2. Classical Estimation Theory

2.1 Mean-Square Estimation

Define,

$X \in R^n$: unknown random vector to be estimated
$Z \in R^p$: available data which is somehow related to X
$\hat{X}(Z)$: estimate of X, given Z
$\tilde{X} = X - \hat{X}$: estimation error

Bayesian estimation: Minimizes the Bayes risk

$$J = E\left[C\left(\tilde{X}\right)\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(x - \hat{x})\ f_{x,z}(x, z) dx\ dz$$

The mean-square estimation is a Bayesian estimation with $C(\tilde{x}) = \tilde{x}^T \tilde{x}$, i.e.,

$$J = E\left\{\tilde{X}^T \tilde{X}\right\} = E\left\{\left(X - \hat{X}\right)^T \left(X - \hat{X}\right)\right\}$$

Mean-square estimation: General case

Known $f_{xz}(x,z)$, find $\hat{x}(z)$.

The cost function for the mean-square estimation is

$$J = \iint [x-\hat{x}]^T [x-\hat{x}] f_{xz}(x,z)\, dxdz$$

According to the definition of conditional density function $(f_{X|Z}(x \mid z) = \dfrac{f_{XZ}(x,z)}{f_Z(z)})$

$$J = \int_{-\infty}^{\infty} f_Z(z) \int_{-\infty}^{\infty} [x-\widehat{x}(z)]^T [x-\hat{x}(z)]\, f_{X|Z}(x \mid z)\, dxdz$$

J will be a minimum with respect to the measurement *z* when the right-hand side integration is minimum since $f_z(z)$ is non-negative. Define the right-hand side integration as *J'*

$$\begin{aligned} J' &= \int_{-\infty}^{\infty} [x-\widehat{x}(z)]^T [x-\hat{x}(z)] f_{X|Z}(x \mid z) dx \\ &= E\left\{[x-\widehat{x}(z)]^T [x-\hat{x}(z)] | z\right\} \\ &= E\left\{x^T x | z\right\} - 2\widehat{x}(z)^T E\left\{x | z\right\} + \widehat{x}(z)^T \hat{x}(z) \end{aligned}$$

$\hat{x}(z)$ minimizing *J'* should satisfy

$$\frac{\partial J'}{\partial \hat{x}(z)} = -2E\left\{x|z\right\} + 2\hat{x}(z) = 0 \tag{2.1-1}$$

Solving Eq. (2.1-1) gives

$$\hat{x}_{MS}(z) = E\left\{x|z\right\} \tag{2.1-2}$$

(This is the so called *a posteriori estimate* or *update*.)

Error Mean and Covariance

The mean of $\tilde{x} = x - \hat{x}_{MS}(z)$

$$\begin{aligned} E\{\tilde{x}\} &= E\left\{x - \hat{x}_{MS}(z)\right\} \\ &= E\left\{x - E\{x \mid z\}\right\} \\ &= E\{x\} - E\{E\{x \mid z\}\} \qquad \text{(unbiased estimate)} \\ &= E\{x\} - E\{x\} \\ &= 0 \end{aligned}$$

(We used the following relationship: $E\{x\} = E\left\{E\left\{x \mid z\right\}\right\}$.)

The covariance of $\tilde{x}$

$$P_{\tilde{X}} = E\left\{\tilde{x}\tilde{x}^T\right\}$$

$$= E\left\{E\left\{\tilde{x}\tilde{x}^T \mid Z\right\}\right\}$$

$$= E\left\{P_{X|Z}\right\}$$

If $P_{X|Z}$ is independent of *Z*,

$$P_{\tilde{X}} = E\left\{\tilde{x}\tilde{x}^T\right\}$$

$$= E\left\{(x - E\{x \mid z\})(x - E\{x \mid z\})^T\right\}$$

$$= P_{X|Z}$$

A Computation of $E\{X \mid Z\}$

$$E\{X \mid Z\} = \int_{-\infty}^{\infty} x\, f_{X|Z}(x \mid z)\, dx$$

$$= \int_{-\infty}^{\infty} x\, \frac{f_{XZ}(x,z)}{f_Z(z)}\, dx$$

$$= \frac{\int_{-\infty}^{\infty} x\, f_{XZ}(x,z)\, dx}{\int_{-\infty}^{\infty} f_{XZ}(x,z)\, dx} \qquad (2.1\text{-}3)$$

$$= \frac{\int_{-\infty}^{\infty} x\, f_{Z|X}(z \mid x)\, f_x(x)\, dx}{\int_{-\infty}^{\infty} f_{Z|X}(z \mid x)\, f_x(x)\, dx}$$

Example 2.1-1: A Nonlinear Mean-Square Estimation

Let an unknown scalar x be distributed uniformly according to

$$f_x(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

a. In the absence of other information, find the a priori estimate.

$$\hat{x}_{MS} = E\{x\} = \int_{-\infty}^{\infty} x\, f_x(x)\, dx$$
$$= \int_0^1 x\, dx = \frac{1}{2}.$$

b. Suppose a z is measured which is related to x by

$$z = \ln\left(\frac{1}{x}\right) + v,$$

where v is noise with exponential distribution

$$f_v(v) = \begin{cases} e^{-v}, & v \ge 0 \\ 0, & v < 0, \end{cases}$$

and *x* and *v* are independent. Find the conditional mean estimate.

