

Lecture 4: Extended Kalman Filter and Statistical Linearization

Simo Särkkä

January 30, 2019

Learning Outcomes

- 1 Summary of the Last Lecture
- 2 Overview of EKF
- 3 Linear Approximations of Non-Linear Transforms
- 4 Extended Kalman Filter
- 5 Statistically Linearized Filter
- 6 Summary

Summary of the Last Lecture

- Probabilistic state space models consist of Markovian dynamic models and conditionally independent measurement models.
- Special cases are, for example, linear Gaussian models and non-linear and non-Gaussian models.
- Bayesian filtering equations form the formal solution to general Bayesian filtering problem.
- The Bayesian filtering equations consist of prediction and update steps.
- Kalman filter is the closed form filtering solution to linear Gaussian models.

Extended Kalman filter (EKF) can be used in models of the form:

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

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k$$

- $\mathbf{x}_k \in \mathbb{R}^n$ is the **state**
- $\mathbf{y}_k \in \mathbb{R}^m$ is the **measurement**
- $\mathbf{q}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$ is the Gaussian **process noise**
- $\mathbf{r}_k \sim \mathcal{N}(0, \mathbf{R}_k)$ 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

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

- Recall the **formal optimal filtering solution**:

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

- **No closed form solution** for non-linear \mathbf{f} and \mathbf{h} .

The Idea of Extended Kalman Filter

- In EKF, the **non-linear functions are linearized** (via the first order **Taylor series expansion**) as follows:

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{m}) + \mathbf{F}_x(\mathbf{m})(\mathbf{x} - \mathbf{m})$$

$$\mathbf{h}(\mathbf{x}) \approx \mathbf{h}(\mathbf{m}) + \mathbf{H}_x(\mathbf{m})(\mathbf{x} - \mathbf{m})$$

where $\mathbf{x} \sim N(\mathbf{m}, \mathbf{P})$, and \mathbf{F}_x , \mathbf{H}_x are the Jacobian matrices of \mathbf{f} , \mathbf{h} , respectively.

- If we replace $\mathbf{f}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ with their linearizations, we get a **linear Gaussian model – Kalman filter can be used**.
- Can also be seen as **approximate transformations of random variables** – let us look into that point of view next.

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

$$\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x})$$

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

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

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

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

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

- First order, that is, **linear approximation**:

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

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

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

- For **covariance** we get the approximation:

$$\begin{aligned}\text{Cov}[\mathbf{g}(\mathbf{x})] &= \mathbb{E} \left[(\mathbf{g}(\mathbf{x}) - \mathbb{E}[\mathbf{g}(\mathbf{x})]) (\mathbf{g}(\mathbf{x}) - \mathbb{E}[\mathbf{g}(\mathbf{x})])^T \right] \\ &\approx \mathbb{E} \left[(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{m})) (\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{m}))^T \right] \\ &\approx \mathbf{G}_x(\mathbf{m}) \mathbf{P} \mathbf{G}_x^T(\mathbf{m})\end{aligned}$$

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 $\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$.
- Consider the **pair of transformations**

$$\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$$

$$\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

$$\mathbf{y}_1 = \mathbf{x}$$

$$\mathbf{y}_2 = \mathbf{g}(\mathbf{x}) + \mathbf{q}.$$

- Applying the linear approximation gives

$$\mathbb{E} \left[\begin{pmatrix} \mathbf{x} \\ \mathbf{g}(\mathbf{x}) + \mathbf{q} \end{pmatrix} \right] \approx \begin{pmatrix} \mathbf{m} \\ \mathbf{g}(\mathbf{m}) \end{pmatrix}$$

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

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 $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$ and $\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is

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

where

$$\boldsymbol{\mu}_L = \mathbf{g}(\mathbf{m})$$

$$\mathbf{S}_L = \mathbf{G}_x(\mathbf{m}) \mathbf{P} \mathbf{G}_x^T(\mathbf{m}) + \mathbf{Q}$$

$$\mathbf{C}_L = \mathbf{P} \mathbf{G}_x^T(\mathbf{m}).$$

Derivation of EKF [1/4]

- Assume that the **filtering distribution** of previous step is **Gaussian**

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

- 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

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

where

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

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

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

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

then

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

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

$$\mathbf{m}_k^- = \mathbf{f}(\mathbf{m}_{k-1})$$

$$\mathbf{P}_k^- = \mathbf{F}_x(\mathbf{m}_{k-1}) \mathbf{P}_{k-1} \mathbf{F}_x^T(\mathbf{m}_{k-1}) + \mathbf{Q}_{k-1}$$

- 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

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

where

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

Derivation of EKF [4/4]

- Recall that if

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

then

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

- Thus we get

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

where

$$\mathbf{m}_k = \mathbf{m}_k^- + \mathbf{P}_k^- \mathbf{H}_x^\top (\mathbf{H}_x \mathbf{P}_k^- \mathbf{H}_x^\top + \mathbf{R}_k)^{-1} [\mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-)]$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{P}_k^- \mathbf{H}_x^\top (\mathbf{H}_x \mathbf{P}_k^- \mathbf{H}_x^\top + \mathbf{R}_k)^{-1} \mathbf{H}_x \mathbf{P}_k^-$$

