ELEC-E8107 - Stochastic models, estimation and control

Arto VISALA, and Issouf OUATTARA

September 26, 2022

Exercises Session 1

Exercise 1

$p(z|x) =$

A parameter x is measured with correlated (rather than independent) additive Gaussian noises, such that:

$$\underline{z_k = x + w_k,}$$

where, $k = 1 \ldots n$ and the mean value of the noise at instant k is $E[w_k] = 0$, while the variances are:

$$E[w_k w_j] = \begin{cases} 1, & \text{if } k = j \\ \rho, & \text{if } |k - j| = 1 \\ 0, & \text{if } |k - j| > 1 \end{cases}$$

$$p(x|z) = \frac{p(z|x)\, p(x)}{p(z)}$$

For $n = 2$:

1. Compute the likelihood function of the parameter x
2. Find the MLE of x.
3. Find the CRLB for the estimation of x.
4. Is the MLE efficient?

Solution Exercise 1

1. The likelihood function

The likelihood function is the probability density function of the measurement conditioned on the parameter of interest. For $n = 2$, the likelihood function is given by:

$$\begin{aligned}\Lambda(x) &= p(z|x) \\ &= p(z_1, z_2|x) \\ &= ce^{-\frac{1}{2}Q(x)}\end{aligned}$$

With:

$$\begin{aligned}Q(x) &= (z-x)^T P^{-1}(z-x) \\ &= \begin{bmatrix} z_1 - x & z_2 - x \end{bmatrix} P^{-1} \begin{bmatrix} z_1 - x \\ z_2 - x \end{bmatrix}\end{aligned}$$

Where P is the covariance matrix associated with the noise vector $w = [w_1 \quad w_2]^T$. P is computed as:

$P = E[(w-\bar{w})(w-\bar{w})^T]$

$E[w_1 w_1] = 1$

$$\begin{aligned}P &= E[ww^T] \\ &= E\left[\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \cdot \begin{bmatrix} w_1 & w_2 \end{bmatrix}\right] \\ &= E\left[\begin{bmatrix} w_1 w_1 & w_1 w_2 \\ w_2 w_1 & w_2 w_2 \end{bmatrix}\right]\end{aligned}$$

From the problem statement we have: $E[w_1 w_1] = E[w_2 w_2] = 1$ and $E[w_1 w_2] = E[w_2 w_1] = \rho$. So

$$P = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

And

$$P^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \qquad (\forall \rho \neq 1)$$

2. The MLE of x

The maximum likelihood estimate of the parameter x is given by:

$$\hat{x}_{ML} = \arg\max_x \Lambda(x) = \arg\min_x Q(x)$$

$$y = f(x)\, g(x)$$

$$\frac{\partial y}{\partial x} = \frac{\partial f(x)}{\partial x} g + f \cdot \frac{\partial g}{\partial x}$$

$$(P^{-1})^T = P^{-1}$$

$\hat{x}_{ML}$ can be found by taking the derivatives of $Q(x)$ with respect to the parameter x and set it to zero:

$$\begin{aligned}
\frac{\partial}{\partial x}Q(x) &= \frac{\partial}{\partial x}\left(\begin{bmatrix} z_1 - x & z_2 - x \end{bmatrix} P^{-1} \begin{bmatrix} z_1 - x \\ z_2 - x \end{bmatrix}\right) \\
&= \begin{bmatrix} \frac{\partial(z_1-x)}{\partial x} & \frac{\partial(z_2-x)}{\partial x} \end{bmatrix} P^{-1} \begin{bmatrix} z_1 - x \\ z_2 - x \end{bmatrix} + \begin{bmatrix} z_1 - x & z_2 - x \end{bmatrix} P^{-1} \begin{bmatrix} \frac{\partial(z_1-x)}{\partial x} \\ \frac{\partial(z_2-x)}{\partial x} \end{bmatrix} \\
&= \begin{bmatrix} -1 & -1 \end{bmatrix} P^{-1} \begin{bmatrix} z_1 - x \\ z_2 - x \end{bmatrix} + \begin{bmatrix} z_1 - x & z_2 - x \end{bmatrix} P^{-1} \begin{bmatrix} -1 \\ -1 \end{bmatrix}
\end{aligned}$$

