

Chapter 3

LINEAR ESTIMATION IN STATIC SYSTEMS

3.1 INTRODUCTION

3.1.1 Outline

This chapter presents the minimum mean square error (MMSE) estimation of Gaussian random vectors (Section 3.2) and the *linear* MMSE estimator for arbitrarily distributed random vectors (Section 3.3). The latter is the estimator constrained to have a linear form. The estimation of unknown constant vectors according to the least squares (LS) criterion is then discussed in Section 3.4, where both the batch and recursive versions are derived. These results are then applied to polynomial fitting in Section 3.5.

Section 3.6 presents the statistical tools for deciding what is the appropriate order of the polynomial when fitting a set of data points. This latter method is especially important because, in practice, the models — for instance, for the motion of targets — are not known *a priori* and have to be inferred from the data.

A realistic example that deals with the localization of a target — a source of radiated energy — based on bearings-only measurements from a platform is presented in Section 3.7. This illustrates the use of nonlinear LS to a practical problem.

3.1.2 Linear Estimation in Static Systems — Summary of Objectives

Present

- MMSE estimation of Gaussian random vectors
- Linear MMSE estimator for arbitrarily distributed random vectors
- LS estimation of unknown constant vectors from *linear* observations

- batch form
- recursive form.

Apply the LS technique to

- Polynomial fitting
- Choice of order of the polynomial when fitting a set of data points.

Illustrate the use of *nonlinear* LS to a practical problem — target (source of radiated energy) localization from bearings-only measurements.

3.2 ESTIMATION OF GAUSSIAN RANDOM VECTORS

3.2.1 The Conditional Mean and Covariance for Gaussian Random Vectors

Consider two random vectors x and z that are *jointly* normally (Gaussian) distributed.

Define the *stacked vector*

$$y \triangleq \begin{bmatrix} x \\ z \end{bmatrix} \quad (3.2.1-1)$$

The notation

$$y \sim \mathcal{N}[\bar{y}, P_{yy}] \quad (3.2.1-2)$$

will indicate that the variable y is ***normally (Gaussian) distributed*** with mean

$$\bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix} \quad (3.2.1-3)$$

and covariance matrix (assumed nonsingular)

$$P_{yy} = \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix} \quad (3.2.1-4)$$

written in partitioned form, where \bar{x} is the mean of x ,

$$P_{xx} = E[(x - \bar{x})(x - \bar{x})'] \quad (3.2.1-5)$$

is its covariance, and

$$P_{xz} = E[(x - \bar{x})(z - \bar{z})'] = P'_{zx} \quad (3.2.1-6)$$

is the covariance between x and z (also called ***cross-covariance***), and so on.

In the above, x is the random variable to be estimated and z is the ***measurement*** or the ***observation***. As shown in (2.4.1-7), the ***estimate*** of the random variable x in terms of z according to the ***minimum mean square***

error (MMSE) criterion — the **MMSE estimator** — is the *conditional mean* of x given z .

For x and z *jointly Gaussian*, as assumed in (3.2.1-2), the *conditional mean* is

$$\hat{x} \triangleq E[x|z] = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z}) \quad (3.2.1-7)$$

and the corresponding **conditional covariance matrix** is

$$P_{xx|z} \triangleq E[(x - \hat{x})(x - \hat{x})'|z] = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx} \quad (3.2.1-8)$$

(see Subsection 1.4.14 for proof).

This follows from the fact that the conditional pdf of x given z is Gaussian with mean (3.2.1-7) and covariance (3.2.1-8).

Note that the optimal estimator (in the MMSE sense) of x in terms of z is a *linear function* of z . This is a consequence of the Gaussian assumption.

Another important property specific to this Gaussian problem is that the *conditional covariance* (3.2.1-8), which measures the “quality” of the estimate, is *independent of the observation z* .

3.2.2 Estimation of Gaussian Random Vectors — Summary

The *MMSE estimate* — the *conditional mean* — of a *Gaussian* random vector in terms of another *Gaussian* random vector (the measurement) is a *linear combination* of

- The prior (unconditional) mean of the variable to be estimated;
- The difference between the measurement and its prior mean.

The *conditional covariance* of one *Gaussian* random vector given another *Gaussian* random vector (the measurement) is *independent* of the measurement.

Both of the above properties hinge strictly on the assumption that the two random vectors under consideration are *jointly Gaussian*.

3.3 LINEAR MINIMUM MEAN SQUARE ERROR ESTIMATION

3.3.1 The Principle of Orthogonality

The *minimum mean square error (MMSE) estimate* of a random variable x in terms of another random variable z is, according to (2.4.1-7), the conditional mean $E[x|z]$.

In many problems the distributional information needed for the evaluation of the conditional mean is not available. Furthermore, even if it were available, the evaluation of the conditional mean could be prohibitively complicated.

In view of this, a method that (1) is simple — yields the estimate as a linear function of the observation(s) and (2) requires little information — only first and second moments, is highly desirable. Such a method, called *linear MMSE estimation*, relies on the *principle of orthogonality*:

The best linear estimate (in the sense of MMSE) of a random variable in terms of another random variable — the observation(s) — is such that

1. The estimate is unbiased — the estimation error has mean zero, and
2. The estimation error is uncorrelated from the observation(s);

that is, they are *orthogonal*.

Linear MMSE Estimation for Zero-Mean Random Variables

The linear MMSE estimation can be formulated in terms of a (normed linear) space of random variables as follows.

The set of real-valued scalar *zero-mean random variables* $z_i, i = 1, \dots, n$, can be considered as *vectors in an abstract vector space* or *linear space*. Such a space is closed under addition of its elements and multiplication by scalars — the linear combination of two random variables is another element in this space.

A (complete) vector space in which one defines an *inner product* is a Hilbert space. The inner product that can be defined is

$$\langle z_i, z_k \rangle = E[z_i z_k] \quad (3.3.1-1)$$

Since the random variables under consideration are zero mean, it is clear that

$$\langle z_i, z_i \rangle = E[z_i^2] = \|z_i\|^2 \quad (3.3.1-2)$$

satisfies the properties of a *norm* and can be taken as such.

With this definition of the norm, *linear dependence* is defined by stating that the norm of a linear combination of vectors is zero

$$E \left[\left(\sum_{i=1}^m \alpha_i z_i \right)^2 \right] = 0 \quad (3.3.1-3)$$

If $\alpha_1 \neq 0$, then z_1 is a linear combination of z_2, \dots, z_m , namely,

$$z_1 = -\frac{1}{\alpha_1} \sum_{i=2}^m \alpha_i z_i \quad (3.3.1-4)$$

that is, it is an element of the *subspace* spanned by z_2, \dots, z_m . Two vectors are *orthogonal*, denoted as $z_i \perp z_k$, if and only if

$$\langle z_i, z_k \rangle = 0 \quad (3.3.1-5)$$

which is equivalent to these zero-mean random variables being *uncorrelated*.

The **linear MMSE estimator** of a zero-mean random variable x in terms of z_i , $i = 1, \dots, n$, is given by

$$\hat{x} = \sum_{i=1}^n \beta_i z_i \quad (3.3.1-6)$$

and has to be such that the norm of the **estimation error**

$$\tilde{x} \triangleq x - \hat{x} \quad (3.3.1-7)$$

is minimum. The linear MMSE estimate is denoted also by a circumflex (“hat”), even though it is not the conditional mean as in (3.2.1-7).

Thus the norm of the estimation error

$$\|\tilde{x}\|^2 = E[(x - \hat{x})^2] = E\left[\left(x - \sum_{i=1}^n \beta_i z_i\right)^2\right] \quad (3.3.1-8)$$

will have to be minimized with respect to β_i , $i = 1, \dots, n$.

Setting the derivative of (3.3.1-8) with respect to β_k to zero

$$-\frac{1}{2} \frac{\partial}{\partial \beta_k} \|\tilde{x}\|^2 = E\left[\left(x - \sum_{i=1}^n \beta_i z_i\right) z_k\right] = E[\tilde{x} z_k] = \langle \tilde{x}, z_k \rangle = 0 \quad k = 1, \dots, n \quad (3.3.1-9)$$

is seen to be equivalent to requiring the following *orthogonality* property:

$$\tilde{x} \perp z_k \quad \forall k \quad (3.3.1-10)$$

This is the **principle of orthogonality**: In order for the error to have minimum norm, it has to be *orthogonal to the observations*. This is equivalent to stating that the estimate \hat{x} has to be the **orthogonal projection** of x into the space spanned by the observations, as illustrated in Fig. 3.3.1-1.

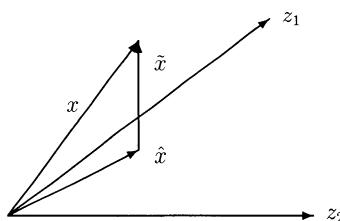


Figure 3.3.1-1: Orthogonal projection of random variable x into the subspace of $\{z_1, z_2\}$.

Linear MMSE Estimation for Nonzero-Mean Random Variables

For a random variable x with nonzero mean \bar{x} , the best linear estimator is of the form (an affine function)

$$\hat{x} = \beta_0 + \sum_{i=1}^n \beta_i z_i \quad (3.3.1-11)$$

Since the MSE is the sum of the square of the mean and the variance

$$E[\hat{x}^2] = (E[\tilde{x}])^2 + \text{var}(\tilde{x}) \quad (3.3.1-12)$$

in order to minimize it, the estimate should have the **unbiasedness property**:

$$E[\tilde{x}] = 0 \quad (3.3.1-13)$$

It can be easily shown that this follows from the fact that β_0 enters only into the mean of the error (3.3.1-13), whose minimum norm is zero. Equation (3.3.1-13) yields

$$\beta_0 = \bar{x} - \sum_{i=1}^n \beta_i \bar{z}_i \quad (3.3.1-14)$$

where

$$\bar{z}_i = E[z_i] \quad (3.3.1-15)$$

Inserting (3.3.1-14) into (3.3.1-11) leads to

$$\hat{x} = \bar{x} + \sum_{i=1}^n \beta_i (z_i - \bar{z}_i) \quad (3.3.1-16)$$

The error corresponding to the estimate (3.3.1-16) is

$$\begin{aligned} \tilde{x} &\triangleq x - \hat{x} \\ &= x - \bar{x} - \sum_{i=1}^n \beta_i (z_i - \bar{z}_i) \end{aligned} \quad (3.3.1-17)$$

This has transformed the nonzero-mean case into the zero-mean case.

The orthogonality principle (3.3.1-10) then yields the coefficients β_i from the following equations:

$$\langle \tilde{x}, z_k \rangle = E[\tilde{x} z_k] = E\left[\left[x - \bar{x} - \sum_{i=1}^n \beta_i (z_i - \bar{z}_i)\right] z_k\right] = 0 \quad k = 1, \dots, n \quad (3.3.1-18)$$

The estimator (3.3.1-16) is also known as the **best linear unbiased estimator**.

3.3.2 Linear MMSE Estimation for Vector Random Variables

Consider the vector-valued random variables x and z , which are not necessarily Gaussian or zero-mean. The “best linear” estimate of x in terms of z is obtained as follows. The criterion for “best” is the MMSE; that is, find the estimator

$$\hat{x} = Az + b \quad (3.3.2-1)$$

that minimizes the *scalar MSE criterion*, which in the multidimensional case is the *expected value of the squared norm* of the estimation error,

$$J \triangleq E[(x - \hat{x})'(x - \hat{x})] \quad (3.3.2-2)$$

According to the previous discussion, the *linear MMSE estimator* is such that the estimation error

$$\tilde{x} = x - \hat{x} \quad (3.3.2-3)$$

is *zero-mean* (the estimate is unbiased) and *orthogonal to the observation z* . In other words, the estimate \hat{x} is the orthogonal projection of the vector x into the space spanned by the (random components of the) observation vector z . The unbiasedness requirement (3.3.1-13) is

$$E[\tilde{x}] = \bar{x} - (A\bar{z} + b) = 0 \quad (3.3.2-4)$$

and it yields

$$b = \bar{x} - A\bar{z} \quad (3.3.2-5)$$

The estimation error is then

$$\tilde{x} = x - \bar{x} - A(z - \bar{z}) \quad (3.3.2-6)$$

The orthogonality requirement is, in the multidimensional case, that *each component* of \tilde{x} be orthogonal to *each component* of z .

The orthogonality requirement can thus be written as

$$E[\tilde{x}z'] = E\{[x - \bar{x} - A(z - \bar{z})]z'\} \quad (3.3.2-7)$$

$$= E\{[x - \bar{x} - A(z - \bar{z})](z - \bar{z})'\} = P_{xz} - AP_{zz} = 0 \quad (3.3.2-8)$$

The subtraction of \bar{z} from z in the transition from (3.3.2-7) to (3.3.2-8) could be done in view of the property (3.3.2-4) that \tilde{x} is zero mean.

The solution for the **weighting matrix A** is thus

$$A = P_{xz}P_{zz}^{-1} \quad (3.3.2-9)$$

The existence of the above requires the invertibility of P_{zz} , i.e., no linear dependence between the measurements (or, equivalently, no redundant measurements). Combining (3.3.2-5) and (3.3.2-9) yields the expression of the **linear MMSE estimator for the multidimensional case** as

$$\boxed{\hat{x} = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z})} \quad (3.3.2-10)$$

which is *identical* to the conditional mean (3.2.1-7) from the Gaussian case.

The **matrix MSE** corresponding to (3.3.2-10) is given by

$$E[\tilde{x}\tilde{x}'] = E[[x - \bar{x} - P_{xz}P_{zz}^{-1}(z - \bar{z})][x - \bar{x} - P_{xz}P_{zz}^{-1}(z - \bar{z})]'] \quad (3.3.2-11)$$

This becomes, after simple manipulations,

$$\boxed{E[\tilde{x}\tilde{x}'] = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx} = P_{xx|z}} \quad (3.3.2-12)$$

that is, *identical* expression to the conditional covariance (3.2.1-8) in the Gaussian case. Note, however, that, strictly speaking, the matrix MSE (3.3.2-12) is not a covariance matrix since (3.3.2-10) is not the conditional mean.

Equations (3.3.2-10) and (3.3.2-12) are the **fundamental equations of linear estimation**.

Remarks

Note the distinction between the scalar MSE criterion (3.3.2-2), an inner product, and the matrix MSE (3.3.2-12), an outer product. The **matrix MSE** is sometimes called, with abuse of language, a **covariance matrix**.

From the above derivations it follows that

- the *best estimator* (in the MMSE sense) for *Gaussian random variables*

is identical to

- the *best linear estimator* for *arbitrarily distributed random variables with the same first- and second-order moments*.

The linear estimator (3.3.2-10) is the overall best if the random variables are Gaussian; otherwise, it is only the *best within the class of linear estimators*.

Gaussian Assumption as “Worst Case” in MMSE Estimation

The following statement can be made: From the point of view of MMSE estimation, one can view the **Gaussian assumption as the worst case**:

- If the random variables are Gaussian, the minimum achievable matrix MSE is (3.2.1-8).
- If they are not Gaussian, but with the same first two moments, one can achieve (with the linear estimator) the matrix MSE (3.3.2-12), which is the same as (3.2.1-8).

However, in the non-Gaussian case,

The conditional mean (if one can compute it), being the *absolute best* as opposed to the *best within the class of linear estimators*, would give a matrix MSE (covariance matrix in this case) *less or equal* to (3.3.2-12).

On the Terminology

The LMMSE estimator is also referred to in the literature as the *least mean square (LMS)* or *minimum variance (MV)* or *least squares (LS)*.

3.3.3 Linear MMSE Estimation — Summary

The *linear MMSE estimator* of one random vector in terms of another random vector (the measurement) is such that the estimation error is

1. Zero-mean (the estimate is unbiased)
2. Uncorrelated from the measurements

These two properties imply that the error is orthogonal to the measurements. This is the *principle of orthogonality*.

The expression of the *linear MMSE estimator* is identical to the expression of the *conditional mean* of Gaussian random vectors if they have the same first two moments.

Similarly, the *matrix MSE* associated with the LMMSE estimator has the same expression as the *conditional covariance* in the Gaussian case.

The *linear MMSE estimator* is

1. The *overall best* if the random variables are *Gaussian*
2. The *best* within the class of *linear* estimators otherwise (i.e., there is room for improvement in this case¹)

3.4 LEAST SQUARES ESTIMATION

3.4.1 The Batch LS Estimation

In the *linear least squares (LS)* problem it is desired to estimate the n_x -vector x , modeled as an *unknown constant*, from the linear observations (n_z -vectors)

$$z(i) = H(i)x + w(i) \quad i = 1, \dots, k \quad (3.4.1-1)$$

such as to minimize the quadratic error

$$J(k) = \sum_{i=1}^k [z(i) - H(i)x]' R(i)^{-1} [z(i) - H(i)x] \quad (3.4.1-2)$$

weighted with the inverses of the positive definite matrices $R(i)$.

