plane with a uniform angular velocity $\omega$, which is set as close to $2\pi/T$ as possible, where $T$ is the period of the orbit;

(b) a horizon sensor mounted in the cylinder, which measures the angle $\alpha(t)$ between the line of sight to the Earth's center and a reference line on the rotating cylinder. This reference line is chosen so that $\alpha(T_0) = 0$; that is, $\alpha \equiv 0$ at perigee. In the absence of measurement error, $\alpha(t)$ is a periodic function, well defined by the parameters $e$, $a$, $T_0$, $\omega$, and $\alpha(T_0)$; e.g., if the orbit is circular, then $\alpha(t) = 0$ if $\omega = 2\pi/T$ exactly. Any deviation of the orbit from a circle will cause the rotating reference line to periodically lead and lag behind the line of sight. The problem here is to improve the estimates of $e$, $a$, $T_0$, $\omega$, and $\alpha(T_0)$, based on noisy measurements of $\alpha(t)$.

The relationship between $\alpha$, at a time $t$, and the parameters $e$, $a$, $T_0$, $\omega$, and $\alpha(T_0)$ is given implicitly as follows:

$$
\alpha = \phi - m; \quad \cos \phi = \frac{\cos E - e}{1 - e \cos E}; \quad M = E - e \sin E;
$$

$$
M = \frac{2\pi(t - T_0)}{T}; \quad m = \omega(t - T_0) - \alpha_0; \quad \alpha_0 = \alpha(T_0);
$$

$$
T = \frac{2\pi}{R \sqrt{g}}. \quad
$$

Here the angles $\phi$, $M$, and $E$ are known as the true, mean, and eccentric anomalies respectively, $g = \text{acceleration of gravity at the Earth's surface}$, and $R = \text{radius of Earth}$. Note that $m = M$ if $\omega = 2\pi/T$ and $\alpha_0 = 0$.

By taking differentials of the relationships above and eliminating $d\phi$, $dM$, $dE$, $dm$, and $dT$, the following relation may be obtained:

$$
da = \frac{\partial \alpha}{\partial a} da + \frac{\partial \alpha}{\partial e} de + \frac{\partial \alpha}{\partial T_0} dT_0 + T_0 d\omega + d\alpha_0; \quad
$$

where

$$
\frac{\partial \alpha}{\partial a} = \frac{3\pi T_0 (1 - e^2) \sin E}{a T \sin \phi (1 - e \cos E)^3}, \quad \frac{\partial \alpha}{\partial e} = \frac{(1 - e^2) E}{\sin \phi (1 - e \cos E)^2},
$$

$$
\frac{\partial \alpha}{\partial T_0} = \omega - \frac{2\pi (1 - e^2) \sin E}{T \sin \phi (1 - e \cos E)^2},
$$

and these partial derivatives are evaluated with the best current estimates of $a$, $e$, $T_0$, $\omega$, $\alpha_0$ at the time of the measurement.

The measurement $z(t)$ is assumed to contain a random error with mean zero.

---

\[
\begin{align*}
\dot{z}(t) &= \alpha(t) + \nu, \\
E(\nu) &= 0, \quad E(\nu^2) = R, \quad \text{and } R \text{ is known. Let } \ddot{\alpha}(t) \text{ be the predicted measurement, using the best current estimates of } a, e, T_0, \omega, \alpha_0, \text{ at time } t. \quad \text{Then we have}
\end{align*}
\]
\[
\begin{align*}
\dot{z}(t) - \ddot{\alpha}(t) &= d\alpha(t) + \nu. \\
\begin{pmatrix}
\frac{\partial \alpha}{\partial a} \\
\frac{\partial \dot{\alpha}}{\partial e} \\
\frac{\partial \dot{\alpha}}{\partial T_0} \\
\frac{\partial \alpha}{\partial T_0}
\end{pmatrix}
\begin{pmatrix}
da \\
de \\
dT_0 \\
d\omega
\end{pmatrix} + \nu.
\end{align*}
\]

The linear relation can then be used to estimate \( da, de, dT_0, d\omega \), and the measurement \( z(t) - \ddot{\alpha}(t) \).

---

1. In Equation (12.2.1), assume that \( x \) and \( \nu \) are independent normal vectors with Gaussian density functions. Show that the joint density function \( p(x, \nu) \) is proportional to \( \exp(-J) \), where \( J \) is as used in Equation (12.2.5). Thus, \( x = \dot{x}, \nu = z - H\dot{x} \) maximize \( p(x, \nu) \), justifying the name “maximum likelihood estimate.”

