Exponential Data Weighting in the Kalman-Bucy Filter*

BRIAN D. O. ANDERSON
Department of Electrical Engineering, University of Newcastle, NSW 2308, Australia.

Communicated by John M. Richardson

ABSTRACT

The use of exponential weighting of data (favoring more recent data) is considered for continuous-time and discrete-time Kalman-Bucy filters. Quantitative consideration is given to matters such as the computation, stability and stationarity of filters designed with data weighting. The problem of divergence or data saturation is explored, and the use of exponential weighting is compared with other techniques for dealing with the problem.

1. INTRODUCTION

Application of the Kalman-Bucy filter theory to practical situations has not always proved straightforward. One of the most commonly observed difficulties is that of *divergence;* this is the name given to the situation arising when the actual estimation errors are statistically inconsistent with the computed estimation errors. In particular, if the computed estimation error covariance approaches zero or ever becomes very small as the measurement interval tends to infinity, the actual error covariance may well be unbounded, approaching infinity as the measurement interval tends to infinity. This phenomenon can arise if there are errors in the modeling of the system upon which filtering is being performed, or errors in the modeling of the statistics of the associated noise. For a full discussion, see Ref. 1.

Several techniques, all with some theoretical basis but perhaps empirical in application, have been suggested for coping with the divergence problem. One technique involves assuming an increase of the noise to the system whose output is being filtered in carrying out the filter design, [1] ; it has been found that the resultant filter performance is less sensitive to modeling errors, and that the filter

*Work supported by Australian Research Grants Committee, Australian-American Educational Foundation in Australia and USAF Contract F44620-68-C-0023 of the Air Force Office of Scientific Research.

has good stability properties—which again implies some sort of ability to recover from the effect of errors.

Other approaches rely on some means for rejecting old data, thus preventing the filter from, to quote Ref. 1, "learning the wrong state too well." More specifically, use of exponential weighting favoring more recent data in discrete time problems, or use of total rejection of data prior to a certain fixed time distance back into the past from the present time, have been suggested. Some results with references are summarized in Ref. 1.

In this paper, we explore the use of exponential weighting, for continuous and discrete time systems. An important result which we show is that use of exponential weighting is equivalent to an assumption of increased input noise (the increase being computable). Use of exponential weighting also possesses the attractive property of yielding a stationary filter when lack of exponential weighting would guarantee a stationary filter. These and other properties are described in the body of the paper.

The layout of the paper is as follows. In Sec. 2, we introduce a quantitative statement of the continuous-time filtering problem with exponential weighting, and in Sec. 3, we derive a formal solution. In Sec. 4, we consider a number of properties, associated with such matters as computation, stability, stationarity, equivalence of exponential weighting and increase of noise input covariance, performance analysis, and the divergence or data saturation problem. In Sec. 5, we review briefly the corresponding discrete time results.

It is interesting to note that the use of exponential weighting for the dual quadratic loss optimal control problem has been covered in the literature, and that many advantages accrue from its use [2]. It was in fact the search for dual results to those of Ref. 2 which led the author to conceive the present material. The problem studied in this paper has also been tackled independently by Sorenson and Sacks, [10], whose work was drawn to the attention of the author after the submission but prior to the publication of this paper. Many results are common to the two papers.

2. PROBLEM STATEMENT

The conventional continuous-time Kalman-Bucy filter problem statement is as follows, [3]. There is prescribed a linear system

$$\dot{x} = F(t)x + G(t)u, \tag{2-1}$$

$$z = H'(t)x + v, \tag{2-2}$$

with $u(\cdot)$ and $v(\cdot)$ zero mean gaussian random processes, $x(t_0)$ a gaussian random variable of mean $\overline{x}_0$, and

$$E[u(t)u'(\tau)] = Q(t)\delta(t-\tau), \quad E[v(t)v'(\tau)] = R(t)\delta(t-\tau),$$
$$E\{[x(t_0)-\overline{x}_0]\,[x(t_0)-\overline{x}_0]'\} = P_0. \tag{2-3}$$

The processes $u(\cdot)$ and $v(\cdot)$ and the initial condition random variable $x(t_0)$ are assumed independent; the matrix $R(t)$ is positive definite for all t. The problem is to find a minimum variance estimate of $x(t)$ given $z(\tau)$, $t_0 \leqslant \tau < t$. In other words, one has to compute $E[x(t)|z(\tau), t_0 \leqslant \tau < t]$ since the minimum variance estimate is the same as the conditional mean [4].

