

THE UNIVERSITY OF MICHIGAN  
COLLEGE OF ENGINEERING

LEAST SQUARES AND LINEAR UNBIASED  
MINIMUM VARIANCE ESTIMATION IN  
EUCLIDEAN SPACE AND HILBERT SPACE

Philip H. Fiske  
William L. Root

ORA Project-011306

supported by:

UNITED STATES AIR FORCE  
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH  
AIR FORCE SYSTEMS COMMAND  
GRANT NO. 72-2328A  
WASHINGTON, D. C.

administered through:

OFFICE OF RESEARCH ADMINISTRATION

ANN ARBOR

January 1974

Engu  
UMR  
1494

## 1. INTRODUCTION

The problem to be discussed is that of estimating the unknown quantity  $h$  in the linear model

$$y = Xh + n, \quad (1.1)$$

where  $y$  represents an observed signal,  $X$  is a known linear transformation, and  $n$  is a noise or error signal. The unknown "vector parameter"  $h$  is an element of a separable Hilbert space  $\mathcal{H}_1$ ;  $y$  is an element of a separable Hilbert space  $\mathcal{H}_2$ ;  $X$  is a bounded linear transformation from  $\mathcal{H}_1$  into  $\mathcal{H}_2$ , and  $n$  is either an  $\mathcal{H}_2$ -valued random variable or an unknown error term in  $\mathcal{H}_2$ . The spaces  $\mathcal{H}_1$  and  $\mathcal{H}_2$  may be real or complex Hilbert spaces, and either or both may be finite-dimensional, although clearly only a finite-dimensional projection of  $h$  is estimatable when  $\mathcal{H}_2$  is finite-dimensional.

The following cases are considered: (i)  $h$  is an unknown, and  $n$  is an unknown error; (ii)  $h$  is an unknown, and  $n$  is a random variable in  $\mathcal{H}_2$  with mean zero and a known covariance operator; (iii)  $h$  is unknown but known to belong to a specified subset of  $\mathcal{H}_1$ , and  $n$  is again a random variable as in (ii). Case (i) is handled, as is conventional when there is no statistical information, by finding a least-squares estimator for  $h$ . This estimator is also put in recursive form for a linear dynamical systems model. For case (ii) a linear unbiased minimum-variance (LUMV) estimator is obtained. For the special case of finite-dimensional parameter and observation spaces in a dynamical systems model, this estimator is also put in recursive

form. For case (iii) a modification of the LUMV estimator is obtained which is no longer unbiased but which yields smaller error under the stated conditions than the LUMV estimator.

The least-squares solution for case (i) is well-known (see, e.g., [B-1]). We are not aware that the recursive form has been given for Hilbert-space-valued parameters, but it follows readily by standard methods. The LUMV estimator for case (ii) is developed in [B-1] and is stated in [R-2] without proof. The recursive form in finite-dimensional spaces is well-known (see, e.g., [S-1]). The modified LUMV estimator for case (iii) is given in [R-2] and [R-3]. Thus, most of the material contained herein appears elsewhere, and some is quite standard; however we know of no place where it has been put together in a reasonably self-contained, unified treatment. The proof of the Gauss-Markov theorem, that is, the theorem yielding an LUMV estimator, is somewhat complicated in the Hilbert space context, and has been put in an Appendix. The proof given is basically the same as that in [B-1], but it is superficially different because we do not use the theory of pseudo-inverses explicitly. The recursive solutions obtained for the least-squares and LUMV estimators are of the same form as the Kalman filtering equations, but they are obtained as solutions to a different minimization problem. Kalman filtering requires that  $h$  be a random variable, and then the expected value of the squared error in the parameter estimate is minimized.

There are, of course, a great many applications of the theory of estimating parameters in a linear model. However, we have in mind application to system identification, considered from a rather general point of view according to which there is no initial parameterization. This application leads to the model (1.1) and the estimation problems based on (1.1) as follows. If we momentarily neglect statistical considerations, the systems to be considered can be characterized by the equation  $y = H(x)$  where  $x$  (the input signal) is an element of a specified set  $\mathcal{X}$ ,  $y$  (the resulting output signal) is an element of a specified set  $\mathcal{Y}$ , and  $H$  (the system transformation) is an element of a specified class  $\mathcal{H}$  of functions from  $\mathcal{X}$  into  $\mathcal{Y}$ . Briefly, the basic identification problem is to find an  $H \in \mathcal{H}$  that adequately represents an unknown system from the results of one or a series of experiments involving the introduction of known input signals  $x$  and measurement of the corresponding output signals  $y$ .

One can regard each  $x \in \mathcal{X}$  as determining a function from  $\mathcal{H}$  into  $\mathcal{Y}$ . To emphasize this point of view, we can write  $y = H(x) = X(H)$ ,  $H \in \mathcal{H}$ , i.e.  $X$  is the mapping that assigns output  $y$  to system  $H$  when  $x$  ( $\sim X$ ) is the input signal and  $y = H(x)$ . Then, obviously, the problem of identifying  $H$  when the input signal is  $x$  is essentially the problem of inverting the mapping  $X$ . If the output space  $\mathcal{Y}$  is a linear space this can always be interpreted as a linear inversion problem as follows. Define addition and scalar multiplication in the way they are conventionally defined for spaces of functions,

i. e. for each  $H_1, H_2, H \in \mathcal{H}$  and all real numbers  $\alpha$  define  $H_1 + H_2$  and  $\alpha H$  by

$$(H_1 + H_2)(x) = H_1(x) + H_2(x), \quad x \in \mathcal{X}$$

$$(\alpha H)(x) = \alpha[H(x)], \quad x \in \mathcal{X}$$

Extend  $\mathcal{H}$  if necessary so that it is closed under linear combinations; then  $\mathcal{H}$  becomes a linear space. This extension is permissible because it only enlarges the basic class of systems being considered. The notation  $\mathcal{H}$  will be retained, but it is henceforth assumed that  $\mathcal{H}$  is a full linear space. Then  $X$  is a linear transformation on  $\mathcal{H}$ , because:

$$\begin{aligned} X(\alpha H_1 + \beta H_2) &= (\alpha H_1 + \beta H_2)(x) \\ &= \alpha H_1(x) + \beta H_2(x) \\ &= \alpha X(H_1) + \beta X(H_2). \end{aligned}$$

The first and third equalities follow from the definition of  $X$ , the second from the definitions of addition and scalar multiplication in  $\mathcal{H}$ . Hence, with the linear structure prescribed for  $\mathcal{H}$ , the mapping  $X$  from  $\mathcal{H}$  into  $\mathcal{Y}$  defined by  $y = X(H) = H(x)$  is linear.

This simple observation is important because it says that the input-output identification problem can always be studied by linear analysis. When output observation noise is added, the problem can be treated as a linear regression problem. The observation takes on practical as well as theoretical importance if, when parameters are introduced, as they must be to permit actual computation, they are introduced in such fashion as to determine linearly the transformations

H. Then the actual computation algorithm will be linear, or, in the usual terminology, the problem will be "linear in the parameters." The only condition necessary for this linearity of  $X$  is that  $\mathcal{Y}$  be a linear space.

The problems applicable to this paper are those in which  $\mathcal{Y}$  and  $\mathcal{H}$  are separable Hilbert spaces (denoted, respectively, by  $\mathcal{H}_2$  and  $\mathcal{H}_1$ ) and  $X$  is also a bounded transformation. The assumption that  $X$  is bounded is chiefly for convenience, but it is realistic for most applications.

## 2. NOTATIONS AND CERTAIN ELEMENTARY FACTS ABOUT LINEAR OPERATORS IN HILBERT SPACE

To establish notations, let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be separable Hilbert spaces, which may be the same space. If  $\mathcal{E} \subset \mathcal{H}_1$ , we denote its closure in  $\mathcal{H}_1$  by  $\bar{\mathcal{E}}$  and its orthogonal complement in  $\mathcal{H}_1$  by  $\mathcal{E}^\perp$ . Let  $X$  be a linear mapping from (part of)  $\mathcal{H}_1$  into  $\mathcal{H}_2$ .  $\mathcal{D}(X) = \{h \in \mathcal{H}_1 : Xh \text{ is defined}\}$  is the domain of  $X$ ;  $\mathcal{N}(X) = \{h \in \mathcal{D}(X) : Xh = 0\}$  is the null-space of  $X$ , and  $\mathcal{R}(X) = \{y \in \mathcal{H}_2 : Xh = y \text{ for some } h\}$  is the range of  $X$ . All are linear sets. We call  $X$  an operator from  $\mathcal{H}_1$  to  $\mathcal{H}_2$ ; the term bounded operator will imply that not only is  $X$  bounded in the usual operator norm, but also that  $\mathcal{D}(X) = \mathcal{H}_1$  unless an exception is explicitly stated.

Let  $X$  be a bounded operator. Its adjoint,  $X^*$ , is the bounded operator from  $\mathcal{H}_2$  to  $\mathcal{H}_1$  defined by  $(Xh, z) = (h, X^*z)$  for all  $h \in \mathcal{H}_1$ ,  $z \in \mathcal{H}_2$ . Note that  $(Xh, z)$  is an inner product in  $\mathcal{H}_2$ , whereas  $(h, X^*z)$  is an inner product in  $\mathcal{H}_1$ . There is no real ambiguity in using the same notation for inner products (or norms) in both spaces.

The following facts are standard, and for the most part can be proven easily. (However, see [B-1], Appendix A). Assume  $X$  is a bounded operator

- 1)  $\mathcal{N}(X)$  and  $\mathcal{N}(X^*)$  are closed subspaces of  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , respectively.
- 2)  $\mathcal{R}(X)$  is closed if it is finite-dimensional, which is necessarily true if  $\mathcal{H}_1$  or  $\mathcal{H}_2$  is finite-dimensional. In general, a necessary and sufficient condition for  $\mathcal{R}(X)$  to be



closed is that  $X$  restricted to  $\mathcal{N}^\perp(X)$ , which is a closed subspace of  $\mathcal{H}_1$ , be bounded from below. That is, there must exist a constant  $c > 0$  such that  $\|Xh\| \geq c \|h\|$  for all  $h \in \mathcal{N}^\perp(X)$ . We denote the restriction of  $X$  to  $\mathcal{N}^\perp(X)$  by  $X_r$ .

- 3)  $\overline{\mathcal{R}(X)} = \mathcal{N}^\perp(X^*)$ , and  $\overline{\mathcal{R}(X^*)} = \mathcal{N}^\perp(X)$ . If we denote direct sum by  $\oplus$ , then

$$\begin{aligned}\mathcal{H}_1 &= \mathcal{N}(X) \oplus \mathcal{N}^\perp(X) = \mathcal{N}(X) \oplus \overline{\mathcal{R}(X^*)} \\ \mathcal{H}_2 &= \mathcal{N}(X^*) \oplus \mathcal{N}^\perp(X^*) = \mathcal{N}(X^*) \oplus \overline{\mathcal{R}(X)}.\end{aligned}$$

- 4)  $\mathcal{N}(X^*X) = \mathcal{N}(X)$ , and  $\mathcal{N}(XX^*) = \mathcal{N}(X^*)$ .

- 5)  $\mathcal{R}(X^*X) \subset \mathcal{R}(X^*) \subset \overline{\mathcal{R}(X^*X)}$ , and  
 $\mathcal{R}(XX^*) \subset \mathcal{R}(X) \subset \overline{\mathcal{R}(XX^*)}$ .

- 6) If any one of the sets  $\mathcal{R}(X)$ ,  $\mathcal{R}(X^*)$ ,  $\mathcal{R}(X^*X)$ ,

$\mathcal{R}(XX^*)$  is closed, so are all of the others, and then

$$\begin{aligned}\mathcal{R}(X) &= \mathcal{R}(XX^*) = \overline{\mathcal{R}(X)}, \text{ and} \\ \mathcal{R}(X^*) &= \mathcal{R}(X^*X) = \overline{\mathcal{R}(X^*)}.\end{aligned}$$

Consider  $A = X^*X$ , which is a bounded self-adjoint, nonnegative operator from  $\mathcal{H}_1$  into  $\mathcal{H}_1$ . Since  $A = A^*$ ,  $\mathcal{N}^\perp(A) = \overline{\mathcal{R}(A)}$ , and we denote the restriction of  $A$  to  $\overline{\mathcal{R}(A)}$  by  $A_r$ . Since  $\mathcal{N}(A_r) = \{0\}$ ,  $A_r$  is 1:1 from  $\overline{\mathcal{R}(A)}$  onto  $\mathcal{R}(A) = \mathcal{R}(A_r) \subset \overline{\mathcal{R}(A)}$ . Consequently,  $A_r^{-1}$  exists as an operator (not in general bounded) from  $\overline{\mathcal{R}(A)}$  to  $\overline{\mathcal{R}(A)}$ , with  $\mathcal{D}(A_r^{-1}) = \mathcal{R}(A)$ .

