PROGRESS REPORT

Research in and Application of State Variable Feedback Design of Guidance Control Systems for Aerospace Vehicles

by

T. L. Williams

R. T. Stefani

S. Yakowitz

March, 1970

Prepared under Contract NGR-03-002-115

ENGINEERING EXPERIMENT STATION

COLLEGE OF ENGINEERING

THE UNIVERSITY OF ARIZONA

TUCSON, ARIZONA

Progress Report

Research in and Application of State Variable Feedback Design of Guidance Control Systems for Aerospace Vehicles

by

T. L. Williams

R. T. Stefani

S. Yakowitz

March, 1970

Prepared under Contract NGR-03-002-115

The University of Arizona
Electrical Engineering Department
Tucson, Arizona

for

Lewis Research Center
National Aeronautics and Space Administration

Table of Contents

Nomenclature

A - Matrix chosen to provide an unbiased estimate of H

E{ } - Expected value operator

H - Parameter matrix

J - Performance index

M - Weighting matrix

M_α - Aerodunamic stability derivative (°/sec^2/° of α)

M_δ - Fin effectiveness gain (°/sec^2/° of δ)

m_{ij} - Weighting coefficient (element of M)

N - Additive noise on X_e

N_a - Normal acceleration (g's)

P - Covariance matrix of the estimation error

R - Covariance matrix of the noise V

S - Covariance matrix of the noise N

X_e - Sensed value for the system state vector

V - Additive noise on Y_e

Y_e - Exact value for the product X_e H

Y_s - Sensed value for Y_e

Z_α - Normal force coefficient (g's/° of α)

α - Angle of attack (°)

δ - Fin deflection angle (°)

ζ - Mean

$\ddot{\theta}$ - Angular acceleration (°/sec^2)

σ^2 - Variance

Σ - Summation

$\|e\|_M$ - Norm = $e^T Me$ (defines squared error for error vector e)

Weighted Least Squares Parameter Estimation in the Presence of Noise

1. Introduction

Discussed in this report is a generalization of conventional weighted least squares state (or parameter) estimation techniques to the problem of parameter estimation where all sensor measurements are noisy.

This problem is illustrated in Figure A. Sensors provide noisy measurement vectors X_s and Y_s. These vectors may contain many repititions of some data measurement. At any rate, the parameter matrix H relating the exact (deterministic) states X_e and Y_e is considered to be unknown. The problem then is to provide an unbiased estimate of H in a manner that is optimal in a least squares sense. Thus it is necessary to select a least squares performance index and some associated weighting matrix.

The transition from conventional weighted least squares theory is made. An estimation equation is derived which provides an unbiased parameter estimate as well as least squared error. The weighting matrix is selected so as to provide minimum variance of the estimation error. The selection of the weighting matrix is shown to be more complicated than for the conventional least squares case. Two examples are given in which the weighting matrix is selected for specific problems. A procedure for operating on repeated measurements is discussed. A simulation is presented in which the derived estimation equations do provide unbiased estimates and the variance of the estimation error is calculated as a check on the analytical results.

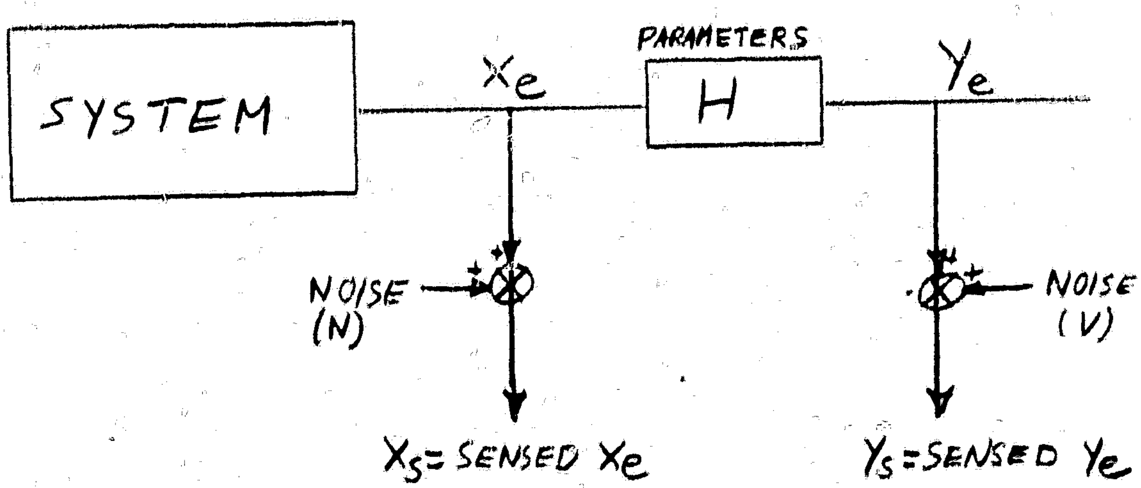


Figure A - Basic Block Diagram

2. Conventional Weighted Least Squares Parameter Estimation

2.1 Conventional State Estimation

The conventional least squares state estimation problem may be defined as follows. Consider an exact equation

$$Y_e = H X_e \tag{1}$$

where Y_e and X_e are states and H are known parameters. Suppose a sensor provides noisy measurements of Y_e

$$Y_s = Y_e + V, \quad \{E(v(t)) = 0;\ E(V(t+\tau)V^T(t)) = R\delta(t-\tau)\} \tag{2}$$

Then select $\hat{X}_e$ such that one minimizes

$$= \|Y_s - H\hat{X}_e\|_M \tag{3}$$

Subject to the requirement of an unbiased estimate

$$E(\hat{X}_e) = X_e \tag{4}$$

The well-known answer is that

$$\hat{X}_e = (H^T M H)^{-1} H^T M Y_s \tag{5}$$

The variance of the estimation error is

$$\begin{aligned} P &= E[(\hat{X}_e - X_e)(\hat{X}_e^T - X_e^T)] \\ &= (H^T M H)^{-1}(H^T M R M H)(H^T M H)^{-1} \end{aligned} \tag{6}$$

The variance of the error is minimum if $M = R^{-1}$

$$P_{min} = [H^T R^{-1} H]^{-1} \tag{7}$$

The above problem can easily be rephrased as a parameter estimation problem (estimate H with Y_s being a noisy measurement of Y_e. The general problem is to estimate H with Y_s being a noisy measurement of Y_e and also X_s being a noisy measurement of X_e. First, let us consider the former case by simply rewriting the above equations.

2.2 Conventional Parameter Estimation

Consider an exact equation

$$Y_e = X_e H \tag{8}$$

where Y_e and X_e are states and H contains unknown parameters. Suppose a sensor provides noisy measurements of Y_e

$$Y_s = Y_e + V \quad \{E(V(t)) = 0;\ E[V(t+\tau)V^T(t)] = R\delta(t-\tau)\} \tag{9}$$

Then select $\hat{H}$ such that one minimizes

$$= \|Y_s - X_e \hat{H}\|_M \tag{10}$$

Subject to the requirement of an unbiased estimate

$$E(H) = \hat{H} \tag{11}$$

In this case, the answer follows by rewriting Equations 5-7

$$\begin{aligned} \hat{H} &= (X_e^T M X_e)^{-1} X_e^T M Y_s \\ P &= E[(\hat{H} - H)(\hat{H}^T - H^T)] \\ &= (X_e^T M X_e)^{-1} (X_e^T M R M X_e) (X_e^T M X_e)^{-1} \\ P_{min} \text{ (for } M = R^{-1}) &= (X_e^T R^{-1} X_e)^{-1} \end{aligned} \tag{12}$$

There are two interesting degeneracies associated with Equation set 12.

1) Suppose M is a scalar. Then R must also be a scalar (let $R = \sigma_v^2$). Then **Equations** 12 are

$$\hat{H} = (X_e^T X_e)^{-1} X_e^T Y_s$$
$$P = \sigma_v^2 (X_e^T X_e)^{-1} = P_{min} \qquad (13)$$

Thus P is immediately minimum variance since M cancels out and P is independent of M.

2. Suppose X_e is a non singular matrix. Then Equations 12 are

$$\hat{H} = X_e^{-1} Y$$
$$P = X_e^{-1} R(X_e^T)^{-1} = P_{min} \qquad (14)$$

Thus, P again is minimum variance since again M cancels out of all the equations and P is independent of M.

With the above review accomplished, let us now consider a more general problem.

3. Weighted Least Squares Parameter Estimation - All sensors Noisy

3.1 Problem Statement and Approach to Solution

Consider an exact equation

$$Y_e = X_e H \qquad (15)$$

where Y_e and X_e are states and H contains unknown parameters. Suppose sensors provide noisy measurements of Y_e and X_e.

$$Y_s = Y_e + V \{E(v(t)) = 0;\ E[V(t+\tau)\ V^T(t)] = R\delta(t-\tau)\}$$

$$X_s = X_e + N \{E(N(t)) = 0;\ E[N^T(t+\tau)\ N(t)] = S\delta(t-\tau)\} \tag{16}$$

Then select $\hat{H}$ such that one minimizes (since X_e is not known)

$$J = \| Y_s - \hat{X}_e \hat{H} \|_M \tag{17}$$

Subject to the requirement of an unbiased estimate

$$E(\hat{H}) = H \tag{18}$$

If the objective function of Equation 17 was chosen as an artifice such that a parameter estimator could be derived, then one may additionally select M such that the variance (P) of the estimation error is a minimum.

$$P = E[(\hat{H}-H)\ (\hat{H}^T - H^T)] \tag{19}$$

Now the following procedure will be followed in solving this problem.

Step 1. Choose $\hat{H}$ to minimize the weighted squared error J.

Step 2. Select $\hat{X}_e = X_s A$ and choose A such that the estimate is unbiased $(E(\hat{H}) = H)$ (20)

Step 3. Select M for minimum P, if one is free to choose M. If M is pre-selected, then only Steps 1 and 2 are carried out.

3.2 Derivation of Parameter Estimator and Variance of the Estimation Error

Step 1. Let us solve for $\hat{H}$ such that $\frac{\partial J}{\partial \hat{H}} = 0$. Then the result is

$$\hat{H} = (\hat{X}_e^T M \hat{X}_e)^{-1} \hat{X}_e^T M Y_s \qquad (21)$$

Step 2. Let us evaluate the expected value of $\hat{H}$ for $\hat{X}_e = X_s A$, substituting the sensor Equations 16 into 21. Assuming that A is non singular we get, noting that the noise terms have zero mean and may be eliminated wherever they appear as cross products:

$$E(\hat{H}) = E\{A^{-1}(X_s^T \; M \; X_s)^{-1} (X_e^T M X_e) \; H\} \qquad (22)$$

Now for an unbiased estimate, set Equation 22=H. Suppose we select $A = (X_s^T M X_s)^{-1}[X_s^T M X_s - E\{N^T MN\}]$. Then we have an unbiased estimate of H. Finally let us substitute the chosen value for $\hat{X}_e$ into Equation 21, since the result is on unbiased estimate of H.

