DEPARTMENT OF MATHEMATICS

CLEMSON UNIVERSITY
Clemson, South Carolina

Clemson University
Department of Mathematics

APPLICATION OF GENERALIZED EXPONENTIAL SMOOTHING AND MATRIX PSEUDOINVERSES TO ESTIMATION IN DYNAMIC SYSTEMS

March 1969

Final Report on Contract NAS 8-11259
Advanced Studies Office
Astrionics Laboratory
Marshall Space Flight Center
National Aeronautics and Space Administration
Huntsville, Alabama 35812

CLEMSON UNIVERSITY INVESTIGATORS

W. B. Owen

B. J. Prochaska

TABLE OF CONTENTS

ABSTRACT

The investigating members of the Mathematics Department of Clemson University, under an extension of NASA Contract NAS 8-11259, studied applications of pseudoinverses of matrices to the statistical filter problem. A new Kalman-type filter is presented based on these studies.

SUMMARY

A review of the filtering problem and the Kalman solution are presented. The assumptions are emphasized.

To obtain a new Kalman-type filter, generalized exponential smoothing is presented and the relevant results of pseudoinverses of matrices are reviewed.

The new filtering equations are introduced and briefly studied with respect to unbiasness.

LIST OF SYMBOLS BY SECTIONS

Section 2

t_n	times
X_n	state vector
Y_n	observation vector
$\phi_{n+1,n}$	transition matrix
M_n	mapping matrix
U_n	error vector
Q_n	covariance matrix of U_n
$E[\ ]$	mathematical expectation
$\hat{X}_n$	initial estimate of X_n
P_n	covariance matrix of $\hat{X}_n$
$\hat{X}^*_n$	an "updated" estimate of X_n
P^*_n	covariance matrix of $\hat{X}^*_n$
$\hat{Y}_n$	prediction of Y_n
ψ	matrix

Section 3

$x(n)$	univariate stochastic process
μ	mean of $x(i)$
$\hat{x}(n)$	estimate of μ
α	constant
$e(n)$	estimation error
σ_n^2	variance of $x(n)$
$k(n)$	sequence of real numbers
γ_n	sequence of coefficients

X_n		p-variate stochastic process
α_i		constants
$k_i(n)$		sequences of real numbers
$E(n)$		estimation error, a vector

Section 4

Y	$m \times 1$	vector of observables
M	$m \times p$	matrix of known constants
X	$p \times 1$	vector of unknowns
M^+		unique pseudoinverse of M
C'		the transpose of the matrix C
$r(A)$		rank of the matrix A
D		$n \times n$ diagonal matrix
I_m		the identity matrix of order m
$\hat{X}$		any solution to $MX = Y$
Z		an arbitrary $p \times 1$ vector
e		unobservable vector of random deviations
X_0		Best Approximate Solution (BAS) to system of equations $MX = Y$
$tr(A)$		the sum of the main diagonal elements of the square matrix A.

Section 5

X_n	$p \times 1$ state vector
Y_n	$q_n \times 1$ observation vector
M_n	mapping matrix

$\hat{X}^{-}{}_{n}$	sequence of estimators of X_n
A_n	matrix
$k_i(n)$	sequences of real numbers
α_i	constants
$\tilde{X}_n$	matrix
$M^{+}{}_{n}$	pseudoinverse of M_n
I	identity matrix
Z_n	arbitrary vector
$\hat{X}_n$	estimates of X_n
$\phi_{n+1,n}$	transition matrix
$\hat{Y}_n$	predictor of Y_n
Y	vector
M	matrix
X	vector
M^1	submatrix of M
X^1	subvector of X
M^*	submatrix of M
X^*	subvector of X
$\hat{X}^{*}{}_{n}$	subvector of $\hat{X}_n$

Section 1: Introduction

Partly because of space navigation and guidance problems, estimation of the state vector in a dynamic system has become very important. An estimation procedure due to R. E. Kalman [1] has been used extensively in recent years. This statistical "filter" requires the prior knowledge of the values of many statistical parameters (variances and covariances). By assuming exact knowledge of the dynamic system and of the parameters, Kalman proved that his procedure was "best" in a certain statistical sense.

In some situations it is unreasonable to assume that one knows the values of the parameters required in the Kalman filter. Furthermore, accurate estimation of the parameters may be impossible. This paper introduces and briefly studies a new filtering scheme which combines generalized exponential smoothing and the pseudoinverse of a matrix. The new procedure requires fewer assumptions than does the Kalman filter. In particular, none of the parameters have to be known. Obviously, one would not expect the accuracy of this new method to be as great as that of the Kalman filter, but preliminary studies on an elementary dynamic model are encouraging.

The main disadvantage of this new method of filtering is that of exponential smoothing (EWMA) in general. The smoothing constant must be determined numerically. However, in the analysis of economic time series it has been found that the smoothing constant is relatively stable in the sense that small amounts of new data generally do not significantly affect the value of the constant. The main use of EWMA is in short term forcasting which is frequently the role of the Kalman filter. Past records could also be used to estimate the constants.

Section 2 of this report reviews one useful form of the Kalman filter as presented by Solloway [2]. Generalized exponential smoothing is developed in Section 3. Pertinent results from the theory of matrix pseudoinverses are discussed in Section 4. The general form of the new filter is presented and studied in Section 5.

Section 2: Review of Kalman Filter

Kalman's statistical filter is a particular formulation of a solution to the Weiner problem [1]. These results appeared in [1] for a particular dynamic system--one with random error in the difference equation of a discrete dynamic system, but without random errors in the observation equations. Solloway [2] presents similar results for a system with undisturbed difference equations and observation equations with additive random error. Solloway's model is more applicable to free-flight trajectory analysis and is thus presented below.