It is thus possible to write  $(X^*X)^{-1}g$  for all  $g \in \mathcal{R}(X^*X)$ , where  $(X^*X)^{-1}$  means  $A_r^{-1}$  (the subscript is dropped to simplify the notation).

If  $X$  has closed range, then it follows from (6) that  $A_r$  is 1:1 from  $\overline{\mathcal{R}(A_r)}$  onto  $\mathcal{R}(X) = \overline{\mathcal{R}(X)} = \overline{\mathcal{R}(A)}$ , and  $A_r^{-1}$  is then a bounded operator on  $\overline{\mathcal{R}(A)}$ .

These observations, not all of which are usually stated explicitly in standard works on Hilbert space (see, e.g. [R-1]), coupled with some basic results that are given in almost any reference, should suffice for the material in the body of the report. If one wishes to restrict consideration to finite-dimensional space only, the formal calculations in what follows can be read as just matrix calculations. In Appendix A it is necessary to deal with unbounded closed operators, and considerations similar to those above but somewhat more intricate are necessary.

### 3. LEAST-SQUARES ESTIMATION

In the model (1.1) we take  $n$  simply to be an unknown error;  $n$  is not assumed to have any statistical properties. It is desired to find a best estimate  $\hat{h}$  of  $h$  according to the criterion that the corresponding error  $n = y - X\hat{h}$  be of minimum norm. This is the classical criterion of least-squares estimation, and is usually stated: find  $\hat{h}$  such that

$$\|X\hat{h} - y\| \leq \|Xh - y\| \quad \forall h \in \mathcal{H}_1.$$

To see what  $\hat{h}$  should be we first do a formal calculation that consists essentially of completing a square. Precise statements and proofs are given subsequently, but it may be noted that the following calculation already yields a proof for the finite-dimensional case.

The expansion of  $\|Xh - y\|^2$  is

$$\begin{aligned} \|Xh - y\|^2 &= (Xh - y, Xh - y) \\ &= (X^* Xh, h) - 2 \operatorname{Re}(h, X^* y) + \|y\|^2. \end{aligned} \quad (3.1)$$

Let  $A = X^* X$ . Since  $A$  is a nonnegative self-adjoint operator, it has a nonnegative self-adjoint square root,  $A^{1/2}$  ([R-1, p. 265]).  $(A^{1/2})^{-1}$  exists as a (perhaps unbounded) operator on  $\overline{\mathcal{R}(A^{1/2})}$  with domain equal to  $\mathcal{R}(A^{1/2})$ ; denote  $(A^{1/2})^{-1}$  by  $A^{-1/2}$ .<sup>(1)</sup> Suppose that  $A^{-1/2} X^* y$  is defined, and put  $g = A^{-1/2} X^* y$ . Then

<sup>(1)</sup>We note that  $(A^{1/2})^{-1} = (A^{-1})^{1/2} = A^{-1/2}$ , and that  $A^{-1/2}$  is also a nonnegative self-adjoint operator.

$$\begin{aligned} \|A^{1/2}h-g\|^2 &= (A^{1/2}h, A^{1/2}h) - 2 \operatorname{Re}(A^{1/2}h, g) + \|g\|^2 \\ &= (Ah, h) - 2 \operatorname{Re}(h, X^*y) + \|g\|^2 \end{aligned} \quad (3.2)$$

since  $(A^{1/2}h, g) = (h, A^{1/2}g) = (h, A^{1/2}A^{-1/2}X^*y) = (h, X^*y)$ . But

(3.1) can be rewritten

$$\begin{aligned} \|Xh-y\|^2 &= (Ah, h) - 2 \operatorname{Re}(h, X^*y) + \|g\|^2 + (\|y\|^2 - \|g\|^2) \\ &= \|A^{1/2}h-g\|^2 + (\|y\|^2 - \|g\|^2). \end{aligned} \quad (3.3)$$

Obviously (3.3) is minimized by  $h = \hat{h}$  such that  $A^{1/2}\hat{h} = g$  if such  $\hat{h}$

exists. That is, the minimum is given formally by

$$\begin{aligned} \hat{h} &= A^{-1/2}g = A^{-1/2}A^{-1/2}X^*y \\ &= (X^*X)^{-1}X^*y. \end{aligned} \quad (3.4)$$

The expression on the right side of (3.4) is a classical formula for the least-squares estimate in finite-dimensional spaces.

We now develop rigorously the rather simple basic facts about least-squares estimates.

Proposition 3.1: Put  $y = y_1 + y_2$  where  $y_1 \in \overline{\mathcal{R}(X)}$ ,  $y_2 \in \mathcal{N}(X^*)$ .

Then a necessary and sufficient condition for there to exist an  $h$  such that  $\|Xh-y\|$  is a minimum is that  $y_1 \in \mathcal{R}(X)$ . If  $y_1 \in \mathcal{R}(X)$  then any  $h$  such that  $Xh = y_1$  provides a minimum.

Proof: Since  $(y_1, y_2) = 0$ ,

$$\begin{aligned} \|Xh-y\|^2 &= \|Xh-y_1-y_2\|^2 \\ &= \|Xh-y_1\|^2 + \|y_2\|^2. \end{aligned} \quad (3.5)$$

If  $y_1 \in \mathcal{R}(X)$ ,  $\min \|Xh-y\|^2$  exists and is equal to  $\|y_2\|^2$ . Clearly

any  $h$  such that  $Xh = y_1$  provides a minimum. Conversely, if  $\min$

$\|Xh-y\|^2$  exists it must be equal to  $\|y_2\|^2$ , which implies that

$y_1 \in \mathcal{R}(X)$ .  $\square$

If  $X$  has closed range, i.e.  $\mathcal{R}(X) = \overline{\mathcal{R}(X)}$ ,  $\min \|Xh - y\|^2$  always exists. As mentioned in the previous Section, a necessary and sufficient condition that  $X$  have closed range is that  $X$  restricted to  $\mathcal{N}^\perp(X)$  be bounded from below. If the range of  $X$  is finite-dimensional it is always closed and hence a minimum always exists. Finally, it is worth noting that since the estimate is based on an observation in the output space, if  $X$  is not 1:1, the minimizing  $\hat{h}$  is not unique. In this case a further criterion must be introduced to fix  $\hat{h}$ ; usually  $\hat{h}$  is taken to be itself of minimum norm, which amounts to requiring that  $h \in \mathcal{N}^\perp(X)$ .

Proposition 3.2: If  $y_1 \in \mathcal{R}(X)$  then  $(X^* X)^{-1} X^* y$  is defined and provides a minimum.

Proof:

$$\begin{aligned} (X^* X)^{-1} X^* y &= (X^* X)^{-1} X^* (y_1 + y_2) \\ &= (X^* X)^{-1} X^* y_1 \end{aligned}$$

Since  $y_1 \in \mathcal{R}(X)$ , there exists  $h$  such that  $Xh = y_1$ . Hence

$$(X^* X)^{-1} X^* y = (X^* X)^{-1} X^* (Xh) = h$$

which **implies that**  $(X^* X)^{-1} X^* y$  is defined. From Proposition 3.1 and the fact that

$$X[(X^* X)^{-1} X^* y] = y_1$$

we have that  $(X^* X)^{-1} X^* y$  provides a minimum.

Proposition 3.3: If  $(X^* X)^{-1} X^* y$  is defined then a minimum exists (i.e.,  $y_1 \in \mathcal{R}(X)$ ) and is provided by  $(X^* X)^{-1} X^* y$ .

Proof: Since  $(X^* X)^{-1} X^* y$  is defined we can write

$$\begin{aligned} (X^* X)^{-1} X^* y &= (X^* X)^{-1} X^* (y_1 + y_2) \\ &= (X^* X)^{-1} X^* y_1 \end{aligned}$$

Hence  $X^* y_1 \in \mathcal{D}[(X^* X)^{-1}]$  so in fact  $X^* y_1 \in \mathcal{R}(X^* X)$ . Thus there exists  $h \in \mathcal{N}_1$  (in fact, we can choose  $h \in \mathcal{N}_1^\perp(X^* X)$ ) such that

$$X^* Xh = X^* y_1$$

Since  $Xh \in \mathcal{N}_1^\perp(X^*)$  it follows that  $Xh = y_1$  (i. e.,  $y_1 \in \mathcal{R}(X)$ ).

Since  $y_1 \in \mathcal{R}(X)$ , we know that a minimum exists from Proposition 1.1. Proposition 1.2 yields the fact that  $(X^* X)^{-1} X^* y$  provides a minimum. |||

It is perhaps worth noting that if  $y_1 \in \mathcal{R}(X)$  then the steps in the heuristic derivation of the least-squares estimation formula are justified. Recall that this amounts to showing that  $(X^* X)^{-1/2} X^* y$  and  $(X^* X)^{-1/2} (X^* X)^{-1/2} X^* y$  are defined, where  $(X^* X)^{-1/2}$  means the inverse of the restriction of  $(X^* X)^{1/2}$ . From Proposition 3.2 we have that  $(X^* X)^{-1} X^* y$  is defined, so  $X^* y \in \mathcal{R}(X^* X)$  and we can write

$$z = X^* y = X^* Xh_1, \quad h_1 \in \mathcal{N}_1^\perp(X^* X).$$

Then  $h_1 \in \mathcal{N}_1^\perp[(X^* X)^{1/2}]$  since  $\mathcal{N}_1^\perp(X^* X) \subset \mathcal{N}_1^\perp[(X^* X)^{1/2}]$ .

$$\begin{aligned} \text{But, } (X^* X)^{1/2} h_1 &= [(X^* X)^{1/2}]^{-1} (X^* X)^{1/2} (X^* X)^{1/2} h_1 \\ &= (X^* X)^{-1/2} z \\ &= (X^* X)^{-1/2} X^* y, \end{aligned}$$

so  $(X^* X)^{-1/2} X^* y$  is defined. Also

$$[(X^* X)^{1/2}]^{-1} (X^* X)^{1/2} h_1 = (X^* X)^{-1/2} (X^* X)^{-1/2} X^* y$$

is defined. The last expression can be shown to be just  $(X^* X)^{-1} X^* y$ , of course.

We comment briefly on the situation when no minimum exists (i. e.,  $y_1 \notin \mathcal{R}(X)$ ). First, one can always find elements  $\tilde{h}$  that provide an approximation to a minimum. In fact, given  $\epsilon > 0$ , one can find  $\tilde{y}_1 \in \mathcal{R}(X)$  such that  $\|\tilde{y}_1 - y_1\|^2 \leq \epsilon$ . Write  $\tilde{y}_1 = X\tilde{h}$ , then

$$\begin{aligned} \|X\tilde{h} - y\|^2 &= \|X\tilde{h} - y_1 - y_2\|^2 \\ &\leq \epsilon + \|y_2\|^2. \end{aligned}$$

However, this fact is of limited usefulness because of the result of the next Proposition.

Proposition 3.4: If  $y_1 \notin \mathcal{R}(X)$ , then any sequence  $\{h_n\}$  such that

$\lim_{n \rightarrow \infty} Xh_n = y_1$  is unbounded.

