Lecture #3

Gustaf Hendeby
gustaf.hendeby@liu.se

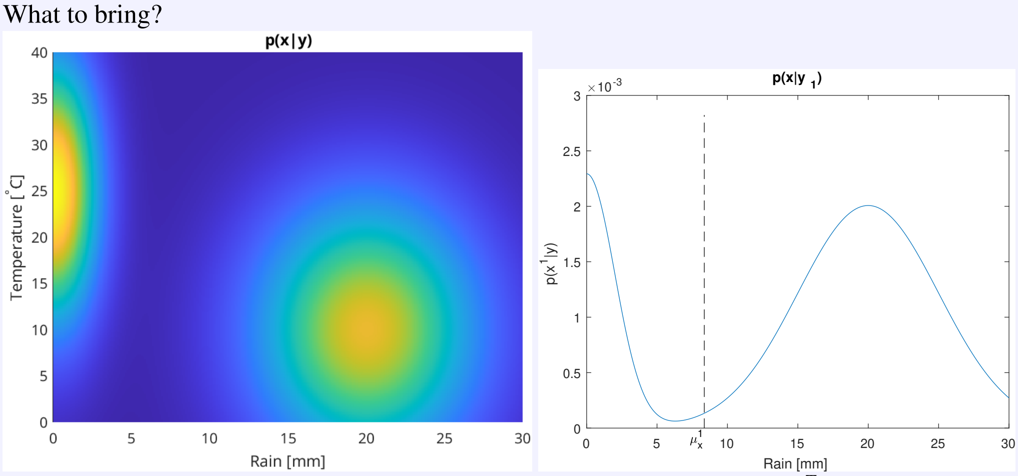
Version: 2023-11-11

Lecture 1 and 2 dealt with point estimates of x given the observations $Y = \{y_1, \dots, y_N\}$, *i.e.*, $\hat{x} = g(Y)$ and achieves MSE.

Ex: Planned travel to Swedish coast:

- lse. predicted temperature 15 °C
- lse. predicted rain 15 mm

What to bring?



More useful forecasting given the posterior pdf $p(x|y)$, $x = \begin{pmatrix} \text{rain} & \text{temp.} \end{pmatrix}^T$. Note the multimodal distributions.

1 Bayesian Approach

- Instead of a point estimate, calculate the posterior pdf $p(x|y)$. Steps:
  1. Prior distribution $p(x)$.
  2. Apply Bayes' rule:

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

where the likelihood $p(y|x)$ (comprising the information in the measurement) is combined with the prior $p(x)$, and $p(y) = \int p(y,x)\,dx$ is a normailizing constant.

3. Point estimate (if required):

$$\hat{x}^{\text{MMSE}} = \int x p(x|y)\,dx$$
$$\hat{x}^{\text{MAP}} = \arg\max_x p(x|y),$$

where: MAP — *maximum a aposteriori* and MMSE — *minimum mean squared error*.

- Generally computationally demanding to calculate the posterior distribution.
- How to handle the case when x is time-varying and observations are received sequentially?

Le 2: Casual Wiener Filter:

$$\xrightarrow{Y_k = \{y_i\}_{i=-\infty}^{k}} \boxed{H(z)} \xrightarrow{\hat{x}_k} \quad \text{llse. given } Y_k$$

Signal model: $\Phi_{yy}(z) \longrightarrow T(z) \longrightarrow$
$\Phi_{xy}(z) \longrightarrow H(z) = \frac{1}{T(z)}\left[\frac{\Phi xy(z)}{T^*(z^{-*})}\right]_+$

Where $T(z)$ is the spectral factorization.
$H(z)$ — IIR filter for sequential estimation of $\hat{x}_k$.

2 State-Space Models (SSM)

Linear discrete-time SSM:

$$x_{k+1} = F_k x_k + G_k w_k$$
$$y_k = H_k x_k + v_k$$

- x_k $(n \times 1)$ state vector
- w_k $(m \times 1)$ process noise
- v_k $(p \times 1)$ observation noise
- y_k $(p \times 1)$ observation vector
- x_0 initial state
- F_k $(n \times n)$ system matrix
- G_k $(n \times m)$ noise gain matrix
- H_k $(p \times n)$ observation matrix

x_0, w_k, and v_k are stochastic quantities with:

- $\mathsf{E}[x_0] = 0$, $\mathsf{E}[w_k] = 0$, $\mathsf{E}[v_k] = 0 \quad \forall k$
- $\mathsf{E}[x_0 x_0^*] = P_0$, $\mathsf{E}[x_0 v_k^*] = 0$
- $\mathsf{E}\left[\begin{pmatrix} w_k \\ v_k \end{pmatrix}\begin{pmatrix} w_l \\ v_l \end{pmatrix}^*\right] = \begin{pmatrix} Q_k & S_k \\ S_k^* & R_k \end{pmatrix}\delta_{kl}$ (δ_{kl} Kronecker's delta function)

NB: If x_0, v_k, and w_k are jointly Gaussian, so are x_k and y_k, due to the linear model.