Let observations be made at times $t_0 < t_1 < \cdot \cdot \cdot < t_n \cdot \cdot \cdot$. A discrete dynamic system can be represented by

$$X_{n+1} = \Phi_{n+1,n} X_n \tag{2.1}$$

and

$$Y_n = M_n X_n + U_n, \tag{2.2}$$

where

X_n is the <u>state vector</u> at time t_n,

Y_n is the <u>observation vector</u> at time t_n,

$\Phi_{n+1,n}$ is the <u>transition matrix</u> for the interval $[t_n, t_{n+1}]$,

M_n is the <u>mapping matrix</u> at time t_n,

and

U_n is an <u>error vector</u> with zero mean and <u>covariance matrix</u> Q_n.

Equation 2.1 illustrates the dynamics of the system, in that it shows how to obtain the state vector at time t_{n+1}, from the state vector at time t_n. Unfortunately X_n is not usually known and therefore must be estimated. Equation 2.2 illustrates the mathematical relationship between the observables at time t_n, Y_n, and the state vector at time t_n, X_n.

The mathematical theory of dynamic systems and their connection with differential equations have been studied extensively. For further references see, for instance, [3], [4], and [5].

The sequential estimation procedure derived by the Kalman method assumes that:

(i) $$E[U_n \; U'_m] = 0 \qquad n \neq m$$

(The prime notation on a matrix is used to indicate the matrix transpose);

(ii) An initial estimate of X_0, denoted by $\hat{X}_0$, is available;

(iii) The initial estimate is unbiased and has known covariance matrix P_0;

(iv) The covariance matrices of the "noise" in the observations are all known. That is, $E[U_n \; U'_n] = Q_n$ is known for $n = 0, 1, \cdots$.

The Kalman filter is a sequential procedure which "updates" a previous estimate after a new vector of observations is observed. More precisely, the problem is to find a best estimate $\hat{X}_{n+1}$ of the state vector X_{n+1}, by using the observation Y_n and a previous estimate $\hat{X}_n$ which is based on the observations

Y_0, Y_1, . . ., Y_{n-1}. The procedure also assumes that P_n, the covariance matrix of the estimate $\hat{X}_n$, is known. The criteria used in establishing the formula for the updated estimate $\hat{X}_{n+1}$ is to minimize the trace of the covariance matrix of the new estimate.

Under the above assumptions the best estimate in the above sense, denoted by $\hat{X}_{n+1}$, is given by

$$\hat{X}_{n+1} = \hat{\Phi}_{n+1,n} \hat{X}^*_n, \tag{2.3}$$

where

$$\hat{X}^*_n = \hat{X}_n + \psi_n(Y_n - \hat{Y}_n), \tag{2.4}$$

$$\psi_n = P_n M'_n (M_n P_n M'_n + Q_n)^{-1}, \tag{2.5}$$

and

$$\hat{Y}_n = M_n \hat{X}_n. \tag{2.6}$$

(If A is a nonsingular matrix, then A^{-1} denotes its inverse.) The covariance matrix of $\hat{X}^*_n$, denoted by P^*_n, is given by

$$P^*_n = P_n - \psi_n M_n P_n.$$

The covariance matrix of $\hat{X}_{n+1}$, denoted by P_{n+1}, is given by

$$P_{n+1} = \Phi_{n+1,n} P^*_n \Phi'_{n+1,n}.$$

It is easily verified that if $\hat{X}_0$ is unbiased then so is $\hat{X}^*_n$ and $\hat{X}_{n+1}$, for each n.

The following observations are made for future reference.

(1) The change made in the estimate $\hat{X}_n$ to obtain $\hat{X}^*_n$ is a <u>weighted</u> difference in the observation and a prediction of what the observation should be based on the past observations.

(2) The dynamics of the system is used only to project the updated estimate at time t_n, $\hat{X}^*_n$, up to an estimate of the state vector at time t_{n+1}.

(3) Of paramount importance is the fact that regardless of the dimension of X_n and Y_n the filtering technique is applicable. That is, even if there are fewer observables than components in the state vector, one is able to obtain a new improved estimate of the state vector. Also, the routine does not require that the same quantities be observed at each time instant.

For an extensive discussion of the Kalman filter see, for example, [1], [2], [3], and [6].

Section 3: Generalized Exponential Smoothing

The method of exponential smoothing of time series (exponentially weighted moving averages, EWMA) has been used extensively in inventory control and production planning [7, 8]. EWMA is a method by which data is weighted in such a manner that the last observation is "used" more than the previous ones, the next to last observation more than the ones which preceded it, etc. This is a plausible method of handling time dependent economic data because the last observation possesses more information about the present economic environment than do the previous observations. On the other hand, the previous observations do contain relevant information in decreasing degrees and thus should be used accordingly in any prediction procedure.

Let $x(i)$, $i = 0, 1, \cdots, n, \cdots$ be a one-dimensional stochastic process with mean μ. Let $\hat{x}(0)$ denote an initial estimate of the mean and for $n \geq 0$, $\hat{x}(n+1)$ will be the estimate of μ using $x(0), x(1), \cdots, x(n)$. The basic EWMA model is

$$\hat{x}(n + 1) = \hat{x}(n) + \alpha e(n) \tag{3.1}$$

where α is a constant such that $0 \leq \alpha \leq 1$ and $e(n) = x(n) - \hat{x}(n)$.

