Empirical Bayes Estimation of Observation Error Variances in Linear Systems

HARRY F. MARTZ JR.*
Texas Tech University, Lubbock, Texas

AND

GEORGE H. BORN†
NASA Manned Spacecraft Center, Houston, Texas

A filter set is developed for estimating the state vector and observation error variances in a discrete-time linear system by use of empirical Bayes techniques. The error variances are assumed to be random and to vary over time. No initial conditions or distributional assumptions are required for the error variances, but all other assumptions for the Kalman filter are assumed to hold. The treatment is analytical, and a Monte Carlo simulation is used to verify the results. Graphs are presented which compare performance with the ideal case of known variances. The filter was found to converge fairly rapidly for the examples considered.

Introduction

OF useful and widespread interest in trajectory estimation is the problem of estimating the state in a linear discrete-time dynamic system by means of observation data. Consider the linear system

$$\mathbf{x}(t_i) = \Phi(t_i,t_{i-1})\mathbf{x}(t_{i-1}) + \mathbf{u}(t_{i-1})$$

augmented by the linear state-observation equation

$$\mathbf{y}(t_i) = H(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i), \qquad i = 1,2, \ldots$$

where $\mathbf{u}$ and $\mathbf{v}$ are random vectors with completely known and specified distributions. For notational simplicity, we shall refer to time epoch t_i by use of the single subscript i. For estimating the state $\mathbf{x}$ by means of past as well as current observations $\mathbf{y}$, one of the most widely accepted solution techniques is provided by the Kalman filter.[1,2]

Frequently, all or some of the distributional assumptions necessary for the filter to be applied are not known to hold in practice. Let us restrict consideration to the situation in which the variances of the observation error $\mathbf{v}$ are unknown and are to be estimated in addition to the state of the system. Several techniques currently exist to accomplish this task. If the unknown variances can be assumed to remain constant over time, then it is possible to annex the Kalman filter equations with an equation which recursively estimates these variances. The technique of maximum likelihood is commonly used to perform this estimation.[3] If the error variances are allowed to vary deterministically with time, that is, if they vary in a known fashion but are not precisely known at any time, then a Bayesian technique currently exists for this situation.[3] This technique requires that nominal values of the variances be known over time, but that these nominal values remain biased away from the respective true variances each by an unknown scale factor. Each observation type has a unique scale factor associated with it which is sequentially estimated over time as observations of that type become available for processing. Each of these unknown scale factors is considered to be a value of a random variable having an inverted-gamma distribution with unknown parameters which are sequentially estimated by means of the observation data of that type. Once these scale factors have been initially realized, they are then assumed to remain constant over time. The nominal variance values may be thought of as preliminary estimates which give the orders of magnitude of the actual variance values. The performance of the resulting filter depends upon the specific choice of a prior distribution for each of the scale factors. This Bayesian technique is presented in the Appendix.

Empirical Bayes estimation theory is used here to develop a procedure for estimating the unknown random scale factors. The empirical Bayes theory does not depend upon the choice of a specific prior distribution for each of the scale factors and is applicable regardless of the actual prior distributions involved. Thus fewer total assumptions are required to implement the filter. In addition, the scale factors are permitted to vary randomly over time. This provides a generalization of the situation previously described. The empirical Bayes variance equations are then adjoined to the usual Kalman filter equations for use in the state estimation phase of the problem.

Statement of the Problem

Consider the following linear discrete dynamic system

$$\mathbf{x}_i = \Phi_i\mathbf{x}_{i-1} + \mathbf{u}_{i-1} \tag{1}$$

with the linear observation-state relationship

$$\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i, \qquad i = 1,2, \ldots$$

where i is a time index, $\mathbf{x}_i$ is a $p \times 1$ system state vector, $\mathbf{y}_i$ is a $q \times 1$ observation vector, Φ_i is a $p \times p$ state transition matrix, and H_i is a $q \times p$ matrix relating $\mathbf{x}_i$ to $\mathbf{y}_i$. In addition, it is assumed that 1) $\mathbf{u}_i$ dist $N_p(\mathbf{0},Q_i)$(p-normal with mean vector $\mathbf{0}$ and covariance matrix Q_i); 2) $\mathrm{Cov}(\mathbf{u}_i,\mathbf{u}_j) = \delta_{ij}Q_i$; 3) $\mathbf{v}_i$, conditional on R_i, dist $N_q(\mathbf{0},R_i)$; 4) $\mathrm{Cov}(\mathbf{v}_i,\mathbf{v}_j|R_i,R_j) = \delta_{ij}R_i$; 5) Φ_i, H_i, Q_i are known; and 6) $\mathbf{x}_0$ dist $N_p(\mathbf{0},P_0)$, P_0 known.

