Kalman Filter

A rigorous but painless introduction for Roboticists

shuqi

September 25, 2023

xiaosq2000 xiaosq2000@gmail.com

TL; DR

Kalman filter is an optimal, recursive, linear-quardratic estimator[1] for dynamic systems under certain assumptions.

> *In summary, the following assumptions are made about random processes: Physical random phenomena may be thought of as due to primary random sources exciting dynamic systems. The primary sources are assumed to be independent Gaussian random processes with zero mean; the dynamic systems will be linear.*[1]
>
> — *Rudolf E. Kálmán*

[1] alias LQE

Prerequistes

Undergrate-level

- Probability and Stochasitic Process
- Linear Algebra
- Matrix Caculus

Outline

1. Underlying dynamic system model
2. Linear quardratic estimation
3. Derivation of the optimal Kalman gain
4. Summary
5. Variants
6. Example application

Underlying dynamic system model

Underlying dynamic system model i

A classic state-space representation of dynamic system[2].

$\in \mathbb{R}^n$, state vector

$\in \mathcal{M}(n,p)$, input/control matrix

$\in \mathcal{M}(n,n)$, system matrix

$\in \mathbb{R}^p$, input/control vector

$\in \mathbb{R}$, time

$$
\begin{cases}
\dot{\mathrm{x}}(t) = \mathrm{A}(t)\mathrm{x}(t) + \mathrm{B}(t)\mathrm{u}(t) & (1)\\
\mathrm{y}(t) = \mathrm{C}(t)\mathrm{x}(t) + \mathrm{D}(t)\mathrm{u}(t) & (2)
\end{cases}
$$

$\in \mathcal{M}(q,p)$, feedforward matrix

$\in \mathbb{R}^q$, output/measurement vector

$\in \mathcal{M}(q,n)$, output/measurement matrix

Figure 1: continuous-time, deterministic, time-variant, linear dynamic system

Underlying dynamic system model ii

Definition (the underlying model of Kalman filter)

A discrete-time, linear dynamic system[2], with additive Gaussian white noise[3].

$$\begin{cases} x_k = A_{\overbrace{k}^{\in \mathbb{N},\ \text{index/timestamp}}} x_{k-1} + B_k u_k + \overbrace{v_k}^{\sim \mathcal{N}(0, Q_k),\ \text{process noise}} & (3)\\ y_k = C_k x_k + \underbrace{w_k}_{\sim \mathcal{N}(0, R_k),\ \text{measurement noise}} & (4) \end{cases}$$

Figure 2: discrete-time, random, time-variant, linear dynamic system

Underlying dynamic system model iii

Corollary

$$\forall\, k \in \mathbb{N}\, (\mathsf{x}_k \sim \mathcal{N}, \mathsf{y}_k \sim \mathcal{N}) \tag{5}$$

$$\begin{cases} \mathrm{E}\left(\mathsf{x}_k\right) = \mathsf{A}_k\, \mathrm{E}\left(\mathsf{x}_{k-1}\right) + \mathsf{B}_k \mathsf{u}_k & (6) \\ \mathrm{Cov}\left(\mathsf{x}_k\right) = \mathsf{A}_k\, \mathrm{Cov}\left(\mathsf{x}_{k-1}\right) \mathsf{A}_k^{\mathrm{T}} + \mathsf{Q}_k & (7) \end{cases}$$

$$\begin{cases} \mathrm{E}\left(\mathsf{y}_k\right) = \mathsf{C}_k\, \mathrm{E}\left(\mathsf{x}_k\right) & (8) \\ \mathrm{Cov}\left(\mathsf{y}_k\right) = \mathsf{C}_k\, \mathrm{Cov}\left(\mathsf{x}_k\right) \mathsf{C}_k^{\mathrm{T}} + \mathsf{R}_k & (9) \end{cases}$$

[2] the feedforward part is omitted for convenience

Linear quardratic estimation

Problem formulation

Given the state-space dynamic model and

$\hat{\mathrm{x}}_0$	prior estimated distribution of the initial state
$\mathrm{Q}_0, \cdots, \mathrm{Q}_{k-1}$	prior knowledge of the process noise
$\mathrm{R}_0, \cdots, \mathrm{R}_{k-1}$	prior knowledge of the measurement noise
$\mathrm{u}_0, \cdots, \mathrm{u}_{k-1}$	control signals (deterministic) up to now
$\mathrm{y}_0, \cdots, \mathrm{y}_{k-1}$	measurements up to now

to estimate x_k and its uncertainties in a linear and recursive method.

