Notes on Kalman Filter
(KF, EKF, ESKF, IEKF, IESKF)

Gyubeom Edward Im*

February 9, 2024

Contents

*blog: alida.tistory.com, email: criterion.im@gmail.com

1 Recursive Bayes Filter

State estimation problems generally refer to estimating the current state (e.g., position, speed, rotation) based on given control inputs and observations. Given control inputs and observations, the degree of reliability of the current state $\mathbf{x}_t$, or the Belief $\text{bel}(\mathbf{x}_t)$ about $\mathbf{x}_t$, is defined as follows:

$$\text{bel}(\mathbf{x}_t) = p(\mathbf{x}_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \tag{1}$$

- $\mathbf{x}_t$: State variable at time t
- $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \cdots, \mathbf{z}_t\}$: Observations from time 1 to t
- $\mathbf{u}_{1:t} = \{\mathbf{u}_1, \cdots, \mathbf{u}_t\}$: Control inputs from time 1 to t
- $\text{bel}(\mathbf{x}_t)$: Belief of $\mathbf{x}_t$, which represents the probability (degree of belief) that the robot is at $\mathbf{x}_t$ based on sensor observations $\mathbf{z}_{1:t}$ and control inputs $\mathbf{u}_{1:t}$ from the start time to t seconds.

$\text{bel}(\cdot)$ is expressed and developed according to the Bayesian rule, so it is also called the Bayes filter. Using the Markov assumption and the Bayesian rule, a recursive filter can be induced, and this is called the recursive Bayes filter.

$$\begin{aligned} \text{bel}(\mathbf{x}_t) &= \eta \cdot p(\mathbf{z}_t \mid \mathbf{x}_t)\overline{\text{bel}}(\mathbf{x}_t) \\ \overline{\text{bel}}(\mathbf{x}_t) &= \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_t)\text{bel}(\mathbf{x}_{t-1})d\mathbf{x}_{t-1} \end{aligned} \tag{2}$$

- $\eta = 1/p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t})$: A normalization factor to maintain the definition of the probability distribution by normalizing its width to 1
- $p(\mathbf{z}_t \mid \mathbf{x}_t)$: Observation model
- $\int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_t)d\mathbf{x}_{t-1}$: Motion model

The Recursive Bayes filter is thus called a recursive filter because it calculates the current step's $\text{bel}(\mathbf{x}_t)$ from the previous step's $\text{bel}(\mathbf{x}_{t-1})$.

1.1 Derivation of Recursive Bayes Filter

The formulation of the Recursive Bayes Filter is derived as follows:

$$\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \text{bel}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \overline{\text{bel}}(\mathbf{x}_{t-1})
\end{aligned} \tag{3}$$

Steps applied include Bayesian rule, Markov Assumption, and Marginalization to derive the recursive filter process.

Step 1:

$$\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t})
\end{aligned} \tag{4}$$

Apply the Bayesian rule: $p(x|y) = \frac{p(y|x)p(x)}{p(y)}$

Step 2:

$$\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t})
\end{aligned} \tag{5}$$

Apply the Markov Assumption that the current state depends only on the immediately previous state.

Step 3:

$$\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1}
\end{aligned} \tag{6}$$

Apply the Law of Total Probability or Marginalization: $p(x) = \int_y p(x|y) \cdot p(y)dy$

Step 4:

$$\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1}
\end{aligned} \tag{7}$$

Apply the Markov Assumption that the current state depends only on the immediately previous state.

Step 5:

$$
\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1}) d\mathbf{x}_{t-1}
\end{aligned} \tag{8}
$$

Apply the Markov Assumption that the current state depends only on the immediately previous state.

Step 6:

$$
\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot \text{bel}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1}
\end{aligned} \tag{9}
$$

$p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1})$ is substituted by $\text{bel}(\mathbf{x}_{t-1})$ as they are equivalent.

Step 7:

$$
\begin{aligned}
\text{bel}(\mathbf{x}_t) &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot \text{bel}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \\
&= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t) \cdot \overline{\text{bel}}(\mathbf{x}_{t-1})
\end{aligned} \tag{10}
$$

The integral $\int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \cdot \text{bel}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1}$ is replaced by $\overline{\text{bel}}(\mathbf{x}_{t-1})$.

1.2 Gaussian Belief Case

If $\text{bel}(\mathbf{x}_t)$ follows a Gaussian distribution, this is specifically referred to as the Kalman filter.

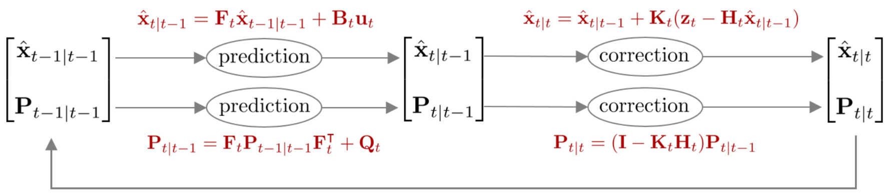
$$
\begin{aligned}
&\overline{\text{bel}}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}) \quad \text{(Kalman Filter Prediction)} \\
&\text{bel}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \quad \text{(Kalman Filter Correction)}
\end{aligned} \tag{11}
$$

Mean and covariance are also denoted as $(\hat{\mathbf{x}}, \mathbf{P})$ or $(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$, representing the same concept with different notations.

2 Kalman Filter (KF)

NOMENCLATURE of Kalman Filter

- Scalars are denoted by lowercase letters, e.g., a
- Vectors are denoted by bold lowercase letters, e.g., **a**
- Matrices are denoted by bold uppercase letters, e.g., **R**
- prediction: $\overline{\text{bel}}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$
  - $\hat{\mathbf{x}}_{t|t-1}$: Mean at step t given the correction value at step $t-1$. Some literature also denotes this as $\mathbf{x}_t^-$.
  - $\hat{\mathbf{P}}_{t|t-1}$: Covariance at step t given the correction value at step $t-1$. Some literature also denotes this as $\mathbf{P}_t^-$.
- correction: $\text{bel}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$
  - $\hat{\mathbf{x}}_{t|t}$: Mean at step t given the prediction value at step t. Some literature also denotes this as $\mathbf{x}_t^+$.
  - $\hat{\mathbf{P}}_{t|t}$: Covariance at step t given the prediction value at step t. Some literature also denotes this as $\mathbf{P}_t^+$.



At time t, the robot's position is denoted by $\mathbf{x}_t$, the measurements from the robot's sensor by $\mathbf{z}_t$, and the robot's control input by $\mathbf{u}_t$. Using these, we can define the motion model and observation model. The models are constrained by the requirement that they must be linear.

$$\begin{aligned} \text{Motion Model:} \qquad & \mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{B}_t\mathbf{u}_t + \mathbf{w}_t \\ \text{Observation Model:} \qquad & \mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t \end{aligned} \tag{12}$$

- $\mathbf{x}_t$: Model state variable
- $\mathbf{u}_t$: Model input
- $\mathbf{z}_t$: Model measurement
- $\mathbf{F}_t$: Model state transition matrix
- $\mathbf{B}_t$: Matrix that transforms input $\mathbf{u}_t$ to the state variable when given
- $\mathbf{H}_t$: Model observation matrix
- $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$: Motion model noise. $\mathbf{Q}_t$ denotes the covariance matrix of $\mathbf{w}_t$.
- $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R}_t)$: Observation model noise. $\mathbf{R}_t$ denotes the covariance matrix of $\mathbf{v}_t$.

Assuming all random variables follow Gaussian distributions, $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_t), p(\mathbf{z}_t \mid \mathbf{x}_t)$ can be represented as follows.

$$\begin{aligned} p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_t) \quad & \sim \mathcal{N}(\mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{B}_t\mathbf{u}_t, \mathbf{Q}_t) \\ & = \frac{1}{\sqrt{\det(2\pi\mathbf{Q}_t)}} \exp\Big(-\frac{1}{2}(\mathbf{x}_t - \mathbf{F}_t\mathbf{x}_{t-1} - \mathbf{B}_t\mathbf{u}_t)^\intercal \mathbf{Q}_t^{-1}(\mathbf{x}_t - \mathbf{F}_t\mathbf{x}_{t-1} - \mathbf{B}_t\mathbf{u}_t)\Big) \end{aligned} \tag{13}$$

$$\begin{aligned} p(\mathbf{z}_t \mid \mathbf{x}_t) \quad & \sim \mathcal{N}(\mathbf{H}_t\mathbf{x}_t, \mathbf{R}_t) \\ & = \frac{1}{\sqrt{\det(2\pi\mathbf{R}_t)}} \exp\Big(-\frac{1}{2}(\mathbf{z}_t - \mathbf{H}_t\mathbf{x}_t)^\intercal \mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{H}_t\mathbf{x}_t)\Big) \end{aligned} \tag{14}$$

Next, the $\overline{\text{bel}}(\mathbf{x}_t), \text{bel}(\mathbf{x}_t)$ that need to be computed through the Kalman filter can be represented as follows.

$$\begin{aligned} \overline{\text{bel}}(\mathbf{x}_t) &= \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_t)\text{bel}(\mathbf{x}_{t-1})d\mathbf{x}_{t-1} \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}) \\ \text{bel}(\mathbf{x}_t) &= \eta \cdot p(\mathbf{z}_t \mid \mathbf{x}_t)\overline{\text{bel}}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \end{aligned} \tag{15}$$

As seen in (15), the Kalman filter operates by first computing the predicted value $\overline{\text{bel}}(\mathbf{x}_t)$ using the previous step's value and motion model in the prediction phase, then obtaining the corrected value $\text{bel}(\mathbf{x}_t)$ using the measurement and observation model in the correction phase. By substituting (13) and (14) into (15), we can derive the means and covariances for the prediction and correction steps $(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}), (\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$ respectively. For a detailed derivation process, refer to Section 3.2.4 "Mathematical Derivation of the KF" in [[13]].

The initial value $\text{bel}(\mathbf{x}_0)$ is given as follows.

$$\text{bel}(\mathbf{x}_0) \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0) \tag{16}$$

- $\hat{\mathbf{x}}_0$: Typically set to 0
- $\mathbf{P}_0$: Typically set to a small value (<1e-2).

2.1 Prediction step

Prediction involves computing $\overline{\text{bel}}(\mathbf{x}_t)$. Since $\overline{\text{bel}}(\mathbf{x}_t)$ follows a Gaussian distribution, we can compute the mean $\hat{\mathbf{x}}_{t|t-1}$ and variance $\mathbf{P}_{t|t-1}$ as follows.

$$\boxed{\begin{aligned} &\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}_t\mathbf{u}_t \\ &\mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^\intercal + \mathbf{Q}_t \end{aligned}} \tag{17}$$

2.2 Correction step

Correction involves computing $\text{bel}(\mathbf{x}_t)$. Since $\text{bel}(\mathbf{x}_t)$ also follows a Gaussian distribution, we can compute the mean $\hat{\mathbf{x}}_{t|t}$ and variance $\mathbf{P}_{t|t}$ as follows. Here, $\mathbf{K}_t$ represents the Kalman gain.

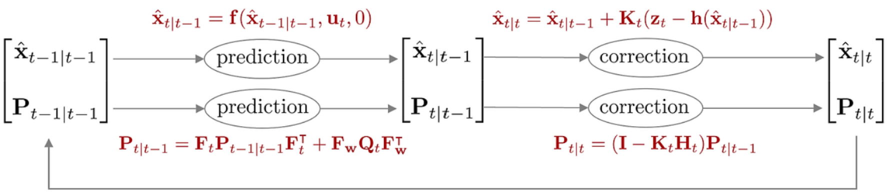
$$\boxed{\begin{aligned} &\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal + \mathbf{R}_t)^{-1} \\ &\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}) \\ &\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1} \end{aligned}} \tag{18}$$

2.3 Summary

The Kalman Filter can be expressed as a function as follows.

$$\boxed{\begin{aligned} &\text{KalmanFilter}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}, \mathbf{u}_t, \mathbf{z}_t)\{ \\ &\quad \text{(Prediction Step)} \\ &\quad \hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}_t\mathbf{u}_t \\ &\quad \mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^\intercal + \mathbf{Q}_t \\ & \\ &\quad \text{(Correction Step)} \\ &\quad \mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal + \mathbf{R}_t)^{-1} \\ &\quad \hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}) \\ &\quad \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1} \\ &\quad \text{return} \;\; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t} \\ &\} \end{aligned}} \tag{19}$$

3 Extended Kalman Filter (EKF)



Kalman filters assume that both the motion model and the observation model are linear to estimate the state. However, since most phenomena in the real world are modeled non-linearly, applying the Kalman filter as previously defined would not work properly. The extended Kalman filter (EKF) was proposed to use the Kalman filter even in non-linear motion and observation models. **EKF uses the first-order Taylor approximation to approximate the non-linear model to a linear model before applying the Kalman filter.** The motion model and observation model of the EKF are as follows.

$$\begin{aligned} \text{Motion Model:} \quad & \mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{w}_t) \\ \text{Observation Model:} \quad & \mathbf{z}_t = \mathbf{h}(\mathbf{x}_t, \mathbf{v}_t) \end{aligned} \tag{20}$$

