfrak{M}_0 \tag{25}$$

the summands are mutually orthogonal and contained in $\mathfrak{M}_0$; it follows that each one is closed, as we had to show.

We return to the remote past of $\{e_2\chi_\lambda\}$. Translating (25),

$$\mathfrak{M}_\tau = e_1\mathfrak{M}_\tau \oplus e_2\mathfrak{M}_\tau \oplus e_3\mathfrak{M}_\tau. \tag{26}$$

We have to show that the projection of $e_2\chi_0$ on $e_2\mathfrak{M}_\tau$ has norm as small as we please, if τ is large enough. From (26), this projection is the same as its projection on $\mathfrak{M}_\tau$ itself. But the intersection of all $\mathfrak{M}_\tau$ is $\mathfrak{H}_3$, to which $e_2\chi_0$ is orthogonal, and it is a simple exercise in geometry to verify then that $e_2\chi_0$ has small projection in $\mathfrak{M}_\tau$ for large τ. This completes the proof for $j=2$.

Statement 3: $j=3$.—We are to show that $\{e_3\chi_\lambda\}$ is deterministic. It suffices to prove that linear combinations of $e_3\chi_\lambda$ $(\lambda>0)$ are dense in the manifold spanned by $e_3\chi_\lambda$ (all λ), since by translation the same will hold for the span of $e_3\chi_\lambda$ $(\lambda>\tau)$, no matter how large τ. In other words, if

$$\int f(e_3\bar{\chi}_\lambda)\,d\mu = 0 \tag{27}$$

for all $\lambda>0$, then the same relation is to hold for all λ. But (27) can be written

$$\int (e_3 f)\bar{\chi}_\lambda\, d\mu = 0, \tag{28}$$

and this for $\lambda>0$ means that $e_3 f$ is orthogonal to $\mathfrak{M}_0$ in L^2_μ. Then $e_3 f$ is orthogonal to $\mathfrak{H}_3$ contained in $\mathfrak{M}_0$; but $e_3 f$ itself belongs to $\mathfrak{H}_3$, and therefore must be zero. Hence (28) holds for all λ.

This completes the proof of Theorem 1.

We mention here a problem about processes of type 2, whose solution would contribute a good deal to our knowledge about function theory on B. Let μ be absolutely continuous with respect to σ, so that $d\mu = w\,d\sigma$ for some summable function w. Assume the canonical process in L^2_w is evanescent. We know from definition that

$$\bigcup_{\lambda>0} \mathfrak{M}_\lambda \tag{29}$$

is dense in $\mathfrak{M}_0$, but it is conceivable that the subspace

$$\mathfrak{N}_0 = \bigcap_{\lambda<0} \mathfrak{M}_\lambda \tag{30}$$

is larger than $\mathfrak{M}_0$. If $\mathfrak{N}_0$ is larger than $\mathfrak{M}_0$, let $\mathfrak{R}_0$ be the complement of $\mathfrak{M}_0$ in $\mathfrak{N}_0$. The same definition at τ gives a subspace $\mathfrak{R}_\tau$ at each τ, and it is easy to see that $\mathfrak{R}_\tau = S_\tau \mathfrak{R}_0$. These subspaces are mutually orthogonal. We can prove, using theorems presented later in this paper, that each subspace $\mathfrak{R}_\tau$ is one-dimensional, and that together they span L^2_μ. If f is a representative of $\mathfrak{R}_0$, then $\chi_\tau f$ is in $\mathfrak{R}_\tau$, and the set $\{\chi_\tau f\}$ is a complete orthonormal system in L^2_μ. Furthermore $\mathfrak{M}_0$ is exactly the closed linear span of the elements $\chi_\lambda f$ $(\lambda > 0)$. So in a new sense the canonical process in L^2_w is an innovation process with innovation f at time 0.

The problem is to decide whether $\mathfrak{N}_0$ is sometimes or always larger than $\mathfrak{M}_0$, and what properties of w decide between the two cases if they can both occur. If $\mathfrak{N}_0$ is larger than $\mathfrak{M}_0$, then w has the form $|g|^2$, where g belongs to H^2 and is *outer* in a generalized sense, although the logarithmic integral of (7) evidently diverges. As we shall show, there are functions g in H^2 for which that integral diverges; for $w = |g|^2$, the canonical process in L^2_w is clearly evanescent, but we do not know whether it can be or must be an innovation process in the new sense.

In the case of a general measure μ, the evanescent part of the Wold decomposition divides further into a part which is continuous from the left as well as from the right, in the sense that (29) is dense in (30), and a part which is an innovation process in the new sense. We shall not be interested in this refinement.

4. Analysis of the decomposition

The decomposition theorem itself does not tell how to find the component measures μ_1, μ_2, μ_3, or equivalently the subspaces $\mathfrak{H}_1$, $\mathfrak{H}_2$, $\mathfrak{H}_3$ from knowledge of μ. The purpose of this section is to set down what we can from known results or standard arguments.

In the case of the circle group (with its dual, the integers) only $\mathfrak{H}_1$ and $\mathfrak{H}_3$ appear ([7], Chapter XII). Write μ (now a measure on the circle) as a sum

$$d\mu = w\,d\sigma + d\mu_s \tag{31}$$

of absolutely continuous and singular parts. The main result is this: *if* $\int \log w\,d\sigma > -\infty$, *then* $d\mu_1 = w\,d\sigma$ *and* $d\mu_3 = d\mu_s$; *otherwise* $d\mu_1 = 0$ *and* $d\mu_3 = d\mu$.

We shall prove a natural generalization of this theorem to the group B; but whereas the result is complete on the circle, it cannot differentiate between μ_2 and μ_3 on B.

THEOREM 2. *If* $d\mu = w\,d\sigma + d\mu_s$, *where* w *is a non-negative summable function on* B *and* μ_s *is singular with respect to* σ, *and if*

$$\int \log w\,d\sigma > -\infty, \tag{32}$$

then $d\mu_1 = w\,d\sigma$. *If* (32) *is false,* $d\mu_1 = 0$.

Proof. The part of χ_0 orthogonal to $\mathfrak{M}_0$ was called $1 + H$ in [10], and was shown not to be the null function under the hypothesis (32). Indeed $1 + H$ was different from zero almost everywhere for σ, but vanished for μ_s. Therefore e_1 is the function equal to 1 almost everywhere for σ and zero almost everywhere for μ_s. This proves that $d\mu_1 = w\,d\sigma$.

If (32) fails, then χ_0 can be approximated by linear combinations of χ_λ with $\lambda > 0$, and the process has no innovation. Therefore $d\mu_1 = 0$.

By Theorem 2, if μ has absolutely continuous component satisfying (32) then this part of μ is exactly the summand μ_1 of the Wold decomposition. Moreover $\mathfrak{H}_1$ is naturally identified with the Hilbert space L^2_w, and the process $\{e_1\chi_\lambda\}$ in L^2_μ is isomorphic with the canonical process $\{\chi_\lambda\}$ in L^2_w. The orthogonality and invariance of the spaces $\mathfrak{H}_j$ can be expressed informally by saying that second-order questions about the prediction of the canonical process in L^2_μ decompose under the Wold decomposition into the analogous questions for the orthogonal subprocesses; in the case we are discussing, this means that the absolutely continuous and the singular parts of μ can be treated separately, provided that (32) is true.

The next theorem shows that this simplification does not depend on (32), and that it suffices to study $w\,d\sigma$ and $d\mu_s$ separately even when (32) fails.

THEOREM 3. *Let* e *and* e' *be functions in* L^2_μ *satisfying*

$$e = 1 \text{ a.e. } (d\sigma); \quad e' = 1 \text{ a.e. } (d\mu_s); \quad e \cdot e' \equiv 0. \tag{33}$$

Then $e\mathfrak{M}_\tau$ *and* $e'\mathfrak{M}_\tau$ *are closed subspaces of* $\mathfrak{M}_\tau$ *for each* τ.

Proof. For simplicity consider $\tau = 0$. Suppose that f belongs to L^2_μ and is orthogonal to $\mathfrak{M}_0$:

$$\int \chi_\lambda \bar{f}\, d\mu = 0 \quad (\text{all } \lambda > 0). \tag{34}$$

This expresses the fact that $\bar{f}\,d\mu$ has Fourier-Stieltjes series of analytic type. By Theorem 16 of [10], the same is true separately of $\bar{f}w\,d\mu$ and $\bar{f}\,d\mu_s$, so that ef and $e'f$ are separately orthogonal to $\mathfrak{M}_0$. This is the same as to say that f is orthogonal to $e\mathfrak{M}_0$ and to $e'\mathfrak{M}_0$. Since f was an arbitrary function orthogonal to $\mathfrak{M}_0$, it follows that $e\mathfrak{M}_0$ and $e'\mathfrak{M}_0$ are contained in $\mathfrak{M}_0$.

In the decomposition

$$\mathfrak{M}_0 = e\mathfrak{M}_0 \oplus e'\mathfrak{M}_0, \tag{35}$$

which is analogous to (25) in the proof of the Wold decomposition, the summands are mutually orthogonal and contained in $\mathfrak{M}_0$, and it is obvious that each must be closed.

The archimedean property of R_d has not been used to this point, but it is involved in the discussion which follows and in succeeding sections.

Suppose now that the process $\{\chi_\lambda\}$ in L^2_μ depends continuously on λ, or what is equivalent, merely that the inner product (χ_0, χ_λ) is a continuous function of λ in the ordinary sense. This is the kind of process treated by Hanner [8], Karhunen [13], Wiener [19], and others, and complete results have been obtained for the questions we are considering [7]. In order that the process be of this type it is necessary and sufficient (as one proves without difficulty) that μ be supported by B_0, so that in particular μ is singular with respect to σ. If σ_1 denotes linear measure on B_0, then μ has a decomposition

$$d\mu = w\,d\sigma_1 + d\mu_s, \tag{36}$$

where now w is summable for $d\sigma_1$ and $d\mu_s$ is singular with respect to σ_1 but carried on B_0. *The process is deterministic if and only if*

$$\int_{-\infty}^{\infty} \frac{\log w(x)}{1+x^2}\, d\sigma_1(x) = -\infty. \tag{37}$$

If the integral in (37) *is finite, the remote past is the subspace of* L^2_μ *consisting of functions which vanish almost everywhere for* $d\sigma_1$.

