THE OPERATOR PSEUDOINVERSE IN CONTROL

AND SYSTEMS IDENTIFICATION

by

Frederick J. Beutler
William L. Root

Computer, Information & Control Engineering Program
The University of Michigan

September, 1973

To be presented at the Advanced Seminar on Generalized Inverses and Applications (Mathematics Research Center, University of Wisconsin, October, 1973)

TABLE OF CONTENTS

ii

PART 0

INTRODUCTION

Extension of the concept of the pseudoinverse of a linear mapping in finite dimensional linear space to a linear operator in Hilbert and even Banach space has been of interest for some years [B-3] [B-4] [D-1] [N-1] [P-3]. Our concern in this paper is with the (Moore-Penrose) pseudoinverse of a linear transformation between Hilbert spaces. Part of our work deals with basic theory, which we then use for applications to systems identification and the quadratic regulator problem.

Since the finite dimensional (matrix) pseudoinverse can easily be interpreted geometrically in terms of certain orthogonal projections on Euclidean spaces, generalization to Hilbert spaces is completely natural; indeed, if one limits consideration to bounded operators with closed range, much of the theory of matrix pseudoinverses generalizes directly to Hilbert space [D-1] [B-3]. However, widening the class of transformations to encompass densely defined closed linear operators introduces substantive complications; domains of definition must be carefully treated, and account taken of the fact that neither the sum nor the product of densely defined closed operators necessarily enjoys the same property. Nevertheless, a good part of the theory carries over in satisfying fashion.

There are genuine difficulties, however, when an operator has non-closed range, for then the pseudoinverse is unbounded and only densely defined; moreover, many of the usual pseudoinverse characterizations become invalid. In case the range of the operator is not closed, the pseudoinverse usually fails to provide an answer to real problems (e. g. , in statistical estimation), since one cannot tolerate a "solution" which is only sometimes meaningful. To circumvent this impediment, we consider (1) shrinking the image Hilbert space and endowing it with a new topology which insures that the range of the operator is closed, and (2) replacing the original operator or its pseudo-inverse by a better behaved approximation. Either procedure can be applied as desired, but since some change is necessarily engendered thereby in the figure of merit which is optimized, the changed formulation may no longer be legitimate for its intended application. Roughly speaking, changes in the Hilbert space or modifying the original operator is related to proper mathematical modelling of a physical situation, while approximations to the pseudoinverse are more closely connected to pseudoinverse theory as such.

There is good reason to analyze pseudoinverses of arbitrary densely defined linear closed operators, rather than just bounded operators and/or those whose ranges are closed. Our examples (and doubtless many others) will indicate this. On the other hand, it seems evident that we cannot easily drop the requirements that an operator be closed and densely defined.

Part I of this work is devoted to a self-contained exposition of the basic theory of the pseudoinverse for densely defined linear closed operators with arbitrary range. It is self-contained in the sense that no results are borrowed from the literature on pseudoinverses, finite-dimensional or otherwise. Much of the material covered in Part I is available elsewhere in some form, but we believe some of our results for unbounded operators represent extensions. The theory we develop includes the case of operators that do not have closed range, but any questions regarding approximations and change in topologies are deferred until later. Appendix A consists of a collection of lemmas ("exercises for the reader") relevant to unbounded operators; they are intended to facilitate the reading of Part I and of the later sections.

In Part II, a Gauss-Markov theorem on statistical estimation is proved under the hypothesis that both the quantity to be estimated and the observations are elements of Hilbert spaces. This theorem is an improved version of a theorem stated earlier in [R-5], where the proof was not given; the proof given here suffices for the previously stated result. Our theorem applies only to non-singular covariance operators, and reduces to the classical Gauss-Markov theorem when the spaces are finite dimensional. To obtain our result (in Hilbert spaces) it becomes necessary to treat the inversion of an unbounded non-closed operator, which carries us beyond the subject matter of Part I. Consequently, an additional hypothesis and some further argument are required.

Appendix B (needed for Part II) is a brief development of the more elementary aspects of the theory of Hilbert-space-valued random variables. The use of Bochner integrals enables us to establish the needed facts very quickly, thus sparing the reader the chore of studying the literature on the subject.

Two applications of engineering interest are discussed; the first is a problem in unknown system identification (Part III), while the second (Part IV) is an optimal control problem known to workers in the field as the quadratic regulator problem. In Part III, Volterra-Frechet integral polynomials are taken to represent a class of unknown systems in input-output form, and the Gauss-Markov theorem of Part II is applied to the estimation of their kernels. There is also in this section some discussion of the applicability of the pseudoinverse to identification problems in general.

In one classical form of the quadratic regulator problem, it is required to find the minimum energy input which will move a system from some initial state to the origin at a designated time. Part IV offers a reformulation which generalizes this problem to admit a greater variety of linear constraints, possibly including some which are incompatible and/or unattainable by the system. The solution always exists as a pseudoinverse (even if the system is described by an unbounded operator), and reduces to the classical result if the system is capable of meeting the constraint. Another related quadratic regulator problem, usually called the free endpoint problem, is also treated in Part IV. Here, a quadratic loss function involving both input

and output is to be minimized by appropriate choice of input. We show this loss function to be conveniently represented if we change the topology on the image Hilbert space of the input-output operator; moreover, the new topology guarantees the range of this operator to be closed. The solution to a generalization of the free endpoint quadratic regulator problem can then be directly obtained by applying the pseudoinverse. In short, the pseudoinverse is a unifying influence on extensions of some quadratic regulator problems, and as such, provides an insight not generally obtained through the more classical approaches.

In both Parts III and IV we encounter as a recurrent theme the necessity of inverting, in some acceptable sense, operators with non-closed range. The technique of changing the topology on a reduced image space is employed in both places, but not in identical fashion. In particular, Part III describes a general approach which may also be useful in dealing with problems other than those considered here.

Approximations to the pseudoinverse constitutes the subject matter of Part V. Some known matrix approximations are examined in the context of unbounded operators whose range may not be closed. Properties of these approximations, and of other techniques suggested for operators of non-closed range, are investigated. Systematic use is made of the polar decomposition of operators, and of the functional calculus applicable to the resulting positive operators expressed as a spectral representation. We believe the material of this section to be largely original, especially in their coverage of unbounded operators.

Part V also contains some results on the limits to be expected from any choice of approximations.

It may be useful to the reader to mention that the entire paper is dependent on Part I, but that Parts IV and V are independent of one another, and of Parts II and III.

PART I

PSEUDOINVERSE OPERATORS IN HILBERT SPACE

The pseudoinverse of a matrix has a rich literature, and has become sufficiently well recognized to constitute the subject of two recent books [A-2] [R-1]. Much of the underlying theory is phrased in matrix-theoretic concepts, even though some of the principal optimization applications are more clearly motivated by the "best approximation" property, which the pseudoinverse matrix possesses with respect to Euclidean norms. In particular, it has been observed that application of the pseudo-inverse matrix solves certain optimal control (quadratic regulator [K-1]) and minimum mean square estimation [A-1] problems.

The emphasis on norm minimization suggests a function analytic rather than an algebraic approach to the pseudoinverse. Indeed, it seems natural to attempt to extend pseudoinverses to a Hilbert space context [D-1], since Hilbert space itself is no more than a generalization of finite dimensional linear vector space with Euclidean norm. The necessary extension is in fact easily accomplished for bounded linear operators of closed range [D-1] [B-2]. This class of course includes all bounded linear operators of finite dimensional range, and a fortiori, operators on finite dimensional spaces. Thus, the Hilbert space pseudoinverse theory is not only a legitimate extension of the matrix theory, but also represents an approch eminently suitable to optimization questions.

If we abandon the assumption that an operator is necessarily bounded and/or equipped with closed range, new complexities are encountered. These must be faced squarely if application to control or estimation theory are to be contemplated in any degree of generality. One may recall, for instance, that the differentiation operator in an L_2 space is unbounded [R-2], so that the input-output relation of a dynamical system described by differential equations is represented by an operator whose range is not a closed set. On the other hand, we shall see that the Gauss-Markov estimation problem in separable Hilbert space is equivalently formulated in terms of unbounded operators.

In order to avoid a complete chaos of pathological behaviors, we shall suppose that we seek the pseudoinverse of a linear operator $A: H_1 \to H_2$, where A is a closed operator ([R-2], Section 115) densely defined on the Hilbert space H_1, with Hilbert space H_2 as range space; such an operator will be called DDC (densely defined closed). Roughly speaking, the pseudoinverse A^+ of A will solve the following minimization problem: given $z \in H_2$, A^+ delivers $x_0 = A^+ z$, where $x_0 \in H_1$ minimizes the norm of $z - Ax$ over $x \in H_1$, and where x_0 is the element of least norm accomplishing the minimization. These notions will be made precise, and we shall also attempt to determine the conditions under which A^+ exists, what its properties are, and how it may be characterized.

We begin by defining the symbols we shall need for our analysis. $D(A)$ is the domain of A, and $R(A)$ its range; an overline denotes closure, so $\overline{R(A)}$ is the least subspace containing the range of A. By a

subspace we shall consistently mean a closed linear manifold in the appropriate Hilbert space. Thus, the null space N(A) of A is a subspace because A is DDC. An orthogonal complement is indicated by the standard symbol $\perp$, as for example $N(A)^{\perp}$ is the complement of the null space of A. We use P to stand for an orthogonal projection, and particularly P_R for the projection on $\overline{R(A)}$ and P_M for the projection on $N(A)^{\perp}$. In addition, standard symbols are employed for norm, adjoint, inverse and the like.

For the minimization problem, it is convenient to define some special symbols which are used recurrently. Thus, let

$$\delta_z = \inf_{x \in D(A)} ||z - Ax||, \qquad z \in H_2 \qquad (1.1)$$

$$K_z = \{x: Ax = P_R z\} \qquad \text{(preimage of } P_R z \text{ under A)} \qquad (1.2)$$

$$F_A = \{z: P_R z \in R(A)\} \qquad (1.3)$$

Finally, the subscript $_r$ indicates the restriction of any operator on H_1 to the subspace $N(A)^{\perp}$, or the restriction to $\overline{R(A)}$ of an operator with domain space H_2. Specifically, $A_r: N(A)^{\perp} \to \overline{R(A)}$ and $A^*_r: \overline{R(A)} \to N(A)^{\perp}$. Properties of restrictions of operators and their combinations are summarized in Appendix A to the extent that they are required in what follows.

As we have noted, our applications of the pseudoinverse rest on its "best approximation" property. In intuitive terms, the pseudo-inverse, when applied to $z \in H_2$, must give us the best approximate solution $x_o = A^+ z$ to the functional equation

$$z = Ax \qquad (1.4)$$

We use BAS as the acronym for "best approximate solution," the term

being more precisely described by

<u>Definition 1.1</u>: $x_0 \in H_1$ is a BAS of (1.4) if

$$|| z - Ax_0 || = \delta_z \qquad (1.5)$$

and

$$|| x_0 || < || x || \qquad (1.6)$$

for any other x which also attains the infimum (1.6).

<u>Remark</u>: Uniqueness of the BAS constitutes part of its definition, being

an immediate consequence of the strict inequality (1.6). The definition

does <u>not</u> assert the existence of the BAS; indeed, there is always a

sequence $\{x_n\} \in D(A)$ approaching the infimum (1.1), but without any

element attaining it as required by Definition 1.1. However, the ques-

tion of existence of the BAS is quickly settled by

<u>Theorem 1.1</u>: A BAS exists iff $z \in F_A$ [cf. (1.3)]; whenever a BAS

x_0 exists, it is unique and satisfies

$$Ax_0 = P_R z \qquad (1.7)$$

and

$$x_0 \in N(A)^\perp. \qquad (1.8)$$

Conversely, assume x_0 satisfies (1.7) and (1.8). Then x_0 is the BAS,

and is the only element of H_1 satisfying these two equations.

Proof: We know $\inf_{y \in R(A)} || z - y || = \delta_z$, the infimum being uniquely at-

tained by $z_1 = P_R z$. It follows that (1.5) is met iff x is such that

$Ax = P_R z$. But x of this type exists only if $z \in F_A$, in which case K_z

[see (1.2)] is non-empty, and (1.5) and (1.7) are both equivalent to

$x \in K_z$.

Suppose now x', $x'' \in K_z$. Then $Ax' = Ax''$, or $(x'-x'') \in N(A)$.

Hence any $x \in K_z$ has the orthogonal decomposition

$$x = x_o + x_1 \tag{1.9}$$

where $x_o \in N(A)^\perp$ is the same for every $x \in K_z$, and $x_1 \in N(A)$. To

verify $x_o \in K_z$, we use Lemma A.1 to argue $x_o \in D(A)$, and obtain

$Ax = Ax_o$ from (1.9) and $x_1 \in N(A)$. Now apply the Pythagorean theorem

to (1.9); the strict inequality (1.6) follows unless $x_1 = \theta$ and conse-

quently $x = x_o$. Thus x_o is a BAS satisfying (1.8).

For the converse, note (1.7) means $z \in F_A$, whence a BAS x'_o

exists and satisfies (1.7) and (1.8). But $(x_o - x'_o) \in N(A)$ from (1.7),

whereas $(x_o - x'_o) \in N(A)^\perp$ by (1.8). Therefore $x_o = x'_o$, or x_o is seen

to be the BAS as claimed. The same argument shows that at most one

vector can satisfy (1.7) and (1.8). $|||$

Theorem 1.1 is readily specialized to operators whose ranges

are closed sets, viz.

Corollary 1.1: If $R(A)$ is a closed set, the BAS always exists.

Proof: For any $z \in H_2$, $P_R z \in \overline{R(A)} = R(A)$, so $F_A = H_2$. $|||$

We have seen that every $z \in F_A$ has associated with it a unique

BAS $x_o \in H_1$, thereby suggesting an operator which transform elements

of H_2 to elements of H_1. More formally:

Definition 1.2: An operator $A^+ : H_2 \to H_1$ is called a pseudoinverse

operator (henceforth abbreviated PI) if $D(A^+) = F_A$, and if, for each

$z \in F_A$

$$x_o = A^+ z \tag{1.10}$$

is the BAS.

How can we recognize an operator as a PI? The following criterion is often helpful.

Lemma 1.1: The linear operator $A^+: H_2 \to H_1$ is a PI iff $D(A^+) = F_A$ and

$$N(A^+) = R(A)^\perp \tag{1.11}$$

$$\overline{R(A^+)} = N(A)^\perp \tag{1.12}$$

and

$$AA^+ y = y \qquad \text{for all } y \in R(A). \tag{1.13}$$

Proof: (Sufficiency) For $z \in F_A$, again call $P_R z = z_1$. Then $z_1 \in R(A)$, and (1.13) asserts $A^+ z_1 \in K_z$, i.e. $Ax_o = P_R z$ with $x_o = A^+ z_1$. Second, this $x_o \in N(A)^\perp$ by (1.12), so we conclude x_o is the BAS [compare (1.8) in Theorem 1.1]. To complete the proof, we need only show $A^+ z = A^+ z_1$. To this end, we decompose $z = z_1 + z_2$, where z_1 is as before and $z_2 \in N(A^+)$. Now $z \in D(A^+) = F_A$, and $A^+ z_2 = \theta$, whence $A^+ z = A^+ z_1 = x_o$ as required by (1.10).

(Necessity) Clearly, $R(A^+) \subset N(A)^\perp$, for otherwise (1.8) fails. To prove $R(A^+)$ dense in $N(A)^\perp$, observe $D(A) \cap N(A)^\perp$ to be dense in $N(A)^\perp$ (cf. Lemma A.4), and each $x \in [D(A) \cap N(A)^\perp]$ is a BAS for $y = Ax$ (Theorem 1.1). Since then $x = A^+ y$, $[D(A) \cap N(A)^\perp] \subset R(A^+)$. Thus (1.12) is necessary.

If (1.13) is not satisfied, $A^+ z \notin K_z$ for some $z \in R(A) \subset F_A$, so so (1.7) does not hold. Respecting (1.11), consider that the BAS for $z \perp R(A)$ is perforce the null vector from (1.7) and (1.8), requiring that $N(A^+) \supset R(A)^\perp$ if A^+ is to be a PI. On the other hand, (1.13) demands $N(A^+) \subset R(A)^\perp$, which then leads to the validity of (1.11). |||

Remark: The conditions (1.11) and (1.13) of Lemma 1.1 already in-

sure that $R(A)^{\perp}$ and $R(A)$ are in the domain of A^+, wherefrom $F_A \subset$

$D(A^+)$. To see this, observe that A^+ is linear by hypothesis, and that

$$F_A = R(A) \oplus R(A)^{\perp} \qquad (1.14)$$

is then in $D(A^+)$. We note F_A is dense in H_2, which implies A^+ to be

densely defined because $F_A \subset D(A^+)$. However, if $D(A^+)$ is larger

than F_A we cannot claim A^+ to be DDC, whereas (as we shall prove

below) A^+ must be a closed operator if $D(A^+) = F_A$.

There are many alternative forms for a matrix PI [R-1], but these

fail to carry over to unbounded operators in direct fashion. The source

of difficulty is that combinations of unbounded operators need not be

DDC, and may in fact be defined only on the null vector; hence, man-

ipulations of such operators is anything but routine. Nevertheless, we

can demonstrate some of the properties of the PI by an explicit construc-

tion, which (inter alia) exhibits the PI as a linear DDC operator.

Theorem 1.2: The PI exists as the uniquely defined linear DDC opera-

tor

$$A^+ = A_r^{-1} P_R . \qquad (1.15)$$

Proof: Since a unique BAS corresponds to each $z \in F_A$, the PI is prop-

erly and uniquely described on F_A. It then suffices to prove that the

right side of (1.15) constitutes a linear DDC operator which meets the

conditions set forth in Lemma 1.1.

In the first place, Lemma A.6 asserts A_r^{-1} to be defined and DDC.

Then, since P_R is bounded, $A_r^{-1} P_R$ is likewise a closed operator.

$A_r^{-1} P_R$ is also clearly linear. To show $A_r^{-1} P_R$ densely defined, we

note its domain includes both its null space $R(A)^{\perp}$ and the domain of

A_r^{-1}, namely $R(A)$ [see Lemma A. 2]. By linearity, we therefore have

$D(A_r^{-1}P_R) \supset F_A$, the latter having the form (1. 14); since F_A is dense

in H_2, so is $D(A_r^{-1}P_R)$.

We have proved that A^+ [as given by (1. 15)] is linear DDC, with

$D(A^+) \supset F_A$. To show $D(A^+) = F_A$, consider $z \notin F_A$. Then

$P_R z \notin R(A) = D(A_r^{-1})$, which means $z \notin D(A_r^{-1}P_R)$.

Finally, we must verify (1. 11), (1. 12) and (1. 13) of Lemma 1. 1.

From (1. 15), $N(A^+) \supset R(A)^{\perp}$, but also $A_r^{-1} y \neq \theta$ for non-null

$y \in R(A)$; in fact, $A_r^{-1} : \overline{R(A)} \to N(A)^{\perp}$ has domain $R(A)$ and possesses

an inverse A_r. Consequently, the A^+ of (1. 15) satisfies (1. 11).

Next, we see from Lemma A. 4 that $\overline{R(A_r^{-1})} = N(A)^{\perp}$. Moreover,

$R(A^+) = R(A_r^{-1})$, since each range depends only on preimages in $R(A)$,

and for any $y \in R(A)$, $A^+ y = A_r^{-1} y$. Thus, (1. 12) has been proven.

For any $y \in R(A)$, we now obtain

$$AA^+ y = AA_r^{-1} y = A_r A_r^{-1} y = y \qquad (1. 16)$$

where the middle equality results from the definition of A_r as the re-

striction of A to $N(A)^{\perp}$, which is precisely the range space of A_r^{-1}.

Therefore, (1. 13) is validated also, and the proof of the theorem is

complete. |||

Corollary 1. 2: A^+ is bounded iff $R(A)$ is closed.

Proof: A^+ is bounded iff A_r^{-1} is bounded, and the latter is equivalent to

$R(A)$ closed by Lemma A. 8. |||

The matrix pseudoinverse is often defined not by its best approximation property (Definition 1.2), but rather as the (unique) matrix satisfying the identities [G-2]

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad AA^+ = P_R, \quad A^+A + P_M. \quad (1.17)$$

These same identities—suitably modified for DDC operators—are relevant to the PI as per Definition 1.2. Indeed, the modified identities (1.17) follow from our definition. Conversely, a DDC operator for which a weaker form of (1.17) is valid must be a PI. We now proceed to state precisely and prove these relationships.

__Theorem 1.3:__ Let A^+ be the PI for the DDC operator A. Then

$$AA^+A = A \quad\quad\quad (1.18)$$

$$A^+AA^+ = A^+ \quad\quad\quad (1.19)$$

$$AA^+ \subset P_R, \text{ with } D(AA^+) = F_A \quad\quad\quad (1.20)$$

and $\quad A^+A \subset P_M, \text{ with } D(A^+A) = D(A). \quad\quad\quad (1.21)$

Proof: (1.20) is a direct consequence of (1.11) and (1.13) in Lemma 1.1. In view of (1.11), both sides of (1.19) yield the null vector when applied to $R(A)^\perp$, so we need consider only $y \in R(A)$, and demonstrate $A^+AA^+y = A^+y$ for (1.19). But $AA^+y = y$ by (1.13), and so (1.19) follows.

To prove (1.21), decompose $x \in D(A)$ as

$$x = x_1 + x_2, \quad x_1 \in N(A)^\perp, \quad x_2 \in N(A). \quad\quad\quad (1.22)$$

Then $x_1 \in D(A)$ by Lemma A.1, and

$$A^+Ax = A^+Ax_1 = A_r^{-1}P_R A_r x_1 = A_r^{-1}A_r x_1 = x_1 \quad\quad\quad (1.23)$$

which shows A^+A acts like P_M for every $x \in D(A)$. With this result it is easy to prove (1.18), since from (1.23) and (1.22) above

$$AA^+Ax = Ax_1 = Ax \quad\quad\quad (1.24)$$

for all $x \in D(A)$. $|||$

Although A^+A and AA^+ are DD according to (1.21) and (1.20), one cannot assume these to be closed operators. In fact, A^+A (respectively AA^+) closed corresponds precisely to the boundedness of A (respectively A^+), as is seen from

<u>Corollary 1.3</u>: Let A^+ be the PI of the DDC operator A. Then A is bounded iff A^+A is a closed operator; A^+ is bounded [or equivalently, R(A) is a closed set] iff AA^+ is a closed operator.

