Lecture 3: Bayesian Optimal Filtering
Equations and Kalman Filter

Simo Särkkä

Department of Biomedical Engineering and Computational Science
Aalto University

February 10, 2011
Contents

1 Probabilistics State Space Models
2 Bayesian Optimal Filter
3 Kalman Filter
4 Examples
5 Summary and Demonstration
General probabilistic state space model:

measurement model: \( y_k \sim p(y_k \mid x_k) \)

dynamic model: \( x_k \sim p(x_k \mid x_{k-1}) \)

\( x_k = (x_{k1}, \ldots, x_{kn}) \) is the state and \( y_k = (y_{k1}, \ldots, y_{km}) \) is the measurement.

Has the form of hidden Markov model (HMM):

observed: \( y_1 \quad y_2 \quad y_3 \quad y_4 \)

hidden: \( x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow \cdots \)
Example (Gaussian random walk)

Gaussian random walk model can be written as

\[ x_k = x_{k-1} + w_{k-1}, \quad w_{k-1} \sim N(0, q) \]
\[ y_k = x_k + e_k, \quad e_k \sim N(0, r), \]

where \( x_k \) is the hidden state and \( y_k \) is the measurement. In terms of probability densities the model can be written as

\[
p(x_k | x_{k-1}) = \frac{1}{\sqrt{2\pi q}} \exp \left( -\frac{1}{2q} (x_k - x_{k-1})^2 \right)
\]
\[
p(y_k | x_k) = \frac{1}{\sqrt{2\pi r}} \exp \left( -\frac{1}{2r} (y_k - x_k)^2 \right)
\]

which is a discrete-time state space model.
Example (Gaussian random walk (cont.))
Linear Gauss-Markov model:

\[ x_k = A_{k-1} x_{k-1} + q_k \]
\[ y_k = H_k x_k + r_k, \]

Gaussian driven non-linear model:

\[ x_k = f(x_{k-1}, q_k) \]
\[ y_k = h(x_k, r_k). \]

Hierarchical and/or non-Gaussian models

\[ q_k \sim \text{Dirichlet}(q_k \mid \alpha) \]
\[ x_k = f(x_{k-1}, q_k) \]
\[ \sigma_k^2 \sim \text{InvGamma}(\sigma_k^2 \mid \sigma_{k-1}^2, \gamma) \]
\[ r_k \sim \mathcal{N}(0, \sigma_k^2 I) \]
\[ y_k = h(x_k, r_k). \]
The dynamic model $p(x_k \mid x_{k-1})$ is Markovian:

1. Future $x_k$ is independent of the past given the present (here “present” is $x_{k-1}$):

$$p(x_k \mid x_{1:k-1}, y_{1:k-1}) = p(x_k \mid x_{k-1}).$$

2. Past $x_{k-1}$ is independent of the future given the present (here “present” is $x_k$):

$$p(x_{k-1} \mid x_{k:T}, y_{k:T}) = p(x_{k-1} \mid x_k).$$

The measurements $y_k$ are conditionally independent given $x_k$:

$$p(y_k \mid x_{1:k}, y_{1:k-1}) = p(y_k \mid x_k).$$
Bayesian Optimal Filter: Principle

- **Bayesian optimal filter** computes the distribution

  \[ p(x_k \mid y_{1:k}) \]

- Given the following:
  1. Prior distribution \( p(x_0) \).
  2. State space model:

    \[
    x_k \sim p(x_k \mid x_{k-1}) \\
    y_k \sim p(y_k \mid x_k),
    \]

  3. Measurement sequence \( y_{1:k} = y_1, \ldots, y_k \).

- Computation is based on **recursion rule** for incorporation of the new measurement \( y_k \) into the posterior:

  \[
  p(x_{k-1} \mid y_{1:k-1}) \longrightarrow p(x_k \mid y_{1:k})
  \]
Assume that we know the posterior distribution of previous time step:

\[ p(x_{k-1} \mid y_{1:k-1}) \].

The joint distribution of \( x_k, x_{k-1} \) given \( y_{1:k-1} \) can be computed as (recall the Markov property):

\[
p(x_k, x_{k-1} \mid y_{1:k-1}) = p(x_k \mid x_{k-1}, y_{1:k-1}) p(x_{k-1} \mid y_{1:k-1}) \\
= p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{1:k-1}),
\]

Integrating over \( x_{k-1} \) gives the Chapman-Kolmogorov equation

\[
p(x_k \mid y_{1:k-1}) = \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{1:k-1}) \, dx_{k-1}.
\]

This is the prediction step of the optimal filter.
Now we have:

1. **Prior distribution** from the Chapman-Kolmogorov equation
   \[ p(x_k | y_{1:k-1}) \]

2. **Measurement likelihood** from the state space model:
   \[ p(y_k | x_k) \]

The posterior distribution can be computed by the **Bayes’ rule** (recall the conditional independence of measurements):

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

\[
= \frac{1}{Z_k} p(y_k | x_k) p(x_k | y_{1:k-1})
\]

This is the **update step** of the optimal filter.
Optimal filter

- **Initialization:** The recursion starts from the prior distribution $p(x_0)$.