This result can be proved from the corresponding theorem about discrete processes by a change of variable ([1], p. 263), or it can be obtained with a certain complication by the methods of [10]. The analogy with discrete processes is pressed further and more deeply by Hanner. But we shall not be concerned with processes of this type, whose innovation components always vanish in the Wold decomposition. Instead we consider the other extreme type: processes with absolutely continuous spectral measures, which cannot possibly be continuous.

5. Cauchy measures

A certain family of measures on B is intimately connected with the problems we are studying, as Malliavin pointed out to us, and this section sets down their definition and relevant properties.

For each r $(0 < r < 1)$ consider the measure μ_r on B whose Fourier-Stieltjes transform is

$$\hat{\mu}_r(\lambda) = r^{|\lambda|}. \tag{38}$$

The function on the right side is positive definite on R_d, so μ_r exists on B and is non-negative, with total mass one. But the transform is continuous in the ordinary topology of the line, and one proves easily that this is the case if and only if the measure is carried on B_0. Therefore each μ_r is carried on B_0, and hence is singular with respect to σ. These properties are also easy to verify:

$$\mu_r * \mu_s = \mu_{rs}; \quad \lim_{r\to 0} \mu_r = \sigma; \quad \lim_{r\to 1} \mu_r = \delta, \tag{39}$$

where in the last relations the limit refers to the weak star-topology of measures, and δ is the unit mass at the identity of B.

The explicit form of μ_r is easy to give, although it is not necessary to our work; $d\mu_r$ is the measure

$$\frac{y\,dt}{\pi(t^2+y^2)}, \quad (r = e^{-y}), \tag{40}$$

where t is the linear coordinate on B_0. This is the classical Cauchy kernel, and therefore we call the μ_r *Cauchy measures.*

Every segment of B_0 carries part of the mass of μ_r, and since B_0 is dense in B, it follows that μ_r assigns positive measure to every non-empty open set in B. Much more than this simple statement is true, and the next three lemmas give further essential information.

LEMMA 3. *If g is the characteristic function of a set of positive measure in B, then $\mu_r * g > 0$ almost everywhere for σ.*

This result is surely not new, but we have found it with Malliavin. If it were false, then for a certain value of r and a certain function g we should have $\mu_r * g = 0$ on a set of positive measure, whose characteristic function we call h. Let $h'(x) = h(-x)$; then

$$0 = \int h(x)\mu_r * g(x)\,d\sigma(x) = \mu_r * g * h'(0). \tag{41}$$

Now $g*h'$ is a continuous non-negative function, and we know that μ_r has positive mass in every non-empty open set. Therefore (41) is only possible if $g*h'$ vanishes identically. But

$$\int g*h'\,d\sigma = \int g\,d\sigma \cdot \int h'\,d\sigma \neq 0. \tag{42}$$

This contradiction proves that $\mu_r * g$ is positive almost everywhere.

LEMMA 4. *There exists a set F having positive measure in B with characteristic function g such that $\mu_r * g$ is not essentially bounded from zero.*

Since B_0 has measure zero, there is an open set G of small measure containing B_0. Moreover a translate G_x of G by the element x of B still contains virtually all the mass of μ_r provided x is close to the identity, because nearly all the mass of μ_r is carried on a compact segment of B_0 which remains in G under small translations. Therefore the characteristic function g of F, the negative of the complement of G, has the required property.

LEMMA 5. *There is a non-negative function u defined on B such that*

$$\int u\,d\sigma = \infty; \quad \mu_r * u < \infty \text{ a.e.} \quad (0 < r < 1). \tag{43}$$

Let t, r, and s satisfy the relations

$$0 < t < r < 1; \quad t = rs. \tag{44}$$

We require the set F and the function g of Lemma 4, with the observation that they were constructed independent of r. Since $\mu_t * g$ is not bounded from zero we can find a non-negative function u which is not summable but which satisfies

$$\int \mu_t * g(x)\,u(-x)\,d\sigma(x) = \mu_t * g * u(0) < \infty. \tag{45}$$

Using the associativity of the convolution operation a second time, this is equivalent to

$$\int_F \mu_t * u(-x)\,d\sigma(x) < \infty, \tag{46}$$

so that $\mu_t * u(-x)$ is finite almost everywhere on F. We conclude that $\mu_r * u$ is finite almost everywhere on B from the representation

$$\mu_t * u = \mu_s * (\mu_r * u), \tag{47}$$

because by Lemma 3 convolution with μ_s would detect the set where $\mu_r * u$ is infinite with probability one. Since r was arbitrary, the lemma is proved.

Actually the relation

$$\mu_r * u < \infty \text{ a.e.} \tag{48}$$

for a non-negative function u is true or false simultaneously for all values of r, and it is impossible for the convolution to converge on a set of positive measure unless it converges almost everywhere. These facts can be derived from (40), but they are contained in our later results, and we shall not prove them here.

The importance of the Cauchy measures lies in this property:

MALLIAVIN'S THEOREM. *For any function f of H^1 we have*

$$\log|\mu_r * f| \leqslant \mu_r * \log|f| \quad \text{a.e.} \tag{49}$$

Of course this is a generalization of the classical fact that a subharmonic function in the unit circle is dominated by the Poisson integral of its boundary values. In our general context, however, the functions $\mu_r * f$ do not need to be continuous, and they exist only as summable functions almost everywhere. The theorem of Malliavin is closely related to this result which we shall use as well, and which was found independently for the class of double power series by M. Rosenblum:

THEOREM 4. *If f is an outer function in H^1, then* (49) *is almost everywhere equality.*

The proof of these theorems belongs to the function-theoretic part of this work, and is postponed until section seven. We proceed meanwhile to the solution of the prediction problem which is the main result of these sections.

6. The prediction problem

THEOREM 5. *For every summable non-negative weight function w the process $\{\chi_\lambda\}$ in L^2_w is pure. It is of type one if*

$$\int \log w \, d\sigma > -\infty; \tag{50}$$

of type two if (50) *fails but*

$$\mu_r * \log w > -\infty \text{ a.e.} \quad (0 < r < 1); \tag{51}$$

and of type three if (51) *fails, when necessarily*

$$\mu_r * \log w = -\infty \text{ a.e.} \quad (0 < r < 1). \tag{52}$$

The proof will take the rest of this section, and for the sake of clarity we shall break it into a number of simple lemmas.

Let $d(\tau)$ denote the distance in L^2_w from χ_0 to $\mathfrak{M}_\tau$, the manifold spanned by linear combinations of χ_λ with $\lambda > \tau$. For $\varepsilon > 0$, $d_\varepsilon(\tau)$ is the corresponding distance in $L^2_{w_\varepsilon}$, where $w_\varepsilon = \max(w, \varepsilon)$. These distances of course are zero for $\tau < 0$; and we know [10] that $d(0) > 0$ if and only if (50) holds.

The condition $d(0) > 0$ evidently means that the process has non-trivial innovation component, but it is easy to go further and show that the process is *purely* of type one. Consider $\mathfrak{H}_1$, the invariant subspace spanned by the innovation elements y_τ in L^2_w. Since these elements are mutually orthogonal, we have from (15)

$$\int \chi_\lambda |y_0|^2 w \, d\sigma = 0 \quad (\lambda \neq 0), \tag{53}$$

so that y_0 has constant modulus different from zero. Therefore in the decomposition (16) only e_1 can be different from the null function, and the process is pure as we asserted.

Now in the rest of the proof we may assume that (50) fails, and the process has no innovation component.

LEMMA 6. *If $d(\tau) = 0$ for some $\tau > 0$, then the same is true for every τ.*

For if χ_0 belongs to $\mathfrak{M}_\tau$ for some $\tau > 0$, then more generally χ_λ belongs to $\mathfrak{M}_{\lambda+\tau}$ for every λ. It follows that all the subspaces $\mathfrak{M}_\tau$ are identical, and so χ_0 is in $\mathfrak{M}_\tau$ for every τ.

LEMMA 7. $d(\tau) = \lim_{\varepsilon \to 0} d_\varepsilon(\tau)$ *for each* $\tau > 0$.

We have for each ε

$$d(\tau)^2 = \inf_P \int |1 + \chi_\tau P|^2 w \, d\sigma \leqslant \inf_P \int |1 + \chi_\tau P|^2 w_\varepsilon \, d\sigma, \tag{54}$$

where P ranges over all analytic trigonometric polynomials (6) with mean value zero. Taking the limit in ε we find $d(\tau) \leqslant \lim d_\varepsilon(\tau)$.

Conversely, for each fixed P

$$\int |1 + \chi_\tau P|^2 w \, d\sigma = \lim_\varepsilon \int |1 + \chi_\tau P|^2 w_\varepsilon \, d\sigma \geqslant \lim_\varepsilon d_\varepsilon(\tau)^2. \tag{55}$$

Taking the infimum over P gives $d(\tau) \geqslant \lim_\varepsilon d_\varepsilon(\tau)$, as we had to show.

