

# APPROXIMATION AND VARIANCE

BY

ARTHUR SARD

**1. Introduction.** A *process* is an operator  $T$  on a Hilbert space  $\mathcal{H}$  (described in the next section) to  $\mathcal{H}$ . We use a process  $T$  to approximate a function  $f=f(x)\in\mathcal{H}$  in the following way: Given  $f+\delta f\in\mathcal{H}$ , where  $\delta f$  is the *error*, we *approximate*  $f$  by  $T(f+\delta f)$ .

We shall study three processes. Each is defined relative to a given fixed closed linear manifold  $\mathcal{M}\subset\mathcal{H}$ . In a particular instance,  $\mathcal{M}$  might, for example, be the manifold of solutions of a linear homogeneous equation related to the source of our interest in approximating  $f$ .

An operator  $T$  is a *least square process* if, for each  $f$  in  $\mathcal{H}$ , the inner product  $(W[f-Tf], f-Tf)$  is minimal among all operators  $T$  such that  $Tf\in\mathcal{M}$ , where  $W$ , the *weight*, is a nonnegative definite operator proper on  $\mathcal{M}$  (§§2, 5). The least square process exists, is unique, and is bounded and linear (Theorem 3).

The least square process is particularly useful because of its linearity and continuity. It is sensitive to the weight  $W$ , however. The choice of  $W$  affects the approximation  $T(f+\delta f)$ . What  $W$  is appropriate? Gauss suggested, in a particular case, that an appropriate  $W$  be one which would make the least square process have the additional property of being an approximation process. Approximation processes are defined as follows.

Assume that the error  $\delta f=\delta f(x, \omega)$  is a function of  $(x, \omega)$ , where  $\omega$  is a point of a probability space  $\Omega$ , and that the *expected value*  $E\delta f$  vanishes identically in  $x$ . (The complete hypothesis on  $\delta f$  is given in §3.) For almost each  $\omega$  in  $\Omega$ ,  $\delta f(x, \omega)=\delta f$  is a possible error in the datum  $f+\delta f$ . Assume also that  $f=f(x)$  is an arbitrary element of  $\mathcal{H}$ . In any particular use of the process  $T$ ,  $\delta f$  will be determined by an element  $\omega$  in  $\Omega$  and  $f$  will be determined in some fashion unrelated to  $\omega$ <sup>(1)</sup>.

---

Presented to the International Congress of Mathematicians, September 4, 1950, under the title *Least square error and variance*; received by the editors November 16, 1951.

(<sup>1</sup>) Alternatively, the results of the present paper apply to certain cases in which both the function  $f$  and the error  $\delta f$  are stochastic and in which the distribution of  $\delta f$  is related to the value of  $f$ . An exact description of one such case is as follows. Let  $X$  be the real line  $-\infty < x < \infty$ . Let  $\Omega_1, \Omega_2$  be probability spaces, with probability functions  $p_1, p_2$ , respectively. Denote expected value on  $\Omega_1 \times \Omega_2$  by  $E$ ; on  $\Omega_2$  by  $E_2$ . Assume that  $f=f(x)=f(x, \omega_1)$  is a complex function of  $(x, \omega_1)$  measurable  $mp_1$  on  $X \times \Omega_1$  ( $m$  is described in §2 below); that  $f(x, \omega_1)\in\mathcal{M}$  for almost each  $\omega_1$  in  $\Omega_1$ , where  $\mathcal{M}$  is a closed linear manifold in  $\mathcal{H}$ ; that the space essentially spanned by  $f(x, \omega_1), \omega_1\in\Omega_1$ , is dense in  $\mathcal{M}$  (i.e., the space spanned by  $f(x, \omega_1), \omega_1\in\Omega_1^*$ , is dense in  $\mathcal{M}$  whenever  $\Omega_1^*$  is a subset of  $\Omega_1$  of probability unity). Assume that  $\delta f=\delta f(x)=\delta f(x, \omega_1, \omega_2)$  is a complex function of  $(x, \omega_1, \omega_2)$  measurable  $mp_1p_2$  on  $X \times \Omega_1 \times \Omega_2$ ; that  $E_2\delta f=0$  on almost all  $\Omega_1$ ; that  $E\|\delta f\|^2 < \infty$ . These assumptions are to replace (i), (ii), (iii) of §3. In the definitions of approximation process and curve fitting process, the condition " $ET(f+\delta f)=f$  or  $E(g+\delta g)=f$  whenever  $f\in\mathcal{M}$ " is to be replaced by " $E_2T(f+\delta f)=f$  or  $E_2(g+\delta g)=f$  on almost all  $\Omega_1$ ." Then the theorems and lemmas of the present paper remain true providing merely that, in Lemma 3,  $mp$  be replaced by  $mp_1p_2$  and  $E\delta g$  by  $E_2\delta g$ . The modifications of the proofs are slight.

Let  $\mathcal{A}$  be the class of all bounded linear operators  $T$  on  $\mathcal{H}$  to  $\mathcal{H}$  which are such that<sup>(2)</sup>

$$ET(f + \delta f) = f \quad \text{whenever } f \in \mathcal{M}.$$

An *approximation process* is an operator  $T$  in  $\mathcal{A}$  which minimizes  $E\|T\delta f\|^2$ . One might describe an approximation process as a bounded linear operator which produces the minimal response to errors, subject to the constraint of being unbiased on  $\mathcal{M}$ .

Let  $V$  be the *variance* of  $\delta f$  (§3). A *necessary and sufficient condition that the unique least square process be itself an approximation process* is that  $VW\mathcal{M} \subset \mathcal{M}$  (Theorem 4). Heretofore only the following has been known [Gauss 3; Aitken 1, 2; Kolmogorov 7]: The least square process is an approximation process if  $W = V^{-1}$ , providing that  $\mathcal{H}$  is finite-dimensional and  $V$  is invertible. Now the condition  $VW\mathcal{M} \subset \mathcal{M}$  includes the condition  $W = V^{-1}$  and is of greater scope: In the infinite-dimensional case, for example,  $V$  will never be invertible when  $E\|\delta f\|^2$  is finite (§11).

The approximation process seems the most important of the three processes that we study. Theorem 1 gives a necessary and sufficient condition that an operator be an approximation process. Theorem 2 gives a necessary and sufficient condition that approximation processes exist. If approximation processes exist, then least square processes which are approximation processes exist (§8). Our principle results do not require that  $V$  be positive definite. Accordingly we admit any degree of dependence among the errors whatever.

The *curve fitting processes* (§9) turn out to be approximation processes; their study illuminates the case in which  $\mathcal{M}$  is finite-dimensional.

Kolmogorov and Wiener, in [6] and [11] respectively, consider problems involving least square error, prediction, and approximation. Both authors take the datum  $f + \delta f$  to be a stationary stochastic process, with mean identically zero. Our investigation is essentially different from theirs.

**2. The space  $\mathcal{H}$ .** Let  $m = m(x)$  be a nondecreasing function defined for all real  $x$  and not identically constant. The *space*  $\mathcal{H}$  is the space of all complex functions  $f = f(x)$  of the real variable<sup>(3)</sup>  $x$  which are measurable relative to  $m$  and are such that  $\|f\|^2 = \int_{-\infty}^{\infty} |f(x)|^2 dm(x) < \infty$ . Here and elsewhere integrals are to be understood in the sense of Lebesgue-Stieltjes. Two functions which are equal almost everywhere  $m$  correspond to the same element, if any, of  $\mathcal{H}$ . The reader may consult references [4; 8; 9; 10] for definitions and details omitted from our discussion.

