Lecture 4: Extended Kalman Filter, Statistically Linearized Filter and Fourier-Hermite Kalman Filter

Simo Särkkä

Department of Biomedical Engineering and Computational Science
Aalto University

February 9, 2012
Contents

1 Overview of EKF
2 Linear Approximations of Non-Linear Transforms
3 Extended Kalman Filter
4 Statistically Linearized Filter
5 Fourier-Hermite Kalman Filter
6 Summary
Basic EKF filtering model is of the form:

\[ x_k = f(x_{k-1}) + q_{k-1} \]
\[ y_k = h(x_k) + r_k \]

- \( x_k \in \mathbb{R}^n \) is the state
- \( y_k \in \mathbb{R}^m \) is the measurement
- \( q_{k-1} \sim N(0, Q_{k-1}) \) is the Gaussian process noise
- \( r_k \sim N(0, R_k) \) is the Gaussian measurement noise
- \( f(\cdot) \) is the dynamic model function
- \( h(\cdot) \) is the measurement model function
The EKΦ model is clearly a special case of probabilistic state space models with

\[
p(x_k | x_{k-1}) = N(x_k | f(x_{k-1}), Q_{k-1})
\]
\[
p(y_k | x_k) = N(y_k | h(x_k), R_k)
\]

Recall the formal optimal filtering solution:

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

No closed form solution for non-linear \( f \) and \( h \).
In EKF, the non-linear functions are linearized as follows:

\[ f(x) \approx f(m) + F_x(m)(x - m) \]
\[ h(x) \approx h(m) + H_x(m)(x - m) \]

where \( x \sim N(m, P) \), and \( F_x, H_x \) are the Jacobian matrices of \( f, h \), respectively.

Only the first terms in linearization contribute to the approximate means of the functions \( f \) and \( h \).

The second term has zero mean and defines the approximate covariances of the functions.

Let’s take a closer look at transformations of this kind.
Consider the transformation of $x$ into $y$:

$$x \sim N(m, P)$$

$$y = g(x)$$

The probability density of $y$ is now non-Gaussian:

$$p(y) = |J(y)| \ N(g^{-1}(y) | m, P)$$

Taylor series expansion of $g$ on mean $m$:

$$g(x) = g(m + \delta x) = g(m) + G_x(m) \delta x$$

$$+ \sum_i \frac{1}{2} \delta x^T G_{xx}^{(i)}(m) \delta x e_i + \ldots$$

where $\delta x = x - m$. 

Simo Särkkä Lecture 4: EKF, SLF and FHKF
First order, that is, linear approximation:

\[ g(x) \approx g(m) + G_x(m) \delta x \]

Taking expectations on both sides gives approximation of the mean:

\[ E[g(x)] \approx g(m) \]

For covariance we get the approximation:

\[
\text{Cov}[g(x)] = E \left[ (g(x) - E[g(x)]) (g(x) - E[g(x)])^T \right] \\
\approx E \left[ (g(x) - g(m)) (g(x) - g(m))^T \right] \\
\approx G_x(m) P G_x^T(m)
\]
In EKF we will need the joint covariance of $x$ and $g(x) + q$, where $q \sim N(0, Q)$.

Consider the pair of transformations

$$
\begin{align*}
x & \sim N(m, P) \\
q & \sim N(0, Q) \\
y_1 &= x \\
y_2 &= g(x) + q.
\end{align*}
$$

Applying the linear approximation gives

$$
\begin{align*}
E \left[ \begin{pmatrix} x \\ g(x) + q \end{pmatrix} \right] & \approx \begin{pmatrix} m \\ g(m) \end{pmatrix} \\
Cov \left[ \begin{pmatrix} x \\ g(x) + q \end{pmatrix} \right] & \approx \begin{pmatrix} P & \text{PG}_x^T(m) \\ \text{G}_x(m)P & \text{G}_x(m)P\text{PG}_x^T(m) + Q \end{pmatrix}
\end{align*}
$$
The linear Gaussian approximation to the joint distribution of \( x \) and \( y = g(x) + q \), where \( x \sim N(m, P) \) and \( q \sim N(0, Q) \) is

\[
\begin{pmatrix} x \\ y \end{pmatrix} \sim N \left( \begin{pmatrix} m \\ \mu_L \end{pmatrix}, \begin{pmatrix} P & C_L \\ C_L^T & S_L \end{pmatrix} \right),
\]

where

\[
\mu_L = g(m) \\
S_L = G_x(m) P G_x^T(m) + Q \\
C_L = P G_x^T(m).
\]
Assume that the filtering distribution of previous step is Gaussian

