# Bayes Optimal Filter Recursion

## Sensor Fusion

**Fredrik Gustafsson**
fredrik.gustafsson@liu.se

Gustaf Hendeby
gustaf.hendeby@liu.se

Linköping University

# Purpose

Derive the Bayes optimal filtering recursion

- Dynamic state-space models: $x_{k+1} = f(x_k, v_k)$.
- Measurement model: $y_k = h(x_k, e_k)$.
- General Bayesian solution
- Derivation based on Bayes law $p(A, B) = p(A|B)p(B)$ in different variations!

# State-Space Models

We will derive the Bayes' optimal filter for the general nonlinear model, here defined by the conditional probability $p(x_{k+1}|y_{1:k})$ (Markov model) and $p(y_k|x_k)$ (likelihood):

$$
\begin{aligned}
x_{k+1} &= f(x_k, v_k) &\quad \text{or} \quad & x_{k+1}|x_k \sim p(x_{k+1}|x_k)\\
y_k &= h(x_k, e_k) &\quad \text{or} \quad & y_k|x_k \sim p(y_k|x_k)
\end{aligned}
$$

To make it more concrete, consider the special case of a linear Gaussian model with $v_k \sim \mathcal{N}(0, Q_k)$, $e_k \sim \mathcal{N}(0, R_k)$ and $x_0 \sim \mathcal{N}(0, P_0)$

$$
\begin{aligned}
x_{k+1} &= Fx_k + G_k v_k &\quad \text{or} \quad & x_{k+1}|x_k \sim p(x_{k+1}|x_k) = \mathcal{N}\left(x_{k+1} - Fx_k, GQG^T\right)\\
y_k &= Hx_k + e_k &\quad \text{or} \quad & y_k|x_k \sim p(y_k|x_k) = \mathcal{N}\left(y_k - H(x_k), R\right)
\end{aligned}
$$

# Bayesian estimation approach: MAP estimate

- Maximum *a posteriori* (MAP) estimate defined by

$$\hat{x} = \arg\max_x p(x|y) = \arg\max_x \frac{p(y|x)p(x)}{p(y)}$$

- Often (for symmetrical distributions), the MAP estimate coincides with the Mimimum Mean Square Error (MMSE) estimate.
- Combine current information in likelihood with the prior $p(x)$.
- By letting the prior come from past observations, a natural recursive algorithm is obtained.

# Bayesian estimation approach: Measurement Update (1/2)

Consider Bayes' theorem in the form

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

Assume that

$$\begin{aligned} A &= x_k \\ B &= y_k \\ C &= y_{1:k-1} = \{y_1, \dots, y_{k-1}\} \\ B, C &= y_{1:k} \end{aligned}$$

then

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k, y_{1:k-1})p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}$$

# Bayesian estimation approach: Measurement Update (2/2)

Components in

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k, y_{1:k-1})p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}$$

- $p(y_k|x_k, y_{1:k-1}) = p(y_k|x_k) = p_e(y_k - h(x_k))$ (last step additive!)
- $p(y_k|x_k) =$ measurement likelihood
- $p(x_k|y_{1:k-1}) =$ prior state information
- $p(y_k|y_{1:k-1}) = x$ independent normalizing constant $= \alpha$

# Bayesian estimation approach: Time Update (1/2)

Consider Bayes' theorem, now in another form

$$p(A, B|C) = p(A|B, C)p(B|C)$$

Identifying terms gives

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

with the components

$$\begin{aligned} p(x_{k+1}|x_k, y_{1:k}) &= p(x_{k+1}|x_k) \quad \text{(Markov property)} \\ &= p_v\big(x_{k+1} - f(x_k)\big) \quad \text{(Additive process noise)} \end{aligned}$$

$$p(x_k|y_{1:k}) \quad \text{Given by the measurement update}$$

# Bayesian estimation approach: Time Update (2/2)

We need one more step to *marginalize* $p(x_{k+1}, x_k | y_{1:k})$ to get rid of the $x_k$ quantity. Marginalization is in Bayesian notation defined by

$$p(A|C) = \int p(A, B|C)\, dB \Rightarrow$$

This gives

$$\begin{aligned} p(x_{k+1}|y_{1:k}) &= \int p(x_{k+1}, x_k|y_{1:k})\, dx_k \\ &= \int p(x_{k+1}|x_k)p(x_k|y_{1:k})\, dx_k \end{aligned}$$