The space  $\mathcal{H}$  is a Hilbert space of a countable infinity of dimensions or a finite-dimensional complex Euclidean space. The latter case occurs if and

<sup>(2)</sup> The symbol  $E$  operates on its argument as a function of  $\omega$  alone. Thus, for example,  $ET(f + \delta f) = E(Tf + T\delta f) = Tf + ET\delta f$ , since  $Tf$  is independent of  $\omega$ .

<sup>(3)</sup> More generally,  $x$  may be a point of real Euclidean  $N$ -space  $R_N$ ;  $m(x)$  a measure, not identically zero, on  $R_N$ ; and  $f = f(x)$  a complex function of the  $N$  real variables  $x$ .

only if  $m(x)$  is constant except for a finite number of jumps. We refer to the latter as the tabular case. In the tabular case  $f(x)$  is characterized by a table of its values at the finite number of points  $x$  at which  $m(x)$  has jumps. The jumps in  $m(x)$  may be thought to measure the importance of the respective tabular values. Indeed in all cases  $dm(x)$  may be thought to measure the importance of  $f+\delta f$  at  $x$ . All three processes that we consider depend on  $m$  in an essential way since the inner product  $(f, g) = \int_{-\infty}^{\infty} f(x)\bar{g}(x)dm(x)$  depends on  $m$ .

That elements of  $\mathcal{H}$  need be given only almost everywhere  $m$  is an essential point. In the tabular case, for example, the table of values of  $f+\delta f$  is adequate because it is a description of  $f+\delta f$  almost everywhere  $m$ .

By a *bounded linear operator* we mean one which is bounded and linear on all of  $\mathcal{H}$ . We say that an operator  $T$  is *nonnegative definite* if  $T$  is bounded, linear, self-adjoint, and if  $(Tf, f) \geq 0$  for all  $f$  in  $\mathcal{H}$ . We say that  $T$  is *positive definite on* a closed linear manifold  $\mathcal{M}$  if  $T$  is nonnegative definite and if  $(Tf, f) = 0, f \in \mathcal{M}$  imply that  $f = 0$ . We say that  $T$  is *positive definite* if  $T$  is positive definite on  $\mathcal{H}$ .

LEMMA 1. *Suppose that  $T$  is a bounded linear operator and  $\mathcal{M}$  a closed linear manifold. Then the conditions*

$$T\mathcal{M} = 0 \quad \text{and} \quad T^*\mathcal{H} \subset \mathcal{M}^\perp$$

*are equivalent.*

Here we write  $T\mathcal{M}$  for the set of all elements  $Tf, f \in \mathcal{M}$ . The adjoint of  $T$  is  $T^*$ . The orthogonal complement of  $\mathcal{M}$  is  $\mathcal{M}^\perp$ .

**Proof.** Let  $P$  be the operator of projection onto  $\mathcal{M}$ . Then  $T\mathcal{M} = 0$  is equivalent to  $TP = 0$ ; hence to  $(TP)^* = 0 = P^*T^* = PT^*$ ; hence to  $T^*\mathcal{H} \subset \mathcal{M}^\perp$ , since  $\mathcal{M}^\perp, \mathcal{M}$  are the sets of elements  $f \in \mathcal{H}$  such that  $Pf = 0, Pf = f$ , respectively.

**3. The error  $\delta f$ .** Let  $\Omega$  be a probability space of points  $\omega$  with probability function  $p$ . If  $z$  is a function of  $\omega$ , the expected value of  $z$  is  $Ez = \int_{\Omega} z dp$ . We assume of the error  $\delta f$  that

(i)  $\delta f = \delta f(x) = \delta f(x, \omega)$  is a complex function of  $(x, \omega)$  and is measurable  $mp$  on the product space of the real line and  $\Omega$ .

(ii)  $E\delta f = 0$ .

(iii)  $E\|\delta f\|^2 < \infty$ .

Thus  $\delta f$  is a stochastic process. For each  $x$  except those on a set of  $m$ -measure zero,  $\delta f$  is a random variable. For each  $\omega$  except those on a set of  $p$ -measure zero,  $\delta f$  is an error in  $f + \delta f$ .

The *variance*  $V$  of  $\delta f$  is the operator which carries  $h \in \mathcal{H}$  into

$$Vh = E(h, \delta f)\delta f.$$

The operator  $V$  is the integral operator whose kernel is the covariance func-

tion of  $\delta f$ . In the tabular case, the matrix of  $V$  is sometimes called the variance-covariance matrix.

LEMMA 2. *The variance  $V$  is nonnegative definite; also*

$$\text{trace } V = E\|\delta f\|^2 < \infty.$$

**Proof.** The existence of  $Vh$ ,  $h \in \mathcal{H}$ , can be established and the inversions below of orders of operations can be justified by Fubini's theorem, Schwarz's inequality, and Lebesgue's theorem on series of nonnegative terms.

First, the definition of  $V$  implies its linearity.

Second,  $V$  is bounded. Thus, for  $h \in \mathcal{H}$ ,

$$\begin{aligned} Vh &= E(h, \delta f)\delta f = E'(h, \delta f')\delta f'; \\ \|Vh\|^2 &= (Vh, Vh) = (E(h, \delta f)\delta f, E'(h, \delta f')\delta f') \\ &= EE'(h, \delta f)(\delta f', h)(\delta f, \delta f') \\ &\leq EE'|(h, \delta f)(h, \delta f')| \|\delta f\| \|\delta f'\| = [E|(h, \delta f)| \|\delta f\|]^2. \end{aligned}$$

Hence

$$\|Vh\| \leq E|(h, \delta f)| \|\delta f\| \leq E\|h\| \|\delta f\| \|\delta f\| = B\|h\|,$$

where  $B = E\|\delta f\|^2 < \infty$ .

Third,  $V = V^*$ . Thus, for  $h, h' \in \mathcal{H}$ ,

$$\begin{aligned} (Vh, h') &= (E(h, \delta f)\delta f, h') = E(h, \delta f)(\delta f, h'); \\ (h, Vh') &= \overline{(Vh', h)} = \overline{E(h', \delta f)(\delta f, h)} = (Vh, h'). \end{aligned}$$

Fourth,  $V$  is nonnegative definite. Thus, for  $h \in \mathcal{H}$ ,

$$(Vh, h) = E(h, \delta f)(\delta f, h) = E|(h, \delta f)|^2 \geq 0.$$

Last, consider trace  $V$ . Suppose that  $\{e_r\}$ ,  $r=1, 2, \dots$ , is a complete orthonormal set in  $\mathcal{H}$ . Then [9, pp. 93-101],

$$\text{trace } V = \sum_r (Ve_r, e_r) = \sum_r E|(\delta f, e_r)|^2 = E \sum_r |(\delta f, e_r)|^2 = E\|\delta f\|^2 < \infty.$$

REMARK. Any nonnegative definite operator of finite trace is the variance of some stochastic process  $\delta f$  satisfying conditions (i), (ii), (iii) above.