Proof: Suppose that there exists a constant  $C$  such that  $\|h_n\| \leq C$  for all  $n$ . Since every bounded point set in a Hilbert space is weakly compact [A-1, p. 46], there is a subsequence of  $\{h_n\}$ , which we denote by  $\{h_{n_i}\}$ , such that  $(h_{n_i}, f) \rightarrow (h, f)$  for all  $f \in \mathcal{H}_1$ . Pick an arbitrary  $z \in \mathcal{H}_2$  (without loss of generality we can choose a  $z$  such that  $\|z\| = 1$ ) and consider

$$\begin{aligned} (Xh - y_1, z) &= ((Xh - Xh_{n_i}) + (Xh_{n_i} - y_1), z) \\ &= (X(h - h_{n_i}), z) + (Xh_{n_i} - y_1, z) \\ &= (h - h_{n_i}, X^* z) + (Xh_{n_i} - y_1, z) \end{aligned}$$

Since  $\|z\| = 1$ ,

$$|(Xh - y_1, z)| \leq |(h - h_{n_i}, X^* z)| + \|Xh_{n_i} - y_1\|.$$

Given any  $\epsilon > 0$ , there exists  $n_0$  such that for  $n_i \geq n_0$ ,  $\|Xh_{n_i} - y_1\| \leq \epsilon/2$  and  $|(h - h_{n_i}, X^* z)| \leq \epsilon/2$ . Hence  $|(Xh - y_1, z)| \leq \epsilon$ , but  $\epsilon$  is arbitrary, so  $(Xh - y_1, z) = 0$ . Since this relation holds for all  $z \in \mathcal{H}_2$ , we must have  $Xh = y_1$ . But  $y_1 \notin \mathcal{R}(X)$ , so a contradiction results and thus  $\{h_n\}$  must be unbounded. |||

In the finite-dimensional case where  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are Euclidean spaces,  $\mathcal{R}(X^* X) = \mathcal{R}(X^*) = \overline{\mathcal{R}(X^*)}$ ,  $X^* X$  is a 1:1 mapping from  $\mathcal{R}(X^*)$  onto  $\mathcal{R}(X^*)$ , and  $(X^* X)^{-1}$  is well defined as an operator from  $\mathcal{R}(X^*)$  to  $\mathcal{R}(X^*)$ . Also, of course,  $\mathcal{N}(X^*) = \mathcal{N}^\perp(A)$ . The formula

$$\hat{h} = (X^* X)^{-1} X^* y \quad (3.6)$$

which gives the least-squares estimate of  $h$  is always defined. In what might be regarded as the conventional situation, at least from the statistical point of view,  $\mathcal{N}(X) = 0$ , and  $\mathcal{R}(X) = \mathcal{N}^\perp(X^*)$  is a proper subspace of  $\mathcal{B}_1$ , usually of much lower dimension. In other words, there is redundancy in the observations, and the least-squares formula uses this redundancy to minimize the errors. At the other extreme, the dimension of  $\mathcal{B}_2$  exceeds the dimension of  $\mathcal{B}_1$ , and  $\mathcal{R}(X) = \mathcal{B}_1$ . There is then not only no redundancy in the observations, but part of  $h$  cannot be estimated because  $\mathcal{N}(X)$  is necessarily non-zero. In this case  $h$  minus the projection of  $h$  on  $\mathcal{N}^\perp(X)$  is treated as error and its norm is set to zero by the estimator  $\hat{h}$ . The formula for the least-squares estimator thus gives a pseudo-inverse. In general, neither  $\mathcal{N}(X)$  nor  $\mathcal{N}(X^*)$  are zero, so part of  $h$  cannot be estimated, while on the other hand, there is redundancy



available for reducing the error in estimating the projection of  $h$  on  $\mathcal{R}_1(X)$ .

In the infinite-dimensional case the situations regarding redundancy in the observations and a non-estimatable part of  $h$  are essentially the same. However,  $\mathcal{R}_1(X)$  need not be closed, and in case it is not, least-squares estimation does not make much sense practically because only for some observations (those that project into the range of  $X$ ) does the method apply. If  $\mathcal{R}_1(X)$  is not closed it is always possible to introduce a new Hilbert space  $\tilde{\mathcal{H}}_2$  which contains the elements of  $\mathcal{R}_1(X)$  and in which  $\mathcal{R}_1(X)$  is closed. In many instances  $\tilde{\mathcal{H}}_2$  will give a satisfactory model for least-squares estimation, but we do not discuss this matter here (see [B-1, Section 3]).

For the remainder of this Section we consider the problem of obtaining a recursive solution for the least-squares estimation problem in the case where  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are Hilbert spaces. The procedure followed for converting the basic solution into a recursive one is standard for the finite-dimensional problem, and as will be seen, there is no difficulty in extending it to Hilbert spaces.

We consider the system

$$\begin{aligned}
 y_1 &= B_1 h_1 + n_1 \stackrel{d}{=} B_1 h + n_1 \\
 y_2 &= B_2 \phi_{2,1} h_1 + n_2 \stackrel{d}{=} B_2 \phi_{2,1} h + n_2 \\
 y_3 &= B_3 \phi_{3,2} \phi_{2,1} h_1 + n_3 \stackrel{d}{=} B_3 \phi_{3,1} h + n_3 \\
 &\vdots \\
 y_n &= B_n \phi_{n,n-1} \phi_{n-1,n-2} \cdots \phi_{2,1} h_1 + n_n \stackrel{d}{=} B_n \phi_{n,1} h + n_n
 \end{aligned}$$

which equivalently can be written in the state variable form

$$\begin{aligned} h_{i+1} &= \phi_{i+1,i} h_i \\ y_i &= B_i h_i + n_i \quad i=1, \dots, n \\ h_1 &\stackrel{d}{=} h \end{aligned} \quad (3.7)$$

We assume that  $B_i$  and  $\phi_{i+1,i}$ ,  $i=1, \dots, n$ , are all bounded operators with zero null-space and closed range, and in fact that the  $\phi_{i+1,i}$  are 1:1 from  $\mathcal{H}_1$  onto  $\mathcal{H}_1$ . We can then define  $X_1 \stackrel{d}{=} B_1$ ,  $X_2 \stackrel{d}{=} B_2 \phi_{2,1}$ ,  $\dots$ ,  $X_n \stackrel{d}{=} B_n \phi_{n,1}$  so that

$$\begin{aligned} y_1 &= X_1 h + n_1 \\ y_2 &= X_2 h + n_2 \\ &\vdots \\ y_n &= X_n h + n_n \end{aligned} \quad (3.8)$$

where the  $X_i$  are bounded operators from  $\mathcal{H}_1$  to  $\mathcal{H}_2$  with  $\mathcal{N}(X_i) = 0$  and  $\mathcal{R}(X_i)$  closed in  $\mathcal{H}_2$ . The  $n_i$  are taken to be unknown errors lying in  $\mathcal{H}_2$ .

The system of equations in (3.8) can also be written in the form

$$\underline{y}_n = \underline{X}_n h + \underline{n}_n \quad (3.9)$$

where

$$\underline{y}_n \stackrel{d}{=} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \underline{X}_n = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad \underline{n}_n = \begin{bmatrix} n_1 \\ \vdots \\ n_n \end{bmatrix}$$

i.e.,  $\underline{y}_n$  and  $\underline{n}_n$  are elements of  $\mathcal{H}_2$ , the n-fold direct sum of copies of  $\mathcal{H}_2$ , and  $\underline{X}_n$  is a bounded operator from  $\mathcal{H}_1$  into  $\mathcal{H}_2$ . The "vector-matrix" notation is self-evident.

From Proposition 3.3 it follows that the least-squares estimator of  $h$  is given by

$$\hat{h}_n = (\underline{X}_n^* \underline{X}_n)^{-1} \underline{X}_n^* y_n \quad (3.10)$$

if the expression on the right is defined. Put  $P_n = (\underline{X}_n^* \underline{X}_n)^{-1}$  and  $b_n = \underline{X}_n^* y_n$ . Note that since  $\mathcal{R}(X_i)$  is closed, so is  $\mathcal{R}(X_i^* X_i)$ .

Thus

$$\mathcal{R}(X_i^* X_i) = \overline{\mathcal{R}(X_i^* X_i)} = \mathcal{R}(X_i^*) = \overline{\mathcal{R}(X_i^*)} = \mathcal{R}^\perp(X_i) = \mathcal{H}_1$$

$$\mathcal{N}(X_i^* X_i) = \mathcal{R}[(X_i^* X_i)^*]^\perp = \mathcal{R}(X_i^* X_i)^\perp = \mathcal{H}_1^\perp = 0.$$

So  $X_i^* X_i$  is a 1:1 map with  $\mathcal{R}(X_i^* X_i) = \mathcal{R}(X_i^* X_i) = \mathcal{H}_1$ .

Consider  $\mathcal{R}(P_n^{-1}) = \mathcal{R}(\underline{X}_n^* \underline{X}_n)$ . Suppose  $z \perp \mathcal{R}(\underline{X}_n^* \underline{X}_n)$ . Then

$$((\underline{X}_n^* \underline{X}_n) y, z) = 0 \quad \text{for every } y \in \mathcal{H}_1$$

But

$$\begin{aligned} ((\underline{X}_n^* \underline{X}_n) y, z) &= ((X_1^* X_1) y, z) + \dots + ((X_n^* X_n) y, z) \\ &= (X_1 y, X_1 z) + \dots + (X_n y, X_n z) \end{aligned}$$

Since the above is true for all  $y \in \mathcal{H}_1$ , it must be true for  $y = z$ .

Hence we have

$$((\underline{X}_n^* \underline{X}_n) z, z) = \|X_1 z\|^2 + \dots + \|X_n z\|^2 = 0$$

This implies that we must have  $X_1 z = \dots = X_n z = 0$  which in turn

implies that  $z = 0$  since  $\mathcal{R}(X_i) = 0$ ,  $i=1, \dots, n$ . Hence  $\overline{\mathcal{R}(\underline{X}_n^* \underline{X}_n)} =$

$\mathcal{H}_1$ . Since

$$\underline{X}_n^* \underline{X}_n = X_1^* X_1 + \dots + X_n^* X_n$$

is self-adjoint,  $\mathcal{N}(\underline{X}_n^* \underline{X}_n) = \mathcal{H}_1^\perp = 0$  by an argument similar to the one presented above.

We have just established that  $\mathcal{D}(P_n) = \mathcal{D}[(\tilde{X}_n^* \tilde{X}_n)^{-1}] = \mathcal{N}_1$ .  
 The operator  $(\tilde{X}_n^* \tilde{X}_n)^{-1}$  is self-adjoint, hence it is also closed.  
 However, a closed operator defined everywhere is bounded. So  
 $P_n = (\tilde{X}_n^* \tilde{X}_n)^{-1}$  is bounded and self-adjoint. Of course,  $P_n^{-1} = \tilde{X}_n^* \tilde{X}_n$  is also bounded and self-adjoint.

We can now write

$$\begin{aligned} P_n^{-1} &= \tilde{X}_n^* \tilde{X}_n = X_1^* X_1 + \dots + X_n^* X_n \\ &= \tilde{X}_{n-1}^* \tilde{X}_{n-1} + X_n^* X_n \\ &= P_{n-1}^{-1} + X_n^* X_n \end{aligned} \quad (3.11)$$

Also,

$$\begin{aligned} b_n &\stackrel{d}{=} \tilde{X}_n^* y_n = X_1^* y_1 + X_2^* y_2 + \dots + X_n^* y_n \\ &= \tilde{X}_{n-1}^* y_{n-1} + X_n^* y_n \\ &= b_{n-1} + X_n^* y_n \end{aligned} \quad (3.12)$$

From equation (3.11),

$$P_n P_n^{-1} P_{n-1} = P_n [P_{n-1}^{-1} + X_n^* X_n] P_{n-1}$$

Hence,

$$P_{n-1} = P_n + P_n X_n^* X_n P_{n-1} \quad (3.13)$$

and

$$\begin{aligned} P_{n-1} X_n^* &= P_n X_n^* + P_n X_n^* X_n P_{n-1} X_n^* \\ &= P_n X_n^* [I + X_n P_{n-1} X_n^*] \end{aligned}$$

We now note that  $[I + X_n P_{n-1} X_n^*]^{-1}$  exists since  $[I + X_n P_{n-1} X_n^*]^{-1} = [I + (X_n P_{n-1}^{1/2})(X_n P_{n-1}^{1/2})^*]^{-1}$ , which is known to exist [R-1, p. 307].

Furthermore, the norm of this operator is bounded by one.

