Interim Report

ADAPTIVE FILTERING

Andrew H. Jazwinski
Ann E. Bailie

Report No. 67-6
Contract NAS 5-9085

March 1967

Washington Office
Analytical Mechanics Associates, Inc.
9301 Annapolis Road
Lanham, Maryland 20801

ABSTRACT

Applications of the Kalman filter in orbit determination have sometimes encountered a difficulty which has been referred to as divergence. The phenomenon is a growth in the residuals; the state and its estimate diverge. This problem can often be traced to insufficient accuracy in modeling the dynamics used in the filter. Although more accurate dynamics is an obvious solution, it is often an impractical one. Model inaccuracies are here approximated by a Gaussian white noise input, and its covariance (Q) is determined so as to produce most probable sequences of residuals. The method is desirably non-Bayesian and adaptive, with direct feedback from the residuals. An explicit solution to the problem of when, and how much, Q to add is given.

I Introduction

In the determination of orbits via the Kalman filter (or, for that matter, by other methods as well), practical considerations place a limitation on the accuracy to which the dynamics are modeled in the filter. This produces an inherent inconsistency between the filter and the real observations. As long as the uncertainty in the estimation error of the dynamical state is large, this inconsistency is not very significant. The extremely precise observations which are taken tend to rapidly reduce the uncertainty in the estimation error, to the point where the modeling errors become significant. The commonly observed phenomenon is a growth in the residuals; the state and its estimate diverge.

It is relatively clear that, in order to improve prediction accuracy, the dynamics have to be modeled more accurately and perhaps the nonlinearity of the problem need be taken into account. It does not appear that "short-cut" methods are available for the prediction problem.

There do exist possible "short-cut" methods of improving the filter. Residuals are available and the filter gain can be "manipulated" to produce satisfactory performance. One has to abandon the idea of optimal filtering and adopt the pragmatic engineering philosophy of constructing a filter that "works". The test to determine the efficiency of the filter now must be the size of the residuals, since this is the only enduring link between the analysis and the real world.

A very natural approach to account for modeling errors is to provide a noise input to the system and add its covariance Q to P, the covariance of the estimation errors. The problem is when and how much Q to add.

Implicit in this sort of procedure is the assumption that the noise input is zero-mean, white and Gaussian. This type noise input may reasonably account for modeling errors whose effect is periodic with short period. Nonzero-mean noise might conceivably be used to account for modeling errors of a secular nature but, in this case, the effect of the mean value of the noise on the state estimate itself must be taken into account. Looking at it in a non-theoretical but pragmatic way, Q may be thought of as representing some extra degrees of freedom which can be used to make the filter "work".

Bayesian-type estimation of Q is possible, but this is not acceptable as a solution to the problem of growing residuals. Such estimates asymptotically approach constant values, and the degrees of freedom which Q represents are lost. Since the modeled noise input does not exactly represent the modeling errors, the residuals will start growing again. The problem is not really solved; its onset is only somewhat retarded.

It is believed that any scheme for determining a Q (or any other artifice) must deal directly with the actual residuals themselves. It is desired that the sequence of residuals considered in any scheme be long enough to be statistically significant. On the other hand, the sequence of residuals (actually the time-arc over which the sequence is defined) should be sufficiently short so that an adequate "fit" can be realized with the limited degrees of freedom which Q represents. These are, unfortunately, conflicting objectives. If not statistically significant, the scheme must be adaptive or self-correcting. It is doubtful that, in any orbit determination problem, a sufficient number of residuals is available over reasonably short time-arcs to represent a statistically significant sample. Thus a decision (Q) may have no statistical significance, and subsequent decisions must correct previous errors.

A scheme to determine Q which appears to have such adaptive features is presented here. The approach is to determine that Q which produces the most probable sequence of residuals. It involves numerous computations if the sequence of residuals considered is long. However, this scheme appears promising even when only one residual is considered at a time, in which case the computations are negligible. Numerical evaluation of the method is in progress.

The problem which is considered here, that is, filter divergence due to modeling errors, was first considered by Schmidt [1]. His approach was to impose an a priori lower bound on a certain projection of P, which led to some interesting results. The approach taken here is somewhat more direct and, in a special case, a part of Schmidt's result is also obtained.