- $\mathbf{x}_t$: state variable of the model
- $\mathbf{u}_t$: input of the model
- $\mathbf{z}_t$: measurement of the model
- $\mathbf{f}(\cdot)$: non-linear motion model function
- $\mathbf{h}(\cdot)$: non-linear observation model function
- $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$: noise in the motion model. $\mathbf{Q}_t$ represents the covariance matrix of $\mathbf{w}_t$
- $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R}_t)$: noise in the observation model. $\mathbf{R}_t$ represents the covariance matrix of $\mathbf{v}_t$

In the above formula, $\mathbf{f}(\cdot)$ means a non-linear motion model and $\mathbf{h}(\cdot)$ means a non-linear observation model. When the first-order Taylor approximation is performed on $\mathbf{f}(\cdot), \mathbf{h}(\cdot)$, it becomes as follows.

$$\begin{aligned} \mathbf{x}_t &\approx \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) + \mathbf{F}_t(\mathbf{x}_{t-1} - \hat{\mathbf{x}}_{t-1|t-1}) + \mathbf{F}_\mathbf{w}\mathbf{w}_t \\ \mathbf{z}_t &\approx \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0) + \mathbf{H}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{H}_\mathbf{v}\mathbf{v}_t \end{aligned} \tag{21}$$

At this time, $\mathbf{F}_t$ represents the Jacobian matrix of the motion model calculated at $\hat{\mathbf{x}}_{t-1|t-1}$, and $\mathbf{H}_t$ represents the Jacobian matrix of the observation model calculated at $\hat{\mathbf{x}}_{t|t-1}$. And $\mathbf{F}_\mathbf{w}, \mathbf{H}_\mathbf{v}$ represent the Jacobian matrix of the noise when $\mathbf{w}_t = 0, \mathbf{v}_t = 0$ respectively. For more information on the Jacobian, refer to this post.

$$\mathbf{F}_t = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}_{t-1}}\right|_{\mathbf{x}_{t-1}=\hat{\mathbf{x}}_{t-1|t-1}} \qquad \mathbf{F}_\mathbf{w} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{w}_t}\right|_{\substack{\mathbf{x}_{t-1}=\hat{\mathbf{x}}_{t-1|t-1} \\ \mathbf{w}_t=0}} \tag{22}$$

$$\mathbf{H}_t = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}_t}\right|_{\mathbf{x}_t=\hat{\mathbf{x}}_{t|t-1}} \qquad \mathbf{H}_\mathbf{v} = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{v}_t}\right|_{\substack{\mathbf{x}_t=\hat{\mathbf{x}}_{t|t-1} \\ \mathbf{v}_t=0}} \tag{23}$$

(21) If the formula is expanded, it becomes as follows.

$$\boxed{\begin{aligned} \mathbf{x}_t &= \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) - \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{F}_\mathbf{w}\mathbf{w}_t \\ &= \mathbf{F}_t\mathbf{x}_{t-1} + \tilde{\mathbf{u}}_t + \tilde{\mathbf{w}}_t \end{aligned}} \tag{24}$$

- $\tilde{\mathbf{u}}_t = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) - \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1}$ - $\tilde{\mathbf{w}}_t = \mathbf{F}_\mathbf{w}\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal)$

$$\boxed{\begin{aligned} \mathbf{z}_t &= \mathbf{H}_t\mathbf{x}_t + \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0) - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1} + \mathbf{H}_\mathbf{v}\mathbf{v}_t \\ &= \mathbf{H}_t\mathbf{x}_t + \tilde{\mathbf{z}}_t + \tilde{\mathbf{v}}_t \end{aligned}} \tag{25}$$

- $\tilde{\mathbf{z}}_t = \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0) - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}$ - $\tilde{\mathbf{v}}_t = \mathbf{H}_\mathbf{v}\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_v^\intercal)$

Assuming all random variables follow a Gaussian distribution, $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t), p(\mathbf{z}_t|\mathbf{x}_t)$ can be expressed as follows.

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t) \quad & \sim \mathcal{N}(\mathbf{F}_t\mathbf{x}_{t-1} + \tilde{\mathbf{u}}_t, \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal) \\ & = \frac{1}{\sqrt{\det(2\pi\mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal)}} \exp\bigg(-\frac{1}{2}(\mathbf{x}_t - \mathbf{F}_t\mathbf{x}_{t-1} - \tilde{\mathbf{u}}_t)^\intercal(\mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal)^{-1}(\mathbf{x}_t - \mathbf{F}_t\mathbf{x}_{t-1} - \tilde{\mathbf{u}}_t)\bigg) \end{aligned} \tag{26}$$

$$\begin{aligned} p(\mathbf{z}_t|\mathbf{x}_t) \quad & \sim \mathcal{N}(\mathbf{H}_t\mathbf{x}_t + \tilde{\mathbf{z}}_t, \mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal) \\ & = \frac{1}{\sqrt{\det(2\pi\mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal)}} \exp\bigg(-\frac{1}{2}(\mathbf{z}_t - \mathbf{H}_t\mathbf{x}_t - \tilde{\mathbf{z}}_t)^\intercal(\mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal)^{-1}(\mathbf{z}_t - \mathbf{H}_t\mathbf{x}_t - \tilde{\mathbf{z}}_t)\bigg) \end{aligned} \tag{27}$$

$\mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal$ represents the noise in the linearized motion model, and $\mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal$ represents the noise in the linearized observation model. Next, $\overline{\text{bel}}(\mathbf{x}_t), \text{bel}(\mathbf{x}_t)$ that need to be found through the Kalman filter can be expressed as follows.

$$\begin{aligned} \overline{\text{bel}}(\mathbf{x}_t) &= \int p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t)\text{bel}(\mathbf{x}_{t-1})d\mathbf{x}_{t-1} \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}) \\ \text{bel}(\mathbf{x}_t) &= \eta \cdot p(\mathbf{z}_t|\mathbf{x}_t)\overline{\text{bel}}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \end{aligned} \tag{28}$$

EKF operates in a similar manner to KF, using the values from the previous step and the motion model in the prediction to first obtain the predicted value $\overline{\text{bel}}(\mathbf{x}_t)$, and then using the observed values and observation model in the correction to obtain the corrected value $\text{bel}(\mathbf{x}_t)$. Substitute $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t)$, $p(\mathbf{z}_t|\mathbf{x}_t)$ into the above equation to obtain the mean and covariance of the prediction and correction steps $(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}), (\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$ respectively. Refer to section 3.2.4 "Mathematical Derivation of the KF" in [[13]] for detailed derivation.

The initial value $\text{bel}(\mathbf{x}_0)$ is given as follows.

$$\text{bel}(\mathbf{x}_0) \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0) \tag{29}$$

- $\hat{\mathbf{x}}_0$: Typically set to 0
- $\mathbf{P}_0$: Typically set to a small value (<1e-2)

3.1 Prediction step

Prediction involves calculating $\overline{\text{bel}}(\mathbf{x}_t)$. The covariance matrix is obtained using the linearized Jacobian matrix $\mathbf{F}_t$.

$$\boxed{\begin{aligned} & \hat{\mathbf{x}}_{t|t-1} = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) \\ & \mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^\intercal + \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal \end{aligned}} \tag{30}$$

3.2 Correction step

Correction involves calculating $\text{bel}(\mathbf{x}_t)$. The Kalman gain and covariance matrix are obtained using the linearized Jacobian matrix $\mathbf{H}_t$.

$$\boxed{\begin{aligned} & \mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal + \mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal)^{-1} \\ & \hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0)) \\ & \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1} \end{aligned}} \tag{31}$$

3.3 Summary

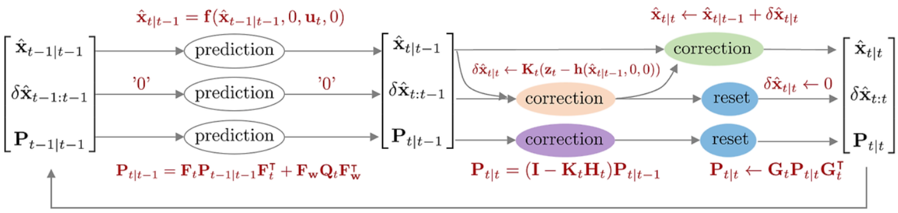
The Extended Kalman Filter can be expressed as a function as follows.

$$
\begin{array}{|l|}
\hline
\text{ExtendedKalmanFilter}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}, \mathbf{u}_t, \mathbf{z}_t)\{ \\
\quad \text{(Prediction Step)} \\
\quad \hat{\mathbf{x}}_{t|t-1} = \mathbf{f}_t(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) \\
\quad \mathbf{P}_{t|t-1} = \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^\intercal + \mathbf{F}_\mathbf{w} \mathbf{Q}_t \mathbf{F}_\mathbf{w}^\intercal \\
\\
\quad \text{(Correction Step)} \\
\quad \mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t^\intercal (\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\intercal + \mathbf{H}_\mathbf{v} \mathbf{R}_t \mathbf{H}_\mathbf{v}^\intercal)^{-1} \\
\quad \hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1}, 0)) \\
\quad \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_{t|t-1} \\
\quad \text{return} \;\; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t} \\
\} \\
\hline
\end{array}
\tag{32}
$$

4 Error-state Kalman Filter (ESKF)

NOMENCLATURE of Error-State Kalman Filter

- prediction: $\overline{\text{bel}}(\delta \mathbf{x}_t) \sim \mathcal{N}(\delta \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$
  - $\delta \hat{\mathbf{x}}_{t|t-1}$: The mean at step t given the correction value at step $t-1$. Some literature also denotes this as $\delta \mathbf{x}_t^-$.
  - $\hat{\mathbf{P}}_{t|t-1}$: The covariance at step t given the correction value at step $t-1$. Some literature also denotes this as $\mathbf{P}_t^-$.
- correction: $\text{bel}(\delta \mathbf{x}_t) \sim \mathcal{N}(\delta \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$
  - $\delta \hat{\mathbf{x}}_{t|t}$: The mean at step t given the prediction value at step t. Some literature also denotes this as $\delta \mathbf{x}_t^+$.
  - $\hat{\mathbf{P}}_{t|t}$: The covariance at step t given the prediction value at step t. Some literature also denotes this as $\mathbf{P}_t^+$.



The error-state Kalman filter (ESKF) is a Kalman filter algorithm that estimates the mean and variance of the error state variable $\delta \mathrm{x}_t$, unlike the EKF which estimates the mean and variance of the conventional state variable x_t. Since it estimates through the error state rather than directly estimating the state variable, it is also called an indirect Kalman filter. It is also known as the error-state extended Kalman filter (ES-EKF). In the ESKF, the conventional state variable is called the true state variable, and it is represented as a sum of the nominal (nominal) state and the error (error) state.

$$
\mathbf{x}_{\text{true},t} = \mathbf{x}_t + \delta \mathbf{x}_t \tag{33}
$$

- $\mathbf{x}_{\text{true},t}$: True state variable updated in conventional KF, EKF at step t
- $\mathbf{x}_t$: Nominal state variable at step t. Represents a state without error
- $\delta \mathbf{x}_t$: Error state variable at step t

If we interpret the above formula, it means that the actual (true) state variable $\mathbf{x}_t$ we want to estimate can be represented as a sum of the general (or nominal) state $\hat{\mathbf{x}}_t$ without errors and the error state $\delta\mathbf{x}_t$ arising from model and sensor noise. Here, the nominal state has (relatively) large values and is nonlinear. In contrast, the error state has small values near zero and is linear. **While the conventional EKF linearizes the nonlinear true (nominal + error) state variable for filtering, resulting in slow speed and accumulated errors over time, the ESKF linearizes only the error state for filtering, thus having faster speed and accuracy.** The advantages of ESKF over conventional EKF are as follows (Madyastha et al., 2011):

- The representation of the orientation error state involves minimal parameters. That is, it has the minimum number of parameters as degrees of freedom, so phenomena like singularities due to over-parameterization do not occur.
- The error state system only operates near the origin, making it easy to linearize. Therefore, parameter singularities like gimbal lock do not occur, and it is always possible to perform linearization.
- Since error states typically have small values, terms of second order and higher can be ignored. This helps perform Jacobian operations quickly and easily. Some Jacobians are also used as constants.

However, while the prediction speed of the ESKF is fast, the correction step where the nominal state with generally nonlinear large values is processed is slow.