If *x* and *v* are independent,

$$f_{z|x}(z \mid x) = f_v\left[z - \ln\left(\frac{1}{x}\right)\right]$$

$$= \begin{Bmatrix} e^{-(z-\ln(1/x))}, & z - \ln\frac{1}{x} \geq 0 \\ 0, & z - \ln\frac{1}{x} < 0 \end{Bmatrix}$$

$$= \begin{Bmatrix} \frac{1}{x}e^{-z}, & x \geq e^{-z} \\ 0, & x < e^{-z} \end{Bmatrix}.$$

Now using Eq. (2.1-3)

$$\hat{x}_{MS}(z) = \frac{\int_{e^{-z}}^{1} e^{-z} dx}{\int_{e^{-z}}^{1} \frac{1}{x} e^{-z} dx}$$

$$= \frac{\left| x e^{-z} \right|_{e^{-z}}^{1}}{\left| e^{-z} \ln x \right|_{e^{-z}}^{1}}$$

$$= \frac{\left(1 - e^{-z}\right) e^{-z}}{z e^{-z}}$$

$$= \frac{1 - e^{-z}}{z}.$$

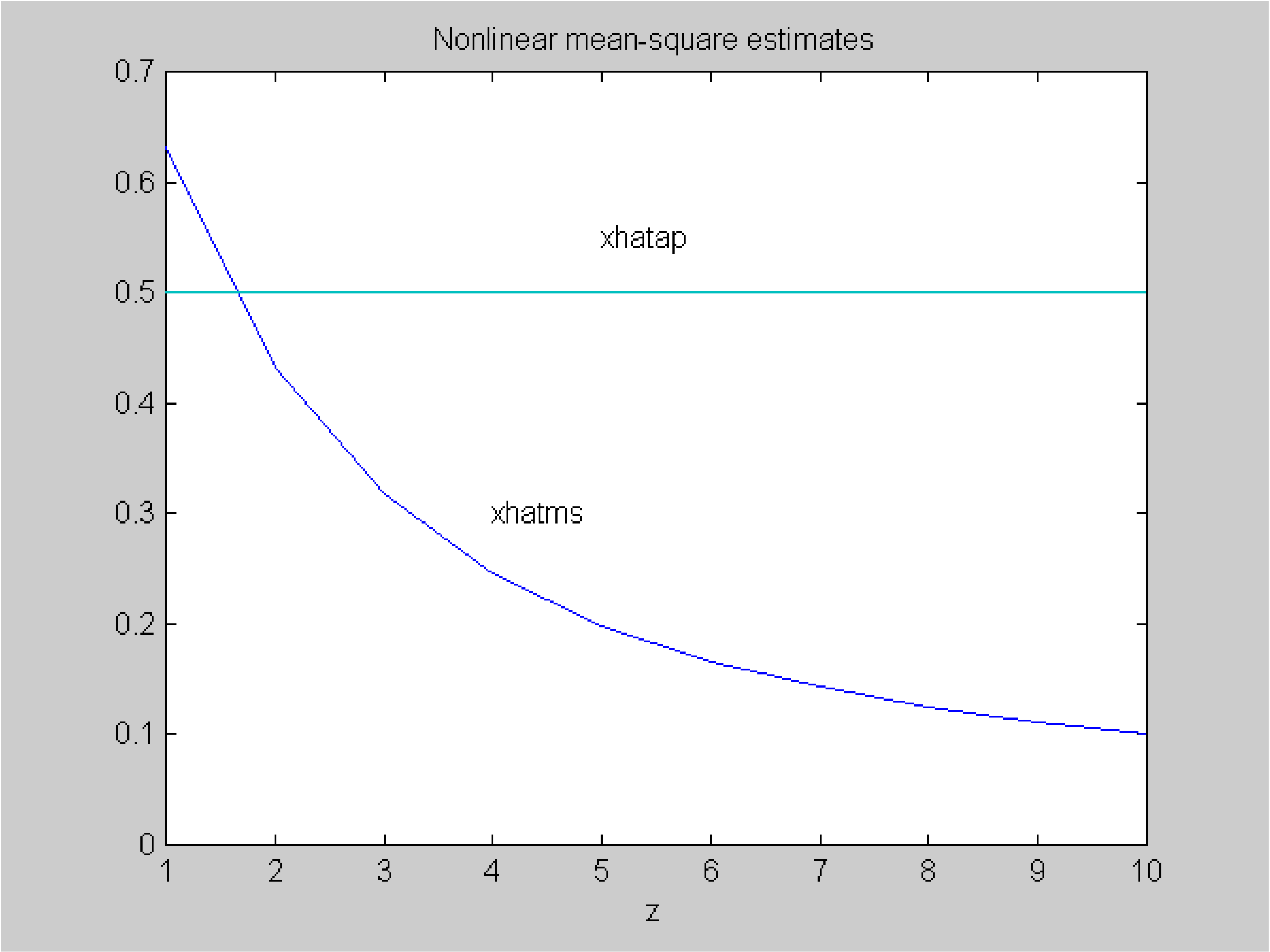


Fig. 2.1-1. Nonlinear Mean-Square Estimate (1)

Gaussian conditional mean and covariance

Let

$$x \sim N(\bar{x}, P_x)$$

$$z \sim N(\bar{z}, P_z)$$

$$y = \begin{bmatrix} x \\ z \end{bmatrix}, \ \bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix}, \ P_y = \begin{bmatrix} P_x & P_{xz} \\ P_{zx} & P_z \end{bmatrix}$$

then,

$$f_y(y) = \frac{1}{\sqrt{(2\pi)^{n+p} \left| P_y \right|}} e^{-(1/2)(y-\bar{Y})^T P_Y^{-1}(y-\bar{Y})}$$

where

$$P_Y^{-1} = \begin{bmatrix} D^{-1} & -D^{-1} P_{xz} P_z^{-1} \\ -P_z^{-1} P_{zx} D^{-1} & P_z^{-1} + P_z^{-1} P_{zx} D^{-1} P_{xz} P_z^{-1} \end{bmatrix}$$

$$D = P_x - P_{xz} P_z^{-1} P_{zx} \quad (\textit{Schur complement})$$

(Refer to Simon 1.1.2)