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

The joint distribution of \( x_{k-1} \) and \( x_k = f(x_{k-1}) + q_{k-1} \) is non-Gaussian, but can be approximated linearly as

\[ p(x_{k-1}, x_k, \mid y_{1:k-1}) \approx N \left( \begin{bmatrix} x_{k-1} \\ x_k \end{bmatrix} \mid m', P' \right) , \]

where

\[
\begin{align*}
m' &= \begin{pmatrix} m_{k-1} \\ f(m_{k-1}) \end{pmatrix} \\
P' &= \begin{pmatrix} P_{k-1} & P_{k-1} F_x^T(m_{k-1}) \\ F_x(m_{k-1}) P_{k-1} & F_x(m_{k-1}) P_{k-1} F_x^T(m_{k-1}) + Q_{k-1} \end{pmatrix} .
\end{align*}
\]
Recall that if $x$ 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

$$
y \sim N(b, B)
$$

Thus, the approximate predicted distribution of $x_k$ given $y_{1:k-1}$ is Gaussian with moments

$$
\begin{align*}
\mathbf{m}_k^- &= f(m_{k-1}) \\
\mathbf{P}_k^- &= F_x(m_{k-1}) \mathbf{P}_{k-1} F_x(m_{k-1})^T + \mathbf{Q}_{k-1}
\end{align*}
$$
The joint distribution of $x_k$ and $y_k = h(x_k) + r_k$ is also non-Gaussian, but by linear approximation we get

$$p(x_k, y_k \mid y_{1:k-1}) \approx N \left( \begin{bmatrix} x_k \\ y_k \end{bmatrix} \mid m'', P'' \right),$$

where

$$m'' = \begin{pmatrix} m_k^- \\ h(m_k^-) \end{pmatrix},$$

$$P'' = \begin{pmatrix} P_k^- & P_k^- H_x^T(m_k^-) \\ H_x(m_k^-) P_k^- & H_x(m_k^-) P_k^- H_x^T(m_k^-) + R_k \end{pmatrix}.$$
Recall that if
\[
\begin{pmatrix} x \\ y \end{pmatrix} \sim N \left( \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} A & C \\ C^T & B \end{pmatrix} \right),
\]
then
\[
x \mid y \sim N(a + CB^{-1}(y - b), A - CB^{-1}C^T).
\]

Thus we get
\[
\rho(x_k \mid y_k, y_{1:k-1}) \approx N(x_k \mid m_k, P_k),
\]
where
\[
m_k = m_k^- + P_k^- H_x^T (H_x P_k^- H_x^T + R_k)^{-1} [y_k - h(m_k^-)] \\
P_k = P_k^- - P_k^- H_x^T (H_x P_k^- H_x^T + R_k)^{-1} H_x P_k^-
\]
**Extended Kalman filter**

- **Prediction:**
  
  $$m_k^- = f(m_{k-1})$$
  
  $$P_k^- = F_x(m_{k-1}) P_{k-1} F_x^T(m_{k-1}) + Q_{k-1}.$$  

- **Update:**
  
  $$v_k = y_k - h(m_k^-)$$
  
  $$S_k = H_x(m_k^-) P_k^- H_x^T(m_k^-) + R_k$$
  
  $$K_k = P_k^- H_x^T(m_k^-) S_k^{-1}$$
  
  $$m_k = m_k^- + K_k v_k$$
  
  $$P_k = P_k^- - K_k S_k K_k^T.$$
Pendulum with mass \( m = 1 \), pole length \( L = 1 \) and random force \( w(t) \):

\[
\frac{d^2 \theta}{dt^2} = -g \sin(\theta) + w(t).
\]

In state space form:

\[
\begin{bmatrix}
\frac{d}{dt} \theta \\
\frac{d}{dt} \frac{d\theta}{dt}
\end{bmatrix} =
\begin{bmatrix}
\frac{d\theta}{dt} \\
-g \sin(\theta)
\end{bmatrix} +
\begin{bmatrix}
0 \\
w(t)
\end{bmatrix}
\]

Assume that we measure the \( x \)-position:

\[
y_k = \sin(\theta(t_k)) + r_k,
\]
If we define state as \( \mathbf{x} = (\theta, d\theta/dt) \), by Euler integration with time step \( \Delta t \) we get

\[
\begin{pmatrix}
x_k^1 \\
x_k^2 \\
x_k^1 \\
\end{pmatrix} = \underbrace{\begin{pmatrix}
x_{k-1}^1 + x_{k-1}^2 \Delta t \\
x_{k-1}^2 - g \sin(x_{k-1}^1) \Delta t \\
\end{pmatrix}}_{f(\mathbf{x}_{k-1})} + \begin{pmatrix}
0 \\
q_{k-1} \\
\end{pmatrix}
\]

\( y_k = \sin(x_k^1) + r_k \),

The required Jacobian matrices are:

\[
F_{x}(\mathbf{x}) = \begin{pmatrix}
1 & \Delta t \\
-g \cos(x^1) \Delta t & 1
\end{pmatrix}, \quad H_{x}(\mathbf{x}) = \begin{pmatrix}
\cos(x^1) & 0
\end{pmatrix}
\]
Advantages of EKF

- Almost same as basic Kalman filter, easy to use.
- Intuitive, engineering way of constructing the approximations.
- Works very well in practical estimation problems.
- Computationally efficient.
- Theoretical stability results well available.
Limitations of EKF

- Does not work in **considerable non-linearities**.
- **Only Gaussian** noise processes are allowed.
- Measurement model and dynamic model functions need to be **differentiable**.
- Computation and programming of **Jacobian matrices** can be quite **error prone**.
The Idea of Statistically Linearized Filter

- In SLF, the non-linear functions are statistically linearized as follows:

\[
\begin{align*}
    f(x) & \approx b_f + A_f (x - m) \\
    h(x) & \approx b_h + A_h (x - m)
\end{align*}
\]

where \( x \sim N(m, P) \).

- The parameters \( b_f, A_f \) and \( b_h, A_h \) are chosen to minimize the mean squared errors of the form

\[
\begin{align*}
    \text{MSE}_f(b_f, A_f) &= E[||f(x) - b_f - A_f \delta x||^2] \\
    \text{MSE}_h(b_h, A_h) &= E[||h(x) - b_h - A_h \delta x||^2]
\end{align*}
\]

where \( \delta x = x - m \).

- Describing functions of the non-linearities with Gaussian input.
Again, consider the transformations
\[\begin{align*}
x &\sim N(m, P) \\
y &= g(x).
\end{align*}\]

Form linear approximation to the transformation:
\[g(x) \approx b + A \delta x,\]
where \(\delta x = x - m\).

Instead of using the Taylor series approximation, we minimize the mean squared error:
\[\text{MSE}(b, A) = E[(g(x) - b - A \delta x)^T (g(x) - b - A \delta x)].\]
Expanding the MSE expression gives:

\[
\text{MSE}(b, A) = E[g^T(x) g(x) - 2 g^T(x) b - 2 g^T(x) A \delta x \\
+ b^T b - 2 b^T A \delta x + \delta x^T A^T A \delta x] = 0
\]

Derivatives are:

\[
\frac{\partial \text{MSE}(b, A)}{\partial b} = -2 E[g(x)] + 2 b
\]

\[
\frac{\partial \text{MSE}(b, A)}{\partial A} = -2 E[g(x) \delta x^T] + 2 A P
\]
Setting derivatives with respect to $b$ and $A$ zero gives

\[ b = E[g(x)] \]
\[ A = E[g(x)\delta x^T]P^{-1}. \]

Thus we get the approximations

\[ E[g(x)] \approx E[g(x)] \]
\[ \text{Cov}[g(x)] \approx E[g(x)\delta x^T]P^{-1}E[g(x)\delta x^T]^T. \]

- The mean is exact, but the covariance is approximation.
- The expectations have to be calculated in closed form!
Statistical linearization

The statistically linearized Gaussian approximation to the joint distribution of $\mathbf{x}$ and $y = g(\mathbf{x}) + \mathbf{q}$ where $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$ and $\mathbf{q} \sim \mathcal{N}(0, \mathbf{Q})$ is given as

$$
\begin{pmatrix}
\mathbf{x} \\
y
\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}
\mathbf{m} \\
\mu_S
\end{pmatrix}, \begin{pmatrix}
\mathbf{P} & \mathbf{C}_S \\
\mathbf{C}_S^T & \mathbf{S}_S
\end{pmatrix}\right),
$$

where

$$
\mu_S = \mathbb{E}[g(\mathbf{x})],
$$

$$
\mathbf{S}_S = \mathbb{E}[g(\mathbf{x}) \delta \mathbf{x}^T] \mathbf{P}^{-1} \mathbb{E}[g(\mathbf{x}) \delta \mathbf{x}^T]^T + \mathbf{Q},
$$

$$
\mathbf{C}_S = \mathbb{E}[g(\mathbf{x}) \delta \mathbf{x}^T]^T.
$$
The statistically linearized filter (SLF) can be derived in the same manner as EKF.

