TABLE I
METHODS OF AVERAGING SINGLE-STAGE ESTIMATES

Type of Average	Summation Form	Recursive Form
Straight Average	$\bar{p}_k = \frac{1}{k}\sum_1^k \hat{p}_i$	$\bar{p}_k = \frac{k-1}{k}\bar{p}_{k-1} + \frac{1}{k}\hat{p}_k, \quad \bar{p}_0 = 0$
Fading-Memory Average ($\beta < 1$)	$\bar{p}_k = \left(\sum_1^k \beta^{k-i}\right)^{-1} \sum_1^k \beta^{k-i}\hat{p}_i$	$\bar{p}_k = \frac{C_{k-1}}{C_k}\bar{p}_{k-1} + \frac{1}{C_k}\hat{p}_k, \quad \bar{p}_0 = 0$ $C_k = \beta C_{k-1} + 1, \quad C_0 = 0$
Weighted Average	$\bar{p}_k = \dfrac{\sum_1^k \operatorname{tr} \hat{P}_i W_i}{\sum_1^k \operatorname{tr} P_0^i W_i}$	$\bar{p}_k = \frac{\lambda_{k-1}}{\lambda_k}\bar{p}_{k-1} + \frac{\alpha_k}{\lambda_k}\hat{p}_k, \quad \bar{p}_0 = 0$ $\alpha_k = \operatorname{tr} P_0^i W_i$ $\lambda_k = \sum_1^k \alpha_i = \lambda_{k-1} + \alpha_k, \quad \lambda_0 = 0$
Faded-Memory Weighted Average ($\beta < 1$)	$\bar{p}_k = \dfrac{\sum_1^k \beta^{k-i}\alpha_i\hat{p}_i}{\sum_1^k \beta^{k-i}\alpha_i}$	Same as Weighted Average, But $\lambda_k = \beta\lambda_{k-1} + \alpha_k, \quad \lambda_0 = 0$

following four candidates are proposed; formulas for the smoothed estimates are given in Table I.

1) *Straight Averaging:* This is the analog of (5).

2) *Fading-Memory Averaging* [6]: This allows tracking of a slowly-varying p, and helps to discard poor estimates that develop early in the estimation process when the Δ_k are poorly known.

3) *Weighted Average:* The weighting coefficients in the average are chosen to minimize the variance of $\bar{p}_k$, assuming the $\mathbf{v}_k$ to be uncorrelated, a situation which, hopefully, occurs once p has been satisfactorily identified.

4) *Fading-Memory Weighted Average:* Fading memory is incorporated in the weighted average using Sorenson and Sacks' approach [13].

APPLICATION TO FADING-MEMORY FILTERING

These algorithms may be used to adaptively choose the fading factor in the faded-memory Kalman filter [12] by taking

$$\Delta_k = H_k Q_{k-1} H_k' + R_k$$
$$P_0^k = H_k A_{k-1} \Sigma_{k-1} A_{k-1}' H_k'$$

and p as the fading factor.

REFERENCES

[1] R. K. Mehra, "Approaches to adaptive filtering," in *Proc. 9th IEEE Symp. Adaptive Processes*, Austin, Texas, Dec. 7–9, 1970.
[2] I. M. Weiss, "A survey of discrete Kalman–Bucy filtering with unknown noise covariances," presented at the AIAA Guidance, Control, and Flight Mechanics Conf., Santa Barbara, Calif., Aug. 1970, paper 70-955.
[3] A. P. Sage and G. W. Husa, "Algorithms for sequential adaptive estimation of prior statistics," in *Proc. 8th IEEE Symp. on Adaptive Processes*, Pennsylvania State University, State College, Pa., Nov. 17–19, 1969.
[4] Andrew H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.
[5] H. W. Sorenson, "Kalman filtering techniques," in *Advances in Control Systems*, vol. 3, C. T. Leondes, Ed. New York: Academic, 1966, pp. 219–292.
[6] J. C. Shellenbarger, "Estimation of covariance parameters for an adaptive Kalman filter," in *Proc. National Electronics Conf.*, Chicago, Ill., 1966, pp. 698–702.
[7] A. P. Sage and G. W. Husa, "Adaptive filtering with unknown prior statistics," in *Proc. 1969 Joint Automatic Control Conf.*, Boulder, Colo., Aug. 1969, pp. 760–769.
[8] B. D. Tapley and G. H. Born, "Sequential estimation of the state and the observation-error covariance matrix," *AIAA J.*, vol. 9, pp. 212–217, Feb. 1971.
[9] L. J. Levy, "Adaptive estimation algorithms," Iowa State Univ. Eng. Res. Inst., Ames, Iowa, Tech. Rep. ERI-93400, Dec., 1970.
[10] M. Athans and F. C. Schweppe, "Gradient matrices and matrix calculations," Mass. Inst. Technol., Lincoln Lab., Cambridge, Mass., Tech. Note 1965-53, Nov. 17, 1965.
[11] H. Hotelling, "A generalized *T* test and measure of multivariate dispersion," in *Proc. 2nd Berkeley Symp. Math. Statistics and Probability*. Berkeley, Calif.: Univ. of California Press, 1951, pp. 23–42.
[12] H. W. Sorenson and J. E. Sacks, "Recursive fading-memory filtering," *Inform. Sci.*, vol. 3, pp. 101–119, 1971.