The above can be rewritten in a compact form as

$$J(k) = [z^k - H^k x]' (R^k)^{-1} [z^k - H^k x] \quad (3.4.1-3)$$

¹If one is willing and capable of “going nonlinear” to obtain the optimal (nonlinear) estimator.

where

$$z^k = \begin{bmatrix} z(1) \\ \vdots \\ z(k) \end{bmatrix} \quad (3.4.1-4)$$

is the **stacked vector** of measurements (of dimension $kn_z \times 1$),

$$H^k = \begin{bmatrix} H(1) \\ \vdots \\ H(k) \end{bmatrix} \quad (3.4.1-5)$$

is the **stacked measurement matrix** (of dimension $kn_z \times n_x$),

$$w^k = \begin{bmatrix} w(1) \\ \vdots \\ w(k) \end{bmatrix} \quad (3.4.1-6)$$

is the stacked vector of the measurement errors, and

$$R^k = \begin{bmatrix} R(1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R(k) \end{bmatrix} = \text{diag}[R(i)] \quad (3.4.1-7)$$

Note that the matrix (3.4.1-7) is a block-diagonal positive definite $kn_z \times kn_z$ matrix. If in (3.4.1-1) there is a nonlinear function of the unknown vector x , then one has a **nonlinear LS** problem (see Subsection 3.4.4 and Section 3.7).

The **LS estimator** that minimizes (3.4.1-3) is obtained by setting its gradient with respect to x to zero

$$\nabla_x J(k) = -2H^{k'}(R^k)^{-1}[z^k - H^kx] = 0 \quad (3.4.1-8)$$

which yields

$$\hat{x}(k) = [H^{k'}(R^k)^{-1}H^k]^{-1}H^{k'}(R^k)^{-1}z^k \quad (3.4.1-9)$$

assuming the required inverse exists.

It can be easily shown that since R^k , defined in (3.4.1-7), is positive definite, the Hessian of (3.4.1-3) with respect to x is positive definite, and consequently the extremum point (3.4.1-9) is a minimum.

Note that (3.4.1-9) is a **batch estimator** — the entire data have to be processed simultaneously for every k .

Remark

In this approach, x is an unknown constant. The estimate $\hat{x}(k)$ is a random variable if the disturbances (measurement errors) $w(i)$ are modeled as random.

Relationship to the Maximum Likelihood (ML) Estimator

If the measurement errors $w(i)$ are *independent Gaussian* random variables with mean zero and covariance $R(i)$, then minimizing the LS criterion (3.4.1-2) is equivalent to maximizing the likelihood function

$$\begin{aligned}\Lambda_k(x) &= p[z^k|x] = \prod_{i=1}^k p[z(i)|x] \\ &= c e^{-\frac{1}{2} \sum_{i=1}^k [z(i) - H(i)x]' R(i)^{-1} [z(i) - H(i)x]}\end{aligned}\quad (3.4.1-10)$$

that is, the LS and ML estimators coincide in this case.

The LS criterion (3.4.1-2) implicitly assumes that $w(i)$ are independent and zero mean with covariance $R(i)$, and it leads to the minimization of the sum of their weighted norms. Since this is equivalent to the maximization of the likelihood function under the additional Gaussian assumption, the LS is clearly a “disguised” ML technique.

Properties of the LS Estimator

With the assumption that $w(i)$ are uncorrelated, zero-mean random variables with covariance $R(i)$, but without any further distributional assumptions, the LS estimator (3.4.1-9) is *unbiased*, that is,

$$E[\hat{x}(k)] = [H^k'(R^k)^{-1}H^k]^{-1}H^k'(R^k)^{-1}E[H^kx + w^k] = x \quad (3.4.1-11)$$

The estimation error is

$$\tilde{x}(k) = x - \hat{x}(k) = -[H^k'(R^k)^{-1}H^k]^{-1}H^k'(R^k)^{-1}w^k \quad (3.4.1-12)$$

Thus, the **covariance matrix of the LS estimator** is

$$\begin{aligned}P(k) &\triangleq E[\{\hat{x}(k) - E[\hat{x}(k)]\}\{\hat{x}(k) - E[\hat{x}(k)]\}'] \\ &= E[[\hat{x}(k) - x][\hat{x}(k) - x]'] \\ &= E[\tilde{x}(k)\tilde{x}(k)'] \\ &= [H^k'(R^k)^{-1}H^k]^{-1}H^k'(R^k)^{-1}R^k(R^k)^{-1}H^k[H^k'(R^k)^{-1}H^k]^{-1}\end{aligned}\quad (3.4.1-13)$$

where in the last line above use was made of (3.4.1-12) and the fact that based on (3.4.1-7) one has

$$E[w^k w^{k'}] = R^k \quad (3.4.1-14)$$

Equation (3.4.1-13) yields, after cancellations,

$$P(k) = [H^k'(R^k)^{-1}H^k]^{-1} \quad (3.4.1-15)$$

Note that when carrying out the expectation in (3.4.1-11) and (3.4.1-13) it is over w^k .

Existence of the Solution — Parameter Observability

The *existence of the inverse* of $H'R^{-1}H$ required in (3.4.1-9) is equivalent to having the covariance of the error (3.4.1-15) *finite*. This amounts to requiring the parameter x to be **observable** — that is, it can be estimated from the observations. (See also problem 3-3.)

Small eigenvalues of $H'R^{-1}H$ (near-singularity) which lead to large eigenvalues in (3.4.1-15) are an indication of **marginal (poor) observability**. The corresponding eigenvector(s) give the direction(s) in the parameter space where one has poor observability [Ham83].

3.4.2 The Recursive LS Estimator

A useful feature of the LS estimator (3.4.1-9) is that it can be rewritten in recursive form (i.e., suitable for sequential rather than batch processing). In this case, k is interpreted as “discrete time.”

When $z(k+1)$ is obtained, one can write the following partitioned forms

$$z^{k+1} = \begin{bmatrix} z^k \\ z(k+1) \end{bmatrix} \quad (3.4.2-1)$$

$$H^{k+1} = \begin{bmatrix} H^k \\ H(k+1) \end{bmatrix} \quad (3.4.2-2)$$

$$w^{k+1} = \begin{bmatrix} w^k \\ w(k+1) \end{bmatrix} \quad (3.4.2-3)$$

$$R^{k+1} = \begin{bmatrix} R^k & 0 \\ 0 & R(k+1) \end{bmatrix} \quad (3.4.2-4)$$

The Recursion for the Inverse Covariance

Expression (3.4.1-15) at $k+1$ can be expressed recursively as

$$\begin{aligned} P(k+1)^{-1} &= H^{k+1'}(R^{k+1})^{-1}H^{k+1} \\ &= \left[H^{k'} \ H(k+1)' \right] \left[\begin{array}{cc} R^k & 0 \\ 0 & R(k+1) \end{array} \right]^{-1} \left[\begin{array}{c} H^k \\ H(k+1) \end{array} \right] \\ &= H^{k'}(R^k)^{-1}H^k + H(k+1)'R(k+1)^{-1}H(k+1) \end{aligned} \quad (3.4.2-5)$$

or

$$P(k+1)^{-1} = P(k)^{-1} + H(k+1)'R(k+1)^{-1}H(k+1) \quad (3.4.2-6)$$

This can be interpreted as follows: The **information** (in the sense of Fisher, that is, the inverse covariance) at $k+1$ equals the sum of the information

at k and the new information about x obtained from $z(k+1)$. The information is *additive* here because of the following:

1. The problem is static — the parameter is fixed.
2. The observations are modeled as *independent*.

Using the matrix inversion lemma (1.3.3-11), (3.4.2-6) can be rewritten as

$$\begin{aligned} P(k+1) &= [P(k)^{-1} + H(k+1)'R(k+1)^{-1}H(k+1)]^{-1} \\ &= P(k) - P(k)H(k+1)'[H(k+1)P(k)H(k+1)' \\ &\quad + R(k+1)]^{-1}H(k+1)P(k) \end{aligned} \tag{3.4.2-7}$$

The Residual Covariance and the Update Gain

Denote the matrices

$$S(k+1) \triangleq H(k+1)P(k)H(k+1)' + R(k+1) \tag{3.4.2-8}$$

$$W(k+1) \triangleq P(k)H(k+1)'S(k+1)^{-1} \tag{3.4.2-9}$$

which, as will be seen later, have the interpretations of **covariance of the residual** and parameter **update gain**, respectively.

The Recursion for the Covariance

With (3.4.2-8) and (3.4.2-9), recursion (3.4.2-7) can be rewritten more compactly as

$$P(k+1) = [I - W(k+1)H(k+1)]P(k) \tag{3.4.2-10}$$

that is,

$$P(k+1) = P(k) - W(k+1)S(k+1)W(k+1)' \tag{3.4.2-11}$$

which is an alternative to (3.4.2-7).

Alternative Expression for the Gain

Using (3.4.2-7), one has the following identity

$$\begin{aligned} P(k+1)H(k+1)'R(k+1)^{-1} &= \{P(k)H(k+1)' - P(k)H(k+1)' \\ &\quad \cdot [H(k+1)P(k)H(k+1)' + R(k+1)]^{-1} \\ &\quad \cdot H(k+1)P(k)H(k+1)'\}R(k+1)^{-1} \\ &= P(k)H(k+1)' \\ &\quad \cdot [H(k+1)P(k)H(k+1)' + R(k+1)]^{-1} \end{aligned}$$

$$\begin{aligned}
& \cdot \{H(k+1)P(k)H(k+1)' + R(k+1) \\
& - H(k+1)P(k)H(k+1)'\}R(k+1)^{-1} \\
= & P(k)H(k+1)'S(k+1)^{-1} \\
= & W(k+1)
\end{aligned} \tag{3.4.2-12}$$

This gives an alternative expression for the update gain (3.4.2-9) as

$$W(k+1) = P(k+1)H(k+1)'R(k+1)^{-1} \tag{3.4.2-13}$$

The Recursion for the Estimate

The batch estimation equation (3.4.1-9) for $k+1$ is rewritten as

$$\begin{aligned}
\hat{x}(k+1) &= P(k+1)H^{k+1'}(R^{k+1})^{-1}z^{k+1} \\
&= P(k+1) \begin{bmatrix} H^{k'} & H(k+1)' \end{bmatrix} \begin{bmatrix} R^k & 0 \\ 0 & R(k+1) \end{bmatrix}^{-1} \begin{bmatrix} z^k \\ z(k+1) \end{bmatrix} \\
&= P(k+1)H^{k'}(R^k)^{-1}z^k + P(k+1)H(k+1)'R(k+1)^{-1}z(k+1) \\
&= [I - W(k+1)H(k+1)]P(k)H^{k'}(R^k)^{-1}z^k + W(k+1)z(k+1) \\
&= [I - W(k+1)H(k+1)]\hat{x}(k) + W(k+1)z(k+1)
\end{aligned} \tag{3.4.2-14}$$

where (3.4.1-9), (3.4.1-15), (3.4.2-10) and (3.4.2-13) were used.

The above is the **recursive parameter estimate updating equation** — the **recursive LS estimator**, written as

$$\hat{x}(k+1) = \hat{x}(k) + W(k+1)[z(k+1) - H(k+1)\hat{x}(k)] \tag{3.4.2-15}$$

The new (updated) estimate $\hat{x}(k+1)$ is therefore equal to the previous one plus a **correction term**. This correction term consists of the **gain** $W(k+1)$ multiplying the **residual** — the difference between the observation $z(k+1)$ and the **predicted value** of this observation from the previous k measurements.

Remark

Since this is a recursive scheme, **initialization** is required. This can be done by using a batch technique on a small number of initial measurements or by using an “a priori” initial estimate and an associated covariance.

The Residual Covariance

It can be easily shown that $S(k+1)$ defined in (3.4.2-8) is the **covariance of the residual** from (3.4.2-15), that is,

$$E[[z(k+1) - H(k+1)\hat{x}(k)][z(k+1) - H(k+1)\hat{x}(k)]'] = S(k+1) \tag{3.4.2-16}$$

See also problem 3-6.

3.4.3 Examples and Incorporation of Prior Information

The Sample Mean

Consider noisy observations on a constant scalar x

$$z(i) = x + w(i) \quad i = 1, \dots, k \quad (3.4.3-1)$$

For the batch LS formulation, one has

$$H^k = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad (3.4.3-2)$$

a k -dimensional vector, and let

$$R^k = I\sigma^2 \quad (3.4.3-3)$$

where I is the $k \times k$ identity matrix.

Then, using (3.4.1-9) with (3.4.3-2) and (3.4.3-3) from above, one has the LS estimate

$$\begin{aligned} \hat{x}(k) &= [H^{k'}(R^k)^{-1}H^k]^{-1}H^{k'}(R^k)^{-1}z^k \\ &= \left\{ [1 \ \cdots \ 1] (I\sigma^2)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\}^{-1} [1 \ \cdots \ 1] (I\sigma^2)^{-1} \begin{bmatrix} z(1) \\ \vdots \\ z(k) \end{bmatrix} \\ &= \frac{1}{k} \sum_{i=1}^k z(i) \end{aligned} \quad (3.4.3-4)$$

that is, the **sample mean**. The variance of this estimate, assuming $w(i)$ to be a sequence of independent and identically distributed random variables that are zero mean and with variance σ^2 , follows from (3.4.1-15) as

$$P(k) = [H^{k'}(R^k)^{-1}H^k]^{-1} = \frac{\sigma^2}{k} \quad (3.4.3-5)$$

Note that this is the same result as obtained by direct evaluation in (2.6.3-1).

To obtain the recursive form of the LS estimation, using (3.4.2-8) and (3.4.2-9) one has

$$S(k+1) = H(k+1)P(k)H(k+1)' + R(k+1) = \frac{\sigma^2}{k} + \sigma^2 = \frac{k+1}{k}\sigma^2 \quad (3.4.3-6)$$

$$W(k+1) = P(k)H(k+1)'S(k+1)^{-1} = \frac{\sigma^2}{k} \left(\frac{k+1}{k}\sigma^2 \right)^{-1} = \frac{1}{k+1} \quad (3.4.3-7)$$

and thus, using (3.4.2-15)

$$\boxed{\hat{x}(k+1) = \hat{x}(k) + \frac{1}{k+1}[z(k+1) - \hat{x}(k)]} \quad (3.4.3-8)$$

The above recursion could have also been obtained directly from the batch expression by the following simple algebraic manipulation

$$\begin{aligned}\hat{x}(k+1) &= \frac{1}{k+1} \sum_{i=1}^{k+1} z(i) \\ &= \frac{1}{k+1} \left[\sum_{i=1}^k z(i) + z(k+1) \right] \\ &= \frac{1}{k+1} [k\hat{x}(k) + z(k+1) - \hat{x}(k) + \hat{x}(k)] \\ &= \hat{x}(k) + \frac{1}{k+1}[z(k+1) - \hat{x}(k)] \quad k = 1, \dots \quad (3.4.3-9)\end{aligned}$$

with the initial condition $\hat{x}(1) = z(1)$.

Estimation of the Mean with Prior Information

The previous example will be reconsidered with x now assumed a *random variable* with **prior information** consisting of the mean \bar{x} and the variance

$$P_{xx} = \sigma_0^2 \quad (3.4.3-10)$$

The estimation will be cast in the framework of the LMMSE estimation discussed in Section 3.3.

The measurements are as in (3.4.3-1) and $w(i)$ are i.i.d. zero-mean, with variance σ^2 and independent of x . Thus (dropping the superscript of z)

$$z = H^k x + w^k \quad (3.4.3-11)$$

Averaging over x and w , one has

$$\begin{aligned}P_{zz} &= E[(z - \bar{z})(z - \bar{z})'] = E[[H^k(x - \bar{x}) + w^k)][H^k(x - \bar{x}) + w^k]'] \\ &= H^k \sigma_0^2 H^{k'} + I \sigma^2 \quad (3.4.3-12)\end{aligned}$$

and, similarly,

$$P_{xz} = E[(x - \bar{x})(z - \bar{z})'] = E[(x - \bar{x})[H^k(x - \bar{x}) + w^k]'] = \sigma_0^2 H^{k'} \quad (3.4.3-13)$$

The inverse of (3.4.3-12), using the matrix inversion lemma (1.3.3-12), is

$$\begin{aligned}P_{zz}^{-1} &= I \sigma^{-2} - \sigma^{-2} H^k (H^{k'} \sigma^{-2} H^k + \sigma_0^{-2})^{-1} H^{k'} \sigma^{-2} \\ &= I \sigma^{-2} - \frac{\sigma^{-4}}{k \sigma^{-2} + \sigma_0^{-2}} H^k H^{k'} \quad (3.4.3-14)\end{aligned}$$

The LMMSE estimate (3.3.2-10) of x is then

$$\begin{aligned}
 \hat{x}(k) &= \bar{x} + P_{xz}P_{zz}^{-1}(z - H^k\bar{x}) \\
 &= \bar{x} + \sigma_0^2 H^{k'} \left(I\sigma^{-2} - \frac{\sigma^{-4}}{k\sigma^{-2} + \sigma_0^{-2}} H^k H^{k'} \right) (z - H^k\bar{x}) \\
 &= \bar{x} + \sigma_0^2 \left(\sigma^{-2} - \frac{\sigma^{-4}}{k\sigma^{-2} + \sigma_0^{-2}} H^{k'} H^k \right) H^{k'} (z - H^k\bar{x}) \\
 &= \bar{x} + \sigma_0^2 \left(\sigma^{-2} - \frac{k\sigma^{-4}}{k\sigma^{-2} + \sigma_0^{-2}} \right) \sum_{i=1}^k [z(i) - \bar{x}] \\
 &= \bar{x} + \frac{\sigma^{-2}}{k\sigma^{-2} + \sigma_0^{-2}} \sum_{i=1}^k [z(i) - \bar{x}]
 \end{aligned} \tag{3.4.3-15}$$

or

$$\hat{x}(k) = \frac{\sigma_0^{-2}}{k\sigma^{-2} + \sigma_0^{-2}} \bar{x} + \frac{\sigma^{-2}}{k\sigma^{-2} + \sigma_0^{-2}} \sum_{i=1}^k z(i) \tag{3.4.3-16}$$

Effect of Diffuse Prior

Note that if $\sigma_0 \rightarrow \infty$, the prior information becomes **diffuse**, which amounts to lack of prior information — this is the motivation of the term **noninformative**. In this case (3.4.3-16) becomes

$$\hat{x}(k) = \frac{1}{k} \sum_{i=1}^k z(i) \tag{3.4.3-17}$$

which is the LS or ML estimate, as expected. (See also problem 3-2.)

3.4.4 Nonlinear LS — An Example

In this section the use of the LS technique for estimating parameters from a set of nonlinear measurements is presented. The **Nonlinear Least Squares (NLS)** estimator is illustrated on the scenario in Figure 3.4.4-1, where **angle-only measurements** of a stationary target are obtained from a moving platform at different times.²

In Figure 3.4.4-1, the target position is denoted by $x = [\xi \ \eta]$ and the sensor platform position at time i is denoted by $x_p(i) = [\xi_p(i) \ \eta_p(i)]$. The measurement $z(i)$ at time i is given by

$$\begin{aligned}
 z(i) &= \tan^{-1} \left(\frac{\eta - \eta_p(i)}{\xi - \xi_p(i)} \right) + v(i) \\
 &= h(x, x_p(i)) + v(i) \quad i = 1, 2, 3
 \end{aligned} \tag{3.4.4-1}$$

²This problem is identical to the situation where simultaneous measurements are made from the same number of sensors. It is also known as **multilateration**, a generalized version of **triangulation**.

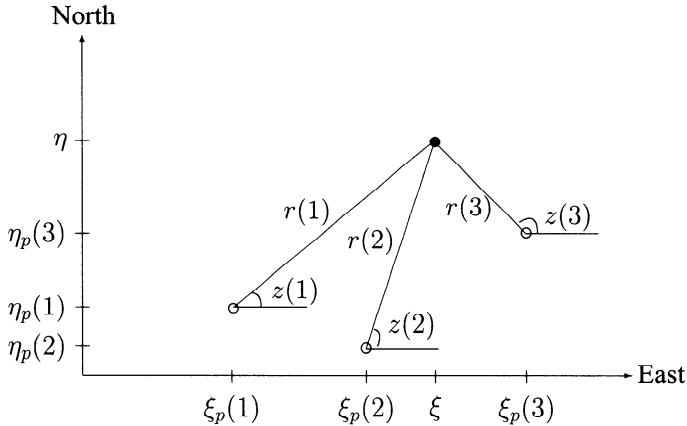


Figure 3.4.4-1: Sensor platform-target geometry for NLS estimation (○ — platform, • — target).

where $v(i)$ is the zero-mean white Gaussian measurement noise with variance σ^2 .

The stacked vector z of measurements obtained for $i = 1, 2, 3$ is denoted by

$$\begin{aligned} z &= \begin{bmatrix} z(1) \\ z(2) \\ z(3) \end{bmatrix} \\ &= \mathbf{h}(x, x_p) + w \end{aligned} \quad (3.4.4-2)$$

where

$$\mathbf{h}(x) = \begin{bmatrix} h(x, x_p(1)) \\ h(x, x_p(2)) \\ h(x, x_p(3)) \end{bmatrix} \quad (3.4.4-3)$$

and w is the stacked vector of measurement noises.

Iterated Least Squares Estimation

The **Iterated Least Squares (ILS)** estimator is a technique for iteratively improving the current estimate using the measurements until convergence (or up to a certain maximum number of iterations) based on the LS principle [Fletcher87, Blackman99].

Given the estimate \hat{x}_j^{ILS} at the end of iteration j , the updated ILS estimate $\hat{x}_{j+1}^{\text{ILS}}$ is obtained as

$$\hat{x}_{j+1}^{\text{ILS}} = \hat{x}_j^{\text{ILS}} + (J'_j R^{-1} J_j)^{-1} J'_j R^{-1} [z - \mathbf{h}(\hat{x}_j^{\text{ILS}}, x_p)] \quad (3.4.4-4)$$

where

$$J_j = \left. \frac{\partial \mathbf{h}(x, x_p)}{\partial x} \right|_{x=\hat{x}_j^{\text{ILS}}} \quad (3.4.4-5)$$

is the Jacobian matrix and the measurement covariance matrix R is given by $R = \text{diag}(\sigma^2, \sigma^2, \sigma^2)$.

The ILS estimator progresses by applying a correction term to the current estimate based on the measurement residual. Note that (3.4.4-4) is similar to the recursive LS estimator obtained in (3.4.2-14).

The Jacobian matrix J_j is given by

$$J_j = \left. \begin{bmatrix} \frac{\partial h(x, x_p(1))}{\partial \xi} & \frac{\partial h(x, x_p(1))}{\partial \eta} \\ \frac{\partial h(x, x_p(2))}{\partial \xi} & \frac{\partial h(x, x_p(2))}{\partial \eta} \\ \frac{\partial h(x, x_p(3))}{\partial \xi} & \frac{\partial h(x, x_p(3))}{\partial \eta} \end{bmatrix} \right|_{x=\hat{x}_j^{\text{ILS}}} \quad (3.4.4-6)$$

with

$$\begin{aligned} \frac{\partial h(x, x_p(i))}{\partial \xi} &= -\frac{\eta - \eta_p(i)}{(\xi - \xi_p(i))^2 + (\eta - \eta_p(i))^2} \\ \frac{\partial h(x, x_p(i))}{\partial \eta} &= \frac{\xi - \xi_p(i)}{(\xi - \xi_p(i))^2 + (\eta - \eta_p(i))^2} \end{aligned} \quad (3.4.4-7)$$

An initial estimate \hat{x}_0^{ILS} for the ILS estimator can be obtained from the intersection of any two of the three angle measurements. The iteration (3.4.4-4) is carried out until the result converges to within a certain accuracy.

The derivation of the ILS recursion is as follows. Using a first order series expansion about \hat{x}_j one has

$$z = \mathbf{h}(x, x_p) + w \approx \mathbf{h}(\hat{x}_j, x_p) + J_j(x - \hat{x}_j) + w \quad (3.4.4-8)$$

which can be rewritten (using now an equal sign, even though it is an approximation) as

$$z - \mathbf{h}(\hat{x}_j, x_p) = J_j(x - \hat{x}_j) + w \quad (3.4.4-9)$$

The above can be seen as a linear LS problem for $x - \hat{x}_j$ with observations $z - \mathbf{h}(\hat{x}_j, x_p)$, the solution to which is

$$\hat{x}_{j+1} - \hat{x}_j = (J'_j R^{-1} J_j)^{-1} J'_j R^{-1} [z - \mathbf{h}(\hat{x}_j, x_p)] \quad (3.4.4-10)$$

which immediately yields (3.4.4-4).

The mean square error of the final estimate \hat{x} can be obtained (approximately) from the above with $\hat{x}_{j+1} = \hat{x}$ and x instead of \hat{x}_j , i.e.,

$$\hat{x}_{j+1} - x = (J' R^{-1} J)^{-1} J' R^{-1} [z - \mathbf{h}(x, x_p)] = (J' R^{-1} J)^{-1} J' R^{-1} w \quad (3.4.4-11)$$

which then yields

$$E[(\hat{x}_{j+1} - x)(\hat{x}_{j+1} - x)'] = (J'R^{-1}J)^{-1} \quad (3.4.4-12)$$

where J is the Jacobian, evaluated at the final estimate. Note that the above is not a covariance but a **mean square error (MSE) matrix** because one cannot claim that the estimate is unbiased. Furthermore, it is an approximate one because it hinges on the accuracy of the expansion (3.4.4-8) on which the ILS is based.

ML Estimation with Newton-Raphson Search

The ML estimate \hat{x}^{ML} is obtained by maximizing the likelihood function $\Lambda(x)$ given by

$$\begin{aligned} \Lambda(x) &= \prod_{i=1}^3 p(z(i)|x) \\ &= \prod_{i=1}^3 ce^{-\frac{1}{2\sigma^2}[z(i)-h(x,x_p(i))]^2} \end{aligned} \quad (3.4.4-13)$$

where c is a normalizing constant. That is,

$$\begin{aligned} \hat{x}^{\text{ML}} &= \arg \max_x \Lambda(x) = \arg \min_x \{-\ln[\Lambda(x)]\} \\ &\triangleq \arg \min_x \lambda(x) \end{aligned} \quad (3.4.4-14)$$

where

$$\lambda(x) = \frac{1}{2\sigma^2} \sum_{i=1}^3 [z(i) - h(x, x_p(i))]^2 \quad (3.4.4-15)$$

Note that this is the same as the NLS problem formulation.

The minimization has to be carried out using a numerical search, for example, with the Newton-Raphson method, where the estimate \hat{x}_j^{ML} in the j th iteration is related to the next estimate $\hat{x}_{j+1}^{\text{ML}}$ via

$$\hat{x}_{j+1}^{\text{ML}} = \hat{x}_j^{\text{ML}} - H_j^{-1} \nabla_x \lambda(x) \Big|_{x=\hat{x}_j^{\text{ML}}} \quad (3.4.4-16)$$

where the Hessian matrix H_j is given by

$$H_j = \left[\begin{array}{cc} \frac{\partial^2 \lambda(x)}{\partial^2 \xi} & \frac{\partial^2 \lambda(x)}{\partial \xi \partial \eta} \\ \frac{\partial^2 \lambda(x)}{\partial \xi \partial \eta} & \frac{\partial^2 \lambda(x)}{\partial^2 \eta} \end{array} \right] \Big|_{x=\hat{x}_j^{\text{ML}}} \quad (3.4.4-17)$$

with

$$\begin{aligned}
 \frac{\partial \lambda(x)}{\partial \xi} &= \frac{1}{\sigma^2} \sum_{i=1}^3 [z(i) - h(x, x_p(i))] \frac{\partial h(x, x_p(i))}{\partial \xi} \\
 \frac{\partial \lambda(x)}{\partial \eta} &= \frac{1}{\sigma^2} \sum_{i=1}^3 [z(i) - h(x, x_p(i))] \frac{\partial h(x, x_p(i))}{\partial \eta} \\
 \frac{\partial^2 \lambda(x)}{\partial^2 \xi} &= \frac{1}{\sigma^2} \sum_{i=1}^3 [z(i) - h(x, x_p(i))] \frac{\partial^2 h(x, x_p(i))}{\partial^2 \xi} - \left(\frac{\partial h(x, x_p(i))}{\partial \xi} \right)^2 \\
 \frac{\partial^2 \lambda(x)}{\partial^2 \eta} &= \frac{1}{\sigma^2} \sum_{i=1}^3 [z(i) - h(x, x_p(i))] \frac{\partial^2 h(x, x_p(i))}{\partial^2 \eta} - \left(\frac{\partial h(x, x_p(i))}{\partial \eta} \right)^2 \\
 \frac{\partial^2 \lambda(x)}{\partial \xi \partial \eta} &= \frac{1}{\sigma^2} \sum_{i=1}^3 [z(i) - h(x, x_p(i))] \frac{\partial^2 h(x, x_p(i))}{\partial \xi \partial \eta} - \\
 &\quad \frac{\partial h(x, x_p(i))}{\partial \xi} \frac{\partial h(x, x_p(i))}{\partial \eta}
 \end{aligned} \tag{3.4.4-18}$$

and

$$\frac{\partial^2 h(x, x_p(i))}{\partial^2 \xi} = \frac{2(\eta - \eta_p(i))(\xi - \xi_p(i))}{[(\eta - \eta_p(i))^2 + (\xi - \xi_p(i))^2]^2} \tag{3.4.4-19}$$

$$\frac{\partial^2 h(x, x_p(i))}{\partial^2 \eta} = -\frac{2(\eta - \eta_p(i))(\xi - \xi_p(i))}{[(\eta - \eta_p(i))^2 + (\xi - \xi_p(i))^2]^2} \tag{3.4.4-20}$$

$$\frac{\partial^2 h(x, x_p(i))}{\partial \xi \partial \eta} = \frac{(\eta - \eta_p(i))^2 - (\xi - \xi_p(i))^2}{[(\eta - \eta_p(i))^2 + (\xi - \xi_p(i))^2]^2} \tag{3.4.4-21}$$

Note that the Hessian may lose its positive definiteness due to numerical approximations — it is impossible to ensure that the Newton step will be taken in a direction decreasing in value. To rectify this, quasi-Newton or variable metric methods, which use an approximate inverse Hessian to guarantee downhill progress toward the minimum in each Newton step, should be used in practice [Press92].

The Cramer-Rao Lower Bound

The next step is to quantify the accuracy obtainable with an (unbiased) estimator. This is done by evaluating the CRLB (see Section 2.7.2)

$$E [(x - \hat{x})(x - \hat{x})'] \geq J^{-1} \tag{3.4.4-22}$$

where J is the Fisher information matrix (FIM) given by

$$\begin{aligned}
 J &= E \{ [\nabla_x \ln \Lambda(x)] [\nabla_x \ln \Lambda(x)]' \} |_{x=x_{\text{true}}} \\
 &= E \{ [\nabla_x \lambda(x)] [\nabla_x \lambda(x)]' \} |_{x=x_{\text{true}}}
 \end{aligned} \tag{3.4.4-23}$$

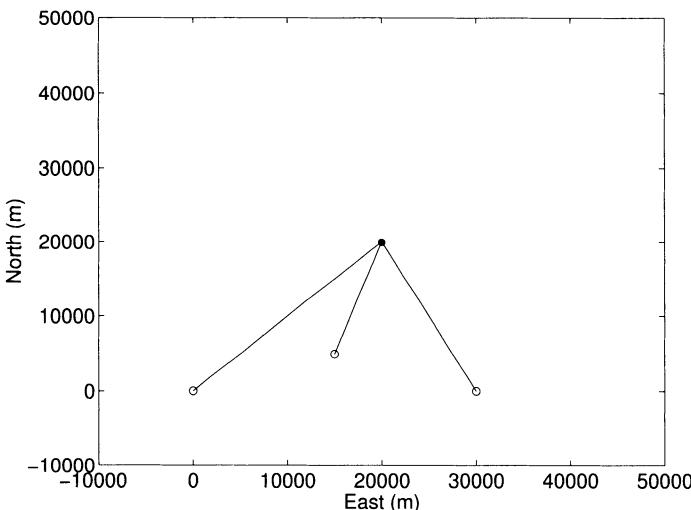


Figure 3.4.4-2: Scenario with good observability (○ — platform, • — target).

For this angle-only tracking problem, the FIM can be shown to be (see Section 3.7.4)

$$J = \frac{1}{\sigma^2} \sum_{i=1}^3 [\nabla_x h(x, x_p(i))] [\nabla_x h(x, x_p(i))]' \Big|_{x=x_{\text{true}}} \quad (3.4.4-24)$$

Simulation Results

The estimation using the ILS and ML estimators is illustrated on two different scenarios, one with a “good” sensor-to-target geometry and the other with a “poor” geometry. That is, these two scenarios have good and poor observabilities, indicated by a low and a high CRLB values, respectively. The scenarios are shown in Figs. 3.4.4-2 and 3.4.4-3, respectively.

The RMS estimation errors obtained over 350 Monte Carlo runs, together with the corresponding CRLB values, are given in Table 3.4.4-1 and 3.4.4-2 for the above scenarios. The measurement standard deviation σ in both cases is 1° . The ILS and ML iterations terminate when the norm of the difference between two consecutive estimates is less than 0.5 m or when the number of iterations exceeds 50. Typically, the algorithms converge in about 5 iterations. Both algorithms are initialized with the intersection of the first and the third line-of-sight measurements.

It can be seen from Tables 3.4.4-1 and 3.4.4-2 that both estimators yield nearly identical results in the first scenario with good observability. However, in

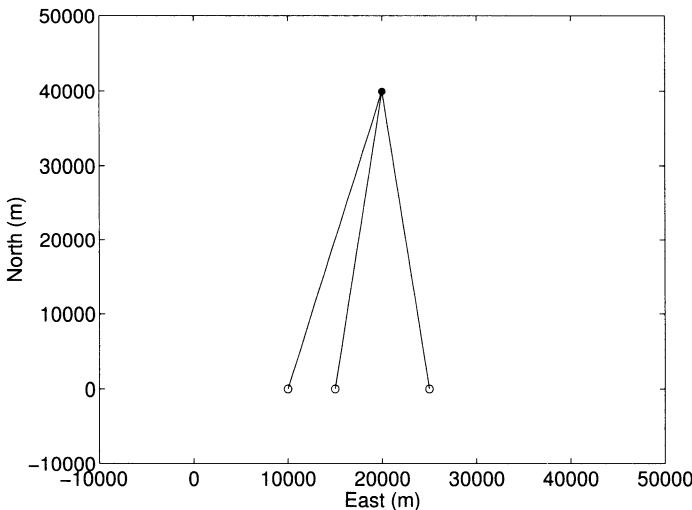


Figure 3.4.4-3: Scenario with poor observability (\circ — platform, \bullet — target).

the second scenario with poor observability, the ILS estimator outperforms the ML estimator, which requires the Hessian for its calculations. This is because, with a poor geometry, the numerical evaluation of the Hessian is not accurate enough to guarantee its positive definiteness.

Table 3.4.4-1: Simulation results for the scenario in Figure 3.4.4-2.

Algorithm	ξ (m)	$\bar{\xi}$ (m)	$\sigma_{\xi}^{\text{CRLB}}$ (m)	$\hat{\sigma}_{\xi}$ (m)	η (m)	$\bar{\eta}$ (m)	$\sigma_{\eta}^{\text{CRLB}}$ (m)	$\hat{\sigma}_{\eta}$ (m)
ILS	20,000	20,000.9	244.7	268.9	20,000	20,001.1	494.7	512.2
ML	20,000	20,000.9	244.7	268.9	20,000	20,001.1	494.7	512.2

The estimation results from a single run on the first scenario are shown in Figs. 3.4.4-4 and 3.4.4-5 to illustrate the convergence of the estimates in both algorithms. It can be seen that the estimates converge to their final values within a few iterations.

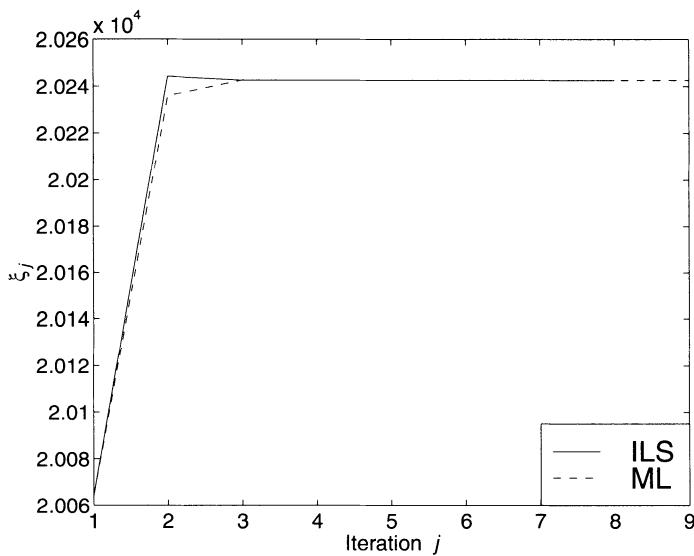
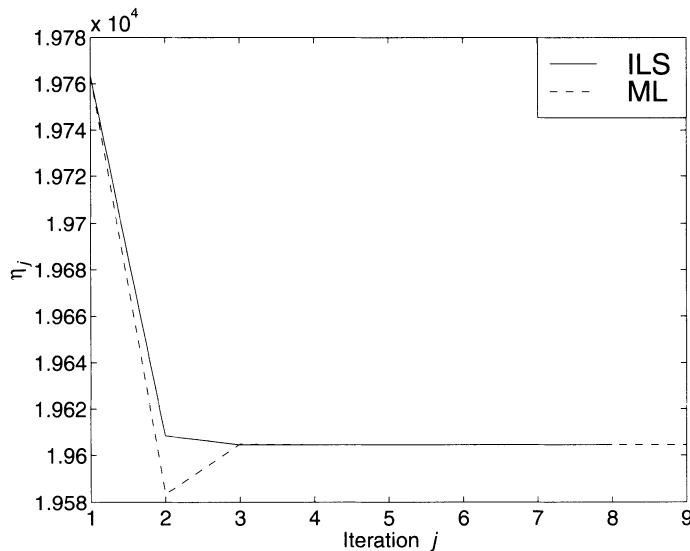
Figure 3.4.4-4: Convergence of iterative estimates $\hat{\xi}_j$.Figure 3.4.4-5: Convergence of iterative estimates $\hat{\eta}_j$.

Table 3.4.4-2: Simulation results for the scenario in Figure 3.4.4-3.

Algorithm	ξ_{true} (m)	$\bar{\xi}$ (m)	$\sigma_{\xi}^{\text{CRLB}}$ (m)	$\hat{\sigma}_{\xi}$ (m)	η_{true} (m)	$\bar{\eta}$ (m)	$\sigma_{\eta}^{\text{CRLB}}$ (m)	$\hat{\sigma}_{\eta}$ (m)
ILS	20,000	19,997.9	465.1	480.3	40,000	40,088.4	2671.4	2876.5
ML	20,000	20,003.5	465.1	520.2	40,000	40,065.8	2671.4	2928.1

Geometric Dilution of Precision

The **Geometric Dilution of Precision (GDOP)** is defined as the degradation of the overall position estimate RMSE vs. (in this case) the crossrange accuracy of the sensor. That is,

$$\text{GDOP} = \frac{\text{Position RMSE}}{\text{Distance to target} \times \sigma} \quad (3.4.4-25)$$

Note that the angle standard deviation σ has to be in radians.

The GDOP is very sensitive to the sensor-target geometry — a higher value of GDOP indicates a poor geometry. In addition, the lower the GDOP the better the estimation algorithm. For the scenarios in Figs. 3.4.4-2 and 3.4.4-3, the GDOP values are in the range 1.0–1.3 and 3.9–4.0, respectively. As one would expect, the GDOP values are higher for the scenario with poor observability.

3.4.5 LS Estimation — Summary

The LS estimator based on a set (or a sequence) of linear measurements of an *unknown constant parameter* is a linear function of the *stacked measurement vector* — this is the *batch* form of the LS.

The LS criterion is really a disguised ML criterion under suitable Gaussian assumptions on the observation noises (zero mean, uncorrelated, and with the covariance given by the inverses of the LS criterion weighting matrices).

The LS estimator can be rewritten in a *recursive* form:

The estimate based on a given set of measurements is a *linear combination* of

1. the *previous estimate* (available prior to the latest measurement) and
2. the *latest measurement*.

The inverse of the covariance of the parameter estimate (the information matrix in the sense of Fisher) is also obtained *recursively* as the sum of:

1. the inverse covariance of the estimate prior to the latest measurement — the *old information* — and

2. the information about the parameter in the latest measurement — the ***new information***.

3.5 POLYNOMIAL FITTING

3.5.1 Fitting a First-Order Polynomial to Noisy Measurements

Assume that one measures, in the presence of additive noise, the position of an object moving in one dimension with constant velocity, that is,

$$z(i) = x_0 + \dot{x}_0 t_i + w(i) \quad i = 1, \dots \quad (3.5.1-1)$$

This motion is characterized by the unknown parameter

$$x = [x_0 \ \dot{x}_0]' \quad (3.5.1-2)$$

consisting of the object's *initial position and velocity* at $t_0 = 0$, the **reference time**.

Equation (3.5.1-1), which is known in statistics as a first-order **regression**, can be written as

$$z(i) = H(i)x + w(i) \quad (3.5.1-3)$$

where

$$H(i) = [1 \ t_i] \quad (3.5.1-4)$$

The problem of estimating the parameter x amounts to **polynomial fitting**: in this case, fitting a first-order polynomial — a straight line — to a set of noisy measurements.

If the noises $w(i)$ are independent, identically distributed, zero mean, and with variances σ^2 , then according to (3.4.1-7), the covariance matrix of the stacked measurement noise vector is

$$R^k = I\sigma^2 \quad (3.5.1-5)$$

where I denotes the identity matrix of dimension k , not indicated for simplicity.

The batch solution is, noting that σ^2 cancels,

$$\begin{aligned} \hat{x}(k) &= [H^k(R^k)^{-1}H^k]^{-1}H^k(R^k)^{-1}z^k = [H^k(I\sigma^2)^{-1}H^k]^{-1}H^k(I\sigma^2)^{-1}z^k \\ &= (H^kH^k)^{-1}H^kz^k \end{aligned} \quad (3.5.1-6)$$

Because of the special form of the noise covariance (3.5.1-5), the parameter estimate covariance can be written as

$$P(k) = [H^k(R^k)^{-1}H^k]^{-1} = [H^k(I\sigma^2)^{-1}H^k]^{-1} = (H^kH^k)^{-1}\sigma^2 \quad (3.5.1-7)$$

The recursive solution is

$$\hat{x}(k+1) = \hat{x}(k) + W(k+1)[z(k+1) - H(k+1)\hat{x}(k)] \quad (3.5.1-8)$$

The **initialization** of recursion (3.5.1-8) is to be done using a batch estimate. This requires a minimum of two observations in (3.5.1-6), because otherwise the required inversion cannot be done. For $k = 2$, one has

$$\begin{aligned} P(2) &= (H^{2'} H^2)^{-1} \sigma^2 = \left\{ \begin{bmatrix} 1 & 1 \\ t_1 & t_2 \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \end{bmatrix} \right\}^{-1} \sigma^2 \\ &= \frac{1}{(t_2 - t_1)^2} \begin{bmatrix} t_1^2 + t_2^2 & -t_1 - t_2 \\ -t_1 - t_2 & 2 \end{bmatrix} \sigma^2 \end{aligned} \quad (3.5.1-9)$$

and

$$\begin{aligned} \hat{x}(2) &= P(2) H^{2'} \sigma^{-2} z^2 = \frac{1}{(t_2 - t_1)^2} \begin{bmatrix} t_1^2 + t_2^2 & -t_1 - t_2 \\ -t_1 - t_2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ t_1 & t_2 \end{bmatrix} \begin{bmatrix} z(1) \\ z(2) \end{bmatrix} \\ &= \frac{1}{t_2 - t_1} \begin{bmatrix} z(1)t_2 - z(2)t_1 \\ z(2) - z(1) \end{bmatrix} \end{aligned} \quad (3.5.1-10)$$

The covariance matrix of the estimation error of the parameter vector (3.5.1-2) is, from $k + 1$ measurements, given by

$$\begin{aligned} P(k+1) &= (H^{k+1'} H^{k+1})^{-1} \sigma^2 = \left\{ \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_{k+1} \\ \vdots & & \vdots \\ 1 & & t_{k+1} \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_{k+1} \end{bmatrix} \right\}^{-1} \sigma^2 \\ &= \begin{bmatrix} s_0 & s_1 \\ s_1 & s_2 \end{bmatrix}^{-1} = \frac{1}{s_0 s_2 - s_1^2} \begin{bmatrix} s_2 & -s_1 \\ -s_1 & s_0 \end{bmatrix} \sigma^2 \end{aligned} \quad (3.5.1-11)$$

where

$$s_j \triangleq \sum_{i=1}^{k+1} (t_i)^j \quad j = 0, 1, 2 \quad (3.5.1-12)$$

denote compactly the three functions of the sampling times needed in (3.5.1-11).

Then, the gain (a two-dimensional vector) is obtained from (3.4.2-12) as

$$\begin{aligned} W(k+1) &= P(k+1) H(k+1)' R(k+1)^{-1} \\ &= \frac{1}{s_0 s_2 - s_1^2} \begin{bmatrix} s_2 - s_1 t_{k+1} \\ -s_1 + s_0 t_{k+1} \end{bmatrix} \end{aligned} \quad (3.5.1-13)$$

These expressions are general — the sampling times are arbitrary. They can be simplified for uniformly spaced samples, as will be seen next.

If the samples are uniformly spaced with sampling interval (period) T , that is,

$$t_i = iT \quad i = 1, 2, \dots \quad (3.5.1-14)$$

then the three functions in (3.5.1-12) have the following closed-form expressions:

$$s_0 = k + 1 \quad (3.5.1-15)$$

$$s_1 = \frac{(k+1)(k+2)}{2} T \quad (3.5.1-16)$$

$$s_2 = \frac{(k+1)(k+2)(2k+3)}{6} T^2 \quad (3.5.1-17)$$

In this case the gain (3.5.1-13) is

$$W(k+1) = \begin{bmatrix} -\frac{2}{(k+1)} \\ \frac{6}{(k+1)(k+2)T} \end{bmatrix} \quad (3.5.1-18)$$

The explicit expression of the covariance (3.5.1-11) is

$$\begin{aligned} P(k+1) &= \frac{\sigma^2}{s_0 s_2 - s_1^2} \begin{bmatrix} s_2 & -s_1 \\ -s_1 & s_0 \end{bmatrix} \\ &= \frac{\sigma^2}{k(k+1)} \begin{bmatrix} 2(2k+3) & -\frac{6}{T} \\ -\frac{6}{T} & \frac{12}{(k+2)T^2} \end{bmatrix} \end{aligned} \quad (3.5.1-19)$$

Remark

The gain (3.5.1-18) tends to zero as $k \rightarrow \infty$ since the covariance $P(k)$ tends to zero as more observations are made. The reason for the gain tending to zero is that *zero variance implies perfect estimate*, in which case *there is no more need to update the estimate*.

Note

In statistics, regression is used to relate one variable, z , to a set of variables, y_j

$$z(i) = \sum_{j=1}^n a_j y_j(i) + w(i) \quad (3.5.1-20)$$

with the variance of $w(i)$ assumed, in general, *unknown*, to be estimated together with the coefficients a_j . In the case discussed above we had $n = 2$, $y_1(i) = 1$, $y_2(i) = t_i$, $a_1 = x_0$, $a_2 = \dot{x}_0$, and the variance of $w(i)$ was assumed to be known. This last assumption is reasonable if we know that the measurements are made with a sensor with known accuracy, which is the case in (most of the) engineering problems.

3.5.2 Fitting a General Polynomial to a Set of Noisy Measurements

Assume that the evolution of the position of an object is modeled as a polynomial in time, that is,

$$\xi(t) = \sum_{j=0}^n a_j \frac{t^j}{j!} \quad (3.5.2-1)$$

with the parameters being the polynomial coefficients a_j , $j = 0, 1, \dots, n$, to be estimated.

The coefficient a_j is the j th derivative of the position at the **reference time** $t = 0$.

The LS technique from Section 3.4.1 will be used to estimate these parameters via **polynomial fitting** — of order n in this case.

The noisy measurements of the position (3.5.2-1) can be written as

$$z(i) = h(i)' a + w(i) \quad i = 1, \dots, k \quad (3.5.2-2)$$

where

$$a = [a_0 \ a_1 \ \dots \ a_n]' \quad (3.5.2-3)$$

is the $(n+1)$ -dimensional parameter vector to be estimated and the row vector

$$h(i)' = \left[1 \ t_i \ \dots \ \frac{t_i^n}{n!} \right] \quad (3.5.2-4)$$

plays the role of $H(i)$ from (3.4.1-1).

