Kalman filter

intuition: scalar case

$$\text{KG} = \text{Kalman Gain} = \frac{E_{est}}{E_{est} + E_{meas}} \qquad 0 \leq KG \leq 1$$

$$est_t = est_{t-1} + KG[mea - est_{t-1}]$$

E_{est} = error in the estimate. If the KG is large (close to 1), that means E_{meas} is small (accurate measurement, unstable estimate) therefore we want to move est closer towards mea. On the other hand, if KG is small that means E_{meas} is large (inaccurate measurement, stable estimates) we don't want to take much of the difference, $mea - est_{t-1}$.

$$E_{est_t} = \frac{E_{mea} E_{est_{t-1}}}{E_{mea} + E_{est_{t-1}}} \Rightarrow E_{est_t} = [1 - KG] E_{est_{t-1}}$$

In the above error in the estimate update, we observe that if KG is large ($1-KG$ is small) then E_{meas} converge faster. On the other hand, if KG is small, it takes longer time for E_{meas} to converge (larger $(1 - KG)$).
recap:

1. calculate Kalman Gain (KG)
2. calculate current estimate
3. update estimate error

model state space

The Kalman filter assumes that we have a state vector $\mathbf{x} \in \mathbb{R}^n$ which evolves in the following way:

$$x_k = A_{k-1} x_{k-1} + B_{k-1} u_{k-1} + q_{k-1}$$

where $A_{k-1} \in \mathbb{R}^{n \times n}$ is the state transition matrix, $q_{k-1} \in \mathbb{R}^n$ is process noise. $B_{k-1} \in \mathbb{R}^{n \times r}$ is a matrix that multiplies the control vector $u \in \mathbb{R}^{r \times 1}$. In addition we need to have some measurements $y_k \in \mathbb{R}^m$, modeled as:

$$y_k = H x_k + r_k$$

Here $H \in \mathbb{R}^{m \times n}$ denotes the measure matrix, and it being constant here means we measure the same linear combinations of the state-vector components at each time. The vector r_k is another noise, which models measurement errors.

In order to use the Kalman filter, we need to suppose that the initial distribution x_0 from which it starts is Gaussian, and that the noises are Gaussian zero-mean:

$$x_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$$

$$q_k \sim \mathcal{N}(0, \Sigma_q)$$

$$r_k \sim \mathcal{N}(0, \Sigma_r)$$

we also assume that the noises are not temporally correlated, this means $E[q_k q_j^T] = \Sigma_q \delta_{kj}$ and $E[r_k r_j^T] = \Sigma_r \delta_{kj}$, and that the noises and the state are mutually statistically independent, $E[x_k r_j^T] = 0 \quad \forall k, j$ and $E[x_k q_j^T] = 0 \quad \forall j > k$ (the state is correlated with previous noises, $j < k$ since the noises appears in the state evolution equation). Finally the two noises are required to be mutually independent: $E]r_k q_j^T] = 0 \quad \forall k, j$

propagation of state error

Estimating the state x_k is optimally performed by iterating between uncertainty propagation and measurement updates. The mean and covariance after propagation from step $k-1$ to step k are denoted by $(x_{k|k-1}, \Sigma_{k|k-1})$. The mean and covariance after a measurement update at step k are denoted by $(x_{k|k}, \Sigma_{k|k})$. Given the discrete linear, time-varying (LTV) model of x, we define the state prediction as:

$$\hat{x}_{k|k-1} = A_{k-1}\hat{x}_{k-1,k-1} + B_{k-1}u_{k-1}$$

which is equivalent to the mean,

$$E[x_k|x_{k-1}] = A_{k-1}\bar{x}_{k-1} + B_{k-1}u_{k-1} + E[q_k] = A_{k-1}\bar{x}_{k-1} + B_{k-1}u_{k-1}$$

note that the control matrix and vector are, for our purposes considered as constants since they only vary over time but are deterministic at each time-step. This means that u_k is not drawn from a probability distribution, and $E[u_k] = u_k \qquad \forall k$. We have the covariance of this state estimate,

$$\Sigma_{k|k-1} = E[(x_{k|k-1} - \hat{x}_{k|k-1})(x_{k|k-1} - \hat{x}_{k|k-1})^T]$$

$$= E[(A_{k-1}(x_{k-1,k-1} - \hat{x}_{k-1,k-1}) + q_{k-1})(A_{k-1}(x_{k-1,k-1} - \hat{x}_{k-1,k-1}) + q_{k-1})^T]$$

$$= E[A_{k-1}(x_{k-1,k-1} - \hat{x}_{k-1,k-1})(x_{k-1,k-1} - \hat{x}_{k-1,k-1})A_{k-1}^T + A_{k-1}(x_{k-1,k-1} - \hat{x}_{k-1,k-1})q_{k-1}$$

