Recursive Estimation

Raffaello D’Andrea

Spring 2019

Problem Set 3:
Extracting Estimates from Probability Distributions

Last updated: March 27, 2019

Notes:

• **Notation:** Unless otherwise noted, \( x, y, \) and \( z \) denote random variables, \( p_x \) denotes the probability density function of \( x \), and \( p_{x|y} \) denotes the conditional probability density function of \( x \) conditioned on \( y \). Note that shorthand (as introduced in Lecture 1 and 2) and longhand notation is used. The expected value of \( x \) and its variance is denoted by \( \mathbb{E}[x] \) and \( \text{Var}(x) \), and \( \text{Pr}(Z) \) denotes the probability that the event \( Z \) occurs. A normally distributed random variable \( x \) with mean \( \mu \) and variance \( \sigma^2 \) is denoted by \( x \sim \mathcal{N}(\mu, \sigma^2) \).

• Please report any errors found in this problem set to the teaching assistants (hofermat@ethz.ch or csferrazza@ethz.ch).
Problem Set

Problem 1
Let $x$ and $w$ be scalar CRVs, where $x, w$ are independent: $p(x, w) = p(x) \cdot p(w)$, and $w \in [-1, 1]$ is uniformly distributed. Let

$$ z = x + 3w. $$

a) Calculate $p(z|x)$ using the change of variables.

b) Let $x = 1$. Verify that $Pr(z \in [1, 4]|x = 1) = Pr(w \in [0, 1]|x = 1)$.

Problem 2
Let $w \in \mathbb{R}^2$ and $x \in \mathbb{R}^4$ be CRVs, where $x$ and $w$ are independent: $p(x, w) = p(x) \cdot p(w)$. The two elements of the measurement noise vector $w$ are $w_1, w_2$, and the PDF of $w$ is uniform:

$$ p_w(w) = p_{w_1, w_2}(w_1, w_2) = \begin{cases} 1/4 & \text{for } -1 \leq w_1 \leq 1 \text{ and } -1 \leq w_2 \leq 1 \\ 0 & \text{otherwise}. \end{cases} $$

The measurement model is

$$ z = Hx + Gw $$

where

$$ H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1/2 & 0 \\ 0 & -3 \end{bmatrix}. $$

Calculate the conditional joint PDF $p_{z|x}$ using the change of variables.

Problem 3
Find the maximum likelihood (ML) estimate $\hat{x}_{\text{ML}} := \arg \max_x p(z|x)$ when

$$ z_i = x + w_i, \quad i = 1, 2 $$

where $x, w_i$ are mutually independent, with $w_i \sim \mathcal{N}(0, \sigma_i^2)$ and $z_i, w_i, x \in \mathbb{R}$.

Problem 4
Solve Problem 3 if the joint PDF of $w$ is

$$ p_w(\bar{w}) = p_{w_1, w_2}(\bar{w}_1, \bar{w}_2) = \begin{cases} 1/4 & \text{for } -1 \leq \bar{w}_1 \leq 1 \text{ and } -1 \leq \bar{w}_2 \leq 1 \\ 0 & \text{otherwise}. \end{cases} $$
Problem 5

Let \( x \in \mathbb{R}^n \) be an unknown, constant parameter. Furthermore, let \( w \in \mathbb{R}^m, n \leq m \), represent the measurement noise with its joint PDF defined by the multivariate Gaussian distribution

\[
p_w(\bar{w}) = \frac{1}{(2\pi)^{m/2} (\det(\Sigma))^{1/2}} \exp \left( -\frac{1}{2} \bar{w}^T \Sigma^{-1} \bar{w} \right)
\]

with mean \( \mathbb{E}[w] = 0 \) and variance \( \mathbb{E}[ww^T] = \Sigma = \Sigma^T \in \mathbb{R}^{m \times m} \). Assume that \( x \) and \( w \) are independent. Let the measurement be

\[ z = Hx + w \]

where the matrix \( H \in \mathbb{R}^{m \times n} \) has full column rank.

a) Calculate the ML estimate \( \hat{x}_{\text{ML}} := \arg \max_x p(z|x) \) for a diagonal variance

\[
\Sigma = \begin{bmatrix}
\sigma_1^2 & 0 & \cdots & 0 \\
0 & \sigma_2^2 & \cdots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & \sigma_m^2
\end{bmatrix}
\]

Note that in this case, the elements \( w_i \) of \( w \) are mutually independent.

b) Calculate the ML estimate for non-diagonal \( \Sigma \), which implies that the elements \( w_i \) of \( w \) are correlated.

Hint: The following matrix calculus is useful for this problem (see, for example, the Matrix Cookbook for reference):

Recall the definition of a Jacobian matrix from the lecture: Let \( g(x) : \mathbb{R}^n \to \mathbb{R}^m \) be a function that maps the vector \( x \in \mathbb{R}^n \) to another vector \( g(x) \in \mathbb{R}^m \). The Jacobian matrix is defined as

\[
\frac{\partial g(x)}{\partial x} := \begin{bmatrix}
\frac{\partial g_1(x)}{\partial x_1} & \frac{\partial g_1(x)}{\partial x_2} & \cdots & \frac{\partial g_1(x)}{\partial x_n} \\
\frac{\partial g_2(x)}{\partial x_1} & \frac{\partial g_2(x)}{\partial x_2} & \cdots & \frac{\partial g_2(x)}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial g_m(x)}{\partial x_1} & \frac{\partial g_m(x)}{\partial x_2} & \cdots & \frac{\partial g_m(x)}{\partial x_n}
\end{bmatrix} \in \mathbb{R}^{m \times n}
\]

where \( x_j \) is the \( j \)-th element of \( x \) and \( g_i(x) \) the \( i \)-th element of \( g(x) \).