The measurement disturbances $w(i)$ are assumed to be a zero-mean white sequence with a known variance σ^2 .

The stacked measurement matrix (3.4.1-5) is

$$H^k = \begin{bmatrix} h(1)' \\ \vdots \\ h(k)' \end{bmatrix} \quad (3.5.2-5)$$

Then, since

$$R^k = \sigma^2 I \quad (3.5.2-6)$$

one has

$$\begin{aligned} H^{k'} (R^k)^{-1} H^k &= \sigma^{-2} H^{k'} H^k = \sigma^{-2} [h(1) \ \dots \ h(k)] \begin{bmatrix} h(1)' \\ \vdots \\ h(k)' \end{bmatrix} \\ &= \sigma^{-2} \sum_{i=1}^k h(i) h(i)' \end{aligned} \quad (3.5.2-7)$$

The estimate of the parameter vector a is then, using (3.4.1-9),

$$\hat{a}(k) = \left\{ \sum_{i=1}^k h(i) h(i)' \right\}^{-1} \sum_{i=1}^k h(i) z(i) \quad (3.5.2-8)$$

with the covariance matrix

$$P(k) = \sigma^2 \left\{ \sum_{i=1}^k h(i)h(i)' \right\}^{-1} \quad (3.5.2-9)$$

Note that because the term to be inverted in (3.5.2-8) and in (3.5.2-9) is an $(n + 1) \times (n + 1)$ matrix that is the sum of k dyads, one needs $k \geq n + 1$ in order for the inverse to exist; that is, *at least as many measurements as the number of the parameters are needed.*

Using (3.5.2-4) one can write the i th dyad in (3.5.2-8) or (3.5.2-9) as

$$h(i)h(i)' = \begin{bmatrix} 1 & t_i & \cdots & t_i^n/n! \\ t_i & t_i^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ t_i^n/n! & \cdots & (t_i^n/n!)^2 \end{bmatrix} \quad (3.5.2-10)$$

With this, explicit expressions can be obtained for (3.5.2-8) and (3.5.2-9) if the samples are uniformly spaced.

Let

$$t_i = \frac{2i - k - 1}{2}T \quad i = 1, \dots, k \quad (3.5.2-11)$$

where T is the sampling period and the sampling times are centered around $t = 0$ for convenience.

Then the parameter a_j is the j th derivative of the position *at the center of the batch*, which is the **reference time**.

The mapping of these estimates to an *arbitrary time* is presented in Subsection 3.5.3.

The explicit expressions of the parameter estimates and their covariances for polynomials of order $n = 1, 2, 3$ are given next.

First-Order Polynomial

For $n = 1$, which corresponds to a **constant velocity motion (straight-line fit)**, one has

$$P(k) = \frac{\sigma^2}{k} \begin{bmatrix} 1 & 0 \\ 0 & \frac{12}{(k-1)(k+1)T^2} \end{bmatrix} \quad (3.5.2-12)$$

and

$$\begin{bmatrix} \hat{a}_0(k) \\ \hat{a}_1(k) \end{bmatrix} = \sigma^{-2} P(k) \begin{bmatrix} \sum_{i=1}^k z(i) \\ \sum_{i=1}^k z(i)t_i \end{bmatrix} \quad (3.5.2-13)$$

In the above \hat{a}_0 and \hat{a}_1 are the position and velocity estimates, respectively, at the *center of the batch*, which corresponds to the reference time $t = 0$.

Note that (3.5.2-12) and (3.5.2-13) are equivalent to the results of the example from Subsection 3.5.1. The only difference is that in the latter the parameters were the initial position and velocity.

Second-Order Polynomial

For $n = 2$, which corresponds to a **constant acceleration motion (parabolic fit)**, one has

$$P(k) = \sigma^2 \begin{bmatrix} \frac{3(3k^2-7)}{4k(k^2-4)} & 0 & \frac{-30}{k(k^2-4)T^2} \\ 0 & \frac{12}{k(k^2-1)T^2} & 0 \\ \frac{-30}{k(k^2-4)T^2} & 0 & \frac{720}{k(k^2-1)(k^2-4)T^4} \end{bmatrix} \quad (3.5.2-14)$$

and

$$\hat{a}(k) = \begin{bmatrix} \hat{a}_0(k) \\ \hat{a}_1(k) \\ \hat{a}_2(k) \end{bmatrix} = \sigma^{-2} P(k) \begin{bmatrix} \sum_{i=1}^k z(i) \\ \sum_{i=1}^k z(i)t_i \\ \sum_{i=1}^k z(i)t_i^2/2 \end{bmatrix} \quad (3.5.2-15)$$

In the above \hat{a}_0 , \hat{a}_1 , and \hat{a}_2 are the position, velocity, and acceleration estimates, respectively, for the center of the batch ($t = 0$).

Third-Order Polynomial

For $n = 3$, which corresponds to a **constant jerk motion (cubic fit)**, one has the following parameter estimate covariance:

$$P(k) = \sigma^2 \begin{bmatrix} \frac{3(3k^2-7)}{4k(k^2-4)} & 0 & \frac{-30}{k(k^2-4)T^2} & 0 \\ 0 & \frac{25(3k^4-18k^2+31)}{k(k^2-1)(k^2-4)(k^2-9)T^2} & 0 & \frac{-840(3k^2-7)}{k(k^2-1)(k^2-4)(k^2-9)T^4} \\ \frac{-30}{k(k^2-4)T^2} & 0 & \frac{720}{k(k^2-1)(k^2-4)T^4} & 0 \\ 0 & \frac{-840(3k^2-7)}{k(k^2-1)(k^2-4)(k^2-9)T^4} & 0 & \frac{100,800}{k(k^2-1)(k^2-4)(k^2-9)T^6} \end{bmatrix} \quad (3.5.2-16)$$

The parameter estimates are given by

$$\hat{a}(k) = \begin{bmatrix} \hat{a}_0(k) \\ \hat{a}_1(k) \\ \hat{a}_2(k) \\ \hat{a}_3(k) \end{bmatrix} = \sigma^{-2} P(k) \begin{bmatrix} \sum_{i=1}^k z(i) \\ \sum_{i=1}^k z(i)t_i \\ \sum_{i=1}^k z(i)t_i^2/2 \\ \sum_{i=1}^k z(i)t_i^3/6 \end{bmatrix} \quad (3.5.2-17)$$

Remark

Comparing (3.5.2-12), (3.5.2-14) and (3.5.2-16), it can be seen that as the order of the polynomial fit increases, the parameter variances increase. This is because there is *less information per parameter* when more parameters are fitted to the same number of data points.

This increase takes place (for the present choice of the parameters — position and its derivatives at the center of the batch) for every other parameter. For example, comparing (3.5.2-14) with (3.5.2-16) indicates that

$$P_{11}(k, n = 3) = P_{11}(k, n = 2) \quad (3.5.2-18)$$

$$P_{22}(k, n=3) > P_{22}(k, n=2) \quad (3.5.2-19)$$

A second argument, indicating the order of the polynomial fit, has been used in the above equations to distinguish between the various covariance matrices.

3.5.3 Mapping of the Estimates to an Arbitrary Time

Assume a constant acceleration motion characterized by the 3-dimensional vector consisting of position, velocity, and acceleration at the reference time (center of the batch) estimated in (3.5.2-15). The estimate of the corresponding position-velocity-acceleration vector $x(t)$ at an arbitrary time t , that is, the ***prediction*** (or ***extrapolation***) based on k measurements uniformly spaced according to (3.5.2-11), is

$$\hat{x}(t|k) = \Phi(t)\hat{a}(k) \quad (3.5.3-1)$$

where $\hat{a}(k)$ is given in (3.5.2-15), and

$$\Phi(t) = \begin{bmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} \quad (3.5.3-2)$$

The corresponding covariance is

$$P(t|k) = \Phi(t)P(k)\Phi(t)' \quad (3.5.3-3)$$

where $P(k)$ is given in (3.5.2-14).

Similar transformations are used for lower- or higher-dimensional parameter vectors.

As an example, for the 2-dimensional case (straight line fitting, i.e., constant velocity assumption) it can be shown that the closed-form expression of the covariance matrix for the ***one-step prediction*** ($t = t_{k+1} = (k + 1)T/2$) is

$$P(t_{k+1}|k) \triangleq P(k+1|k) = \frac{\sigma^2}{(k-1)k} \begin{bmatrix} 2(2k+1) & 6/T \\ 6/T & \frac{12}{(k+1)T^2} \end{bmatrix} \quad (3.5.3-4)$$

Based on (3.5.3-4), Table 3.5.3-1 shows the values of the normalized one-step position prediction variances for a linear extrapolation based on k uniformly spaced observations and the corresponding velocity estimate variances.

In tracking, when the motion of the object of interest is, say, with constant velocity, and one has k measurements, a ***prediction of the location*** of the next measurement $z(k+1)$ at $t = t_{k+1} = (k + 1)T/2$ is made according to the expression

$$\hat{x}_1(t_{k+1}|k) \triangleq \hat{x}_1(k+1|k) = \hat{a}_1 + \hat{a}_2 t_{k+1} \quad (3.5.3-5)$$

Table 3.5.3-1: One-step prediction variances for linear extrapolation from k measurements.

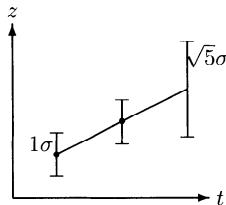
k	2	3	4	5	6
$P_{11}(k+1 k)/\sigma^2$	5	2.33	1.5	1.1	0.867
$T^2 P_{22}(k+1 k)/\sigma^2$	2	0.5	0.2	0.1	0.057

The measurement $z(k+1)$ will be, with a certain probability, in a *region around this predicted location*, called **gate**, determined by the variance associated with the prediction, $P_{11}(k+1|k)$, given above.

In practice one can use a “ 3σ gate.” Then the measurement will be in this gate with a probability of 99.8% under the Gaussian assumption (see table 1.5.4-1). Such a gate is the interval

$$\left[\hat{x}_1(k+1|k) - 3\sqrt{P_{11}(k+1|k)}, \hat{x}_1(k+1|k) + 3\sqrt{P_{11}(k+1|k)} \right] \quad (3.5.3-6)$$

Figure 3.5.3-1 illustrates the **one sigma region** corresponding to the one-step position prediction based on two observations.

**Figure 3.5.3-1:** Uncertainty for a one-step position prediction (straight-line motion).

Similarly, the position variance for an n -step prediction via linear extrapolation is

$$P_{11}(k+n|k) = 2\sigma^2 \frac{(k-1)(2k-1) + 6n(k-1) + 6n^2}{(k-1)(k+1)k} \quad (3.5.3-7)$$

For example,

$$P_{11}(4|2) = 13\sigma^2 \quad P_{11}(5|3) = \frac{29}{6}\sigma^2 \quad (3.5.3-8)$$

These values are useful in **track continuation** when the motion of the target is assumed to be described by a straight line. For example, if the target was not detected at t_{k+1} , then one has to wait until t_{k+2} . Note the increase of the two-step prediction variances compared to the one-step prediction variances from Table 3.5.3-1.

3.5.4 Polynomial Fitting — Summary

Polynomials in time can be fitted to a set of arbitrarily spaced data points (noisy position measurements) via the least squares method.

Explicit expressions have been presented for the coefficients of polynomials up to third order fitted to a set of uniformly spaced data points.

The coefficients, as presented, yield estimates of the position and its derivatives at the center of the batch of measurements (the reference time).

The covariances associated with estimates also have explicit expressions.

A simple linear mapping transforms these coefficients to the corresponding estimates to an arbitrary time — this yields the extrapolation or prediction of the motion. The same transformation matrix can be used to obtain the covariance of the position and its derivatives at an arbitrary time.

The predicted position can be used as the center of the region in which the next measurement will be with a high probability, called a *gate* or a *spatial window*. The size of the gate is determined from the variance of this prediction.

3.6 GOODNESS-OF-FIT AND STATISTICAL SIGNIFICANCE OF PARAMETER ESTIMATES

3.6.1 Hypothesis Testing Formulation of the Problem

When a set of parameters is estimated to fit a polynomial to a number of data points, there is always the following question:

What is the appropriate order of the polynomial?

The following fundamental result of estimation/polynomial fitting is relevant to this question:

Theorem. Through *any three points* on a sheet of paper, one can pass a straight line.³

Given a set of data points (scalar measurements), when fitting a polynomial to these points, one can encounter the following situations:

- If the order of the polynomial is too low — that is, *underfitting* — then the fit will be poor,

or, at the other extreme,

- If the order of the polynomial is too high — that is, *overfitting* — the estimates of some of the parameters are not “statistically significant,” that is, they are “noise.”

³Proof: Left as a trivial exercise to the reader. (*Hint:* Use a pencil that is thick enough.)

As shown in the next subsection, the sum of the squares of the residuals in an LS estimation problem — the minimized value of the LS criterion (3.4.1-2) — also called the *goodness-of-fit* or *fitting error*,

$$J^*(k) = [z^k - H^k \hat{x}(k)]' (R^k)^{-1} [z^k - H^k \hat{x}(k)] \quad (3.6.1-1)$$

has, if the noises are Gaussian, a chi-square distribution with $kn_z - n_x$ degrees of freedom. In the above, the notations (3.4.1-4) to (3.4.1-7) and (3.4.1-9) have been used and k is the number of measurements of dimension n_z , while n_x is the dimension of the parameter vector.

The matrix R^k , consisting of the noise covariances, is assumed to be known — a similar result is available for the situation where this matrix is unknown and estimated together with the “regression coefficients” x .

Test for Underfitting

The order of the polynomial fit to a set of data points is *too low* if the fit is not “good enough,” that is,

$$J^* > c = \chi_{kn_z - n_x}^2(1 - \alpha) \quad (3.6.1-2)$$

where c is obtained from Table 1.5.4-1 such that the probability of a $kn_z - n_x$ degrees of freedom chi-square random variable exceeding it is α (usually 5% or 1%).

If the first choice of the polynomial is too low, one can increase it until an “acceptable” fit is obtained, that is, the resulting J^* falls below the maximum allowed.

Test for Overfitting

If the order of the polynomial is *too high*, then the estimate (usually of the highest power coefficient) will be *statistically insignificant*.

Assuming the noises to be normal with zero mean and known variances, the estimate of the i -th component of the parameter vector is

$$\hat{x}_i(k) \sim \mathcal{N}[x_i, P_{ii}(k)] \quad (3.6.1-3)$$

that is, it is normal with mean equal to the unknown true value x_i and with variance $P_{ii}(k)$.

The *parameter estimate significance test* is the test between the following hypotheses:

$$H_0 : x_i = 0 \quad (3.6.1-4)$$

$$H_1 : x_i \neq 0 \quad (3.6.1-5)$$

subject to

$$P\{\text{accept } H_1 | H_0 \text{ true}\} = \alpha \quad (3.6.1-6)$$

Then one accepts H_1 (i.e., that the parameter is nonzero) if and only if

$$\frac{|\hat{x}_i(k)|}{[P_{ii}(k)]^{1/2}} > c' = \mathcal{G} \left(1 - \frac{\alpha}{2} \right) \quad (3.6.1-7)$$

The above implies a **two-sided probability region**. The threshold c' is obtained from the tables of Subsection 1.5.4 such that the probability of a standard Gaussian random variable exceeding it is $\alpha/2$. For example, for $\alpha = 5\%$, one has $c' = 1.96$.

If (3.6.1-7) does not hold, then the estimate of the parameter is *statistically insignificant* and it is better to accept H_0 . Then the problem is solved again for a lower-dimension parameter obtained by deleting the component found insignificant.

3.6.2 The Fitting Error in a Least Squares Estimation Problem

Consider the LS estimation of the n_x -dimensional vector x based on k measurements $z(i)$ of dimension n_z described in (3.4.1-1). The stacked vector of measurements, of dimension kn_z , is

$$z^k = H^k x + w^k \quad (3.6.2-1)$$

and the estimate was obtained as (the superscripts will be dropped for simplicity in the sequel)

$$\hat{x} = (H'R^{-1}H)^{-1}H'R^{-1}z \quad (3.6.2-2)$$

We want to evaluate the minimized value of the criterion (3.4.1-2), that is, the **fitting error**, or the **goodness-of-fit**, or the **norm of the residual**

$$J^* \triangleq (z - H\hat{x})' R^{-1} (z - H\hat{x}) \quad (3.6.2-3)$$

Note that the above, which is the sum of the squares of the normalized residuals, is a scalar and a (physically) **dimensionless quantity**.

