* *Based in part on the author's Ph.D. dissertation at Yale University.*

PROJECTION WITH THE WRONG INNER PRODUCT AND ITS APPLICATION TO REGRESSION WITH CORRELATED ERRORS AND LINEAR FILTERING OF TIME SERIES*

by

William S. Cleveland

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 673

MARCH 1970

Projection with the Wrong Inner Product and its Application to Regression with Correlated Errors and Linear Filtering of Time Series*

by William S. Cleveland
Department of Statistics
University of North Carolina at Chapel Hill

1. INTRODUCTION

In many places in statistics one wants to calculate the orthogonal projection Px of some vector x on a subspace $\mathcal{P}$. Oftentimes the inner product function is specified by the unknown covariances C of a set of random variables. The usual procedure is to estimate C by C^* and approximate Px by P^*x, the orthogonal porjection with respect to C^*; that is, x is projected on $\mathcal{P}$ using a wrong inner product. There is, therefore, interest in knowing when P^* will be a good approximation of P.

In Section 2, the question of calculating orthogonal projections with the wrong inner product in a general Hilbert space is investigated. The results are then applied to the problem of regression with correlated errors in Section 3 and to linear filtering operations on multi-channel, wide-sense stationary, stochastic processes in Section 4.

2. PROJECTION WITH THE WRONG INNER PRODUCT IN A GENERAL HILBERT SPACE

Let $\mathcal{H}$ be a Hilbert space with inner product $(\cdot,\cdot)$ and norm $||\cdot|| = (\cdot,\cdot)^{\frac{1}{2}}$. Let $[\cdot,\cdot]$, which will be thought of as the wrong inner

---

* *Based in part on the author's Ph.D. dissertation at Yale University.*

product, be a bilinear functional with the following properties:

(1) $[\cdot,\cdot]$ is defined on $\mathcal{D}\times\mathcal{D}$ where $\mathcal{D}$ is a linear subset of $\mathcal{H}$ whose closure is $\mathcal{H}$, $[x,y] = \overline{[y,x]}$, and for fixed x the linear functional $[x,\cdot]$ on $\mathcal{D}$ is bounded.

(2) If z_n is a sequence of vectors in $\mathcal{H}$ such that $z_n \to z$ and z_n is a $[\cdot,\cdot]$ Cauchy sequence, that is $[z_n - z_m,\ z_n - z_m] \to 0$ as $n,m \to \infty$, then $z\in\mathcal{D}$.

In (1), it has not been assumed that $[\cdot,\cdot]$ is a bounded bilinear functional. Assumption (2) has been made to ensure that $[\cdot,\cdot]$ is defined everywhere it is possible to do so and maintain the properties in (1).

From (1), for fixed $x\in\mathcal{D}$, the definition of the linear functional $[x,\cdot]$ may be extended boundedly to all of $\mathcal{H}$. From the Riesz Representation Theorem (Halmos, 1957, p.31) there exists $y\in\mathcal{H}$ such that $[x,\cdot] = (y,\cdot)$. Let B be the mapping defined by $Bx = y$. B is linear and self-adjoint, and B is bounded if and only if $[\cdot,\cdot]$ is a bounded bilinear functional.

Let $\mathcal{P}$ be a subspace of $\mathcal{H}$ and P the orthogonal projection operator onto $\mathcal{P}$. Let $\mathcal{P}^* = \mathcal{P}\cap\mathcal{D}$ and let $\mathcal{Q}^*$ be the set of all $x\in\mathcal{D}$ such that $[x,y] = 0$ for all $y\in\mathcal{P}^*$. Let $\mathcal{D}^*$ be the set of all $x\in\mathcal{D}$ such that $x = p+q$ with $p\in\mathcal{P}^*$ and $q\in\mathcal{Q}^*$. This decomposition of x will be unique if and only if $[\cdot,\cdot]$ is positive definite on $\mathcal{P}^*\times\mathcal{P}^*$. For, suppose $[\cdot,\cdot]$ is not positive definite, that is there is a $z\in\mathcal{P}^*$ with $z \neq 0$ such that $[z,z] = 0$. Since the Cauchy-Schwartz inequality holds for $[\cdot,\cdot]$ (Helmberg, 1969, p.10), $[z,y] = 0$ for all $y\in\mathcal{P}^*$. Thus $z\in\mathcal{Q}^*$ and $x = (p+z)+(q-z)$ so that the decomposition is not unique. Suppose $[\cdot,\cdot]$ is positive definite and $x = p'+q'$ with $p'\in\mathcal{P}^*$ and $q'\in\mathcal{Q}^*$. Then

$$0 = [p'+q'-p-q,\ p'+q'-p-q] = [p'-p,\ p'-p] + [q'-q,\ q'-q]$$

which implies $p' = p$ and $q' = q$.