Therefore, we can now write

$$P_{n-1} X_n^* [I + X_n P_{n-1} X_n^*]^{-1} X_n P_{n-1} = P_n X_n^* X_n P_{n-1} \quad (3.14)$$

Using equations (3.13) and (3.14) we obtain

$$P_n = P_{n-1} - P_{n-1} X_n^* [I + X_n P_{n-1} X_n^*]^{-1} X_n P_{n-1} \quad (3.15)$$

Using equations (3.12) and (3.15) in the expression for  $\hat{h}_n$  yields

$$\begin{aligned} \hat{h}_n &= \{P_{n-1} - P_{n-1} X_n^* [I + X_n P_{n-1} X_n^*]^{-1} X_n P_{n-1}\} \{b_{n-1} + X_n^* y_n\} \\ &= P_{n-1} b_{n-1} - P_{n-1} X_n^* [I + X_n P_{n-1} X_n^*]^{-1} X_n P_{n-1} b_{n-1} \\ &\quad + P_{n-1} X_n^* y_n - P_{n-1} X_n^* [I + X_n P_{n-1} X_n^*]^{-1} X_n P_{n-1} X_n^* y_n \end{aligned} \quad (3.16)$$

We now observe that if we have an operator  $A$  such that  $(I+A)^{-1}$  exists and is bounded, then

$$[I - (I+A)^{-1}A] = (I+A)^{-1}[I+A-A] = (I+A)^{-1}$$

Thus we can now write

$$\hat{h}_n = \hat{h}_{n-1} - P_{n-1} X_n^* [I + X_n P_{n-1} X_n^*]^{-1} [X_n \hat{h}_{n-1} - y_n]$$

or

$$\hat{h}_n = \hat{h}_{n-1} + K_n [y_n - X_n \hat{h}_{n-1}] \quad (3.17)$$

where

$$K_n = P_{n-1} X_n^* [I + X_n P_{n-1} X_n^*]^{-1} \quad (3.18)$$

and

$$P_n = P_{n-1} - K_n X_n P_{n-1} \quad (3.19)$$

Hence, to obtain the recursive least-squares estimator, we start

with

$$\hat{h}_1 = (X_1^* X_1)^{-1} X_1^* y_1$$

$$P_1 = (X_1^* X_1)^{-1}$$

and apply the recursive relations for  $n=2, 3, \dots$

#### 4. LUMV ESTIMATION

Consider again the linear model

$$y = Xh + n \tag{4.1}$$

where as before  $h$  belongs to a separable Hilbert space  $\mathcal{H}_1$  (which may be finite-dimensional) and is the vector of parameters to be estimated,  $X$  is a bounded linear transformation from  $\mathcal{H}_1$  to a separable Hilbert space  $\mathcal{H}_2$ , but where  $n$  is now a random variable taking values in  $\mathcal{H}_2$  (which may also be finite-dimensional). Again  $y$  is a vector of observations in  $\mathcal{H}_2$ , but now, for each  $h$ , it is a random variable taking values in  $\mathcal{H}_2$ . We assume that  $n$  has mean zero, that its covariance operator  $R$  exists and is strictly positive definite and that  $R^{1/2}$  is Hilbert-Schmidt. The covariance operator  $R$  is the unique bounded, self-adjoint operator, when it exists, satisfying  $E(f, n)(n, g) = (Rf, g)$ . If  $R^{1/2}$  is Hilbert-Schmidt, then  $R$  has finite trace. If  $\mathcal{H}_2$  is finite-dimensional all these conditions are satisfied automatically of course except the one that  $R$  (which is just the usual covariance matrix) be strictly positive definite. <sup>(1)</sup>

For further discussion of covariance operators for Hilbert-space-valued random variables, see [B-1, Appendix B].

We use the notation

$$h = h_1 + h_2, \quad h_1 \in \mathcal{N}_\perp(X), \quad h_2 \in \mathcal{N}(X)$$

---

<sup>(1)</sup>We need this condition for our proof. In the finite-dimensional case there are proofs that do not require it, see [A-2].

so that

$$y = Xh_1 + n.$$

The problem is to find an unbiased linear estimator for  $h_1$  that is of minimum "variance," if such exists. That is, find (if possible) a bounded linear operator  $C$  from  $\mathcal{H}_2$  to  $\mathcal{R}^\perp(X)$ , defined on all of  $\mathcal{H}_2$ , such that the estimate  $\hat{h} \stackrel{d}{=} Cy$  has the properties

$$(1) E\hat{h} = h_1$$

$$(2) E\|\hat{h} - E\hat{h}\|^2 = E\|\hat{h} - h_1\|^2 \text{ exists and is a minimum.}$$

We ask only for an estimator  $C$  for  $h_1$  because clearly the observation  $y$  tells us nothing about  $h_2$ .

Before proceeding to this problem we discuss first the possibility of unbiased finite-variance estimators. The unbiasedness condition (1) can be rewritten

$$E\hat{h} = E(CXh_1 + Cn) = CXh_1 = h_1 \quad (4.2)$$

or,  $CX_r h_1 = h_1$ , where  $X_r$  is the restriction of  $X$  to  $\mathcal{R}^\perp(X)$ . Consequently, the restriction of  $C$  to  $\overline{\mathcal{R}(X)}$  must be  $X_r^{-1}$ . Now we note that

$$\begin{aligned} \|C\| &= \sup_{\|y\|=1} \|Cy\| \geq \sup_{\substack{\|y\|=1 \\ y \in \mathcal{R}(X)}} \|Cy\| = \sup_{\substack{\|y\|=1 \\ y \in \mathcal{R}(X)}} \|X_r^{-1}y\| \\ &= \|X_r^{-1}\|. \end{aligned}$$

Thus  $C$  cannot be bounded unless  $X_r^{-1}$  is bounded, or equivalently,  $X$  has closed range. Hence we must have the condition that  $X$  has closed range.

The mean-squared error, or "variance," is given by

$$E\|\hat{h}-h_1\|^2 = E\|CXh_1 + Cn - h_1\|^2 = E\|Cn\|^2 \quad (4.3)$$

from the unbiased condition. Let  $\{\phi_i\}$  be any c. o. n. s. for  $\mathcal{H}_1$ .

Then

$$\|Cn\|^2 = \sum_i |(Cn, \phi_i)|^2 = \sum_i (C^* \phi_i, n)(n, C^* \phi_i) \quad (4.4)$$

Formally we have

$$\begin{aligned} E\|Cn\|^2 &= \sum_i E(C^* \phi_i, n)(n, C^* \phi_i) = \sum_i (RC^* \phi_i, C^* \phi_i) \\ &= \sum_i (CRC^* \phi_i, \phi_i) = \text{Trace } (CRC^*) \end{aligned} \quad (4.5)$$

In fact, if  $C$  is such that  $\text{Trace } (CRC^*)$  exists, the interchange of expectation and summation is justified (by the Beppo Levi theorem), and equation (4.5) holds. Now if  $R^{1/2}$  is Hilbert-Schmidt, then  $CR^{1/2}$  and  $R^{1/2}C^*$  are Hilbert-Schmidt since  $C$  is bounded, and  $CR^{1/2} \cdot R^{1/2}C^* = CRC^*$  has finite trace [G-1].

Thus if we assume that  $R^{1/2}$  is Hilbert-Schmidt,  $C$  can be any bounded operator that satisfies the unbiasedness condition, and a finite-variance estimator results.

In this Section we shall derive the LUMV estimator only for the classical case where  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are finite-dimensional, with the statement and proof for the Hilbert-space case relegated to the Appendix because of some fussy details required. The structure of the proof remains the same in the Hilbert space case. However, there are two preparatory lemmas which we now state and prove, which, because they will be used in the Appendix and because they are of some interest in themselves, are given in Hilbert space form. In the following two propositions A need be only a closed, densely defined



linear operator, but if one does not need the results for so general an operator,  $A$  can be taken as bounded and everywhere defined.

Proposition 4.1: Let  $\mathcal{B}$  and  $\mathcal{K}$  be Hilbert spaces and let  $A$  be a bounded (or merely closed and densely defined) linear operator from  $\mathcal{B}$  to  $\mathcal{K}$ . If  $z \in \mathcal{B}$  and  $b \in \mathcal{K}$  then a necessary and sufficient condition that a minimum exist for the problem

$$\begin{aligned} \min \|z\|^2 \\ \text{subject to } Az = b \end{aligned}$$

is that  $b \in \mathcal{R}(A)$ . If a minimum exists it is unique and belongs to  $\mathcal{N}^\perp(A)$ . Also, if  $\hat{z} \in \mathcal{N}^\perp(A)$  and  $A\hat{z} = b$  then  $\hat{z}$  is a minimum.

Proof: Clearly,  $b \notin \mathcal{R}(A)$  there exists no  $z \in \mathcal{B}$  satisfying  $Az = b$ . Suppose  $b \in \mathcal{R}(A)$  and let  $z$  be any solution of  $Az = b$ . Put

$$z = z_1 + z_2, \quad z_1 \in \mathcal{N}^\perp(A), \quad z_2 \in \mathcal{N}(A).$$

If  $z'$  is any other solution, put

$$z' = z'_1 + z'_2, \quad z'_1 \in \mathcal{N}^\perp(A), \quad z'_2 \in \mathcal{N}(A)$$

Then

$$Az - Az' = A(z - z') = 0$$

which implies that  $(z - z') \in \mathcal{N}(A)$ . Since

$$z - z' = (z_1 - z'_1) + (z_2 - z'_2) \in \mathcal{N}(A)$$

With  $z_1, z'_1 \in \mathcal{N}^\perp(A)$ , we must have  $z_1 = z'_1$ . That is,  $z_1$  is uniquely determined. Since  $Az_1 = b$ , and  $\|z_1\|^2 \leq \|z\|^2$ ,  $z_1$  gives the desired minimum.

Now, suppose  $\hat{z} \in \mathcal{N}^\perp(A)$  and  $A\hat{z} = b$ . Then  $\hat{z}$  must be the same as the  $z_1$  just determined. |||

Proposition 4.2: If  $A$  is as defined in Proposition 4.1, then a sufficient condition that the problem

$$\begin{aligned} \min \|z\|^2 \\ \text{subject to } Az = b \end{aligned}$$

has a solution is that  $b \in \mathcal{R}(AA^*)$ . The minimum value is then given by  $\hat{z} = A^*(AA^*)^{-1}b$ .

Proof: Recall that by  $(AA^*)^{-1}$  we really mean  $(AA^*)_{\mathcal{R}}^{-1}$  where  $(AA^*)_{\mathcal{R}}$  is the restriction of  $AA^*$  to  $\mathcal{N}^{\perp}(AA^*)$ . Since  $\mathcal{R}(AA^*) \subset \mathcal{R}(A)$  the condition is sufficient by Proposition 4.1. Since  $\mathcal{D}(AA^*) \subset \mathcal{D}(A^*)$  and  $\mathcal{R}[(AA^*)^{-1}] = \mathcal{R}(AA^*) \cap \mathcal{N}^{\perp}(AA^*)$  the expression for  $\hat{z}$  is meaningful. Note that

$$A\hat{z} = A[A^*(AA^*)^{-1}]b = b$$

From Proposition 4.1,  $\hat{z}$  provides a minimum if and only if  $\hat{z} \in \mathcal{N}^{\perp}(A) = \overline{\mathcal{R}(A^*)}$ . But  $\hat{z} \in \mathcal{R}(A^*) \subset \overline{\mathcal{R}(A^*)}$ , so  $\hat{z}$  must provide a minimum. |||

For the remainder of this Section  $\mathcal{R}'_1$  and  $\mathcal{R}'_2$  are to be finite-dimensional Euclidean spaces. For now let  $\mathcal{R}'_1$  and  $\mathcal{R}'_2$  be  $N$  and  $M$ -dimensional, respectively, so that  $y$  and  $n$  are random  $M$ -vectors and  $h$  is an  $N$ -vector. Let  $\{\phi_i\}$ ,  $i=1, \dots, N$ , be a c.o.n.s. for  $\mathcal{R}'_1$  with the special property that  $\{\phi_1, \dots, \phi_p\}$  spans  $\mathcal{N}^{\perp}(X)$ , where  $p$  ( $p \leq M, N$ ) is the rank of  $X$ . Then

$$E\|Cn\|^2 = \sum_{i=1}^N (RC^*\phi_i, C^*\phi_i).$$

Put  $c_i = C^*\phi_i$ , so that

$$E\|Cn\|^2 = \sum_{i=1}^N (Rc_i, c_i) = \sum_{i=1}^N \|R^{1/2}c_i\|^2. \quad (4.6)$$

Now the condition (4.2) that  $CXh_1 = h_1$  implies that  $CXh = CX(h_1 + h_2) = h_1$ , so  $CX$  is the orthogonal projection on  $\mathcal{R}^\perp(X)$ . But then  $CX = (CX)^* = X^*C^*$  so that

$$\begin{aligned} X^*C^*\phi_i &= CX\phi_i = \phi_i, & i=1, \dots, p \\ &= 0, & i=p+1, \dots, N \end{aligned}$$

or,

$$\begin{aligned} X^*c_i &= \phi_i, & i=1, \dots, p \\ &= 0, & i=p+1, \dots, N. \end{aligned} \quad (4.7)$$

The problem of finding the LUMV estimator  $C$  is thus equivalent to finding  $c_i$ ,  $i=1, \dots, N$ , to solve the problem:

$$\begin{aligned} &\text{minimize } \sum_{i=1}^N \|R^{1/2}c_i\|^2 \\ &\text{subject to } X^*c_i = \phi_i, & i=1, \dots, p \\ & &= 0, & i=p+1, \dots, N. \end{aligned}$$

We can individually minimize each term  $\|R^{1/2}c_i\|^2$  subject to  $X^*c_i = \phi_i$ ,  $i=1, \dots, p$ , as we shall show immediately, and we can certainly minimize each term  $\|R^{1/2}c_i\|^2$ ,  $i=p+1, \dots, N$ , by setting  $c_i = 0$ . The  $c_i$  that yields these minima will then determine the estimator  $C$ .

