The Annals of Probability
1985, Vol. 13, No. 1, 226–233

ON SERIES REPRESENTATIONS FOR LINEAR PREDICTORS

BY PETER BLOOMFIELD[1]

North Carolina State University

The series expressions for the linear predictors of a stationary process have been known for a long time, but necessary and sufficient conditions for the mean square convergence of these series are still not available. It is shown that an equivalent problem is to find necessary and sufficient conditions for the invertibility of the infinite moving average representation of the process. Two known sufficient conditions are discussed, and a more general condition that includes both as special cases is given. The process that arises from fractional differencing of a random walk is discussed as an example.

1. Introduction. Suppose that $\{x(t)\}$ is a discrete time weakly stationary stochastic process. The ν-step linear prediction problem consists of finding a linear combination of $x(t)$, $x(t-1)$, $\cdots$ that is close to $x(t+\nu)$, $\nu > 0$, in the sense of mean squared error.

In practice we have only a finite number of observations $x(t)$, $x(t-1)$, $\cdots$, $x(t-n)$ from which to construct the predictor, and so the minimum mean squared error linear predictor can be found as the linear regression of $x(t+\nu)$ on $x(t)$, $x(t-1)$, $\cdots$, $x(t-n)$. However, finding this optimal predictor generally involves solving an $n+1$ by $n+1$ system of equations, and it is often more convenient to solve the problem as if the entire past were available, and then use an approximation involving only the finite past.

Suppose that the best ν-step predictors based on the finite and infinite past are

$$\textstyle\sum_{r=0}^{n} b_{\nu,r}^{(n)} x(t-r) \quad \text{and} \quad \sum_{r=0}^{\infty} b_{\nu,r} x(t-r)$$

respectively, where we assume that the latter series converges in mean square. Then

$$\begin{aligned} &E\{x(t+\nu) - \textstyle\sum_{r=0}^{\infty} b_{\nu,r} x(t-r)\}^2 \\ &\quad \le E\{x(t+\nu) - \textstyle\sum_{r=0}^{n} b_{\nu,r}^{(n)} x(t-r)\}^2 \le E\{x(t+\nu) - \sum_{r=0}^{n} b_{\nu,r} x(t-r)\}^2 \\ &\quad = E\{x(t+\nu) - \textstyle\sum_{r=0}^{\infty} b_{\nu,r} x(t-r)\}^2 + E\{\sum_{r=n+1}^{\infty} b_{\nu,r} x(t-r)\}^2. \end{aligned}$$

In general the nth partial sum of the infinite predictor is not as good as the best finite predictor. From the above it follows that

$$\begin{aligned} &E\{x(t+\nu) - \textstyle\sum_{r=0}^{n} b_{\nu,r} x(t-r)\}^2 - E\{x(t+\nu) - \sum_{r=0}^{n} b_{\nu,r}^{(n)} x(t-r)\}^2 \\ &\qquad\qquad \le E\{\textstyle\sum_{r=n+1}^{\infty} b_{\nu,r} x(t-r)\}^2 \to 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

and consequently the partial sum is nearly as good as the best predictor when n

---

Received August 1983; revised May 1984.

[1] This work was carried out while the author was on leave from Princeton University at the University of North Carolina, Chapel Hill.

AMS 1980 *subject classifications.* Primary 60G25; secondary 42A20.

Key words and phrases. Linear prediction, mean square convergence, invertibility.

is large. This is the justification for using the truncated infinite predictor, but clearly the argument depends on the mean square convergence of the series representation of the infinite predictor.

2. Prediction and inversion. We assume that $\{x(t)\}$ is purely linearly indeterministic (Anderson, 1971, page 421), and hence possesses an infinite moving average representation

(1) $$x(t) = \sum_{r=0}^{\infty} a_r \varepsilon(t-r),$$

where $\{\varepsilon(t)\}$ is white noise, $a_0 = 1$, and $\sum a_r^2 < \infty$. We note that $\varepsilon(t)$ is the one-step prediction error and hence is uncorrelated with $x(u)$, $u < t$. Suppose that (1) may be inverted to give

