The Annals of Statistics
1983, Vol. 11, No. 3, 837–847

CANONICAL CORRELATIONS OF PAST AND FUTURE FOR TIME SERIES: DEFINITIONS AND THEORY

BY NICHOLAS P. JEWELL[1] AND PETER BLOOMFIELD[2]

University of California, Berkeley and Princeton University

The concepts of canonical correlations and canonical components are familiar ideas in multivariate statistics. In this paper we extend these notions to stationary time series with a view to determining the most predictable aspect of the future of a time series. We relate properties of the canonical description of a time series to well known structural properties of the series such as (i) rational spectra (i.e., ARMA series), (ii) strong mixing, (iii) absolute regularity, etc.

1. Introduction. The problem of predicting the future of a weakly stationary time series, knowing its values up to and including the present, is one that arises naturally in many fields. If the series is denoted by $\{x(t)\}$, and the present and past are the values $x(0)$, $x(-1)$, $\cdots$, then we may define the (linear) predictability of a (linear) function of future values, $\sum_{r=1}^{\infty} a_r x(r)$, in terms of the ratio of its minimum mean squared prediction error to its variance. Specifically one can use

$$1 - [\inf_{\{b_r\}} E[\{\textstyle\sum_{r=1}^{\infty} a_r x(r) - \sum_{r=0}^{\infty} b_r x(-r)\}^2]/E[\{\sum_{r=1}^{\infty} a_r x(r)\}^2]]$$

as a measure of the predictability of $\Sigma a_r x(r)$.

The first problem of this type to receive attention was the one-step prediction problem, in which $x(1)$ is the future value to be predicted. The first general result was obtained by Szego (1920, 1921) and later in a more complete form by Kolmogorov and Krein. It states that if the spectral distribution function of $\{x(t)\}$ is $F(\omega)$ then the predictability of $x(1)$ is

$$(1) \qquad 1 - 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log F'(\omega)\, d\omega\right\}$$

where $F'(\omega)$ represents the Radon-Nikodym derivative of the spectral measure dF with respect to Lebesgue measure $d\omega$. A proof of this result may be found in Doob (1953).

It is clear that the one-step predictor may be adapted to provide multistep predictions. We may describe the predictor as the projection of the future value onto the past in a suitable geometry and thus a multistep predictor may be found by successive projections one step at a time. Stated in an alternative way one can thus find the (linear) functional $\hat{x}(s)$ of the values $\{x(t): t \leq 0\}$ which is the least squares approximation of $x(s)$ where s is any positive integer. In this case the predictability of $x(s)$ can be described in terms of the correlation coefficient

$$(2) \qquad \rho(s) = [1 - \sigma^2(s)/E[x(s)^2]]^{1/2}$$

between $x(s)$ and $\hat{x}(s)$, where $\sigma^2(s)$ is the mean square error of the prediction; (here and later we consider, without loss of generality, only processes $\{x(t)\}$ with $Ex(t) = 0$). Neither the predictor nor the mean squared error for the s-step prediction problem possess such simple expressions as those for the one-step predictor.

We may likewise construct predictors for *any given* linear function of future values, and

---

Received April 1981; revised January 1983.

[1] Supported in part by an NSF grant.

[2] Supported in part by DOE Contract DE-AC02-81ER10841.

AMS 1980 *subject classifications.* Primary 62M15; secondary 60G25, 47B35.

Key words and phrases. Time series, spectrum, prediction theory, canonical correlations, strong mixing.

an expression, perhaps complicated, for its mean squared error. Thus, in one sense, solving the one-step prediction problem provides solutions for *all* prediction problems. However certain questions are left unanswered. While the expression for the one-step prediction error yields much insight into the characteristics of a series with a given spectral distribution function it does not help us with problems such as *how to find* the most predictable aspect (that is, linear function) of the future. In this paper we shall address the latter problem.

A related problem that has received a great deal of attention is to find conditions (on dF) under which there is no aspect of the future which is completely predictable, or even arbitrarily close to being so. A result of Helson and Szego (1960) along this line is discussed in Section 2. This result characterizes those spectral measures dF for which there is a bound less than 1 on the predictability of aspects of the future. The exact value of this upper bound may be described mathematically in several ways which shall be discussed in Section 3. None are as directly computable as expression (1) for the one-step predictability. Similarly, the coefficients of the most predictable aspect and of its predictor (should these exist) may be mathematically described but again they lack the relatively direct computability of the one-step predictor coefficients.

It is clear that the questions being raised here would, in the context of multivariate analysis, be answered by examining *canonical correlations*. In Section 3 we adopt an infinite dimensional version of this theory in order to obtain a different view of the predictability problems already discussed, and one which may be more familiar to statisticians. As might be expected, we shall see that the canonical correlations turn out to be eigenvalues of a certain bounded operator on a Hilbert space. This operator belongs to a certain class of operators called *Hankel operators*. Independently Grenander (1981) has observed that the maximum canonical correlation is the norm of a Hankel operator. However we believe that this is the first comprehensive study of the canonical structure of a time series which appears in the literature.

Our approach enables us to derive some interesting qualitative results concerning the canonical structure of the past and future of a time series. In the second paper of this series, (Jewell, Bloomfield and Bartmann, 1983) we will show how to use the definition of canonical correlations and components given in Section 3 in their computation and also look at some examples where we have computed estimates of the values of the canonical components, etc. The second paper also contains some simple methods for giving bounds on the largest canonical correlation when the spectrum takes a certain form.

It is hoped that the ideas of canonical analysis should be of value in the study of a time series, much as they are in multivariate analysis. This usefulness is not always in the sense of one set of random variables providing predictors of another set, but rather in giving the canonical components which provide insight into the structure and relationship of the two sets of random variables. As we shall see, the canonical structure of a time series yields more subtle information on the predictive properties of the series than is usually available from other standard procedures or model fitting. In this way, examining the canonical structure information may be valuable in checking the adequacy of fit of models to the series. As is shown in the second paper, the computation of the canonical structure is relatively straightforward and so this procedure could be added to the tools time series analysts already use to study the structure of a series.

We have already hinted that, in what follows, we shall restrict our attention to predictors or functionals which are *linear* combinations of the values $\{x(t): t \leq 0\}$. The justification for this is twofold. First, in practical applications, such combinations are easily handled. Second and more significantly, for Gaussian processes the best predicted value for $x(s)$ is the best linear prediction. Thus we can say that the problem of linear least squares prediction of the stationary process $x(t)$ is the wide sense version of the general problem of least squares prediction.

