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 estimator¹ 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

¹alias 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].

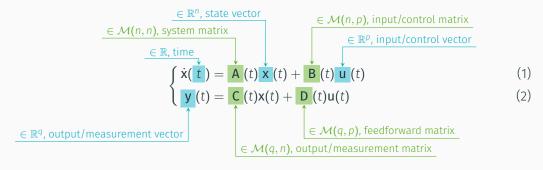


Figure 1: continuous-time, deterministic, time-variant, linear dynamic system

Underlying dynamic system model ii

Definition (the underlying model of Kalman filter)

A discrete-time, linear dynamic system², with additive Gaussian white noise^[3].

$$\begin{cases} x_{k} = A_{k} x_{k-1} + B_{k} u_{k} + v_{k} \\ y_{k} = C_{k} x_{k} + w_{k} \\ & &$$

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

Underlying dynamic system model iii

Corollary

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

$$\begin{cases} E(x_{k}) = A_{k} E(x_{k-1}) + B_{k} u_{k} & (6) \\ Cov(x_{k}) = A_{k} Cov(x_{k-1}) A_{k}^{T} + Q_{k} & (7) \end{cases} \qquad \begin{cases} E(y_{k}) = C_{k} E(x_{k}) & (8) \\ Cov(y_{k}) = C_{k} Cov(x_{k}) C_{k}^{T} + R_{k} & (9) \end{cases}$$

²the feedforward part is omitted for convenience

Linear quardratic estimation

Problem formulation

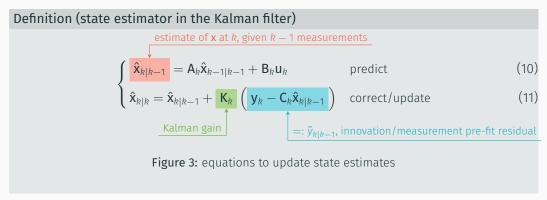
Given the state-space dynamic model and

$\hat{\mathbf{x}}_0$	prior estimated distribution of the initial state
Q_0, \cdots, Q_{k-1}	prior knowledge of the process noise
R_0, \cdots, R_{k-1}	prior knowledge of the measurement noise
$\mathbf{u}_0,\cdots,\mathbf{u}_{k-1}$	control signals (deterministic) up to now
$\mathbf{y}_0, \cdots, \mathbf{y}_{k-1}$	measurements up to now

to estimate \mathbf{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.



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, $E(\Box)$ with the estimate, $\hat{\Box}$.

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.

Predict & Correct iii

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., $\forall k \in \mathbb{N}$,

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

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

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

Predict & Correct iv

Corollary (prior error covariance update)

Cov
$$(\hat{\mathbf{x}}_{k-1|k-1} - \mathbf{x}_{k-1})$$
, last posterior error covariance
$$\tilde{\mathbf{P}}_{k|k-1} = \mathbf{A}_k \tilde{\mathbf{P}}_{k-1|k-1} \mathbf{A}_k^{\mathrm{T}} + \mathbf{R}_k$$

$$\hat{\mathbf{c}} := \mathsf{Cov} (\hat{\mathbf{x}}_{k|k-1} - \mathbf{x}_k), \textit{prior error covariance}$$
(15)

Proof.

$$\hat{\mathbf{x}}_{k|k-1} - \mathbf{x}_k = \mathbf{A}_k \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k \mathbf{u}_k - (\mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{v}_k)$$
(16)

$$Cov (\hat{\mathbf{x}}_{k|k-1} - \mathbf{x}_k) = \mathbf{A}_k Cov (\hat{\mathbf{x}}_{k-1|k-1} - \mathbf{x}_{k-1}) \mathbf{A}_k^{\mathrm{T}} + Cov (\mathbf{v}_k)$$
(17)

$$\tilde{\mathbf{P}}_{k|k-1} = \mathbf{A}_k \tilde{\mathbf{P}}_{k-1|k-1} \mathbf{A}_k^{\mathrm{T}} + \mathbf{R}_k \tag{18}$$



Predict & Correct v

Corollary (innovation covariance)

$$\mathbf{S}_{k|k-1} = \mathbf{C}_k \tilde{\mathbf{P}}_{k|k-1} \mathbf{C}_k^{\mathrm{T}} + \mathbf{R}_k$$

$$:= \operatorname{Cov} (\tilde{\mathbf{y}}_{k|k-1})$$
(19)

Proof.

$$Cov (\tilde{\mathbf{y}}_{k|k-1}) = Cov (C_k \mathbf{x}_k + \mathbf{w}_k - C_k \hat{\mathbf{x}}_{k|k-1})$$

$$= C_k Cov (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) C_k^T + Cov (\mathbf{w}_k)$$

$$= C_k \tilde{\mathbf{P}}_{k|k-1} C_k^T + \mathbf{R}_k$$
(20)
$$(21)$$

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 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.,

$$\mathbf{K}_{k} := \underset{\mathbf{K}_{k}}{\operatorname{argmin}} \mathbf{E} \left(\| \tilde{\mathbf{X}}_{k} \|_{2} \right) \\
\underline{=: \hat{\mathbf{X}}_{k} - \mathbf{X}_{k}, \operatorname{error}} \tag{23}$$

