A note on Kalman Filter

C. R. Rao[†]

Statistics Department, Pennsylvania State University, University Park, PA 16802

The Kalman Filter commonly employed by control engineers and other physical scientists has been successfully used in such diverse areas as the processing of signals in aerospace tracking and underwater sonar, and statistical quality control. More recently, it has been used in some nonengineering applications such as short-term forecasting, time series, survival analysis, and so on. In all of these situations, we have a set of equations governing the true state of a system and another set connecting the observations made at any given time on the system with its true state. The problem is that of predicting the true state of the system at any given time point based on available observations. The solution proposed in the vast literature of the subject depends on the assumptions made on the initial state of the system. In this paper, a method that is independent of the initial state is proposed. This is useful when the *a priori* information on the initial state is not available. The method is also applicable when some observations are missing.

1. Dynamical and Observational Equation

There is extensive literature on Kalman Filtering (KF) and its applications arising out of the seminal contributions by Kalman (1) and Kalman and Bucy (2).

The general problem can be stated as follows. We have a time sequence of p and q-vector valued random variables $\{x(t), y(t)\}$, $t = 1, 2, \ldots$ related through structural equations

$$x(t) = Fx(t - 1) + \xi(t) \qquad [1.1]$$

$$y(t) = Hx(t) + \eta(t), \qquad [1.2]$$

where F and H are $p \times p$ and $q \times p$ matrices, respectively, and the following hold, where E stands for expectation and C for covariance.

$$E[\xi(t)] = 0, \ E[\eta(t)] = 0, \ C[\xi(t), \eta(s)] = 0$$

$$C[x(s), \xi(t)] = 0, \ t > s, \ C[x(s), \eta(t)] = 0, \ t \geq s$$

$$C[\xi(t), \xi(t)] = V, \ C(\eta(t), \eta(t)) = W$$

$$C[\xi(t), \xi(s)] = 0, \ C[\eta(t), \eta(s)] = 0, \ t \neq s. \qquad [1.3]$$

We have observations only on $y(1), \ldots, y(t)$ up to time t, and the problem is to estimate $x(s)$ given $y(1), \ldots, y(t)$, or a subset of them.

In the engineering literature, estimation of $x(t)$ is called smoothing if $s < t$, filtering if $s = t$, and prediction if $s > t$. An excellent review of KF and its applications to a wide variety of statistical problems can be found in refs. 1–5.

In the literature on KF, the conditions are usually stated in terms of independence of the variables involved instead of zero correlations as stated in **1.3**. If we are interested in linear estimates of observations, we need only require the knowledge of the variances and covariances of the variables involved.

The theory described in the paper can be extended to the case where the matrices F and H in **1.1** and **1.2** and V and W in **1.3** depend on time.

2. Reduction to Linear Equations

From **1.1**,

$$x(t) = Fx(t - 1) + \xi(t) = F^t x(0) + e_t \qquad [2.1]$$

with $e_t = F_t \xi_t$, where

$$F_t = (F^{t-1} : F^{t-2} : \ldots : I_p), \ \xi_t' = (\xi'(1), \ldots, \xi'(t)),$$

and I_p is the identity matrix of order p. A simple calculation yields

$$C[e_t, e_t] = V_{tt} = F_t(I_p \otimes V)F_t'$$

$$C[e_t, e_s] = V_{ts} = F_t \begin{pmatrix} I_s & \otimes & V \\ & 0 & \end{pmatrix} F_s', \quad \text{if } t \geq s$$

$$= F_t(I_t \otimes V{:}0)F_s', \qquad \text{if } t \leq s, \qquad [2.2]$$

where $I_s \otimes V$ is the Kronecker product, and 0 is a null matrix of appropriate order.