Let us further assume that different observation types are mutually statistically independent; thus, R_i is a diagonal matrix. In accordance with the discussion in the Introduction and similar to what was done in Ref. 3, define

$$R_i = \mathrm{diag}\ (r_{i1}^2/\theta_{i1},r_{i2}^2/\theta_{i2}, \ldots, r_{iq}^2/\theta_{iq}) \tag{2}$$

where r_{ij}^2, $j = 1,2, \ldots, q$, is a known nominal value of the

Presented as Paper 70-1058 at the AAS/AIAA Astrodynamics Conference, Santa Barbara, Calif., August 19–21, 1970; submitted September 14, 1970. This research was supported by NASA Grant NGR 44-011-031.

* Associate Professor of Industrial Engineering and Statistics.
† Special Assistant, Theory and Analysis Office, Computation and Analysis Division; now at Jet Propulsion Laboratories, Pasadena, Calif. Member AIAA.

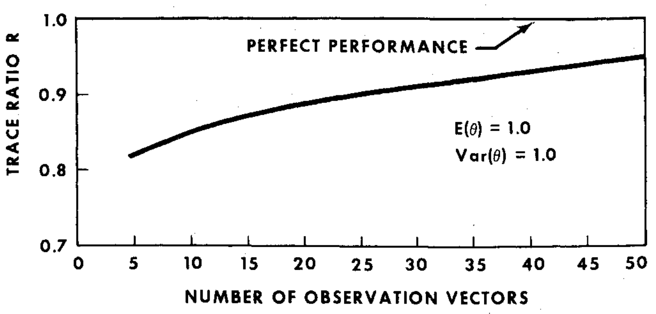

Fig. 1 The ratio R of the square root of the trace of the average squared error matrices of the pure Kalman filter to the Kalman, empirical Bayes filter.

observation error variance associated with the jth observation type at time epoch i. Also, $\{\theta_{ij}: i = 1,2, \dots\}$ are independent realizations of a random variable Θ_j having a completely unknown and unspecified distribution $G_j(\theta)$ which is zero on the negative real numbers and which may be different for each observation type. If we estimate θ_{ij} based on all available observation data of type j, it is possible to estimate the "true" error variance r_{ij}^2/θ_{ij}. This is accomplished by "adjusting" (multiplying) the nominal value r_{ij}^2 by the estimated scale factor $1/\hat{\theta}_{ij}$, where $\hat{\theta}_{ij}$ is the estimate of θ_{ij}.

Empirical Bayes Variance Estimation

Before proceeding with the solution of the problem, the major aspects of empirical Bayes estimation theory are described. A complete description of empirical Bayes estimation is provided in Refs. 4–7 and is summarized here. Consider the estimation of an unknown value θ of the random parameter Θ with regard to a squared error loss function. A value z of the random variable Z is observed, which has a known conditional density function $f(z|\theta)$. First, z_1 is observed from $f(z|\theta_1)$ and θ_1 is to be estimated; next, z_2 is observed from $f(z|\theta_2)$ and θ_2 is to be estimated. This continues until the observation z_n is obtained from $f(z|\theta_n)$ and θ_n is to be estimated. The estimation stages are assumed to be mutually independent so that by the time an estimate of θ_n is required, the sequence of observations $z_1,z_2, \dots, z_n$ is available. Each θ_i represents a realization of Θ according to the unknown and unspecified distribution $G(\theta)$. The standard empirical Bayes technique is to express the Bayes estimator $E(\theta|z_n)$ as a function of $f(z_n)$ where

$$f(z_n) = \int f(z_n|\theta)dG(\theta)$$

and to use the observations $z_1,z_2, \dots, z_n$ to estimate $f(z_n)$. The estimate of $f(z_n)$ is usually denoted by $f_n(z_n)$. Empirical Bayes estimators are thus approximations to Bayes estimators. An empirical Bayes estimate is referred to as asymptotically optimal if it converges in a certain statistical sense to the corresponding Bayes estimator as n tends to infinity.