II Analysis

The following linear dynamical system model is assumed:

$$x_{k+1} = \Phi_{k+1,\,k}\, x_k + G_k\, w_k\,, \qquad k = 0,\ 1,\ \ldots,\ x_0 \sim N(\hat{x}_0,\ P_0),$$

$$y_k = M_k\, x_k + v_k \tag{1}$$

where,

x_k - n x 1 vector state,

$\Phi_{k+1,\,k}$ - n x n state transition matrix,

G_k - n x r input noise coefficient matrix,

w_k - r x 1 vector noise input,

y_k - scalar observation,

M_k - 1 x n matrix,

v_k - scalar observation noise.

$\{w_k\}$ is a zero-mean, white Gaussian noise sequence with

$$\mathcal{E}\left\{w_k\, w_\ell^T\right\} = q\, I\, \delta_{k\ell}, \qquad q \geq 0 \text{ scalar},\ I \quad r \times r \tag{2}$$

and $\{v_k\}$ is a zero-mean, white Gaussian noise sequence with

$$\mathcal{E}\left\{v_k\, v_\ell\right\} = r\, \delta_{k\ell}, \qquad r > 0 \text{ scalar} \tag{3}$$

We have assumed, without loss of generality, that the observations are scalar. A particularly simple form for the input noise statistics (2) is assumed, thus requiring the modeling of the matrix function G_k . The analysis which follows can be readily carried through with a more general form for $\mathcal{E}\left\{w_k\, w_\ell^T\right\}$, which would in turn require less modeling of the function G_k . As indicated in (1), the initial condition x_0 is assumed Gaussian with mean $\hat{x}_0$ and covariance matrix P_0 .

The well known Kalman-Bucy filter for this model is

$$\hat{x}_{k+} = \hat{x}_{k-} + P_{k-}\, M_k^T\, (M_k\, P_{k-}\, M_k^T + r)^{-1}\, (y_k - M_k\, \hat{x}_{k-})$$

$$P_{k+} = P_{k-} - P_{k-}\, M_k^T\, (M_k\, P_{k-}\, M_k^T + r)^{-1}\, M_k\, P_{k-} \qquad (4)$$

$$\hat{x}_{(k+1)-} = \Phi_{k+1,\,k}\, \hat{x}_{k+}$$

$$P_{(k+1)-} = \Phi_{k+1,\,k}\, P_{k+}\, \Phi_{k+1,\,k}^T + q\, G_k\, G_k^T \qquad (5)$$

with initial conditions $\hat{x}_0$ and P_0 .

Let $\mathcal{Y}_k = \left\{\ldots, y_{k-1}, y_k\right\}$. $\mathcal{Y}_k$ includes the prior data and all the observations up to and including y_k . We define the following residuals:

$$r_k \equiv y_k - \mathcal{E}\left\{y_k \mid \mathcal{Y}_{k-1}\right\} \qquad (6)$$

In view of (1),

$$\begin{aligned} r_k &= y_k - M_k \hat{x}_{k-} \\ &= M_k (x_k - \hat{x}_{k-}) + v_k \end{aligned} \tag{7}$$

The statistics of these residuals are particularly nice (and this motivated our choice of these residuals):

$$\mathcal{E}\{r_k\} = \mathcal{E}\{\mathcal{E}\{r_k | \mathcal{Y}_{k-1}\}\} = 0 \tag{8}*$$

$$k > \ell, \quad \mathcal{E}\{r_\ell r_k\} = \mathcal{E}\{\mathcal{E}\{(y_\ell - \mathcal{E}\{y_\ell | \mathcal{Y}_{\ell-1}\})(y_k - \mathcal{E}\{y_k | \mathcal{Y}_{k-1}\})|\mathcal{Y}_{k-1}\}\} \tag{9}*$$

$$= \mathcal{E}\{(y_\ell - \mathcal{E}\{y_\ell | \mathcal{Y}_{\ell-1}\})\mathcal{E}\{(y_k - \mathcal{E}\{y_k | \mathcal{Y}_{k-1}\})|\mathcal{Y}_{k-1}\}\}$$

$$= 0$$

Thus residuals $\{r_k\}$ are uncorrelated and, since they are clearly Gaussian, are also independent. It is easy to compute

$$\mathcal{E}\{r_k^2\} = M_k P_{k-} M_k^T + r \tag{10}$$

---

* We have used "Adam's rule". See p. 34 of [2], for example. These results can also be obtained by straightforward but lengthy computations using (7).

