7.0 FORMAL DERIVATION OF THE DISCRETE-TIME KALMAN FILTER

Now that we have a feeling for the operation of the discrete-time Kalman filter we will present a formal, more rigorous development of the discrete-time Kalman filter equations. The development will consider the general, linear, time varying vector case. Before proceeding we will present the general mean-squared estimation problem and then develop the orthogonality condition for linear mean-squared estimators.

The development below makes the tacit assumption that the initial state. This and the assumption that the system disturbance, \( w(k) \), is zero-mean means that the state, \( x(k) \) is zero-mean. In fact, the only time the zero-mean assumption is needed is when we show that the linear, minimum mean-square estimator is the same as the minimum, mean-square estimator when the initial state, measurement noise and system disturbance are Gaussian, and the system is linear. All of the rest of the derivations do not invoke the restriction that the mean of \( x(k) \) be zero. The appendix contains a means of avoiding the zero-mean restriction even when showing that the linear, minimum mean-square estimate is the same as the minimum mean-square estimate, under the aforementioned restrictions.

7.1 General Linear Mean-squared Estimation

We present the general mean-squared estimation problem in the form of a theorem which we will prove.

**Theorem 7-1** Given a set of measurements \( y(0), y(1), \ldots y(k) \) that are functions of the states, \( x(0), x(1), \ldots x(k) \), the vector function

\[
\hat{x}(k) = g(y(0), y(1), \ldots y(k)) \tag{7-1}
\]

that minimizes the mean-squared error,

\[
E \left[ \left[ x(k) - \hat{x}(k) \right]^T \left[ x(k) - \hat{x}(k) \right] \right],
\]

is the mean of \( x(k) \) conditioned on the measurements \( y(0), y(1), \ldots y(k) \). That is

\[
\hat{x}(k) = E \{ x(k) \mid y(0), y(1), \ldots y(k) \}. \tag{7-2}
\]

**PROOF:**

We can write the joint density of \( x(k) \) and the measurements as the product of a density of \( x(k) \) conditioned on \( y(0), y(1), \ldots y(k) \) and a joint density of the measurements. Specifically, we can write

\[
f(x(k), y(0), y(1), \ldots y(k))
\]

\[
= f(x(k) \mid y(0), y(1), \ldots y(k)) f(y(0), y(1), \ldots y(k)) \tag{7-3}
\]
Also, since the mean-squared error is a function of the measurements through Equation 7-1 we can write the mean-squared error as
\[
e = E \left[ (x(k) - \hat{x}(k))^T (x(k) - \hat{x}(k)) \right] \quad (7-4)
\]
\[
e = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[ (x(k) - \hat{x}(k))^T (x(k) - \hat{x}(k)) \right] f(x(k), y(0), y(1), \ldots y(k)) dy(0) \cdots dy(k) dx(k)
\]
If we substitute Equation 7-3 into Equation 7-4 we get
\[
e = \int_{y(0), \ldots, y(k)} \cdots \int_{x(k)} f(x(k), y(0), y(1), \ldots y(k)) dx(k) dy(0) \cdots dy(k)
\]
\[
e = \int_{x(k)} \cdots \int_{x(k)} \left[ (x(k) - \hat{x}(k))^T (x(k) - \hat{x}(k)) \right] f(x(k)) dx(k) dy(0) \cdots dy(k)
\]
Since \( f(y(0), y(1), \ldots y(k)) \) is fixed (because \( y(0), y(1), \ldots y(k) \) are specified random variables with a fixed joint density) and positive, \( e \) is minimized by minimizing the multiple integral over \( x(k) \). However, this integral is equal to (see Papoulis, Reference 8) \( E \left[ (x(k) - \hat{x}(k))^T (x(k) - \hat{x}(k)) \right] y(0), y(1), \ldots y(k) \).
Furthermore this conditional expectation is minimized (again, see Papoulis) when \( \hat{x}(k) \) is equal to the mean of \( x(k) \) conditioned on the measurements.
With this we conclude that the estimate, \( \hat{x}(k) \), that minimizes the mean-squared error is given by
\[
\hat{x}(k) = g(y(0), y(1), \ldots y(k)) = E \left[ x(k) \mid y(0), y(1), \ldots y(k) \right]
\]
and the theorem is proved.

Equations 7-6 and 7-2 tell us that if we could compute the conditional expectation we would have the form of the minimum mean-squared estimator. Unfortunately, computing the conditional expectation, with today’s mathematical tools, is virtually impossible for the general case. The value of these equations will come later when we show that \( g(y(0), y(1), \ldots y(k)) \) is a linear function of \( y(0), y(1), \ldots y(k) \) for the case where \( x(k), y(0), y(1), \ldots y(k) \) are zero-mean, Gaussian random variables.

7.2 Orthogonality Condition

Since the general minimum mean-squared estimator is difficult to find we turn our attention to finding a linear estimator that minimizes the mean-squared error between the state and its estimate. In this case we present what we could call a specialized form of the general mean-squared estimation problem. This is normally called the Orthogonality Condition since the result of the theorem states that the optimum linear estimator is the one that makes the error between the actual state and the estimate orthogonal to all of the
measurements. We will state the orthogonality conditions in the form of a theorem which we will prove.

\textit{Theorem 7-2: Orthogonality Condition} Given a set of measurements \( y(0), y(1), \ldots, y(k) \) that are functions of the states, \( x(0), x(1), \ldots, x(k) \), the linear, vector function

\[ \hat{x}(k) = L(y(0), y(1), \ldots, y(k)) \tag{7-7} \]

that minimizes the mean-squared error

\[ E\left[ \left( x(k) - \hat{x}(k) \right)^T \left( x(k) - \hat{x}(k) \right) \right] \]

satisfies the condition

(orthogonality condition)

\[ E\left[ \left( x(k) - \hat{x}(k) \right) y^T(l) \right] = 0 \quad \forall l = 0, 1, \ldots, k. \tag{7-8} \]

Furthermore, the minimum error is given by

\[ e_{\min} = E\left[ \left( x(k) - \hat{x}(k) \right)^T x(k) \right]. \tag{7-9} \]

\textbf{PROOF}

In the proof we use a vector random variable property that states that if \( z \) is a vector random variable then \( E\{z^T z\} = \text{tr}\left( E\{zz^T\} \right) \) where \( \text{tr}(\cdot) \) denotes the trace of a square matrix and is the sum of the diagonal elements. The derivation of this equation is left as an exercise for the reader. Henceforth we denote \( L(y(0), y(1), \ldots, y(k)) \) by \( L(y(k)) \) to save space.

Suppose \( L'(y(k)) \) (the prime does not denote derivative) is a linear operator that minimizes the mean-squared error and \( L(y(k)) \) is a linear operator that satisfies the orthogonality condition of the theorem. We will show that \( L'(y(k)) \) and \( L(y(k)) \) are equal and thus the same linear operator. We can write

\[ x(k) - L'(y(k)) = x(k) - L(y(k)) + L(y(k)) - L'(y(k)) \]

\[ = x(k) - L(y(k)) + L^*(y(k)) \tag{7-10} \]

where

\[ L^*(y(k)) = L(y(k)) - L'(y(k)). \tag{7-11} \]

\textit{Equation 7-11} follows from the fact that the difference of linear operators is another linear operator.

Since \( L'(y(k)) \) minimizes the mean-squared error we can write

\[ e_{\min} = \text{tr}\left( E\left[ \left( x(k) - L'(y(k)) \right) \left( x(k) - L'(y(k)) \right)^T \right] \right) \tag{7-12} \]
or using \( Equation \ 7-10 \)

\[
e_{\min} = \text{tr}\left( E\left[ \left( x(k) - L(y(k)) + L'(y(k)) \right)\left( x(k) - L(y(k)) + L'(y(k)) \right)^T \right] \right).
\]  

(7-13)

If we expand \( Equation \ 7-13 \) we get

\[
e_{\min} = \text{tr}\left( E\left[ \left( x(k) - L(y(k)) \right)\left( x(k) - L(y(k)) \right)^T \right] \right)
+ \text{tr}\left( E\left[ \left( x(k) - L(y(k)) \right)\left( L'(y(k)) \right)^T \right] \right)
+ \text{tr}\left( E\left[ L'(y(k))\left( x(k) - L(y(k)) \right)^T \right] \right)
+ \text{tr}\left( E\left[ L'(y(k))\left( L'(y(k)) \right)^T \right] \right).
\]  

(7-14)

