CHAPTER V - FILTERING THEORY

JOSEPH G. CONLON

1. Introduction

Suppose we wish to predict the value of a random variable $X$ whose probability distribution we know. We might wish to choose the predicted value $u$ so that

\begin{equation}
\text{mean square} = E\{ (X - u)^2 \} \text{ is minimum.}
\end{equation}

We can think here of $u$ as being a control and the mean square (1.1) as being the cost function. It is easy to solve the minimization problem (1.1) exactly since

\begin{equation}
E\{ (X - u)^2 \} = E\{ (X - E[X])^2 \} + (E[X] - u)^2.
\end{equation}

Hence the optimal mean square predictor $u$ is the mean $u = E[X]$. Of course if we choose a different cost function we get a different value for the optimal $u$.

We can generalize the above considerations as follows: Suppose we have made observations of random variables $X_1, X_2, \ldots, X_n$, and wish to predict from these observations the most likely value of a random variable $X$. We are assuming here that the joint pdf of $(X_1, \ldots, X_n, X)$ is known. Suppose this predicted value is a function $g(X_1, \ldots, X_n)$ of the observed values of $X_1, \ldots, X_n$. We choose the function $g$ to minimize

\begin{equation}
\text{mean square} = E\{ (X - g(X_1, \ldots, X_n))^2 \}.
\end{equation}

As in (1.2) we have that

\begin{equation}
E\{ (X - g(X_1, \ldots, X_n))^2 \mid X_1, \ldots, X_n \} =
E\{ (X - E[X|X_1, \ldots, X_n])^2 \mid X_1, \ldots, X_n \} + \{E[X|X_1, \ldots, X_n] - g(X_1, \ldots, X_n)\}^2.
\end{equation}

Now using the fact that

\begin{equation}
E\{ (X - g(X_1, \ldots, X_n))^2 \mid X_1, \ldots, X_n \} = E[ E[ (X - g(X_1, \ldots, X_n))^2 \mid X_1, \ldots, X_n ]] \text{,}
\end{equation}

we conclude that

\begin{equation}
g(X_1, \ldots, X_n) = E[X|X_1, \ldots, X_n].
\end{equation}

The problem with the predictor (1.6) is that it may be difficult to compute. Instead of (1.6) we could make an ansatz for the function $g(\cdot)$, say that $g(\cdot)$ is linear in the variables $X_1, \ldots, X_n$. Thus we set $g(\cdot)$ to be given by the formula

\begin{equation}
g(X_1, \ldots, X_n) = u_0 + u_1 X_1 + \cdots + u_n X_n,
\end{equation}

where the parameters $u_0, \ldots, u_n$ are to be determined by minimizing a cost function. If we use the mean square cost function then we need to solve the minimization problem

\begin{equation}
\text{minimize } E\{ (X - u_0 - u_1 X_1 - \cdots - u_n X_n)^2 \} \text{ as a function of } u_0, \ldots, u_n.
\end{equation}
Observe that the problem (1.8) is reminiscent of the LQ problems we encountered in classical control theory. Evidently the solution to (1.8) is obtained by setting the first partials in $u_0, \ldots, u_n$ to 0, which gives a linear system of equations,

\begin{align}
(1.9) \
 u_0 + \sum_{j=1}^n u_j E[X_j] &= E[X], \\
 u_0 E[X_i] + \sum_{j=1}^n u_j E[X_i X_j] &= E[X_i X], \quad 1 \leq i \leq n.
\end{align}

If we solve (1.9) for $u_0, \ldots, u_n$, then (1.7) gives the best approximation to $X$ in a least squares sense by a linear function of $X_1, \ldots, X_n$.

We can explicitly solve (1.9) in a simple but significant case, where the variables $X_1, \ldots, X_n$ are identically distributed with positive correlation. Thus suppose that there is a random variable $\Theta$, and $X, X_1, \ldots, X_n$ conditioned on $\Theta$ are i.i.d. with mean and variance given by the formulas

\begin{align}
(1.10) \
 E[X|\Theta] &= \mu(\Theta), \quad \text{var}[X|\Theta] = \sigma^2(\Theta).
\end{align}

Then we have that

\begin{align}
(1.11) \
 E[X_i X_j] &= E[E[X_i X_j|\Theta]] = E[E[X_i|\Theta]E[X_j|\Theta]] = E[\mu(\Theta)^2] \quad \text{if } i \neq j, \\
 E[X_i^2] &= E[E[X_i^2|\Theta]] = E[\sigma^2(\Theta) + \mu(\Theta)^2] = E[\sigma^2(\Theta)] + E[\mu(\Theta)^2].
\end{align}