The operator P^*, orthogonal projection with respect to $[\cdot,\cdot]$, will be defined in the following manner. Let $\mathcal{N}$ be the set of vectors $y \in \mathcal{D}$ such that $[y,y] = 0$. Clearly $\mathcal{N}$ is a linear space. Let z_n be a Cauchy sequence in $\mathcal{N}$ and let z be the limit of z_n. From (2), $z \in \mathcal{D}$ and $[z,z] = (Bz,z) = \lim_{n\to\infty} (Bz,z_n) = \lim_{n\to\infty} (z,Bz_n) = 0$. Thus $\mathcal{N}$ is a subspace. Let N be the orthogonal projection operator onto $\mathcal{N}$. Let $x \in \mathcal{D}^*$ have the decomposition $x = p+q$. Then define P^*x to be $p-Np$. We could have taken P^*x to be any vector in $p+\mathcal{N}$ or even all of them. However, in the special cases of Sections 3 and 4, it is seen that $p-Np$, the vector in $p+\mathcal{N}$ with the smallest norm, is a natural assignment.

(3) THEOREM: Let $x \in \mathcal{D}$. Then $x \in \mathcal{D}^*$ (i.e., P^* is defined for x) if and only if there exists a $p \in \mathcal{P}^*$ such that $[x-p,x-p] = \inf\{[x-y,x-y]\colon y \in \mathcal{P}^*\}$, in which case $P^*x = p-Np$.

PROOF: The proof proceeds exactly the same as the analogous theorem for orthogonal projections.

(4) THEOREM: If

$$\inf\left\{\frac{[y,y]}{\|y-Ny\|^2} : \quad y \in \mathcal{P}^*\right\} = \theta$$

is positive, then $\mathcal{D} = \mathcal{D}^*$.

PROOF: Let $x \in \mathcal{D}$ and let

$$\zeta = \inf\{[x-y,x-y]\colon \ y \in \mathcal{P}^*\}.$$

We may proceed exactly as in (Halmos, 1957, p.23, Theorem 1) to show that there is a $[\cdot,\cdot]$ Cauchy sequence $y_n \in \mathcal{P}^*$ with $[x-y_n,x-y_n] \to \zeta$. Let $z_n = y_n-Ny_n$. Then $z \in \mathcal{P}^*$, $[x-z_n,x-z_n] \to \zeta$, and z_n is a $[\cdot,\cdot]$ Cauchy sequence. From the hypothesis, $[z_n-z_m,z_n-z_m] \geq \theta(z_n-z_m,z_n-z_m)$ so that z_n is $(\cdot,\cdot)$ Cauchy. Therefore, there exists $z\epsilon P$ with $z_n \to z$. From (2), $z\epsilon P^*$. Let $v_n = z-z_n$ then

$$\begin{aligned} 0 &= \lim_{n,m\to\infty} [v_n-v_m,\ v_n-v_m] \\ &= \lim_{n\to\infty} (Bv_n,v_n) - \lim_{m\to\infty}\lim_{n\to\infty} (Bv_m,v_n) \\ &\quad - \lim_{n\to\infty}\lim_{m\to\infty} (Bv_n,v_m) + \lim_{m\to\infty} (Bv_m,v_m) \\ &= \lim_{n\to\infty} [v_n,v_n] + \lim_{m\to\infty} [v_m,v_m], \end{aligned}$$

which proves that $[z-z_n,\ z-z_n] \to 0$. Thus

$$\begin{aligned} 0 &= \lim_{n\to\infty} [z-z_n,z-z_n] \\ &= [z,z] - \lim_{n\to\infty} (Bz,z_n) - \lim_{n\to\infty} (z_n,Bz) + \lim_{n\to\infty} [z_n,z_n] \\ &= [z,z] - \lim_{n\to\infty} [z_n,z_n]. \end{aligned}$$