The motion model and observation model of the ESKF are as follows.

$$\begin{aligned} &\text{Error-state Motion Model:} && \mathbf{x}_t + \delta\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \delta\mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{w}_t) \\ &\text{Error-state Observation Model:} && \mathbf{z}_t = \mathbf{h}(\mathbf{x}_t, \delta\mathbf{x}_t, \mathbf{v}_t) \end{aligned} \tag{34}$$

- $\mathbf{x}_t$: Model's nominal state variable
- $\delta\mathbf{x}_t$: Model's error state variable
- $\mathbf{u}_t$: Model's input
- $\mathbf{z}_t$: Model's measurement
- $\mathbf{f}(\cdot)$: Nonlinear motion model function
- $\mathbf{h}(\cdot)$: Nonlinear observation model function
- $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$: Noise of the error state model. $\mathbf{Q}_t$ denotes the covariance matrix of $\mathbf{w}_t$
- $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R}_t)$: Noise of the observation model. $\mathbf{R}_t$ denotes the covariance matrix of $\mathbf{v}_t$

When first-order Taylor approximations are performed for $\mathbf{f}(\cdot), \mathbf{h}(\cdot)$, the following expansions result. This development refers to [5].

$$\begin{aligned} &\mathbf{x}_t + \delta\mathbf{x}_t \approx \mathbf{f}(\mathbf{x}_{t-1|t-1}, \delta\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) + \mathbf{F}_t(\delta\mathbf{x}_{t-1} - \delta\hat{\mathbf{x}}_{t-1|t-1}) + \mathbf{F}_\mathbf{w}\mathbf{w}_t \\ &\mathbf{z}_t \approx \mathbf{h}(\mathbf{x}_{t|t-1}, \delta\hat{\mathbf{x}}_{t|t-1}, 0) + \mathbf{H}_t(\delta\mathbf{x}_t - \delta\hat{\mathbf{x}}_{t|t-1}) + \mathbf{H}_\mathbf{v}\mathbf{v}_t \end{aligned} \tag{35}$$

Note that both Jacobians, $\mathbf{F}_t, \mathbf{H}_t$, are for the error state $\delta\mathbf{x}_t$, not the true state $\mathbf{x}_{\mathbf{true},t}$. This is the most significant difference from EKF in the Jacobian part.

$$\boxed{\mathbf{F}_t = \frac{\partial \mathbf{f}}{\partial \delta\mathbf{x}_{t-1}}\bigg|_{\delta\mathbf{x}_{t-1}=\delta\hat{\mathbf{x}}_{t-1|t-1}} \qquad \mathbf{F}_\mathbf{w} = \frac{\partial \mathbf{f}}{\partial \mathbf{w}_t}\bigg|_{\substack{\delta\mathbf{x}_{t-1}=\delta\hat{\mathbf{x}}_{t-1|t-1} \\ \mathbf{w}_t=0}}} \tag{36}$$

$$\boxed{\mathbf{H}_t = \frac{\partial \mathbf{h}}{\partial \delta\mathbf{x}_t}\bigg|_{\delta\mathbf{x}_t=\delta\hat{\mathbf{x}}_{t|t-1}} \qquad \mathbf{H}_\mathbf{v} = \frac{\partial \mathbf{h}}{\partial \mathbf{v}_t}\bigg|_{\substack{\delta\mathbf{x}_t=\delta\hat{\mathbf{x}}_{t|t-1} \\ \mathbf{v}_t=0}}} \tag{37}$$

$\mathbf{H}_t$ can be expressed using the chain rule as follows.

$$\mathbf{H}_t = \frac{\partial \mathbf{h}}{\partial \delta\mathbf{x}_t} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\text{true},t}} \frac{\partial \mathbf{x}_{\text{true},t}}{\partial \delta\mathbf{x}_t} \tag{38}$$

In this, the front part $\frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\text{true},t}}$ is the same Jacobian obtained in EKF, but a Jacobian for the error state variable $\frac{\partial \mathbf{x}_{\text{true},t}}{\partial \delta\mathbf{x}_t}$ has been added. For more details on this, see the post Quaternion kinematics for the error-state Kalman filter summary.

Expanding (35) results in the following.

$$\mathbf{x}_t + \delta\mathbf{x}_t = \mathbf{F}_t\delta\mathbf{x}_{t-1} + \mathbf{f}(\mathbf{x}_{t-1|t-1}, \delta\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) - \mathbf{F}_t\delta\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{F}_\mathbf{w}\mathbf{w}_t \tag{39}$$

Since $\delta\hat{\mathbf{x}}_{t-1|t-1} = 0$ is always initialized at the previous correction step, 0 is substituted for the related terms.

$$\mathbf{x}_t + \delta\mathbf{x}_t = \mathbf{F}_t\delta\mathbf{x}_{t-1} + \mathbf{f}(\mathbf{x}_{t-1|t-1}, 0, \mathbf{u}_t, 0) + \mathbf{F}_\mathbf{w}\mathbf{w}_t \tag{40}$$

The nominal state variable $\mathbf{x}_t$ is identical to $\mathbf{f}(\mathbf{x}_{t-1}, 0, \mathbf{u}_t, 0)$ by definition, so they cancel each other out.

$$\boxed{\begin{aligned}\delta\mathbf{x}_t &= \mathbf{F}_t\delta\mathbf{x}_{t-1} + \mathbf{F}_\mathbf{w}\mathbf{w}_t \\ &= \mathbf{F}_t\delta\mathbf{x}_{t-1} + \tilde{\mathbf{w}}_t \\ &= 0 + \tilde{\mathbf{w}}_t\end{aligned}} \tag{41}$$

- $\tilde{\mathbf{w}}_t = \mathbf{F}_\mathbf{w}\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal)$

The observation model function is similarly developed by substituting the error state variable values with 0.

$$\boxed{\begin{aligned}\mathbf{z}_t &= \mathbf{H}_t\delta\mathbf{x}_t + \mathbf{h}(\mathbf{x}_{t|t-1}, \delta\hat{\mathbf{x}}_{t-1|t-1}, 0) - \mathbf{H}_t\delta\hat{\mathbf{x}}_{t|t-1} + \mathbf{H}_\mathbf{v}\mathbf{v}_t \\ &= \mathbf{h}(\mathbf{x}_{t|t-1}, 0, 0) + \mathbf{H}_\mathbf{v}\mathbf{v}_t \\ &= \tilde{\mathbf{z}}_t + \tilde{\mathbf{v}}_t\end{aligned}} \tag{42}$$

- $\tilde{\mathbf{z}}_t = \mathbf{h}(\mathbf{x}_{t|t-1}, 0, 0)$
- $\tilde{\mathbf{v}}_t = \mathbf{H}_\mathbf{v}\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}^\intercal)$

Assuming all random variables follow a Gaussian distribution, $p(\delta\mathbf{x}_t \mid \delta\mathbf{x}_{t-1}, \mathbf{u}_t), p(\mathbf{z}_t \mid \delta\mathbf{x}_t)$ can be represented as follows.

$$\begin{aligned}p(\delta\mathbf{x}_t \mid \delta\mathbf{x}_{t-1}, \mathbf{u}_t) \quad &\sim \mathcal{N}(0, \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal) \\ &= \frac{1}{\sqrt{\det(2\pi\mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal)}} \exp\bigg(-\frac{1}{2}(\delta\mathbf{x}_t)^\intercal(\mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal)^{-1}(\delta\mathbf{x}_t)\bigg)\end{aligned} \tag{43}$$

$$\begin{aligned}p(\mathbf{z}_t \mid \delta\mathbf{x}_t) \quad &\sim \mathcal{N}(\tilde{\mathbf{z}}_t, \mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal) \\ &= \frac{1}{\sqrt{\det(2\pi\mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal)}} \exp\bigg(-\frac{1}{2}(\mathbf{z}_t - \tilde{\mathbf{z}}_t)^\intercal(\mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal)^{-1}(\mathbf{z}_t - \tilde{\mathbf{z}}_t)\bigg)\end{aligned} \tag{44}$$

$\mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal$ denotes the noise of the linearized error state motion model, and $\mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal$ denotes the noise of the linearized observation model. Next, the Kalman filter needs to derive $\overline{\text{bel}}(\delta\mathbf{x}_t), \text{bel}(\delta\mathbf{x}_t)$ as follows.

$$\begin{aligned}\overline{\text{bel}}(\delta\mathbf{x}_t) &= \int p(\delta\mathbf{x}_t \mid \delta\mathbf{x}_{t-1}, \mathbf{u}_t)\text{bel}(\delta\mathbf{x}_{t-1})d\mathbf{x}_{t-1} \sim \mathcal{N}(\delta\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}) \\ \text{bel}(\delta\mathbf{x}_t) &= \eta \cdot p(\mathbf{z}_t \mid \delta\mathbf{x}_t)\overline{\text{bel}}(\delta\mathbf{x}_t) \sim \mathcal{N}(\delta\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})\end{aligned} \tag{45}$$

ESKF also operates in the same way as EKF, where the prediction first derives the predicted value $\overline{\text{bel}}(\delta\mathbf{x}_t)$ using the previous step's value and motion model, and then the correction derives the corrected value $\text{bel}(\delta\mathbf{x}_t)$ using the measurement value and observation model. **Substituting $p(\delta\mathbf{x}_t|\delta\mathbf{x}_{t-1}, \mathbf{u}_t), p(\mathbf{z}_t|\delta\mathbf{x}_t)$ into the above formula allows you to derive the mean and covariance for the prediction and correction steps, $(\delta\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}), (\delta\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$, respectively. The detailed derivation process can be found in section 3.2.4 "Mathematical Derivation of the KF" of [[13]].**

The initial value $\text{bel}(\delta\mathbf{x}_0)$ is given as follows.

$$\text{bel}(\delta\mathbf{x}_0) \sim \mathcal{N}(0, \mathbf{P}_0) \tag{46}$$

- $\delta\hat{\mathbf{x}}_0 = 0$: Always has a value of 0
- $\mathbf{P}_0$: Typically set to small values (<1e-2).

4.1 Prediction step

Prediction is the process of deriving $\overline{\text{bel}}(\delta\mathbf{x}_t)$. The linearized Jacobian matrix $\mathbf{F}_t$ is used to derive the covariance matrix.

$$\boxed{\begin{aligned}&\delta\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t\delta\hat{\mathbf{x}}_{t-1|t-1} = 0 \quad \leftarrow \text{Always 0}\\ &\hat{\mathbf{x}}_{t|t-1} = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, 0, \mathbf{u}_t, 0)\\ &\mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^\intercal + \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal\end{aligned}} \tag{47}$$

Note that $\mathbf{F}_t$ is the Jacobian for the error state. Since the error state variable $\delta\hat{\mathbf{x}}$ is reset to 0 at every correction step, multiplying it by the linear Jacobian $\mathbf{F}_t$ still results in 0. **Therefore, the value of $\delta\hat{\mathbf{x}}$ always remains 0 in the prediction step.** Thus, the error state $\delta\hat{\mathbf{x}}_{t|t-1}$ remains unchanged in the prediction step, while only the nominal state $\hat{\mathbf{x}}_{t|t-1}$ and the error state covariance $\mathbf{P}_{t|t-1}$ are updated.

4.2 Correction step

Correction is the process of deriving $\text{bel}(\delta\mathbf{x}_t)$. The Kalman gain and covariance matrix are derived using the linearized Jacobian matrix $\mathbf{H}_t$.

$$\boxed{\begin{aligned}&\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal + \mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v})^{-1}\\ &\delta\hat{\mathbf{x}}_{t|t} = \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0, 0))\\ &\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \delta\hat{\mathbf{x}}_{t|t}\\ &\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}\\ &\text{reset } \delta\hat{\mathbf{x}}_{t|t}\\ &\qquad\delta\hat{\mathbf{x}}_{t|t} \leftarrow 0\\ &\qquad\mathbf{P}_{t|t} \leftarrow \mathbf{G}\mathbf{P}_{t|t}\mathbf{G}^\intercal\end{aligned}} \tag{48}$$

Note that in the above formula, $\mathbf{H}_t$ is the Jacobian for the observation model of the error state $\delta\mathbf{x}_t$, not the true state $\mathbf{x}_{\mathbf{true},t}$. The symbols $\mathbf{P}_{t|t-1}, \mathbf{K}_t$ are only similar to those in EKF, but the actual values are different. Thus, while the overall formulas are the same as in EKF, the matrices $\mathbf{F}, \mathbf{H}, \mathbf{P}, \mathbf{K}$ denote values for the error state $\delta\mathbf{x}_t$, which is a key difference.

4.2.1 Reset

Once the nominal state is updated, the next step is to reset the value of the error state to 0. The reset is necessary because it is necessary to represent a new error for the new nominal state. The reset updates the covariance of the error state $\mathbf{P}_{t|t}$.

If the reset function is denoted as $\mathbf{g}(\cdot)$, it can be represented as follows. For more details on this, see the chapter 6 content in the post Quaternion kinematics for the error-state Kalman filter summary.

$$\delta\mathbf{x} \leftarrow \mathbf{g}(\delta\mathbf{x}) = \delta\mathbf{x} - \delta\hat{\mathbf{x}}_{t|t-1} \tag{49}$$

The reset process in ESKF is as follows.

$$\begin{aligned}&\delta\hat{\mathbf{x}}_{t|t} \leftarrow 0\\ &\mathbf{P}_{t|t} \leftarrow \mathbf{G}\mathbf{P}_{t|t}\mathbf{G}^\intercal\end{aligned} \tag{50}$$

$\mathbf{G}$ denotes the Jacobian for the reset defined as follows.