2. Establish the relations
\[
\begin{align*}
P &= M - MHH^T + R, \\
PH^TR^{-1} &= MHH^T + R.
\end{align*}
\]

That these relations involve inverting matrices of smaller dimension than \( P \) and \( K \) than Equation (12.2.8) if \( R \) is of smaller dimension than \( P \), i.e., if \( p < n \). Equations (a) and (12.2.8) are known as the “matrix inversion pair” (see Problem 4, Section 1.3).

3. Complete the square in Equation (12.2.5) and show that
\[
J = \frac{1}{2} [x - \ddot{x} - PH^TR^{-1}(z - H\ddot{x})]^T P^{-1} [x - \ddot{x} - PH^TR^{-1}(z - H\ddot{x})] \\
+ \frac{1}{2} (z - H\ddot{x})^T R^{-1} (z - H\ddot{x}).
\]

Thus, \( J \) is minimized by choosing \( x = \ddot{x} \), where \( \ddot{x} = \ddot{x} + PH^TR^{-1}(z - H\ddot{x}) \),

is in agreement with Equation (12.2.7).

---

4. Given two correlated Gaussian random vectors \( x \) and \( z \), mean values \( \bar{x}, \bar{z} \) and covariance matrices \( P_{xx}, P_{zz} \), respectively,
and correlation $E[(x - \tilde{x}) (z - \tilde{z})^T] = P_{xz}$, show that the conditional density function $p(x/z)$ is gaussian, with

$$E(x/z) = \tilde{x} + P_{xz} P_{zz}^{-1} (z - \tilde{z}) = \tilde{x},$$

$$E\{[(x - \tilde{x}) (x - \tilde{x})^T]/z\} = P_{xx} - P_{xz} P_{zz}^{-1} P_{xz}^T.$$

Problem 5. In Problem 4, let $z = Hx + v$, where $H$ is a known matrix and $v$ is independent of $x$, with mean value zero and covariance $R$. Let $P_{xz} = M$ and show that

$$P_{zz} = R + MHz^T, \quad P_{xz} = MH^T, \quad \tilde{z} = H\tilde{x}.$$

Using these relations in Problem 4, verify Equations (12.2.7) and (12.2.8) (you will also need the results of Problem 2). Note that $K = P_{xz} P_{zz}^{-1}$, a most reasonable result!

Problem 6. In Problem 4, show that the gaussian random vectors $e = E(x/z) - x$ and $z - \tilde{z}$ are independent; i.e., we have $E[e(z - \tilde{z})] = 0$.

Problem 7. Suppose that the number of theoretical relationships among measured variables $z$ and state variables $x$ is less than the number of measured variables; for example, we have

$$Az = Hx + Av,$$

where

$A$ is a $(q \times p)$-matrix, $q < p$, $H$ is a $(q \times n)$-matrix,

$E(v) = 0$, $E(vv^T) = R$ is a $(p \times p)$-matrix.

Show that the estimation procedure of this section applies with $A$ replaced by $Az$ and $R$ by $ARA^T$.

Problem 8. Consider the usual problem of least square fit, i.e., determining $x$ to minimize

$$J = \frac{1}{2}||z - Hx||^2.$$

Show that the error of the fit $e = z - H\tilde{x}$ is orthogonal to the fit $\tilde{z} = H\tilde{x}$, in the sense $e^T\tilde{z} = 0$.

Problem 9. In Example 2, suppose that initial estimates of an orbit are $a = 6,000$ miles, $\bar{e} = 1/6$, $\bar{\omega} = 0$, $\bar{\alpha} = 0$. Take $A = 2000$ miles, $g = 32.2$ ft sec$^{-2}$, and make an improved estimate of $a, e, \bar{\omega}, \alpha$, using the single measurement

$$z(t) = 16.7 \text{ deg at } t = 1,357 \text{ sec},$$

where

$$E(e^2) = 10^{-2} \text{ (deg^2)}, \quad E(T - T_e) = 0,$$

and all covariances are based on the estimated covariance rather than Equations (11.2.7) and (12.2.7).
Optimal Filtering and Prediction - 359

10^{-2} (deg)^2, \quad E(a - \hat{a})^2 = 10^{-4} (miles)^2, \quad E(e - \hat{e})^2 = 10^{-4},
E(T_{o} - \hat{T}_{o})^2 = 10^2 (sec)^2, \quad E(\omega - \hat{\omega})^2 = 10^{-10} (sec)^{-2},
E(\alpha_{o} - \hat{\alpha}_{o})^2 = 10^{-2} (deg)^2,

all covariances are zero. [HINT: Use Equation (a) of Problem 2, rather than Equation (12.2.8), to find \bar{P}]