$$+ q_{k-1}(x_{k-1,k-1} - \hat{x}_{k-1,k-1})^T A_{k-1}^T + q_{k-1}q_{k-1}^T]$$

$x_{k-1} - \bar{x}_{k-1}$ and q_{k-1} i.e. state error and state are uncorrelated implies

$$E[(x_{k-1,k-1} - \hat{x}_{k-1,k-1})q_{k-1}^T] = E(x_{k-1,k-1} - \hat{x}_{k-1,k-1})E(q_{k-1}^T) = 0$$

also

$$E[q_{k-1}(x_{k-1,k-1} - \hat{x}_{k-1,k-1})^T] = 0$$

two of the summands in the equation for $\Sigma_{k|k-1}$ vanish. The recursive covariance equation reduces to

$$\Sigma_{k|k-1} = A_{k-1}\Sigma_{k-1,k-1}A_{k-1}^T + \Sigma_q$$

which is actually how the covariance of a variable resulting from the affine transformation (from the model's linearity assumption) of x, i.e. $y = Ax + b$, will propagate over time.

finding the optimum x

we are given measurements y, whose error is Σ_r. How can we get the best estimate $\hat{x}$ which minimizes the errors weighted by the accuracy (inverse of Σ_r, Σ_r^{-1})?. But we are also given a priori estimate of the state, $\hat{x}_{k|k-1}$ with covariance $\Sigma_{k|k-1}$. This can be achieved by minimizing the quadratic cost function

$$J(\hat{x}) = \frac{1}{2}\Big((y_k - H\hat{x}_k)^T\Sigma_r^{-1}(y_k - H\hat{x}_k) + (\hat{x}_{k|k-1} - \hat{x}_k)^T\Sigma_{k|k-1}^{-1}(\hat{x}_{k|k-1} - \hat{x}_k)\Big)$$

The idea is to use the old information together with this newly observed information to arrive at the best estimate of x. Intuitively a small standard deviation means bigger error, which means it will more importance in the cost function.

$$\begin{aligned}
\frac{\partial J}{\partial \hat{x}} &= -H^T\Sigma_r^{-1}(y_k - H\hat{x}_k) - \Sigma_{k|k-1}^{-1}(\hat{x}_{k|k-1} - \hat{x}_k) = 0 \\
-H^T\Sigma_r^{-1}y_k + H^T\Sigma_r^{-1}H\hat{x}_k &- \Sigma_{k|k-1}^{-1}\hat{x}_{k|k-1} + \Sigma_{k|k-1}^{-1}\hat{x}_k = 0 \\
(H^T\Sigma_r^{-1}H + \Sigma_{k|k-1}^{-1})\hat{x}_k &= H^T\Sigma_r^{-1}y_k + \Sigma_{k|k-1}^{-1}\hat{x}_{k|k-1} \\
\hat{x}_k &= (H^T\Sigma_r^{-1}H + \Sigma_{k|k-1}^{-1})^{-1}(H^T\Sigma_r^{-1}y_k + \Sigma_{k|k-1}^{-1}\hat{x}_{k|k-1})
\end{aligned}$$

Woodbury matrix identity:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

utilizing this identity,

$$(H^T\Sigma_r^{-1}H + \Sigma_{k|k-1}^{-1})^{-1} = \Sigma_{k|k-1} - \Sigma_{k|k-1}H^T(\Sigma_r + H\Sigma_{k|k-1}H^T)^{-1}H\Sigma_{k|k-1}$$

$$\begin{aligned}
\hat{x}_k =& \Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k + \hat{x}_{k|k-1} - \Sigma_{k|k-1}H^T(\Sigma_r + H\Sigma_{k|k-1}H^T)^{-1}H\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k \\
&- \Sigma_{k|k-1}H^T(\Sigma_r + H\Sigma_{k|k-1}H^T)^{-1}H\hat{x}_{k|k-1} \\
=& \hat{x}_{k|k-1} + (\mathbb{I} - \Sigma_{k|k-1}H^T(\Sigma_r + H\Sigma_{k|k-1}H^T)^{-1}H)\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k \\
&- \Sigma_{k|k-1}H^T(\Sigma_r + H\Sigma_{k|k-1}H^T)^{-1}H\hat{x}_{k|k-1}
\end{aligned}$$

define K_k,

$$K_k = \Sigma_{k|k-1}H^T(\Sigma_r + H\Sigma_{k|k-1}H^T)^{-1}$$

$$\hat{x}_k = \hat{x}_{k|k-1} + (I - KH)\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k - KH\hat{x}_{k|k-1} \tag{1}$$

analyse the second term,

$$(I - KH)\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k = \Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k - KH\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k \tag{2}$$

from the definition of K,

$$K(\Sigma_r + H\Sigma_{k|k-1}H^T) = \Sigma_{k|k-1}H^T \tag{3}$$

substituting this to (2),

$$\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k - KH\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k = K(\Sigma_r + H\Sigma_{k|k-1}H^T)\Sigma_r^{-1}y_k - KH\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k$$
$$= K\Sigma_r\Sigma_r^{-1}y_k + KH\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k - KH\Sigma_{k|k-1}H^T\Sigma_r^{-1}y_k = Ky_k$$