This last result together with the fact that

$$\begin{aligned} \lim_{n\to\infty} [z_n,x] &= \lim_{n\to\infty} (Bz_n,x) \\ &= \lim_{n\to\infty} (z_n,Bx) \\ &= (z,Bx) \\ &= [z,x] \end{aligned}$$

proves

$$\zeta = \lim_{n\to\infty} [z_n-x,z_n-x] = [z-x,z-x].$$

Thus from (3), $x \in \mathcal{D}^*$.

(5) **THEOREM:** Suppose $\mathcal{D}^* = \mathcal{D}$. Then $P = P^*$ on $\mathcal{D}$ if and only if $\overline{B\mathcal{P}^*} = \mathcal{P}$

PROOF: Suppose $P = P^*$ on $\mathcal{D}$. For all $x \in \mathcal{D}$ and $y \in \mathcal{P}^*$

$$(x,By) = (Bx,y) = [x,y] = [Px,y] = (BPx,y) = (Px,By) .$$

Since P and $(\cdot,By)$ are continuous and $\mathcal{D}$ is dense in $\mathcal{H}$, $(x,By) = (Px,By)$ for all $x \in \mathcal{H}$. In particular it holds for all vectors in the orthogonal complement of $\mathcal{P}$ in $\mathcal{H}$, which implies $By \in \mathcal{P}$; hence $B\mathcal{P}^* \subset \mathcal{P}$. To show equality, we first prove $\overline{\mathcal{P}^*} = \mathcal{P}$ and B is one-to-one on $\mathcal{P}^*$. Let $y \in \mathcal{P}$. Since $\mathcal{D}$ is dense in $\mathcal{H}$, there is a sequence $y_n \in \mathcal{D}$ with $y_n \to y$; thus $Py_n \to Py = y$. But $Py_n = P^* y_n \in \mathcal{P}^*$ so that $\overline{\mathcal{P}^*} = \mathcal{P}$. Suppose that for some $x \in \mathcal{P}^*$, $Bx = 0$. Then $[x,y] = 0$ for all $y \in \mathcal{P}^*$ which implies $P^*x = 0$. But $P^*x = Px = x$. Thus B is one-to-one on $\mathcal{P}^*$. The facts that $\overline{\mathcal{P}^*} = \mathcal{P}$ and B is a one-to-one self-adjoint transformation on $\mathcal{P}^*$ with $B\mathcal{P}^* \subset \mathcal{P}$ together with (Helmberg, 1969, Theorem 6, p.121), imply $\overline{B\mathcal{P}^*} = \mathcal{P}$.

Suppose $\overline{B\mathcal{P}^*} = \mathcal{P}$. Let $x \in \mathcal{D}$ then for all $y \in \mathcal{P}^*$, $(x,By) = (Bx,y) = [x,y] = [P^*x,y] = (BP^*x,y) = (P^*x,By)$. Thus $(x,z) = (P^*x,z)$ for all $z \in B\mathcal{P}^*$. From the hypothesis, this last equality may be extended to all $z \in \mathcal{P}$, which implies $P^*x = Px$.

The following inequality is due to Kantorovich (1948, p.142).

(6) **LEMMA:** Let $0 < a_1 \le a_2 \dots \le a_n < \infty$ and $c_i \ge 0$ for $i = 1, \dots, n$. Then

$$\sum_{i=1}^{n} c_i a_i \sum_{i=1}^{n} c_i a_i^{-1} \le \frac{(a_1+a_n)^2}{4a_1 a_n} \sum_{i=1}^{n} c_i ,$$

and equality holds for the proper choice of c_1.