Optimal filtering for single-stage linear transitions

Consider a system that makes a discrete transition from state 0 to

\[ x_1 = \Phi x_0 + \Gamma \omega_0, \quad (12.3.1) \]

where \( \Phi \) is a known \((n \times n)\) transition matrix, \( \Gamma \) is a known \((n \times r)\)-

transition vector, \( \omega_0 \) is thus a random vector, with mean \( \bar{\omega}_0 \) and co-

covariance \( \bar{Q}_o \). The state \( x_0 \) is also a random vector, with mean \( \bar{x}_0 \) and co-

covariance \( \bar{P}_o \); that is, we have \( E(x_0) = \bar{x}_0 \), \( E((x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T) = \bar{P}_o \). (12.3.3)

Furthermore, \( x_1 \) and \( \omega_0 \) are independent. From this information, it 

follows from (12.3.5) that, on the average, the effect of the uncertain \( \omega_0 \) in a transition of the type (12.3.1) is to increase the uncertainty in our knowledge of the state \( x_1 \).† This is to be contrasted with the result (12.2.8), where it was shown that measurements, on 

the average, decrease the uncertainty in our knowledge of the state.†† Hence, we may write measurements, as in Section 12.2, after the transition from state 1. Then, from Equations (12.2.7) and (12.2.8), the estimate of \( x_1 \) is given by \( \hat{x}_1 = \hat{x}_1 + P_1 H_1^T R_1^{-1}(z_1 - H_1 \hat{x}_1) \), (12.3.6)
\[ \begin{align*}
P_i &= (M_i^{-1} + H_i^T R_i^{-1} H_i)^{-1} = M_i - M_i H_i^T (H_i M_i H_i^T + R_i)^{-1} H_i M_i, \\
\end{align*} \tag{12.3.7} \]

Here \( \tilde{x}_i \) and \( M_i \) are as given in (12.3.4) and (12.3.5). Note that \( \tilde{x}_i \) is the estimate of \( x_i \) before measurement, whereas \( \hat{x}_i \) is the estimate after measurement. Similarly, \( M_i \) is the error covariance matrix before measurement and \( P_i \) is the error covariance matrix after measurement. Symbolically, we can describe this process as follows:

\[ \begin{align*}
\tilde{\omega}_i & \rightarrow \tilde{x}_i \\
\downarrow & \downarrow \\
\text{mean:} & \quad \hat{x}_i \rightarrow \hat{x}_i, \\
Q_o & \rightarrow R_i \\
\downarrow & \downarrow \\
\text{covariance:} & \quad P_o \rightarrow M_i \rightarrow P_i. 
\end{align*} \]

\[ \text{Sec. 12.4 \cdot Linear prediction for linear multistage processes} \]

12.4 Optimal filtering and prediction for linear multistage processes

Consider the linear, stochastic, multistage process described by

\[ x_{i+1} = \Phi x_i + \Gamma w_i, \quad i = 0, \ldots, N - 1, \tag{12.4.1} \]

where

\[ \begin{align*}
E(x_0) &= \hat{x}_0, \\
E(w_i) &= \tilde{w}_i, \\
E(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T &= M_o, \\
E(w_i - \tilde{w}_i)(w_j - \tilde{w}_j)^T &= Q_i \delta_{ij}, \\
E(w_i - \tilde{w}_i)(x_0 - \hat{x}_0)^T &= 0. 
\end{align*} \tag{12.4.2-12.4.5} \]

Measurements \( z_i \) are made while the system is in stage \( i \), and are linearly related to the state \( x_i \) by

\[ z_i = H_i x_i + v_i, \quad i = 0, \ldots, N, \tag{12.4.6} \]

where

\[ \begin{align*}
E(v_i) &= 0, \\
E(v_i v_j^T) &= R_i \delta_{ij}, \\
E(w_i - \tilde{w}_i) v_j^T &= 0, \quad \text{and} \quad E(x_0 - \hat{x}_0) v_j^T = 0. 
\end{align*} \tag{12.4.7-12.4.10} \]

It is reasonable to expect (see the derivation in Section 12.7 and Chapter 13) that the weighted-least-square or maximum-likelihood estimate of the state \( x_i \), using only the measurements \( z_i \), is given by the sequence of the previous section

\[ \hat{x}_i = \tilde{x}_i + K_i (z_i - H_i \hat{x}_i), \tag{12.4.11} \]

where

\[ \begin{align*}
P_0 &= (M_i^{-1} + H_i^T R_i^{-1} H_i)^{-1} \\
\end{align*} \tag{12.4.12} \]

This is the Kalman filter formula (Kalman, 1960). Notice that \( P_i \) is the covariance of the system \( (12.4.2) \), and \( H_i \) is the difference between measurement \( H_i \hat{x}_i \). The dot represents the ratio between measurements \( R_i \) and the transformation matrix \( H_i \).