(2) $$\varepsilon(t) = \sum_{r=0}^{\infty} b_r x(t-r),$$

where the series on the right-hand side converges in mean square. (In the remainder of this paper, convergence of a series of random variables will always be in mean square). Now (1) implies that

$$E\{\varepsilon(t)x(t)\} = E\{\varepsilon(t)^2\}$$

and (2) implies that

$$E\{\varepsilon(t)^2\} = b_0 E\{\varepsilon(t)x(t)\},$$

and hence $b_0 = 1$. Thus (2) may be rewritten

$$x(t+1) = \varepsilon(t+1) - \sum_{r=1}^{\infty} b_r x(t+1-r).$$

Since $\varepsilon(t+1)$ is uncorrelated with $x(u)$, $u \le t$, it follows that the best one-step predictor of $x(t+1)$ is

(3) $$\hat{x}_1(t+1) = -\sum_{r=1}^{\infty} b_r x(t+1-r).$$

Thus convergence of (2) implies the existence of the convergent representation (3) for the one-step predictor. Conversely, if the one-step predictor $\hat{x}_1(t+1)$ has the mean-square convergent representation (3), then the one-step prediction error satisfies

$$\varepsilon(t+1) = x(t+1) - \hat{x}_1(t+1) = \sum_{r=0}^{\infty} b_r x(t+1-r)$$

with $b_0 = 1$ and the sum convergent in mean square. Thus we have shown that a necessary and sufficient condition for the existence of (3) as a mean square limit is the existence and convergence of (2). In the next section we shall find sufficient conditions for the latter. However, we show first that when (2) converges, we can find convergent representations for all the ν-step predictors, $\nu \ge 1$. This fact was apparently overlooked by Akutowicz (1957), who having found a sufficient condition for the convergence of (2), introduced further assumptions to assure the existence of convergent series representation for the predictors.

THEOREM 1. *The ν-step predictor of $x(t+\nu)$ possesses a convergent series representation for all $\nu \ge 1$ if and only if $\varepsilon(t)$ has the convergent series representation* (2).

Proof. In view of the above remarks, we need only show sufficiency of the convergence of (2). Now (2) implies that for $s > 0$,

$$0 = E\{\varepsilon(t-s)\varepsilon(t)\} = \sum_{r=0}^{s} b_r E\{\varepsilon(t-s)x(t-r)\},$$

while (1) with t replaced by $t - r$ implies that for $0 \le r \le s$,

$$E\{\varepsilon(t-s)x(t-r)\} = a_{s-r}E\{\varepsilon(t-s)^2\}.$$

Thus

$$\sum_{r=0}^{s} b_r a_{s-r} = 0, \quad s > 0.$$

Hence again using (2),

$$\begin{aligned}\sum_{j=1}^{\nu} a_{\nu-j}\varepsilon(t+j) &= \sum_{j=1}^{\nu} a_{\nu-j} \sum_{r=0}^{\infty} b_r x(t+j-r)\\ &= \sum_{j=1}^{\nu} a_{\nu-j}\{\sum_{r=0}^{j-1} b_r x(t+j-r) + \sum_{r=j}^{\infty} b_r x(t+j-r)\}\\ &= \sum_{j=1}^{\nu} a_{\nu-j} \sum_{s=1}^{j} b_{j-s}x(t+s) + \sum_{j=1}^{\nu} a_{\nu-j} \sum_{s=0}^{\infty} b_{j+s}x(t-s)\\ &= \sum_{s=1}^{\nu} x(t+s) \sum_{j=s}^{\nu} a_{\nu-j}b_{j-s} + \sum_{s=0}^{\infty} x(t-s) \sum_{j=1}^{\nu} a_{\nu-j}b_{j+s}.\end{aligned}$$

Now

$$\sum_{j=s}^{\nu} a_{\nu-j}b_{j-s} = \sum_{k=0}^{\nu-s} a_{\nu-s-k}b_k = \begin{cases} 1 & \nu = s \\ 0 & \nu > s, \end{cases}$$

and hence

$$\sum_{j=1}^{\nu} a_{\nu-j}\varepsilon(t+j) = x(t+\nu) - \sum_{s=0}^{\infty} b_{\nu,s}x(t-s),$$

where

$$b_{\nu,s} = -\sum_{j=1}^{\nu} a_{\nu-j}b_{j+s}.$$

Since the left-hand side is uncorrelated with $x(u)$ $u \le t$, it follows that the best ν-step predictor of $x(t+\nu)$ is

$$\hat{x}_\nu(t+\nu) = \sum_{r=0}^{\infty} b_{\nu,r}x(t-r). \tag{4}$$

This is a well-known expression (see, for instance, Wiener, 1981). Our point in rederiving it is to show that the convergence of the series follows from its construction as a finite linear combination of tail sums of convergent series.