Hence (1.9) becomes

\begin{align}
(1.12) \
 u_0 + \sum_{j=1}^n u_j E[\mu(\Theta)] &= E[\mu(\Theta)], \\
 u_0 E[\mu(\Theta)] + \sum_{j=1}^n u_j E[\mu(\Theta)^2] + u_i E[\sigma^2(\Theta)] &= E[\mu(\Theta)^2], \quad 1 \leq i \leq n.
\end{align}

By symmetry all the $u_j$ for $j = 1, \ldots, n$, are equal, so (1.12) becomes

\begin{align}
(1.13) \
 u_0 + nu_1 E[\mu(\Theta)] &= E[\mu(\Theta)], \\
 u_0 E[\mu(\Theta)] + nu_1 E[\mu(\Theta)^2] + u_1 E[\sigma^2(\Theta)] &= E[\mu(\Theta)^2].
\end{align}

We conclude that $u_0, u_1$ are given by the formulas

\begin{align}
(1.14) \
 u_0 &= \frac{E[\mu(\Theta)]E[\sigma^2(\Theta)]}{n \text{var}[\mu(\Theta)] + E[\sigma^2(\Theta)]}, \quad u_1 = \frac{\text{var}[\mu(\Theta)]}{n \text{var}[\mu(\Theta)] + E[\sigma^2(\Theta)]}.
\end{align}

Observe that when the $X, X_1, \ldots, X_n$ are i.i.d. then $\text{var}[\mu(\Theta)] = 0$, whence the predictor (1.7) is just $E[\mu(\Theta)] = E[X]$.

There are some cases where the exact solution of the least squares problem (1.3) is linear as in (1.7), so

\begin{align}
(1.15) \
 E[X|X_1, \ldots, X_n] &= u_0 + u_1 X_1 + \cdots + u_n X_n,
\end{align}

for suitable values of $u_0, \ldots, u_n$. As an example we consider the situation just mentioned, and take $\Theta$ to be a beta random variable. Thus $0 < \Theta < 1$ with probability 1 and the pdf of $\Theta$ is

\begin{align}
(1.16) \
 \text{pdf of beta variable} &= \theta^{a-1}(1-\theta)^{b-1} / \text{normalization}.
\end{align}
The parameters of the beta variable are \( a, b > 0 \), and the normalization is given by the beta function

\[
\beta(a, b) = \int_0^1 \theta^{a-1} (1 - \theta)^{b-1} \, d\theta.
\]

There is a well known relationship between the beta and Gamma function,

\[
\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} \, dx, \quad k > 0.
\]

The \( \Gamma \) function satisfies the recurrence relation \( \Gamma(k+1) = k\Gamma(k) \), and hence \( \Gamma(k) = (k-1)! \) for integer \( k \geq 1 \). It is easy to see from the normalization constant \( \sqrt{2\pi} \) for the standard normal variable that \( \Gamma(1/2) = \sqrt{\pi} \), and so one can obtain a formula for \( \Gamma(k) \) if \( k > 0 \) is a non-negative integer plus 1/2. The relation between the \( \beta \) and \( \Gamma \) functions is given by the formula

\[
\beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b), \quad a, b > 0.
\]

Next we let the variable \( X \) be such that \( X \) conditioned on the \( \beta \) variable \( \Theta \) is binomial with parameters \( (N, \Theta) \). Thus

\[
P(X = m | \Theta) = \binom{N}{m} \Theta^m (1 - \Theta)^{N-m} \quad \text{for} \quad 0 \leq m \leq N.
\]

We have now that

\[
E[X|X_1,\ldots,X_n] = E[E[X|X_1,\ldots,X_n,\Theta]|X_1,\ldots,X_n] = NE[\Theta|X_1,\ldots,X_n].
\]

Note that in (1.21) we have used the fact that the variables \( X, X_1, \ldots, X_n \) conditioned on \( \Theta \) are independent. We show that the variable \( \Theta \) conditioned on \( X_1,\ldots,X_n \) is also a \( \beta \) variable with parameters depending on \( X_1,\ldots,X_n \). To see this observe that for any functions \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) and \( g : \mathbb{R} \rightarrow \mathbb{R} \),

\[
E[f(X_1,\ldots,X_n)g(\Theta)] = E[E[f(X_1,\ldots,X_n)g(\Theta)|X_1,\ldots,X_n]].
\]