# Bayesian estimation approach: summary

Bayes' Solution: nonlinear model with additive noise

$$\alpha = \int_{\mathbb{R}^{n_y}} p_{e_k}\big(y_k - h(x_k)\big)\, p(x_k|y_{1:k-1})\, dx_k,$$

$$p(x_k|y_{1:k}) = \frac{1}{\alpha} p_{e_k}\big(y_k - h(x_k)\big)\, p(x_k|y_{1:k-1})$$

$$p(x_{k+1}|y_{1:k}) = \int_{\mathbb{R}^{n_x}} p_{v_k}\big(x_{k+1} - f(x_k)\big)\, p(x_k|y_{1:k})\, dx_k$$

To get analytical solution, we need a model that keeps the same functional form of the posterior during:

- the nonlinear transformation $f(x_k)$.
- the addition of $f(x_k)$ and $v_k$.
- the inference of $x_k$ from $y_k$ done in the measurement update.

# Practical Cases with Analytic Solution

Bayes solution can be represented with finite dimensional statistics analytically in the following cases:

- Linear Gaussian model (Kalman filter)
- Hidden Markov model (HMM)
- Linear-Gaussian mixture (Kalman filter filterbank; however exponential complexity in time)

# The Kalman filter

Consider a linear Gaussian model. Then the posterior distributions will also be Gaussian in the Bayes optimal filter, and the optimal filter becomes recursions for the mean and covariance, respectively.

- Prediction density $p(x_{k+1}|y_{1:k}) = \mathcal{N}(\hat{x}_{k+1|k}, P_{k+1|k})$, where

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

- Filtering density $p(x_k|y_{1:k}) = \mathcal{N}(\hat{x}_{k|k}, P_{k|k})$, where

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

# General Approximation Approaches

In most practical problems, no analytical solution exists, and the Bayes optimal filter must be approximated.

1. Approximate the model to a case where an optimal algorithm exists.
   - *Extended KF* (EKF) which approximates the model with a linear one.
   - *Unscented KF* (UKF) and EKF2 that apply higher order approximations.
2. Approximate the optimal nonlinear filter for the original model.
   - *Point-mass filter* (PMF) which uses a *regular* grid of the state space and applies the Bayesian recursion.
   - *Particle filter* (PF) which uses a *random* grid of the state space and applies the Bayesian recursion.

# A General Bayesian Filter Framework

Interpretation of Bayes' solution in terms of the sensor fusion formula:

$$p(x_k|y_{1:k}) = \frac{1}{\alpha} p_{e_k}(y_k - h(x_k))\, p(x_k|y_{1:k-1})$$

$$p(x_{k+1}|y_{1:k}) = \int_{\mathbb{R}^{n_x}} p_{v_k}(x_{k+1} - f(x_k))\, p(x_k|y_{1:k})\, dx_k$$

1. **Estimation:** Provides the complete distribution $p(x_k|y_k)$.
2. **Fusion:** Estimated information $p(x_k|y_k)$ is merged with the prior information $p(x_k|y_{1:k-1})$ to obtain $p(x_k|y_{1:k})$.
3. **Transformation:** Propagate information through the dynamics $z = f(x_k, u_k)$. This gives $p(z|y_{1:k})$.
4. **Diffusion:** Add uncertainty from the process noise. This gives $p(x_{k+1}|y_{1:k})$.

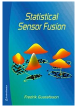
# Summary

- Bayes optimal filter in most general form

$$p(x_k|y_{1:k}) = \frac{1}{\alpha} p_{e_k}\big(y_k - h(x_k)\big) p(x_k|y_{1:k-1})$$
$$p(x_{k+1}|y_{1:k}) = \int_{\mathbb{R}^{n_x}} p_{v_k}\big(x_{k+1} - f(x_k)\big) p(x_k|y_{1:k})\, dx_k$$

- and for state space models with additive noise

$$p(x_k|y_{1:k}) = \frac{1}{\alpha} p_{e_k}\big(y_k - h(x_k)\big) p(x_k|y_{1:k-1})$$
$$p(x_{k+1}|y_{1:k}) = \int p_{v_k}\big(x_{k+1} - f(x_k)\big) p(x_k|y_{1:k})\, dx_k.$$




Section 6.3