Some of the results we discuss are not new and, as we shall see, have been around in the literature for some time. The approach of Section 3 and some of the characterizations there appear to be new at least in the form we describe. Several of the known results we make use of have appeared in quite different contexts than that of canonical correlations.

Other work related to the problems discussed here include the paper of Akaike (1975) on Markov representations of ARMA processes. Other authors have considered the problem of predictions where we allow *non-linear* functionals of the past to be used as predictors of non-linear functions of future values—see, for example, Gelfand and Yaglom (1959), and Hannan (1961). Yaglom (1965) considered the first canonical component and correlation of stationary processes with rational spectra.

Throughout the paper we will make use of standard results on analytic functions on the open unit disk in $\mathbf{C}$ (denoted by Δ) and their extensions to the unit circle in $\mathbf{C}$ (denoted by C). We shall also assume familiarity with the Hardy space, H^2, of functions in L^2 of the unit circle in $\mathbf{C}$ which possess analytic extensions into the disk. A basic reference for this material is Hoffman (1962). We shall also use some operator theory results and terminology. Halmos (1974) is a good source for most of these ideas.

2. Characterization of processes. We recall that a weakly stationary stochastic process $x(t)$ has a spectral representation. Namely the covariances $\gamma_t = E\{x(0)x(t)\}$ form a positive definite sequence and so are Fourier coefficients of a finite positive measure F on C, i.e., $\gamma_t = \int_C e^{it\omega} dF(\omega)$. Let P be the probability measure of the process. Then we obtain an isometry between $L^2(dF)$ and the span in $L^2(dP)$ of the functions $\{x(t): t \in Z\}$ by mapping $x(t)$ to the function $e^{it\omega}$ and extending by linearity and continuity. Thus the process $\{x(t)\}$ is represented by a sequence of exponentials $\{e^{it\omega}: t \in Z\}$ in the Hilbert space $L^2(dF)$.

We now make some definitions which we shall use in the following discussion. The *past* of the process is the σ-algebra generated by $\{x(t): t \leq 0\}$. In the spectral representation the past is the span in $L^2(dF)$ of the exponentials $e^{it\omega}$ with $t \leq 0$ and we denote it by $\mathscr{P}$. For future reference $\mathscr{P}_k$ is defined to be the span in $L^2(dF)$ of the exponentials $e^{it\omega}$ with $t \leq k$. Similarly the *future beyond time* s is the σ-algebra generated by $\{x(t): t \geq s\}$. In the spectral representation this is the span in $L^2(dF)$ of the exponentials $e^{it\omega}$ with $t \geq s$ and we denote it by $\mathscr{F}_s$. For the case $s = 1$ we simply refer to $\mathscr{F}_1$ as the *future* of the process.

From this point on we restrict our attention to *Gaussian processes* for reasons mentioned in the introduction. The process is called *deterministic* if its past determines the future, i.e., for each $t > 0$, $x(t)$ is measurable with respect to the past. This is translated in the spectral representation to the property that $\mathscr{P} = L^2(dF)$. A necessary and sufficient condition for this to occur is that $\log(dF/d\omega)$ be not integrable. Conversely the process is *indeterministic* if $\log(dF/d\omega)$ is integrable. A stronger restriction that indeterminism is that the process is *purely indeterministic* or *regular*. This is an asymptotic condition which in the spectral representation becomes the condition that $\rho(s) \to 0$ as $s \to \infty$ where $\rho(s)$ is given by (2). Alternatively this is equivalent to $\bigcap_{s=1}^{\infty} \mathscr{F}_s = \{0\}$, a condition which is often referred to by saying that the process has trivial remote future. Results of Szego (1920, 1921), Kolmogorov (1941a) and Krein (1944) show that $\{x(t)\}$ is regular if and only if dF is absolutely continuous with respect to Lebesgue measure and $\log(dF/d\omega)$ is integrable. We can then write $dF = w\,d\omega = |h|^2 d\omega$, where w is known as the spectrum of the process and h is an outer function in H^2.

A stronger property than that of regularity is *minimality*. Introduced by Kolmogorov (1941a), this property says that a process is *minimal* if the value of the random variable $x(0)$ cannot be predicted without error from the values of the random variables $\{x(t): t \neq 0\}$. In other words, a process is not minimal if it is possible to perfectly interpolate any value of the process from knowledge of the remaining values of the process. Kolmogorov (1941b) proved that a regular Gaussian process is minimal if and only if w^{-1} is in $L^1(d\omega)$.

In what follows we shall be interested in correlations between elements of $\mathscr{P}$ and $\mathscr{F}_1$. It is trivial that if $\mathscr{P}$ and $\mathscr{F}_1$ have non-empty intersection, we can obtain a correlation of 1 between elements of $\mathscr{P}$ and $\mathscr{F}_1$. It is possible however, even if $\mathscr{P} \cap \mathscr{F}_1$ is empty, to obtain correlations between past and future elements arbitrarily close to 1. To discuss this, we introduce the notion of *the angle between* $\mathscr{P}$ *and* $\mathscr{F}_1$. The maximum correlation between the elements of $\mathscr{P}$ and $\mathscr{F}_1$ is given by $\lambda_1 = \sup\{|\langle f, g\rangle| : f \in$ unit ball of $\mathscr{P}$, $g \in$ unit ball of $\mathscr{F}_1\}$ where the inner product is in the space $L^2(dF)$. The angle α between $\mathscr{P}$ and $\mathscr{F}_1$ is defined by $\cos\alpha = \lambda_1$. It follows that a yet stronger condition of independence between $\mathscr{P}$ and $\mathscr{F}_1$ than regularity is the requirement that $\mathscr{P}$ and $\mathscr{F}_1$ are at a positive angle, i.e., $\alpha > 0$ or $\lambda_1 < 1$. It is also easy to verify that the condition of a positive angle is strictly stronger than minimality.