LEMMA 3. *Suppose that  $\delta g = A\delta f$ , where  $A$  is a bounded linear operator. Then  $\delta g$  is itself a stochastic process,  $\delta g$  is measurable  $mp$ ,  $E\delta g = 0$ , and  $E\|\delta g\|^2 < \infty$ . The variance of  $\delta g$  is  $AVA^*$ .*

**Proof.** The theory of the direct product of two Hilbert spaces shows that  $\delta g$  is indeed defined almost everywhere  $mp$  and is measurable  $mp$ ; also that

$$E\delta g = EA\delta f = AE\delta f = 0.$$

Now

$$E\|\delta g\|^2 \leq E\|A\|^2\|\delta f\|^2 = \|A\|^2 E\|\delta f\|^2 < \infty.$$

Finally, the variance of  $\delta g$  is the operator which carries  $h \in \mathcal{H}$  into

$$E(h, \delta g)\delta g = E(h, A\delta f)A\delta f = EA(A^*h, \delta f)\delta f = AE(A^*h, \delta f)\delta f = AVA^*h.$$

**4. Approximation processes.** Let  $\mathcal{M}$  be a fixed closed linear manifold of  $\mathcal{H}$ . Let  $\delta f$  be the stochastic process of the preceding section. Suppose that  $A$  is a bounded linear operator. For any  $f$  in  $\mathcal{H}$ , put

$$g + \delta g = A(f + \delta f), \quad g = Af, \quad \delta g = A\delta f.$$

Let  $\mathcal{A}$  be the class of bounded linear operators  $A$  such that

$$(1) \quad E(g + \delta g) = f \quad \text{whenever } f \in \mathcal{M}.$$

Among the operators in  $\mathcal{A}$ , there may be some which minimize the quantity  $E\|\delta g\|^2$ . We call such operators *approximation processes*.

Note that

$$E(g + \delta g) = Eg + E\delta g = g + 0 = Af.$$

Hence condition (1) amounts to this: That  $Af = f$  whenever  $f \in \mathcal{M}$ ; that is, that

$$(2) \quad (A - I)\mathcal{M} = 0,$$

where  $I$  is the identity operator. Condition (1) therefore implies that the error  $g + \delta g - f$  in the approximation of  $f$  by  $g + \delta g$  is  $\delta g$  whenever  $f \in \mathcal{M}$ .

**LEMMA 4.** *A bounded linear operator  $A$  is an approximation process if and only if  $\text{trace } AVA^* = E\|\delta g\|^2$  is minimal, subject to the constraint (2).*

**Proof.** Lemmas 2, 3 imply that  $\text{trace } AVA^* = E\|\delta g\|^2$ . Conditions (1) and (2) are equivalent.

**THEOREM 1.** *A bounded linear operator  $A$  is an approximation process if and only if*

$$(3) \quad VA^*\mathcal{H} \subset \mathcal{M} \quad \text{and} \quad (A - I)\mathcal{M} = 0.$$

*The approximation process is unique if  $V$  is positive definite on  $\mathcal{M}^\perp$ .*

Note that the condition  $VA^*\mathcal{H} \subset \mathcal{M}$  is equivalent to  $AV\mathcal{M}^\perp = 0$ , by Lemma 1.

**Proof.** Suppose that a bounded linear operator  $A$  satisfies the constraint (2). A bounded linear operator  $A + \Delta A$  will also satisfy that constraint:

$$(A + \Delta A - I)\mathcal{M} = 0,$$

if and only if  $\Delta A\mathcal{M} = 0$ , that is,

$$(4) \quad \Delta A^* \mathcal{H} \subset \mathcal{M}^\perp,$$

by Lemma 1. Assume (4).

Lemmas 2, 3 imply that

$$v = \text{trace } A V A^* = E \|\delta g\|^2 < \infty;$$

also that  $\text{trace } (A + \Delta A) V (A + \Delta A)^*$  and  $\text{trace } \Delta A V \Delta A^*$  are finite. Hence

$$(5) \quad \begin{aligned} \Delta v &= \text{trace } (A + \Delta A) V (A + \Delta A)^* - \text{trace } A V A^* \\ &= \text{trace } \Delta A V \Delta A^* + \text{trace } (A V \Delta A^* + \Delta A V A^*). \end{aligned}$$

We now prove the sufficiency of (3). Assume (3). Then  $A V \mathcal{M}^\perp = 0$ , and

$$A V \Delta A^* \mathcal{H} \subset A V \mathcal{M}^\perp = 0,$$

by (4); hence

$$A V \Delta A^* = 0 = (A V \Delta A^*)^* = \Delta A V A^*.$$

Hence, by (5),

$$(6) \quad \Delta v = \text{trace } \Delta A V \Delta A^* = \sum_r (\Delta A V \Delta A^* e_r, e_r) = \sum_r (V \Delta A^* e_r, \Delta A^* e_r),$$

where  $\{e_r\}$  is a complete orthonormal set in  $\mathcal{H}$ . As  $V$  is nonnegative definite, each term on the right is nonnegative and is minimal for  $\Delta A = 0$ . Hence  $\Delta v \geq 0$ ,  $v$  is minimal, and  $A$  is an approximation process. (Indeed  $A$  minimizes  $(A V A^* e, e)$ , subject to the constraint (2), where  $e$  is any fixed element of  $\mathcal{H}$ .)

We now prove the necessity of (3). Suppose that  $A$  is an approximation process. Then  $\Delta v \geq 0$  for all bounded linear operators  $\Delta A$  for which (4) holds. Hence (4) must imply that

$$\text{trace } (A V \Delta A^* + \Delta A V A^*) = 0;$$

that is,

$$(7) \quad \sum_r (A V \Delta A^* e_r + \Delta A V A^* e_r, e_r) = 2 \sum_r \text{Real } (A V \Delta A^* e_r, e_r) = 0.$$

(Otherwise we could make  $\Delta v < 0$  by putting  $\gamma \Delta A$  for  $\Delta A$  in (5) and letting the real number  $\gamma$  be sufficiently small and of proper sign.)

It follows that (4) must imply that

$$(8) \quad (A V \Delta A^* e_r, e_r) = 0 \quad \text{for each } r.$$

To establish this assertion, consider another  $\Delta A$ , viz.  $\Delta A_1$ , defined as follows. Let  $r_0$  be fixed; put

$$\Delta A_1 h = (\Delta A h, e_{r_0}) e_{r_0} = (h, \Delta A^* e_{r_0}) e_{r_0}, \quad h \in \mathcal{H}.$$

Then

$$\Delta A_1^* h = (h, e_{r_0}) \Delta A^* e_{r_0};$$

and (4) holds with  $\Delta A$  replaced by  $\Delta A_1$ . Hence (7) must hold with  $\Delta A$  replaced by  $\Delta A_1$ ; that is,

$$0 = \sum_r \operatorname{Real} (AV \Delta A_1^* e_r, e_r) = \operatorname{Real} (AV \Delta A^* e_{r_0}, e_{r_0}),$$

since  $\Delta A_1^* e_r = (e_r, e_{r_0}) \Delta A^* e_{r_0} = \delta_{r,r_0} \Delta A^* e_{r_0}$ . The entire argument holds with  $\Delta A$  replaced by  $(-1)^{1/2} \Delta A$ . Hence  $(AV \Delta A^* e_{r_0}, e_{r_0}) = 0$ . Since this holds for any  $r_0$ , (8) holds.

We conclude that  $AV \mathcal{M}^\perp = 0$ . Thus, consider  $f \in \mathcal{M}^\perp$  and the particular bounded linear operator  $\Delta A$ , defined as follows:

$$\Delta A h = (h, f) e_{r_0}, \quad h \in \mathcal{H}, r_0 \text{ fixed.}$$

Then  $\Delta A^* h = (h, e_{r_0}) f \in \mathcal{M}^\perp$ , and

$$(AV \Delta A^* e_r, e_r) = (AV(e_r, e_{r_0}) f, e_r) = \delta_{r,r_0} (AV f, e_r) = 0,$$

by (8). Hence

$$(AV f, e_{r_0}) = 0.$$

Since this holds for any  $r_0$ ,  $AV f = 0$ .