The result is after simplification

$$\boxed{\hat{H} = [X_s^T MX_s - E(N^T MN)]^{-1} X_s^T MY_s} \qquad (23)$$

In summary Equation 23 is an unbiased estimation of the parameter vector H and the estimate also causes the weight least squared error to be a minimum.

We are now in a position to calculate the variance of the estimation error. Since the estimates of H are unbiased then $E(\hat{H}H^T) = E(\hat{H})H^T = HH^T$. It follows that

$$P = E\,[(\hat{H}-H)(\hat{H}^T-H^T)] = -HH^T + E[\hat{H}\hat{H}^T] \qquad (24)$$

We can now substitute Equations 16 and 23 into 24 noting that all cross product terms involving the noises can be eliminated. The result of obtaining the indicated expected values is

$$P = (X_e^T MX_e)^{-1}[X_e^T MRMX_e + E\{N^T M(Y_e Y_e^T + R)MN\}](X_e^T MX_e)^{-1} \qquad (25)$$

With regard to Equations 23 and 25, there are three degeneracies of definite interest.

1. $N(t) = 0$ for all t. Then $X_s = X_e$ and Equations 23 and 25 compare with $\hat{H}$ and P in Equation 12 for the conventional parameter estimator.
2. Suppose M is a scalar. Then R is a scalar σ_V^2 and Y is a scalar. Then we obtain

$$H = [X_s^T X_s - S]^{-1} X_s^T Y_s$$

$$P = \sigma_V^2 (X_e^T X_e)^{-1} + (Y_e^2 + \sigma_V^2)(X_e^T X_e)^{-1} S (X_e^T X_e)^{-1} \qquad (26)$$

In this case, since P and $\hat{H}$ are independent of M we can then assume that the above value of P is the minimum variance of the estimation error for the case where we are free to choose M. Note that if S = 0 then we have Equations 13 for the conventional parameter estimator.

3. Suppose we can write

$$E\{N^T M(Y_e Y_e^T + R) MN\} = X_e^T M Q M X_e$$

Then P becomes

$$P = (X_e^T M X_e)^{-1} [X_e^T M (R + Q) M X_e] (X_e^T M X_e)^{-1} \quad (27)$$

Comparing Equations 27 and 12 we suspect that for the case where we are free to choose M, P_{min} occurs when $M = (R + Q)^{-1}$.

Let us now investigate the case where we are free to choose M. We are then interested in the conditions under which minimum variance of the estimation error occurs as a function of the weighting matrix M.

3.3 Derivation of the Weighting Matrix to Minimize the Variance of the Error

Minimization of each element of P is equivalent to minimizing the trace of P. Let us accomplish this by writing the optimum value of M as $\hat{M}$. We observe that any other M can be written

$$M = \hat{M} + \varepsilon\eta$$

where η is an arbitrary matrix. Then optimality demands

$$\left.\frac{\partial \text{ trace } P}{\partial \varepsilon}\right|_{\varepsilon=0} = 0 \qquad \text{for all } \eta.$$

From Equation 25 we obtain the following equation by taking the above derivative.

$$0 = \text{trace} \left\{ [X_e^T \hat{M} X_e]^{-1} \left[\begin{array}{l} -[X_e^T \eta X_e][X_e^T \hat{M} X_e]^{-1}[X_e^T \hat{M} R \hat{M} X_e \\ \qquad + E\{N^T \hat{M} (Y_e Y_e^T + R) \hat{M} N\}] \\ +[X_e^T \eta R \hat{M} X_e] + E\{N^T \eta (Y_e Y_e^T + R) \hat{M} N\} \end{array} \right] \times [X_e^T \hat{M} X_e]^{-1} \right\} \tag{28}$$

Then any $\hat{M}$ satisfying Equation 28 results in P being a minimum.

Three degeneracies concerning Equation 28 tend to enhance the understanding of this rather involved expression.

1. $N(t) = 0$ for all t. Then Equation 28 is zero for $\hat{M} = R^{-1}$ which is verified by Equation 12.
2. Suppose M is a scalar. Then so are $R = \sigma_v^2$, Y, and η. In this case Equation 28 is identically zero for all M and any M results in P_{min} as suggested in Equations 13 and 26.
3. Suppose we can write

   $$E\{N^T M(Y_e Y_e^T + R) MN\} = X_e^T M Q M X_e.$$

   Then Equation 28 becomes

   $$0 = \text{trace} \left\{ (X_e^T \hat{M} X_e)^{-1} \left[\begin{array}{l} -(X_e^T \eta X_e)(X_e^T \hat{M} X_e)^{-1} X_e^T \hat{M} (R+Q) \hat{M} X_e \\ + X_e^T \eta (R+Q) \hat{M} X_e \end{array} \right] (X_e^T \hat{M} X_e)^{-1} \right\} \tag{29}$$

   If we choose $\hat{M} = (R+Q)^{-1}$, Equation 29 is zero, hence $\hat{M} = (R+Q)^{-1}$ is optimal as suggested in the last section by Equation 27.

Although Equation 28 is rather involved, general statements can be made. If either $N(t) = 0$ or M is scalar then Equations 12 or 26 provide an estimator with minimum variance of the estimation error and solving Equation 28 is unnecessary. If, however, $N(t) \neq 0$ and M is not scalar then Equation 28 must be solved. The optimal matrix $\hat{M}$ contains, say, n unknown weights. We wish to minimize the trace of P which is then one equation in n variables. Hence the optimal solution is, in general, a dependence of any one weight on all other weights. That is, any one weight can be considered a dependent variable and all other weights are independent. Now since we write $M = \hat{M} + \varepsilon\eta$, then for the ith weight $m_i = \hat{m}_i + \varepsilon\eta_i$. Thus Equation 28 provides n equations which must equal zero, one for each η_i. By selecting any one of these equations and then treating the corresponding value of $\hat{m}_i = f(\hat{m}_j)$ $j \neq i$, then any set of values $\hat{m}_i$, i = 1 to n which satisfy this equation must be an optimal set and therefore all equations multiplying all other η_j, $j \neq 1$ must go to zero. As an example, degenerary 1 calls for $\hat{M} = R^{-1}$. Suppose

$$\hat{M} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix} = \begin{pmatrix} \hat{m}_1 & 0 \\ 0 & \hat{m}_2 \end{pmatrix} \quad (30)$$

If we also investigate the zeros of the equation multiplying η, and treat $\hat{m}_1$ as being dependent, then we obtain a first order equation whose solution is

$$\hat{m}_1 = f(\hat{m}_2) = \frac{(\hat{m}_2 \ \sigma_2^2)}{\sigma_1^2} \quad (31)$$

Now the equation multiplying η_2 might have been used with the result that

$$\hat{m}_2 = f(\hat{m}_1) = \frac{(\hat{m}_1\ \sigma_1^2)}{\sigma_2^2} \tag{32}$$

Now not only does the solution of Equation 30 work, suppose that we chose $\hat{m}_2 = 1$, forcing $\hat{m}_1 = \sigma_2^2/\sigma_1^2$ to satisfy Equation 31. Note that this same selection satisfies also Equations 32. The point to be made is that, to solve Equation 28 we need only obtain one equation multiplying, say, η_i, treat the corresponding $\hat{m}_i$ as being a dependent variable, set the equation equal to zero, and then obtain $\hat{m}_i = f(\hat{m}_j)\ j \neq i$. The equation which must be solved to obtain the desired functionality is first order when degeneracy 3 occurs (assuming, of course, that $N(t) \neq 0$), and fourth order in the general case. The solution of a quartic is by no means pleasant, hence no general answer has yet been obtained. However, two special cases have been investigated and the results appear in the following examples.

3.3.1 Example with a 2 x 2 Weighting Matrix and a Scalar Error Variance

The simplest non degenerate case involves y of dimension 2 x 1 (hence M is not scalar) and correspondingly X is of dimension 2 x 1 (hence singular). Then there is but one parameter so P is a scalar. Thus we have an exact equation

$$\begin{bmatrix} Y_{1e} \\ Y_{2e} \end{bmatrix} = \begin{bmatrix} X_{1e} \\ X_{2e} \end{bmatrix} h$$

There are then available sensor measurements

$$\begin{bmatrix} Y_{1s} \\ Y_{2e} \end{bmatrix} = \begin{bmatrix} Y_{1e} \\ Y_{2e} \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad \left\{ \begin{aligned} & E\begin{bmatrix} V_1(t) \\ V_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \; E\left\{\begin{bmatrix} V_1(t+\tau) \\ V_2(t+\tau) \end{bmatrix} \begin{bmatrix} V_1(t+\tau) & V_2(t+\tau) \end{bmatrix}\right\} \\ & = \begin{bmatrix} \sigma_{V1}^2 & 0 \\ 0 & \sigma_{V2}^2 \end{bmatrix} \delta(t-\tau) \end{aligned} \right.$$

$$\begin{bmatrix} X_{1s} \\ X_{2s} \end{bmatrix} = \begin{bmatrix} X_{1e} \\ X_{2e} \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \quad \left\{ \begin{aligned} & E\begin{bmatrix} N_1(t) \\ N_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \; E\left\{\begin{bmatrix} N_1(t+\tau) & N_2(t+\tau) \end{bmatrix} \begin{bmatrix} N_1(t+\tau) \\ N_2(t+\tau) \end{bmatrix}\right\} \\ & = (\sigma_{N1}^2 + \sigma_{N2}^2)\, \delta(t-\tau) \end{aligned} \right.$$

The problem is to select $\hat{h}$ to minimize, subject to $E(\hat{h})=h$

$$J = \left\| \begin{bmatrix} Y_{1s} \\ Y_{2s} \end{bmatrix} - \begin{bmatrix} \hat{X}_{1e} \\ X_{2e} \end{bmatrix} \hat{h} \right\|_{\begin{vmatrix} m_1 & 0 \\ 0 & m_2 \end{vmatrix}}$$

Equation 23 gives the estimator

$$\hat{h} = \frac{X_{1s}\, m_1\, Y_{1s} + X_{2s}\, m_2\, Y_{2s}}{(X_{1s}^2 - \sigma_{N1}^2)\, m_1 + (X_{2s}^2 - \sigma_{N2}^2)\, m_2}$$

Equation 25 gives the variance of the estimation error

$$P = \frac{m_1^2\, [X_{1e}^2\, \sigma_{V1}^2 + \sigma_{N1}^2\, [Y_{1e}^2 + \sigma_{V1}^2]] + m_2^2\, [X_{2e}^2\, \sigma_{V2}^2 + \sigma_{N2}^2 [Y_{2e}^2 + \sigma_{V2}^2]]}{(X_{1e}^2\, m_1 + X_{2e}^2\, m_2)^2} \tag{33}$$