It is not immediately obvious how this problem statement should be modified to allow for exponential weighting of the measurement data. So we turn to another problem statement which is nontrivially equivalent to the first. A good demonstration of the equivalence may be found in Ref. 1.

The same data is assumed to apply. Instead of attempting to directly compute $E[x(t)|z(\tau), t_0 \leqslant \tau < t]$ we attempt to minimize (with respect to $x(\tau)$, $t_0 \leqslant \tau < t$ and $u(\tau)$, $t_0 \leqslant \tau < t$) the functional

$$J_t = \frac{1}{2}(x(t_0) - \overline{x}_0)' P_0^{\#}(x(t_0) - \overline{x}_0)$$

$$+ \frac{1}{2}\int_{t_0}^{t} [z(\tau) - H'(\tau)x(\tau)]' R^{-1}(\tau)[z(\tau) - H'(\tau)x(\tau)]\, d\tau$$

$$+ \frac{1}{2}\int_{t_0}^{t} u'(\tau) Q^{\#}(\tau) u(\tau) d\tau\,, \tag{2-4}$$

subject to the constraint

$$\frac{d}{d\tau}\, x(\tau) = F(\tau)x(\tau) + G(\tau)u(\tau). \tag{2-5}$$

The notation $A^{\#}$ for a square matrix A denotes the Moore-Penrose pseudo-inverse of A.

Let $\hat{x}(\cdot)$ denote the minimizing trajectory $x(\cdot)$. As shown in Ref. 1, the quantity $\hat{x}(t)$ computed in this minimization process is the same as $E[x(t)|z(\tau), t_0 \leqslant \tau < t]$. What is really happening can be described by noting the following formula, which can be proved via an application of Bayes' rule: with $x_i = x(\tau_i)$ and $z_i = z(\tau_i)$, $t_0 \leqslant \tau_i < t$

$$\lim_{n\to\infty} \{Kp(x_1, x_2, \ldots, x_n | z_1, z_2, \ldots, z_n) p(z_1, z_2, \ldots, z_n)\}$$

$$= \exp(-J_t) \qquad |\tau_{i+1} - \tau_i| \to 0. \tag{2-6}$$

Here, K is a normalizing constant depending on the τ_i [1]. The trajectory $\hat{x}(\tau)$, $t_0 \leqslant \tau \leqslant t$, which minimizes J_t must therefore maximize the limit of the a posteriori probability densities $p(x_1, x_2, \ldots, x_n | z_1, z_2, \ldots, z_n)$. It is a nontrivial fact, following from the Markovian and Gaussian nature of the $x(\cdot)$ process, that $\hat{x}(t)$ maximizes $p(x(t)|z(\tau), t_0 \leqslant \tau < t)$, and it is also nontrivial, and also follows from the Gauss-Markov nature of the x process, that

$$\arg\max p(x(t)|z(\tau), t_0 \leqslant \tau < t) = E[x(t)|z(\tau), t_0 \leqslant \tau < t]. \tag{2-7}$$

Using the revised problem statement it is clear how we can introduce exponential weighting favouring more recent data. Let α be a positive constant. Instead of trying to minimize J_t, we try to minimize [with the same constraint Eq. (2-5) as before]

$$\begin{aligned} J_t^\alpha = & \frac{1}{2}(x(t_0) - \bar{x}_0)' P_0^\#(x(t_0) - \bar{x}_0) \\ & + \frac{1}{2}\int_{t_0}^{t} [z(\tau) - H'(\tau)x(\tau)]' e^{2\alpha(\tau - t_0)} R^{-1}(\tau)[z(\tau) - H'(\tau)x(\tau)]\, d\tau \\ & + \frac{1}{2}\int_{t_0}^{t} u'(\tau) e^{2\alpha(\tau - t_0)} Q^\#(\tau) u(\tau)\, d\tau. \end{aligned} \tag{2-8}$$

We no longer have the reinterpretation of this minimization problem as one of maximizing an a posteriori density. But we can argue that the minimization problem associated with Eq. (2-4) is a *least squares* minimization problem with specially selected weighting matrices; then in Eq. (2-8), these weighting matrices are adjusted so as to favor the more recent data.

As indicated by material in Ref. 10, it is possible to tolerate time-varying α. This means that the terms $\exp[2\alpha(\tau - t_0)]$ appearing in Eq. (2-8) will be replaced by $\exp[2\int_{t_0}^{\tau} \alpha(\sigma)d\sigma]$, with a set of consequential changes appearing in later equations. There does however seem little advantage in making the generalization.

3. SOLUTION OF PROBLEM WITH EXPONENTIAL WEIGHTING

Our aim is to minimize with respect to $x(\tau)$ and $u(\tau)$, for appropriate τ the functional Eq. (2-8), and to take as an estimate of $x(t)$ the value $\hat{x}(t)$ of the minimizing trajectory at time t. Initially, our line of attack depends on the observation that Eq. (2-8) is the same as a functional in the class of functionals Eq. (2-4), so that we can use known procedures for computing $\hat{x}(t)$; let us define

$$R_\alpha(\tau) = e^{-2\alpha(\tau - t_0)} R(\tau), \quad Q_\alpha(\tau) = e^{-2\alpha(\tau - t_0)} Q(\tau). \tag{3-1}$$

Then Eq. (2-8) is the same as

$$J_t = \frac{1}{2}(x(t_0) - \bar{x}_0)' P_0^\#(x(t_0) - \bar{x}_0)$$

$$+\frac{1}{2}\int_{t_0}^{t}[z(\tau)-H'(\tau)x(\tau)]'R_\alpha^{-1}(\tau)[z(\tau)-H'(\tau)x(\tau)]\,d\tau$$

$$+\frac{1}{2}\int_{t_0}^{t}u(\tau)'Q_\alpha^{\#}(\tau)u(\tau)d\tau, \tag{3-2}$$

and Eq. (3-2) has the same form as Eq. (2-4). Thus the quantity $\hat{x}(t)$ we are seeking can be regarded as $E[x(t)|z(\tau), t_0 \leqslant \tau < t]$ where x and z are still defined as in Eqs. (2-1) and (2-2), where $u(\cdot)$ and $v(\cdot)$ are still zero mean Gaussian processes and $x(t_0)$ is still a Gaussian random variable of mean $\bar{x}_0$, but now

$$E[u(t)u'(\tau)] = Q_\alpha(t)\delta(t-\tau), \quad E[v(t)v'(\tau)] = R_\alpha(t)\delta(t-\tau). \tag{3-3}$$

The covariance of $x(t_0)$ is as before; we continue to assume that $u(\cdot)$, $v(\cdot)$ and $x(t_0)$ are independent.

In view of Eq. (3-1), we see that *exponential weighting in favor of the more recent measurement is equivalent*, as is intuitively reasonable, *to exponential weighting of the noise covariance matrices*, lowering the values at more recent times. This means that we can compute the modified $\hat{x}(t)$ by a standard procedure [3]. We form the Riccati equation

$$\frac{dP_\alpha}{dt} = P_\alpha(t)F'(t) + F(t)P_\alpha(t) - P_\alpha(t)H(t)R_\alpha^{-1}(t)H'(t)P_\alpha(t)$$
$$+ G(t)Q_\alpha(t)G'(t), \quad P_\alpha(t_0) = P_0, \tag{3-4}$$

and obtain $\hat{x}(t)$ via

$$\frac{d}{dt}\hat{x}(t) = F(t)\hat{x}(t) - P_\alpha(t)H(t)R_\alpha^{-1}(t)[H'(t)\hat{x}(t) - z(t)] \qquad \hat{x}(t_0) = \bar{x}_0. \tag{3-5}$$

4. PROPERTIES OF THE ESTIMATING PROCEDURE

COMPUTATION

Let us note an interesting but minor variation in the computation of P_α and $P_\alpha HR^{-1}$. The Riccati equation (3-4) is

$$\frac{dP_\alpha}{dt} = P_\alpha(t)F'(t) + F(t)P_\alpha(t) - e^{2\alpha(t-t_0)}P_\alpha(t)H(t)R^{-1}(t)H'(t)P_\alpha(t)$$
$$+ e^{-2\alpha(t-t_0)}G(t)Q(t)G'(t), \quad P_\alpha(t_0) = P_0.$$

Let us set

$$\Pi_\alpha = e^{2\alpha(t-t_0)}P_\alpha. \tag{4-1}$$

Then

$$\dot{\Pi}_\alpha = 2\alpha\Pi_\alpha + e^{+2\alpha(t-t_0)}\dot{P}_\alpha ,$$
$$= \Pi_\alpha(F' + \alpha I) + (F + \alpha I)\Pi_\alpha - \Pi_\alpha H R^{-1} H' \Pi_\alpha + GQG', \quad \Pi_\alpha(t_0) = P_0 . \qquad (4\text{-}2)$$

Observe also that

$$\Pi_\alpha(t)H(t)R^{-1}(t) = P_\alpha(t)H(t)R_\alpha^{-1}(t) . \qquad (4\text{-}3)$$

It follows from Eqs. (4-2) and (4-3) that *computation of the filter with exponential weighting can proceed not by changing the noise covariances, but by changing* F *to* F + αI *in the "variance," but not the state estimate, equation.* This may be an advantage if F, G, H, Q, R and R^{-1} are bounded.

