Kalman Filter

shuqi

September 21, 2023

xiaosq2000 xiaosq2000@gmail.com

Outline

1. State-space representation of dynamic systems
2. Linear quardratic estimation
3. Derivation of the optimal Kalman gain
4. Example application

TL; DR

Kalman filter is a recursive, linear-quardratic estimator[1] for a dynamic system, which is optimal 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] thus, alias LQE

State-space representation of dynamic systems

State-space representation[2] of dynamic systems i

The classic one,

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

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

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

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

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

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

State-space representation[2] of dynamic systems ii

The underlying dynamic model of Kalman filter is a time-discreted one with additive Gaussian white noises[2].

$\in \mathbb{N}$, index/timestamp

$\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$, process noise

$$
\begin{cases}
\mathbf{x}_k = \mathbf{A}_{\boxed{k}} \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \boxed{\mathbf{v}_k} & (3) \\
\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \boxed{\mathbf{w}_k} & (4)
\end{cases}
$$

$\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$, measurement noise

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

State-space representation[2] of dynamic systems iii

It's a huge transition that the states and measurements are Gaussian distributions now.

$$\begin{cases} \mathrm{E}\left(\mathbf{x}_k\right) = \mathbf{A}_k \, \mathrm{E}\left(\mathbf{x}_{k-1}\right) + \mathbf{B}_k \mathbf{u}_k & (5) \\ \mathrm{Cov}\left(\mathbf{x}_k\right) = \mathbf{A}_k \, \mathrm{Cov}\left(\mathbf{x}_{k-1}\right) \mathbf{A}_k^{\mathrm{T}} + \mathbf{Q}_k & (6) \end{cases}$$

$$\begin{cases} \mathrm{E}\left(\mathbf{y}_k\right) = \mathbf{C}_k \, \mathrm{E}\left(\mathbf{x}_k\right) & (7) \\ \mathrm{Cov}\left(\mathbf{y}_k\right) = \mathbf{C}_k \, \mathrm{Cov}\left(\mathbf{x}_k\right) \mathbf{C}_k^{\mathrm{T}} + \mathbf{R}_k & (8) \end{cases}$$

[2] the 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
$\mathbf{Q}_0, \cdots, \mathbf{Q}_{k-1}$	prior knowledge of the process noise
$\mathbf{R}_0, \cdots, \mathbf{R}_{k-1}$	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[3] in a linear and quardratic method.

[3] Attributes about randomness, i.e., expectation and covariance in this Gaussian model. Actually, the state estimate is based on the estimated expectation of the state, so only the covariance is viewed as a measure of uncertainty

Predict & Correct i

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

$$
\begin{cases}
\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_k \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k \mathbf{u}_k & \text{predict} \quad (9) \\
\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \underbrace{\boxed{\mathbf{K}_k}}_{\text{Kalman gain}} \Big(\underbrace{\boxed{\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_{k|k-1}}}_{=:\, \tilde{y}_{k|k-1},\ \text{innovation/measurement pre-fit residual}} \Big) & \text{update/correct} \quad (10)
\end{cases}
$$

Figure 3: the estimated state update equations in the two phases of Kalman filter

Predict & Correct ii

Remark on the Kalman gain

Intuitively, it's 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.

It certainly cannot be chosen arbitrarily. How? Stay tuned.

For the predict-phase,

$$\begin{cases} \hat{\mathrm{E}}\left(\mathbf{x}_{k|k-1}\right) = \mathbf{A}_k \hat{\mathrm{E}}\left(\mathbf{x}_{k-1|k-1}\right) + \mathbf{B}_k \mathbf{u}_k & (11) \\ \underbrace{\boxed{\hat{\mathrm{Cov}}\left(\mathbf{x}_{k|k-1}\right)}}_{=:\ \hat{\mathbf{P}}_{k|k-1},\ \text{priori/predicted covariance}} = \mathbf{A}_k \hat{\mathrm{Cov}}\left(\mathbf{x}_{k-1|k-1}\right) \mathbf{A}_k^{\mathrm{T}} + \mathbf{Q}_k & (12) \end{cases}$$