We first consider the second and third terms of \( Equation \ 7-14 \). According to our previous suppositions, \( L(y(k)) \) satisfies the orthogonality condition. This means that \( x(k) - L(y(k)) \) is orthogonal to \( y(0), y(1), \cdots y(k) \). However, since \( L'(y(k)) \) is a linear combination of \( y(0), y(1), \cdots y(k), x(k) - L(y(k)) \) is also orthogonal to \( L'(y(k)) \). Thus, the second and third terms are zero and \( Equation \ 7-14 \) reduces to

\[
e_{\min} = \text{tr}\left( E\left[ \left( x(k) - L(y(k)) \right)\left( x(k) - L(y(k)) \right)^T \right] \right).
\]  

(7-15)

By the vector random variable property indicated earlier, both terms of \( Equation \ 7-15 \) are mean-squared values and are thus positive. Furthermore, the first term is recognized as the mean-squared error using the linear operator \( L(y(k)) \). If the second term is anything but zero, \( Equation \ 7-15 \) states that the linear operator \( L'(y(k)) \) does not minimize the minimum mean-squared error. This contradicts our previous supposition. Thus we conclude that the second term must be zero.

The expectation in the second term is recognized as a correlation. From random variable theory, (see Papoulis) if the correlation is zero, the random variables are either orthogonal or one is zero. For the first statement to be true \( L'(y(k)) \) would need to be orthogonal to itself, which is impossible. Thus we conclude that \( L'(y(k)) = 0 \) and thus that \( L(y(k)) = L'(y(k)) \). This, in turn, states that the linear operator that satisfies the orthogonality condition is equal to the linear operator that minimizes the mean-squared error. With this, the first part of the theorem is proved.
The second part of the theorem can be proved by direct computation. Specifically, given a linear operator, \( L(y(k)) \), that satisfies the orthogonality condition, the minimum error is given by

\[
e_{\text{min}}^2 = \text{tr} \left( E \left[ \left( x(k) - L(y(k)) \right) \left( x(k) - L(y(k)) \right)^T \right] \right) = \text{tr} \left( E \left[ \left( x(k) - L(y(k)) \right) x^T(k) \right] \right) - \text{tr} \left( E \left[ \left( x(k) - L(y(k)) \right) L(y(k)) \right] \right).
\] (7-16)

However, the last term is zero by the orthogonality condition so that the minimum mean-squared error is given by

\[
e_{\text{min}}^2 = \text{tr} \left( E \left[ \left( x(k) - L(y(k)) \right) x^T(k) \right] \right) \quad (7-17)
\]

and the second part of the theorem is proved.

As we will see shortly, the orthogonality condition is very powerful in that it gives us the criterion for designing linear, minimum mean-squared estimators. Furthermore, the criterion is one that we can directly use to develop the estimator equations. Before Kalman, the estimators were formulated as explicit linear combinations of the measurements. That is, the formulations are of the form

\[
\hat{x}(k) = \sum_{l=0}^{k} A_{k,l} y(l). \quad (7-18)
\]

This formulation is very cumbersome because a new set of \( A_{k,l} \) must be computed every time a new measurement is made. As the number of measurements increases, the number of \( A_{k,l} \) increases, which imposes a significant computational burden in terms of finding the \( A_{k,l} \) and then implementing Equation 7-18. As we have already noted, Kalman avoided this pitfall by deriving a recursive filter. Further, he showed, as will we, that the recursive filter still minimizes the mean-squared error. That is, it was the optimum linear estimator in a minimum mean-squared sense.

### 7.3 Gaussian Noise Ramifications

So far in our developments of Kalman filters we have not discussed the form of the density functions for the states, measurements, system disturbances and measurement noise. In this section we show that the linear estimator is the optimum, minimum mean-squared estimator if we have a linear system model, a linear measurement model, if the initial state is a zero-mean Gaussian random variable and if the system disturbance and measurement noises are uncorrelated, zero-mean, white, Gaussian random processes that are uncorrelated with the initial state. The difference in the conditions of the previous sentence and those that we have discussed previously is that we have now added the additional restriction that the initial state, measurement noise
and system disturbance be Gaussian\(^1\). The fact that we can claim that the linear estimator is the optimum estimator means that it is the best of all estimators, including any non-linear estimators we could derive (for a linear system).

We begin by stating and proving another theorem and then use the results of the theorem to establish our final results.