Now, try to find $f_{x|z}$

$$f_{x|z}(x \mid z) = \frac{f_{xz}(x,z)}{f_z(z)} = \frac{f_Y(y)}{f_z(z)}$$

$$= \frac{\sqrt{(2\pi)^p \left|P_z\right|}}{\sqrt{(2\pi)^{n+p} \left|P_Y\right|}} \exp\left\{-1/2\left[(y-\overline{Y})^T P_Y^{-1}(y-\overline{Y}) - (z-\overline{z})^T P_z^{-1}(z-\overline{z})\right]\right\}$$

{ } can be written

$$\begin{bmatrix} x-\overline{x} \\ z-\overline{z} \end{bmatrix}^T P_Y^{-1} \begin{bmatrix} x-\overline{x} \\ z-\overline{z} \end{bmatrix} - (z-\overline{z})^T P_z^{-1}(z-\overline{z})$$

$$= (x-\overline{x})^T D^{-1}(x-\overline{x}) - (x-\overline{x})^T D^{-1} P_{xz} P_z^{-1}(z-\overline{z}) - (z-\overline{z})^T P_z^{-1} P_{zx} D^{-1}(x-\overline{x})$$

$$+ (z-\overline{z})^T (P_z^{-1} + P_z^{-1} P_{zx} D^{-1} P_{xz} P_z^{-1})(z-\overline{z}) - (z-\overline{z})^T P_z^{-1}(z-\overline{z}) \qquad (2.1\text{-}4)$$

Define the mean and covariance as follows,

$$E\{x \mid z\} = \overline{x} + P_{xz} P_z^{-1}(z-\overline{z}) \qquad (2.1\text{-}5)$$

$$P_{x|z} = P_x - P_{xz} P_z^{-1} P_{zx} \qquad (2.1\text{-}6)$$

Eq. (2.1-4) may be simplified,

$$(x - E\{x \mid z\})^T P_{x|z}{}^{-1}(x - E\{x \mid z\}).$$

Then,

$$f_{x|z}(x \mid z) = \frac{1}{\sqrt{(2\pi)^n \left|P_{x|z}\right|}} \exp\{-\frac{1}{2}(x - E\{x \mid z\})^T P_{x|z}{}^{-1}(x - E\{x \mid z\})\}.$$

This says, $x \mid z$ is a Gaussian process with the above PDF.

If *x* and *z* are Gaussian, $\hat{x}_{MS}(z)$ is given by Eq. (2.1-5).

Conditional mean for linear Gaussian measurements

Consider that the Gaussian processes x and z are related by the following linear function, viz.,

$$z = Hx + v$$

where $x \sim N(\bar{x}, P_x), v \sim N(0, R), E\{xv^T\} = 0.$

Compute first the mean and covariance of z,

$$E\{z\} = E\{Hx + v\} = HE\{x\} \qquad (2.1\text{-}7)$$

$$\begin{aligned} P_z &= E\{(z-\bar{z})(z-\bar{z})^T\} \\ &= E\{(Hx+v-H\bar{x})(Hx+v-H\bar{x})^T\} \\ &= HP_xH^T + E\{Hxv^T\} - H\bar{x}\bar{v}^T + E\{vx^TH^T\} - \bar{v}\bar{x}^TH^T + R \\ &= HP_xH^T + R \end{aligned} \qquad (2.1\text{-}8)$$

The cross-covariance P_{xz} is

$$
\begin{aligned}
P_{xz} &= E\{(x-\bar{x})(z-\bar{z})^T\} \\
&= E\{(x-\bar{x})(Hx+v-H\bar{x})^T\} \\
&= P_x H^T + E\{xv^T\} - \bar{x}\bar{v}^T \\
&= P_x H^T
\end{aligned}
\qquad (2.1\text{-}9)
$$

Inserting Eqs. (2.1-8) and (2.1-9) into Eq. (2.1-6), we can obtain the estimation error covariance

$$
\begin{aligned}
P_{\tilde{x}} = P_{x|z} &= P_x - P_{xz} P_z^{-1} P_{zx} \\
&= P_x - P_x H^T (H P_x H^T + R)^{-1} H P_x
\end{aligned}
\qquad (2.1\text{-}10)
$$

Eq. (2.1-10) can be simplified by employing the matrix inversion lemma,

$$
P_{\tilde{x}} = P_{x|z} = (P_x^{-1} + H^T R^{-1} H)^{-1} \qquad (2.1\text{-}11)
$$

If we apply Eqs. (2.1-8) and (2.1-9) to Eq. (2.1-5), we obtain the conditional mean estimate,

$$\hat{x}(z) = E\{x \mid z\} = E\{x\} + P_{xz} P_z^{-1} (z - \overline{z})$$
$$= E\{x\} + P_x H^T (H P_x H^T + R)^{-1} (z - H\overline{x}) \qquad (2.1\text{-}12)$$

or

$$\hat{x}(z) = E\{x\} + P_{x|z} H^T R^{-1} (z - H\overline{x}). \qquad (2.1\text{-}13)$$

Note:

1. $P_{x|z} \leq P_x$ according to Eq. (2.1-10). (improvement on the certainty of *x*)
2. $P_{x|z}$ is independent of *z* and can be computed off-line.
3. Eqs. (2.1-12) and (2.1-13) can be considered as a combination of a priori and a posteriori estimates.