From **1.2**,

$$y(t) = Hx(t) + \eta(t) = HF^t x(0) + \delta(t) \qquad [2.3]$$

where $\delta(t) = He_t + \eta(t)$. A simple calculation gives

$$C[\delta(t), \delta(t)] = W_{tt} = HV_{tt}H' + W,$$

$$C[\delta(t), \delta(s)] = W_{ts} = HV_{ts}H',$$

$$C[e_t, \delta(s)] = w_{ts} = V_{ts}H'. \qquad [2.4]$$

Writing $Y'(t) = (y'(1), \ldots, y'(t))$ and

$$G_t = \begin{pmatrix} HF^1 \\ \vdots \\ HF^t \end{pmatrix}, \ \delta_t = \begin{pmatrix} \delta(1) \\ \vdots \\ \delta(t) \end{pmatrix},$$

we have the composite linear model for observations up to time t

$$Y(t) = G_t x(0) + \delta_t. \qquad [2.5]$$

To estimate $x(s)$ given $Y(t)$ for any s and t, we need to consider the equations

$$x(s) = F^s x(0) + e_s$$

$$Y(t) = G_t x(0) + \delta_t \qquad [2.6]$$

with the variance-covariance matrix of (e_s, δ_t),

$$\begin{pmatrix} V_{ss} & w_{s1} & \ldots & w_{st} \\ w_{1s} & W_{11} & \ldots & W_{1t} \\ \vdots & \cdot & \ldots & \cdot \\ w_{ts} & W_{t1} & \ldots & W_{tt} \end{pmatrix} = \begin{pmatrix} V_{ss} & u' \\ u & U \end{pmatrix} \text{ say,} \qquad [2.7]$$

where V_{ss}, w_{rs}, and W_{ij} are as defined in **2.2** and **2.4**, and $u' = (w_{s1}, \ldots, w_{st})$ and $U = (W_{ij})$.

STATISTICS

3. Different Assumptions on $x(0)$

3.1. Given Prior Distribution of $x(0)$.
Let $x(0)$ have a prior mean μ and variance-covariance matrix P. Then we can write the model **2.6** as

$$x(s) - F^s\mu = a(s)$$

$$Y(t) - G_t\mu = b(t) \qquad \textbf{[3.1]}$$

with

$$C[a(s), a(s)] = F^sPF^{s'} + V_{ss} = R$$

$$C[a(s), b(t)] = F^sPG_t' + u' = S$$

$$C[b(t), b(t)] = G_tPG_t' + U = T. \qquad \textbf{[3.2]}$$

Then, using the least squares theory (ref. 6, pp. 544–547), the estimate of $x(s)$ is

$$\hat{x}(s) = F^s\mu + ST^{-1}(Y(t) - G_t\mu) \qquad \textbf{[3.3]}$$

with the minimum dispersion error matrix

$$C[x(s) - \hat{x}(s), x(s) - \hat{x}(s)] = R - ST^{-1}S'. \qquad \textbf{[3.4]}$$

3.2. $x(0)$ Is a Known Constant μ.
The solution in this case is the same as **3.4** with P in **3.2** replaced by the null matrix. Then the estimate of $x(s)$ is

$$\hat{x}(s) = F^s\mu + u'U^{-1}(Y(t) - G_t\mu) \qquad \textbf{[3.5]}$$

with the minimum error dispersion matrix

$$C[x(s) - \hat{x}(s), x(s) - \hat{x}(s)] = V_{ss}u'U^{-1}u.$$

If $x(0)$ is a constant, but its value is not known, then we may estimate $x(0)$ from the second equation in **2.6** using least squares theory provided the rank of G_t is p. In such a case, the least squares estimate of $x(0)$ is

$$\hat{x}(0) = (G_t'U^{-1}G_t)^{-1}G_t'U^{-1}Y(t). \qquad \textbf{[3.6]}$$

Substituting $\hat{x}(0)$ for μ in **3.5**, we have the empirical estimate

$$\hat{x}(s) = (F^s - u'U^{-1}G_t)\hat{x}(0) + u'U^{-1}Y(t). \qquad \textbf{[3.7]}$$

The estimate **3.7** and its dispersion matrix are derived in a more natural way in the next section.

3.3. No Information Is Available on $x(0)$.
Consider a linear function $L'Y(t)$ as an estimate of $x(s)$ such that

$$E[x(s) - L'Y(t)] = 0 \text{ independently of } x(0)$$

$$\Rightarrow F^s - L'G_t = 0. \qquad \textbf{[3.8]}$$

The variance-covariance matrix of $x(s) - L'Y(t)$ is

$$V_{ss} - L'u - u'L + L'UL, \qquad \textbf{[3.9]}$$

where V_{ss}, u, and U are as defined in **2.7**.