Derivative of a quadratic form:

\[
\frac{\partial}{\partial x} x^T A x = x^T (A + A^T) \in \mathbb{R}^{1 \times n}, \quad A \in \mathbb{R}^{n \times n}.
\]

Chain rule:

\[
u = g(x), \quad x \in \mathbb{R}^n, u \in \mathbb{R}^n
\]

\[
\frac{\partial}{\partial x} u^T A u = u^T (A + A^T) \frac{\partial u}{\partial x}, \quad \frac{\partial u}{\partial x} = \frac{\partial g(x)}{\partial x} \in \mathbb{R}^{n \times n}.
\]
Problem 6
Find the maximum a posteriori (MAP) estimate \( \hat{x}_{\text{MAP}} := \arg \max_x p(z|x) p(x) \) when
\[
z = x + w, \quad x, z, w \in \mathbb{R}
\]
where \( x \) and \( w \) are independent, \( x \) is exponentially distributed,
\[
p_x(\bar{x}) = \begin{cases} 0 & \text{for } \bar{x} < 0 \\ e^{-\bar{x}} & \text{for } \bar{x} \geq 0 \end{cases}
\]
and \( w \sim \mathcal{N}(0, 1) \).

Problem 7
In the derivation of the recursive least squares algorithm, we used the following result:
\[
\frac{\partial \text{trace} (ABA^T)}{\partial A} = 2AB \quad \text{if } B = B^T
\]
where \( \text{trace} (\cdot) \) is the sum of the diagonal elements of a matrix. Prove that the above holds for the two-dimensional case, i.e. \( A, B \in \mathbb{R}^{2 \times 2} \).

Problem 8
In the derivation of the recursive least squares algorithm, we also used the following result
\[
\frac{\partial \text{trace} (AB)}{\partial A} = B^T.
\]
Prove that the above holds for \( A, B \in \mathbb{R}^{2 \times 2} \).

Problem 9
Consider the observation model, which we discussed in class,
\[
z(k) = H(k)x + w(k) \quad \text{with } z(k), w(k) \in \mathbb{R}^m, \ x \in \mathbb{R}^n,
\]
\[
\hat{x}_0 := E [x], \ P_x := E [(x - \hat{x}_0)(x - \hat{x}_0)^T] = \text{Var} [x],
\]
\[
E [w(k)] = 0, \ R(k) := \text{Var} [w(k)],
\]
where \( x \) does not change (our knowledge of it changes, however). We derived the standard Weighted Least Squares solution to the above estimation problem, but we neglected prior knowledge on \( x \). Incorporate the prior knowledge on \( x \) and formulate the estimation problem as a standard Weighted Least Squares problem.

Problem 10
Consider the recursive least squares algorithm presented in lecture, where the estimation error at step \( k \) was shown to be
\[
e(k) = \left( I - K(k)H(k) \right)e(k-1) - K(k)w(k).
\]
Furthermore, \( E [w(k)] = 0, \ E [e(k)] = 0, \) and \( \{x, w(1), \ldots\} \) are mutually independent. Show that the measurement error variance \( P(k) = \text{Var} [e(k)] = E [e(k)e^T(k)] \) is given by
\[
P(k) = \left( I - K(k)H(k) \right)P(k-1)\left( I - K(k)H(k) \right)^T + K(k)R(k)K^T(k)
\]
where \( R(k) = \text{Var} [w(k)] \).
Problem 11

Use the results from Problems 7 and 8 to derive the gain matrix $K(k)$ for the recursive least squares algorithm that minimizes the mean squared error (MMSE) by solving

$$0 = \frac{\partial}{\partial K(k)} \text{trace} \left( \left( I - K(k)H(k) \right) P(k-1) \left( I - K(k)H(k) \right)^T + K(k)R(k)K^T(k) \right).$$

Problem 12

Apply the recursive least squares algorithm when

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

where

$$E[x] = 0, \ Var[x] = 1, \quad \text{and} \quad E[w(k)] = 0, \ Var[w(k)] = 1.$$\hspace{1cm} (2)

The random variables \{x, w(1), w(2), \ldots\} are mutually independent and $z(k), w(k), x \in \mathbb{R}$.

a) Express the gain matrix $K(k)$ and variance $P(k)$ as explicit functions of $k$.

b) Simulate the algorithm for normally distributed continuous random variables $w(k)$ and $x$ with the above properties (2).

c) Simulate the algorithm for $x$ and $w(k)$ being discrete random variables that take the values 1 and $-1$ with equal probability. Note that $x$ and $w(k)$ satisfy (2). Simulate for values of $k$ up to 10000.

Problem 13

Consider the random variables and measurement equation from Problem 11 c), i.e.