(7) THEOREM: Let $x \in \mathcal{D}^*$. Let

$$\alpha_1 = \inf\left\{\frac{[z,z]}{(z,z)} : z \in \mathcal{D}\right\}$$

and

$$\alpha_2 = \sup\left\{\frac{[z,z]}{(z,z)} : z \in \mathcal{D}\right\}.$$

Then

$$\frac{||x-P^*x||^2}{||x-Px||^2} \le \frac{(\alpha_1+\alpha_2)^2}{4\alpha_1\alpha_2} ,$$

where the right side of the inequality is assigned the value ∞ if either $\alpha_1 = 0$ or $\alpha_2 = \infty$. The inequality is best in the sense that the left side can be made arbitrarily close to the right by the proper choice of $\mathcal{P}$ and x.

PROOF: If $\alpha_1 = 0$ or $\alpha_2 = \infty$ the inequality is clearly true. Thus assume $\alpha_1 > 0$ and $\alpha_2 < \infty$. From (2) and (4), $\mathcal{D}^* = \mathcal{D} = \mathcal{H}$. Let $z_1 = x-Px$ and $z_2 = P^*z_1$. Let R be the orthogonal projection operator onto the space spanned by z_2, and R^* the orthogonal projection operator with respect to $[\cdot,\cdot]$. Then $z_1-R^*z_1 = x-P^*x$ and $z_1-Rz_1 = x-Px$. Thus there is no loss of generality in assuming $\mathcal{P}$ is one-dimensional, x is orthogonal to $\mathcal{P}$, and $||x|| = 1$. Let y be a basis for $\mathcal{P}$, where $||y|| = 1$. Let $\mathcal{S}$ be the space spanned by x and y. Then $Px = 0$ and $P^*x = ([x,y]/[y,y])y$ and

$$\alpha = \frac{||x-P^*x||^2}{||x-Px||^2} = 1 + \frac{|[x,y]|^2}{[y,y]^2} .$$

Since the quadratic forms $Q_1(v) = (v,v)$ and $Q_2(v) = [v,v]$ for $v \in S$ are nonsingular with respect to each other, we may coordinatize S so that if v corresponds to the two-tuple v_1, v_2 then

$$Q_1(v) = |v_1|^2 + |v_2|^2$$

and

$$Q_2(v) = \beta_1 |v_1|^2 + \beta_2 |v_2|^2 ,$$

where $\beta_1 = \min\{Q_1(v)/Q_2(v): v \in S\}$ and β_2 is defined similarly with min replaced by max. Let x correspond to x_1, x_2 and y to y_1, y_2 then

$$\alpha = 1 + \left(\frac{\beta_1 x_1 \bar{y}_1 + \beta_2 x_2 \bar{y}_2}{\beta_1 |y_1|^2 + \beta_2 |y_2|^2} \right)^2 .$$

Since $||x|| = ||y|| = 1$ and $(x,y) = 0$, we have $|x_1| = |y_2|$ which implies

$$\alpha = \frac{\beta_1^2 |y_1|^2 + \beta_2^2 |y_2|^2}{(\beta_1 |y_1|^2 + \beta_2 |y_2|^2)^2} .$$

Letting $a_j = \beta_j$ and $c_j = \beta_j |y_j|^2 / (\beta_1 |y_1|^2 + \beta_2 |y_2|^2)$ we may apply (6) with the result that

$$(8) \qquad \alpha \le \frac{(\beta_1 + \beta_2)^2}{4 \beta_1 \beta_2} .$$

The inequality of the theorem now follows from this and the fact that $\beta_1 \ge \alpha_1$ and $\beta_2 \le \alpha_2$.

It will now be shown that the inequality cannot be improved. (We are still assuming $\alpha_1 > 0$ and $\alpha_2 < \infty$.) For $\alpha_1 = \alpha_2$ it is obvious. Thus assume $\alpha_1 < \alpha_2$. For $\varepsilon > 0$, let $u, v \in H$ be such that

$$\frac{[u,u]}{(u,u)} < \alpha_1 + \varepsilon$$

and

$$\frac{[v,v]}{(v,v)} > \alpha_2 - \varepsilon.$$

Clearly u,v are linearly independent for ε small enough. Now let S be the space spanned by u and v. Then $\beta_1<\alpha_1+\varepsilon$ and $\beta_2>\alpha_2-\varepsilon$. But from (6) equality holds in (8) for the proper choice of y_1,y_2, which shows the inequality cannot be improved for the case $\alpha_1>0$ and $\alpha_2<\infty$.

