Kalman Filter

Tsang-Kai Chang
tsangkaichang@ucla.edu

Nov., 2020

Goal

1. to understand the theory behind Kalman filter
2. to be able to use EKF to solve problems

Summary

1. Multivariate Gaussian random vector
2. Kalman filter and its Properties
3. