COMPUTATIONAL EXPERIMENTS IN CONTINUOUS NONLINEAR FILTERING

by

B.W. Licht
and

S.K. Mitter

Systems Research Center
Case Western Reserve University
Cleveland, Ohio 44106

Summary

Five relatively simple continuous nonlinear filtering problems are considered and computational results are presented. For each problem an upper performance bound is computed by numerically evaluating the conditional probability density p(x/Z) via Bucy's representation theorem at certain time instants for a number of Monte Carlo simulations. Numerous approximate nonlinear filtering techniques were investigated and for each of the five problems the performance achieved with the relinearized Kalman-Bucy (K-B) filter approached most closely that of the upper performance bound. Two additional problems are considered for both of which the relinearized K-B filter proves ineffective and alternate filtering methods are given.

I. Introduction

Consider the continuous dynamic process (the signal process) the behavior of which can be adequately represented, or modelled, by the differential equation

$$\frac{dx}{dt} = f(x,t) + g(x,t)w(t) \qquad (1)$$

where x is an n-dimensional state vector, f(x,t) is an n-vector of state functions, g(x,t) is an n x m matrix of disturbance functions, and w(t) is an m-dimensional vector of uncorrelated white-noise processes having unity variances (i.e., $E\{w(t)w(s)\} = I\,\delta(t-s)$). The initial state vector, x(0), is a random variable which is considered to be statistically independent of the disturbance variables w(t), and the probability density at time zero, p(x(0)), is assumed known.

Suppose that certain process variable measurements are made continuously in time and that these measurements are in some sense related to the state vector by the observation equation

$$z(t) = h(x,t) + v(t) \qquad (2)$$

where z(t) is a q-dimensional vector of process measurement variables, h(x,t) is a q-vector of observation functions, and v(t) is a white-noise process vector independent of w(t) and having a covariance matrix R(t) (i.e., $E\{v(t)w(s)\} = 0$ and $E\{v(t)v(s)\} = R(t)\,\delta(t-s)$). The minimum-variance continuous nonlinear filtering problem asks what, given the measurements $\{z(s): 0 \leq s \leq t\}$ of the system represented by relations (1) and (2) and the conditions stated above, is the minimum-variance estimate of the state vector at time t and how does one realistically generate this estimate continuously in time? Physical examples of the nonlinear filtering problem are quite numerous and to stimulate industrial interest in the problem we mention one of them.

A demonstrative example of nonlinear filtering and one having tremendous commercial value is associated with the basic oxygen furnace (BOF) steel-making process. In the BOF process very close-tolerance control of the steel temperature and carbon concentration is desired near the end of the melt since the properties of the final steel product are largely determined by these two variables. The physics of the BOF process, however, makes it extremely difficult to obtain real-time, direct measurements of melt temperature and carbon content, with only the flue gases being accessible for continuous analysis. A nonlinear filter could be constructed for this problem and would utilize a simplified mathematical model of the steel melt dynamics and any available measurement data (e.g., flue gas properties, bomb calorimeter or continuous melt temperature estimates, oxygen lance position, etc.) to generate statistically optimal real-time estimates of steel melt temperature and carbon content. The economic benefits of such a device might be quite significant as reduced batch-times and increased steel qualities could conceivably result from its application.

We have stated the basic problem and attempted to provide some motivation for considering its solution. Happily, the general solution to the continuous minimum-variance nonlinear filtering problem is and has been well known for some period of time[1,2,3]. Unhappily, however, the solution is infinite dimensional in nature and for any practical applications appropriate approximations must be introduced to obtain a finite-dimensional filter. Fortunately, many approximate methods of minimum-variance nonlinear filtering have been proposed, each having its own advantages and disadvantages, some of which are not at all clear. Many questions remain to be answered before the worths of those approximate filtering techniques can be accurately determined; most important, for example, is the question of how far from optimal is the performance of a suboptimal nonlinear filter. We consider this latter question in this paper and hope that by way of examples, some insight can be achieved regarding its answer.