If $[\cdot,\cdot]$ is positive definite on $\mathcal{D}\times\mathcal{D}$ and either $\alpha_1 = 0$ or $\alpha_2 = \infty$ then by a proof analogous to that in the previous paragraph it may be shown that the inequality cannot be improved.

Suppose $x\epsilon\mathcal{D}$ is such that $[x,x] = 0$ and $x \neq 0$. Let $\mathcal{P}$ be the space spanned by x then $Px = x$ and $P^*x = 0$. Thus the inequality cannot be improved in this case.

(9) **COROLLARY:**

$$||Px-P^*x||^2 \le \frac{(\alpha_1-\alpha_2)^2}{4\alpha_1\alpha_2}||x-Px||^2 \le \frac{(\alpha_1-\alpha_2)^2}{4\alpha_1\alpha_2}||x-P^*x||^2 .$$

PROOF: The first inequality follows easily from (7) and the equality $||x-P^*x||^2 = ||x-Px||^2 + ||Px-P^*x||^2$. The second follows from the fact that $||x-Px||^2 \le ||x-P^*x||^2$.

3. REGRESSION WITH CORRELATED ERRORS

Suppose $x = x_1,\ldots,x_n$ has a multivariate normal distribution with mean m and nonsingualr covariance matrix C. Suppose m lies in $\mathcal{P}$, a subspace of Euclidean n-space $\mathcal{E}$. The norm $(y,y)^{\frac{1}{2}} = ||y|| = (yC^{-1}y')^{\frac{1}{2}}$, where $y \in \mathcal{E}$, is oftentimes referred to as the Mahalanobis distance. It provides a measure of the distance of two normal populations with the same covariance matrix C, in the sense that the closer the means of the two populations under this norm, the more difficult it is to discriminate them. (Rao, 1965, Section 8e.)

Thus if $\hat{m}$ and m^* are estimates of m, a measure of their proximity is $||\hat{m}-m^*||$. In fact, if $\hat{m}$ denotes the posterior mean of m under the assumption that the prior on m in uniform, then $\hat{m}$ is the vector in $\mathcal{P}$ which minimizes $||x-p||$ as p ranges over $\mathcal{P}$. ($\hat{m}$ is also the Gauss-Markov estimate even if normality is not assumed.) That is, $\hat{m} = Px$ where P is the orthogonal projection operator onto $\mathcal{P}$.

Suppose, however, that C is not known but C^* is available where C^* is a nonsingular approximation or estimate of C or perhaps just a convenient choice, as in least squares estimation. Now suppose m is estimated as above with C replaced by C^*. That is, $[y,y] = y(C^*)^{-1}y'$ and m is estimated by $m^* = P^*x$, where P^* is the orthogonal projection operator on $\mathcal{P}$ with respect to $[\cdot,\cdot]$.

In the notation of Section 2, let $\mathcal{H}$ be the subspace of $\mathcal{E}$ spanned by x and $\mathcal{P}$. Let α_1 and α_2 be defined as in (7). Then (7) provides a bound for $||x-m^*||/||x-\hat{m}||$ and (9) gives a bound for $||\hat{m}-m^*||$, the Mahalanobis distance between $\hat{m}$ and m^*. (Compare this with (Watson, 1967, p.1685).)

Instead of choosing $\mathcal{H}$ as above we can take $\mathcal{H} = \mathcal{E}$. This, in general, results in less sharp bounds for a particular x and $\mathcal{P}$, but bounds which now hold whatever the choice of x and $\mathcal{P}$. From (7), these bounds are in fact best among all bounds which do not depend on x and P. The operator B of Section 2 is $(C^*)^{-1}C$ and α_1 and α_2 are the minimum and maximum eigenvalues of this operator. Clearly $(C^*)^{-1}CP = \overline{(C^*)^{-1}CP}$ so that from (4) and (5), $m^* = \hat{m}$ if and only if $(C^*)^{-1}CP = P$, a result first proved by Kruskal (1968).

4. LINEAR FILTERING OPERATIONS

The calculation of projections, in particular linear predictions, interpolations, and signal extractions, in the space spanned by a multi-channel, wide-sense stationary time series X_t has received great attention since the initial works of Wiener (1949) and Kolmogorov (1941). The theory is based on the assumption that the spectral distribution matrix of X_t is known. But it is not really known in practice, and the traditional remedy is to estimate the matrix and then calculate the projections as though the estimate were the true matrix.