Helson and Szego (1960) supplied the characterization of those processes for which $\mathscr{P}$ and $\mathscr{F}_1$ are at a positive angle. Their necessary and sufficient condition on dF is that it be absolutely continuous w.r.t. Lebesgue measure and $w = e^{u+\tilde{v}}$ where u, v are real $L^\infty(d\omega)$ functions and $\|v\|_\infty < \pi/2$. ($\tilde{v}$ represents the Hilbert transform of v.) A totally different characterization of spectra which yield processes for which $\mathscr{P}$ and $\mathscr{F}_1$ are at a positive angle was found by Hunt, Muckenhoupt and Wheeden (1973). This condition is more appealing since it gives a structural condition on w. The result is that $\mathscr{P}$ and $\mathscr{F}_1$ are at a positive angle if and only if w satisfies the following inequality:

$$\sup_I\left(\frac{1}{|I|}\int_I w\,d\omega\right)^{1/2}\left(\frac{1}{|I|}\int_I w^{-1}d\omega\right)^{1/2} < \infty,$$

where I ranges over all subarcs of C.

The problem of when the angle between $\mathscr{P}$ and $\mathscr{F}_k$ is positive is related to the characterization of processes which satisfy Rosenblatt's (1956) *strong mixing condition.* The linear version of this is that the predictability of all aspects of the future beyond time k from the past should converge to zero as $k \to \infty$. Helson and Sarason (1967) showed that a necessary and sufficient condition for this is that dF is absolutely continuous w.r.t. Lebesgue measure and $w = |p|^2 e^{u+\tilde{v}}$ where p is a trigonometric polynomial and u and v are real continuous functions on C. (The final version of this characterization is due to Sarason, 1972.)

A class of strong mixing processes which will appear later are those which are known as *absolutely regular*. Without going into details, they are processes for which the amount of "information" in $\mathscr{P}$ about $\mathscr{F}_k$ converges to zero as $k \to \infty$. The necessary and sufficient condition on the spectrum for absolute regularity is that $w = |p|^2 f$ where p is a polynomial with zeros on C and $\log f$ has Fourier series $\sum_{-\infty}^{\infty} f_j e^{ij\omega}$ such that $\sum_{-\infty}^{\infty} |j|\,|f_j|^2 < \infty$—see Ibragimov and Rozanov (1978) for details.

3. Canonical correlations. Many of the questions raised in the last section would, in the context of multivariate analysis, be answered by examining canonical correlations which involves looking at the eigenvalues of a correlation matrix. In this section we want to extend the finite-dimensional approach to consider canonical correlations of the past and future. As might be expected, we shall see that the correlation matrix will be replaced by an operator between infinite-dimensional spaces.

The correlation between an element $f \in \mathscr{F}_1$ and an element $g \in \mathscr{P}$ is defined to be

$$\frac{\langle f, g\rangle}{\|f\|\,\|g\|} = \left(\int_C f\bar{g}\,dF\right)\Big/\left(\int_C |f|^2 dF\right)^{1/2}\left(\int_C |g|^2 dF\right)^{1/2}.$$

Since we are restricting our attention to regular processes we can write $\operatorname{corr}(f, g) = \int f\bar{g}\,w\,d\omega$ where we only consider elements f, g of norm 1 in $L^2(dF)$.

Before going any further, we digress to introduce some machinery which is crucial in what follows. Recall that for regular processes we can write $dF = w\,d\omega = |h|^2 d\omega$ where h is outer in H^2. Consider the mapping $T: L^2(df) \to L^2(d\omega)$ given by $Tf = hf$. It is easily verified that T is an isometry of $L^2(dF)$ into $L^2(d\omega)$. Also T maps $\mathscr{F}_1$ onto H_0^2, the set of functions in H^2 whose analytic extension into the disk vanishes at the origin, and T maps $\mathscr{P}$ onto $(h/\bar{h})\bar{H}^2$. Using the isometry T we thus have

$$\operatorname{corr}(f, g) = \int_C f\bar{g}|h|^2\,d\omega = \int_C (fh)(\bar{g}h)(\bar{h}/h)\,d\omega = \int_C FG(\bar{h}/h)\,d\omega$$

where $F \in H_0^2$, $G \in H^2$, $\|F\|_2 = 1$, $\|G\|_2 = 1$ where $\|\cdot\|_2$ represents the norm in L^2.

Hence $\operatorname{corr}(f, g) = \langle(\bar{h}/h)G, \bar{F}\rangle$ where the inner product is in L^2

$= \langle(\bar{h}/h)G, (I-P)\bar{F}\rangle$ where P is the orthogonal projection of L^2 onto H^2

$= \langle(I-P)(\bar{h}/h)G, \bar{F}\rangle = \langle H_{\bar{h}/h}G, \bar{F}\rangle$ where $H_{\bar{h}/h}G$ is the Hankel operator with symbol $\bar{h}/h$.

[If ϕ is a function on the unit circle in L^∞ then the *Hankel operator with symbol* ϕ is defined as a bounded linear operator from H^2 to $L^2 \ominus H^2$ by $H_\phi(f) = (I-P)(\phi f)$ for $f \in H^2$. The norm of the Hankel operator with symbol ϕ is given by

$$\|H_\phi\| = \inf\{\|\phi - h\| : h \in H^\infty\}$$

where H^∞ is the closed subspace of L^∞ given by L^∞ functions which are the boundary values of bounded analytic functions on the open unit disk. For a proof of this and other results on Hankel operators, see Power (1980).]

Now since T maps $\mathscr{F}_1$ *onto* H_0^2 and $\mathscr{P}$ *onto* $(h/\bar{h})\bar{H}^2$, we have that the maximum correlation between elements in $\mathscr{P}$ and $\mathscr{F}_1$ is given by

$$\begin{aligned}
&\sup\{|\operatorname{corr}(f, g)| : f \in \mathscr{F}_1, \|f\| = 1; g \in \mathscr{P}, \|g\| = 1\}\\
&\quad = \sup\{|\operatorname{corr}(f, g)| : f \in \mathscr{F}_1, \|f\| \le 1; g \in \mathscr{P}, \|g\| \le 1\}\\
&\quad = \sup\{|\langle H_{\bar{h}/h}G, \bar{F}\rangle| : F \in H_0^2, \|F\|_2 = 1; G \in H^2, \|G\|_2 = 1\}\\
&\quad = \|H_{\bar{h}/h}\| = \inf\{\|\bar{h}/h - u\|_\infty : u \in H^\infty\}, \text{ i.e., } d(\bar{h}/h, H^\infty).
\end{aligned}$$

Alternatively,

$$\begin{aligned}
\|H_{\bar{h}/h}\|^2 &= \|H^*_{\bar{h}/h}H_{\bar{h}/h}\|\\
&= \sup\{|\mu| : \mu I - H^*_{\bar{h}/h}H_{\bar{h}/h} \text{ is not invertible}\}\\
&= \sup\{\mu : \mu \in \operatorname{sp}(H^*_{\bar{h}/h}H_{\bar{h}/h})\} = r(H^*_{\bar{h}/h}H_{\bar{h}/h}),
\end{aligned}$$

where $r(S)$ is the *spectral radius of* S

[The *spectrum* of a bounded linear operator T on a Hilbert space is the set in $\mathbf{C}$ given by $\operatorname{sp}(T) = \{\lambda : \lambda I - T \text{ is not invertible}\}$. Of course if the Hilbert space is finite dimensional this is just the set of eigenvalues of T. The spectral radius of T is given by $r(T) = \sup\{|\lambda| : \lambda \in \text{spectrum of } T\}$. See Halmos (1974).]