Statistical linearization is used instead of Taylor series based linearization.

Requires closed form computation of the following expectations for arbitrary $x \sim N(m, P)$:

$$
E[f(x)] \\
E[f(x) \delta x^T] \\
E[h(x)] \\
E[h(x) \delta x^T],
$$

where $\delta x = x - m$. 
### Statistically Linearized Filter

#### Prediction (expectations w.r.t. $x_{k-1} \sim N(m_{k-1}, P_{k-1})$):

- $m_k^- = E[f(x_{k-1})]$
- $P_k^- = E[f(x_{k-1}) \delta x_{k-1}^T] P_{k-1}^{-1} E[f(x_{k-1}) \delta x_{k-1}^T]^T + Q_k$, 

#### Update (expectations w.r.t. $x_k \sim N(m_k^-, P_k^-)$):

- $v_k = y_k - E[h(x_k)]$
- $S_k = E[h(x_k) \delta x_{k-1}^T] (P_k^-)^{-1} E[h(x_k) \delta x_{k-1}^T]^T + R_k$
- $K_k = E[h(x_k) \delta x_{k-1}^T]^T S_k^{-1}$
- $m_k = m_k^- + K_k v_k$
- $P_k = P_k^- - K_k S_k K_k^T$. 

---

**Simo Särkkä**

Lecture 4: EKF, SLF and FHKF
If the function $g(x)$ is differentiable, we have

$$E[g(x)(x - m)^T] = E[G_x(x)] P,$$

where $G_x(x)$ is the Jacobian of $g(x)$, and $x \sim N(m, P)$.

In practice, we can use the following property for computation of the expectation of the Jacobian:

$$\mu(m) = E[g(x)]$$

$$\frac{\partial \mu(m)}{\partial m} = E[G_x(x)].$$

The resulting filter resembles EKF very closely.

Related to replacing Taylor series with Fourier-Hermite series in the approximation.
Recall the discretized pendulum model

\[
\begin{pmatrix}
    x^1_k \\
    x^2_k
\end{pmatrix} =
\begin{pmatrix}
    x^1_{k-1} + x^2_{k-1} \Delta t \\
    x^2_{k-1} - g \sin(x^1_{k-1}) \Delta t
\end{pmatrix} +
\begin{pmatrix}
    0 \\
    q_{k-1}
\end{pmatrix}
\]

\[f(x_{k-1})\]

\[
y_k = \sin(x^1_k) + r_k, \]

\[h(x_k)\]

If \(x \sim N(m, P)\), by brute-force calculation we get

\[
E[f(x)] = \begin{pmatrix}
    m_1 + m_2 \Delta t \\
    m_2 - g \sin(m_1) \exp(-P_{11}/2) \Delta t
\end{pmatrix}
\]

\[E[h(x)] = \sin(m_1) \exp(-P_{11}/2)\]
The required cross-correlation for prediction step is

\[
E[f(x)(x - m)^T] = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},
\]

where

\[
c_{11} = P_{11} + \Delta t P_{12} \\
c_{12} = P_{12} + \Delta t P_{22} \\
c_{21} = P_{12} - g \Delta t \cos(m_1) P_{11} \exp(-P_{11}/2) \\
c_{22} = P_{22} - g \Delta t \cos(m_1) P_{12} \exp(-P_{11}/2)
\]

The required term for update step is

\[
E[h(x)(x - m)^T] = \begin{pmatrix} \cos(m_1) P_{11} \exp(-P_{11}/2) \\ \cos(m_1) P_{12} \exp(-P_{11}/2) \end{pmatrix}
\]
Advantages of SLF

- **Global approximation**, linearization is based on a range of function values.
- Often more **accurate** and more **robust** than EKF.
- **No differentiability or continuity requirements** for measurement and dynamic models.
- **Jacobian matrices do not** need to be computed.
- Often computationally **efficient**.
Limitations of SLF

- Works only with **Gaussian noise** terms.
- **Expected values** of the non-linear functions have to be computed in **closed form**.
- Computation of expected values is **hard and error prone**.
- If the expected values cannot be computed in closed form, there is not much we can do.
We can generalize statistical linearization to higher order polynomial approximations:

\[ g(x) \approx b + A \delta x + \delta x^T C \delta x + \ldots \]

where \( x \sim N(m, P) \) and \( \delta x = x - m \).