We also have that

\[
E[f(X_1,\ldots,X_n)g(\Theta)] = E[g(\Theta)E[f(X_1,\ldots,X_n)|\Theta]].
\]

\[
E[f(x_1,\ldots,x_n)\left(\binom{N}{x_1} \cdots \binom{N}{x_n}\right) \Theta^{x_1}(1-\Theta)^{N-x_1} \cdots \Theta^{x_n}(1-\Theta)^{N-x_n}] =
\]

\[
\sum_{x_1,\ldots,x_n} f(x_1,\ldots,x_n) \left(\binom{N}{x_1} \cdots \binom{N}{x_n}\right) \int_0^1 d\theta \theta^{a-1} (1-\theta)^{b-1} \theta^{x_1}(1-\theta)^{N-x_1} \cdots \theta^{x_n}(1-\theta)^{N-x_n}/\beta(a,b).
\]

We conclude from (1.22), (1.23) that

\[
E[g(\Theta)|X_1,\ldots,X_n] = \int_0^1 g(\theta) \rho(\theta) \, d\theta,
\]

where \( \rho(\cdot) \) is the probability density for the \( \beta \) variable with parameters

\[
a + \sum_{j=1}^n X_j, \quad b + \sum_{j=1}^n (N - X_j).
\]

For a \( \beta \) variable \( Y \) with parameters \( a, b \), it is easy to see that \( E[Y] = a/(a+b) \).

We conclude then from (1.21), (1.25), that

\[
E[X|X_1,\ldots,X_n] = \frac{N[a + \sum_{j=1}^n X_j]}{a + b + nN}.
\]
The beta and binomial variables are said to be *conjugate* because of the above property. We can find other sets of conjugate variables by taking limits of the beta and binomial variables. One such limit yields the case where both $\Theta$ and $X$ are Gaussian.

2. The Kalman Filter

We consider a system which evolves as

$$
(2.1) \quad x(i+1) = Ax(i) + Bu(i) + \xi_1(i).
$$

In (2.1) the vector $x(i) \in \mathbb{R}^n$ is an $n$ dimensional vector, and $u(i) \in \mathbb{R}^m$ is $m$ dimensional, with the matrix $A$ being $n \times n$ and $B$ being $n \times m$. The variables $\xi_1(i) \in \mathbb{R}^n, i = 1, 2, \ldots$, are independent Gaussian variables in $\mathbb{R}^n$ with zero mean and $n \times n$ symmetric positive definite covariance matrix $\Sigma_1$. Thus if $\xi_1(i) = [\xi_1^{(1)}(i), \ldots, \xi_1^{(n)}(i)]$, then

$$
(2.2) \quad E[\xi_1^{(1)}(r)\xi_1^{(1)}(s)] = \Sigma_1(r, s).
$$

Let us assume that $x(i)$ is Gaussian with mean $m(i)$ and covariance matrix $\Sigma(i)$. Then since $\xi_1(i)$ is independent of $x(i)$, it follows from (2.1) that $x(i+1)$ is also Gaussian-with mean $m(i+1)$ and covariance matrix $\Sigma(i+1)$ to be determined. Evidently (2.1) implies that

$$
(2.3) \quad m(i+1) = A m(i) + B u(i).
$$

To find the covariance matrix $\Sigma(i+1)$, let $\eta = [\eta_1, \ldots, \eta_n]$ be a Gaussian variable with zero mean and covariance matrix $\Sigma(i)$. If $A = [a_{r,k}], 1 \leq r, k \leq n$, then

$$
(2.4) \quad \Sigma(i+1)(r,s) = E \left[ \sum_{1 \leq k, k' \leq n} a_{r,k} \eta_k a_{s,k'} \eta_{k'} \right] + \Sigma_1(r,s)
$$

$$
= \sum_{1 \leq k, k' \leq n} a_{r,k} \Sigma(i)(k,k') a_{s,k'} + \Sigma_1(r,s) = (A \Sigma(i) A^*)_{r,s} + \Sigma_1(r,s),
$$

where $A^*$ is the adjoint of $A$. We conclude from (2.4) that

$$
(2.5) \quad \Sigma(i+1) = A \Sigma(i) A^* + \Sigma_1.
$$

Next suppose we make observations on the variables $x(i)$. These observations yield a $p$ dimensional vector $y(i)$, and are given by the formula

$$
(2.6) \quad y(i) = C x(i) + \xi_2(i),
$$

where $C$ is an $p \times n$ matrix. The random variable $\xi_2(i) \in \mathbb{R}^p$ measures the uncertainty in the observation and is assumed to be Gaussian with zero mean and covariance matrix $\Sigma_2$. The variables $\xi_2(i), i = 1, 2, \ldots$, are also assumed to be i.i.d and independent of the variables $\xi_1(i), i = 1, 2, \ldots$, which measure the uncertainty in the dynamics (2.1).