Figure 4: the priori covariance update equation in the predict-phase

Remark

Be cautious. It's easy to confuse "the estimator of the covariance" and "the covariance of the estimator". Take the predict-phase for example,

$$\begin{cases} \operatorname{E}\left(\hat{\mathbf{x}}_{k|k-1}\right) = \mathbf{A}_k \operatorname{E}\left(\hat{\mathbf{x}}_{k-1|k-1}\right) + \mathbf{B}_k \mathbf{u}_k & (13) \\ \operatorname{Cov}\left(\hat{\mathbf{x}}_{k|k-1}\right) = \mathbf{A}_k \operatorname{Cov}\left(\hat{\mathbf{x}}_{k-1|k-1}\right) \mathbf{A}_k^{\mathrm{T}} & (14) \end{cases}$$

Figure 5: the expectation/covariance of the estimator

Predict & Correct v

For the correct-phase,

$$\boxed{\mathbf{S}_{k|k-1}} := \mathrm{Cov}\left(\tilde{\mathbf{y}}_{k|k-1}\right) = \mathrm{Cov}\left(\mathbf{C}_k\mathbf{x}_k + \mathbf{w}_k - \mathbf{C}_k\hat{\mathbf{x}}_{k|k-1}\right) \tag{15}$$

$$= \mathbf{C}_k \,\mathrm{Cov}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right) \mathbf{C}_k^{\mathrm{T}} + \mathrm{Cov}\left(\mathbf{w}_k\right) \tag{16}$$

$$=: \mathbf{C}_k \underbrace{\boxed{\tilde{\mathbf{P}}_{k|k-1}}}_{\text{priori error covariance}} \mathbf{C}_k^{\mathrm{T}} + \mathbf{R}_k \tag{17}$$

Figure 6: the innovation covariance

Derivation of the optimal Kalman gain

Remark

We are actually designing an estimator with a well-formed structure as mentioned above，but the parameter $\mathbf{K}_k$ hasn't been decided yet.

Derivation of the optimal Kalman gain ii

MMSE (Minimum mean square error) is chosen as our rule for optimality, i.e., the optimal Kalman gain

$$\mathbf{K}_k := \underset{\mathbf{K}_k}{\operatorname{argmin}} \mathrm{E}\left(\| \underbrace{\boxed{\tilde{\mathbf{x}}_k}}_{=: \hat{\mathbf{x}}_k - \mathbf{x}_k, \text{ error}} \|_2 \right) \tag{18}$$

Figure 7: minimum expectation of L2-norm of the error from a Bayesian approach

Derivation of the optimal Kalman gain iii

Some tricks here,

$$\underset{\mathbf{K}_k}{\operatorname{argmin}} \operatorname{E}\left(\|\tilde{\mathbf{x}}_k\|_2\right) = \underset{\mathbf{K}_k}{\operatorname{argmin}} \operatorname{E}\left(\|\tilde{\mathbf{x}}_k\|_2^2\right) \tag{19}$$

$$\begin{aligned}
\operatorname{E}\left(\|\tilde{\mathbf{x}}_k\|_2^2\right) &= \operatorname{E}\left(\tilde{\mathbf{x}}_k^{\mathrm{T}}\tilde{\mathbf{x}}_k\right) && (20)\\
&= \operatorname{E}\left(\operatorname{Tr}\left(\tilde{\mathbf{x}}_k^{\mathrm{T}}\tilde{\mathbf{x}}_k\right)\right) && (21)\\
&= \operatorname{E}\left(\operatorname{Tr}\left(\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^{\mathrm{T}}\right)\right) && (22)\\
&= \operatorname{Tr}\left(\operatorname{E}\left(\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^{\mathrm{T}}\right)\right) && (23)
\end{aligned}$$

