Introduction to
Recursive Filtering and Estimation

Spring 2010

Problem Set:
Introduction to estimation

Notes:

• **Notation:** Unless otherwise noted, $x$, $y$, and $z$ denote random variables, $f_x(x)$ (or the short hand $f(x)$) denotes the probability density function of $x$, and $f_{x|y}(x|y)$ (or $f(x|y)$) denotes the conditional probability density function of $x$ conditioned on $y$. The expected value is denoted by $E[\cdot]$, the variance is denoted by $\text{Var}(\cdot)$ 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 error that you may find to the teaching assistants (strimpe@ethz.ch or aschoellig@ethz.ch).
Problem Set

Problem 1
Find the maximum likelihood (ML) estimate \( \hat{x}_{ML} := \arg \max_x f(z|x) \) when
\[
z_i = x + w_i, \quad i = 1, 2,
\]
where \( w_i \) are assumed to be independent with \( w_i \sim \mathcal{N}(0, \sigma_i^2) \) and \( z_i, w_i, x \in \mathbb{R} \).

Problem 2
Solve Problem 1 for independent \( w_i \) uniformly distributed on \([-1, 1]\).

Problem 3
Find the ML estimate \( \hat{x}_{ML} \) when
\[
z = H x + w, x, z, w \in \mathbb{R},
\]
where \( x \in \mathbb{R}^n \) and \( z, w \in \mathbb{R}^m, n \leq m \). The matrix \( H \) has full rank.
The components of the vector \( w \) are assumed to be independent with \( w_i \sim \mathcal{N}(0, \sigma_i^2), \quad i = 1, \ldots, m \).
How does this relate to weighted least squares?

Problem 4
Find the maximum a posteriori (MAP) estimate \( \hat{x}_{MAP} := \arg \max_x f(z|x)f(x) \) when
\[
z = x + w, \quad x, z, w \in \mathbb{R},
\]
where \( x \) is exponentially distributed,
\[
f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}
\]
with \( \lambda = 1 \) and \( w \sim \mathcal{N}(0, 1) \).

Problem 5
Prove that
\[
\frac{\partial}{\partial A} \text{Tr}(ABA^T) = 2AB \quad \text{if} \quad B = B^T,
\]
for \( A, B \in \mathbb{R}^{2 \times 2} \), where \( \text{Tr}(\cdot) \) denotes the trace of a matrix.

Problem 6
Prove that
\[
\frac{\partial}{\partial A} \text{Tr}(AB) = B^T,
\]
for \( A, B \in \mathbb{R}^{2 \times 2} \).
Problem 7

Apply the recursive least squares algorithm when

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

where

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

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

a) Solve for \( K(k) \) and \( P(k) \).

b) Simulate the algorithm for normally distributed continuous random variables \( w(k) \) and \( x \) with the above properties (1). Choose \( k \) up to 10000.

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

d) For the distribution in part c), what is the optimal minimum mean-squared error (MMSE) estimate?
Problem Set 3

Problem 1

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

(proportional)

\[ f(z_{a1}, z_{a2} | x) = f(z_{a1} | x) f(z_{a2} | x) \quad \text{(conditional independent)} \]

\[ \propto \exp \left( -\frac{1}{2} \left( \frac{(z_{a1} - x)^2}{\sigma_{a1}^2} + \frac{(z_{a2} - x)^2}{\sigma_{a2}^2} \right) \right) \]

\[
\Rightarrow \text{Differentiate with respect to } x \text{ and set to 0:} \\
\left( \frac{(z_{a1} - x)}{\sigma_{a1}^2} \right) + \left( \frac{(z_{a2} - x)}{\sigma_{a2}^2} \right) = 0 \iff \hat{x} = \frac{z_{a1} \sigma_{a2}^2 + z_{a2} \sigma_{a1}^2}{\sigma_{a1}^2 + \sigma_{a2}^2}
\]

Interpretation:
\[ \sigma_{a1}^2 = 0 : \hat{x} = z_{a1} \]
\[ \sigma_{a2}^2 = 0 : \hat{x} = z_{a2} \]

Problem 2

\[ w_i = \begin{cases} 1/2 & \text{if } w_i \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \]

\[ f(z_{a1} | x) = \begin{cases} 1/2 & \text{for } -1 \leq z_{a1} - x \leq 1 \text{ or } z_{a1} - 1 \leq x \leq z_{a1} + 1 \\ 0 & \text{otherwise} \end{cases} \]

\[ f(z_{a1}, z_{a2} | x) = f(z_{a1} | x) f(z_{a2} | x) \]
Consider different cases:

I) $|z_1 - z_2| > 2$

$f(z_1, z_2 | x) = 0$ \implies no value for $x$ can explain $z_1, z_2$
given the model $z_c = x + w_i$

II) $|z_1 - z_2| \leq 2$

$f(z_1, z_2 | x) = \begin{cases} \mathbf{1/4} & \text{for } x \in [z_1-1, z_1+1] \cap [z_2-1, z_2+1] \\ 0 & \text{otherwise} \end{cases}$

\implies $x \in [z_1-1, z_1+1] \cap [z_2-1, z_2+1]$

all equally optimal

**Problem 3**

Proceed similarly as in class: $z, w \in \mathbb{R}^m, x \in \mathbb{R}^n$

$$H = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix}, \quad H_i = \begin{bmatrix} h_{i1} & \cdots & h_{in} \end{bmatrix}, \quad \text{scalars}$$
\[ f(z|\mathbf{x}) \propto \exp \left( -\frac{1}{2} \sum_{i=1}^{n} \frac{(z_i - h(x))^2}{\sigma_i^2} \right) \]

Differentiate w.r.t. \( x_j \), set to 0:

\[ \frac{\partial}{\partial x_j} \sum_{i=1}^{n} \frac{z_i - h(x)}{\sigma_i^2} h_{ij} = 0 \quad j = 1, 2, \ldots, n \]

\[ \text{Lin} \quad h_{ij} \quad \begin{bmatrix} \frac{\partial z_i}{\partial x_j} & 0 \\ 0 & 1/\sigma_i^2 \end{bmatrix} \begin{bmatrix} 1/\sigma_i^2 \\ 0 \end{bmatrix} (z - h(x)) = 0 \quad j = 1, \ldots, n \]

\[ \text{Column of } H \quad \begin{bmatrix} \mathbf{w} \end{bmatrix} \]

\[ \Rightarrow \quad H^T \mathbf{w} (z - h(x)) = 0 \quad \text{with } \mathbf{w} \text{ weight matrix} \]

\[ \Rightarrow \quad \mathbf{x} = \left( H^T \mathbf{w} H \right)^{-1} H^T \mathbf{w} z \]

Weighted least square:

\[ w(x) = z - h(x) \quad x \text{ again} \quad w(x)^T \mathbf{w} \text{ would weight of measurements} \]

**Problem 4**

Exponential distribution:

\[ f(x) = \begin{cases} 0 & x < 0 \\ \exp(-x) & x \geq 0 \end{cases} \]

\[ f(z|x) = \frac{4}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( z - x \right)^2 \right) \]

\[ \int f(z|x) f(x) = \begin{cases} 0 & x < 0 \\ \frac{4}{\sqrt{2\pi}} \exp(-x) \exp \left( -\frac{1}{2} (z - x)^2 \right) & x \geq 0 \end{cases} \]
Differentiate \((x)\) w.r.t. \(x\) and set to 0:

\[-1 + (z-x) = 0 \implies \frac{\hat{z}}{z} = z-1 \quad \text{if } z \geq 1\]

\[\implies \frac{\hat{z}}{z} = 0 \quad \text{if } z < 1 \quad \text{(cf. f(y,w))}\]

**Problem 5**

\[
A = [a_{ij}] \quad \text{with } a_{ij} = 1,2
\]

\[
B = [b_{ij}] \quad \text{with } b_{ij} = b_{ij}
\]

\[
(AB)_{ij} = \sum_{r=1}^{N^2} a_{ir} b_{rj}
\]

\[
(ABA^T)_{ij} = \sum_{s=1}^{N^2} (\sum_{r=1}^{N^2} a_{is} b_{rs}) a_{js}
\]

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

\[
= \sum_{s=1}^{N^2} \sum_{r=1}^{N^2} \sum_{t=1}^{N^2} a_{is} b_{rs} a_{ts}
\]

\[
\frac{\partial \text{Tr}(ABA^T)}{\partial a_{ij}} = \sum_{s=1}^{N^2} b_{js} a_{is} + \sum_{r=1}^{N^2} a_{ir} b_{rj} \frac{\partial }{\partial a_{ij}} b_{rj}
\]

\[
\implies \frac{\partial \text{Tr}(ABA^T)}{\partial A} = 2AB \quad \text{q.e.d.}
\]
Problem 6

Similar way of proceeding:

\[ \text{Tr}(AB) = \frac{2}{s} \sum_{r=1}^{2} a_{sr} b_{rs} \]

\[ \frac{\partial \text{Tr}(AB)}{\partial A} = \text{b}_{ji} \]

\[ \Rightarrow \frac{\partial \text{Tr}(AB)}{\partial A} = \text{b}_{ji} T = [\text{b}_{ji}] \]