BOUNDEDNESS OF Π_α *(t)*

Sufficient conditions for $P(t)$ to be bounded are known [1, 5, and 6]. These are

(a) F, G, H, Q, R and R^{-1} are bounded,

(b) $[F, H]$ is uniformly completely observable.

Note: Refs. 1 and 5 appear to demand further conditions, but these are inessential as pointed out in Ref. 6.

If F is replaced by $F + \alpha I$, neither of (a) or (b) is varied. [To check that (b) is unvaried requires a small calculation.] Therefore (a) and (b) imply that $\Pi_\alpha(t)$ is bounded.

STABILITY

From Ref. 5, sufficient conditions for the exponential asymptotic stability of the filter derived with no exponential weighting are known to be

(a) F, G, H, Q, R and R^{-1} are bounded,

(b) $[F, H]$ is uniformly completely observable,

(c) $[F, GD]$ is uniformly completely controllable for any D such that $DD' = Q$.

From Ref. 6, sufficient conditions for asymptotic stability, but not necessarily exponential asymptotic stability, of the filter derived with no exponential weighting are known to be (a) and (b) above, and (c) relaxed to:

$$(c') \quad P_0 + \int_{t_0}^{t_1} \Phi(t_0, t)G(t)Q(t)G'(t)\Phi'(t_0, t)dt ,$$

is nonsingular for some t_1, where $\Phi(\cdot,\cdot)$ is the transition matrix of $\dot{y} = Fy$.

Condition (c') is a modified form of complete controllability condition. Note that it is satisfied if P_0 is nonsingular, irrespective of F, G, and Q. It is also satisfied if condition (c) is satisfied.

For constant α, it is easy to show that $[F+\alpha I, GD]$ is uniformly completely controllable if $[F, GD]$ is. Likewise $[F+\alpha I, H]$ is uniformly completely observable if $[F, H]$ has this property. Also, the transition matrix of $\dot{y} = (F+\alpha I)y$ is $e^{\alpha t}\Phi(t,\tau)e^{-\alpha\tau}$ which means that if condition (c$'$) holds, it will continue to hold if $\Phi(t,\tau)$ is replaced by $e^{\alpha t}\Phi(t,\tau)e^{-\alpha\tau}$. So if conditions (a), (b) and (c), or (a), (b) and (c$'$) hold for a certain set of F, G, etc., they will also hold if F is replaced by $F+\alpha I$. Assuming (a), (b) and (c$'$) hold, it follows that the filter derived for the system

$$\dot{x} = (F+\alpha I) + Gu,$$
$$z = H'x + v, \tag{4-4}$$

(with u, v and $x(t_0)$ satisfying Eq. (2-3) and the associated remarks, there being no exponential weighting) will be asymptotically stable. This filter is obtained by solving Eq. (4-2) and is

$$\dot{\hat{x}} = (F+\alpha I)\hat{x} - \Pi_\alpha HR^{-1}[H'\hat{x} - z] \ . \tag{4-5}$$

Thus $F+\alpha I - \Pi_\alpha HR^{-1}H'$ is an asymptotically stable matrix. It follows immediately that $F - \Pi_\alpha HR^{-1}H'$ is a matrix such that

$$\dot{\hat{x}} = F\hat{x} - \Pi_\alpha HR^{-1}[H'\hat{x} - z] \ , \tag{4-6}$$

has (exponential) degree of stability α. But in view of Eq. (4-3), Eq. (4-6) is the filter equation we derived in Eq. (3-5) for the original system with exponential weighting. So *the introduction of exponential weighting increases the degree of stability of the filter by α, and will convert nonexponential asymptotic stability to exponential asymptotic stability.*

STATIONARITY

Suppose F, G, H, Q and R are constant and that t_0 is $-\infty$. Suppose also that $[F, H]$ is completely observable. Then the filter associated with the original system without exponential weighting exists and is stationary [5]. By replacing F by $F+\alpha I$, we see that *the filter designed with exponential weighting also exists and is stationary* .

