The Annals of Probability
1988, Vol. 16, No. 4, 1854–1859

ON THE WEINER–MASANI ALGORITHM FOR FINDING THE GENERATING FUNCTION OF MULTIVARIATE STOCHASTIC PROCESSES[1]

BY A. G. MIAMEE

Hampton University

It is shown that the algorithms for determining the generating function and prediction error matrix of multivariate stationary stochastic processes developed by Weiner and Masani (1957, 1958) and later by Masani (1960) will work in some more general setting.

1. Introduction. In their papers [3] and [4] Wiener and Masani developed some algorithms for determining the generating function, the prediction error matrix and an autoregressive representation for the linear least-squares predictor of a multivariate stationary stochastic process $\mathbf{f}_n$, $-\infty < n < \infty$. Their algorithms were obtained under the requirement that the spectral distribution matrix $\mathbf{F}$ of the process $\mathbf{f}_n$ is absolutely continuous with density $\mathbf{f}$ and

(1) $$c \le \mathbf{f}(e^{i\theta}) \le d,$$

where c and d are two positive numbers. Later Masani [2] showed that their algorithms for determining the generating function and the prediction error matrix work under some weaker condition, namely,

(2) $$\mathbf{f}^{-1} \in \mathbf{L}_1 \quad \text{and} \quad \mu/\lambda \in L_1,$$

where μ and λ are the largest and smallest eigenvalues of $\mathbf{f}$, respectively, and that the autoregressive series for the predictor converges under the condition

(3) $$\mathbf{f}^{-1} \in \mathbf{L}_1 \quad \text{and} \quad \mathbf{f} \in \mathbf{L}_\infty,$$

which is stronger than (2) but still weaker than (1).

In the present paper we show that Wiener and Masani's algorithms for finding the generating function and the prediction error matrix can be adjusted to work when the spectral density $\mathbf{f}$ can be factored as

$$\mathbf{f} = \mathbf{P}(e^{i\theta})\mathbf{g}(e^{i\theta})\mathbf{P}^*(e^{i\theta}),$$

where $\mathbf{g}$ is a new spectral density satisfying Masani's condition (2) mentioned previously and $\mathbf{P}$ is a certain kind of polynomial (for precise conditions see our assumption and theorem in Section 3).

2. Preliminaries. Let $(\Omega, \mathscr{F}, P)$ be a probability space and H denote the Hilbert space $L_0^2(\Omega, \mathscr{F}, P)$ of all complex-valued random variables on Ω with

Received July 1984; revised November 1987.
[1]This work has been partially supported by NASA Grant NAG-1-768.
AMS 1980 *subject classifications*. Primary 60G10, 60G25; secondary 62M10.
Key words and phrases. Multivariate stationary stochastic process, spectral density, generating function, prediction error matrix.

zero expectation and finite variance with the usual inner product (,) and norm $\| \;\|$.

Following [3] for $q \geq 1$, H^q denotes the Cartesian product of q copies of H, i.e., the set of all column vectors $\mathbf{f} = (f^1, \ldots, f^q)^T$, with $f^i \in H$. For $\mathbf{f}$ and $\mathbf{g}$ in H^q their *Gramian matrix* $(\mathbf{f}, \mathbf{g})$ is defined to be the $q \times q$ matrix $[(f^i, g^j)]$. It is well known [3] that H^q is a Hilbert space under the inner product $((\mathbf{f}, \mathbf{g})) = \operatorname{trace}(\mathbf{f}, \mathbf{g})$ and norm $\|\mathbf{f}\| = ((\mathbf{f}, \mathbf{f}))$, provided their linear combinations are formed with matrix coefficients. Two elements $\mathbf{f}$ and $\mathbf{g}$ in H^q are said to be *orthogonal* (denoted by $\mathbf{f} \perp \mathbf{g}$) if $(\mathbf{f}, \mathbf{g}) = 0$, which is the same as saying $f^i \perp g^j$, for all i, j.