In order to select m_1 and m_2 to minimize P we can proceed in two ways. First, we might note that the above value of P resulting from Equation 25 could be written

$$P = (X_e^T M X_e)^{-1} \left\{ X_e^T M \overbrace{\begin{bmatrix} \sigma_{V1}^2 & 0 \\ 0 & \sigma_{V2}^2 \end{bmatrix}}^{R} M X_e + X_e^T M \overbrace{\begin{bmatrix} \frac{\sigma_{N1}^2}{X_{1e}^2}(Y_{1e}^2+\sigma_{V1}^2) & 0 \\ 0 & \frac{\sigma_{N2}^2}{X_{2e}^2}(Y_{2e}^2+\sigma_{V2}^2) \end{bmatrix}}^{Q} M X_e \right\}$$

$$\times (X_e^T M X_e)^{-1}\}$$

as pointed out in degeneracy 3 in the last section, the correct answer for the above results by making a comparison with Equation 12 for the conventional weighted least squares estimator. The optimal M is simply

$$\hat{M} = (R + Q)^{-1}$$

So

$$\hat{m}_1 = \frac{1}{\sigma_{V1}^2 + \frac{\sigma_{N1}^2}{X_{1e}^2}(Y_{1e}^2 + \sigma_{V1}^2)}$$

$$\hat{m}_2 = \frac{1}{\sigma_{V2}^2 + \frac{\sigma_{N2}^2}{X_{1e}^2}(Y_{2e}^2 + \sigma_{V2}^2)} \qquad (34)$$

Now the more difficult approach which is more straightforward is to evaluate Equation 28. The resulting equation (which must be true for all η_1 and η_2) is

$$0 = \frac{+\eta_1 \left\{ \begin{array}{l} \hat{M}_1\hat{M}_2 X_{2e}^2 [X_{1e}^2\sigma_{v1}^2+\sigma_{N1}^2(Y_{1e}^2+\sigma_{v1}^2)] \\ - \hat{M}_2^2 X_{1e}^2 [X_{2e}^2\sigma_{v2}^2+\sigma_{N2}^2(Y_{2e}^2+\sigma_{v2}^2)] \end{array} \right\} +\eta_2 \left\{ \begin{array}{l} \hat{M}_2\hat{M}_1 X_{1e}^2 [X_{2e}^2\sigma_{v2}^2+\sigma_{N2}^2(Y_{2e}^2+\sigma_{v2}^2)] \\ - \hat{M}_1^2 X_{2e}^2 [X_{1e}^2\sigma_{v1}^2+\sigma_{N1}^2(Y_{1e}^2+\sigma_{v1}^2)] \end{array} \right\}}{(\hat{X}_{1e}^2\hat{M}_1 + X_{2e}^2\hat{M}_2)^3}$$

Now one could force the product multiplying η_1 to be zero by selecting $\hat{M}_1$ as a dependent variable and one obtains

$$\hat{M}_1 = \hat{M}_2 \frac{X_{1e}^2[X_{2e}^2\sigma_{v2}^2+\sigma_{N2}^2(Y_{2e}^2+\sigma_{v2}^2)]}{X_{2e}^2[X_{1e}^2\sigma_{v1}^2+\sigma_{N1}^2(Y_{1e}^2+\sigma_{v1}^2)]}$$

or equivalently, one could force the product multiplying η_2 to be zero by selecting $\hat{M}_2$ as a dependent variable and obtain

$$\hat{M}_2 = \hat{M}_1 \frac{X_{2e}^2[X_{1e}^2\sigma_{v1}^2+\sigma_{N1}^2(Y_{1e}^2+\sigma_{v1}^2)]}{X_{1e}^2[X_{2e}^2\sigma_{v2}^2+\sigma_{N2}^2(Y_{2e}^2+\sigma_{v2}^2)]}$$

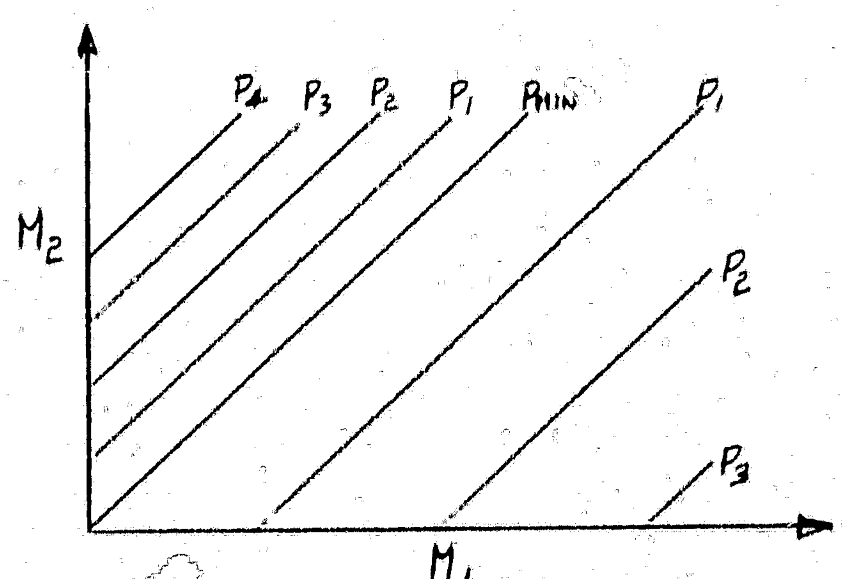
as pointed out in the last section, it is only necessary to force any one product term on η_i to be zero and all other product terms go to zero. (i.e., the ratios $\hat{M}_1/\hat{M}_2$ is the same for both the above equations). Note again that since this problem falls under the category of degeneracy 3, equation 28 results in a first order equation in order to solve for $\hat{M}_i$. The above equations are certainly solved using equations set 34. It is most interesting to note with regard to Equation 34 that large measurements y_{ie} and y_{2e} by multiplying the signal to noise ratios σ_{N1}^2/X_{1e}^2 and σ_{N2}^2/X_{2e}^2 could easily be much larger than the variances σ_{v1}^2 and σ_{v2}^2. Thus, if one selected $\hat{m}_1$ and $\hat{m}_2$ as per conventional least squares theory

$$\hat{m}_1 = \frac{1}{\sigma_{V1}^2} \qquad \hat{m}_2 = \frac{1}{\sigma_{V2}^2}$$

one could select grossly incorrect weightings. The optimal value of P, resulting from the above correct selection is

$$P_{min} = \frac{1}{\dfrac{X_{1e}^2}{\sigma_{V1}^2 + \dfrac{\sigma_{N1}^2}{X_{1e}^2}(Y_{1e}^2 + \sigma_{V1}^2)} + \dfrac{X_{2e}^2}{\sigma_{V2}^2 + \dfrac{\sigma_{N2}^2}{X_{2e}^2}(Y_{2e}^2 + \sigma_{V2}^2)}} \tag{35}$$

A computer program was used to check out the optimum choice of m_1 and m_2. The variance P was evaluated at all combinations of 100 values of m_1 and 100 values of m_2. The values of m_1 and m_2 giving the minimum was selected. It is interesting to note that P depends on the ratio of m_1/m_2 so that the contours of P in the m_1, m_2 plane are straight lines (a plot of m_1/m_2 = constant). The figure below illustrates this



In the above $P_4 > P_3 > P_2 > P_1 > P_{min}$. Now the following table summarizes the two runs made for this problem.

DATA									RESULT	
Case	X_{1e}	X_{2e}	Y_{1e}	Y_{2e}	σ_{V1}^2	σ_{V2}^2	σ_{N1}^2	σ_{N2}^2	$\hat{m}_1/\hat{m}_2$	P_{min}
1	1	2	1	2	100	1000	100	10	.348	808
2	1	2	1	2	100	1000	1	1000	1230.	200

By substituting the above data into Equation 34, the observed optimum ratio $\hat{m}_1/\hat{m}_2$ is verified and by substituting the data into Equation 35, the observed minimum P is verified.

3.32 Example with a 2 x 2 Weighting Matrix and a 2 x 2 Error Variance Matrix

A second problem, similar but more complicated than the last can now be presented. Consider the exact equation

$$\begin{bmatrix} Y_{1e} \\ Y_{2e} \end{bmatrix} = \begin{bmatrix} X_{1e} & X_{2e} \\ X_{3e} & X_{4e} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

There are available, the following sensor outputs

$$\begin{bmatrix} Y_{1S} \\ Y_{2S} \end{bmatrix} = \begin{bmatrix} Y_{1e} \\ Y_{2e} \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \left\{ E \begin{bmatrix} V_1(t) \\ V_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};\ E\left\{ \begin{bmatrix} V_1(t+\tau) \\ V_2(t+\tau) \end{bmatrix} \begin{bmatrix} V_1(t+\tau) V_2(t+\tau) \end{bmatrix} \right\} = \right.$$

$$\begin{bmatrix} \sigma_{V1}^2 & 0 \\ 0 & \sigma_{V1}^2 \end{bmatrix} \delta(t-\tau)$$

$$\begin{bmatrix} X_{1S} & X_{2S} \\ X_{3S} & X_{4S} \end{bmatrix} = \begin{bmatrix} X_{1e} & X_{2e} \\ X_{3e} & X_{4e} \end{bmatrix} + \begin{bmatrix} N_1 & N_2 \\ N_3 & N_4 \end{bmatrix} \left\{ \begin{array}{l} E \begin{bmatrix} N_1(t) & N_2(t) \\ N_3(t) & N_4(t) \end{bmatrix} = \begin{bmatrix} 00 \\ 00 \end{bmatrix}; \\ E \left\{ \begin{bmatrix} N_1(t+\tau) & N_3(t+\tau) \\ N_2(t+\tau) & N_4(t+\tau) \end{bmatrix} \times \begin{bmatrix} N_1(t) & N_2(t) \\ N_3(t) & N_4(t) \end{bmatrix} \right\} = \\ \begin{bmatrix} (\sigma_{N1}{}^2 + \sigma_{N3}{}^2) & \\ 0 & (\sigma_{N2}{}^2 + \sigma_{N4}{}^2) \end{bmatrix} \times \\ \delta(t-\tau) \end{array} \right.$$

The problem is to select $\hat{H}$ to minimize, subject to $E(\hat{H}) = H$

$$J = \left\| \begin{matrix} Y_{1S} \\ Y_{2S} \end{matrix} - \begin{bmatrix} \hat{X}_{1e} & \hat{X}_{2e} \\ \hat{X}_{3e} & \hat{X}_{4e} \end{bmatrix} \begin{matrix} \hat{h}_1 \\ \hat{h}_2 \end{matrix} \right\|_{\begin{bmatrix} m_1 & Q \\ 0 & m_2 \end{bmatrix}},$$

Equation 23 gives the estimator; Equation 25 gives the variance of the estimation error and since X_e is a non singular matrix Equation 28 can be simplified as follows. This provides a means for selecting $\hat{m}_1$ and $\hat{m}_2$ so P is minimized.

$$0 = \text{trace} \left\{ (X_e{}^T \hat{M} X_e)^{-1} \begin{bmatrix} -(X_e^T \eta X_e)(X_e^T \hat{M} X_e)^{-1} E\{N^T \hat{M}(Y_e{}^T Y_e + R)\hat{M}N\} \\ + E\{N^T \eta (Y_e{}^T Y_e + R)\hat{M}N \end{bmatrix} (X_e{}^T \hat{M} X_e)^{-1} \right\} \quad (36)$$