Predict & Correct i

The Kalman filter is usually conceptualized as two distinct phases, of which names are enough enlightening.

Definition (state estimator in the Kalman filter)

$$\begin{cases} \overbrace{\hat{\mathbf{x}}_{k|k-1}}^{\text{estimate of } \mathbf{x} \text{ at } k,\ \text{given } k-1 \text{ measurements}} = \mathrm{A}_k \hat{\mathbf{x}}_{k-1|k-1} + \mathrm{B}_k \mathbf{u}_k & \text{predict} \qquad (10) \\ \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \underbrace{\mathrm{K}_k}_{\text{Kalman gain}} \left(\underbrace{\mathbf{y}_k - \mathrm{C}_k \hat{\mathbf{x}}_{k|k-1}}_{=:\ \tilde{y}_{k|k-1},\ \text{innovation/measurement pre-fit residual}} \right) & \text{correct/update} \qquad (11) \end{cases}$$

Figure 3: equations to update state estimates

Predict & Correct ii

Remark (motivation of prediction)

It is straightforward and reasonable that the predict-phase state update (eq. 11) is derived from the state transition (eq. 4), by replacing the expectation, $\mathrm{E}(\square)$ with the estimate, $\hat{\square}$.

Remark (motivation of correction)

Intuitively, the Kalman gain is like a weight coefficient to emphasize or weaken the information from real measurements, and the weighted difference between prediction and measurements (measurement pre-fit residual) is added to correct the prediction.

The Kalman gain certainly cannot be chosen arbitrarily. How? Stay tuned.

Corollary (ideal invariants)

Our prior knowledge is assumed to be 100% accurate, not only the model, but also the initial estimate ($\hat{\mathbf{x}}_0 = \mathbf{x}_0$), then there are some invariants during the whole process, e.g., $\forall\, k \in \mathbb{N}$,

$$\mathrm{E}\left(\tilde{\mathbf{x}}_{k|k-1}\right) = \mathbf{0} \tag{12}$$

$$\mathrm{E}\left(\tilde{\mathbf{y}}_{k|k-1}\right) = \mathbf{0} \tag{13}$$

$$\mathrm{E}\left(\tilde{\mathbf{x}}_{k|k}\right) = \mathbf{0} \tag{14}$$

Corollary (prior error covariance update)

$$\underbrace{\tilde{P}_{k|k-1}}_{:= \mathrm{Cov}\left(\hat{x}_{k|k-1} - x_k\right),\ \textit{prior error covariance}} = A_k \overbrace{\tilde{P}_{k-1|k-1}}^{\mathrm{Cov}\left(\hat{x}_{k-1|k-1} - x_{k-1}\right),\ \textit{last posterior error covariance}} A_k^{\mathrm{T}} + R_k \tag{15}$$

Proof.

$$\hat{x}_{k|k-1} - x_k = A_k \hat{x}_{k-1|k-1} + B_k u_k - (A_k x_{k-1} + B_k u_k + v_k) \tag{16}$$

$$\mathrm{Cov}\left(\hat{x}_{k|k-1} - x_k\right) = A_k \,\mathrm{Cov}\left(\hat{x}_{k-1|k-1} - x_{k-1}\right) A_k^{\mathrm{T}} + \mathrm{Cov}\left(v_k\right) \tag{17}$$

$$\tilde{P}_{k|k-1} = A_k \tilde{P}_{k-1|k-1} A_k^{\mathrm{T}} + R_k \tag{18}$$

$\square$

Predict & Correct v

Corollary (innovation covariance)

$$\underbrace{\boxed{S_{k|k-1}}}_{:= \operatorname{Cov}\left(\tilde{y}_{k|k-1}\right)} = C_k \boxed{\tilde{P}_{k|k-1}} C_k^{\mathrm{T}} + R_k \tag{19}$$

Proof.