Linear mean-square estimation

Consider (a linear estimate restriction)

$$\hat{x}(z) = Az + b \tag{2.1-14}$$

where A and b are constant. Z is not necessarily Gaussian.

Define J such that

$$\begin{aligned} J &= E\{(x-\hat{x})^T(x-\hat{x})\} \\ &= tr\ E\{(x-\hat{x})(x-\hat{x})^T\} \\ &= tr\ E\{(x-Az-b)(x-Az-b)^T\} \\ &= tr\ E\{[(x-\bar{x})-(Az+b-\bar{x})][(x-\bar{x})-(Az+b-\bar{x})]^T\} \\ &= tr\,[P_x + A(P_z+\bar{z}\bar{z}^T)A^T + (b-\bar{x})(b-\bar{x})^T + 2A\bar{z}(b-\bar{x})^T - 2AP_{zx}] \end{aligned}$$

$$\frac{\partial J}{\partial b} = 2(b-\bar{x}) + 2A\bar{z} = 0 \tag{2.1-15}$$

$$\frac{\partial J}{\partial A} = 2A(P_z+\bar{z}\bar{z}^T) - 2P_{xz} + 2(b-\bar{x})\bar{z}^T = 0 \tag{2.1-16}$$

A and *b* are found from Eqs. (2.1-15) and (2.1-16). Inserting *A* and *b* into Eq. (2.1-14) provides the linear mean-square estimate

$$\hat{x}_{LMS}(z) = E\{x\} + P_{xz}P_z^{-1}(z - \bar{z}) \tag{2.1-17}$$

The mean of the estimate is

$$\begin{aligned} E\{\hat{x}_{LMS}(z)\} &= E\{E\{x\} + P_{xz}P_z^{-1}(z - \bar{z})\} \\ &= E\{x\} \end{aligned}$$

and consequently

$$E\{\tilde{x}\} = 0$$

The covariance of the estimate is

$$
\begin{aligned}
P_{\tilde{x}} &= E\{(\tilde{x}-\overline{\tilde{x}})(\tilde{x}-\overline{\tilde{x}})^T\} \\
&= E\{\tilde{x}\tilde{x}^T\} \\
&= E\{(x-\hat{x})(x-\hat{x})^T\} \\
&= E\{[(x-\bar{x})-P_{xz}P_z^{-1}(z-\bar{z})][(x-\bar{x})-P_{xz}P_z^{-1}(z-\bar{z})]^T\} \\
&= P_x - P_{xz}P_z^{-1}P_{zx} - P_{xz}P_z^{-1}P_{zx} + P_{xz}P_z^{-1}P_zP_z^{-1}P_{zx} \\
&= P_x - P_{xz}P_z^{-1}P_{zx}
\end{aligned}
\tag{2.1-18}
$$

Note:

1. Eqs. (2.1-17) and (2.1-18) are equivalent to Eqs. (2.1-6) and (2.1-7), respectively. That means "In Gaussian case, the optimal mean-square estimate is linear, since $\hat{x}_{MS} = E\{x \mid z\}$ and $\hat{x}_{LMS}$ are the same."
2. $\hat{x}_{LMS} = a\ priori\ estimate + a\ posteriori\ estimate$
3. If P_{xz} is small (less correlated), there is not much improvement in a posteriori estimate.
4. If there is no confidence in z (P_z large), there will be less influence from $(z-\bar{z})$. If there is much confidence in z (P_z small), there will be a large influence from $(z-\bar{z})$.

Ex. 2.1-2

Find $\hat{x}_{LMS}$ for the experiment of Ex. 2.1-1.

Solution:

$$f_x(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

$$z = \ln\frac{1}{x} + v$$

$$f_v(v) = \begin{cases} e^{-v}, & v \ge 0 \\ 0, & v < 0 \end{cases}$$

The following are found in Ex. (2.1-1)

$$\overline{x} = E\{x\} = 1/2$$

$$f_{z|x}(z \mid x) = \begin{cases} \frac{1}{x}e^{-z}, & x \ge e^{-z} \\ 0, & x < e^{-z} \end{cases}$$

$$\hat{x}_{MS}(z) = \frac{1 - e^{-z}}{z}$$

Now, we want to find $\hat{x}_{LMS}(z) = \bar{x} + P_{xz} P_z^{-1} (z - \bar{z})$.

$$f_{xz}(x,z) = f_{z|x}(z \mid x) f_x(x)$$
$$= \begin{Bmatrix} \frac{1}{x} e^{-z}, & for\ e^{-z} \leq x \leq 1,\ z \geq 0 \\ 0, & otherwise \end{Bmatrix}$$

$$f_z(z) = \int_{e^{-z}}^{1} \frac{1}{x} e^{-z}\ dx,\ z \geq 0$$
$$= e^{-z} \ln \frac{1}{x} \Big|_{e^{-z}}^{1} = \begin{Bmatrix} ze^{-z}, & z \geq 0 \\ 0, & z < 0 \end{Bmatrix}$$
$$= \begin{Bmatrix} ze^{-z}, & z \geq 0 \\ 0, & z < 0 \end{Bmatrix}$$

$$\bar{z} = \int_0^{\infty} z^2 e^{-z}\ dz = 2$$

$$E\{xz\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xz \; f_{xz}(x,z)\, dx\, dz$$