Let us select the matrix η as

$$\eta = \begin{bmatrix} \eta_1 & 0 \\ 0 & 0 \end{bmatrix}$$

Hence we let $\hat{m}_1$ be a dependent variable and evaluate Equation 36. By combining the above equations, factoring out common terms, and writing the result in terms of $(\hat{m}_1/\hat{m}_2)$ the following quartic equation results. (In the general case, with $\hat{m}_i$ being the dependent variable, the equation would be a quartic in $\hat{m}_i$ with the various coefficients being functions of all the other $\hat{m}_j$, $j \neq i$.)

$$0 = \left(\frac{\hat{m}_1}{\hat{m}_2}\right)^4 (X_{1e}^2 + X_{2e}^2)[\alpha X_{2e}^2 + \gamma X_{1e}^2] + \left(\frac{\hat{m}_1}{\hat{m}_2}\right)^3 [X_{1e} X_{3e} + X_{2e} X_{4e}]$$

$$[X_{2e} X_{4e} \alpha + X_{1e} X_{3e} \gamma] - \left(\frac{\hat{m}_1}{\hat{m}_2}\right) [X_{2e} X_{4e} \beta + X_{1e} X_{3e} \delta]$$

$$[X_{2e} X_{4e} + X_{1e} X_{3e}] - [X_{3e}^2 + X_{4e}^2][X_{4e}^2 \beta + X_{3e}^2 \delta] \quad (37)$$

where

$$\alpha = \sigma_{N1}^2 (Y_{1e}^2 + \sigma_{V1}^2)$$
$$\beta = \sigma_{N3}^2 (Y_{2e}^2 + \sigma_{V2}^2)$$
$$\gamma = \sigma_{N2}^2 (Y_{1e}^2 + \sigma_{V1}^2)$$
$$\delta = \sigma_{N3}^2 (Y_{2e}^2 + \sigma_{V2}^2)$$

General procedures exist for solving quartic equations but the work is rather tedious. For the present, it suffices by example to show that optimal solutions do solve the above equation. For specific choices of the variables listed above the trace of

P was calculated on a computer for all combinations of 100 values of m_1 and 100 values of m_2. The resulting (trace of P) is a function of the ratio m_1/m_2, so contours of constant (trace P) in the $(m_1 - m_2)$ plane appear as straight lines (plots of m_1/m_2 = constant). The figure shown in the last section represents the contours. The following table summarizes the results of the computer study.

	DATA												RESULT	
case	X_{1e}	X_{2e}	X_{3e}	X_{4e}	Y_{1e}	Y_{2e}	σ^2_{N1}	σ^2_{N2}	σ^2_{N3}	σ^2_{N4}	σ^2_{V1}	σ^2_{V2}	$\frac{\hat{m}_1}{\hat{m}_1}$	trace P_{min}
1	1	2	3	4	1	2	100	1	1000	10	100	1000	7.7	5×10^6
2	1	2	3	4	1	2	1	10	10	1000	10	10	9.2	2.8×10^4
3	1	2	3	4	1	2	1	0	0	1000	100	1000	25.8	6.1×10^5

For example, if the data of case 3 is used, we desire the solution of

$$0 = \left(\frac{\hat{m}_1}{m_2}\right)^4 (2020) + \left(\frac{\hat{m}_1}{m_2}\right)^3 (8888) - \left(\frac{\hat{m}_1}{m_2}\right) 33(10)^6 - 225(10)^6 \quad (38)$$

The observed optimal value of 25.8 when substituted into the above satisfies the equation, demonstrating that the optimum ratio is indeed a root of the above. Similarly, the optimal solution for cases 1 and 2 also satisfy Equation 37.

Additional computer runs were made for the above 3 cases where the matrix M was chosen

$$M^1 = \begin{bmatrix} 1 & m_2^1 \\ m_2^1 & m_1^1 \end{bmatrix}$$

rather than

$$M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \quad (40)$$

It is interesting to note (using case 3 as an example) that the optimum was $\hat{m}_2^1 = 0$, $\hat{m}_1^1 = .0383$. Then the ratio $1/\hat{m}_1^1$ is comparable to $\hat{m}_1/\hat{m}_2$. Then

$$\frac{1}{\hat{m}_1} = \frac{1}{.0383} = 26.1$$

which compare with $\hat{m}_1/\hat{m}_2 = 25.8$. The point is that nothing is gained by considering a non-diagonal M matrix since the end result appears to reduce to a diagonal matrix. This result is true because the measurement noises are all independent and the covariance matrices are diagonal. Hence, since no cross product terms appear in the covariance matrices, no cross product terms are necessary in the weighting matrix.

4. Repeated Data

The dimensions of the weighted least squares estimation problem follow directly from the exact equation

$$\underset{m\times 1}{Y_e} = \underset{(m\times n)}{X_e} \; \underset{n\times 1}{H} \quad (41)$$

The dimensions of all other variables follow Equation 41. For example, the variance of the estimation error is a matrix P of dimension n×n, the dimension being dictated by the parameter matrix H. Similarly, the weighting matrix M is m×m, this dimension being determined by Y.

Suppose that Equation 41 represents the result of making p repetitions of the same measurement. Then the dimensions could be written

$$\underset{(pm'\times 1)}{Y_e} = \underset{(pm'\times n)}{X_e} \quad \underset{(n\times 1)}{H}$$

After many such repetitions, the vectors and matrices could become exceedingly large in size. Fortunately, calculating the estimation equation and the variance of the estimation error involves products such

$$\underset{(n\times pm')}{X_e^T} \quad \underset{(pm'\times pm')}{M} \quad \underset{(pm'\times n)}{X_e}$$

Now this product is, in effect, the sum of taking p products of the form

$$\underset{(n\times m')}{X_{e_i}^T} \underset{(m'\times m')}{M_i} \underset{(m'\times n)}{X_{e_i}} \qquad 1 \leq i \leq p$$

The subscript i refers to the measurement number, hence X_{e_i} is a matrix of dimension (m'×n), representing the data from the ith repetition. Then we can write

$$X_e^T M X_e = \sum_{i=1}^{p} (X_{e_i}^T M X_{e_i})$$

In an analogous manner, all the operations previously described can be translated directly into a summation format. Now the simulation in Section 5 involves repetitions of a single measurement Y_e (i.e., $m'=1$). In this case the matrix M (of dimension p×p) can be reduced to a scalar if the weightings are not time dependent, but rather, the matrices M_i are all equal. As previously discussed, when M is a scalar the weights cancel out of the estimation equation for $\hat{H}$ and of the variance of the estimation error P (see Equation 26). For use in the next section, Equations 26 are translated into summation format. For simplicity, the subscripts i are dropped and the summations are assumed to be over all p repetitions of the measurement.

$$\underset{(n\times 1)}{\hat{H}} = \underset{(n\times n)}{[\Sigma(X_s^T X_s S)]^{-1}} \; \underset{(n\times 1)}{[\Sigma X_s^T Y_s]} \tag{42}$$

$$\underset{(n\times n)}{P} = \underset{(1\times 1)}{\sigma_v^2} \underset{(n\times n)}{[\Sigma X_e^T X_e]^{-1}} + \underset{(1\times 1)}{[(\Sigma(Y_e^2 + \sigma_v^2)]} \underset{(n\times n)}{[\Sigma X_e^T X_e]^{-1}} \underset{(n\times n)}{S} \underset{(n\times n)}{[\Sigma X_e^T X_e]^{-1}}$$

Note that all resulting matrices are of dimension n. Hence, no matter how large the matrices of Equation 41 become, the matrices which must actually be used to compute $\hat{H}$ and P are always of dimension n (the number of parameters to be estimated).

5. Simulation Demonstrating the Parameter Estimator

Consider the following exact equation

$$\ddot{\theta}_e = [\alpha_e \;\; \delta_e] \begin{bmatrix} M_\alpha \\ M_\delta \end{bmatrix}$$

where θ_e, α_e, and δ_e are measureable states. Suppose sensors provide noisy measurements

$$\ddot{\theta}_s = \ddot{\theta}_e + V \qquad E(V(t)) = 0;\ E[V(t+\tau)V(t)] = \sigma_V^{\,2}\,\delta(t-\tau)$$

$$\begin{aligned} \alpha_s &= \alpha_e + N_\alpha \\ \delta_s &= \delta_e + N_\delta \end{aligned} \qquad E\begin{bmatrix} N_\alpha(t) \\ N_\delta(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad E\left\{[N_\alpha(t+\tau)\,N_\delta(t+\tau)]\begin{bmatrix} N_\alpha(t) \\ N_\delta(t) \end{bmatrix}\right\} = (\sigma_\alpha^2 + \sigma_\delta^2)\ \delta\ (t-\tau)$$

It is assumed that p repetitions of this measurement are to be made. However, as discussed previously, since $\ddot{\theta}_e$ is a scalar and since the weighting matrix is the same for each measurement, then the estimation equation and the variance of the estimation error become independent of M, hence minimum variance follows once the unbiased estimator is arrived at. In this context, weighted least squares yields the same results as non-weighted ($\underset{\sim}{M} = I$) least squares. Hence, it suffices to consider selecting $\begin{bmatrix} \hat{M}_\alpha \\ \hat{M}_\delta \end{bmatrix}$ to minimize the following objective function. Summation format is now used

$$J = \left|\left|\Sigma\ \left(\ddot{\theta}_s - [\hat{\alpha}_s\ \ \hat{\delta}_s]\begin{bmatrix} \hat{M}_\alpha \\ \hat{M}_\delta \end{bmatrix}\right)\right|\right|$$

Minimization is subject to $E\begin{bmatrix} \hat{M}_\alpha \\ \hat{M}_\delta \end{bmatrix} = \begin{bmatrix} M_\alpha \\ M_\delta \end{bmatrix}$.