Proof: Because of the duality evident in Theorem 1.3, we need only consider the first assertion. By (1.21), A^+A is bounded and DD, so A^+A is closed iff it is defined on all H_1 as a bounded operator. Of course, A (a DDC operator now defined on all H_1) is then bounded also. On the other hand, A bounded and A^+ DDC implies A^+A to be a closed operator.

We could state a direct converse to Theorem 1.3, but is is noteworthy that less restrictive requirements can be substituted for (1.20) and (1.21). More specifically, we have

<u>Theorem 1.4</u>: Suppose $A^+: H_2 \to H_1$ is a linear closed operator with $D(A^+) \subset F_A$. Let this operator satisfy (1.18) and (1.19), and assume further

$$A^+A \quad \text{is a symmetric operator} \tag{1.25}$$

and

$$AA^+ \quad \text{is a symmetric operator.} \tag{1.26}$$

Then A^+ is the PI of the (linear DDC) operator A.

Proof: We again turn to the sufficiency conditions of Lemma 1.1, starting with (1.13). If $y \in R(A)$, we may take $x \in D(A)$ such that $Ax = y$,

whence (1.18) becomes $AA^+y = y$ for each $y \in R(A)$. This disposes of (1.13).

Our next task is to prove (1.11), or $N(A^+) = R(A)^\perp$. Actually, $N(AA^+) = R(A)^\perp$ suffices, because $N(A^+) = N(AA^+)$ by the chain

$$N(A^+) \subset N(AA^+) \subset N(A^+AA^+) = N(A^+) \tag{1.27}$$

The usual argument on ranges and null spaces produces

$$R(AA^+)^\perp = N([AA^+]*) \subset N(AA^+), \tag{1.28}$$

the right hand inclusion following from (1.26). However, the containment in (1.28) is even an equality, because $D(AA^+)$ is dense in $D([AA^+]*)$, so that $N(AA^+)$ is not only dense in $N([AA^+]*)$, but is also closed by (1.27). We then have

$$R(AA^+)^\perp = N(AA^+) \tag{1.29}$$

Finally, $R(A) = R(AA^+)$ from

$$R(A) \supset R(AA^+) \supset R(AA^+A) = R(A). \tag{1.30}$$

The combination of these equalities therefore yields $N(A^+) = R(A)^\perp$, as was to be shown.

The remaining equality (1.12) of Lemma 1.1 can be written as $R(A^+)^\perp = N(A)$, which is like (1.11) except that the symbols A and A^+ are interchanged. But the hypotheses on A and A^+ are symmetrical, so that the desired identity can be demonstrated in the same fashion as $R(A)^\perp = N(A^+)$.

We must still show A^+ to have the correct domain. The domain clearly includes $R(A)$, since $AA^+y = y$ for all $y \in R(A)$. Also, the null space of A^+ is $R(A)^\perp$, so by the linearity of A^+

$$D(A^+) \supset R(A) \,\oplus\, R(A)^{\perp} = F_A; \qquad (1.31)$$

in view of the original hypothesis $D(A^+) \subset F_A$, we then have $D(A^+) =$ F_A. |||

Remark: In the literature on the matrix PI (1.25) and (1.26) are re- placed by the stronger hypotheses that A^+A and AA^+ are self-adjoint.

Such assumptions are inappropriate here, since (by Corollary 1.3),

they would limit the applicability of Theorem 1.4 to bounded operators

A and A^+. Moreover, even the (apparently) weaker assumptions A^+A

and AA^+ closed symmetric already imply $A^+A = P_M$, $AA^+ = P_R$, with

both A and A^+ bounded.

Many identities have been developed for the matrix pseudoinverse

[R-1] [A-2]. Most of these equalities are retained in a weaker ver- sion when applied to unbounded operators. The two presented below,

however, are generalized without change in their statement, although

the proofs necessarily become sophisticated beyond mere modifications

of the usual matrix arguments.

__Theorem 1.5:__ $(A^+)^+ = A.$ $\qquad (1.32)$

Proof: We construct $(A^+)^+$ by the method of Theorem 1.2. In the first

place, $(A^+)_r = A_r^{-1}$ from (1.15). Secondly, $\overline{R(A^+)} = N(A)^{\perp}$ by (1.12).

But then $(A^+)_r^{-1} = A_r$ and the projection on the closure of the range of A^+

is P_M, so that the application of (1.15) to A^+ yields

$$(A^+)^+ = A_r P_M = A P_M = A . \qquad (1.33)$$

The right hand equality is here justified by Lemma A.1 and the ortho- gonal decomposition (A.1). |||

Theorem 1.6: $(A^+)^* = (A^*)^+$.   (1.34)

Proof: The construction of Theorem 1.2 exhibits the PI of A^* to be

$$(A^*)^+ = (A^*_r)^{-1} P_M;$$ (1.35)

by Lemma 1.1, this PI has null space $N[(A^+)^*] = R(A^+)^\perp = N(A)$, with

the right hand equality being obtained from (1.12). Since both operators

in (1.34) have the same null space, it suffices to compare their restric-

tions $(A^+)^*_r$ and $[(A^*)^+]_r$ to $N(A)^\perp$. Using successively Lemma A.9

and the representation of Theorem 1.2, we obtain

$$(A^+)^*_r = (A^+)^*_r = (A_r^{-1})^* .$$ (1.36)

But also

$$[(A^*)^+]_r = (A^*_r)^{-1} = (A^*_r)^{-1} = (A_r^{-1})^* ;$$ (1.37)

here the equalities follow respectively from (1.35), Lemma A.9 and

the interchange of inversion and the adjoint operation (see [R-2], Sec-

tion 117). |||

It is also characteristic of the matrix pseudoinverse that it may

be expressed in various forms involving combinations of other matrices

related to A [R-1] [A-2]. For example, among the alternative formula-

tions of the matrix pseudoinverse we have $A^+ = A^* (AA^*)_r^{-1}$ and

$A^+ = (A^* A)_r^{-1} A^*$. Since such forms also occur in applications to infi-

nite dimensional problems (e.g., Gauss-Markov estimation), it would

be desirable if they were also valid in the more general case of DDC

operators. But again, operators A which are unbounded and/or have

non-closed range lead to complications which must be taken into account.

To this end, we need

Definition 1.3: $A^X : H_2 \rightarrow H_1$ is a <u>restricted pseudoinverse operator</u>

(acronym: RPI) for the linear DDC operator A if $A^X \subset A^+$.

<u>Remark:</u> A^X need be neither densely defined nor bounded to be a RPI.

<u>Theorem 1.7:</u> $A' = A^*(AA^*)_r^{-1} P_R$ (1.38)

is a RPI with domain

$$D(A') = R(AA^*) \;\oplus\; R(A)^{\perp}. \qquad (1.39)$$

Proof: By Lemma A.14, $(AA^*)_r$ is invertible, and $(AA^*)_r^{-1} = (A_r^*)^{-1}A_r^{-1}$.

Thus (1.38) makes sense and

$$A' = A^*[(A_r^*)^{-1}A_r^{-1}]P_R. \qquad (1.40)$$

Now $A^*(A_r^*)^{-1} = A_r^*(A_r^*)^{-1}$ is a restriction of the identity. In other words,

(1.30) implies

$$A' \subset A_r^{-1}P_R = A^+. \qquad (1.41)$$

Let us determine $D(A')$. It is clear A' is linear and well defined

on $R(A)^{\perp}$. In $\overline{R(A)}$, A' is at best defined on $R(A)$ because of (1.41)

and the extent of $D(A^+)$; we therefore limit consideration to $y \in R(A)$.

Now $D[(AA^*)_r^{-1}] = R[(AA^*)_r] = R(AA^*)$, the equality on the right follow-

ing from Lemmas A.2 and A.10 and Corollary A.2. Hence $y \in R(A)$ is

in the domain of $(AA^*)_r^{-1}$ iff $y \in R(AA^*)$. Further, such a y leads to

$$(AA^*)_r^{-1}y \in R[(AA^*)_r^{-1}] = D[(AA^*)_r] = D(A_r A_r^*) \subset D(A_r^*) \subset D(A^*). \qquad (1.42)$$

Here we have again used Lemma A.12 for the equality on the right.

Since $(AA^*)_r^{-1}y$ automatically falls in $D(A^*)$ as shown by (1.42), and

since $R(A) \supset R(AA^*)$, $D(A')$ is correctly given by (1.39) and A' is an RPI. |||

<u>Corollary 1.4:</u> $A' = A^+$ iff $R(AA^*) = R(A)$; in particular, A' is the PI if

$R(A)$ is closed (or equivalently, any of the conditions of Lemmas A.8 or

A.15 are met).

Proof: From (1.39) and (1.14), $D(A') = D(A^+)$ iff

$$R(AA^*) = R(A). \quad ||| \qquad (1.43)$$

Should $R(A)$ be closed (which is true if any of the conditions of Lemmas A.8 or A.15 are satisfied), Corollary A.4 states (1.43) to be valid.

Remark: A' is densely defined in any event because $R(AA^*) = \overline{R(A)}$ as indicated by Corollary A.3. Also, (1.41) verifies the existence of a closed extension for A'; it is plausible that A^+ is actually the minimal closed extension of A', although neither proof nor counterexample has been found to elucidate the question.

Analogous to A', but differing in its domain, is

$$A'' = (A^*A)^{-1}A^* . \qquad (1.44)$$

If A is a matrix, or even if A is bounded and $R(A)$ closed, A' and A'' are both bounded and coincide with A^+, and there is little advantage of one with respect to the other. But when A is an unbounded operator of closed range, we can guarantee only $A' = A^+$ (see Corollary 1.4 above) and conclude nothing further on A'', whereas A bounded (with arbitrary range) gives rise to $A'' = A^+$ without any simplification of A'.

Theorem 1.8: A'' is a RPI with domain

$$D(A'') = [D(A^*) \cap R(A)] \oplus R(A)^{\perp}. \qquad (1.45)$$

Proof: The proof is much like that of Theorem 1.7, so we shall omit some of the details. Since $\overline{R(A)} = N(A^*)^{\perp}$ we write

$$A'' = (A^*A)_r^{-1}A_r^*P_R \qquad (1.46)$$

$$= A_r^{-1}[(A_r^*)^{-1}A_r^*]P_R . \qquad (1.47)$$

The term $[(A_r^*)^{-1}A_r^*]$ is a restriction of the identity on $\overline{R(A)}$ to $D(A_r^*)$, so that

$$A'' \subset A_r^{-1} P_R = A^+. \tag{1.48}$$

The domain of A'' evidently contains $R(A)^{\perp}$. On $\overline{R(A)}$, this

domain is limited to vectors in $R(A)$, and [by (1.46)] even to

$y \in [R(A) \cap D(A^*)]$. For such y, $A_r y \in R(A_r^*) = D[(A_r^*)^{-1}]$, and

$[(A_r^*)^{-1} A_r^*] y = y$. But $y \in R(A) = D(A_r^{-1})$, whence $A_r^{-1}[(A_r^*)^{-1} A_r^*] y$ is

defined also. Thus $y \in [R(A) \cap D(A)]$ implies $y \in D(A'')$, and $D(A'')$

is described by (1.45) by the linearity of A''. |||

Corollary 1.5: $A'' = A^+$ iff $R(A) \subset D(A^*)$; in particular, A'' is the PI

if A is bounded (or equivalently, if any of A^*, AA^* or A^*A are bounded).

Proof: If $R(A) \subset D(A^*)$, the $D(A'')$ of (1.45) coincides with $D(A^+)$. The

second assertion of the corollary is then immediate, since any of the

hypotheses imply $D(A^*) = H_2$. |||

Remark: Attention is called to a limited duality between A' and A''.

If any of A_r^{-1}, $(A_r^*)^{-1}$, $(AA^*)_r^{-1}$ or $(A^*A)_r^{-1}$ are bounded, A' is the PI A^+,

and is thus a bounded operator defined on all H_2. On the other hand, if

any of A, A^*, AA^* or A^*A [i.e., A_r, A_r^*, $(AA^*)_r$ or $(A^*A)_r$] are

bounded, A'' is the same as the PI A^+, and is thus a DDC operator on

H_2.

PART II

A GAUSS-MARKOV THEOREM FOR HILBERT SPACE

A natural extension of the classical theory of linear unbiased min-
imum-variance estimators (LUMV estimators) is to the case that both
the vector of unknown parameters and the vector of observations are
infinite-dimensional in the sense that both are elements of separable
Hilbert spaces. Although one may well argue that in any actual data
processing scheme only a finite set of parameters will be estimated
and only a finite set of data will be used, a Hilbert space version of
the theory is of interest as providing characterization of limit cases.
Indeed, there are many estimation problems which are inherently
infinite-dimensional, and for which reduction to finite sets of observa-
tions and parameters represents an approximation. The subject of
identification of unknown systems to be discussed briefly in Part III
provides examples of such problems. In this Part we state and prove
a Gauss-Markov theorem for Hilbert space, but we leave comments
about its application to Part III.

Necessary and sufficient conditions for the existence of an LUMV
estimator in Hilbert space, and a characterization of the estimator in
terms of reproducing-kernel Hilbert space concepts, is given in [P-2].
Our objective here is somewhat different; we wish to exhibit reason-
able sufficient conditions for the existence of an LUMV estimator, that
also guarantee an extension of the standard classical formula for the

estimator. Let H_1 and H_2 be separable Hilbert spaces. Consider the linear model

$$z = Bx + w \qquad\qquad (2.1)$$

where $x \in H_1$, where B is a continuous linear transformation from all of H_1 into H_2, and where w is an H_2-valued random variable. Hilbert-space-valued random variables are discussed below and in Appendix B, where definitions are given for the mathematical expectation of an H-valued random variable w and for the covariance operator of w, as well as conditions for the existence of the covariance. We assume the expected value of w, Ew, exists and equals zero and that a covariance operator, K, for w exists. The element x represents the vector of unknown parameters to be estimated, and z represents the vector of observations. A <u>linear estimator</u> C is defined to be a continuous linear transformation from all of H_2 into H_1; $\hat{x} = Cz$ is then an <u>estimate</u> of x. Since z is an H_2-valued random variable, Cz is an H_1-valued random variable (see Appendix B). We call $E||Cz - E(Cz)||^2 = E||Cw||^2$ the variance of the estimator C; it will be seen below that if K has finite trace the variance is always finite. We say C is an <u>unbiased</u> estimator for x if $E[Cz]$ exists and equals x. If C is unbiased, then the variance equals the mean-squared error, $E||Cz - x||^2$. C^o is <u>LUMV</u> if it is linear, unbiased and of finite variance, and $E||C^o z - x||^2 \le E||Cz - x||^2$ for all linear unbiased estimators C. In the classical case H_1 and H_2 are finite-dimensional Euclidean spaces. Then, if K is strictly positive definite, and if $N(B) = 0$, an LUMV estimator always exists and

is given, e.g., by the formula[1]

$$C = (B^* K^{-1} B)_r^{-1} B^* K^{-1}. \tag{2.2}$$

It is this formula we extend to the case H_1 and H_2 are Hilbert spaces. If one puts $A = K^{-1/2} B$, then (2.2) can be written

$$\hat{x} = (A^* A)_r^{-1} A^* (K^{-1/2} z) \tag{2.3}$$

$$= A^+ (K^{-1/2} z)$$

The standard proof of (2.2) is carried out by first proving that C is in fact the PI when $K = I$, then transforming z in the model (2.1) by $K^{-1/2}$, that is by re-norming the observation space H_2. Eseentially, this[2] is what we do below, although it is not quite literally what we do because of certain technical considerations.

Before getting to the theorem, some further elucidation of the conditions already imposed and their implications is warranted. The discussion to follow also points out the reasonableness, and sometimes the necessity, of certain additional conditions. First, there are various ways of establishing the existence of a quantity w that can reasonably be called a Hilbert-space-valued random variable, that satisfies $E(w) = 0$, and that has associated with it a bounded, self-adjoint, non-negative-definite operator K such that

---

[1] The notation is consistent with that of Part I.

[2] It would be interesting to extend a Gauss-Markov theorem to Hilbert space with the condition that K be nonsingular removed, but we have not done this. The classical method and formula we use does not appear to be suitable. It would appear more promising, for example, to try to extend the method and formula in [A-2].

$$(Ky, z) = E[(y, w)\overline{(z, w)}]. \tag{2.4}$$

K is called the covariance operator (since the matrix form of (2.4) characterizes the covariance matrix of a finite-dimensional random vector). An outline of a simple way of defining w is relegated to Appendix B, since the material involved is largely foreign to the rest of the paper. It will be observed that the construction given in Appendix B automatically implies that K is nuclear (i.e., is compact with finite trace). This condition is also reasonable as a requirement imposed on the estimation problem; indeed, if $E||w||^2 < \infty$ and the probability measure is countably additive, K must be nuclear (see Appendix B), and $\text{Tr } K = E||w||^2$. The condition that K be strictly definite is necessary for the forthcoming formula to be meaningful; see the footnote p. 25.

The definition of an LUMV estimator is lifted from the classical one for the finite-dimensional case. The now additional restriction that the estimator be continuous as well as linear is almost necessary for the estimation problem to make sense. In fact, the estimator must be everywhere-defined or it is useless; for technical reasons we want it to be a closed operator. Hence it must be bounded. The condition that B be continuous is certainly a technical convenience. As far as the modelling of real problems is concerned, it seems natural to require B to be continuous.

Now if the unknown element x in (2.1) has a non-zero component in the nullspace of B, that component in no way affects the observation z, and it cannot possibly be estimated. Hence we can only estimate the

orthogonal projection of x on $N(B)^{\perp}$. In a notation consistent with that used in Part 1, we let P_M be the projection on $N(B)^{\perp}$, B_r be the restriction of B to $N(B)^{\perp}$ and put $x_1 = P_M x$. Henceforth we are concerned only with estimates of x_1.

Since w and z are H_2-valued random variables, Cw and Cz are H_1-valued random variables for any (bounded linear) estimator C (see Appendix B). Further, the mathematical expectations and covariance operators exist, and we have

$$E[Cz] = E[CBx + Cw] = CBx \qquad (2.5)$$

$$= CBx_1$$

The unbiasedness condition (for x_1) then becomes $CBx_1 = x_1$, or

$$CB_r x_1 = x_1 . \qquad (2.6)$$

Thus, the restriction of C to R(B) is necessarily B_r^{-1}, which is another way of stating the unbiasedness condition. From this observation we can make the simple but important inference that B must have closed range. In fact,

$$\|C\| = \sup_{\|y\|=1} \|Cy\| \geq \sup_{\substack{\|y\|=1 \\ y \in R(B)}} \|Cy\|$$

$$= \sup_{\substack{\|y\|=1 \\ y \in R(B)}} \|B_r^{-1} y\| = \|B_r^{-1}\| .$$

Thus C cannot be bounded unless B_r^{-1} is bounded, but this is equivalent to B having closed range (Lemma A. 8). The condition that B have closed range becomes an essential hypothesis.

The variance of a finite variance linear unbiased estimator C for x_1 is

$$E||Cz - x_1||^2 = E||CBx + Cw - x_1||^2$$

$$= E||Cw||^2. \tag{2.7}$$

Let $\{\phi_n\}$, $n=1,2,\ldots$, be a c.o.n.s. for H_1, chosen so that a subsequence of the ϕ_n's exactly spans $N(B)^{\perp}$. Since K, the covariance of w (see Appendix B), is required to be a nuclear operator, we have for C unbiased,

$$E||Cz - x_1||^2 = E||Cw||^2$$

$$= E[\sum_{}^{\infty} |(w, C^*\phi_n)|^2] = \sum_{}^{\infty} E(C^*\phi_n, w)\overline{(C^*\phi_n, w)}$$

$$= \sum_{}^{\infty} (K C^*\phi_n, C^*\phi_n) = \sum_{}^{\infty} (CKC^*\phi_n, \phi_n)$$

$$= Tr(CKC^*) \tag{2.8}$$

In fact, CKC^* is nuclear since K is nuclear and C is bounded, and so $Tr(CKC^*)$ exists (cf. [G-1], Chapter 1). We see that the conditions on K and C guarantee that the estimator is of finite variance. We summarize these comments and make an obvious addition in

Lemma 2.1: Let the B of (2.1) be a DDC linear operator and let w be an H_2-valued random variable with a nuclear covariance operator K and with Ew = 0. Then a necessary condition that a linear unbiased estimator C for x_1 exist is that B have closed range. If B has closed range a finite variance linear unbiased estimator does exist, in particular $C = B^+$ is one. For these assertions to hold K does not have to be strictly definite and B does not have to be bounded.

Proof: The first assertion has been proved above. By Corollary 1.2, B^+ is a bounded operator. It meets the condition for unbiasedness,

(2.6), and the error variance is $\text{Tr}(B^+ K B^{+*})$, which is finite as has been pointed out above. |||

Remark: By Corollaries 1.4 and 1.5 it follows that B^+ can be expressed in the form B', and if B is bounded, in the form B''. B^+ provides an estimator whose range is $N(B)^\perp$. However, it is fairly obvious that if $\overline{R(B)} \neq H_2$ and $N(B) \neq \theta$ finite variance unbiased estimators C exist such that $R(C) \cap N(B) \neq \theta$.

We proceed now to a consideration of LUMV estimation. The key step in getting the LUMV estimator is solving a minimization problem — a slightly complicated variant of the basic minimization problem that led to the pseudoinverse. We treat this problem first out of context.

Let H_1 and H_2 be Hilbert spaces (not necessarily separable). Let K be a bounded self-adjoint, strictly positive-definite linear operator on H_2. Let B be a bounded linear operator with $D(B) = H_1$ and closed range $R(B) \subset H_2$. Let ϕ be an arbitrary element of $N(B)^\perp$. We call the following minimization problem, problem I:

$$\text{find } c \in H_2 \text{ to minimize } (Kc, c)$$

$$\text{subject to } B^* c = \phi \tag{I}$$

It is convenient to change the form of this problem. Since K is self-adjoint and strictly positive-definite, it has a self-adjoint strictly positive-definite square root $K^{1/2}$. Put $\xi = K^{1/2} c$, from which we have $c = K^{-1/2} \xi$. Then problem II is defined:

$$\text{find } \phi \in H_2 \text{ to minimize } \|\xi\|^2$$

$$\text{subject to } B^* K^{-1/2} \xi = \phi \tag{II}$$

I and II are exactly equivalent in the following sense: if a ξ^o exists

that solves II, then $K^{-1/2}\xi_o$ is defined and $c^o = K^{-1/2}\xi_o$ solves I;

if a c^o exists that solves I, then $\xi^o = K^{1/2}c_o$ is such that $B^* K^{-1/2}\xi^o$

is defined, and ξ^o solves II. Consequently, we can restrict our atten-

tion to II. Except for one thing we could apply the results on BAS di-

rectly to problem II; the difficulty is that the (in general unbounded)

operator $B^* K^{-1/2}$ need not be closed, although it is densely defined.