**Theorem 7-3** If the states \(x(0), x(1), \ldots, x(k)\), and the measurements \(y(0), y(1), \ldots, y(k)\) of Theorems 7-1 and 7-2 are zero-mean, jointly Gaussian, random variables then the general, minimum mean-squared estimate of Theorem 7-1 reduces to the linear, minimum mean-squared estimate of Theorem 7-2. That is

\[
E\{x(k)|y(0), y(1), \ldots, y(k)\} = L(y(k)).
\]  
(7-19)

**PROOF**

From **Theorem 7-2** \(L(y(k))\) satisfies the orthogonality condition. That is, \(x(k) - L(y(k))\) is orthogonal to \(y(0), y(1), \ldots, y(k)\). But, since \(x(k)\) and \(y(0), y(1), \ldots, y(k)\) are zero mean, \(L(y(k))\) is also zero mean and \(x(k) - L(y(k))\) is uncorrelated with \(y(0), y(1), \ldots, y(k)\). Further, since \(y(0), y(1), \ldots, y(k)\) are Gaussian random variables, \(L(y(k))\) is a Gaussian random variable. Also, since \(x(k)\) and \(L(y(k))\) are Gaussian random variables, \(x(k) - L(y(k))\) is a Gaussian random variable. Finally, since \(x(k) - L(y(k))\) and \(y(0), y(1), \ldots, y(k)\) are Gaussian and uncorrelated they are also independent.

The last statement of the previous paragraph allows us to write

\[
E\{x(k) - L(y(k))\}y(0), y(1), \ldots, y(k)\} = E\{x(k) - L(y(k))\}] = 0
\]  
(7-20)

where the first equality comes from the fact that \(x(k) - L(y(k))\) and \(y(0), y(1), \ldots, y(k)\) are independent and the second equality comes from the fact that \(x(k)\) and \(L(y(k))\) are zero mean. **Equation 7-20** allows us to write

\[
E\{x(k)|y(0), y(1), \ldots, y(k)\} = E\{L(y(k))|y(0), y(1), \ldots, y(k)\}.
\]  
(7-21)

Since the right side of **Equation 7-21** is the expectation of a function of \(y(0), y(1), \ldots, y(k)\) given \(y(0) = y(0), y(1) = y(1), \ldots, y(k) = y(k)\) it reduces to \(L(y(k))\), which completes the proof of the theorem.

\(^1\) And that the mean of the initial state is zero.
We now want to extend this to our particular application. First, from Chapter 4 we know that the state, \( x(k) \) is a linear combination of the initial state, \( x(0) \) and the input disturbances \( w(0), w(1), \ldots w(k-1) \). Therefore, since \( x(0) \) and the \( w(0), w(1), \ldots w(k-1) \) are zero-mean, Gaussian random variables, the state \( x(k) \) is also a zero-mean, Gaussian random variable. This holds for any \( k \). The measurements, \( y(0), y(1), \ldots y(k) \) are linear combinations of the states \( x(0), x(1), \ldots x(k) \) and the measurement noise random variables \( v(0), v(1), \ldots v(k) \). Therefore, since the states and measurement noise random variables are zero-mean and Gaussian, the measurements \( y(0), y(1), \ldots y(k) \) are zero-mean and Gaussian. Based on these observations we conclude that the state, \( x(k) \), and measurements \( y(0), y(1), \ldots y(k) \) satisfy the conditions of Theorem 7-3. Therefore, we conclude that the linear, minimum mean-squared estimator is the optimum minimum mean-squared estimator for this case. Stated another way, the Kalman filter is the optimum, minimum mean-squared estimator, for the assumption of a linear system and all of the other assumptions on the initial state, measurement noise and system disturbance.

7.4 Kalman Filter Derivation

We now proceed to derive the Kalman filter equations. The derivation we have chosen is based on a paper by Kailath (References 3 and 4) and employs the concept of an innovations process. The reader will recall (see Papoulis) that the concept of an innovations filter arose when studying whitening filters. In fact, a key element of the innovations approach of Kailath is to define a process that converts the measurements to white noise. This step greatly simplifies the use of the orthogonality condition in deriving the Kalman filter. We will start our development by carefully presenting our problem statement. We then proceed to define the whitening filter and show that its output is white. Finally, we use these results, along with the orthogonality condition to derive the Kalman filter equations.

7.4.1 Problem Statement

The system model that provides the basis for the Kalman filter is defined by the state equation

\[
x(k+1) = F(k)x(k) + G(k)w(k)
\]  

(7-22)

where \( F(k) \) and \( G(k) \) are known matrices and \( w(k) \) is a white, zero mean, random process with a covariance of \( Q(k) \). That is

\[
E\{w(k)w^T(l)\} = Q(k)\delta(l-k)
\]  

(7-23)
where $\delta_s(k-l)$ is the Kronecker delta function.

The initial state is random variable with a mean of $\eta_s(0)$ and a covariance of $P_0$. It is uncorrelated with $w(k)$. Thus

\[
E\{x(0)\} = \eta_s(0),
\]

\[
E\left\{ \left( \left[ x(0) - \eta_s(0) \right] \left[ x(0) - \eta_s(0) \right]^T \right) \right\} = P_0,
\]

\[
E\{x(0)w^T(k)\} = 0 \quad \forall k, \text{ and}
\]

\[
E\{w(k)x^T(0)\} = 0 \quad \forall k.
\]

From Equations 7-23 and 7-26, and the results of Chapter 4 we have that

\[
E\{x(k)w^T(l)\} = 0 \quad \forall l \geq k.
\]

Likewise, from Equations 7-23 and 7-27, and their results of Chapter 4 we have that

\[
E\{w(l)x^T(k)\} = 0 \quad \forall l \geq k.
\]

The measurement model for our Kalman filter design is given by

\[
y(k+1) = H(k+1)x(k+1) + v(k+1)
\]

where $H(k+1)$ is a know matrix and $v(k+1)$ is a white, zero-mean, random process with a covariance of $R(k+1)$. That is

\[
E\{v(k)v^T(l)\} = R(k)\delta_s(k-l).
\]

The processes $w(k)$ and $v(k)$ are uncorrelated and $v(k)$ is uncorrelated with $x(0)$. Thus,

\[
E\{w(k)v^T(l)\} = 0 \quad \forall k,l,
\]

\[
E\{v(l)w^T(k)\} = 0 \quad \forall k,l,
\]

\[
E\{v(k)x^T(0)\} = 0 \quad \forall k, \text{ and}
\]

\[
E\{x(0)v^T(k)\} = 0 \quad \forall k.
\]

The above properties can be combined to yield the following properties.

\[
E\{y(k+1)v^T(l)\} = 0 \quad \forall l > k,
\]
\[ E\{w(l)y^T(k+1)\} = 0 \ \forall l > k , \quad (7-37) \]
\[ E\{x(k)v^T(l)\} = 0 \ \forall k,l , \quad (7-38) \]
\[ E\{v(l)x^T(k)\} = 0 \ \forall k,l , \quad (7-39) \]
\[ E\{y(k)v^T(l)\} = 0 \ \forall l \neq k , \text{ and} \]
\[ E\{v(l)y^T(k)\} = 0 \ \forall l \neq k . \quad (7-41) \]

Given the system and attendant limiters described we wish to find the linear estimator
\[ \hat{x}(k+1) = L(y(k+1)) \quad (7-42) \]
that minimizes the mean-squared error
\[ e = \text{tr}\left( P(k+1) \right) = \text{tr}\left( E\left[ (x(k+1) - \hat{x}(k+1))(x(k+1) - \hat{x}(k+1))^T \right] \right) \quad (7-43) \]

In Equation 7-43, \( P(k+1) \) is the covariance of the state estimation error.

The estimator defined by Equation 7-42 does not provide for a means of estimating \( \hat{x}(0) \), if we assume that the indexing on \( k \) starts at zero. Since we desire \( \hat{x}(0) \), and we will need it and \( P(0) \) to get the Kalman filter started, we will set \( \hat{x}(0) \) to the reasonable value of \( E\{x(0)\} = \eta_0 \). With this we also have that
\[ P(0) = E\left[ (x(0) - \hat{x}(0))(x(0) - \hat{x}(0))^T \right] \]
\[ = E\left[ (x(0) - \eta_0)(x(0) - \eta_0)^T \right] = P_0 . \quad (7-44) \]

As you showed in an earlier homework, this initial state estimate minimizes the mean-squared error given no measurements.