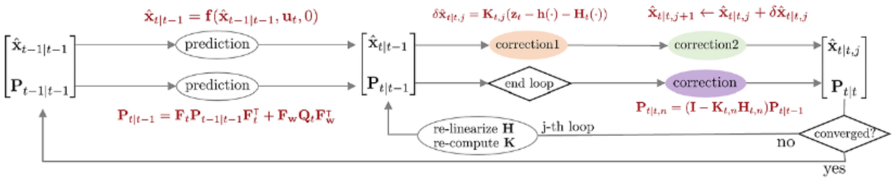
$$\mathbf{G} = \left.\frac{\partial \mathbf{g}}{\partial \delta\mathbf{x}}\right|_{\delta\mathbf{x}_t = \delta\hat{\mathbf{x}}_{t|t}} \tag{51}$$

4.3 Summary

If ESKF is represented as a function, it is as follows.

$$
\begin{array}{l}
\text{ErrorStateKalmanFilter}(\hat{\mathbf{x}}_{t-1|t-1}, \delta\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}, \mathbf{u}_t, \mathbf{z}_t)\{ \\
\quad \text{(Prediction Step)} \\
\quad \delta\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t \hat{\mathbf{x}}_{t-1|t-1} = 0 \quad \leftarrow \text{Always 0} \\
\quad \hat{\mathbf{x}}_{t|t-1} = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, 0, \mathbf{u}_t, 0) \\
\quad \mathbf{P}_{t|t-1} = \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^\intercal + \mathbf{F}_\mathbf{w} \mathbf{Q}_t \mathbf{F}_\mathbf{w}^\intercal \\
\\
\quad \text{(Correction Step)} \\
\quad \mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t^\intercal (\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\intercal + \mathbf{H}_\mathbf{v} \mathbf{R}_t \mathbf{H}_\mathbf{v}^\intercal)^{-1} \\
\quad \delta\hat{\mathbf{x}}_{t|t} = \mathbf{K}_t (\mathbf{z}_t - \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1}, 0, 0)) \\
\quad \hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \delta\hat{\mathbf{x}}_{t|t} \\
\quad \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_{t|t-1} \\
\quad \text{reset } \delta\hat{\mathbf{x}}_{t|t} \\
\qquad \delta\hat{\mathbf{x}}_{t|t} \leftarrow 0 \\
\qquad \mathbf{P}_{t|t} \leftarrow \mathbf{G} \mathbf{P}_{t|t} \mathbf{P}^\intercal \\
\quad \text{return } \; \delta\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t} \\
\}
\end{array}
\tag{52}
$$

5 Iterated Extended Kalman Filter (IEKF)



The Iterated Extended Kalman Filter (IEKF) is an algorithm that repetitively performs the correction step of the Extended Kalman Filter (EKF). Since EKF linearizes non-linear functions to estimate state variables, errors inevitably occur during the linearization process. IEKF reduces these linearization errors by repeatedly performing the linearization if the update change $\delta\hat{\mathbf{x}}_{t|t,j}$ after the correction step is sufficiently large.

In this context, $\delta\hat{\mathbf{x}}$ is interpreted as the update change, not an error state variable. That is, $\hat{\mathbf{x}}_{t|t,j+1} \leftarrow \hat{\mathbf{x}}_{t|t,j} + \delta\hat{\mathbf{x}}_{t|t,j}$ is used solely to update the j-th posterior value to the j+1 posterior value. In other words, it is not the target of state estimation.

5.1 Compare to EKF

5.1.1 Commonality 1

In IEKF, the motion and observation models are as follows, which are exactly the same as in EKF.

$$
\begin{array}{ll}
\text{Motion Model:} & \mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{w}_t) \\
\text{Observation Model:} & \mathbf{z}_t = \mathbf{h}(\mathbf{x}_t, \mathbf{v}_t)
\end{array}
\tag{53}
$$

- $\mathbf{x}_t$: state variable of the model - $\mathbf{u}_t$: input of the model - $\mathbf{z}_t$: measurement of the model - $\mathbf{f}(\cdot)$: non-linear motion model function - $\mathbf{h}(\cdot)$: non-linear observation model function - $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$: noise in the motion model. $\mathbf{Q}_t$ represents the covariance matrix of $\mathbf{w}_t$ - $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R}_t)$: noise in the observation model. $\mathbf{R}_t$ represents the covariance matrix of $\mathbf{v}_t$

5.1.2 Commonality 2

Next, the linearization process is also exactly the same as EKF's equations (24), (25).

$$\boxed{\begin{aligned}\mathbf{x}_t &= \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) - \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{F}_\mathbf{w}\mathbf{w}_t \\ &= \mathbf{F}_t\mathbf{x}_{t-1} + \tilde{\mathbf{u}}_t + \tilde{\mathbf{w}}_t\end{aligned}} \tag{54}$$

- $\tilde{\mathbf{u}}_t = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) - \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1}$ - $\tilde{\mathbf{w}}_t = \mathbf{F}_\mathbf{w}\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal)$

$$\boxed{\begin{aligned}\mathbf{z}_t &= \mathbf{H}_t\mathbf{x}_t + \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0) - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1} + \mathbf{H}_\mathbf{v}\mathbf{v}_t \\ &= \mathbf{H}_t\mathbf{x}_t + \tilde{\mathbf{z}}_t + \tilde{\mathbf{v}}_t\end{aligned}} \tag{55}$$

- $\tilde{\mathbf{z}}_t = \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0) - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}$ - $\tilde{\mathbf{v}}_t = \mathbf{H}_\mathbf{v}\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{H}_\mathbf{v}\mathbf{R}_t\mathbf{H}_\mathbf{v}^\intercal)$

5.1.3 Commonality 3

IEKF, like EKF, uses the matrices $\mathbf{F}, \mathbf{H}, \mathbf{K}$ true to the state variables $\mathbf{x}_{\text{true}}$.

5.1.4 Difference 1

EKF: The correction value is obtained from the prediction value in one step.
IEKF: The correction value becomes the prediction value again and the correction step is repeated iteratively.

$$\boxed{\begin{aligned}&\text{EKF Correction :} && \hat{\mathbf{x}}_{t|t-1} \rightarrow \hat{\mathbf{x}}_{t|t} \\ &\text{IEKF Correction :} && \hat{\mathbf{x}}_{t|t,j} \leftrightarrows \hat{\mathbf{x}}_{t|t,j+1}\end{aligned}} \tag{56}$$

- j: j-th iteration

5.1.5 Difference 2

Expanding equation (55) to create the innovation term $\mathbf{r}_t$ gives the following.

$$\mathbf{r}_t = \mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0) - \mathbf{H}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) \tag{57}$$

In EKF, looking at the second line of correction step (31), it can be seen that the Kalman gain $\mathbf{K}_t$ is multiplied by $\mathbf{r}_t$ to compute the posterior.

$$\begin{aligned}\text{EKF correction :} \quad \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{r}_t) \\ &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0))\end{aligned} \tag{58}$$

EKF: The $\mathbf{H}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})$ part is eliminated when $\mathbf{x}_t = \hat{\mathbf{x}}_{t|t-1}$ is substituted, and only the remaining parts are used.

IEKF: Linearization for a new state (new operating point) is performed at each moment of the correction step, so this part is not eliminated.

$$\boxed{\begin{aligned}&\text{EKF innovation term :} && \mathbf{r}_t = \mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0) \\ &\text{IEKF innovation term :} && \mathbf{r}_{t,j} = \mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}, 0) - \mathbf{H}_{t,j}(\hat{\mathbf{x}}_{t|t-1} - \hat{\mathbf{x}}_{t|t,j})\end{aligned}} \tag{59}$$

If it is the first iteration $j = 0$, $\hat{\mathbf{x}}_{t|t,0} = \hat{\mathbf{x}}_{t|t-1}$ so it is eliminated resulting in a formula similar to EKF, but from $j = 1$ onwards, different values are used so it is not eliminated.

5.2 Prediction step

Prediction involves the process of determining $\overline{\text{bel}}(\mathbf{x}_t)$. The linearized Jacobian matrix $\mathbf{F}_t$ is used when computing the covariance matrix. This is identical to the prediction step in EKF.

$$\boxed{\begin{aligned}&\hat{\mathbf{x}}_{t|t-1} = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) \\ &\mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^\intercal + \mathbf{F}_\mathbf{w}\mathbf{Q}_t\mathbf{F}_\mathbf{w}^\intercal\end{aligned}} \tag{60}$$

5.3 Correction step

Correction involves the process of determining $\text{bel}(\mathbf{x}_t)$. The Kalman gain and covariance matrix are computed using the linearized Jacobian matrix $\mathbf{H}_t$. The IEKF's correction step is performed iteratively until the update change $\delta\hat{\mathbf{x}}_{t|t,j}$ is sufficiently small.

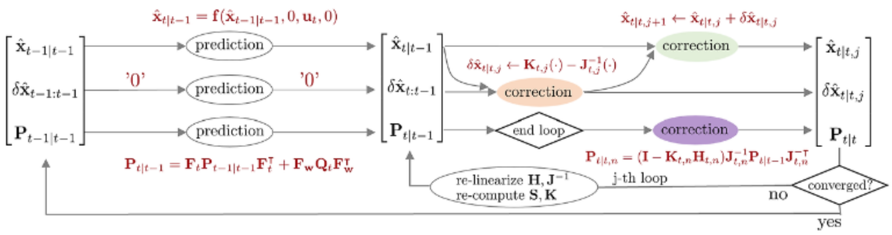
$$
\boxed{
\begin{array}{l}
\text{set } \epsilon \\
\text{start j-th loop} \\
\quad \mathbf{K}_{t,j} = \mathbf{P}_{t|t-1}\mathbf{H}_{t,j}^{\intercal}(\mathbf{H}_{t,j}\mathbf{P}_{t|t-1}\mathbf{H}_{t,j}^{\intercal} + \mathbf{H}_{\mathbf{v},j}\mathbf{R}_t\mathbf{H}_{\mathbf{v},j}^{\intercal})^{-1} \\
\quad \delta\hat{\mathbf{x}}_{t|t,j} = \mathbf{K}_{t,j}(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}, 0) - \mathbf{H}_t(\hat{\mathbf{x}}_{t|t-1} - \hat{\mathbf{x}}_{t|t,j})) \\
\quad \hat{\mathbf{x}}_{t|t,j+1} = \hat{\mathbf{x}}_{t|t,j} + \delta\hat{\mathbf{x}}_{t|t,j} \\
\quad \text{iterate until } \delta\hat{\mathbf{x}}_{t|t,j} < \epsilon. \\
\text{end loop} \\
\mathbf{P}_{t|t,n} = (\mathbf{I} - \mathbf{K}_{t,n}\mathbf{H}_{t,n})\mathbf{P}_{t|t-1}
\end{array}
}
\tag{61}
$$

5.4 Summary

If IEKF is represented as a function, it would be as follows.

$$
\boxed{
\begin{array}{l}
\text{IteratedExtendedKalmanFilter}(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}, \mathbf{u}_t, \mathbf{z}_t)\{ \\
\quad \text{(Prediction Step)} \\
\quad \hat{\mathbf{x}}_{t|t-1} = \mathbf{f}_t(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{u}_t, 0) \\
\quad \mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^{\intercal} + \mathbf{F}_{\mathbf{w}}\mathbf{Q}_t\mathbf{F}_{\mathbf{w}}^{\intercal} \\
\\
\quad \text{(Correction Step)} \\
\quad \text{set } \epsilon \\
\quad \text{start j-th loop} \\
\qquad \mathbf{K}_{t,j} = \mathbf{P}_{t|t-1}\mathbf{H}_{t,j}^{\intercal}(\mathbf{H}_{t,j}\mathbf{P}_{t|t-1}\mathbf{H}_{t,j}^{\intercal} + \mathbf{H}_{\mathbf{v},j}\mathbf{R}_t\mathbf{H}_{\mathbf{v},j}^{\intercal})^{-1} \\
\qquad \delta\hat{\mathbf{x}}_{t|t,j} = \mathbf{K}_{t,j}(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}, 0) - \mathbf{H}_t(\hat{\mathbf{x}}_{t|t-1} - \hat{\mathbf{x}}_{t|t,j})) \\
\qquad \hat{\mathbf{x}}_{t|t,j+1} = \hat{\mathbf{x}}_{t|t,j} + \delta\hat{\mathbf{x}}_{t|t,j} \\
\qquad \text{iterate until } \delta\hat{\mathbf{x}}_{t|t,j} < \epsilon. \\
\quad \text{end loop} \\
\quad \mathbf{P}_{t|t,n} = (\mathbf{I} - \mathbf{K}_{t,n}\mathbf{H}_{t,n})\mathbf{P}_{t|t-1} \\
\quad \text{return } \ \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t} \\
\}
\end{array}
}
\tag{62}
$$

6 Iterated Error-State Kalman Filter (IESKF)



While IEKF estimates the true state variable $\mathbf{x}_{\text{true}}$ iteratively in the correction step, IESKF is an algorithm that iteratively estimates the error state variable $\delta\mathbf{x}_t$. The relationship between these state variables is as follows.

$$\mathbf{x}_{\text{true},t} = \mathbf{x}_t + \delta\mathbf{x}_t \tag{63}$$

- $\mathbf{x}_{\text{true},t}$: true state variable at step t updated in the conventional KF, EKF
- $\mathbf{x}_t$: nominal state variable at step t. Represents the state without error
- $\delta\mathbf{x}_t$: error state variable at step t

The iterative update process in IESKF involves understanding $\mathbf{x}_{\text{true},t}$ as the post-update state, nominal $\mathbf{x}_t$ as the pre-update state, and the error state variable $\delta\mathbf{x}_t$ as the update amount.