EQUIVALENCE OF EXPONENTIAL WEIGHTING AND INCREASING THE INPUT NOISE

One of the techniques for improving the computational stability of filters is the artificial raising of the input noise [1]. Let us now note how exponential weighting is equivalent to raising the input noise.

We may rewrite Eq. (4-2) as

$$\dot{\Pi}_\alpha = \Pi_\alpha F + F'\Pi_\alpha - \Pi_\alpha HR^{-1}H'\Pi_\alpha + (GQG' + 2\alpha\Pi_\alpha). \quad (4\text{-}7)$$

Thus Π_α satisfies the same Riccati equation as the matrix P which would define a filter design without exponential weighting, save that the covariance matrix GQG' is increased to $GQG' + 2\alpha\Pi_\alpha$. *Thus use of exponential weighting amounts to increasing the input noise by an amount determinable only after the design with exponential weighting.*

The filter that results from exponential weighting could have resulted from assumption of a measurement noise covariance which was too large coupled with a standard filter design, with no weighting. By a theorem of Nishimura [7], this means that the filter when used on the system defined by the original Eqs. (2-1) through (2-3) and the associated remarks will, provided the system description is accurate, lead to an actual error variance $E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}$ which is larger than that which would be achieved were the correct filter used.

PERFORMANCE ANALYSIS

Let us compute the error variance $E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}$ resulting from a filter design with exponential weighting. In [1], an analysis is performed computing $E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}$ when the filter producing $\hat{x}(t)$ is designed assuming a larger input noise covariance matrix than is actually present. Now, the matrix

$$\mathcal{E}(t) = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\} \quad (4\text{-}8)$$

resulting from an exponential weighting assumption in the filter design is the same as the error variance matrix resulting from assuming $GQG' + 2\alpha\Pi_\alpha$ rather than GQG' as the input noise term. Formulas of Ref. 1 thus apply to yield

$$\dot{\mathcal{E}} = [F - \Pi_\alpha HR^{-1}H']\,\mathcal{E}(t) + \mathcal{E}(t)\,[F - \Pi_\alpha HR^{-1}H']' + GQG' + \Pi_\alpha HR^{-1}H'\Pi_\alpha, \quad (4\text{-}9)$$

with initial condition $\mathcal{E}(t_0) = P_0$. This is a linear differential equation, and thus perhaps easier to solve than a Riccati differential equation. In the event that the plant and filter are stationary, $\mathcal{E}(t)$ will be constant and will satisfy

$$[F - \Pi_\alpha HR^{-1}H']\,\mathcal{E} + \mathcal{E}\,[F - \Pi_\alpha HR^{-1}H']' = -GQG' - \Pi_\alpha HR^{-1}H'\Pi_\alpha. \quad (4\text{-}10)$$

An upper bound on $\mathcal{E}(t)$ which may be useful, since solution of Eq. (4-9) may not then be required, is as follows:

$$\mathcal{E}(t) \leqslant \Pi_\alpha(t). \quad (4\text{-}11)$$

This follows readily from Eqs. (4-2) and (4-9).

COPING WITH DIVERGENCE OR DATA SATURATION

Inaccuracies in modeling and round off errors may cause disastrous performance of an apparently well designed Kalman filter; actual error variances, though predicted to be bounded, may be unbounded in practice. For example, suppose that a system is modelled by

$$\dot{x} = 0,$$
$$y = x + v, \tag{4-12}$$

with x a scalar, $E[v(t)v(\tau)] = \delta(t - \tau), E[x^2(0)] = 1, E[x(0)] = 0$ and $x(0)$ and $v(\cdot)$ independent. The optimum filter is

$$\dot{\hat{x}} = \frac{1}{t+1}(y - \hat{x}),$$

and the theoretical error variance is

$$p(t) = \frac{1}{t+1}$$

As the theory predicts, the filter is asymptotically stable. It is not, however, exponentially asymptotically stable.

Now suppose that in reality, the system is

$$\dot{x} = u$$
$$y = x + v.$$

If u a zero mean Gaussian random process with covariance $E[u(t)u(\tau)] = \epsilon\delta(t - \tau)$ for some constant ϵ, and if the same filter is used, the actual error variance is

$$\frac{\epsilon}{3}(t+1) + \frac{1}{t+1} - \frac{\epsilon}{3(t+1)^2}$$

which is an unbounded function of time for any nonzero ϵ. Again, suppose that u is now a nonzero constant ϵ; then use of the original filter will lead to an actual error covariance of

$$\frac{1}{t+1} + \frac{1}{2}\epsilon(t+1),$$

which again is unbounded. Both these situations are examples of the divergence problem.