Equivalently, equation 3.1 can be written as

$$\hat{x}(n + 1) = (1 - \alpha)\hat{x}(n) + \alpha x(n). \tag{3.2}$$

By substitution in 3.2, one easily obtains

$$\hat{x}(n + 1) = (1 - \alpha)^{n+1}\hat{x}(0)$$
$$+ \alpha[x(n) + (1 - \alpha)x(n - 1) + \cdots + (1 - \alpha)^n x(0)].$$

Thus the most recent observation receives a weight of α, the next to last $\alpha(1 - \alpha)$, $\cdots$, and the first observation $\alpha(1 - \alpha)^n$. For $\alpha \neq 0$, the initial estimate is used less as the number of observations increases, since its weight is $(1 - \alpha)^{n+1}$.

Since there is no closed form solution for α, the choice of α is a numerical one. The number α is chosen so that

$$\sum_{i=0}^{n} e(i)^2 = \sum_{i=0}^{n} [x(i) - \hat{x}(i)]^2 \tag{3.3}$$

is minimized.

As will be seen shortly, EWMA yields an unbiased estimate of μ for each n if $\hat{x}(0)$ is an unbiased estimate of μ. If one knows only that the variances of $x(0), x(1), \cdots, x(n)$, denoted by ${\sigma_0}^2, {\sigma_1}^2, \cdots, {\sigma_n}^2$, form a decreasing sequence, i.e., ${\sigma_0}^2 \geq {\sigma_1}^2 \geq \cdots \geq {\sigma_n}^2 \geq \cdots$, then equation 3.1 is a more reasonable method of estimating μ than using the sample mean. (Obviously if the variances are known exactly, then one would use the minimum variance unbiased estimator.)

If the variances are increasing instead of decreasing, then EWMA would not give an appropriate estimation model. In space navigation and guidance problems it is frequently the case that observations become more variable as time progresses since the distance over which the measurements are made become greater [3]. If certain measurements were made on a spacecraft in an elliptic orbit, then one would expect the variances to alternately increase and decrease. These considerations lead to "generalized exponential smoothing" (GEWMA).

Let $x(i)$, $i = 0, 1, \cdots, n, \cdots$ be the one-dimensional stochastic process described above and let $k(i)$, $i = 0, 1, \cdots, n, \cdots$ be a sequence of real numbers. For α a real number in the unit interval and $\hat{x}(0)$ an initial estimate consider the sequence of estimators $\hat{x}(1), \hat{x}(2), \cdots, \hat{x}(n+1), \cdots$ given by

$$\hat{x}(n + 1) = \hat{x}(n) + \alpha^{k(n)} e(n), \tag{3.4}$$

where

$$e(n) = x(n) - \hat{x}(n).$$

By substitution one sees that

$$\begin{aligned}\hat{x}(n + 1) = &\prod_{i=0}^{n} (1 - \alpha^{k(i)})\hat{x}(0) \\ &+ \alpha^{k(n)} x(n) + \alpha^{k(n-1)}(1 - \alpha^{k(n)})x(n - 1) \\ &+ \cdots + \alpha^{k(0)} \prod_{i=0}^{n} (1 - \alpha^{k(i)})x(0).\end{aligned} \tag{3.5}$$

If $k(i) = 1$, for each i, then it is clear that equation 3.4 reduces to equation 3.1. Thus 3.4 is a generalization of EWMA.

For the case where $k(i) = i$, for each i, one sees that (1) no initial estimate is needed and (2) the latest observation is used less and less. This would be a reasonable model if the variances of the observations were increasing in some definite pattern.

It will now be demonstrated that GEWMA is an unbiased estimator of μ, if $\hat{x}(0)$ is unbiased (hence EWMA gives unbiased estimates). From equation 3.5 one sees that $\hat{x}(n + 1)$ is a linear combination of the observations $x(0), x(1), \cdot \cdot \cdot, x(n)$ and $\hat{x}(0)$. Thus one needs to show that the sum of the coefficients is equal to one for each $n \geq 0$.

Define $\gamma_{-1}, \gamma_0, \cdot \cdot \cdot, \gamma_n$ as follows:

$$\gamma_{-1} = \prod_{i=0}^{n} (1 - \alpha^{k(i)}),$$

$$\gamma_n = \alpha^{k(n)},$$

and

$$\gamma_j = \alpha^{k(j)} \prod_{i=j+1}^{n} (1 - \alpha^{k(i)})$$

for $0 \leq j < n$.

Thus the sum of the coefficients is

$$\sum_{j=-1}^{n} \gamma_j = \begin{cases} \gamma_{-1} + \gamma_n + \sum_{j=0}^{n-1} \gamma_j & n \geq 1 \\ \gamma_{-1} + \gamma_0. & n = 0 \end{cases}$$

It will suffice to show that

$$\sum_{j=0}^{n-1} \gamma_j = 1 - \gamma_{-1} - \gamma_n$$

or

$$\sum_{j=0}^{n-1} \alpha^{k(j)} \prod_{i=j+1}^{n} (1 - \alpha^{k(i)}) = 1 - \prod_{i=0}^{n} (1 - \alpha^{k(i)}) - \alpha^{k(n)} \qquad (3.6)$$

for $n \geq 1$.

The proof will be by mathematical induction.