Extended Kalman filter

- Prediction:

$$\mathbf{m}_k^- = \mathbf{f}(\mathbf{m}_{k-1})$$

$$\mathbf{P}_k^- = \mathbf{F}_x(\mathbf{m}_{k-1}) \mathbf{P}_{k-1} \mathbf{F}_x^T(\mathbf{m}_{k-1}) + \mathbf{Q}_{k-1}.$$

- Update:

$$\mathbf{v}_k = \mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-)$$

$$\mathbf{S}_k = \mathbf{H}_x(\mathbf{m}_k^-) \mathbf{P}_k^- \mathbf{H}_x^T(\mathbf{m}_k^-) + \mathbf{R}_k$$

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_x^T(\mathbf{m}_k^-) \mathbf{S}_k^{-1}$$

$$\mathbf{m}_k = \mathbf{m}_k^- + \mathbf{K}_k \mathbf{v}_k$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T.$$

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

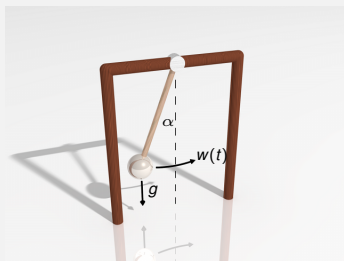
$$\frac{d^2\alpha}{dt^2} = -g \sin(\alpha) + w(t).$$

- 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:

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



- 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:

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

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 with **strong 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**:

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{b}_f + \mathbf{A}_f (\mathbf{x} - \mathbf{m})$$

$$\mathbf{h}(\mathbf{x}) \approx \mathbf{b}_h + \mathbf{A}_h (\mathbf{x} - \mathbf{m})$$

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

- Corresponds to replacing Taylor series in EKF with **Fourier–Hermite series expansion**.
- 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

$$\text{MSE}_f(\mathbf{b}_f, \mathbf{A}_f) = \mathbb{E}[||\mathbf{f}(\mathbf{x}) - \mathbf{b}_f - \mathbf{A}_f \delta\mathbf{x}||^2]$$

$$\text{MSE}_h(\mathbf{b}_h, \mathbf{A}_h) = \mathbb{E}[||\mathbf{h}(\mathbf{x}) - \mathbf{b}_h - \mathbf{A}_h \delta\mathbf{x}||^2]$$

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

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

- Again, consider the transformations

$$\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}).$$

- 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**:

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

- Expanding the MSE expression gives:

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

- Derivatives are:

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

- Setting derivatives with respect to \mathbf{b} and \mathbf{A} zero gives

$$\mathbf{b} = E[\mathbf{g}(\mathbf{x})]$$

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

- Thus we get the approximations

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

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

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

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

where

$$\boldsymbol{\mu}_S = \mathbb{E}[\mathbf{g}(\mathbf{x})]$$

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

$$\mathbf{C}_S = \mathbb{E}[\mathbf{g}(\mathbf{x}) \delta \mathbf{x}^T]^T.$$

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

$$E[\mathbf{f}(\mathbf{x})]$$

$$E[\mathbf{f}(\mathbf{x}) \delta \mathbf{x}^T]$$

$$E[\mathbf{h}(\mathbf{x})]$$

$$E[\mathbf{h}(\mathbf{x}) \delta \mathbf{x}^T],$$

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

Statistically linearized filter

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

$$\mathbf{m}_k^- = \mathbb{E}[\mathbf{f}(\mathbf{x}_{k-1})]$$

$$\mathbf{P}_k^- = \mathbb{E}[\mathbf{f}(\mathbf{x}_{k-1}) \delta \mathbf{x}_{k-1}^T] \mathbf{P}_{k-1}^{-1} \mathbb{E}[\mathbf{f}(\mathbf{x}_{k-1}) \delta \mathbf{x}_{k-1}^T]^T + \mathbf{Q}_{k-1},$$

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

$$\mathbf{v}_k = \mathbf{y}_k - \mathbb{E}[\mathbf{h}(\mathbf{x}_k)]$$

$$\mathbf{S}_k = \mathbb{E}[\mathbf{h}(\mathbf{x}_k) \delta \mathbf{x}_k^T] (\mathbf{P}_k^-)^{-1} \mathbb{E}[\mathbf{h}(\mathbf{x}_k) \delta \mathbf{x}_k^T]^T + \mathbf{R}_k$$

$$\mathbf{K}_k = \mathbb{E}[\mathbf{h}(\mathbf{x}_k) \delta \mathbf{x}_k^T]^T \mathbf{S}_k^{-1}$$

$$\mathbf{m}_k = \mathbf{m}_k^- + \mathbf{K}_k \mathbf{v}_k$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T.$$

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

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

where $\mathbf{G}_x(\mathbf{x})$ is the Jacobian of $\mathbf{g}(\mathbf{x})$, and $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$.

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

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

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

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

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

Statistically Linearized Filter: Example [2/2]

- The required cross-correlation for prediction step is

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

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

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

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**.
- **BUT** – sigma-point filters such as UKF can be seen as modified numerical-integration versions of SLF, we come back to this later.
- If the expected values cannot be computed in closed form, there is not much we can do.

- **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 \mathbf{f} and \mathbf{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.
 - But, there is a connection to **sigma-point filters** (e.g., UKF).