$$= \int_{0}^{\infty} \int_{e^{-z}}^{1} ze^{-z}\, dx\, dz = \frac{3}{4}$$

$$P_{xz} = E\{(x-\overline{x})(z-\overline{z})\} = E\{xz\} - E\{x\}\overline{z} - \overline{x}E\{z\} + \overline{xz}$$

$$= E\{xz\} - \overline{xz} = \frac{3}{4} - \frac{1}{2} \cdot 2 = -\frac{1}{4}$$

$$E\{z^2\} = \int_{-\infty}^{\infty} z^2 \; f_z(z)\, dz = \int_{0}^{\infty} z^3 \, e^{-z} \, dz = 6$$

$$P_z = E\{(z-\overline{z})^2\} = E\{z^2\} - 2E\{z\}\overline{z} + \overline{z}^2 = E\{z^2\} - \overline{z}^2 = 6 - 4 = 2$$

$$\hat{x}_{LMS}(z) = \overline{x} + P_{xz} P_z^{-1} (z - \overline{z}) = \frac{1}{2} - \frac{1}{4} \cdot \frac{1}{2}(z-2) = \frac{3}{4} - \frac{1}{8} z \, .$$

2.2 Maximum-likelihood estimation

Given a measurement Z assuming that a priori information on x are not available (e.g., $\hat{x}_0 = 0$, $P_0 = \infty$), estimate x to maximize the likelihood function $f_{z|x}$, i.e.,

$$\left.\frac{\partial f_{z|x}\left(z \mid x\right)}{\partial x}\right|_{\hat{x}_{ML}} = 0 \qquad (2.2\text{-}1)$$

or

$$\left.\frac{\partial \ln f_{z|x}\left(z \mid x\right)}{\partial x}\right|_{\hat{x}_{ML}} = 0 \qquad (2.2\text{-}2)$$

Example 2.2-1

Let an unknown scalar random variable x is measured by

$$z = \ln\left(\frac{1}{x}\right) + v,$$

where v is noise with exponential distribution

$$f_v(v) = \begin{cases} e^{-v}, & v \geq 0 \\ 0, & v < 0, \end{cases}$$

and x and v are independent.

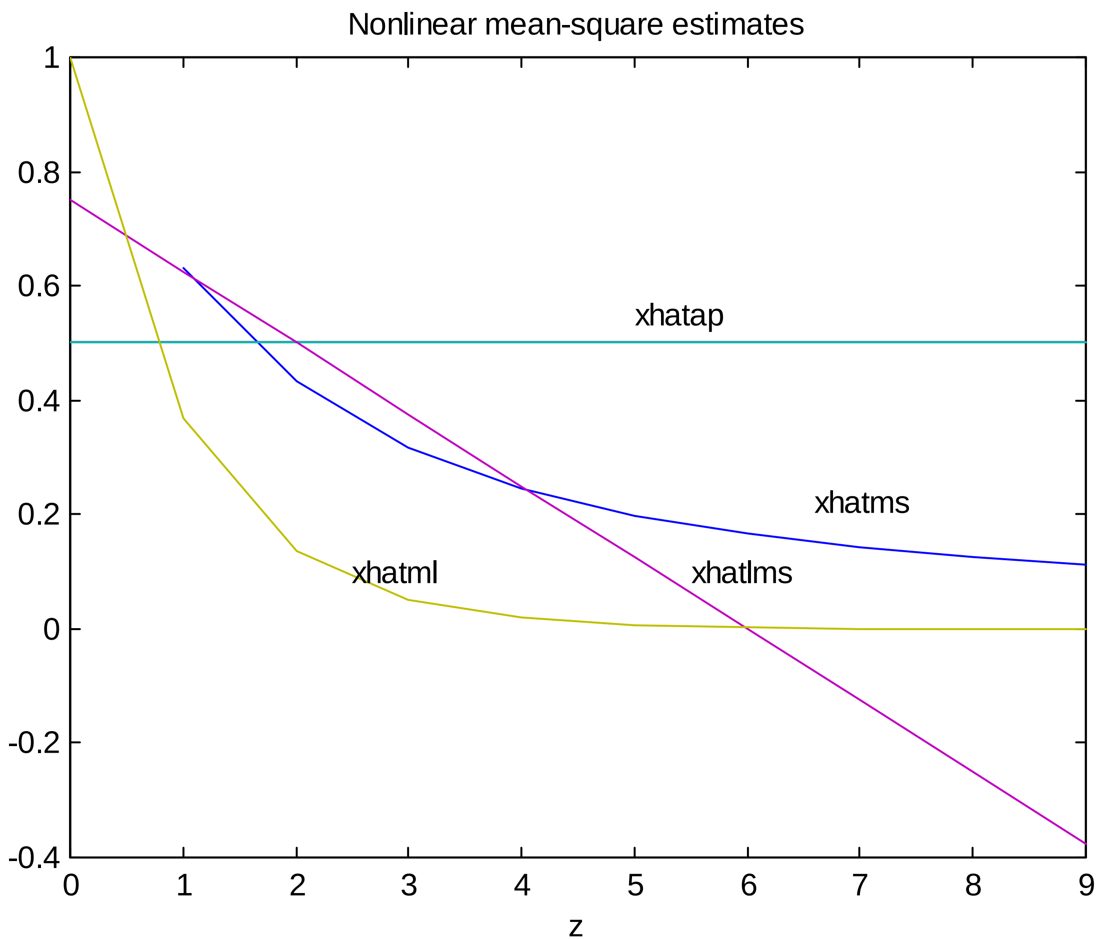
Then, the likelihood function is given by

$$\begin{aligned} f_{z|x}(z \mid x) &= f_v\left[z - \ln\left(\frac{1}{x}\right)\right] \\ &= \begin{cases} e^{-(z-\ln(1/x))}, & z - \ln\dfrac{1}{x} \geq 0 \\ 0, & z - \ln\dfrac{1}{x} < 0 \end{cases} \\ &= \begin{cases} \dfrac{1}{x}e^{-z}, & x \geq e^{-z} \\ 0, & x < e^{-z} \end{cases} \end{aligned}$$

Find x that maximizes the likelihood function $f_{z/x}(z/x)$.

$$\hat{x}_{ML} = e^{-z}.$$

Note that "xhatml" is sensitive to z in the following figure.