To complete the proof of the theorem, we show that the approximation process is unique if  $V$  is positive definite on  $\mathcal{M}^\perp$ . Suppose that  $A$  is an approximation process. Consider any bounded linear operator  $A + \Delta A$  such that (4) holds. Then (6) holds. Now  $\Delta A^* e_r \in \mathcal{M}^\perp$  by (4). If  $V$  is positive definite on  $\mathcal{M}^\perp$ ,  $\Delta v > 0$  unless  $\Delta A^* e_r = 0$  for all  $r$ , by (6); that is,  $\Delta v > 0$  unless  $\Delta A = 0$ . Hence only  $A$  in  $\mathcal{A}$  minimizes  $v$ : The approximation process is unique.

**COROLLARY.** *Suppose that  $A$  is an approximation process. Then so is  $A^q$ , where  $q$  is any positive integer. Also*

$$AVA^* = AV = VA^*.$$

*The variance of  $\delta f - \delta g$  is*

$$(I - A)V(I - A)^* = V - AV,$$

*so that the variance of  $\delta f$  is the sum of the variances of  $\delta g$  and  $\delta f - \delta g$ .*

**Proof.** If  $A$  satisfies conditions (3), so does  $A^q$ . Also

$$AVA^* = A(VA^*) = VA^*$$

by (3); and  $VA^* = AV$  since  $AVA^*$  is self-adjoint. That  $(I - A)V(I - A)^*$  is the variance of  $\delta f - \delta g = (I - A)\delta f$  follows from Lemma 3.

**5. Operators proper on  $\mathcal{M}$ .** We say that an operator  $U$  is *proper* on  $\mathcal{M}$  if a positive constant  $b$  exists such that

$$(9) \quad \|PUf\| \geq b\|f\| \quad \text{whenever } f \in \mathcal{M},$$

where  $P$  is the projection onto  $\mathcal{M}$ . Thus a necessary condition that  $U$  be proper on  $\mathcal{M}$  is that  $Uf \in \mathcal{M}^\perp, f \in \mathcal{M}$  imply that  $f=0$ . This condition is sufficient, if  $U$  is bounded and linear and  $\mathcal{M}$  is finite-dimensional (see Lemma 17).

**LEMMA 5.** *If a bounded linear operator  $U$  is proper on  $\mathcal{M}$ , then  $U\mathcal{M}$  and  $PU\mathcal{M}$  are closed linear manifolds.*

**Proof.** This elementary lemma is an immediate consequence of [4, Theorem 2, pp. 37–38], since  $\|Uf\| \geq \|PUf\|, f \in \mathcal{M}$ .

Now consider a bounded linear operator  $U$ . We denote by  $U^\sim$  any bounded linear operator such that

$$(10) \quad (U^\sim PU - I)\mathcal{M} = 0.$$

Thus  $U^\sim$  is a left inverse of  $PU$  on  $\mathcal{M}$ .

**LEMMA 6.** *A necessary and sufficient condition that a bounded linear operator  $U$  be proper on  $\mathcal{M}$  is that an operator  $U^\sim$  exist.*

This lemma is essentially Toeplitz's theorem [5, p. 1430] applied to  $PU$  on  $\mathcal{M}$ . For simplicity we give the proof. Suppose that  $U$  is proper on  $\mathcal{M}$ . Define  $U^\sim$  on the closed linear manifold  $\mathcal{N}=PU\mathcal{M}$  as follows. Suppose that  $g \in \mathcal{N}$ . Then at least one element  $f \in \mathcal{M}$  exists such that  $PUf=g$ . That element  $f$  is unique. For, if  $PUf'=g, f' \in \mathcal{M}$ , then  $PU(f-f')=0$  and  $f-f'=0$ , by (9). Put

$$U^\sim g = f, \quad g \in \mathcal{N}.$$

Then  $U^\sim$  is linear on  $\mathcal{N}$ . Also

$$\|U^\sim g\| = \|f\| \leq b^{-1}\|PUf\| = b^{-1}\|g\|, \quad g \in \mathcal{N};$$

hence  $U^\sim$  is bounded on  $\mathcal{N}$ .

To complete the definition,  $U^\sim$  may be an arbitrary bounded linear operator on  $\mathcal{N}^\perp$ . Then  $U^\sim$  is a bounded linear operator on  $\mathcal{H}$ , and  $U^\sim PUf=f, f \in \mathcal{M}$ . Since  $U^\sim \mathcal{N}=\mathcal{M}$ , it is always possible to define  $U^\sim$  so that  $U^\sim \mathcal{H}=\mathcal{M}$ .

Conversely, suppose that an operator  $U^\sim$  exists. Then for  $f \in \mathcal{M}$ ,

$$\|U^\sim PUf\| = \|f\| \leq B\|PUf\|, \quad B > 0,$$

since  $U^\sim$  is bounded. Hence (9) holds with  $b=1/B$ .

**6. Companions.** A bounded linear operator  $U$  is a *companion* (of  $V$  relative to  $\mathcal{M}$ ) if  $U$  is proper on  $\mathcal{M}$  and if

$$(11) \quad VU^*\mathcal{M} \subset \mathcal{M}.$$

**LEMMA 7.** *If  $U$  is a companion,  $A=U^\sim PU$  is an approximation process.*



**Proof.** By Theorem 1, it is sufficient to show that  $(A - I)\mathcal{M} = 0$  and  $VA^*\mathcal{H} \subset \mathcal{M}$ . The first is precisely (10). Also

$$VA^*\mathcal{H} = VU^*PU^*\mathcal{H} \subset VU^*\mathcal{M} \subset \mathcal{M},$$

by (11) and the fact that  $P$  is the projection onto  $\mathcal{M}$ .

If in addition  $U^*\mathcal{M} \subset \mathcal{M}$ , then  $(U^*PU)^2 = U^*PU$ , by (10) since  $U^*PU\mathcal{H} \subset \mathcal{M}$ .

**LEMMA 8.** *If an approximation process exists, companions exist. Indeed, if  $A$  is an approximation process,*

$$U = XA + Y$$

*is a companion, providing that  $X$  and  $Y$  are bounded linear operators,  $X$  is proper on  $\mathcal{M}$ , and  $Y^*\mathcal{M} = 0$ .*

**Proof.** Such operators  $X, Y$  exist; one pair is:  $X = I, Y = 0$ . (Thus  $A$  is itself a companion, as is evident directly. But  $A^*PA$  need not be  $A$ .)

We now show that  $U = XA + Y$  is a companion. Thus, by Theorem 1,

$$VU^*\mathcal{M} = VA^*X^*\mathcal{M} + VY^*\mathcal{M} = VA^*X^*\mathcal{M} + 0 \subset VA^*\mathcal{H} \subset \mathcal{M}.$$

Also, if  $f \in \mathcal{M}$ ,

$$PUf = PXAf + PYf = PXf,$$

since  $Af = f, Y\mathcal{H} \subset \mathcal{M}^\perp$ ; hence

$$\|PUf\| = \|PXf\| \geq b\|f\|, \quad b > 0,$$

since  $X$  is proper on  $\mathcal{M}$ .

**LEMMA 9.** *If an approximation process exists, positive definite companions exist. Indeed, if  $A$  is an approximation process,*

$$(12) \quad U = A^*ZA + Y$$

*is a positive definite companion, providing that the operator  $Z$  is proper on  $\mathcal{M}$  and positive definite, and the operator  $Y$  is positive definite on  $\mathcal{M}^\perp$  and  $Y\mathcal{M} = 0$ . All positive definite companions are of the form (12), with  $A, Z, Y$  satisfying the stated conditions.*

**Proof.** Such operators  $Z, Y$  exist; one pair is:  $Z = I, Y = I - P$ , the projection onto  $\mathcal{M}^\perp$ .

Suppose that  $A$  is an approximation process. Then, by Theorem 1,

$$(A - I)P = 0 = [(A - I)P]^* = P(A^* - I).$$

Hence,  $PA^* = P$  and  $PA^*Z = PZ$ . Hence  $A^*Z$  is proper on  $\mathcal{M}$ , since  $Z$  is. By Lemma 8, then,  $U$  is a companion, since  $Y^*\mathcal{M} = Y\mathcal{M} = 0$ .