For each ε, let g_ε be the outer function in H^2 such that

$$|g_\varepsilon|^2 = w_\varepsilon, \quad \int g_\varepsilon \, d\sigma > 0. \tag{56}$$

(This is the representation obtained in Theorem 3 of [10].) We have

$$g_\varepsilon(x) \sim \sum_{\lambda \geqslant 0} a_\varepsilon(\lambda) \chi_\lambda(x); \quad \sum |a_\varepsilon(\lambda)|^2 = \int |g_\varepsilon|^2 d\sigma = \int w_\varepsilon \, d\sigma. \tag{57}$$

LEMMA 8. *In order for the process to be of type three it is necessary and sufficient that*

$$\lim_{\varepsilon\to 0}\sum_{0\leqslant\lambda\leqslant\tau}|a_\varepsilon(\lambda)|^2=0 \tag{58}$$

for every positive τ. This is the case if the condition holds for a single value of τ.

It is obvious that the process is of type three just if $d(\tau)=0$ for all τ, or by Lemma 6, even for a single τ. We shall show that the sum in (58) is exactly $d_\varepsilon(\tau)^2$. Therefore, using Lemma 7, the limit is $d(\tau)$, and (58) is necessary and sufficient for the process to be deterministic.

Since g_ε is outer, functions of the form Pg_ε where P is an analytic trigonometric polynomial are dense in H^2. We conclude easily that functions $\chi_\tau P g_\varepsilon$, where now P is analytic with mean value zero and τ is fixed, can approximate any function in H^2 whose coefficients vanish for indices less than or equal to τ. In particular, if we approximate the function

$$-\sum_{\lambda>\tau}a_\varepsilon(\lambda)\chi_\lambda \tag{59}$$

then in the expansion

$$(1+\chi_\tau P)g_\varepsilon\sim\sum_{0\leqslant\lambda\leqslant\tau}a_\varepsilon(\lambda)\chi_\lambda+\sum_{\lambda>\tau}b_\varepsilon(\lambda)\chi_\lambda \tag{60}$$

the second sum has norm as small as we please. Therefore

$$d_\varepsilon(\tau)^2=\inf_P\int|1+\chi_\tau P|^2|g_\varepsilon|^2\,d\sigma=\sum_{0\leqslant\lambda\leqslant\tau}|a_\varepsilon(\lambda)|^2 \tag{61}$$

as we had to prove.

LEMMA 9. *In order for* (58) *to hold it is necessary and sufficient that $\mu_r * g_\varepsilon$ tend to zero in the norm of H^2 as ε tends to 0, for each $r\,(0<r<1)$. It suffices that this be true for a single value of r.*

The lemma is obvious from the equality

$$\|\mu_r * g_\varepsilon\|^2=\sum|a_\varepsilon(\lambda)|^2r^{2\lambda}, \tag{62}$$

together with the fact that

$$\sum|a_\varepsilon(\lambda)|^2 \tag{63}$$

is uniformly bounded as ε tends to 0.

LEMMA 10. *$\mu_r * g_\varepsilon$ tends to zero in H^2 if and only if $\mu_r * \log w=-\infty$ almost everywhere.*

From Theorem 4 we have almost everywhere

$$\log|\mu_r * g_\varepsilon|^2=\mu_r * \log w_\varepsilon, \tag{64}$$

and so

$$\int|\mu_r * g_\varepsilon|^2\,d\sigma=\int\exp(\mu_r * \log w_\varepsilon)\,d\sigma. \tag{65}$$

The left side is finite, so the integrand on the right is at any rate summable; since w_ε decreases to w the lemma follows.

From Lemmas 8, 9 and 10 we see that *the process is deterministic if and only if* (52) *holds; and* (52) *is true for every r if it holds for a single one.* Therefore the process has non-trivial evanescent component just if (52) fails. We assert that in this case (51) holds. Indeed, suppose for some r there is a set of positive measure on which $\mu_r * \log w$ is finite. We know, from (47) as before, that if this convolution is infinite on a set of positive measure, then $\mu_t * \log w = -\infty$ almost everywhere for each $t < r$, and hence for all t, contrary to hypothesis. Therefore (51) is the only alternative to (52).

If (51) holds, the process has some evanescent component. We complete the proof of the theorem by showing that the canonical process in L^2_w is pure, so that it must be purely evanescent.

LEMMA 11. *If w vanishes on a set of positive measure, the process is deterministic.*

Indeed, if w vanishes on a set of positive measure, then by Lemma 3 (52) holds.

(This result was proved in a direct way in our conversation with Malliavin, before we knew the criterion (52) for determinacy.)

Now suppose the canonical process in L^2_w has both remote past and evanescent component. By the decomposition theorem we have

$$w = w_2 + w_3, \quad w_2 \cdot w_3 \equiv 0, \tag{66}$$

where the process $\{\chi_\lambda\}$ in $L^2_{w_2}$ is purely evanescent, and that in $L^2_{w_3}$ is deterministic. But according to (66), w_2 must vanish on a set of positive measure. Lemma 11 asserts that this is impossible. Therefore one or the other component in (66) must have been null. This shows that the canonical process in L^2_w is pure, and the theorem is completely proved.

It is important to remark, finally, that conditions (50) and (51) really are different, or in other words that the canonical process in L^2_w can be evanescent. If $\log w$ is summable, so that (50) holds, then indeed (51) is true. To show that the converse implication is false, let $w = e^{-u}$ where u is the function of Lemma 5. Then w is summable and satisfies (51) but not (50).

Now we leave prediction theory and turn to the study of analytic functions on B.

7. Outer functions

In order to prove Malliavin's Theorem and Theorem 4, the most natural method (and the one followed by Malliavin) is to refer the problem to the complex plane by means of the canonical image B_0 of the line in B. This technique (which has been conspicuously exploited by Bochner) gives complete results about B at least in principle, because B can be constructed directly out of R so that R becomes B_0. But the details of proof by this method are often formidable, and we prefer to present an intrinsic function theory on B which furnishes proofs which seem to us closer to the subject matter. In this section we offer proofs of Malliavin's Theorem and Theorem 4 in this spirit. First we have to extend Beurling's notion of *outer function* further than we did in [10], of which the results were summarized in section two.

Let u and v be real functions defined on B. We say that v is *conjugate* to u if $u + iv$ is analytic in some suitably general sense, and v has mean value zero. If u is in L^2, there is exactly one function v in L^2 having mean value zero such that $u + iv$ belongs to H^2; but if u is merely summable, there may be no summable function v such that $u + iv$ is in H^1. We do have this weaker result [9]: for $0 < p < 1$ there exists a constant K_p such that

$$(\int |v|^p d\sigma)^{1/p} \leqslant K_p \int |u| d\sigma \tag{67}$$

for every trigonometric polynomial u with conjugate v. If a sequence $\{u_n\}$ of trigonometric polynomials converges to a function u in the metric of L^1, then the conjugate functions form a fundamental sequence in the metric space L^p for each $p < 1$ and so converge in measure to a unique limit function v. By definition v is the function conjugate to u. It will be important in the proofs to follow that v is besides the *pointwise* limit almost everywhere of a suitably chosen subsequence of $\{v_n\}$.

The next theorem gives a characterization of outer functions which is really an abstract version of the integral representation formula of Beurling.

THEOREM 6. *If u is a real function such that u and e^u are summable, then e^{u+iv} (where v is conjugate to u) is an outer function in H^1. Conversely, if a summable outer function f has the representation e^{u+iv} with u and v real, then u is summable with e^u and v is equal to its conjugate modulo 2π, aside from an additive constant.*

Proof. Suppose that u and e^u are summable and v is conjugate to u. We have to show that $f = e^{u+iv}$ belongs to H^1, and then that it is outer. First suppose that u (and so also v) are trigonometric polynomials. The power series expansion for the exponential is uniformly convergent, and the analyticity of f is therefore obvious. The power series also shows that (7) holds:

$$\int f\, d\sigma = \sum_0^\infty \int \frac{(u + i\, v)^n}{n!}\, d\sigma = \sum_0^\infty \frac{1}{n!} \left[\int (u + i\, v)\, d\sigma \right]^n = e^{\int (u+iv)d\sigma} = e^{\int \log|f| d\sigma}. \tag{68}$$

More generally, if both u and v are bounded, $u + iv$ can be approximated *boundedly* by trigonometric polynomials, and the two properties of f persist in the limit.

Two more limit processes are required to treat the general case, and they use the result about conjugate functions just stated. Suppose u is bounded, but not necessarily v. Choose trigonometric polynomials u_n converging boundedly to u and such that the conjugate trigonometric polynomials v_n tend to v almost everywhere. Then

$$\int |e^{u+iv} - e^{u_n+iv_n}| \, d\sigma \tag{69}$$

tends to zero by the theorem on dominated convergence. As the limit in norm of elements of H^1, f itself belongs to H^1. Moreover in the equation

$$\int u_n \, d\sigma = \log \int e^{u_n+iv_n} \, d\sigma \tag{70}$$

which expresses the fact that each function $e^{u_n+iv_n}$ is outer, we can pass to the limit on both sides to obtain the same result for f.

Finally, for an unbounded function u define

$$u_n = \begin{cases} u & \text{where} \quad -n \leqslant u \leqslant n \\ n & \text{where} \quad u > n \\ -n & \text{where} \quad u < -n. \end{cases} \tag{71}$$

Once more we can assume that the conjugate functions v_n tend to v pointwise, and the theorem on dominated convergence shows that the norm in (69) tends to zero, and that the equation (70) is valid in the limit.

Conversely, let $f = e^{u+iv}$ be an outer function in H^1. We know from (7) that u is summable; let v' be its conjugate function. The part of our theorem already proved states that $g = e^{u+iv'}$ is analytic and outer. We have to show that outer functions with the same modulus can differ only by a constant factor.

Set $w = |f| = |g|$. In L^2_w the trigonometric polynomials $1 + P$, where P is analytic with mean value zero, form a convex subset in whose closure there is a unique element $1 + H$ of minimal norm. We have shown [10] that $1 + H$ is not the null function, and further that $|1 + H|^2 w$ is almost everywhere equal to $\omega = \exp\left(\int \log w \, d\sigma\right)$. Hence $(1 + H)^2 f$ and $(1 + H)^2 g$ have modulus constant and equal to ω.