Note that Q(x) is a scalar. And the derivative with respect to a scalar is also a scalar. In the last expression of the derivative, the two terms are scalars and transpose of each other (It is the case because P is a symmetric matrix, $P = P^T$). We can write the following:

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} z_1 - x \\ z_2 - x \end{bmatrix}$$

$$\begin{bmatrix} 1-\rho & 1-\rho \end{bmatrix} \begin{bmatrix} z_1 - x \\ z_2 - x \end{bmatrix} = (1-\rho)(z_1 - x + z_2 - x)$$

$$\begin{aligned}
\frac{\partial}{\partial x}Q(x) &= -2 \begin{bmatrix} 1 & 1 \end{bmatrix} P^{-1} \begin{bmatrix} z_1 - x \\ z_2 - x \end{bmatrix} = 0 \\
&= -2(z_1 + z_2 - 2x)\frac{1-\rho}{1-\rho^2} = 0
\end{aligned}$$

$$\frac{\partial^2 \phi(x)}{\partial x^2} = 4\,\frac{1-\rho}{1-\rho^2} = 4\,\frac{1-\rho}{(1-\rho)(1+\rho)} = 4\,\frac{1}{1+\rho}$$

Then $\hat{x}_{ML} = \frac{1}{2}(z_1 + z_2) \qquad \forall |\rho| \neq 1$

3. The CRLB of the estimation According to the lecture slide, and the textbook of the course (page 109), the CRLB can be computed by taking the second derivative of the likelihood function with respect to the variable that is being estimated.

$$\frac{\partial^2(-\frac{1}{2}\phi(x))}{\partial x^2} = \frac{-2}{1+\rho}$$

$$\begin{aligned}
J &= -E\left[\frac{\partial^2 ln\Lambda(x)}{\partial x^2}\right]|_{x=x_0} \\
J &= -E\left[-\cancel{\tfrac{1}{2}Q}(x)\right] = -E\left[\frac{\partial^2(-\frac{1}{2}\phi(x))}{\partial x^2}\right] \\
J &= \frac{2}{1+\rho} \\
J^{-1} &= \frac{1+\rho}{2}
\end{aligned}$$

4. Efficiency of the MLE

$$\begin{aligned}
\hat{x}_{ML} &= \frac{1}{2}(z_1 + z_2) \\
&= \frac{1}{2}(2x + w_1 + w_2) \\
&= x + \frac{1}{2}(w_1 + w_2)
\end{aligned}$$

By taking the mean of the above we have $E[\hat{x}_{ML}] = E[x]$, the estimator is unbiased. The variance of the estimator is:

$$\sigma^2 = E[\cancel{\frac{1}{2}(w_1 + w_2)^2}] \qquad E = \left[\left(\frac{1}{2}(w_1+w_2)\right)^2\right]$$

$$= \frac{1}{4}E[(w_1 + w_2)^2] \quad = \frac{1}{4}\left(E[w_1^2) + E[w_2^2] + 2E[w_1 w_2]\right) \quad \text{(see the definition of P for each terms}$$

$$= \frac{1+\rho}{2} = J^{-1} \qquad \frac{1}{4}(1+1+2\rho)$$

The MLE is efficient.

Exercise 2

Given $z = x + w$, where all the variables are n-vectors, with:

$$w \sim \mathcal{N}(0, P) \qquad x \sim \mathcal{N}(\bar{x}, P_0)$$

x and w are independent. Find the MAP estimator of x in terms of z and the covariance of this estimator.