Now  $U$  is positive definite. For,  $U^* = U$ . Also

$$(Uf, f) = (A^*ZAf, f) + (Yf, f) = (ZAf, Af) + (Yf, f).$$

Each of the last terms is nonnegative. And  $(ZAf, Af) > 0$  unless  $Af = 0$ . Now if  $Af = 0$ , then  $(Yf, f) > 0$  unless  $f = 0$ . Thus, write

$$f = f_1 + f_2, \quad f_1 \in \mathcal{M}, f_2 \in \mathcal{M}^\perp.$$

Then

$$(Yf, f) = (Yf_1, f_1) + (Yf_1, f_2) + (Yf_2, f_1) + (Yf_2, f_2) = (Yf_2, f_2),$$

since  $Yf_1 = Y^*f_1 = 0$ . Also

$$A(f_1 + f_2) = 0 = Af_1 + Af_2 = f_1 + Af_2,$$

and

$$f_1 = -Af_2.$$

By hypothesis  $(Yf_2, f_2) = 0$  implies that  $f_2 = 0$ ; hence that

$$f_1 = 0 = f.$$

We omit the proof that all positive definite companions are of the form (12).

Lemmas 7, 8, 9 imply the following theorem.

**THEOREM 2.** *Approximation processes exist if and only if companions exist. If  $U$  is a companion,  $U^*PU$  is an approximation process. If approximation processes exist, positive definite companions exist.*

An example in which no approximation process exists is given in §12. An example in which a unique approximation process exists even though  $V$  has no bounded inverse is given in §11.

**7. Least square processes.** Let  $W$  be a fixed nonnegative definite operator, proper on  $\mathcal{M}$ . By a *least square process* we mean an operator  $C$  which assigns to each  $f$  in  $\mathcal{H}$  an element  $g = Cf$  with the following properties:

- (i)  $g \in \mathcal{M}$ ;
- (ii) For each fixed  $f$  in  $\mathcal{H}$ ,  $(W[f - g], f - g)$  is minimal among all elements  $g \in \mathcal{M}$ .

A series of lemmas on  $W$  leads to Theorem 3 below.

**LEMMA 10.** *The weight  $W$  is positive definite on  $\mathcal{M}$ .*

**Proof.** Since  $W$  is nonnegative definite,  $(Wf, f) = 0$  only if  $Wf = 0$  [9, p. 53]. Suppose that  $f \in \mathcal{M}$ ,  $(Wf, f) = 0$ . Then  $\|PWf\| = 0 \leq b\|f\|$ ,  $b > 0$ ; hence  $f = 0$ . The proof is complete.

The manifold  $W\mathcal{M}$  is closed and linear, by Lemma 5. Put

$$\mathcal{L} = (W\mathcal{M})^\perp.$$

Then  $(\mathcal{L}, W\mathcal{M}) = 0 = (W\mathcal{L}, \mathcal{M})$  and

$$W\mathcal{L} \subset \mathcal{M}^\perp.$$

LEMMA 11. *If*

$$f + g = f' + g'; \quad f, f' \in \mathcal{M}; g, g' \in \mathcal{L};$$

*then*

$$f = f', \quad g = g'.$$

**Proof.** We shall show that

$$f + g = 0, \quad f \in \mathcal{M}, g \in \mathcal{L},$$

imply that  $f = g = 0$ . Thus

$$\begin{aligned} 0 &= (W[f + g], f + g) = (Wf, f) + (Wf, g) + (Wg, f) + (Wg, g) \\ &= (Wf, f) + (Wg, g), \end{aligned}$$

since  $(W\mathcal{M}, \mathcal{L}) = 0$ . Now each of the last terms in nonnegative. Hence

$$(Wf, f) = (Wg, g) = 0.$$

But  $W$  is positive definite on  $\mathcal{M}$ , by Lemma 10. Hence  $f = 0$ . Hence  $g = 0$ .

LEMMA 12. *The manifold  $\{\mathcal{M}, \mathcal{L}\}$  spanned by  $\mathcal{M}$  and  $\mathcal{L}$  is closed.*

**Proof.** Suppose that

$$f_\nu \in \mathcal{M}, \quad g_\nu \in \mathcal{L}, \quad f_\nu + g_\nu \rightarrow h \text{ as } \nu \rightarrow \infty.$$

We shall show that  $h = f + g$ ,  $f \in \mathcal{M}$ ,  $g \in \mathcal{L}$ . Thus

$$PW(f_\nu + g_\nu) = PWf_\nu + PWg_\nu = PWf_\nu,$$

since  $W\mathcal{L} \subset \mathcal{M}^\perp$ . Hence

$$W^\sim PW(f_\nu + g_\nu) = W^\sim PWf_\nu = f_\nu.$$

Now  $W^\sim PW(f_\nu + g_\nu) \rightarrow W^\sim PWh = h'$ , say, since  $W^\sim PW$  is continuous. Hence  $f_\nu \rightarrow h'$  and  $h' \in \mathcal{M}$ , since  $\mathcal{M}$  is closed. Hence  $g_\nu \rightarrow h - h' \in \mathcal{L}$ , since  $\mathcal{L}$  is closed. Hence

$$h = h' + (h - h') \in \{\mathcal{M}, \mathcal{L}\}.$$

LEMMA 13. *Any  $f$  in  $\mathcal{H}$  can be written*

$$(13) \quad f = f_1 + f_2, \quad f_1 \in \mathcal{M}, f_2 \in \mathcal{L}.$$

**Proof.** It will be sufficient to show that  $\{\mathcal{M}, \mathcal{L}\}^\perp = 0$ , for then  $\{\mathcal{M}, \mathcal{L}\}^{\perp\perp} = \mathcal{H} = \{\mathcal{M}, \mathcal{L}\}$ , since  $\{\mathcal{M}, \mathcal{L}\}$  is closed. Suppose that

$$f \in \mathcal{M}^\perp \quad \text{and} \quad f \in \mathcal{L}^\perp = (W\mathcal{M})^{\perp\perp} = W\mathcal{M}.$$

Then  $f = Wg \in \mathcal{M}^\perp$ , where  $g \in \mathcal{M}$ . Hence  $g = 0$ , since  $W$  is proper on  $\mathcal{M}$ . Hence  $f = 0$ . This completes the proof.

The elements  $f_1$  and  $f_2$  of Lemma 13 are unique, by Lemma 11. Define the operator  $C$  as follows:

$$Cf = f_1, \quad f \in \mathfrak{K},$$

where  $f_1$  is determined by (13). Then

$$(14) \quad Cf = 0 \quad \text{if and only if } f \in \mathcal{L}.$$

LEMMA 14. The operator  $C$  is bounded and linear; it is uniquely characterized by the additional properties:

$$(15) \quad (C - I)\mathcal{M} = 0,$$

$$(16) \quad C\mathfrak{K} \subset \mathcal{M},$$

$$(17) \quad WC = C^*W.$$

**Proof.** Clearly  $C$  is linear on  $\mathfrak{K}$ . Suppose that  $f_r \rightarrow 0$ ,  $f_r \in \mathfrak{K}$ . Then  $Cf_r \rightarrow 0$ . Thus

$$f_r = Cf_r + (I - C)f_r, \quad Cf_r \in \mathcal{M}, (I - C)f_r \in \mathcal{L}.$$

Then  $PWf_r = PWCf_r$ , since  $PW\mathcal{L} = 0$ ; and

$$W \sim PWf_r = W \sim PWCf_r = Cf_r,$$

since  $Cf_r \in \mathcal{M}$ . Hence  $Cf_r \rightarrow 0$ , since  $W \sim PW$  is continuous. Hence  $C$  is bounded.