Suppose X_t is an r-channel process. Let $F(\lambda) = [F_{jk}(\lambda)]$ be the $r \times r$ spectral distribution matrix of X_t, where $-\infty<\lambda<\infty$ if X_t is a continuous parameter process and $0\le\lambda\le 1$ if it is discrete parameter. Let $F^*(\lambda) = [F^*_{jk}(\lambda)]$ be an estimate of $F(\lambda)$. Assume $\sum_{j=1}^{r} F^*_{jj}$ is absolutely continuous with respect to $\sum_{j=1}^{r} F_{jj}$.

Any projection in the Hilbert space spanned by the process can be described as a projection in $L_2(F)$, the Hilbert space of vector functions $v(\lambda) = v_1(\lambda),\ldots,v_r(\lambda)$ such that

$$||v||^2 = \sum_{j,k=1}^{r} \int \overline{v_j(\lambda)}\, v_k(\lambda)\, dF_{jk}(\lambda) < \infty,$$

where the range of integration is $-\infty$ to ∞ if X_t is continuous parameter and 0 to 1 if it is discrete (Rosanov, 1967, p.28). Let m be any measure such that $\sum_{j=1}^{r} F_{jj}$ is absolutely continuous with respect to m. Let $f(\lambda) = \frac{dF}{dm}(\lambda)$ and $f^*(\lambda) = \frac{dF^*}{dm}(\lambda)$. Then

$$||v||^2 = \int \overline{v(\lambda)}\, f(\lambda) v'(\lambda) dm(\lambda) ,$$

where $v'(\lambda)$ denotes the transpose of $v(\lambda)$.

To apply the results of Section 2, we let $H = L_2(F)$ and define $[\cdot,\cdot]$ by

$$[u,v] = \int \overline{u(\lambda)}\, f^*(\lambda) v'(\lambda) dm(\lambda) ,$$

so that $\mathcal{D}$ is the set of $v \in L_2(F)$ with $[v,v] < \infty$. Using F^* in place of F to calculate a projection is equivalent to using $[\cdot,\cdot]$ in place of $(\cdot,\cdot)$. It is easily seen that assumptions (1) and (2) hold, and the operator B maps v to u, where u at the point λ is

$$u(\lambda) = \left\{\frac{\overline{v(\lambda)} f^*(\lambda) v'(\lambda)}{\overline{v(\lambda)} f(\lambda) v'(\lambda)}\right\} v(\lambda) .$$

With these definitions, the results of Section 2 may now be applied.

The values α_1 and α_2 in (7) can be written in terms of f and f^* as shown by the following theorem.

(10) THEOREM: Let

$$\Delta_1(\lambda) = \min \left\{\frac{\bar{c} f^*(\lambda) c'}{\bar{c} f(\lambda) c'} : c = c_1, \ldots, c_r, \text{ a complex r-tuple}\right\}$$

for all λ. Define $\Delta_2(\lambda)$ similarly with min replaced by max. Then Δ_1 and Δ_2 are m-measurable. Let

$\zeta_1 = \text{ess}_\lambda \inf \Delta_1(\lambda)$ and $\zeta_2 = \text{ess}_\lambda \sup \Delta_2(\lambda)$, where ess inf and sup are with respect to m. Then $\alpha_1 = \zeta_1$ and $\alpha_2 = \zeta_2$. If $f(\lambda)$ is nonsingular, $\Delta_1(\lambda)$ is the minimum eigenvalue of $f^*(\lambda)f^{-1}(\lambda)$ and $\Delta_2(\lambda)$ the maximum.

PROOF: Since the complex r-tuples with rational real and imaginary parts are dense in the space of r-tuples with norm $(\bar{c}f(\lambda)c')^{\frac{1}{2}}$, the minimum in the definition of $\Delta_1(\lambda)$ may be taken over such rational c. Thus Δ_1 is m-measurable since it is the minimum of a countable number of m-measurable functions. Similarly, Δ_2 is m-measurable.

