Monograph:
Billings, S.A. (1980) Introduction to Kalman Filters. Research Report. ACSE Report 127 . Department of Control Engineering, University of Sheffield, Mappin Street, Sheffield

Introduction to Kalman Filters

S. A. Billings, B.Eng.,Ph.D.,C.Eng.
MIEE,AFIMA,M.Inst.MC

Research Report No.127

October 1980

IEE Colloquium on "Kalman Filters and their Applications",
17th November 1980.

INTRODUCTION TO KALMAN FILTERS

S. A. Billings

Consider a linear, discrete time-invariant multivariable system defined by

$$x(k+1) = \phi x(k) + Bu(k) + Cw(k) \qquad (1)$$

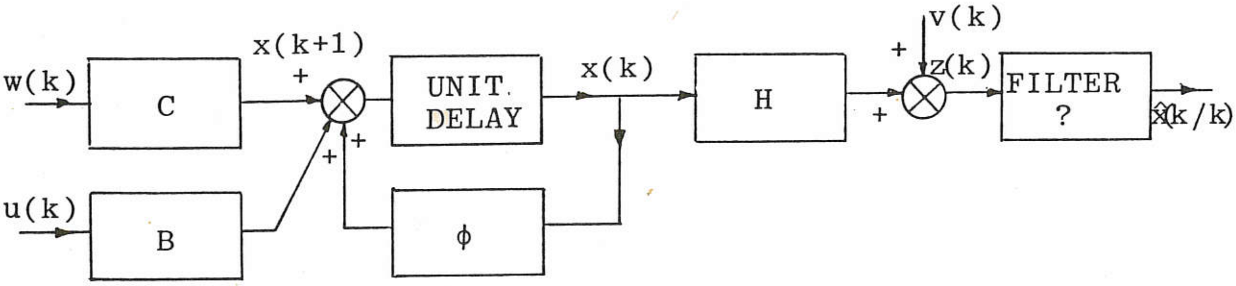
$$z(k+1) = Hx(k+1) + v(k+1) \qquad (2)$$

where $z(\cdot)$ is a vector of outputs or measurements, $u(\cdot)$ is the vector of control inputs and $w(\cdot)$ is a vector of disturbances acting on the system. The measurements are corrupted by noise and round-off errors, represented by the additive vector $v(\cdot)$.

The problem to be solved is to estimate the value of the state vector $x(k+1)$ given all the available data up to the current instant, i.e. $z(k+1), z(k), \ldots . z(0)$ and $u(k), \ldots . u(0)$. This is illustrated schematically in the diagram below. The notation $\hat{x}(k/j)$ is used to mean the estimate of $x(k)$ based on all the information up to and including the time interval j.

If $k>j$, the problem is one of PREDICTION, if $k = j$ one of FILTERING, and if $k<j$ one of SMOOTHING or interpolation.

Only the filtering problem will be treated here and the derivation is accomplished in a manner that relies more upon physical intuition than upon mathematical sophistication.



The variables $w(k)$ and $v(k)$ are assumed to be zero mean, stationary white sequences with the following properties,

$$E\left[w(k)w^T(j)\right] = Q\delta_{jk} \begin{array}{ll} = Q, & j = k \\ = 0, & j \neq k \end{array} \qquad (3)$$

$$E\left[v(k)v^T(j)\right] = R\delta_{jk} \qquad (4)$$

S. A. Billings is with the Department of Control Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD.

where $w(k)$ and $v(k)$ are assumed to be independent

$$E\left[v(k)w^T(j)\right] = 0 \forall k,j \tag{5}$$

Given the preceding model determine an estimate $\hat{x}(k+1/k+1)$ of the system state at time $(k+1)$ that is a linear combination of the previous system state and the measurement data $z(k+1)$. The Kalman-Bucy filter is designed such that the estimate is 'best' in the sense that the expected value of the sum of squares of the error in the estimate is minimum. That is, the $\hat{x}(k+1/k+1)$ is to be chosen so that

$$E\left[(x(k+1)-\hat{x}(k+1/k+1))(x(k+1)-\hat{x}(k+1/k+1))^T\right] = \text{minimum}$$

Assume that the estimate $\hat{x}(k/k)$ is available. What is the best estimate of $x(k+1)$ given the measurements up to the time interval (k)? From estimation theory it can be shown (see Appendix I) that the best estimate is given by the minimum variance estimate or the Conditional Expectation $E\left[x(k+1)/k\right]$.