For $n = 0$,

$$\gamma_{-1} + \gamma_0 = \prod_{i=0}^{0} (1 - \alpha^{k(i)}) + \alpha^{k(0)}$$

$$= 1 - \alpha^{k(0)} + \alpha^{k(0)}$$

$$= 1.$$

Assume that 3.6 holds for $n = m$. That is,

$$\sum_{j=0}^{m-1} \alpha^{k(j)} \prod_{i=j+1}^{m} (1 - \alpha^{k(i)}) = 1 - \alpha^{k(m)} - \prod_{i=0}^{m} (1 - \alpha^{k(i)}). \qquad (3.7)$$

For $n = m + 1$,

$$\sum_{j=0}^{m} \alpha^{k(j)} \prod_{i=j+1}^{m+1} (1 - \alpha^{k(i)})$$

$$= \alpha^{k(m)} \prod_{i=m+1}^{m+1} (1 - \alpha^{k(i)}) + \sum_{j=0}^{m-1} \alpha^{k(j)} \prod_{i=j+1}^{m+1} (1 - \alpha^{k(i)})$$

$$= \alpha^{k(m)} (1 - \alpha^{k(m+1)})$$

$$+ (1 - \alpha^{k(m+1)}) \left[\sum_{j=0}^{m-1} \alpha^{k(j)} \prod_{i=j+1}^{m} (1 - \alpha^{k(i)}) \right].$$

By the inductive hypothesis the quantity in the brackets is given in equation 3.7. Therefore,

$$\sum_{j=0}^{m} \alpha^{k(j)} \prod_{i=j+1}^{m+1} (1 - \alpha^{k(i)})$$

$$= \alpha^{k(m)}(1 - \alpha^{k(m+1)})$$

$$+ (1 - \alpha^{k(m+1)})\left[1 - \alpha^{k(m)} - \prod_{i=0}^{m} (1 - \alpha^{k(i)})\right]$$

$$= \alpha^{k(m)}(1 - \alpha^{k(m+1)})$$

$$+ (1 - \alpha^{k(m)})(1 - \alpha^{k(m+1)}) - \prod_{i=0}^{m+1} (1 - \alpha^{k(i)})$$

$$= (1 - \alpha^{k(m+1)})(\alpha^{k(m)} + 1 - \alpha^{k(m)}) - \prod_{i=0}^{m+1} (1 - \alpha^{k(i)})$$

$$= 1 - \alpha^{k(m+1)} - \prod_{i=0}^{m+1} (1 - \alpha^{k(i)}).$$

It therefore follows that the expected value of $\hat{x}(n + 1)$ is μ whenever $\hat{x}(0)$ is unbiased. That is,

$$E\hat{x}(n + 1) = E\left[\sum_{j=0}^{n} \gamma_j x(j) + \gamma_{-1}\hat{x}(0)\right]$$

$$= \sum_{j=0}^{n} \gamma_j \, E[x(j)] + \gamma_{-1} \, E[\hat{x}(0)]$$

$$= \mu\left[\sum_{j=-1}^{n} \gamma_j\right]$$

$$= \mu.$$

In the case of a p-variate stochastic process GEWMA could be applied to each component thus giving the model

$$\hat{X}(n+1) = \hat{X}(n) + A(n)\,E(n), \tag{3.8}$$

where

$$A(n) = \begin{pmatrix} \alpha_1 k_1(n) & 0 & \cdots & 0 \\ 0 & \alpha_2 k_2(n) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \alpha_p k_p(n) \end{pmatrix},$$

α_i is a constant for each i,

$k_i(n)$ is a sequence for each i,

and $E(n) = X(n) - \hat{X}(n)$.

The estimate $\hat{X}(n)$ is a $p \times 1$ vector, as is $X(n)$.

The model in equation 3.8 does not take into account the statistical dependency (if any) of the different components of the vector X(n). But the exact dependency (partially expressed by covariance matrices) is frequently unknown. Furthermore when this technique is applied to estimation in a dynamic system in section 5, the dynamics of the model will aid in introducing the appropriate interactions.

Section 4: The Matrix Pseudoinverse and Its Applications

Consider a system of linear equations

$$Y = MX \tag{4.1}$$

where

Y is an $m \times 1$ vector.

M is an $m \times p$ matrix of known constants.

X is a $p \times 1$ vector of unknowns.

It is well known that if M is square and of rank p, then a unique solution-vector X exists and is given by $X = M^{-1}Y$. When M is non-square or when M is square but singular, there may still be a solution(s) to the system and a unified theory to treat both the case when M^{-1} exists and when it doesn't is desirable. One such theory involves the use of the pseudoinverse of a matrix, which will be discussed in this section.

<u>DEFINITION 4.1</u>: Let M be an $m \times p$ matrix. If a matrix M^+ exists that satisfies the four following conditions, then it will be called a pseudoinverse of M:

$$\begin{aligned} &\text{(i)} \quad MM^+ \text{ is symmetric} \\ &\text{(ii)} \quad M^+M \text{ is symmetric} \\ &\text{(iii)} \quad MM^+M = M \\ &\text{(iv)} \quad M^+MM^+ = M^+ \end{aligned} \tag{4.2}$$

If M is square and nonsingular, then it is clear that M^{-1} exists and satisfies the four conditions of (4.2). If M is square but singular, however, or if M is not square, then the problem remains as to whether a matrix M^+ exists that will satisfy (4.2) It will now be shown that for any matrix M, a pseudoinverse M^+ exists and is in fact unique. Note that if a pseudoinverse exists, it must have order $p \times m$, since MM^+ must be symmetric and hence square. Also note that if M is the null matrix of order $m \times p$, denoted by $\phi_{m\times p}$, then $\phi_{p\times m}$ satisfies the four conditions required in (4.2).

<u>THEOREM 4.1</u>: For every $m \times p$ matrix M, there is a matrix M^+ satisfying the four conditions of (4.2) i.e., every matrix has a pseudoinverse.