We consider the problem of finding the variables $x(i)$ conditioned on all observations $y(j)$ for $j \leq i$. Let us assume this variable is Gaussian with mean $m(i)$ and covariance matrix $\Sigma(i)$. We put now

$$
(2.7) \quad \zeta = x(i) \bigg| y(j) \text{ for } j \leq i, \quad \eta = x(i+1) \bigg| y(j) \text{ for } j \leq i.
$$
From (2.1) we have that
\begin{equation}
\eta = A\zeta + B\mathbf{u}(i) + \xi_1(i) \,.
\end{equation}

From (2.3), (2.5) it follows that \( \eta \) is a Gaussian variable with mean and covariance given by
\begin{align}
\text{mean} = & \begin{bmatrix} m_0 \end{bmatrix} = A\mathbf{m}(i) + B\mathbf{u}(i) , \\
\text{covariance} = & \Sigma_0 = A\Sigma(i)A^* + \Sigma_1 .
\end{align}

We are interested in finding the variable \( \mathbf{x}(i+1) \) conditioned on \( \mathbf{y}(j) \) for \( j \leq i+1 \). This is the same as the variable \( \eta \) conditioned on \( \mathbf{y}(i+1) \), which from (2.6) is the variable \( \eta \) conditioned on \( \mathbf{C}\eta + \varepsilon_2(i+1) \). If \( U \subset \mathbb{R}^n \) and \( V \subset \mathbb{R}^p \) are arbitrary open sets, then
\begin{equation}
P(\eta \in U \mid \mathbf{C}\eta + \varepsilon_2(i+1) \in V) = \frac{P(\eta \in U, \mathbf{C}\eta + \varepsilon_2(i+1) \in V)}{P(\mathbf{C}\eta + \varepsilon_2(i+1) \in V)} .
\end{equation}

Since \( \varepsilon_2(i+1) \) is independent of \( \eta \) we have
\begin{equation}
P(\mathbf{C}\eta + \varepsilon_2(i+1) \in V \mid \eta) = \frac{1}{\text{normalization}} \int_V \exp \left[ -\frac{1}{2}(\xi - C\eta)\Sigma_2^{-1}(\xi - \mathbf{C}\eta) \right] d\xi .
\end{equation}

We conclude that
\begin{equation}
P(\eta \in U, \mathbf{C}\eta + \varepsilon_2(i+1) \in V) = \frac{1}{\text{normalization}} \int_U \exp \left[ -\frac{1}{2}(\eta - \mathbf{m}_0)\Sigma_0^{-1}(\eta - \mathbf{m}_0) \right] d\eta \int_V \exp \left[ -\frac{1}{2}(\xi - C\eta)\Sigma_2^{-1}(\xi - \mathbf{C}\eta) \right] d\xi .
\end{equation}

Now we let the set \( V \) contract to the point \( \mathbf{y} = \mathbf{y}(i+1) \) to conclude that the pdf of the variable \( \eta \) conditioned on \( \mathbf{C}\eta + \varepsilon_2(i+1) = \mathbf{y} \) is given by the formula
\begin{equation}
\frac{1}{\text{normalization}} \exp \left[ -\frac{1}{2}(\eta - \mathbf{m}_0)\Sigma_0^{-1}(\eta - \mathbf{m}_0) - \frac{1}{2}(\mathbf{y} - C\eta)\Sigma_2^{-1}(\mathbf{y} - \mathbf{C}\eta) \right] .
\end{equation}

Hence the variable \( \mathbf{x}(i+1) \) conditioned on \( \mathbf{y}(j) \) for \( j \leq i+1 \), is Gaussian with covariance \( \Sigma(i+1) \) given by the formula
\begin{equation}
\Sigma(i+1) = \{ \Sigma_0^{-1} + C^*\Sigma_2^{-1}C \}^{-1} .
\end{equation}

The mean of this variable can be obtained from the linear term in \( \eta \) in the exponential (2.13), which is
\begin{equation}
\eta\Sigma_0^{-1}\mathbf{m}_0 + \eta C^*\Sigma_2^{-1}\mathbf{y} = \eta \left[ \Sigma_0^{-1} + C^*\Sigma_2^{-1}C \right] \left[ \Sigma_0^{-1} + C^*\Sigma_2^{-1}C \right]^{-1} \left[ \Sigma_0^{-1}\mathbf{m}_0 + C^*\Sigma_2^{-1}\mathbf{y} \right] .
\end{equation}