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

for mutually independent $x$ and $w(k)$ that take the values 1 and $-1$ with equal probability.

a) What is the MMSE estimate of $x$ at time $k$ given all past measurements $\bar{z}(1), \bar{z}(2), \ldots, \bar{z}(k)$? Remember that the MMSE estimate is given by

$$\hat{x}^{\text{MMSE}}(k) := \arg\min_{\hat{x}} \mathbb{E}_{x \mid z(1), z(2), \ldots, z(k)} \left[ (\hat{x} - x)^2 \right].$$

b) Now consider the MMSE estimate that is restricted to the sample space of $x$, i.e.

$$\hat{x}^{\text{MMSE2}}(k) := \arg\min_{\hat{x} \in \mathcal{X}} \mathbb{E}_{x \mid z(1), z(2), \ldots, z(k)} \left[ (\hat{x} - x)^2 \right].$$

where $\mathcal{X}$ denotes the sample space of $x$. Calculate the estimate of $x$ at time $k$ given all past measurements $z(1), z(2), \ldots, z(k)$.\hspace{1cm} (3)
Sample solutions

Problem 1

a) The change of variables formula for conditional PDFs when $z = g(x, w)$ is

$$p_{z|x}(\bar{z}|\bar{x}) = \left| \frac{\partial g}{\partial w}(\bar{x}, h(\bar{z}, \bar{x})) \right| p_{w|x}(h(\bar{z}, \bar{x})|\bar{x}).$$

where $\bar{w} = h(\bar{z}, \bar{x})$ is the unique solution to $\bar{z} = g(\bar{x}, \bar{w})$. We solve $\bar{z} = g(\bar{x}, \bar{w})$ for $\bar{w}$ to find $h(\bar{z}, \bar{x})$

$$\bar{w} = \frac{\bar{z} - \bar{x}}{3} = h(\bar{z}, \bar{x})$$

and apply the change of variables:

$$p_{z|x}(\bar{z}|\bar{x}) = \left| \frac{\partial g}{\partial w}(\bar{x}, h(\bar{z}, \bar{x})) \right| = \frac{1}{3} p_{w|x} \left( \frac{1}{3}(\bar{z} - \bar{x}) \right).$$

With the given uniform PDF of $w$, the PDF of $p(z|x)$ can then be stated explicitly:

$$p_{z|x}(\bar{z}|\bar{x}) = \begin{cases} \frac{1}{6} & \text{for } \bar{x} - 3 \leq \bar{z} \leq \bar{x} + 3 \\ 0 & \text{otherwise} \end{cases}$$

b) We begin by computing $\Pr(z \in [1,4]|x = 1)$:

$$\Pr(z \in [1,4]|x = 1) = \int_1^4 p_{z|x}(\bar{z}|1) d\bar{z} = \int_1^4 \frac{1}{6} d\bar{z} = \frac{1}{2}.$$

The probability $\Pr(w \in [0,1]|x = 1)$ is

$$\Pr(w \in [0,1]|x = 1) = \Pr(w \in [0,1]) = \int_0^1 p_w(\bar{w}) d\bar{w} = \int_0^1 \frac{1}{2} d\bar{w} = \frac{1}{2}$$

which is indeed identical to $\Pr(z \in [1,4]|x = 1)$.

Problem 2

Analogous to the lecture notes, the multivariate change of variables formula for conditional PDFs is

$$p_{z|x}(\bar{z}|\bar{x}) = p_{w|x}(h(\bar{z}, \bar{x})|\bar{x}) |\text{det}(G)|^{-1}$$

where $\bar{w} = h(\bar{z}, \bar{x})$ is the unique solution to $\bar{z} = H\bar{x} + G\bar{w}$. The function $h(\bar{z}, \bar{x})$ is given by

$$\bar{w} = G^{-1} (\bar{z} - H\bar{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -\frac{1}{3} \end{bmatrix} \left( \bar{z} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \bar{x} \right)$$

$$= \begin{bmatrix} 2 \bar{z}_1 - \bar{x}_1 \\ -\frac{1}{3} (\bar{z}_2 - \bar{x}_3) \end{bmatrix}$$
where we used $x = (x_1, x_2, x_3, x_4)$ and $z = (z_1, z_2)$. Applying the change of variables, it follows that

\[
p_{z|\bar{x}}(\bar{z}|\bar{x}) = p_{w|x}(h(\bar{z}, \bar{x})|\bar{x}) |\det(G)|^{-1}
= p_{w|x}(G^{-1}(\bar{z} - H\bar{x})|\bar{x}) \left| -\frac{3}{2} \right|^{-1} = \frac{2}{3} p_w(2(\bar{z}_1 - \bar{x}_1), -\frac{1}{3}(\bar{z}_2 - \bar{x}_3)) .
\]

Using the PDF $p_w$, this can then be expressed as

\[
p_{z|\bar{x}}(\bar{z}|\bar{x}) = \begin{cases} 
\frac{1}{6} & \text{for } \bar{x}_1 - \frac{1}{2} \leq \bar{z}_1 \leq \bar{x}_1 + \frac{1}{2} \text{ and } \bar{x}_3 + 3 \geq \bar{z}_2 \geq \bar{x}_3 - 3 \\
0 & \text{otherwise.}
\end{cases}
\]

**Problem 3**

For a normally-distributed random variable with $w_i \sim N(0, \sigma_i^2)$, we have

\[
p(w_i) \propto \exp \left( -\frac{1}{2} \frac{w_i^2}{\sigma_i^2} \right) .
\]