Hence, substituting back the second term to (1),

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k y_k - K_k H\hat{x}_{k|k-1} \\ &= \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1})\end{aligned}$$

where $K_k \in \mathbb{R}^{n\times m}$ denotes the Kalman Gain.

covariance update

posteriori error,

$$\begin{aligned}x_{k|k-1} - \hat{x}_{k|k} &= x_{k|k-1} - (\hat{x}_{k|k-1} + Ky_k - KH\hat{x}_{k|k-1}) \\ &= x_{k|k-1} - (\hat{x}_{k|k-1} + K(Hx_{k|k-1} + r_k) - KH\hat{x}_{k|k-1}) \\ &= x_{k|k-1} - \hat{x}_{k|k-1} - KHx_{k|k-1} - Kr_k + KH\hat{x}_{k|k-1} \\ &= (\mathbb{I} - KH)(x_{k|k-1} - \hat{x}_{k|k-1}) - Kr_k\end{aligned}$$

$$\Sigma_{k|k} = E[((\mathbb{I}-KH)(x_{k|k-1}-\hat{x}_{k|k-1})-Kr_k)((\mathbb{I}-KH)(x_{k|k-1}-\hat{x}_{k|k-1})-Kr_k)^T]$$
$$= E[(\mathbb{I}-KH)(x_{k|k-1}-\hat{x}_{k|k-1})(x_{k|k-1}-\hat{x}_{k|k-1})^T(\mathbb{I}-KH)^T - (\mathbb{I}-KH)(x_{k|k-1}-\hat{x}_{k|k-1})r_k^TK^T$$
$$-Kr_k(x_{k|k-1} - \hat{x}_{k|k-1})^T(1 - KH)^T + Kr_kr_k^TK^T]$$
$$= (\mathbb{I} - KH)\Sigma_{k|k-1}(I - KH)^T + K\Sigma_rK^T$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1}H^TK^T - KH\Sigma_{k|k-1} + KH\Sigma_{k|k-1}H^TK^T + K\Sigma_rK^T$$

substituting (3),

$$\Sigma_{k|k} =$$
$$\Sigma_{k|k-1} - K(\Sigma_r + H\Sigma_{k|k-1}H^T)K^T - KH\Sigma_{k|k-1} + KH\Sigma_{k|k-1}H^TK^T + K\Sigma_rK^T$$
$$= \Sigma_{k|k-1} - K\Sigma_rK^T - KH\Sigma_{k|k-1}H^TK^T - KH\Sigma_{k|k-1} + KH\Sigma_{k|k-1}H^TK^T + K\Sigma_rK^T$$
$$= \Sigma_{k|k-1} - KH\Sigma_{k|k-1} = (\mathbb{I} - KH)\Sigma_{k|k-1}$$

kalman filter algorithm

$$\begin{aligned}
\hat{x}_{k|k-1} &= A_{k-1}\hat{x}_{k-1,k-1} + B_{k-1}u_{k-1} && \text{state prediction} \\
\Sigma_{k|k-1} &= A_{k-1}\Sigma_{k-1,k-1}A_{k-1}^T + \Sigma_q && \text{prediction covariance} \\
K_k &= \Sigma_{k|k-1}H^T(\Sigma_r + H\Sigma_{k|k-1}H^T)^{-1} && \text{calculate Kalman Gain} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1}) && \text{state correction/update} \\
\Sigma_{k|k} &= (\mathbb{I} - KH)\Sigma_{k|k-1} && \text{update covariance matrix}
\end{aligned}$$

references:

1. Michael Van Biezen, youtube.

2. Marin Kobilarov, Applied Optimal Control EN530.603, Lecture 10: Optimal State Estimation, 2021.

3. G.S. Chirikjian, A.B. Kyatkin, Engineering Applications of Noncommutative Harmonic Analysis, CRC Press, Boca Raton, FL, 2001.

4. Christopher D'souza, Fundamentals of Kalman Filtering and Estimation in Aerospace Engineering, 2013.

5. Annica Andersson, Estimation of the compliance of the arterial walls through Kalman filtering, 2011.

6. G.L. Plett, Applied Kalman Filtering ECE5550, Lecture Notes, 2018.