Derivation of the optimal Kalman gain iv

And let's assume that our prior knowledge is 100% accurate, the dynamic model, $\hat{\mathbf{x}}_0 = \mathbf{x}_0$ etc., then there are some **invariants**,

$$\begin{cases} \mathrm{E}\left(\tilde{y}_{k|k-1}\right) = \mathbf{0} & (24) \\ \mathrm{E}\left(\tilde{\mathbf{x}}_k\right) = \mathbf{0} & (25) \end{cases}$$

thus,

$$\mathrm{E}\left(\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^{\mathrm{T}}\right) = \underbrace{\boxed{\mathrm{Cov}\left(\tilde{\mathbf{x}}_k\right)}}_{:= \tilde{\mathbf{P}}_k \text{ or } \tilde{\mathbf{P}}_{k|k}\text{, (posterior) error covariance}} \qquad (26)$$

Derivation of the optimal Kalman gain v

The optimal Kalman gain becomes the solution of

$$\frac{\partial}{\partial \mathbf{K}_k} \operatorname{Tr}\left(\tilde{\mathbf{P}}_k\right) = \mathbf{0} \tag{27}$$

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 vi

$$
\begin{aligned}
\tilde{\mathbf{P}}_k &= \operatorname{Cov}\left(\hat{\mathbf{x}}_k - \mathbf{x}_k\right) && (28)\\
&= \operatorname{Cov}\left(\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{C}_k\hat{\mathbf{x}}_{k|k-1}\right) - \mathbf{x}_k\right) && (29)\\
&= \operatorname{Cov}\left(\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\left(\mathbf{C}_k\mathbf{x}_k + \mathbf{w}_k - \mathbf{C}_k\hat{\mathbf{x}}_{k|k-1}\right) - \mathbf{x}_k\right) && (30)\\
&= \operatorname{Cov}\left(\left(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k\right)\hat{\mathbf{x}}_{k|k-1} - \left(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k\right)\mathbf{x}_k + \mathbf{K}_k\mathbf{w}_k\right) && (31)\\
&= \operatorname{Cov}\left(\left(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k\right)\hat{\mathbf{x}}_{k|k-1} - \left(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k\right)\mathbf{x}_k\right) + \operatorname{Cov}\left(\mathbf{K}_k\mathbf{w}_k\right) && (32)\\
&= \boxed{\left(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k\right)\tilde{\mathbf{P}}_{k|k-1}\left(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k\right)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}}} && (33)\\
&= \tilde{\mathbf{P}}_{k|k-1} - \mathbf{K}_k\mathbf{C}_k\tilde{\mathbf{P}}_{k|k-1} - \tilde{\mathbf{P}}_{k|k-1}\mathbf{C}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k\mathbf{C}_k\tilde{\mathbf{P}}_{k|k-1}\mathbf{C}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}} && (34)\\
&= \tilde{\mathbf{P}}_{k|k-1} - \mathbf{K}_k\mathbf{C}_k\tilde{\mathbf{P}}_{k|k-1} - \tilde{\mathbf{P}}_{k|k-1}\mathbf{C}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k\mathbf{S}_{k|k-1}\mathbf{K}_k^{\mathrm{T}} && (35)
\end{aligned}
$$

Derivation of the optimal Kalman gain vii

Ready to go,

$$\frac{\partial \operatorname{Tr}\left(\tilde{\mathbf{P}}_k\right)}{\partial \mathbf{K}_k} = -2\tilde{\mathbf{P}}_{k|k-1}\mathbf{C}_k^{\mathrm{T}} + 2\mathbf{S}_{k|k-1}\mathbf{K}_k = \mathbf{0} \tag{36}$$

Finally,

$$\boxed{\mathbf{K}_k = \tilde{\mathbf{P}}_{k|k-1}\mathbf{C}_k^{\mathrm{T}}\mathbf{S}_{k|k-1}^{-1}} \tag{37}$$

Example application

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.