12 Optimal filtering and prediction

12.1 Introduction

In the previous chapters we have seen that optimal control of a dynamic system requires a knowledge of the state of that system. In practice, the individual state variables cannot be determined exactly by direct measurements; instead, we usually find that the measurements that can be made are *functions* of the state variables and that these measurements contain random errors. The system itself may also be subjected to random disturbances. In many cases, we have too few measurements at a given time to infer the state variables at that time, even if the measurements were quite precise. On occasion, we have more than enough measurements, so that the state variables are overdetermined. Thus, we are faced with the problem of making good *estimates* of the state variables from either too few or too many measurements, which are imprecise and only functions of the state variables, knowing, too, that the system itself is subjected to random disturbances.

If we believe that we understand the dynamics of the ideal system (with perfect and complete measurements and no random disturbances), and if we believe that we have some knowledge of the degree of uncertainty in the measurements and of the degree of intensity of the random disturbances to the system, then, on the basis of all the measurements up to the present time, we can determine the most likely values of the state variables. The process of determining these most likely values is called smoothing, filtering, or prediction, depending on whether we are finding past, present, or future values of the state variables, respectively. In this chapter, the filtering and prediction problems are treated. The results will be directly applicable to stochastic control problems. The smoothing problem is dealt with in Chapter 13.

have seen that optimal
owledge of the state of
variables cannot be
tead, we usually find that
functions of the state
random errors. The
disturbances. In many
n time to infer the state
ents were quite precise.
surements, so that the
re faced with the problem
bles from either too
mprecise and only function
the system itself is subjected

erstand the dynamics of the
measurements and no
t we have some knowledge
rements and of the degree
the system, then, on the
esent time, we can dete
iables. The process of de
smoothing, filtering, or
finding past, present, or
ively. In this chapter, the
ated. The results will be
oblems. The smoothing

12.2 Estimation of parameters using weighted least-squares

Suppose that we wish to estimate the n-component state vector x of a static system using the p-component measurement vector, z, containing random errors, v, which are independent of the state x, where

$$z = Hx + v, \tag{12.2.1}$$

and

$$H = \text{a known } \ (p \times n)\text{-matrix},$$

$$E(v) = 0\,, \tag{12.2.2}$$

$$E(vv^T) = R\,, \text{a known } \ (p \times p) \ \text{ positive matrix}. \tag{12.2.3}$$

Let us also suppose that we had an estimate of the state, before the measurements were made, which we will call $\bar{x}$, where

$$E[(x - \bar{x})(x - \bar{x})^T] = M\,, \text{a known } \ (n \times n) \ \text{ positive matrix}. \tag{12.2.4}$$

One very reasonable estimate of x, taking into account the measurements, z, is the *weighted-least-squares estimate*, which we shall call $\hat{x}$. For this estimate we choose $\hat{x}$ as the value of x that minimizes the quadratic form

$$J = \tfrac{1}{2}[(x - \bar{x})^T M^{-1}(x - \bar{x}) + (z - Hx)^T R^{-1}(z - Hx)]\,. \tag{12.2.5}$$

Note that the weighting matrixes, M^{-1} and R^{-1}, are the inverse matrices of the prior expected values of $(x - \bar{x})(x - \bar{x})^T$ and $(z - Hx)(z - Hx)^T$, respectively. With this choice of weighting matrices, the weighted-least-squares estimate is identical to the conditional-expected-value estimate assuming gaussian distributions of x and v; that is, we have $\hat{x} = E(x/z)$, which, in turn, is identical to the maximum-likelihood or minimum-variance estimate (see Problems 1 through 6 at the end of this section).

To determine $\hat{x}$, consider the differential of (12.2.5):†

$$dJ = dx^T[M^{-1}(x - \bar{x}) - H^T R^{-1}(z - Hx)]\,. \tag{12.2.6}$$

In order that $dJ = 0$ for arbitrary dx^T, the coefficient of dx^T in (12.2.6) must vanish:

$$(M^{-1} + H^T R^{-1} H)\hat{x} = M^{-1}\bar{x} + H^T R^{-1} z$$
$$= (M^{-1} + H^T R^{-1} H)\bar{x} + H^T R^{-1}(z - H\bar{x})$$

or

$$\hat{x} = \bar{x} + PH^T R^{-1}(z - H\bar{x})\,, \tag{12.2.7}$$

† A better elementary derivation of $\hat{x}$ can be made by completing the square in Equation (12.2.5), since $\hat{x}$ can then be determined by inspection, see Problem 3.