If ϕ is a function on the unit circle in L^∞ then the *Toeplitz operator with symbol* ϕ is defined as a bounded linear operator from H^2 to H^2 by $T_\phi(f) = P(\phi f)$ for $f \in H^2$. The norm of a Toeplitz operator is given by $\|T_\phi\| = \|\phi\|_\infty$. Further details and information on Toeplitz operators can be found in Chapter 7 of a book by R. Douglas (1972). Toeplitz operators are related to Hankel operators in many ways, the simplest given by the following algebraic identity which is easily established:

$$H_{\bar{f}}^* H_g = T_{fg} - T_f T_g \quad (f, g \in L^\infty). \tag{3}$$

Using this identity we can describe the maximum correlation in terms of Toeplitz operators. We substitute $f = h/\bar{h}$, $g = \bar{h}/h$ into (3) and obtain $H^*_{\bar{h}/h}H_{\bar{h}/h} = I - T^*_{\bar{h}/h}T_{\bar{h}/h}$. For simplicity we write $H \equiv H_{\bar{h}/h}$ and $T \equiv T_{\bar{h}/h}$. Thus the maximum correlation is given by

$$\begin{aligned}
\sqrt{\sup\{\mu : \mu \in \operatorname{sp}(H^*H)\}} &= \sqrt{\sup\{\mu : \mu \in \operatorname{sp}(I - T^*T)\}}\\
&= \sqrt{1 - \inf\{\mu : \mu \in \operatorname{sp}(T^*T)\}} = \sqrt{1 + r(-T^*T)}.
\end{aligned}$$

Thus we have the following theorem.

THEOREM 1. *The maximum correlation between $\mathscr{P}$ and $\mathscr{F}_1$ is given by* $d(\bar{h}/h, H^\infty) = \sqrt{r(H^*H)} = \sqrt{1 - \inf\{\mu : \mu \in \operatorname{sp}(T^*T)\}}$.

Since one of our descriptions of the maximum correlation involves the spectrum of the operator H^*H we want to consider the properties of this set briefly. Trivially H^*H is a positive, self-adjoint operator and $\| H^*H \| \le 1$ since $\| \bar{h}/h \|_\infty = 1$. Hence $\operatorname{sp}(H^*H)$ is a compact subset of $[0, 1]$; it may be a "continuous" set (i.e., not discrete) but since H^*H is self-adjoint any discrete (i.e. isolated) points in the spectrum correspond to eigenvalues (possibly of infinite multiplicity).

We now look at how we shall extend the finite-dimensional concept of canonical correlations. If the largest point in the spectrum, λ_1^2, of H^*H is a positive eigenvalue of finite multiplicity we shall say that λ_1 is the first canonical correlation, ($\lambda_1 > 0$). We now show how to obtain canonical components in this situation. Let $f \in H^2$ be a unit eigenvector for H^*H corresponding to λ_1^2; then $Hf \in L^2 \ominus H^2$. Consider the correlation between $f_1 = \overline{Hf}/\lambda_1 h$ and $f_2 = \bar{f}/\bar{h}$. Clearly $f_1 \in \mathscr{F}_1$, $f_2 \in \mathscr{P}$ and $\| f_1 \| = \| f_2 \| = 1$. Also

$$|\operatorname{corr}(f_1, f_2)| = \left| \int_C \frac{\overline{Hf}}{\lambda_1 h} \frac{f}{h} |h|^2 \, d\omega \right| = \left| \int_C \frac{\overline{Hf}}{\lambda_1} f\left(\frac{\bar{h}}{h}\right) d\omega \right|$$

$$= \frac{1}{\lambda_1} \left\langle \frac{\bar{h}}{h} f, Hf \right\rangle = \frac{1}{\lambda_1} \langle Hf, Hf \rangle = \frac{1}{\lambda_1} \langle H^*Hf, f \rangle = \lambda_1.$$

Thus the elements $f_1 \in \mathscr{F}_1$, $f_2 \in \mathscr{P}$ have (absolute) correlation equal to λ_1. They are called the *first canonical components.*

By analogy with the finite-dimensional definition of successive canonical correlations we are next led to consider

$$\begin{aligned} \lambda_2 = \sup\{ |\operatorname{corr}(f, g)| : \ & f \in \mathscr{F}_1, \text{ orthogonal to 1st canonical component in } \mathscr{F}_1 \\ & g \in \mathscr{P}, \text{ orthogonal to 1st canonical component in } \mathscr{P} \\ & \| f \| = \| g \| = 1 \}. \end{aligned}$$

If λ_1 is an eigenvalue of finite multiplicity greater than one, then we can choose f^1 in H^2 which is a unit eigenvector for H^*H corresponding to λ_1^2 and is orthogonal to f in H^2. In this case λ_1 is also the second canonical correlation and the second canonical components are $\overline{Hf^1}/\lambda_1 h \in \mathscr{F}_1$ and $\overline{f^1}/\bar{h} \in \mathscr{P}$. (It is easy to see that $\overline{Hf^1}/\lambda_1 h$ and $\overline{f^1}/\bar{h}$ are orthogonal to the appropriate 1st canonical components in $\mathscr{F}_1$ and $\mathscr{P}$ respectively.)