Three remedies have been suggested to cope with this sort of problem [1].

They are

(1) Artificial increase of the input noise covariance,
(2) Use of exponential weighting,
(3) Use of limited memory filtering.

(Actually, use of the second remedy has only been proposed for discrete time problems.) We have already argued that use of exponential weighting is equivalent to a certain increase of the input noise covariance. We can also argue, at least qualitatively, that use of exponential weighting is equivalent to use of limited memory filtering. Consider the functional (2-8). At least if the processes in question are stationary and Q and R are constant, we could argue that we are discarding all measurements from times earlier than $t - (5/\alpha)$ where t is the present time. Certainly too if the processes are time-varying and R and Q nonconstant, we would expect that with reasonable sorts of variation, an approximate sort of limited memory filtering was in effect.

Let us now note more precisely one of the ways in which use of exponential weighting copes with data saturation. The conditions (a), (b) and (c') listed earlier which guarantee asymptotic stability of the filter designed with no weighting guarantee exponential asymptotic stability of the filter designed with weighting. They also guarantee boundedness of $\Pi_\alpha(t)$ and the nonsingularity of $P(t)$ for $t \geqslant t_1$ [6], and thus of $\Pi_\alpha(t)$ for $t \geqslant t_1$, since $\Pi_\alpha(t) \geqslant P(t) > 0$. Now we have argued that the use of exponential weighting is equivalent to raising the input noise covariance term GQG' to $GQG' + 2\alpha\Pi_\alpha$. Hence for α nonzero, the effective input noise covariance is guaranteed to be nonsingular. Since a special bounded nonsingular input noise covariance is known to yield a bounded estimation error variance, it follows that any input noise with bounded covariance will yield a bounded error variance. Also, any input noise with nonzero mean will yield a bounded error variance. In other words, variation of the input noise away from the design figure will not cause disastrous filter performance.

5. DISCRETE TIME RESULTS

Exponential weighting for discrete-time filtering problems has been suggested in Refs. 1 and 8, but virtually no analysis is made concerning its use. Analysis as for the continuous case is possible, but rather than repeating the foregoing material in detail, we shall indicate results in outline form.

There is prescribed a linear system

$$x(k+1) = F(k)x(k) + G(k)u(k+1), \tag{5-1}$$

$$z(k) = H'(k)x(k) + v(k), \tag{5-2}$$

with $u(\cdot)$ and $v(\cdot)$ independent zero mean Gaussian processes and

$$E[u(k)u'(l)] = Q(k)\delta(k-l), \quad E[v(k)v'(l)] = R(k)\delta(k-l). \tag{5-3}$$

Also, $x(k_0)$ is a Gaussian random variable, independent of $u(\cdot)$ and $v(\cdot)$, with mean $\overline{x}_0$ and covariance

$$E\{[x(k_0)-\overline{x}_0][x(k_0)-\overline{x}_0]'\}=P_0. \tag{5-4}$$

The matrices $F(k)$ and $R(k)$ are assumed nonsingular for all k.

The probability density $p(x(k_0),\ldots,x(k)\,|\,z(k_0+1),\ldots,z(k))$ is [1],

$$\frac{a\exp(-J_k)}{p(z(l),k_0+1\leqslant l\leqslant k)}, \tag{5-5}$$

where a is a constant and

$$J_k=\frac{1}{2}[x(k_0)-\overline{x}_0]'P_0^{\#}[x(k_0)-\overline{x}_0]$$
$$+\frac{1}{2}\sum_{l=k_0+1}^{k}\{[z(l)-H'(l)x(l)]'R^{-1}(l)[z(l)-H'(l)x(l)]$$
$$+u'(l)Q^{\#}(l)u(l). \tag{5-6}$$

(The constraint equation (5-1) is assumed to apply).

Finding the trajectory $\hat{x}(k_0),\ldots,\hat{x}(k)$ which maximizes the a posteriori density $p(x(k_0),\ldots,x(k)\,|\,z(k_0+1),\ldots,z(k))$ is the same as finding the trajectory $\hat{x}(k_0),\ldots,\hat{x}(k)$ which minimizes J_k. It can be shown that $\hat{x}(k)$ also maximizes $p(x(k)\,|\,z(k_0+1),\ldots z(k))$.