Thus from equation (1)

$$\hat{x}(k+1/k) = E(x(k+1)/k) = E\left[(\phi x(k)+Bu(k)+Cw(k))/k\right]$$
$$= \phi E\left[x(k)/k\right] + BE\left[u(k)/k\right] + CE\left[w(k)/k\right]$$

But by definition $E\left[x(k)/k\right] = \hat{x}(k/k)$, $E\left[u(k)/k\right] = u(k)$ because $u(k)$ is not a random vector and by hypothesis, the noise $w(k)$ is independent of the state at all times earlier than $(k+1)$ and is also independent of the measurement noise. Thus because $w(k)$ is a white noise sequence

$$E\left[w(k)/k\right] = 0$$

Hence the BEST estimate of $x(k+1)$ based on the estimate at time (k) is

$$\hat{x}(k+1/k) = \phi\hat{x}(k/k) + Bu(k) \tag{6}$$

and similarly

$$\hat{z}(k+1) = H\hat{x}(k+1/k) \tag{7}$$

If a measurement at time $(k+1)$ is now made available the error $\tilde{z}(k+1)$ in the predicted measurement can be defined as

$$\tilde{z}(k+1) = z(k+1) - \hat{z}(k+1) \tag{8}$$

The prediction of equation (6) can be improved by using the information available at time $(k+1)$ and adding a proportion of $\tilde{z}$ to each element of $\hat{x}(k+1/k)$,

$$\hat{x}(k+1/k+1) = \hat{x}(k+1/k) + K\tilde{z}(k+1) \tag{9}$$

The estimate of equation (9) is thus of the 'predictor-corrector' type which tries to drive $\tilde{z}$ to zero.

Substituting equations (7) and (8) into (9) gives

$$\hat{x}(k+1/k+1) = (I-KH)\hat{x}(k+1/k) + Kz(k+1) \qquad (10)$$

$$= (I-KH)(\phi\hat{x}(k/k)+Bu(k)) + Kz(k+1) \qquad (11)$$

Equation (11) is the defining equation of the state estimator which, estimates x(k+1) from only the current measurements z(k+1) and the previous estimate. Thus equation (11) defines a recursive filter. (If w(k) and v(k+1) are zero for all k, equation (11) defines the LUENBERGER OBSERVER and $\hat{x}$ approaches x for all K). The problem now remains to find a suitable value of the filter gain matrix K.

Define the estimation error as

$$\tilde{x}(k+1) = x(k+1) - \hat{x}(k+1/k+1) \qquad (12)$$

Substituting from equations (1), (2) and (11) and manipulating gives

$$\tilde{x}(k+1) = (I-KH)(\phi\tilde{x}(k) + Cw(k)) - Kv(k+1) \qquad (13)$$

The filter gain K must be chosen such that

$$E[(x(k+1)-\hat{x}(k+1/k+1))(x(k+1)-\hat{x}(k+1/k+1))^T]$$

$$= E[\tilde{x}(k+1)\tilde{x}^T(k+1)] = \text{minimum}$$

Define the covariance of the estimation error as

$$P(k+1) = E[\tilde{x}(k+1)\tilde{x}^T(k+1)] \qquad (14)$$

so that "squaring" equation (13), and taking the expectation, noting that the cross-product terms such as $E[x(k)w^T(k)]$, $E[w(k),v^T(k+1)]$, etc are zero gives

$$P(k+1) = (I-KH)\phi P(k)\phi^T(I-KH)^T+(I-KH)CQC^T(I-KH)^T+KRK^T$$

$$= (I-KH)P^*(k+1)(I-KH)^T + KRK^T \qquad (15)$$

where

$$P^*(k+1) = \phi P(k)\phi^T + CQC^T \qquad (16)$$

Note that since $P(\cdot)$ and Q are symmetric, so is $P^*(\cdot)$.