Using the change of variables, and by the conditional independence of $z_1, z_2$ given $x$, we can write

\[
p(z_1, z_2|x) = p(z_1|x) p(z_2|x) \propto \exp \left( -\frac{1}{2} \left( \frac{(z_1 - x)^2}{\sigma_1^2} + \frac{(z_2 - x)^2}{\sigma_2^2} \right) \right) .
\]

Differentiating with respect to $x$, and setting to 0, leads to

\[
\frac{\partial p(z_1, z_2|x)}{\partial x} \bigg|_{x=\hat{x}_{ML}} = 0 \iff \left( \frac{\hat{z}_1 - \hat{x}_{ML}}{\sigma_1^2} \right) + \left( \frac{\hat{z}_2 - \hat{x}_{ML}}{\sigma_2^2} \right) = 0 \iff \hat{x}_{ML} = \frac{\hat{z}_1 \sigma_2^2 + \hat{z}_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}
\]

which is a weighted sum of the measurements. Note that

- for $\sigma_2^2 = 0$: $\hat{x}_{ML} = \hat{z}_1$
- for $\sigma_1^2 = 0$: $\hat{x}_{ML} = \hat{z}_2$

**Problem 4**

Using the total probability theorem, we find that

\[
p_{w_i}(\bar{w}_i) = \begin{cases} 
\frac{1}{2} & \text{for } \bar{w}_i \in [-1, 1] \\
0 & \text{otherwise.}
\end{cases}
\]

and $p(w_1, w_2) = p(w_1) p(w_2)$. Using the change of variables,

\[
p_{z_1|x}(\bar{z}_1|\bar{x}) = \begin{cases} 
\frac{1}{2} & \text{if } -1 \leq \bar{z}_1 - \bar{x} \leq 1 \\
0 & \text{otherwise.}
\end{cases}
\]

We need to solve for $p_{z_1, z_2|x}(\bar{z}_1, \bar{z}_2|\bar{x}) = p_{z_1|x}(\bar{z}_1|\bar{x}) p_{z_2|x}(\bar{z}_2|\bar{x})$. Consider the different cases:

- $|\bar{z}_1 - \bar{z}_2| > 2$, for which the likelihoods do not overlap, see the following figure:
Since no value for \( x \) can explain both, \( z_1 \) and \( z_2 \), given the model \( z_i = x + w_i \), we find
\[
p_{z_1, z_2|x}(\bar{z}_1, \bar{z}_2|\bar{x}) = 0.
\]

- \( |\bar{z}_1 - \bar{z}_2| \leq 2 \), where the likelihoods overlap, which is shown in the following figure:

We find
\[
p_{z_1, z_2|x}(\bar{z}_1, \bar{z}_2|\bar{x}) = \begin{cases} 
\frac{1}{4} & \text{for } \bar{x} \in [\bar{z}_1 - 1, \bar{z}_1 + 1] \cap [\bar{z}_2 - 1, \bar{z}_2 + 1] \\
0 & \text{otherwise}
\end{cases}
\]

where \( A \cap B \) denotes the intersection of the sets \( A \) and \( B \). Therefore the estimate is \( \hat{x}_{\text{ML}} \in [\bar{z}_1 - 1, \bar{z}_1 + 1] \cap [\bar{z}_2 - 1, \bar{z}_2 + 1] \). All values in this range are equally likely, and any value in the range is therefore a valid maximum likelihood estimate \( \hat{x}_{\text{ML}} \).

**Problem 5**

a) We apply the multivariate change of coordinates:
\[
p_{z|x}(\bar{z}|\bar{x}) = p_{w|x}(h(\bar{z}, \bar{x})|\bar{x}) |\det(G)|^{-1} = p_{w|x}(\bar{z} - H\bar{x}|\bar{x}) |1|^{-1} = p_w(\bar{z} - H\bar{x}) \\
= \frac{1}{(2\pi)^{m/2} (\det(\Sigma))^{1/2}} \exp\left( -\frac{1}{2} (\bar{z} - H\bar{x})^T \Sigma^{-1} (\bar{z} - H\bar{x}) \right).
\]

We proceed similarly as in class, for \( z, w \in \mathbb{R}^m \) and \( x \in \mathbb{R}^n \):
\[
H = \begin{bmatrix}
H_1 \\
H_2 \\
\vdots \\
H_m
\end{bmatrix}, \quad H_i = \begin{bmatrix} h_{i1} & h_{i2} & \cdots & h_{in} \end{bmatrix}, \quad \text{where } h_{ij} \in \mathbb{R}.
\]

Using
\[
p_{z|x}(\bar{z}|\bar{x}) \propto \exp\left( -\frac{1}{2} \sum_{i=1}^m \frac{(\bar{z}_i - H_i \bar{x})^2}{\sigma_i^2} \right)
\]
differentiating with respect to $\bar{x}_j$, and setting to zero yields

$$\sum_{i=1}^{m} \frac{\bar{z}_i - H_i \hat{x}^{ML}_{\bar{x}}}{\sigma_i^2} h_{ij} = 0, \quad j = 1, 2, \ldots, n$$

$$\Leftrightarrow \begin{bmatrix} h_{1j} & \cdots & h_{mj} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ \frac{1}{\sigma_2^2} & \ddots \\ 0 & \ddots & \frac{1}{\sigma_m^2} \end{bmatrix} \begin{bmatrix} \bar{z} - H\hat{x}^{ML} \end{bmatrix} = 0, \quad j = 1, \ldots, n.$$  