By reinterpreting the minimization of Eq. (5-6) as a least squares problem, it then makes sense to exponentially weight the data, weighting more recent data more heavily, by replacing $R(k)$ and $Q(k)$ by

$$R_\alpha(k)=\alpha^{-2(k-k_0)}R(k),\quad Q_\alpha(k)=\alpha^{-2(k-k_0)}Q(k) \tag{5-7}$$

for some $\alpha>1$.

The filter for the system governed by Eqs. (5-1) through (5-3) is given by Ref. 9,

$$\hat{x}(k)=F(k-1)\hat{x}(k-1)+K(k)[z(k)-H'(k)F(k-1)\hat{x}(k-1)], \tag{5-8}$$

where

$$K(k)=\hat{P}(k)H(k)[H'(k)\hat{P}(k)H(k)+R(k)]^{-1}, \tag{5-9}$$

$$\hat{P}(k)=F(k-1)P(k-1)F'(k-1)+G(k-1)Q(k)G'(k-1), \tag{5-10}$$

and

$$P(k)=[I-K(k)H'(k)]\hat{P}(k)[I-K(k)H'(k)]'+K(k)R(k)K'(k), \tag{5-11}$$

with the iteration started by $\hat{x}(k_0)=\overline{x}_0$ and $P(k_0)=P_0$. The matrix $P(k)$ is $E\{[x(k)-\hat{x}(k)][x(k)-\hat{x}(k)]'\}$.

The filter for the system governed by Eqs. (5-1), (5-2) and (5-7) may be found by replacing $R(k)$ and $Q(k)$ in Eqs. (5-9) through (5-11) above by $R_\alpha(k)$ and $Q_\alpha(k)$. We shall denote the corresponding replacements of $K(k), \hat{P}(k)$ and $P(k)$ by $K_\alpha(k), P_\alpha(k)$ and $\hat{P}_\alpha(k)$.

Let us define

$$\hat{\Pi}_\alpha(k) = \alpha^{2(k-k_0)}\hat{P}_\alpha(k), \tag{5-12}$$

and $\Pi_\alpha(k)$ similarly. Then we can obtain equations for $K_\alpha(k)$, $\hat{\Pi}_\alpha(k)$ and $\Pi_\alpha(k)$ as

$$K_\alpha(k) = \hat{\Pi}_\alpha(k)H(k)[H'(k)\hat{\Pi}_\alpha(k)H(k) + R(k)]^{-1}, \tag{5-13}$$

$$\hat{\Pi}_\alpha(k) = [\alpha F(k-1)]\,\Pi_\alpha(k-1)\,[\alpha F'(k-1)] + G(k-1)Q(k)G'(k-1), \tag{5-14}$$

$$\Pi_\alpha(k) = [I - K_\alpha(k)H'(k)]\,\hat{\Pi}_\alpha(k)\,[I - K_\alpha(k)H'(k)]' + K_\alpha(k)R(k)K'_\alpha(k). \tag{5-15}$$

By comparing Eqs. (5-13) through (5-15) with Eqs. (5-9) through (5-11), we see that the optimal gain matrix $K_\alpha(k)$ resulting from a filter design with exponential weighting might just as well have resulted from a filter design with no exponential weighting, but with $F(k)$ replaced by $\alpha F(k)$ for all k.

In general, $P_\alpha(k)$ will be unbounded as $k \to \infty$ while $\Pi_\alpha(k)$ will remain bounded. So there is computational advantage in using Eqs. (5-13) through (5-15) to obtain the filter.

By rewriting Eq. (5-14) as

$$\begin{aligned}\hat{\Pi}_\alpha(k) = F(k-1)\Pi_\alpha(k)F'(k-1) + \{&G(k-1)Q(k)G'(k-1)\\ &+ (\alpha^2-1)F(k-1)\Pi_\alpha(k)F'(k-1)\},\end{aligned} \tag{5-16}$$

we see too that the design with exponential weighting might just as well have resulted from a filter design with no exponential weighting but with the noise covariance matrix $G(k-1)Q(k)G'(k-1)$ increased to $G(k-1)Q(k)G'(k-1) + (\alpha^2-1)F(k-1)\Pi_\alpha(k)F'(k-1)$ for each k.

As for the continuous time case, we can establish various significant properties. Let us outline these.

(1) $\Pi_\alpha(k)$ is bounded if F, F^{-1}, G, H, Q, R and R^{-1} are bounded and $[F, H]$ is uniformly completely observable.