We can introduce hypotheses to guarantee that it is closable, but then

just replacing it with its closure is not good enough, because a mini-

mizing element ξ_o in the domain of the closure would not need to belong

to $D(B^* K^{-1/2})$. The difficulty is circumvented by working with the

adjoint.

<u>Theorem 2.1</u>: Problem II has a unique solution ξ_o if

(1) $K^{-1/2}B$ is densely defined on H_1

(2) $B_r^{-1}K^{1/2}P_R K^{-1/2}$ is a bounded operator, where P_R

is the projection on $\overline{R(K^{-1/2}B)}$ (This operator is well-

defined as will be shown).

The solution ξ_o is given by $\xi_o = K^{1/2}S^*\phi$, where S is the bounded

continuous extension of $B_r^{-1}K^{1/2}P_R K^{-1/2}$ to all of H_2 (it will be seen

that $B_r^{-1}K^{1/2}P_R K^{-1/2}$ is densely defined and such an extension exists).

Further, S has closed range, $R(S) = N(B)^{\perp}$.

Proof: $K^{-1/2}B$ is closed since $K^{-1/2}$ is closed and B is bounded; it is

densely defined by hypothesis (1). Thus, if we put $A = K^{-1/2}B$, A is

DDC and admits a PI, A^+. By Theorem 1.2, $A^+ = A_r^{-1}P_R$, where P_R

is the projection on $\overline{R(A)}$ and A_r is the natural restriction of A to

$N(A)^{\perp}$. We note certain preliminary facts about the transformation

A: first,

$$N(A) = N(B), \qquad (2.9)$$

or, equivalently,

$$\overline{R(A^*)} = N(A)^{\perp} = N(B)^{\perp}; \qquad (2.10)$$

second,

$$A_r = K^{-1/2}B_r \qquad (2.11)$$

where B_r is as before the restriction of B to $N(B)^{\perp}$; third, A_r^{-1} is

bounded, or, equivalently, A has closed range. The equality (2.9) is

valid because $N(A) = N(K^{-1/2}B)$ and the only element $K^{-1/2}$ carries

into the null element is the null element. Then (2.11) follows imme-

diately from (2.10). To show that A_r^{-1} is bounded one can verify direc-

tly from (2.11) that $A_r^{-1} = B_r^{-1}K^{1/2}$, and note that B_r^{-1} is bounded.

To show that $B_r^{-1}K^{1/2}P_R K^{-1/2}$ is well-defined we must show that

$D(B_r^{-1}) \supset R(K^{1/2}P_R K^{-1/2})$. In fact, $\overline{R(K^{1/2}P_R K^{-1/2})} \subset \overline{R(K^{1/2}P_R)} =$

$\overline{R(K^{1/2}K^{-1/2}B)} = \overline{R(B)} = R(B) = R(B_r)$. The one perhaps non-obvious

step in this chain is the first equality, which one can verify straightfor-

wardly using the fact that $K^{1/2}$ is bounded.

Now consider the transformation $A^+K^{-1/2} = A_r^{-1}P_R K^{-1/2} =$

$B_r^{-1}K^{1/2}P_R K^{-1/2}$. Since A^+ is closed and bounded, $D(A^+K^{-1/2}) =$

$D(K^{-1/2})$, which is dense in H_2. By hypothesis (2) $A^+K^{-1/2}$ is bounded.

Consequently, $A^+K^{-1/2}$ can be extended by continuity to a closed bounded

linear transformation S defined on all of H_2.

We show that SB is the orthogonal projection P_M on $N(B)^{\perp}$. Let $x \in D(K^{-1/2}B)$ and put $x = x_1 + x_2$, $x_1 \in N(B)^{\perp}$, $x_2 \in N(B)$. Then,

$$SBx = A^{+}K^{-1/2}B(x_1 + x_2) = A^{+}A(x_1 + x_2) = x_1$$

by Lemma A.1 and Theorem 1.3. Since SB is continuous and everywhere defined and agrees with P_M on a dense subset of H_1, $SB = P_M$. But then SB is self-adjoint, and so $B^{*}S^{*} = P_M$.

We can now verify that $\xi^{o} = K^{1/2}S^{*}\phi$ solves problem II. First, since $\phi \in N(B)^{\perp}$

$$B^{*}K^{-1/2}\xi_o = B^{*}K^{-1/2}K^{1/2}S^{*}\phi = B^{*}S^{*}\phi = \phi$$

so the linear constraint is satisfied. Now, $A\phi$ is closed and $A^{*} \supset B^{*}K^{-1/2}$, since for all $x \in D(A)$ and $y \in D(K^{-1/2})$,

$$(Ax, y) = (K^{-1/2}Bx, y) = (x, B^{*}K^{-1/2}y).$$

Consequently, any ξ satisfying $B^{*}K^{-1/2}\xi = \phi$ satisfies $A^{*}\xi = \phi$, and $A^{*}(\xi - \xi_o) = \theta$. Thus, $\xi - \xi_o \in N(A*)$. It will follow that $||\xi_o||^2$ is a minimum if $\xi_o \in N(A*)^{\perp}$, for then

$$||\xi||^2 = ||\xi - \xi_o||^2 + ||\xi_o||^2 \geq ||\xi_o||^2.$$

But ξ_o does belong to $N(A^{*})^{\perp}$. For $SK^{1/2} \supset (A^{+}K^{-1/2})K^{1/2} = A^{+}$, and since A^{+} is everywhere defined, $SK^{1/2} = A^{+}$. Then $K^{1/2}S^{*} = (A^{+})^{*}$, so $\xi_o \in \overline{R((A^{+})^{*})}$. But by Theorem 1.6 and Lemma 1.1, $\overline{R((A^{+})^{*})} = \overline{R((A^{*})^{+})} = M(A^{*})^{\perp}$.

To establish uniqueness, let us note first that $\phi \in R(A^{*})$, and hence by Theorem 1.1 the problem of minimizing $||\xi||^2$ subject to $A^{*}\xi = \phi$ has a unique solution ξ'. ξ' is characterized as that one element which satisfies $A^{*}\xi' = \phi$ and $\xi' \in N(A*)^{\perp}$. Since ξ_o has been shown to satisfy both these conditions, $\xi' = \xi_o$. Then, a fortiori, ξ_o is a unique solution to problem II.

Finally, we observe that $R(S) \supset R(SB) = R(P_M) = N(B)^{\perp}$, while

on the other hand, $\overline{R(S)} = \overline{R(A^+ K^{-1/2})} \subset \overline{R(A^+)} = N(A)^{\perp} = N(B)^{\perp}$.

Hence $R(S) = N(B)^{\perp}$. $\|\|$

Corollary 2.1: The conclusion of Theorem 2.1 holds if hypothesis (2)

is replaced by any of the following:

(2') $A^+ K^{-1/2}$ is bounded;

(2") $A^*(AA^*)_r^{-1} P_R K^{-1/2}$ is bounded;

(2''') $(A^* A)_r^{-1} A^* K^{-1/2}$ is bounded.

Correspondingly, S may be defined as the continuous extension of any

of the operators defined in (2'), (2") or (2''').

Proof: From the proof of Theorem 2.1 we have $A^+ = B_r^{-1} K^{1/2} P_R$,

and from Corollary 1.4 $A^+ = A^*(AA^*)_r^{-1} P_R$, so the assertions involving

(2') and (2") follow. For (2'''), we note that $(A^* A)_r^{-1} A^* \subset A^+$ by Theorem

1.8.

Furthermore,

$$(A^* A)_r^{-1} A^* K^{-1/2} \supset (A^* A)_r^{-1} B^* K^{-1/2} K^{-1/2} = (A^* A)_r^{-1} B^* K^{-1}.$$

But $R(B^* K^{-1}) \subset R(A^*)$, and by Corollary A.4 $D[(A^* A)_r^{-1}] = R(A^* A)$

$= R(A^*)$, so $D[(A^* A_r)^{-1} A^* K^{-1/2}] = D(K^{-1/2})$, which is dense in H_2.

Thus the continuous extension of the operator of (2''') is the same S as

before. $\|\|$

Theorem 2.1, or equivalently Corollary 2.1, can now be applied

"coordinate-wise" to give a Gauss-Markov theorem. For this, we need

again that the Hilbert spaces be separable. We retain the notation of

Theorem 2.1.

Theorem 2.2[1]: In the linear model given by (2.1) let H_1 and H_2 be

separable Hilbert spaces, and let w be an H_2-valued random variable

with mean zero and strictly positive definite, nuclear covariance opera-

tor K. Assume:

(1) B is a bounded linear operator with $D(B) = H_1$, $R(B) \subset H_2$, and

   $R(B) = \overline{R(B)}$.

(2) $K^{-1/2}B$ is densely defined on H_1.

(3) $A^+K^{-1/2}$ is a bounded operator (or equivalently the operators defined

   by (2), (2'') or (2''') of Theorem 2.1 and Corollary 2.1 are bounded).

Then C = S is an LUMV estimator for x_1, where $x_1 = P_M x$. The

solution C = S is unique.

Proof: For C to be an LUMV estimator of x_1 it must be bounded, satis-

fy $CBx_1 = x_1$, and minimize $\sum\limits_{i}^{\infty} (CKC^*\phi_i, \phi_i)$ over the class of all C

satisfying the other conditions. From Theorem 2.1, S is bounded and

$SB = P_M$, so S = C satisfies the first two conditions.

Now the condition $CBx_1 = x_1$ implies $CB = P_M$. Hence $CB = B^*C^*$,

and the condition $CBx_1 = x_1$ can be rewritten as $B^*C^*x_1 = x_1$. Recall

that the c.o.n.s. $\{\phi_i\}$ was chosen so that a subsequence exactly spans

$N(B)^\perp$. Put $c_i = C^*\phi_i$, i=1,2,.... . Then the unbiasedness condition

becomes

$$B^*c_i = \phi_i \text{ for those } \phi_i \in N(B)^\perp \qquad (2.12)$$

and the expression to be minimized is

----

[1] See also [R-5].

$$\sum_{i=1}^{\infty} (K c_i, c_i). \qquad\qquad (2.13)$$

If we can find bounded C so that the corresponding c_i satisfy (2.12), so that all c_i corresponding to $\phi_i \in N(B)$ are zero, and so that each non-zero term of the sum in (2.13) is minimized, then C is LUMV, provided the sum in (2.13) is finite. However, the finiteness of the sum is automatic since C is bounded (see (2.8) and the succeeding comments). But c_i minimizes (Kc_i, c_i) subject to $B^* c_i = \phi_i$ if $c_i = K^{-1/2}(K^{1/2}S^*\phi_i)$ $= S^*\phi_i$ by Theorem 2.1 and the equivalence of problems I and II. Furthermore, $S^*\phi_i = \theta$ for all $\phi_i \in N(B)$ because $N(S^*) = \overline{R(S)}^{\perp} = N(B)$.

Thus $C^* = S^*$ provides a minimizing set of c_i's, and $C = S$ is LUMV. The uniqueness follows from Theorem 2.1. |||

Corollary 2.1: If the subspace R(B) is invariant under K, then conditions (2) and (3) of Theorem 2.2 are automatically satisfied, and B^+ is the unique LUMV estimator.

Proof: Since K is self-adjoint, $R(B)^{\perp}$ as well as R(B) is invariant under K, that is $K[R(B)] \subset R(B)$ and $K[R(B)^{\perp}] \subset R(B)^{\perp}$. Then since K is nonsingular on H_2, its restriction K_B to R(B) is a 1:1 mapping of R(B) into R(B). K_B retains the properties of self-adjointness, strict positivity and nuclearity; K_B^{-1} is defined, and $D(K_B^{-1})$ is dense in R(B). By Lemma A.8, B_r is bounded from below, hence $D(K^{-1}B_r)$ is dense in $N(B)^{\perp}$, whence by Lemma A.1, $D(K^{-1}B)$ is dense in H_1. This implies condition (2) of Theorem 2.2.

We now consider $B_r^{-1}K^{1/2}P_R K^{-1/2}$, where P_R, it will be recalled, is the projection on $R(K^{-1/2}B)$. Let P_B be the projection on R(B). It

is immediate from the hypothesis that P_B commutes with K, from

which it follows that P_B commutes with $K^{-1/2}$ (c.f. $\lfloor R\text{-}2 \rfloor$, p. 143).

Then, since B has closed range, $\overline{R(K^{-1/2}B)} = \overline{R(K^{-1/2}P_B)} = \overline{R(P_B K^{-1/2})}$

$= \overline{R(B)}$. Thus $P_R = P_B$, and $B_r^{-1}K^{1/2}P_R K^{-1/2} = B_r^{-1}K^{1/2}P_B K^{-1/2}$

$= B_r^{-1}P_B K^{1/2}K^{-1/2} = B_r^{-1}K^{1/2}K^{-1/2}$. This operator is bounded, so

condition (3) is satisfied, and its continuous extension is B^+. $|||$

Remarks: (1) Although in the case of Corollary 2.1 the LUMV estimator

does not depend on K, the error variance, as given by (2.8), does. The

result of Corollary 2.1 is to be expected, of course, from the classical

case. Intuitively, Corollary 2.1 is saying that if the noise in $R(B)^{\perp}$

is uncorrelated with the noise in $R(B)$, one may as well first project

the observations on $R(B)$.

(2) The hypotheses of Theorem 2.2 are presumably somewhat awkward

to verify in many instances. An example where they are satisfied, in-

volving conditions similar to but somewhat less restrictive than the

commutativity condition of Corollary 2.1, can be constructed as follows.

The verifications are routine and will not be given. Let B be written in

the form DU where D is a self-adjoint operator on H_2 and U is par-

tially isometric from H_1 into H_2. Suppose D has discrete spectrum,

so $D\psi_i = \lambda_i \psi_i$, where $\{\psi_i\}$ is a c.o.n.s. for $R(B)$ and $\lambda_i > 0$. Let $\{\eta_i\}$

be a c.o.n.s. for $R(B)^{\perp}$. Let $K = K_1 P_1 + K_2 P_2 + K_3 P_3$, where: K_1

is a nonsingular covariance operator on $M_1 = V\{\psi_1, \ldots, \psi_N\}$;

$\eta_1, \ldots, \eta_M\}^1$, K_2 is a nonsingular covariance operator on

---

$^1V\{\ldots\}$ denotes the subspace spanned by the vectors.

$M_2 = V\{\psi_{N+1}, \psi_{N+2}, \ldots\}$, K_3 is a nonsingular covariance operator

on $M_3 = V\{\eta_{M+1}, \eta_{M+2}, \ldots\}$, and P_i is the orthogonal projection on M_i.

Then conditions (2) and (3) of Theorem 2.2 are satisfied.

(3) The condition that B have closed range is, as we have seen,

essential, but it is often not satisfied in the kinds of applications one

would like to consider. In some instances, one can replace the "natu-

ral" H_2, which is typically ℓ_2 or L_2, by a smaller Hilbert space H'_2

which still contains all elements $y \in R(B) \subset H_2$, but in which the range

of B becomes a closed set. This cannot be done satisfactorily unless

the noise is small enough in an appropriate sense. We shall describe

this procedure in connection with the example of Part III.

PART III

AN APPLICATION TO SYSTEM IDENTIFICATION

To identify an unknown input-output system means to determine a suitable mathematical model for the unknown system from incomplete prior knowledge of the system, using data obtained by measurement of outputs and either measurement or prior knowledge of corresponding inputs. A model is suitable if (i) it will reproduce the behavior of the system well enough, according to some set criterion, when the system is stimulated by any one of the class of inputs of interest, and (ii) it is in a useful form. Both the criterion of fit and the usefulness of the model will depend to some degree on what the identification is to be used for, e.g., to permit control of the system, to allow transmission of information through the system, to yield predictions of future behavior, etc. There is usually no reason why there should be only one acceptable model. A definition as general as the one above encompasses a tremendous variety of problems; any two of them may have very little in common with each other, and acceptable solutions may involve disparate mathematical methods. However, at bottom, identification problems are inversion problems, as will be pointed out specifically below, so it is not surprising that certain examples involve generalized inverses. Often the measurements are noisy enough that the basic inversion required becomes a problem of statistical estimation.

To fix ideas, let u denote an input to a system, and y the corresponding output. We suppose $u \in \mathcal{U}$, the class of inputs of interest, and $y \in \mathcal{Y}$, any fixed class of outputs containing all y corresponding to $u \in \mathcal{U}$. The sets $\mathcal{U}$ and $\mathcal{Y}$ will be assigned mathematical structures as seem appropriate (of course, in modelling a problem, these structures are not unique). We assume there is a functional relationship from u to y, that is, for each input $u \in \mathcal{U}$ there is one corresponding output y[1]; then we have

$$y = F(u), \quad u \in \mathcal{U}. \qquad (3.1)$$

The function F characterizes the system in question completely, of course, as far as input and output data are concerned (as long as we think of the system as a "black box"), and we shall refer to the mapping F as the "system." This simple terminology carries with it the implication that what one is perhaps accustomed to calling one dynamical system, but with different initial states, here becomes a collection of systems.

Now suppose the system F is unknown and we want to identify it. First, either from prior knowledge, or purely as a working hypothesis, we postulate a class of systems to which (we hope) the unknown system must belong. In different language, we postulate a set $\mathcal{F}$ of mappings

---

[1]There has to be some relation between inputs and outputs or identification makes no sense. There are more general situations of course, e.g., an input u might determine a probability distribution on outputs.

from $\mathcal{U}$ into $\mathcal{Y}$ which we assume contains the unknown F. Then we carry out if possible a set of experiments to yield data to fix F closely enough within the class $\mathcal{F}$. Temporarily we can deal with the mapping F as an abstract entity, but eventually it must be represented concretely. The final representation of the estimate of F is the identification. Often an unknown dynamical system is represented in terms of a differential equation, the parameters of which are to be determined by the identification. This really amounts to representing F^{-1}. Another type of representation is directly in terms of integral operators, and this is the kind of representation we work with here.

Suppose now that $\mathcal{U}$, $\mathcal{Y}$ and $\mathcal{F}$ are given as sets and that in addition $\mathcal{Y}$ is a linear space. We do not yet need to require any mathematical structure for $\mathcal{U}$. There is then a linear structure imposed on $\mathcal{F}$ in the ordinary way; i.e., one defines αF, α a scalar, and $F_1 + F_2$, F, F_1, $F_2 \in \mathcal{F}$, by the equations

$$[\alpha F](u) = \alpha F(u) \tag{3.2}$$

$$[F_1 + F_2](u) = F_1(u) + F_2(u). \tag{3.3}$$

Let $\mathcal{L}$ be the linear space of mappings from $\mathcal{U}$ into $\mathcal{Y}$ generated by $\mathcal{F}$. Consider first the problem of noise-free identification of F in which the outputs y = F(u), $u \in \mathcal{U}$, can be known exactly. We interchange the roles of u and F and regard u as determining a mapping from $\mathcal{F}$, and even from $\mathcal{L}$, into $\mathcal{Y}$. So we can write

$$y = U(F) = F(u), \quad F \in \mathcal{L} \tag{3.4}$$

for each $u \in \mathcal{U}$, where U is the mapping corresponding to the input u.

The problem of finding F is now the problem of inverting U. U is always a linear mapping. In fact:

$$U(\alpha F_1 + \beta F_2) = [\alpha F_1 + \beta F_2](u) = \alpha F_1(u) + \beta F_2(u)$$

$$= \alpha \, U(F_1) + \beta \, U(F_2).$$

Thus, in a basic sense, noise-free identification is always a linear problem. If the output of the system can be observed only in the presence of noise w, then the model (3.1) is replaced by

$$z = F(u) + w \tag{3.5}$$

and (3.4) is replaced by

$$z = U(F) + w. \tag{3.6}$$

Equation (3.6) is an abstract version of the usual linear model for statistical estimation.

This simple observation that identification is basically a linear problem, under fairly general conditions, is often not appreciated. It indicates the potential applicability of generalized inverses to identification, quite apart from the particular example to follow. We note, obviously, that the linearity can be lost by using representations of the unknown systems which do not preserve it, and such representations are often used, sometimes for good reasons.

In the generic example to follow, a situation is to be considered in which $\mathcal{U}$ is a Hilbert space, so that w and z are Hilbert-space-valued random variables. $\mathcal{H}$, the space of systems, will also be a Hilbert space, and U, which is necessarily linear, will be a bounded operator, sometimes with closed range. The model used in the example has been chosen so that we can illustrate the application of theorems

in Parts I and II, and hopefully provide some insight into certain aspects of identification theory. However, the model is not typically used in practice, partly because of the computational complexity it introduces. A number of comments need to be made to place what we are doing in better perspective, but these are postponed till after the example.

We model the class of systems to which the unknown system is to belong with a class of Volterra-Frechet polynomials. In particular, consider transformations of the form

$$y(t) = [H(u)](t) = \sum_{n=1}^{N} \int_0^\tau \ldots \int_0^\tau k_n(\nu_1, \ldots, \nu_n) u(t-\nu_1) \ldots u(t-\nu_n) d\nu_1 \ldots d\nu_n,$$

$$t \in T \qquad (3.7)$$

where T is a finite interval in R^1, τ is a positive number or $+ \infty$, N is a fixed positive integer, and the functions u, $k_1, \ldots, k_N$, y are real or complex-valued. Let $u \in L_2(R^1)$ and $k_n \in L_2([0, \tau]^n)$, $n=1, \ldots, n$; that is, let k_n satisfy

$$\int_0^\tau \ldots \int_0^\tau |k_n(\nu_1, \ldots, \nu_n)|^2 d\nu_1, \ldots, d\nu_n < \infty. \qquad (3.8)$$

Clearly, if one permutes the arguments ν_i of the k_n's the integrals in (3.7) are unchanged. Consequently, one can symmetrize each k_n, i.e., replace $k_n(\nu_1, \ldots, \nu_n)$ by $\frac{1}{n!} \sum_\pi k_n(\nu_{\pi(1)}, \ldots, \nu_{\pi(n)})$ where the sum is over all permutations π of n integers, and we suppose this done. The symmetric kernels of n arguments form a closed linear subspace of $L_2([0, \tau]^n)$ which we denote $\mathcal{L}_n$.

