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

Simo Särkkä

February 24, 2016

Contents

EKF Filtering Model

Basic EKF filtering model is of the form:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1}$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k$$

- $\mathbf{x}_k \in \mathbb{R}^n$ is the state
- $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement
- $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}_{k-1})$ is the Gaussian process noise
- $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R}_k)$ is the Gaussian measurement noise
- $\mathbf{f}(\cdot)$ is the dynamic model function
- $\mathbf{h}(\cdot)$ is the measurement model function

Bayesian Optimal Filtering Equations

- The EKF model is clearly a special case of probabilistic state space models with

$$
\begin{aligned}
p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) &= \mathrm{N}(\mathbf{x}_k \mid \mathbf{f}(\mathbf{x}_{k-1}), \mathbf{Q}_{k-1}) \\
p(\mathbf{y}_k \mid \mathbf{x}_k) &= \mathrm{N}(\mathbf{y}_k \mid \mathbf{h}(\mathbf{x}_k), \mathbf{R}_k)
\end{aligned}
$$

- Recall the formal optimal filtering solution:

$$
\begin{aligned}
p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) &= \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \, p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}) \, \mathrm{d}\mathbf{x}_{k-1} \\
p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) &= \frac{1}{Z_k} p(\mathbf{y}_k \mid \mathbf{x}_k) \, p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1})
\end{aligned}
$$

- No closed form solution for non-linear **f** and **h**.

The Idea of Extended Kalman Filter

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

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\approx \mathbf{f}(\mathbf{m}) + \mathbf{F}_{\mathbf{x}}(\mathbf{m})\,(\mathbf{x} - \mathbf{m}) \\ \mathbf{h}(\mathbf{x}) &\approx \mathbf{h}(\mathbf{m}) + \mathbf{H}_{\mathbf{x}}(\mathbf{m})\,(\mathbf{x} - \mathbf{m}) \end{aligned}$$

  where $\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$, and $\mathbf{F}_{\mathbf{x}}$, $\mathbf{H}_{\mathbf{x}}$ are the Jacobian matrices of $\mathbf{f}$, $\mathbf{h}$, respectively.
- Only the first terms in linearization contribute to the approximate means of the functions $\mathbf{f}$ and $\mathbf{h}$.
- 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.

Linear Approximations of Non-Linear Transforms [1/4]

- Consider the transformation of $\mathbf{x}$ into $\mathbf{y}$:

$$\begin{aligned} \mathbf{x} &\sim \mathrm{N}(\mathbf{m}, \mathbf{P}) \\ \mathbf{y} &= \mathbf{g}(\mathbf{x}) \end{aligned}$$

- The probability density of $\mathbf{y}$ is now non-Gaussian:

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

- Taylor series expansion of $\mathbf{g}$ on mean $\mathbf{m}$:

$$\begin{aligned} \mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{m} + \delta\mathbf{x}) &= \mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\, \delta\mathbf{x} \\ &\quad + \sum_i \frac{1}{2} \delta\mathbf{x}^{\mathsf{T}}\, \mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})\, \delta\mathbf{x}\, \mathbf{e}_i + \ldots \end{aligned}$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{m}$.

- First order, that is, linear approximation:

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\,\delta\mathbf{x}$$

- Taking expectations on both sides gives approximation of the mean:

$$\mathrm{E}[\mathbf{g}(\mathbf{x})] \approx \mathbf{g}(\mathbf{m})$$

- For covariance we get the approximation:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{g}(\mathbf{x})] &= \mathrm{E}\left[(\mathbf{g}(\mathbf{x}) - \mathrm{E}[\mathbf{g}(\mathbf{x})])\,(\mathbf{g}(\mathbf{x}) - \mathrm{E}[\mathbf{g}(\mathbf{x})])^{\mathsf{T}}\right] \\
&\approx \mathrm{E}\left[(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{m}))\,(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{m}))^{\mathsf{T}}\right] \\
&\approx \mathbf{G}_{\mathbf{x}}(\mathbf{m})\,\mathbf{P}\,\mathbf{G}_{\mathbf{x}}^{\mathsf{T}}(\mathbf{m})
\end{aligned}$$