Nonlinear mean-square estimates
xhatap
xhatms
xhatml
xhatlms
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
1
2
3
4
5
6
7
8
9
z

The Cramer-Rao Bound

Theorem: (Cramer-Rao Bound)

If $\hat{X}$ is any unbiased estimate of a deterministic, scalar variable *X* based on measurement *Z*, then the covariance of the estimation error $\tilde{X} = X - \hat{X}$ is bounded by

$$P_{\tilde{X}} \geq J_F^{-1} \tag{2.2-3}$$

where the Fisher information matrix is given by

$$J_F = E\left[\frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X}\right]^2 = -E\left\{\frac{\partial^2 \ln f_{Z|X}\left(z \mid x\right)}{\partial X^2}\right\}. \tag{2.2-4}$$

It is assumed that $\frac{\partial f_{Z|X}}{\partial X}$ and $\frac{\partial^2 f_{Z|X}}{\partial X^2}$ exist and are absolutely integrable.

Proof: (Ref. H.L. Van Trees, pp. 66-68)

Because $\hat{X}(Z)$ is unbiased,

$$E\left[\widehat{X}(Z) - X\right] = \int_{-\infty}^{\infty}\left[\hat{X}(Z) - X\right] f_{Z|X}\left(z \mid x\right) dZ = 0. \tag{2.2-5}$$

Differentiating both sides with respect to *X*, we have

$$\frac{d}{dX}\int_{-\infty}^{\infty}\left[\widehat{X}(Z) - X\right] f_{Z|X}\left(z \mid x\right) dZ$$

$$= \int_{-\infty}^{\infty} \frac{\partial}{\partial X} \left\{ \left[\hat{X}(Z) - X \right] f_{Z|X}\left(z \mid x\right) \right\} dZ = 0, \qquad (2.2\text{-}6)$$

where the assumption in the theorem allows us to bring the differentiation inside the integral. Then,

$$-\int_{-\infty}^{\infty} f_{Z|X}\left(z \mid x\right) dZ + \int_{-\infty}^{\infty} \frac{\partial f_{Z|X}\left(z \mid x\right)}{\partial X} \left[\hat{X}(Z) - X \right] dZ = 0. \qquad (2.2\text{-}7)$$

The first integral is just 1. Now observe that

$$\begin{aligned} \frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X} &= \frac{1}{f_{Z|X}\left(z \mid x\right)} \frac{\partial f_{Z|X}\left(z \mid x\right)}{\partial X} \\ \frac{\partial f_{Z|X}\left(z \mid x\right)}{\partial X} &= \frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X} f_{Z|X}\left(z \mid x\right). \end{aligned} \qquad (2.2\text{-}8)$$

Substituting Eq. (2.2-8) into Eq. (2.2-7), we have

$$\int_{-\infty}^{\infty} \frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X} f_{Z|X}\left(z \mid x\right) \left[\hat{X}(Z) - X \right] dZ = 1. \qquad (2.2\text{-}9)$$

Rewriting, we have

$$\int_{-\infty}^{\infty} \left[\frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X} \sqrt{f_{Z|X}\left(z \mid x\right)} \right] \left[\sqrt{f_{Z|X}\left(z \mid x\right)} \left[\hat{X}(Z) - X \right] \right] dZ = 1, \qquad (2.2\text{-}10)$$

And, using the Schwarz inequality, we have

$$\int_{-\infty}^{\infty} \left[\frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X} \right]^2 f_{Z|X}\left(z \mid x\right) dZ \int_{-\infty}^{\infty} \left[\hat{X}(Z) - X \right]^2 f_{Z|X}\left(z \mid x\right) dZ \geq 1, \qquad (2.2\text{-}11)$$

where we recall from the derivation of the Schwarz inequality that equality holds if and only if

$$\frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X} = \left[\hat{X}(Z) - X\right]k(X), \tag{2.2-12}$$

for all *Z* and *X*. We see that the two terms of the left side of Eq. (2.2-11) are the expectations in the theorem. Thus,

$$E\left\{\left[\hat{X}(Z) - X\right]^2\right\} \geq \left\{E\left[\frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X}\right]^2\right\}^{-1}. \tag{2.2-13}$$

To prove another equality in Eq. (2.2-4) we observe

$$\int_{-\infty}^{\infty} f_{Z|X}\left(z \mid x\right) dZ = 1.$$

Differentiating with respect to *X*, we have

$$\int_{-\infty}^{\infty} \frac{\partial f_{Z|X}\left(z \mid x\right)}{\partial X} dZ = \int_{-\infty}^{\infty} \frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X} f_{Z|X}\left(z \mid x\right) dZ = 0.$$

Differentiating again with respect to *X* and applying Eq. (2.2-8), we obtain

$$\int_{-\infty}^{\infty} \frac{\partial^2 \ln f_{Z|X}\left(z \mid x\right)}{\partial X^2} f_{Z|X}\left(z \mid x\right) dZ + \int_{-\infty}^{\infty} \left(\frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X}\right)^2 f_{Z|X}\left(z \mid x\right) dZ = 0$$

or

$$E\left[\frac{\partial \ln f_{Z|X}\left(z \mid x\right)}{\partial X}\right]^2 = -E\left\{\frac{\partial^2 \ln f_{Z|X}\left(z \mid x\right)}{\partial X^2}\right\}, \tag{2.2-14}$$

which together with Eq. (2.2-13) completes the proof.