In Section II of this paper we present a very brief summary of the theory of optimal continuous nonlinear filtering and some of the better known methods of approximate filtering. A technique is discussed in Section III which was used to establish upper performance bounds for nonlinear filtering problems by numerical evaluations of the Bucy representation theorem. Five filtering problems are also investigated in Section III and computational results from these investigations are presented. Since the relinearized K-B filter proves to be so misleadingly effective for the five problems, in Section IV two additional problems are considered which demonstrate some of the potential pitfalls of applying the relinearized K-B filtering algorithm.

II. Continuous Filtering: Basic Theory

Optimal Solution

No matter what particular criterion for optimality is specified, the formal solution to the optimal nonlinear filtering problem requires that the conditional probability density $p(x(t)/Z)$ be evaluated, where

$$Z = \{z(s):\ 0 \leq s \leq t\}.$$

For example, the maximum-likelihood estimate of the state vector, $x(t)$, is that vector X which maximizes the density function $p(x(t)=X/Z)$. Similarly, the minimum-variance estimate of $x(t)$ is given as the conditional mean

$$m = \int X\ p(x(t)=X/Z)\,dX\ , \qquad (3)$$

and it is the minimum-variance optimality criterion with which we are primarily concerned in this paper.

Bucy's[3,4] representation theorem, proven by a straightforward application of Bayesian probability theory, states that the conditional probability density satisfies the expression

$$p(x(t)=X/Z) = \frac{E\{\exp\Phi / x(t)=X\}\,p(x(t)=X)}{E\{\exp\Phi\}} \qquad (4)$$

where

$$\Phi = \int_0^t h^T(x(s),s)R^{-1}(s)\left[z(s) - \frac{1}{2}h(x(s),s)\right]ds\ , \qquad (5)$$

$p(x(t))$ is the a'priori probability density of $x(t)$, and the observations, $z(s)$, are held fixed during the averaging operations in Eq. (5). Since it is impractical to have to retain all the observation data, a sequential expression for $p(x/Z)$ is desirable and towards this end it was first shown by Stratonovich[1] and then by Kushner[2] that

$$\frac{\partial p(x/Z)}{\partial t} = \mathcal{L}^*[p(x/Z)] + p(x/Z)(z(t) - \widehat{h(x,t)})^T R^{-1}(t)(h(x,t) - \widehat{h(x,t)}) \qquad (6)$$

where $\mathcal{L}^*$ is the adjoint of the differential generator

$$\mathcal{L}[\cdot] = \sum_i f_i(x,t)\frac{\partial\,\cdot}{\partial x_i} + \frac{1}{2}\sum_{i,j}[gg^T]_{ij}\frac{\partial^2\cdot}{\partial x_i \partial x_j}$$

and the hatted term in Eq. (6) is defined by

$$\widehat{h(x,t)} = \int h(X,t)p(X/Z)\,dX\ . \qquad (7)$$

The many and great difficulties associated with the practical solution of partial-differential-integral equations, such as Eq. (6), make it necessary to parameterize the conditional density function, $p(x/Z)$, and to determine and solve the differential equations satisfied by the parameters. For the minimum-variance filter problem, it is most advantageous to select the central moments for these parameters, and from Eq. (6) one may show (see Kushner[5]) that the first- and second-order central moment parameters for $p(x/Z)$ (i.e., the mean and covariance) satisfy the differential equations

$$\frac{dm_i}{dt} = \widehat{f_i(x,t)} + (\widehat{x_i h(x)} - m_i\widehat{h(x)})^T R^{-1}(z(t) - \widehat{h(x)}) \qquad (8)$$

and

$$\frac{dm_{ij}}{dt} = \overline{\mathcal{L}[(x_i - m_i)(x_j - m_j)]} - (\widehat{x_i h} - m_i\widehat{h})^T R^{-1}(\widehat{x_j h} - m_j\widehat{h}) + (z - \widehat{h})^T R^{-1}\,\overline{(h - \widehat{h})(x_i - m_i)(x_j - m_j)} \qquad (9)$$

where if $\phi(x)$ is any function of x, then $\widehat{\phi(x)} = \int \phi(X)p(X/Z)\,dX$, $m_i = \int X_i p(X/Z)\,dX$, and $m_{ij} = \int (X_i - m_i)(X_j - m_j)p(X/Z)\,dX$.