On the other hand $(1 + H)^2 f$ is the limit in H^1 of functions of the form $(1 + P)^2 f$, each of which has mean value equal to the mean value of f. Therefore, using (7), we have the two relations

$$|(1+H)^2 f| \equiv \omega; \quad \left|\int (1+H)^2 f \, d\sigma\right| = \left|\int f \, d\sigma\right| = \omega. \tag{72}$$

It follows that $(1 + H)^2 f$ is almost everywhere equal to a constant.

The same reasoning applies to g, and the new function $1 + H$ is the same as the old, because it depends only on w. Therefore $(1 + H)^2 g$ is almost everywhere constant as well, and since $1 + H$ almost never vanishes, f and g differ at most by a constant factor. This completes the proof of the theorem.

Proof of Theorem 4. The Cauchy measures μ_r have the characteristic property

$$\mu_r * e^h = \exp(\mu_r * h) \tag{73}$$

if h is analytic and suitably restricted. From this formula, (49) with equality follows in a formal way for the function $f = e^h$:

$$\log|\mu_r * e^h| = \mathrm{Re}(\mu_r * h) = \mu_r * \mathrm{Re}(h)h = \mu_r * \log|e^h|. \tag{74}$$

Unfortunately (73) is not easy to prove as generally as it is true, and it is not generally enough true to prove our theorem because the right side is not defined when h is not summable, even if e^h belongs to H^1.

Suppose however that h is a trigonometric polynomial. If P and Q are any analytic trigonometric polynomials, then

$$\mu_r * (PQ) = (\mu_r * P)(\mu_r * Q), \tag{75}$$

as one sees by comparing coefficients in the Fourier expansions of left and right sides. In particular, $\mu_r * P^2 = (\mu_r * P)^2$, with the same equation for higher exponents by induction. Hence

$$\sum_0^\infty \frac{\mu_r * h^n}{n!} = \sum_0^\infty \frac{(\mu_r * h)^n}{n!}, \tag{76}$$

and this is exactly (73). By (74), then, Theorem 4 holds for outer functions $f = e^h$ where h is any analytic trigonometric polynomial.

According to the characterization of outer functions given by Theorem 6, we have to prove that

$$\log|\mu_r * e^{u+iv}| = \mu_r * \log u \tag{77}$$

for each real function u, summable together with e^u, with conjugate function v. We have just established this fact if u is a trigonometric polynomial. Now we imitate the approximation argument used in the proof of Theorem 6 to establish (77) for larger classes of functions. The details are neither novel nor very tedious, and so we shall not reproduce them.

The relation (77) is a kind of generalization of the defining property (7) of outer functions. If $f = e^{u+iv}$ is continuous and bounded from zero, we can let r tend to 0 in (77) and obtain

$$\log\left|\int f\,d\sigma\right| = \int \log|f|\,d\sigma. \tag{78}$$

In the function theory of the unit circle, where μ_r is simply the Poisson kernel, the general form (77) can be obtained from its special case (78) by appropriate conformal transformations of the interior of the circle. But on B, μ_r is singular with respect to Haar measure, and the limiting case $r=0$ is distinguished from the regular ones $r>0$. Therefore some more complicated device, such as the one used in our proof, must be used to derive (77) from (78).

Proof of Malliavin's Theorem. Let f belong to H^1, and suppose its mean value is different from zero. Then [10, p. 178] $f=g\cdot h$, where g is outer and h is an analytic function with constant modulus equal to 1. The result to be proved takes the form

$$\log|\mu_r * (g\cdot h)| \leqslant \mu_r * \log|g|. \tag{79}$$

Now (75) is still true, by a standard limit process, if P and Q are replaced by g and h; and since μ_r has unit total mass we have

$$|\mu_r * h| \leqslant 1 \quad \text{a.e.} \tag{80}$$

Therefore using Theorem 4 for the outer function g,

$$\log|\mu_r * f| = \log|\mu_r * g| + \log|\mu_r * h| \leqslant \log|\mu_r * g| = \mu_r * \log|g|. \tag{81}$$

Thus (79) is proved, under the assumption that the mean value of f is not zero.

If the integral of f does vanish, we have (49) anyway for $f+\varepsilon$ in place of f. Fatou's Lemma shows that the inequality is preserved in the limit.

8. Analytic functions

THEOREM 7. *If a function f in H^1 vanishes on a set of positive measure in B, or indeed if merely*

$$\mu_r * \log|f| = -\infty \quad \text{a.e.}, \tag{82}$$

then f vanishes identically. Otherwise under either condition

$$\int f\,d\sigma \neq 0 \tag{83}$$

or

$$f \text{ is continuous} \tag{84}$$

we have

$$\int \log|f|\,d\sigma > -\infty. \tag{85}$$

Nevertheless there are non-null functions in H^∞ for which (85) *is false. More generally, if w is a non-negative summable function such that*

$$\mu_r * \log w > -\infty \quad \text{a.e.} \tag{86}$$

then there is some f in H^1 satisfying

$$0 < |f| \leqslant w \quad \text{a.e.} \tag{87}$$

This theorem contains most of what we know about analytic functions on B. The proof may be clearer if we comment on its various statements before we begin. On the circle group (85) holds for non-null analytic functions subject to very mild growth conditions inside the circle. The fact that (85) follows from (83) in general is an elementary result from [10]; the same conclusion from (84) was proved by Arens [3] in two ways, but neither one is easy. We shall give a new proof of Arens' Theorem which relates it to the other statements of Theorem 7.

In spite of these positive results, (85) is not true in general even for analytic functions which are *bounded*. This fact follows from the last assertion of the theorem if we take w to be bounded, and satisfy (86) but not

$$\int \log w \, d\sigma > -\infty. \tag{88}$$

The existence of such weight functions is asserted by Lemma 5.

If we ask only that the counterexample belong to H^2, then the existence of functions violating (85) follows easily from Theorem 5, and this simple proof is given before the more complicated discussion of (87).

Proof of Theorem 7. Suppose f belongs to H^1 but satisfies (82); this will be the case in particular if f vanishes on a set of positive measure, by Lemma 3. From Malliavin's Inequality

$$-\infty = \mu_r * \log |f| \geqslant \log |\mu_r * f|. \tag{89}$$

Therefore $\mu_r * f$ vanishes identically. But the Fourier-Stieltjes coefficients (38) of μ_r vanish nowhere, so that f must itself be the null function, as we had to prove.

To prove Arens' Theorem, let f be a continuous analytic function which is not everywhere zero. Then $\mu_r * f$ is continuous and not identically zero. Find an open set E on which $\mu_r * f$ is bounded from zero, and construct a continuous function h such that $h(-x)$ is non-negative everywhere, positive somewhere on E, and zero outside E. Then we have

$$h * \log |\mu_r * f|(0) > -\infty. \tag{90}$$

Using Malliavin's Inequality and the associativity of convolution,

$$(h * \mu_r) * \log |f|(0) \geqslant h * \log |\mu_r * f|(0) > -\infty. \tag{91}$$

But $h * \mu_r$ is a continuous non-negative function, which moreover can never vanish, because h is non-negative and μ_r has mass in every open set. Hence $h * \mu_r \geqslant \eta > 0$ on B, and (91) implies

$$\eta \int \log |f| \, d\sigma > -\infty. \tag{92}$$

Our proof to this point was found in conversation with Malliavin. The rest of the proof depends on Theorem 5, and has a different character.

We are to construct an analytic function f not identically zero but such that (85) is false. For this purpose find a weight function w such that (86) holds but not (88); w can moreover be taken bounded. By Theorem 5 there is a non-null element g orthogonal to $\mathfrak{M}_0$ in L^2_w:

$$\int \chi_\lambda \bar{g} w \, d\sigma = 0 \quad (\text{all } \lambda > 0), \tag{93}$$

so that $f = \bar{g} w$ is summable and analytic. Define

$$W = |g|^2 w, \tag{94}$$

an element of L^1; then

$$|g| = (W/w)^{1/2}, \quad |f| = |g| w = (Ww)^{1/2}. \tag{95}$$

From (95) we see that f is square-summable (because w is bounded), and so belongs to H^2, and we have

$$\int \log |f| \, d\sigma = \frac{1}{2} \int \log \, W \, d\sigma + \frac{1}{2} \int \log \, w \, d\sigma. \tag{96}$$

The first term on the right side is finite or negatively infinite, and by the choice of w the second term is $-\infty$. Hence the left side diverges also, and f has the properties sought.

We come to the final assertion of the theorem. Suppose that w satisfies (86); once more by Theorem 5 we can find g in L^2_w not identically zero such that (93) holds. Then $|g| w^{1/2}$ is in L^2, and so

$$U = \min (1, \, |g|^{-1} w^{-1/2}) \tag{97}$$

satisfies

$$\int \log U \, d\sigma > -\infty. \tag{98}$$

Using Theorem 6, find h in H^∞ so that $U = |h|$ almost everywhere. (We do not need the fact that h can be chosen to be outer.) Now $\bar{g} w$ belongs to H^1 and h to H^∞, so the product $h \bar{g} w$ belongs to H^1, and evidently

$$|h \bar{g} w| = \min (|g| w, \quad w^{1/2}) \leq w^{1/2}. \tag{99}$$

Neither analytic factor of $h \bar{g} w$ vanishes on a set of positive measure, and so this function is a non-null element of H^2. Therefore $f = (h \bar{g} w)^2$ belongs to H^1 and satisfies (87). All parts of the theorem have now been proved.

We do not know whether (87) can be improved to

$$|f| = w \quad \text{a.e.}; \tag{100}$$

this question is related to interesting problems with prediction-theoretic import which lie deeper than Theorem 5. A result we are publishing elsewhere [11] can be mentioned in this connection, although it does not directly extend our knowledge about analytic functions.