<u>Proof</u>: If $M = \phi$, then from the above remarks, $M^+ = \phi_{p\times m}$. Thus assume $M \neq \phi$. If M has rank r, it can be factored (nonuniquely) as

$$M = BC \tag{4.3}$$

where B is $m \times r$ of rank r, and C is $r \times p$ of rank r. Since B and C are full rank it follows that BB' and CC' are each nonsingular. Let M^+ be defined as follows:

$$M^+ = C'(CC')^{-1}(BB')^{-1}B'. \tag{4.4}$$

It is easily shown that M^+ is a pseudoinverse of M. The factorization of M in (4.3) is not unique, but the pseudoinverse as given in (4.4) is unique as shown by the following theorem.

THEOREM 4.2: For every matrix M, there exists a unique matrix M^+ that satisfies the four conditions of (4.2), i.e., every matrix has a unique pseudoinverse.

Proof: Assume that M^+_1 and M^+_2 are two distinct pseudoinverses of M. To show that this implies $M^+_1 = M^+_2$, it will first be shown that $MM^+_1 = MM^+_2$. Multiply $M = MM^+M$ on the right by M^+_2, obtaining

$$MM^+_2 = MM^+_1MM^+_2$$

From condition (i) of (4.2), MM^+_2 is symmetric and thus so is $MM^+_1MM^+_2$:

$$MM^+_1MM^+_2 = (MM^+_1MM^+_2)'$$

Thus, since both MM^+_2 and MM^+_1 are symmetric,

$$\begin{aligned} MM^+_2 = (MM^+_1MM^+_2)' &= (MM^+_2)'(MM^+_1)' \\ &= (MM^+_2)(MM^+_1) \\ &= (MM^+_2M)M^+_1 \\ &= MM^+_1. \end{aligned}$$

Similarly, if we multiply $M = MM^+_1M$ on the left by M^+_2, we obtain

$$M^+_1M = M^+_2M.$$

This results, together with $MM^+_2 = MM^+_1$, give

$$M^+_1 = M^+_1MM^+_1 = M^+_2MM^+_1 = M^+_2MM^+_2 = M^+_2$$

and the proof is complete.

Several additional theorems concerning pseudoinverses will be stated without proof. In these results, note the similarities between a pseudoinverse and the ordinary inverse of a square nonsingular matrix.

<u>THEOREM 4.3</u>: Let $r(A)$ denote the rank of the matrix A. Then

$$r(M) = r(M^+) = r(MM^+) = r(M^+M) = r(MM^+M) = r(M^+MM^+).$$

<u>THEOREM 4.4</u>: $(M')^+ = (M^+)'$.

<u>THEOREM 4.5</u>: $(M^+)^+ = M$.

<u>THEOREM 4.6</u>: $(M'M)^+ = M^+M'^+ = M^+M^{+\prime}$.

<u>THEOREM 4.7</u>: $(MM^+)^+ = MM^+$ and $(M^+M)^+ = M^+M$.

THEOREM 4.8: Let P be an $m \times m$ orthogonal matrix, let Q be a $p \times p$ orthogonal matrix, and let M be any $m \times p$ matrix. Then

$$(PMQ)^+ = Q'M^+P'$$

THEOREM 4.9: If $M = M'$, then $M^+ = M^{+\prime}$.

THEOREM 4.10: If $M = M'$, then $MM^+ = M^+M$.

THEOREM 4.11: If M is symmetric idempotent (i.e., $M^2 = M$, $M = M'$), then $M^+ = M$, i.e., a symmetric idempotent matrix is its own pseudoinverse.

THEOREM 4.12: Let D be an $n \times n$ diagonal matrix with diagonal elements d_{ii}, $i = 1, 2, \cdot \cdot \cdot, n$. The pseudoinverse D^+ of D is a diagonal matrix with diagonal elements d_{ii}^{-1} if $d_{ii} \neq 0$, and zero otherwise.

THEOREM 4.13: If M is an $m \times p$ matrix of rank m, then $M^+ = M'(MM')^{-1}$ and $MM^+ = I_m$. If the rank of M is p, then $M^+ = (M'M)^{-1}M'$ and $M^+M = I_p$.

THEOREM 4.14: The matrices MM^+, M^+M, $I - MM^+$ and $I - M^+M$ are each symmetric idempotent.

Let us now use the pseudoinverse of a matrix to investigate the system of equations $MX = Y$. In particular we shall be concerned with the consistency of the system, and when consistent, with finding the general solution.

THEOREM 4.15: The system of equations $Y = MX$ is consistent (has at least one solution) if and only if

$$MM^{+}Y = Y.$$

Proof: Assume the system is consistent, and let X_1 be a vector satisfying the system, i.e., $MX_1 = Y$. Multiply on the left by MM^{+} and get

$$MM^{+}MX_1 = MM^{+}Y.$$

But the left hand side is

$$(MM^{+}M)X_1 = MX_1 = Y.$$

Thus $MM^{+}Y = Y$.

For the converse, assume $MM^{+}Y = Y$. Let $X^* = M^{+}Y$. X^* obviously satisfies $MX = Y$, and thus the equations are consistent.

The next theorem is quite useful and will be used often in sections to follow.

THEOREM 4.16: If the system of equations $MX = Y$ is consistent, then the general solution is

$$\hat{X} = M^{+}Y + (M^{+}M - I)Z, \tag{4.5}$$

where the vector Z is arbitrary.

Proof: Substitute (4.5) into $MX = Y$.

Suppose the system of equations $Y = MX$ is not consistent, i.e., there is no vector X that satisfies the system. Then we may write

$$Y - MX = e$$

where e is a vector of deviations. Since there is no vector X such that $e = \phi$, then it would be desirable to find a vector X_0 such that e is minimized in some sense. One useful index to consider is minimization of the sum of squared deviations, $\sum e^{2}{}_{i}$, which in matrix form may be written as $e'e = (Y - MX)'(Y - MX)$. This leads to the following definition.