**7.4.2 Innovations Process**

There are many ways to solve the linear, minimum, mean-squared estimation problem so as to arrive at the Kalman filter equations. The one we will use is based on a technique by Kailath (References 3 and 4). With this technique we define an innovations process and then proceed to show that it is white. After this, we use the orthogonality condition to derive the Kalman filter equations. The step of finding the innovations process, and showing it is white, greatly simplifies the derivation of the actual Kalman filter equations.
Let \( \hat{x}(k) \) be an estimate of the state, \( x(k) \), at stage \( k \) that satisfies the orthogonality condition (i.e., the conditions of Theorem 7.2). Given this we form a random variable, \( a(k+1) \), as

\[
a(k+1) = y(k+1) - H(k+1)F(k)\hat{x}(k). \tag{7-45}
\]

We now show that the random process formed by all of the \( a(k+1) \) is white. That is, we show that

\[
E\{a(k+1)a^T(l+1)\} = S(k+1)\delta_l(k-l). \tag{7-46}
\]

For future reference we recall that \( \hat{x}(k) = \sum_{l=0}^{k} A_{k,l}y(l) \) (Equation 7-18).

If we use Equations 7-22 and 7-30 to replace \( y(k+1) \), Equation 7-45 becomes

\[
a(k+1) = H(k+1)F(k)[x(k) - \hat{x}(k)] + H(k+1)G(k)w(k) + v(k+1). \tag{7-47}
\]

With this, \( E\{a(k+1)a^T(l+1)\} \) can be written as

\[
E\{a(k+1)a^T(l+1)\} = H(k+1)F(k)E\left\{[x(k) - \hat{x}(k)][x(l) - \hat{x}(l)]^T\right\}F^T(l)H^T(l+1) \\
+ H(k+1)F(k)E\left\{[x(k) - \hat{x}(k)]w^T(l)\right\}G^T(l)H^T(l+1) \\
+ H(k+1)F(k)E\left\{[x(k) - \hat{x}(k)]v^T(l+1)\right\} \\
+ H(k+1)G(k)E\left\{w(k)[x(l) - \hat{x}(l)]^T\right\}F^T(l)H^T(l+1) \\
+ H(k+1)G(k)E\left\{w(k)w^T(l)\right\}G^T(l)H^T(l+1) \\
+ H(k+1)G(k)E\left\{w(k)v^T(l+1)\right\} \\
+ E\left\{v(k+1)[x(l) - \hat{x}(l)]^T\right\}F^T(l)H^T(l+1) \\
+ E\left\{v(k+1)w^T(l)\right\}G^T(l)H^T(l+1) \\
+ E\left\{v(k+1)v^T(l+1)\right\} \tag{7-48}
\]

Since the processes \( w(k) \) and \( v(k) \) are uncorrelated, the sixth and eighth terms in Equation 7-48 are zero for all \( k \) and \( l \). Thus, we need not consider them any further.

Let's first consider the case where \( k > l \). Using Equations 7-27 and 7-37, we conclude that the fourth term is zero. Also, since the process \( w(k) \) is white, the fifth term is also zero. Using Equations 7-39 and 7-41, we conclude that the
seventh term is also zero. Finally, since the process $v(k)$ is white, the last term is zero. With this, and some manipulation, Equation 7-48 reduces to

$$
E\{a(k+1)a^T(l+1)\} = H(k+1)F(k)E\left[\left(x(k) - \hat{x}(k)\right)\left(x(k) - \hat{x}(k)\right)^T\right]F^T(l)H^T(l+1)
$$

(7-49)

Since $l < k$, $l + 1$ is at most $k$ and, by the orthogonality condition, the first term is zero. The second term is also zero because of the orthogonality condition. From this we conclude that

$$
E\{a(k+1)a^T(l+1)\} = 0 \forall k > l
$$

(7-50)

Next we consider the case where $k < l$. From Equations 7-28 and 7-36, the second term of Equation 7-48 is zero. Also, from Equations 7-38 and 7-40, the third term is zero. As before, since the processes $w(k)$ and $v(k)$ are white the fifth and last terms are zero. With this, Equation 7-48 reduces to

$$
E\{a(k+1)a^T(l+1)\} = E\left\{y(k+1)\left[x(l) - \hat{x}(l)\right]^T\right\}F(l)H^T(l+1)
$$

$$
- H(k+1)F(k)E\left\{\hat{x}(k)\left[x(l) - \hat{x}(l)\right]^T\right\}F^T(l)H^T(l+1)
$$