Equations similar to (8) and (9) can be derived for higher than second-order central moment parameters of $p(x/Z)$ (note that the set of differential equations for the moment parameters constitute, in effect, a continuous minimum-variance nonlinear filter).

Approximate Solutions

Unfortunately, in all but the linear Gaussian case, most applications of Eq's. (8) and (9) and their higher-order counterparts result in an infinite set of coupled integro-differential equations (note that dm_{ij}/dt is a function of m_{iij} if $h(x) = x_i$), making their solution by an optimal nonlinear filter a somewhat "sticky" proposition. Clearly, to successfully apply the optimal filter theory to physical problems some judicious engineering approximations must be made and the effects of such approximations evaluated.

Of the many approximate methods of optimal nonlinear filtering which have appeared in the literature, the following techniques are those which we feel to be among the most common and potentially useful.

Nominal Linearization. The process and observation functions are linearized about a nominally chosen state trajectory (e.g., $h(x) \cong h(x_{nom}) + \nabla h(x_{nom})(x - x_{nom}))$ and the conditional density $p(x/Z)$ is assumed to be Gaussian.

Relinearization. The process and observation functions are linearized about the present best estimate—the conditional mean—of the state vector (e.g., $h(x) \cong h(m) + \nabla h(m)(x - m))$ and $p(x/Z)$ is assumed Gaussian.

Second-Order Expansion.[6] The process and observation functions are expanded in a Taylor series about the conditional mean and all terms of higher than second order are dropped from the series with $p(x/Z)$ assumed to be Gaussian.

Wide-Sense K-B.[7] By a suitable transformation of state variables the nonlinear filtering problem is converted to a linear problem and the initial state density is assumed Gaussian; e.g., if $dx/dt = -x$, $z = x + x^3 + v$, and $p(x(0)) = N(m_o, \sigma_o^2)$, then letting $y_1 = x$ and $y_2 = x^3$ yields the linear equations

$$dy_1/dt = -y_1, \quad dy_2/dt = -3y_2, \quad z = y_1 + y_2 + v$$

with the moments of $p(y_1(0), y_2(0))$ determined from those of $p(x(0))$.

Assumed-Form Density.[5] The differential equations for the moments of $p(x/Z)$ (e.g., Eq's. (8) and (9)) are evaluated by assuming some known form, finite parameter density function for $p(x/Z)$. The number of moments required equals the number of parameters for the density chosen.

How does one decide which particular approximate method of optimal filtering to choose for a particular application? The choice at this time is not at all clear and many questions concerning these approximation techniques remain unanswered. For example; How far from optimal are the performances of these suboptimal filtering schemes? and How influential are the higher than second-order central-moment parameters with regards to filter performance? By computing so-called upper performance bounds for five relatively simple nonlinear filtering problems, we hoped to obtain at least partial answers to these questions.

III. Upper Performance Bounds

From the previous section one may conclude that there exist a number of nonlinear filtering techniques from which to choose from, the appropriate choice not being particularly clear. Each technique differs in complexity and performance from that of any other technique. As a result, the need for trading off complexity and performance is apparent in any nonlinear filter design and for this purpose an upper performance bound is most desirable; for example, one would probably not consider a more complex filter design if the maximum possible increase in performance was only one percent.

The Bucy representation theorem (see Eq. (4), Sect. II) was used in this study to obtain an upper bound on performance (the minimum variance) for nonlinear filters. An approximate numerical evaluation of $p(x/Z)$ was obtained from Eq's. (4) and (5) as follows:

(i) At a point X in the state space at time t let $x(t)=X$ and integrate the process equation (1) backward in time to $t = 0$.