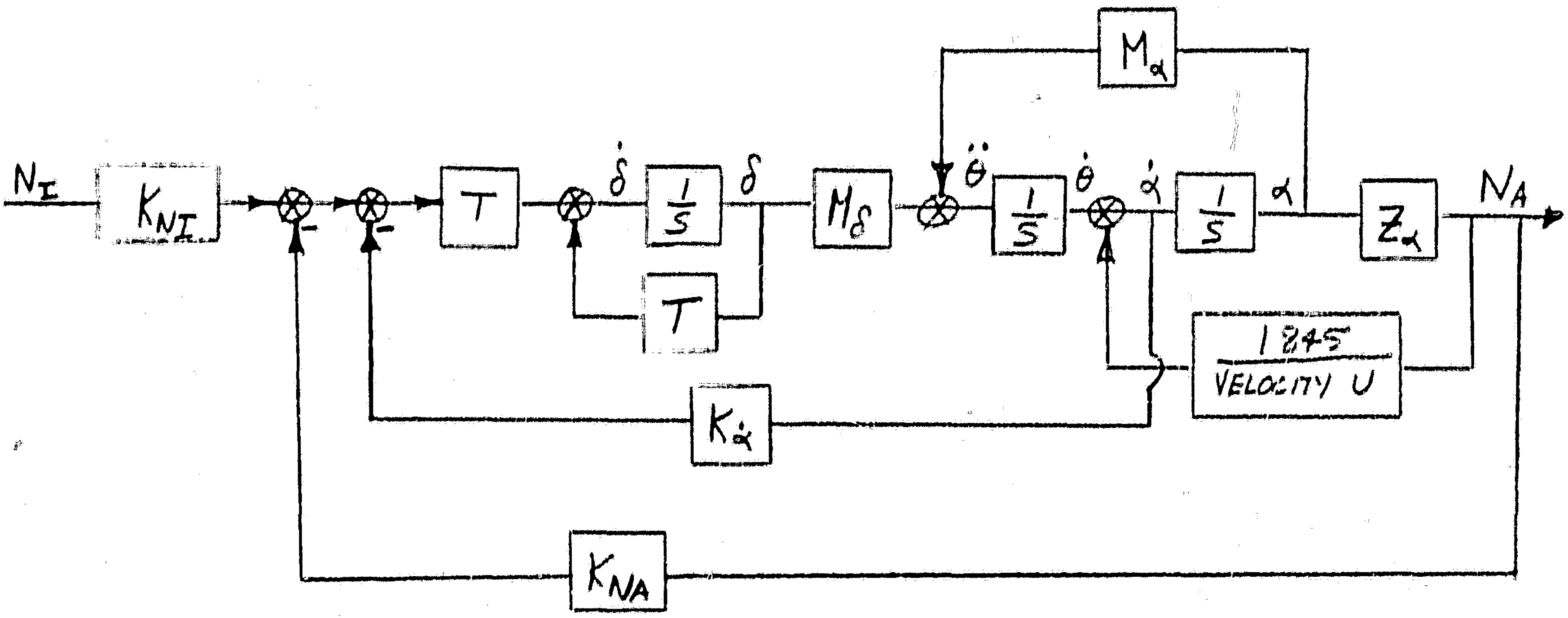
The solution to this problem is presented in general in Section 3.1. For the case of scalar M, we have Equations 26. In summation format, we have Equations 42. Then the above equations may be substituted directly into Equations 42. The result is an unbiased estimator for M_α and M_δ. The variance of the estimation error is also as small as possible for this class of unbiased parameter estimators.

$$\begin{bmatrix} \hat{M}_\alpha \\ \hat{M}_\delta \end{bmatrix} = \begin{bmatrix} \Sigma(\alpha_s^2 - \sigma_\alpha^2) & (\Sigma\ \alpha_s\ \delta_s) \\ (\Sigma\ \alpha_s\ \delta_s) & (\Sigma(\delta_s^2 - \sigma_\delta^2)) \end{bmatrix}^{-1} \begin{bmatrix} \Sigma\ \alpha_s\ \ddot{\theta}_s \\ \Sigma\ \delta_s\ \ddot{\theta}_s \end{bmatrix} \tag{43}$$

$$P = \frac{\sigma_v^{\ 2}}{(\Sigma\alpha_e^2)(\Sigma\delta_e^2)-(\Sigma\alpha_e\delta_e)^2} \begin{bmatrix} \Sigma\ \delta_e^2 & -\Sigma\ \alpha_e\delta_e \\ -\Sigma\ \alpha_e\delta_e & \Sigma\ \alpha_e^2 \end{bmatrix} + \frac{\Sigma(\ddot{\theta}_e^2 + \sigma_v^2)}{[(\Sigma\alpha_e^2)(\Sigma\delta_e^2)-(\Sigma\alpha_e\delta_e)^2]^2}$$

$$\times \begin{bmatrix} [\sigma_\alpha^2\ (\Sigma\delta_e^2) + \sigma_\delta^2(\Sigma\alpha_e\delta_e)^2] & \begin{bmatrix} -(\Sigma\alpha_e\delta_e)[\sigma_\alpha^2(\Sigma\delta_e^2) \\ +\sigma_\delta^2(\Sigma\alpha_e)^2 \end{bmatrix} \\ \begin{bmatrix} -(\Sigma\alpha_e\delta_e)[\sigma_\alpha^2(\Sigma\delta_e)^2 \\ +\sigma_\delta^2\ (\Sigma\alpha_e)^2 \end{bmatrix} & \begin{bmatrix} -\sigma_\delta^2(\Sigma\alpha_e^2)^2 \\ +\sigma_\alpha^2\ (\Sigma\alpha_e\delta_e)^2 \end{bmatrix} \end{bmatrix}$$

In order to test Equations 43, the control system shown in Figure 2 was used. Essentially, it is a third order pitch plane normal acceleration control system with first order actuator dynamics and second order pitch plane (plant) dynamics.



$$\frac{N_A}{N_I} = \frac{M_\delta \; Z_\alpha \; K_{NI} \; T}{S^3 + S^2 \left[-Z_\alpha \frac{1845}{U} + T\right] + S \left[Z_\alpha \frac{1845}{U} T + K_{\dot{\alpha}} M_\delta T - M_\alpha\right] + \left[M_\delta Z_\alpha K_{NA} T - M_\alpha T\right]}$$

Let $$K_{NI} = \frac{C_2}{M_\delta Z_\alpha T} \quad ; \quad T >> \frac{Z_\alpha 1845}{u}$$

$$K_{NA} = \frac{C_2 + M_\alpha T}{M_\delta Z_\alpha T}$$

$$K_{\dot{\alpha}} = \frac{C_1 - Z_\alpha \frac{1845}{U} T + M_\alpha}{M_\delta T}$$

Then $$\frac{N_A}{N_I} = \frac{C_2}{S^3 + T S^2 + C_1 S + C_2}$$

Figure 1 - Third Order Pitch Plane Normal Acceleration Controller

In Figure 1, the following symbols are used:

M_α = Aerodynamic stability derivative ($°/sec^2/°$ of α)

M_δ = Fin effectiveness gain ($°/sec^2/°$ of δ)

T = Actuator time constant (1/sec)

U = Velocity along center line (ft/sec)

Z_α = Normal force coefficient (g's/° of α)

α = Angle of attack (°)

δ = Fin deflection (°)

θ = Attitude angle (°)

As shown in the figure, the parameters M_α, M_δ, and Z_α are utilized to calculate control gains such that system response characteristics to command are constant (corresponding to a second order system which is .6 damped with a 2 cps bandwidth).

Two basic modes of operation are possible--off line and on line. In the off line mode, all system gains are updated using actual parameter values while the estimator processes the resulting data in an attempt to reconstruct the parameter values. The off line mode was used in the simulation so that nominal data results, data unaffected by parameter estimates. Hence, the covariance of the estimation error is given by Equation 43. It is assumed that the parameter Z_α is known whereas M_α and M_δ are not known. The second mode of operation is the on-line mode wherin parameter estimates are used to calculate the control system gains. In this case, the variance of the estimation error is more complicated since data such as Y(t) becomes Y(t,H), that is, a function of both time and previous parameter estimates. For the present then, the off-line mode is used to derive data for use by the estimator.

A word should be said about the nature of the estimation process. When no noise is present, then $\hat{M}_\alpha$ and $\hat{M}_\delta$ are deterministic time functions $\hat{M}_\alpha(t)$ and $\hat{M}_\delta(t)$. When noise is present, $\hat{M}_\alpha$ and $\hat{M}_\delta$ are stochastic processes $\hat{M}_\alpha(t, \xi_i)$, $\hat{M}_\delta(t, \xi_i)$, that is, functions of both time and the outcome of the i^{th} probabalistic experiment, here the selection of a sequence of noise values. Hence, for any selection of a sequence of noise values, there follows time functions $\hat{M}_\alpha(t)$ and $\hat{M}_\delta(t)$. Similarly, at any one time point, $\hat{M}_\alpha(\xi)$ and $\hat{M}_\delta(\xi)$ are random variables with a mean $\zeta_{\hat{M}_\alpha}$ and $\zeta_{\hat{M}_\delta}$ and variances $\sigma^2_{\hat{M}_\alpha}$ (the (1,1) term of the P matrix) and $\sigma^2_{\hat{M}_\delta}$ (the (2,2) term of the P matrix).

In order to verify the predicted variances $\sigma^2_{\hat{M}_\alpha}$ and $\sigma^2_{\hat{M}_\delta}$ each case involving noise was repeated 20 times and unbiased estimates of the mean and variance of both $\hat{M}_\alpha$ and $\hat{M}_\delta$ were made. For an N sample ensemble, one can estimate the mean ζ of a random variable X by the sample mean $\bar{X}$.

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} \tag{44}$$

The variance is estimated by

$$\hat{\sigma}^2_X = \frac{\sum_{i=1}^{N} X_i^2}{N-1} - \left(\frac{N}{N-1}\right)\bar{X}^2 \tag{45}$$

These estimates are unbiased; that is

$$E(\bar{X}) = \zeta \qquad\qquad E(\hat{\sigma}^2_X) = \sigma^2_X$$

Equations 44 and 45 were used to verify that unbiased estimates of $\hat{M}_\alpha$ and $\hat{M}_\delta$ resulted and that the variance of the estimation error for $\hat{M}_\alpha$ and $\hat{M}_\delta$ was as predicted above.

Two cases were run with each case being repeated 20 times. In both cases

$$M_\alpha = 500$$

$$M_\delta = 1000$$

$$Z_\alpha = 0$$

$$\text{INITIAL g's} = 10$$

In each case, the plant dynamics were simulated by taking a set of first order linear differential equations

$$\dot{X}(t) = A\ X(t) + B\ U(t)$$

and transforming these equations into the form

$$X(t+\tau) = \phi(\tau)\ X(t) + \phi(\tau)\ U(t)$$

The transition matrices were calculated from the Taylor series expansion for an exponential. In all 40 passes (that is 20 repetitions of the 2 cases), 100 iterations were made, each simulating a .01 second duration. Additional 10 g commands were called for every 10 iterations (.1 seconds). This forced the plant to be quite active. The results are shown in the following table. The data included here, refers to the final or 100th iteration.

Case	Data			Mean		Predicted	Observed	Predicted	Observed
	σ_v^2	σ_α^2	σ_δ^2	$\bar{M}_\alpha$	$\bar{M}_\delta$	$\sigma^2_{\hat{M}_\alpha}$	$\sigma^2_{\hat{M}_\alpha}$	$\sigma^2_{\hat{M}_\delta}$	$\sigma^2_{\hat{M}_\delta}$
1	100	0	0	500.2	1000.3	56.4	58.5	238.6	248.5
2	100	.006	.0015	540.1	1082.6	4737	7627	30045	31828

From the above data it is apparent that adding some noise to measurements of α and δ has an appreciable effect on the accuracy of the estimator. The simple rational for this occurrence follows from consideration of the equation

$$Y_s = X_e H$$

If X_e is square and invertible we could write

$$\hat{H} = X_s^{-1} Y_s$$

Now if Y_s contains some noise, $\hat{H}$ might be slightly inaccurate. On the other hand, since X_s must be inverted, a relatively small error in each element of X_s could cause a disproportionately large error in $\hat{H}$, which is the case above.