To minimize  $\|R^{1/2}c_i\|^2$  subject to  $X^*c_i = \phi_i$  we first put  $z_i \stackrel{d}{=} R^{1/2}c_i$ , so that  $c_i = R^{-1/2}z_i$  (which is defined since  $R$  is strictly positive definite). Then the problem is to minimize  $\|z_i\|^2$  subject to  $X^*R^{-1/2}z_i = \phi_i$ . In the finite-dimensional case which we are

considering,  $\hat{\mathcal{R}}(X^* R^{-1/2}) = \overline{\hat{\mathcal{R}}(X^* R^{-1/2})} = \hat{\mathcal{R}}(X^* R^{-1} X)$   
 $= \overline{\mathcal{R}(X^* R^{-1} X)}$ , so Proposition 4.2 applies with  $A = X^* R^{-1/2}$ ,

and

$$\begin{aligned}\hat{z}_i &= R^{-1/2} X (X^* R^{-1/2} R^{-1/2} X)^{-1} \phi_i \\ &= R^{-1/2} X (X^* R^{-1} X)^{-1} \phi_i, \quad i=1, \dots, p.\end{aligned}\quad (4.8)$$

Thus, the minimizing  $\hat{c}_i$  are

$$\begin{aligned}\hat{c}_i &= R^{-1/2} \hat{z}_i = R^{-1} X (X^* R^{-1} X)^{-1} \phi_i, \quad i=1, \dots, p \\ &= 0, \quad i=p+1, \dots, N.\end{aligned}\quad (4.9)$$

Since  $\hat{c}_i = C^* \phi_i$ ,

$$\hat{C}^* = R^{-1} X (X^* R^{-1} X)^{-1} P$$

where  $P$  is the orthogonal projection on  $\mathcal{N}^\perp(X) = \mathcal{R}(X^*)$   
 $= \mathcal{R}(X^* R^{-1} X)$ . Then

$$\begin{aligned}\hat{C} &= P (X^* R^{-1} X)^{-1} X^* R^{-1} \\ &= (X^* R^{-1} X)^{-1} X^* R^{-1}.\end{aligned}\quad (4.10)$$

This development is summarized in the following proposition.

Proposition 4.3: If  $\mathcal{Y}_1$  and  $\mathcal{Z}_2$  are finite-dimensional and  $R$  is strictly positive definite there always exists a unique LUMV estimate for  $h_1$  given by

$$\hat{h} = (X^* R^{-1} X)^{-1} X^* R^{-1} y. \quad (4.11)$$

The error variance is

$$E \|\hat{h} - h_1\|^2 = \text{Trace} [(X^* R^{-1} X)^{-1}]. \quad (4.12)$$

Proof: That  $\hat{h}_1$  as given by (4.12) is LUMV has already been proved.

The minimizing element  $\hat{z}_i$  given by (4.9) is unique by Proposition 4.1, and the uniqueness of  $\hat{C}$  and hence  $\hat{h}$  follows obviously. Finally,

from (4.5)

$$\begin{aligned} E\|\hat{h} - h_1\|^2 &= E\|\hat{C}n\|^2 = \text{Trace}(\hat{C}R\hat{C}^*) \\ &= \text{Trace}[(X^*R^{-1}X)^{-1}X^*R^{-1}X(X^*R^{-1}X)^{-1}] \\ &= \text{Trace}[(X^*R^{-1}X)^{-1}] \end{aligned}$$

where by  $(X^*R^{-1}X)^{-1}$  we mean as always the inverse of the restriction to  $\mathcal{R}(X^*R^{-1}X)$ . |||

Corollary 1: The proposition still holds if only  $\mathcal{A}_2$  is finite-dimensional.

Proof: In this case  $\mathcal{N}_\perp(X)$  is necessarily finite dimensional, and the proof is unchanged except that there are now an infinite number of  $c_i$  to be set equal to zero, corresponding to those  $\phi_i$  that lie in  $\mathcal{N}_\perp(X)$ . |||

The case where only  $\mathcal{A}_1$  is finite-dimensional is not so simple since  $R^{-1/2}$  becomes an unbounded operator. For this case it is necessary to go to the more general theorem of the Appendix.

Corollary 2: Not only is

$$\text{Trace}(\hat{C}R\hat{C}^*) \leq \text{Trace}(CRC^*) \quad (4.13)$$

where  $C$  is any unbiased estimator, but also,

$$\hat{C}R\hat{C}^* \leq CRC^* \quad (4.14)$$

in the usual ordering of nonnegative self-adjoint operators.

Proof: The inequality (4.13) is just one way of saying that  $\hat{C}$  is of minimum variance. The inequality (4.14) says that  $(\hat{C}R\hat{C}^* f) \leq (CRC^* f, f)$  for any  $f \in \mathcal{A}'_1$ . But we actually proved that  $(R\hat{C}_i, \hat{C}_i) = (\hat{C}R\hat{C}^* \phi_i, \phi_i)$  was a minimum for each  $\phi_i$  amongst the class of unbiased estimators  $C$ . Since  $\{\phi_i\}$  is a c.o.n.s. this establishes (4.14). |||

Corollary 2 says that we actually proved a little more than we set out to show. It has the following useful implication.

Corollary 3: Let  $B$  be a linear transformation from  $\mathcal{R}_1$  into  $\mathcal{R}_3$ . The unbiased linear estimate  $\hat{h}$  of  $h_1$  that minimizes  $E\|Bh_1 - \hat{h}\|^2$  is given by  $\hat{h} = h$ , the LUMV estimate. In particular,  $\hat{h}$  minimizes  $E\|Xh_1 - Xh'\|^2$  over the class of linear unbiased estimates.

Proof: Let  $h'$  be an unbiased linear estimate,  $h' = Cy$ . Then

$$\begin{aligned} CXh_1 &= h_1 \text{ and} \\ E\|Bh_1 - Bh'\|^2 &= E\|Bh_1 - BCXh_1 - BCn\|^2 \\ &= E\|BCn\|^2 \\ &= \text{Trace}(BCRC^*B^*). \end{aligned} \tag{4.15}$$

Now, for any  $f \in \mathcal{R}_1$ , by Corollary 2

$$\begin{aligned} (BCRC^*B^*f, f) &= (CRC^*B^*f, B^*f) \\ &\geq (\hat{C}R\hat{C}^*B^*f, B^*f) \\ &= (B\hat{C}R\hat{C}^*B^*f, f) \end{aligned}$$

Hence  $BCRC^*B^* \geq B\hat{C}R\hat{C}^*B^*$ , which implies that  $\text{Trace}(BCRC^*B^*) \geq \text{Trace}(B\hat{C}R\hat{C}^*B^*)$ . |||

Remark. The formula (4.11) for  $\hat{h}$  involves  $(X^*R^{-1}X)^{-1}$  which is meaningful as the inverse of the restriction of a transformation.

However it cannot be directly interpreted as a matrix inverse unless  $\mathcal{N}(X^*R^{-1}X) = 0$ . Since  $\mathcal{N}^\perp(X^*R^{-1}X) = \mathcal{R}(X^*R^{-1}X) = \mathcal{R}(X^*R^{-1/2})$ ,  $\mathcal{N}(X^*R^{-1}X) = \mathcal{N}(R^{-1/2}X) = \mathcal{N}(X)$ . If  $\mathcal{N}(X) \neq 0$ , an orthogonal transformation of coordinates in  $\mathcal{R}_1$  can be used to put the matrix representing  $X^*R^{-1}X$  in diagonal form, from which a matrix representation of (4.11) is obtained easily.

Let us now consider the problem of obtaining a recursive solution for the LUMV estimation problem in the finite-dimensional case.

We consider the system

$$\begin{aligned} y_1 &= B_1 h_1 + n_1 \stackrel{d}{=} B_1 h + n_1 \\ y_2 &= B_2 \phi_{2,1} h_1 + n_2 \stackrel{d}{=} B_2 \phi_{2,1} h + n_2 \\ y_3 &= B_3 \phi_{3,2} \phi_{2,1} h_1 + n_3 \stackrel{d}{=} B_3 \phi_{3,1} h + n_3 \\ &\vdots \\ y_n &= B_n \phi_{n,n-1} \phi_{n-1,n-2} \cdots \phi_{2,1} h_1 + n_n \stackrel{d}{=} B_n \phi_{n,1} h + n_n \end{aligned}$$

which equivalently can be written in the state variable form

$$\begin{aligned} h_{i+1} &= \phi_{i+1,i} h_i \\ y_i &= B_i h_i + n_i \quad i=1, \dots, n \\ h_1 &\stackrel{d}{=} h \end{aligned}$$

(Note that the subscripts no longer refer to projections.) We assume that each  $h_i$  is an  $m$ -vector, each  $B_i$  is a  $k \times m$  matrix, each  $\phi_{i+1,i}$  is an  $m \times m$  matrix, and each  $n_i$  is a random  $k$ -vector. The noise vectors  $n_i$  are taken to be uncorrelated, to have mean zero, and to have the same covariance operator (or matrix)  $R$ , where  $R$  is strictly positive definite. We can define  $X_1 \stackrel{d}{=} B_1$ ,  $X_2 \stackrel{d}{=} B_2 \phi_{2,1}, \dots, X_n \stackrel{d}{=} B_n \phi_{n,1}$  so that

$$\begin{aligned} y_1 &= X_1 h + n_1 \\ y_2 &= X_2 h + n_2 \\ &\vdots \\ y_n &= X_n h + n_n \end{aligned} \tag{4.16}$$

The system of equations on (4.16) can also be written in the form

$$\underline{y}_n = \underline{X}_n h + \underline{n}_n$$

where<sup>(1)</sup>

$$\underline{y}_n \stackrel{d}{=} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \underline{X}_n \stackrel{d}{=} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad \underline{n}_n = \begin{bmatrix} n_1 \\ \vdots \\ n_n \end{bmatrix}$$

The LUMV estimate for  $h$  is given by

$$\hat{\underline{h}}_n = (\underline{X}_n^* \underline{R}_n^{-1} \underline{X}_n)^{-1} \underline{X}_n^* \underline{R}_n^{-1} \underline{y}_n \quad (4.17)$$

where

$$\underline{R}_n = \begin{bmatrix} R & & & \bigcirc \\ & \ddots & & \\ & & \ddots & \\ \bigcirc & & & R \end{bmatrix}$$

$$\underline{R}_n^{-1} = \begin{bmatrix} R^{-1} & & & \bigcirc \\ & \ddots & & \\ & & \ddots & \\ \bigcirc & & & R^{-1} \end{bmatrix}$$

(1) The "partitioned matrix" notation is to be understood as a shorthand notation for defining  $\underline{y}_n$ ,  $\underline{X}_n$ ,  $\underline{n}_n$ , as in Section 3. Thus  $\underline{X}_n$  is a linear transformation from  $m$ -dimensional space into  $k \times n$ -dimensional space, defined in an obvious way. With this interpretation,  $(\underline{X}_n^* \underline{R}_n^{-1} \underline{X}_n)^{-1}$  is meaningful, as before, as the inverse of the restriction of an operator.