(ii) Using the observations $\{z(s): 0 \leq s \leq t\}$ and the computed state trajectory from (i) evaluate the functional $\Phi(x)$ in Eq. (5) and its exponential value.

(iii) Repeat steps (i) and (ii) a number of times using different disturbance noise sequences $w(t)$ and compute the arithmetic mean of $\exp \Phi(x)$ (skip this step if $w(t)=0$).

(iv) Repeat (i), (ii), and (iii) at a finite number of points X and determine $p(x/Z)$ from Eq. (4). The mean value of the resulting density function represents the minimum-variance estimate of the state vector at time t.

To compute the upper performance bound for a problem, multiple Monte Carlo simulations were conducted of that problem and the statistics of the filtering errors resulting from the minimum-variance estimates computed as described above, were determined. All integrals were evaluated via rectangular integration and all computations were done on a GE 4060 process control digital computer using single precision (24-bit word length) arithmetic.

Nonlinear filters were designed and

evaluated for five example problems. The five problems are defined as follows:

Problem 1.

$$dx_1/dt = -x_1 \quad ; \quad z = x_1 + x_1^3 + v$$

$$p(x_1(0)) = N(1.0, 1.0)$$

Problem 2.

$$dx_1/dt = f(x_1) \quad ; \quad z = x_1 + v$$

$$f(x_1) = \begin{cases} -0.25\, x_1 & , \ x_1 \geq 0 \\ -x_1 & , \ x_1 \leq 0 \end{cases}$$

$$p(x_1(0)) = N(0.0, 1.0)$$

Problem 3.

$$dx_1/dt = f(x_1) \quad ; \quad z = x_1 + v$$

$$f(x_1) = \begin{cases} 2 + x_1 & , \ x_1 \leq -1 \\ -x_1 & , \ -1 \leq x_1 \leq 1 \\ -2 + x_1 & , \ 1 \leq x_1 \end{cases}$$

$$p(x_1(0)) = \begin{cases} 0.25 & , \ -2 \leq x_1(0) \leq 2 \\ 0.0 & , \ 2 < x_1(0) < -2 \end{cases}$$

Problem 4.

$$dx_1/dt = x_1 x_2$$

$$dx_2/dt = 0 \quad ; \quad z = x_1 + v$$

$$p(x_1(0), x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_1(0)) = N(0,1) \ ; \ p(x_2(0)) = N(-1, 1/4)$$

Problem 5.

$$dx_1/dt = x_2$$

$$dx_2/dt = -x_1 - 2.5x_2 \ ; \ z = x_1 x_2 + v$$

$$p(x_1(0), x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_1(0)) = N(1,1) \ ; \ p(x_2(0)) = N(0,1)$$

For each of the above five problems a particular filtering algorithm was evaluated by computing the average squared estimation errors, $(x_i - m_i)^2$, for a sizeable number (N) of Monte Carlo simulations. Identical noise sequences were utilized for each filter studied and time step sizes (Δt) were made small enough to guarantee convergence in statistics. Some results of these studies are presented in the average squared-error vs. time curves of Figures (1) - (6), each of which includes the upper performance bound for the problem of interest, as determined via a numerical evaluation of Bucy's representation for $p(x/Z)$.

Without delving into the particulars of each filter investigated (for a more detailed account consult Reference 8), we can make this general observation from the experimental results presented in Fig's. (1) - (6). For all five of the problems investigated the relinearized K-B filtering scheme proved to be the most effective, yielding almost optimal performance with a relatively simple computational algorithm. More complex filtering methods were considered but in almost every case resulted in a lesser performance figure than that of the less complex relinearized technique.