Case 2 was rerun using a conventional parameter estimator (that is, one in which σ_α^2 and σ_δ^2 are zero in Equation 43). The result, as expected, was a biased estimator

$$\bar{M}_\alpha = 308 \qquad \hat{\sigma}^2_{\hat{M}_\alpha} = 478$$

$$\bar{M}_\delta = 606 \qquad \hat{\sigma}^2_{\hat{M}_\delta} = 19600$$

6. Outline for Future Work

With regard to the problem presented in this paper, it would follow that one should investigate ways of obtaining the noise variances assuming that they are unknown. Then, an on-line simulation should be made to investigate the ability of the estimator to provide invariant response characteristics.

A further extension would be to time varying parameters. One should consider

1) Given the plant whose equations are

$$\dot{X}(t) = A(t)\ X(t) + B(t)u(t)$$

How can control gains K(t) be selected such that the feedback signal $u(t) = K(t)\ X(t)$ renders the response time invariant? What if u(t) must change only at fixed time points?

2) Suppose the parameters contained in H are time dependent in that they depend linearly on certain states. Thus $H = Z_e H'$. Then, given the exact equation

$$Y_e = Z_e H' X_e$$

where Y_e, X_e, and Z_e may be sensed with noise, how can H' be estimated to obtain an unbiased estimate? What must be done to ensure minimum variance of the estimation error?

3) Can the estimator of 2) be used to provide the necessary control gains for 1) such that nearly constant response characteristics follow?

It is suggested that the above could be applied to a linear control system problem.

Kalman Filter Problem - No measurement noise

Abstract

Suppose that the measurement noise for a Kalman filter approaches zero. It is shown in this discussion that the result is a filter suggested by conventional state variable feedback and estimation techniques.

State Variable Approach

Consider the linear constant parameter closed loop system shown in Figure 1. With control gains on all states, one can then place the closed loop poles anywhere in the complex plane. The closed loop transfer function (assuming zero initial conditions) is

$$\frac{Y}{Y_C}(S) = [SI - A + BK_1]^{-1} BK_2 \qquad (1)$$

Suppose that not all the states (X) are sensed, but rather only certain measurements (X_S) are available. If these measurments are not noisy, Figure 2 suggests using a filter to reconstruct (estimate) the states X using the measurements X_S. This is the conventional state variable approach. One may use the results of the modified observer design included in the previous report.

As an example, Figure 3 shows a second order, unity gain, system with feedback gains on the states X_1 and $\dot{X}_1$. With these gains the closed loop transfer function is

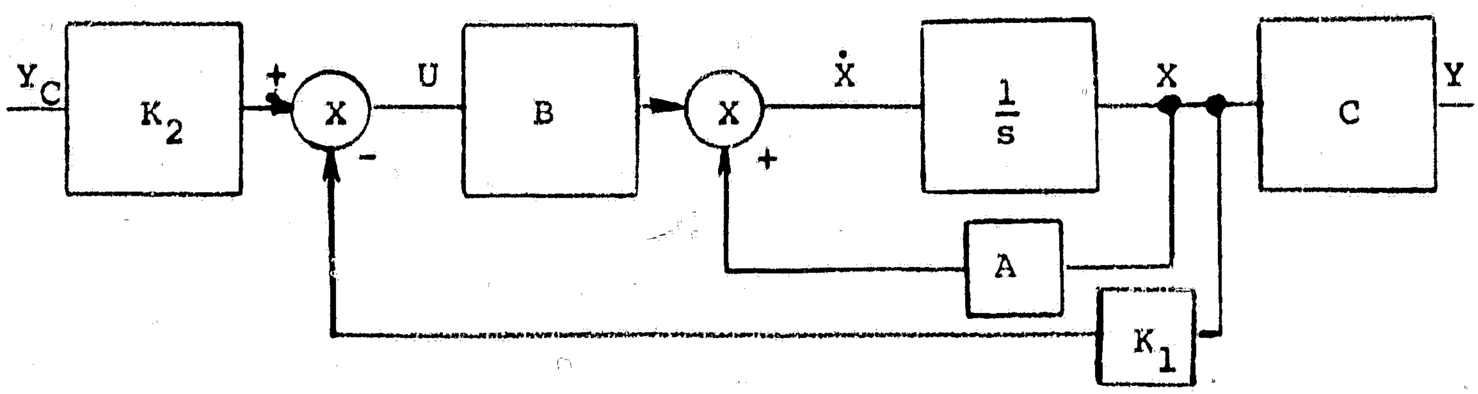


FIGURE 1 - Linear Constant Parameter Closed Loop System

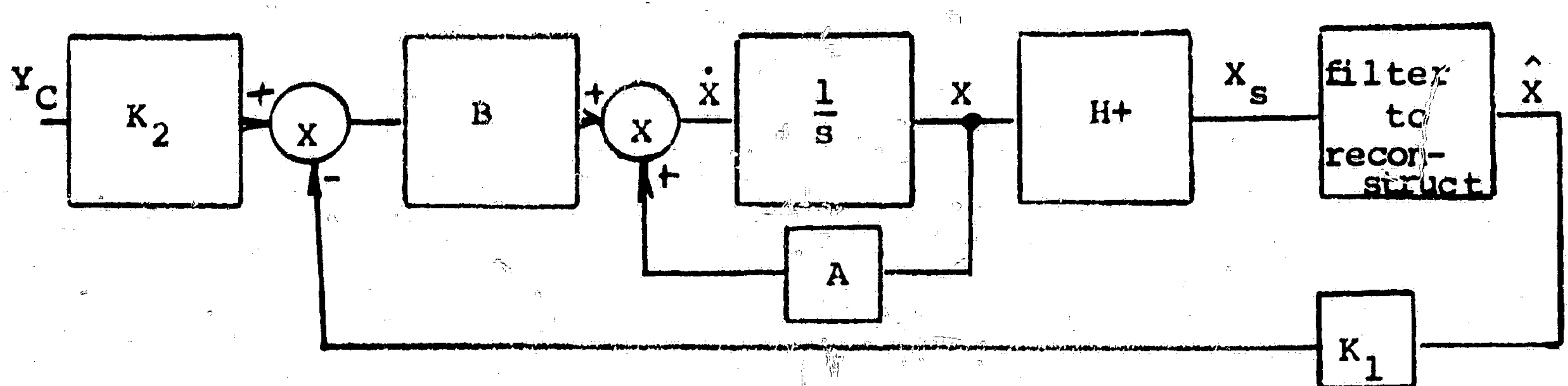


FIGURE 2 - State Variable Estimator

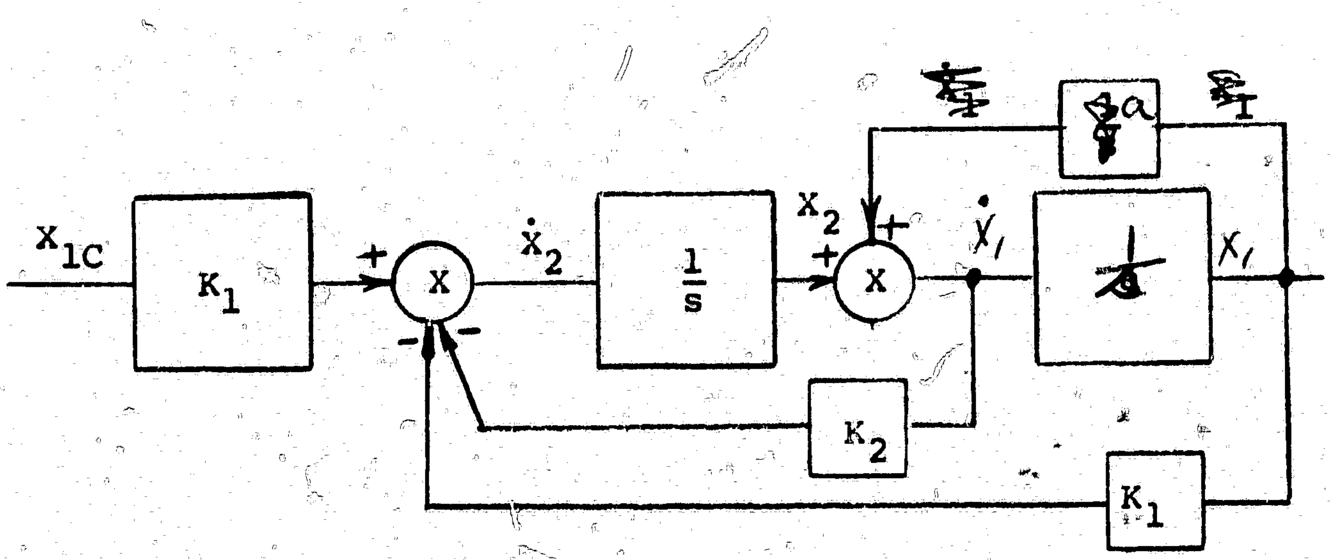


FIGURE 3 - Second Order Unity Gain System

$$\frac{X_1}{X_{1c}}(S) = \frac{K_1}{S^2 + (K_2 - a)S + K_1} \tag{2}$$

Suppose that only X_1 is available as a measurement and that we must estimate $\dot{X}_1$ for feedback purposes. The other quantity $(\dot{X}_2)$ is THEN known since K_1, K_2, X, and X_{1C} are all known. Simple block diagram manipulations can be used to obtain Figure 4. As usual some approximation such as $\frac{ps}{s+p}$ as $p \to \infty$ must be used to estimate $\dot{X}_1$.