$$\begin{aligned}
\operatorname{Cov}\left(\tilde{y}_{k|k-1}\right) &= \operatorname{Cov}\left(C_k x_k + w_k - C_k \hat{x}_{k|k-1}\right) && (20)\\
&= C_k \operatorname{Cov}\left(x_k - \hat{x}_{k|k-1}\right) C_k^{\mathrm{T}} + \operatorname{Cov}\left(w_k\right) && (21)\\
&= C_k \tilde{P}_{k|k-1} C_k^{\mathrm{T}} + R_k && (22)
\end{aligned}$$

$\square$

Derivation of the optimal Kalman gain

Derivation of the optimal Kalman gain i

As mentioned above, we've already designed an estimator with a well-formed structure, except that the parameter $\mathbf{K}_k$ hasn't been decided yet.

Definition (the optimal Kalman gain)

MMSE (minimum mean square error) is chosen as our rule for optimality, i.e.,

$$\mathbf{K}_k := \underset{\mathbf{K}_k}{\operatorname{argmin}}\, \mathrm{E}\left(\| \underbrace{\boxed{\tilde{\mathbf{x}}_k}}_{=:\, \hat{\mathbf{x}}_k - \mathbf{x}_k,\ \text{error}} \|_2 \right) \tag{23}$$

Figure 4: minimum expectation of L2-norm of the error random variable from a Bayesian approach

Derivation of the optimal Kalman gain ii

Some little tricks here,

$$\operatorname*{argmin}_{K_k} E\left(\|\tilde{x}_k\|_2\right) = \operatorname*{argmin}_{K_k} E\left(\|\tilde{x}_k\|_2^2\right) \tag{24}$$

$$\begin{aligned} E\left(\|\tilde{x}_k\|_2^2\right) &= E\left(\tilde{x}_k^T \tilde{x}_k\right) && (25)\\ &= E\left(\mathrm{Tr}\left(\tilde{x}_k^T \tilde{x}_k\right)\right) && (26)\\ &= E\left(\mathrm{Tr}\left(\tilde{x}_k \tilde{x}_k^T\right)\right) && (27)\\ &= \mathrm{Tr}\left(E\left(\tilde{x}_k \tilde{x}_k^T\right)\right) && (28) \end{aligned}$$

Since eq. 14,

$:= \tilde{P}_k$ or $\tilde{P}_{k|k}$, posterior error covariance

$$E\left(\tilde{x}_k \tilde{x}_k^T\right) = \boxed{\mathrm{Cov}\left(\tilde{x}_k\right)} \tag{29}$$

The optimal Kalman gain becomes the solution of

$$\frac{\partial}{\partial \mathrm{K}_k} \mathrm{Tr}\left(\tilde{\mathrm{P}}_k\right) = 0 \tag{30}$$

Hold on! I know you can't wait to roll up your sleeves, expand out the terms and play with the messy matrix calculus, but the following Joseph form of the posterior error covariance will save you a lot of trouble.

$$
\begin{aligned}
\tilde{P}_k &= \mathrm{Cov}\left(\hat{x}_k - x_k\right) && (31)\\
&= \mathrm{Cov}\left(\hat{x}_{k|k-1} + K_k\left(y_k - C_k\hat{x}_{k|k-1}\right) - x_k\right) && (32)\\
&= \mathrm{Cov}\left(\hat{x}_{k|k-1} + K_k\left(C_k x_k + w_k - C_k\hat{x}_{k|k-1}\right) - x_k\right) && (33)\\
&= \mathrm{Cov}\left(\left(I - K_kC_k\right)\hat{x}_{k|k-1} - \left(I - K_kC_k\right)x_k + K_k w_k\right) && (34)\\
&= \mathrm{Cov}\left(\left(I - K_kC_k\right)\hat{x}_{k|k-1} - \left(I - K_kC_k\right)x_k\right) + \mathrm{Cov}\left(K_k w_k\right) && (35)\\
&= \boxed{\left(I - K_kC_k\right)\tilde{P}_{k|k-1}\left(I - K_kC_k\right)^{T} + K_kR_kK_k^{T}} && (36)\\
&= \tilde{P}_{k|k-1} - K_kC_k\tilde{P}_{k|k-1} - \tilde{P}_{k|k-1}C_k^{T}K_k^{T} + K_kC_k\tilde{P}_{k|k-1}C_k^{T}K_k^{T} + K_kR_kK_k^{T} && (37)\\
&= \tilde{P}_{k|k-1} - K_kC_k\tilde{P}_{k|k-1} - \tilde{P}_{k|k-1}C_k^{T}K_k^{T} + K_kS_{k|k-1}K_k^{T} && (38)
\end{aligned}
$$