The transformation defined by (3.7) represents a system which, in systems engineering terminology, is causal with finite memory. Clearly, by changing the interval of integration, the causality condition

can be removed. If $T = [0, \tau]$ and if $u(t) = 0$ for $t < 0$, then the integrals can be rewritten with the variable upper limit t without changing the transformation.

It is convenient to introduce the notations: F_n for the integral operator with kernel k_n, and y_n for the nth integral in (3.7). Then (3.7) can be written

$$y = \sum_{n=1}^{N} F_n(u) = \sum_{n=1}^{N} y_n \qquad (3.9)$$

We identify the operators F_n with their symmetric kernels k_n, and define $||F_n||$ to be the L_2 norm of k_n. Thus, the F_n form a Hilbert space that is isometrically isomorphic to $\mathscr{L}_n$ under the correspondence $F_n \longleftrightarrow k_n$, and to cut down the verbiage we say simply that $F_n \in \mathscr{L}_n$. We define a norm for F by $||F||^2 = \sum_{n=1}^{N} ||F_n||^2 = \sum_{n=1}^{N} ||k_n||^2$, and, with again an abuse of notation, regard F as an element of $\mathscr{L}_1 \oplus \ldots \oplus \mathscr{L}_N$, which we denote by $\mathscr{L}$.

An application of the Schwarz inequality shows that

$$|y_n(t)|^2 \leq ||k_n||^2 \, ||u||^{2n}, \qquad (3.10)$$

and hence that $y \in L_2(T)$, with

$$||y_n||^2 \leq m \, ||k_n||^2 \, ||u||^{2n} \qquad (3.11)$$

where m is the length of the interval T. Thus F is a mapping from $L_2(R^1)$ into $L_2(T)$, and it is bounded on bounded sets. It can be shown without much difficulty that if $y_n = F_n(u)$, $y_n' = F_n(u')$ then

$$||y_n - y_n'||^2 \leq C(n)||u - u'||^2[\max(||u||, ||u'||)]^{2(n-1)} \qquad (3.12)$$

(c.f., [R-4]), where C(n) is a constant for fixed n. From (3.12) it follows that H is a continuous mapping from $L_2(R^1)$ into $L_2(T)$. We do not need this fact in what follows, so we do not bother to prove (3.12).

We are now ready to discuss identification of systems modeled by (3.7) in accordance with the ideas leading to (3.6). Let $\mathcal{H}$, the class of systems to which the unknown system is to belong, be $\mathcal{L}$. Reinterpret (3.7) as $y = U(F)$, where $F = \{k_1, \ldots, k_N\} \in \mathcal{L}$, and U is the operator determined by the right side of (3.7) with fixed u. U is a linear operator from $\mathcal{L}$ into $L_2(T)$. It follows easily from (3.11) that U is a bounded linear operator. Actually, U is compact. To see this, let us again introduce notation for each term in the sum in (3.7). Let U_n be nth integral operator in the sum; U_n has kernel $[u(t-\nu_1)\ldots(u(t-\nu_n)]$.

We have $y = \sum_{n=1}^{N} U_n(k_n) = \sum_{n=1}^{N} U_n(F_n)$ with an obvious abuse of notation. Further, if we regard U_n as the operator on $\mathcal{L}$ which carries F_n into y_n and all F_j, $j \neq n$, into zero, we have

$$y = \sum_{n=1}^{N} U_n(F). \tag{3.13}$$

Now each U_n is Hilbert-Schmidt, in fact

$$\int_T \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |u(t-\nu_1)\ldots u(t-\nu_n)|^2 \, d\nu_1 \ldots d\nu_n dt \leq m||u||^{2n}. \tag{3.14}$$

Thus, each U_n is compact, and from (3.13) it follows that U is compact.

Suppose that observations of the output are made in the presence of additive noise w, where w is to be represented as an $L_2(T)$-valued random variable with mean zero and nuclear covariance operator K (see Appendix B). The model is then of the form of (3.6), which we repeat,

$$z = U(F) + w, \tag{3.6}$$

and we wish to estimate $F \in \mathcal{H} = \mathcal{L}$. Consider first (non-statistical) least-squares estimation of F—for this, of course, we disregard the statistical properties of w; w is just an error. U is DDC, so by Theorem 1.1 a BAS exists iff $z \in F_U$, the set of z's for which the projection of z on

$\overline{R(U)}$ belongs to R(U). The BAS is given by U^+z, where $D(U^+) = F_U$.

Since U is bounded, $U^+ = U'' = (U^*U)_r^{-1}U^*$ by Corollary 1.5. These results, though they describe the situation that obtains, are of little practical interest if R(U) is not closed, because an unbounded estimator that is not always defined is of little value. Unfortunately, R(U) is not closed unless U is a degenerate operator, since U is compact. If we consider the estimation of F by linear unbiased estimators from a statistical point of view, the same difficulty obviously arises. In particular, Theorem 2.2 does not guarantee an LUMV estimator unless (aside from other hypotheses) R(U) is closed. Fortunately, this difficulty can be circumvented mathematically by replacing the observation space $L_2(T)$ with a smaller Hilbert space in which the range of U is closed; we do this below. This replacement changes the model of the problem in such fashion that, speaking loosely, some output information is lost. This loss of information may or may not be of consequence, depending on the noise, as we shall see.

We digress from the example temporarily to construct the new Hilbert space that replaces the observation space H_2. This can be done in more generality than is needed for the example.

Lemma 3.1: Let T be a DDC operator on a Hilbert space H. Then the linear space D(T) is itself a (complete) Hilbert space with the inner product

$$(x, y)'_T = (x, y) + (Tx, Ty), \quad x, y \in D(T). \qquad (3.15)$$

If T is bounded from below, D(T) is also a Hilbert space with the inner product

$$(x, y)_T = (Tx, Ty), \quad x, y \in D(T). \qquad (3.16)$$

Proof: Since T is closed, the graph of T, G(T), is a Hilbert space

with the norm, $||\{x, Tx\}||^2 = ||x||^2 + ||Tx||^2$ ([D-2], p. 1186). That is,

(3.15) defines an inner product for G(T), so D(T) is a Hilbert space.

If T is bounded from below there is $\alpha > 0$ such that $||Tx|| > \alpha||x||$,

$x \in D(T)$. Then

$$||x||_T^2 = ||Tx||^2 \leq ||x||^2 + ||Tx||^2 \leq (\alpha^{-1} + 1)||x||_T^2,$$

so the norms given by (3.15) and (3.16) are topologically equivalent. |||

Theorem 3.1: Let B be a DDC linear transformation from the Hilbert

space H_1 into the Hilbert space H_2. Then there is a Hilbert space $\tilde{H}_2$

formed on the linear set $F_B \subset H_2$, but with a different norm, such that

(1) $R(B) \subset \tilde{H}_2$ and $R(B)$ is closed in $\tilde{H}_2$.

(2) The orthogonal complement of $R(B)$ in H_2 is the same set as the

orthogonal complement of $R(B)$ in $\tilde{H}_2$, and furthermore the H_2-

norm of $y \in R(B)^{\perp}$ is the same as the $\tilde{H}_2$-norm.

(3) $\tilde{B}$, the mapping B reinterpreted as a mapping from H_1 into $\tilde{H}_2$, has

closed range, and is bounded iff B is bounded.

Proof: B*, being DDC, has a polar decomposition, B* = JT, where

$T = (BB^*)^{1/2}$ is self-adjoint and J is partially isometric ([D-2], XII

7.7). Then B = B** = T*J* = T-J* since J is bounded ([D-2], XII 1.6),

and J* is also partially isometric ([D-2], XII, 7.6) with initial domain

$\overline{R(B^*)} = N(B)^{\perp}$ and final domain $\overline{R(B)}$.

To avoid some confusion of notation later we put $M = \overline{R(B)}$ and

$N = R(B)^{\perp}$, so that $H_2 = M \oplus N$. Now $\overline{R(T)} = \overline{R(B)} = M$, and the

restriction of T to M, T_r, has a self-adjoint inverse T_r^{-1} by Lemma

A.14. We note further that $D(T_r^{-1}) = R(T) = R(B)$. In fact, from

$B = TJ^*$ it follows that $R(T) \supset R(B)$. To show that $R(B) \supset R(T)$, suppose $y \in R(T)$, then $y = Tz$ for some $z \in N(T)^{\perp} = \overline{R(T)} = \overline{R(B)}$. But $z \in \overline{R(B)}$ is given by $z = J^* x$ for some $x \in N(B)^{\perp}$, since J^* restricted to $N(B)^{\perp}$ is a unitary mapping onto $\overline{R(B)}$. Thus $y = TJ^* x = Bx$, so $y \in R(B)$.

We can now apply Lemma 3.1 to $R(B) = D(T_r^{-1})$, with T_r^{-1} replacing T, and call the resulting Hilbert space $\tilde{M}$. Then we define

$$\tilde{H}_2 = \tilde{M} \oplus N.$$

The assertion (1) is now immediate since $\tilde{M}$ is itself a Hilbert space. (2) is obvious from the definition of $\tilde{H}_2$. $R(\tilde{B}) = \overline{R(\tilde{B})}$ is just a restatement of (1) since $R(B)$ and $R(\tilde{B})$ are the same set. We have left to show that $\tilde{B}$ is bounded iff B is bounded. Let P_1 and P_2 be the projections on $\tilde{M}$ and N respectively. For $x \in H_1$, let x_1 be the projection of x on $N(B)^{\perp}$ and x_2 the projection on $N(B)$. Now, using the norm induced by (3.15) for $\tilde{M}$, and denoting norms in $\tilde{H}_2$ by $|| \cdot ||_{\sim}$, we have for $x \in D(B)$

$$|| \tilde{B}x ||_{\sim}^2 = || Bx ||^2 + || T_r^{-1} Bx ||^2$$

$$= || Bx ||^2 + || T_r^{-1} TJ^* x_1 ||^2$$

$$= || Bx ||^2 + || x_1 ||^2$$

which proves the assertion. $|||$

<u>Remark 1.</u> If B is bounded the norm established by (3.16) can be used for $\tilde{M}$, and then $\tilde{B}$ is partially isometric.

<u>Remark 2.</u> If B is compact, $T^2 = BB^*$ is compact as well as self-adjoint and positive, so that $T^2 e_n = \lambda_n e_n$, $n = 1, 2, \ldots$, where the e_n are orthonormal and span $\overline{R(T)} = \overline{R(B)}$, and where $\lambda_n > 0$ and converge

monotonically to zero. The norm induced by (3.16) is applicable in this case, and we have for $y \in \tilde{H}_2$,

$$||y||_{\sim}^2 = \sum \lambda_n^{-1} |(y, e_n)|^2 + ||P_2 y||^2 \qquad (3.17)$$

where the norm in the last term is the norm of H_2.

Let us return to the identification problem for which the model is given by (3.6), or by (3.4) in the noise-free case, and for which $\hat{\mathcal{U}}$ is the class of Volterra-Frechet polynomials which has been introduced. The operator U now plays the role of the operator B in the preceding theorem; otherwise we shall use the notation of that theorem and of Remark 2. For the noise-free case, since $R(U) \subset \tilde{H}_2$ (as a set), we can clearly replace the observation space H_2 and the equation $y = U(F)$ with the space $\tilde{H}_2$ and the equation $\tilde{y} = \tilde{U}(F)$, where $\tilde{H}_2$ and $\tilde{U}$ are as given by Theorem 3.1 with $B = U$. When there is additive observation noise, however, so that the model is given by (3.6), this replacement may mean taking unjustified liberties with the model. For if $L_2(T)$ is the "actual" observation space, then for some values of $w(\omega)$, z will not lie in $\tilde{H}_2$; changing the mathematical model would in this case actually mean changing the original problem. The only case where no harm is done in changing to $\tilde{H}_2$ is when $w(\omega)$ belongs to the set $\tilde{H}_2$ with probability one.

A condition which guarantees this is given in the theorem to follow. Furthermore, the condition will prove to be necessary in a sense to be specified. Since $w(\omega)$ is an H_2-valued random variable, in the sense defined in Appendix B, $\{(w, e_n)\}$, with the e_n chosen as in Remark 2, is a sequence of scalar-valued random variables. We suppose w has

mean zero and nuclear covariance operator K; then $E[(w, e_n)] = 0$,

$E|(w, e_n)|^2 = (K e_n, e_n) = $ variance $[(w, e_n)]$. From Remark 2 it follows

that

$$||w(\omega)||^2_{\sim} = \sum_{n=1}^{\infty} \lambda_n^{-1} |(w(\omega), e_n)|^2 + ||P_2 w(\omega)||^2 \qquad (3.18)$$

for all ω for which the right side is finite.

Theorem 3.2: If

$$\sum_{n}^{\infty} \lambda_n^{-1} (K e_n, e_n) < \infty \qquad (3.19)$$

then $w(\omega) \in \tilde{H}_2$ (as a subset of H_2) with probability one. If $\tilde{w}(\omega) \in \tilde{H}_2$

is defined to equal $w(\omega)$ for all ω for which $||w(\omega)||_{\sim}$ is finite and zero

otherwise, then $\tilde{w}(\omega)$ is an $\tilde{H}_2$-valued random variable with mean zero

and with a nuclear covariance operator $\tilde{K}$.

Conversely, if $w(\omega) \in \tilde{H}_2$ with probability one and $\tilde{w}$ is defined as

above, then the existence of a nuclear covariance operator $\tilde{K}$ for $\tilde{w}$

implies that (3.19) is satisfied.

Proof: From (3.18) it follows that

$$E||w(\omega)||^2_{\sim} = \sum_{n=1}^{\infty} \lambda_n^{-1}(K e_n, e_n) + E||P_2 w(\omega)||^2. \qquad (3.20)$$

Since the second term of the right-hand side of (3.20) is finite by the

properties of w, the condition (3.19) guarantees the finiteness of (3.20)

which in turn implies that $w(\omega) \in \tilde{H}_2$ with probability one.

With the notation used in Theorem 3.1, $\tilde{H}_2 = M \oplus N$. We

establish a c.o.n.s. for $\tilde{H}_2$ by separately specifying c.o.n. systems

for $\tilde{M}$ and N. The c.o.n.s. for N can be taken arbitrarily; we denote

it by $\{\tilde{f}_m\}$. For $\tilde{M}$ it can immediately be verified, using the fact from

Remark 2 that $T_r^{-2} e_n = \lambda_n^{-1} e_n$, that $\{\tilde{e}_n\}$ is a c.o.n.s. in $\tilde{M}$ if

$\tilde{e}_n = \lambda_n^{1/2} e_n$. Then, with $\tilde{w}$ defined as in the statement of the theorem,

we have with probability one,

$$(\tilde{w}, f_n)_\sim = (w, f_n) \tag{3.21}$$

and
$$
\begin{aligned}
(\tilde{w}, \tilde{e}_n)_\sim &= (T_r^{-1} w, T_r^{-1} \tilde{e}_n) \\
&= (w, T_r^{-2} (\lambda_n^{1/2} e_n)) \\
&= \lambda_n^{-1/2} (w, e_n).
\end{aligned} \tag{3.22}
$$

From (3.21) and (3.22) and the fact that w is weakly measurable it

follows that $(\tilde{w}, f_n)_\sim$ and $(\tilde{w}, \tilde{e}_n)_\sim$ are measurable scalar-valued functions.

Then a limit argument implies that $\tilde{w}$ is weakly measurable, and hence

is an $\tilde{H}_2$-valued random variable (see Appendix B).

It now follows from the fact that $E\|\tilde{w}\|_\sim^2 = E\|w\|_\sim^2 < \infty$, which

has already been established, that $\tilde{w}$ has a bounded (and hence self-

adjoint) covariance operator $\tilde{K}$. The proof is the same as the corres-

ponding proof in Appendix B. To show that $\tilde{K}$ is nuclear we show that

$\tilde{K}^{1/2}$ is Hilbert-Schmidt and use the fact ([G-1], p. 39) that the square

of a Hilbert-Schmidt operator is nuclear. In fact

$$
\sum_{n=1}^{\infty} \|\tilde{K}^{1/2} \tilde{e}_n\|_\sim^2 = \sum_{n=1}^{\infty} (\tilde{K} \tilde{e}_n, \tilde{e}_n)_\sim = \sum_{n=1}^{\infty} E|(\tilde{e}_n, \tilde{w})_\sim|^2
$$

$$
= \sum_{n=1}^{\infty} \lambda_n^{-1} E|(e_n, w)|^2
$$

$$
= \sum_{n=1}^{\infty} \lambda_n^{-1} (K e_n, e_n) < \infty \tag{3.23}
$$

and,

$$\sum_{n=1}^{\infty} ||\tilde{K}^{1/2} f_n||_{\sim}^2 = \sum_{n=1}^{\infty} (\tilde{K} f_n, f_n)_{\sim}$$

$$= \sum_{n=1}^{\infty} (K f_n, f_n) < \infty, \tag{3.34}$$

where we have used (3.21) and (3.22).

The finiteness of the left sides of (3.23) and (3.24) demonstrates that $\tilde{K}^{1/2}$ is Hilbert-Schmidt ([G-1], p. 34). The existence of the mean of $\tilde{w}$ follows as in Appendix B. Equations (3.21) and (3.22) together with the fact that $Ew = \theta$ imply that $E\tilde{w} = 0$.

To prove the converse we suppose $\tilde{K}$ exists and is nuclear. Then $\tilde{K}$ has finite trace, so a fortiori the sum $\sum_{n=1}^{\infty} (\tilde{K} \tilde{e}_n, \tilde{e}_n)_{\sim}$ must be finite. But this sum is the same as $\sum_{n=1}^{\infty} ||\tilde{K}^{1/2} \tilde{e}_n||_{\sim}^2$, and the conclusion follows by (3.23). |||

Thus, in the identification example, if the condition (3.19) is satisfied, we can replace H_2, U and w by $\tilde{H}_2$, $\tilde{U}$ and $\tilde{w}$, respectively, and then apply Theorem 2.2 to the new model with the assurance that the new model is just as faithful a representation of the real-life problem as was the original one. Theorem 2.2 will apply, of course, only if the covariance operator of the noise is such that conditions (2) and (3) of that theorem hold. These conditions do not conflict, at least in general, with (3.18), as consideration of Corollary 2.1 and the remark (2) following it will show.

As indicated earlier there are a number of comments to be made, both regarding the example as such, and its place in identification theory.

Remarks (1). The condition (3.18) can be interpreted as specifying

that a "noise-to-signal-ratio" is finite, or more properly that the sum

of noise-to-signal ratios in each orthogonal component is finite. (Ke_n, e_n)

is the variance of the noise component (w, e_n), and λ_n is the square of

the "signal" component along e_n.

(2) There is no difficulty in extending the system model of (3.7) to the

finite-dimensional vector case where u and y are vector-valued (c.f.

[R-3]). The same results follow.

(3) The restriction that F be time-invariant can be relaxed in the fol-

lowing way with no essential changes resulting. Let the kernels k_n be

of the form:

$$k_n(t, v_1, \ldots, v_n) = \sum_{i=1}^{M} \alpha_i(t) k_n^{(i)}(v_1, \ldots, v_n) \qquad (3.20)$$

With M finite and suitable restrictions on $\alpha_i(t)$, the function spaces are

still Hilbert spaces, and U is still compact (c.f. [R-4]). Certain classes

of time-varying systems can be modeled this way.

(4) The estimate one obtains is, of course, only for F_1, the projection

of F on $N(U)^\perp$. How "large" $N(U)^\perp$ is depends of course on U. If τ

is finite, one can choose a sequence of inputs $u_1, u_2, \ldots,$ of duration

m-τ, say (recall that m is the length of T); then, allowing τ units of

time between inputs (dead-time), make repeated measurements. If p

measurements are made, the output space becomes the p-fold product

of $L_2(T)$, the input space becomes the p-fold product of $L_2(T-\tau)$, and

U is suitably determined by $u_1, \ldots, u_p$. $\mathcal{K}$, however, remains effectively

the same. Clearly, $N(U)^\perp$ is a monotone nondecreasing sequence of

subspaces with increasing p. One would expect to do a better job of

identifying F as more measurements are made, if the u_i are suitably

chosen (see Remark (5)). But making the P repeated measurements

is just one special way of making one measurement where the input

has duration pm-τ and the output has duration pm.

(5) From the point of view of identification theory as reasonably con-

strained by practical considerations, we feel a better approach than

what has been done here is the following. Assume a priori that $\mathcal{B}'$

is a compact subset of $\mathcal{L}$ and that $\mathcal{U}$ is a compact subset of $L_2(R^1)$.

Then, given $\epsilon > 0$, the model (3.4) can be replaced by a finite set of lin-

ear equations with finitely many unknown parameters to be found with the

property that the solutions for the parameters in the finite set of equa-

tions will determine the true system in $\mathcal{B}'$ uniformly to within the speci-

fied $\epsilon > 0$. With the addition of observation noise w, the problem be-

comes that of estimating parameters in a finite-dimensional version of

(3.6). Then the classical theory of pseudoinverses and of LUMV esti-

mation applies (c.f. [R-3], for example. One can at least argue that

the compactness assumptions required are reasonable for most prob-

lems.

When identification is studied from the point of view of finite-

dimensional approximations, the question raised in the previous re-

mark about the number of measurements necessary (or the required

duration of one measurement) is easier to phrase satisfactorily, and is

often answerable in principle at least (c.f. [R-3]).

(6) In fairness to the reader unfamiliar with the area, it should be

pointed out that the application of Volterra polynomials to problems

in system identification has been considered for some time (c. f. [B-2])

but as has been indicated, has not been widely used in practice. A

much more common approach is to estimate parameters directly in a

state-variable model, using techniques from the Kalman-Bucy recur-

sive filtering theory.