DEFINITION 4.2: The vector X_0 will be called the Best Approximate Solution (BAS) to the system of equations $MX = Y$ if

(1) For all X, $(Y - MX_0)'(Y - MX_0) \leq (Y - MX)'(Y - MX)$,

and (2) For those vectors X in which the equality in (1) holds, $X'X \geq X'_0X_0$.

This definition essentially states that the vector X_0 minimizes the sum of squared deviations, and if there is a set of vectors such that each member in the set gives the same minimum sum of squared deviations, then X_0 is chosen from this set if there is no other vector in the set with smaller squared length. We shall now state and prove a theorem which will enable us to use the pseudoinverse of M to find the BAS to Y = MX.

<u>THEOREM 4.17</u>: The Best Approximate Solution (BAS) to the system of equations Y = MX is given by

$$X_0 = M^+Y$$

where M^+ is the pseudo-inverse of M.

<u>Proof</u>: Upon adding and subtracting MM^+Y to $MX - Y$, we obtain

$$\begin{aligned}(Y - MX)'(Y - MX) &= (MX - Y)'(MX-Y)\\ &= (MX - MM^+Y + MM^+Y - Y)'(MX - MM^+Y + MM^+Y - Y)\\ &= [M(X - M^+Y) + (MM^+ - I)Y]'[M(X - M^+Y) + (MM^+ - I)Y]\\ &= [M(X - M^+Y)]'[M(X - M^+Y)] + [(MM^+ - I)Y]'[(MM^+ - I)Y]\end{aligned}$$

since the cross product terms both go to zero. It is obvious that this expression is minimized as a function of X when $X = X_0 = M^+Y$. Thus,

$$\begin{aligned}(Y - MX)'(Y - MX) &\geq [(MM^+ - I)Y]'[(MM^+ - I)Y]\\ &= (MX_0 - Y)'(MX_0 - Y).\end{aligned}$$

The equality holds if and only if

$$[M(X - M^{+}Y)]'[M(X - M^{+}Y)] = \phi,$$

i.e. if and only if $MX = MM^{+}Y$. If this is true, then it must be shown that

$$X'X \geq (M^{+}Y)'(M^{+}Y) = X'_0X_0.$$

Consider the following identity which is true for all vectors X:

$$[M^{+}Y + (I - M^{+}M)X]'[M^{+}Y + (I - M^{+}M)X]$$
$$= (M^{+}Y)'(M^{+}Y) + [(I - M^{+}M)X]'[(I - M^{+}M)X].$$

Upon substituting $MM^{+}Y$ for MX, or equivalently, $M^{+}Y$ for $M^{+}MX$, the identity becomes

$$X'X = (M^{+}Y)'(M^{+}Y) + (X - M^{+}Y)'(X - M^{+}Y)$$

or

$$X'X \geq (M^{+}Y)'(M^{+}Y) = X'_0 X_0$$

and the theorem is proved.

The importance of Theorem 4.17 is best described as follows. In space navigation and orbital determination problems the state vector X describes a particle's location and its instantaneous rates of change in the direction of the coordinate axes. X is not observable, and Y is measured instead. Y, however, is contaminated with noise, so that $Y = MX + e$ is an appropriate representation, where e is a vector of unobservable random errors. Considered as a system of equations, $Y = MX + e$ is therefore inconsistent.

Theorem 4.17, however, states that the best approximate solution to $Y = MX + e$ is given by $X = M^{+}Y$, where M^{+} is the unique pseudoinverse of Y. Thus one arrives at the same solution by solving either the inconsistent set $Y = MX + e$, or the consistent set $Y = MX$, obtained by "dropping" the error term.

In the remaining sections of this report, the pseudoinverse of a matrix will be used primarily in solving a system of linear equations. In particular, a great deal of use is made of the representation of the solution to $Y = MX$ as

$$X = M^{+}Y + (M^{+}M - I)Z,$$

where Z is arbitrary. It will be shown that proper choice of the arbitrary vector Z will yield an estimator of the state vector with certain desirable properties. These topics are discussed in more detail in the following sections. This section concludes with an algorithm for computing the pseudoinverse of a matrix. The trace of a matrix A, say, will be denoted by $\mathrm{tr}(A)$.

<u>THEOREM 4.18</u>: Let A be any $m \times n$ matrix. To compute A^+, proceed as follows:

(i) Compute $B = A'A$

(ii) Define $C_1 = I$. Compute

$$C_{i+1} = \frac{I}{i}\,tr(C_iB) - C_iB, \; i = 1, 2, \cdot \cdot \cdot, r-1$$

Stop when $C_{i+1}B = \phi$. This will occur when $i = r$.

(iii) $$A^+ = \frac{rC_rA'}{tr(C_rB)}\,,\; tr(C_rB) \neq 0.$$

For a proof of this theorem, see (12). Other algorithms are given in (9), (10) and (11).

Section 5: A New Kalman-Type Filter

Attention is now turned to the estimation problem for the dynamic system given in equations 2.1 and 2.2 when there are at most as many observables as there are components in the state vector. Let the dimension of the state vector X_n be $p \times 1$ and the dimension of the observation vector Y_n be $q_n \times 1$ where $q_n \leq p$, for each n. Obviously this is the most difficult estimation case and the situation for which the Kalman filter is the most useful.

The new filtering equation combines the technique of generalized exponential smoothing and the theory of generalized inverses of matrices. In particular, the (unique) pseudoinverse of M_n is used.