IV. Relinearized K-B Filter: Pros & Cons

After considering the results of these five computational experiments, one might be tempted to conclude that the relinearized K-B filtering method is the only such method which needs to be considered. This temptation, though strong, should be repressed for the following reasons:

(i) Generalizations should not be drawn from only five simple example problems; more such examples need to be studied.

(ii) The relinearized K-B filter requires the existence of the partial derivatives in the matrices of first partials, $\Delta h(x)$ and $\nabla f(x)$. What does one do when they don't exist? Consider, for example, the problem of estimating on-line the time-delay of a first-order dynamic system from noisy measurements of that system's step response. In particular, consider the system defined by

Problem 6.

$$dx_1/dt = -x_1 + u(t - x_2) + w(t)$$

$$dx_2/dt = 0 \quad ; \quad z = x_1 + v(t)$$

$$u(t) = \begin{cases} 1 \ , & t \geq 0 \\ 0, & t < 0 \end{cases}$$

$$p(x_1(0), x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_1(0)) = N(0, 1/16) \ ; \ p(x_2(0)) = N(1, \tfrac{1}{4})$$

Since the step function $u(t)$ is nondifferentiable at $t = 0$, the relinearized K-B filter algorithm can not be applied directly to this problem. As one alternative, the time delay was approximated by a first-order Pade approximation and the relinearized filter constructed and tested. The results were completely negative and no reduction in variance was achieved for the time-delay estimate.

A remarkably effective filter was finally constructed for Problem 6 by approximating $p(x/Z)$ by a uniform density-function and utilizing the differential equations for the first and second central moments of $p(x/Z)$. The results are discussed in Reference 8 .

(iii) Finally, a rather simple nonlinear filtering problem exists which can not be effectively handled by applying the relinearized K-B filter, and this fact leads one to question the general applicability of this filtering technique. Consider the problem defined by

Problem 7.

$$dx_1/dt = -x_1 + x_2^{1/2}w(t)$$

$$dx_2/dt = 0 \quad ; \quad z = x_1 + v(t)$$

$$p(x_1(0), x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_i(0)) = N(m_{i0}, \sigma_{i0}^2) \ ; \ i = 1,2 \ .$$

The relinearized K-B filter for this problem can be easily shown to satisfy the equations

$$dm_1/dt = -m_1 + m_{11}R^{-1}(z - m_1)$$

$$dm_2/dt = m_{12}R^{-1}(z - m_1) \qquad (10)$$

$$dm_{11}/dt = -m_{11}^2R^{-1} - 2m_{11} + m_2$$

$$dm_{22}/dt = -m_{12}^2R^{-1}$$

$$dm_{12}/dt = -m_{12}m_{11}R^{-1} - m_{12} \qquad (11)$$

with $m_i(0) = m_{i0}$, $m_{ii}(0) = \sigma_{i0}^2$; $i = 1,2$

$m_{12}(0) = 0$.

Note from Eq.(10) that the estimated value of x_2 (m_2) can only change if the value of m_{12} is nonzero. But since $m_{12}(0) = 0$, we see from Eq.(11) that dm_{12}/dt is initially and always of zero value. As a result, this filter provides no useful information regarding the value of x_2—a rather perplexing observation for such an apparently simple nonlinear filtering problem.

An effective filtering algorithm for Problem 7 was constructed from the differential equations for eight of the central moments of p(x/Z) (i.e., m_1, m_2, m_{11}, m_{22}, m_{111}, m_{112}, m_{122}) by assuming p(x/Z) to be essentially Gaussian. See Reference 8 for a complete discussion of this problem.

V. Conclusions

It would be very satisfying at this point to be able to list a concrete set of general guidelines and recommendations for the designer to follow in constructing a nonlinear filter for a particular application. Realistically, however, this is impossible due to the limited number and simplified nature of the seven problems considered in this study. Nevertheless, these problems do stand for themselves and to the less-than-casual observer should provide a significant amount of "feeling" for the characteristic behavior of continuous optimal nonlinear filters. For more detailed information concerning these problems one should consult Reference 8 .