6.1 Compare to ESKF

6.1.1 Commonality 1

The motion and observation models of IESKF are as follows, which are identical to those of ESKF.

$$\begin{aligned} &\text{Error-state Motion Model:} && \mathbf{x}_t + \delta\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \delta\mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{w}_t) \\ &\text{Error-state Observation Model:} && \mathbf{z}_t = \mathbf{h}(\mathbf{x}_t, \delta\mathbf{x}_t, \mathbf{v}_t) \end{aligned} \tag{64}$$

- $\mathbf{x}_t$: model's nominal state variable
- $\delta\mathbf{x}_t$: model's error state variable
- $\mathbf{u}_t$: model's input
- $\mathbf{z}_t$: model's observation value
- $\mathbf{f}(\cdot)$: nonlinear motion model function
- $\mathbf{h}(\cdot)$: nonlinear observation model function
- $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$: noise in the error state model. $\mathbf{Q}_t$ represents the covariance matrix of $\mathbf{w}_t$
- $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R}_t)$: noise in the observation model. $\mathbf{R}_t$ is the covariance matrix of $\mathbf{v}_t$

6.1.2 Commonality 2

Both Jacobians $\mathbf{F}_t, \mathbf{H}_t$ of IESKF refer to the Jacobians for the error state $\delta\mathbf{x}_t$, similar to ESKF.

$$\boxed{\mathbf{F}_t = \frac{\partial \mathbf{f}}{\partial \delta\mathbf{x}_{t-1}}\bigg|_{\delta\mathbf{x}_{t-1}=\delta\hat{\mathbf{x}}_{t-1|t-1}} \qquad \mathbf{F}_{\mathbf{w}} = \frac{\partial \mathbf{f}}{\partial \mathbf{w}_t}\bigg|_{\substack{\delta\mathbf{x}_{t-1}=\delta\hat{\mathbf{x}}_{t-1|t-1} \\ \mathbf{w}_t=0}}} \tag{65}$$

$$\boxed{\mathbf{H}_t = \frac{\partial \mathbf{h}}{\partial \delta\mathbf{x}_t}\bigg|_{\delta\mathbf{x}_t=\delta\hat{\mathbf{x}}_{t|t-1}} \qquad \mathbf{H}_{\mathbf{v}} = \frac{\partial \mathbf{h}}{\partial \mathbf{v}_t}\bigg|_{\substack{\delta\mathbf{x}_t=\delta\hat{\mathbf{x}}_{t|t-1} \\ \mathbf{v}_t=0}}} \tag{66}$$

$\mathbf{H}_t$ can be expressed through the chain rule as follows.

$$\mathbf{H}_t = \frac{\partial \mathbf{h}}{\partial \delta\mathbf{x}_t} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\text{true},t}} \frac{\partial \mathbf{x}_{\text{true},t}}{\partial \delta\mathbf{x}_t} \tag{67}$$

6.1.3 Commonality 3

The linearized error state variable $\hat{\mathbf{x}}_t$ can also be derived identically to ESKF.

$$\boxed{\begin{aligned} \delta\mathbf{x}_t &= \mathbf{F}_t \delta\mathbf{x}_{t-1} + \mathbf{F}_{\mathbf{w}}\mathbf{w}_t \\ &= \mathbf{F}_t \delta\mathbf{x}_{t-1} + \tilde{\mathbf{w}}_t \\ &= 0 + \tilde{\mathbf{w}}_t \end{aligned}} \tag{68}$$

- $\tilde{\mathbf{w}}_t = \mathbf{F}_{\mathbf{w}}\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{F}_{\mathbf{w}}\mathbf{Q}_t\mathbf{F}_{\mathbf{w}}^\intercal)$

6.1.4 Difference (MAP-based derivation)

ESKF derivation:

The ESKF correction step unfolds the following equation to derive the mean $\delta\hat{\mathbf{x}}_{t|t}$ and covariance $\mathbf{P}_{t|t}$.

$$\text{bel}(\delta\mathbf{x}_t) = \eta \cdot p(\mathbf{z}_t|\delta\mathbf{x}_t)\overline{\text{bel}}(\delta\mathbf{x}_t) \sim \mathcal{N}(\delta\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \tag{69}$$

- $p(\mathbf{z}_t|\delta\mathbf{x}_t) \sim \mathcal{N}(\tilde{\mathbf{z}}_t, \mathbf{H_v}\mathbf{R}_t\mathbf{H_v^\intercal})$: (see 44)
- $\overline{\text{bel}}(\delta\mathbf{x}_t) \sim \mathcal{N}(\delta\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$

$$\boxed{\begin{aligned}&\delta\hat{\mathbf{x}}_{t|t} = \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0, 0))\\ &\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}\end{aligned}} \tag{70}$$

IESKF derivation:

IESKF uses maximum a posteriori (MAP) for deriving the correction step. For MAP estimation, the Gauss-Newton optimization method is utilized. For more details, refer to section 7. This derivation process is based on references [6], [7], [8], [9], [10], and [11], among which [6] provides the most detailed explanation of the IESKF derivation process.

If EKF is derived using the MAP approach, the final equation to optimize is as follows:

$$\arg\min_{\mathbf{x}_t} \quad \left\|\mathbf{z}_t - \mathbf{h}(\mathbf{x}_{\text{true},t}, 0)\right\|_{\mathbf{H_v}\mathbf{R}^{-1}\mathbf{H_v^\intercal}} + \left\|\mathbf{x}_{\text{true},t} - \hat{\mathbf{x}}_{t|t-1}\right\|_{\mathbf{P}^{-1}_{t|t-1}} \tag{71}$$

When expressed in terms of the nominal state $\mathbf{x}_t$ and error state $\delta\mathbf{x}_t$, it becomes the following:

$$\arg\min_{\delta\hat{\mathbf{x}}_{t|t}} \quad \left\|\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t, \delta\mathbf{x}_t, 0)\right\|_{\mathbf{H_v}\mathbf{R}^{-1}\mathbf{H_v^\intercal}} + \left\|\delta\hat{\mathbf{x}}_{t|t-1}\right\|_{\mathbf{P}^{-1}_{t|t-1}} \tag{72}$$

> **Tip**
>
> Prediction error state variable $\delta\hat{\mathbf{x}}_{t|t-1}$ is defined as follows:
>
> $$\begin{aligned}\delta\hat{\mathbf{x}}_{t|t-1} &= \mathbf{x}_{\text{true},t} - \hat{\mathbf{x}}_{t|t-1}\\ &\sim \mathcal{N}(0, \mathbf{P}_{t|t-1})\end{aligned} \tag{73}$$
>
> Posterior (correction) error state variable $\delta\hat{\mathbf{x}}_{t|t}$ is defined as follows:
>
> $$\delta\hat{\mathbf{x}}_{t|t} = \mathbf{x}_{\text{true},t|t} - \hat{\mathbf{x}}_{t|t} \tag{74}$$
>
> If the equation is rearranged, the following formula is established:
>
> $$\mathbf{x}_{\text{true},t|t} = \hat{\mathbf{x}}_{t|t} + \delta\hat{\mathbf{x}}_{t|t} \tag{75}$$

The first part of (72) should be linearized similarly to (42).

$$\mathbf{z}_t - \mathbf{h}(\mathbf{x}_{t|t-1}, \delta\hat{\mathbf{x}}_{t|t-1}, 0) \approx \mathbf{z}_t - \mathbf{h}(\mathbf{x}_{t|t-1}, 0, 0) - \mathbf{H}_t\delta\mathbf{x}_t \tag{76}$$

In ESKF, $\delta\mathbf{x}_t$ is always 0, thus it was removed, but in IESKF it retains a non-zero value, so it is not removed. The next part of (72) with (75) substituted in unfolds as follows:

$$\begin{aligned}\delta\hat{\mathbf{x}}_{t|t-1} &= \mathbf{x}_{\text{true},t} - \hat{\mathbf{x}}_{t|t-1}\\ &= (\hat{\mathbf{x}}_{t|t} + \delta\hat{\mathbf{x}}_{t|t}) - \hat{\mathbf{x}}_{t|t-1}\\ &\approx \hat{\mathbf{x}}_{t|t} - \hat{\mathbf{x}}_{t|t-1} + \mathbf{J}_t\delta\hat{\mathbf{x}}_{t|t}\\ &\sim \mathcal{N}(0, \mathbf{P}_{t|t-1})\end{aligned} \tag{77}$$

Here, $\mathbf{J}_t$ is defined as follows:

$$\mathbf{J}_t = \frac{\partial}{\partial\delta\hat{\mathbf{x}}_{t|t}}\left((\hat{\mathbf{x}}_{t|t} + \delta\hat{\mathbf{x}}_{t|t}) - \hat{\mathbf{x}}_{t|t-1}\right)\bigg|_{\delta\hat{\mathbf{x}}_{t|t}=0} \tag{78}$$

If the third and fourth lines of (77) are rearranged and simplified, the following formula is obtained:

$$\delta\hat{\mathbf{x}}_{t|t} \sim \mathcal{N}(-\mathbf{J}_t^{-1}(\hat{\mathbf{x}}_{t|t} - \hat{\mathbf{x}}_{t|t-1}),\ \mathbf{J}_t^{-1}\mathbf{P}_{t|t-1}\mathbf{J}_t^{-\intercal}) \tag{79}$$

Having calculated (76) and (77), the MAP estimation problem becomes as follows:

$$\begin{aligned}\arg\min_{\delta\hat{\mathbf{x}}_{t|t}} \quad & \left\|\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0, 0) - \mathbf{H}_t\delta\hat{\mathbf{x}}_{t|t}\right\|_{\mathbf{H_v}\mathbf{R}^{-1}\mathbf{H_v^\intercal}} \\ & + \left\|\hat{\mathbf{x}}_{t|t} - \hat{\mathbf{x}}_{t|t-1} + \mathbf{J}_t\delta\hat{\mathbf{x}}_{t|t}\right\|_{\mathbf{P}_{t|t-1}^{-1}}\end{aligned} \tag{80}$$

IESKF iteratively estimates the error state variable $\delta\hat{\mathbf{x}}_{t|t}$ until it converges to a value less than a specific threshold ϵ. The expression for the j-th iteration is as follows:

$$\boxed{\delta\hat{\mathbf{x}}_{t|t,j} \sim \mathcal{N}(-\mathbf{J}_{t,j}^{-1}(\hat{\mathbf{x}}_{t|t,j} - \hat{\mathbf{x}}_{t|t-1}),\ \mathbf{J}_{t,j}^{-1}\mathbf{P}_{t|t-1}\mathbf{J}_{t,j}^{-\intercal})} \tag{81}$$

$$\boxed{\begin{aligned}\arg\min_{\delta\hat{\mathbf{x}}_{t|t,j}} \quad & \left\|\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}, 0, 0) - \mathbf{H}_t\delta\hat{\mathbf{x}}_{t|t,j}\right\|_{\mathbf{H_v}\mathbf{R}^{-1}\mathbf{H_v^\intercal}} \\ & + \left\|\hat{\mathbf{x}}_{t|t,j} - \hat{\mathbf{x}}_{t|t-1} + \mathbf{J}_{t,j}\delta\hat{\mathbf{x}}_{t|t,j}\right\|_{\mathbf{P}_{t|t-1}^{-1}}\end{aligned}} \tag{82}$$

6.2 Prediction step

Prediction involves the process of determining $\overline{\text{bel}}(\delta\mathbf{x}_t)$. The linearized Jacobian matrix $\mathbf{F}_t$ is used when calculating the covariance matrix. This is completely identical to ESKF.

$$\boxed{\begin{aligned}&\delta\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t\delta\hat{\mathbf{x}}_{t-1|t-1} = 0 \quad \leftarrow \text{Always } 0 \\ &\hat{\mathbf{x}}_{t|t-1} = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, 0, \mathbf{u}_t, 0) \\ &\mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^\intercal + \mathbf{F_w}\mathbf{Q}_t\mathbf{F_w^\intercal}\end{aligned}} \tag{83}$$