(7-51)

We note that Equation 7-51 is the transpose of Equation 7-49 with the $l$ and $k$ indices reversed. Using this we conclude that

$$
E\{a(k+1)a^T(l+1)\} = 0 \forall k < l
$$

(7-52)

For the last step we need to show that Equation 7-48 is not zero for $k = l$. For this case we note that, because of Equations 7-28 and 7-36, the second term is zero and because of Equations 7-38 and 7-40, the third term is zero. Since $k = l$, we can use Equations 7-29 and 7-31 to conclude that the fourth term is zero. Also, from Equations 7-39 and 7-41 the seventh term is zero. This leaves us with

$$
E\{a(k+1)a^T(k+1)\} = H(k+1)F(k)E\left[\left[x(k) - \hat{x}(k)\right]\left[x(k) - \hat{x}(k)\right]^T\right]F^T(k)H^T(k+1)
$$

$$
+ H(k+1)G(k)E\left[w(k)w^T(k)\right]G^T(k)H^T(k+1)
$$

$$
+ E\left\{v(k+1)v^T(k+1)\right\}
$$

(7-53)

With this we have shown that, indeed, the process $a(k)$, is white. We also conclude that $a(k)$ has a covariance of
7.4.3 Derivation of the Kalman Filter Equations

With the results of the previous section, we can now proceed to derive the Kalman filter equations. We start by noting, from Equations 7-18 and 7-45 that \( \alpha(m) \) is a function of \( y(l) \) for \( l \in [0, m] \). Therefore, the state estimate at \( k+1 \) given by

\[
\hat{x}(k+1) = \sum_{m=0}^{k-1} B_{k+1,m} \alpha(m)
\]  

(7-55)

is equivalent to the state estimate given by

\[
\hat{x}(k+1) = \sum_{l=0}^{k-1} A_{k+1,l} y(l).
\]  

(7-56)

Furthermore by the uniqueness of the minimum, mean-squared estimator (see Theorem 7-2) if we find the \( B_{k+1,m} \) that satisfy the orthogonality condition for \( \alpha(k) \) and the \( A_{k+1,l} \) that satisfy the orthogonality condition for \( y(k) \), the state estimates given by Equations 7-55 and 7-56 will be exactly the same. Given this, we elect to find the \( B_{k+1,m} \) that satisfy the orthogonality condition for \( \alpha(k) \).

From the orthogonality condition (Theorem 7-2) we have

\[
E \left[ (x(k+1) - \hat{x}(k+1))^T \alpha^T(l) \right] = 0 \quad \forall l \in [0, k+1].
\]  

(7-57)

Substituting from Equation 7-55 and manipulating results in

\[
E \{ x(k+1) \alpha^T(l) \} = \sum_{m=0}^{k-1} B_{k+1,m} E \{ \alpha(m) \alpha^T(l) \} = B_{k+1,l} S(l)
\]  

(7-58)

where the last term follows from the fact that the process \( \alpha(k) \) is white. \( S(l) \) is defined in Equation 7-54. From this we conclude that

\[
B_{k+1,l} = E \{ x(k+1) \alpha^T(l) \} S^{-1}(l)
\]  

(7-59)

and that

\[
\hat{x}(k+1) = \sum_{m=0}^{k-1} B_{k+1,m} \alpha(m) = \sum_{m=0}^{k-1} E \{ x(k+1) \alpha^T(m) \} S^{-1}(m) \alpha(m).
\]  

(7-60)

We can subsequently write this as

\[
\hat{x}(k+1) = \sum_{m=0}^{k} E \{ x(k+1) \alpha^T(m) \} S^{-1}(m) \alpha(m) + K(k+1) \alpha(k+1)
\]  

(7-61)

where

\[
S(k+1) = H(k+1) \left[ F(k) P(k) F^T(k) + G(k) Q(k) G^T(k) \right] H^T(k+1) + R(k+1).
\]  

(7-54)
\[ K(k+1) = E\{x(k+1)\alpha^T(k+1)\} S^{-1}(k+1). \]  
(7-62)

We can manipulate the first term using the system model to obtain

\[
\sum_{m=0}^{k} E\{x(k+1)\alpha^T(m)\} S^{-1}(m)\alpha(m)
= \sum_{m=0}^{k} E\left[\left[F(k)x(k) + G(k)w(k)\right]\alpha^T(m)\right] S^{-1}(m)\alpha(m).
\]  
(7-63)

Substituting this and Equation 7-45 into Equation 7-61 gives us

\[ \hat{x}(k+1) = F(k)\hat{x}(k) + K(k+1)[y(k+1) - H(k+1)F(k)\hat{x}(k)] \]  
(7-64)

which is the form of the Kalman filter we presented in Section 5.