3. Inversion. We have just seen that the key question is the convergence of the inversion formula (2). We now use a well known isomorphism to discuss this in terms of the mean convergence of Fourier series.

Let $\mathcal{M}$ be the Hilbert space found by closing the set of finite linear combinations

$$\sum a_t x(t)$$

with respect to the norm associated with the inner product

$$(a, b) = E(a\bar{b}).$$

Then the map $x(t) \mapsto e^{it\theta}$, $-\pi < \theta \le \pi$, from $\mathcal{M}$ into $L^2(w)$ (the space of functions defined on the interval $(-\pi, \pi]$, square-integrable with respect to the spectral density function of $\{x(t)\}$, $w(\theta)$) is an isomorphism (see, for instance, Doob, 1952, page 483). Hence a series such as

$$\sum_{r=0}^{\infty} b_r x(t-r)$$

converges in mean square if and only if the Fourier series

$$\sum_{r=0}^{\infty} b_r e^{i(t-r)\theta}$$

converges in $L^2(w)$.

Helson and Szegö (1960) showed that every $f \in L^2(w)$ has a Fourier series which is convergent in $L^2(w)$ if and only if w has the representation

$$w(\theta) = \exp\{u(\theta) + \tilde{v}(\theta)\}, \tag{5}$$

where $u \in L^\infty$, $v \in L^\infty$ and $\| v \|_\infty < \pi/2$. Hunt, Muckenhoupt and Wheeden (1973) showed that an equivalent condition is that

$$\left(\frac{1}{|I|}\int_I w(\theta)\, d\theta\right)\left(\frac{1}{|I|}\int_I w(\theta)^{-1}\, d\theta\right) \le K \tag{6}$$

for every interval I, for some constant K. (Here "intervals" are allowed to wrap around from $-\pi$ to π, and $|I|$ is the length of the interval.) Thus these equivalent conditions assure that every random variable in $\mathscr{M}$ possesses a convergent series representation.

The simplest functions that satisfy (5) or (6) are those that are bounded and bounded away from 0: $w \in L^\infty$ and $w^{-1} \in L^\infty$. This condition was studied in the multivariate context by Wiener and Masani (1958), who showed directly that it implies the convergence of (2). Akutowicz (1957) also used the condition that $w \in L^\infty$, but showed that the condition $w^{-1} \in L^\infty$ could be replaced by a condition which is easily seen to be equivalent to $w^{-1} \in L^1$. Masani (1960) showed that an analogous condition is sufficient in the vector case. Masani (1981) attributed the scalar version of the result to Wiener and Kallianpur. Now functions satisfying Akutowicz' conditions

$$w \in L^\infty, \quad w^{-1} \in L^1 \tag{7}$$

do not necessarily satisfy (5) or (6). For instance, one easy consequence of (5) is that w and $w^{-1} \in L^p$ for some $p > 1$, (Zygmund, 1968, page 254), while a function satisfying (7) need not have $w^{-1} \in L^p$ for *any* $p > 1$. On the other hand, the function $w(\theta) = (\cos\theta/2)^\alpha$, $-1 < \alpha < 0$, satisfies (5) (Helson and Szegö, 1960), but fails to satisfy (7) by being unbounded.

Thus we have two different sufficient conditions neither of which can be necessary. Below we derive a more general condition which includes both as special cases. However, it is instructive first to examine a proof of the sufficiency of the Wiener–Kallianpur condition, since the proof of our result is very similar.

We first note that the isometry described above maps ε_t into $e^{it\theta}/\overline{h(\theta)}$, where

$$w(\theta) = (\tau^2/2\pi)\,|h(\theta)|^2, \tag{8}$$

τ^2 is the one-step prediction error, and $h(\theta)$ is an outer function in the Hardy space H^2 (see, for instance, Hannan, 1970, Section III.3). We shall not use any of the properties of h except the identity (8). Under the condition $w^{-1} \in L^1$, (8) implies that $1/h \in L^2$, and hence $1/h$ possesses a Fourier series that converges in L^2. But under the condition $w \in L^\infty$, L^2-convergence implies $L^2(w)$-convergence. Note that we have really only used the property that $w/|h|^2$ is bounded. Thus we have proved

THEOREM 2 (Akutowicz, Masani). *If $w \in L^\infty$ and $w^{-1} \in L^1$, then every f satisfying $w|f|^2 \in L^\infty$ has a Fourier series that converges in $L^2(w)$.*

Our main theorem is proved in essentially the same way. It uses as a critical step:

THEOREM 3 (Hunt, Muckenhoupt and Wheeden, 1973). *If w is a nonnegative function on $(-\pi, \pi]$, and $1 < p < \infty$, then every $f \in L^p(w)$ possesses a Fourier series that converges in $L^p(w)$ if and only if w satisfies the condition*

$$A_p\colon \left(\frac{1}{|I|}\int_I w(\theta)\,d\theta\right)\left(\frac{1}{|I|}\int_I w(\theta)^{-1/(p-1)}\,d\theta\right)^{p-1} \le K$$

for all intervals I and for some finite K independent of I.