We conclude from (2.15) that the mean \( \mathbf{m}(i+1) \) of \( \eta \) conditioned on \( \mathbf{y}(i+1) = \mathbf{y} \), is given by the formula
\begin{equation}
\mathbf{m}(i+1) = \left[ \Sigma_0^{-1} + C^*\Sigma_2^{-1}C \right]^{-1} \left[ \Sigma_0^{-1}\mathbf{m}_0 + C^*\Sigma_2^{-1}\mathbf{y} \right] .
\end{equation}

From (2.9), (2.14) we obtain a recurrence formula for the covariance matrices \( \Sigma(i) \) of the variables \( \mathbf{x}(i) \) conditioned on all observations \( \mathbf{y}(j) \) for \( j \leq i \),
\begin{equation}
\Sigma(i+1) = \{ A\Sigma(i)A^* + \Sigma_1 \}^{-1} + C^*\Sigma_2^{-1}C \}^{-1} .
\end{equation}
We can similarly obtain a recurrence relation for the means \( m(i) \) of these variables. Thus we write (2.16) as

\[
(2.18) \quad m(i + 1) = m_0 + \left[ \Sigma_0^{-1} + C^*\Sigma_2^{-1}C \right]^{-1} C^*\Sigma_2^{-1}(y - Cm_0). 
\]

Hence from (2.9), (2.14), (2.18) we have the recurrence relation

\[
(2.19) \quad m(i + 1) = Am(i) + Bu(i) + \Sigma(i + 1)C^*\Sigma_2^{-1}(y - C[m(i) + Bu(i)]).
\]

The system of equations (2.17), (2.19) are the Kalman filter equations. Thus if our initial guess say at time \( i = 0 \) for \( x(i) \) is a Gaussian variable, then we can update our guess for \( x(i) \) for \( i > 0 \) using the incoming observations by the Kalman equations.

Observe that the Kalman equations solve the linear least squares problem for the dynamics (2.1) with observations (2.6). All we need to assume is that the set of variables \( \xi_1(i), \xi_2(i), i = 0, 1, .., \) are independent with mean 0 and covariances \( \Sigma_1 \) and \( \Sigma_2 \) respectively. In that case \( m(i) \) is the best predictor of \( x(i) \) in the mean square sense for a linear function of the observations. We have solved the least squares problem by finding the exact Bayes conditional expectation for the variable \( x(i) \) conditioned on all observations \( y(j) \) for \( j \leq i \) in the Gaussian case and observed that it is linear in the \( y(j), j \leq i \). Trying to solve the least squares problem directly is significantly more complicated than what we have just done. One can in fact show that the least squares problem is equivalent to an LQ problem of classical control theory, and hence soluble by the methods developed in Chapter I.

**Example 1.** Consider the one dimensional system

\[
(2.20) \quad x(i + 1) = x(i) + c\xi_i, \quad y(i) = x(i) + m\zeta_i,
\]

where \( c, m \) are parameters and the variables \( \xi_i, \zeta_i, i = 0, 1, .., \) are i.i.d. standard normal. The Kalman equations are then

\[
(2.21) \quad \sigma^2(i + 1) = \{[\sigma^2(i) + c^2]^{-1} + 1/m^2\}^{-1}, \\
m(i + 1) = m(i) + \sigma^2(i + 1)[y(i + 1) - m(i)]/m^2.
\]

It is clear that \( \sigma^2(i + 1) < m^2 \). We shall show that

\[
(2.22) \quad \lim_{i \to \infty} \sigma^2(i) = \alpha_0 > 0.
\]

To see this we define a function \( f(\cdot) \) by

\[
(2.23) \quad f(\alpha) = \{[\alpha + c^2]^{-1} + 1/m^2\}^{-1} = \frac{m^2[\alpha + c^2]}{\alpha + m^2 + c^2} = m^2\left\{1 - \frac{m^2}{\alpha + m^2 + c^2}\right\}.
\]

Evidently if (2.22) holds then \( f(\alpha_0) = \alpha_0 \), which we can write as

\[
(2.24) \quad \frac{1}{2} \left[ \frac{\alpha_0 + m^2 + c^2}{m^2} + \frac{m^2}{\alpha_0 + m^2 + c^2} \right] = 1 + \frac{c^2}{2m^2}.
\]