Now we introduce a few concepts and state a theorem which is essential for our study here. For the detail of these and other standard concepts of the prediction theory of multivariate stochastic processes which are used but not formally presented here see [2]–[4].

A bisequence $\{\mathbf{f}_n, -\infty < n < \infty\} \subset H^q$ is said to be a *q-variate stationary* stochastic process (SSP) if their Gramian $(\mathbf{f}_m, \mathbf{f}_n)$ depends only on $m - n$. It can be shown that for such a process one has a spectral representation of the form

$$(4) \qquad (\mathbf{f}_m, \mathbf{f}_n) = (1/2\pi) \int_0^{2\pi} e^{-i(m-n)\theta} \mathbf{F}(d\theta),$$

where $\mathbf{F}$ is a countably additive nonnegative matrix-valued measure called the *spectral distribution* of $\mathbf{f}_n$.

For the q-variate SSP $\mathbf{f}_n$, we define its time domain $\mathbf{M}(+\infty) = \overline{\mathrm{SP}}\{\mathbf{f}_k: -\infty < k \leq \infty\}$, its past $\mathbf{M}(n) = \overline{\mathrm{SP}}\{\mathbf{f}_k: -\infty < k \leq n\}$ and its remote past $\mathbf{M}(-\infty) = \bigcap_n \mathbf{M}(n)$. The process is called *nondeterministic* or *purely nondeterministic* according as $\mathbf{M}(-\infty) \neq \mathbf{M}(+\infty)$ or $\mathbf{M}(-\infty) = \mathbf{0}$. In case $\mathbf{F}$ is absolutely continuous with respect to the Lebesgue measure, its spectral density is given by $\mathbf{f}(\theta) = d\mathbf{F}/d\theta$. In this case the spectral domain of the process denoted by $\mathbf{L}^2(\mathbf{f})$ is defined by $\mathbf{L}^2(\mathbf{f}) = \{\mathbf{\Phi}: \mathbf{\Phi}$ is a $q \times q$ matrix-valued function with $\operatorname{trace} \int \mathbf{\Phi}(\theta)\mathbf{f}(\theta)\mathbf{\Phi}^*(\theta)\, d\theta < \infty\}$. It is well known [3] that $\mathbf{L}^2(\mathbf{f})$ is an inner product space with inner product given by $((\phi, \psi)) = \operatorname{trace}(\mathbf{\Phi}, \mathbf{\Psi})$, where $(\mathbf{\Phi}, \psi) = \int \mathbf{\Phi}(\theta)\mathbf{f}(\theta)\mathbf{\Psi}^*(\theta)\, d\theta$. Now if we consider the map $\mathbf{I}$ sending $\mathbf{f}_n$ to $e^{-in\theta}$, one can see by (4) that $(\mathbf{f}_n, \mathbf{f}_m) = (\mathbf{I}(\mathbf{f}_n), \mathbf{I}(\mathbf{f}_m))$. The well-known Kolmogorov isomorphism theorem proves that this map $\mathbf{I}$ can be extended to an isometric isomorphism between the time domain $\mathbf{M}(+\infty)$ and the spectral domain $\mathbf{L}^2(\mathbf{f})$. Isometric being in the sense that $(\mathbf{g}, \mathbf{f}) = (\mathbf{I}(\mathbf{g}), \mathbf{I}(\mathbf{f}))$ and $((\mathbf{g}, \mathbf{f})) = ((\mathbf{I}(\mathbf{g}), \mathbf{I}(\mathbf{f})))$ for every $\mathbf{g}, \mathbf{f}$ in the time domain.

The *innovation process* $\mathbf{g}_n$ of a multivariate SSP $\mathbf{f}_n$ is defined by $\mathbf{g}_n = \mathbf{f}_n - (\mathbf{f}_n | \mathbf{M}(n-1))$, where $(\mathbf{f}_n | \mathbf{M}(n-1))$ denotes the orthogonal projection of $\mathbf{f}_n$ on $\mathbf{M}(n-1)$. It is easy to see that $(\mathbf{g}_m, \mathbf{g}_n) = \delta_{nm}\mathbf{G}$, where $\mathbf{G} = (\mathbf{g}_0, \mathbf{g}_0)$ is called the prediction error matrix. The process $\mathbf{f}_n$ is nondeterministic iff $\mathbf{G} \neq \mathbf{0}$ and it is called *nondeterministic of full rank* if $\mathbf{G}$ is nonsingular, which is in turn equivalent to the requirement $\log \Delta\mathbf{f} \in L^1$. If the process $\mathbf{f}_n$ is nondeterministic of full rank, then one can see from the Wold decomposition theorem that we have the moving average representation

$$\mathbf{f}_n = \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{h}_{n-j}, \quad \text{with } \sum_{j=0}^{\infty} |\mathbf{C}_j|_E^2 < \infty.$$

Then the function $\mathbf{\Phi}(e^{i\theta}) = \sum_{j=0}^{\infty} \mathbf{C}_j e^{ij\theta}$ is the so-called *generating function* of our process. From the inequality $\sum_{j=0}^{\infty} |\mathbf{C}_i|_E^2 < \infty$, it is clear that each entry of $\mathbf{\Phi}$ belongs to L^2, a fact that we express by $\mathbf{\Phi} \in \mathbf{L}_2$. Furthermore, the negative frequencies of $\mathbf{\Phi}$ are $\mathbf{0}$, hence $\mathbf{\Phi} \in \mathbf{L}_2^{0+}$. Finally, one can easily see that $\mathbf{f}$ admits the factorization $\mathbf{f} = \mathbf{\Phi}\mathbf{\Phi}^*$.

We close this section by stating the following uniqueness theorem due to Wiener and Masani which is needed in the next section.

UNIQUENESS THEOREM ([3], Theorem 8.12). *If a matricial spectral density function* $\mathbf{f}$ *has a factor* $\mathbf{\Phi}$ *in* $\mathbf{L}_2^{0+}$ *of the form*

$$\mathbf{f}(\theta) = \mathbf{\Phi}(\theta)\mathbf{\Phi}^*(\theta)$$

such that $\mathbf{\Phi}^{-1} \in \mathbf{L}_2^{0+}$ *and* $\mathbf{\Phi}_+(0) > \mathbf{0}$, *then* $\mathbf{\Phi}$ *is unique.*

3. Determination of the generating function and the prediction error matrix. As we mentioned in the Introduction, Masani [2] found a series representation for the generating function and hence the prediction error matrix when the spectral density function $\mathbf{f}$ of the SSP satisfies condition (2). Thus, he obtained an algorithm for finding the generating function and the prediction error matrix. We start with a SSP $\mathbf{f}_n$ whose spectral density $\mathbf{f}$ does not necessarily satisfy condition (2) but can be factored as $\mathbf{f} = \mathbf{P}\mathbf{g}\mathbf{P}^*$, where $\mathbf{P}$ is a special kind of polynomial and $\mathbf{g}$ is a new spectral density which does satisfy (2). Thus, one can use Masani's technique to get the generating function and the predictor error matrix corresponding to $\mathbf{g}$. Then we will apply our theorem to get the generating function and the predictor error matrix of the SSP corresponding to $\mathbf{f}$.

To be more precise let us state our assumptions.

ASSUMPTION. We assume that $\mathbf{f}_n$ has an absolutely continuous spectral distribution with density $\mathbf{f}$ such that

$$\mathbf{f}(e^{i\theta}) = \mathbf{P}(e^{i\theta})\mathbf{g}(e^{i\theta})\mathbf{P}^*(e^{i\theta}),$$

where

(a) $\mathbf{P}$ is an optimal polynomial with $\mathbf{P}_+(0) = \mathbf{I}$,

(b) $\mathbf{g}$ is a new density with $\mathbf{g}, \mathbf{g}^{-1} \in \mathbf{L}_1$, and

(c) if $\lambda(e^{i\theta})$ and $\mu(e^{i\theta})$ are the smallest and largest eigenvalues of $\mathbf{g}(e^{i\theta})$, then $\mu/\lambda \in L_1$.

Note that if $\mathbf{f}$ satisfies our assumption, then the conditions on $\mathbf{g}$ are exactly those in (2), which was required by Masani in [2]. Hence, using the algorithm developed there, one can compute the generating function $\mathbf{\Psi}$ and prediction error matrix $\mathbf{K}$ of the process corresponding to $\mathbf{g}$. Using the following theorem, one can find the generating function $\mathbf{\Phi}$ and the prediction error matrix $\mathbf{G}$ of our process $\mathbf{f}_n$.

THEOREM. *Let the q-variate SSP* $\mathbf{f}_n$ *satisfy the previous assumption. Then:*

(a) $\mathbf{f}_n$ *is purely nondeterministic of full rank.*

(b) *If* $\mathbf{\Phi}$ *and* $\mathbf{\Psi}$ *are the generating functions corresponding to the spectral densities* $\mathbf{f}$ *and* $\mathbf{g}$, *and* $\mathbf{G}$ *and* $\mathbf{K}$ *are their prediction error matrices; one has* $\mathbf{\Phi} = \mathbf{P\Psi}$ *and* $\mathbf{K} = \mathbf{G}$.

PROOF. (a) From our assumption one can write

$$\log \Delta \mathbf{f} = 2\log|\Delta \mathbf{P}| + \log \Delta \mathbf{g}.$$

Since the spectral density $\mathbf{g}$ and its inverse $\mathbf{g}^{-1}$ are in $\mathbf{L}_1$ the corresponding SSP is full rank minimal, and therefore nondeterministic of full rank [2, 2.8 and 2.5]. So, $\log \Delta \mathbf{g} \in L_1$. Also since $\Delta \mathbf{P}$ is a nonzero polynomial $\log|\Delta \mathbf{P}| \in L_1$. Hence $\log \Delta \mathbf{f} \in L_1$. Therefore $\mathbf{f}_n$ has full rank and is purely nondeterministic of full rank.

(b) On one hand we have $\mathbf{g} = \mathbf{\Psi\Psi}^*$ and on the other hand we can factor $\mathbf{g}$ as

$$\mathbf{g} = \mathbf{P}^{-1}\mathbf{f}\mathbf{P}^{*-1} = \mathbf{P}^{-1}\mathbf{\Phi\Phi}^*\mathbf{P}^{*-1} = (\mathbf{P}^{-1}\mathbf{\Phi})(\mathbf{P}^{-1}\mathbf{\Phi})^*. \tag{5}$$

In order to show that $\mathbf{\Phi} = \mathbf{P\Psi}$, or equivalently $\mathbf{P}^{-1}\mathbf{\Phi} = \mathbf{\Psi}$, we appeal to the uniqueness theorem presented in Section 2. To do this, we first show that

$$\mathbf{\Psi}, \mathbf{\Psi}^{-1}, \mathbf{P}^{-1}\mathbf{\Phi}, (\mathbf{P}^{-1}\mathbf{\Phi})^{-1} \in \mathbf{L}_2^{0+}.$$

Since by our assumption $\mathbf{g}^{-1}, \mathbf{g} \in \mathbf{L}_1$ and $\mathbf{\Psi}$ is the generating function of $\mathbf{g}$, we know that [cf. Lemma 2.1(c) in [2]] functions $\mathbf{\Psi}, \mathbf{\Psi}^{-1}$ are in $\mathbf{L}_2^{0+}$. Now let $\mathbf{P}(e^{i\theta}) = \mathbf{I} + \mathbf{E}_1 e^{i\theta} + \cdots + \mathbf{E}_k e^{ik\theta}$. From our assumption $\mathbf{P}^*\mathbf{f}^{-1}$ belongs to $\mathbf{L}_2(\mathbf{f})$ so that it has an isomorph χ in $\mathbf{M}(+\infty)$. By the Kolomogorov isomorphism we have

$$(\chi, \chi) = (1/2\pi)\int_0^{2\pi} \mathbf{P}^*(e^{i\theta})\mathbf{f}^{-1}(e^{i\theta})\mathbf{P}(e^{i\theta})\, d\theta \tag{6}$$

and

$$\begin{aligned}(\chi, \mathbf{f}_n) &= (1/2\pi)\int_0^{2\pi} \mathbf{P}^*(e^{i\theta})\mathbf{f}^{-1}(e^{i\theta})\mathbf{f}(e^{i\theta})e^{in\theta}\, d\theta \\ &= (1/2\pi)\int_0^{2\pi} \mathbf{P}^*(e^{i\theta})e^{in\theta}\, d\theta.\end{aligned}$$

Since we have $\mathbf{P}(e^{i\theta}) = \mathbf{I} + \mathbf{E}_1 e^{i\theta} + \mathbf{E}_2 e^{2i\theta} + \cdots + \mathbf{E}_k e^{ik\theta}$, we get

$$(\chi, \mathbf{f}_n) = \begin{cases} \mathbf{E}_n^*, & \text{if } n = 1, \ldots, k, \\ \mathbf{I}, & \text{if } n = 0, \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{7}$$

By (a) our process $\mathbf{f}_n$ is of full rank so that the normalized innovation $\mathbf{h}_n = \mathbf{G}^{-1}\mathbf{g}_n$ is well defined. By (7), $\mathbf{M}(-1) \perp \chi$. But for a purely nondeterministic SSP $\mathbf{f}_n$ which is of full rank, $\mathbf{M}(n) = \overline{\mathrm{SP}}\{\mathbf{h}_k\colon k \le n\}$, for every n, which implies that

$\chi \in \overline{\mathrm{SP}}\{\mathbf{h}_k:\ -\infty < k < \infty\}$ and $\chi \perp \overline{\mathrm{SP}}\{\mathbf{h}_k:\ k \le -1\}$. Since the innovation process $\mathbf{h}_n$ is orthogonal, we get

$$\chi = \sum_{j=0}^{\infty} \mathbf{A}_j \mathbf{h}_j, \qquad \sum_{j=0}^{\infty} |\mathbf{A}_j|_E^2 < \infty. \tag{8}$$

By the moving average representation of Section 2 we have

$$\mathbf{f}_n = \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{h}_{n-j}. \tag{9}$$

(8) and (9) imply $(\chi, \mathbf{f}_n) = \sum_{j=0}^{n} \mathbf{A}_j \mathbf{C}_{n-j}^*$, for each nonnegative integer n. Hence, by (7) we get

$$\sum_{j=0}^{n} \mathbf{A}_j \mathbf{C}_{n-j}^* = \begin{cases} \mathbf{E}_n^*, & \text{if } n = 1, \ldots, k, \\ \mathbf{I}, & \text{if } n = 0, \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{10}$$

Now taking $\mathbf{D}_j$ to be the jth Taylor coefficient of $\Phi_+^{-1}(z)\mathbf{P}(z)$ and noting that

$$\begin{aligned} &\Phi(e^{i\theta})\Phi^{-1}(e^{i\theta})\left(\mathbf{I} + \mathbf{E}_1 e^{i\theta} + \cdots + \mathbf{E}_k e^{ik\theta}\right) \\ &\quad = \mathbf{I} + \mathbf{E}_1 e^{i\theta} + \mathbf{E}_2 e^{2i\theta} + \cdots + \mathbf{E}_k e^{ik\theta}, \end{aligned}$$

we get

$$\sum_{j=0}^{n} \mathbf{C}_{n-j} \mathbf{D}_j = \begin{cases} \mathbf{E}_n, & \text{if } n = 1, \ldots, k, \\ \mathbf{I}, & \text{if } n = 0, \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{11}$$

By taking adjoint from (11) and comparing the result with (10), we get

$$\sum_{j=0}^{n} \mathbf{A}_j \mathbf{C}_{n-j}^* = \sum_{j=0}^{n} \mathbf{D}_j^* \mathbf{C}_{n-j}^*, \quad \text{for all } n \ge 0.$$

Now noting that $\mathbf{C}_0 = \mathbf{G}$ is invertible, a simple inductive argument shows that $\mathbf{A}_n = \mathbf{D}_n^*$, for all $n \ge 0$. This shows that $(\mathbf{P}^{-1}\Phi)^{-1} \in \mathbf{L}_2^{0+}$.

Since $\mathbf{P}$ is a polynomial each entry of $\mathbf{P}^{-1}$ is the quotient of a polynomial to $\Delta(\mathbf{P})$. Now since $\Phi \in \mathbf{L}_2^{0+}$ each of its entries belongs to L_2^{0+}. Thus, each entry $(\mathbf{P}^{-1}\Phi)_{ij}$ of $\mathbf{P}^{-1}\Phi$ has the form

$$\gamma_{ij}(e^{i\theta})/\Delta(\mathbf{P}(e^{i\theta})), \quad \text{with } \gamma_{ij} \in L_2^{0+}. \tag{12}$$

Furthermore, since $\mathbf{g}$ belongs to $\mathbf{L}_1$, from (5) we conclude $\mathbf{P}^{-1}\Phi \in \mathbf{L}_2$, which implies

$$\gamma_{ij}(e^{i\theta})/\Delta(\mathbf{P}(e^{i\theta})) \in L_2, \qquad i, j = 1, \ldots, q. \tag{13}$$

Since by our assumption $\mathbf{P}$ is optimal, one can see that $\Delta(\mathbf{P}(e^{i\theta}))$ is optimal. Using (12) and (13), one can conclude that (cf. [1], page 75)

$$\left(\mathbf{P}^{-1}(e^{i\theta})\Phi(e^{i\theta})\right)_{ij} = \gamma_{ij}(e^{i\theta})/\Delta\left(P(e^{i\theta})\right) \in L_2^{0+},$$

which means $\mathbf{P}^{-1}\Phi$ is also in $\mathbf{L}_2^{0+}$.

Now we check the other requirements of the uniqueness theorem: Note that $(\mathbf{P}^{-1}\boldsymbol{\Phi})_+(0) = \boldsymbol{\Phi}_+(0) = \mathbf{G} > \mathbf{0}$. Now $\boldsymbol{\Psi}_+(0) > \mathbf{0}$ follows from [3], 7.5. Thus, the uniqueness theorem can now be applied to conclude $\boldsymbol{\Psi} = \mathbf{P}^{-1}\boldsymbol{\Phi}$ or $\boldsymbol{\Phi} = \mathbf{P}\boldsymbol{\Psi}$. The last part, namely $\mathbf{K} = \mathbf{G}$, is easy. In fact, $\sqrt{\mathbf{G}} = \boldsymbol{\Phi}_+(0) = (\mathbf{P}\boldsymbol{\Psi})_+(0) = \boldsymbol{\Psi}_+(0) = \sqrt{\mathbf{K}}$. □

REMARK. In conjunction with this work and [2]–[4] there is one more question to be settled, and that is to get an algorithm for finding the best linear predictor.

Acknowledgment. The author wishes to thank the referee for his valuable suggestions which improved the presentation of this article.

REFERENCES

[1] HOFFMAN, K. (1962). *Banach Spaces of Analytic Functions*. Prentice-Hall, Englewood Cliffs, N.J.

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

[3] WIENER, N. and MASANI, P. (1957). The prediction theory of multivariate stochastic processes. I. *Acta Math.* **98** 111–150.

[4] WIENER, N. and MASANI, P. (1958). The prediction theory of multivariate stochastic processes. II. *Acta Math.* **99** 93–137.

DEPARTMENT OF MATHEMATICS
HAMPTON UNIVERSITY
HAMPTON, VIRGINIA 23668