PART IV

ON THE QUADRATIC REGULATOR PROBLEM[1]

In optimal control theory, no problem has received more atten-

tion or attained a greater degree of maturity than the one indicated by

the title of this section (see [I-1] for detailed summaries and biblio-

graphical information). Various versions of the quadratic regulator

problem (hereafter designated QRP) have appeared in standard texts

[A-3] [B-6], perhaps because minimization of a quadratic loss func-

tional applied to a linear dynamical system with linearly constrained

inputs leads to a simple and easily understood solution. Here we shall

exhibit the PI as unifying apparently disparate aspects of the QRP, pro-

viding generalizations beyond its customary form, and having broad

capabilities to solve QRP specified in infinite dimensional spaces and

by unbounded operators.

The optimal control literature generally relates the QRP to the

notion of a linear dynamical system, as described by[2]

$$\dot{x} = C(t)x + B(t)u \qquad\qquad x(0) = x^o. \qquad\qquad (4.1)$$

Here $u(\cdot)$ and $x(\cdot)$—usually referred to as input and output, respective-

ly—are finite dimensional vector functions of time over $[0, T]$, $C(\cdot)$ and

---

[2] The PI was first applied to a QRP for a linear dynamical system in
[K-1].

B($\cdot$) are matrix functions of consistent dimension and assuring that a unique solution $x \in L_2^n$ corresponds to every input $u \in L_2^p$. By L_2^m we mean a Hilbert function space; $v \in L_2^m$ implies that $v(t)$ is an m dimensional vector for each $t \in [0, T]$, with v having measurable components v_i, and being possessed of the L_2^m norm

$$|||v||| = \sqrt{\sum_{i=1}^{p} \int_0^T |v_i(t)|^2 dt} = \sqrt{\int_0^T ||v(t)||^2 dt} . \qquad (4.2)$$

We can now state one form of the QRP, sometimes called the fixed endpoint QRP. It is desired to choose $u \in L_2^p$ which moves the system from $x(0) = x^o$ to a designated $x(T) = x^1$, and among the u satisfying this constraint find the u_o having minimum L_2^p norm. (See [B-6], p. 137). As usually phrased and solved, this QRP presupposes conditions insuring that any x^1 can actually be attained by the system (4.1) at time T; the solution is then said to exist. But actually, this is an unnecessarily narrow interpretation of the fixed endpoint problem due to the limitation of the techniques sometimes used. Application of the PI places this QRP in a natural setting admitting easy generalizations.

Let us begin our analysis by describing a classical approach to the fixed point QRP, and thus exhibiting the limitations resulting therefrom. Since

$$x(T) = \int_0^T W(T, t)B(t)u(t)dt + W(T, 0)x^o \qquad (4.3)$$

in terms of the state transition matrix $W(\cdot, \cdot)$, the constraint $x(T) = x^1$ can be expressed for an n dimensional x^1 by the simultaneous linear equations

$$(u, h_i) = a_i , \qquad\qquad i = 1, 2, \ldots, n \qquad\qquad (4.4)$$

with a_i the ith component of the n vector $x^1 - W(T,0)x^o$, i.e.

$$a_i = [\delta_{1i} \ \delta_{2i} \ \cdots \ \delta_{ni}] \ \{x^1 - W(T, 0)x^o\} ,$$

and $\overline{h_i(\cdot)}$ the p component row vector function with i^{th} entry

$$\overline{h_i(t)} = [\delta_{1i} \ \delta_{2i} \ \cdots \ \delta_{ni}] W(T, t)B(t) . \qquad\qquad (4.5)$$

To complete the classical solution, let M be the subspace of L_2^p

spanned by $h_1, h_2, \ldots, h_n$, and write u in terms of the orthogonal de-

composition

$$u = \sum_1^n \alpha_i h_i + u_1 \qquad\qquad (4.6)$$

with $u_1 \in M^\perp$. If (4.6) is substituted in (4.4), we see u meets the

required constraint $x(T) = x^1$ iff

$$\sum_{j=1}^n \alpha_j (h_j, h_i) = a_i \qquad i = 1, 2, \ldots, n. \qquad\qquad (4.7)$$

The second term u_1 in (4.6) plays no role in meeting the constraint,

and (by the Pythagorean theorem for orthogonal elements in Hilbert

space) only serves to increase $|||u|||$; consequently, $|||u|||$ is mini-

mized uniquely by choosing

$$u_1 = \theta . \qquad\qquad (4.8)$$

Evidently, the system can satisfy the final state constraint

$x(T) = x^1$ for arbitrary x^1 iff the matrix with elements $\{(h_i, h_j)\}$ is

nonsingular. This matrix is Gramian [A-1], which means that its

invertibility is equivalent to the linear-independence of the h_i,

$i = 1, 2, \ldots, n$ in L_2^p. If we say—following standard terminology in

linear systems theory—that the system is controllable over $[0, T]$ if,

for every x^o and x^1 there exists an input $u \in L_2^p$ taking the system from

$x(0) = x^o$ to $x(T) = x^1$, we have attained the following result: the system

(4.1) is controllable over $[0, T]$ iff the n rows of the matrix $W(T, \cdot)B(\cdot)$

are linearly independent in L_2^p. This criterion is well known in sys-

tems theory [C-1]; our arguments constitute a short and elegant proof

of its validity.

We now turn to a PI formulation generalizing the fixed endpoint

QRP in several directions. First, the QRP can be solved even if the

system is not controllable. In fact, the optimal input u_o to (4.1) min-

imizes $\|x(T) - x^1\|$ over all $u \in L_2^p$, regardless of controllability; if

the system (4.1) should be controllable, this norm is zero and the u_o

obtained via the PI model agrees with the optimum resulting from the

classical problem statement. Second, we shall be able to consider

constraints not only on the endpoint $x(T)$, but also on x and u. These

three classes of constraints can all be expressed in terms of linear

functionals on u, so that they can be viewed in unified form. It is also

noteworthy that the PI yields valid optima even when the constraints are

incompatible with one another. Finally, the PI model permits us to

consider dynamical systems (4.1) in which $u(t)$ and $x(t)$ are Hilbert-

space-valued (i.e., infinite dimensional) for each $t \in [0, T]$, although

we will continue to be restricted to a finite set of constraints.

To apply the PI to the fixed endpoint QRP, we need to express

the input-output system relation by a linear transformation, that is

$$Lu = x, \tag{4.9}$$

where u is any element of a Hilbert space H_1, and $L:H_1 \rightarrow H_2$ is a linear bounded operator defined for every such u. In general, a dynamical system such as (4. 1) fails to satisfy (4. 9), but we may write (substituting $\tilde{x}$ for x)

$$\tilde{x}(t) = (Lu)(t) + S(t)x^o ; \tag{4. 10}$$

then, taking $x = \tilde{x} - Sx^o$, we see (4. 9) to be applicable without loss of generality.

The linear constraints are considered next. They represent desired system behavior, but do not preclude the possibility that they cannot be attained; in the latter event, the PI automatically chooses a u_o which causes the constraint values to be approached as closely as possible. As we have mentioned, there are three types of constraints, all of which ultimately lead to constraint equations having the form

$$(u, f_i) = a_i .$$

A linear functional on u is already in the stated form. As for a linear functional on x, we may write $G_i(x) = a_i$ which, by the Riesz representation theorem and the definition $f_i = L^*g_i$ gives rise to

$$G_i(x) = (x, g_i) = (Lu, g_i) = (u, L^*g_i) = (u, f_i) = a_i; \tag{4. 11}$$

the constraint is then in the proper form. Lastly, choosing $f_i = h_i$ as in (4. 5), and defining a_i as in the equation above, (4. 5) renders $(u, f_i) = [x(T)]_i$ and $a_i = x_i^1$, whence $(u, f_i) = a_i$, i=1, 2, ... , n is equivalent to $x(T) = x^1$.

Application of the PI involves an operator equation to which the PI gives the BAS. Thus, the constraints expressed as $(u, f_i) = a_i$ must

be translated into a more appropriate form for this purpose. According-ingly, let H_3 be a separable Hilbert space with a maximal linearly independent set $\{y_n\}$, and define $A : L_2^p \to H_3$ by

$$Au = \sum_i (u, f_i) y_i . \tag{4.12}$$

Then, with

$$z = \sum_i a_i y_i \tag{4.13}$$

we obtain

$$Au = z \tag{4.14}$$

as the operator equation embodying all the desired constraints.

If a finite constraint set is set forth, the range of A is a finite dimensional set in H_3 and perforce closed. By Theorem 1.2 and Corollary 1.2, A possesses a PI A^+ which is bounded and delivers a BAS (cf. Definition 1.1)

$$u_o = A^+ z = \sum_i a_i A^+ y_i , \tag{4.15}$$

which means $u_o \in L_2^p$ is the unique input to the dynamical system satis-fying

$$\inf_{u \,\in\, L_2^p} ||z - Au||^2 = \inf_{u \,\in\, L_2^p} \left\{ \sum_{i,j} [(u, f_i) - a_i][\overline{(u, f_j) - a_j}](y_i, y_j) \right\} = \tag{4.16}$$

$$\sum_{i,j} [(u_o, f_i) - a_i][\overline{(u_o, f_j) - a_j}] (y_i, y_j),$$

and having smallest L_2^p norm (i.e., least energy) attaining the infimum (4.16). By choosing suitable families $\{y_n\}$ we are thus able to prove existence and uniqueness of optimal controls for a variety of quadratic

loss functions. We remark parenthetically that if there exists a

$u \in L_2^p$ exactly satisfying all the constraints $(u, f_i) = a_i$, (4.16) be-

comes zero, and u_o is the element of smallest norm which meets the

constraints. In particular, if the constraints represent $x(T) = x^1$ and

the system (4.1) is controllable, the u_o furnished by the PI (4.15) is

the same as the optimum obtained as the solution of the classical

fixed endpoint QRP.

In applying the PI to the QRP in the manner described, we need

not be concerned with the consistency of the constraints; if the con-

straints are incompatible, the PI mediates among them to produce a

compromise BAS. To illustrate this property, we given an example

involving a simple constraint. Consider therefore the system (4.4)

with n=2, $h_1 = h_2 = h$ and $a_1 \neq a_2$, so that (4.7) is clearly impossible

to satisfy. We may, however, apply (4.15) to determine the BAS u_o.

The latter naturally depends on the choice of y_1 and y_2, as can be seen

from (4.16). If we choose y_1 and y_2 orthonormal (for convenience), we

have $N(A)^{\perp} = V\{h\}$, $R(A) = V\{y_1 + y_2\}$, $Ah = |||h|||^2 (y_1 + y_2)$, and

$z = a_1 y_1 + a_2 y_2$. In view of $P_R z = \frac{1}{2}(a_1 + a_2)(y_1 + y_2)$, the description

of A^+ by Theorem I.2 leads to

$$u_o = A^+ z = \frac{1}{2}(a_1 + a_2) \frac{h}{|||h|||^2} .$$

Thus far, reference in (4.1) has been to dynamical systems in

which $x(t)$ and $u(t)$ are finite dimensional vectors of fixed dimension

for each t. Since we have made no use of the finite dimensionality, we

may as well generalize (4.1) as follows: $U(t) \in K_1$ for almost every

$t \in [0, T]$, K_1 being a separable Hilbert space. Assume $u(\cdot)$ strongly

(Bochner) measurable [H-2], and $u \in H(K_1)$, that is

$$|||u|||^2 = \int_0^T ||u(t)||^2 dt < \infty. \qquad (4.17)$$

Further, in (4.1), let $B(\cdot)$ be an essentially uniformly bounded meas-

urable operator valued function which for each t carries K_1 into a new

Hilbert space K_2, and let $C(\cdot)$ be locally Bochner integrable, with $C(t)$

an endomorphism on K_2 for almost every t. Then, if (4.1) is inter-

preted in integral equation form, there is a unique solution $x(\cdot)$ which

is uniformly bounded (cf. [M-1]) and hence belongs to $H(K_2)$; indeed,

one obtains (4.9) and (4.10) as before, with $L : H(K_1) \to H(K_2)$ and

$S(T) : K_2 \to K_2$ both bounded operators. The remainder of the analysis

is unchanged, provided that there is a finite constraint set, which then

implies $R(A)$ in (4.14) to be finite dimensional.

The finite dimensionality of $R(A)$ [with A specified by (4.12)] be-

comes inconsistent with a constraint of the type $x(T) = x^1$ whenever K_2

is infinite dimensional. At this point we could confine ourselves to a

finite constraint set as represented by the projection of x^1 onto a hyper-

plane of finite dimension. Alternatively, we might attempt to deal more

directly with infinite dimensional $x(T)$ (or any infinite constraint set)

by choosing $\{f_i\}$ and $\{y_i\}$ in

$$Au = \sum_1^\infty (u, f_i) y_i$$

so that A [cf. also (4.12)] is DDC with closed range. Only then is the

PI A^+ defined for arbitrary z (Corollary 1.2) and hence for every choice

of x^1. Moreover, z makes sense only if the sum $\sum_i a_i y_i$ appearing in

(4.13) is convergent in H_3. The requirements on A and z are formidable, and cannot be met in general. For instance, the obvious terminal constraint choice of $\{y_i\}$ as a complete orthonormal set in $H_3 = K_2$ and the a_i as projections of x^1 on $V\{y_i\}$ lead to a convergent sum for z, but in combination with a DDC operator A whose range is not closed.

The limitation of the above PI technique to a finite constraint set is avoided in a somewhat different QRP model, which is termed the free endpoint QRP. As usually described in the optimal control literature, the free endpoint QRP is concerned with moving the system (4.1) from $x(0) = x^o$ toward the origin at time T with minimum expenditure of energy, simultaneously maintaining x as small as possible. More precisely ([B-6], Section 21), one wishes to attain the infimum of the loss function

$$J(u) = \int_0^T [||u(t)||^2 + ||Q_2 x(t)||^2]dt + ||Q_3 x(T)||^2 \qquad (4.18)$$

where the $||\cdot||$ are Euclidian norms on the relevant finite dimensional vector space. Analysis of the free endpoint QRP revolves around the existence and uniqueness of the optimal control [i.e., u attaining the infimum of (4.18)], and the computation of the optimum whenever it does exist.

In what follows, we shall generalize the free endpoint QRP to infinite dimensional spaces, and demonstrate that the optimal input is (uniquely) provided by the PI even for some models involving unbounded operators. Moreover, our loss function

$$I(u) = ||Q_1(u-z_1)||^2 + ||Q_2(x-z_2)||^2 + ||Q_3[x(T)-z_3]||^2 \qquad (4.19)$$

will be more general than the $J(\cdot)$ of (4.18). The z_i in (4.19) may be thought of as target values of the input, output and terminal state sought in the system design; they also reflect the change of variable (4.10), so that (4.19) becomes equivalent to (4.18) if we specify in particular $z_1 = \theta$, $z_2(t) = -C(t)x^o$ and $z_3 = -C(T)x^o$. We should mention, however, that any interpretation of (4.19) in terms of an optimal control problem is merely an intuitive convenience, since we shall make no use of any special properties of the dynamical system (4.1).

For treating the minimization problem incompletely posed by (4.19), we establish the following structure. Let $u \in H_1$ (an arbitrary Hilbert space), and take $L:H_1 \to H_2$, with

$$Lu = x. \qquad (4.20)$$

It will be supposed that L — like all other operators mentioned here — is linear and DDC. We shall assume further that H_2 is a Hilbert function space, so that we are writing x for $x(\cdot)$, and that there is a linear DDC operator $S:H_1 \to H_3$ such that

$$x(T) = Su. \qquad (4.21)$$

We introduce yet another Hilbert space, $K = H_1 \times H_2 \times H_3$; its norm is the standard

$$|| \{u, v, w\} ||^2 = ||u||^2 + ||v||^2 + ||w||^2 . \qquad (4.22)$$

The operator $A: H_1 \to K$ is defined by

$$Au = \{u, Lu, Su\} \qquad (4.23)$$

and we have

$$D(A) = D(L) \cap D(S). \qquad (4.24)$$

Since L and S are both closed, Au is a closed graph in K, and A itself is a closed operator. But $D(A)$ is not necessarily dense in H_1, so A must be assumed densely defined by hypothesis. Lastly, let us define $Q: K \to K$ by the relation

$$Q\{u, v, w\} = \{Q_1 u, Q_2 v, Q_3 w\} \tag{4.25}$$

where each Q_i is not only bounded, but also bounded from below. The latter condition is expressed by the existence of an $\alpha > 0$ such that

$$\|Q_i y_i\| \geq \alpha \|y_i\| \qquad \text{for all } y_i \in H_i . \tag{4.26}$$

We can now state

Theorem 4.1: Let L and S be DDC operators as defined above, with $D(L) \cap D(S)$ dense in H_1. Assume that each Q_i is bounded and bounded below. Then, for each $z = \{z_1, z_2, z_3\} \in K$ there exists a unique $u_o \in H_1$ satisfying

$$I(u_o) = \inf_{u \in H_1} [I(u)], \tag{4.27}$$

with I as in (4.19). This u_o is given by

$$u_o = (QA)^+ Qz . \tag{4.28}$$

Proof: It is clear from (4.25) and the properties of the Q_i that Q is an endomorphism on K, and is bounded below; from this (and that A is DDC) one shows QA to be a closed operator. Furthermore, A is densely defined, so the same is true for QA, and in fact QA is DDC.

The range of QA is closed because [see Lemma A.8(e)] for all $u \in D(A)$

$$\|QAu\| \geq \|Q_1 u\| \geq \alpha \|u\| , \tag{4.29}$$

and the inequality also shows that QA is one-to-one. In the light of

these properties, we can now consider the BAS for the equation

$$(QA)u = Qz ; \qquad\qquad (4.30)$$

since QA is DDC with R(QA) closed, the BAS is furnished for every

(Qz) ϵ K (Corollaries 1. 1 and 1. 2) by the PI. This means the u_o of

(4. 28) satisfies

$$||Qz - (QA)u_o|| = \inf_{u \,\epsilon H_1} ||Qz - (QA)u|| . \qquad\qquad (4.31)$$

If we compare the norm of Qz-(QA)u with the expression (4. 19) for

I(u), we see at once they are identical, so that (4. 31) is equivalent

to (4. 27).

To prove u_o is the unique element attaining the infimum of (4. 31),

note (from the proof of Theorem 1. 1) that $(QA)u_o$ must be the projec-

tion of Qz on R(QA). But QA is one-to-one, so u_o is the only element

of H_1 such that $(QA)u = P_R(Qz)$. |||

Various other combinations of assumptions also yield free end-

point QRP having the same solution (4. 28). Most of these are of little

interest, and can be reproduced by the reader as needed. However,

we may want to consider bounded L and S, especially since these corres-

pond to the dynamical system (4. 1) with infinite dimensional vector

valued functions as per our earlier discussion. We find that the

assumptions on Q_2 and Q_3 can then be relaxed, as is indicated by

Corollary 4. 1:   If L and S are bounded operators, if each Q_i is bounded

and Q_1 is bounded from below, the conclusions of Theorem 4. 1 continue

to hold.

Proof: QA is now a bounded operator which is everywhere defined. The remainder of the proof of Theorem 4.1 remains unchanged, the same assertions being valid throughout. |||

PART V

PSEUDOINVERSE OPERATOR APPROXIMATIONS

In the literature on the matrix PI, it is sometimes proposed that

the PI be obtained by a sequence of approximations involving AA^* or

A^*A (see [A-2], Theorem 3.4 and [L-2], p. 167). Such forms exhibit

computational simplicity, since they substitute the inversion of a strict-

ly positive (Hermitian) matrix for more complicated combinations of

inverses. When approximations of the same type are applied to DDC

operators, the objectives are necessarily different. Computational

simplicity is no longer at issue; we are now concerned with analytical

tractability, particularly for unbounded operators whose range may not

be a closed set.

One goal is to circumvent the problems that inevitably arise when

$R(A)$ is not closed, and A^+ is unbounded and not defined everywhere;

we have already sought solutions in Parts III and IV by changing the

topology, but now we also consider approximating sequences of opera-

tors. The description of A^+ and approximations thereto in terms of the

positive operators AA^* or A^*A not only generalizes matrix approxima-

tions of the same type, but also suggests use of the spectral represen-

tation and the associated functional calculus. We shall find that the PI

can be expressed in terms of positive operators (and the functional

calculus) as a consequence of the use of the polar decomposition ([D-2],

XII. 7.7); indeed, the polar decomposition proves to be an extremely

powerful tool in the analysis of the PI, and will be used often in this section. Finally, we remark that the approximations considered below are closely related to the operators A' and A'' defined by (1.38) and (1.44), respectively; these operators are studied extensively in Part I, and find application in Parts II and III.

Since we are dealing with DDC operators, we shall face questions that fail to arise in the context of the matrix PI. In addition to the boundedness and domain properties as in Part I, we shall need to investigate the mode of convergence for each approximation to the PI. In particular, we shall show that certain approximation sequences always converge to the PI strongly, but in norm iff $R(A)$ is a closed set.

In what follows, we shall freely use the notation and results of Part I and Appendix A, although we shall assist the reader with specific references whenever possible. For future reference, we also call the reader's attention to the well known content of

<u>Lemma 5.1</u>: Every DDC operator A has a polar decomposition

$$A = VS \tag{5.1}$$

where $V:H_1 \to H_2$ is a partial isometry from $N(A)^{\perp}$ onto $\overline{R(A)}$, and S is the positive operator $(A^*A)^{1/2}$ with spectral representation

$$S = (A^*A)^{1/2} = \int_0^\infty \lambda \, dG_\lambda. \tag{5.2}$$

In (5.2), G_0 is the projection on $N(A)$. Alternatively, A may be decomposed as

$$A = TU^*, \tag{5.3}$$

$U:H_2 \to H_1$ being a partial isometry from $\overline{R(A)}$ onto $N(A)^{\perp}$ [and null on

$R(A)^{\perp}]$, and $T:H_2 \to H_1$ the positive operator

$$T = (AA^*)^{1/2} = \int_0^\infty \lambda \, dE_\lambda \; ; \qquad (5.4)$$

here E_0 is the projection on $R(A)^{\perp}$.

Proof: The first decomposition is stated and proved in [D-2], XII. 7. 7, and the spectral representation discussed in XII. 2 of the same reference. Now apply this decomposition to A^*, i.e., $A^* = UT$. On taking the adjoint of A^*, we obtain (5.3); the equality is a consequence of U bounded ([R-2], Section 115).