Put  $P_n = (X_n^* R_n^{-1} X_n)^{-1}$  and  $b_n = X_n^* R_n^{-1} y_n$ . Since

$$X_n^* R_n^{-1} = [X_1^* R^{-1} : X_2^* R^{-1} : \dots : X_n^* R^{-1}]$$

we have that

$$P_n^{-1} = P_{n-1}^{-1} + X_n^* R^{-1} X_n \quad (4.18)$$

and

$$b_n = b_{n-1} + X_n^* R^{-1} y_n \quad (4.19)$$

From equation (4.18)

$$P_n P_n^{-1} P_{n-1} = P_n [P_{n-1}^{-1} + X_n^* R^{-1} X_n] P_{n-1}$$

Hence

$$P_{n-1} = P_n + P_n X_n^* R^{-1} X_n P_{n-1} \quad (4.20)$$

and

$$P_{n-1} X_n^* R^{-1/2} = P_n X_n^* R^{-1/2} [I + R^{-1/2} X_n P_{n-1} X_n^* R^{-1/2}].$$

The inverse to the bracketed expression exists (see Section 3) and

$$\begin{aligned} P_{n-1} X_n^* R^{-1/2} [I + R^{-1/2} X_n P_{n-1} X_n^* R^{-1/2}]^{-1} R^{-1/2} X_n P_{n-1} \\ = P_n X_n^* R^{-1} X_n P_{n-1} \end{aligned} \quad (4.21)$$

Then from equations (4.20) and (4.21) one can obtain

$$P_n = P_{n-1} - P_{n-1} X_n^* R^{-1/2} [I + R^{-1/2} X_n P_{n-1} X_n^* R^{-1/2}]^{-1} R^{-1/2} X_n P_{n-1} \quad (4.22)$$

Using equations (4.19) and (4.22) in the expression for  $\hat{h}_n$  yields

$$\begin{aligned} \hat{h}_n = \left\{ P_{n-1} - P_{n-1} X_n^* R^{-1/2} [I + R^{-1/2} X_n P_{n-1} X_n^* R^{-1/2}]^{-1} \right. \\ \left. \cdot R^{-1/2} X_n P_{n-1} \right\} \left\{ b_{n-1} + X_n^* R^{-1} y_n \right\} \end{aligned}$$

$$\begin{aligned}
&= P_{n-1} b_{n-1} - P_{n-1} X_n^* R^{-1/2} [ \ ]^{-1} R^{-1/2} X_n P_{n-1} b_{n-1} \\
&\quad + P_{n-1} X_n^* R^{-1} y_n - P_{n-1} X_n^* R^{-1/2} [ \ ]^{-1} R^{-1/2} X_n P_{n-1} X_n^* R^{-1} y_n \\
&= \hat{h}_{n-1} - P_{n-1} X_n^* R^{-1/2} [ \ ]^{-1} R^{-1/2} X_n \hat{h}_{n-1} \\
&\quad + P_{n-1} X_n^* R^{-1/2} \{ I - [ \ ]^{-1} R^{-1/2} X_n P_{n-1} X_n^* R^{-1/2} \} R^{-1/2} y_n
\end{aligned}$$

For operators A such that  $[I+A]^{-1}$  exists we have

$$\{I - [I+A]^{-1}A\} = [I+A]^{-1} [I+A-A] = [I+A]^{-1}$$

Therefore,

$$\begin{aligned}
\hat{h}_n &= \hat{h}_{n-1} - P_{n-1} X_n^* R^{-1/2} [ \ ]^{-1} R^{-1/2} X_n \hat{h}_{n-1} \\
&\quad + P_{n-1} X_n^* R^{-1/2} [I + R^{-1/2} X_n P_{n-1} X_n^* R^{-1/2}]^{-1} R^{-1/2} y_n \\
&= \hat{h}_{n-1} - P_{n-1} X_n^* R^{-1/2} [I + R^{-1/2} X_n P_{n-1} X_n^* R^{-1/2}]^{-1} \\
&\quad \cdot [R^{-1/2} X_n \hat{h}_{n-1} - R^{-1/2} y_n]
\end{aligned}$$

or

$$\hat{h}_n = \hat{h}_{n-1} + K_n [y_n - X_n \hat{h}_{n-1}] \quad (4.23)$$

where

$$K_n = P_{n-1} X_n^* [R + X_n P_{n-1} X_n^*]^{-1} \quad (4.24)$$

and from (4.22)

$$P_n = P_{n-1} - K_n X_n P_{n-1} \quad (4.25)$$

Hence, to obtain the recursive LUMV estimator we start with

$$\hat{h}_1 = (X^* R^{-1} X)^{-1} X^* R^{-1} y_1 \quad (4.26)$$

and

$$P_1 = (X^* R^{-1} X)^{-1} \quad (4.27)$$

and apply the recursive relation for  $n=2, 3, \dots$ . The error variance at the  $n^{\text{th}}$  step is

$$\begin{aligned}
\mathbb{E} \|\tilde{\mathbf{h}}_n - \mathbf{h}\|^2 &= \text{Tr} \left\{ (\tilde{\mathbf{X}}_n^* \mathbf{R}_n^{-1} \tilde{\mathbf{X}}_n)^{-1} \right\} \\
&= \text{Tr} \left\{ [\mathbf{X}_1^* \mathbf{R}^{-1} \mathbf{X}_1 + \mathbf{X}_2^* \mathbf{R}^{-1} \mathbf{X}_2 \cdots \right. \\
&\quad \left. + \mathbf{X}_n^* \mathbf{R}^{-1} \mathbf{X}_n]^{-1} \right\}
\end{aligned} \tag{4.28}$$

## 5. A MODIFIED LUMV ESTIMATOR

Each term in the model

$$y = Xh + n \tag{5.1}$$

is to have the same interpretation as in Section 4. It is now supposed, however, that there is prior information that the unknown  $h$  belongs to a known bounded subset of  $\mathcal{H}_1$ . This information is used to modify the LUMV estimate  $\hat{h}$  so as to reduce the mean-squared error. The new estimate,  $\hat{\hat{h}}$ , is biased, so that its mean-squared error becomes a function of the true value of  $h$ , but its mean-squared error will be seen to be not greater than that of  $\hat{h}$ , for any  $h$ .

It will be convenient to use the term rectangular paralleloiped (r. p.), defined as follows. In a separable Hilbert space  $\mathcal{H}$  let  $\{\psi_i\}$  be a c. o. n. s. Any subset  $B$  of  $\mathcal{H}$  of the form

$$B = \{h \in \mathcal{H} : |(h, \psi_i)| \leq b_i, \quad i=1, 2, \dots\} \tag{5.2}$$

where  $\{b_i\}$  is a bounded sequence of nonnegative numbers will be called an r. p. with respect to the  $\psi_i$ .

Suppose now that  $\hat{h} = Cy$  is an LUMV estimate for  $h_1$ ; i. e., either the conditions of Proposition A. 4 are satisfied so that  $C$  is the  $S$  given by that proposition, or, in the finite-dimensional case,  $C$  is simply the  $\hat{C}$  of equation (4. 10). The operator  $CRC^*$  is self-adjoint and compact (since  $C$  is bounded and  $R^{1/2}$  is Hilbert-Schmidt), so by the spectral theorem for such operators one has

$$CRC^* \psi_i = \sigma_i^2 \psi_i, \quad i=1, 2, \dots \tag{5.3}$$

where the eigenvectors form a c. o. n. s. and the eigenvalues  $\sigma_i^2$  are real and

nonnegative (zero eigenvalues are allowed if necessary so that the set  $\{\psi_i\}$  can be complete). Furthermore, since  $CRC^*$  has finite trace,

$$\text{Trace } (CRC^*) = \sum_{i=1}^{\infty} \sigma_i^2 < \infty. \quad (5.4)$$

The LUMV estimate can be written

$$\hat{h}(y) = Cy = CXh + Cn = h_1 + Cn \quad (5.5)$$

and also of course it can be expanded in the  $\psi_i$ ,

$$\hat{h}(y) = \sum_i (\hat{h}(y), \psi_i) \psi_i. \quad (5.6)$$

Equation (5.5) can be interpreted as describing a trivial case of the linear model for which the linear transformation is I, the "observation" is  $\hat{h}(y)$  and the "noise" is Cn. Note that  $\mathcal{H}_1$  and  $\mathcal{H}_2$  have both been replaced by  $\mathcal{H}^\perp(X)$ . It can easily be shown that, even in the infinite-dimensional case, the LUMV estimate for  $h_1$  in the model (5.5) is just the observation. Thus there is no ambiguity in writing  $\hat{h}(y)$  for the observation. The covariance  $\Gamma$  of the new noise Cn is the operator  $CRC^*$ ; in fact,

$$\begin{aligned} (\Gamma u, v) &= E(u, Cn)(Cn, v) \\ &= E(C^* u, n)(n, C^* v) \\ &= (RC^* u, C^* v) = (CRC^* u, v). \end{aligned} \quad (5.7)$$

It follows that

$$E(\psi_i, Cn)(Cn, \psi_j) = (CRC^* \psi_i, \psi_j) = \sigma_i^2 \delta_{ij}. \quad (5.8)$$

Now, starting from the model (5.5) we consider a completely arbitrary linear estimate  $\tilde{h}(y)$  of  $h_1$  and expand it in terms of the  $\psi_i$ .

Thus,

$$\tilde{h}(y) = \sum_{i,j} a_{ij} (\hat{h}(y), \psi_j) \psi_i \quad (5.9)$$

where the  $a_{ij}$  are real numbers. The formula (5.6) is of course a special case of (5.9). The error in  $\tilde{h}(y)$  is

$$\tilde{h}(y) - h_1 = \sum_i [(a_{ii} - 1)(h_1, \psi_i) + \sum_{\substack{j \\ j \neq i}} a_{ij}(h_1, \psi_j) + \sum_j a_{ij}(Cn, \psi_j)] \psi_i. \quad (5.10)$$

From (5.8) and the fact  $Cn$  has mean zero it follows that the mean-squared error is given by

$$E\|\tilde{h}(y) - h_1\|^2 = \sum_i |(a_{ii} - 1)(h_1, \psi_i) + \sum_{\substack{j \\ j \neq i}} a_{ij}(h_1, \psi_j)|^2 + \sum_{i,j} a_{ij}^2 \sigma_j^2. \quad (5.11)$$

The equations (5.9), (5.10), (5.11) still apply to an arbitrary unknown  $h_1$  and to an arbitrary linear estimate. Now we impose the condition that  $h_1$  belong to a known bounded subset of  $\mathcal{H}^\perp(X)$ . Then there is a set  $\mathcal{B}$  which is an r. p. with respect to the  $\psi_i$  such that  $h_1$  must belong to  $\mathcal{B}$  (of course, the "fit" may or may not be very good) and we have  $|(h_1, \psi_i)| \leq b_i, i=1, 2, \dots$

From (5.11) it then follows that

$$E\|\tilde{h}(y) - h_1\|^2 \leq \sum_i [ |a_{ii} - 1| b_i + \sum_{\substack{j \\ j \neq i}} |a_{ij}| b_j ]^2 + \sum_{i,j} a_{ij}^2 \sigma_j^2 \quad (5.12)$$

for all  $h_1 \in \mathcal{B}$ . The upper bound given by (5.12) is minimized by putting  $a_{ij} = 0, i \neq j$ , and  $a_{ii} = b_i^2 / (b_i^2 + \sigma_i^2)$ . The linear estimate that results when these values for  $a_{ij}$  are used is,

$$\hat{h}(y) = \sum_j \frac{b_j^2}{b_j^2 + \sigma_j^2} (Cy, \psi_j) \psi_j, \quad (5.13)$$

and from (5.12),

$$\begin{aligned}
E\|\hat{h}(y)-h_1\|^2 &\leq \sum_i \left( \left| \frac{b_i^2}{b_i^2+\sigma_i^2} - 1 \right| b_i \right)^2 + \sum_i \left( \frac{b_i^2}{b_i^2+\sigma_i^2} \right)^2 \sigma_i^2 \\
&= \sum_i \frac{\sigma_i^2 b_i^2}{b_i^2+\sigma_i^2}
\end{aligned} \tag{5.14}$$

From (5.4) we have that

$$E\|\hat{h}(y)-h_1\|^2 = \text{Tr}(\text{CRC}^*) = \sum_i \sigma_i^2, \tag{5.15}$$

and comparison with (5.14) then gives

$$E\|\hat{h}(y) - h_1\|^2 \leq E\|\hat{h}(y) - h_1\|^2$$

for all  $h_1 \in \mathcal{B}$ .