Note that:

(a) The propagation equations \( (12.4.1-12.4.4) \) determine \( \hat{x}_i \) and \( P_i \).

(b) The computations \( (12.4.5-12.4.10) \) are only for the control covariance. Thus, the computational burden of the optimal filter is very great.

Prediction of the data, when measurements are not available, say starting with \( \hat{x}_m \), is

\[ \hat{x}_{i+1} = \Phi \hat{x}_i + \Gamma_i w_i, \quad i = 0, \ldots, N, \]

where \( \hat{x}_m \) is obtained from \( (12.4.6) \) and \( \hat{x}_i \) is the estimate of \( x_i \) after measurement. Thus, with the filtering matrix \( K_i \), we consider \( R_i = \infty \) (finite-memory) in \( (12.4.11) \), \( (12.4.12) \), and \( (12.4.15) \).

The case in which \( w_i \) and \( v_j \) are correlated is considered in Chapter 13.
given by the sequential use of the single-stage estimation procedure of the previous section:

\[ \hat{x}_i = \hat{x}_i + K_i(z_i - H_i\hat{x}_i), \quad (i = 0, \ldots, k, \text{ where } k \leq N). \quad (12.4.11) \]

\[ \hat{x}_{i+1} = \Phi_i\hat{x}_i + \Gamma_i\hat{w}_i, \quad \hat{x}_o \text{ given.} \quad (12.4.12) \]

\[ K_i = P_iH_i^TR_i^{-1}, \quad (12.4.13) \]

\[ P_i = (M_i^{-1} + H_i^TR_i^{-1}H_i)^{-1} = M_i - M_iH_i^TR_i^{-1}H_iM_i, \quad (12.4.14) \]

\[ M_{i+1} = \Phi_iP_i\Phi_i^T + \Gamma_iQ_i\Gamma_i^T. \quad (12.4.15) \]

This is the Kalman filter for linear multistage processes (see Rosenbrock, 1960). Note that the filter (12.4.11) and (12.4.12) is a model of the system (12.4.1), with a correction term proportional to the difference between the actual measurement \( z_i \) and the predicted measurement \( H_i\hat{x}_i \). The proportionality matrix \( K_i \) in (12.4.13) is essentially the ratio between uncertainty in the state \( P_i \) and the uncertainty in the measurements \( R_i \); the matrix \( H_i \) is simply the state-to-measurement information matrix of (12.4.7).

The propagation of the covariance of the error of the estimate, equations (12.4.14) and (12.4.15), is independent of the measurements \( z_i \). Thus, the covariance matrix can be computed beforehand and updated if the parameters of the system and the observation equation are given.

The computation of the updated estimate, Equations (12.4.11) and (12.4.12), involves only the current measurement and error variance. Thus, it can easily be carried out in real time.

The prediction of the state beyond the stage where measurements are available, say state \( m \), can be done only by repeated use of (12.4.12);

\[ \hat{x}_{i+1} = \hat{x}_{i+1} = \Phi_i\hat{x}_i + \Gamma_i\hat{w}_i; \quad i = m, m + 1, \ldots, \quad (12.4.16) \]

\[ \hat{x}_{m+1} \]

is obtained from the filter (12.4.11) through (12.4.15). In other words, the best prediction we can make uses the expected value of \( \hat{x}_{m} \), namely, \( \hat{x}_{m} \), in the transition relations (12.4.1), starting, however, with the filtering estimate of \( \hat{x}_m \). Another way of seeing this is to consider \( R_i = \infty \) for \( i = m, m + 1, \ldots \). In this case, (12.4.14) and (12.4.12) reduce to

\[ P_{i+1} = \Phi_iP_i\Phi_i^T + \Gamma_iQ_i\Gamma_i^T, \quad (12.4.17) \]

\[ \hat{x}_{m} \]

in (12.4.12) is, of course, to be understood as \( E(x_i/z_i, \ldots, z_m) \) and not