From the equality ([D-2], XII. 2. 6)

$$|| Ty ||^2 = \int_0^\infty \lambda^2 d || E_\lambda y ||^2 \qquad (5.5)$$

and the properties of the resolution of the identity, we see that E_0 is the projection on $N(T) = R(T)^{\perp} = R(AA^*)^{\perp} = R(A)^{\perp}$ (compare Corollary A. 3). The proof that G_0 is the projection on $N(A)$ is similar. |||

The representations (5. 1) and (5. 3) lead to a natural way of expressing approximations to the PI when these involve

$$AA^* = T^2 \qquad \text{or} \qquad A^*A = S^2 . \qquad (5.6)$$

To conveniently compare the approximations with A^+, we then write the latter in terms of T (or S) also. Furthermore, the formulation of A^+ in terms of the polar decomposition may be of some independent interest. Motivated by these considerations, we state and prove

Theorem 5. 1: The PI A^+ of A can be represented as[1]

$$A^+ = U_r T_r^{-1} P_R \tag{5.7}$$

or alternatively

$$A^+ = S_r^{-1} V_r^{-1} P_R . \tag{5.8}$$

Proof: We apply the characterization of $A^+ = A_r^{-1} P_R$ of Theorem 1.2. From the definition of U as a partial isometry, $N(U^*) = N(A)$, so the results of Appendix A on restrictions of operators are applicable. By Lemma A.9, we may write U^* to denote either $(U^*)_r$ or $(U_r)^*$ interchangeably; hence, for $x \in N(A)^\perp$, $A_r x = Ax = TU^* x = TU_r^* x$. Furthermore, U_r^* is a unitary operator on the restricted spaces, with $R(U_r^*) = R(U^*) = \overline{R(A)}$ so that we even have

$$A_r = TU_r^* = T_r U_r^* = T_r U_r^{-1} . \tag{5.9}$$

To compute A_r^{-1} from (5.9)—this inverse exists by Lemma A.6—we first argue that T_r is invertible; this follows because (5.9) implies $A_r U_r = T_r$, with both operators on the left side of the equality having inverses. With the existence of T_r^{-1}, inversion of (5.9) yields $A_r^{-1} = U_r T_r^{-1}$ (c.f., [R-2], Section 114), and this proves (5.7).

The argument on (5.8) is analogous and even easier. $\overline{R(S)} = N(A)^\perp$, which means $A = V_r S$, V_r being a unitary operator from $N(A)^\perp$ to $\overline{R(A)}$. Clearly, $A_r = V_r S_r$ from which $A_r^{-1} = S_r^{-1} V_r^{-1}$. Then A^+ is given by (5.8) as claimed. |||

---

[1] To avoid possible confusion, we again remind the reader that the restrictions applied to any operator are to the nullspace or closure of the range of A. The identities and relationships given in Appendix A apply to U_r, T_r, etc. only because the null and/or range spaces coincide with $N(A)$ and/or $R(A)$.

The restrictions of S and T are easily identified in terms of the spectral representations (5.2) and (5.4). Since the spectral representations will appear again later in this section, we need

Lemma 5.2: Let

$$H_\lambda = G_\lambda - G_o \quad \text{and} \quad F_\lambda = E_\lambda - E_o . \tag{5.10}$$

Then S_r and T_r have the respective spectral representations

$$S_r = \int_0^\infty \lambda \, dH_\lambda \quad \text{and} \quad T_r = \int_0^\infty \lambda \, dF_\lambda \tag{5.11}$$

Proof: The verifications are identical for S_r and T_r, so we need prove the spectral representation only for the first of these. From Theorem 5.1, G_o is the projection on $N(A)$, whence $G_\lambda x = H_\lambda x$ for any $x \in N(A)^{\perp}$; this means the spectral integrals (5.2) and (5.11) are the same when applied to any such x. It remains to show that H_λ is a resolution of the identity on the Hilbert space $N(A)^{\perp}$. Evidently, H_λ is a right continuous increasing family of projections with $H_o = \theta$ and $\lim_{\lambda \to \infty} H_\lambda = I - G_o$. The latter is the projection on $N(A)^{\perp}$, which is the identity in the restricted space. Thus, H_λ satisfies all the conditions required of a resolution of the identity. |||

We are now ready to define the approximations to the PI which constitute the principal objects of our study. For any positive number σ, let

$$A'_\sigma = A^*[\sigma I + (AA^*)]^{-1} . \tag{5.12}$$

In the finite dimensional case ([A-2], Theorem 3.4 and [L-2], p. 167) the limit of A'_σ exists as $\sigma \to 0$, and in fact

$$\lim_{\sigma \to 0} A'_\sigma = A^+ . \tag{5.13}$$

Since one is only required to invert a positive matrix to implement A'_σ,

(5. 12) appears as an attractive formula for approximating the PI.

Some matrix manipulations on (5. 12), followed by a passage to the

limit, yields

$$\lim_{\sigma \to 0} A'_\sigma = A', \tag{5.14}$$

where we have previously defined

$$A' = A^*(AA^*)_r^{-1} P_R . \tag{5.15}$$

Clearly, A' is another form of the PI in finite dimensional spaces; this

also is obtained as a special case of our Corollary 1. 4.

As we shall see, the simple and straightforward results pertain-

ing to finite dimensional spaces continue to hold in the more general

context of operators whose range is closed. In the event $R(A)$ is not

closed, A' is an RPI as shown by Theorem 1. 7, whereas A'_σ behaves

quite differently from A', and moreover fails to converge to A^+ in norm

as $\sigma \to 0$. The comparison of A^+, A and A'_σ for operators of non-closed

range is of particular interest in Part II, where the PI of a Hilbert-

Schmidt operator makes its appearance.

The essential facts regarding A'_σ are summarized in

<u>Theorem 5. 2</u>: The operator A'_σ given by (5. 12) is bounded and defined

on all H_2, and the same is true for

$$AA'_\sigma = (AA^*)[\sigma I + (AA^*)]^{-1} . \tag{5.16}$$

The restriction of A'_σ to $D(A^+)$ converges strongly to A^+. The con-

vergence

$$\lim_{\sigma \to 0} A'_\sigma = A^+ \tag{5.17}$$

is uniform (i.e., in norm) iff $R(A)$ is a closed set.

Remark: Since AA'_σ is defined for all $z \in H_2$, one might conjecture that A'_σ accomplishes what A^+ cannot. As we know, the BAS fails to exist and $z \notin D(A^+)$ if $P_R z \notin R(A)$ (see Theorem 1.1), but it is reasonable to expect that A'_σ delivers something close to the optimum as σ becomes small. These possibilities will be explored later in this section.

Proof: It is known that $[\sigma I + (AA^*)]^{-1}$ is bounded and everywhere defined ([R-2], Section 118), and since A^* is DDC, the operator A'_σ is closed. We also have

$$R\{[\sigma I + (AA^*)]^{-1}\} = D[\sigma I + (AA^*)] = D(AA^*) \subset D(A^*) \qquad (5.18)$$

so A'_σ is everywhere defined and, being closed, must be bounded according to the closed graph theorem. The same reasoning shows AA'_σ bounded and everywhere defined also.

For a comparison of A^+ and A'_σ it suffices to consider the application of these operators to $z \in R(A)$. In fact, $N(A^+) = R(A)^\perp$ as shown in Lemma 1.1, and we now prove $N(A'_\sigma) \supset R(A)^\perp$. Suppose then $z \in R(A)^\perp$, which implies $[\sigma I + (AA^*)]^{-1} z = \sigma^{-1} z \in R(A)^\perp$, or in other words, $R(A)^\perp$ reduces $[\sigma I + (AA^*)]^{-1}$. But $N(A^*) = R(A)^\perp$, so we obtain $A'_\sigma z = \theta$.

To demonstrate the asserted strong convergence of A'_σ to A^+, we apply $(A'_\sigma - A^+)$ to $z \in R(A)$. Since $z = Ax$ for some $x \in N(A)^\perp$, and since $A^+ A x = x$ for such an element (Theorem 1.3), we have

$$|| (A'_\sigma - A^+)z ||^2 = || \{A^*[\sigma I + (AA^*)]^{-1}A\}x - x ||^2 . \qquad (5.19)$$

All operations on the right side of (5.19) take place in the restricted

Hilbert spaces $N(A)^{\perp}$ and $\overline{R(A)}$ because $[\sigma I + (AA^*)]^{-1}$ reduces $\overline{R(A)}$

as we have shown. Thus we obtain

$$A_r^*[\sigma I_r + (AA^*)_r]^{-1}A_r = U_r T_r[\sigma I_r + T_r^2]^{-1}T_r U_r^{-1}. \tag{5.20}$$

Now U_r is unitary over the restricted Hilbert spaces, so that (5.19)

becomes

$$\| \{T_r[\sigma I_r + T_r^2]^{-1}T_r - I_r\}y\|^2 = \int_0^\infty \left| \frac{\sigma}{\sigma + \lambda^2} \right|^2 d\| F_\lambda y\|^2 \tag{5.21}$$

for $U_r^{-1}x = y \in D(T_r)$; the rules applicable to such calculations may be

found in [R-2], Section 128. On putting these last three equations to-

gether we find

$$\lim_{\sigma \to 0} \| A'_\sigma z - A^+z\|^2 = \lim_{\sigma \to 0} \int_0^\infty \left| \frac{\sigma}{\sigma + \lambda^2} \right|^2 d\| F_\lambda y\|^2. \tag{5.22}$$

The measure generated by $\| F_\lambda y\|^2$ is finite, and the integrand in (5.22)

is bounded by unity. Consequently, the Lebesque dominated convergence

theorem insures that the right side of (5.22) is zero, thus yielding the

strong convergence asserted by **the theorem.**

Suppose now $R(A)$ is not closed. Then A'_σ remains bounded but

A^+ is unbounded (Corollary 1.2), so A'_σ cannot converge in norm to A^+.

On the other hand, for $y \in R(A)$ it is always true that

$$\| (A'_\sigma - A^+)y\|^2 = \| U_r[T_r(\sigma I_r + T_r^2)^{-1} - T_r^{-1}]y\|^2 = \| [T_r(\sigma I_r + T_r^2)^{-1} - T_r^{-1}]y\|^2$$

$$\tag{5.23}$$

by the characterization (5.7) of A^+, and the fact that U_r is unitary.

The norm of (5.23) can once more be written in terms of the spectral

representation, viz.

$$\left|\left|\,(A'_{\sigma}-A^{+})y\right|\right|^{2} = \int_{0}^{\infty}\left|\frac{\sigma}{\sigma+\lambda^{2}}-\lambda^{-1}\right|^{2}\!d\left|\left|F_{\lambda}y\right|\right|^{2} = \int_{0}^{\infty}\left|\frac{\lambda^{-1}\sigma}{\sigma+\lambda^{2}}\right|^{2}\!d\left|\left|F_{\lambda}y\right|\right|^{2}.$$

$$(5.24)$$

If $R(A)$ is closed, $R[(AA^{*})_{r}]$ is likewise closed (Corollary A.3), whence T_{r}^{2} and $\underline{a}$ $\underline{fortiori}$ T_{r} have bounded inverses. But then ([R-2], Section 128) there exists an $\alpha > 0$ such that $F_{\lambda} = \theta$ whenever $\lambda < \alpha$. The lower limit of integration in (5.24) can then be changed from zero to α, with the result that the integrand is bounded by $\alpha^{-6}\sigma^{2}$ over the entire interval of integration. Consequently,

$$\left|\left|\,(A'_{\sigma}-A^{+})y\right|\right|^{2} \le \alpha^{-6}\sigma^{2}\left|\left|y\right|\right|^{2} \tag{5.25}$$

and hence A'_{σ} converges to A^{+} in norm as σ tends toward zero. |||

In Section I, we introduced not only $A' = A^{*}(AA^{*})_{r}^{-1}P_{R}$, but also another RPI which enjoyed a limited duality with A'. The latter operator was $A'' = (A^{*}A)_{r}^{-1}A^{*}$, whose domain was quite complicated (see Theorem 1.8) in general; nevertheless, A'' proved useful in connection with the applications of Section II. The relation between A' and A'' is mirrored by a corresponding similarity of A'_{σ} to A''_{σ}, where

$$A''_{\sigma} = [\sigma I + (A^{*}A)]^{-1}A^{*}, \qquad \sigma > 0 \tag{5.26}$$

The principal properties of A''_{σ} are stated in

<u>Theorem 5.3</u>: A''_{σ} is bounded and is possessed of the domain $D(A''_{\sigma}) = D(A^{*})$. The (closed minimal) extension of A''_{σ} to an operator defined on all H_{2} converges strongly (as $\sigma \to 0$) to A^{+} on $D(A^{+})$, and converges to A^{+} in norm iff $R(A)$ is a closed set. [1]

---

[1] Strong convergence of A''_{σ} for bounded A is already asserted in [B-1], p. 60.

Remark: If A (and hence A^*) is unbounded, A_σ'' cannot be closed. To

avoid the ensuing pathologies, we therefore state the convergence parts

of the above theorem in terms of the closed extension of A_σ'' .

Proof: The domain of A_σ'' is indeed D(A*) because $[\sigma I + (A^*A)]^{-1}$ is

bounded and everywhere defined on H_1. To show A_σ'' bounded, observe

that $(A^*)_\sigma' = A[\sigma I + (A^*A)]^{-1}$ is bounded and everywhere defined, as

may be seen by applying the last theorem to A^* in place of A. Then its

adjoint $[(A^*)_\sigma']^*$ is likewise bounded and defined everywhere of H_2. The

adjoint satisfies (see [R-2], Section 115c)

$$[(A^*)_\sigma']^* \supset A_\sigma'' . \tag{5.27}$$

Thus A_σ'' is bounded on $D(A^*)$, and since the latter is dense in H_2, the

minimal closed extension of A_σ'' is the bounded operator $[(A^*)_\sigma']^*$.

We call $[(A^*)_\sigma']^* = A_\sigma^X$. When R(A) is not closed, A^+ is unbounded,

and A_σ^X cannot converge to A^+ in norm. On the other hand, A^+ is

bounded whenever R(A) is a closed set, and then

$$||A_\sigma^X - A^+|| = ||(A_\sigma^X - A^+)^*|| = ||(A^*)_\sigma' - (A^+)^*|| = ||(A^*)_\sigma' - (A^*)^+|| \to 0.$$

$$\tag{5.28}$$

These relations are justified by: (1) a bounded operator and its adjoint

have the same norm ([H-1], Theorem 22.2), (2) $(A^+)^* = (A^*)^+$ by Theo-

rem 1.6, and (3) since $R(A^*)$ is closed also (Corollary A.1) $(A^*)_\sigma'$ con-

verges to $(A^*)^+$ by Theorem 5.2 above.

To complete the proof, we verify the strong convergence of A_σ^X

to A^+ on $D(A^+)$. We note at once that $N(A_\sigma'') = N(A^*) = R(A)^\perp = N(A^+)$,

so (as in the proof of Theorem 5.2) we confine ourselves to z $\in$ R(A),

which we will represent as z = Ax, x $\in$ N(A)$^\perp$. Consequently, it is

sufficient to show

$$\|A_\sigma^X Ax - A^+Ax\|^2 \to 0 . \qquad (5.29)$$

Here again $A^+Ax = x$, and $A_\sigma^X A$ may be replaced by its minimal closed

extension $S^2(\sigma I + S^2)^{-1}$. That the latter is in fact the closed minimal

extension follows from the definition of A_σ^X and

$$(A_\sigma^X A) = A^*(A^*)'_\sigma = S^2(\sigma I + S^2)^{-1}, \qquad (5.30)$$

in which we have used (5.16), and adopted once more the notation

$S^2 = A^*A$. The minimal extension is obtained from (5.30) by taking

the adjoint of both sides of the equation, and noting that

$$S^2(\sigma I + S^2)^{-1} = \int_0^\infty \frac{\lambda^2}{\sigma + \lambda^2} dG_\lambda \qquad (5.31)$$

wherefrom $S^2(\sigma I + S^2)^{-1}$ is self-adjoint. A repetition of the argument

employed in Theorem 5.2 indicates that the restriction of $S^2(\sigma I + S^2)^{-1}$

to $N(A)^{\perp}$ is simply $S_r^2(\sigma I_r + S_r^2)^{-1}$. The left side of (5.29) is then

$$\|(A_\sigma^X A - A^+A)x\|^2 = \|[S_r^2(\sigma I_r + S_r^2)^{-1} - I_r]x\|^2 = \int_0^\infty \left|\frac{\sigma}{\sigma + \lambda^2}\right|^2 d\|H_\lambda x\|^2.$$
$$\qquad (5.32)$$

The desired convergence now follows, because the integral in (5.32)

tends toward zero with σ, as is easily shown through use of the dom-

inated convergence theorem. |||

Of course, A_σ' and A_σ'' are merely two of the many possible approx-

imations to A^+. Except for the convenience of dealing with positive op-

erators, we might as well have chosen

$$A_\sigma^+ = U_r(T_\sigma)_r^{-1} P_R . \qquad (5.33)$$

with

$$(T_\sigma)_r = \int_0^\infty (\lambda + \sigma) dF_\lambda . \qquad (5.34)$$

When so defined, A_σ^+ is bounded, and converges to A^+ in the same modes

as A_σ' (vide Theorem 5.2); proofs of these claims follows the same

lines of argument already used in Theorem 5.2. Alternatively, one

might have taken the T_σ as

$$T_\sigma = \int_{\lambda \geq \sigma} \lambda \, dE_\lambda , \qquad (5.35)$$

or perhaps based on approximation on the decomposition (5.8) as a

starting point. If we apply (5.34), or use either of the above approx-

imations in the context of the other form of the polar decomposition

(5.8), we again obtain strong convergence to A^+ in general, and con-

vergence in norm whenever $R(A)$ is closed. From the practical view-

point, little is gained by applying the approximations mentioned earlier

in this paragraph; no computational advantage is evident. We do, how-

ever, assure that a solution exists for every $z \in H_2$, for the above

approximations have the effect of closing the range of A.

When $R(A)$ is not closed, the introduction of a new topology can

be used to modify the minimization problem (perhaps in undesirable

fashion) to insure that $R(A)$ is closed in the new topology. One such

technique has been analyzed in detail in Part III, and its validity demon-

strated by Theorem 3.1. Another appeared in disguised form as part

of Theorem 4.1. Here we describe an approximation similar to the

latter, but in more explicit form. Let us take K_σ as the Hilbert space

$K_\sigma = H_1 \times H_2$, with norm specified by

$$\| \{x, y\} \|^2 = \sigma \| x \|^2 + \| y \|^2, \qquad \sigma > 0 . \qquad (5.36)$$

Given the DDC operator $A : H_1 \to H_2$, we define the new operator

$A_1 : H_1 \to K_\sigma$ as follows:

$$A_1 x = \{x, Ax\} . \qquad (5.37)$$

Because A is closed, A_1 has closed range ([H-2], Definition 2.11.2).

If we now let $z_1 = \{\theta, z\}$, there always exists a BAS (Definition 1.1),

for the functional equation $A_1 x = z_1$, and the BAS is furnished by the

PI of A_1, i.e., $x_o = A_1^+ z_1$. In this instance, the BAS x_o is the unique

element attaining the infimum of

$$\| A_1 x - z_1 \|^2 = \| \{x, Ax\} - \{\theta, z\} \|^2 = \sigma \| x \|^2 + \| Ax - z \|^2 . \qquad (5.38)$$

Formally speaking, one hopes that for small σ the infimum of (5.38) is

close to the infimum of $\| Ax - z \|$, and that this is accomplished without

excessive growth of $\| x \|$.

Although the above touches on the question, we have been unable

to solve the following: given $z \in H_2$ and $\alpha \geqslant 0$, is there a (unique?)

$x \in H_1$ of smallest norm satisfying $\| Ax - z \| \leq \delta + \alpha$?[1] If the answer

is affirmative, how is the optimum x_o related to z? A partial conjec-

ture suggests a construction of such x_o on a "piece-by-piece" basis.

---

[1]The answer to a related problem is better understood: If A is a bounded
operator and C a closed convex bounded subset of H_1, there exists $x_o \in C$
satisfying $\inf_{x \in C} \| Ax - z \| = \| Ax_o - z \|$. To prove the assertion, consider
$\{x_n\} \in C$ such that $\| Ax_n - z \|$ tends to the infimum. Now C is weakly
compact, so a subsequence of $\{x_n\}$ converges to x_o, say. This subse-
quence has, in turn, another subsequence whose Cesaro sums $\{w_n\}$
converge (in norm) to x_o. Lastly, $\| Aw_n - z \|$ tends to the infimum,
and $Aw_n \to Ax_o$ in norm. This result is stated as an exercise for
compact operators in [B-1], p. 59.

In the first place, this problem (in fact, any BAS linear problem) can be reduced without loss of generality to an equivalent formulation in which $x \in N(A)^{\perp}$, $z \in N(A)^{\perp}$, $\delta=0$, and A is a positive operator on $N(A)^{\perp}$ (we omit the proof). In the new form, the optimization problem can be viewed in terms of the spectral representation of A, whose associated resolution of the identity now divides $N(A)^{\perp}$ into mutually orthogonal subspaces with projections $(E_{\lambda_{j+1}} - E_{\lambda_j})$. Now if we take $\Delta_j x = \lambda_j^{-1} (E_{\lambda_{j+1}} - E_{\lambda_j}) z$, we diminish $||Ax-z||^2$ by $||(E_{\lambda_{j+1}} - E_{\lambda_j}) z||^2$ while enlarging $||x||^2$ the amount $\lambda_j^{-2} ||(E_{\lambda_{j+1}} - E_{\lambda_j}) z||^2$; these are estimates tending to true values as $|\lambda_{j+1} - \lambda_j| \to 0$. To obtain $||Ax-z|| \leq \alpha$, we take as many increments $\Delta_j x$ as are needed to bring this norm down to the desired level. For each increment $\Delta_j x$, we choose for the index j the one (not already used) corresponding to the largest λ_j, and hence the smallest possible $\Delta_j x$. We remark that the orthogonality and reducing property of the incremental operators on z imply

$$||Ax-z|| = ||E_{\lambda_{n+1}} z|| \quad \text{and} \quad ||x||^2 = \sum \lambda_j^{-2} ||(E_{\lambda_{j+1}} - E_{\lambda_j}) z||^2, \quad (5.39)$$

where $x = \sum \Delta_j x$, and there are a total of n terms with the λ_j arranged in decreasing order. If we formally pass to the limit, we find that

$$x = \int_{\beta}^{\infty} \lambda^{-1} dE_{\lambda} z \quad (5.40)$$

in which β is chosen as the largest number consistent with $||Ax-z|| \leq \alpha$. Since β is dependent on z, (5.40) does not define a linear operator, so we must assume that the solution of the optimization problem (if this is indeed the solution) represents a nonlinear operator.

