Weiner-Masani Work on Prediction Theory and Harmonic Analysis

V. Mandrekar
Department of Statistics and Probability
Michigan State University

Abstract

This work is an invited work for the book "Connected at Infinity" (ed. R. Bhatia) which consists of articles that explain to non-specialists an important piece of the work by Indian mathematicians and their influence. In this article, we explain the work of Professor Masani (Joint with Norbert Weiner) on multivariate prediction theory which had a major influence in harmonic analysis, operator theory, and factorization problems. To make the work accessible, we first explain the work of Wiener and Kolmogorov in the univariate case starting with the work of H. Wold. We then explain the Wiener-Masani work as a generalization of that of Wiener-Kolmogorov. The references bring readers to some modern work influenced by them.

I want to thank Mr. Wenning Feng for his Tex typing of this work, with important suggestions about the presentation.

Wiener-Masani Work on Prediction Theory and Harmonic Analysis

V. Mandrekar
Department of Statistics and Probability
Michigan State University

Introduction

The work of Wiener-Masani started at ISI during the visit of Norbert Wiener in 1955-56. Wiener [18] had done substantial work in the univariate case and had partial results in the multivariate case. Kolmogorov [7] had studied the univariate case in detail, with emphasis on a fundamental theorem of H. Wold. In addition, Zasuhin [20] had announced partial results. Because of the connection between prediction theory (of interest to Wiener) and factorization of matrix-valued functions (of interest to Wiener [18] p.150 and Masani [12]), their collaboration produced results and techniques which had a lasting effect in the case of prediction theory, analysis and operator theory (Nagy and Foias, see [13])

Finally, it will be incomplete to give the influence of Masani in Mathematics without mentioning his mentoring of students as described in [4]. This led to these students contributing to Mathematics in general.

In the next section, we start by describing the concepts involved in the prediction theory in the univariate case. The material is based on [9].

We give at the end of references, the book of K. Hoffman on analytic functions, as a general reference.

1 Motivation for Wiener-Masani work

In order to understand the significance and motivation for the fundamental work of Wiener-Masani, it is necessary to describe the work of Kolmogorov

We call $\{X_n, n \in \mathbb{Z}\}$ purely non-deterministic (regular) if $L(X : -\infty) = \{0\}$ and deterministic (singular) if $L(X : -\infty) = L(X : n)$ for all n. Now if X_n is regular, writing

$$X_0 = \sum_{k=-\infty}^{0} a_k \nu_k$$

$$X_n = \sum_{k=-\infty}^{0} a_k U^n \nu_k = \sum_{k=-\infty}^{0} a_k \nu_{n+k} = \sum_{k'=-\infty}^{n} a_{k'-n} \nu_{k'}$$

Major part of Kolmogorov-Wiener work is to get analytic conditions for a process to be regular or singular. This is done by using Bochner Theorem to obtain

$$r(n) = \int_{-\pi}^{\pi} e^{in\lambda} \mathrm{d}F(\lambda)$$

We call $F(\lambda)$ the spectral measure of $\mathbf{X}$, $f(\lambda)$, the density of F *w.r.t.* Lebesgue measure σ, the spectral density and $L^2(F) = L^2((-\pi, \pi], \mathcal{B}(-\pi, \pi], F)$ is called spectral domain. The map $V : X_n \to e^{in\cdot}$ can be extended to a unitary operator from time domain $L(X)$ onto spectral domain $L^2(F)$. One then gets that $\{X_n\}$ is purely non-deterministic iff $f(\lambda) = |\varphi(\lambda)|^2$ where $\hat{\varphi}(k) = 0$ for $k < 0$. If $X_n = \sum_{k=0}^{\infty} a_k \nu_{n-k}$, then $\varphi(\lambda) = \sum_{k=0}^{\infty} \overline{a}_k e^{ik\lambda}$.

If $\mathbf{X}$ is singular, then F is singular *w.r.t.* Lebesgue measure. Thus the Wold decomposition is equivalent to Lebesgue decomposition of F. Let us consider the map from $L^2(T)$ to $H^2(T) = \{\varphi \in L^2(T) : \hat{\varphi}(k) = 0, k < 0\}$ as

$$\varphi = \sum_{n=-\infty}^{\infty} \hat{\varphi}(n) e^{in\lambda} \longmapsto (\varphi)_+ = \sum_{n=0}^{\infty} \hat{\varphi}(n) e^{in\lambda}$$

We observe that $\varphi \in H^2(T)$ does not vanish on a set of positive Lebesgue measure without being identically zero. Now if $f(\lambda) = |\varphi(\lambda)|^2$, $\varphi \in H^2(T)$, then we get that

$$X_n = V^{-1}(e^{in}\varphi(\cdot))$$

gives a stationary process with spectral density f and $\xi_k = V^{-1}(e^{ik})$ giving

$$X_n = \sum_{k=-\infty}^{0} \hat{\varphi}(k) \xi^{\varphi}_{n-k}$$

Now RHS above equals using the map V, $\sigma^2 = E|X_0 - P_{L(X:-1)}X_0|^2$, the prediction error. Since Szegö gives

$$\varphi(z) = \exp\left\{\frac{1}{2}\int_0^{2\pi} \log f(\theta)\frac{e^{i\theta}+z}{e^{i\theta}-z}\mathrm{d}\sigma\right\}$$

we get $|\varphi(0)|^2 = E|X_0 - P_{L(\xi^\varphi:-1)}X_0|^2$. As in general $L(X:-1) \subseteq L(\xi^\varphi:-1)$, we get

$$\sigma^2 \geq |\varphi(0)|^2$$

and φ is maximal, one gets $\sigma^2 = |\varphi(0)|^2$. Thus we get the result of Kolmogorov [8].

Theorem K1. Let $\{X_n, n \in \mathbb{Z}\}$ be a weakly stationary process, then

a) $\sigma^2 = \exp\left\{\int_0^{2\pi} \log f_a(\theta)\mathrm{d}\sigma\right\}$ where f_a is the density of spectral measure F *w.r.t.* σ

b) X is deterministic iff $\int_0^{2\pi} \log f_a(\theta)\mathrm{d}\sigma = -\infty$.

In order to solve the prediction problem, one wants to express ν_n, the innovations, in terms of $\{X_n, n \in \mathbb{Z}\}$. That is done in the next result.

Theorem K2. Let $\{X_n, n \in \mathbb{Z}\}$ be a weakly stationary purely non-deterministic process. Let f be its density and assume $f \in L^\infty(\mathrm{d}\sigma)$. Then

$$\nu_n = \sum_{k=0}^{\infty} d_k X_{n-k} \quad \text{with} \quad \sum_k d_k^2 < \infty$$

iff

$$f^{-1} \in L^1(\sigma)$$

and $d_k = \int_0^{2\pi} e^{-ik\lambda}\frac{1}{\varphi(\lambda)}\mathrm{d}\lambda$ where φ is maximal factor.

As one can see from the review of Kolmogorov-Wiener work in one-dimension, the solving of the problem requires combinations of methods from analytic function theory, geometry of time domain of stationary stochastic processes, isometry between the time and spectral domain. It turns out that, in multivariate case studied by Wiener and Masani, the time domain is a Hilbert module and the spectral domain involves square integrable matrix-valued functions *w.r.t.* a non-negative definite matrix-valued measure.

As in the univariate case, let us define $\mathcal{M}_n = \mathcal{M}(\mathbf{f} : n) = \overline{\text{sp}}\{\mathbf{f}_k, k \leq n\}$, $\mathcal{M}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{M}_n$. We know that $U\mathbf{f}_n^i = \mathbf{f}_{n+1}^i$ is a unitary operator on H and we can write

$$\boldsymbol{U}(\mathbf{f}_n) = (U\mathbf{f}_n^i)_{i=1}^q = \mathbf{f}_{n+1}$$

Then we get $\boldsymbol{U}\mathcal{M}_n = \mathcal{M}_{n+1}$ and hence just as in univariate case we get $\boldsymbol{U}$ acting projection on $\mathcal{M}_n$ equals projection of $\mathcal{M}_{n+1}$ acting on $\boldsymbol{U}$. Define $\mathbf{g}_n = \mathbf{f}_n - (\mathbf{f}_n|\mathcal{M}_{n-1})$ and $\mathcal{W}_n = \overline{\text{sp}}\{\mathbf{g}_n\}$ for each n. Then one gets with $\mathcal{M}_\infty = \overline{\text{sp}}\{\mathbf{f}_n, n \in \mathbb{Z}\}$