If we define the function \( g(\cdot) \) by

\[
(2.25) \quad g(z) = \frac{1}{2} \left[ z + \frac{1}{z} \right], \quad \text{for } z > 0,
\]

then it is clear that there is a unique solution \( z_0 \) to the equation

\[
(2.26) \quad g(z) = 1 + \frac{c^2}{2m^2}, \quad z > 1,
\]
and $z_0 > 1 + c^2/m^2$. Hence there is a unique $\alpha_0 > 0$ satisfying $f(\alpha_0) = \alpha_0$, and it is given by the formula

$$\alpha_0 = m^2 \left\{ z_0 - 1 - \frac{c^2}{m^2} \right\}.$$  

(2.27)

To show that the limit (2.22) holds we write the recurrence (2.21) for the variance in the form

$$z(i+1) = 2 + \frac{c^2}{m^2} - \frac{1}{z(i)} = F(z(i)),$$

where the function $F(\cdot)$ is given by

$$F(z) = 2 + \frac{c^2}{m^2} - \frac{1}{z}.$$  

(2.28)

It is clear that $F(\cdot)$ maps the set $E = \{ z \in \mathbb{R} : z \geq z_0 \}$ into itself and that

$$\sup_{z \geq z_0} |F'(z)| = 1/z_0^2 < 1.$$  

(2.29)

Hence $F : E \to E$ is a contraction mapping and so has a unique fixed point, which is $z_0$.

Let us assume now that $\sigma^2(i)$ takes its asymptotic value $\alpha_0$ and put $\beta = \alpha_0/m^2 < 1$. Then from (2.21) we have the recurrence

$$m(i+1) = (1 - \beta)m(i) + \beta y(i+1).$$  

(2.30)

If we iterate this we obtain a formula for $m(N)$,

$$m(N) = \beta y(N) + \beta(1 - \beta)y(N-1) + \beta(1 - \beta)^2y(N-2) + \cdots + \beta(1 - \beta)^k y(N-k) + \cdots.$$  

(2.31)

Thus the estimate on the variable $x(N)$ is the exponentially weighted moving average of the observations. This method of exponentially weighted moving averages is commonly used, for example in estimating the volatility of a stock from historical data. Note that the variance of our estimate remains bounded and does not go to 0 at large time since at each time step we generate an $O(1)$ amount of noise.

3. The Kalman-Bucy Filter

The continuous time version of the Kalman filter is known as the Kalman-Bucy filter. A continuous time version of the linear dynamics (2.1) is

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t) + W_1(t).$$  

(3.1)

In (3.1) the vector $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$. The noise term $W_1(t) \in \mathbb{R}^n$ is a Gaussian process with mean 0 and symmetric positive definite covariance $n \times n$ matrix $\Sigma_1$. Thus one has

$$E[ W_1(t) ] = 0, \quad E[ \{ v \cdot W_1(t) \} \{ v \cdot W_1(s) \} ] = (v \Sigma_1 v) \delta(t - s),$$  

(3.2)

for any vector $v \in \mathbb{R}^n$, where $\delta(\cdot)$ is the Dirac delta function. Similarly the observation process $y(t) \in \mathbb{R}^p$ satisfies

$$\frac{dy(t)}{dt} = Cx(t) + W_2(t),$$  

(3.3)
where the noise term \( W_2(t) \in \mathbf{R}^p \) is a Gaussian process with mean 0 and symmetric positive definite covariance \( p \times p \) matrix \( \Sigma_2 \).

We assume that \( x(t) \) conditioned on the observations \( y(s) \) for \( s \leq t \) is Gaussian with mean \( \mathbf{m}(t) \) and \( n \times n \) covariance matrix \( \Sigma(t) \). From (3.1) we have that

\[
(3.4) \quad x(t + \Delta t) = x(t) + \int_t^{t+\Delta t} A x(s) + B u(s) + W_1(s) \, ds \approx [I + \Delta t A] x(t) + \Delta t B u(t) + \xi_1(t).
\]

From (3.2) it is clear that the variable \( \xi_1(t) \) is Gaussian with zero mean and covariance \( \Sigma_1 \Delta t \). In the observation process the new observation we make between time \( t \) and \( t + \Delta t \) is

\[
(3.7) \quad y(t + \Delta t) - y(t) \approx C \Delta t x(t) + \int_t^{t+\Delta t} W_2(s) ds = C \Delta t x(t) + \xi_2(t),
\]

where \( \xi_2(t) \) is Gaussian with zero mean and covariance \( \Sigma_2 \Delta t \).