(2) The filter designed with weighting is exponentially asymptotically stable with degree of stability α if the conditions of (1) are satisfied, and if either $[F, GQ^{1/2}]$ is uniformly completely controllable, or the following weaker condition holds:

$$P_0 + \sum_{k=k_0}^{k_1-1} \Phi(k_0, k+1)G(k)Q(k+1)G'(k)\Phi'(k_0, k+1)$$

is nonsingular for some k_1, where $\Phi(k,l) = F(k-1)F(k-2)\ldots F(l)$, $k > l$, $\Phi(l,l) = I$ and $\Phi(k,l) = \Phi^{-1}(l,k)$ for $k < l$.

(3) As already noted, use of exponential weighting is equivalent to assuming a larger input noise covariance than is actually present. As is clear from Eq. (5-16), the amount of the increase is only determinable after the filter design with exponential weighting. When the filter designed with exponential weighting is used on the original system, and the assumed data concerning the original system is correct, the actual error variance

$$\mathcal{E}[(k) = E\{[x(k) - \hat{x}(k)]\,[x(k) - \hat{x}(k)]'\} \tag{5-17}$$

satisfies

$$P(k) \leqslant \mathcal{E}(k) \leqslant \Pi_\alpha(k), \tag{5-18}$$

and it also may be derived from

$$\hat{\mathcal{E}}(k) = F(k-1)\mathcal{E}(k-1)F'(k-1) + G(k'-1)Q(k)G'(k-1), \tag{5-19}$$

$$\mathcal{E}(k) = [I - K_\alpha(k)H'(k)]\,\hat{\mathcal{E}}(k)[I - K_\alpha(k)H'(k)] + K_\alpha(k)R(k)K_\alpha'(k), \tag{5-20}$$

with starting condition $\mathcal{E}(k_0) = P_0$.

(4) The filter designed with exponential weighting provides one technique for coping with the divergence or data saturation problem, at least when $\Pi_\alpha(k)$ is nonsingular—which will be the case if the conditions listed in (2) above are fulfilled. One may regard the use of exponential weighting as being equivalent to the use of a raised input noise covariance matrix or, qualitatively at least, to the use of limited memory filtering. In particular, when the conditions listed in (2) are satisfied, variations in the system input noise should not cause disastrous performance of the filter if the filter is designed with exponential weighting.

6. CONCLUSIONS

The main aim of this paper has been to place the use of exponential weighting of data in a Kalman-Bucy filter design on a theoretical as distinct from empirical basis. Of particular interest in the actual results is that the use of exponential weighting is equivalent to an assumption that the plant input noise is higher than is really the case, since both use of this assumption and of exponential weighting have been proposed as techniques for dealing with the problem of data saturation. Exponential weighting offers no more formal computational difficulties in filter design than occur in a normal filter design, in contrast perhaps to the limited memory filtering technique of Ref. 1, which is a further technique aimed at dealing with the data saturation problem. Stationarity and stability properties are easy to predict, and information is available as to the suboptimality of filters designed using exponential weighting.

ACKNOWLEDGMENT

Part of this work was performed while the author was a Visiting Professor in the Information and Control Sciences Center, Institute of Technology, Southern Methodist University, Dallas, Texas 75222, U.S.A. Sincere thanks are extended to Professor Andrew P. Sage for his hospitality at this time.

REFERENCES

1 A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic, N. Y. (1970).
2 B. D. O. Anderson, and J. B. Moore, Linear system optimization with prescribed degree of stability, *Proc. IEE* **116**, 2083 (1969).
3 R. E. Kalman, and R. S. Bucy, New results in linear filtering and prediction theory, *Trans. ASME, Series D, J. Basic Eng.* **83**, 95 (1961).
4 A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, N. Y. (1965).
5 R. E. Kalman, New methods and results in Linear Filtering and Prediction Theory, *Proceedings of the Symposium on Engineering Applications of Probability and Random Functions*, Wiley, New York (1961).
6 B. D. O. Anderson, Stability Properties of Kalman-Bucy Filters, *J. Franklin Institute* **291**, 137 (1971).
7 T. Nishimura, Error Bounds of Continuous Kalman Filters and the Application to Orbit Determination Problems, *IEEE Transactions on Automatic Control,* **12**, 268 (1967).
8 S. L. Fagin, Recursive Linear Regression Theory, Optimal Filter Theory, and Error Analysis of Optimal Systems, *IEEE Intern. Conv. Record*, 12, 216 (1964).
9 R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Trans. ASME, Series D, J. Basic Eng.* **82**, 35 (1960).
10 H. W. Sorenson, and J. E. Sacks, Recursive Fading Memory Filtering, *Information Science*, **3**, (1971).

Received April 1971