- **Prediction:** by the Chapman-Kolmogorov equation
  \[ p(x_k | y_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | y_{1:k-1}) \, dx_{k-1}. \]

- **Update:** by the Bayes’ rule
  \[ p(x_k | y_{1:k}) = \frac{1}{Z_k} p(y_k | x_k) p(x_k | y_{1:k-1}), \]

  - The normalization constant $Z_k = p(y_k | y_{1:k-1})$ is given as
  \[ Z_k = \int p(y_k | x_k) p(x_k | y_{1:k-1}) \, dx_k. \]
On prediction step the distribution of previous step is propagated through the dynamics.

Prior distribution from prediction and the likelihood of measurement.

The posterior distribution after combining the prior and likelihood by Bayes’ rule.
Kalman Filter: Model

- Gaussian driven linear model, i.e., Gauss-Markov model:
  \[ x_k = A_{k-1} x_{k-1} + q_k \]
  \[ y_k = H_k x_k + r_k, \]

- \( q_k \sim N(0, Q_k) \) white process noise.
- \( r_k \sim N(0, R_k) \) white measurement noise.
- \( A_{k-1} \) is the transition matrix of the dynamic model.
- \( H_k \) is the measurement model matrix.
- In probabilistic terms the model is
  \[
  \rho(x_k \mid x_{k-1}) = N(x_k \mid A_{k-1} x_{k-1}, Q_k) \\
  \rho(y_k \mid x_k) = N(y_k \mid H_k x_k, R_k).
  \]
Gaussian probability density

\[ N(x \mid m, P) = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} \exp \left( -\frac{1}{2}(x - m)^T P^{-1} (x - m) \right), \]

Let \( x \) and \( y \) have the Gaussian densities

\[ p(x) = N(x \mid m, P), \quad p(y \mid x) = N(y \mid Hx, R), \]

Then the joint and marginal distributions are

\[ \begin{pmatrix} x \\ y \end{pmatrix} \sim N \left( \begin{pmatrix} m \\ Hm \end{pmatrix}, \begin{pmatrix} P & PH^T \\ HP & HPH^T + R \end{pmatrix} \right) \]

\[ y \sim N(Hm, HPH^T + R). \]
If the random variables $x$ and $y$ have the joint Gaussian probability density

$\begin{pmatrix} x \\ y \end{pmatrix} \sim N \left( \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} A & C \\ C^T & B \end{pmatrix} \right),$

then the marginal and conditional densities of $x$ and $y$ are given as follows:

$x \sim N(a, A)$

$y \sim N(b, B)$

$x \mid y \sim N(a + CB^{-1} (y - b), A - CB^{-1}C^T)$

$y \mid x \sim N(b + C^T A^{-1} (x - a), B - C^T A^{-1} C).$
Assume that the posterior distribution of previous step is Gaussian

\[ p(x_{k-1} | y_{1:k-1}) = N(x_{k-1} | m_{k-1}, P_{k-1}). \]

The Chapman-Kolmogorov equation now gives

\[
p(x_k | y_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | y_{1:k-1}) \, dx_{k-1}
\]

\[ = \int N(x_k | A_{k-1} x_{k-1}, Q_k) \, N(x_{k-1} | m_{k-1}, P_{k-1}). \]

Using the Gaussian distribution computation rules from previous slides, we get the prediction step

\[ p(x_k | y_{1:k-1}) = N(x_k | A_{k-1} m_{k-1}, A_{k-1} P_{k-1} A_{k-1}^T, Q_k) \]
The joint distribution of $y_k$ and $x_k$ is

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

$$= N \left( \begin{bmatrix} x_k \\ y_k \end{bmatrix} \mid m'', P'' \right),$$

where

$$m'' = \begin{pmatrix} m_k^- \\ H_k m_k^- \end{pmatrix}$$

$$P'' = \begin{pmatrix} P_k^- & P_k^- H_k^T \\ H_k P_k^- & H_k P_k^- H_k^T + R_k \end{pmatrix}.$$
The conditional distribution of \( x_k \) given \( y_k \) is then given as

\[
\rho(x_k \mid y_k, y_{1:k-1}) = \rho(x_k \mid y_{1:k}) = N(x_k \mid m_k, P_k),
\]

where

\[
S_k = H_k P_k^{-1} H_k^T + R_k
\]

\[
K_k = P_k^{-1} H_k^T S_k^{-1}
\]

\[
m_k = m_k^- + K_k [y_k - H_k m_k^-]
\]

\[
P_k = P_k^- - K_k S_k K_k^T.
\]
Kalman Filter

- **Initialization:** \( x_0 \sim N(m_0, P_0) \)
- **Prediction step:**
  \[
  m_k^- = A_{k-1} m_{k-1} \\
  P_k^- = A_{k-1} P_{k-1} A_{k-1}^T + Q_{k-1}.
  \]
- **Update step:**
  \[
  v_k = y_k - H_k m_k^- \\
  S_k = H_k P_k^- H_k^T + R_k \\
  K_k = P_k^- H_k^T S_k^{-1} \\
  m_k = m_k^- + K_k v_k \\
  P_k = P_k^- - K_k S_k K_k^T.
  \]
Kalman Filter: Properties