Thus the joint probability density function of $r_{k+1}, r_{k+2}, \ldots, r_{k+N}$ is

$$p(r_{k+1}, r_{k+2}, \ldots, r_{k+N}) = p(r_{k+1}) \cdot p(r_{k+2}) \cdots p(r_{k+N}) \tag{11}$$

where

$$p(r_{k+j}) = \frac{1}{(2\pi)^{1/2}} \left(M_{k+j} P_{(k+j)-} M_{k+j}^T + r\right)^{-1/2} e^{-\frac{1}{2}\left\{\frac{r_{k+j}^2}{M_{k+j} P_{(k+j)-} M_{k+j}^T + r}\right\}} \tag{12}$$

As an aside, it is interesting to note that

$$p(r_\ell) = p(y_\ell | \vartheta_{\ell-1}) \tag{13}$$

so that

$$p(r_{k+1}, \ldots, r_{k+N}) = p(y_{k+1} | \vartheta_k) \cdots p(y_{k+N} | \vartheta_{k+N-1}) \tag{14}$$

Suppose that the filter is at time $k+$, so that $\hat{x}_{k+}$ and P_{k+} are available and (5) is about to be used to compute the $(k+1)-$ values. Fix N. The q value to be used (call it $\hat{q}_N$) is determined by the operation

$$\max_{q \geq 0} \quad p(r_{k+1}, \ldots, r_{k+N}) \tag{15}$$

That is, $\hat{q}_N$ is that value of q which produces the most likely sequence of residuals $\{r_{k+1}, \ldots, r_{k+N}\}$. In view of (4) and (5), it is seen that

$$p(r_{k+1}, \ldots, r_{k+N}) = \text{fcn}\ (y_{k+1}, \ldots, y_{k+N}; \hat{x}_{k+}; P_{k+}; r; q) \tag{16}$$

The operation (15) involves a one-dimensional search for the maximum. Now $p \geq 0$ everywhere and $p \to 0$ as $q \to \infty$. This should make the search for the maximum easier. A new $\hat{q}_N$ could be determined at each time k. Alternatively, the same $\hat{q}_N$ could be used for $\ell \leq N$ time steps before a new one is found.

The adaptive features of this scheme are seen in the $N = 1$ example of the following section.

III Estimate of q from One Residual

In the case of only one residual ($N = 1$), it is easy to get a closed-form solution to the problem

$$\max_{q \geq 0} p(r_{k+1}) \tag{17}$$

Let

$$\bar{q} \equiv \frac{r_{k+1}^2 - \left[M_{k+1} \Phi_{k+1,k} P_{k+} \Phi_{k+1,k}^T M_{k+1}^T + r \right]}{M_{k+1} G_k G_k^T M_{k+1}^T} \tag{18}$$

then

$$\hat{q}_1 = \begin{cases} \bar{q}, & \text{if } \bar{q} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

We have, of course, assumed $M_{k+1} G_k G_k^T M_{k+1}^T \neq 0$. $\bar{q}$, if non-negative, provides the only local maximum of p on $[0, \infty)$, and $\frac{\partial^2 p}{\partial q^2}(\bar{q}) = -C r_{k+1}^{-4}$, $C > 0$, so $\bar{q}$ provides the global maximum on $[0, \infty)$. If $\bar{q}$ is negative, then $\frac{\partial p}{\partial q} < 0$ for all $q \geq 0$, so $\hat{q}_1 = 0$.

To analyze this result, let us compute

$$\mathcal{E}\left\{ r_{k+1}^2 \mid q = 0 \right\} = \mathcal{E}\left\{ \left[M_{k+1}\left(x_{k+1} - \hat{x}_{(k+1)-} \right) + v_{k+1} \right] \left[M_{k+1}\left(x_{k+1} - \hat{x}_{(k+1)-} \right) + v_{k+1} \right]^T \mid q=0 \right\} \tag{20}$$

$$= M_{k+1} \Phi_{k+1,k} P_{k+} \Phi_{k+1,k}^T M_{k+1}^T + r$$

Thus,

$$\bar{q} = \frac{r_{k+1}^2 - \mathcal{E}\left\{r_{k+1}^2 \mid q = 0\right\}}{M_{k+1} G_k G_k^T M_{k+1}^T} \tag{21}$$