Since for each λ, $\Delta_2(\lambda)$ is the maximum eigenvalue of $f^*(\lambda)$ with respect to $f(\lambda)$, there exists an eigenvector $e(\lambda) = e_1(\lambda),\ldots,e_r(\lambda)$ such that

$$e(\lambda)f^*(\lambda) = \Delta_2(\lambda)e(\lambda)f(\lambda).$$

It is easy to specify a routine for choosing $e(\lambda)$ to ensure that e is m-measurable. Also $e(\lambda)$ may be chosen so that $|e_j(\lambda)| < 1$, since a constant times an eigenvector is also an eigenvector.

Now if $v \epsilon \mathcal{D}$,

$$\begin{aligned}[v,v] &= \int \overline{v(\lambda)}\ f^*(\lambda)\ v'(\lambda)\ dm(\lambda) \\ &= \int \frac{\overline{v(\lambda)}f^*(\lambda)v'(\lambda)}{\overline{v(\lambda)}f(\lambda)v'(\lambda)}\ \overline{v(\lambda)}\ f(\lambda)\ v'(\lambda)\ dm(\lambda) \\ &\le \zeta_2||v||^2.\end{aligned}$$

That $\zeta_2 = \alpha_2$ will be proved by exhibiting a sequence $v_n \epsilon L_2(F)$ such that $||v_n|| = 1$ and $[v_n,v_n] \to \zeta_2$. Define the set S_n as follows: if $\zeta_2 = \infty$ then

$$S_n = \{\lambda: \Delta_2(\lambda) > n\} ;$$

if $\zeta_2<\infty$ then

$$S_n = \{\lambda: \Delta_2(\lambda) > \zeta_2 - \frac{1}{n}\}.$$

There exists an m-measurable subset T_n of S_n with positive m measure such that T_n is contained in a bounded interval. Let T_n also denote the indicator function of the set T_n and define $v_n(\lambda) = ||T_n e||^{-1} T_n(\lambda)e(\lambda)$. $v_n \epsilon L_2(F)$ since e and T_n are m-measurable, $|e_j(\lambda)| < 1$, and T_n lies in a bounded interval. Now

$$[v_n,v_n] = \int \Delta_2(\lambda)\, \overline{v_n(\lambda)} f(\lambda)\, v_n'(\lambda)\, dm(\lambda)$$

so that

$$\operatorname*{ess\,inf}_{\lambda\epsilon T_n} T_n(\lambda)\, \Delta_2(\lambda) \le [v_n,v_n] \le \operatorname*{ess\,sup}_{\lambda} T_n(\lambda)\, \Delta_2(\lambda) \le \zeta_2 .$$

Since the left side tends to ζ_2, so does $[v_n,v_n]$.

A similar proof shows $\alpha_1 = \zeta_1$. The final statement of the theorem is a well-known fact (Rao, 1965, p.15).

A special case of F and F* deserves comment because of its frequent occurrence in practice. Suppose that X_t is a discrete parameter process, m is Lebesgue measure, and

$$p_1 I \le f(\lambda) \le p_2 I$$

for all λ where I is the r×r identity matrix and p_1,p_2 are positive constants. A common way of obtaining f* is to assume X_t is a k-th order autoregression

$$A_0X_t + A_1X_{t-1} + \dots + A_kX_{t-k} = W_t$$

where the $k\times k$ matrices A_j are such that the roots of the polynomial in z

$$\det[A_0 + A_1z + \dots + A_kz^k]$$

lie outside the unit circle, and W_t is r-channel white noise; that is, $E\, W_{t_1}W'_{t_2} = I$ for $t_1 = t_2$ and 0 for $t_1 \neq t_2$. For such a model

$$f(\lambda) = (A_0+A_1e^{2\pi i\lambda}+\dots+A_ke^{2\pi ik\lambda})^{-1}(A'_0+A'_1e^{-2\pi i\lambda}+\dots+A'_ke^{-2\pi ik\lambda})^{-1}.$$

(cf. Whittle, 1953, p.125). The parameters $A_0,\dots,A_k$ are estimated by $A^*_0,\dots,A^*_k$ and $f^*(\lambda)$ is formed by replacing A_j by A^*_j in the above expression for $f(\lambda)$.

