# A Bayesian derivation of the Kalman filter

Sensor fusion & nonlinear filtering

Lars Hammarstrand

# LINEAR AND GAUSSIAN STATE SPACE MODELS

- Consider a linear and Gaussian model:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1}, & \mathbf{q}_{k-1} &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}) \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{r}_k, & \mathbf{r}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k) \end{aligned}$$

  where $\mathbf{x}_0$ is Gaussian with mean $\hat{\mathbf{x}}_{0|0}$ and covariance matrices $\mathbf{P}_{0|0}$.

- We can also express this model as

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

**Objective (in this video)**

- Derive analytical expressions for $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ and $p(\mathbf{x}_k | \mathbf{y}_{1:k})$.

# A BRUTE FORCE DERIVATION

- It is possible to derive the Kalman filter equations from the filtering equations

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1})\, d\mathbf{x}_{k-1}$$

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1})$$

- Unfortunately, the derivation involves various matrix manipulations and is rather tedious.

- **Standard derivations** instead make use of “well known” results regarding Gaussian distributions. We use this approach below.

# PREDICTION STEP

## Prediction step

- Objective is to compute $p(\mathbf{x}_k \big| \mathbf{y}_{1:k-1})$ using

$$\begin{cases} p(\mathbf{x}_{k-1} \big| \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \\ \mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \end{cases}$$

## Background theory (well know results)

- if $\mathbf{z}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1)$ and $\mathbf{z}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Lambda}_2)$ are independent

$$\begin{aligned} \Rightarrow \mathbf{z} &= \mathbf{B}_1\mathbf{z}_1 + \mathbf{B}_2\mathbf{z}_2 \\ &\sim \mathcal{N}(\mathbf{B}_1\boldsymbol{\mu}_1 + \mathbf{B}_2\boldsymbol{\mu}_2, \mathbf{B}_1\boldsymbol{\Lambda}_1\mathbf{B}_1^T + \mathbf{B}_2\boldsymbol{\Lambda}_2\mathbf{B}_2^T). \end{aligned}$$

$p(x_k|y_{1:k-1}) =$

$$\mathcal{N}(x_k; \underbrace{A_{k-1}\hat{x}_{k-1|k-1}}_{\hat{x}_{k|k-1}}, \underbrace{A_{k-1} P_{k-1|k-1} A_{k-1}^T + Q_{k-1}}_{P_{k|k-1}})$$

# A LEMMA FOR THE UPDATE STEP

## Conditional distribution of Gaussian variables

- If $\mathbf{x}$ and $\mathbf{y}$ are two Gaussian random variables with the joint probability density function

  E{x} Cov{x} Cov{x,y}

  $$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{xx} & \mathbf{P}_{xy} \\ \mathbf{P}_{yx} & \mathbf{P}_{yy} \end{bmatrix} \right)$$

  E{y} $P_{xy}^T$ Cov{y}

  then the conditional density of $\mathbf{x}$ given $\mathbf{y}$ is

  $$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_x + \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \mathbf{P}_{xx} - \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}\mathbf{P}_{yx})$$

Note:

- $\mathbf{P}_{xy} = \mathbf{0} \Rightarrow p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_x, \mathbf{P}_{xx})$ .
- $\text{Cov}\{\mathbf{x}|\mathbf{y}\} \leq \mathbf{P}_{xx}$.

# THE UPDATE STEP

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} I \\ H_k \end{bmatrix} x_k + \begin{bmatrix} 0 \\ I \end{bmatrix} r_k$$

- We have a predicted density $\mathbf{x}_k \middle| \mathbf{y}_{1:k-1} \sim \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$ and observe a measurement $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{r}_k$

$$\Rightarrow \begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{bmatrix} \middle| \mathbf{y}_{1:k-1} \sim \mathcal{N}\left( \begin{bmatrix} \overset{\mu_x}{\hat{\mathbf{x}}_{k|k-1}} \\ \underset{\mu_y}{\mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}} \end{bmatrix}, \begin{bmatrix} \overset{P_{xx}}{\mathbf{P}_{k|k-1}} & \overset{P_{xy}}{\mathbf{P}_{k|k-1} \mathbf{H}_k^T} \\ \underset{P_{yx} = P_{xy}^T}{\mathbf{H}_k \mathbf{P}_{k|k-1}} & \underbrace{\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k}_{P_{yy} = S_k} \end{bmatrix} \right)$$

$$\boxed{S_k \cdot S_k^{-1} = I}$$

- We use the notation $p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$ where

$$\hat{x}_{k|k} = \mu_x + P_{xx} \cdot P_{yy}^{-1} (y - \mu_y)$$

$$= \hat{x}_{k|k-1} + \underbrace{P_{k|k-1} \cdot H_k^T S_k^{-1}}_{K_k} \underbrace{(y_k - H_k \hat{x}_{k|k-1})}_{v_k}$$

$$= \hat{x}_{k|k-1} + K_k \cdot v_k$$

$$P_{k|k} = P_{xx} - P_{xy} \cdot P_{yy}^{-1} \cdot P_{xy}^T = P_{k|k-1} - P_{k|k-1} \cdot H_k^T S_k^{-1} (P_{k|k-1} H_k^T)^T$$

$$= P_{k|k-1} - \underbrace{P_{k|k-1} \cdot H^T S_k^{-1}}_{K_k} S_k \underbrace{S_k^{-T} \cdot (P_{k|k-1} H_k^T)^T}_{K_k^T} = P_{k|k-1} - K_k \cdot S_k \cdot K_k^T$$

## SELF ASSESSMENT

With an ideal sensor we would have $y_k = x_k$. (We consider a scalar case here for simplicity.) Under that assumption, which of the following apply?

- $p(y_k | y_{1:k-1}) = \mathcal{N}(y_k; \hat{x}_{k|k-1}, \mathbf{P}_{k|k-1})$
- $p(y_k | y_{1:k-1}) = \delta(y_k - x_k)$
- $p(x_k, y_k | y_{1:k-1}) = \mathcal{N}\left( \begin{bmatrix} x_k \\ y_k \end{bmatrix} \middle| \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{\mathbf{x}}_{k|k-1} \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & P_{k|k-1} \\ P_{k|k-1} & P_{k|k-1} \end{bmatrix} \right)$
- $p(x_k, y_k | y_{1:k-1}) = \mathcal{N}\left( \begin{bmatrix} x_k \\ y_k \end{bmatrix} \middle| \begin{bmatrix} \hat{x}_{k|k-1} \\ x_k \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & 0 \\ 0 & P_{k|k-1} \end{bmatrix} \right)$