6.3 Correction step

Correction involves the process of determining $\text{bel}(\mathbf{x}_t)$. The Kalman gain and covariance matrix are calculated using the linearized Jacobian matrix $\mathbf{H}_t$. The correction step of IESKF is performed iteratively until the update amount $\delta\hat{\mathbf{x}}_{t|t,j}$ becomes sufficiently small. The derived formula (82) is differentiated and set to zero as follows:

$$\boxed{\begin{aligned}&\text{set } \epsilon \\ &\text{start j-th loop} \\ &\quad \mathbf{S}_{t,j} = \mathbf{H}_{t,j}\mathbf{J}_{t,j}^{-1}\mathbf{P}_{t|t-1}\mathbf{J}_{t,j}^{-\intercal}\mathbf{H}_{t,j}^\intercal + \mathbf{H}_{\mathbf{v},j}\mathbf{R}_t\mathbf{H}_{\mathbf{v},j}^\intercal \\ &\quad \mathbf{K}_{t,j} = \mathbf{J}_{t,j}^{-1}\mathbf{P}_{t|t-1}\mathbf{J}_{t,j}^{-\intercal}\mathbf{H}_{t,j}^\intercal\mathbf{S}_{t,j}^{-1} \\ &\quad \delta\hat{\mathbf{x}}_{t|t,j} = \mathbf{K}_{t,j}\Big(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}, 0, 0) + \mathbf{H}_{t,j}\mathbf{J}_{t,j}^{-1}(\hat{\mathbf{x}}_{t|t,j} - \hat{\mathbf{x}}_{t|t-1})\Big) - \mathbf{J}_{t,j}^{-1}(\hat{\mathbf{x}}_{t|t,j} - \hat{\mathbf{x}}_{t|t-1}) \\ &\quad \hat{\mathbf{x}}_{t|t,j+1} = \hat{\mathbf{x}}_{t|t,j} \boxplus \delta\hat{\mathbf{x}}_{t|t,j} \\ &\quad \text{iterate until } \delta\hat{\mathbf{x}}_{t|t,j} < \epsilon. \\ &\text{end loop} \\ &\mathbf{P}_{t|t,n} = (\mathbf{I} - \mathbf{K}_{t,n}\mathbf{H}_{t,n})\mathbf{J}_{t,n}^{-1}\mathbf{P}_{t|t-1}\mathbf{J}_{t,n}^{-\intercal}\end{aligned}} \tag{84}$$

6.4 Summary

IESKF can be expressed as a function as follows:

$$
\begin{array}{|l|}
\hline
\text{IteratedErrorStateKalmanFilter}(\hat{\mathbf{x}}_{t-1|t-1}, \delta\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}, \mathbf{u}_t, \mathbf{z}_t)\{ \\
\quad \text{(Prediction Step)} \\
\quad \delta\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t \hat{\mathbf{x}}_{t-1|t-1} = 0 \quad \leftarrow \text{Always 0} \\
\quad \hat{\mathbf{x}}_{t|t-1} = \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}, 0, \mathbf{u}_t, 0) \\
\quad \mathbf{P}_{t|t-1} = \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^\intercal + \mathbf{F}_\mathbf{w} \mathbf{Q}_t \mathbf{F}_\mathbf{w}^\intercal \\
\\
\quad \text{(Correction Step)} \\
\quad \text{set } \epsilon \\
\quad \text{start j-th loop} \\
\qquad \mathbf{S}_{t,j} = \mathbf{H}_{t,j} \mathbf{J}_{t,j}^{-1} \mathbf{P}_{t|t-1} \mathbf{J}_{t,j}^{-\intercal} \mathbf{H}_{t,j}^\intercal + \mathbf{H}_{\mathbf{v},j} \mathbf{R}_t \mathbf{H}_{\mathbf{v},j}^\intercal \\
\qquad \mathbf{K}_{t,j} = \mathbf{J}_{t,j}^{-1} \mathbf{P}_{t|t-1} \mathbf{J}_{t,j}^{-\intercal} \mathbf{H}_{t,j}^\intercal \mathbf{S}_{t,j}^{-1} \\
\qquad \delta\hat{\mathbf{x}}_{t|t,j} = \mathbf{K}_{t,j} \Big(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}, 0, 0) + \mathbf{H}_{t,j} \mathbf{J}_{t,j}^{-1} (\hat{\mathbf{x}}_{t|t,j} - \hat{\mathbf{x}}_{t|t-1}) \Big) - \mathbf{J}_{t,j}^{-1} (\hat{\mathbf{x}}_{t|t,j} - \hat{\mathbf{x}}_{t|t-1}) \\
\qquad \hat{\mathbf{x}}_{t|t,j+1} = \hat{\mathbf{x}}_{t|t,j} \boxplus \delta\hat{\mathbf{x}}_{t|t,j} \\
\qquad \text{iterate until } \delta\hat{\mathbf{x}}_{t|t,j} < \epsilon. \\
\quad \text{end loop} \\
\quad \mathbf{P}_{t|t,n} = (\mathbf{I} - \mathbf{K}_{t,n} \mathbf{H}_{t,n}) \mathbf{J}_{t,n}^{-1} \mathbf{P}_{t|t-1} \mathbf{J}_{t,n}^{-\intercal} \\
\quad \text{return } \ \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t} \\
\} \\
\hline
\end{array} \tag{85}
$$

7 MAP, GN, and EKF relationship

7.1 Traditional EKF derivation

Let's consider the observation model function for the EKF given as follows. For the convenience of development, the observation noise $\mathbf{v}_t$ is positioned outside.

$$
\text{Observation Model:} \quad \mathbf{z}_t = \mathbf{h}(\mathbf{x}_t) + \mathbf{v}_t \tag{86}
$$

The correction step of the EKF unfolds the following formula to derive the mean $\hat{\mathbf{x}}_{t|t}$ and the covariance $\mathbf{P}_{t|t}$.

$$
\text{bel}(\mathbf{x}_t) = \eta \cdot p(\mathbf{z}_t \mid \mathbf{x}_t) \overline{\text{bel}}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \tag{87}
$$

- $p(\mathbf{z}_t \mid \mathbf{x}_t) \sim \mathcal{N}(\mathbf{h}_t(\mathbf{x}_t), \mathbf{R}_t)$
- $\overline{\text{bel}}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$

$$
\boxed{\begin{aligned} \hat{\mathbf{x}}_{t|t} &= \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1})) \\ \mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_{t|t-1} \end{aligned}} \tag{88}
$$

7.2 MAP-based EKF derivation

7.2.1 Start from MAP estimator

As a method for deriving the correction step, one can use the maximum a posteriori (MAP) estimation that maximizes the probability of the posterior. The details were written with reference to [12].

$$
\begin{aligned}
\hat{\mathbf{x}}_{t|t} &= \arg\max_{\mathbf{x}_t} \operatorname{bel}(\mathbf{x}_t) \\
&= \arg\max_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \quad \cdots \text{posterior} \\
&\propto \arg\max_{\mathbf{x}_t} p(\mathbf{z}_t|\mathbf{x}_t)\overline{\operatorname{bel}}(\mathbf{x}_t) \quad \cdots \text{likelihood} \cdot \text{prior} \\
&\propto \arg\max_{\mathbf{x}_t} \exp\bigg(-\frac{1}{2}\Big[(\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t))^\intercal \mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t)) \\
&\qquad\qquad + (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\intercal \mathbf{P}_{t|t-1}^{-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})\Big]\bigg)
\end{aligned}
\tag{89}
$$

- $p(\mathbf{z}_t \mid \mathbf{x}_t) \sim \mathcal{N}(\mathbf{h}_t(\mathbf{x}_t), \mathbf{R}_t)$
- $\overline{\operatorname{bel}}(\mathbf{x}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$

Removing the negative sign turns the maximization problem into a minimization problem, and it can be organized into the following optimization formula.

$$
\begin{aligned}
\hat{\mathbf{x}}_{t|t} \propto \arg\min_{\mathbf{x}_t} \exp\bigg(&(\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t))^\intercal \mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t)) \\
&+ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\intercal \mathbf{P}^{-1}{}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})\bigg)
\end{aligned}
\tag{90}
$$

$$
\boxed{\hat{\mathbf{x}}_{t|t} = \arg\min_{\mathbf{x}_t} \|\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t)\|_{\mathbf{R}_t^{-1}} + \left\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}\right\|_{\mathbf{P}_{t|t-1}^{-1}}}
\tag{91}
$$

- $\|\mathbf{a}\|_{\mathbf{B}} = \mathbf{a}^\intercal\mathbf{B}\mathbf{a}$: Mahalanobis norm

Expanding the formula within (91) and defining it as a cost function $\mathbf{C}_t$ results in the following. Note that switching the order of $\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}$ does not affect the overall value.

$$
\mathbf{C}_t = (\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t))^\intercal \mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{h}(\mathbf{x}_t)) + (\hat{\mathbf{x}}_{t|t-1} - \mathbf{x}_t)^\intercal \mathbf{P}_{t|t-1}^{-1}(\hat{\mathbf{x}}_{t|t-1} - \mathbf{x}_t)
\tag{92}
$$

This can be expressed in matrix form as follows.

$$
\mathbf{C}_t = \begin{bmatrix} \hat{\mathbf{x}}_{t|t-1} - \mathbf{x}_t \\ \mathbf{z}_t - \mathbf{h}(\mathbf{x}_t) \end{bmatrix}^\intercal \begin{bmatrix} \mathbf{P}_{t|t-1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_t^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{t|t-1} - \mathbf{x}_t \\ \mathbf{z}_t - \mathbf{h}(\mathbf{x}_t) \end{bmatrix}
\tag{93}
$$

7.2.2 MLE of new observation function

A new observation function satisfying the above equation can be defined as follows.

$$
\boxed{\begin{aligned} \mathbf{y}_t &= \mathbf{g}(\mathbf{x}_t) + \mathbf{e}_t \\ &\sim \mathcal{N}(\mathbf{g}(\mathbf{x}_t), \mathbf{P}_\mathbf{e}) \end{aligned}}
\tag{94}
$$

- $\mathbf{y}_t = \begin{bmatrix} \hat{\mathbf{x}}_{t|t-1} \\ \mathbf{z}_t \end{bmatrix}$
- $\mathbf{g}(\mathbf{x}_t) = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{h}(\mathbf{x}_t) \end{bmatrix}$
- $\mathbf{e}_t \sim \mathcal{N}(0, \mathbf{P}_\mathbf{e})$
- $\mathbf{P}_\mathbf{e} = \begin{bmatrix} \mathbf{P}_{t|t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_t \end{bmatrix}$

The nonlinear function $\mathbf{g}(\mathbf{x}_t)$ can be linearized as follows.

$$
\begin{aligned}
\mathbf{g}(\mathbf{x}_t) &\approx \mathbf{g}(\hat{\mathbf{x}}_{t|t-1}) + \mathbf{J}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) \\
&= \mathbf{g}(\hat{\mathbf{x}}_{t|t-1}) + \mathbf{J}_t\delta\hat{\mathbf{x}}_{t|t-1}
\end{aligned}
\tag{95}
$$

The Jacobian $\mathbf{J}_t$ is as follows.

$$
\begin{aligned}
\mathbf{J}_t &= \frac{\partial \mathbf{g}}{\partial \mathbf{x}_t}\bigg|_{\mathbf{x}_t=\hat{\mathbf{x}}_{t|t-1}} \\
&\frac{\partial \begin{bmatrix} \mathbf{x}_t \\ \mathbf{h}(\mathbf{x}_t) \end{bmatrix}}{\partial \mathbf{x}_t}\Bigg|_{\mathbf{x}_t=\hat{\mathbf{x}}_{t|t-1}} \\
&\frac{\partial \begin{bmatrix} \mathbf{x}_t \\ \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}) + \mathbf{H}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) \end{bmatrix}}{\partial \mathbf{x}_t}\Bigg|_{\mathbf{x}_t=\hat{\mathbf{x}}_{t|t-1}} \\
&= \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_t \end{bmatrix}
\end{aligned} \tag{96}
$$

Therefore, the likelihood for $\mathbf{y}_t$ can be developed as follows.

$$
\begin{aligned}
p(\mathbf{y}_t|\mathbf{x}_t) &\sim \mathcal{N}(\mathbf{g}(\mathbf{x}_t), \mathbf{P}_\mathbf{e}) \\
&= \eta \cdot \exp\Big(-\frac{1}{2}(\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t))^\intercal \mathbf{P}_\mathbf{e}^{-1}(\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t)) \Big)
\end{aligned} \tag{97}
$$

즉, 기존의 bel($\mathbf{x}_t$)에 대한 maximum a posteriori(MAP) 문제는 $p(\mathbf{y}_t|\mathbf{x}_t)$에 대한 maximum likelihood estimation(MLE) 문제를 푸는 것으로 귀결된다. (97) 식을 MLE로 풀면 다음과 같다.

$$
\begin{aligned}
\hat{\mathbf{x}}_{t|t} &= \arg\max_{\mathbf{x}_t} p(\mathbf{y}_t|\mathbf{x}_t) \\
&\propto \arg\min_{\mathbf{x}_t} -\ln p(\mathbf{y}_t|\mathbf{x}_t) \\
&\propto \arg\min_{\mathbf{x}_t} \frac{1}{2}(\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t))^\intercal \mathbf{P}_\mathbf{e}^{-1}(\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t)) \\
&\propto \arg\min_{\mathbf{x}_t} \|\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t)\|_{\mathbf{P}_\mathbf{e}^{-1}}
\end{aligned} \tag{98}
$$

7.2.3 Gauss-Newton Optimization

(98) 식은 최소제곱법의 형태를 지닌다. 특히 가중치 $\mathbf{P}_\mathbf{e}^{-1}$가 중간에 곱해지므로 weighted least squares(WLS) 라고도 부른다. 식을 선형화한 후 다시 정리하면 아래와 같다.