The condition (6) stated earlier and its consequence are the special case $p = 2$. We can now prove

THEOREM 4. *If the weight function w may be written as $w = w_1 w_2$, where for some p, $1 \le p < \infty$, we have*

(a) *$w_1^{-1} \in L^p$ and $w_1 \in L^{p'}$, where $1/p' + 1/p = 1$, and*

(b) *w_2^p satisfies A_{2p},*

then every f satisfying $w|f|^2 \in L^\infty$ possesses a Fourier series that converges in $L^2(w)$.

PROOF.

$$\text{(i)} \quad \int |f(\theta)|^{2p} w_2(\theta)^p\,d\theta = \int \{w(\theta)|f(\theta)|^2\}^p w_1(\theta)^{-p}\,d\theta$$

$$\le (\| w|f|^2 \|_\infty)^p . \int w_1(\theta)^{-p}\,d\theta,$$

so $w|f^2| \in L^\infty$ and $w_1^{-1} \in L^p$ imply that $f \in L^{2p}(w_2^p)$.

(ii) Since w_2^p satisfies A_{2p}, $f \in L^{2p}(w_2^p)$ implies that f has a Fourier series that converges in $L^{2p}(w_2^p)$ (Theorem 3).

(iii) For any $g \in L^2(w)$,

$$\int |g(\theta)|^2 w(\theta)\, d\theta = \int |g(\theta)|^2 w_1(\theta) w_2(\theta)\, d\theta$$
$$\le \left\{\int w_1(\theta)^{p'}\, d\theta\right\}^{1/p'} \left\{\int |g(\theta)|^{2p} w_2(\theta)^p\, d\theta\right\}^{1/p}.$$

Thus convergence in $L^2(w)$ is implied by convergence in $L^{2p}(w_2^p)$. This completes the proof.

4. Special cases. Note first that in the case $p = 1$, $p' = \infty$ and we have the factorization $w = w_1 w_2$, where

(a) $w_1^{-1} \in L^1$ and $w_1 \in L^\infty$, and

(b) w_2 satisfies A_2.

Thus the product of a function satisfying (6) with a function satisfying (7) meets the criteria of Theorem 4. Since constant functions satisfy (7) and A_2 (indeed, A_p), we see that both previous sufficient conditions also emerge as special cases.

Another special case of some interest is where $w_2 \equiv 1$. We then have that it is sufficient that $w \in L^{p'}$ and $w^{-1} \in L^p$, for some p, $1 \le p < \infty$.

It is interesting that this simple condition, (6) and (7) all imply that $w^{-1} \in L^1$, or in other words that the process $\{x(t)\}$ is minimal (Rozanov, 1967, page 99). However, not all weight functions satisfying Theorem 4 have this property. The most we can say is that $w^{-1} \in L^q$ for some $q > ½$, which follows from Holder's inequality and the fact that since w_2^p satisfies A_{2p}, it also satisfies A_r for some $r < 2p$, $r > 1$ (Hunt, Muckenhoupt, and Wheeden, 1973) and therefore $w_2^{-p} \in L^{1/(r-1)}$ for this r. For example, $w(\theta) = (\cos \theta/2)^\alpha$ satisfies (5) (and hence A_2) for $-1 < \alpha < 1$ (Helson and Szego, 1960), and satisfies (7) for $0 \le \alpha < 1$. Thus it may be factorized as in Theorem 4 with $p = 1$ for $-1 < \alpha < 2$. (This extends the range of values given by Hosking, 1981, Theorem 1.) For the values $1 \le \alpha < 2$, however, $w^{-1} \notin L^1$, and hence the process is not minimal. Note that for $\alpha \le -1$, $w \notin L^1$ and hence cannot be a spectral density function. Also if $\alpha = 2$, $\varepsilon(t)$ does not have a convergent series representation (Topsoe, 1977). In fact for $\alpha \ge 2$, $1/h \notin L^1$ and hence $1/h$ does not possess a Fourier series. Thus Theorem 4 gives complete information about this example.