A Note on Kalman–Bucy Filters with Zero Measurement Noise

P. J. MOYLAN

Abstract—**The limiting form of the Kalman–Bucy filter as measurement noise tends to zero does not, in general, correspond to the optimal filter derived assuming zero measurement noise. This may be considered to be due to a difference in initial conditions.**

A standard problem in linear filtering theory is the following: given the process equations

$$\dot{x}(t) = Fx(t) + Gu(t)$$
$$y(t) = H'x(t) + v(t)$$

where $u(t)$ and $v(t)$ are sample functions from independent Gaussian white noise processes, find a filter generating the conditional mean

$$\hat{x}(t) = E[x(t)|y(\tau), t_0 \leqslant \tau \leqslant t].$$

It is assumed that $x(t_0)$ is a Gaussian random variable, of zero mean and known covariance P_0, and that the noise covariances are also given as

$$E[u(t)u'(\tau)] = Q\delta(t - \tau)$$

and

$$E[v(t)v'(\tau)] = R\delta(t - \tau).$$

If the matrix R is positive definite, the solution is simply a Kalman–Bucy filter [1]. If R is singular, however, the problem becomes more complicated. In the particular case where $R = 0$, there are at least two plausible approaches.

1) Assuming $R = \epsilon R_0$, with R_0 any positive definite matrix, and ϵ a positive scalar parameter, write down the equations for the Kalman–Bucy filter; then take the limit as $\epsilon \to 0$. This approach is implicit in, for example, the treatment of Kwakernaak and Sivan [2].

2) Differentiate the measurement process $y(t)$. At least in the case where $H'GQG'H$ is nonsingular, the new measurement vector $\dot{y}(t)$ will contain a nonsingular white-noise component, and again a

Manuscript received November 5, 1973.
The author is with the Department of Electrical Engineering, University of Newcastle, New South Wales, Australia.

Kalman–Bucy filter may be used. This approach has been studied by Bryson and Johansen [3].

In this note it is shown that these two approaches do not lead to the same answer. In fact, the limit as $\epsilon \to 0$ for case 1) is not well-defined—there appears to be a discontinuity in the solution at $\epsilon = 0$.

Consider first the case where $R = \epsilon R_0$. The optimal filter is described [1] by the state equations

$$\dot{\hat{x}}(t) = F\hat{x}(t) + K_\epsilon(t)[y(t) - H'\hat{x}(t)]$$

where

$$K_\epsilon(t) = \frac{1}{\epsilon} P_\epsilon(t)HR_0^{-1}$$

and

$$P_\epsilon(t) = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}$$

is the error covariance. The precise form of $P_\epsilon(t)$ is of no importance here (it may be determined by solving a Riccati equation), but we note that

$$P_\epsilon(t_0) = P_0$$

for any $\epsilon > 0$. This implies that

$$K_\epsilon(t_0) = \frac{1}{\epsilon} P_0HR_0^{-1}.$$

If P_0 is nonsingular—as is usually the case—then $\|K_\epsilon(t_0)\|$ will increase without limit as $\epsilon \to 0$. (It will normally also be true that $K_\epsilon(t)$, for any $t > t_0$, contains unbounded entries, but this is more difficult to prove.) The limiting case therefore fails to have a solution, except in the sense that the optimal filter contains infinite feedback gains.