The vector residual is

$$\begin{aligned} z - H\hat{x} &= Hx + w - H(H'R^{-1}H)^{-1}H'R^{-1}(Hx + w) \\ &= [I - H(H'R^{-1}H)^{-1}H'R^{-1}]w \end{aligned} \quad (3.6.2-4)$$

where I in (3.6.2-4) denotes the $kn_z \times kn_z$ identity matrix. This follows from the fact that the stacked vectors z and w are of dimension kn_z .

Using (3.6.2-4) in (3.6.2-3) yields

$$\begin{aligned} J^* &= w'[I - H(H'R^{-1}H)^{-1}H'R^{-1}]'R^{-1}[I - H(H'R^{-1}H)^{-1}H'R^{-1}]w \\ &= w'[R^{-1} - R^{-1}H(H'R^{-1}H)^{-1}H'R^{-1}]w \\ &= w'R^{-1/2}[I - R^{-1/2}H(H'R^{-1}H)^{-1}H'R^{-1/2}]R^{-1/2}w \\ &\triangleq \omega' A \omega \end{aligned} \quad (3.6.2-5)$$

where $R^{-1/2}$ denotes a *square root* of R^{-1} (see (1.3.2-15)),

$$\omega \triangleq R^{-1/2}w \quad (3.6.2-6)$$

is a vector of dimension kn_z , and

$$A \triangleq I - R^{-1/2}H(H'R^{-1}H)^{-1}H'R^{-1/2} \quad (3.6.2-7)$$

Assuming

$$w \sim \mathcal{N}(0, R) \quad (3.6.2-8)$$

it follows that

$$\omega \sim \mathcal{N}(0, I) \quad (3.6.2-9)$$

that is, the components of ω are *independent standardized Gaussians*.

It can be easily verified that the symmetric matrix A defined in (3.6.2-7) is *idempotent* (see (1.3.2-14)), that is,

$$AA = A \quad (3.6.2-10)$$

Such a matrix can have eigenvalues equal to 0 or 1 only. Furthermore,

$$\begin{aligned} \text{tr}(A) &= \text{tr}[I - R^{-1/2}H(H'R^{-1}H)^{-1}H'R^{-1/2}] \\ &= \text{tr}(I) - \text{tr}[R^{-1/2}H(H'R^{-1}H)^{-1}H'R^{-1/2}] \end{aligned} \quad (3.6.2-11)$$

Note that, since the dimension of I is $kn_z \times kn_z$, one has

$$\text{tr}(I) = kn_z \quad (3.6.2-12)$$

Using circular permutations for matrices multiplying each other under the trace operator (1.3.2-13) yields

$$\begin{aligned} \text{tr}[R^{-1/2}H(H'R^{-1}H)^{-1}H'R^{-1/2}] &= \text{tr}[(H'R^{-1}H)^{-1}H'R^{-1/2}R^{-1/2}H] \\ &= \text{tr}[I_{n_x}] = n_x \end{aligned} \quad (3.6.2-13)$$

where I_{n_x} is the $n_x \times n_x$ identity matrix. This follows from the fact that the dimension of $H'R^{-1}H$ is $n_x \times n_x$.

Thus, combining (3.6.2-12) and (3.6.2-13) into (3.6.2-11) results in

$$\text{tr}(A) = kn_z - n_x \quad (3.6.2-14)$$

that is, the matrix A in (3.6.2-5), which has dimension $kn_z \times kn_z$, has $kn_z - n_x$ unity eigenvalues and n_x zero eigenvalues.

Now, since ω , defined in (3.6.2-6), is zero mean normal with identity covariance matrix (of dimension $kn_z \times kn_z$) and A has $kn_z - n_x$ unity eigenvalues (and the rest zero), the **fitting error**

$$J^* = \omega' A \omega \quad (3.6.2-15)$$

is, as shown next, the sum of the squares of $kn_z - n_x$ independent scalar random variables that are normal with mean zero and unity variance.

Proof

The spectral representation (1.3.6-12) of the symmetric matrix A is

$$A = \sum_{i=1}^{kn_z} \lambda_i u_i u_i' \quad (3.6.2-16)$$

where λ_i are the eigenvalues of A , and u_i are its normalized eigenvectors that are orthogonal to each other, that is,

$$u_i' u_j = \delta_{ij} \quad (3.6.2-17)$$

Such vectors are called **orthonormal**.

Using (3.6.2-16) in (3.6.2-5), one has

$$\begin{aligned} \omega' A \omega &= \omega' \sum_{i=1}^{kn_z} \lambda_i u_i u_i' \omega = \sum_{i=1}^{kn_z} \lambda_i \omega' u_i u_i' \omega \\ &= \sum_{i=1}^{kn_z} \lambda_i u_i' \omega u_i' \omega = \sum_{i=1}^{kn_z} \lambda_i (u_i' \omega)^2 \\ &\triangleq \sum_{i=1}^{kn_z} \lambda_i \xi_i^2 \end{aligned} \quad (3.6.2-18)$$

where

$$\xi_i \triangleq \text{col}(\xi_i) \sim \mathcal{N}(0, I) \quad (3.6.2-19)$$

since ξ_i is a linear combination of the components of ω and

$$\begin{aligned} E[\xi_i \xi_j] &= E[u_i' \omega u_j] = E[u_i' \omega \omega' u_j] \\ &= u_i' E[\omega \omega'] u_j = u_i' I u_j = u_i' u_j = \delta_{ij} \end{aligned} \quad (3.6.2-20)$$

that is, ξ_i , $i = 1, \dots, kn_z$, are *independent standardized Gaussians*.

Therefore, since in (3.6.2-18) there are $kn_z - n_x$ unity eigenvalues λ_i and the rest are zero, it follows that it is the sum of $kn_z - n_x$ independent standardized Gaussians squared; that is, it has a chi-square distribution with $kn_z - n_x$ degrees of freedom. Therefore,

$$\omega' A \omega = \sum_{i=1}^{kn_z} \lambda_i \xi_i^2 = \sum_{i=1}^{kn_z - n_x} \xi_i^2 \quad (3.6.2-21)$$

and, consequently, from (3.6.2-5)

$$J^* \sim \chi_{kn_z - n_x}^2 \quad (3.6.2-22)$$

Remark

The meaning of **degrees of freedom** can be seen as being the number of (scalar) observations, kn_z , minus the number of parameters estimated, n_x .

3.6.3 A Polynomial Fitting Example

Consider the following numerical example. The true parameter vector is assumed to be of dimension 3; that is, a constant acceleration motion (polynomial of order 2) models the truth.

The measurements are given as in (3.5.2-2) by

$$z(i) = x_1 + x_2 t_i + x_3 \frac{t_i^2}{2} + w(i) \quad (3.6.3-1)$$

where

$$x_1 = 10 \quad (3.6.3-2)$$

$$x_2 = 1 \quad (3.6.3-3)$$

$$x_3 = 0.2 \quad (3.6.3-4)$$

$$t_i = \frac{1}{2}(2i - k - 1) \quad i = 1, \dots, k; \quad k = 15 \quad (3.6.3-5)$$

with the noise sequence white and

$$w(i) \sim \mathcal{N}(0, 1) \quad (3.6.3-6)$$

Table 3.6.3-1 presents the results of the fitting of a sequence of measurements generated according to the above model with polynomials of order 1, 2, and 3.

Table 3.6.3-1: Fitting of various-order polynomial models.

<i>Assumed model Order</i> n_x	<i>Linear</i> 2	<i>Quadratic</i> 3	<i>Cubic</i> 4
\hat{x}_1	11.97	10.34	10.34
$\sqrt{P_{11}}$	0.258	0.388	0.388
$ \hat{x}_1 /\sqrt{P_{11}}$	46.38	26.60	26.60
\hat{x}_2	0.996	0.997	1.114
$\sqrt{P_{22}}$	0.059	0.059	0.151
$ \hat{x}_2 /\sqrt{P_{22}}$	16.67	16.67	7.341
\hat{x}_3		0.174	0.174
$\sqrt{P_{33}}$		0.031	0.031
$ \hat{x}_3 /\sqrt{P_{33}}$		5.611	5.611
\hat{x}_4			0.021
$\sqrt{P_{44}}$			0.025
$ \hat{x}_4 /\sqrt{P_{44}}$			0.843
J^*	44.67	13.18	12.47
$\chi^2_{k-n_x}(95\%)$	22.4	21.0	19.7

For the linear model ($n_x = 2$) the fitting error J^* is too large — well above the threshold, chosen as 95% point from the chi-square tables.

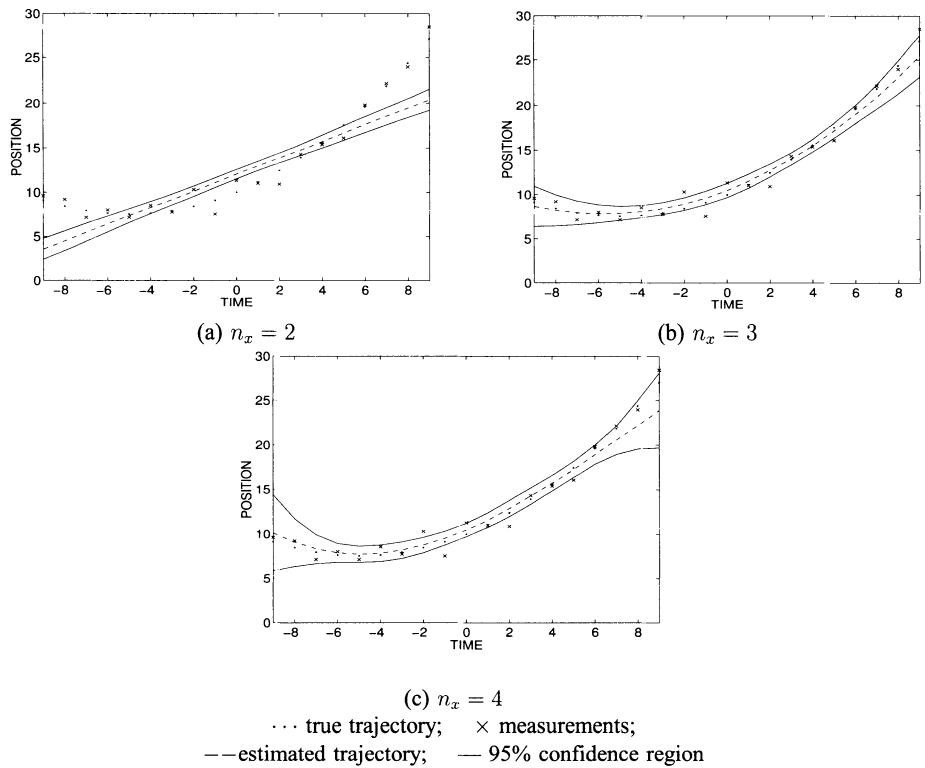


Figure 3.6.3-1: Fitting of various-order polynomials to a constant acceleration motion.

For the quadratic model ($n_x = 3$), J^* is below the threshold (i.e., acceptable) and all the parameter estimates are significant.

For the cubic model ($n_x = 4$) the last parameter (third derivative of the position — jerk) estimate is statistically insignificant: $0.843 < \mathcal{G}(97.5\%) = 1.96$; the fit has only slightly improved (the so-called F test would indicate that this improvement is statistically insignificant — this is equivalent to the parameter estimate significance test).

Thus $n_x = 4$ is clearly an “overparametrization.” It can also be seen that for $n_x = 4$ the velocity standard deviation $\sqrt{P_{22}}$ is *much larger* than for $n_x = 3$ because of the additional parameter to be estimated.

Figure 3.6.3-1 illustrates the resulting fit with these three polynomials. Also shown are the “uncertainty tubes” — the 2σ confidence region (95%) around the predicted position. Note how much more rapidly this widens for the third-order polynomial fit ($n_x = 4$) in comparison with the second-order ($n_x = 3$) case.

3.6.4 Order Selection in Polynomial Fitting — Summary

Fitting of polynomials to noisy measurements, which is a particular case of regression, consists of LS estimation of its coefficients.

The *order* of the polynomial chosen for fitting is

1. *too low* if the *fitting error* is too high (sum of squares of the residuals is too large);
2. *too high* if some estimates of coefficients are *statistically insignificant* (“buried in noise”).

The fitting error is, under the Gaussian assumption, chi-square distributed with number of degrees of freedom equal to the number of measurements minus the number of estimated parameters (this is the origin of the term *degrees of freedom*). The fitting error (“goodness-of-fit”) has to be *below a threshold in order to be acceptable*.

The statistical significance of a parameter estimate is measured by the ratio of the *magnitude of the parameter estimate* to its *standard deviation*. This ratio has to be *above a threshold* for the estimate to be significant.

Fitting a polynomial (or, in general, a model) of *unnecessarily high order* decreases the accuracy of some of the estimated coefficients (model parameters) — this *wastes information*.

3.7 USE OF LS FOR A NONLINEAR PROBLEM: BEARINGS-ONLY TARGET MOTION ANALYSIS

3.7.1 The Problem

The *nonlinear least squares* problem, defined in Subsection 2.4.1, will be used to estimate the motion parameters of a constant velocity target — *target motion analysis* — based on observations from a passive sensor that measures only the direction of arrival of a signal emitted by the target. The emitted energy can be acoustic or electromagnetic. This problem is also called *passive localization* or *passive ranging*.

The target, which is moving in a plane, is observed from a platform with a known position $[\xi_p(k), \eta_p(k)]$ in the same plane. The target “localization parameter” is the vector of dimension $n_x = 4$ consisting of its initial position and velocity in Cartesian coordinates

$$x \triangleq [x_1 \ x_2 \ x_3 \ x_4]' \triangleq [\xi(0) \ \eta(0) \ \dot{\xi} \ \dot{\eta}]' \quad (3.7.1-1)$$

The position of the target at time t_k is

$$\xi(k) \triangleq \xi(t_k) = \xi(0) + \dot{\xi}t_k = x_1 + x_3 t_k \quad (3.7.1-2)$$

$$\eta(k) \triangleq \eta(t_k) = \eta(0) + \dot{\eta}t_k = x_2 + x_4 t_k \quad (3.7.1-3)$$

It is assumed that the available measurements are bearings (“line of sight” angles with respect to some reference direction) only, given by

$$z(k) \triangleq z(t_k) = h(k, x) + w(k) \quad k = 1, \dots, n \quad (3.7.1-4)$$

where

$$h(k, x) \triangleq \tan^{-1} \frac{\eta(k) - \eta_p(k)}{\xi(k) - \xi_p(k)} \quad (3.7.1-5)$$

and $w(k)$ is the measurement noise, assumed to be a zero-mean Gaussian white sequence with known variance r , that is,

$$E[w(k)w(j)] = r\delta_{kj} \quad (3.7.1-6)$$

3.7.2 Observability of the Target Parameter in Passive Localization

A question of interest is the effect of the motion of the platform on the ability to estimate the target parameters. It will be shown that if the platform moves with a constant velocity, then the target motion parameter cannot be estimated — it is *unobservable*.

If we denote the platform velocity components as $\dot{\xi}_p$ and $\dot{\eta}_p$, then the true bearing to the target at time t is

$$\begin{aligned} \tan^{-1} \frac{\eta(t) - \eta_p(t)}{\xi(t) - \xi_p(t)} &= \tan^{-1} \frac{\eta(0) + \dot{\eta}t - \eta_p(0) - \dot{\eta}_p t}{\xi(0) + \dot{\xi}t - \xi_p(0) - \dot{\xi}_p t} \\ &= \tan^{-1} \frac{\eta(0) - \eta_p(0) + (\dot{\eta} - \dot{\eta}_p)t}{\xi(0) - \xi_p(0) + (\dot{\xi} - \dot{\xi}_p)t} \\ &= \tan^{-1} \frac{[\eta(0) - \eta_p(0)]\alpha + (\dot{\eta} - \dot{\eta}_p)\alpha t}{[\xi(0) - \xi_p(0)]\alpha + (\dot{\xi} - \dot{\xi}_p)\alpha t} \quad \forall \alpha \neq 0 \end{aligned} \quad (3.7.2-1)$$

The above is seen to hold for all α , that is, one will obtain the same sequence of true bearings if the relative position and relative velocity are multiplied by an arbitrary constant α .

Thus, an infinite number of values of the target parameter vector can yield the same observations and, consequently, its parameter vector cannot be estimated in this case — the observability requirement of full and unique recovery of the initial state is not satisfied.