From which we have

$$H^TW(\bar{z} - H\hat{x}^{ML}) = 0$$

and

$$\hat{x}^{ML} = (H^TWH)^{-1}H^TW\bar{z}.$$  

Note that this can also be interpreted as a weighted least squares,

$$w(x) = z - Hx \quad \hat{x} = \arg \min_x w(x)^TWw(x)$$

where $W$ is a weighting of the measurements. The result in both cases is the same!

b) We find the maximizing $\hat{x}^{ML}$ by setting the derivative of $p(z|x)$ with respect to $x$ to zero. With the given facts about matrix calculus, we find

$$0 = \frac{\partial}{\partial \bar{x}} p(z|\bar{x}) \propto \frac{\partial}{\partial \bar{x}} \exp \left( -\frac{1}{2}(\bar{z} - H\bar{x})^T \Sigma^{-1} (\bar{z} - H\bar{x}) \right)$$

$$\propto -\frac{1}{2}(\bar{z} - H\bar{x})^T \left( \Sigma^{-1} + (\Sigma^{-1})^T \right) \frac{\partial (\bar{z} - H\bar{x})}{\partial \bar{x}} = -(\bar{z} - H\bar{x})^T \Sigma^{-1}(-H)$$

where we use the fact that

$$(\Sigma^{-1})^T = (\Sigma T)^{-1} = \Sigma^{-1}.$$

Finally,

$$(\bar{z} - H\bar{x})^T \Sigma^{-1} H = 0$$

$$\bar{z}^T \Sigma^{-1} H = \bar{x}^T H^T \Sigma^{-1} H$$

and after we transpose the left and right hand sides, we find

$$H^T \Sigma^{-1} \bar{z} = H^T \Sigma^{-1} H \bar{z}.$$  

It follows that

$$\hat{x}^{ML} = (H^T \Sigma^{-1} H)^{-1} H^T \Sigma^{-1} \bar{z}.$$
Problem 6

We calculate the conditional probability density function,

$$p_{z|x}(\bar{z}|\bar{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\bar{z} - \bar{x})^2\right).$$

Both the measurement likelihood $p_{z|x}(\bar{z}|\bar{x})$ and the prior $p_x(\bar{x})$ are shown in the following figure.

We find

$$p_{z|x}(\bar{z}|\bar{x}) p_x(\bar{x}) = \begin{cases} 0 & \text{for } \bar{x} < 0 \\ \frac{1}{\sqrt{2\pi}} \exp(-\bar{x}) \exp\left(-\frac{1}{2} (\bar{z} - \bar{x})^2\right) & \text{for } \bar{x} \geq 0. \end{cases}$$

We calculate the MAP estimate $\hat{x}_{\text{MAP}}$ that maximizes the above function for a given $\bar{z}$. We note that $p_{z|x}(\bar{z}|\bar{x}) p_x(\bar{x}) > 0$ for $\bar{x} \geq 0$, and the maximizing value is therefore $\hat{x}_{\text{MAP}} \geq 0$. Since the function is discontinuous at $\bar{x} = 0$, we have to consider two cases:

1. The maximum is at the boundary, $\hat{x}^1_{\text{MAP}} = 0$. It follows that

$$p_{z|x}(\bar{z}|\hat{x}^1_{\text{MAP}}) p_x(\hat{x}^1_{\text{MAP}}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \bar{z}^2\right). \quad (4)$$

2. The maximum is where the derivative of the likelihood $p_{z|x}(\bar{z}|\bar{x}) p_x(\bar{x})$ with respect to $\bar{x}$ is zero, i.e.

$$0 = \left. \frac{\partial}{\partial \bar{x}} (p_{z|x}(\bar{z}|\bar{x}) p_x(\bar{x})) \right|_{\bar{x} = \hat{x}^2_{\text{MAP}}} = -1 + (\bar{z} - \hat{x}^2_{\text{MAP}})$$

from which follows

$$\hat{x}^2_{\text{MAP}} = \bar{z} - 1$$

at the maximum. Substituting back into the likelihood, we obtain

$$p_{z|x}(\bar{z}|\hat{x}^2_{\text{MAP}}) p_x(\hat{x}^2_{\text{MAP}}) = \frac{1}{\sqrt{2\pi}} \exp(-\bar{z} + 1) \exp\left(-\frac{1}{2}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\bar{z} + 1)\right). \quad (5)$$

We note that $p_{z|x}(\bar{z}|\hat{x}^1_{\text{MAP}}) p_x(\hat{x}^1_{\text{MAP}}) \leq p_{z|x}(\bar{z}|\hat{x}^2_{\text{MAP}}) p_x(\hat{x}^2_{\text{MAP}})$ for $\bar{z} \geq 1$, and the ML estimate is therefore

$$\hat{x}_{\text{MAP}} = \begin{cases} 0 & \text{for } \bar{z} < 1 \\ \bar{z} - 1 & \text{for } \bar{z} \geq 1. \end{cases}$$
Problem 7

We begin by evaluating the individual elements of \((AB)\) and \((ABA^T)\) with \((A)_{ij}\) denoting the \((i,j)\)th element of matrix \(A\) (element in \(i\)th row and \(j\)th column):

\[
(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j}
\]

\[
(ABA^T)_{ij} = (a_{i1}b_{11} + a_{i2}b_{21}) a_{j1} + (a_{i1}b_{12} + a_{i2}b_{22}) a_{j2}
\]