It cannot be claimed from what has been done that  $\hat{h}(y)$  is optimum in any very meaningful **sense**, because the upper bound on mean-squared error given by (5.12) may not be tight, and because the elementary minimization carried out was on an upper bound, not on the error itself. Still,  $\hat{h}(y)$  would appear to be a good estimate. Since  $\hat{h}(y)$  is a linear estimate it necessarily will yield values for some observations  $y$  that do not belong to  $\mathcal{B}$ . It can be improved, as can  $h(y)$ , by being truncated to  $\mathcal{B}$ , but this of course makes it nonlinear. A truncated estimate  $\hat{h}_b$  that has uniformly smaller error variance than  $\hat{h}$  is defined by:

$$\hat{h}_b(y) = \sum_i \alpha_i \psi_i \tag{5.16}$$

where,

$$\alpha_i = \left\{ \begin{array}{l} b_j \text{ if } \frac{b_j^2}{b_j^2 + \sigma_j^2} (Cy, b_j) \geq b_j \\ - b_j \text{ if } \frac{b_j^2}{b_j^2 + \sigma_j^2} (Cy, b_j) \leq - b_j \\ \frac{b_j^2}{b_j^2 + \sigma_j^2} \text{ otherwise.} \end{array} \right.$$

The estimate  $\hat{h}$  is discussed in [R-2] and [R-3], where an application is made to identification.



## APPENDIX

In this Appendix we obtain a solution to an infinite-dimensional LUMV estimation problem in a Hilbert space setting. A proof that is fundamentally the same but different in detail is given in [B-1], where the problem is attached to a systematic study of pseudoinverses.

In what follows the operator  $A$  is required only to be closed and densely defined, instead of closed and bounded as in most of what precedes. If  $A$  is closed and densely defined then it is readily verified that  $\mathcal{N}(A)$  is a closed linear subspace, and also that  $A$  is densely defined in  $\mathcal{N}^\perp(A)$ . We do not give specific references, but the operator-theoretic background for all that follows is in [R-1]. To begin with we again note that if  $A$  is only closed and densely defined, the conclusions of Propositions 4.1 and 4.2 are still valid since the respective proofs do not require that  $A$  be bounded, but merely that  $\mathcal{N}(A)$  be closed.

Before going further, we need a preparatory lemma:

Proposition A.1: If  $A$  is a closed, densely defined operator from  $\mathcal{H}_1$  to  $\mathcal{H}_2$ , then  $AA^*$  restricted to  $\overline{\mathcal{R}(AA^*)}$  is self-adjoint and 1:1, so its inverse exists and is self-adjoint.

Proof: We have the fact that  $AA^* = A^{**}A^*$  is self-adjoint and hence closed and densely defined in  $\mathcal{H}_2$ , so we can write

$$\mathcal{H}_2 = \overline{\mathcal{R}(AA^*)} \oplus \mathcal{N}(AA^*)$$

Let  $f \in \mathcal{D}(AA^*)$  and write

$$f = f_1 + f_2, \quad f_1 \in \overline{\mathcal{R}(AA^*)}, \quad f_2 \in \mathcal{N}(AA^*)$$

Then  $f_1 \in \mathcal{D}(AA^*)$  and  $g \stackrel{d}{=} AA^* f_1$  belongs to  $\overline{\mathcal{R}(AA^*)} \subset \mathcal{R}(AA^*)$ . Hence  $AA^*$  provides a 1:1 mapping from  $\mathcal{D}(AA^*) \cap \overline{\mathcal{R}(AA^*)}$  onto

$\mathcal{R}(AA^*) \subset \overline{\mathcal{R}(AA^*)}$ . Thus  $(AA^*)^{-1}$  is defined on  $\mathcal{R}(AA^*)$ , which is dense in  $\overline{\mathcal{R}(AA^*)}$ , and  $\mathcal{D}[(AA^*)^{-1}] = \mathcal{D}(AA^*) \cap \overline{\mathcal{R}(AA^*)}$ .  $(AA^*)^{-1}$  will always mean the inverse of  $AA^*$  restricted to  $\overline{\mathcal{R}(AA^*)}$  as indicated.

Let  $T$  be the restriction of  $AA^*$  to  $\overline{\mathcal{R}(AA^*)}$ . Then  $T^{-1} = (AA^*)^{-1}$  as defined above.  $T$  is also closed and densely defined. Let  $f$  and  $g \in \mathcal{D}(T) = \mathcal{D}(AA^*) \cap \overline{\mathcal{R}(AA^*)}$ .  $T$  is symmetric since  $(Tf, g) = (f, Tg)$ . Thus  $T \subset T^*$ . We now want to show that  $T \supset T^*$ . We can start by observing that

$$(Tf, g) = (f, T^*g)$$

for all  $f \in \mathcal{D}(T)$ ,  $g \in \mathcal{D}(T^*)$ . Also, if  $f \in \mathcal{D}(T)$

$$(Tf, g) = (AA^*f, g)$$

Let  $f'$  be an arbitrary element in  $\mathcal{D}(AA^*)$  and write

$$f' = f'_1 + f'_2, \quad f'_1 \in \mathcal{D}(T), \quad f'_2 \in \mathcal{N}(AA^*).$$

Then

$$(AA^*f', g) = (AA^*f'_1, g) = (Tf'_1, g).$$

Suppose  $(Tf, g) = (f, g^*)$  holds for some  $g \notin \mathcal{D}(T)$  and  $g^* \in \overline{\mathcal{R}(AA^*)}$ . Since we are working in the Hilbert space  $\overline{\mathcal{R}(AA^*)}$ , we must have  $g \in \overline{\mathcal{R}(AA^*)}$  for every  $f \in \mathcal{D}(T)$ . Then for every  $f' \in \mathcal{D}(AA^*)$ ,

$$\begin{aligned} (AA^*f', g) &= (AA^*f'_1, g) = (Tf'_1, g) = (f'_1, g^*) \\ &= (f' - f'_2, g^*) \\ &= (f', g^*) - (f'_2, g^*). \end{aligned}$$

Since  $g^* \in \overline{\mathcal{R}(AA^*)}$  and  $\overline{\mathcal{R}(AA^*)} = \mathcal{N}(AA^*)^\perp$ , we must have

$$(f'_2, g^*) = 0$$

Thus

$$(AA^*f', g) = (f', g^*)$$

where  $g \notin \mathcal{D}(T)$ , but  $g \in \overline{\mathcal{R}(AA^*)} = \mathcal{N}^\perp(AA^*)$  for all  $f'$ . But since  $\mathcal{D}(T) = \mathcal{D}(AA^*) \cap \overline{\mathcal{R}(AA^*)}$ , we must have  $g \notin \mathcal{D}(AA^*)$  which is a contradiction since  $AA^*$  is self-adjoint. |||

The covariance operator  $R$  for the Hilbert-space-valued noise  $r$  in equation (4.1) is a bounded self-adjoint, nonnegative definite operator on  $\mathcal{H}_2$ , which, as mentioned before, is assumed in addition to be strictly positive definite. For further discussion see [B-1, Appendix B]. We note that  $R^{1/2}$  exists and is self-adjoint, and that  $(R^{1/2})^{-1}$  also exists and is self-adjoint (hence closed and densely defined).  $(R^{1/2})^{-1}$  may be unbounded. In fact we eventually make one more requirement on  $R$ , that  $R^{1/2}$  be Hilbert-Schmidt (so that  $R$  is trace class), which ensures that  $(R^{1/2})^{-1}$  is unbounded. Since  $(R^{1/2})^{-1} = (R^{-1})^{1/2}$  we can write  $R^{-1/2} = (R^{1/2})^{-1}$ .

In the material to follow we refer to three related minimization problems, listed below as (A), (B) and (C). Solving problem (B) is the essential step in getting the LUMV estimator. The problems are:

$$\left. \begin{array}{l} \min \|z\|^2 \\ \text{subject to } Az = b \end{array} \right\} \quad (\text{A})$$

$$\left. \begin{array}{l} \min (R\xi, \xi) \\ \text{subject to } D\xi = b \end{array} \right\} \quad (\text{B})$$

$$\left. \begin{array}{l} \min \|z\|^2 \\ \text{subject to } DR^{-1/2}z = b \end{array} \right\} \quad (\text{C})$$

where  $R$  is a bounded, self-adjoint, strictly positive definite operator and  $D$  is a bounded operator from  $\mathcal{H}_2$  to  $\mathcal{H}_1$ . Note that (B) is equivalent to the problem

$$\min \|R^{1/2}\xi\|^2$$

subject to  $D\xi = b$ .

Proposition A.2: Problems (B) and (C) are equivalent in that if either has a solution, so does the other. If  $\xi_0$  and  $z_0$  are the respective solutions, they are related by

$$z_0 = R^{1/2}\xi_0, \quad \xi_0 = R^{-1/2}z_0.$$

Proof: Suppose (B) has a solution  $\xi_0$ . Put  $z_0 = R^{1/2}\xi_0$ . Then

$$DR^{-1/2}z_0 = DR^{-1/2}R^{1/2}\xi_0 = D\xi_0 = b.$$

Suppose  $z$  also satisfies  $DR^{-1/2}z = b$ . Put  $\xi = R^{-1/2}z$ , which is certainly defined if  $DR^{-1/2}z$  is, and note that  $D\xi = b$ . Then

$$\|z_0\|^2 = \|R^{1/2}\xi_0\|^2 \leq \|R^{1/2}\xi\|^2 = \|z\|^2.$$

So  $z_0 = R^{1/2}\xi_0$  provides a solution for (C). Now, suppose (C) has a solution  $z_0$ . Since  $R^{-1/2}z_0$  is defined, we may put  $\xi_0 = R^{-1/2}z_0$ . Then

$$D\xi_0 = DR^{-1/2}z_0 = b.$$

Suppose  $\xi$  also satisfies  $D\xi = b$ . Put  $z = R^{1/2}\xi$ . Then

$$DR^{-1/2}z = DR^{-1/2}R^{1/2}\xi = b$$

and  $\|R^{1/2}\xi_0\|^2 = \|z_0\|^2 \leq \|z\|^2 = \|R^{1/2}\xi\|^2$

so  $\xi_0 = R^{-1/2}z_0$  provides a solution for (B). |||

Let us now try to relate problem (C) to problem (A). From the facts that  $R^{-1/2}$  is densely defined and  $\mathcal{D}(D) = \mathcal{A}'_2$  we have that  $DR^{-1/2}$  is densely defined; however it need not be closed. On the other hand,  $R^{-1/2}D^*$  need not be densely defined, but it is closed since  $R^{-1/2}$  is closed and  $D^*$  is bounded. We shall have to introduce some additional hypotheses and we choose to require that  $R^{-1/2}D^*$  be densely defined (this is restrictive, but not too much so, see [B-1, Section 2]). Then  $R^{-1/2}D^*$  is closed and densely

defined, so  $A = (R^{-1/2}D^*)^*$  exists and is closed and densely defined, and  $A^* = R^{-1/2}D^*$ .  $A$  is a closed linear extension of  $DR^{-1/2}$  (the condition has effectively guaranteed that  $DR^{-1/2}$  has a closed extension).

Suppose now we consider problem (A) with the operator  $A$  as just defined; this problem can be regarded as a modification of (C). From Proposition 4.1 we know that (A) has a unique solution if  $b \in \mathcal{R}(A)$ . Call this solution  $z_0$  a weak solution for (C). Problem (C) may not actually have a solution because  $DR^{-1/2}z_0$  may not be defined. If a weak solution  $z_0$  happens to be contained in  $\mathcal{D}(DR^{-1/2})$ , then problems (C) and (B) do have proper solutions with  $\xi_0 = R^{-1/2}z_0$ .

From Proposition 4.2 we know that if  $b \in \mathcal{R}(AA^*)$ , then a solution exists to (A) with

$$z_0 = A^*(AA^*)^{-1}b.$$

Furthermore, if  $R^{-1/2}z_0$  is defined, then (B) has a solution

$$\xi_0 = R^{-1/2}z_0 = R^{-1/2}A^*(AA^*)^{-1}b.$$

In a sense this is the answer for problem (B) when the only condition imposed is that  $R^{-1/2}D^*$  be densely defined. It is not satisfactory, however, because a weak solution is not guaranteed to exist or to be a solution when it does exist, and even a weak solution may not be given by the formula if  $\mathcal{R}(AA^*)$  is not closed.

Thus, in order to be sure of getting a solution for problem (B) that holds for all  $b \in \mathcal{N}_1$  we need to introduce further conditions that will make  $\mathcal{R}(A)$  closed (and hence,  $\mathcal{R}(AA^*)$  closed) and will guarantee that  $z_0 \in \mathcal{D}(R^{-1/2})$ .

We have:

Proposition A.3: Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be Hilbert spaces. Assume that:

(a)  $R$  is a bounded, self-adjoint, strictly positive definite operator on  $\mathcal{H}_2$ .

(b)  $D$  is a bounded linear operator from  $\mathcal{H}_2$  to  $\mathcal{H}_1$  with  $\mathcal{D}(D) = \mathcal{H}_2$  and

$$\mathcal{R}(D) = \overline{\mathcal{R}(D)}.$$

(c)  $R^{-1}D^*$  is densely defined on  $\mathcal{H}_1$ .

(d) With  $A = (R^{-1/2}D^*)^*$ , the operator  $(AA^*)^{-1}AR^{-1/2}$  is bounded.

Then  $(AA^*)^{-1}AR^{-1/2}$  has a continuous extension  $S$  to all of  $\mathcal{H}_2$ , and problem

(B) is solved by  $\xi_0 = S^*b$ . This solution is unique. Furthermore,  $\mathcal{N}(S^*)$

$$= \mathcal{N}(D^*).$$

Proof: We already have by Proposition A.1 that  $(AA^*)^{-1}$  (interpreted as the restriction to  $\overline{\mathcal{R}(AA^*)}$ ) is self-adjoint. From (a) and (b) it follows that  $(AA^*)^{-1}$  is a bounded operator on the subspace  $\overline{\mathcal{R}(AA^*)}$ . In fact, since  $D$  has closed range, so also does  $D^*$ . Then  $D^*$  restricted to  $\mathcal{N}^\perp(D^*)$  is bounded from below, and  $R^{-1/2}D^*$  is bounded from below on  $\mathcal{N}^\perp(D^*)$ . But  $\mathcal{N}(R^{-1/2}D^*) = \mathcal{N}(D^*)$  since  $R^{-1/2}D^*y = 0$  implies  $D^*y = 0$ . Thus  $A^* = R^{-1/2}D^*$  restricted to  $\mathcal{N}^\perp(R^{-1/2}D^*)$  is bounded from below and hence has closed range. This implies that  $A$  and  $AA^*$  have closed range (in fact  $\mathcal{R}(A) = \mathcal{R}(AA^*) = \overline{\mathcal{R}(A)}$ ) which in turn implies that  $(AA^*)^{-1}$  is bounded. Since it is closed and has domain dense in  $\overline{\mathcal{R}(AA^*)}$ , it is defined everywhere on  $\overline{\mathcal{R}(AA^*)}$ .

Now,

$$(AA^*)^{-1}AR^{-1/2} \supset (AA^*)^{-1}DR^{-1/2}R^{-1/2} = (AA^*)^{-1}DR^{-1}.$$

But  $\mathcal{D}(DR^{-1})$  is dense in  $\mathcal{H}_2$ , and  $\mathcal{R}(DR^{-1}) \subset \mathcal{R}(DR^{-1/2}) \subset \mathcal{R}(A)$ , so

$(AA^*)^{-1}AR^{-1/2}$  is densely defined. By (d), it then has a continuous extension

$S$  to all of  $\mathcal{H}_2$ .

We show that  $SD^*$ , which is a bounded operator on  $\mathcal{H}_2$ , is in fact the orthogonal projection on  $\overline{\mathcal{R}(AA^*)}$ . We have that

$$SD^* \supset (AA^*)^{-1}AR^{-1/2}D^*, \supset (AA^*)^{-1}DR^{-1/2}R^{-1/2}D^*$$

and that the domain of the operator on the right is equal to  $\mathcal{D}(R^{-1}D^*)$ , which is dense in  $\mathcal{H}_1$  by (c). Let  $g \in \mathcal{D}(R^{-1}D^*) \cap \mathcal{D}(R^{-1/2}D^*)$ , then

$$\begin{aligned} SD^*g &= (AA^*)^{-1}AR^{-1/2}D^*g \\ &= (AA^*)^{-1}AA^*g = g_0 \end{aligned}$$

where  $g_0$  is the projection of  $g$  on  $\mathcal{N}^\perp(AA^*) = \overline{\mathcal{R}(AA^*)}$ . Since  $SD^*$  agrees with the orthogonal projection in question on a dense subset,  $SD^*$  is equal to it. Then  $SD^*$  is self-adjoint, so  $DS^*$  is also the orthogonal projection on  $\overline{\mathcal{R}(AA^*)}$ .

To show that  $\xi_0 = S^*b$  solves (B) we show that  $z_0 = R^{1/2}\xi_0$  solves (C), and invoke Proposition A.2. Clearly,

$$DR^{-1/2}z_0 = DR^{-1/2}R^{1/2}\xi_0 = DS^*b = b,$$

so it remains to show that  $z_0$  provides a minimum. This will be established if it can be shown that  $z_0 \in \mathcal{N}^\perp(A)$ . For then, by Proposition 4.1,  $z_0$  will minimize  $\|z\|^2$  subject to  $Az = b$  and a fortiori  $z_0$  will minimize  $\|z\|^2$  subject to  $DR^{-1/2}z = b$ . So let  $y \in \mathcal{N}(A)$ ; then since  $(AA^*)^{-1}Ay = (AA^*)^{-1}AR^{-1/2}R^{1/2}y = SR^{1/2}y$ , it follows that  $SR^{1/2}y = 0$ . Hence

$$(z_0, y) = (R^{1/2}S^*b, y) = (b, SR^{1/2}y) = 0$$

so that  $z_0 \in \mathcal{N}^\perp(A)$ .

The uniqueness of  $z_0$  follows also, a fortiori, from the uniqueness guaranteed by Proposition 4.1.

Finally, we verify that  $\mathcal{N}(S^*) = \mathcal{N}(D^*)$ . The bounded, densely defined operator  $(AA^*)^{-1}AR^{-1/2}$  is a mapping from  $\mathcal{H}_2$  into  $\overline{\mathcal{R}(AA^*)}$ , since  $\mathcal{R}[(AA^*)^{-1}] = \mathcal{D}(AA^*) \subset \overline{\mathcal{R}(AA^*)}$ . But we have seen earlier that  $\overline{\mathcal{R}(AA^*)} = \overline{\mathcal{R}(A)}$ , and also that  $\mathcal{N}(A^*) = \mathcal{N}(D^*)$ . Consequently,  $\overline{\mathcal{R}(AA^*)} = \overline{\mathcal{R}(A)} = \mathcal{N}^\perp(A^*) = \mathcal{N}^\perp(D^*)$ . Since  $S$  is the continuous extension of  $(AA^*)^{-1}AR^{-1/2}$ , it follows that  $\overline{\mathcal{R}(S)} = \mathcal{N}^\perp(S^*) \subset \mathcal{N}^\perp(D^*)$ , or that  $\mathcal{N}(S^*) \supset \mathcal{N}(D^*)$ . On the other hand  $\overline{\mathcal{R}(S)} \supset \overline{\mathcal{R}(AA^*)} = \mathcal{N}^\perp(D^*)$  since  $SD^*$  is the orthogonal projection on  $\overline{\mathcal{R}(AA^*)}$ . Thus,  $\mathcal{N}(S^*) \subset \mathcal{N}(D^*)$ , which establishes the equality. |||

From the solution to problem (B) as just provided, the LUMV estimator is readily obtained, almost exactly as in the finite-dimensional case discussed in Section 4.

Proposition A. 4: Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be separable Hilbert spaces, and let

$$y = Xh + n$$

where  $h \in \mathcal{H}_1$ ,  $X$  is a bounded operator from  $\mathcal{H}_1$  into  $\mathcal{H}_2$ , and  $n$  is an  $\mathcal{H}_2$ -valued random variable with mean zero and covariance operator  $R$ . With the correspondences  $R \sim R$ ,  $X \sim D^*$ , assume that the conditions (a), (b), (c), (d) of Proposition A. 3 are in force and also that  $R^{1/2}$  is Hilbert-Schmidt.

Then a unique LUMV estimate for  $h_1$  exists and is given by

$$\hat{h}(y) = Sy$$

where  $S$  is the operator defined in Proposition A. 3. The variance of  $\hat{h}(y)$ , i. e.  $E\|\hat{h}(y) - h_1\|^2$ , is equal to  $\text{Trace}(SRS^*)$ .

Proof: Let  $\{\phi_i\}$ ,  $i=1,2,\dots$ , be a c. o. n. s. for  $\mathcal{H}_1$  with the special property that a subsequence of the  $\phi_i$  spans  $\mathcal{N}^\perp(X)$  and the remaining  $\phi_i$ , of course, span  $\mathcal{N}(X)$ . Let  $C$  be an unbiased estimator ( $C$  must be a bounded linear



operator as discussed in Section 4). Then, with  $\tilde{h}(y) = Cy$ , we have as in

(4.5)

$$E\|\tilde{h}(y) - h_1\|^2 = E\|Cn\|^2 = \sum_i (CRC^* \phi_i, \phi_i) < \infty,$$

where  $C$  must satisfy  $CXh_1 = h_1$  for all  $h_1 \in \mathcal{N}^\perp(X)$ . We wish to find

$C = \hat{C}$  such that  $E\|Cn\|^2$  is a minimum. Again we put  $c_i = C^* \phi_i$  and deduce,

as in Section 4, that the problem of finding the LUMV estimator is equivalent

to finding the  $c_i = \hat{c}_i$  that solve the problem:

$$\begin{aligned} &\text{minimize } \sum_i \|R^{1/2} c_i\|^2 \\ &\text{subject to } X^* c_i = \phi_i, \quad \phi_i \in \mathcal{N}^\perp(X) \\ &\quad \quad \quad = 0, \quad \phi_i \in \mathcal{N}(X). \end{aligned}$$

As in the finite-dimensional case, this problem can be solved by minimizing

individually each term  $\|R^{1/2} c_i\|^2$  subject to either  $X^* c_i = \phi_i$  or to

$X^* c_i = 0$ , respectively as  $\phi_i \in \mathcal{N}^\perp(X)$  or  $\phi_i \in \mathcal{N}(X)$ . In fact, for the latter

case we take  $c_i = 0$ , and for the former we apply Proposition A.3. Thus,

the minimizing  $\hat{c}_i$  are

$$\begin{aligned} \hat{c}_i &= S^* \phi_i, & \phi_i \in \mathcal{N}^\perp(X) \\ &= 0, & \phi_i \in \mathcal{N}(X). \end{aligned}$$

But from Proposition A.3 we know also that  $\mathcal{N}(S^*) = \mathcal{N}(X)$ ; hence we can

write simply  $\hat{c}_i = S^* \phi_i$  for all  $i=1, 2, \dots$ . Since  $\hat{c}_i = \hat{C}^* \phi_i$ ,  $\hat{C} = S$  as claimed,

and

$$E\|h(y) - h_1\|^2 = \sum_i (SRS^* \phi_i, \phi_i) = \text{Trace}(SRS^*). \quad |||$$

Remark:  $SRS^* = (AA^*)^{-1}P$ , where  $P$  is the orthogonal projection on  $\mathcal{R}(AA^*)$ . In fact,

$$\begin{aligned} SRS^* &\supset (AA^*)^{-1}AR^{-1/2}RS^* = (AA^*)^{-1}AR^{1/2}S^* \\ &\supset (AA^*)^{-1}X^*R^{-1/2}R^{1/2}S^* = (AA^*)^{-1}X^*S^*. \end{aligned}$$

But  $(AA^*)^{-1}X^*S^*$  is everywhere defined, so  $SRS^* = (AA^*)^{-1}X^*S^*$ . Since  $X^*S^* = P$ , the assertion follows.

Remark: It should be noted that Corollaries 2 and 3 to Proposition 4.3 still hold in the infinite-dimensional case.

## REFERENCES

- [A-1] N. I. Akhiezer & I. M. Glazman, Theory of Linear Operators in Hilbert Space, vols. I and II (tr. by M. Nestell), Frederick Ungar Publishing Co., New York, 1961 and 1963.
- [A-2] A. Albert, Regression and the Moore-Penrose Pseudoinverse, Academic Press, New York, 1972.
- [B-1] F. J. Beutler & W. L. Root, "The operator pseudoinverse in control and systems identification," Proc. of the Advanced Seminar on Generalized Inverses and Applications, Mathematics Research Center, University of Wisconsin, 1973.
- [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.
- [R-1] F. Riesz & B. Sz. -Nagy, Functional Analysis (tr. by L. Boron), Frederick Ungar Publishing Co., New York, 1955.
- [R-2] W. L. Root, "On the modelling and estimation of communication channels," Multivariate Analysis-III, (ed. R. Krishnaiah) Academic Press, New York, 1973.
- [R-3] W. L. Root, "Estimation in identification theory," Proc. Ninth Annual Allerton Conference on Circuit and System Theory, pp. 1-9, 1971.

UNIVERSITY OF MICHIGAN



3 9015 02826 5760