Re-arranging equation (15)

$$P(k+1) = K(HP^*(k+1)H^T+R)K^T-P^*(k+1)H^TK^T-KHP^*(k+1)+P^*(k+1) \qquad (17)$$

Observe that the first three terms of equation (17) have the form of a quadratic matrix polynomial in terms of the unknown filter gain K. Hypothesize the existence of a matrix V such that equation (17) becomes

$$P(k+1) = (K-V)(HP^*(k+1)H^T+R)(K-V)^T-V(HP^*(k+1)H^T+R)V^T+P^*(k+1) \qquad (18)$$

This procedure is the matrix equivalent of completing the square of a quadratic polynomial. Equation (18) is equivalent to equation (17) provided the matrix V satisfies

$$V(HP^*(k+1)H^T+R) = P^*(k+1)H^T \tag{19}$$

For an optimal estimate, K must be chosen so that the estimation error is a minimum. In the minimum variance sense this is achieved when P(k+1) is a minimum (equation 14). Note that the value of K which minimises P(k+1) will not necessarily be the same value that minimises P(k+2) etc. Hence the particular value of K will be denoted K(k+1) etc. By inspection of equation (18), since the only term involving K is quadratic, the minimum P(k+1) is obtained when

$$K = V \tag{20}$$

Hence from (19)

$$K(k+1) = P^*(k+1)H^T(HP^*(k+1)H^T+R)^{-1} \tag{21}$$

Substituting K = V in equation (18) gives the optimal solution as

$$P(k+1) = P^*(k+1)-K(HP^*(k+1)H^T+R)K^T \tag{22}$$

Substituting equation (21) in (22) and using the fact that $P^*(\cdot)$ is symmetric gives

$$P(k+1) = P^*(k+1)(I-H^TK^T) = (I-KH)P^*(k+1) \tag{23}$$

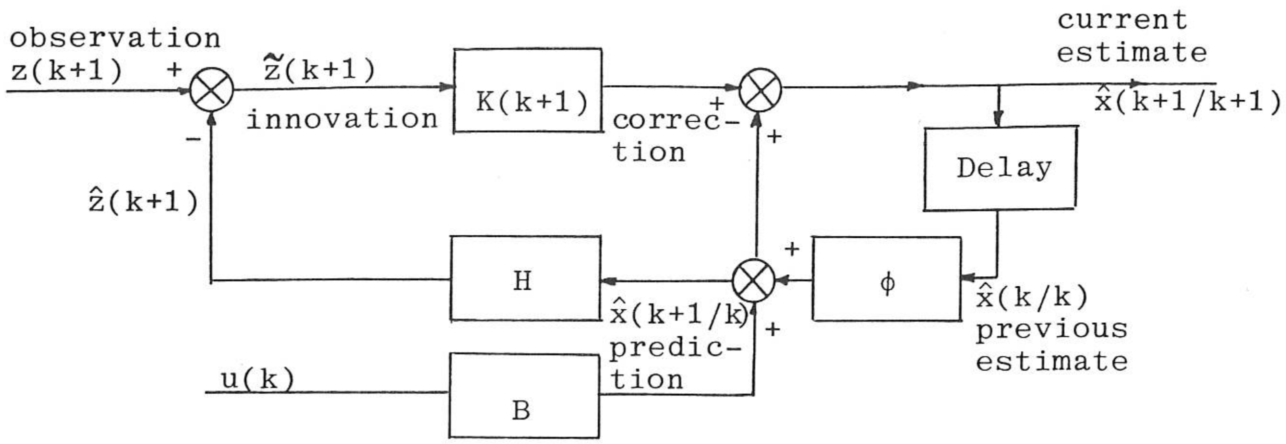
Hence to summarise the KALMAN-BUCY filter is given by

$$P^*(k) = \phi P(k-1)\phi^T+CQC^T \tag{24}$$
$$K(k) = P^*(k)H^T(HP^*(k)H^T+R)^{-1} \tag{25}$$
$$\hat{x}(k/k) = (I-K(k)H)(\phi\hat{x}(k-1/k-1)+Bu(k-1))+K(k)z(k) \tag{26}$$
$$P(k) = (I-K(k)H)P^*(k) \tag{27}$$

Equations (24) through (27) indicate that the optimal estimator is time varying, even for the autonomous process defined by equations (1) and (2). A schematic diagram of the Kalman-Bucy filter is illustrated below. Notice that the equations for K(k), P(k) and P*(k) are independent of the observation sequence and can be precomputed if desired.

In the case where system and measurement dynamics are linear constant coefficient equations and the noise statistics are stationary (as in the case above), the filtering process may reach a steady-state where the covariance and gain matrix are constant. Complete observability is a sufficient condition for the existence of a steady-state solution. Complete controllability will assure that the steady-state solution is unique.

By applying vector-matrix manipulations the Kalman filter equations can be put into a number of equivalent forms. With appropriate values of ϕ, B, C, H, Q and R the equations are still valid for time-varying processes.