If the multiplicity of λ_1 is k we repeat this procedure to obtain the first k canonical correlations and respective canonical components. We next consider

$$\begin{aligned} \lambda_{k+1} = \sup\{ |\operatorname{corr}(f, g)| : \ & f \in \mathscr{F}_1, \text{ orthogonal to first } k \text{ canonical components in } \mathscr{F}_1 \\ & g \in \mathscr{P}, \text{ orthogonal to first } k \text{ canonical components in } \mathscr{P} \\ & \| f \| = \| g \| = 1 \} \\ = \sup\{ |\langle HG, \bar{F} \rangle| : \ & F \in H_0^2, \ \| F \|_2 \le 1, \ G \in H^2, \ \| G \|_2 \le 1 \\ & F, G \perp \text{ eigenspace of } H^*H \text{ corresponding to } \lambda_1^2 \}. \end{aligned}$$

This follows since as $f \in \mathscr{F}_1$ moves through all the elements orthogonal to the first k canonical components in $\mathscr{F}_1$, hf moves through all the elements in H_0^2 orthogonal to the eigenspace of H^*H corresponding to λ_1^2. Similarly as $g \in \mathscr{P}$ moves through all the elements orthogonal to the first k canonical components in $\mathscr{P}$, $h\bar{g}$ moves through all the elements in H^2 orthogonal to the eigenspace of H^*H corresponding to λ_1^2.

It follows that $\lambda_{k+1}^2 = \sup\{\mu : \mu \in \mathrm{sp}(H^*H) \text{ and } \mu \leq \lambda_k\}$ (see Butz, 1974) i.e. λ_{k+1}^2 is the $(k+1)$th largest "element" of $\mathrm{sp}(H^*H)$, counting multiplicities.

If λ_{k+1}^2 is an eigenvalue of finite multiplicity then λ_{k+1} is the $(k+1)$th canonical correlation and we define the appropriate canonical components as before, and continue to seach for further canonical correlations. If λ_{k+1}^2 is not an eigenvalue of finite multiplicity then we stop the process.

In order to describe the canonical correlations more succinctly we introduce some notation from operator theory. We recall that for a bounded operator A on a Hilbert space which is normal the *essential spectrum* consists of the limit points of $\mathrm{sp}(A)$ together with those eigenvalues of A which have infinite multiplicity. Enumerating the "upper part" of the spectrum of $(A^*A)^{1/2}$ we thus have a sequence of eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ counted with multiplicity and then we reach the supremum of the essential spectrum of $(A^*A)^{1/2}$, denoted by λ_∞. Of course there may be only a finite number of eigenvalues of finite multiplicity before λ_∞ is reached. The sequence $\lambda_1, \lambda_2, \cdots$ of eigenvalues of finite multiplicity is thus either finite or countable in which case it converges to λ_∞. λ_j is known as the jth *s-number* of the operator A. (For reference on this material see Gohberg and Krein, 1969.) Our earlier work is thus summarized by the following theorem.

THEOREM 2. *The jth canonical correlation of the process is the jth s-number of the Hankel matrix H. There may be only a finite number of these or a countable number (which converge to λ_∞, the supremum of the essential spectrum of $(H^*H)^{1/2}$).*

λ_∞ will be referred to as the *essential correlation* of the past and future of the process. At this point, however many canonical components with correlation λ_∞ are removed, the correlation of the remainder of the past with the remainder of the future remains λ_∞.

We now return to considering various types of regular processes and characterize them in terms of the operator H rather than the spectrum w.

The necessary and sufficient condition for $\mathscr{P}$ and $\mathscr{F}_1$ to be at a positive angle is easily determined. Recall that the angle α between $\mathscr{P}$ and $\mathscr{F}_1$ is given by $\cos\alpha = \sup\{|\langle f, g\rangle| : f \in \text{unit ball of } \mathscr{F}_1, g \in \text{unit ball of } \mathscr{P}\} = \|H\|$. Thus $\mathscr{P}$ and $\mathscr{F}_1$ are at a positive angle if and only if $\|H\| < 1$.

This simple result also extends to considering the angle between $\mathscr{P}$ and $\mathscr{F}_n$, call it α_n. We have

$$\begin{aligned}\cos(\alpha_n) &= \sup\{|\langle f, g\rangle| : f \in \text{unit ball of } \mathscr{F}_n, g \in \text{unit ball of } \mathscr{P}\}\\ &= \int_C z^{n-1}FG(\bar{h}/h)\,d\omega \quad \text{as before,}\\ &\qquad \text{where} \quad F \in H_0^2, \quad G \in H^2 \quad \text{and} \quad \|F\|_2 = 1, \quad \|G\|_2 = 1\\ &= \|H_{z^{n-1}\bar{h}/h}\| = d(z^{n-1}\bar{h}/h, H^\infty).\end{aligned}$$

Hence $\cos(\alpha_n) \to 0$ as $n \to \infty$ if and only if $d(z^{n-1}\bar{h}/h, H^\infty) \to 0$ as $n \to \infty$. This is equivalent to $d(\bar{h}/h, H^\infty + K) = 0$ where K is the space of the continuous functions on C. But $H^\infty + K$ is a closed subspace of L^∞—see Helson and Sarason (1967). Thus $\bar{h}/h \in H^\infty + K$ and this is equivalent to H being compact by a theorem of Hartman (1958). Thus the process is strong mixing if and only if H is compact. Since compact operators are exactly those operators whose spectrum consists of a countable (perhaps finite) number of non-zero eigenvalues of finite multiplicity together with perhaps a point at 0 the above discussion shows that strong mixing processes are exactly those processes with a countable (perhaps finite) number of non-zero canonical correlations and essential correlation equal to zero. This gives perhaps the simplest extension of the finite dimensional theory in the sense that all the correlation between the past and future can be described using a countable number of canonical components.

What about processes with only a finite number of canonical correlations and essential correlation zero? For such processes $(H^*H)^{1/2}$ is finite rank and so H must be finite rank. A theorem of Kronecker (1881) shows that a Hankel matrix is finite rank if and only if its symbol is the sum of a rational function and an H^∞ function. (By a rational function we mean the quotient of two polynomials in z where the zeros of the denominator all have absolute value less than 1.) Now suppose $\bar{h}/h$ is of this form i.e., $\bar{h}/h = (p/q) + f$ where p, q are polynomials and the zeros of q all lie in the open unit disc, and $f \in H^\infty$. Then $\bar{h}q = ph + fqh$. The LHS $\in z^k\overline{H^2}$ where k is the degree of q, and the RHS $\in H^2$. Hence both sides are polynomials of degree k; i.e., $\bar{h}q = r$ where r is a polynomial of degree k. This implies $w = |\bar{h}|^2 = |r|^2/|q|^2$, i.e. the spectrum is the quotient of two positive polynomials the zeros of the denominator all lying in the open unit disc. If w is such a function, it is straightforward to show that $\bar{h}/h$ is rational. We have thus shown that the process possesses only a finite number of non-zero canonical correlations and has essential correlation zero if and only if the spectrum is the quotient of two positive polynomials, the zeros of the denominator all lying in the open unit disc, i.e. the process is an ARMA process. This result can also be easily derived from the Yule-Walker equations. In fact, for an ARMA (p, q) process the number of non-zero canonical correlations is $\max(p, q)$.