Kalman Filter Approach

State estimation in the presence of noise, especially white gaussian noise, suggests the use of a Kalman filter. Shown in Figure 5 are the Kalman gains (K_K) the feedback gains K_1, and the Kalman filter configuration for a linear constant coefficient control system. The net result is that the transfer function of Equation 1 is realized as in Figures 1 and 2. The performance of the filter portion of the system is best analyzed by considering the transfer function $\frac{\hat{X}}{u}(s)$. To derive this transfer function the forward-path part of Figure 5 can be written

$$\begin{bmatrix} \dot{X} \\ \dot{\hat{X}} \end{bmatrix} = \begin{bmatrix} A & 0 \\ K_K H & A-K_K H \end{bmatrix} \begin{bmatrix} X \\ \hat{X} \end{bmatrix} + \begin{bmatrix} B \\ B \end{bmatrix} U \tag{5}$$

$$\hat{X} = [0 \quad I] \begin{bmatrix} X \\ \hat{X} \end{bmatrix}$$

Then it immediately follows that

$$\frac{\hat{X}}{U}(S) = [0 \ I] \begin{bmatrix} (SI-A) & 0 \\ -K_K H & (S-A+K_K H) \end{bmatrix}^{-1} \begin{bmatrix} B \\ B \end{bmatrix} \tag{6}$$

which can be written more simply as

$$\frac{\hat{X}}{U}(S) = [SI + K_K H - A]^{-1} K_K [SI-A]^{-1} B + [SI+K_K H-A]^{-1} B \tag{7}$$

This result follows from the identity

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1}$$

where

$$C_{22} = [A_{22} - A_{21}A_{11}^{-1}A_{12}]^{-1}$$

$$C_{12} = -A_{11}^{-1}A_{12}C_{22}$$

$$C_{21} = -C_{22}A_{21}A_{11}^{-1}$$

$$C_{11} = -A_{11}^{-1}(I_{11} - A_{12}C_{21})$$

Now comparing Equation 7 with Figure 5 note that the first term on the right side of Equation 7 represents the lower part of

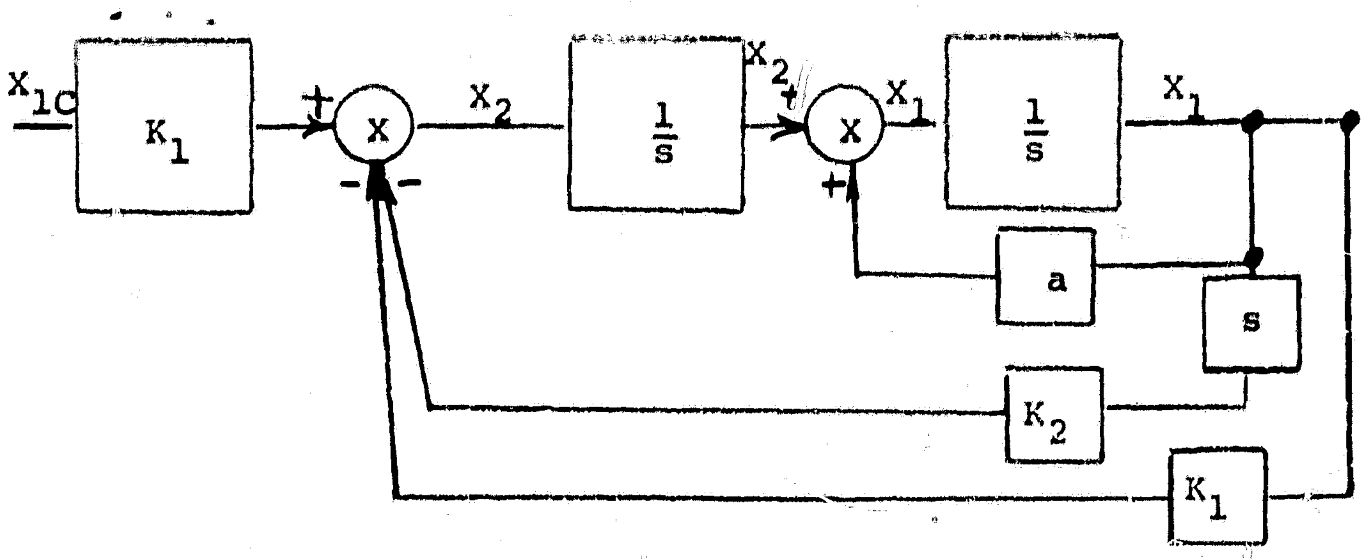


FIGURE 4 - Use of a State Variable Estimator

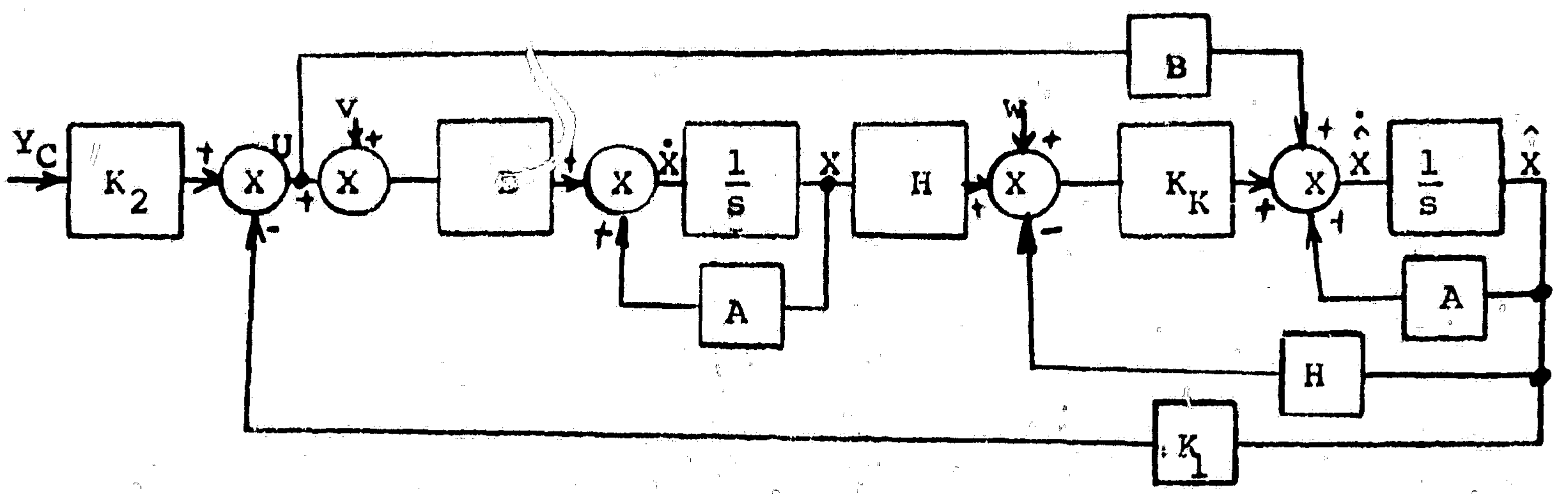


FIGURE 5 - Kalman Filter for a Closed Loop System

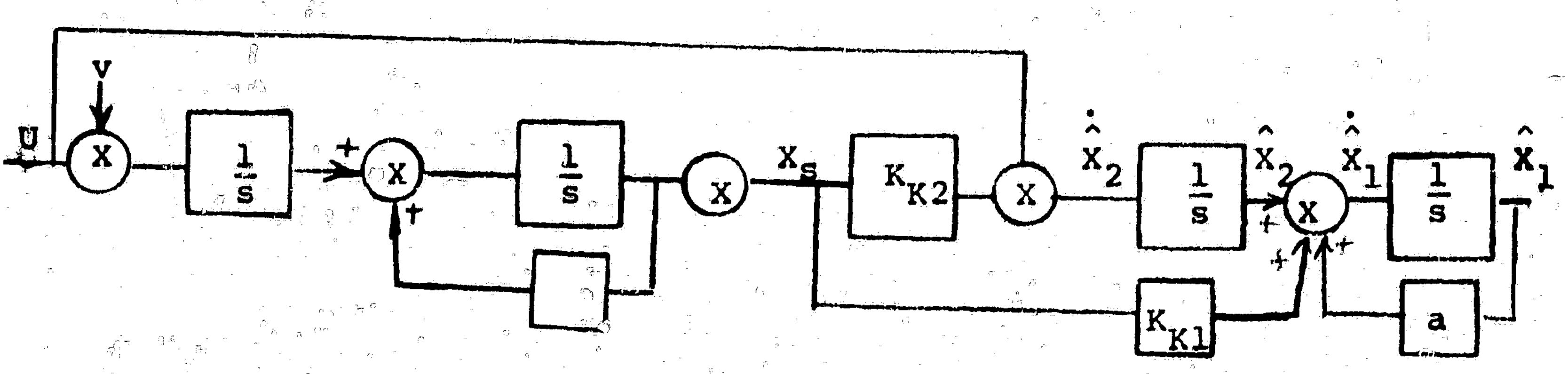


FIGURE 6 - Kalman Filter for the Second Order System of Figure 3

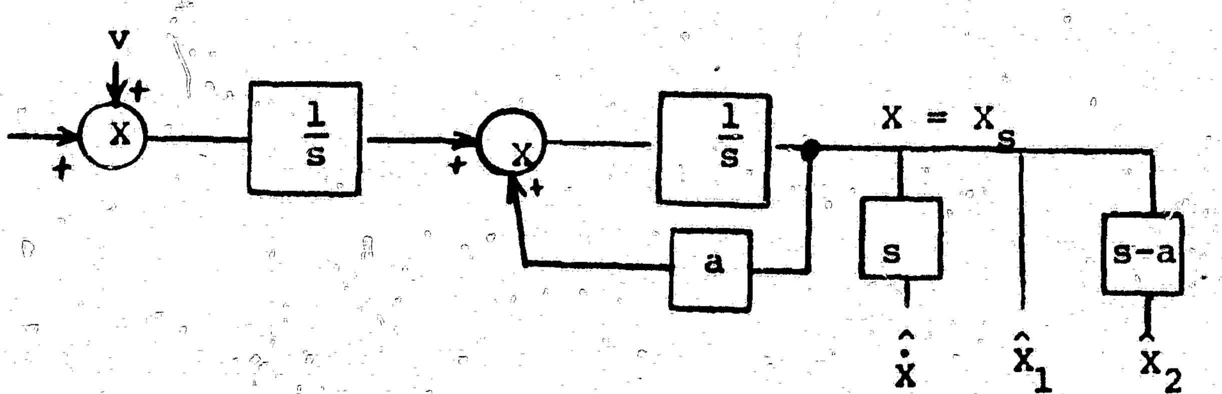


FIGURE 7 - Kalman Filter for No Measurement Noise

the forward path (that part including the plant and the model) while the second term on the right side of Equation 7 represents the upper part of the forward path (that part including only the model).

Suppose we write

$$\frac{\hat{X}}{U}(s) = \left(\frac{\hat{X}}{U}\right)_{LOWER} + \left(\frac{\hat{X}}{U}\right)_{UPPER}$$

Now when the noise w is present but the noise v is absent, an examination of the matrix Riccati equation reveals that the Kalman gains (K_K) approach zero. Therefore

$$\left(\frac{\hat{X}}{u}\right)_{LOWER} = 0$$

$$\frac{\hat{X}}{U}(s) = \left(\frac{\hat{X}}{U}\right)_{UPPER} = (sI - A)^{-1}B$$

In this case the Kalman filter considers the measurement X_s to be noisy, hence the measurement is completely ignored and the parameter estimates follow directly from the model.

When noises w and v are present, both the lower and upper paths are activated. We can combine these as follows

$$\frac{X}{U}(s) = [sI + K_K H - A]^{-1}[K_K H(sI - A)^{-1} + (sI - A)(sI - A)^{-1}]B$$

$$= [sI - A]^{-1}B$$

Thus we also get $(sI - A)^{-1}B$ with both paths being excited.