Block diagram of the discrete filter

Continuous-Time Formulation

$$\dot{x} = Fx(t) + Du(t) + Gw(t) \qquad (28)$$

$$z(t) = Hx(t) + v(t) \qquad (29)$$

If the unit interval for the discrete case solution is allowed to approach zero, we obtain the solution for the continuous case; viz

$$\dot{\hat{x}} = F\hat{x} + K(t)[z(t)-H\hat{x}] + Du(t) \qquad (30)$$

$$K(t) = P(t/t)H^TR^{-1} \qquad (31)$$

$$\dot{P} = FP + PF^T - PH^TR^{-1}HP+GQG^T \quad \text{RICCATI EQUATION} \qquad (32)$$

The limiting process is a matter of some delicacy because of the presence of Dirac delta functions. However, the procedure can be made mathematically legitimate by a sufficiently sophisticated analysis.

The Extended Kalman Filter

The Kalman filter can be used to estimate the state of a non-linear system by linearising the system equations around a nominal solution. Thus if

$$\dot{x}_i = f_i(x,u,w,t) \qquad (33)$$

and x^o is a nominal trajectory, expanding eqn (33) in a Taylor series about the nominal value neglecting all except the first order terms yields

$$\delta x = x-x^o \quad ; \quad \delta u = u-u^o$$

$$\delta\dot{x} = \Lambda\delta x + \Delta\delta u + \Omega w(t) \qquad (34)$$

where

$$\Lambda_{ij} = \frac{\delta f_i}{\delta x_j} \quad ; \quad \Delta = \frac{\delta f_i}{\delta u_j} \quad ; \quad \Omega_{ij} = \frac{\delta f_i}{\delta w_j}\Bigg|_{\substack{x=x^o \\ u=u^o}} \qquad (35)$$

Equation (34) can be converted to discrete-time form in the usual manner and the estimate $\delta\hat{x}$ can be generated by a Kalman filter using this discrete form where $\hat{x} = x^o + \delta\hat{x}$. When the nominal trajectory is defined as the best previous estimate $x^o(k)=\hat{x}(k/k)$ the resulting estimator is called an Extended Kalman Filter. Note that the gain K, the linearisation and discretisation must be recomputed at each step.

References

1. JAZWINSKI, A.H.: "Stochastic Processes and Filtering Theory", Academic Press, 1970.

2. MEDITCH, J.S.: "Stochastic Optimal Linear Estimation and Control", McGraw-Hill, 1969.

3. Advances in Control Systems Vol. 3, Edited by C.T.Leondes, (Kalman Filtering Techniques; H.W.Sorenson).

4. NAHI, N.E.: "Estimation Theory and Applications", John Wiley, 1969.

Appendix 1

Minimum Variance Estimates

Suppose that a vector random variable x is to be estimated from the given observations $z_1, z_2 \ldots z_q$. Only the case where x and z are scalars will be considered to simplify the derivation. Assume that the conditional probability density function P(X/Z) is known and consider the variance of an estimate $\hat{x}$ of x determined as a function of the observations (i.e. $\hat{x} = \hat{x}(z)$).

For a given set of observations we can define

$$\mathrm{var}\{\hat{x}(z)\} = \int_{-\infty}^{\infty} (\hat{x}(z)-X)^2 P(X/Z)dX \tag{36}$$

where X is the true value of $\hat{x}$. Differentiating w.r.t. $\hat{x}$ yields

$$d/d\hat{x}\{\mathrm{var}\{\hat{x}(z)\}\} = 2\int_{-\infty}^{\infty} (\hat{x}(z)-X)P(X/Z)dX \tag{37}$$

and at the extremum which can be shown to be a minimum

$$\hat{x}(z)\int_{-\infty}^{\infty} P(X/Z)dX = \int_{-\infty}^{\infty} XP(X/Z)dX$$

But $\int_{-\infty}^{\infty} P(X/Z)dX = 1$ and the integral on the r.h.s. is the expected value

$$\therefore \quad \hat{x}(z) = E[X/Z] \tag{38}$$

That is the conditional expectation of the unknown random variable x given the observation z is a minimum variance estimate.

It can readily be shown that if the cost to be minimised is quadratic in the estimation error, the optimal estimate is E[X/Z] regardless of the form of P(X/Z). If P(X/Z) is Gaussian a linear