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,

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

$$\mathsf{E}\left(\|\tilde{\mathbf{x}}_{k}\|_{2}^{2}\right) = \mathsf{E}\left(\tilde{\mathbf{x}}_{k}^{\mathrm{T}}\tilde{\mathbf{x}}_{k}\right) \tag{25}$$

$$= \mathsf{E}\left(\mathsf{Tr}\left(\tilde{\mathsf{x}}_{\mathsf{R}}^{\mathsf{T}}\tilde{\mathsf{x}}_{\mathsf{R}}\right)\right) \tag{26}$$

$$= \mathsf{E}\left(\mathsf{Tr}\left(\tilde{\mathsf{X}}_{k}\tilde{\mathsf{X}}_{k}^{\mathsf{T}}\right)\right) \tag{27}$$

$$= \operatorname{Tr}\left(\mathsf{E}\left(\tilde{\mathsf{x}}_{k}\tilde{\mathsf{x}}_{k}^{\mathrm{T}}\right)\right) \tag{28}$$

Since eq. 14,

$$= \tilde{\mathbf{p}}_{k} \text{ or } \tilde{\mathbf{p}}_{k|k}, \text{ posterior error covariance}$$

$$= \mathbf{E}\left(\tilde{\mathbf{x}}_{k}\tilde{\mathbf{x}}_{k}^{\mathrm{T}}\right) = \mathbf{Cov}\left(\tilde{\mathbf{x}}_{k}\right)$$

$$(29)$$

Derivation of the optimal Kalman gain iii

The optimal Kalman gain becomes the solution of

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

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.

Derivation of the optimal Kalman gain iv

$$\tilde{P}_{k} = \text{Cov} (\hat{x}_{k} - x_{k}) \tag{31}$$

$$= \text{Cov} (\hat{x}_{k|k-1} + K_{k} (y_{k} - C_{k} \hat{x}_{k|k-1}) - x_{k}) \tag{32}$$

$$= \text{Cov} (\hat{x}_{k|k-1} + K_{k} (C_{k} x_{k} + w_{k} - C_{k} \hat{x}_{k|k-1}) - x_{k}) \tag{33}$$

$$= \text{Cov} ((I - K_{k} C_{k}) \hat{x}_{k|k-1} - (I - K_{k} C_{k}) x_{k} + K_{k} w_{k}) \tag{34}$$

$$= \text{Cov} ((I - K_{k} C_{k}) \hat{x}_{k|k-1} - (I - K_{k} C_{k}) x_{k}) + \text{Cov} (K_{k} w_{k}) \tag{35}$$

$$= (I - K_{k} C_{k}) \tilde{P}_{k|k-1} (I - K_{k} C_{k})^{T} + K_{k} R_{k} K_{k}^{T} \tag{36}$$

$$= \tilde{P}_{k|k-1} - K_{k} C_{k} \tilde{P}_{k|k-1} - \tilde{P}_{k|k-1} C_{k}^{T} K_{k}^{T} + K_{k} C_{k} \tilde{P}_{k|k-1} C_{k}^{T} K_{k}^{T} + K_{k} R_{k} K_{k}^{T} \tag{37}$$

$$= \tilde{P}_{k|k-1} - K_{k} C_{k} \tilde{P}_{k|k-1} - \tilde{P}_{k|k-1} C_{k}^{T} K_{k}^{T} + K_{k} S_{k|k-1} K_{k}^{T} \tag{38}$$

Derivation of the optimal Kalman gain v

Ready to go,

$$\frac{\partial \operatorname{Tr}\left(\tilde{\mathsf{P}}_{k}\right)}{\partial \mathsf{K}_{k}} = -2\tilde{\mathsf{P}}_{k|k-1}\mathsf{C}_{k}^{\mathrm{T}} + 2\mathsf{S}_{k|k-1}\mathsf{K}_{k} = \mathbf{0}$$
(39)

Finally,

$$\mathbf{K}_{k} = \tilde{\mathbf{P}}_{k|k-1} \mathbf{C}_{k}^{\mathsf{T}} \mathbf{S}_{k|k-1}^{-1} \tag{40}$$



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 $\hat{x}_{\text{010}}, \tilde{P}_{\text{010}}$
- 3. Predict

$$\mathbf{\hat{x}}_{k|k-1} = \mathbf{A}_k \mathbf{\hat{x}}_{k-1|k-1} + \mathbf{B}_k \mathbf{u}_k, \quad \tilde{\mathbf{P}}_{k|k-1} = \mathbf{A}_k \mathbf{P}_{k-1|k-1} \mathbf{A}_k^T + \mathbf{R}_k$$

4. Caculate the Kalman gain

$$S_k = C_k \tilde{P}_{k|k-1} C_k^T + R_k, \quad K_k = \hat{P}_{k|k-1} C_k^T S_k^{-1}$$

5. Correct

$$\hat{\mathbf{x}}_{k|k} = \left(\mathbf{I} - \mathbf{K}_k \mathbf{C}_k\right) \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \mathbf{y}_k, \quad \tilde{\mathbf{P}}_{k|k} = \left(\mathbf{I} - \mathbf{K}_k \mathbf{C}_k\right) \tilde{\mathbf{P}}_{k|k-1} \left(\mathbf{I} - \mathbf{K}_k \mathbf{C}_k\right)^{\mathrm{T}} + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^{\mathrm{T}}$$

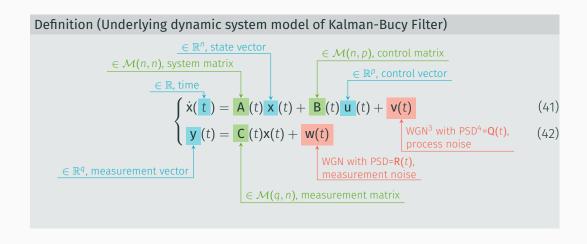


Variants

Variants

Kalman-Bucy Filter

Underlying dynamic system model i



Underlying dynamic system model ii

Additional information on white noise

Take the process noise for example,

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

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

³white Gaussian noise

⁴power spectral density

Example application

Appendix

References i

- [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.