Linear Approximations of Non-Linear Transforms [3/4]

- In EKF we will need the joint covariance of $\mathbf{x}$ and $\mathbf{g}(\mathbf{x}) + \mathbf{q}$, where $\mathbf{q} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q})$.
- Consider the pair of transformations

$$
\begin{aligned}
\mathbf{x} &\sim \mathrm{N}(\mathbf{m}, \mathbf{P}) \\
\mathbf{q} &\sim \mathrm{N}(\mathbf{0}, \mathbf{Q}) \\
\mathbf{y}_1 &= \mathbf{x} \\
\mathbf{y}_2 &= \mathbf{g}(\mathbf{x}) + \mathbf{q}.
\end{aligned}
$$

- Applying the linear approximation gives

$$
\begin{aligned}
\mathrm{E}\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{g}(\mathbf{x}) + \mathbf{q} \end{pmatrix}\right] &\approx \begin{pmatrix} \mathbf{m} \\ \mathbf{g}(\mathbf{m}) \end{pmatrix} \\
\mathrm{Cov}\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{g}(\mathbf{x}) + \mathbf{q} \end{pmatrix}\right] &\approx \begin{pmatrix} \mathbf{P} & \mathbf{P}\,\mathbf{G}_{\mathbf{x}}^{\mathsf{T}}(\mathbf{m}) \\ \mathbf{G}_{\mathbf{x}}(\mathbf{m})\,\mathbf{P} & \mathbf{G}_{\mathbf{x}}(\mathbf{m})\,\mathbf{P}\,\mathbf{G}_{\mathbf{x}}^{\mathsf{T}}(\mathbf{m}) + \mathbf{Q} \end{pmatrix}
\end{aligned}
$$

Linear Approximation of Non-Linear Transform

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

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{m} \\ \boldsymbol{\mu}_L \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{C}_L \\ \mathbf{C}_L^\mathsf{T} & \mathbf{S}_L \end{pmatrix} \right),$$

where

$$\begin{aligned} \boldsymbol{\mu}_L &= \mathbf{g}(\mathbf{m}) \\ \mathbf{S}_L &= \mathbf{G}_\mathbf{x}(\mathbf{m})\, \mathbf{P}\, \mathbf{G}_\mathbf{x}^\mathsf{T}(\mathbf{m}) + \mathbf{Q} \\ \mathbf{C}_L &= \mathbf{P}\, \mathbf{G}_\mathbf{x}^\mathsf{T}(\mathbf{m}). \end{aligned}$$

Derivation of EKF [1/4]

- Assume that the filtering distribution of previous step is Gaussian

$$p(\mathbf{x}_{k-1} \,|\, \mathbf{y}_{1:k-1}) \approx \mathrm{N}(\mathbf{x}_{k-1} \,|\, \mathbf{m}_{k-1}, \mathbf{P}_{k-1})$$

- The joint distribution of $\mathbf{x}_{k-1}$ and $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1}$ is non-Gaussian, but can be approximated linearly as

$$p(\mathbf{x}_{k-1}, \mathbf{x}_k, \,|\, \mathbf{y}_{1:k-1}) \approx \mathrm{N}\left(\begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{x}_k \end{bmatrix} \,\middle|\, \mathbf{m}', \mathbf{P}' \right),$$

where

$$\mathbf{m}' = \begin{pmatrix} \mathbf{m}_{k-1} \\ \mathbf{f}(\mathbf{m}_{k-1}) \end{pmatrix}$$

$$\mathbf{P}' = \begin{pmatrix} \mathbf{P}_{k-1} & \mathbf{P}_{k-1}\,\mathbf{F}_x^\mathsf{T}(\mathbf{m}_{k-1}) \\ \mathbf{F}_x(\mathbf{m}_{k-1})\,\mathbf{P}_{k-1} & \mathbf{F}_x(\mathbf{m}_{k-1})\,\mathbf{P}_{k-1}\,\mathbf{F}_x^\mathsf{T}(\mathbf{m}_{k-1}) + \mathbf{Q}_{k-1} \end{pmatrix}.$$