Problem 7

a) \( P(0) = 1 \), \( \gamma'(0) = 0 \)

\( P(k) = (1 - K(k)) P(k-1) + K(k)^2 \)

\( = \frac{P(k-1)}{(1+P(k-1))^2} + \frac{P(k-1)}{(1+P(k-1))^2} \times \frac{K(k)^2}{(1+P(k-1))^2} \)

\( = \frac{P(k-1)}{1+P(k-1)} \) (**)\n
\( P(0) = 1 \), \( P(1) = \frac{1}{2} \), \( P(2) = \frac{1}{3} \) \ldots \( P(k) = \frac{1}{1+k} \)

proof by induction - see below

\[ \Rightarrow P(k) = \frac{1}{1+k} \]

\( K(k) = \frac{1}{1+k} \) from (**)
Proof by induction.
 * Assume \( P(k) = \frac{1}{1+k} \) (1)
 * Start with \( P(0) = \frac{1}{1+0} = 1 \)
 * Show \( P(k+1) = \frac{1}{1+k+1} = \frac{1}{k+2} = P(k) \) from (4 * k)

\[
\frac{1}{1+k} = \frac{1}{k+2} \quad \text{q.e.d.}
\]

Assumption (4)

b) see code attached...
   - recursive least squares works, needs around 3500 iterations (measurements) to converge
   - recursive least squares is optimal estimator for Gaussian noise

c) see code attached
   - recursive least squares work, but we can do much better with the proposed nonlinear estimator

d) The only possible combinations are

<table>
<thead>
<tr>
<th>( x )</th>
<th>( w )</th>
<th>( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

\( x_{MMSE} = E(x|z) \)

if \( z = -2 \) or \( z = 2 \), we know \( x \) precisely: \( x = -1 \) or \( x = 1 \), respectively. if \( z = 0 \) it is equally likely that \( x = 1 \) or \( x = -1 \)
Note that we do not have continuous random variables; that is, we have to use the original definition.

\[ \hat{x}_{\text{MMSE}} := \arg\min_x \mathbb{E} \left[ (z - x)^2 | z \right] \] for scalar case.

* for \( z = 2 \), \( \hat{x}_{\text{MMSE}} = 1 \) since \( f(x = -1 | z = 2) = 0 \)

* for \( z = -2 \), \( \hat{x}_{\text{MMSE}} = 1 \) since \( f(x = 1 | z = -2) = 0 \)

* for \( z = 0 \), \( \hat{x}_{\text{MMSE}} = \arg\min_x \left( f(x = 1 | z = 0) \cdot (x - 1)^2 + f(x = -1 | z = 0) \cdot (x + 1)^2 \right) \)

\[ = \arg\min_x \left( \frac{4}{2} (x - 1)^2 + \frac{4}{2} (x + 1)^2 \right) \]

\[ = \pm 1 \]
clear

% Use 0 for normally distributed, CRV, 1 for DRV equally likely at +-1
DISTRIBUTION_TYPE = 0;

% Set the state of the normally distributed random number generator.
randn('state',0);

% Number of time samples. 100, 1000, or 10000
N = 1e4;
xEstLin = zeros(N+1,1);
xEstNL = zeros(N+1,1);

% Generate the true value of x
if DISTRIBUTION_TYPE == 0
    xActual = randn;
else
    xActual = sign(randn);
end

% Simulation
xEstLin(1) = 0;
xEstNL(1) = 0;
for k = 1:N
    if DISTRIBUTION_TYPE == 0
        w = randn;
    else
        w = sign(randn);
    end

    % Generate simulated measurement
    z = xActual + w;

    % The optimal linear estimator
    K = 1/(1+k);
xEstLin(k+1) = xEstLin(k) + K*(z - xEstLin(k));

% The nonlinear estimator, which only makes sense for discrete random
% variables. Case 1: we have already determined what x is
if (abs(xEstNL(k)) > 0.5)
    xEstNL(k+1) = xEstNL(k);
else
    % Otherwise, we have to look at the measurement. If z = 0, then it
    % is equally likely that x = 1 or -1. If z = 2, then x must be 1.
    % If z = -2, then x must be -1
    if (abs(z) < 1)
        xEstNL(k+1) = 0;
    elseif (z > 1)
        xEstNL(k+1) = 1;
    else
        xEstNL(k+1) = -1;
    end
else
end

% Plot the results
figure(1)
plot(0:N,xEstLin,'*',0:N,xEstNL,'o',0:N,xActual,'r.);
legend('Linear Estimate','Non-linear Estimate','Actual');

findfigs