(2.2) a) $\mathcal{M}_n = \sum_{k=0}^{\infty} \oplus \mathcal{W}_{n-k} \oplus \mathcal{M}_{-\infty}$

b) $\mathcal{M}_\infty = \sum_{k=-\infty}^{+\infty} \oplus \mathcal{W}_k \oplus \mathcal{M}_{-\infty}$

As in one-dimensional case, we call [19] a process purely non-deterministic if $\mathcal{M}_{-\infty} = \{0\}$ and deterministic if $\mathcal{M}_n = \mathcal{M}_{-\infty}$. Using this, Wiener-Masani prove Wold Decomposition for a stationary process $\{\mathbf{f}_n, n \in \mathbb{Z}\}$.

Wold Decomposition ([19] Thm. 6.11). Let $G = ((\mathbf{g}_0, \mathbf{g}_0))$. Then

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

where $A_k G = ((\mathbf{f}_0, \mathbf{g}_{-k}))$, $A_0 \mathbf{g}_0 = \mathbf{g}_0$ and $\sum_{k=0}^{\infty} |A_k\sqrt{G}|_E^2 < \infty$, $A_0\sqrt{G} = \sqrt{G}$. Note that A_k are not necessarily unique but $A_k G$ and $A_k\sqrt{G}$ are unique.

The above result gives the Wold Decomposition in terms of purely non-deterministic part $\left(\sum_{k=0}^{\infty} A_k \mathbf{g}_{n-k}\right)$ and deterministic part $((\mathbf{f}_n|\mathcal{M}_{-\infty}))$ and the moving average representation of the purely non-deterministic part. The process $\{\mathbf{g}_n, n \in \mathbb{Z}\}$ is stationary and is the innovation process of $\{\mathbf{f}_n, n \in \mathbb{Z}\}$. Earlier, Doob [2] has given this result under full-rank assumption, i.e. G being invertible. The following theorem is a generalization of Kolmogorov Theorem ([7], Thm. 19).

Each of the following conditions are equivalent which can be easily derived from the above theorem by using $\varphi_n = \mathbf{g}_n$ and $K = G$ to get (b)⇒(a) and (a)⇒(b) by using orthogonality $\{\varphi_n\}$ and shift $\boldsymbol{U}\varphi_n = \varphi_{n+1}$ and $\mathcal{M}_n^f \subseteq \mathcal{M}_n^\varphi$. Kolmogorov defines **regularity** of a stationary process by showing $(\mathbf{f}_0|\mathcal{M}_n^f) \to 0$ as $n \to \infty$, which easily follows as $\mathcal{M}_n^f \subseteq \mathcal{M}_n^\varphi$.

Theorem 3.1. (a) The moving average process $(\mathbf{f}_n)_{-\infty}^{\infty}$:

$$\mathbf{f}_n = \sum_{-\infty}^{+\infty} A_k \mathbf{g}_{n-k}, \quad (\mathbf{g}_i, \mathbf{g}_j) = \delta_{ij} G, \quad \sum_{k=-\infty}^{+\infty} |A_k G|_E^2 < \infty$$

has absolutely continuous spectral distribution F such that

$$F'(e^{i\theta}) = \Phi(e^{i\theta}) \cdot \Phi^*(e^{i\theta}), \quad \Phi(e^{i\theta}) = \sum_{k=-\infty}^{+\infty} A_k G^{1/2} e^{ik\theta} \quad a.e.$$

(b) If for this process $A_k = 0$ for $k < 0$, then

$$\Phi(e^{i\theta}) = \sum_{k=0}^{\infty} A_k G^{1/2} e^{ik\theta}$$

and either $\Delta\Phi_+$ vanishes identically or $\log \Delta F' \in L^1$ on C, the complex numbers, and

$$\log \Delta(A_0 G A_0^*) \leq \frac{1}{2\pi} \int_0^{2\pi} \log\left(\Delta(F'(\theta))\right) \mathrm{d}\theta$$

Here $F_+(z) = \sum_{n=0}^{\infty} A_n z^n$ for $|z| < 1$ with $F_+(e^{i\theta})$ as boundary value, and Δ denotes the determinant.