Evidently (3.4), (3.7) for \( t = i \Delta t \) with \( i = 0, 1, 2, \ldots \), give the same dynamics and observations as for the discrete case (2.1), (2.6). Comparing then to the Kalman equation (2.17) we have that

\[
(3.8) \quad \Sigma(t + \Delta t) = \left\{ [I + \Delta t A] \Sigma(t) [I + \Delta t A]^* + \Sigma_1 \Delta t \right\}^{-1} + C^* \Sigma_2^{-1} C \Delta t^{-1}.
\]

Observe now that

\[
(3.9) \quad [I + \Delta t A] \Sigma(t) [I + \Delta t A]^* + \Sigma_1 \Delta t = [\Sigma(t) + \{ A \Sigma(t) + \Sigma(t) A^* + \Sigma_1 \} \Delta t + O((\Delta t)^2) ]^{-1} \Delta t = [\Sigma(t) + \{ A \Sigma(t) + \Sigma(t) A^* + \Sigma_1 \} \Delta t + O((\Delta t)^2) ]^{-1} \Delta t = \Sigma(t) - \{ \Sigma(t)^{-1} A + A^* \Sigma(t)^{-1} + \Sigma(t)^{-1} \Sigma_1 \Sigma(t)^{-1} \} \Delta t + O((\Delta t)^2).
\]

Substituting the last formula in (3.9) into the RHS of (3.8) we see that

\[
(3.10) \quad \Sigma(t + \Delta t) = \left\{ \Sigma(t) - \{ \Sigma(t)^{-1} A + A^* \Sigma(t)^{-1} + \Sigma(t)^{-1} \Sigma_1 \Sigma(t)^{-1} - C^* \Sigma_2^{-1} C \} \Delta t + O((\Delta t)^2) \right\}^{-1} \Delta t = \left\{ I - \{ A + \Sigma(t) A^* \Sigma(t)^{-1} + \Sigma(t) C^* \Sigma_2^{-1} C \} \Delta t + O((\Delta t)^2) \right\}^{-1} \Sigma(t) = \Sigma(t) + \{ A \Sigma(t) + \Sigma(t) A^* + \Sigma(t) C^* \Sigma_2^{-1} C \Sigma(t) \} \Delta t + O((\Delta t)^2).
\]
If we divide (3.10) by $\Delta t$ and let $\Delta t \to 0$, then we obtain the matrix Riccati differential equation

\begin{equation}
\frac{d\Sigma(t)}{dt} = A\Sigma(t) + \Sigma(t)A^* + \Sigma_1 - \Sigma(t)C^*\Sigma_2^{-1}C\Sigma(t).
\end{equation}

We can obtain the differential equation for $m(t)$ corresponding to the discrete equation (2.19) by a similar argument. Thus we have that

\begin{equation}
m(t + \Delta t) = \{I + \Delta tA\}m(t) + \Delta tBu(t) + \Sigma(t + \Delta t)C^*\Sigma_2^{-1}\{y(t + \Delta t) - y(t) - \Delta tC[(I + \Delta tA)m(t) + \Delta tBu(t)]\}.
\end{equation}

If we let $\Delta t \to 0$ in (3.12) we obtain the differential equation

\begin{equation}
\frac{dm(t)}{dt} = Am(t) + Bu(t) + \Sigma(t)C^*\Sigma_2^{-1}\{\frac{dy(t)}{dt} - Cm(t)\}.
\end{equation}

Evidently (3.13) is a linear equation for the mean $m(t)$ of the variable $x(t)$ conditioned on the observations $y(s)$, $s \leq t$. Equations (3.11), (3.13) are the Kalman-Bucy equations for the continuous time linear filter. Note that in this situation both the dynamical equation (3.1) and the observation equation (3.3) are linear.

One can generalize the above analysis to non-linear filters, where the dynamical equation corresponding to (3.1) or the observation equation (3.3) can be nonlinear [2]. Just as in control theory when one departs from LQ problems, explicit formulas for the solutions to non-linear filtering problems are not easily obtained.

**Example 2** (Noisy observation of population growth). We consider the dynamical and observation equations

\begin{equation}
\frac{dx(t)}{dt} = rx(t), \quad \frac{dy(t)}{dt} = x(t) + mW(t),
\end{equation}

where $W(\cdot)$ is white noise. Note that the dynamics in this example is deterministic but the observations are noisy. We assume the initial population $x(0)$ is Gaussian with mean $b$ and variance $\sigma^2$. Letting $\Sigma(t)$ be the variance of $x(t)$ conditioned on the observations, then (3.11) yields the equation

\begin{equation}
\frac{d\Sigma(t)}{dt} = 2r\Sigma(t) - \frac{\Sigma(t)^2}{m^2}, \quad \Sigma(0) = \sigma^2.
\end{equation}