Consider the sequence of estimators $\hat{X}^-_0, \hat{X}^-_1, \cdots, \hat{X}^-_n, \cdots$ generated by

$$\hat{X}^-_n = \hat{X}_n + A_n(\tilde{X}_n - \hat{X}_n), \tag{5.1}$$

where

(1)
$$A_n = \begin{pmatrix} \alpha_1^{k_1(n)} & 0 & \cdots & 0 \\ 0 & \alpha_2^{k_2(n)} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \alpha_p^{k_p(n)} \end{pmatrix},$$

(2) $k_i(n)$ is a sequence of real numbers for each $i = 1, \cdots, p$,

(3) $0 \leq |\alpha_i| \leq 1$ for each $i = 1, \cdots, p$,

(4) $\tilde{X}_n = M^+_n Y_n + (M^+_n M_n - I) Z_n$, (5.2)

(5) $\hat{X}_n$ is a previous estimate of X_n based on the observations $Y_0, Y_1, \cdot \cdot \cdot, Y_{n-1}$, or is an initial estimate $\hat{X}_0$.

The collection of filters defined by equation 5.1 has the same general form as the updated estimate $\hat{X}^*_n$ in the Kalman filter (equation 2.4) and the multidimensional generalized exponential smoothing model (equation 3.8).

As in the case of the Kalman filter (equation 2.3), the dynamics of the system are used only to obtain an estimate of the state vector at t_{n+1}, Precisely,

$$\hat{X}_{n+1} = \Phi_{n+1,n} \hat{X}^-_n. \tag{5.3}$$

The linear filters do not require a knowledge of the covariance matrices of the error vector. The choice of the sequence $k_i(n)$ would depend on one's knowledge of the variability of the data, but this is not nearly as restrictive as requiring that the covariance matrices be known. For certain choices of $k_i(n)$ an initial estimate $\hat{X}_0$ is not even required. The choice of constants α_i is a numerical one.

The vector Z_n in equation 5.2 is arbitrary and will determine the exact form of the estimator.

<u>LEMMA 5.1</u>: If $Z_0, Z_1, \cdot \cdot \cdot, Z_n, \cdot \cdot \cdot$ is any sequence of random vectors which satisfies the sequence of systems of equations

$$(M^+_n M_n - I)Z_n = (M^+_n M_n - I)(-\hat{X}_n) \tag{5.4}$$

and $\hat{X}_0$ is an unbiased estimate of X_0, then the sequence $\hat{X}^-_0, \hat{X}^-_1, \cdot \cdot \cdot, \hat{X}^-_n, \cdot \cdot \cdot$ is an unbiased sequence of estimators of the sequence of state vectors $X_0, X_1, \cdot \cdot \cdot, X_n, \cdot \cdot \cdot$.

<u>Proof</u>: For n = 0, if Z_0 satisfies 5.4, then

$$\hat{X}^-_0 = \hat{X}_0 + A_0(\tilde{X}_0 - \hat{X}_0)$$

$$= \hat{X}_0 + A_0(M^+_0 Y_0 - M^+_0 M_0 \hat{X}_0).$$

Thus

$$E\hat{X}^-_0 = E\hat{X}_0 + A_0(M^+_0 EY_0 - M^+_0 M_0 E\hat{X}_0)$$

$$= X_0 + A_0(M^+_0 M_0 X_0 - M^+_0 M_0 X_0)$$

$$= X_0.$$

Assume that $\hat{X}^-_k$ is unbiased. Then

$$\hat{X}_{k+1} = \Phi_{k+1,k} \hat{X}^-_k$$

and

$$E\hat{X}_{k+1} = \Phi_{k+1,k} E\hat{X}^-_k$$

$$= \Phi_{k+1,k} X_k$$

$$= X_{k+1}.$$

Therefore,

$$E\hat{X}^{-}_{k+1} = X_{k+1} + A_{k+1}(M^{+}_{k+1} M_{k+1} X_{k+1} - M^{+}_{k+1} M_{k+1} X_{k+1})$$

$$= X_{k+1}.$$

If $\hat{X}^{-}_{k}$, $k = 0, 1, \cdot \cdot \cdot, n, \cdot \cdot \cdot$ is an unbiased sequence, then obviously the sequence $\hat{X}_k$, $k = 1, \cdot \cdot \cdot, n, \cdot \cdot \cdot$ given by equation 5.3 is an unbiased sequence.

The result expressed in Lemma 5.1 is not too surprising after the following Lemma is known.

<u>LEMMA 5.2</u>: If $Z_0, Z_1, \cdot \cdot \cdot, Z_n, \cdot \cdot \cdot$ is a sequence of vectors which satisfies the conditions of Lemma 5.2, then equation 5.1 reduces to

$$\hat{X}^{-}_{n} = \hat{X}_n + A_n M^{+}_{n}(Y_n - \hat{Y}_n) \tag{5.5}$$

which is then independent of Z_n. The vector $\hat{Y}_n$ is given by

$$\hat{Y}_n = M_n \hat{X}_n. \tag{5.6}$$

Proof: By substitution,

$$\hat{X}^{-}_{n} = \hat{X}_n + A_n(M^{+}_{n} Y_n + (M^{+}_{n} M_n - I)(-\hat{X}_n) - \hat{X}_n)$$

$$= \hat{X}_n + A_n(M^{+}_{n} Y_n - M^{+}_{n} M_n \hat{X}_n)$$

$$= \hat{X}_n + A_n M^{+}_{n}(Y_n - \hat{Y}_n),$$

where $\hat{Y}_n = M_n \hat{X}_n$.

Much weaker conditions for unbiasness than those of Lemma 5.1 are given in the following lemma. This next results states that the random Z vectors only have to satisfy the systems of equations 5.2 in the mean.