Ready to go,

$$\frac{\partial \operatorname{Tr}\left(\tilde{P}_k\right)}{\partial K_k} = -2\tilde{P}_{k|k-1}C_k^{\mathrm{T}} + 2S_{k|k-1}K_k = \mathbf{0} \tag{39}$$

Finally,

$$\boxed{K_k = \tilde{P}_{k|k-1}C_k^{\mathrm{T}}S_{k|k-1}^{-1}} \tag{40}$$

Summary

Procedure

1. Model the random dynamics of the system
   $A_k, B_k, C_k, Q_k, R_k \quad (\forall k \in \mathbb{N})$

2. Estimate the initial state and error covariance
   $\hat{x}_{0|0}, \tilde{P}_{0|0}$

3. Predict
   $\hat{x}_{k|k-1} = A_k \hat{x}_{k-1|k-1} + B_k u_k, \quad \tilde{P}_{k|k-1} = A_k P_{k-1|k-1} A_k^{\mathrm{T}} + R_k$

4. Caculate the Kalman gain
   $S_k = C_k \tilde{P}_{k|k-1} C_k^{\mathrm{T}} + R_k, \quad K_k = \hat{P}_{k|k-1} C_k^{\mathrm{T}} S_k^{-1}$

5. Correct
   $\hat{x}_{k|k} = (I - K_k C_k)\, \hat{x}_{k|k-1} + K_k y_k, \quad \tilde{P}_{k|k} = (I - K_k C_k)\, \tilde{P}_{k|k-1} \,(I - K_k C_k)^{\mathrm{T}} + K_k R_k K_k^{\mathrm{T}}$

Variants

Variants

Kalman-Bucy Filter

Definition (Underlying dynamic system model of Kalman-Bucy Filter)

$\in \mathbb{R}^n$, state vector

$\in \mathcal{M}(n, p)$, control matrix

$\in \mathcal{M}(n, n)$, system matrix

$\in \mathbb{R}^p$, control vector

$\in \mathbb{R}$, time

$$\begin{cases} \dot{\mathrm{x}}(t) = \mathrm{A}(t)\mathrm{x}(t) + \mathrm{B}(t)\mathrm{u}(t) + \mathrm{v}(t) & (41) \\ \mathrm{y}(t) = \mathrm{C}(t)\mathrm{x}(t) + \mathrm{w}(t) & (42) \end{cases}$$

WGN[3] with PSD[4]=$\mathbf{Q}(t)$, process noise

$\in \mathbb{R}^q$, measurement vector

WGN with PSD=$\mathbf{R}(t)$, measurement noise

$\in \mathcal{M}(q, n)$, measurement matrix

Underlying dynamic system model ii

Additional information on white noise

Take the process noise for example,

$$\mathrm{E}\left(\mathbf{v}(t)\right) = \mathbf{0} \tag{43}$$

$$\mathrm{E}\left(\mathbf{v}(t+\tau)\mathbf{v}^{\mathrm{T}}(t)\right) = \mathbf{Q}(t) = \mathbf{Q}\delta(\tau) \tag{44}$$

[3] white Gaussian noise
[4] power spectral density

Example application

Appendix

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1, 1960, ISSN: 0021-9223. DOI: 10.1115/1.3662552. [Online]. Available: https://asmedigitalcollection.asme.org/fluidsengineering/article/82/1/35/397706/A-New-Approach-to-Linear-Filtering-and-Prediction.

[2] Wikipedia contributors, *State-space representation — Wikipedia, the free encyclopedia*, https://en.wikipedia.org/w/index.php?title=State-space_representation&oldid=1136534469, [Online; accessed 26-February-2023], 2023.

[3] Wikipedia contributors, *Additive white gaussian noise — Wikipedia, the free encyclopedia*, https://en.wikipedia.org/w/index.php?title=Additive_white_Gaussian_noise&oldid=1164245413, [Online; accessed 23-September-2023], 2023.