9. Moving averages

We come now to the second part of the paper, having to do with multivariate stochastic processes and the related theory of matrix-valued functions. This section is devoted to the notion of *moving average*, and more generally to the prediction-theoretic point of view. The results and proofs are not new, and we shall only sketch the development; a complete exposition has been given by Masani and Wiener [16]. Beginning with the next section it will be *essential* that we study the group of integers, rather than ordered groups of more general type to which the matrix theory of our first paper applied.

First we set down the basic definitions. Functions are defined now on the unit circle, or more rarely inside it. The measure $dx/2\pi$ on $(0, 2\pi)$ is denoted by $d\sigma$. An integer N is given and fixed, and reference is usually made to complex Euclidean space of N dimensions. Matrices with N columns operate on the column vectors of this space by multiplication on the left. The trace function is normalized so that I, the identity matrix of any order, has trace 1. We say that two matrices have the *same shape* if they have the same number of rows and the same number of columns.

We consider functions F defined on the circle and taking matrices of various shapes as values:

$$F(e^{ix}) = (F_{jk}(e^{ix})). \tag{101}$$

The entries F_{jk} are always measurable complex scalar functions. If F and G are constant matrices, their inner product is by definition

$$(F, G) = \operatorname{tr}(FG^*), \tag{102}$$

meaningful whenever F and G have the same shape. This definition is extended to matrix functions by integration:

$$(F, G) = \int \operatorname{tr}(FG^*)\, d\sigma. \tag{103}$$

If F and G have p rows, then evidently

$$(F, G) = p^{-1} \sum_{j,k} \int F_{jk}\, \bar{G}_{jk}\, d\sigma. \tag{104}$$

By the *range* of F we mean the linear set of column vectors FX, where X ranges over the constant column vectors. The range of $F(e^{ix})$ generally depends on e^{ix}, and so F determines a mapping from the circle to the class of subspaces of the given Euclidean space. Such a mapping will be called a *range function*, and much is to be said about range functions hereafter.

If the range of F is orthogonal to the range of G at each point, then $(F, G) = 0$. Indeed, for any constant vectors X and Y we have

$$0 = (FX, GY) = (G^* FX, Y); \tag{105}$$

therefore $G^* F = 0$, and

$$0 = N\int \operatorname{tr}(G^* F)\, d\sigma = p\int \operatorname{tr}(FG^*)\, d\sigma = p(F, G). \tag{106}$$

From (104) we see that $(F, F) = \|F\|^2$ is finite if and only if each entry F_{jk} is square-summable. The functions F with $\|F\| = 0$ form a subspace of those with finite norm, and the Hilbert space $\mathbf{L}^2$ is obtained as for scalar functions by identifying F and G if $F - G$ is a null function.

Actually we have a space $\mathbf{L}^2$ for matrices of each shape, and the number of rows and columns must be given in each context.

For functions of any shape belonging to $\mathbf{L}^2$ (or merely having summable entries) we have the notions of Fourier coefficient and Fourier series:

$$F_n = \int F(e^{ix})\, e^{-nix}\, d\sigma(x); \quad F(e^{ix}) \sim \sum_{-\infty}^{\infty} F_n\, e^{nix}. \tag{107}$$

A function and its coefficients are related by the Parseval equality:

$$\|F\|^2 = \sum \operatorname{tr}(F_n F_n^*). \tag{108}$$

Each space $\mathbf{L}^2$ contains a distinguished subspace $\mathbf{H}^2$ consisting of those functions F such that $F_n = 0$ for $n = -1, -2 \ldots$. A function F in $\mathbf{H}^2$ admits an analytic extension to the interior of the circle:

$$F(z) = \sum_{n=0}^{\infty} F_n z^n. \tag{109}$$

Our method systematically avoids mention of such extensions, but we shall apply the word *analytic* to functions in $\mathbf{H}^2$ without misunderstanding.

Now let M be a square matrix of order N whose entries are complex Borel measures M_{jk}, and suppose that M is positive semi-definite in the sense that $M(E)$ is a positive semi-definite matrix for each Borel set E. For continuous functions, at least, we can define a new inner product

$$(F, G)_M = \int \operatorname{tr}(F\, dM\, G^*) \tag{110}$$

and associated norm $\| F \|_M$. We identify functions differing by a null function, and complete the space so obtained. Then we have a Hilbert space $\mathbf{L}^2(M)$ whose elements indeed can be represented as matrix functions; but F and G may represent the same element of $\mathbf{L}^2(M)$ even if their entries are quite different, unless M is a measure of particular type. In general the elements of the space have no Fourier series. It is necessary to specify the number of rows in the matrix of a function in $\mathbf{L}^2(M)$, but the number of columns can only be N if (110) is to have meaning.

A *multivariate stationary process* is a family of N stationary processes $\{X_n^1\}, \ldots, \{X_n^N\}$ in the same Hilbert space, depending (for our purposes) on the integer parameter n, and mutually stationarily correlated:

$$(X_m^j, X_n^k) = \varrho_{jk}(m-n) \quad (j, k = 1, \ldots, N). \tag{111}$$

(The inner product refers to the Hilbert space of the processes.) There is a positive semi-definite matrix-valued measure M with component measures M_{jk} $(j, k = 1, \ldots, N)$ such that

$$\varrho_{jk}(n) = \int e^{-nix}\, d M_{jk}(x). \tag{112}$$

We shall construct an isomorphic process in $\mathbf{L}^2(M)$. Let I^j be the jth row of the identity matrix I of order N $(j = 1, \ldots, N)$. Then $\tilde{X}_m^j = e^{-mix} I^j$ is a function belonging to $\mathbf{L}^2(M)$ formed with row vectors, and for each j, k, m, and n we have the inner product relations

$$(\tilde{X}_m^j, \tilde{X}_n^k)_M = \int e^{-(m-n)ix} I^j\, dM\, I^{k*} = \int e^{-(m-n)ix}\, d M_{jk}(x) = \varrho_{jk}(m-n) = (X_m^j, X_n^k). \tag{113}$$

If we map X_m^j onto $\tilde{X}_m^j$ and extend the correspondence in a linear way, we obtain by (113) a unitary operator which carries the smallest subspace containing all the X_m^j onto $\mathbf{L}^2(M)$, formed with row vectors. Obviously every inner product relation among the X_m^j holds as well for the $\tilde{X}_m^j$. Therefore we abandon the distinction between X and $\tilde{X}$, and study the process in $\mathbf{L}^2(M)$. This process is called the *canonical process* in that space, to distinguish it from other processes going on in the same space.

The canonical process in $\mathbf{L}^2(M)$ consists of the N row vectors

$$e^{-mix} I^j \quad (j = 1, \ldots, N) \tag{114}$$

depending on the variable e^{ix} and further on the integer parameter m. If we write the row vectors together we obtain the matrix function

$$X_m(e^{ix}) = e^{-mix} I. \tag{115}$$

In some problems it is convenient to study this single matrix function in place of the N vector functions (114).

Conversely, suppose that X_m is a square matrix function belonging to $\mathbf{L}^2(M)$ for each integer m, whose rows are X_m^j $(j = 1, \ldots, N)$. In order for its rows to form a multivariate stationary process in the space $\mathbf{L}^2(M)$ of row vectors, it is necessary and sufficient according to (111) that

$$(X_m^j, X_n^k)_M = \varrho_{jk}(m, n) \tag{116}$$

should depend on m and n only through their difference $m - n$. Of course the isomorphic canonical process may lie in a different measure space.

If the correlation matrix ϱ is the identity matrix for $n = 0$ and zero otherwise, the process is said to be *orthonormal*. Then every pair of vectors X_m^j and X_n^k is orthogonal unless $j = k$ and $m = n$.

Suppose $X = \{X_m^j\}$ $(j = 1, \ldots, N)$ is a multivariate process, and A is a matrix function of p rows and N columns in $\mathbf{L}^2$. Denote the Fourier coefficients of A by $A_n = (A_n^{jk})$. Then we can define a new process $Y = \{Y_m^j\}$ $(j = 1, \ldots, p)$ by the formula (convergence being assumed)

$$Y_m^j = \sum_{i,k} A_i^{jk} X_{m-i}^k \quad (i = 0, \pm 1, \ldots;\ k = 1, \ldots, N). \tag{117}$$

If X is given as a matrix function in $\mathbf{L}^2(M)$, this can be written more concisely as

$$Y_m = \sum_i A_i X_{m-i}, \tag{118}$$

where A_i and X_{m-i} are combined by ordinary matrix multiplication, or still more briefly as

$$Y = A \ast X. \tag{119}$$

Y is called a *moving average* of X.

If A and B are two functions of proper shapes in $\mathbf{L}^2$, it is easy to see (convergence questions aside) that

$$A \ast (B \ast X) = (A\,B) \ast X. \tag{120}$$

The case when X is orthonormal is particularly important. Then (117) is always convergent in norm, and the covariance matrix M of $Y = A \ast X$ is the square matrix of order p given by

$$dM = A A^* d\sigma = W d\sigma. \tag{121}$$

If A belongs to $\mathbf{H}^2$ instead of merely to $\mathbf{L}^2$, we call Y a *one-sided moving average* of X, and then W has the special form $A A^*$ for an analytic function A. Conversely, if a process

Y has covariance measure (121), then Y and $A * X$ (where X is any orthonormal process) are both isomorphic to the canonical process in $\mathbf{L}^2(W)$, and hence to each other.

We should like to say more exactly that Y *is* a moving average in its own Hilbert space, but for that we have to choose a particular function A satisfying (121) and use a more complicated argument. For each integer n, let $\mathfrak{M}_n$ be the smallest closed manifold containing all the elements Y_m^j ($m \leqslant n$; $j = 1, \ldots, p$). $\mathfrak{M}_0$ is known familiarly as *the past*, and

$$\mathfrak{M}_{-\infty} = \bigcap_n \mathfrak{M}_n \tag{122}$$

as the *remote past*. (The manifolds $\mathfrak{M}_n$ now increase with n, whereas in the first part of the paper they decreased. The change results from the choice of sign in (114). The convention now adopted is better from the prediction-theoretic point of view.)