Clearly  $C$  has properties (15) and (16). To establish (17), suppose that  $f = f_1 + f_2$ ,  $g = g_1 + g_2$ ,  $f_1, g_1 \in \mathcal{M}$ ,  $f_2, g_2 \in \mathcal{L}$ . Then

$$\begin{aligned} (WCf, g) &= (Wf_1, g) = (f_1, Wg) = (f_1, Wg_1) + (f_1, Wg_2) = (f_1, Wg_1) \\ &= (f_1, Wg_1) + (f_2, Wg_1) = (f, Wg_1) = (f, Wg), \end{aligned}$$

since  $(\mathcal{M}, W\mathcal{L}) = 0$ . Hence  $(WC)^* = WC = C^*W$ .

Conversely, suppose that  $C$  is a bounded linear operator with properties (15), (16), (17). Then

$$f = Cf + (I - C)f, \quad f \in \mathfrak{K}.$$

It will be sufficient to show that  $Cf \in \mathcal{M}$ ,  $(I - C)f \in \mathcal{L}$ , by Lemma 11. Now  $Cf \in \mathcal{M}$ , by (16). Let  $g$  be any element of  $W\mathcal{M}$ ; that is,  $g = Wg'$ ,  $g' \in \mathcal{M}$ . Then

$$\begin{aligned} ((I - C)f, g) &= (f, g) - (Cf, g) = (f, Wg') - (Cf, Wg') \\ &= (f, Wg') - (WCf, g') = (f, Wg') - (f, WCg') \\ &= (f, Wg') - (f, Wg') = 0, \end{aligned}$$

by (17) and (15). Hence  $(I - C)f \in (W\mathcal{M})^\perp = \mathcal{L}$ .

**THEOREM 3.** *The least square process exists, is unique, and is bounded and linear. A necessary and sufficient condition that a bounded linear operator  $C$  be the least square process is that<sup>(4)</sup>*

$$(C - I)\mathcal{M} = 0, \quad C\mathcal{K} \subset \mathcal{M}, \quad WC = C^*W.$$

**Proof.** Given  $f \in \mathcal{K}$ ,  $g$  is to be an element of  $\mathcal{M}$  which minimizes

$$y = (W[f - g], f - g).$$

Write  $f = f_1 + f_2$ ,  $f_1 \in \mathcal{M}$ ,  $f_2 \in \mathcal{L}$ . Then

$$\begin{aligned} y &= (W[f_1 - g], f_1 - g) + (W[f_1 - g], f_2) + (Wf_2, [f_1 - g]) + (Wf_2, f_2) \\ &= (W[f_1 - g], f_1 - g) + (Wf_2, f_2), \end{aligned}$$

since  $(W\mathcal{M}, \mathcal{L}) = 0$ . Now the second term is independent of  $g$ ; and the first term has a minimum uniquely at  $g = f_1$ , by Lemma 10.

Lemma 14 completes the proof. We note in passing that  $(WCf, Cf) = (WCf, f) = (Wf, Cf)$ . Hence the minimal  $y$  is  $(Wf, f) - (WCf, f)$ .

### 8. Approximation and least square processes.

**THEOREM 4.** *The unique least square process is an approximation process if and only if*

$$VW\mathcal{M} \subset \mathcal{M}.$$

Otherwise put, the least square process is an approximation process if and only if the weight  $W$  is a companion. Since positive definite companions exist whenever approximation processes do (Lemma 9), *it is always possible, given the existence of one approximation process, to construct a least square process (with positive definite weight) which is also an approximation process.*

**Proof of Theorem 4.** Suppose that  $VW\mathcal{M} \subset \mathcal{M}$ . To show that the least square process  $C$  is an approximation process, it is sufficient to show that  $(C - I)\mathcal{M} = 0$  and  $VC^*\mathcal{K} \subset \mathcal{M}$ , by Theorem 1. Now the first is true, by Theorem 3. Also  $W\mathcal{M} = \mathcal{L}^\perp$ , and  $C\mathcal{L} = 0$ , by (14). Hence  $C^*\mathcal{K} \subset \mathcal{L}^\perp$ , by Lemma 1. Hence

$$VC^*\mathcal{K} \subset V\mathcal{L}^\perp = VW\mathcal{M} \subset \mathcal{M}.$$

Conversely, suppose that the least square process  $C$  is an approximation process. Then  $VC^*\mathcal{K} \subset \mathcal{M}$  and  $CV\mathcal{M}^\perp = 0$ . Hence  $V\mathcal{M}^\perp \subset \mathcal{L}$ , by (14). Since  $W\mathcal{M} = \mathcal{L}^\perp$ ,

$$(V\mathcal{M}^\perp, W\mathcal{M}) = 0 = (\mathcal{M}^\perp, VW\mathcal{M}).$$

Hence  $VW\mathcal{M} \subset \mathcal{M}$ .

**9. Curve fitting processes.** The curve fitting processes will be defined in

<sup>(4)</sup> Thus  $C$  is a pseudo-projection: If  $W$  were positive definite,  $C$  would be the projection onto  $\mathcal{M}$  relative to the new inner product  $((f, f')) = (Wf, f')$ ,  $f, f' \in \mathcal{K}$ .

terms of independent elements  $\mu_1, \mu_2, \dots$  that span  $\mathcal{M}$  rather than in terms of  $\mathcal{M}$  itself. By requiring that  $\mu_1, \mu_2, \dots$  be orthonormal and complete, one could treat the infinite-dimensional case. We consider it of more interest to admit  $\mu_1, \mu_2, \dots$  which are not necessarily orthonormal.

In this and the next section,  $\mathcal{M}$  will be assumed to be  $n$ -dimensional. The indices  $i, j, k, l$  will range over  $1, \dots, n$  ( $1 \leq n < \infty$ ). Assume that  $\mu_1, \dots, \mu_n$  are fixed (independent) elements of  $\mathcal{K}$  that span  $\mathcal{M}$ .

Let  $\phi_1^*, \dots, \phi_n^*$  be a set of functionals, each bounded and linear on  $\mathcal{K}$ . For each  $f$  in  $\mathcal{K}$ , put

$$\begin{aligned}\alpha_i + \delta\alpha_i &= \phi_i^*(f + \delta f), & \alpha_i &= \phi_i^*f, \\ g + \delta g &= \left( \sum_i \mu_i \phi_i^* \right) (f + \delta f) = \sum_i (\alpha_i + \delta\alpha_i) \mu_i, \\ g &= \left( \sum_i \mu_i \phi_i^* \right) f = \sum_i \alpha_i \mu_i.\end{aligned}$$

Consider those sets of functionals  $\{\phi_i^*\}$  which are such that

$$E(g + \delta g) = f \quad \text{whenever } f \in \mathcal{M}.$$

Among those sets of functionals there may be some which minimize the quantity  $E \sum_i |\delta\alpha_i|^2$ . By a *curve fitting process*, we mean an operator  $\sum_i \mu_i \phi_i^*$  in which  $\{\phi_i^*\}$  is such a minimizing set. Thus a curve fitting process is unbiased on  $\mathcal{M}$  and minimizes the expected response of the coefficients of combination  $\alpha_i$  to the error  $\delta f$ .