We would, none the less, hazard to make the following general comments regarding nonlinear filtering. For one, it appears that the relinearized K-B filter is a remarkably effective filtering technique for a large class of problems. Secondly, we would point out that the mathematical approximations to the optimal nonlinear filtering problem, as considered in this paper—as well as most of the pertinent literature—are but one source of error in nonlinear filtering. Sufficiently accurate process models and statistical data are also essential ingredients to an effective filtering algorithm and the importance of each should not be minimized by the prospective filter designer.

VI. References

1. Stratonovich, R. L., "Conditional Markov Processes," Theory of Prob. & Applic., Vol. 5, No. 2, 1960, pp. 156-178.
2. Kushner, H. J., "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes with Applications", Jour. S.I.A.M. Control, Vol. 2, No. 1, 1964, pp. 106-119.
3. Bucy, R. S., "Recent Results in Linear and Nonlinear Filtering", Symposium: Stoch. Probs. in Cont., J.A.C.C., 1968, pp. 87-105.
4. Bucy, R. S., and Joseph, P. D., Filtering for Stochastic Processes with Appliations, Interscience Pub., N.Y., 1968.
5. Kushner, H. J., "Approximations to Optimal Nonlinear Filters", IEEE Trans/Auto. Cont, Vol. 12, No. 5, Oct. 1967, pp. 546-556.
6. Bass, R. W., Norum, V. D., and Schwartz, L., "Optimal Multichannel Nonlinear Filtering", Jour. Math. Anal. & Applic., Vol. 16, No. 1, Oct. 1966, pp. 152-164.
7. Lo, J. T., Finite Dimensional Sensor Orbits and Optimal Nonlinear Filtering, Ph.D. Thesis, Univ. South. Cal., August 69.
8. Licht, B. W., Approximations in Optimal Nonlinear Filtering, Ph.D. Thesis, Case Western Reserve Univ., June 70.

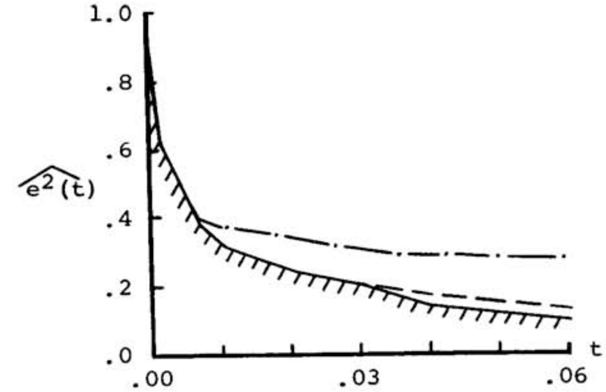

Fig.1 - Problem 1: $\widehat{e^2}$ vs. time for 100 Runs (R = 0.1)

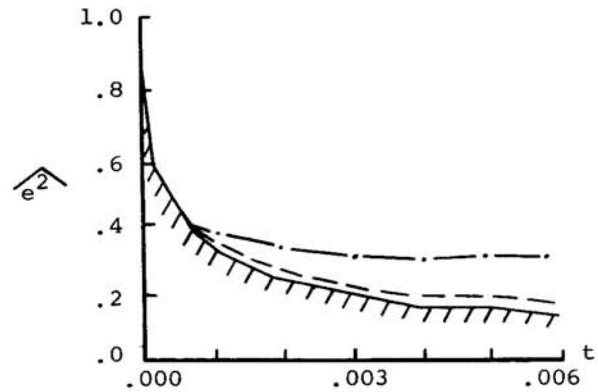

Fig.2 - Problem 1: $\widehat{e^2}$ vs. time for 175 Runs (R = 0.01)