We say that Y has *no* remote past if, more accurately, the remote past consists of the null element alone. It is almost immediate that Y has no remote past if its covariance measure has the form (121), with A in $\mathbf{H}^2$ having any number of columns. We shall show conversely that Y is a moving average, and of a particular type, if it has no remote past.

For each n let $\mathfrak{R}_n$ be the orthogonal complement of $\mathfrak{M}_{n-1}$ in $\mathfrak{M}_n$. The subspaces $\mathfrak{R}_n$ are mutually perpendicular, and satisfy

$$\mathfrak{R}_{n+1} = e^{-ix} \mathfrak{R}_n. \tag{123}$$

Moreover since Y has no remote past, they span together the same manifold as all the $\mathfrak{M}_n$. The dimension q of $\mathfrak{R}_0$ is the *rank* of the process, evidently at most equal to p. We choose an orthonormal basis $\{X_0^j\}$ ($j = 1, \ldots, q$) for $\mathfrak{R}_0$, and associated bases

$$X_n^j = e^{-nix} X_0^j \tag{124}$$

for the other spaces $\mathfrak{R}_n$. Then $\{X_m^j\}$ is an orthonormal process having the same past, and the same innovation manifolds $\mathfrak{R}_n$, as Y itself. Writing Y_m^j as a sum of the vectors X_n^k we obtain the representation (117), where the coefficients A_i^{jk} are independent of m because Y is stationary, and because X satisfies (124). The sum contains only terms with $i \geqslant 0$, so that Y is a one-sided moving average of X.

The result whose proof has been outlined is that *these conditions are equivalent: the process Y is a moving average of an orthonormal process; the covariance measure of Y has the form* (121), *where A is in* $\mathbf{H}^2$ *with (necessarily) p rows but any number of columns; the remote past of Y is null.*

This interesting function-theoretic corollary can be mentioned: *if W has the form AA^* for some A in* $\mathbf{H}^2$ *with any number of columns, then A can be replaced by an analytic function B having at most as many columns as rows.*

It will be profitable now to leave the prediction model and study analytic functions directly.

10. Outer functions

DEFINITION. *A function A in $\mathbf{H}^2$ (with p rows and N columns) is a* left-outer function *if the convex set of functions*

$$(I+P)A \tag{125}$$

where I is the identity matrix of order p and P ranges over all trigonometric polynomials of order p having the form

$$P(e^{ix}) = \sum_{n>0} P_n e^{nix} \tag{126}$$

contains a constant matrix in its closure.

Evidently the constant matrix can only be the constant term of A.

A *right-outer function* is defined in the same way, replacing the functions (125) by

$$A(I+P), \tag{127}$$

where I and P are square of order N. The properties of one kind of outer function can be deduced from those of the other by this observation, whose proof is very simple: *a function A in $\mathbf{H}^2$ with coefficients A_n is left-outer if and only if $\tilde{A}$ is right-outer, where $\tilde{A}$ is the analytic function defined by*

$$\tilde{A}(e^{ix}) = A(e^{-ix})^* \sim \sum_0^\infty A_n^* e^{nix}. \tag{128}$$

In the last section we obtained a representation for a process Y without remote past as a one-sided moving average $A*X$ in a particular way, with the result that the innovation spaces $\mathcal{R}_n$ for the process Y are identical with the corresponding spaces for X. This property implies that the analytic function A whose coefficients are the matrices A_n of the moving average is left-outer. We pause to show why this fact is so. Let Y be a canonical process of N vectors with rank q; then X is an orthonormal process of q vectors, and A belongs to $\mathbf{H}^2$ with N rows and q columns. For any square analytic trigonometric polynomial Q of order N, we have by (120)

$$(Q*Y)_0 = [(QA)*X]_0. \tag{129}$$

The matrix functions which appear on the left, as Q varies, are exactly the matrices whose rows are linear combinations of vectors Y_m^j ($m \leqslant 0$; $j = 1, \ldots, N$). Since the past of Y is the same as the past of X, the matrices on the right are dense in the set of all matrices whose rows belong to the past of X. But this implies, because X is orthonormal, that the functions QA are dense in $\mathbf{H}^2$. It follows immediately that functions of the form (125) are dense in the subset of $\mathbf{H}^2$ containing all functions with the same constant term as A, and so A is outer.

THEOREM 8. *Suppose A is a square matrix in $\mathbf{H}^2$ with non-singular constant coefficient A_0. Then A is left-outer if and only if*

$$\int \log |\det A| d\sigma = \log |\det A_0|. \tag{130}$$

It follows that a function A is right-outer if and only if it is left-outer, if A_0 is non-singular, because A and $\tilde{A}$ satisfy (130) at the same time.

Proof. The theorem is closely related to Theorem 8 of our first paper, and we shall use the method developed there. For a fixed matrix B_0 with determinant 1, and any summable function W with positive semi-definite values, we consider the convex set in $\mathbf{L}^2(W)$ obtained by closing the set of functions $B_0 + P$, where P ranges over the trigonometric polynomials of the form (126). If we denote the unique element of smallest norm in the set by $B_0 + H$, then $(B_0 + H) W (B_0 + H)^*$ is a constant positive semi-definite matrix C. We have moreover $\det C = \exp \int \log \det W d\sigma$, from (44) of the first paper and the elementary equality $(\det C)^{1/N} = \inf \operatorname{tr} [D C D^*]$, where D varies over constant matrices of determinant 1.

Now take $W = AA^*$, $B_0 = I$. From the definition it follows directly that A is left-outer if and only if $(I + H)A = A_0$. If this is the case, obviously $C = A_0 A_0^*$; and if $(I + H)A = A_0 + \dots$ is not constant, then $C - A_0 A_0^*$ is positive semi-definite and not 0. Therefore A is left-outer if and only if $C = A_0 A_0^*$, and in any case $A_0 A_0^* \leqslant C$.

From these facts we have

$$\int \log \det (AA^*) d\sigma = \det C \geqslant \det (A_0 A_0^*) \tag{131}$$

with equality if and only if $C = A_0 A_0^*$, which is to say if A is left-outer. This is equivalent to the statement of the theorem.

The argument shows incidentally that the generalization of Szegö's Theorem proved by Masani and Wiener ([16], I, p. 145) is essentially the same as ours.

Our method of proof suggests that it is possible to view the minimum problem in the partially ordered system of Hermitian matrix functions, even though this system is not a lattice, instead of using the trace function to define a norm with scalar values. The proof just given can be modified to carry out this idea, or one can appeal more directly to the orthogonality relations (49, 50, 51) of our first paper.

The next two theorems express properties of outer functions which are essential for our purpose. We state only the version pertaining to left-outer functions.

THEOREM 9. *Let A be a left-outer function. Almost everywhere the null-space $\mathfrak{N}(e^{ix})$ of $A(e^{ix})$ is equal to the null-space of A_0, and so in particular is independent of x and has constant dimension.*

Proof. Suppose that $B^{(n)} = (I + P^{(n)})A$ is a sequence of functions of the form (125) converging to A_0 in the norm of $\mathbf{L}^2$. The entries of $B^{(n)}$ are square-summable scalar functions which converge to the corresponding constant entries of A_0 in the scalar space $\mathbf{L}^2$. Choose a subsequence converging pointwise at each entry for almost all x. If Y is a constant column vector such that $A(e^{ix})\,Y = 0$ for some e^{ix} in the set of convergence, then also $B^{(n)}(e^{ix})\,Y = 0$ for each n of the subsequence, and so in the limit also $A_0 Y = 0$. Therefore, with the exception of a set of points having measure zero, $\mathfrak{N}(e^{ix})$ is contained in the null-space of A_0.

In the other direction, we must show that $A(e^{ix})\,Y = 0$ almost everywhere if Y is any constant vector such that $A_0 Y = 0$. Let $Z(e^{ix})$ be the analytic vector function $A(e^{ix})\,Y$. Evidently $B^{(n)}Z$ will converge in the space $\mathbf{L}^2$ of column-vector functions to $A_0 Y = 0$; or what is the same, $-P^{(n)}Z$ will tend to Z. Assuming that Z is not identically zero, let p be the smallest positive integer (certainly greater than one) such that $e^{-pix}Z$ is *not* analytic. (The existence of p depends on the fact that we are dealing with the group of integers.) By the argument just given, $e^{-pix}Z$ is the limit of functions $-e^{-pix}P^{(n)}Z$, each of which is analytic because the analytic trigonometric polynomials $P^{(n)}$ have no constant term. The limit function must itself be analytic, contrary to assumption, and the contradiction shows that Z was the null function, as we had to prove.

If the left-outer function A has non-trivial null-space $\mathfrak{N}$, we can restrict A to the complement of $\mathfrak{N}$ and so obtain a left-outer function with the same range as A but trivial null-space. It is nearly always convenient to perform this reduction in applications, so that the outer function maps a fixed domain space onto a variable range space of the same dimension.

THEOREM 10. *Let A be a left-outer function with null-space $\mathfrak{N}$. Denote by* $\mathbf{S}$ *the subspace of* $\mathbf{H}^2$ *obtained by closing the linear set of functions QA, where Q ranges over all analytic trigonometric polynomials. Then* $\mathbf{S}$ *consists exactly of those functions in* $\mathbf{H}^2$ *which vanish on $\mathfrak{N}$.*

Proof. We shall prove the theorem for the case where $\mathfrak{N}$ is trivial, and the general case follows easily. By definition, $\mathbf{S}$ contains the constant function A_0. Since also $\mathbf{S}$ is invariant under multiplication on the left by analytic trigonometric polynomials, $\mathbf{S}$ contains each function QA_0. Using the fact that A_0 has full rank, it is easy to see that every analytic trigonometric polynomial of proper shape has this form, so that indeed $\mathbf{S}$ coincides with $\mathbf{H}^2$.