Some interesting properties appear when we take into account the fact that A_p functions satisfy integrability conditions. Suppose, for instance, that $w \in A_3$. Then w and $w^{-1/2}$ are both integrable. Thus we can factorize w as $w_1 w_2$ where

(a) $w_1 = w^{1/3}$, whence $w_1^{-1} \in L^{3/2}$ and $w_1 \in L^3$, and

(b) $w_2 = w^{2/3}$, whence $w_2^{3/2}$ satisfies A_3.

Thus the conditions of Theorem 4 are met, with $p = 3/2$.

Since $w \in A_q$, $q \ge 1$, implies that $w \in A_r$ for any $r \ge q$ (Hunt, Muckenhoupt, and Wheeden, 1973), it follows that $w \in A_q$, $1 \le q \le 3$, is a sufficient condition for Theorem 4 to hold, with $p = 3/2$. However, this seems to be the highest value of p for which $w \in A_{2p}$ is enough by itself to satisfy the theorem. If $p > 3/2$, an additional condition is required, such as $w^{-1} \in L^{p-1}$.

It is also interesting to note that a simple consequence of the factorization in Theorem 4 is that

$$\left(\frac{1}{|I|}\int_I w(\theta)\,d\theta\right)\left(\frac{1}{|I|}\int_I w(\theta)^{-1/2}\,d\theta\right)^2 \leq \frac{K}{|I|}$$

for all intervals I and for some finite K independent of I. This may be compared with

$$A_3\colon (1/|I|_I w(\theta)\,d\theta)(1/|I|_I w(\theta)^{-1/2}\,d\theta)^2 \leq K.$$

It follows from the above discussion that these are respectively a necessary and a sufficient condition for the existence of the factorization in Theorem 4.

5. Discussion. Theorem 4 provides a sufficient condition for the convergence of (2), and as we showed in Theorem 1, we then are assured of the convergence of other series such as those of the ν-step linear predictors. However, these other series may not be covered by Theorem 4 directly. For instance, if $w \notin L^\infty$ and the Fourier series of h^{-1} converges in $L^2(w)$, then the Fourier series of $1 - h^{-1}$ also converges. But $w|1 - h^{-1}|^2 \notin L^\infty$, and hence $1 - h^{-1}$ is not covered by Theorem 4. It would be of interest to characterize the random variables in $\mathscr{M}$ that possess convergent representations.

Little is known about necessary conditions for the convergence of (2). Akutowicz (1957) showed essentially that if $w \in L^\infty$, then $w^{-1} \in L$ is necessary and sufficient for the convergence of (2) with $\sum b_r^2 < \infty$. Our example has shown that $w^{-1} \in L$ is *not* necessary merely for the convergence of (2), even if $w \in L^\infty$. Thus we are still far from solving the problem raised by Wiener and Masani (1958), of characterizing the spectra of processes for which (2) converges.

Acknowledgements. I am grateful to the referee for his helpful comments, and for providing the reference to Topsoe (1977).

REFERENCES

AKUTOWICZ, E. J. (1957). On an explicit formula in linear least squares prediction. *Math. Scand.* **5** 261–266.

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

COIFMAN, R. R. and FEFFERMAN, C. (1974). Weighted norm inequalities for maximal functions and singular integrals. *Studia Math.* **51** 241–250.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HELSON, H. and SZEGÖ, G. (1960). A problem in prediction theory. *Ann. Mat. Pura Appl.* **51** 107–138.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.

HUNT, R., MUCKENHOUPT, B. and WHEEDEN, R. (1973). Weighted norm inequalities for the conjugate function and Hilbert transform. *Trans. Amer. Math. Soc.* **176** 227–251.

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

MASANI, P. (1981). Commentary on the prediction-theoretic papers. In *Norbert Wiener: Collected Works*, Vol. III, pages 276–306.

ROZANOV, YU. A. (1967). *Stationary Random Processes*. Holden Day, San Francisco.

TOPSOE, F. (1977). Banach algebra methods in prediction theory. *Manuscripta Math.* **23** 19–55.

WIENER, N. (1981). Prevision statistique et filtrage des ordes. In *Norbert Wiener: Collected Works*, Vol. III, pages 101–108. MIT Press, Cambridge, Mass.

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

ZYGMUND, A. (1977). *Trigonometric Series*, Vol. I. Cambridge University Press, Cambridge.

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27650