We could then find the coefficients by minimizing

\[ \text{MSE}_g(b, A, C, \ldots) = \mathbb{E}[||g(x) - b - A \delta x - \delta x^T C \delta x - \ldots ||^2] \]

Possible, but calculations will be quite tedious.

A better idea is to use Hilbert space theory.
Let’s define an inner product for scalar functions $g$ and $f$ as follows:

$$\langle f, g \rangle = \int f(x) g(x) \, N(x \mid m, P) \, dx$$

$$= E[f(x) g(x)],$$

Form the Hilbert space of functions by defining the norm

$$\|g\|_H^2 = \langle g, g \rangle.$$  

There exists a polynomial basis of the Hilbert space — the polynomials are multivariate Hermite polynomials

$$H_{[a_1, \ldots, a_p]}(x; m, P) = H_{[a_1, \ldots, a_p]}(L^{-1}(x - m)),$$

where $L$ is a matrix such that $P = LL^T$ and

$$H_{[a_1, \ldots, a_p]}(x) = (-1)^p \exp(||x||^2/2) \frac{\partial^n}{\partial x_{a_1} \cdots \partial x_{a_p}} \exp(-||x||^2/2).$$
We can expand a function $g(x)$ into a Fourier-Hermite series as follows:

$$g(x) = \sum_{k=0}^{\infty} \sum_{a_1, \ldots, a_k = 1}^{n} \frac{1}{k!} E[g(x) H[a_1, \ldots, a_k](x; m, P)] \times H[a_1, \ldots, a_k](x; m, P).$$

The error criterion can be expressed also as follows:

$$MSE_g = E[\|g(x) - \hat{g}_p(x)\|^2] = \sum_{i} \|g_i(x) - \hat{g}_i^p(x)\|_H$$

where

$$\hat{g}_p(x) = b - A \delta x - \delta x^T C \delta x - \ldots \quad \text{(up to order } p)$$

But the Hilbert space theory tells us that the optimal $\hat{g}_p(x)$ is given by truncating the Fourier–Hermite series to order $p$. 
Fourier-Hermite Kalman filter (FHKF) is like the statistically linearized filter, but uses a higher order series expansion.

In practice, we can express the series in terms of expectations of derivatives by using:

$$E[g(x) H_{[a_1, \ldots, a_k]}(x; m, P)] = \sum_{b_1, \ldots, b_k=1}^{n} E \left[ \frac{\partial^k g(x)}{\partial x_{b_1} \cdots \partial x_{b_k}} \right] \prod_{m=1}^{k} L_{b_m, a_m}$$

The expectations of derivatives can be computed analytically by differentiating the following w.r.t. to mean $m$:

$$\hat{g}(m, P) = E[g(x)] = \int g(x) \ N(x \mid m, P) \, dx$$
Properties of Fourier-Hermite Kalman Filter

- **Global approximation**, based on a range of function values.
- **No differentiability or continuity** requirements.
- **Exact up to an arbitrary polynomials of order** $p$.
- **The expected values** of the non-linearities needed in closed form.
- **Analytical derivatives** are needed in computing the series coefficients.
- **Works only in Gaussian noise case.**
EKF, SLF and FHKF can be applied to filtering models of the form

\[
\begin{align*}
    x_k &= f(x_{k-1}) + q_{k-1} \\
    y_k &= h(x_k) + r_k,
\end{align*}
\]

- **EKF** is based on Taylor series expansions of \( f \) and \( h \).
  - **Advantages:** Simple, intuitive, computationally efficient
  - **Disadvantages:** Local approximation, differentiability requirements, only for Gaussian noises.

- **SLF** is based on statistical linearization:
  - **Advantages:** Global approximation, no differentiability requirements, computationally efficient
  - **Disadvantages:** Closed form computation of expectations, only for Gaussian noises.

- **FHKF** is a generalization of SLF into higher order polynomials approximations.