It will be useful when constructing bounds for the maximum canonical correlation to consider absolutely regular processes which were defined at the end of Section 2. We now characterize these processes in terms of H. Recall that the process is absolutely regular if and only if $w = |p|^2 f$ where p is a polynomial with zeros on C and $\log f$ has Fourier series $\sum_{-\infty}^{\infty} f_j e^{ij\omega}$ where $\sum_{-\infty}^{\infty} |j||f_j|^2 < \infty$. In fact it is easier to prove (see Ibragimov and Rozanov (1978), page 129) that absolute regularity is equivalent to $\bar{h}/h$ having a Fourier series $\sum_{-\infty}^{\infty} c_j e^{ij\omega}$ with the property that $\sum_{-\infty}^{0} |j||c_j|^2$ is convergent. It is well-known and easy to prove that a Hankel operator with symbol ϕ is Hilbert-Schmidt if and only if the Fourier series of ϕ, $\sum_{-\infty}^{\infty} \phi_j e^{ij\omega}$ has this same property, i.e. $\sum_{-\infty}^{0} |j||\phi_j|^2$ converges. Hence the process is absolutely regular if and only if H is Hilbert-Schmidt (a bounded linear operator T is Hilbert-Schmidt if it is compact and its sequence of countable non-zero eigenvalues is square-summable). The Hilbert-Schmidt property of H is equivalent to demanding that $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$ where λ_j is the jth canonical correlation of the past and future.

We finish this section by establishing some results on the canonical correlations of the interpolation error process connected with the process $\{x(t)\}$. Suppose $\{x(t)\}$ is a minimal process. Let $\hat{x}(t)$ be the projection of $x(t)$ onto the subspace spanned by $\{x(s): s = t \pm 1, t \pm 2, \cdots\}$ i.e. $\hat{x}(t)$ is the best prediction (in terms of mean squared error) of $x(t)$ from all the other random variables $x(t \pm 1)$, $x(t \pm 2)$, $\cdots$. Let $y(t)$ denote the error in this prediction, i.e., $y(t) = x(t) - \hat{x}(t)$. It is well-known that $y(t)$ is then a regular (in fact minimal) weakly stationary Gaussian process whose spectrum is given by w^{-1}.

THEOREM 3. *The interpolation error process* $\{y(t)\}$ *has an identical canonical correlation structure to the original minimal process* $\{x(t)\}$.

PROOF. We can write $w^{-1} = |h|^{-2}$, where h^{-1} is also an outer function which is in H^2 due to the minimality of $\{x(t)\}$. By Theorem 2 the canonical correlations and components are described by the spectrum of the operator $H^*_{h/\bar{h}}H_{h/\bar{h}}$. By (3) this equals $I - T^*_{h/\bar{h}}T_{h/\bar{h}}$. But for Toeplitz operators $T^*_\phi = T_{\bar{\phi}}$ (this property does not hold for Hankel operators). Hence $H^*_{h/\bar{h}}H_{h/\bar{h}} = I - T_{\bar{h}/h}T_{h/\bar{h}} = I - T_{\bar{h}/h}T^*_{\bar{h}/h}$. Now for any pair of bounded linear operators A, B on a Hilbert space $\mathrm{sp}(AB)\backslash\{0\} = \mathrm{sp}(BA)\backslash\{0\}$. Hence the spectrum of the operator $H^*_{h/\bar{h}}H_{h/\bar{h}}$ is the same as the spectrum of $H^*_{\bar{h}/h}H_{\bar{h}/h}$ except for possibly the point 1. We claim that $1 \in \mathrm{sp}(H^*_{\bar{h}/h}H_{\bar{h}/h})$ if and only if $1 \in \mathrm{sp}(H_{\bar{h}/h}H^*_{\bar{h}/h})$. Recall that $\|H_{\bar{h}/h}\|^2 = \|H^*_{\bar{h}/h}H_{\bar{h}/h}\| = \sup\{\lambda : \lambda \in \mathrm{sp}(H^*_{\bar{h}/h}H_{\bar{h}/h})\}$. Hence $1 \in \mathrm{sp}(H^*_{\bar{h}/h}H_{\bar{h}/h}) \Rightarrow \|H_{\bar{h}/h}\| = 1$. This implies that the past and the future are at zero angle for the process $\{x(t)\}$. But this implies that the past and the future are at zero angle for the process $\{y(t)\}$ (for either the Helson-Szego characterization or the Hunt-Muckenhoupt-Wheeden result show that, if $w^{-1} \in L^1$, the process $\{x(t)\}$ with spectrum w has the property that the past and future are at positive angle if and only if the process $\{y(t)\}$ with spectrum w^{-1} has the same property). Thus $\| H_{h/\bar{h}} \| = 1$ and this in turn implies that $\sup\{\lambda : \lambda \in \mathrm{sp}(H^*_{\bar{h}/\bar{h}} H_{h/\bar{h}})\} = 1$ and thus $1 \in \mathrm{sp}(H^*_{\bar{h}/\bar{h}} H_{h/\bar{h}})$ since the spectrum of any bounded operator on a Hilbert space is a compact set. We can reverse this reasoning and thus establish our claim. We have now shown that $\mathrm{sp}(H^*_{\bar{h}/h} H_{\bar{h}/h}) = \mathrm{sp}(H^*_{\bar{h}/\bar{h}} H_{h/\bar{h}})$. To complete the proof of the theorem, we need to show that eigenvalues of finite multiplicity for $H^*_{\bar{h}/h} H_{\bar{h}/h}$ correspond to eigenvalues of the same multiplicity for $H^*_{h/\bar{h}} H_{h/\bar{h}}$. Suppose λ is an eigenvalue of the operator A^*A (where A is a bounded linear operator on Hilbert space) of multiplicity k; another way of putting this is that $\ker(\lambda I - A^*A)$ is a k-dimensional subspace. Now it is trivial to check that $A(\ker(\lambda I - A^*A)) \subseteq \ker(\lambda I - AA^*)$. Substituting $T_{\bar{h}/h}$ for A we have $T_{\bar{h}/h}(\ker(\lambda I - T^*_{\bar{h}/h} T_{\bar{h}/h})) \subseteq \ker(\lambda I - T^*_{h/\bar{h}} T_{h/\bar{h}})$ or $T_{\bar{h}/h}(\ker(H^*_{\bar{h}/h} H_{\bar{h}/h} - (1 - \lambda)I)) \subseteq \ker(H^*_{h/\bar{h}} H_{h/\bar{h}} - (1 - \lambda)I)$. Now $T_{\bar{h}/h}$ is one-one for otherwise there exists a non-zero function $f \in H^2$ such that $T_{\bar{h}/h} f = 0$. This would imply that $H^*_{\bar{h}/h} H_{\bar{h}/h} f = (I - T^*_{\bar{h}/h} T_{\bar{h}/h}) f = f$. Writing $g = f/\| f \|_2$ we then have $1 = \| H_{\bar{h}/h} g \|_2 = \| (I - P)(\bar{h}/h) g \|_2 \leq \| (\bar{h}/h) g \|_2 \leq \| g \|_2 = 1$. Thus $(\bar{h}/h) g \in L^2 \ominus H^2$. It follows easily from this that $\bar{g}/\bar{h} \in \mathscr{P} \cap \mathscr{F}_1$. Hence $\mathscr{P} \cap \mathscr{F}_1 \neq \{0\}$. But the process $\{x(t)\}$ is minimal and thus, by Proposition 3 of Bloomfield, Hayashi, Jewell (1983), $\mathscr{P} \cap \mathscr{F}_1 = \{0\}$. Hence $T_{\bar{h}/h}$ is one-one and so $\dim \ker(H^*_{\bar{h}/h} H_{\bar{h}/h} - \mu I) = \dim \ker(H^*_{h/\bar{h}} H_{h/\bar{h}} - \mu I)$ for any $\mu \in \mathbf{C}$. This completes the proof.