The adequacy of the autoregressive model is then checked by calculating the fitted residuals

$$W^*_t = A^*_0X_t + A^*_1X_{t-1} + \dots + A^*_kX_{t-k}$$

and checking them for the white noise assumption. If f^* is to be used for calculating projections then from (7) and (10) the model is adequate if

$$\alpha = \frac{(\alpha_1+\alpha_2)^2}{4\alpha_1\alpha_2} = \frac{(\zeta_1+\zeta_2)^2}{4\zeta_1\zeta_2}$$

is nearly 1. Let $h(\lambda)$ be the derivative of the spectral distribution matrix of W^*_t with respect to Lebesgue measure. It will be shown below that if $\Phi_1(\lambda)$ is the minimum eigenvalue of $h(\lambda)$ and $\Phi_2(\lambda)$ the maximum, then $\Phi_1(\lambda) = \Delta_2^{-1}(\lambda)$ and $\Phi_2(\lambda) = \Delta_1^{-1}(\lambda)$. Thus letting $\phi_1 = \mathrm{ess}_\lambda \inf \Phi_1(\lambda)$ and $\phi_2 = \mathrm{ess}_\lambda \sup \Phi_2(\lambda)$ we have $\phi_1 = \zeta_2^{-1}$ and $\phi_2 = \zeta_1^{-1}$ so that

$$\alpha = \frac{(\phi_1+\phi_2)^2}{4\phi_1\phi_2} .$$

Thus α can be estimated by calculating the fitted residuals and estimating successively $h(\lambda)$, $\Phi_1(\lambda)$ and $\Phi_2(\lambda)$, ϕ_1 and ϕ_2 and finally α.

To see $\Phi_1(\lambda) = \Delta_2^{-1}(\lambda)$, let

$$A^*(\lambda) = A_0^* + A_1^* e^{2\pi i\lambda} + \dots + A_k^* e^{2\pi ik\lambda}.$$

It is easily seen that

$$h(\lambda) = A^*(\lambda) f(\lambda) A^{*\prime}(-\lambda).$$

Now $\Phi_1^{-1}(\lambda)$ is the maximum eigenvalue of

$$h^{-1}(\lambda) = (A^{*\prime}(-\lambda))^{-1} f^{-1}(\lambda) (A^*(\lambda))^{-1},$$

which has the same maximum eigenvalue as

$$(A^*(\lambda))^{-1} (A^{*\prime}(-\lambda))^{-1} f^{-1}(\lambda) = f^*(\lambda) f^{-1}(\lambda),$$

which from (10) is $\Delta_2(\lambda)$. A similar argument shows $\Phi_2(\lambda) = \Delta_1^{-1}(\lambda)$.

As in Section 3, we can define $\mathcal{H}$ to be the space spanned by x and $\mathcal{P}$ rather than all of $L_2(F)$. The new α_1 and α_2 will give better bounds, but bounds which now depend on x and $\mathcal{P}$ and which might not be easily expressed in terms of f and f*.

ACKNOWLEDGMENT

I would like to thank Leonard J. Savage for several helpful suggestions.

REFERENCES

HALMOS, P.R. (1957) *Hilbert Space*. Chelsea, New York.

HELMBERG, GILBERT (1969) *Spectral Theory in Hilbert Space*. Wiley, New York.

KANTOROVICH, L.W. (1948) Functional analysis and applied mathematics (Russian). *Uspehi Mat. Nauk.*, 3(6), 89-189. English translation issued by National Bureau of Standards, Los Angeles, 1952.

KOLMOGOROV, A.N. (1941) Interpolation and extrapolation of stationary random sequences (Russian). *Izv. Akad. Nauk. SSSR. Ser. Mat.* 9, 3-14.

KRUSKAL, W. (1968) When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Stat.*, 39, 70-75.

RAO, C.R. (1965) *Linear Statistical Inference*. Wiley, New York.

WATSON, G.S. (1967) Linear least squares regression, *Ann. Math. Stat.*, 38, 1679-1699.

WHITTLE, P. (1953) The analysis of multiple stationary time series. *J.R.S.S., (B)*, 15, 125-139.

WIENER, N. (1949) *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. M.I.T. Press, Cambridge.