The problem described in the preceding section can now be considered. Since all observation types are assumed to be independent, there is clearly no information about θ_{ij} except in observations of type j. The sequence of observations of type j is available and is given by

$$y_{ij} = \mathbf{h}_{ij}\mathbf{x}_i + v_{ij}, \qquad i = 1,2, \dots, n_j$$

where $\mathbf{h}_{ij}$ is the jth row of H_i, and n_j is the total number of observations of type j available up to the present time. For notational simplicity, let $n_j = n$, keeping in mind that observations of each type are not required to be taken at each time epoch i. The subscript j is dropped for the remainder of this section.

Assume temporarily that $\mathbf{x}_i$, $i = 1, \dots, n$ is exactly known. From assumption 3 following Eq. (1), observe that

$$z_i = (y_i - \mathbf{h}_i\mathbf{x}_i)/r_i \tag{3}$$

conditional on θ_i, has a normal distribution with mean zero and variance $1/\theta_i$. Note that z_i is the observation noise weighted by the factor $1/r_i$. Thus

$$f(z|\theta_n) = \theta_n^{1/2}/(2\pi)^{1/2} \exp[-\theta_n z^2/2] \tag{4}$$

The Bayes estimator for θ_n is given by the posterior mean of θ_n, which is expressed as

$$E(\theta_n|z_n) = \frac{\int \theta_n f(z_n|\theta_n)dG(\theta_n)}{\int f(z_n|\theta_n)dG(\theta_n)} \tag{5}$$

From Eq. (4),

$$\frac{1}{f(z_n|\theta_n)} \frac{\partial f(z_n|\theta_n)}{\partial z} = -\theta_n z_n \tag{6}$$

where $\partial f(z_n|\theta_n)/\partial z$ represents the partial derivative of $f(z|\theta_n)$ with respect to z, which is then evaluated at z_n. From Eq. (6)

$$\theta_n = -(1/z_n)[\partial f(z_n|\theta_n)/\partial z][1/f(z_n|\theta_n)]$$

and, upon substituting this expression into Eq. (5), interchanging the order of integration and differentiation, and simplifying, obtain

$$E(\theta_n|z_n) = -\frac{1}{z_n}\frac{[df(z_n)/dz]}{f(z_n)} = -\frac{1}{z_n}\frac{f'(z_n)}{f(z_n)} \tag{7}$$

where

$$f(z_n) = \int f(z_n|\theta_n)dG(\theta_n) \tag{8}$$

From Eq. (8), $f(z)$ is observed to be unimodal and symmetric about zero so that the RHS of Eq. (7) is always positive for any z_n.

The Bayes estimator given by Eq. (7) is now approximated by estimating $f(z_n)$ and $f'(z_n)$ from the data $z_1, \dots, z_n$. By estimating f and f', a need to explicitly identify $G(\theta_n)$ is essentially bypassed, since $G(\theta_n)$ is only implicitly contained in Eq. (7). To accomplish this, an estimator for a density function as well as its derivative is required. Reference 9 provides a technique for estimating a density function which has been widely applied to problems in empirical Bayes. Suppose $z_1,z_2, \dots, z_n$ are independent and identically distributed random variables with common probability density function f.

Reference 9 provides a class of kernel estimators

$$f_n(z) = \frac{1}{nh(n)} \sum_{i=1}^{n} K\left(\frac{z - z_i}{h(n)}\right)$$

for estimating $f(z)$, where K and $h(n)$ satisfy certain boundedness, regularity, and convergence conditions. Provided these conditions are satisfied, it was shown in Ref. 9 that $f_n(z)$ converges in squared error to $f(z)$ at all points of continuity of $f(z)$. Other studies[6,13] indicate that by taking

$$K(w) = 1/2\pi[\sin(w/2)/(w/2)]^2$$

and

$$h(n) = n^{-1/5}$$

the resulting estimated density function satisfactorily estimates $f(z)$ for small values of n and provides useful results in empirical Bayes estimation. The above choice of K and $h(n)$ satisfies the necessary conditions set forth in Ref. 9. Thus, the estimate of $f(z_n)$ in Eq. (7) is

$$f_n(z_n) = \frac{1}{2\pi n^{4/5}} \sum_{i=1}^{n} \frac{\sin^2(A_i)}{A_i^2}$$

where

$$A_i = (z_n - z_i)/2n^{-1/5}$$

Note that this estimator has not been developed to include the unimodality and symmetry about the origin of $f(z)$ in Eq. (8). Consequently, negative estimates of the scale factor may be obtained from the approximation to Eq. (7). This shortcoming is rectified in a later section of the paper.