LEMMA 5.3: Let $Z_0, Z_1, \cdots, Z_n, \cdots$ be any sequence of random vectors whose means satisfy the following system of equations:

$$(M^{+}_{n} M_n - I)\, EZ_n = (M^{+}_{n} M_n - I)(-X_n)\,. \tag{5.7}$$

If $\hat{X}_0$ is an unbiased estimate of X_0, then the sequence $\hat{X}^{-}_{0}, \hat{X}^{-}_{1}, \cdots, \hat{X}^{-}_{n}, \cdots$ is an unbiased sequence of estimators of the sequence of state vectors $X_0, X_1, \cdots, X_n, \cdots$.

Proof: The proof follows the same lines as that of Lemma 5.1 and is omitted.

It is obvious that one choice of Z_n which would satisfy Lemma 5.1 is $Z_n = -\hat{X}_n$. In this case, Lemma 5.2 states that we could have used equation 5.5 as the filtering equation.

Lemma 5.3 is less restrictive in the choice of Z_n. We digress at this point to note that any $\tilde{X}_n$ for which $E(\tilde{X}_n - X_n) = 0$ will also produce a sequence of unbiased estimators, if $\hat{X}_0$ is unbiased. Notice also that selecting a Z_n is equivalent to selecting an $\tilde{X}_n$ (See Theorem 4.16).

In general if $Y = MX$ is a consistent system of q equations in p unknowns, $q \leq p$, and the rank of M is q, then we may partition (and rearrange) the system so that

$$Y = [M^1 \quad M^*] \begin{bmatrix} X^1 \\ X^* \end{bmatrix},$$

where M^1 is $q \times q$ and nonsingular. The choice of M^1 is not necessarily unique. Then

$$\tilde{X} = \begin{bmatrix} (M^1)^{-1}Y - (M^1)^{-1}M^* X^* \\ X^* \end{bmatrix}$$

is a solution where X^* is arbitrary. The point is, in the estimation problem we have a previous estimate at each stage and could substitute the estimated values of X^* into the equation.

<u>LEMMA 5.4</u>: For

$$\tilde{X}_n = \begin{bmatrix} (M^1)^{-1}Y - (M^1)^{-1}M^* \hat{X}^*{}_n \\ \hat{X}^*{}_n \end{bmatrix},$$

the sequence of estimators $\hat{X}^-{}_0, \hat{X}^-{}_1, \cdot\cdot\cdot, \hat{X}^-{}_n, \cdot\cdot\cdot$ given by equation 5.1 is unbiased whenever $\hat{X}_0$ is unbiased. The notation $X^*{}_n$ denotes those estimates of $X^*{}_n$ in $\hat{X}_n$.

<u>Proof</u>: The proof follows from noticing that

$$\begin{aligned} \tilde{X}_n - \hat{X}_n &= \begin{bmatrix} (M^1{}_n)^{-1}Y_n - (M^1{}_n)^{-1}M^*{}_n \hat{X}^*{}_n - \hat{X}^1{}_n \\ \hat{X}^*{}_n - \hat{X}^*{}_n \end{bmatrix} \\ &= \begin{bmatrix} (M^1{}_n)^{-1}Y_n - (M^1{}_n)^{-1}M^*{}_n \hat{X}^*{}_n - \hat{X}^1{}_n \\ 0 \end{bmatrix} \end{aligned}$$

and

$$EY_n = M^1{}_n X^1{}_n + M^*{}_n X^*{}_n.$$

Section 6: Future Studies

In this report a new family of statistical filters has been introduced and briefly studied with respect to unbiasedness. The motivation for these new filters was furnished by the assumptions, form and role of the Kalman filter.

The authors of this report feel that much valuable work is yet to be done in this area of statistical filtering. The proper choice of the vector Z_n in equation 5.2 could possibly result in a minimum variance unbiased estimator in the class of filters given by equation 5.1. It would not be unreasonable to expect certain asymptotic results to hold. This area should be fruitful for extensive additional investigation.

BIBLIOGRAPHY

1. Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems. Trans. ASME, Ser. F, Jour. Basic Engr., Vol. 82, No. 1, March, 1960, pp. 34-45.

2. Solloway, C. B.: Dynamic Filtering. Jet Propulsion Laboratory Space Programs Summary No. 37-21, Pasadena, California, March, 1963, pp. 7-10.

3. Liebelt, P. B.: An Introduction To Optimal Estimation. Addison-Wesley Pub. Co., 1967.

4. Kalman, R. E.: Mathematical Description of Dynamical Systems. Jour. Soc. Indust. Appl. Math., Ser. A, Control., Vol. 1, No. 2, 1963, pp. 152-192.

5. Kalman, R. E., Y. C. Ho, and K. S. Narendra: Controllability of Linear Dynamical Systems. Contributions to Differential Equations, Vol. 1, No. 2, 1962.

6. Deutsch, R.: Estimation Theory. Prentice Hall Pub. Co., 1965.

7. Brown, R. G.: Statistical Forecasting. Arthur D. Little, Inc., 1958.

8. Contie, G. A., et al: Short-Term Forecasting Techniques of Production Control, Monograph No. 2. D. Van Nostrand Co., Inc., 1966.

9. Graybill, F. A.: Introduction to Matrices With Applications in Statistics. Wadsworth Pub. Co., 1969.

10. Rao, C. R.: Linear Statistical Inference and Its Applications. John Wiley and Sons, 1965.

11. Penrose, R. A.: A Generalized Inverse for Matrices. Proc. Camb. Phil. Soc., Vol. 51, 1955, pp. 406-413.

12. Decell, H. P.: An Application of the Cayley-Hamilton Theorem to Generalized Matrix Inversion. SIAM Review, Vol. 7, No. 4, 1965.