Theorem 3 leads one to suspect that there must be an appropriate finite dimensional analog. There is and we include it since it seems to be little known. Let $X_1, \cdots, X_n$ be a finite set of Gaussian random variables with covariance matrix Σ. Let $Z_1, \cdots, Z_n$ be the set of random variables given by the "regression errors" when a single X is predicted from all the others i.e., $Z_j = X_j - E(X_j \mid \{X_i : i \neq j\})$.

THEOREM 4. *When partitioned in the same way, the X set of random variables and the Z set of random variables have the same canonical correlations.*

PROOF. Let $\mathbf{Y} = \Sigma^{-1}\mathbf{X}$ where $\mathbf{X} = (X_1, \cdots, X_n)^T$. Then $\mathrm{cov}(\mathbf{X}, \mathbf{Y}) = E(\mathbf{X}\mathbf{Y}^T) = \mathbf{I}_n$ and $\mathrm{cov}(\mathbf{Y}, \mathbf{Y}) = E(\mathbf{Y}\mathbf{Y}^T) = \Sigma^{-1}$. Now when partitioned the same way the covariance matrices Σ and Σ^{-1} possess the same canonical correlation structure. For suppose Σ possesses canonical correlations $d_1, \cdots, d_k$ when partitioned in a certain way. Then there exist invertible matrices A_1, A_2 such that

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right] = \left[\begin{array}{c|c} A_1 & 0 \\ \hline 0 & A_2 \end{array}\right] \left[\begin{array}{c|c} I & D \\ \hline D & I \end{array}\right] \left[\begin{array}{c|c} A_1 & 0 \\ \hline 0 & A_2 \end{array}\right] = A \left[\begin{array}{c|c} I & D \\ \hline D & I \end{array}\right] A^T \tag{4}$$

where D is the matrix

$$\begin{bmatrix} d_1 & & & & & & \\ & d_2 & & & 0 & & \\ & & \ddots & & & & \\ 0 & & & & & & \\ & & & d_k & 0 & \cdots & 0 \end{bmatrix}$$

where the number of zero columns depends on the difference in the size of the two parts of $\mathbf{X}$ given by the partition. Thus there is a unitary matrix P such that

$$\Sigma = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} P D^* P \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}^T$$

where D^* is (2×2) block diagonal with block diagonal elements

$$\begin{bmatrix} 1 & d_1 \\ d_1 & 1 \end{bmatrix}, \cdots, \begin{bmatrix} 1 & d_k \\ d_k & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \cdots, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then it follows that $\Sigma^{-1} = (A^{-1})^T P(D^*)^{-1} P^T A^{-1}$. Now $(D^*)^{-1} = D_1 E^* D_1$, where D_1 is the diagonal matrix with diagonal entries $(1 - d_1^2)^{-1/2}$, $(1 - d_1^2)^{-1/2}$, $\cdots$, $(1 - d_k^2)^{-1/2}$, $(1 - d_k^2)^{-1/2}$, 1, $\cdots$, 1, and E^* is the same as D^* except that each d_j is replaced by $-d_j$. Then there exists another unitary Q such that $(D^*)^{-1} = D_1 Q D^* Q^T D_1$. This implies that Σ^{-1} can be written in the form of (4) where D is the same, but A_1, A_2 may change; i.e. Σ^{-1} has canonical correlations $d_1, \cdots, d_k$ when partitioned in the same way as Σ. To turn **Y** into the regression errors **Z** we merely have to rescale the **Y** entries, i.e.

$$Y_j = \sigma^{jj} X_j + \textstyle\sum_{k \neq j} \sigma^{jk} X_k \quad \text{where} \quad \sigma^{ij} \text{ is the } (i, j)\text{th entry of } \Sigma^{-1}$$

and

$$Z_j = \frac{1}{\sigma^{jj}} Y_j = X_j - \textstyle\sum_{k \neq j} \left(-\left(\dfrac{\sigma^{jk}}{\sigma^{jj}} \right) X_k \right).$$

However the canonical structure of **Y** is invariant under such a rescaling and so, when partitioned in the same way, all three of **X**, **Y** and **Z** possess the same canonical structure.

4. The remaining canonical correlations. The other canonical correlations in which we might be interested are those other than the first between $\mathscr{P}$ and $\mathscr{F}_1$, and those between $\mathscr{P}$ and $\mathscr{F}_n$. We denote the latter (if they exist) by $\lambda_j^{(n)}$, $j = 1, 2, \cdots, n = 1, 2, \cdots$. In the light of the discussion of Section 3, we shall assume in what follows that all necessary canonical correlations are well-defined.