Reference 10 provides a method for estimating $f'(z_n)$ in Eq. (7). Under general conditions, $f_n'(z)$ converges in squared error to $f'(z)$ at all points of continuity of $f'(z)$. Thus, to estimate $f'(z_n)$, it is required to differentiate $f_n(z)$ with respect to z and to evaluate the resulting expression at z_n.

By use of these estimates, consider the approximation to Eq. (7) given by

$$\hat{\theta}_n = E_n(\theta_n | z_n) = -(1/z_n)[f_n'(z_n)/f_n(z_n)] \tag{9}$$

where

$$\frac{f_n'(z_n)}{f_n(z_n)} = \sum_{i=1}^{n-1} \frac{A_i \sin(A_i)\cos(A_i) - \sin^2(A_i)}{A_i^3} \Bigg/ \left[n^{-1/5} \sum_{i=1}^{n-1} \frac{\sin^2(A_i)}{A_i^2} + n^{-1/5}\right] \tag{10}$$

and where A_i is previously defined.

For each observation type j, $\hat{\theta}_n$ is obtained from the data $z_1, \ldots, z_n$ by means of Eq. (9). This is done for $n = 2,3, \ldots$. For a single observation z_n from the density given in Eq. (4), the maximum likelihood estimator for θ_n becomes $1/z_n^2$. Since Eq. (10) is undefined for $n = 1$, $\hat{\theta}_1 = 1/z_1^2$ is defined based on the maximum likelihood results.

State Estimation

We proceed to use the empirical Bayes estimates $\hat{\theta}_{nj}$, $j = 1, \ldots, q$ given by Eq. (9) for each observation type in the usual Kalman filter equations.[2] These equations are

$$\hat{\mathbf{x}}_n = \bar{\mathbf{x}}_n + \bar{P}_n H_n^T(\hat{R}_n + H_n\bar{P}_nH_n^T)^{-1}(\mathbf{y}_n - H_n\bar{\mathbf{x}}_n) \tag{11}$$

$$\hat{P}_n = \bar{P}_n - \bar{P}_nH_n^T(\hat{R}_n + H_n\bar{P}_nH_n^T)^{-1}H_n\bar{P}_n \tag{12}$$

$$\bar{\mathbf{x}}_n = \Phi_n\hat{\mathbf{x}}_{n-1} \tag{13}$$

$$\bar{P}_n = \Phi_n\hat{P}_{n-1}\Phi_n + Q_{n-1} \tag{14}$$

where

$$\hat{R}_n = \operatorname{diag}(r_{n1}^2/\hat{\theta}_{n1}, r_{n2}^2/\hat{\theta}_{n2}, \ldots, r_{nq}^2/\hat{\theta}_{nq}) \tag{15}$$

Equations (11)–(15) in conjunction with Eq. (9) are referred to as the Kalman, Empirical Bayes filter equations.

Recall that the entire development in the preceding section was undertaken on the assumption that the true state vector $\mathbf{x}_i$ is known for all $i = 1, \ldots, n$. In particular, $\mathbf{x}_i$ was used in obtaining z_i given by Eq. (3). This is clearly not the case, and the use of a suitable estimate for each $\mathbf{x}_i$ must be considered. The "best" estimate available for $\mathbf{x}_i$ is $\hat{\mathbf{x}}_i$ given by Eq. (11). However, $\hat{\mathbf{x}}_n$ is unattainable at the current observation processing stage until $\hat{R}_n$ is available, but $\hat{R}_n$ depends upon knowledge of $\hat{\mathbf{x}}_n$. Thus, $\bar{\mathbf{x}}_n$ from Eq. (13) is used in lieu of $\mathbf{x}_n$ in computing z_n in Eq. (3). The success of such approximations is demonstrated in the next section, and a general discussion of the properties of the Kalman, Empirical Bayes filter is presented later.

Illustration of Performance

Monte Carlo simulation was employed to examine the effectiveness and performance characteristics of the Kalman, Empirical Bayes filter. A dynamic system with observations given by Eq. (1) was simulated for various parameter combinations. For simplicity and convenience, no attempt was made to obtain the state transition matrix Φ_i as the numerical solution of a system of differential equations; instead, Φ_i was taken to be a matrix of constants for all i. Likewise, H_i was taken to be a matrix of constants, and the state error covariance matrix Q_i as well as the nominal observation error variances r_{ij}^2 were assumed to be constant over time. All scale factors θ_{ij} were independently sampled from a common distribution, which was a member of the Pearson family of distributions,[11] by means of a Pearson pseudo-random number generator available at Texas Tech University.[12] This family permits Monte Carlo sampling from a variety of density function shapes such as bell, gamma, J, L, and U.