From this we obtain the trace as

\[
\text{trace}(ABA^T) = (ABA^T)_{11} + (ABA^T)_{22}
\]

\[
= (a_{11}b_{11} + a_{12}b_{21}) a_{11} + (a_{11}b_{12} + a_{12}b_{22}) a_{12}
\]

\[
+ (a_{21}b_{11} + a_{22}b_{21}) a_{21} + (a_{21}b_{12} + a_{22}b_{22}) a_{22}
\]

and the derivatives

\[
\frac{\partial \text{trace}(ABA^T)}{\partial a_{11}} = 2a_{11}b_{11} + a_{12}b_{21} = 2(AB)_{11}
\]

\[
\frac{\partial \text{trace}(ABA^T)}{\partial a_{12}} = 2a_{12}b_{22} + a_{11}b_{21} = 2(AB)_{12}
\]

\[
\frac{\partial \text{trace}(ABA^T)}{\partial a_{21}} = 2a_{21}b_{11} + a_{22}b_{21} = 2(AB)_{21}
\]

\[
\frac{\partial \text{trace}(ABA^T)}{\partial a_{22}} = 2a_{22}b_{22} + a_{21}b_{12} = 2(AB)_{22}
\]

We therefore proved that

\[
\frac{\partial \text{trace}(ABA^T)}{\partial A} = 2AB
\]

holds for \(A, B \in \mathbb{R}^{2 \times 2}\).

Problem 8

We proceed similarly to the previous problem:

\[
\text{trace}(AB) = (AB)_{11} + (AB)_{22}
\]

\[
= a_{11}b_{11} + a_{12}b_{21} + a_{21}b_{12} + a_{22}b_{22}
\]

\[
\frac{\partial \text{trace}(AB)}{\partial a_{11}} = b_{11} = (B)_{11} = (B^T)_{11}
\]

\[
\frac{\partial \text{trace}(AB)}{\partial a_{12}} = b_{21} = (B)_{21} = (B^T)_{12}
\]

\[
\frac{\partial \text{trace}(AB)}{\partial a_{21}} = b_{12} = (B)_{12} = (B^T)_{21}
\]

\[
\frac{\partial \text{trace}(AB)}{\partial a_{22}} = b_{22} = (B)_{22} = (B^T)_{22}
\]

that is

\[
\frac{\partial \text{trace}(AB)}{\partial A} = B^T
\]

for \(A, B \in \mathbb{R}^{2 \times 2}\).
Problem 9

We use the same variables as introduced in class:

\[
\begin{align*}
    z(1:k) &= H(1:k)x + w(1:k) \\
    z(1:k) &= \begin{bmatrix} z(1) \\ \vdots \\ z(k) \end{bmatrix}, \quad H(1:k) = \begin{bmatrix} H(1) \\ \vdots \\ H(k) \end{bmatrix}, \quad w(1:k) = \begin{bmatrix} w(1) \\ \vdots \\ w(k) \end{bmatrix}, \quad R(1:k) = \begin{bmatrix} R(1) & \cdots & 0 \\ \vdots \\ 0 & \cdots & R(k) \end{bmatrix}.
\end{align*}
\]

Prior knowledge on \(x\) is included by introducing an extended system,

\[
y = Mx + q \quad \text{with} \quad y := \begin{bmatrix} \hat{x}_0 \\ z(1:k) \end{bmatrix}, \quad M := \begin{bmatrix} I \\ H(1:k) \end{bmatrix}, \quad q := \begin{bmatrix} 0 \\ w(1:k) \end{bmatrix}, \quad S := \begin{bmatrix} P_x & 0 \\ 0 & R(1:k) \end{bmatrix}.
\]

The standard Weighted Least Squares solution (including prior knowledge of \(x\)) is then given by:

\[
\hat{x}^\text{WLS}(k) = \arg\min_{\hat{x}} (y - M\hat{x})^T S^{-1} (y - M\hat{x}) = (M^T S^{-1} M)^{-1} M^T S^{-1} y.
\]

Problem 10

By definition,

\[
P(k) = \mathbb{E}_{e(k)} \left[ e(k) e^T(k) \right].
\]

Substituting the estimation error

\[
e(k) = \left( I - K(k)H(k) \right) e(k-1) - K(k) w(k)
\]

in the definition, we obtain

\[
P(k) = \mathbb{E}_{e(k-1),w(k)} \left[ \left( \left( I - K(k)H(k) \right) e(k-1) - K(k) w(k) \right) \left( \left( I - K(k)H(k) \right) e(k-1) - K(k) w(k) \right)^T \right] \\
= \mathbb{E}_{e(k-1),w(k)} \left[ \left( I - K(k)H(k) \right) e(k-1) e^T(k-1) \left( I - K(k)H(k) \right)^T + K(k) w(k) w^T(k) K^T(k) \right. \\
- \left( I - K(k)H(k) \right) e(k-1) w^T(k) K^T(k) - K(k) w(k) e^T(k-1) \left( I - K(k)H(k) \right)^T \right] \\
= \left( I - K(k)H(k) \right) P(k-1) \left( I - K(k)H(k) \right)^T + K(k) R(k) K^T(k) \\
- \left( I - K(k)H(k) \right) \mathbb{E}_{e(k-1),w(k)} \left[ e(k-1) w^T(k) \right] K^T(k) \\
- K(k) \mathbb{E}_{e(k-1),w(k)} \left[ w(k) e^T(k-1) \right] \left( I - K(k)H(k) \right)^T.
\]