where

$$P^{-1} = M^{-1} + H^T R^{-1} H\,. \tag{12.2.8}$$

The quantity P in Equation (12.2.8) is the covariance matrix of the error in the estimate $\hat{x}$; that is, we have

$$P = E[(\hat{x} - x)(\hat{x} - x)^T]\,. \tag{12.2.9}$$

To show this, let

$$e = \hat{x} - x = \text{error in the estimate.}$$

Then we have

$$e = \bar{x} - x + \hat{x} - \bar{x} = \bar{x} - x + K[v - H(\bar{x} - x)]\,,$$

where

$$K = PH^T R^{-1}\,,$$

or

$$e = (I - KH)(\bar{x} - x) + Kv\,. \tag{12.2.10}$$

Since $\bar{x} - x$ and v are independent, it follows from (12.2.10) that

$$E(ee^T) = (I - KH)M(I - KH)^T + KRK^T\,. \tag{12.2.11}$$

Premultiplying (12.2.8) by P and postmultiplying by M, we have

$$M = P + PH^T R^{-1} HM$$

or

$$(I - KH)M = P\,. \tag{12.2.12}$$

Substituting (12.2.12) into (12.2.11), we have

$$E(ee^T) = P - PH^T K^T + KRK^T = P - PH^T R^{-1} HP + PH^T R^{-1} HP$$

or

$$E(ee^T) = P\,. \tag{12.2.13}$$

Since M is the error covariance matrix *before* measurement, it is apparent from (12.2.8) that P, the error covariance matrix after measurement is never larger than M, since $H^T R^{-1} H$ is at least a positive-semidefinite matrix.† Thus, the act of measurement, on the *average*, decreases (more precisely, it never increases) the uncertainty in our knowledge of the state x.

Another noteworthy property of the estimate is the fact that

---

†The matrix P is said to be smaller than M if, for all nonzero vectors x, the scalar quantity $x^T P x < x^T M x$.

$$E(e\hat{x}^T) = E[(I - KH)(\bar{x} - x) + Kv][\bar{x} - KH(\bar{x} - x) + Kv]^T$$
$$= -(I - KH)MH^TK^T + KRK^T = -PH^TK^T + KRK^T = 0\,; \qquad (12.2.14)$$

Thus, the estimate and the error of the estimate are uncorrelated. In the case where x and v are gaussian, this implies that $\hat{x}$ and e are independent (see Chapter 10). We may regard (12.2.14) as a *definition* of optimal estimate in the sense that it contains all the information in z, and no improvement in e can be obtained by knowledge of z.

From (12.2.5), the prior expected value of J is $(n + p)/2$. This is easily shown, since (12.2.5) can be written as

$$J = \tfrac{1}{2}tr[M^{-1}(x - \bar{x})(x - \bar{x})^T] + \tfrac{1}{2}tr[R^{-1}(z - Hx)(z - Hx)^T]\,,$$

where $tr(\;)$ means "trace of ( )"; i.e., the sum of the elements on the principal diagonal of ( ). Then we have

$$E(J) = \tfrac{1}{2}tr\{M^{-1}E[(x - \bar{x})(x - \bar{x})^T]\} + \tfrac{1}{2}tr\{R^{-1}E[(z - Hx)(z - Hx)^T]\}$$
$$= \tfrac{1}{2}tr(M^{-1}M) + \tfrac{1}{2}tr(R^{-1}R)\,.$$

But $M^{-1}M$ is an $(n \times n)$ identity matrix and $R^{-1}R$ is a $(p \times p)$ identity matrix, so we have $tr(M^{-1}M) = n$, $tr(R^{-1}R) = p$. Hence we have

$$E(J) = \tfrac{1}{2}(n + p)\,.$$

The prior knowledge of M and R is sometimes vague and uncertain. As a check, after the estimation process has been completed, the actual value of J in Equation (12.2.5) should be computed with $x = \hat{x}$; call it J_o. The prior expected value of J_o can be calculated and is found to be $E(J_o) = p/2$. If the actual J_o is not close to $p/2$, the elements of M and R should be multiplied by the scale factor $J_o/p/2$, which adjusts the value of $J_o/p/2$. Note that this also multiplies P by the same scale factor [see Equation (12.2.8)], but it does *not* change K [in Equation (12.2.7) the scale factor appears in both P and R and, hence, cancels out]. For this reason it is only necessary to establish the relative magnitude of the elements in M and R; the scale factor can then be applied *post facto* to obtain values of P, M, and R consistent with the scatter in the data.

Many estimation problems are *nonlinear* rather than linear; i.e., instead of (12.2.1), we have

$$z = h(x) + v\,,$$

where $h(x)$ is a known nonlinear function of x. In this case, we may apply the foregoing technique to the linearized version of $z = h(x) + v$:

$$dz \triangleq z - \bar{z} \cong \left.\frac{\partial h}{\partial x}\right|_{x=\bar{x}} (x - \bar{x}) + v \triangleq \frac{\partial h}{\partial x}dx + v\,.$$

This is illustrated in the two examples below.

In some estimation problems, the relationship between the parameters to be estimated and the available measurements is known only implicitly; i.e., we may not be able to write down explicitly the relationship $z(t) = h(x,v,t)$. On the other hand, we may still be able to determine the differential relationship $dz = (\partial h/\partial x)\, dx + v$ directly and solve the linearized estimation problem.

Also, by appropriate formulation, some dynamic estimation problems can be reduced to parameter estimation problems. Example 2 illustrates this point.

Example 1. *Position estimation from angle measurements.* We wish to estimate the location (x,y) of a point A in a plane by angle measurements z_i from several points B_i $(i = 1,2,\ldots,n)$ on a base line (see Figure 12.2.1).

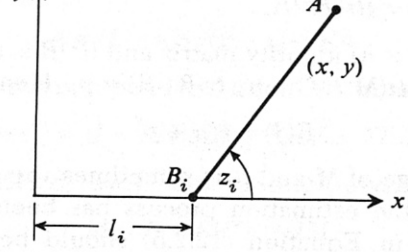


Figure 12.2.1. Position estimation using angle measurements on a base line.

The angle measurements z_i are related to the location of A and B_i by the *nonlinear* relations

$$z_i = \tan^{-1}\frac{y}{x - \ell_i} + v_i\,, \qquad (12.2.15)$$

where v_i is a random error made in the angle measurement. We assume that

$$E(v_i) = 0\,, \qquad E(v_i v_j) = \begin{cases} r_i\,, & i = j\,,\\ 0\,, & i \neq j\,. \end{cases} \qquad (12.2.16)$$

We now *linearize* (12.2.15) about a prior estimate of (x,y), which we will call $(\bar{x},\bar{y})$:

$$z_i - \bar{z}_i = [H_{1i} H_{2i}] \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} + v_i\,,$$

where

$$\bar{z}_i = \tan^{-1}\frac{\bar{y}}{\bar{x} - \ell_i};$$

$$H_{1i} = \left(\frac{\partial z_i}{\partial x}\right)_{x=\bar{x},\, y=\bar{y}}, \qquad H_{2i} = \left(\frac{\partial z_i}{\partial y}\right)_{x=\bar{x},\, y=\bar{y}}.$$

We will let the estimate $\hat{x},\hat{y}$ be independent of prior data by considering the error covariance matrix to be infinite; i.e., we have $M \to \infty$ or $M^{-1} = 0$. Let

$$dz = \begin{bmatrix} z_1 - \bar{z}_1 \\ \cdot \\ \cdot \\ \cdot \\ z_n - \bar{z}_n \end{bmatrix}, \quad H = \begin{bmatrix} H_{1i}, H_{2i} \\ \cdot \\ \cdot \\ \cdot \\ H_{1n}, H_{2n} \end{bmatrix}, \quad dx = \begin{bmatrix} x - \bar{x} \\ \\ y - \bar{y} \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{bmatrix}.$$

Then we have

$$dz = H\,dx + v,$$

where

$$R = \begin{bmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_n \end{bmatrix}, \qquad P = (H^T R^{-1} H)^{-1}.$$

If $d\hat{x}$ is markedly different from zero, we should repeat the procedure, linearizing about $(\hat{x},\hat{y}) = (\bar{x} + d\hat{x}, \bar{y} + d\hat{y})$.

In principle, we should repeat the procedure until $d\hat{x} \cong 0$. The eigenvalues and eigenvectors of P determine the 39 percent likelihood ellipse around $(\hat{x},\hat{y})$.