Solution Exercise 2

The posterior pdf is given by:

$$C_1 e^{-\frac{1}{2}(z-x)^T P^{-1}(z-x)}$$

$$C_2 e^{-\frac{1}{2}(x-\bar{x})^T P_0^{-1}(x-\bar{x})}$$

$$\begin{aligned}
p(x|z) &= \frac{p(z|x)p(x)}{p(z)} \\
&= \frac{\mathcal{N}(z, x, P)\mathcal{N}(x, \bar{x}, P_0)}{p(z)} \\
&\propto \frac{exp\left(\frac{-1}{2}\left((z-x)^T P^{-1}(z-x) + (x-\bar{x})^T P_0^{-1}(x-\bar{x})\right)\right)}{p(z)} \\
&\propto \frac{exp\left(\frac{-1}{2}Q(x)\right)}{p(z)}
\end{aligned}$$

with $Q(x) = (z-x)^T P^{-1}(z-x) + (x-\bar{x})^T P_0^{-1}(x-\bar{x})$.
The MAP is defined as:

$$\begin{aligned}
\hat{x}_{MAP} &= \arg\max_x p(x|z) \\
&= \arg\min Q(x)
\end{aligned}$$

$$P_0(P_0+P)^{-1}$$

$$z = x + w$$

$$E[z] = E[x] + \overset{0}{E[w]}$$

$$x_{MAP} = \bar{x} + P_{xz}P_{zz}^{-1}(z-\bar{z})$$

$$x_{MAP} = \bar{x} + P_0(P+P_0)^{-1}(z-\bar{x})$$

$$\begin{aligned}
P_{zz} &= E[(z-\bar{z})(z-\bar{z})^T] \\
&= E[(x+w-\bar{x})(x+w-\bar{x})^T] \\
&= E[((x-\bar{x})+w)((x-\bar{x})+w)^T] = E[\overbrace{(x-\bar{x})(x-\bar{x})^T}^{P_0} + \overbrace{(x-\bar{x})w^T}^{0} + \overbrace{w(x-\bar{x})^T}^{0} + \overbrace{ww^T}^{P}]
\end{aligned}$$

$\cdot z^{-1} + \underline{Px + P_0^{-1}x} - P_0^{-1}\bar{x} = (P^{-1}+P_0^{-1})x - P^{-1}z - P_0^{-1}\bar{x} = 0$

$\Rightarrow (P^{-1}+P_0^{-1})x = P^{-1}z + P_0^{-1}\bar{x}$

$\hat{x}_{MAP} =$

$\hat{x} = (P^{-1}+P_0^{-1})^{-1}(P^{-1}z + P_0^{-1}\bar{x})$

This can be achieved by taking the gradient of $Q(x)$ with respect to x and setting it to zero.

$$\nabla_x Q(x) = -P^{-1}(z-x) + P_0^{-1}(x-\bar{x}) = 0$$

Which results in:

$$\hat{x}_{MAP} = (P^{-1} + P_0^{-1})^{-1}(P^{-1}z + P_0^{-1}\bar{x}) = \bar{x} + P_0(P+P_0)^{-1}(z-\bar{x})$$

The covariance is given by (see 1.4.14-18):

$$\begin{aligned} cov(\hat{x}_{MAP}) &= P_{xx} - P_{xz}P_{zz}^{-1}P_{zx} \\ &= P_0 - P_0(P+P_0)^{-1}P_0 \end{aligned}$$

Exercise 3

The model for a vehicle moving at a constant speed is $y_i = vt_i + e_i$. The position is measured as a function of time as shown in the Table below:

Time	0	1	2	3	4	10	12	18
Distance	4.71	9	15	19	20	45	55	78

The noise e_i are such that $E[e_i^2] = R_i = 0.9^{8-i}$. Use the batch least square method to estimate the velocity. (Write a Matlab script)

Solution Exercise 3

The observation model $y_i = vt_i + e_i$ can be written as in the lecture textbook:

$$z(i) = H(i)x + w(i)$$

According to (3.4.1-9, page 130 textbook), the solution of the batch least square estimate is given by:

$$\hat{x}(k) = \underline{\left[H^{k'}(R^k)^{-1}H^k\right]^{-1} H^{k'}(R^k)^{-1}z^k}$$