General SSM:

(in order of increasing generality)

- Nonlinear:
$$\begin{aligned} x_{k+1} &= f_k(x_k, w_k) \\ y_k &= h_k(x_k) + v_k \end{aligned}$$

- Implicit:
$$\begin{aligned} f_k(x_{k+1}, x_k, w_k) &= 0 \\ h_k(y_k, x_k, v_k) &= 0 \end{aligned}$$

- pdf:
$$\begin{aligned} x_{k+1|k}|x_k &\sim p(x_{k+1}|x_k) \\ y_k|x_k &\sim p(y_k|x_k) \end{aligned}$$

2.1 Markov process

If v_k and w_k are white, then

$$p(x_{k+1}|x_{1:k}) = p(x_{k+1}|x_k, X_{k-1}) = p(x_{k+1}|x_k)$$

where the last equality utilizes the Markov property. That is, everything worth knowing about the past is available in the last sample!

2.2 General Bayesian Solution

Goal: Recursively calculate $p(x_k|Y_k)$, where $Y_k := \{y_i\}_{i=1}^k$.

Bayes' rule:

$$p(A|B,C) = \frac{p(B|A,C)p(A|C)}{p(B|C)}$$

Measurement update

$$\begin{aligned} p(x_k|Y_k) &= \frac{p(y_k|x_k, Y_{k-1})p(x_k|Y_{k-1})}{p(y_k|Y_{k-1})} \\ &= \frac{\overbrace{p(y_k|x_k, Y_{k-1})}^{\text{likelihood specified by meas. model}} \quad \overbrace{p(x_k|Y_{k-1})}^{\text{priori (predicted) dist.}}}{\int p(y_k|x_k)p(x_k|Y_{k-1})\,dx_k} \end{aligned}$$

Time update

$$\begin{aligned} p(x_k|y_{k-1} &= \int p(x_k, x_{k-1}|Y_{k-1})\,dx_{k-1} \\ &= \int p(x_k|x_{k-1}, Y_{k-1})p(x_{k-1}|Y_{k-1})\,dx_{k-1} = \Big/\text{Markov process}\Big/ \\ &= \int \underbrace{p(x_k|x_{k-1})}_{\text{specified by state transition model}} p(x_{k-1}|Y_{k-1})\,dx_{k-1} \end{aligned}$$

Problems:

- How to represent the distributions?
- How to calculate the integrals?

Representations of the distributions:

- Parametric. Only possible for Gaussian (Kalman filter) and a a few others.
- Discrete representation.
  - Fixed grid (point-mass filter)
  - Stochastic grid (particle filter)
- Gaussian mixture model, *i.e.*, linear combinations of Gaussian distributions

2.3 Linear Gaussian Case — Kalman Filter

$$\begin{aligned} x_{k+1} &= F_k x_k + G_k w_k \\ y_k &= H_k x_k + v_k \end{aligned} \qquad \mathsf{E}\left[\begin{pmatrix} w_k \\ v_k \end{pmatrix} \begin{pmatrix} w_l^* & v_l^* \end{pmatrix}\right] = \begin{pmatrix} Q_k & 0 \\ 0 & R_k \end{pmatrix} \delta_{k-l}$$

x_0, w_k, and v_k jointly Gaussian. ($S_k = 0$ for simplicity.) This results in all relevant distributions being Gaussian,

$$\begin{aligned} p(x_k|y_k) &= \mathcal{N}(x_k; \overbrace{\hat{x}_{k|k}}^{\text{mean}}, \overbrace{P_{k|k}}^{\text{covariance}}) \\ p(y_k|Y_{k-1}) &= \mathcal{N}(y_k; \hat{y}_{k|k-1}, R_{e,k}) \\ p(x_k|Y_{k-1} &= \mathcal{N}(x_k; \hat{x}_{k|k-1}, P_{k|k-1}) \end{aligned}$$

Now:

$$\begin{aligned} \hat{x}_{k|l} &= \mathsf{E}\big[x_k|Y_l\big] \\ P_{k|l} &= \mathsf{E}\big[(x_k - \hat{x}_{k|l})(x_k - \hat{x}_{k|l})^*\big] \end{aligned}$$

Goal: Recursively calculate $\hat{x}_{k|k}$ and $P_{k|k}$.

Recall: If u and z are jointly Gaussian, then

$$p_{u|z}(u|z) = \mathcal{N}(u; \mu_{u|z}, \Sigma_{u|z}),$$

where

$$\begin{aligned} \mu_{u|z} &= \mu_u + \Sigma_{uz}\Sigma_{zz}^{-1}(z - \mu_z) \\ \Sigma_{u|z} &= \Sigma_{uu} - \Sigma_{uz}\Sigma_{zz}^{-1}\Sigma_{zu} \end{aligned}$$

2.4 Measurement update $(\hat{x}_{k|k-1}, P_{k|k-1}) \to (\hat{x}_{k|k}, P_{k|k})$

$$
\begin{aligned}
\left(\begin{pmatrix} x_k \\ y_k \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \hat{x}_{k|k-1} \\ \hat{y}_{k|k-1} \end{pmatrix}, \begin{pmatrix} P_{k|k-1} & A \\ A^* & R_{e,k} \end{pmatrix}\right)\right) \\
\hat{y}_{k|k-1} &= \mathsf{E}\big[y_k|Y_{k-1}\big] = \mathsf{E}\big[H_k x_k + v_k|Y_{k-1}\big] = H_k \hat{x}_{k|k-1} \\
A &= \mathsf{E}\big[(x_k - \hat{x}_{k|k-1})(y_k - \hat{y}_{k|k-1})^*\big] = \mathsf{E}\big[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^*\big] H_k^* = P_{k|k-1} H_k^* \\
R_{e,k} &= \mathsf{E}\big[(y_k - \hat{y}_{k|k-1})(y_k - \hat{y}_{k|k-1})^*\big] = H_k P_{k|k-1} H_k^* + R_k
\end{aligned}
$$

yielding

$$
\begin{aligned}
\hat{x}_{k|k} &= \overbrace{\hat{x}_{k|k-1}}^{\mu_u} + \overbrace{P_{k|k-1} H_k^*}^{\Sigma_{uv}} \overbrace{(H_k P_{k|k-1} H_k^* + R_k)}^{\Sigma_{zz}}{}^{-1} (y_k - \overbrace{\hat{y}_{k|k-1}}^{\mu_z}) \\
P_{k|k} &= \underbrace{P_{k|k-1}}_{\Sigma_{uu}} - \underbrace{P_{k|k-1} H_k^*}_{\Sigma_{uz}} \underbrace{(H_k P_{k|k-1} H_k^* + R_k)}_{\Sigma_{zz}}{}^{-1} \underbrace{H_k P_{k|k-1}}_{\Sigma_{zu}}
\end{aligned}
$$

2.5 Time update $(\hat{x}_{k|k}, P_{k|k}) \to (\hat{x}_{k+1|k}, P_{k+1|k})$

$$
\begin{aligned}
\hat{x}_{k+1|k} &= \mathsf{E}\big[x_{k+1}|Y_k\big] = \mathsf{E}\big[F_k x_k + G_k w_k|Y_k\big] = \Big/ w_l \perp y_k, l \le k \Big/ = F_k \hat{x}_{k|k} \\
P_{k+1|k} &= \mathsf{E}\big[(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^*\big] \\
&= \mathsf{E}\big[(F_k x_k + G_k w_k - F_k \hat{x}_{k|k})(F_k x_k + G_k w_k - F_k \hat{x}_{k|k})^*\big] = F_k P_{k|k} F_k^* + G_k Q_k G_k^*
\end{aligned}
$$

Summary Kalman Filter:

Initialization

$$
\begin{aligned}
\hat{x}_{0|-1} &= x_0 \\
P_{0|-1} &= \Pi_0
\end{aligned}
$$

Measurement Update:

$$
\begin{aligned}
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + P_{k|k-1} H_k^* (H_k P_{k|k-1} H_k^* + R_k)^{-1} (y_k - H_k H_k \hat{x}_{k|k-1}) \\
P_{k|k} &= P_{k|k-1} - P_{k|k-1} H_k^* (H_k P_{k|k-1} H_k^* + R_k)^{-1} H_k P_{k|k-1}
\end{aligned}
$$

Time Update

$$
\begin{aligned}
x_{k+1|k} &= F_k \hat{x}_{k|k} \\
P_{k+1|k} &= F_k P_{k|k} F_k^* + G_k Q_k G_k^*
\end{aligned}
$$