Derivation of EKF [2/4]

- Recall that if $\mathbf{x}$ and $\mathbf{y}$ have the joint Gaussian probability distribution

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

then

$$\mathbf{y} \sim \mathrm{N}(\mathbf{b}, \mathbf{B})$$

- Thus, the approximate predicted distribution of $\mathbf{x}_k$ given $\mathbf{y}_{1:k-1}$ is Gaussian with moments

$$\begin{aligned} \mathbf{m}_k^- &= \mathbf{f}(\mathbf{m}_{k-1}) \\ \mathbf{P}_k^- &= \mathbf{F}_x(\mathbf{m}_{k-1})\,\mathbf{P}_{k-1}\,\mathbf{F}_x^\mathsf{T}(\mathbf{m}_{k-1}) + \mathbf{Q}_{k-1} \end{aligned}$$

Derivation of EKF [3/4]

- The joint distribution of $\mathbf{x}_k$ and $\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k$ is also non-Gaussian, but by linear approximation we get

$$p(\mathbf{x}_k, \mathbf{y}_k \mid \mathbf{y}_{1:k-1}) \approx \mathrm{N}\left(\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{bmatrix} \,\middle|\, \mathbf{m}'', \mathbf{P}'' \right),$$

where

$$\mathbf{m}'' = \begin{pmatrix} \mathbf{m}_k^- \\ \mathbf{h}(\mathbf{m}_k^-) \end{pmatrix}$$

$$\mathbf{P}'' = \begin{pmatrix} \mathbf{P}_k^- & \mathbf{P}_k^- \, \mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-) \\ \mathbf{H}_{\mathbf{x}}(\mathbf{m}_k^-) \, \mathbf{P}_k^- & \mathbf{H}_{\mathbf{x}}(\mathbf{m}_k^-) \, \mathbf{P}_k^- \, \mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-) + \mathbf{R}_k \end{pmatrix}$$

Derivation of EKF [4/4]

- Recall that if

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

  then

$$\mathbf{x} \,|\, \mathbf{y} \sim \mathrm{N}(\mathbf{a} + \mathbf{C}\,\mathbf{B}^{-1}\,(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\,\mathbf{B}^{-1}\mathbf{C}^\mathsf{T}).$$

- Thus we get

$$p(\mathbf{x}_k \,|\, \mathbf{y}_k, \mathbf{y}_{1:k-1}) \approx \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k, \mathbf{P}_k),$$

  where

$$\begin{aligned}
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{P}_k^-\,\mathbf{H}_\mathbf{x}^\mathsf{T}(\mathbf{H}_\mathbf{x}\,\mathbf{P}_k^-\,\mathbf{H}_\mathbf{x}^\mathsf{T} + \mathbf{R}_k)^{-1}[\mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-)] \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{P}_k^-\,\mathbf{H}_\mathbf{x}^\mathsf{T}\,(\mathbf{H}_\mathbf{x}\,\mathbf{P}_k^-\,\mathbf{H}_\mathbf{x}^\mathsf{T} + \mathbf{R}_k)^{-1}\,\mathbf{H}_\mathbf{x}\,\mathbf{P}_k^-
\end{aligned}$$

EKF Equations

Extended Kalman filter

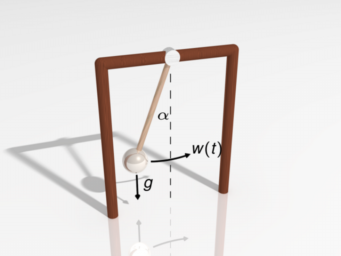
- Prediction:

$$\begin{aligned}\mathbf{m}_k^- &= \mathbf{f}(\mathbf{m}_{k-1}) \\ \mathbf{P}_k^- &= \mathbf{F}_{\mathbf{x}}(\mathbf{m}_{k-1})\,\mathbf{P}_{k-1}\,\mathbf{F}_{\mathbf{x}}^T(\mathbf{m}_{k-1}) + \mathbf{Q}_{k-1}.\end{aligned}$$

- Update:

$$\begin{aligned}\mathbf{v}_k &= \mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-) \\ \mathbf{S}_k &= \mathbf{H}_{\mathbf{x}}(\mathbf{m}_k^-)\,\mathbf{P}_k^-\,\mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-) + \mathbf{R}_k \\ \mathbf{K}_k &= \mathbf{P}_k^-\,\mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-)\,\mathbf{S}_k^{-1} \\ \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\,\mathbf{v}_k \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T.\end{aligned}$$

EKF Example [1/2]



- Pendulum with mass $m = 1$, pole length $L = 1$ and random force $w(t)$:

$$\frac{d^2\alpha}{dt^2} = -g\,\sin(\alpha) + w(t).$$

- In state space form:

$$\frac{d}{dt}\begin{pmatrix}\alpha \\ d\alpha/dt\end{pmatrix} = \begin{pmatrix}d\alpha/dt \\ -g\,\sin(\alpha)\end{pmatrix} + \begin{pmatrix}0 \\ w(t)\end{pmatrix}$$

- Assume that we measure the x-position:

$$y_k = \sin(\alpha(t_k)) + r_k,$$

EKF Example [2/2]

- If we define state as $\mathbf{x} = (\alpha, d\alpha/dt)$, by Euler integration with time step Δt we get

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

$$y_k = \underbrace{\sin(x_k^1)}_{\mathbf{h}(\mathbf{x_k})} + r_k,$$

- The required Jacobian matrices are:

$$\mathbf{F}_x(\mathbf{x}) = \begin{pmatrix} 1 & \Delta t \\ -g \, \cos(x^1) \, \Delta t & 1 \end{pmatrix}, \qquad \mathbf{H}_x(\mathbf{x}) = \begin{pmatrix} \cos(x^1) & 0 \end{pmatrix}$$

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:

$$
\begin{aligned}
\mathbf{f}(\mathbf{x}) &\approx \mathbf{b}_f + \mathbf{A}_f\,(\mathbf{x} - \mathbf{m}) \\
\mathbf{h}(\mathbf{x}) &\approx \mathbf{b}_h + \mathbf{A}_h\,(\mathbf{x} - \mathbf{m})
\end{aligned}
$$

  where $\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$.

- The parameters $\mathbf{b}_f$, $\mathbf{A}_f$ and $\mathbf{b}_h$, $\mathbf{A}_h$ are chosen to minimize the mean squared errors of the form

$$
\begin{aligned}
\mathrm{MSE}_f(\mathbf{b}_f, \mathbf{A}_f) &= \mathrm{E}[||\mathbf{f}(\mathbf{x}) - \mathbf{b}_f - \mathbf{A}_f\,\delta\mathbf{x}||^2] \\
\mathrm{MSE}_h(\mathbf{b}_h, \mathbf{A}_h) &= \mathrm{E}[||\mathbf{h}(\mathbf{x}) - \mathbf{b}_h - \mathbf{A}_h\,\delta\mathbf{x}||^2]
\end{aligned}
$$

  where $\delta\mathbf{x} = \mathbf{x} - \mathbf{m}$.

- Describing functions of the non-linearities with Gaussian input.

- Again, consider the transformations

$$\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$$
$$\mathbf{y} = \mathbf{g}(\mathbf{x}).$$

- Form linear approximation to the transformation:

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{b} + \mathbf{A}\,\delta\mathbf{x},$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{m}$.

- Instead of using the Taylor series approximation, we minimize the mean squared error:

$$\mathrm{MSE}(\mathbf{b}, \mathbf{A}) = \mathrm{E}[(\mathbf{g}(\mathbf{x}) - \mathbf{b} - \mathbf{A}\,\delta\mathbf{x})^{\mathsf{T}}(\mathbf{g}(\mathbf{x}) - \mathbf{b} - \mathbf{A}\,\delta\mathbf{x})]$$

- Expanding the MSE expression gives:

$$\begin{aligned}\mathrm{MSE}(\mathbf{b},\mathbf{A}) &= \mathrm{E}[\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{g}(\mathbf{x}) - 2\,\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{b} - 2\,\mathbf{g}^\mathsf{T}(\mathbf{x})\,\mathbf{A}\,\delta\mathbf{x} \\ &\quad + \mathbf{b}^\mathsf{T}\,\mathbf{b} - \underbrace{2\,\mathbf{b}^\mathsf{T}\,\mathbf{A}\,\delta\mathbf{x}}_{=0} + \underbrace{\delta\mathbf{x}^\mathsf{T}\,\mathbf{A}^\mathsf{T}\,\mathbf{A}\,\delta\mathbf{x}}_{\mathrm{tr}\{\mathbf{A}\,\mathbf{P}\,\mathbf{A}^\mathsf{T}\}}]\end{aligned}$$

- Derivatives are:

$$\begin{aligned}\frac{\partial \mathrm{MSE}(\mathbf{b},\mathbf{A})}{\partial \mathbf{b}} &= -2\,\mathrm{E}[\mathbf{g}(\mathbf{x})] + 2\,\mathbf{b} \\ \frac{\partial \mathrm{MSE}(\mathbf{b},\mathbf{A})}{\partial \mathbf{A}} &= -2\,\mathrm{E}[\mathbf{g}(\mathbf{x})\,\delta\mathbf{x}^\mathsf{T}] + 2\,\mathbf{A}\,\mathbf{P}\end{aligned}$$

Statistical Linearization of Non-Linear Transforms [3/4]

- Setting derivatives with respect to **b** and **A** zero gives

$$\begin{aligned}\mathbf{b} &= \mathrm{E}[\mathbf{g}(\mathbf{x})] \\ \mathbf{A} &= \mathrm{E}[\mathbf{g}(\mathbf{x})\,\delta\mathbf{x}^\mathsf{T}]\,\mathbf{P}^{-1}.\end{aligned}$$

- Thus we get the approximations

$$\begin{aligned}\mathrm{E}[\mathbf{g}(\mathbf{x})] &\approx \mathrm{E}[\mathbf{g}(\mathbf{x})] \\ \mathrm{Cov}[\mathbf{g}(\mathbf{x})] &\approx \mathrm{E}[\mathbf{g}(\mathbf{x})\,\delta\mathbf{x}^\mathsf{T}]\,\mathbf{P}^{-1}\;\mathrm{E}[\mathbf{g}(\mathbf{x})\,\delta\mathbf{x}^\mathsf{T}]^\mathsf{T}.\end{aligned}$$

- 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 $\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{q}$ where $\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$ and $\mathbf{q} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q})$ is given as

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

where

$$\begin{aligned} \boldsymbol{\mu}_S &= \mathrm{E}[\mathbf{g}(\mathbf{x})] \\ \mathbf{S}_S &= \mathrm{E}[\mathbf{g}(\mathbf{x})\,\delta\mathbf{x}^\mathsf{T}]\,\mathbf{P}^{-1}\,\mathrm{E}[\mathbf{g}(\mathbf{x})\,\delta\mathbf{x}^\mathsf{T}]^\mathsf{T} + \mathbf{Q} \\ \mathbf{C}_S &= \mathrm{E}[\mathbf{g}(\mathbf{x})\,\delta\mathbf{x}^\mathsf{T}]^\mathsf{T}. \end{aligned}$$

Statistically Linearized Filter [1/3]

- 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 $\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$:

$$
\begin{aligned}
&\mathrm{E}[\mathbf{f}(\mathbf{x})] \\
&\mathrm{E}[\mathbf{f}(\mathbf{x})\, \delta\mathbf{x}^{\mathsf{T}}] \\
&\mathrm{E}[\mathbf{h}(\mathbf{x})] \\
&\mathrm{E}[\mathbf{h}(\mathbf{x})\, \delta\mathbf{x}^{\mathsf{T}}],
\end{aligned}
$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{m}$.

