1 Baysian Filter

Let x_t, z_t be the state and measurement over time. The Bayesian belief $g(x_t)$ at the time t is the probability of the hidden state x_t given by the history of measurements $z_{1:t} = z_1, \dots, z_t$.

$$g(x_t) := p(x_t | z_{1:t})$$

By using Bayes' theorem on

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \implies p(A|B,C) = \frac{p(B|A,C)p(A|C)}{p(B|C)},$$

we have

$$g(x_t) = p(x_t|z_{1:t}) = p(x_t|z_t, z_{1:t-1}) = \frac{p(z_t|x_t, z_{1:t-1})p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})}.$$

Since the true state x is assumed to be a Markov process, we can simplify the equation above into

$$g(x_t) = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})}.$$

In practice, we would design some algorithm to obtain $p(z_t|x_t)$ and $p(x_t|z_{1:t-1})$, so the denominator $p(z_t|z_{1:t-1})$ can be considered as a coefficient to make $g(x_t)$ to be a valid probability.
Or we may write

$$g(x_t) \propto p(z_t|x_t)p(x_t|z_{1:t-1}).$$

Next, let's expand the term $p(x_t|z_{1:t-1})$ w.r.t x_{t-1}

$$\begin{aligned}
p(x_t|z_{1:t-1}) &= \int p(x_t|x_{t-1}, z_{1:t-1})p(x_{t-1}|z_{1:t-1})dx_{t-1} \\
&= \int p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1})dx_{t-1}, \text{(again, the Markov assumption)} \\
&= \int p(x_t|x_{t-1})g(x_{t-1})dx_{t-1}.
\end{aligned}$$

Finally we have a recursive update process,

$$g(x_t) \leftarrow \eta \cdot p(z_t|x_t) \int p(x_t|x_{t-1})g(x_{t-1})dx_{t-1},$$

with the terms $\eta, p(z_t|x_t), p(x_t|x_{t-1})$ representing some coefficient, the measurements, and the transition model, respectively.

2 Kalman Filter

2.1 Matrix Calculus Background

Trace Formula

$$tr(AB) = A_{ij}B_{ji}$$

$$tr(ABC) = A_{ij}(BC)_{ji} = A_{ij}B_{jk}C_{ki}$$

$$d(trA) = tr(dA)$$

$$F(A) = tr(AB), F'(A) = B^T$$

$$d(tr(AB)) = tr(dA\ B) = dA_{ij}\ B_{ji} \implies F'(A) = B^T$$

$$F(A) = tr(BA^T) = tr(AB^T) \implies F'(A) = B$$

2.2 Model

A physical model

$$\begin{aligned} x_t &= F_t x_{t-1} + w_t \\ z_t &= H_t x_t + v_t, \end{aligned}$$

where

$$\begin{aligned} \text{process noise: } W_t &\sim N(0, Q_t) \\ \text{observation noise: } V_t &\sim N(0, R_t) \\ Q_t &:= \text{cov}(W_t) \\ R_t &:= \text{cov}(V_t) \end{aligned}$$

A posteriori state estimate at time t_1 given observations up to and including at time t_2 is denoted as $\hat{x}_{t_1|t_2}$.
And the other state variable in the filter is the posterior covariance of the estimated accuracy

$$P_{t_1|t_2} := \text{cov}(x_{t_1} - \hat{x}_{t_1|t_2})$$

In predict stage, we can use our priori knowledge $\hat{x}_{t-1|t-1}$ to predict the next state estimate

$$\hat{x}_{t|t-1} = F_t\hat{x}_{t-1|t-1} + B_t u_t$$

and the next estimate covariance

$$
\begin{aligned}
P_{t|t-1} &= \operatorname{cov}(x_t - \hat{x}_{t|t-1}) = \operatorname{cov}(F_t x_{t-1} + W_t - (F_t \hat{x}_{t-1|t-1} + B_t u_t)) \\
&= \operatorname{cov}\left(F_t(x_{t-1} - \hat{x}_{t-1|t-1}) + W_t - B_t u_t\right) \\
&= \operatorname{cov}\left(F_t(x_{t-1} - \hat{x}_{t-1|t-1})\right) + \operatorname{cov}(W_t) + \cancelto{0}{\operatorname{cov}(B_t u_t)} \\
&\text{(separate cov terms due to independence)} \\
&= F_t P_{t-1|t-1} F_t^T + Q_t.
\end{aligned}
$$