- Kalman filter can be applied only to linear Gaussian models, for non-linearities we need e.g. EKF or UKF.
- If several conditionally independent measurements are obtained at a single time step, update step is simply performed for each of them separately.
- \( \Rightarrow \) If the measurement noise covariance is diagonal (as it usually is), no matrix inversion is needed at all.
- The covariance equation is independent of measurements – the gain sequence could be computed and stored offline.
- If the model is time-invariant, the gain converges to a constant \( K_k \rightarrow K \) and the filter becomes stationary:

\[
m_k = (A - KH) m_{k-1} + Ky_k
\]
Example (Kalman filter for Gaussian random walk)

Filtering density is Gaussian

\[ p(x_{k-1} \mid y_{1:k-1}) = \mathcal{N}(x_{k-1} \mid m_{k-1}, P_{k-1}). \]

The Kalman filter prediction and update equations are

\[
\begin{align*}
  m_k^- &= m_{k-1} \\
  P_k^- &= P_{k-1} + q \\
  m_k &= m_k^- + \frac{P_k^-}{P_k^- + r} (y_k - m_k^-) \\
  P_k &= P_k^- - \frac{(P_k^-)^2}{P_k^- + r}.
\end{align*}
\]
Example (Kalman filter for Gaussian random walk (cont.))
The dynamic model of the car tracking model from the first lecture can be written in discrete form as follows:

\[
\begin{pmatrix}
  x_k \\
  y_k \\
  \dot{x}_k \\
  \dot{y}_k
\end{pmatrix} =
\begin{pmatrix}
  1 & 0 & \Delta t & 0 \\
  0 & 1 & 0 & \Delta t \\
  0 & 0 & 1 & 0 \\
  0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  x_{k-1} \\
  y_{k-1} \\
  \dot{x}_{k-1} \\
  \dot{y}_{k-1}
\end{pmatrix} + q_{k-1}
\]

where \(q_k\) is zero mean with a covariance matrix \(Q\).

\[
Q =
\begin{pmatrix}
  q_1^c \Delta t^3 / 3 & 0 & q_1^c \Delta t^2 / 2 & 0 \\
  0 & q_2^c \Delta t^3 / 3 & 0 & q_2^c \Delta t^2 / 2 \\
  q_1^c \Delta t^2 / 2 & 0 & q_1^c \Delta t & 0 \\
  0 & q_2^c \Delta t^2 / 2 & 0 & q_2^c \Delta t
\end{pmatrix}
\]
The measurement model can be written in form

\[ y_k = H \begin{pmatrix} x_k \\ y_k \\ \dot{x}_k \\ \dot{y}_k \end{pmatrix} + e_k, \]

where \( e_k \) has the covariance

\[ R = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \]
The Kalman filter prediction equations:

\[
\begin{align*}
\mathbf{m}_k^- &= \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{m}_{k-1} \\
\mathbf{P}_k^- &= \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{P}_{k-1} \\
&+ \begin{pmatrix} q_1^c \Delta t^3 / 3 & 0 & q_1^c \Delta t^2 / 2 & 0 \\ 0 & q_2^c \Delta t^3 / 3 & 0 & q_2^c \Delta t^2 / 2 \\ q_1^c \Delta t^2 / 2 & 0 & q_1^c \Delta t & 0 \\ 0 & q_2^c \Delta t^2 / 2 & 0 & q_2^c \Delta t \end{pmatrix}
\end{align*}
\]
The Kalman filter update equations:

\[ S_k = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} P_k^- \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}^T + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \]

\[ K_k = P_k^- \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}^T S_k^{-1} \]

\[ m_k = m_k^- + K_k \left( y_k - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} m_k^- \right) \]

\[ P_k = P_k^- - K_k S_k K_k^T \]
Probabilistic state space models are generalizations of hidden Markov models.

Special cases of such HMMs are e.g. linear Gaussian models, non-linear filtering models.

Bayesian optimal filtering equations form the formal solution to general optimal filtering problem.

The optimal filtering equations consist of prediction and update steps.

Kalman filter is the closed form filtering solution to linear Gaussian models.
[Kalman filter for car tracking model]