We can compute the Kalman gain from Equations 7-62 and 7-45. Specifically,

\[
K(k+1) = E\{x(k+1)\alpha^T(k+1)\} S^{-1}(k+1)
= E\left\{\left[F(k)x(k) + G(k)w(k)\right]\left[y^T(k+1) - \hat{x}^T(k)F^T(k)H^T(k+1)\right]\right\} S^{-1}(k+1)
= E\left\{\left[F(k)x(k) + G(k)w(k)\right]\left[\left[x(k) - \hat{x}(k)\right]^T F^T(k)H^T(k+1) + w^T(k)G^T(k)H^T(k+1) + v^T(k+1)\right]\right\} S^{-1}(k+1)
\]  
(7-65)

or

\[
K(k+1) = P(k+1|k)H^T(k+1)[H(k+1)P(k+1|k)H^T(k+1) + R(k+1)]^{-1}
\]  
(7-66)

where

\[
P(k+1|k) = F(k)P(k)F^T(k) + G(k)Q(k)G^T(k).
\]  
(7-67)

Finally, we can compute \(P(k+1)\) from the last part of the orthogonality condition. Specifically,

\[
P(k+1) = E\{[x(k+1) - \hat{x}(k+1)]^2\}
= [I - K(k+1)H(k+1)]P(k+1|k).
\]  
(7-68)

If we extend the scalar results of Chapter 5 we can arrive at the following. We define

\[
\hat{x}(k+1|k) = F(k)\hat{x}(k).
\]  
(7-69)

With this we can write
Thus, we note that if we term $\hat{x}(k+1|k)$ then $P(k+1|k)$ is the covariance on the error between this predicted state and the actual state.

If we summarize the results of this section we obtain the Kalman filter equations presented in Section 5.3. Specifically

$$P(k+1|k) = F(k)P(k)F^T(k) + G(k)Q(k)G^T(k)$$ (7-71)

$$K(k+1) = P(k+1|k)H^T(k+1)[H(k+1)P(k+1|k)H^T(k+1) + R(k+1)]^{-1}$$ (7-72)

$$\hat{x}(k+1|k) = F(k)\hat{x}(k)$$ (7-73)

$$\hat{x}(k+1) = \hat{x}(k+1|k) + K(k+1)[y(k+1) - H(k+1)\hat{x}(k+1|k)]$$ (7-74)

$$P(k+1) = [I - K(k+1)H(k+1)]P(k+1|k)$$ (7-75)

### 7.5 Control Theoretic Form of the Kalman Filter

The above is the tracking form of the Kalman filter. The control theoretic formulation can be derived by defining

$$\hat{x}(k+1) = \sum_{l=0}^{k} A_{k,l}y(l)$$ (7-76)

and

$$\alpha(k) = y(k) - H(k)\hat{x}(k).$$ (7-77)

The details are left to the reader.
7.6 PROBLEMS

1. Show that, if \( z \) is a vector random process, \( \mathbb{E}\{z^T z\} = \text{tr}\left(\mathbb{E}\{zz^T\}\right) \).

2. Show that the sum of two linear operators on a process \( y(k) \) is itself a linear operator.

3. Fill-in the steps behind Equation 7-20.

4. Show that if \( x(0) \) and the \( w(0), w(1), \ldots w(k-1) \) are zero-mean, Gaussian random variables, the state \( x(k) \) is also a zero-mean, Gaussian random variable.

5. Prove Equations 7-28, 7-29, and 7-36 through 7-41.


7. Derive Equations 7-53 and 7-54.


9. Derive Equation 7-68.

10. Derive the control theoretic form of the Kalman filter using the innovations approach of this chapter. Use the hints provided by Equations 7-76 and 7-77.