The minimum value of **3.9** subject to **3.8** is attained at

$$L_* = U^{-1}(u + G_tC), \quad C = (G_t'U^{-1}G_t)^{-1}(F^{s'} - G_t'U^{-1}u), \qquad \textbf{[3.10]}$$

giving the same expression for the estimate of $x(s)$ as the empirical estimate derived in **3.7**. The minimum dispersion error matrix is

$$(V_{ss} - uU^{-1}u') + C'G_t'U^{-1}G_tC. \qquad \textbf{[3.11]}$$

The second expression in **3.11** is the extra error due to the imposed condition **3.8**. We call the estimate $L_*'Y(t)$ the constrained linear predictor (CLP) of $x(s)$.

We prove the following result, which will be used in Section 4. Let M be a matrix such that $E(M'Y(t)) = 0$ independently of $x(0)$, which implies that $M'G_t = 0$. Then, with L_* as defined in **3.10**,

$$C[x(s) - L_*'Y(t), M'Y(t)]$$

$$= u'M - L_*'UM = (u' - L_*'U)M$$

$$= u' - (u + G_tC)'M = -C'G_t'M = 0. \qquad \textbf{[3.12]}$$

4. Recursive Estimation

In the application of KF, it is customary to estimate $x(s)$ using the available observations $(y'(1), \ldots, y'(r)) = Y_1$ (say), up to a time point r, and update the estimate when additional observations $(y'(r + 1), \ldots, y'(t)) = Y_2$ (say) are acquired. The methodology of such recursive estimation is well known in KF theory when unconstrained estimates (such as those discussed in Sections 3.1 and 3.2) are used.

We develop a similar method when CLPs as derived in Section 3.3 are required.

Let $L_{1*}'Y_1$ be the CLP of $x(s)$ based on Y_1, and $L_{2*}'Y_1$ be the CLP of Y_2 based on Y_1. We consider a general linear function of Y_1, Y_2

$$M_1'Y_1 + M_2'(Y_2 - L_{2*}'Y_1) \qquad \textbf{[4.1]}$$

and its deviation from $x(s)$ as

$$(x(s) - L_{1*}'Y_1) - (M_1 - L_{1*})'Y_1 - M_2'(Y_2 - L_{2*}'Y_1)$$

$$= A - B - M_2'D \text{ say.} \qquad \textbf{[4.2]}$$

The condition of unbiasedness of **4.1** independently of $x(0)$ implies $E(B) = 0$, since $E(A)$ and $E(C) = 0$. Then by the Condition **3.12**,

$$C(A, B) = 0, \quad C(D, B) = 0. \qquad \textbf{[4.3]}$$

Hence

$$C(A - B - M_2'D, A - B - M_2'D)$$

$$= C(A - M_2'D, A - M_2'D) + C(B, B). \qquad \textbf{[4.4]}$$

The second term in **4.4** vanishes if we choose $M_1 = L_{1*}$. The best choice of M_2 which minimizes the first term is

$$M_{2*} = [C(D, D)]^{-1}C(D, A). \qquad \textbf{[4.5]}$$

To derive explicit expressions for A, D and their covariance, we write the basic linear model **2.6** by splitting $Y(t)$ into the two components Y_1 and Y_2.

$$x(s) = F^sx(0) + e_s$$

$$Y_1 = G_{(1)}x(0) + \delta_{(1)}$$

$$Y_2 = G_{(2)}x(0) + \delta_{(2)}, \qquad \textbf{[4.6]}$$

where $G_{(1)}$, $G_{(2)}$ are partitions of G_t and $\delta_{(1)}$, $\delta_{(2)}$ are partitions of δ_t. The corresponding covariance matrix of $x(s)$, Y_1, Y_2 is

$$\begin{pmatrix} V_{ss} & u_1' & u_2' \\ u_1 & U_{11} & U_{12} \\ u_2 & U_{21} & U_{22} \end{pmatrix}, \qquad \textbf{[4.7]}$$

obtained by partitioning the matrices u and U in **2.7**.