For each bounded linear functional  $\phi^*$  on  $\mathcal{K}$ , there exists a unique element  $\phi$  of  $\mathcal{K}$  such that  $\phi^*f = (f, \phi)$ ,  $f \in \mathcal{K}$ ; and vice versa. The functional  $\phi^*$  and the element  $\phi$  are adjoints of one another. Henceforth we shall characterize a curve fitting process by the set of  $n$  elements  $\phi_i$  which are respectively the adjoints of the functionals  $\phi_i^*$ . We shall say that the elements  $\{\phi_i\}$  and the curve fitting process *correspond* to one another<sup>(5)</sup>. The curve fitting process that corresponds to  $\{\phi_i\}$  carries  $f \in \mathcal{K}$  into  $\sum_i (f, \phi_i) \mu_i$ .

The omitted proofs of the following lemmas and theorem are similar to proofs in earlier sections.

LEMMA 15. Suppose that

$$\delta\alpha_i = (\delta f, \phi_i), \quad \phi_i \in \mathcal{K}.$$

Then  $E\delta\alpha_i = 0$ , and the variance matrix of  $\{\delta\alpha_i\}$  is<sup>(6)</sup>

<sup>(5)</sup> Let  $x_0$  be a fixed value of  $x$ . The functional  $\phi^*f = f(x_0)$  is bounded and linear on  $\mathcal{K}$  if and only if  $x_0$  is a point of positive  $m$ -mass. Now it is precisely then that  $f(x_0)$  seems a natural element of a procedure of approximation.

<sup>(6)</sup> In our symbol for a square matrix, the index which comes earlier in the alphabet is the row index. For example,  $i$  is the row index in both  $\{a_{i,j}\}$  and  $\{b_{j,i}\}$ .

$$\{E\delta\alpha_i\overline{\delta\alpha_j}\} = \{(V\phi_j, \phi_i)\}.$$

LEMMA 16. The  $n$  elements  $\{\phi_i\}$  of  $\mathcal{H}$  correspond to a curve fitting process if and only if

$$\text{trace } \{(V\phi_j, \phi_i)\} = \sum_i (V\phi_i, \phi_i) = E \sum_i |\delta\alpha_i|^2$$

is minimal, subject to the constraint

$$(\mu_i, \phi_j) = \delta_{i,j}.$$

THEOREM 5. The  $n$  elements  $\{\phi_i\}$  of  $\mathcal{H}$  correspond to a curve fitting process if and only if

$$V\phi_i \in \mathcal{M} \quad \text{and} \quad (\mu_i, \phi_j) = \delta_{i,j}.$$

The functions  $\{\phi_i\}$  and the curve fitting process are unique if  $V$  is positive definite on  $\mathcal{M}^\perp$ .

COROLLARY. Curve fitting processes exist if and only if approximation processes exist. Any curve fitting process is an approximation process. If  $A$  is an approximation process and if

$$(18) \quad \phi_i = A^* \sum_j \rho_{j,i} \mu_j, \quad \{\rho_{i,j}\} = \{(\mu_l, \mu_k)\}^{-1},$$

then  $\{\phi_i\}$  correspond to a curve fitting process.

REMARK. If  $V$  is positive definite on  $\mathcal{M}^\perp$ , each above process is unique; hence either both processes exist and are the same or else neither exists.

**Proof of Corollary.** Suppose that  $\{\phi_i\}$  correspond to a curve fitting process. We shall show that  $A = \sum_i \mu_i \phi_i^*$  is an approximation process.

If  $f \in \mathcal{M}$ , then  $f = \sum_j \beta_j \mu_j$  and

$$Af = \sum_i (f, \phi_i) \mu_i = \sum_{i,j} \beta_j (\mu_j, \phi_i) \mu_i = \sum_{i,j} \beta_j \delta_{j,i} \mu_i = \sum_j \beta_j \mu_j = f.$$

Hence  $(A - I)\mathcal{M} = 0$ .

Also  $A^* = \sum_i \phi_i \mu_i^*$ , and, for any  $f$  in  $\mathcal{H}$ ,

$$VA^*f = V \sum_i (f, \mu_i) \phi_i = \sum_i (f, \mu_i) V\phi_i \in \mathcal{M},$$

since  $V\phi_i \in \mathcal{M}$ . Hence  $VA^*\mathcal{H} \subset \mathcal{M}$ , and  $A$  is indeed an approximation process (with the following special property:  $A\mathcal{H} = \mathcal{M}$ ).

Next, suppose that  $A$  is an approximation process. Define  $\phi_i$  by (18). Since the elements  $\mu_i$  are independent, the Gram matrix  $\{(\mu_l, \mu_k)\}$  is invertible and  $\{\rho_{i,j}\}$  is defined. Then

$$V\phi_i = VA^* \sum_j \rho_{j,i} \mu_j \in \mathcal{M},$$

since  $VA^*\mathcal{H}\subset\mathcal{M}$ . Also

$$\begin{aligned}(\mu_i, \phi_j) &= \left( \mu_i, A^* \sum_k \rho_{k,j} \mu_k \right) = \sum_k \rho_{j,k} (A\mu_i, \mu_k) = \sum_k \rho_{j,k} (\mu_i, \mu_k) \\ &= \delta_{j,i} = \delta_{i,j},\end{aligned}$$

since  $\bar{\rho}_{k,j} = \rho_{j,k}$  and  $(A-I)\mathcal{M}=0$ . Hence  $\{\phi_i\}$  correspond to a curve fitting process (which, however, is not necessarily  $A$ ).

**10. Companions and curve fitting processes.** In this section, as in the preceding,  $\mathcal{M}$  is  $n$ -dimensional. The elements  $\mu_1, \dots, \mu_n$  span  $\mathcal{M}$ .

**LEMMA 17.** *A bounded linear operator  $U$  is proper on  $\mathcal{M}$  if and only if the  $n \times n$  matrix  $\{(U\mu_j, \mu_i)\}$  is invertible.*

**Proof.** The operator  $U$  is proper on  $\mathcal{M}$  if and only if

$$\|PUf\| \geq b > 0 \quad \text{whenever } f \in \mathcal{M}, \|f\| = 1.$$

Assume that  $f \in \mathcal{M}$ ,  $\|f\| = 1$  throughout this proof.

Put

$$b = \min \|PUf\|;$$

this minimum exists, since  $f$  varies on a compact set. Now  $b=0$  if and only if there is an element

$$f = \sum_j \beta_j \mu_j, \quad \|f\| = 1$$

such that  $PUf=0$ ; that is,  $Uf \in \mathcal{M}^\perp$ ; that is,

$$(Uf, \mu_i) = \sum_j \beta_j (U\mu_j, \mu_i) = 0 \quad \text{for all } i;$$

that is, the  $n \times n$  matrix  $\{(U\mu_j, \mu_i)\}$  is singular.

**LEMMA 18.** *If  $U$  is a companion and*

$$(19) \quad \phi_i = U^* \sum_j \tau_{j,i} \mu_j, \quad \{\tau_{i,j}\} = \{(U\mu_1, \mu_k)\}^{-1},$$

*then  $\{\phi_i\}$  correspond to a curve fitting process.*

*If  $\{\phi_i\}$  correspond to a curve fitting process, then*

$$(20) \quad U = Y + \sum_{i,j} \mu_i c_{i,j} \phi_j^*$$

*is a companion, providing that the operator  $Y$  is bounded and linear,  $Y^*\mathcal{M}=0$ , and the  $n \times n$  matrix  $\{c_{i,j}\}$  is invertible. All companions are of the form (20). Also*

$$(21) \quad U = Y + \sum_{i,j} \phi_i c_{i,j} \phi_j^*$$



is a positive definite companion, providing that the operator  $Y$  is positive definite on  $\mathcal{M}^\perp$ ,  $Y\mathcal{M}=0$ , and the  $n \times n$  matrix  $\{c_{i,j}\}$  is Hermitian and positive definite. All positive definite companions are of the form (21).