To gain further insight into the behavior of PI approximations when $P_R z \notin R(A)$, let us study an example involving A'_σ. Let us call

$$x_\sigma = A'_\sigma z , \qquad P_R z \notin R(A) , \qquad (5.41)$$

and evaluate $||Ax_\sigma - z||$ and $||x_\sigma||$. It will become apparent that $||x_\sigma|| \to \infty$ as $||Ax-z|| \to \delta$; as we shall show later in Theorem 5.4, this undesirable behavior of x_σ is inevitable. However, by analyzing the specific case (5.41) we will obtain sharper results on the variation of $||Ax_\sigma - z||$ and $||x_\sigma||$ as a function of σ.

As we saw in the proof of Theorem 5.2, $N(A'_\sigma) = R(A)^\perp$, and not only A'_σ but also AA'_σ is defined on all H_2. Therefore, the first norm becomes

$$|| Ax_\sigma - z ||^2 = || AA'_\sigma z - P_R z ||^2 + || (I-P_R)z ||^2 = \delta_z^2 + \int_0^\infty \left| \frac{\sigma}{\sigma + \lambda^2} \right|^2 d|| F_\lambda P_R z ||^2 ,$$

$$(5.42)$$

in which we have used the fact that $AA'_\sigma = T^2[\sigma I + T^2]^{-1}$, thus giving rise to a spectral representation and corresponding form for norms. From the right hand integral in (5.42), one infers that $||Ax_\sigma - z||$ decreases monotonically with σ, possessing the limit[1]

$$\lim_{\sigma \to 0} ||Ax_\sigma - z|| = \delta_z . \qquad (5.43)$$

---

[1] Equation (5.43) is actually a simple corollary to the strong convergence of A'_σ asserted in Theorem 5.2. Obviously, A''_σ then produces convergence in the sense of (5.43) also.

Thus, we can approach the infimum (1.1) as closely as desired by choosing σ sufficiently small. The convergence does not require $P_R z \in R(A)$; in fact, A'_σ can be applied to $z \in H_2$ which leaves the PI A^+ undefined, and retains the desirable property of minimizing $\|Ax-z\|$ even for such z.

But our analysis remains incomplete without consideration of the norm of x_o. Since $A'_\sigma = U_r T_r (\sigma I + T_r^2)^{-1} P_R$ a direct computation shows

$$\|x_\sigma\|^2 = \int_0^\infty \left| \frac{\lambda}{\sigma + \lambda^2} \right|^2 d \|F_\lambda P_R z\|^2 . \qquad (5.44)$$

Since the integrand is monotonically increasing as σ tends towards zero, the same is true of $\|x_\sigma\|$. In fact, if $P_R z \in R(A)$, $P_R z$ is in the domain of T_r^{-1} [see (5.9) and the argument following], and

$$\|x_\sigma\| \nearrow \|T_r^{-1} P_R z\| . \qquad (5.45)$$

From (5.7), we recognize the right side of (5.45) as the norm of the BAS x_o. Accordingly, as σ tends towards zero

$$\|Ax_\sigma - z\| \searrow \|Ax_o - z\| \quad \text{and} \quad \|x_\sigma\| \nearrow \|x_o\| \qquad (5.46)$$

for all z such that $P_R z \in R(A)$; the norms are strictly decreasing and increasing, respectively.

Whenever $P_R z \notin R(A)$, the right side of (5.44) tends toward infinity (see [D-2], Theorem XII.2.6) because then $P_R z \notin D(T_r^{-1})$ from (5.9). We may therefore conclude that

$$\|x_\sigma\| \nearrow \infty \quad \text{for} \quad P_R z \notin R(A) . \qquad (5.47)$$

The rate at which this norm tends to infinity can be bounded from above by majorizing the right side of (5.44). Indeed, we always have

$$\| x_\sigma \| < \sigma^{-1} , \tag{5.48}$$

whether or not $P_R z$ belongs to the range of A.

That $\| Ax_\sigma - z \| \to \delta_z$ is accompanied by $\| x_\sigma \| \to \infty$ is not merely a coincidence attributable to a poor choice of operator or sequence $\{ x_\sigma \}$. Rather, the comparative behavior of the two norms is immutable, as is indicated by

<u>Theorem 5.4</u>: Suppose $P_R z \notin R(A)$, and $\{ x_n \} \in D(A)$ such that

$$\| Ax_n - z \| \to \delta_z . \tag{5.49}$$

Then

$$\| x_n \| \to \infty . \tag{5.50}$$

Proof: By an argument identical with that leading to the left hand equality in (5.42), we can replace (5.49) by

$$\| Ax_n - y \| \to 0 \tag{5.51}$$

if we take $y = P_R z$. Let us then suppose $\{ x_n \}$ bounded and reason by contradiction. If $\{ x_n \}$ is bounded, it is weakly compact, and we may assume without loss of generality that the entire sequence converges weakly, say

$$x_n \xrightarrow{w} x_o . \tag{5.52}$$

Then $\{ x_n \}$ has another subsequence—which we again denote by $\{ x_n \}$ for brevity in notation—such that ([R-2], Section 38)

$$w_m = m^{-1} \sum_1^m x_j \tag{5.53}$$

and

$$w_m \longrightarrow x_o \quad \text{in norm.} \tag{5.54}$$

But also, the triangle inequality yields

$$\|Aw_m - y\| \leq m^{-1} \sum_1^m \|Ax_j - y\| \to 0 , \qquad (5.55)$$

with the right hand convergence following from (5.51). But A is a

closed operator, and $\|Aw_m - y\| \to 0$, together with (5.54), implies

$x_o \in D(A)$ and $Ax_o = y$. This contradicts the original hypothesis

$y \notin R(A)$. |||

APPENDIX A

SOME PROPERTIES OF OPERATORS IN HILBERT SPACES

In the classical literature on linear unbounded operators in Hilbert spaces [D-2] [R-2] [A-1], one is generally confined to densely defined closed (DDC) operators; operators which are not DDC fail to yield significant useful results, and are therefore seldom considered. But even the DDC assumption leads to a much more complex structure than is encountered for bounded operators (which can be trivially extended to DDC operators). For instance, if one of A and B is bounded and the other DDC, AB need not be DDC. Thus, applications to pseudoinverses requires some knowledge on the characteristics of certain compositions of DDC operators and their restrictions to subspaces. Consequently, we offer here a series of technical lemmas, to which we shall refer in the body of work from time to time. Although each lemma is quite easy, and perhaps not interesting in itself, we have not found these results conveniently accessible in the standard literature.

Our notation is consistent throughout the paper. A is a linear DDC operator from a Hilbert space H_1 to another, H_2. The operator A has domain $D(A)$, and is endowed with range $R(A)$ and nullspace $N(A)$; its restriction to $N(A)^\perp$ will be called A_r. Projections on $N(A)$ and $N(A)^\perp$ are denoted by P_N and P_M, respectively. Analogously, $B:H_2 \to H_1$ will have a restriction B_r to $\overline{R(A)} = N(A*)^\perp$, and the corresponding projection on this substance is P_R.

Lemma A. 1: $x \in D(A)$ iff $P_M x \in D(A)$.

Proof: Consider the orthogonal decomposition

$$x = P_M x + x_1, \qquad\qquad (A. 1)$$

where $x_1 \in N(A) \subset D(A)$. It is clear from the linearity of A that

$P_M x \in D(A)$ implies $x \in D(A)$; conversely, if $x \in D(A)$, so is

$P_M x = x - x_1$. |||

Lemma A. 2: $R(A_r) = R(A)$.

Proof: Since $R(A_r) \subset R(A)$ is obvious, we need only prove the set inclu-

sion in the opposite direction. To this end, take any $z \in R(A)$, so that

there exists an $x \in D(A)$ such that $Ax = z$. The decomposition (A. 1)

then yields $A_r(P_M x) = A(P_M x) = Ax = z$ because $Ax_1 = \theta$. |||

The next lemma applies to __any__ restriction of A, and we take A_r

to be such. Of course, our interest here is the specialization of the

result below to $K = N(A)^{\perp}$.

Lemma A. 3:   Any linear operator $A_r \subset A$ has a closed extension $\overline{A_r}$.

If, in particular, $D(A_r) = D(A) \cap K$, where K is a subspace, A_r is a

closed operator.

Proof: The graph G(A) of A is a Hilbert space under the usual norm,

and $G(A_r) \subset G(A)$. Clearly, A_r and hence $G(A_r)$ can be extended to a

linear manifold dense in some subspace of G(A). We may therefore

suppose $G(A_r)$ already extended in this fashion.   Since $\{\theta, y\} \notin \overline{G(A_r)}$

for any $y \neq \theta$, $\overline{A_r}$ can be defined uniquely in terms of $\overline{G(A_r)}$ ([H-2],

Section 2. 11), and $\overline{A_r}$ is a closed operator because its graph is closed.

Assume now K is a subspace; we show $G(A_r)$ is then complete in

G(A). In fact, if $\{\{x_n, A_r x_n\}\}$ is a Cauchy sequence in $G(A_r)$, $\{x_n\} \in D(A_r)$,

and since K is closed, $x_n \to x_0 \in K$. Moreover, $\{x_n, A_r x_n\} = \{x_n, Ax_n\}$ $\to \{x_0, Ax_0\} \in G(A)$. Thus $x_0 \in G(A) \cap K = D(A_r)$, which means $\{x_0, Ax_0\} \in G(A_r)$. This completes the proof. $|||$

We return to the notation of A_r as the restriction of A to the sub-space $N(A)^{\perp}$. By virtue of the lemmas above already proven we see A_r is a closed linear operator from the Hilbert space $N(A)^{\perp} \subset H_1$ into the Hilbert space $\overline{R(A)} \subset H_2$. Its domain $D(A_r)$ consists precisely of the set $D(A_r) = \{u : u = P_M x, \; x \in D(A)\}$, and its range coincides with that of A. That A_r is in fact a DDC operator now follows from

Lemma A. 4: $A_r : N(A)^{\perp} \to \overline{R(A)}$ is densely defined.

Proof: Since A is densely defined, there is for each $x \in N(A)^{\perp}$ a sequence $\{x_n\} \in D(A)$ such that $x_n \to x$. But $|| x - P_M x_n || \leq || x - x_n || \to 0$, and $\{P_M x_n\} \in D(A_r)$. $|||$

It sometimes becomes necessary to deduce the properties of an operator from those of its restriction, or to extend an operator from $N(A)^{\perp}$ to H_1 [$\overline{R(A)}$ to H_2]. That the extended operator is well be-haved is assured by

Lemma A. 5: The linear operator $A : H_1 \to H_2$ is DDC if

    (a) $N(A)$ is a closed set

    (A. 2)

and   (b) its restriction A_r is DDC on $N(A)^{\perp}$.

Proof: Since A is linear, $N(A)$ is a subspace. Let A_q be the restric-tion of A to $N(A)$, and take $G(A_q)$ to be the corresponding graph, con-sisting of elements $\{x, \theta\}$ with $x \in N(A)$. Then $G(A_q)$ is a subspace in $H_1 \times H_2$, and is orthogonal to $G(A_r)$. We observe further [cf. Lemma A. 1 and the linearity of A] that $G(A) = G(A_r) \oplus G(A_q)$, which exhibits

G(A) to be closed. To demonstrate that A is densely defined, first note

that $D(A) = D(A_r) \oplus N(A)$. If now $x \in H_1$, there exists $\{x_{n1}\} \in D(A_r)$

such that $x_{n1} \to P_M x$, whence $x_{n1} + P_N x = x_n \to x$. As $\{x_n\} \in D(A)$, this

shows $D(A)$ dense in H_1. $|||$

We next exhibit some results on inverses of operators related to

A. If A is invertible with range dense in H_2, there is little content to

the theory of the pseudoinverse. We therefore do not suppose A has

an inverse, but rather subsume it under the more general results that

follow. The first of these is

Lemma A.6: A_r has a DDC inverse A_r^{-1}.

Proof: The inverse exists if A_r is one-to-one. For $x_1, x_2 \in N(A)^{\perp}$

$A_r x_1 = A_r x_2$ is equivalent to $Ax_1 = Ax_2$, which shows $x_1 - x_2 \in N(A)$.

But $x_1 - x_2 \in N(A)^{\perp}$ also, which means $x_1 = x_2$; thus, $A_r : N(A)^{\perp} \to \overline{R(A)}$

is one-to-one. That A_r^{-1} is closed follows from the fact that A_r is

closed (cf. Lemma A.3 and [H-2], Theorem 2.11.15). Finally,

$D(A_r^{-1}) = R(A_r)$, and the latter is dense in $\overline{R(A)}$ from Lemma A.2. $|||$

Lemma A.7: A_r^* has a DDC inverse, and

$$(A_r^*)^{-1} = (A_r^{-1})^* . \tag{A.3}$$

Proof: A_r^* and $(A_r^{-1})^*$ are themselves DDC, whereupon the desired

conclusions follow from a well known argument based on $G(A_r)$; see

[R-2], Section 117 for details. $|||$

Lemma A.8: The following conditions are equivalent:

(a) A_r^{-1} is bounded.

(b) $(A_r^*)^{-1}$ is bounded.

(c) $R(A)$ is a closed set.

(d) $A_r A_r^{-1}$ is a closed operator.

(e) There exists $m > 0$ such that $m||x|| \leq ||A_r x||$ for all $x \in D(A_r)$.

(f) There exists $k > 0$ such that $k||y|| \leq ||A_r^* y||$ for all $y \in D(A_r^*)$.

Proof: Since $D(A_r^{-1}) = R(A)$ and A_r^{-1} is a closed operator (cf. Lemma A.6), condition (c) implies (a) by the closed graph theorem [R-2], Section 117). Conversely, the domain of a closed bounded operator is a closed set, whence (c) follows from (a) and Lemma A.2.

To prove (b) from (a), consider (A.3) and the fact that (a) implies $(A_r^{-1})^*$ bounded. On the other hand, (b) requires $(A_r^{-1})^*$ bounded, and (a) follows.

Respecting (e), observe $D(A_r) = R(A_r^{-1})$, so we may put $x = A_r^{-1} y$, with y uniquely defined and ranging over $D(A_r^{-1})$ as x ranges over $D(A_r)$. This means $||A_r^{-1} y|| \leq m^{-1}||y||$ for all $y \in D(A_r^{-1})$, i.e. $||A_r^{-1}|| \leq m^{-1}$. For the converse, one proves (a) from (e) by retracing the above steps in reverse order. The equivalence of (b) and (f) is similarly demonstrated.

We turn our attention to the relation between (a) and (d). $A_r A_r^{-1}$: $\overline{R(A)} \to \overline{R(A)}$ is the restriction in this space of the identity operator to its domain $D(A_r A_r^{-1}) = R(A)$. Thus $A_r A_r^{-1}$ is bounded and densely defined. Consequently, $A_r A_r^{-1}$ is a closed operator iff its domain $R(A)$ is a closed set. $|||$

It is tempting to assert that (b) is equivalent to yet another condition, namely

(g) $R(A^*)$ is a closed set. (A. 4)

However, the proof given in the first paragraph (of the proof of the lemma) fails because the connection between $(A^*)_r$ and A_r^* is not clear. We recall here that by B_r for $B: H_2 \to H_1$ we mean the restriction of B to $\overline{R(A)}$.

__Lemma A. 9:__ $(A^*)_r = A_r^*$ (A. 5)

Proof: For any $x \in D(A)$, $y \in D(A^*)$ we have

$$(Ax, y) = (x, A^*y) = (x_1, A^*y) = (x_1, A^*y_1) = (x_1, (A^*)_r y_1) \qquad (A. 6)$$

where $x_1 = P_M x$ and $y_1 = P_R y$. At the same time

$$(Ax, y) = (Ax_1, y_1) = (A_r x_1, y_1) . \qquad (A. 7)$$

The right sides of (A. 6) and (A. 7) are valid for all $x_1 \in D(A_r)$ from Lemma A. 1. Therefore, $A_r^* \supset (A^*)_r$, the latter being defined at least on $D(A^*) \cap N(A^*)^{\perp}$. To demonstrate equality, consider any $y \in D(A_r^*)$. For such y and all $x_1 \in D(A_r)$

$$(A_r x_1, y) = (x_1, A_r^* y) = (x_1, y^*), \qquad (A. 8)$$

in which $y^* \in R(A_r^*)$, and $R(A_r^*) = \overline{D(A_r)}$ because [see Lemma A. 7] $(A_r^*)^{-1}$ is densely defined. But any $x \in D(A_r)$ has the orthogonal decomposition $x = x_1 + x_2$ with $x_1 \in D(A_r)$ and $x_2 \in N(A)$ (cf. again Lemma A. 1). Hence in (A. 8) $A_r x_1 = Ax_1 = Ax$, while $(x_1, y^*) = (x, y^*)$ because y^* is orthogonal to $N(A)$. In other words, $(Ax, y) = (x, y^*)$, or $y \in D(A^*)$. The proof is completed by noting that the inclusion $D(A_r^*) \subset D(A^*)$ just obtained yields

$$D(A^*_r) = D(A^*_r) \cap \overline{R(A)} \subset D(A^*) \cap \overline{R(A)} = D[(A^*)_r]. \ |||$$

The validity of Lemma A.9 permits us to use the notation A^*_r to replace both $(A^*)_r$ and A^*_r. This lemma may also be applied to strengthen the results of Lemma A.8, viz.

<u>Corollary A.1</u>: Lemma A.8 remains true if condition (g) of (A.4) is added; i.e., conditions (a) through (g) are equivalent.

Proof: $R(A^*) = R[(A^*)_r]$, an assertion whose proof we defer for the moment. Since $R(A^*_r)$ is closed or not according as $(A^*_r)^{-1}$ is bounded or unbounded, (b) and (g) hold simultaneously. By Lemma A.8, (b) is equivalent to the other properties (a) to (f), so the proof is finished. $|||$

<u>Lemma A.10</u>: (Duality of A_r and A^*_r) Lemmas A.1, A.2, A.4, A.5 and A.8 continue to hold if symbols are consistently interchanged as follows:

A with A^*, A_r with A^*_r. P_M with P_R, and H_1 with H_2. (A.9)

Proof: $A^*: H_2 \to H_1$ is DDC, and A^*_r is its restriction to $\overline{R(A)}$, where $\overline{R(A)} = N(A^*)^\perp$. The arguments of Lemmas A.1, A.2, A.4 and A.5 therefore lead to the desired results in terms of the sumbol substitution (A.9). No proof is required for Lemma A.8, in view of its inherent symmetry with the addition of Corollary A.1. $|||$

Our next objects of study are the restrictions of A^*A and AA^*. As we have pointed out earlier, compositions of DDC operators need be neither closed nor densely defined. However, it is known that A^*A and AA^* are not only DDC, but even self-adjoint positive operators

([R-2], Section 119). The restriction $(A^*A)_r : N(A)^{\perp} \to N(A)^{\perp}$ is then

closed by Lemma A. 3, and similarly for $(AA^*)_r : \overline{R(A)} \to \overline{R(A)}$. It is less

obvious that $(A^*A)_r$ and $(AA^*)_r$ are densely defined and self-adjoint.

We now prove these facts, together with other properties of $(A^*A)_r$

and $(AA^*)_r$ useful elsewhere in this work.

Lemma A. 11: $N(A) = N(A^*A);\ N(A^*) = N(AA^*).$ (A. 10)

Proof: We prove only the first set equality. Clearly, $N(A^*A) \supset N(A)$.

Consider then x $\in N(A^*A)$; we compute

$$0 = (A^*Ax, x) = ||Ax||^2,$$ (A. 11)

which demonstrates that x $\in N(A)$. |||

Lemma A. 11 above enables us to write $(A^*A)_r : N(A^*A)^{\perp} \to N(A^*A)^{\perp}$,

showing through use of identical arguments that

Corollary A. 2: Lemmas A. 1, A. 2, A. 4 and A. 6 remains valid if A is

replaced by A^*A and A_r by $(A^*A)_r$.

In particular, $(A^*A)_r$ is densely defined and possesses a DDC inverse.

Corollary A. 3: $\overline{R[(A^*A)_r]} = \overline{R(A^*A)} = \overline{R(A^*)};\ \overline{R[(AA^*)_r]} = \overline{R(AA^*)} = \overline{R(A)}.$

(A. 12)

Proof: Lemma A. 2 yields $\overline{R[(A^*A)_r]} = \overline{R(A^*A)}$, while $R(A^*A)^{\perp} = R(A^*)^{\perp}$

follows from the standard result $N(A) = R(A^*)^{\perp}$ and (A. 10). The second

set of equalities is similarly obtained. |||

To verify the self-adjointness of $(A^*A)_r$ [as well as $(AA^*)_r$], and

to provide a useful alternative form, we state first

Lemma A. 12: $(A^*A)_r = A_r^* A_r$ and $(AA^*)_r = A_r A_r^*.$ (A. 13)

Proof: The left side of the first equality satisfies

$(A^*A)_r x = A^*Ax$ for any $x \in D(A^*A) \cap N(A)^\perp$. For such x,

$$A^*_r A_r x = A^*_r(Ax) = A^*Ax; \qquad (A.14)$$

the latter equality is true because $Ax \in \overline{R(A)}$, and A^*_r is the restriction

of A^* to $\overline{R(A)}$. We have thus shown $(A^*A)_r \subset A^*_r A_r$. To complete the

proof, we demonstrate that $D(A^*_r A_r) \subset D[(A^*A)_r]$. In fact, $A^*_r A_r x$ is

defined only if $x \in D(A) \cap N(A)^\perp$ and $Ax \in D(A^*) \cap \overline{R(A)}$. But

$Ax \in \overline{R(A)}$ in any case, so the second condition becomes merely

$Ax \in D(A^*)$. That is, $x \in D(A^*A) \cap N(A)^\perp = D[(A^*A)_r]$. $|||$

Remark: The proof makes implicit use of Lemma A.9, since we de-

duce the domain of A^*_r as that of $(A^*)_r$, whereas the applications of

(A.13) interpret A^*_r as A_r^*.

Lemma A.13: $(A^*A)_r$ and $(AA^*)_r$ are self-adjoint operators on $N(A)^\perp$

and $R(A)$, respectively.