For example, consider the following dynamic system

$$\mathbf{x}_i = \begin{bmatrix} 0.15 & 0.22 & 0.13 & 0.20 & 0.15 & 0.15 \\ 0.20 & 0.15 & 0.16 & 0.18 & 0.15 & 0.16 \\ 0.14 & 0.18 & 0.18 & 0.16 & 0.20 & 0.14 \\ 0.10 & 0.20 & 0.23 & 0.13 & 0.15 & 0.19 \\ 0.25 & 0.15 & 0.15 & 0.15 & 0.15 & 0.15 \\ 0.18 & 0.12 & 0.20 & 0.15 & 0.20 & 0.15 \end{bmatrix} \mathbf{x}_{i-1} + \mathbf{u}_{i-1}$$

$$\mathbf{y}_i = \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 & 0 \end{bmatrix} \mathbf{x}_i + \mathbf{v}_i$$

$$i = 1,2, \ldots$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are normally distributed, each with mean vector $\mathbf{0}$ and covariance matrices

$$Q_i = \operatorname{diag}(0.5, 0.63, 0.37, 0.05, 0.037, 0.5)$$

and

$$R_i = \operatorname{diag}(2.0/\theta_{i1}, 2.0/\theta_{i2}, 2.0/\theta_{i3}, 2.0/\theta_{i4}, 2.0/\theta_{i5})$$

respectively. The values 2.0 in R_i are the nominal variance values discussed earlier. The initial covariance matrix P_0 was assumed to be the identity matrix. This situation was repeated 100 times for $i = 1,2, \ldots, 50$ using different random numbers from the same distributions for $\mathbf{u}$, $\mathbf{v}$, and $\boldsymbol{\theta}$ in each run; that is, 50 sets of observations were simulated 100 times. The Kalman, Empirical Bayes filter equations were employed in forming the state estimate $\hat{\mathbf{x}}_i$, $i = 1,2, \ldots, 50$, and estimated squared-error matrices involving $(\mathbf{x} - \hat{\mathbf{x}}_i)$ were obtained by averaging over the 100 replications. Thus, $\hat{E}(\mathbf{x}_i - \hat{\mathbf{x}}_i)(\mathbf{x}_i - \hat{\mathbf{x}}_i)^T$ was obtained by averaging $(\mathbf{x}_i - \hat{\mathbf{x}}_i)(\mathbf{x}_i - \hat{\mathbf{x}}_i)^T$ for each i over the 100 replications. For comparison, the usual Kalman filter equations were applied to the observation data where the true (but random) observation error variances were used in the calculations. This corresponds to the situation in which the observation error variances are exactly known and thus represents the ultimate in filter performance. The estimated squared-error matrices were then computed as discussed previously. A scalar measure of performance was desired, and the ratio R of the square root of the trace of the average squared-error matrices for $\hat{\mathbf{x}}_i$ in this known situation, compared to that obtained when using the Kalman, Empirical Bayes filter, was formed for $i = 1,2, \ldots, 50$. A ratio of one indicates perfect performance when using the Kalman, Empirical Bayes filter. Figure 1 gives the performance for the example previously described for the indicated mean and variance of the scale factors. In practice, however, these moments are generally unknown and are used here as an index only. The filter performance increases as the number of observations increases, and the maximum gradient is obtained for the first 10 or 15 observation sets. Note that by forming the ratio R, the ability of the filter to naturally increase in accuracy has in effect been "cancelled"; thus, the indicated increase in performance is a result of the empirical Bayes portion of the filter.

The interesting question arises of how the usual Kalman filter performs if the nominal variance values are taken as the true observation error variances. This situation was examined for the same data as in Fig. 1. The $\hat{\theta}$ values in Eq. (15) were set equal to one, and filter Eqs. (11–14) were applied. The corresponding performance ratio R remained essentially constant at 0.80 for 50 sets of observation data. This represents a significant decrease in the average squared-error performance of the filter for estimating $\mathbf{x}_i$ when com-

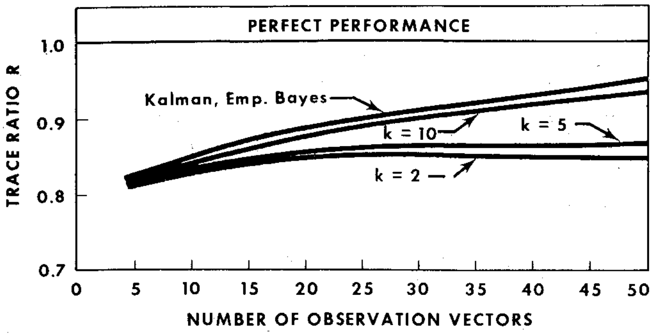

Fig. 2 Performance of Kalman, empirical Bayes, finite-memory adaptation.