Proof of (a) follows by computing $(\mathbf{f}_n, \mathbf{f}_0)$ and representing it in spectral representation. One then observes that $\sum_{k=-\infty}^{+\infty} |A_k G|_E^2 < \infty$ implies that Φ, $\Phi^* \in L^2$, and uniqueness of Fourier transform for $\Phi\Phi^*$ gives the result. Now Φ being one-sided expandable gives that each of its elements are in Hardy space H^2 on the circle. So $\Delta\Phi_+$ vanishes identically or $\log \Delta F' \in L^1$ and

$$\log |\Delta(A_0 G^{1/2})| = \log |\Delta\Phi_+(0)| \leq \frac{1}{2\pi} \int_0^{2\pi} \log |\Delta\Phi(e^{i\theta})| \mathrm{d}\theta$$

using Szegö Theorem. Observe $\Delta(A_0 G A_0^*) = |\Delta(A_0 G)|^2$ and $\Delta F' = |\Delta\Phi|^2$ to complete the proof of (b).

Using the Wold decomposition and expression for $(\mathbf{f}_n, \mathbf{f}_0) = (\mathbf{u}_n + \boldsymbol{\nu}_n, \mathbf{u}_0 + \boldsymbol{\nu}_0)$, we get that the spectral measure $F = F_u + F_\nu$ where $\{\mathbf{u}_n\}$ is purely nondeterministic and $\boldsymbol{\nu}$ is deterministic. Thus we get $F_u'(e^{i\theta}) = \Phi(e^{i\theta})\Phi^*(e^{i\theta})$, We shall call Φ satisfying the above condition a generating function. This was used by Wiener-Masani to give the form of innovations in terms of the observed process under the boundedness condition

$$\lambda I \prec F(e^{i\theta}) \prec \lambda' I, \quad 0 < \lambda \leq \lambda' < \infty \ \ \forall \theta.$$

giving analogue of Theorem K2 in Section 1 ([19], Part II, Theorem 5.5).

Independent of Wiener-Masani, Helson and Lowdenslager [5] studied the prediction problem giving the right of precedence to Wiener-Masani. They used the technique of invariant subspaces of shift due to Beurling. In a subsequent paper, Masani [11] gave an "elegant and unifying" treatment of the two approaches ([3]) using a generalization of a theorem of Halmos on isometries. This is generalized in ([6], [16]) for random fields and proper generalization of Beurling theorem ([10]) follows. As stated in ([5], p.181), this was a difficult problem and exact analogue was not possible ([14]). But Masani's paper inspired the solution.

For other influences of their work, we refer the reader to *Norbert Wiener Collected Works*, III (ed. Masani) and papers of Salehi [15] and Muhly [13].

Acknowledgements

I am thankful to Professors Bhatia and Ajit Iqbal Singh for giving me the opportunity to write about my Guru, without whom I may not have been a mathematician.

References

[1] Cramér, H., On the theory of stationary random processes, *Ann. of Math.*, **41** (1940), 215-230.

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

[3] Doob, J. L., Review MR0140930 (25#4344), *Math. Reviews.*

[4] Gangolli, R., Some Recollections of Pesi Masani's Years in Bombay (Pesi Masani Volume), *Prediction Theory and Harmonic Analysis*, (Ed. V. Mandrekar and H. Salehi), North Holloand, Amsterdam, 1983.

[5] Helson, H. and Lowdenslager, D., Prediction theory and Fourier analysis in several variables, I, *Acta Math.*, **99** (1958) 165-202, and II, *Acta Math.*, **106** (1961), 175-213.

Acta Math., **99** (1958), 93-137, Part III, *Acta Math.*, **104** (1960), 141-162 (Masani)

[20] Zasuhin, V., On the theory of multidimensional stationary random processes (Russian), *Dokl. Acad. Sci.*, **33** (1941), 435-437.

[21] Hoffman, K., *Banach spaces of analytic functions*, Prentice Hall, Inc. N.J., 1961