If the  $U$  of (20) or (21) is substituted in (19), the  $\phi_i$  produced by (19) are precisely the  $\phi_i$  in (20) or (21).

**11. The inverse  $V^{-1}$  need not exist.** When  $\mathcal{H}$  is infinite-dimensional, the condition  $E\|\delta f\|^2 < \infty$  implies that  $V$  cannot have a bounded inverse. For, trace  $V = E\|\delta f\|^2$  and  $V$  is therefore completely continuous [9, p. 98]. Hence a complete orthonormal set exists relative to which the matrix of  $V$  is diagonal:

$$V = \{\delta_{r,s}v_r\}.$$

Since trace  $V = \sum_r v_r < \infty$ ,  $v_r \rightarrow 0$  as  $r \rightarrow \infty$ , and  $1/v_r$  cannot be bounded. But if  $V^{-1}$  were to exist, its matrix would be  $\{\delta_{r,s}/v_r\}$ , which does not have bounded rows. Approximation processes can exist nonetheless.

A particular instance is the following. Let  $\{e_r\}$  be a complete orthonormal set in an infinite-dimensional space  $\mathcal{H}$ . Let  $\mathcal{M}$  be the manifold spanned by  $e_1, \dots, e_n$ . Let  $V$  be the operator whose matrix is

$$\{(Ve_s, e_r)\} = \{\delta_{r,s}/r^2\}, \quad r, s = 1, 2, \dots.$$

Then  $V$  is a positive definite operator of finite trace. Hence  $V$  is a variance. Let  $P$  be the projection onto  $\mathcal{M}$ . Then  $A = P$  is the unique approximation process. Thus,  $(A - I)\mathcal{M} = 0$  and  $VA^*\mathcal{H} = VP^*\mathcal{H} = VP\mathcal{H} = V\mathcal{M} \subset \mathcal{M}$ , by construction.

**12. An example in which no approximation process exists.** Let  $\{e_\lambda\}$ ,  $\lambda = 0, 1, \dots$ , be a complete orthonormal set in an infinite-dimensional space  $\mathcal{H}$ . Let  $\mathcal{M}$  be the one-dimensional manifold spanned by  $e_0$ . Let  $V$  be the operator whose matrix is

$$\{(Ve_\nu, e_\lambda)\} = \begin{pmatrix} 1 & v_1 & v_2 & v_3 & \cdots \\ v_1 & w_1 & 0 & 0 & \cdots \\ v_2 & 0 & w_2 & 0 & \cdots \\ v_3 & 0 & 0 & w_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \lambda, \nu = 0, 1, \dots.$$

We shall show that the numbers  $v_r, w_r, r = 1, 2, \dots$ , can be chosen so that  $V$  is positive definite, of finite trace, and so that no approximation process exists. One such choice is

$$v_r = r^{1/2}3^{-r}, \quad w_r = r3^{-r}, \quad r = 1, 2, \dots.$$

Throughout this section  $r, s$  range over the positive integers;  $\lambda, \nu$  range over the nonnegative integers; summations are taken over  $r$ . We denote the coordinates of an element  $f$  in  $\mathcal{H}$  by  $f_\lambda = (f, e_\lambda)$ .

For  $V$  to be bounded, self-adjoint, of finite trace, it is sufficient that

$$v_r > 0, \quad w_r > 0, \quad \sum v_r < \infty, \quad \sum w_r < \infty.$$

Suppose these conditions satisfied.

For  $V$  to be positive definite, it is necessary and sufficient that

$$\begin{aligned} (Vf, f) &= |f_0|^2 + \bar{f}_0 \sum v_r f_r + f_0 \sum v_r \bar{f}_r + \sum w_r |f_r|^2 \\ &= |f_0 + \sum v_r f_r|^2 + \sum w_r |f_r|^2 - |\sum v_r f_r|^2 > 0 \end{aligned}$$

whenever  $0 < |f_0|^2 + \sum |f_r|^2 < \infty$ ; that is,

$$\sum w_r |f_r|^2 - |\sum v_r f_r|^2 > 0 \quad \text{whenever } 0 < \sum |f_r|^2 < \infty.$$

But Schwarz's inequality implies that

$$|\sum v_r f_r|^2 \leq (\sum w_r |f_r|^2)(\sum v_r^2/w_r),$$

with equality attainable. Hence  $V$  is positive definite if and only if

$$\sum v_r^2/w_r < 1.$$

Assume this.

Now suppose that an operator  $A$  were an approximation process. Then  $AP = P$ , since  $(A - I)P = 0$ ; and  $AV(I - P) = 0$ , since  $AV\mathcal{M}^\perp = 0$ . Denote the matrix of  $A$  by

$$\{(Ae_\nu, e_\lambda)\} = \{a_{\lambda, \nu}\}.$$

The matrix of  $P$  consists entirely of zeros except that its leading element is unity. Hence  $AP = P$  if and only if

$$a_{\lambda, 0} = \delta_{\lambda, 0};$$

and  $AV(I - P) = 0$ , in addition, if and only if

$$a_{r, s} w_s = 0, \quad v_s + a_{0, s} w_s = 0;$$

that is,

$$a_{r, s} = 0, \quad a_{0, s} = -v_s/w_s.$$

But  $A$  is surely not bounded if the sum of the absolute squares of the elements of the first row of its matrix is infinite, that is, if

$$\sum v_r^2/w_r = \infty.$$

In this case, then, there can be no approximation process.

#### REFERENCES

1. A. C. Aitken, *Studies in practical mathematics*. IV. On linear approximation by least squares, Proceedings of the Royal Society of Edinburgh Sec. A, vol. 62 (1945) pp. 138-146.

2. ———, *On least squares and linear combinations of observations*, *ibid.* vol. 55 (1935) pp. 42–48.
3. C. F. Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae* (1821, 1823, 1826), *Werke*, vol. 4, pp. 1–94, or *Abhandlungen zur Methode der kleinsten Quadrate*, Berlin, 1887, pp. 1–91.
4. P. R. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity*, New York, 1951.
5. E. Hellinger and O. Toeplitz, *Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten*, *Encyklopädie der Mathematischen Wissenschaften*, vol. II, 3, 2 (1927) pp. 1335–1597.
6. A. N. Kolmogorov, *Interpolation und Extrapolation von stationären zufälligen Folgen*, *Izvestiya Akad. Nauk SSSR. Ser. Mat.* vol. 5 (1941) pp. 3–14. (Russian, German summary.)
7. ———, *On the proof of the method of least squares*, *Uspehi Matem. Nauk. (N.S.)* vol. 1 (11) no. 1 (1946) pp. 57–70. (Russian.)
8. F. J. Murray, *Linear transformations in Hilbert space*, Princeton, 1941.
9. J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Berlin, 1932, pp. 18–101. (Reprint: New York, 1943.)
10. M. H. Stone, *Linear transformations in Hilbert space*, New York, 1932.
11. N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series* (with two appendices by N. Levinson), Cambridge, Mass., 1949.

QUEENS COLLEGE,  
FLUSHING, N. Y.