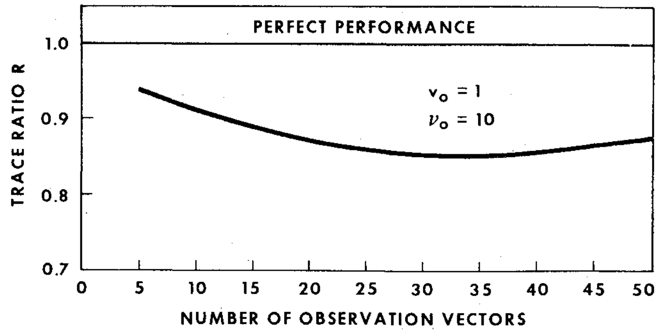

Fig. 3 Performance of Bayes filter.

pared to the Kalman, Empirical Bayes filter performance in Fig. 1. If θ is tightly distributed about a mean of one, this procedure is expected to compare favorably with the method developed here.

Another method is suggested. Suppose θ is considered to be an unknown constant even though it is a random variable. At observation stage n, the approximate maximum likelihood estimate for θ_n becomes

$$n \Big/ \sum_{i}^{n} z_i^2$$

where z_i is given by Eq. (3) with $\hat{\mathbf{x}}_i$ replacing $\mathbf{x}_i$ for $i = 1, \ldots, n - 1$ and $\bar{\mathbf{x}}_n$ replacing $\mathbf{x}_n$. These estimates were used in Eq. (15), and the filter Eqs. (11–14) were applied. For the same data as in Fig. 1, the corresponding performance ratio R leveled out to the value 0.86 after approximately five observation sets. This is a decrease in performance compared to the Kalman, Empirical Bayes results presented in Fig. 1. However, for θ values which are tightly distributed, this technique can be expected to compare favorably with the method presented in this paper.

As previously discussed in the section Empirical Bayes Variance Estimation, the estimates obtained from Eq. (9) are not always positive. Although the estimated marginal density function $f_n(z)$ will usually be unimodal, it will not in general be symmetric with a mode at zero. The "mode" of $f_n(z)$ is defined to be the value of z (e.g., $z_n{}^*$) where $f_n(z)$ attains its maximum value. In effect, we can translate the mode to zero by evaluating Eq. (10) with

$$A_i{}^* = (z_n + z_n{}^* - z_i)/2n^{-1/5} \tag{16}$$

instead of A_i. The mode $z_n{}^*$ of $f_n(z)$ is easily found by means of a computer. When this modal correction was applied, the estimates given by Eq. (9) became positive. Note that the symmetry requirement has not been met by $f_n(z)$, but that this requirement is overtly less critical to the performance of Eq. (9) than the requirement of a zero mode.

Even after the modal correction was made, the Monte Carlo study showed that the estimators given by Eq. (9) almost consistently underestimated the corresponding true scale values. This is partially explainable. The Bayes estimator $E(\theta|z)$ is "unbiased" in the sense that $E\{E(\theta|z)\} = E(\theta)$. Consequently, asymptotically optimal empirical Bayes estimators are "asymptotically unbiased." On the basis of practical experience with empirical Bayes estimators, for small values of n, empirical Bayes estimates of a scale parameter are frequently biased in the sense of frequent underestimation or overestimation of the corresponding parameter. This bias can be attributed to the estimates used for the marginal density and its derivative when approximating the Bayes estimator. Although some of this bias has been removed by the modal correction, some bias still remains.

The use of residuals as the data rather than the weighted observation noise as described in the theoretical development also introduces some degree of bias. The true conditional variance of z_n is larger than the assumed value $1/\theta_n$ due to the use of $\bar{\mathbf{x}}_n$ instead of $\mathbf{x}_n$ in Eq. (3). Since something larger than $1/\theta_n$ is being approximated by $1/\theta_n$, the corresponding estimate $\hat{\theta}_n$ can be expected to be smaller than the estimate obtainable if the weighted observation noise is known.