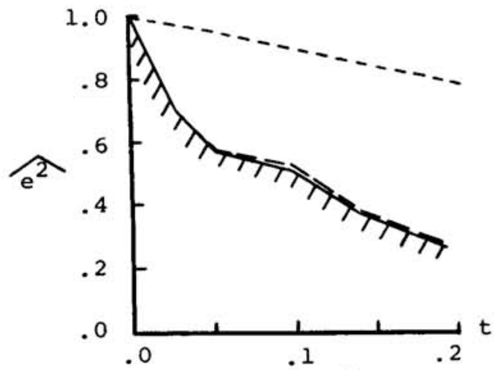

Fig.3 - Problem 2: $\widehat{e^2}$ vs. time for 100 Runs (R = 0.1)

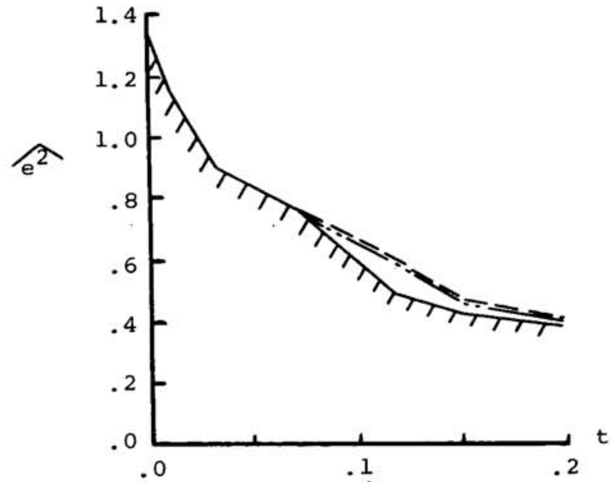

Fig.4 - Problem 3: $\widehat{e^2}$ vs. time for 100 Runs (R = 0.1)

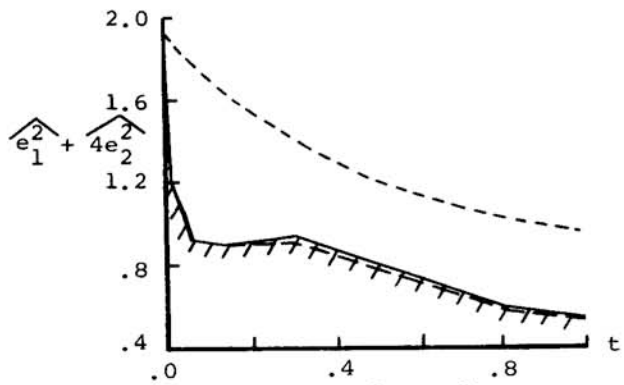

Fig.5 - Problem 4: $\widehat{e_1^2} + 4\widehat{e_2^2}$ vs. time for 50 Runs (R = 0.01)

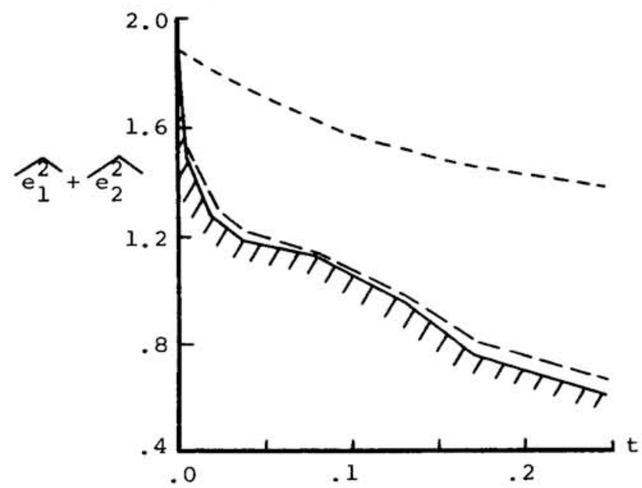

Fig.6 - Problem 5: $\widehat{e_1^2} + \widehat{e_2^2}$ vs. time for 50 Runs (R = 0.01)

Key

Upper Performance Bound

Relinearized K-B Filter

A'Priori Estimator

Wide-Sense K-B Filter

Uniform-Density Filter