COROLLARY. *Let A be a left-outer function in* $\mathbf{H}^2$, *with coefficients A_n. For any non-negative integer n,*

$$B^{(n)}(e^{ix}) = A_0 + A_1 e^{ix} + \ldots + A_n e^{nix} \tag{132}$$

can be approximated by functions $(I + e^{(n+1)ix}Q)A$, where Q ranges over all analytic trigonometric polynomials.

For $n = 0$ this is exactly the definition of outer function. In general, each coefficient A_j of A vanishes on $\mathcal{N}$, the null-space of A, because it is an average of values $A(e^{ix})$, each of which vanishes on $\mathcal{N}$. Hence both $B^{(n)}$ and $A - B^{(n)}$ vanish on $\mathcal{N}$. The analytic function $e^{-(n+1)ix}(A - B^{(n)})$ can be approximated by functions QA, according to the theorem, so that

$$A - e^{(n+1)ix} QA \tag{133}$$

approximates $B^{(n)}$, and this is equivalent to the statement of the corollary.

A scalar outer function is determined by its modulus, up to multiplication by a constant of modulus 1. Our next theorem generalizes this result to matrix functions.

THEOREM 11. *If A and B are left-outer functions in $\mathbf{H}^2$ such that $AA^* = BB^*$ almost everywhere, then there exists a constant unitary matrix U such that $B = AU$.*

Proof. Let $\mathbf{S}_A$ and $\mathbf{S}_B$ denote the subspaces of $\mathbf{H}^2$ spanned by A and B, respectively, in the manner of Theorem 10. It follows from the hypothesis of this theorem that the mapping which carries QA onto QB establishes a unitary transformation of $\mathbf{S}_A$ onto $\mathbf{S}_B$. Furthermore, A_0 corresponds to B_0 under this transformation; for A_0 and B_0 are characterized by the fact that their norm is minimal in $\mathbf{S}_A$ and $\mathbf{S}_B$ respectively. It follows that QA_0 and QB_0 correspond, and in particular CA_0 and CB_0, if C is any constant matrix. From the relation

$$\|CA_0\| = \|CB_0\| \tag{134}$$

valid for every constant matrix C, the definition (102) implies that $A_0A_0^* = B_0B_0^*$. If D is the positive semi-definite square root of this product, we can find constant unitary matrices T and U such that $A_0 = DT$ and $B_0 = DU$, or

$$B_0 = A_0(T^*U). \tag{135}$$

Now T^*U is a unitary matrix, and $A(T^*U)$ is an outer function whose leading coefficient is the same as that of B. Changing the notation, we shall complete the proof by showing that left-outer functions A and B are identical if they satisfy the hypothesis of the theorem and have the same constant term.

Let $W = AA^* = BB^*$. Find a sequence of trigonometric polynomials $P^{(n)}$ of the form (126) such that $(I + P^{(n)})A$ converges to A_0 in $\mathbf{H}^2$. A standard convexity argument [10] shows that $I + P^{(n)}$ converges in $\mathbf{L}^2(W)$ to a limit function $I + H$ such that $(I + H)A = A_0$. Since A and A_0 have the same null-space, this equality implies that $I + H$ is non-singular on the range of A. Now the correspondence between $\mathbf{S}_A$ and $\mathbf{S}_B$ which has been defined carries $(I + P^{(n)})A$ onto $(I + P^{(n)})B$, which therefore converges to $(I + H)B = B_0 = A_0$. As before $(I + H)$ is non-singular on the range of B, and so these two results show that $A = B$ on the complement of the null-space of A_0. But A and B vanish themselves on the null-space of A_0, so the proof is complete.

Our left-outer functions are identical with the *optimal factors* of Masani and Wiener, and their properties are therefore not entirely new. The class of outer functions we introduce now has not, so far as we know, been considered before.

DEFINITION. *A function V in $\mathbf{L}^2$ (of any shape) is a* partial isometry *if it preserves norm as a mapping of the complement of its null-space onto its range.*

In order for V to be a partial isometry it is necessary and sufficient that $V^* V$ be the orthogonal projection on the complement of the null-space of V (in the domain space), and that $V V^*$ be the projection on the range of V (in the range space).

To a matrix function A we associate the *range function* $\mathfrak{H}$ whose value at e^{ix} is the range of the transformation $A(e^{ix})$. If A belongs to $\mathbf{H}^2$ (of any shape) we say that $\mathfrak{H}$ is *analytic*. The principal result of this development, from which the matrix factorization theorems will be derived, is

THEOREM 12. *Every analytic range function is the range of a left-outer partial isometry.*

Proof. Let A belong to $\mathbf{H}^2$ with p rows and N columns, and let $\mathfrak{H}$ be its range function. Then $W = AA^*$ is a positive semi-definite matrix function of order p with the same range function $\mathfrak{H}$, and $W(e^{ix})$ maps $\mathfrak{H}(e^{ix})$ onto itself for each e^{ix}. We have shown that there is a left-outer function B such that $W = BB^*$, and B can be chosen to have q columns, where q is the rank of the canonical process in $\mathbf{L}^2(W)$. The range of B is the same range function $\mathfrak{H}$, and so every analytic range function is the range of a left-outer function. Now we have to find an outer partial isometry with the same range.

B is of rank q at each point. For otherwise we could restrict B to the complement of its null-space (which is constant because B is outer) and find an analytic function C with fewer than q columns such that $CC^* = W$. Then the canonical process in $\mathbf{L}^2(W)$ would have rank smaller than q, which is false.

Consider now $W' = B^* B$, a positive semi-definite function of order q and full rank. We have

$$\tilde{W}' = \tilde{B}\tilde{B}^* = CC^*, \tag{136}$$

where $C = \tilde{B}$ belongs to $\mathbf{H}^2$ with q rows and p columns. Applying previous results once more, we can find D with at most q columns so that $\tilde{W}' = DD^*$; and D can be chosen to be left-outer. Since $\tilde{W}'$ has full rank D must have exactly q columns and be of full rank. Passing backwards,

$$W' = \tilde{D}^* \tilde{D} = B^* B. \tag{137}$$

Define
$$V = B\tilde{D}^{-1}, \tag{138}$$
a transformation from the q-dimensional domain space of B to $\mathfrak{H}$. From (137) it is obvious that $V^* V$ is the identity matrix of order q. Therefore $V V^*$ is the identity when it acts on the variable subspace $\mathfrak{H}$; in the complement of $\mathfrak{H}$, where B^* is zero, $V^* = (\tilde{D}^{-1})^* B^*$ and so also $V V^*$ vanishes. Therefore V is a partial isometry whose range at e^{ix} is $\mathfrak{H}(e^{ix})$.

Moreover V is analytic. Indeed, $V\tilde{D} = B$, where $\tilde{D}$ is right-outer (and left-outer by Theorem 8, but we do not need that fact) and B is analytic. By Theorem 10, rephrased for right-outer functions, $\tilde{D}Q$ ranges over a dense subset of $\mathbf{H}^2$ as Q varies over the analytic trigonometric polynomials. If $\tilde{D}Q$ approximates I, then BQ approximates V, which consequently must be analytic.

It is not quite easy to prove that V is left-outer, although that is true, but the difficulty can be avoided. Since V is analytic, we can write $V V^*$ as $U U^*$, where U is left-outer with q columns (because the range of $U U^*$ has dimension q); and obviously $U^* U$ is the identity of order q. Therefore U is a left-outer partial isometry whose range is $\mathfrak{H}$, and this completes the proof of the theorem.

11. Factorization theorems

There are two kinds of factorization theorems in prediction theory. The first kind expresses a positive function (or a positive definite matrix function) as a product $|A|^2$ or $A A^*$ (for scalar and matrix functions, respectively), where A is analytic. The definitive theorem of this type for scalar functions was proved by Szegö [18]. Much later Wiener called attention in his writings on prediction theory to the importance of this factorization for the subject [19], and he began the search for extensions to matrix functions [20].

The second kind of factorization theorem has its origin in a representation theorem of Nevanlinna, but really took shape in the well-known paper of Beurling [6]. Now an *analytic* function A is given and is to be written as a product of simpler factors, which since Beurling are called inner factors and outer factors. Lax [14] has extended Beurling's Theorem to matrix functions, giving suitable generalizations of inner and outer functions.

In our first paper we derived again the Wiener Theorem of first kind, and showed how it leads to a theorem of the second kind, although the opposite deduction has not been made. Our method was limited to functions of full rank. Now we shall apply the function-theoretic results obtained in the last pages to reduce the first factorization problem to the same problem for functions of full rank. Our result, Theorem 13, is new. We shall also obtain Lax's factorization of the second kind in a more precise form made possible by our analysis of outer functions.

THEOREM 13. *Let W be a positive semi-definite matrix function with summable entries. In order for W to have the form AA^* for some function A in $\mathbf{H}^2$ it is necessary and sufficient that the range $\mathfrak{H}$ of W be an analytic range function, and that*

$$\int \log \Delta W d\sigma > -\infty, \tag{139}$$

where ΔW is the determinant of W as a transformation on its range.

Proof. The necessity of the first condition is obvious. Suppose then that $W = AA^*$; we have shown that A can be chosen to be left-outer with trivial null-space, and further to have the special form VB, where V is a left-outer partial isometry mapping a fixed space of q dimensions onto the range $\mathfrak{H}$ of W, and B (previously called $\tilde{D}$) is an outer function of order q and full rank. We have then

$$\log \Delta W = \log \det (BB^*), \tag{140}$$

whose integral is finite by Theorem 8. (We use the fact that B_0 is non-singular if B is outer and of full rank.) Therefore the necessity of (139) is established.