Proof: Although it was shown earlier that $(A^*A)_r$ is DDC, this follows

again from Lemma A.12 and the fact that $A_r^*A_r:N(A)^\perp \to N(A)^\perp$ is

DDC. Indeed, $A_r^*A_r$ is self-adjoint on this space ([R-2], Section 119),

so the same is true for $(A^*A)_r$ by (A.13). As usual, the assertion

regarding $(AA^*)_r$ is analogously proven. $|||$

Corollary A.2 already claims that $(A^*A)_r$ and $(AA^*)_r$ possess

DDC inverses. We now repeat and sharpen this result via

Lemma A.14: $(A^*A)_r: N(A)^\perp \to N(A)^\perp$ and $(AA^*)_r: \overline{R(A)} \to \overline{R(A)}$

have self-adjoint inverses given respectively by

$$(A^*A)_r^{-1} = A_r^{-1}A_r^{*-1} \quad \text{and} \quad (AA^*)_r^{-1} = A_r^{*-1}A_r^{-1}. \qquad (A.15)$$

Proof: We recall that $(A^*A)_r = A_r^*A_r$ by (A. 13). Then, since A_r and

A_r^* have inverses from Lemmas A. 6 and A. 7, respectively, $(A^*A)_r$

has the inverse stated in (A. 15) (see [R-2], Section 114h). The inverse

of a self-adjoint operator is likewise self-adjoint ([R-2], Section 119),

and so the self-adjointness of $(A^*A)_r^{-1}$ is a consequence of Lemma A. 13. |||

Finally, we give the conditions under which the inverses of $(A^*A)_r$

and $(AA^*)_r$ are bounded, and their ranges closed.

Lemma A. 15: Suppose

(a) any one of the ranges of A, A^*, (A^*A) or (AA^*) is closed

or (b) any of the inverses A_r^{-1}, A_r^{*-1}, $(A^*A)_r^{-1}$, or $(AA^*)_r^{-1}$ bounded.

Then all the ranges in (a) are closed, and all inverses (b) bounded.

Proof: By virtue of Lemma A. 2, and its application to adjoint and com-

posite operators via Lemma A. 10 and Corollary A. 2, we consider the

ranges of the restrictions of the operators appearing in (a). Then the

arguments of Lemma A. 8 demonstrate the desired result for A and A^*,

and also yield the relation between range and inverse for (A^*A) and (AA^*).

If now A_r^{-1} and A_r^{*-1} are bounded, formula (A. 15) verifies the

boundedness of $(A^*A)_r^{-1}$ and $(AA^*)_r^{-1}$. Conversely, suppose $(A^*A)_r^{-1}$

bounded. Then $R(A^*A)$ is a closed set, and we have from (A. 12) and

$R(A^*A) \subset R(A^*)$

$$R(A^*) = R(A^*A) \subset R(A^*) \subset \overline{R(A^*)}; \qquad (A. 16)$$

this means $R(A^*) = R(A^*A)$, or $R(A^*)$ is closed. The latter enables us

to show that all the other conditions of (a) and (b) are met. Likewise,

$R(AA^*)$ closed or $(AA^*)_r^{-1}$ leads to $R(A)$ closed, and hence the truth of

each condition (a) and (b). |||

In the course of the above proof, we have additionally deduced

<u>Corollary A. 4</u>:   If any of the conditions (a) or (b) of Lemma A. 15 are

satisfied,

$$R(A^*A) = R(A^*) \quad \text{and} \quad R(AA^*) = R(A), \tag{A.17}$$

each of these ranges being a closed set. |||

It is well known that if A_1 and A_2 are DDC, $(A_1A_2)^* \supset A_2^*A_1^*$, but

this sheds little light on the closed extension of the right side. However,

if either A_1 or A_2 are bounded, at least partial results can be secured

([R-2], Section 115). Our interest lies in taking A_2 bounded, since

this occurs in Gauss-Markov estimation. More specifically, we have

<u>Lemma A. 16</u>: Suppose

    a. A_1 is densely defined and closable, with closure $\overline{A}_1$.

    b. A_2 is bounded and everywhere defined.

    c. $\overline{A}_1A_2$ is densely defined.

Then the closed extension of $A_2^*A_1^*$ is $(\overline{A}_1A_2)^*$.

Proof: Since A_2^* is bounded and A_1 is densely defined, $A_2^*A_1^*$ is

densely defined and hence has an adjoint; in fact, A_2^* being bounded

implies (see again [R-2]., Section 115)

$$(A_2^*A_1^*)^* = \overline{A}_1A_2. \tag{A.18}$$

Taking adjoints once more then yields the conclusion, viz.

$$\overline{A_2^*A_1^*} = (A_2^*A_1^*)^{**} = (\overline{A}_1A_2)^*. \, ||| \tag{A.19}$$

<u>Remark</u>:   It does not follow that $A_2^*A_1^* = (\overline{A}_1A_2)^*$, because the left side

need not be a closed operator. This can be demonstrated via a counter-

example in which A_1 and A_2 are each self-adjoint, and A_2 has finite

dimensional range.

APPENDIX B

HILBERT-SPACE-VALUED RANDOM VARIABLES

In discussing statistical estimation in Hilbert space we deal with random quantities taking values in a Hilbert space. There is an extensive literature concerning Hilbert-space-valued random variables, or what is essentially the same thing, probability measures in a Hilbert space (see [P-1]). However, what we need is relatively simple and can be presented briefly in a form that directly fits our requirements.

Let $(\Omega, \mathcal{C}, P)$ be a complete probability space, where $\mathcal{C}$ is a σ-algebra of subsets of Ω and P is a probability measure on $\mathcal{C}$. Let $\{w_i\}$, $i=1,2,\ldots$, be a sequence of (real or complex-valued) random variables on $(\Omega, \mathcal{C})$ such that

$$\sum^{\infty} E|w_i|^2 = \sum^{\infty} \int_{\Omega} |w_i(\omega)|^2 dP(\omega) < \infty .\tag{B.1}$$

We note that, by the Fubini theorem, the inequality (B.1) implies

$$E\left(\sum^{\infty} |w_i|^2\right) = \int_{\Omega} \sum^{\infty} |w_i(\omega)|^2 dP(\omega)$$

$$= \sum^{\infty} E|w_i|^2 < \infty .\tag{B.2}$$

Now let $\{\phi_n\}$, $n=1,2,\ldots$, be a complete orthonormal system (c.o.n.s.) in H_2 and define $w(\omega)$ by

$$w(\omega) = \sum^{\infty} w_i(\omega)\phi_i, \quad \omega \in \Lambda$$

$$= 0, \quad \omega \in \Lambda^c \tag{B.3}$$

where $\Lambda \subset \Omega$ is the set of all ω for which $\sum^{\infty} |w_i(\omega)|^2 < \infty$. By (B.2)

and the assumption of completeness of P, $\Lambda^c \in \mathcal{C}$, and $P(\Lambda^c) = 0$. Thus, for all ω, $w(\omega)$ is an element of H_2, and for a.e. ω

$$||w(\omega)||^2 = \sum_{i}^{\infty} ||w_i(\omega)||^2. \tag{B.4}$$

We can now define $\tilde{w}_n(\omega) = (w(\omega), \phi_n)$ for all ω. The $\tilde{w}_n$ are measurable $(\mathcal{C})$ and are equal to the w_n except on a fixed set of probability zero. It should cause no confusion if we drop the tilde and henceforth simply denote $(w(\omega), \phi_n)$ as w_n. Then (B.4) holds for all ω. Since the (w, ϕ_n) are all measurable it is easy to establish that w is weakly measurable, that is that (w, y) is a measurable scalar-valued function for each $y \in H_2$. Then, since H_2 is separable, $w(\cdot)$ is also strongly measurable (see [H-2] Section 3.5 for both definition and proof). We call w a Hilbert-space-valued random variable.

The condition (B.1) is sufficient to establish both a mathematical expectation of w and a covariance operator. Since, by (B.2) and (B.4),

$$E||w||^2 = \int_{\Omega} ||w(\omega)||^2 \, dP(\omega) < \infty \tag{B.5}$$

then,

$$\int_{\Omega} ||w(\omega)|| \, dP(\omega) < \infty. \tag{B.6}$$

The strong measurability of w and (B.6) imply that $w(\omega)$ is Bochner-integrable $(\Omega, \mathcal{C}, P)$ (see [H-2], Section 3.7). We define

$$Ew = \int_{\Omega} w(\omega) dP(\omega) \tag{B.7}$$

where the integral is a Bochner integral. An application of a basic theorem of Hille ([H-2], Theorem 3.7.12) gives

$$(y, Ew) = \int_{\Omega} (y, w) dP.$$

Hence, if the w_i have mean zero it follows that $Ew = \phi$.

Also, we have from (B. 4) that

$$E|(y,w)\overline{(z,w)}| \leq ||y|| \cdot ||z|| \, E||w||^2 < \infty . \tag{B. 8}$$

Since $\psi(y,z) = E\{(y,w)\overline{(z,w)}\}$ is a Hermitian symmetric conjugate

bilinear form, which by (B. 8) is bounded, there exists a bounded sym-

metric operator K defined on all of H_2 such that

$$(Ky, z) = \psi(y, z). \tag{B. 9}$$

K is obviously nonnegative definite; it is called the covariance opera-

tor of w. The assumption (B. 1) further implies that K is nuclear, that

is that it is compact and has finite trace.[1] One can verify directly from

(B. 1) that the bounded operator K is Hilbert-Schmidt, and hence com-

pact, and then, using the compactness, verify that it has finite trace.

---

[1] We refer the reader to [G-1], pp. 26-47. A nonnegative self-adjoint

compact operator T has an o.n.s. of eigenvectors $\{\eta_n\}$ that span $\overline{R(T)}$,

and nonnegative eigenvalues λ_n; $T\eta_n = \lambda_n$. By definition, T is

nuclear if

$$\sum_{n=1}^{\infty} (T\eta_n, \eta_n) = \sum_{n=1}^{\infty} \lambda_n < \infty.$$

If T is nuclear,

$$\sum_{n=1}^{\infty} (T\phi_n, \phi_n) = \sum_{n=1}^{\infty} (T\eta_n, \eta_n) = Tr(T),$$

where $\{\phi_n\}$ is any c.o.n.s.

However, a shorter proof is as follows. Since K is self-adjoint and nonnegative, it has a nonnegative square root $K^{1/2}$. Then

$$\sum_{\infty} \| K^{1/2} \phi_n \|^2 = \sum_{\infty} (K\phi_n, \phi_n) = \sum_{\infty} E | (\phi_n, w) |^2$$

$$= \sum_{\infty} E | w_n |^2 < \infty \qquad (B.10)$$

The convergence of $\sum \| K^{1/2} \phi_n \|^2$ ensures that $K^{1/2}$ is Hilbert-Schmidt (cf. [G-1], p. 34). Then $K = K^{1/2} \cdot K^{1/2}$ is nuclear (cf. [G-1], p. 39).

Finally, we observe that K is strictly positive definite if and only if $(K\phi_n, \phi_n) > 0$ for all n, that is, if and only if $E | w_n |^2 > 0$ for all n.

The preceding remarks may be summarized as a theorem:

Theorem B. 1: Let $(\Omega, \mathcal{Q}, P)$ be a complete probability space and $\{w_n\}$, $n = 1, 2, \ldots$, a sequence of scalar-valued random variables satisfying (B. 1). Let $\{\phi_n\}$ be a c. o. n. s. in the separable Hilbert space H. Then w as defined by (B. 3) is an H-valued random variable, i.e., a strongly measurable H-valued function on $(\Omega, \mathcal{Q})$. The expectation of w as defined by the Bochner integral in (B. 7) exists, and a unique bounded self-adjoint, nonnegative-definite covariance operator K exists that satisfies

$$(Ky, z) = \overline{E(y, w)(z, w)}, \qquad y, z \in H.$$

Further, K is nuclear.

K is strictly positive definite if and only if $E | w_n |^2 > 0$, for all n, and $Ew = 0$ if and only if $Ew_n = 0$ for all n.

Remark: The nuclearity of K is essential in the following sense. Suppose v is an H-valued random variable satisfying $E||v||^2 < \infty$. Then a covariance operator K for v exists and is nuclear. In fact, the existence of K follows exactly as above, and K is bounded, self-adjoint and nonnegative definite. Then for any c. o. n. s. $\{\phi_n\}$ the first two equalities in (B. 10) hold. But

$$\sum_{n}^{\infty} E|(\phi_n, v)|^2 = E||v||^2,$$

which is finite. The rest of the argument follows as before.

One standard situation is for the observations to be functions in an L_2 space. This leads us to discuss stochastic processes whose sample functions belong almost surely to L_2. Let $w(t,\omega)$ be a measurable, separable, real or complex-valued stochastic process with t belonging to the parameter set $T \subset R^1$, and with probability space $(\Omega, \mathcal{A}, P)$ (c. f. [D-3], Chapter 2). T can be an interval, or all of R^1, or indeed any measurable subset of R^1 with positive measure. $w(t, \cdot)$ is a random variable on $(\Omega, \mathcal{A}, P)$. As is conventional, we usually suppress the probability variable and write $w(t)$ instead of $w(t, \omega)$. Let $w(t)$ satisfy

$$\int_T E|w(t)|^2 dt < \infty. \tag{B. 11}$$

Then $R(t, s) \overset{d}{=} E\, w(t)\overline{w(s)}$ exists for a.e. t and a.e. s, and

$$\int_T |w(t, \omega)|^2\, dt < \infty$$

for all ω in some Λ, $P(\Lambda) = 1$. We put $R(t, s) = 0$ for those values of

t, s for which $Ew(t)\overline{w(s)}$ is not defined. We define $w(\omega) \in L_2(T)$ to be

the Lebesgue equivalence class of functions $w(\cdot, \omega)$ for each $\omega \in \Lambda$,

and to be the zero element otherwise. Then we have

Theorem B. 2: $w(\omega)$ as just defined is an H-valued random variable

$(H = L_2(T))$. It has a nuclear covariance operator K given by

$$[Ky](t) = \int_T R(t, s)y(s)ds \qquad (B. 12)$$

for $y \in L_2(T)$. The expected value of $w(\omega)$ exists and is characterized

by

$$(Ew, y) = E(w, y) = \int_T E \, w(t) \, \overline{y(t)} \, dt \qquad (B. 13)$$

for all $y \in L_2(T)$.

Proof: The assertion that w is an H-valued random variable means

that it is strongly measurable. Since $L_2(T)$ is separable this is equiv-

alent to weak measurability. We have for any $y \in L_2(T)$,

$$(w(\omega), y) = \int_T w(t, \omega)\overline{y(t)}dt, \quad \omega \in \Lambda \qquad (B. 14)$$

$$= 0, \quad \omega \in \Lambda^c$$

The integral in (B. 14) is a measurable function of ω(c. f. [D-3], Theorem

2. 7) hence (w, y) is measurable. Thus w is an $L_2(T)$-valued random

variable. From (B. 11) we have $E||w||^2 < \infty$. Hence, as in the pre-

ceding theorem and the remark following, w has a nuclear covariance

operator K.

$$(Ky, z) = E[(y, w)\overline{(z, w)}]$$

$$= E\left\{ \int_T y(t)\overline{w(t)}dt \int_T \overline{z(s)}w(s)ds \right\}$$

$$= \int_T \int_T R(s, t)\overline{z(s)}\, y(t)dt \qquad (B.15)$$

where the interchange of integrals is justified by the Fubini theorem after the following calculation:

$$\int_T \int_T E|w(s)\overline{w(t)}|\; |y(t)|\; |z(s)|\; dtds$$

$$\leq \int_T \int_T [R(s, s)]^{1/2}[R(t, t)]^{1/2}\; |y(t)|\; |z(s)|\; dtds$$

$$\leq \left(\int_T R(s, s)ds \right) ||y||\; ||z|| < \infty$$

by the assumption (B.11). Since (B.15) holds for all $y, z \in L_2(T)$, (B.12) is proved. The expected value of w is given by the Bochner **integral**

$$Ew = \int_\Omega w(\omega)dP(\omega) \qquad (B.16)$$

which exists because

$$\int_\Omega ||w(\omega)||\; dP(\omega) < \infty, \qquad (B.17)$$

which in turn follows from the square integrability guaranteed by (B.11). (B.13) follows from (B.16) and a theorem used previously ([H-2], Theorem 3.7.12). |||

Now let C be a bounded linear transformation from all of H_2 into H_1, and let w be an H_2-valued random variable satisfying $E||w||^2 < \infty$.

Then Cw is an H_1-valued random variable. Clearly Cw(ω) is defined

for all ω. Also Cw is weakly (and hence strongly) measurable since

$(Cw, x) = (w, C*x)$, $x \in H_1$, and w is weakly measurable. Finally,

$$\int_\Omega ||Cw(\omega)||^2 dP(\omega) \leq ||C||^2 \int_\Omega ||w(\omega)||^2 dP(\omega) < \infty .$$

Thus $E[Cw]$ is defined, and a covariance operator K_2 for Cw is defined.

We have, furthermore, again by the Hille theorem,

$$E[Cw] = \int_\Omega Cw(\omega)dP(\omega) = C \int_\Omega w(\omega)dP(\omega) = C(Ew)$$

and

$$(K_2, u, x) = E[(u, Cw)\overline{(x, Cw)}]$$
$$= E[(C^*u, w)\overline{(C^*x, w)}] = (KC^*u, C^*x),$$

from which $K_2 = CKC^*$.

REFERENCES

[A-1]   N. I. Akhiezer & I. M. Glazman, <u>Theory of Linear Operators in Hilbert Space</u>, vol. I and II (tr. by M. Nestell), Frederick Ungar Publishing Co., New York, 1961 and 1963.

[A-2]   A. Albert, <u>Regression and the Moore-Penrose Pseudoinverse,</u> Academic Press, New York, 1972.

[A-3]   M. Athans & P. L. Falb, <u>Optimal Control</u>, McGraw-Hill, New York, 1966.

[B-1]   A. V. Balakrishnan, <u>Introduction to Optimization Theory in a Hilbert Space</u>, Lecture Notes in Operations Research and Mathematical Systems, vol. 42, Springer-Verlag, New York, 1970.

[B-2]   A. V. Balakrishnan, "Determination of nonlinear systems from input-output data, " <u>Proc. of the Princeton University Conference on Identification Problems in Communcation and Control Systems</u> (1963), 31-49.

[B-3]   F. J. Beutler, "The operator theory of the pseudo-inverse I. Bounded operators, " <u>J. Math. Anal. and Appl.</u>, 10 (1965), 451-470.

[B-4]   F. J. Beutler, "The operator theory of the pseudo-inverse II. Unbounded operators with arbitrary range, " <u>J. Math. Anal. and Appl.</u>, 10 (1965), 472-493.

[B-5]   R. Bouldin, "The pseudo-inverse of a product, " <u>SIAM J. Appl. Math.</u> 24 (1973), 489-495.

[B-6]   R. W. Brockett, <u>Finite Dimensional Linear Systems</u>, Wiley, New York, 1970.

[C-1]   C. T. Chen, <u>Introduction to Linear System Theory</u>, Holt, Rinehart & Winston, New York, 1970.

[D-1]   C. Desoer and B. Whalen, "A note on pseudoinverses, " <u>J. SIAM</u>, 11 (1963), 442-447.

[D-2]   N. Dunford & J. T. Schwartz, <u>Linear Operators</u>, vol. I and II, Interscience Publishers, New York, 1958 and 1963.

[D-3]   J. L. Doob, <u>Stochastic Processes</u>, Wiley, New York, 1952.

[G-1]   I. M. Gelfand & N. Ya. Vilenkin, Generalized Functions, vol. 4: Applications of Harmonic Analysis (tr. by A. Feinstein), Academic Press, New York, 1964.

[G-2]   T. N. Greville, "The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations," J. Soc. Indus. Appl. Math., 1 (1959), 38-43.

[H-1]   P. R. Halmos, Introduction to Hilbert Space and the Theory of Spectral Multiplicity, Chelsea, New York, 1951.

[H-2]   E. Hille & R. S. Phillips, Functional Analysis and Semi-Groups (rev. ed.), Am. Math. Soc. Colloq. Pub., Am. Math. Soc., Providence, R.I., 1958.

[I-1]   IEEE Trans. Automatic Control, Special Issue on Linear-Quadratic-Gaussian Problem, AC-16 (1971).

[K-1]   R. Kalman, Y. Ho & K. Narendra, "Controllability of linear dynamical systems," in Contribution to Differential Equations, vol. 1, Wiley, New York, 1962; pp. 189-213.

[L-1]   W. S. Loud, "Generalized inverses and generalized Green's functions," SIAM J. Appl. Math., 12 (1970).

[L-2]   D. G. Luenberger, Optimization by Vector Space Methods, Wiley, New York, 1969.

[M-1]   J. L. Massera & J. J. Schaffer, "Linear differential equations and functional analysis," Ann. of Math., 67 (1958), 517-573.

[M-2]   N. Minamide and K. Nakamura, "A restricted pseudoinverse and its application to constrained minima," SIAM J. Appl. Math., 19 (1970), 167-177.

[N-1]   M. Z. Nashed, A Retrospective and Prospective Survey of Generalized Inverses of Operators, Mathematics Research Center Report no. 1125, University of Wisconsin, Madison, 1971.

[P-1]   K. R. Parthasarathy, Probability Measures in Metric Spaces, Academic Press, New York, 1967.

[P-2]   E. Parzen, "An approach to time series analysis," Ann. Math. Stat., 32 (1961), 951-989.

[P-3]   W. A. Porter, Modern Foundations of System Engineering, Macmillan, New York, 1966.

[P-4]   W. A. Porter & J. P. Williams, "Extensions of the minimum effort control problem, " J. Math. Anal. and Appl., 10 (1966), 536-549.

[R-1]   C. R. Rao & S. K. Mitra, Generalized Inverse of Matrices and Its Application, Wiley, New York, 1971.

[R-2]   F. Riesz & B. Sz.-Nagy, Functional Analysis (tr. by L. Boron), Frederick Ungar Publishing Co., New York, 1955.

[R-3]   W. L. Root, On the Modelling of Systems for Identification, Report, University of Michigan, Ann Arbor, 1972.

[R-4]   W. L. Root, "On the structure of a class of system identification problems, " Automatica, 7 (1971), 219-231.

[R-5]   W. L. Root, "On the modelling and estimation of communication channels, " Multivariate Analysis-III, (ed. R. Krishnaiah) Academic Press, New York, 1973.