The solution to (3.15) is given by the formula

\begin{equation}
\Sigma(t) = \frac{2rm^2}{1 + Ke^{-2rt}}, \quad \text{where} \quad K = \frac{2rm^2}{\sigma^2} - 1.
\end{equation}

Observe that $\lim_{t \to \infty} \Sigma(t) = 2rm^2$, so the variance remains bounded at large time. The evolution $m(t)$ of the mean of $x(t)$ conditioned on $y(s)$, $s \leq t$, is given from (3.13) by the equation

\begin{equation}
\frac{dm(t)}{dt} = \left[r - \frac{\Sigma(t)}{m^2}\right]m(t) + \frac{\Sigma(t)}{m^2}\frac{dy(t)}{dt}, \quad m(0) = b.
\end{equation}

If we take now $\Sigma(t)$ to have its asymptotic value $\Sigma(t) = 2rm^2$, then (3.17) implies that

\begin{equation}
\frac{d}{dt} [e^{rt}m(t)] = 2re^{rt}\frac{dy(t)}{dt},
\end{equation}

\textit{MATH 626-2011 9}
where \( m(t) \) is given by the formula
\[
(3.19) \quad m(t) = e^{-rt} \left[ \int_0^t 2re^{r(s)} dy(s) + b \right].
\]
Note that (3.19) gives \( m(t) \) as an exponentially weighted moving average just as in the discrete time example.

**Example 3 (Estimation of a parameter).** We estimate the value of a parameter \( \theta \) based on observations \( y(t) \) satisfying
\[
(3.20) \quad \frac{dy(t)}{dt} = \theta M(t) + N(t)W(t),
\]
where \( M(\cdot), N(\cdot) \) are known functions and \( W(\cdot) \) is white noise. In this case the dynamical equation for \( \theta(t) \) is trivial i.e.
\[
(3.21) \quad \frac{d\theta(t)}{dt} = 0.
\]
If \( \Sigma(t) \) is the variance of \( \theta \) given the observations \( y(s), s \leq t \), then (3.11) implies that
\[
(3.22) \quad \frac{d\Sigma(t)}{dt} = -\left[ M(t)\Sigma(t) \right]^2,
\]
whence we conclude
\[
(3.23) \quad \Sigma(t) = \left[ \Sigma(0)^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds \right]^{-1}.
\]
Equation (3.13) yields
\[
(3.24) \quad \frac{dm(t)}{dt} = \frac{M(t)\Sigma(t)}{N(t)^2} \left[ \frac{dy(t)}{dt} - M(t)m(t) \right].
\]
Using (3.22) we can write (3.24) as
\[
(3.25) \quad \frac{d}{dt} \left[ \frac{m(t)}{\Sigma(t)} \right] = \frac{M(t) dy(t)}{N(t)^2} dt.
\]
Thus we obtain the formula
\[
(3.26) \quad m(t) = \Sigma(t) \left[ \frac{m(0)}{\Sigma(0)} + \int_0^t \frac{M(s)}{N(s)^2} dy(s) \right].
\]
If we have no initial knowledge of the parameter \( \theta \) then we should take \( \Sigma(0) \to \infty \). The limit of (3.26) in this case is then given by
\[
(3.27) \quad m(t) = \int_0^t \frac{M(s)}{N(s)^2} dy(s) / \int_0^t \frac{M(s)^2}{N(s)^2} ds.
\]

The formula (3.27) coincides with the maximum likelihood estimate of the parameter \( \theta \). To see this we set \( t = N\Delta t \) and note from (3.20) that the observations \( y(n\Delta t) - y((n-1)\Delta t), n = 1, 2, \ldots, N, \) are independent Gaussian variables with mean \( M((n-1)\Delta t)\theta \Delta t \) and variance \( N((n-1)\Delta t)^2 \Delta t \). The maximum likelihood estimate for \( \theta \) therefore consists of minimizing the quadratic function of \( \theta \),
\[
(3.28) \quad \sum_{n=1}^N \frac{[y(n\Delta t) - y((n-1)\Delta t) - M((n-1)\Delta t)\theta \Delta t]^2}{N((n-1)\Delta t)^2 \Delta t}.
\]
If we let $\Delta t \to 0$, the sum in (3.28) becomes an integral

\begin{equation}
(3.29) \quad \int_0^t \left[ \frac{dy(s)}{ds} - M(s) \theta \right]^2 \frac{ds}{N(s)^2}.
\end{equation}

The minimizer $\theta$ for (3.29) is given by the RHS of (3.27).

References


University of Michigan, Department of Mathematics, Ann Arbor, MI 48109-1109
E-mail address: conlon@umich.edu