Consider now the case where $R = 0$. The measurement process now contains no white noise component, being given by

$$y(t) = H'x(t).$$

However, the same information is gained by measuring $\dot{y}(t)$, together with $y(t_0)$. Formally,

$$\dot{y}(t) = H_1'x(t) + v_1(t)$$

where $H_1 = F'H$, and $v_1(t)$ is a zero-mean Gaussian process with covariance

$$E[v_1(t)v_1'(\tau)] = H'GQG'H\delta(t - \tau) \triangleq R_1\delta(t - \tau).$$

At least in the case where R_1 is positive definite, a standard Kalman–Bucy filter may be derived [3]. The precise details are not important for our present purposes; it suffices to note that there exists an estimator of the form

$$\dot{\hat{x}}(t) = F\hat{x}(t) + K(t)[\dot{y}(t) - H_1'\hat{x}(t)]$$

with $K(t)$ well-defined and continuous for all $t \geqslant t_0$. The need for the differentiation may be avoided by defining a new filter state

$$z(t) \triangleq \hat{x}(t) - K(t)y(t).$$

The filter equations are then

$$\dot{z}(t) = [F - K(t)H_1']z(t) + [FK(t) - K(t)H_1'K(t) - \dot{K}(t)]y(t)$$

and

$$\hat{x}(t) = z(t) + K(t)y(t).$$

A number of further simplifications are possible. The main point, however, is that the optimal estimate is generated by a filter whose internal gains are all bounded. The filter cannot therefore be equivalent, in any but a restricted sense, to the limiting case of the Kalman–Bucy filter as the measurement noise tends to zero.

The essential difference between the two problems is in the initial state estimate. For the case of no measurement noise, this is given by

$$\hat{x}(t_0) = E[x(t_0)|y(t_0)]$$

and it is easy to show that this is in fact

$$\hat{x}(t_0) = (H')^{\#}y(t_0) \tag{1}$$

where $(\;)^{\#}$ denotes a pseudo-inverse. For the Kalman–Bucy filter, on the other hand, $y(t_0)$ contains no useful information, since it is a point sample from a random process containing a nonsingular white noise component. The best initial estimate is therefore simply the *a priori* expectation of the initial state, which is zero.

On physical grounds, one would expect the estimates $\hat{x}(t)$ from the two filters to approach each other with probability one; it is then reasonable to expect a large initial transient in the low-noise Kalman–Bucy filter, as $\hat{x}(t)$ is in effect adjusted from a poor initial estimate to a more accurate state estimate. This would account for the unbounded filter gains.

It is interesting in this connection to note an apparent discrepancy between the results outlined above and those of [3]. The authors of [3] report a discontinuity in $\hat{x}(t)$ at $t = t_0$, but the above analysis shows no such discontinuity. The difference, again, is in the initial conditions: Bryson and Johansen [3] assume that $y(t_0)$ is unavailable for measurement, so that (1) may not be used. On the other hand, they do derive an expression similar to (1) for $\hat{x}(t_0^+)$, so that the discrepancy is more apparent than real.

References

[1] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *ASME Trans.* (J. Basic Eng.), vol. 83D, pp. 95–108, Mar. 1961.
[2] H. Kwakernaak and R. Sivan, "The maximally achievable accuracy of linear optimal regulators and linear optimal filters," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 79–86, Feb. 1972.
[3] A. E. Bryson and D. E. Johansen, "Linear filtering for time-varying systems using measurements containing colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-10, pp. 4–10, Jan. 1965.

Optimum Measurements for Estimation

K. D. HERRING AND J. L. MELSA

Abstract—**This correspondence considers the class of estimation applications in which a choice of measurement devices is possible. A cost is associated with each measurement device. An optimization procedure is presented which minimizes the measurement cost while maximizing estimation accuracy.**

I. Introduction

In many estimation applications, the measurement or observation model is specified except for certain parameters. The optimal selection of such parameters has been investigated by several authors [1]–[3]. In a special case of this general problem [4], these parameters represent switching functions which, when optimized, determine which single measurement device should be used at each instant of time during the measurement interval to achieve the best estimation results. This correspondence generalizes these results to allow the selection at each instant of time of the best combination of devices, as opposed to the best single device.

Manuscript received March 19, 1973; revised September 4, 1973.
K. D. Herring was with the Information and Control Sciences Center, Southern Methodist University, Dallas, Tex. He is now with the Frank J. Seiler Research Laboratory, U.S. Air Force Academy, Colo. 80840.
J. L. Melsa was with the Information and Control Sciences Center, Southern Methodist University, Dallas, Tex. He is now with the University of Notre Dame, South Bend, Ind. 46556.