By arguments which parallel those of Section 3 we may show that $\lambda_1^{(n)} = d(\bar{h}/h, \bar{z}^{n-1} H^\infty)$. Every function in $\bar{z}^{n-1} H^\infty$ may be expressed as a sum $r + h$ where $r \in R_n^{(0)}$, $h \in H^\infty$ and $R_n^{(0)}$ is the set of functions on C which are boundary values of a rational function on Δ with a single pole of order n at the origin. The Helson-Sarason theorem states that $\lambda_1^{(n)} < 1$ if and only if w admits a representation of $w = |p|^2 \exp(u + \tilde{v})$ with p, a polynomial of degree less than n, and u, v real-valued functions in L^∞ with $\| v \|_\infty < \pi/2$.

We now return to the set of canonical correlations of $\mathscr{P}$ and $\mathscr{F}_1$. Adamjan, Arov and Krein (1971) have shown that $\lambda_j = d(\bar{h}/h, R_j + H^\infty)$ when R_j is the set of functions on C which are boundary values of a rational function on Δ with poles only in Δ, the number of which (with regard to their order) does not exceed j. Similarly $\lambda_j^{(n)} = d(z^n \bar{h}/h, R_j + H^\infty)$. Thus $\lambda_j^{(n)} = d(z^n \bar{h}/h, R_j + H^\infty) = d(\bar{h}/h, \bar{z}^n H^\infty + \bar{z}^n R_j)$. But clearly for $0 < m < j$, $(\bar{z})^{n+m} R_{j-m} \subset (\bar{z})^n R_j$ so that $\lambda_{j-m}^{(n+m)} \geq \lambda_j^{(n)}$. In particular $\lambda_j = \lambda_j^{(1)} \leq \lambda_{j-1}^{(2)} \leq \cdots \leq \lambda_1^{(j)}$. These inequalities correspond to the familiar interlaced properties of canonical correlations of finite sets of random variables. Also it follows from work of Adamjan, Arov, and Krein (1971) that

$$\lambda_\infty^{(n)} = d(\bar{h}/h, H^\infty + K) \quad (n \geq 1).$$

5. Extensions. In Jewell, Bloomfield and Bartmann (1983) we discuss how to derive some simple bounds for the canonical correlations, especially the largest. We also discuss ways of computing the canonical correlations and components and illustrate these ideas on some simple time series and real data.

We are also presently investigating extensions of these ideas beyond the single discrete time series case. Of most interest are the extensions to (i) discrete vector time series including the situation where a single component of a vector series in the future is predicted using knowledge of all components of the vector in the past; (ii) discrete processes in multidimensional time; (iii) continuous time processes. There are also several important problems relating to the statistical properties of estimates of the canonical correlations and components from a finite set of observations. Details of work on these topics will appear elsewhere.

REFERENCES

ADAMJAN, V. M., AROV, D. Z. and KREIN, M. G. (1971). Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem. *Mat. Sb.* **86** 34–75; *Math. U.S.S.R. Sb.*, 31–73.

AKAIKE, H. (1975). Markovian representation of stochastic processes by canonical variables. *Siam. J. Control* **13** 162–173.

BLOOMFIELD, P., HAYASHI, E. and JEWELL, N.P. (1983). Characterizations of completely nondeterministic stochastic processes. *Pacific J. Math.* **106,** to appear.

BUTZ, J. (1974). *s*-numbers of Hankel matrices. *J. Functional Anal.* **15** 297–305.

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

DOUGLAS, R. G. (1972). *Banach Algebra Techniques in Operator Theory.* Academic, New York.

GELFAND, I. M. and YAGLOM, A. M. (1959). Calculation of the amount of information about a random function contained in another such function. *Amer. Math. Soc. Trans.* **12** 199–246.

GOHBERG, I. C. and KREIN, M. G. (1969). *Introduction to the theory of linear nonselfadjoint operators.* Translation of Mathematical Monographs, Vol. 18, Amer. Math. Soc., Providence, R.I.

GRENANDER, U. (1981). *Abstract Inference.* Wiley, New York.

HALMOS, P. R. (1974). *A Hilbert Space Problem Book.* Springer-Verlag, New York.

HANNAN, E. J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J. Austral. Math. Soc.* **2** 229–242.

HARTMAN, P. (1958). On completely continuous Hankel matrices. *Proc. Amer. Math. Soc.* **9** 862–866.

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

HELSON, H. and SARASON, D. E. (1967). Past and future. *Math. Scand.* **21** 5–16.

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

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

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Guassian Random Processes.* Springer-Verlag, New York.

JEWELL, N. P., BLOOMFIELD, P. and BARTMANN, F. C. (1983). Canonical correlations of past and future for time series: bounds and computation. *Ann. Statist.* **11** 848–855.

KOLMOGOROV, A. N. (1941a). Interpolation und extrapolation von stationaren zufalligen folgen. *Izv. Akad. Nauk. SSSR. Ser. Mat* **5** 3–14.

KOLMOGOROV, A. N. (1941b). Stationary sequences in Hilbert space. *Bull. Moscow State Univ.* **2** 1–40.

KREIN, M. G. (1944). On an extrapolation problem of A. N. Kolmogorov. *Dokl. Akad. Nauk SSSR* **46** 306–309.

KRONECKER, L. (1881). Zur theorie der elimination einer variablen aus zwei algebraischen gleichungen. *Monatsber. Zonigl. Preussichen Akad. Wiss.* (Berlin), 535–600.

POWER, S. C. (1980). Hankel operators on Hilbert space. *Bull. London Math. Soc.* **12** 422–442.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43–47.

SARASON, D. E. (1972). An addendum to "Past and future." *Math. Scand.* **30** 62–64.

SZEGO, G. (1920). Beitrage zur theorie der toeplitzschen formen I. *Math. Z.* **6** 167–202.

SZEGO. G. (1921). Beitrage zur theorie der toeplitzschen formen II. *Math. Z.* **9** 167–190.

YAGLOM, A. M. (1965). Stationary Gaussian processes satisfying the strong mixing condition and best predictable functionals. *Proc. Int. Research Seminar of the Statistical Laboratory, Univ. of Calif., Berkeley,* 1963, 241–252; Springer-Verlag, New York.

PROGRAM IN BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
BOX 5457
RALEIGH, NORTH CAROLINA 27650