For all cases considered, it was found that the maximum gradient in R was obtained for the first 10 or 15 sets of observation data. This suggests that a finite-memory adaption might be possible in which the estimates in Eq. (9) are successively computed as n increases by use of a fixed and finite number of past observations for each observation type. This was done, and Fig. 2 gives the trace ratio R for several possibilities using the same data as in Fig. 1. For example, the plot for $k = 5$ was made by basing the estimate in Eq. (9) on five sets of observation data for each observation type. Hence the summation index i in Eq. (10) ranged from a lower limit of $(n - 4)$ and in general from $(n - k + 1)$. It is observed from Fig. 2 that the successive use of 10 sets of past observation data gives a filter performance which is only slightly inferior to that obtained by the successive use of all observation data in the Empirical Bayes portion of the filter. This phenomenon was repeatedly observed in all examples, and it is generally recommended that this finite memory adaptation be made in situations in which computer core storage is limited.

It also was observed that the form of the prior Pearson distribution had little effect on the ratio R. That is, the coefficients of skewness and kurtosis[11] were varied with the mean and variance of the distribution of Θ held constant and the ratio R was observed to change very little. The results agree with the research findings of other empirical Bayes investigations.[6,13]

Discussion of Optimality

It is pertinent to discuss the optimal (or suboptimal) performance characteristics of the Kalman, Empirical Bayes filter developed here.

Since the observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are used for estimating the state as well as the observation error variances, the Bayes estimator for $\mathbf{x}_n$ is not independent of the Bayes estimator for the error variances as has been tacitly assumed. Thus, the two-phase approach used here introduces some degree of suboptimality. This sacrifice has been made for the mathematical convenience in obtaining the procedure developed here, and the simulations have verified that the resulting filter performs satisfactorily. In addition, the empirical Bayes estimators in Eq. (9) have been further stripped of any asymptotic optimal character due to estimates being used for $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, and a modal sign correction. Usually, empirical Bayes estimators are such that they converge in some well-defined sense to the corresponding Bayes estimator as the number of observations tends toward infinity.[7] Even though the modal correction is removed for large values of n, the empirical Bayes estimator may not converge to the Bayes estimator due to the use of residuals as the data rather than the weighted observation noise. Assuming the residuals tend to more closely approximate the observation noise as n increases, the empirical Bayes estimator should more closely approximate the Bayes estimator as n increases, due to the convergence of the estimated marginal density function and its derivative. However, convergence of the empirical Bayes estimator is not guaranteed and no claims can be made. The effectiveness of the filter has been demonstrated for small numbers of observations, and it is for this practical reason that we advocate its use. Note that even under the Bayesian assumption, where $G_j(\theta)$ is known, the resulting filter is not known to be optimal[3]; hence, this fact must be tolerated in either case.

Conclusions

It is emphasized that these conclusions are based on the numerical simulations described earlier. Although only one set of initial conditions was discussed here, numerous simulations were run with varied initial conditions for all quantities except P_0. Accordingly, the conclusions are summarized as follows.

1) A filter has been developed for use in the situation in which observation error variances are unknown, vary randomly with time, and are to be estimated in addition to the state of the system. The recent developments in empirical Bayes estimation theory provide a new contribution to the existing methodology on trajectory estimation, and it is within this framework that the filter is developed. The Kalman, Empirical Bayes filter requires no assumption about the form of the distribution of each variance. The order of magnitude of each variance is required, which is then "adjusted" to provide the variance estimate as observation data becomes available for processing.

2) After a sufficiently large number of observations, the performance of the Kalman, Empirical Bayes filter is nearly as good as the pure Kalman filter with known observation error variances.

3) The performance of the filter was found to be independent of the form of the distribution of the error variances, but dependent on the mean and variance.

4) A finite storage-finite memory adaptation was considered for use in situations where limited computer core storage is available. For running variance estimates based upon the immediately preceding 10 observations, the resulting filter performance is nearly as good as the filter utilizing all preceding observations for estimating the unknown variances.