Now assume the conditions of the theorem are satisfied. Let V be the left-outer partial isometry which maps a fixed space of q dimensions onto the range of W. Then $V^* W V$ is a positive semi-definite operator in q-space, and by (139)

$$\int \log \det (V^* W V) d\sigma > -\infty. \tag{141}$$

Therefore the factorization theorem for functions of full rank [10, 20] shows that $V^* W V$ has the form BB^*, for some analytic (in fact outer) function B of order q and full rank. We conclude that $W = (VB)(VB)^*$, as we had to prove.

Our factorization theorem of the second kind is this:

THEOREM 14. *Each function A in $\mathbf{H}^2$ (with N rows and p columns) has a representation VBU, where V is a left-outer partial isometry, U is a right-outer partial isometry, and B is a non-singular square analytic function of order q equal to the rank of A. Every other such factorization has the form $(VT)(T^* BR)(R^* U)$ for some constant unitary matrices T and R.*

Proof. Find left-outer partial isometries V and $\tilde{U}$ having the same range as A and $\tilde{A}$, respectively. These functions can be chosen to have q columns, but not fewer. Then U is right-outer, and it annihilates the same subspace as A at each point. If we set $B = V^* A U^*$, then $VBU = A$. Indeed, this fact is obvious on the null-space of A, where U vanishes as well; and on the complement of that space it follows from the definition of B. We have to show now that B is analytic.

Because V and U are outer in their respective senses it follows that $V_0 B U_0$ is analytic, and so also $(V_0^* V_0) B (U_0 U_0^*)$. Now V has trivial null-space, so that Theorem 9 implies that $V_0^* V_0$ is non-singular. The same holds for $U_0 U_0^*$. Therefore we can cancel the constant factors to conclude that B is analytic.

The unicity of the representation is a consequence of Theorem 11.

The theorem just proved exhibits an arbitrary function A in $\mathbf{H}^2$ as a product of outer functions with a *kernel* B, which is an arbitrary square analytic function of full rank. The structure of kernels has been studied before. We can, however, write down the main result ([10, p. 195; 14] easily, and we include it for completeness.

DEFINITION. *An* inner function *is an analytic function whose values are unitary matrices.*

By definition, an inner function is square and of full rank. Theorem 11 implies that a function which is both inner and outer is constant.

THEOREM 15. *Each function A which is square and of full rank in $\mathbf{H}^2$ has the form CD, where C is outer and D is inner.*

Proof. There is an outer function C such that $CC^* = AA^*$. Let $D = C^{-1}A$. Then $A = CD$ as required, and we only have to prove that D is inner. Evidently

$$DD^* = C^{-1}AA^*(C^*)^{-1} = I, \tag{142}$$

so that D is unitary. Each function $QA = QCD$ (where Q is an analytic trigonometric polynomial) is analytic; if QC approximates the identity, which is possible because C is outer, then QA approximates D at least in the norm of H^1, so that D is analytic. This completes the proof.

12. Analytic range functions

Theorem 13 suggests the importance of characterizing in some independent way those range functions $\mathfrak{H}$ which are analytic. We have no such characterization, but in this section we present further results on left-outer partial isometries and their range functions.

THEOREM 16. *Let A belong to $\mathbf{H}^2$ with N rows and p columns, and have rank p. Let V be the left-outer partial isometry with p columns onto the range of A. Then we have*

$$\int \log \det (A^*A)\, d\sigma \geqslant \log \det (A_0^* A_0) - \log \det (V_0^* V_0) \tag{143}$$

with equality if and only if A is left-outer.

Proof. The null-space of A is trivial, and therefore the factor U does not appear in the representation for A given by Theorem 14. We have then $A = VB$, where B is square, with order p and full rank. Since V is a partial isometry, $A^*A = B^*B$, and

$$\int \log \det (A^*A)\, d\sigma = \int \log \det (B^*B)\, d\sigma \geqslant \log \det (B_0^* B_0)$$
$$= \log \det (A_0^* A_0) - \log \det (V_0^* V_0). \quad (144)$$

(The inequality is valid for square analytic functions of full rank [10], and the last equality holds because $A_0 = V_0 B_0$.) Now A is outer if and only if B is outer; and the inequality of (144) becomes equality precisely in that case, by Theorem 8. This proves the theorem.

Theorem 16 is a generalization of Theorem 8 to singular analytic functions. The quantity

$$-d = \log \det (V_0^* V_0) \quad (145)$$

is negative unless V is constant, because

$$1 = \int \operatorname{tr} (V^* V)\, d\sigma = \sum \operatorname{tr} (V_n^* V_n), \quad (146)$$

and by the inequality for the geometric and arithmetic means

$$[\det (V_0^* V_0)]^{1/p} \leqslant \operatorname{tr} (V_0^* V_0). \quad (147)$$

Thus perhaps d measures the deviation of the range of V from constancy. This idea finds content in the following result, in which however the trace function appears instead of the determinant:

THEOREM 17. *Let $\mathfrak{H}$ be an analytic range function. Then*

$$\inf_A \| I - A \|^2 = 1 - \operatorname{tr} (V_0 V_0^*), \quad (148)$$

where I is the identity of order N, V is the left-outer partial isometry with range $\mathfrak{H}$ and trivial null-space, and A ranges over the square functions of $\mathbf{H}^2$ of order N having range in $\mathfrak{H}$.

Proof. Suppose the dimension of $\mathfrak{H}$ is p, so that V has p columns. Then every function A with range in $\mathfrak{H}$ has the form VB, where B belongs to $\mathbf{H}^2$ with p rows and N columns. Let us write $I = VV^* + J$, where VV^* is the orthogonal projection on $\mathfrak{H}$, and J the projection on the orthogonal complement of $\mathfrak{H}$. Since the range of $VV^* - A$ is orthogonal to that of J,

$$\| I - A \|^2 = \| VV^* - VB \|^2 + \| J \|^2. \quad (149)$$

Now J is a projection on a subspace with dimension $N - p$, and it follows that $\|J\|^2 = (N-p)/N$. For the other term we have

$$\int \operatorname{tr}\,[V(V^* - B)(V - B^*)\,V^*]\,d\sigma = \int \operatorname{tr}\,[(V - B^*)(V^* - B)]\,d\sigma, \tag{150}$$

where the factors have been permuted in the bracket on the left, and V^*V omitted as redundant, in order to obtain the right side. From the Parseval equality (108) it is obvious that the analytic function B approximates the conjugate-analytic function V^* best when $B = V_0^*$, a constant function. Therefore the minimal function A in (148) is $V V_0^*$. Evaluating (150) with this choice of B gives $p/N - \operatorname{tr}(V_0 V_0^*)$. We add this to $\|J\|^2 = (N-p)/N$ to obtain the right side of (148).

Theorem 17 presents the outer partial isometries as solutions of a minimal problem. It is possible that one could obtain them by this property in the first place, and thus develop the subject from a different point of view.

In the statement of Theorem 17 it is not necessary to assume that $\mathfrak{H}$ is an *analytic* range function. There is still a minimal function $A = V V_0^*$, where the range of V is contained in $\mathfrak{H}$, and deviates least from it in the metric sense described by the theorem.

Even obvious questions about analytic range functions lead to difficult function-theoretic problems. For example, we do not know which analytic range functions have complements of the same type. We believe that more is to be said in this subject.

References

[1]. N. I. Achieser, *Vorlesungen über Approximationstheorie*. Berlin, 1953.

[2]. R. Arens, A Banach algebra generalization of conformal mappings of the disc. *Trans. Amer. Math. Soc.* 81 (1956), 501–513.

[3]. ——, The boundary integral of log $|\Phi|$ for generalized analytic functions. *Trans. Amer. Math. Soc.* 86 (1957), 57–69.

[4]. R. Arens & I. M. Singer, Function values as boundary integrals. *Proc. Amer. Math. Soc.* 5 (1954), 735–745.

[5]. ——, Generalized analytic functions. *Trans. Amer. Math. Soc.* 81 (1956), 379–393.

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

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

[8]. O. Hanner, Deterministic and non-deterministic stationary random processes. *Ark. Mat.* 1 (1950), 161–177.

[9]. H. Helson, Conjugate series and a theorem of Paley. *Pacific J. Math.* 8 (1958), 437–446.

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

[11]. ——, Invariant subspaces. *Proc. International Symposium on Linear Spaces*, Jerusalem, 1960 (to appear).

[12]. K. Hoffman, Boundary behavior of generalized analytic functions. *Trans. Amer. Math. Soc.* 87 (1958), 447–466.

[13]. K. Karhunen, Über die Struktur stationärer zufälliger Funktionen. *Ark. Mat.* 1 (1950), 141–160.

[14]. P. D. LAX, Translation invariant subspaces. *Acta Math.* 101 (1959), 163–178.
[15]. P. MASANI, The prediction theory of multivariate stochastic processes, III. *Acta Math.* 104 (1960), 141–162.
[16]. P. MASANI & N. WIENER, The prediction theory of multivariate stochastic processes. *Acta Math.* 98 (1957), 111–150; *idem*, II, *Acta Math.* 99 (1958), 93–137.
[17]. ——, On bivariate stationary processes and the factorization of matrix-valued functions. *Teoria veroyatnosti i ee primenenie* (*Moscow*) 4 (1959), 322–331.
[18]. G. SZEGÖ, Über die Randwerte einer analytischen Funktion. *Math. Ann.* 84 (1921), 232–244.
[19]. N. WIENER, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* New York, 1949.
[20]. ——, On the factorization of matrices. *Comment. Math. Helv.* 29 (1955), 97–111.
[21]. H. WOLD, *A Study in the Analysis of Stationary Time Series.* Uppsala, 1938.

Received February 10, 1961