This is a very reasonable result. Except for the proper scaling factor, $\hat{q}_1$ is the excess, if any, of the residual squared over the expected value of that residual squared, under the assumption of no input noise. $\hat{q}_1$ is an unbiased estimate, since an easy computation shows that

$$\mathcal{E}\{\bar{q}\} = q \tag{22}$$

This scheme is adaptive in the following sense. As long as the residuals are small ($\bar{q} \leq 0$), $\hat{q}_1 = 0$, and the noise input is zero. This is as it should be, since the residuals are small and the filter is "working". When the residuals become large ($\bar{q} > 0$), a $\hat{q}_1$ is put into the filter, equation (5). If it should happen that a particular residual is large only because of a wild observation noise sample and this scheme puts too much q into equation (5), no permanent damage is done. This would make P too large and would result in larger corrections to the estimate at subsequent observations, which is consistent with the requirement that the filter "keep on filtering". Subsequent observations will drive P down again. Legitimately large residuals (due to model errors) clearly produce the desirably large $\hat{q}_1$ which will keep the filter "open".

Since, strictly speaking, ten residuals do not have much more statistical significance than one residual, and a thousand residuals are neither available nor feasibly analyzable, $\hat{q}_1$ appears as a serious candidate for the real orbit determination program. Very large q's, due to wild observation noise samples, can be avoided by putting an a priori upper bound on q, and performing the maximization in (15), (17) over $q^* \geq q \geq 0$.

IV More General Input Noise Statistics

Suppose that the noise input w_k (1) has the more general statistic

$$\mathcal{E}\left\{w_k\, w_\ell^T\right\} = Q\,\delta_{k\ell}\,, \qquad Q \quad r \times r\,,\ Q \geq 0 \tag{23}$$

Then the second of equation (5), in the Kalman-Bucy filter, is replaced by

$$P_{(k+1)-} = \Phi_{k+1,\,k}\, P_{k+}\, \Phi_{k+1,\,k}^T + G_k\, Q\, G_k^T \tag{24}$$

Now let us attempt to determine Q from one residual by the operation

$$\max_{Q \geq 0} \; p\,(r_{k+1}) \tag{25}$$

Differentiating p with respect to an element of Q, say q_{ij}, gives

$$\left[-\left(M_{k+1} P_{(k+1)-} M_{k+1}^T + r\right)^{-1} + r_{k+1}^2 \left(M_{k+1} P_{(k+1)-} M_{k+1}^T + r\right)^{-2}\right] \frac{p}{2}\, M_{k+1}\, \frac{\partial}{\partial q_{ij}} \left(G_k Q G_k^T\right) M_{k+1}^T \tag{26}$$

Now assuming $M_{k+1}\, \dfrac{\partial}{\partial q_{ij}}\, (G_k\, Q\, G_k^T)\, M_{k+1}^T \neq 0$ for some i and j,* we get as the single scalar necessary condition for an unconstrained maximum of p the vanishing of the quantity in brackets in (26). This reduces to

$$M_{k+1}\, G_k\, Q\, G_k^T M_{k+1}^T = r_{k+1}^2 - \mathcal{E}\left\{r_{k+1}^2 \,\middle|\, Q = 0\right\} \equiv \epsilon \tag{27}$$

---

* That is, $G_k^T M_{k+1}^T M_{k+1} G_k$ has at least one non-zero element.

Now equation (27) has infinitely many solutions $\bar{Q}$. It can be shown* that the most general solution is given by

$$\bar{Q} = \frac{\epsilon}{\left(M_{k+1} G_k G_k^T M_{k+1}^T\right)^2} G_k^T M_{k+1}^T M_{k+1} G_k + Z$$

$$- \frac{1}{\left(M_{k+1} G_k G_k^T M_{k+1}^T\right)^2} G_k^T M_{k+1}^T M_{k+1} G_k Z G_k^T M_{k+1}^T M_{k+1} G_k \tag{28}$$

where Z is an arbitrary $r \times r$ matrix. It can also be shown, at least for Q diagonal, that

$$\left[\frac{\partial^2 p(\bar{Q})}{\partial q_{ii} \partial q_{jj}} \right] = - C r_{k+1}^{-4} A \tag{29}$$

where $c > 0$ and A is a positive semi-definite matrix, for all solutions $\bar{Q}$. Thus all diagonal $\bar{Q}$ are candidates for local maxima of p .

Some particular solutions will now be exhibited. Let $Z = \epsilon I / M_{k+1} G_k G_k^T M_{k+1}^T$. Then we have

$$\bar{Q}_e = \frac{\epsilon}{M_{k+1} G_k G_k^T M_{k+1}^T} I \tag{30}$$

---

* See [3], for example.