Because \(e(k-1)\) and \(w(k)\) are independent,

\[
\mathbb{E}_{e(k-1),w(k)} \left[ w(k) e^T(k-1) \right] = \mathbb{E}_{w(k)} \left[ w(k) \right] \mathbb{E}_{e(k-1)} \left[ e^T(k-1) \right].
\]
Since $E[w(k)] = 0$, $E[w(k)] E[e^{T}(k-1)] = 0$, and it follows that
\[
P(k) = \left( I - K(k)H(k) \right) P(k-1) \left( I - K(k)H(k) \right)^{T} + K(k)R(k)K^{T}(k) .
\]

**Problem 11**

First, we expand
\[
\left( I - K(k)H(k) \right) P(k-1) \left( I - K(k)H(k) \right)^{T} + K(k)R(k)K^{T}(k)
= P(k-1) - K(k)H(k)P(k-1) - P(k-1)H^{T}(k)K^{T}(k) + K(k) \left( H(k)P(k-1)H^{T}(k) + R(k) \right) K^{T}(k).
\]

We use the results from problems 7 and 8 and evaluate the summands individually, since $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$. We find
\[
\frac{\partial}{\partial K(k)} \text{trace}(P(k-1)) = 0
\]
\[
\frac{\partial}{\partial K(k)} \text{trace}(-K(k)H(k)P(k-1)) = -P(k-1)H^{T}(k), \quad \text{note that } P(k-1) = P^{T}(k-1)
\]
\[
\frac{\partial}{\partial K(k)} \text{trace}(-P(k-1)H^{T}(k)K^{T}(k)) = -P(k-1)H^{T}(k), \quad \text{note that } \text{trace}(A) = \text{trace}(A^{T})
\]
\[
\frac{\partial}{\partial K(k)} \text{trace} \left( K(k) \left( H(k)P(k-1)H^{T}(k) + R(k) \right) K^{T}(k) \right) = 2K(k) \left( H(k)P(k-1)H^{T}(k) + R(k) \right).
\]

Note that $R^{T}(k) = R(k)$ and $\left( H(k)P(k-1)H^{T}(k) \right)^{T} = H(k)P(k-1)H^{T}(k)$. Finally,
\[
K(k) \left( H(k)P(k-1)H^{T}(k) + R(k) \right) = P(k-1)H^{T}(k)
\]
\[
K(k) = P(k-1)H^{T}(k) \left( H(k)P(k-1)H^{T}(k) + R(k) \right)^{-1}.
\]

**Problem 12**

a) We initialize $P(0) = 1$, $\dot{x}(0) = 0$, $H(k) = 1$ and $R(k) = 1$ $\forall k = 1, 2, \ldots$, and get:
\[
K(k) = \frac{P(k-1)}{P(k-1) + 1}
\]
\[
P(k) = (1 - K(k))^{2} P(k-1) + K(k)^{2}
\]
\[
= \frac{P(k-1)}{(1 + P(k-1))^{2}} + \frac{P(k-1)^{2}}{(1 + P(k-1))^{2}}
\]
\[
= \frac{P(k-1)}{1 + P(k-1)} .
\]

From (7) and $P(0) = 1$, we have $P(1) = \frac{1}{2}$, $P(2) = \frac{1}{3}$, $\ldots$, and we can show by induction (see below) that
\[
P(k) = \frac{1}{1 + k},
\]
\[
K(k) = \frac{1}{1 + k} \quad \text{from (6)}.
\]

Proof by induction:
• Assume
\[ P(k) = \frac{1}{1+k}, \]  
(8)

• Start with \( P(0) = \frac{1}{1+0} = 1 \), which satisfies (8).

• Now we show for \((k+1)\):
\[
P(k+1) = \frac{1}{1+k+1} = \frac{1}{k+2} = \frac{P(k)}{1+P(k)} \quad \text{from (7)}
\]
\[
= \frac{\frac{1}{1+k}}{1+\frac{1}{1+k}} \quad \text{by assumption (8)}
\]
\[
= \frac{1}{1+(k+1)} \quad \text{q.e.d.}
\]

b) The Matlab code is available on the class web page.

• Recursive least squares works, and provides a reasonable estimate after several thousand iterations.

• The recursive least squares estimator is the optimal estimator for Gaussian noise.

c) The Matlab code is available on the class web page.

• Recursive least squares works, but we can do much better with the proposed non-linear estimator derived in Problem 12.