Statistically Linearized Filter [2/3]

Statistically linearized filter

- Prediction (expectations w.r.t. $\mathbf{x}_{k-1} \sim \mathrm{N}(\mathbf{m}_{k-1}, \mathbf{P}_{k-1})$):

$$
\begin{aligned}
\mathbf{m}_k^- &= \mathrm{E}[\mathbf{f}(\mathbf{x}_{k-1})] \\
\mathbf{P}_k^- &= \mathrm{E}[\mathbf{f}(\mathbf{x}_{k-1})\, \delta\mathbf{x}_{k-1}^\mathsf{T}]\, \mathbf{P}_{k-1}^{-1}\; \mathrm{E}[\mathbf{f}(\mathbf{x}_{k-1})\, \delta\mathbf{x}_{k-1}^\mathsf{T}]^\mathsf{T} + \mathbf{Q}_{k-1},
\end{aligned}
$$

- Update (expectations w.r.t. $\mathbf{x}_k \sim \mathrm{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$):

$$
\begin{aligned}
\mathbf{v}_k &= \mathbf{y}_k - \mathrm{E}[\mathbf{h}(\mathbf{x}_k)] \\
\mathbf{S}_k &= \mathrm{E}[\mathbf{h}(\mathbf{x}_k)\, \delta\mathbf{x}_k^\mathsf{T}]\, (\mathbf{P}_k^-)^{-1}\; \mathrm{E}[\mathbf{h}(\mathbf{x}_k)\, \delta\mathbf{x}_k^\mathsf{T}]^\mathsf{T} + \mathbf{R}_k \\
\mathbf{K}_k &= \mathrm{E}[\mathbf{h}(\mathbf{x}_k)\, \delta\mathbf{x}_k^\mathsf{T}]^\mathsf{T}\, \mathbf{S}_k^{-1} \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\, \mathbf{v}_k \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\, \mathbf{S}_k\, \mathbf{K}_k^\mathsf{T}.
\end{aligned}
$$

Statistically Linearized Filter [3/3]

- If the function $\mathbf{g}(\mathbf{x})$ is differentiable, we have

$$\mathrm{E}[\mathbf{g}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^\mathsf{T}] = \mathrm{E}[\mathbf{G}_x(\mathbf{x})]\,\mathbf{P},$$

  where $\mathbf{G}_x(\mathbf{x})$ is the Jacobian of $\mathbf{g}(\mathbf{x})$, and $\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$.
- In practice, we can use the following property for computation of the expectation of the Jacobian:

$$\begin{aligned}
\mu(\mathbf{m}) &= \mathrm{E}[\mathbf{g}(\mathbf{x})] \\
\frac{\partial \mu(\mathbf{m})}{\partial \mathbf{m}} &= \mathrm{E}[\mathbf{G}_x(\mathbf{x})].
\end{aligned}$$

- The resulting filter resembles EKF very closely.
- Related to replacing Taylor series with Fourier-Hermite series in the approximation.