$$
\begin{aligned}
\hat{\mathbf{x}}_{t|t} &= \arg\min_{\mathbf{x}_t} \|\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t)\|_{\mathbf{P}_\mathbf{e}^{-1}} \\
&= \arg\min_{\mathbf{x}_t} \big\|\mathbf{y}_t - \mathbf{g}(\hat{\mathbf{x}}_{t|t-1}) - \mathbf{J}_t \delta\hat{\mathbf{x}}_{t|t-1}\big\|_{\mathbf{P}_\mathbf{e}^{-1}} \\
&= \arg\min_{\mathbf{x}_t} \big\|\mathbf{J}_t \delta\hat{\mathbf{x}}_{t|t-1} - (\mathbf{y}_t - \mathbf{g}(\hat{\mathbf{x}}_{t|t-1}))\big\|_{\mathbf{P}_\mathbf{e}^{-1}} \\
&= \arg\min_{\mathbf{x}_t} \big\|\mathbf{J}_t \delta\hat{\mathbf{x}}_{t|t-1} - \mathbf{r}_t\big\|_{\mathbf{P}_\mathbf{e}^{-1}}
\end{aligned} \tag{99}
$$

- $\delta\hat{\mathbf{x}}_{t|t-1} = \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}$: 표현의 편의를 위해 $\mathbf{x}_t$를 true 상태로 표현

선형화된 residual term $\mathbf{r}_t$는 다음과 같다.

$$
\begin{aligned}
\mathbf{r}_t &= \mathbf{y}_t - \mathbf{g}(\hat{\mathbf{x}}_{t|t-1}) \\
&= \mathbf{J}_t \delta\hat{\mathbf{x}}_{t|t-1} + \mathbf{e} \\
&\sim \mathcal{N}(0, \mathbf{P}_\mathbf{e})
\end{aligned} \tag{100}
$$

GN의 정규방정식으로 통해 해를 구하면 다음과 같다.

$$
(\mathbf{J}_t^\intercal \mathbf{P}_\mathbf{e}^{-1} \mathbf{J}_t)\delta\hat{\mathbf{x}}_{t|t-1} = \mathbf{J}_t^\intercal \mathbf{P}_\mathbf{e}^{-1} \mathbf{r}_t \tag{101}
$$

$$
\boxed{\therefore \delta\hat{\mathbf{x}}_{t|t-1} = (\mathbf{J}_t^\intercal \mathbf{P}_\mathbf{e}^{-1} \mathbf{J}_t)^{-1} \mathbf{J}_t^\intercal \mathbf{P}_\mathbf{e}^{-1} \mathbf{r}_t} \tag{102}
$$

위 식에서 $(\mathbf{J}_t^\intercal \mathbf{P}_\mathbf{e}^{-1} \mathbf{J}_t)$ 부분을 일반적으로 근사 헤시안(approximate hessian) 행렬 $\tilde{\mathbf{H}}$이라고 부른다.

Posterior covariance matrix $\mathbf{P}_{t|t}$:

$\mathbf{P}_{t|t}$는 다음과 같이 구할 수 있다.

$$\boxed{\begin{aligned}\mathbf{P}_{t|t} =& \mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\intercal] \\ =& \mathbb{E}(\delta\hat{\mathbf{x}}_{t|t-1}\delta\hat{\mathbf{x}}_{t|t-1}^\intercal) \\ =& \mathbb{E}\Big[(\mathbf{J}_t^\intercal \mathbf{P}_\mathbf{e}^{-1}\mathbf{J}_t)^{-1}\mathbf{J}_t^\intercal\mathbf{P}_\mathbf{e}^{-1}\mathbf{r}_t\mathbf{r}_t^\intercal\mathbf{P}_\mathbf{e}^{-\intercal}\mathbf{J}_t(\mathbf{J}_t^\intercal\mathbf{P}_\mathbf{e}^{-1}\mathbf{J}_t)^{-\intercal}\Big] \\ =& (\mathbf{J}_t^\intercal \mathbf{P}_\mathbf{e}^{-1}\mathbf{J}_t)^{-1}\mathbf{J}_t^\intercal\mathbf{P}_\mathbf{e}^{-1}\mathbb{E}(\mathbf{r}_t\mathbf{r}_t^\intercal)\mathbf{P}_\mathbf{e}^{-\intercal}\mathbf{J}_t(\mathbf{J}_t^\intercal\mathbf{P}_\mathbf{e}^{-1}\mathbf{J}_t)^{-\intercal} \quad \leftarrow \mathbb{E}(\mathbf{r}_t\mathbf{r}_t^\intercal) = \mathbf{P}_\mathbf{e} \\ =& \left(\mathbf{J}_t^\intercal\mathbf{P}_\mathbf{e}^{-1}\mathbf{J}_t\right)^{-1} \\ =& \left(\begin{bmatrix}\mathbf{I} & \mathbf{H}_t^\intercal\end{bmatrix}\begin{bmatrix}\mathbf{P}_{t|t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_t\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{I} \\ \mathbf{H}_t\end{bmatrix}\right)^{-1} \\ =& \left(\mathbf{P}_{t|t-1}^{-1} + \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\mathbf{H}_t\right)^{-1} \\ =& \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^\intercal + \mathbf{R}_t)^{-1}\mathbf{H}_t\mathbf{P}_{t|t-1} \quad \leftarrow \text{matrix inversion lemmas} \\ =& (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}\end{aligned}} \tag{103}$$

위 식의 다섯번 째 줄에서 보다시피 GN을 통해 구한 근사 헤시안 행렬의 역함수 $\tilde{\mathbf{H}}^{-1}$와 EKF의 posterior 공분산 $\mathbf{P}_{t|t}$는 같은 값을 가지는 것을 알 수 있다.

$$\tilde{\mathbf{H}}^{-1} = (\mathbf{J}_t^\intercal\mathbf{P}_\mathbf{e}^{-1}\mathbf{J}_t)^{-1} = \mathbf{P}_{t|t} \tag{104}$$

Posterior mean $\mathbf{x}_{t|t}$:

GN을 반복적으로 수행함에 따라 j번째 $\mathbf{x}_{t|t,j}$는 다음과 같이 구할 수 있다.

$$\boxed{\begin{aligned}\hat{\mathbf{x}}_{t|t,j+1} =& \hat{\mathbf{x}}_{t|t,j} + \delta\hat{\mathbf{x}}_{t|t,j} \\ =& \hat{\mathbf{x}}_{t|t,j} + (\mathbf{J}_t^\intercal\mathbf{P}_{t|t-1}^{-1}\mathbf{J}_t)^{-1}(\mathbf{J}_t^\intercal\mathbf{P}_{t|t-1}^{-1}\mathbf{r}_t) \\ =& (\mathbf{J}_t^\intercal\mathbf{P}_{t|t-1}^{-1}\mathbf{J}_t)^{-1}\mathbf{J}_t^\intercal\mathbf{P}_{t|t-1}^{-1}(\mathbf{y}_t - \mathbf{g}(\hat{\mathbf{x}}_{t|t,j}) + \mathbf{J}_t\hat{\mathbf{x}}_{t|t,j}) \quad \leftarrow \mathbf{r}_t = \mathbf{y}_t - \mathbf{g}(\hat{\mathbf{x}}_{t|t,j}) \\ =& \left(\mathbf{P}_{t|t-1}^{-1} + \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\mathbf{H}_t\right)^{-1}\begin{bmatrix}\mathbf{P}_{t|t-1}^{-1} & \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}_{t|t-1} \\ \mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}) + \mathbf{H}_t\hat{\mathbf{x}}_{t|t,j}\end{bmatrix} \quad \leftarrow \text{expand } \mathbf{J}_t \\ =& \left(\mathbf{P}_{t|t-1}^{-1} + \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\mathbf{H}_t\right)^{-1}\left(\mathbf{H}_t^\intercal\mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}) + \mathbf{H}_t\hat{\mathbf{x}}_{t|t,j}) + \mathbf{P}_{t|t-1}^{-1}\hat{\mathbf{x}}_{t|t-1}\right) \\ =& \left(\mathbf{P}_{t|t-1}^{-1} + \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\mathbf{H}_t\right)^{-1}\left(\mathbf{H}^\intercal\mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}) - \mathbf{H}_t(\hat{\mathbf{x}}_{t|t-1} - \hat{\mathbf{x}}_{t|t,j}) + \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}) + \mathbf{P}_{t|t-1}^{-1}\hat{\mathbf{x}}_{t|t-1}\right) \\ =& \left(\mathbf{P}_{t|t-1}^{-1} + \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\mathbf{H}_t\right)^{-1}\left(\mathbf{H}^\intercal\mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}) - \mathbf{H}_t(\hat{\mathbf{x}}_{t|t-1} - \hat{\mathbf{x}}_{t|t,j})) + \underbrace{(\mathbf{P}_{t|t-1}^{-1} + \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\mathbf{H}_t)\hat{\mathbf{x}}_{t|t-1}}_{\hat{\mathbf{x}}_{t|t-1}}\right) \\ =& \hat{\mathbf{x}}_{t|t-1} + \left(\mathbf{P}_{t|t-1}^{-1} + \mathbf{H}_t^\intercal\mathbf{R}_t^{-1}\mathbf{H}_t\right)^{-1}\mathbf{H}_t^\intercal\mathbf{R}_t^{-1}(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}) - \mathbf{H}_t(\hat{\mathbf{x}}_{t|t-1} - \hat{\mathbf{x}}_{t|t,j})) \\ =& \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t,j}) - \mathbf{H}_t(\hat{\mathbf{x}}_{t|t-1} - \hat{\mathbf{x}}_{t|t,j}))\end{aligned}} \tag{105}$$

위 식은 IEKF의 (61) 식과 동일하다. 마지막 식에서 보다시피 Gauss-Newton을 통해 EKF의 해를 추정하는 것과 IEKF를 통해 해를 추정하는 것은 동일한 의미를 지닌다. 만약 처음 iteration $j = 0$인 경우 $\hat{\mathbf{x}}_{t|t,0} = \hat{\mathbf{x}}_{t|t-1}$이 되어서 식은 다음과 같이 정리된다.

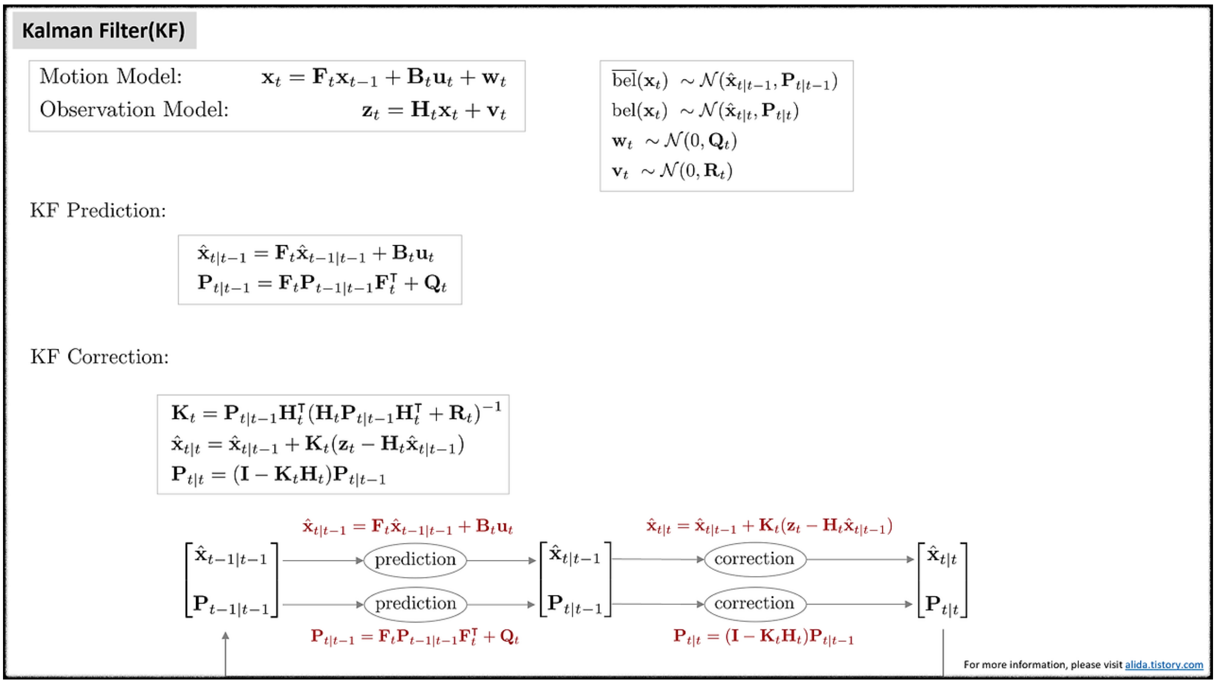
$$\boxed{\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1}))} \tag{106}$$

이는 EKF의 해와 동일하다. 즉, EKF는 GN iteration=1과 동일한 의미를 지니며 IEKF는 GN과 동일한 연산을 수행하는 것을 알 수 있다.