Using the formula **3.10**, the CLP of $x(s)$ based on Y_1 is

$$L'_{1*}Y_1 = (u_1 + G_1C_1)'U_{11}^{-1}Y_1,$$

$$C_1 = (G'_1U_{11}^{-1}G_1)^{-1}(F^s - u'_1U_{11}^{-1}G_1)'. \qquad \textbf{[4.8]}$$

The CLP of Y_2 based on Y_1 is

$$L'_{2*}Y_1 = (U_{21} + G_1C_2)'U_{11}^{-1}Y_1,$$

$$C_2 = (G'_1U_{11}^{-1}G_1)^{-1}(G_2 - U_{21}U_{11}^{-1}G_1)'. \qquad \textbf{[4.9]}$$

Then

$$A = x(s) - L'_{1*}Y_1 = e_s - L'_{1*}\delta_{(1)}$$

$$D = Y_2 - L'_{2*}Y_1 = \delta_{(2)} - L'_{2*}\delta_{(1)}$$

$$C(A, A) = V_{ss} + L'_{1*}U_{11}L_{1*} - u'_1L_{1*} - L'_{1*}u_1$$

$$C(D, D) = U_{22} + L'_{2*}U_{11}L_{2*} - U_{21}L_{2*} - L'_{2*}U_{12}$$

$$C(A, D) = u'_2 - L'_{1*}U_{12} - u'_1L_{2*} + L'_{1*}U_{11}L_{2*}. \qquad \textbf{[4.10]}$$

The updated CLP of $x(s)$ is

$$L'_{1*}Y_1 + C(A, D)[C(D, D)]^{-1}(Y_2 - L'_{2*}Y_1) \qquad \textbf{[4.11]}$$

with the minimum dispersion error

$$C(A, A) - C(A, D)[C(D, D)]^{-1}C(D, A). \qquad \textbf{[4.12]}$$

Note 1: The second equation in **2.6** is in the form of a mixed linear model, which enables us to estimate $x(0)$, and also the matrices V and W if they are unknown by using MINQE method described in ref. 7.

Note 2: The linear model **2.6** is used when we want to estimate $x(s)$ based on the data $y(1), \ldots, y(t)$. A more general problem is the estimation of $x(s)$ based on $y(r_1), \ldots, y(r_n)$, observations made at n time points $r_1, \ldots, r_n$. A particular situation is when observations on y are missing at certain time points, or some observations on y are discarded as not reliable or outliers. The linear equations appropriate for this is

$$x(s) = F^s x(0) + e_s$$

$$Y = Gx(0) + \delta \qquad \textbf{[4.13]}$$

where

$$Y' = (y'(r_1), \ldots, y'(r_n)), \ \delta' = (\delta'(r_1), \ldots, \delta'(r_n))$$

$$G' = (F^{r_1}H' : \ldots : F^{r_n}H')$$

$$C[\delta(r_i), \delta(r_j)] = HV_{r_ir_j}H'$$

$$C[\delta(r_i), \delta(r_i)] = HV_{r_ir_i}H' + W$$

$$C[e_s, \delta(r_i)] = V_{sr_i}H',$$

where V_{ij} are as defined in **2.2**. Then the general methods of Sections 3 and 4 apply.

1. Kalman, R. E. (1960) *J. Basic Eng.* **82,** 34–45.
2. Kalman, R. E. & Bucy, R. S. (1961) *J. Basic Eng.* **83,** 95–108.
3. Meinhold, R. J. & Singpurwalla, N. D. (1983) *Am. Stat.* **37,** 123–127.
4. Meinhold, R. J. & Singpurwalla, N. D. (1987) *Am. Stat.* **41,** 101–106.
5. Rao, C. R. & Toutenburg, H. (1999) *Linear Models: Least Squares and Alternatives* (Springer, New York), 2nd Ed.
6. Rao, C. R. (1973) *Linear Statistical Inference and Its Applications* (Wiley, New York).
7. Rao, C. R. & Kleffe, J. (1988) *Estimation of Variance Components and Its Applications* (North–Holland, Amsterdam).

STATISTICS