Statistically Linearized Filter: Example [1/2]

- Recall the discretized pendulum model

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

$$y_k = \underbrace{\sin(x_k^1)}_{\mathbf{h}(\mathbf{x_k})} + r_k,$$

- If $\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$, by brute-force calculation we get

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

$$\mathrm{E}[h(\mathbf{x})] = \sin(m_1) \, \exp(-P_{11}/2)$$

Statistically Linearized Filter: Example [2/2]

- The required cross-correlation for prediction step is

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

where

$$\begin{aligned}
c_{11} &= P_{11} + \Delta t\, P_{12} \\
c_{12} &= P_{12} + \Delta t\, P_{22} \\
c_{21} &= P_{12} - g\,\Delta t\,\cos(m_1)\,P_{11}\,\exp(-P_{11}/2) \\
c_{22} &= P_{22} - g\,\Delta t\,\cos(m_1)\,P_{12}\,\exp(-P_{11}/2)
\end{aligned}$$

- The required term for update step is

$$\mathrm{E}[h(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^\mathsf{T}] = \begin{pmatrix} \cos(m_1)\,P_{11}\,\exp(-P_{11}/2) \\ \cos(m_1)\,P_{12}\,\exp(-P_{11}/2) \end{pmatrix}$$

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.

Fourier-Hermite Series [1/3]

- We can generalize statistical linearization to higher order polynomial approximations:

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{b} + \mathbf{A}\,\delta\mathbf{x} + \delta\mathbf{x}^{\mathsf{T}}\mathbf{C}\,\delta\mathbf{x} + \ldots$$

  where $\mathbf{x} \sim \mathrm{N}(\mathbf{m}, \mathbf{P})$ and $\delta\mathbf{x} = \mathbf{x} - \mathbf{m}$.
- We could then find the coefficients by minimizing

$$\mathrm{MSE}_{g}(\mathbf{b}, \mathbf{A}, \mathbf{C}, \ldots) = \mathrm{E}[||\mathbf{g}(\mathbf{x}) - \mathbf{b} - \mathbf{A}\,\delta\mathbf{x} - \delta\mathbf{x}^{\mathsf{T}}\mathbf{C}\,\delta\mathbf{x} - \ldots||^2]$$

- Possible, but calculations will be quite tedious.
- A better idea is to use Hilbert space theory.

Fourier-Hermite Series [2/3]

- Let's define an inner product for scalar functions g and f as follows:

$$\begin{aligned} \langle f, g \rangle &= \int f(\mathbf{x})\, g(\mathbf{x})\, \mathrm{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P})\, \mathrm{d}\mathbf{x} \\ &= \mathrm{E}[f(\mathbf{x})\, g(\mathbf{x})], \end{aligned}$$

- 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

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

where $\mathbf{L}$ is a matrix such that $\mathbf{P} = \mathbf{L}\,\mathbf{L}^\mathsf{T}$ and

$$H_{[a_1,\ldots,a_p]}(\mathbf{x}) = (-1)^p \exp(||\mathbf{x}||^2/2) \frac{\partial^n}{\partial x_{a_1} \cdots \partial x_{a_p}} \exp(-||\mathbf{x}||^2/2).$$

Fourier-Hermite Series [3/3]

- We can expand a function $\mathbf{g}(\mathbf{x})$ into a Fourier-Hermite series as follows:

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

- The error criterion can be expressed also as follows:

$$\mathrm{MSE}_g = \mathrm{E}[||\mathbf{g}(\mathbf{x}) - \hat{\mathbf{g}}_p(\mathbf{x})||^2] = \sum_i ||g_i(\mathbf{x}) - \hat{g}_i^p(\mathbf{x})||_H$$

  where

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

- But the Hilbert space theory tells us that the optimal $\hat{\mathbf{g}}^p(\mathbf{x})$ is given by truncating the Fourier–Hermite series to order p.

Idea of Fourier-Hermite Kalman Filter

- 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:

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

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

$$
\hat{\mathbf{g}}(\mathbf{m}, \mathbf{P}) = \mathrm{E}[\mathbf{g}(\mathbf{x})] = \int \mathbf{g}(\mathbf{x})\, \mathrm{N}(\mathbf{x} \,|\, \mathbf{m}, \mathbf{P})\, \mathrm{d}\mathbf{x}
$$

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.

Summary

- EKF, SLF and FHKF can be applied to filtering models of the form

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1}$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k,$$

- 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.