Linear Gaussian Measurements

Find $\hat{x}_{ML}$ for

$$z = Hx + v$$

where

$$E\{v\} = 0$$

$$f_v(v) = \frac{1}{\sqrt{(2\pi)^p |R|}} e^{-1/2 v^T R^{-1} v}. \tag{2.2-15}$$

The likelihood function is given by

$$\begin{aligned} f_{z|x}(z \mid x) &= f_v(z - Hx) \\ &= \frac{1}{\sqrt{(2\pi)^p |R|}} e^{-1/2(z-Hx)^T R^{-1}(z-Hx)} \end{aligned} \tag{2.2-16}$$

$f_{z|x}$ will be maximized when J becomes a minimum

$$J = 1/2(z - Hx)^T R^{-1}(z - Hx). \tag{2.2-17}$$

Find x that minimizes Eq. (2.1-17)

$$\frac{\partial J}{\partial x} = H^T R^{-1}(z - Hx) = 0.$$

The solution is the maximum likelihood estimate of x

$$\hat{x}_{ML} = (H^T R^{-1} H)^{-1} H^T R^{-1} z. \qquad (2.2\text{-}18)$$

Its estimation error is given by

$$\begin{aligned}\tilde{x} = x - \hat{x} &= x - (H^T R^{-1} H)^{-1} H^T R^{-1}(Hx + v) \\ &= -(H^T R^{-1} H)^{-1} H^T R^{-1} v.\end{aligned}$$

The error mean is

$$E\{\tilde{x}\} = -(H^T R^{-1} H)^{-1} H^T R^{-1} E\{v\} = 0.$$

The error covariance is

$$\begin{aligned}P_{\tilde{x}} &= E\{\tilde{x}\tilde{x}^T\} \\ &= (H^T R^{-1} H)^{-1} H^T R^{-1} E\{vv^T\} R^{-1} H (H^T R^{-1} H)^{-1} \qquad (2.2\text{-}19) \\ &= (H^T R^{-1} H)^{-1}.\end{aligned}$$

Recursive estimation

ML from *k* measurements

$$Z_k = H_k x + v_k, \quad v_k \approx N(0, R_k), \quad R_k = diag\{\sigma_1, \sigma_2, \cdots, \sigma_k\}$$

$$P_k = \left(H_k^T R_k^{-1} H_k\right)^{-1} \tag{2.2-20}$$

$$\hat{x}_k = P_k H_k^T R_k^{-1} Z_k$$

In case of *k+1* measurements,

$$Z_{k+1} = H_{k+1} x + v_{k+1}, \quad v_{k+1} \approx N(0, R_{k+1})$$

$$Z_{k+1} = \begin{bmatrix} H_k \\ h_{k+1}^T \end{bmatrix} x + v_{k+1}, \quad v_{k+1} \approx N\left(0, \begin{bmatrix} R_k & 0 \\ 0 & \sigma_{k+1}^2 \end{bmatrix}\right)$$

$$P_{k+1} = \left(\begin{bmatrix} H_k \\ h_{k+1}{}^T \end{bmatrix}^T \begin{bmatrix} R_k & 0 \\ 0 & \sigma_{k+1}{}^2 \end{bmatrix}^{-1} \begin{bmatrix} H_k \\ h_{k+1}{}^T \end{bmatrix} \right)^{-1}$$

$$= \left(H_k{}^T R_k{}^{-1} H_k + \frac{h_{k+1} h_{k+1}{}^T}{\sigma_{k+1}{}^2} \right)^{-1} \qquad (2.2\text{-}21)$$

$$= \left(P_k{}^{-1} + \frac{h_{k+1} h_{k+1}{}^T}{\sigma_{k+1}{}^2} \right)^{-1}$$

Or, using the matrix inversion lemma,

$$P_{k+1} = P_k - P_k h_{k+1} \left(h_{k+1}{}^T P_k h_{k+1} + \sigma_{k+1}{}^2 \right)^{-1} h_{k+1}{}^T P_k$$

On the other hand,

$$
\begin{aligned}
\hat{x}_{k+1} &= P_{k+1} H_{k+1}{}^T R_{k+1}{}^{-1} Z_{k+1} \\
&= P_{k+1} \begin{bmatrix} H_k \\ h_{k+1}{}^T \end{bmatrix}^T \begin{bmatrix} R_k & 0 \\ 0 & \sigma_{k+1}{}^2 \end{bmatrix}^{-1} \begin{bmatrix} Z_k \\ z_{k+1} \end{bmatrix} \\
&= P_{k+1} \left(H_k{}^T R_k{}^{-1} Z_k + \frac{h_{k+1}}{\sigma_{k+1}{}^2} z_{k+1} \right) \\
&= P_{k+1} \left(P_k{}^{-1} \hat{x}_k + \frac{h_{k+1}}{\sigma_{k+1}{}^2} z_{k+1} \right)
\end{aligned}
\qquad (2.2\text{-}22)
$$