Therefore, in order to estimate the target parameter vector, the platform has to have at least an acceleration. A constant speed platform with a change of course (heading) satisfies this requirement. In general, the platform trajectory has to have *one more nonzero derivative than the target trajectory*.

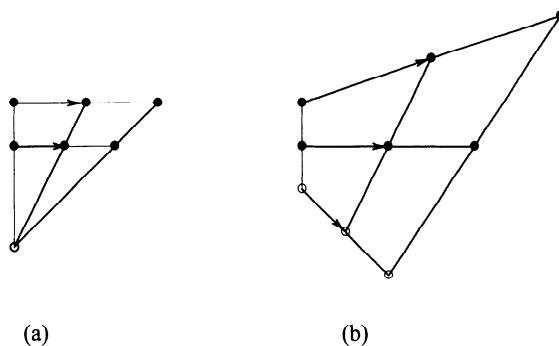


Figure 3.7.2-1: Unobservable target (● target; ○ platform).

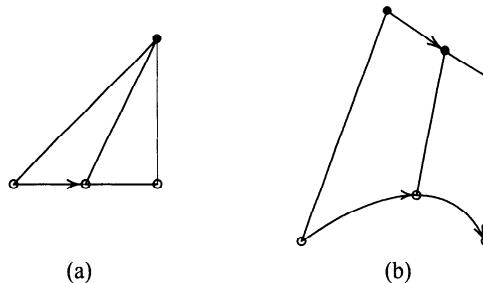


Figure 3.7.2-2: Observable target (● target; ○ platform).

Figure 3.7.2-1 illustrates two cases where the target localization parameter is not observable from angle-only measurements: (a) moving target and fixed sensor platform and (b) constant velocity target and constant velocity platform.

Figure 3.7.2-2 illustrates two observable cases: (a) fixed target and moving platform and (b) constant velocity target and platform with acceleration.

3.7.3 The Likelihood Function for Target Parameter Estimation

The likelihood function of the target parameter vector (3.7.1-1) is

$$\Lambda(x) = p(Z^n|x) = p[z(1), \dots, z(n)|x] = \prod_{k=1}^n p[z(k)|x] \quad (3.7.3-1)$$

where, in view of (3.7.1-4),

$$p[z(k)|x] = \mathcal{N}[z(k); h(k, x), r] = ce^{-\frac{1}{2r}[z(k) - h(k, x)]^2} \quad (3.7.3-2)$$

and $h(k, x)$, given in (3.7.1-5), is the expected value (average) of the observation $z(k)$ for a given target parameter vector x , and r is its variance. This follows from the assumption (3.7.1-6) that the measurement noises are white, zero mean, and with variance r .

In view of (3.7.3-2), the maximization of (3.7.3-1), which yields the maximum likelihood estimate (MLE), is equivalent to the following **nonlinear least squares (NLS)** problem:

$$\hat{x} = \arg \max_x \Lambda(x) = \arg \min_x \lambda(x) \quad (3.7.3-3)$$

where

$$\lambda(x) \triangleq \frac{1}{2r} \sum_{k=1}^n [z(k) - h(k, x)]^2 \quad (3.7.3-4)$$

is the **negative log-likelihood function** with the irrelevant additive constants omitted. The above expression of the log-likelihood function is clearly that of an NLS criterion.

The minimization of the log-likelihood function (3.7.3-4) can be carried out via one of the many existing numerical optimization algorithms. The Newton-Raphson or quasi-Newton techniques are the most effective in this case.

3.7.4 The Fisher Information Matrix for the Target Parameter

The **Cramer-Rao lower bound (CRLB)** on the covariance matrix of the target parameter estimate \hat{x} is (assuming this estimate to be unbiased)

$$E[(\hat{x} - x)(\hat{x} - x)'] \geq J^{-1} \quad (3.7.4-1)$$

where J is the **Fisher information matrix (FIM)**

$$\begin{aligned} J &= E\{[\nabla_x \ln \Lambda(x)][\nabla_x \ln \Lambda(x)]'\}|_{x=x_0} \\ &= E\{[\nabla_x \lambda(x)][\nabla_x \lambda(x)]'\}|_{x=x_0} \end{aligned} \quad (3.7.4-2)$$

The FIM is to be evaluated at the true value of the parameter x_0 ; in practice, when this is not known the evaluation is done at the estimate.

The gradient of the log-likelihood function (3.7.3-4) is

$$\nabla_x \lambda(x) = -r^{-1} \sum_{k=1}^n [\nabla_x h(k, x)][z(k) - h(k, x)] \quad (3.7.4-3)$$

which, when inserted into (3.7.4-2), yields (see problem 3-4) the sum of dyads

$$J = r^{-1} \sum_{k=1}^n h_x(k, x)h_x(k, x)'|_{x=x_0} \quad (3.7.4-4)$$

where

$$h_x(k, x) \triangleq \nabla_x h(k, x) \quad (3.7.4-5)$$

Parameter Observability and the FIM

In order to have **parameter observability** (i.e., to allow its estimation without the ambiguity discussed in Subsection 3.7.2), the FIM (3.7.4-2) *must be invertible*. If the FIM is not invertible (i.e., it is singular), then the lower bound (3.7.4-1) will not exist (actually, it will have one or more infinite eigenvalues, which means total uncertainty in a subspace of the parameter space, i.e., ambiguity). This amounts to the information limit being insufficient for the estimation problem at hand, i.e., **insufficient information in the data**.

The Normalized Estimation Error and Estimation Efficiency

Assume that the parameter estimation error

$$\tilde{x} \triangleq x - \hat{x} \quad (3.7.4-6)$$

is (approximately) Gaussian distributed, with covariance given by the Cramer-Rao lower bound, that is,

$$P \triangleq E[\tilde{x}\tilde{x}'] = J^{-1} \quad (3.7.4-7)$$

where J defined in (3.7.4-4) is assumed invertible — its rank is n_x (the dimension of x). In this case the inverse, being a covariance matrix, is *positive definite*.

The **normalized estimation error squared (NEES)** for the parameter x , defined as

$$\epsilon_x \triangleq \tilde{x}' P^{-1} \tilde{x} = \tilde{x}' J \tilde{x} \quad (3.7.4-8)$$

is chi-square distributed with n_x degrees of freedom, that is,

$$\epsilon_x \sim \chi_{n_x}^2 \quad (3.7.4-9)$$

The NEES can be used in simulations to check whether the estimator is efficient, that is, *the errors “match” the covariance given by the Cramer-Rao lower bound* — this is the **efficiency check**.

Let g be such that

$$P\{\epsilon_x \leq g^2\} = 1 - Q \quad (3.7.4-10)$$

where Q is a small “tail” probability. The NEES will have to be less than g^2 with a high probability $1 - Q$ if the estimator is efficient — this is a *one-sided probability region* for ϵ_x .

Thus, the efficiency check of the estimator (in simulations — because this is the only situation where \tilde{x} is available) consists of verifying whether (3.7.4-10) holds. This is a “single run” test — the multiple run (Monte Carlo) test is discussed later.

Confidence Region for the Target Parameter Vector

From (3.7.4-10) one can see that, given the true parameter x , one has, if the estimator is efficient, a “maximum distance” (squared) from it to the estimated parameter — the squared norm of the error. Conversely, when x is not known and \hat{x} has been obtained, one can say that x should be within some neighborhood of \hat{x} , determined by (3.7.4-10) — the **confidence region** for the true parameter.

The confidence region for the parameter x follows from (3.7.4-8) and from (3.7.4-10) as the (inside of the) “ g -sigma” **ellipsoid**

$$(x - \hat{x})' P^{-1} (x - \hat{x}) = g^2 \quad (3.7.4-11)$$

The above is an ellipsoid (actually a hyperellipsoid of dimension n_x) since P is positive definite; its semiaxes are g times the square roots of the eigenvalues of P . This region is the **probability concentration ellipsoid**, obtained by cutting the tails of the multivariate Gaussian density. The left-hand side of (3.7.4-11), which is a “normalized” distance, is also known as the **Mahalanobis distance**.

Position Estimate at an Arbitrary Time and its Confidence Region

Similarly to Subsection 3.5.3, the 2-dimensional **position estimate for an arbitrary time** t is given by the following transformation of the target parameter vector

$$\hat{x}_p(t) = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \end{bmatrix} \hat{x} \triangleq \Phi_p(t) \hat{x} \quad (3.7.4-12)$$

where Φ_p follows from the constant velocity assumption for the target.

The covariance corresponding to the above is

$$P_p(t) = \Phi_p(t) P \Phi_p(t)' \quad (3.7.4-13)$$

and the **position confidence region** is determined by the ellipse

$$[x_p - \hat{x}_p(t)]' P_p^{-1}(t) [x_p - \hat{x}_p(t)] = g_p^2 \quad (3.7.4-14)$$

where g_p is chosen for the desired probability concentration or **gate probability**.

Remarks

The Cramer-Rao lower bound quantifies in this case the **stochastic observability**, or **estimability**, which is not a binary property as in the deterministic case:

1. Lack of invertibility of J (in practice, ill-conditioning) indicates that the parameter is ***unobservable***. This happens if the condition number (1.3.6-13) of J is too large.
2. If J is invertible but the position confidence region (3.7.4-14) is “large,”⁴ one has ***marginal observability***. This occurs if the gradient vectors in (3.7.4-4) are “nearly collinear.”
3. A “small” confidence region — ***good observability*** — is obtained if the n gradient vectors in (3.7.4-4) span “well” the n_x -dimensional space, that is, they are “far” from being collinear. The actual measure of this is the condition number of J .

Expressions of the Gradient Vector Components

From (3.7.1-5) one has the following expressions for the components of the gradient vector entering into the FIM

$$h_{x_1}(k, x) = -\frac{\eta(k) - \eta_p(k)}{[\xi(k) - \xi_p(k)]^2 + [\eta(k) - \eta_p(k)]^2} \quad (3.7.4-15)$$

$$h_{x_2}(k, x) = \frac{\xi(k) - \xi_p(k)}{[\xi(k) - \xi_p(k)]^2 + [\eta(k) - \eta_p(k)]^2} \quad (3.7.4-16)$$

$$h_{x_3}(k, x) = t_k h_{x_1}(k, x) \quad (3.7.4-17)$$

$$h_{x_4}(k, x) = t_k h_{x_2}(k, x) \quad (3.7.4-18)$$

3.7.5 The Goodness-of-Fit Test

A test based on data from a *single run*, which can be used with real data is presented next. This test does not require knowledge of the true parameter.

Similarly to the linear LS, the minimized value of the log-likelihood function (3.7.3-4), multiplied by 2 for convenience, is

$$\lambda^* \triangleq \lambda(\hat{x}) = \frac{1}{r} \sum_{k=1}^n [z(k) - h(k, \hat{x})]^2 \quad (3.7.5-1)$$

This is the ***normalized sum of the squares of the residuals*** or the ***fitting error***, and it can be used as a measure of the goodness of fit. Note that (3.7.5-1) is a physically dimensionless quantity.

In the *linear LS* case, under the *Gaussian noise assumptions*, the fitting error was shown to be chi-square distributed in Subsection 3.6.2. In the present *nonlinear LS* problem, the same result can be assumed to hold approximately. Then, with n being the number of (scalar) measurements, one has

$$\lambda^* \sim \chi_{n-n_x}^2 \quad (3.7.5-2)$$

⁴In the eye of the engineer, like beauty in the eye of the beholder.

and a suitable probability region check can be made to ascertain that the model used for the problem is valid. Namely, λ^* should be, with 95% probability, below the threshold $\chi_{n-n_x}^2(0.95)$, with the notation (1.5.4-3).

This test can also be used with the results of **Monte Carlo runs**. In this case, by summing up the fitting error from N runs with *independent random variables*, one obtains a total error that is chi-square distributed with $N(n - n_x)$ degrees of freedom and has to be below a threshold obtained similarly to the one discussed above.

3.7.6 Testing for Efficiency with Monte Carlo Runs

The practical procedure to check the estimator efficiency is using **Monte Carlo simulations** as follows. Let ϵ_x^i be the NEES (3.7.4-8) in run i , $i = 1, \dots, N$, and let the sample average NEES from N independent such runs be

$$\bar{\epsilon}_x = \frac{1}{N} \sum_{i=1}^N \epsilon_x^i \quad (3.7.6-1)$$

The quantity $N\bar{\epsilon}_x$ is chi-square distributed with Nn_x degrees of freedom. Let the $1 - Q$ **two-sided probability region** for $N\bar{\epsilon}_x$ be the interval $[\epsilon'_1, \epsilon'_2]$. Using the notation from (1.5.4-3), we obtain

$$\epsilon'_1 = \chi_{Nn_x}^2 \left(\frac{Q}{2} \right) \quad (3.7.6-2)$$

$$\epsilon'_2 = \chi_{Nn_x}^2 \left(1 - \frac{Q}{2} \right) \quad (3.7.6-3)$$

For example, for $n_x = 4$, $N = 25$, $Q = 5\%$, one has $\epsilon'_1 = 74$, $\epsilon'_2 = 130$. The $1 - Q$ ($= 95\%$) two-sided probability region for $\bar{\epsilon}_x$ is

$$[\epsilon_1, \epsilon_2] = [3, 5.2] \quad (3.7.6-4)$$

where, in view of the division by N in (3.7.6-1), one has

$$\epsilon_i = \frac{\epsilon'_i}{N} \quad i = 1, 2 \quad (3.7.6-5)$$

Thus, if the estimator is *efficient*, one has to have

$$P\{\bar{\epsilon}_x \in [\epsilon_1, \epsilon_2]\} = 1 - Q \quad (3.7.6-6)$$

Remarks

The division by N in (3.7.6-1) is a convenience: It yields results in the neighborhood of n_x regardless of N . Note that the interval (3.7.6-4) is not symmetric about the mean $n_x = 4$. For a single run ($N = 1$) the corresponding interval is $[0.5, 11.1]$; also note how much smaller is (3.7.6-4), which corresponds to $N = 25$. This illustrates how Monte Carlo runs *reduce the variability*.

Another option is to use a *one-sided probability region* and check only for the upper limit (the lower limit is zero).

3.7.7 A Localization Example

The following example, using the software *BearDATTM*, illustrates the problem of estimating the localization parameter of a constant velocity target based on noisy bearings-only observations. The target parameter vector is

$$x = [10,000 \text{ m} \ 20,000 \text{ m} \ 0 \text{ m/s} \ -10 \text{ m/s}]' \quad (3.7.7-1)$$

Bearing measurements with standard deviation of 1° are made from a sensor moving on a platform, every 4 s over a total period of 900 s. The measurement noises are independent, identically distributed Gaussian with mean zero.

Two scenarios are depicted in Fig. 3.7.7-1 — they differ in the trajectory of the sensor platform. The target moves over the period of 900 s from its initial location, designated as “ \times ”. At the same time, the sensor platform moves in a two-leg constant speed trajectory starting from point “ \circ ”. This sensor motion makes the target localization parameter observable.

The 95% probability mass ellipses around the target’s initial and final position based on the Cramer-Rao lower bound, also shown, quantify the stochastic observability of the target’s localization parameter. The target’s final position uncertainty ellipse has been obtained using the transformation technique discussed in Section 3.4.

Figure 3.7.7-1b illustrates the sequences of bearing measurements obtained in one realization of each scenario.

In scenario 2, the resulting target localization uncertainty is larger than in scenario 1: The change of course in the platform trajectory leads to less information from the measurements.

Monte Carlo simulations can be used to verify estimator efficiency, that is, the *validity of the Cramer-Rao lower bound as the actual parameter estimate’s covariance matrix*. This makes it possible to obtain a *confidence region for the location of the target* at any time (under the constant velocity motion assumption).

Figure 3.7.7-2 shows the scattering of the initial and final position estimates for 100 runs of scenario 1. As can be seen, the 95% ellipses for the position indeed contain all but three of the estimated positions.

The sample average of the NEES (3.7.4-8) was obtained from these 100 runs as 3.87, which is very close to the theoretical value of $n_x = 4$. This confirms the validity of the **CRLB as the actual covariance** for the present problem.

3.7.8 Passive Localization — Summary

The technique of least squares has been illustrated for the problem of estimating the localization parameter of a constant velocity target based on noisy bearings-only observations. The noises were assumed independent and identically distributed Gaussian with zero mean and known variance.