Numerical example. Suppose that $n = 3$ and the data are

i	1	2	3	Units
ℓ_i	0	500	1,000	ft
z_i	30.1	45.0	73.6	deg
r_i	.01	.01	.04	deg².

From a rough graph (see Figure 12.2.2), we estimate $\bar{x} = 1{,}210$ ft, $\bar{y} = 700$ ft. From this, it follows that

$$dz = \begin{bmatrix} .05 \\ .40 \\ .29 \end{bmatrix} \text{deg}, \qquad H = \begin{bmatrix} -.0205, .0354 \\ -.0403, .0409 \\ -.0751, .0225 \end{bmatrix} \text{deg ft}^{-1},$$

$$P = [H^T R^{-1} H]^{-1} = \begin{bmatrix} 11.26, 10.31 \\ 10.31, 12.75 \end{bmatrix} \text{ft}^2 ,$$

$$K = PH^T R^{-1} = \begin{bmatrix} 13.4, -3.1, -15.3 \\ 24.0, \ 10.6, -12.2 \end{bmatrix} \text{ft(deg)}^{-1} ,$$

$$d\hat{x} = K\, dz = \begin{bmatrix} -5.01 \\ +1.89 \end{bmatrix}, \qquad \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} 1205.0 \\ 701.9 \end{bmatrix}.$$

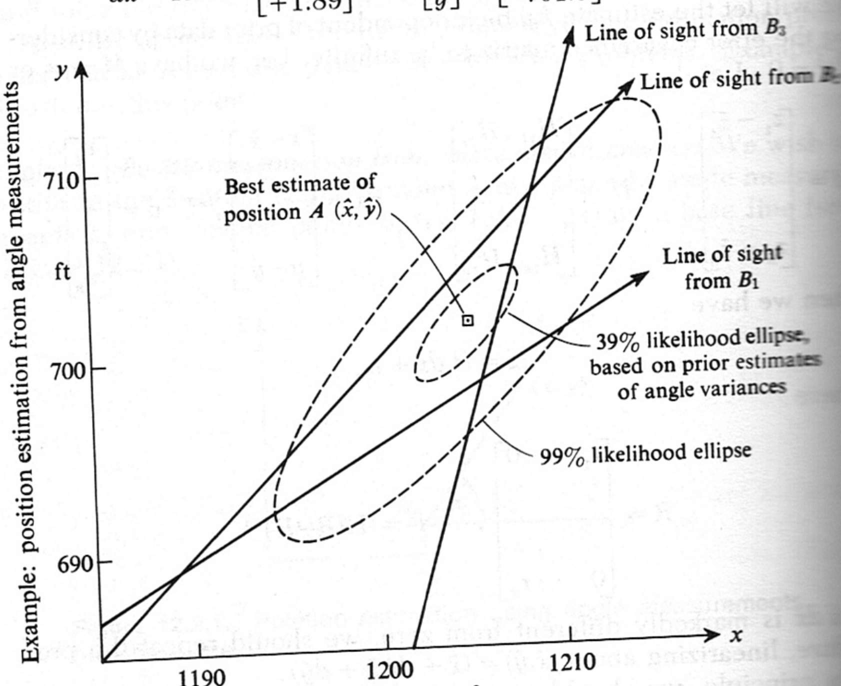


Figure 12.2.2. Numerical example of position estimation by angle measurements.

The eigenvalues and eigenvectors of P are found to be

Eigenvalue	Eigenvector direction
22.34 ft²	47.0°
1.66 ft²	−43.0° .

Using the square root of the eigenvalues (4.72 ft and 1.29 ft) as semi-axes, measured along the eigenvectors, we can sketch the 39 percent likelihood ellipse with center at $(\hat{x},\hat{y})$ (see Figure 12.2.2). The 99 percent likelihood ellipse is three times the size of the 39 percent ellipse in linear dimension. The lines of sight from B_1, B_2, and B_3 are also shown.