Applying Eq. (2.2-21), we obtain,

$$
\begin{aligned}
P_{k+1} P_k{}^{-1} &= P_{k+1} \left(P_{k+1}{}^{-1} - \frac{h_{k+1} h_{k+1}{}^T}{\sigma_{k+1}{}^2} \right) = I - P_{k+1} \frac{h_{k+1} h_{k+1}{}^T}{\sigma_{k+1}{}^2} \\
\hat{x}_{k+1} &= \hat{x}_k + P_{k+1} \frac{h_{k+1}}{\sigma_{k+1}{}^2} \left(z_{k+1} - h_{k+1}{}^T \hat{x}_k \right)
\end{aligned}
\qquad (2.2\text{-}23)
$$

Eqs. (2.2-21) and (2.2-23) represent the Recursive ML.

Initial conditions, $P_0^{-1} = 0,\ \hat{x}_0 = 0$ (Complete ignorance of the a priori statistics of *x*.)

Note: Eqs. (2.2-21) and (2.2-23) with initial conditions $\hat{x}_0 = \bar{x},\ P_0 = P_x$ represent the recursive MS.

Prewhitening of data

Suppose *R* in (2.2-20) is nonsingular but non-diagonal, representing colored noise. We need to diagonalize (prewhitening) *R* before applying the recursive ML.

Let *S* be any square root of R^{-1} so that $R^{-1} = S^T S$. (S is a square matrix.)

Multiply *S* to the measurement equation to obtain,

$$S\,z = S\,H\,x + S\,v$$

and define new quantities by

$$z^W = S\,z$$
$$H^W = S\,H$$
$$v^W = S\,v$$

so that

$$z^W = H^W x + v^W .$$

The covariance of the prefiltered noise v^W is then

$$R^W = E\left\{Sv(Sv)^T\right\} = E\left\{S\left(vv^T\right)S^T\right\}$$
$$= SRS^T = S\left(S^T S\right)^{-1} S^T = I$$

Now derive Eqs. (2.2-21) and (2.2-22) using $z^W, H^W \ and \ R^W = I$ instead of *z, H, R*.

2.3 Maximum *a posteriori* Estimation (MAP)

Suppose that we have

$$\text{z} = g(\text{x},\text{v})$$

where z, x, and v are random variables. Given the observation z, the MAP is given by

$$\hat{x}_{MAP} = \text{value of } x \text{ that maximizes } f_{\text{x}}\left(x \middle| \text{z} = z\right). \qquad (2.3\text{-}1)$$

Assuming $f_{\text{x}}\left(x \middle| \text{z} = z\right)$ is differentiable and has a unique maximum in the interior of its domain, we have

$$\hat{x}_{MAP} = \text{value of } x \text{ for which } \frac{\partial f_{\text{x}}\left(x \middle| \text{z} = z\right)}{\partial x} = 0. \qquad (2.3\text{-}2)$$

By Bayes' formula,

$$f_{\text{x}}\left(x \middle| \text{z} = z\right) = \frac{f_{\text{z}}\left(z \mid \text{x} = x\right) f_{\text{x}}(x)}{f_{\text{z}}(z)}.$$

Since z is given, $f_{\text{z}}(z)$ is constant and thus independent of x, we may express $\hat{x}_{MAP}$ as

$$\hat{x}_{MAP} = \text{value of } x \text{ that maximizes } f_{\text{z}}\left(z \middle| \text{x} = x\right) f_{\text{x}}(x). \qquad (2.3\text{-}3)$$

MAP Estimation with Gaussian Noise

Consider the additive-noise case $\mathrm{z} = \mathrm{x} + \mathrm{v}$, where $\mathrm{v} \sim N(0, \sigma_{\mathrm{v}}^{2})$ and $\mathrm{v} \sim N(0, \sigma_{\mathrm{v}}^{2})$. Then,

$$f_{\mathrm{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-\eta_x)^2/2\sigma_x^{2}},$$

and

$$f_{\mathrm{z}}\left(z \middle| \mathrm{x} = x\right) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-(z-x)^2/2\sigma_v^{2}}.$$

Hence,

$$f_{\mathrm{z}}\left(z \middle| \mathrm{x} = x\right) f_{\mathrm{x}}(x) = \frac{1}{2\pi\sigma_x\sigma_v} \exp\left[-\frac{(z-x)^2}{2\sigma_v^{2}} - \frac{(x-\eta_x)^2}{2\sigma_x^{2}} \right].$$

The above is maximized when the bracketed term is minimized. Differentiating the term with respect to *x* and setting it to zero gives

$$\hat{x}_{MAP} = \frac{\sigma_v^{\ 2}}{\sigma_v^{\ 2} + \sigma_x^{\ 2}} \eta_x + \frac{\sigma_x^{\ 2}}{\sigma_v^{\ 2} + \sigma_x^{\ 2}} z$$

or

$$\hat{x}_{MAP} = \eta_x + \frac{\sigma_x^{\ 2}}{\sigma_v^{\ 2} + \sigma_x^{\ 2}} (z - \eta_x). \qquad (2.3\text{-}4)$$

Observe that when $\sigma_v^{\ 2} \ll \sigma_x^{\ 2}$ (the noise power is much less than the signal power), the MAP estimate becomes

$$\hat{x}_{MAP} = \eta_x + \frac{\sigma_x^{\ 2}}{\sigma_x^{\ 2}} (z - \eta_x) = z = \hat{x}_{ML}.$$