Appendix: A Bayesian Technique

It is assumed that θ_{nj}, $j = 1,2, \ldots, q$ has a gamma-2 distribution (Ref. 14, p. 226) with density defined by

$$g(\theta_{nj}) = \frac{e^{-(1/2)\nu v \theta_{nj}}(\frac{1}{2}\nu v \theta_{nj})^{\nu/2-1}}{\Gamma(\nu/2)}\left(\frac{1}{2}\nu v\right) \qquad \begin{matrix}\theta_{nj} \geq 0 \\ \nu, v > 0\end{matrix}$$

To estimate the state vector $\mathbf{x}_n$, we annex the usual Kalman filter Eqs. (11–15) with the additional equations

$$\nu_n = \nu_{n-1} + 1(\nu_0 \text{ known}) \tag{A1}$$

and

$$v_n = \frac{v_{n-1}}{\nu_{n-1}+1}[\nu_{n-1} + (y_{nj} - \mathbf{h}_{nj}\hat{\mathbf{x}}_{n-1})^2 \times (\mathbf{h}_{nj}\bar{P}_n\mathbf{h}_{nj}{}^T + v_{n-1}r_{nj}{}^2)^{-1}] \tag{A2}$$

$$(v_0 \text{ known})$$

from which we then compute

$$\hat{\theta}_{nj} = 1/v_n$$

This estimate is then used in Eq. (15) to form $\hat{R}_n$. We note that each observation type j has its own ν and v which are updated by means of Eqs. (A1) and (A2) each time an observation of that same type becomes available for processing. This situation was essentially considered in Ref. 3.

Figure 3 presents the ratio R, as defined for Fig. 1, as a function of the number of observation vectors in which all θ_j,s were assumed to have a common gamma-2 distribution, and where initially $v_0 = 1(\hat{\theta}_{0j} = 1)$ and $\nu_0 = 10$. The same data was used in Figs. 1 and 3. Equations (A1) and (A2) were developed under the assumption that θ_{nj}, once realized according to a gamma-2 distribution, remains constant over time. Since this paper permits θ_{nj} to vary randomly over time, a fair comparison cannot be made for the case of random θ values. We do observe from Figs. 1 and 3 that for the case of random θ values, the Kalman, Empirical Bayes filter monotonically increases in performance whereas the Bayes filter does not. For the same situation in which $\nu_0 = 2$ instead of 10, the resulting curve was only slightly above the curve reported in Fig. 3. This corresponds to lesser precision in the initial knowledge of the scale factors.

References

[1] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, 1960, pp. 35–50.

[2] Bucy, R. S. and Joseph, P. D., *Filtering for Stochastic Processes with Applications to Guidance*, Interscience, New York, 1968, pp. 56–57.

[3] Smith, G. L., "Sequential Estimation of Observation Error Variances in a Trajectory Estimation Problem," *AIAA Journal*, Vol. 5, No. 11, Nov. 1967, pp. 1964–1970.

[4] Robbins, H., "An Empirical Bayes Approach to Statistics," *Proceedings of the Third Berkeley Symposium on Probability and Statistics*, Vol. 1, 1955, pp. 157–163.

[5] Atchison, T. A. and Martz, H. F., eds., *Proceedings of the Symposium of Empirical Bayes Estimation and Computing in Statistics*, Texas Tech Press, Lubbock, Texas, 1970.

[6] Martz, H. F. and Krutchkoff, R. G., "Empirical Bayes Estimators in a Multiple Linear Regression Model," *Biometrika*, Vol. 56, No. 2, 1969, pp. 367–374.

[7] Wasan, M. T., *Parametric Estimation*, McGraw-Hill, New York, 1970.

[8] Smith, G. L., "On the Theory and Methods of Statistical Inference," TR R-251, April 1967, NASA.

[9] Parzen, E., "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, Vol. 33, 1962, pp. 1065–1076.

[10] Martz, H. F., "A Note on Estimating the Derivative of a Density Function with Empirical Bayes Applications," *Proceedings of the Symposium on Empirical Bayes Estimation and Computing in Statistics*, Texas Tech Press, Lubbock, 1970, pp. 141–149.

[11] Elderton, W. P., *Frequency Curves and Correlation*, Cambridge University Press, London, 1953.

[12] Thomas, D. G., "Computer Methods for Generating Pseudo-Random Numbers from Pearson Distributions and Mixtures of Pearson and Uniform Distributions," M.S. thesis, 1966, Virginia Polytechnic Institute.

[13] Clemmer, B. A. and Krutchkoff, R. G., "The Use of Empirical Bayes Estimators in a Linear Regression Model," *Biometrika*, Vol. 55, No. 3, 1968, pp. 525–534.

[14] Raiffa, H. and Schlaifer, R., *Applied Statistical Decision Theory*, Harvard University Press, Boston, 1961.