Note that we ha

$(\hat{z}_1 -$

$(\hat{z}_2 -$

$(\hat{z}_3 -$

Hence, we have

$\frac{1}{2}\sum_{i=1}^{3}$

The prior expecte
the limited sample
gle variances by t
39 percent likeliho

Example 2. *Orbit es*
elliptic orbit in a
are known (see Fig

a = semi-majo

T_o = time of pe

θ_o = angle betw

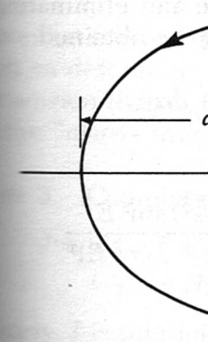


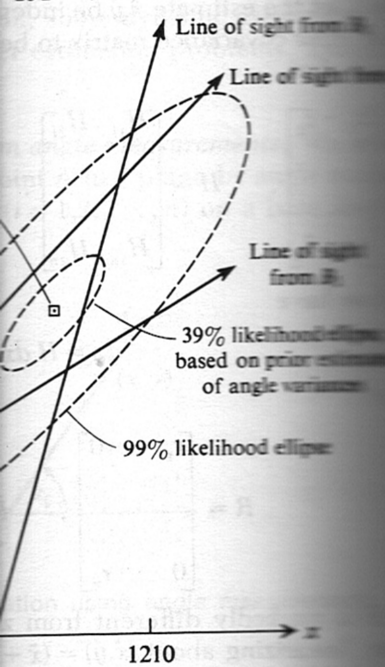
Figure 12.2.3.
sensor measure

A satellite is eq
(a) a cylinder ro

$$\begin{bmatrix} 6, 10.31 \\ 1, 12.75 \end{bmatrix} \text{ft}^2,$$

$$\begin{bmatrix} .1, -15.3 \\ .6, -12.2 \end{bmatrix} \text{ft(deg)}^{-1},$$

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} 1205.0 \\ 701.9 \end{bmatrix}.$$



xample of position estimation

rs of P are found to be

Eigenvector direction
47.0°
−43.0°.

envalues (4.72 ft and 1.29 ft
ectors, we can sketch the
t $(\hat{x},\hat{y})$ (see Figure 12.2.2).
ree times the size of the
e lines of sight from B

Note that we have

$$(\hat{z}_1 - z_1)^2 = (30.22 - 30.1)^2 = (.12)^2 = .0144,$$
$$(\hat{z}_2 - z_2)^2 = (44.88 - 45.0)^2 = (.12)^2 = .0144,$$
$$(\hat{z}_3 - z_3)^2 = (73.73 - 73.6)^2 = (.13)^2 = .0169.$$

Hence, we have

$$\frac{1}{2}\sum_{i=1}^{3}\frac{(\hat{z}_i - z_i)^2}{r_i} = \frac{1.44 + 1.44 + .42}{2} = \frac{3.30}{2}.$$

The prior expected value of this quantity was 3.00/2; thus, based on the limited sample of three measurements, we might scale up the angle variances by the factor 3.30/3 = 1.10. This would scale up the percent likelihood ellipse by the factor $\sqrt{1.10} = 1.05$.

Example 2. *Orbit estimation from horizon sensor measurements.* An elliptic orbit in a plane is specified if the following four parameters are known (see Figure 12.2.3):

a = semi-major axis of ellipse, e = eccentricity of ellipse,

T_p = time of perigee passage,

ϕ_p = angle between perigee and a reference line.

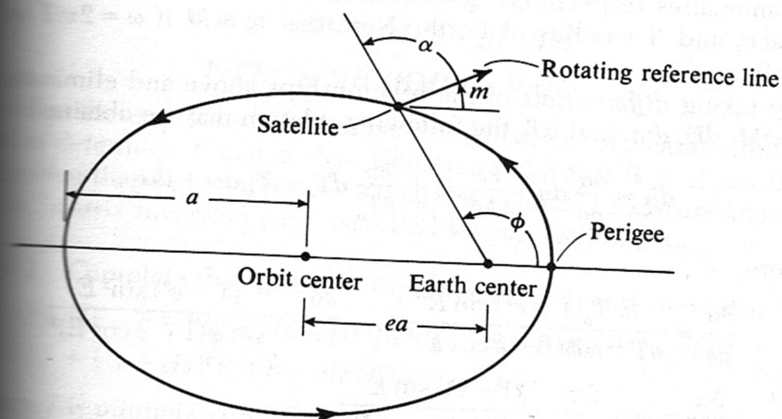


Figure 12.2.3. Nomenclature for orbit estimation using horizon sensor measurements.

The satellite is equipped with a measurement system consisting of a cylinder rotating about an axis perpendicular to the orbital