Note that the coefficient in (30) is the $\bar{q}$ of Section III. $\bar{Q}_e$ is an "equal weights" solution. If Q is 1×1, that is, G_k is $n \times 1$, then

$$\bar{Q} = \frac{\epsilon}{M_{k+1} G_k G_k^T M_{k+1}^T} \tag{31}$$

for all Z, and provides the global maximum of p. A family of "relative weights" solutions is given by

$$\bar{Q}_i = \frac{\epsilon}{\sum_{j=1}^{r} \left[M_{k+1} (G_k)_j \right]^{i+2}} \begin{bmatrix} \left[M_{k+1} (G_k)_1 \right]^i & & 0 \\ & \ddots & \\ 0 & & \left[M_{k+1} (G_k)_r \right]^i \end{bmatrix}, \quad i = 0, 2, 4, \ldots \tag{32}$$

where $(G_k)_j$ are the columns of G_k, as can be verified directly in (27). $\bar{Q}_0 = \bar{Q}_e$.

Now all the solutions exhibited contain ϵ, so that estimates of Q based upon these quantities have properties similar to those of $\hat{q}_1$ of Section III.

$$\hat{Q}_i = \begin{cases} \bar{Q}_i, & \epsilon > 0 \\ 0, & \text{otherwise} \end{cases} \tag{33}$$

are positive semi-definite and are proportional to the excess, if any, of r_{k+1}^2 over its expected value, with no input noise. All these estimates are, however, biased estimates.

Suppose that it is not feasible to model the matrix function G_k and, therefore, desirable to determine the correction $G_k Q G_k^T$ to $P_{(k+1)}$ - [equation (24)] directly by the operation

$$\max_{G_k Q G_k^T \geq 0} p(r_{k+1}) \tag{34}$$

In that case, $G_k Q G_k^T$ is treated as the unknown in (27). The most general solution is, therefore,

$$\overline{G_k Q G_k^T} = \frac{\epsilon}{\left(M_{k+1} M_{k+1}^T\right)^2} M_{k+1}^T M_{k+1} + Y - \frac{1}{\left(M_{k+1} M_{k+1}^T\right)^2} M_{k+1}^T M_{k+1} Y M_{k+1}^T M_{k+1} \tag{35}$$

with Y an arbitrary $n \times n$ matrix. It can also be shown that for $G_k Q G_k^T$ diagonal

$$\left[\frac{\partial^2 p\left(\overline{G_k Q G_k^T}\right)}{\partial q_{ii} \partial q_{jj}}\right] = - d\, r_{k+1}^{-4} B \tag{36}$$

where $d > 0$ and B is a positive semi-definite matrix, for all solutions $\overline{G_k Q G_k^T}$. The q_{ii} now denote the diagonal elements of $G_k Q G_k^T$.

Some particular solutions may now be exhibited. With $Y = 0$ or $Y = M_{k+1}^T M_{k+1}$, we obtain

$$\overline{G_k Q G_k^T} = \frac{\epsilon}{\left(M_{k+1} M_{k+1}^T\right)^2} M_{k+1}^T M_{k+1} \tag{37}$$

This solution resembles the correction to P derived by Schmidt [1], except that our correction is to $P_{(k+1)-}$ and his is to $P_{(k+1)+}$. Also, we do not modify the estimate itself directly, while he does. Our strategy, based on (37), would be

$$\widehat{G_k Q G_k^T} = \begin{cases} \overline{G_k Q G_k^T}, & \epsilon > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{38}$$

which contains no undetermined scaling parameter, while Schmidt's scheme does. In fact, the scaling parameter is explicitly given, namely,

$$\frac{\epsilon}{\left(M_{k+1} M_{k+1}^T\right)^2} \tag{39}$$

Let $Y = \epsilon I / M_{k+1} M_{k+1}^T$. Here, we have another solution

$$\overline{G_k Q G_k^T} = \frac{\epsilon}{M_{k+1} M_{k+1}^T} I \tag{40}$$

Again, the estimates based on these solutions are biased estimates.

Actually, it is felt that $M_{k+1}^T M_{k+1}$ [equation (37)] has as little to do with model errors as does I [equation (40)], and neither solution is therefore recommended. One can certainly do better with even a crude modeling of G_k. It has not been proven (or disproven) that the $\bar{Q}_i$ (32) maximize p, nor has the global maximum of p(Q) been found. As a result, it is preferable to use the model and estimate given in Section III, if the estimate is to be based on only one residual.