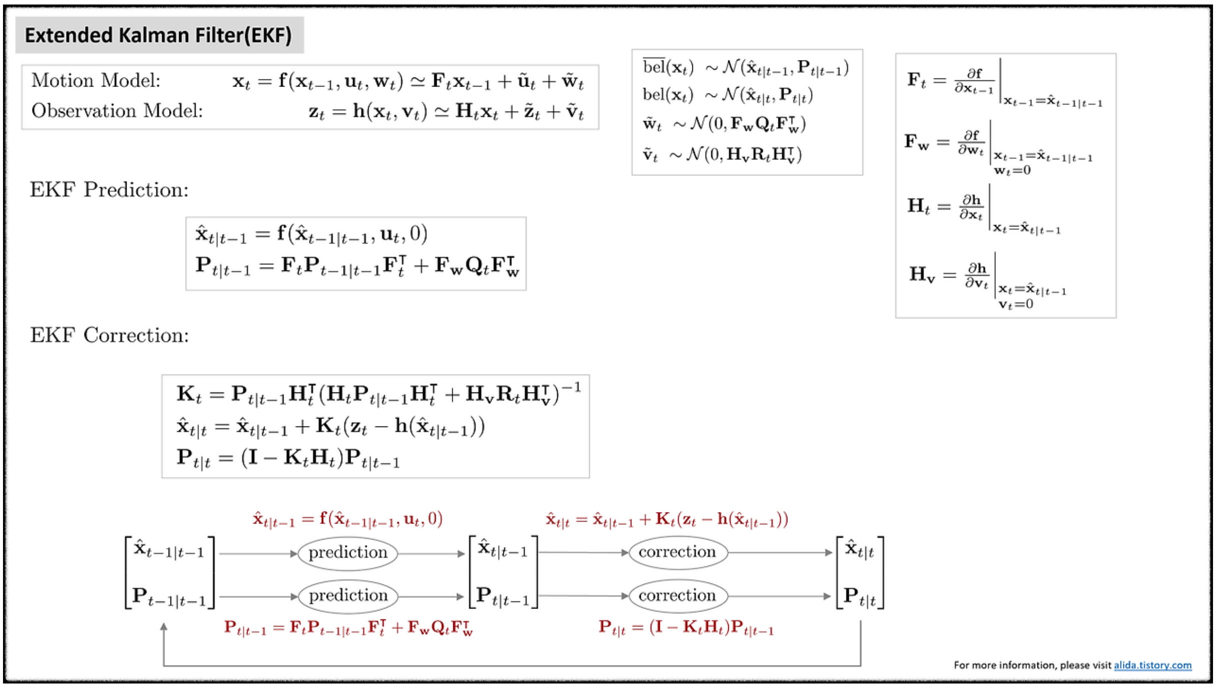
8 Wrap-up

지금까지 설명한 KF, EKF, ESKF, IEKF, IESKF를 한 장의 슬라이드로 표현하면 다음과 같다. 클릭하면 큰 그림으로 볼 수 있다.

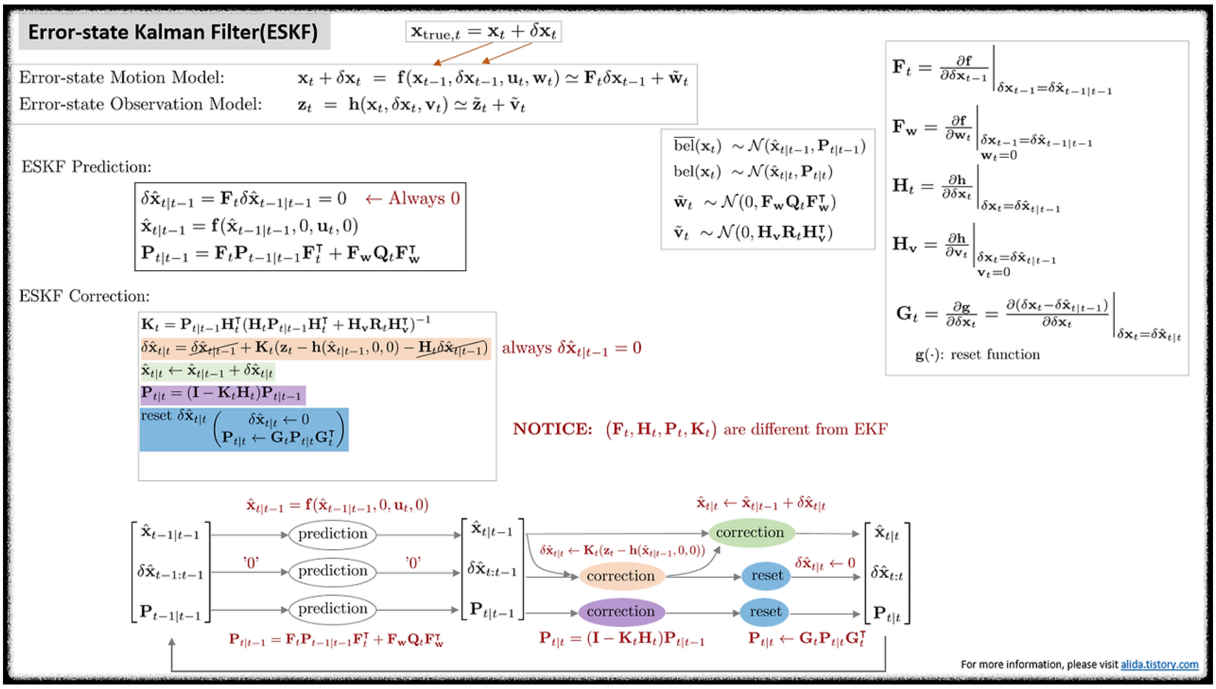
8.1 Kalman Filter (KF)



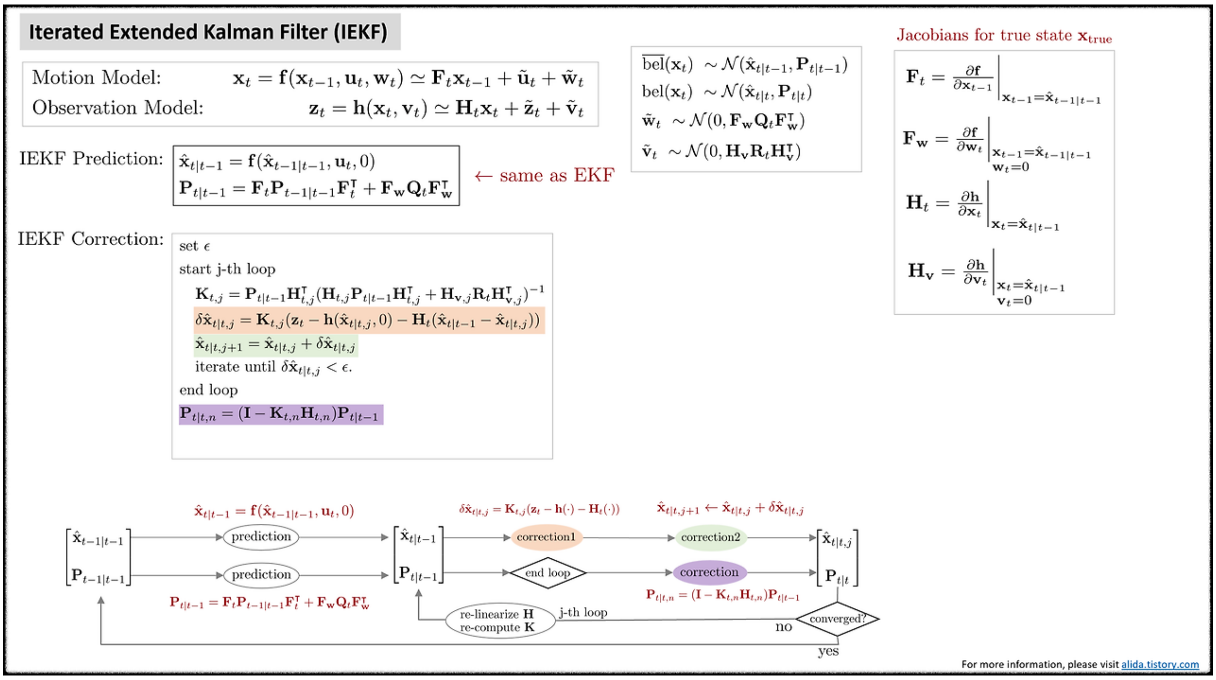
8.2 Extended Kalman Filter (EKF)



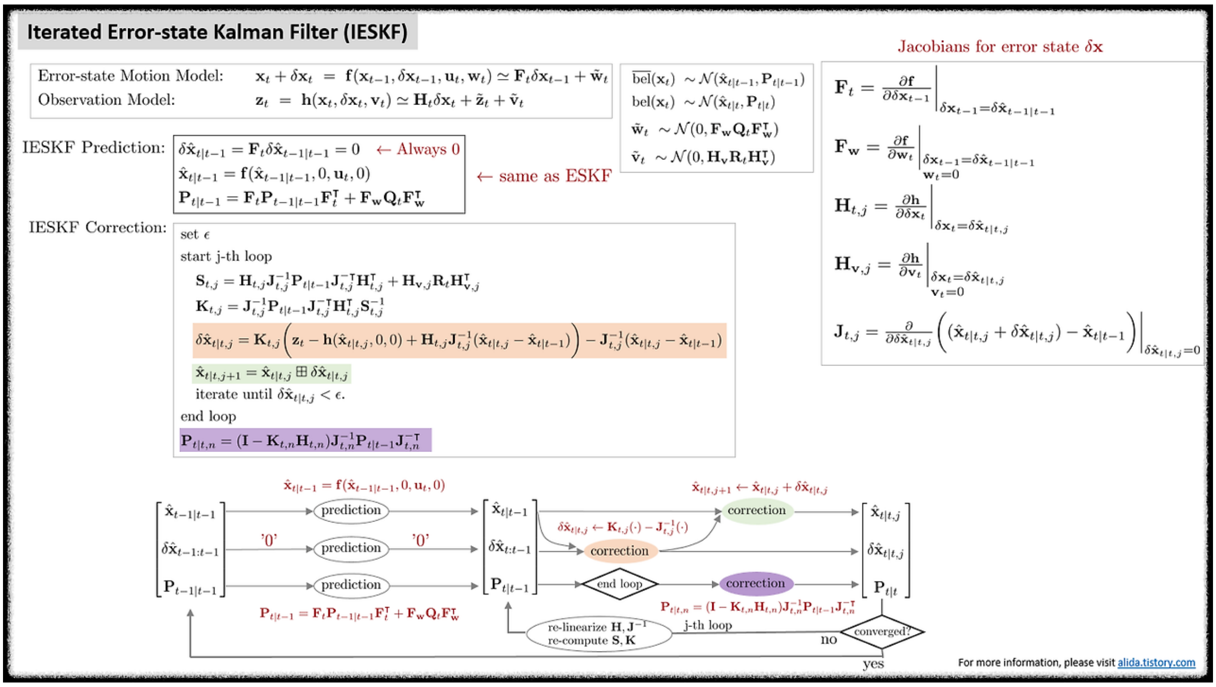
8.3 Error-state Kalman Filter (ESKF)



8.4 Iterated Extended Kalman Filter (IEKF)



8.5 Iterated Error-state Kalman Filter (IESKF)



9 Reference

[1] (Wiki) Kalman Filter

[2] (Paper) Sola, Joan. "Quaternion kinematics for the error-state Kalman filter." arXiv preprint arXiv:1711.02508 (2017).

[3] (Youtube) Robot Mapping Coure - Freiburg Univ

[4] (Blog) [SLAM] Kalman filter and EKF(Extended Kalman Filter) - Jinyong Jeong

[5] (Blog) Error-State Kalman Filter understanding and formula derivation - CSDN

[6] (Paper) He, Dongjiao, Wei Xu, and Fu Zhang. "Kalman filters on differentiable manifolds." arXiv preprint arXiv:2102.03804 (2021).

[7] (Book) SLAM in Autonomous Driving book (SAD book)

[8] (Paper) Xu, Wei, and Fu Zhang. "Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter." IEEE Robotics and Automation Letters 6.2 (2021): 3317-3324.

[9] (Paper) Huai, Jianzhu, and Xiang Gao. "A Quick Guide for the Iterated Extended Kalman Filter on Manifolds." arXiv preprint arXiv:2307.09237 (2023).

[10] (Paper) Bloesch, Michael, et al. "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback." The International Journal of Robotics Research 36.10 (2017): 1053-1072.

[11] (Paper) Skoglund, Martin A., Gustaf Hendeby, and Daniel Axehill. "Extended Kalman filter modifications based on an optimization view point." 2015 18th International Conference on Information Fusion (Fusion). IEEE, 2015.

[12] (Blog) From MAP, MLE, OLS, GN to IEKF, EKF

[13] (Book) Thrun, Sebastian. "Probabilistic robotics." Communications of the ACM 45.3 (2002): 52-57.

10 Revision log

- 1st: 2020-06-23
- 2nd: 2020-06-24
- 3rd: 2020-06-26
- 4th: 2023-01-21
- 5th: 2023-01-31
- 6th: 2023-02-02
- 7th: 2023-02-04
- 8th: 2024-02-08
- 9th: 2024-02-09