This can be computed using the following matlab script:

```
%% Batch LS Estimator.
time = [0 1 2 3 4 10 12 18]'; H = time;
Dist = [4.71 9 15 19 20 45 55 78]'; z = Dist;
invR = diag(0.9.^(7:-1:0));
% The estimate of the speed.
v_batch = (H'*invR*H)^-1*H'*invR*z;
```

Exercise 4

Use the recursive least square method to estimate the velocity in the above problem. (Write a Matlab script)

Solution Exercise 4

The recursive least square algorithm is given in lecture slide 2b and in the textbook (page 132 to 134). The following matlab script compute the recursive least square solution:

```
%% Recursive LS Estimator
P = 1e6; % Initial covariance.
v_recursive = 0; % First estimate.
v_recursive_plot = zeros(1,8);
P_vector = zeros(1,8);
% LS estimator algorithm
for i = 1:8
    S = H(i,:)*P*H(i,:)' + inv(invR(i,i));
    W = P*H(i,:)'*S^-1;
    P = P - W*S*W';
    P_vector(i) = P;
    v_recursive = v_recursive + W*(z(i)-H(i,:)*v_recursive);
    % Recursive speed estimation each recursion
    v_recursive_plot(i) = v_recursive;
end
```

Figure 1 shows the comparison of the two different techniques to compute the LS estimator. In the recursive algorithm, the **information** at instant $k+1$ equals the sum of the information at k and the new information about x obtained from new measurement $z(k+1)$.

Notice that the LS estimate $\hat{x}$ values from both algorithms become similar at the last time instant. Also, the recursive algorithm requires an initial values of the $\hat{x}$ and covariance matrix P. In this case, we selected $\hat{x}_{\text{init}} = 0$ with $P = 1.0 \times 10^6$. Setting a high initial covariance (of the LS estimate) means that we do not know anything about the process at the start.

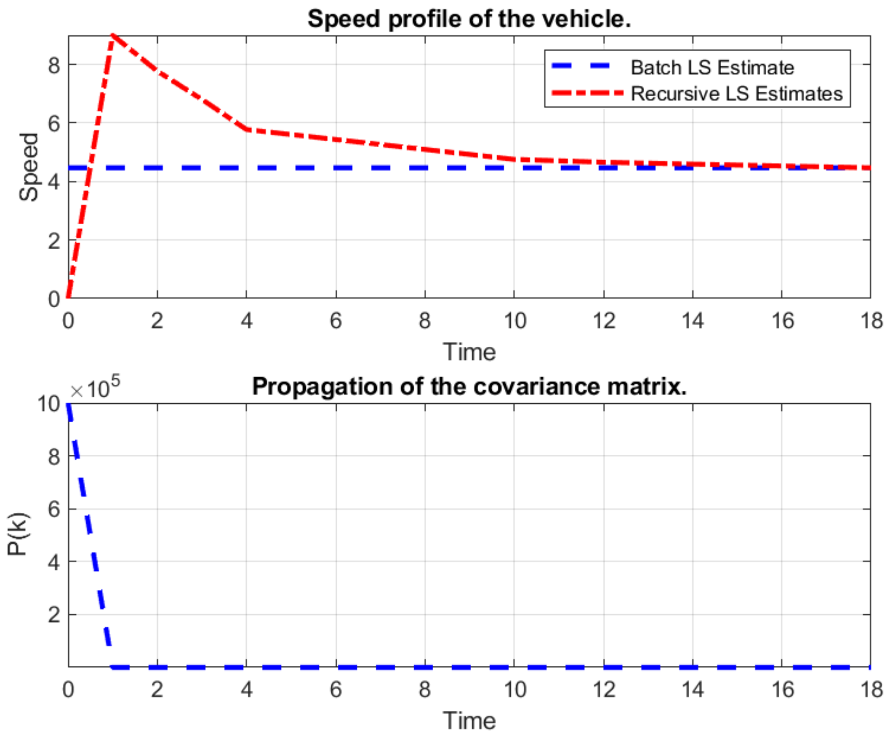

Figure 1: Top: Comparison of Batch and Recursive LS estimators. Bottom: Illustration of how the covariance evolves as the LS estimator does the estimation.