V Estimate of r from One Residual

The present analysis can also be applied to the estimation of the observation noise variance r [see (3)]. The "parameter" r is not considered as useful as q in controlling filter divergence, since r can only produce a slower rate of decrease of the covariance matrix P, but cannot cause P to increase. The estimate of r from one residual follows.

At time $k-$, r is determined from

$$\max_{r \geq 0} \quad p(r_k) \tag{41}$$

If

$$\bar{r} \equiv r_k^2 - M_k P_{k-} M_k^T \tag{42}$$

it is easy to compute the solution

$$\hat{r}_1 = \begin{cases} \bar{r}, & \bar{r} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{43}$$

Since

$$M_k P_{k-} M_k^T = \mathcal{E}\{r_k^2 | r = 0\} \tag{44}$$

we have

$$\hat{r}_1 = \begin{cases} r_k^2 - \mathcal{E}\{r_k^2 | r = 0\}, & \text{if positive} \\ 0, & \text{otherwise} \end{cases} \tag{45}$$

Also,

$$\mathcal{E}\{\bar{r}\} = r \tag{46}$$

VI Adaptive Filter Performance

Consider the filter (4, 5) with q replaced by the estimate $\hat{q}_1$ from (19).* Now $\hat{q}_1$ is random since it depends on the observation. As a result, the difference equations for P_k are stochastic. It is no longer possible to show that the estimates $\hat{x}_k$ are unbiased estimates (as they are in the case where q is known). Clearly, $\hat{x}_k$ is not the minimum variance estimate. The difficulty is that the difference equation for P_k is nonlinear and P_k appears nonlinearly in the filter gain (4). Thus it is impossible to compute a recursion for $\mathcal{E}\left\{(x_k - \hat{x}_{k+})\right\}$, which is the expected estimation error, the relevant quantity in the consideration of biasness.

Furthermore, it is impossible to compute $\mathcal{E}\left\{(x_k - \hat{x}_{k+})(x_k - \hat{x}_{k+})^T | \mathcal{Y}_k\right\}$ for the reasons mentioned above. This latter quantity is the matrix of conditional second moments about the estimate. Also,

$$P_{k+} \neq \mathcal{E}\left\{(x_k - \hat{x}_{k+})(x_k - \hat{x}_{k+})^T | \mathcal{Y}_k\right\} \tag{47}$$

where P_{k+} is computed in the adaptive filter. At this point, it is to be noted that the statistical properties of the estimates of q (or Q) obtained in this paper are strictly valid only for the first such estimate, due to (47). The performance of the adaptive filter cannot be evaluated theoretically because

---

* This estimate of q is considered here for concreteness. The results which follow apply to the other estimates of q as well.

it is impossible to compute

$$\mathcal{E}\left\{(x_k - \hat{x}_{k+})(x_k - \hat{x}_{k+})^T\right\}$$

which is an indicator of filter performance.

The above remarks, however disquieting to the theoretician, need not be taken to heart by the engineer. The optimal filter, when applied to orbit determination, often does not work, and does this, to the frustration of the engineer, in an almost optimal way. A practical alternative, not without some theoretical basis, is offered here. It does not possess some of the nice theoretical properties, yet intuition indicates that it may indeed work very well. The "proof" can only come from the "pudding" of numerical experiments.

VII Acknowledgments

The authors wish to express their appreciation for the active interest shown by Mr. R. K. Squires of Goddard Space Flight Center and Dr. H. Wolf of Analytical Mechanics Associates, Inc. in this problem and its solution. The point of view here adopted, of looking at residuals as the most important link between the filter and the real world, is due to Mr. Squires.

VIII References

[1] Schmidt, S. F.; "Estimation of State with Acceptable Accuracy Constraints," Interim Report No. 67-4, Contract NAS 5-9085, Analytical Mechanics Associates, Inc., January 1967.

[2] Jazwinski, A. H.; "Stochastic Processes with Application to Filtering Theory," Report No. 66-6, Final Report Contract NAS 5-9298, Analytical Mechanics Associates, Inc., November 1966.

[3] Newhouse, S. E.; "Introduction to Matrix Generalized Inverses and their Applications," NASA Goddard Space Flight Center Report X-643-66-34, January 1966.