In update stage, the innovation representing the difference between the observation and the forecast is

$$
\hat{y}_t = z_t - H_t \hat{x}_{t|t-1}.
$$

Plus, the covariance of innovation S_t can easily computed by

$$
\begin{aligned}
S_t &= \operatorname{cov}(z_t - H_t \hat{x}_{t|t-1}) = \operatorname{cov}(H_t(x_t - x_{t|t-1}) + v_t) \\
&= H_t \operatorname{cov}(x_t - x_{t|t-1}) H_t^T + \operatorname{cov}(v_t) \\
&\text{(separate due to the independence)} \\
&= H_t P_{t|t-1} H_t^T + R_t
\end{aligned}
$$

And we can use the above difference to weight update the state estimate (the posteriori)

$$
\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \hat{y}_t.
$$

Colloquially, the gain factor K_t control the linear interpolation

$$
\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \hat{y}_t = (1 - K_t H_t)\hat{x}_{t|t-1} + K_t z_t.
$$

A lower K_t means higher error in the sensor.
The goal of the Kalman filter is to find the optimal gain to minimize the error.
Let's derive the posterior estimate covariance matrix

$$
\begin{aligned}
P_{t|t} &= \operatorname{cov}(x_t - \hat{x}_{t|t}) = \operatorname{cov}(x_t - (1 - K_t H_t)\hat{x}_{t|t-1} - K_t z_t) \\
&= \operatorname{cov}(x_t - (1 - K_t H_t)\hat{x}_{t|t-1} - K_t H_t x_t - K_t v_t) \\
&= \operatorname{cov}((1 - K_t H_t)(x_t - \hat{x}_{t|t-1}) - K_t v_t) \\
&\text{(separate terms due to the independence)} \\
&= \operatorname{cov}((1 - K_t H_t)(x_t - \hat{x}_{t|t-1})) + \operatorname{cov}(K_t v_t) \\
&= (1 - K_t H_t) \operatorname{cov}(x_t - \hat{x}_{t|t-1})(1 - K_t H_t)^T + K_t \operatorname{cov}(v_t) K_t^T \\
&= (1 - K_t H_t) P_{t|t-1} (1 - K_t H_t)^T + K_t R_t K_t^T.
\end{aligned}
$$

Now let's find the Kalman gain of the minimizing probelm

$$
\operatorname{argmin}_{K_t} \mathbb{E}\left(\|x_t - \hat{x}_{t|t}\|^2\right),
$$

which is equivalent to minimize $\mathrm{tr}(P_{t|t})$.
Since

$$\begin{aligned}
P_{t|t} &= (1 - K_t H_t) P_{t|t-1} (1 - K_t H_t)^T + K_t R_t K_t^T \\
&= P_{t|t-1} - K_t H_t P_{t|t-1} - P_{t|t-1} H_t^T K_t^T + K_t (H_t P_{t|t-1} H_t^T + R_t) K_t^T \\
&= P_{t|t-1} - K_t H_t P_{t|t-1} - P_{t|t-1} H_t^T K_t^T + K_t S_t K_t^T,
\end{aligned}$$

we can solve the optimal gain by letting

$$\begin{aligned}
\frac{d\mathrm{tr}(P_{t|t})}{K_t} &= 0 = -2(H_t P_{t|t-1})^T + K_t (S_t + S_t^T) \\
&= -2 P_{t|t-1} H_t^T + 2 K_t S_t \\
&\quad \text{(using the symmetry of the covariance matrices } P_{t|t-1}, S_t\text{)},
\end{aligned}$$

which implies

$$K_t = P_{t|t-1} H_t^T S_t^{-1}.$$

And we can further simplify the posterior error covariance by

$$\begin{aligned}
& \because K_t S_t K_t^T = P_{t|t-1} H_t^T K_t^T \\
\Longrightarrow \; P_{t|t} &= P_{t|t-1} - K_t H_t P_{t|t-1} + (-P_{t|t-1} H_t^T K_t^T + K_t S_t K_t^T) \\
&= (1 - K_t H_t) P_{t|t-1}.
\end{aligned}$$