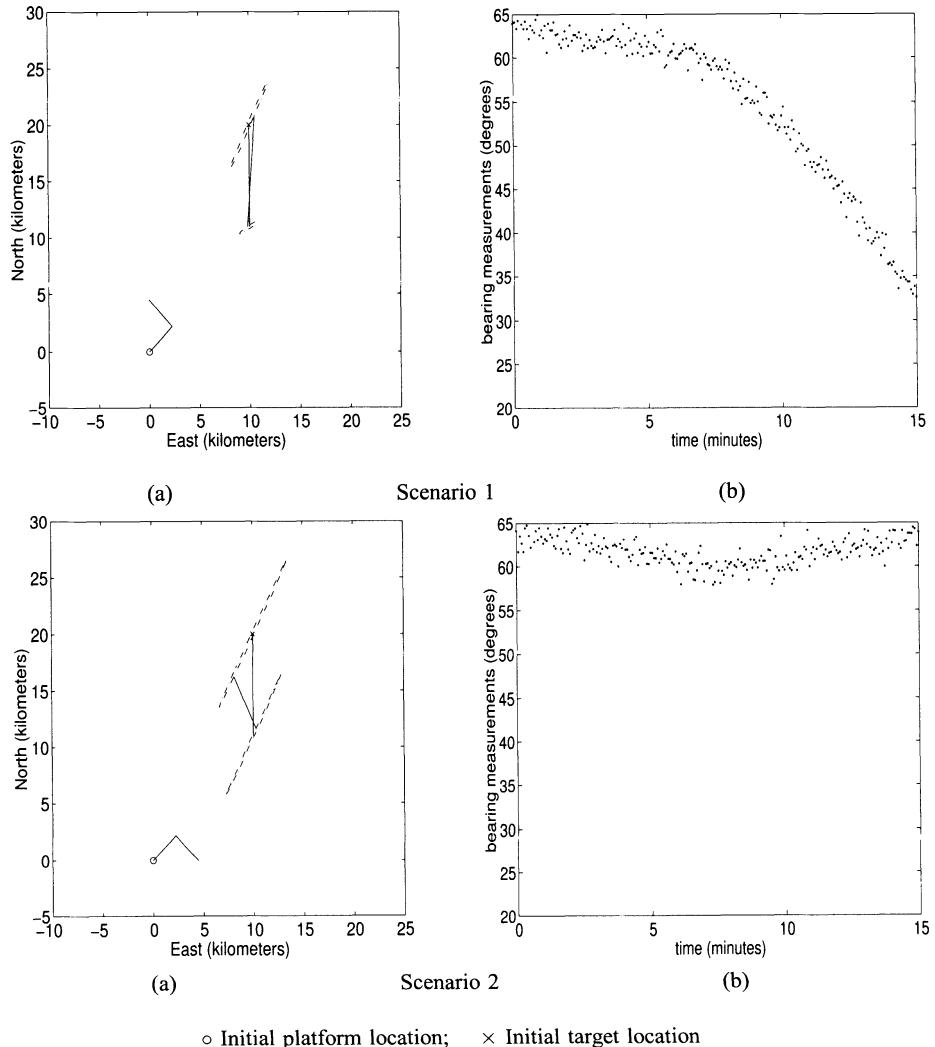


Figure 3.7.7-1: (a) Target and platform trajectories with target localization uncertainty and its estimated trajectory. (b) Sequence of bearing measurements (one realization).

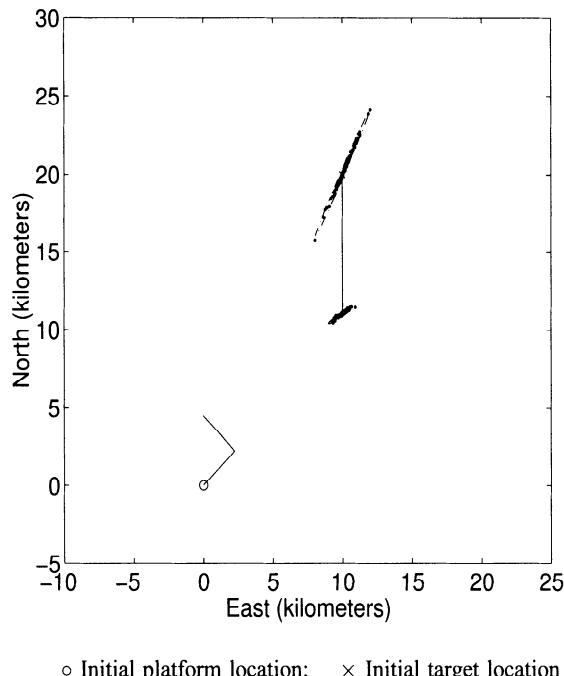


Figure 3.7.7-2: Estimated target positions in 100 runs (scenario 1).

In order for the target parameter vector to be observable, the platform, from which the measurements are made, has to undergo an acceleration. A constant-speed platform motion with a change of course satisfies this requirement.

The likelihood function of the target parameter is a nonlinear function of the parameter, and therefore a numerical search technique (Newton-Raphson or quasi-Newton) is needed to find the MLE.

The FIM for this problem has been obtained, which allows the evaluation of the Cramer-Rao lower bound for the parameter estimate's covariance.

The target's position estimate at an *arbitrary time* and the corresponding covariance matrix have been obtained.

The goodness-of-fit test can be used to ascertain the acceptability of the parameter estimate — the fitting error has to be below a certain threshold.

Monte Carlo simulations can be used to verify estimator efficiency, that is, the *validity of the CRLB as the actual parameter estimate's covariance matrix*. This makes it possible to obtain a *confidence region for the location of the target* at any time (under the constant velocity motion assumption).

3.8 NOTES, PROBLEMS AND A PROJECT

3.8.1 Bibliographical Notes

The static linear estimation material can be also found in [Sage71] and, in more detail — with simultaneous estimation of the parameters and the noise variance — in statistics or econometrics texts, for example, [Johnston72]. The reason the noise variance was assumed known here is that in engineering systems the measurements are obtained from a sensor and signal processor whose accuracy is (usually) known. In contrast, in statistics the noises reflect modeling errors and their variances are not known. The concepts of goodness of fit and statistical significance are treated mainly in statistics and econometrics texts, for example, [Johnston72]. This text (and most other statistics texts) present in detail the general regression theory where the variance is estimated together with the regression coefficients, as well as the t test and the F test used for significance testing in this case.

The software *BearDATTM* is based on [Kirubarajan96]. The general observability conditions for the passive localization of Section 3.7 are discussed in [Fogel88].

3.8.2 Problems

3-1 MMSE estimation with correlated noises. Given the prior information $x \sim \mathcal{N}(\bar{x}, \sigma_0^2)$ and the measurements

$$z(j) = x + w(j) \quad j = 1, 2$$

with the jointly Gaussian measurement noises $w(j) \sim \mathcal{N}(0, \sigma^2)$ independent of x but correlated among themselves, with

$$E[w(1)w(2)] = \rho\sigma^2$$

1. Find the variance of the MMSE estimator of x conditioned on these measurements.
2. What is the “effective number” of measurements (the number of measurements with independent noises with the same variance σ^2 , which yield the same variance $P_{xx|z}$ for the MMSE estimator)?
3. If $\bar{x} = 10$, $\sigma_0 = 1$, and $\rho = 0.5$, how accurate should the measurements be (i.e., find σ) if we want the estimate to be within 10% of the true value with 99% probability?
4. Repeat (3) if $\rho = 0$.

3-2 Sensitivity to incorrect prior variance. The random variable x with prior mean \bar{x} and variance σ_0^2 is measured via

$$z = x + w$$

where w is zero mean, with variance σ^2 , and independent of x .

1. Write the *linear* MMSE estimator \hat{x} of x in terms of z and the MSE σ_1^2 associated with this estimator.
2. Write the estimate x^* of x as above but under the *incorrect* assumption that the prior variance is σ_p^2 .
3. Find the actual MSE, σ_a^2 , associated with (2), and the MSE σ_c^2 computed by the estimator in (2).
4. Let $\sigma_p^2 = s\sigma_0^2$. Assuming $\sigma_0^2 = \sigma^2 = 2$, evaluate the computed and the actual MSE for $s = 0.5, 1, 2, 10, 20$ and compare with the best linear estimator.
5. Indicate the limits as $s \rightarrow 0$ and $s \rightarrow \infty$ and interpret the results.

3-3 Fisher information matrix in the LS problem. Derive the FIM for (3.4.1-10).

3-4 Fisher information matrix in bearings-only target localization. Prove (3.7.4-4).

3-5 Passive localization with direction cosine measurements. Rework the problem of Section 3.7 with measurements

$$h(k, x) = \frac{\xi - \xi_p}{(\xi - \xi_p)^2 + (\eta - \eta_p)^2}$$

3-6 Covariance of the residual. Prove (3.4.2-16).

3-7 Noise correlated to prior error. Consider the random variable x with

$$Ex = \bar{x} \quad \text{var}(x) = \sigma_0^2$$

The observation

$$z = x + w$$

is made, where w is a random variable with

$$Ew = 0 \quad \text{var}(w) = \sigma_w^2 \quad \text{cov}(x, w) = \rho\sigma_0\sigma_w$$

1. Find the LMMSE estimate of x given z .

2. Find the corresponding MMSE.
3. Find ρ that minimizes the MMSE.
4. For what value of ρ does this problem have no solution? Explain.

3-8 A measurement that can be noise only. Consider the random variable $x \sim (\bar{x}, \sigma_x^2)$. The observation

$$z = \alpha x + w$$

is made, where α is a random variable with

$$P\{\alpha = 1\} = p_1 \quad P\{\alpha = 0\} = 1 - p_1$$

(i.e., the measurement can be only noise) and $w \sim (0, \sigma_w^2)$. The random variables x , α and w are mutually independent.

1. Find the LMMSE estimate of x in terms of z and the associated MSE. Evaluate for $\bar{x} = 1$, $\sigma_x^2 = 1$, $p_1 = 1/2$, $\sigma_w^2 = 1$.
2. Find the (optimal) MMSE estimate of x in terms of z and the associated covariance. Assume that x and w are Gaussian with the same moments as before. Evaluate for the same numbers and compare.

3-9 LMMSE estimation for vectors. Let

$$y = Ax_1 + Bx_2$$

$$z = Cy + Dw$$

be vectors of suitable dimensions, with the matrices A , B , C , D known,

$$x_i \sim (\bar{x}_i, P_{ii}), \quad i = 1, 2$$

(i.e., distributed with the indicated mean and covariance)

$$\text{cov}(x_1, x_2) = P_{12}$$

and

$$w \sim (\bar{w}, P_w)$$

orthogonal on x_i .

1. Find the LMMSE estimate of y in terms of z
2. Find the associated matrix MSE.

3-10 Velocity estimation. Find an approximate lower bound on the estimation accuracy of the velocity of an object whose position is measured every 10 s over 70 s (8 observations) with a zero-mean noise with $\sigma = 64$ ft. Express the results in ft/min (the calculation can be done by hand with an accuracy of a few percent).

- 3-11 A position estimation variance.** An object is moving in one dimension with unknown constant velocity. Its position is observed at equal sampling intervals with an additive zero-mean noise with variance σ^2 .

1. Indicate the variance of its last position estimate after n measurements.

Hint: The result is (well) hidden somewhere in the text.

2. Now assume the velocity is known perfectly, but only the position is unknown and observed as before. What is the last position estimate variance after n measurements?
3. Show that the ratio of the position estimation variance with unknown constant velocity to the one with perfectly known velocity is $4(n + 0.5)/(n - 1)$.

- 3-12 Estimation from indirect observations.** Given the scalar random variables x_i , $i = 1, 2, 3$ jointly normal with means μ_i and covariances σ_{ij}^2 . Let

$$y_1 = x_2 + x_3$$

$$y_2 = x_2 - x_3$$

Find

1. $p(x_1|y_1)$
2. $p(x_1|y_2)$

Hint: There is no need to derive the densities. Use the fundamental equations of linear estimation.

- 3-13 Estimation with quadratic observations.** Consider the scalar random variable

$$x \sim \mathcal{N}(\bar{x}, P_{xx})$$

$$z = x + w$$

with

$$w \sim \mathcal{N}(0, P_{ww}) \quad E[xw] = 0$$

and

$$y = z^2$$

Find the LMMSE estimate of x in terms of y and the associated MSE.

- 3-14 Constant altitude trajectory estimation from range-only observations.** An object at unknown location x_0, y_0 at t_0 is moving in a plane with unknown constant velocity \dot{x}_0 . It is known that $\dot{y}(t) = 0 \quad \forall t$.

Measurements $r(t_k)$ of the range (distance to the object) are made at $t_k, k = 0, 1, \dots, N$, by a sensor at the origin with additive Gaussian zero-mean white noises with s.d. σ_r .

- (i) Formulate the ML estimation problem for the parameter vector defining the motion.
- (ii) Provide all the expressions needed (by a programmer) to evaluate the CRLB for this problem.

- 3-15 Nonlinear LS.** Consider the nonlinear LS problem in Subsection 3.4.4. Assume that now the sensor measures the range $r(i)$ to the target at times $i = 1, 2, 3$. That is,

$$z = \begin{bmatrix} r(1) \\ r(2) \\ r(3) \end{bmatrix}$$

with

$$r(i) = \sqrt{(\eta - \eta_p(i))^2 + (\xi - \xi_p(i))^2} + v_r(i) \quad i = 1, 2, 3$$

where $v_r \sim \mathcal{N}[0, (40\text{m})^2]$.

1. Derive and implement the ILS estimator for the two scenarios shown in Figs. 3.4.4-2 and 3.4.4-3.
2. Derive and implement the ML estimator with Newton-Raphson search for the same scenarios.
3. Derive the CRLB and compare it with the performances of the ILS and ML estimators.
4. Evaluate the GDOP of the estimators for both scenarios.

Note that this problem represents a simplified version of the **Global Positioning System (GPS)**. See Chapter 12 for a detailed description of GPS.

- 3-16 LMMSE estimator.** Consider the random vectors $x_i \sim (\bar{x}_i, P_{ii})$, $\text{cov}[x_i, x_j] = P_{ij}$, $i = 1, 2, 3$. Let

$$z = x_1 + x_2 - x_3$$

Find the LMMSE estimator of each x_i in terms of z and the corresponding MSE errors.

- 3-17 Prediction with unequal sampling times and variances.** Let $x = x_0 + v_0 t$. Two measurements $z_i = x(t_i) + w_i$, $i = 1, 2$, are made with zero-mean uncorrelated noises with (different) variances σ_i^2 .

1. Find the estimates of x_0 , v_0 .
2. Find the covariance matrix of their estimation errors.
3. Find the prediction of $x(t_3)$.
4. Find the variance of the error in 3.

Note: The sampling times and the noise variances are arbitrary.

3.8.3 PROJECT: An Interactive Program for Bearings-Only Target Localization

The following project is meant to develop a user-friendly interactive program for **bearings-only target localization**.

Set up an interactive program BEAR.EXE that accepts the following user-specified inputs for the problem of Section 3.7:

A1. True value of the target parameter x corresponding to $t = 0$.

A2. Sampling period T .

A3. Number of samples n .

A4. Measurement noise variance r .

A5. Platform motion — initial position, initial velocity.

A6. Platform maneuver — as velocities in each coordinate with starting and ending time (k_1 and k_2 , with the velocities being in effect from k_1T to k_2T).

A7. An initial estimate of the parameter to start the minimization. A set of default input values should be readable in from file BEAR.DEF.

The program should carry out the following:

B1. Evaluation of the Fisher information matrix corresponding to the true target parameter vector. Based on this, indicate the observability by stating whether the FIM is invertible.

B2. If the inverse can be obtained, it should be displayed.

B3. The target position uncertainty for each sampling time should be computed and the corresponding covariance matrices displayed.

B4. Generate noisy bearing measurements for the configuration defined.

B5. The MLE of the target parameter should be obtained based on the noisy measurements using a minimization algorithm (Newton-Raphson or quasi-Newton recommended) starting from the initial estimate specified as part of the input and the result will be displayed.

B6. Repeat B1–B3 for the estimated parameter.

B7. The fitting error (sum of the squares of the residuals) should be evaluated and displayed together with its number of degrees of freedom.

B8. The “normalized estimation error squared” (NEES) for the parameter vector of interest, $\tilde{x}'J\tilde{x}$, should be calculated and it should be indicated whether it is within its probability limit (e.g., 95%).

Another program, DBEAR.EXE, should display graphically the position uncertainty ellipses for the target, at user-specified times, centered at the estimate and their size according to the covariance matrices from B6. The ellipses will be of user-specified size (number of sigmas).

Document the algorithm that takes a 2×2 symmetric positive definite matrix

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix}$$

and yields the “ g -sigma” ellipse corresponding to it

$$x'P^{-1}x = g^2$$