Problem 13

a) The MMSE estimate is
\[
\hat{x}_{\text{MMSE}}(k) = \arg \min_{\hat{x}} \left\{ \sum_{\bar{x} \in \{-1,1\}} (\hat{x} - \bar{x})^2 p_x|z(1),...,z(k) (\bar{x}|\bar{z}(1), \bar{z}(2), \ldots, \bar{z}(k)) \right\}. \]  
(9)

Let us first consider the MMSE estimate at time \( k = 1 \). The only possible combinations are:

\[
\begin{array}{ccc}
\bar{x} & \bar{w}(1) & \bar{z}(1) \\
-1 & -1 & -2 \\
1 & -1 & 0 \\
1 & 1 & 2 \\
\end{array}
\]

We can now construct the conditional PDF of \( x \) when conditioned on \( z(1) \):

\[
\begin{array}{c|c|c}
\bar{x} & \bar{z}(1) & p_x|z(1)(\bar{x}|\bar{z}(1)) \\
-1 & -2 & 1 \\
-1 & 0 & 1/2 \\
1 & 0 & 1/2 \\
1 & 2 & 1 \\
\end{array}
\]

Therefore, if \( \bar{z}(1) = -2 \) or \( \bar{z}(1) = 2 \), we know \( x \) precisely: \( \bar{x} = -1 \) and \( \bar{x} = 1 \), respectively. If \( \bar{z}(1) = 0 \), it is equally likely that \( \bar{x} = 1 \) or \( \bar{x} = -1 \). The MMSE estimate is

• for \( \bar{z}(1) = 2 \):
\[
\hat{x}_{\text{MMSE}}(1) = 1 \quad \text{since} \quad p_x|z(1)(-1|2) = 0
\]

14
• for $\bar{z}(1) = -2$:
  \[
  \hat{x}^{\text{MMSE}}(1) = -1 \quad \text{since} \quad p_{x|z(1)}(1| -2) = 0
  \]

• for $\bar{z}(1) = 0$:
  \[
  \hat{x}^{\text{MMSE}}(1) = \arg \min_{\hat{x}} \left( p_{x|z(1)}(1|0) (\hat{x} - 1)^2 + p_{x|z(1)}(-1|0) (\hat{x} + 1)^2 \right)
  \]
  \[
  = \arg \min_{\hat{x}} \left( \frac{1}{2} (\hat{x} - 1)^2 + \frac{1}{2} (\hat{x} + 1)^2 \right)
  \]
  \[
  = \arg \min_{\hat{x}} (\hat{x}^2 + 1)
  \]
  \[
  = 0.
  \]

Note that the same result could have been obtained by evaluating
\[
\hat{x}^{\text{MMSE}}(k) = E_{x|z(1)}[x|0] = \sum_{\bar{x} = (-1, 1)} \bar{x} p_{x|z(1)}(\bar{x}|0) = \left( \frac{1}{2} (-1) + \frac{1}{2} (1) \right) = 0.
\]

At time $k = 2$, we could construct a table for the conditional PDF of $x$ when conditioned on $z(1)$ and $z(2)$ as above. However, note that the MMSE estimate is given for all $k$ if one measures either $-2$ or $2$ at time $k = 1$,

for $\bar{z}(1) = 2$, $\hat{x}^{\text{MMSE}}(k) = 1 \quad \forall k = 1, 2, \ldots$

for $\bar{z}(1) = -2$, $\hat{x}^{\text{MMSE}}(k) = -1 \quad \forall k = 1, 2, \ldots$.

On the other hand, a measurement $\bar{z}(1) = 0$ does not provide any information, i.e. $\bar{x}(1) = 1$ and $\bar{x}(1) = -1$ stay equally likely. As soon as a subsequent measurement $z(k)$, $k \geq 2$, has a value of $-2$ or $2$, $x$ is known precisely which is also reflected by the MMSE estimate. As long as we keep measuring $0$, we do not gain additional information and

\[
\hat{x}^{\text{MMSE}}(k) = 0 \quad \text{if} \quad \bar{z}(\ell) = 0 \quad \forall \ell = 1, 2, \ldots, k.
\]

Matlab code for this nonlinear MMSE estimator is available on the class webpage.

b) The solution is identical to part a), except that the estimate is the minimizing value from the sample space of $x$:

\[
\hat{x}^{\text{MMSE}2}(k) = \arg \min_{\bar{x} \in \mathcal{X}} \left\{ \sum_{\bar{x} = (-1, 1)} (\bar{x} - \bar{x})^2 p_{x|z(1), \ldots, z(k)}(\bar{x}|\bar{z}(1), \ldots, \bar{z}(k)) \right\}.
\] (10)

The same argument holds that, if one measurement is either $-2$ or $2$, we know $x$ precisely. However, for $\bar{z}(1) = 0$, the estimate is

\[
\hat{x}^{\text{MMSE}2}(1) = \arg \min_{\bar{x} \in \mathcal{X}} \left( p_{x|z(1)}(1|0) (\hat{x} - 1)^2 + p_{x|z(1)}(-1|0) (\hat{x} + 1)^2 \right)
\]

\[
= \arg \min_{\bar{x} \in \mathcal{X}} \left( \frac{1}{2} (\hat{x} - 1)^2 + \frac{1}{2} (\hat{x} + 1)^2 \right)
\]

\[
= \pm 1.
\]

Note that $\mathcal{X} = \{-1, 1\}$ and therefore, we cannot differentiate and set to zero the right hand side of Equation 10 to find the minimizing value. This implies that the estimate is not necessarily equal to $E_{x|z}[x|0]$, which is the case for measuring $\bar{z} = 0$. In this case $E_{x|z}[x|0]$ is equal 0, whereas the MMMSE estimator yields $\pm 1$ for $\hat{x}^{\text{MMSE}2}$.

The remainder of the solution is identical to the above part a) in that the estimate remains $\hat{x}^{\text{MMSE}2} = \pm 1$ until a measurement is either $-2$ or $2$. 

15