Suppose we now consider the case under discussion (i.e. $w = 0$ and $v \neq 0$). A consideration of the matrix Ricatti equation reveals that the Kalman gains become infinite. Then we see from Equation 7

$$\left(\frac{\hat{X}}{U}\right)_{upper} = 0$$

$$\frac{\hat{X}}{u}(s) = \left(\frac{\hat{X}}{u}\right)_{LOWER} = [K_K H]^{-1}[K_K H][sI - A]^{-1}B = [sI - A]^{-1}B$$

In this case, the Kalman filter considers the measurement X_s to be reliable and only the lower path is activated with the m model being used to reconstruct the unavailable states. Such operation is identical with state variable estimation procedures mentioned earlier. To emphasize this point consider Figure 6, the Kalman filter for the forward path of the second order system of Figure 3. As the noise w approaches zero, the Kalman gains K_{K1} and K_{K2} approach infinity, hence only the lower portion of the system is activated. Let us write the transfer functions relating the estimates to the measurement.

$$\frac{\hat{X}_1}{X_s}(s) = \frac{s\,K_{K1} + K_{K2}}{s^2 + (K_{K1} - a)s + K_{K2}}$$

$$\frac{\hat{X}_2}{X_s}(s) = \frac{K_{K2}(s-a)}{s^2 + (K_{K1} - a)s + K_{K2}}$$

Now, as K_{K1} gets large we note that

$$\underset{K_{K1}\to\infty}{\text{LIMIT}} \frac{\hat{X}_2}{X_s}(s) = \underset{K_{K1}\to\infty}{\text{LIMIT}} \frac{s\,K_{K1}}{s\,K_{K1}} = 1$$

Similarly, as K_{K2} gets large

$$\lim_{K_{K2}\to\infty} \frac{\hat{X}_2}{X_s}(s) = \lim_{K_{K2}\to\infty} \frac{K_{K2}(s - a)}{K_{K2}} = s - a$$

as a result we have

$$X_1 = X_s$$

$$X_2 = (s - a)X_s$$

since $\dot{X}_1 = X_2 + aX_1$ we get

$$\dot{X}_1 = X_2 + aX_1 = sX_s$$

Then Figure 6 approaches Figure 7 as the noise w approaches zero. If we add the control gains to Figure 7 and close the loop, we have Figure 4 which resulted from state variable estimation procedures.

Conclusion

In the limit as the measurement noise goes to zero, the Kalman Filter reduces to a state variable estimator derivable from block diagram manipulations to reconstruct, for feedback purposes, unavailable states.

If the input noise is absent then the output X_s is ignored. In either case, after the states are estimated, the state feedback K_1 is selected to either achieve the desired transfer function or to minimize a quadratic performance index. In the absence of zeros these are equivalent operations (see Schultz and Melsa).

RANDOM SEARCH

ABSTRACT

Many problems arising in engineering and operations research contexts have the following structure: The decision maker is provided with a class $\mathcal{F}$ of functions, whose common domain $\mathcal{X}$, a bounded set is specified. Some mechanism selects a function f from $\mathcal{F}$. The decision maker is not informed of this choice. He would like somehow to find a point x^* at which f assumes its maximum value (denoted by $\|f\|$). Toward this end, the decision maker may sequentially and without constraint select elements $x_1, x_2, \ldots$ from $\mathcal{X}$. Upon choosing x_n, he is informed of the value $f(x_n)$. Thus the decision maker may come to learn certain features of f. Any (perhaps randomized) strategy for choosing x_n on the basis of the pairs $\{(x_j, f(x_j))\}_{j=1}^{n-1}$ will be termed a <u>search procedure</u>. The problem of finding a search procedure S under which, for all $f \in \mathcal{F}$, $f(x_n)$ converges to $\|f\|$, in some specified sense, has generated a lively body of research papers, some of which will be referenced and described in the present paper.

For an example of the sort of engineering question giving rise to a search problem, suppose that an airplane is to fly in a fixed direction and speed. Its fuel efficiency will then be a function of the carburation setting. If x is the relative mixture of fuel and air, and f(x) the associated fuel consumption required to maintain the aircraft's velocity, then the framework for a search problem is present. $\mathcal{X}$ may be taken to be the unit interval and $\mathcal{F}$, perhaps, may be considered to be the set of continuous functions on the unit interval.

Under certain restrictions on $\mathcal{X}$ and $\mathcal{F}$, effective search procedures have been revealed. The most publicized of these is the "gradient method" which, in its simplest form, determines x_{j+1} from x_j by estimating the gradient ∇f of f at x_j (by difference approximations derived from local samples) and then setting $x_{j+1} = x_j + \lambda \nabla f(x_j)$. λ is chosen from heuristic considerations and may vary as the process evolves. If the functions of $\mathcal{F}$ are concave and $\mathcal{X}$ is bounded and sufficiently regular, the gradient method will provide a Cauchy sequency $\{f(x_j)\}$ converging to $\|f\|$. Hadley's book Nonlinear and Dynamic Programming [1] devotes a nicely written chapter to the gradient method and its variations. The review paper by Spang [2] has an extensive bibliography on the gradient method.

J. Kiefer [3,4] has published interesting analyses for the case that $\mathcal{X}$ is a bounded interval in the real line. In particular, under the search procedure he proposes, in n trials (the number n must be specified in advance) the point x^* at which $f(x^*) = \|f\|$ can be located within a distance of $1/L_n$, where L_n is the nth Fibonacci number, provided $\mathcal{F}$ is the set of concave functions on $\mathcal{X}$. Further, the search procedure is minimax in the sense that no non-randomized strategies can improve on this error tolerance uniformly in $\mathcal{F}$. Bellman and Dreyfus [5] devote a chapter to this optimization approach. To this writer's knowledge, and analogous search which also posseses the minimax property has yet to be revealed for multi-dimensional $\mathcal{X}$.

An intriguing search model (which is slightly closer to the path to be followed here in that probabilistic ideas are prominent and multi-modal functions are included in $\mathcal{F}$) was proposed by H. Kushner [6,7] who supposed f to be a sample function from the Brownian motion process on a bounded linear interval, $\mathcal{X}$. An advantage to this viewpoint is that, in addition to including multi-modal functions, ideas from Wiener filter theory can be brought to bear on the problem of designing an optimal search procedure. Kushner points out that numerical evaluation of the optimal procedure is computationally prohibitive, but provides a search procedure under which $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} f(x_i) = \|f\|$, almost surely.

The research report in this paper follows an approach sketched by S. Brooks [8] . Presumably, Brooks took $\mathcal{X}$ to be a finite set, and took the loss associated with the function $f \in \mathcal{F}$ and operating point $x \in \mathcal{X}$ to be

$L(x, f)$ = "proportion" of points $x' \in \mathcal{X}$ such that $f(x') > f(x)$. Then, given any positive numbers c,d, a smallest number N is readily calculated such that if $x_1, x_2, \ldots x_N$ are selected from $\mathcal{X}$ by a randomization which gives equal weight to each element of $\mathcal{X}$, then for any real-valued function f,

$$P\left[\max_{i \leq n} L(x_i, f) > c\right] < d, \qquad \text{for } n > N.$$

Brooks, as well as Kushner, consider the possibility that the measurements $f(x_i)$ may be corrupted by additive noise. These considerations will be detailed, along with a brief review of "stochastic approximation" in a later section (Section 4) of this paper.

Let us summarize the results of this paper. $\mathcal{F}$ will, in all our studies, at least include the set of continuous functions on $\mathcal{X}$, which, for expository reasons, will be the unit interval. Section 2 reveals two random search procedures; the first of these achieves almost sure convergence of $1/n \sum_{i=1}^{n} f(x_i)$ to $\|f\|$ for each $f \in \mathcal{F}$ and the second yields a random sequence $\{f(x_i)\}$ which converges in probability to $\|f\|$. Section 2 concludes with a theorem on the non-existence of a search procedure under which $f(x_n) \to \|f\|$ almost surely for all continuous f.

Section 3 reopens and extends the research path suggested by Brooks [8]. Where Brooks defines the loss associated with operating point x and $f \in \mathcal{F}$ to be "proportion" of $x' \in \mathcal{X}$ such that $f(x') > f(x)$, we define the loss to be

$$L(x,f) = \text{Lebesgue measure } \{x' : f(x') > f(x)\} .$$

It will be verified that this retains the important feature in Brooks' study that, for any positive numbers c and d, one may compute in advance of making measurements, how many measurements N are required so that, for any $f \in \mathcal{F}$, $n > N$,

$$(1.1) \qquad P\left[L(x_n^*, f) > c\right] < d,$$

x_n^* being the element x_i, $1 \leq i \leq n$, which maximizes the measurement $f(x_i)$. Further, random searches S_1 and S_2 and numbers N_1 and N_2 are described such that, for any $f \in \mathcal{F}$, under S_1,

$$(1.2) \qquad P\left[\sup_{n > N_1} 1/n \sum_{i=1}^{n} L(x_i, f) > c\right] < d$$

and under S_2,

$$(1.3) \qquad P\left[L(x_n, f) > c\right] < d \qquad \text{for all } n > N_2.$$

The section concludes with brief consideration of non-uniform searches. It is suggested how Bayesians might adapt the theory to account for *a priori* notions of where the better operating points may be found in $\mathcal{X}$.

If for each $f \in \mathcal{F}$ and each real number a, if $f^{-1}(a)$ has Lebesgue measure 0, and if the x_i's are selected according to the uniform law, then the random variable $n\, f(x_n^*)$ (x_n^* being defined in connection with (1.1)) has the exponential law for a limiting distribution, as is demonstrated in Section 4.

Section 5 studies the case that the measurements $\{f(x_i)\}$ are corrupted by additive noise, which is assumed independent of x_i, the magnitude of $f(x_i)$, and the sampling time, i. With no further assumptions on the noise process, we reveal a search procedure under which the average operating loss, $1/n \sum_{i=1}^{n} L(x_i, f)$, converges in probability to 0 for all Lebesque measureable functions f; however, in the noise case, no lower bounds for the rate of this convergence have been discovered. We compare this problem and the results obtained to the class of problems which are known to yield to the method of stochastic approximation, and also mention related results due to Brooks and Kushner.

References

1 Hadley, G. Nonlinear and Dynamic Programming. Reading, Massachusetts: Addison Wesley, 1964.

2 Spang, H. A. III. "A Review of Minimization Techniques for Nonlinear Functions," S.I.A.M. Review 4. 1962. pp. 343-365.

3 Kiefer, J. "Sequential Minimax Search for a Macimum," Proc. Am. Math. Soc., 4. 1953. pp. 503-506.

4 Kiefer, J. "Optimum Sequential Search and Approximation Under Minimum Regularity Assumptions," S.J.A.M. Journal 5. 1957. pp. 105-136.

5 Bellman, R. and S. Dreyfeus. Appli[illegible] Dynamic Programming. Princeton, New Jersey: Princeton Universi[illegible] [illegible]ess, [illegible].

6 Kushner, H. "A versatile Stochastic Model of a Function of Unknown and Time Varying Form," J. Math. Anal. and Appl., 5. 1962.

7 Kushner, H. "A New Method for Locating the Maximum Point in an Arbitrary Multipeak Curve in the Presence of Noise," A.S.M.E. J. Basic Engr. 86. 1964. pp. 97-106.

8 Brooks, S. "A Discussion of Random Methods for Seeking Maxima," J. Operations Res. Soc. 6, 1958. pp. 244-251.