ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for
Dynamic Systems and Control
IDSC
Institut für Dynamische Systeme
und Regelungstechnik

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 $\mathrm{E}[\cdot]$, the variance is denoted by $\mathrm{Var}(\cdot)$ and $\Pr(Z)$ denotes the probability that the event Z occurs. A normally distributed random variable x with mean μ and variance σ^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

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

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

$$z = Hx + w\,,$$

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 $w_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, \ldots, m$.

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

$$z = x + w\,, \qquad x, z, w \in \mathbb{R}\,,$$

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

$$\frac{\partial\ \mathrm{Tr}(ABA^T)}{\partial A} = 2AB \qquad \text{if } \ B = B^T,$$

for $A, B \in \mathbb{R}^{2\times 2}$, where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix.

Problem 6

Prove that

$$\frac{\partial\ \mathrm{Tr}(AB)}{\partial A} = B^T,$$

for $A, B \in \mathbb{R}^{2\times 2}$.

Problem 7

Apply the recursive least squares algorithm when

$$z(k) = x + w(k)\,, \qquad k = 1, 2, \dots\,,$$

where

$$\mathrm{E}[x] = 0,\;\; \mathrm{Var}(x) = 1, \quad \text{and} \qquad \mathrm{E}[w(k)] = 0,\;\; \mathrm{Var}\,(w(k)) = 1. \tag{1}$$

The random variables $\{x, w(1), w(2), \dots\}$ 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

Angela Schöllig - aschoellig@ethz.ch

Problem 1

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

$\uparrow$ proportional

$f(z_1, z_2 \mid x) = f(z_1|x)\, f(z_2|x)$ (conditional independent)

$\propto \exp\left(-\frac{1}{2}\left(\frac{(z_1-x)^2}{\sigma_1^2} + \frac{(z_2-x)^2}{\sigma_2^2}\right)\right)$

$\Rightarrow$ Differentiate with respect to x and set to 0:

$\frac{(z_1-\hat{x})}{\sigma_1^2} + \frac{(z_2-\hat{x})}{\sigma_2^2} \overset{!}{=} 0 \iff \hat{x} = \frac{z_1\sigma_2^2 + z_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}$

weighted sum!

Interpretation:

$\sigma_1^2 = 0 \;:\; \hat{x} = z_1$

$\sigma_2^2 = 0 \;:\; \hat{x} = z_2$

Problem 2

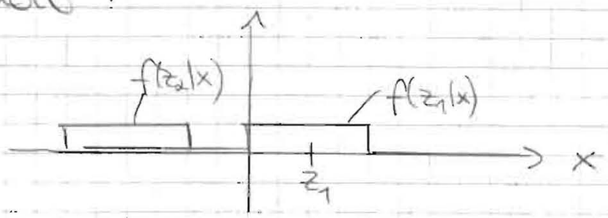
$$w_i = \begin{cases} 1/2 & w_i \in [-1,1] \\ 0 & \text{otherwise} \end{cases}$$

$$f(z_i|x) = \begin{cases} 1/2 & \text{for } -1 \le z_i - x \le 1 \text{ or } z_i - 1 \le x \le z_i + 1 \\ 0 & \text{otherwise} \end{cases}$$

$f(z_1, z_2|x) = f(z_1|x)\, f(z_2|x)$

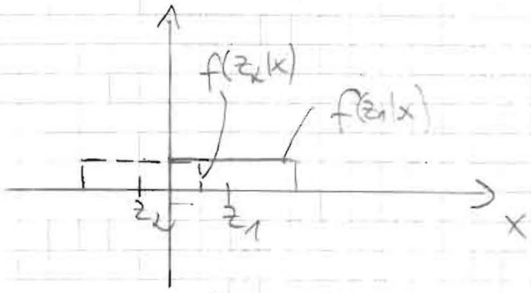
Consider different cases:

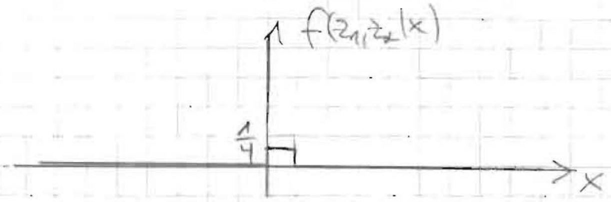
I) $|z_1 - z_2| > 2$



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

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





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

intersection

$\Rightarrow \hat{x} \in [z_1 - 1, z_1 + 1] \cap [z_2 - 1, z_2 + 1]$ all equally optimal

Problem 3

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

$$H = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_m \end{bmatrix} \qquad H_i = [\underbrace{h_{i1}, \ldots, h_{in}}_{\text{scalars}}]$$

$$f(z|x) \propto \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{m} \frac{(z_i - H_i x)^2}{\sigma_i^2}\right)\right)$$

Differentiate w.r.t. x_j, set to 0:

$$\sum_{i=1}^{m} \frac{z_i - H_i x}{\sigma_i^2} h_{ij} = 0 \qquad j = 1, 2, \dots n$$

$$\underbrace{[h_{1j} \dots h_{mj}]}_{\text{column of } H} \underbrace{\begin{bmatrix} 1/\sigma_1^2 & & 0 \\ & 1/\sigma_2^2 & \\ 0 & & 1/\sigma_m^2 \end{bmatrix}}_{W} (z - Hx) = 0 \qquad j = 1, \dots n$$

$\Rightarrow H^T W (z - Hx) = 0$ with W weight matrix

$\Rightarrow \hat{x} = (H^T W H)^{-1} H^T W z$

Weighted least square:

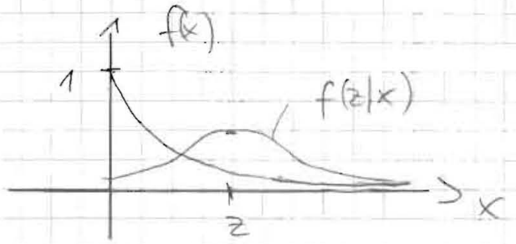
$w(x) = z - Hx$ $\qquad$ $x = \arg\min w(x)^T W w(x)$

weighting of measurements

Problem 4

exponential distribution:

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



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

$$f(z|x)\, f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\sqrt{2\pi}} \exp(-x) \exp\left(-\frac{1}{2}(z-x)^2\right) \quad (*) & x \geq 0 \end{cases}$$

Differentiate (*) w.r.t. x and set to 0:

$$-1 + (z - \hat{x}) \overset{!}{=} 0 \iff \underline{\underline{\hat{x} = z - 1}} \quad \text{if } z \geq 1$$

$$\Rightarrow \underline{\underline{\hat{x} = 0}} \quad \text{if } z < 1 \quad \text{(cf. figure)}$$

Problem 5

$A = [a_{ij}]$ $\quad i,j = 1,2$

element in ith row and jth column

$B = [b_{ij}]$ $\quad i,j = 1,2$, $b_{ij} = b_{ji}$

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

$$(ABA^T)_{ij} = \sum_{s=1}^{2} \left(\sum_{r=1}^{2} a_{ir} b_{rs} \right) a_{js}$$ from A^T

$$\mathrm{Tr}(ABA^T) = (ABA^T)_{11} + (ABA^T)_{22}$$
$$= \sum_{t=1}^{2} \sum_{s=1}^{2} \sum_{r=1}^{2} a_{tr} b_{rs} a_{ts}$$

$$\frac{\partial \mathrm{Tr}(ABA^T)}{\partial a_{ij}} = \sum_{s=1}^{2} \underbrace{b_{js}}_{= b_{sj}} a_{is} + \sum_{r=1}^{2} a_{ir} b_{rj},$$

$$= 2(AB)_{ij}$$

$$\Rightarrow \frac{\partial \mathrm{Tr}(ABA^T)}{\partial A} = 2AB \quad \text{q.e.d.}$$

Problem 6

Similar way of proceeding:

$$Tr(AB) = \sum_{s=1}^{2} \sum_{r=1}^{2} a_{\underset{ij}{sr}} b_{rs}$$

$$\frac{\partial\, Tr(AB)}{\partial a_{ij}} = b_{ji}$$

$$\Rightarrow \frac{\partial\, Tr(AB)}{\partial A} = B^T = [\, b_{ji} \,]$$

Problem 7

a) • $P(0) = 1$, $\hat{x}(0) = 0$,
• $H(k) = 1$, $R(k) = 1$

$$K(k) = \frac{P(k-1)}{P(k-1)+1} \quad (*)$$

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

$$= \frac{P(k-1)}{1+P(k-1)} \quad (**)$$

$$P(0) = 1 \quad P(1) = \frac{1}{2} \quad P(2) = \frac{1}{3} \quad \dots \quad P(k) = \frac{1}{1+k}$$

Proof by induction
→ see below

$$\Rightarrow P(k) = \frac{1}{1+k}, \quad K(k) = \frac{1}{1+k} \text{ from } (*)$$

Proof by induction:

- Assume $P(k) = \frac{1}{1+k}$ (1)
- Start with $P(0) = \frac{1}{1+0} = 1$ ✓
- Show $P(k+1) = \frac{1}{1+k+1} = \frac{1}{k+2} = \frac{P(k)}{1+P(k)}$ from $(\cancel{*})$

$$\underset{\substack{\uparrow \\ \text{use} \\ \text{Assumption (1)}}}{=} \frac{\frac{1}{1+k}}{1+\frac{1}{1+k}} = \frac{1}{k+2} \checkmark \qquad \text{q.e.d.}$$

b) see Code attached

→ recursive least squares works, needs around 3500 iterations / measurements to converge

→ recursive least squares is <u>optimal</u> estimator for Gaussian noise

c) see Code attached

→ recursive least squares works, but we can do much better with the proposed nonlinear estimator

d) The only possible combinations are

x	w	z
-1	-1	-2
-1	1	0
1	-1	0
1	1	2

$\hat{x}_{MMSE} = E[x|z]$

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

$$\hat{x}_{MMSE} := \underset{\hat{x}}{\arg\min} \; \underset{x}{E}\left[(\hat{x}-x)^2 \mid z\right] \quad \text{for scalar case}$$

- for $z = 2$, $\hat{x}_{MMSE} = 1$ since $f(x=-1 \mid z=2) = 0$
- for $z = -2$, $\hat{x}_{MMSE} = -1$ " $f(x=1 \mid z=-2) = 0$
- for $z = 0$, $\hat{x}_{MMSE} = \underset{\hat{x}}{\arg\min}\left(f(x=1 \mid z=0) \cdot (\hat{x}-1)^2 + f(x=-1 \mid z=0) \cdot (\hat{x}+1)^2 \right)$

$$= \underset{\hat{x}}{\arg\min}\left(\frac{1}{2}(\hat{x}-1)^2 + \frac{1}{2}(\hat{x}+1)^2 \right)$$

$$= \pm 1$$

```
% Problem Set3 - Problem 7
%
% *Recursive Least Squares*
%
% Recursive Filtering and Estimation
% Spring 2010
%
% --
% ETH Zurich
% Institute for Dynamic Systems and Control
% Angela Schöllig
% aschoellig@ethz.ch
%
% --
% Revision history
% [01.04.10, AS]    first version
%


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
    end
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
```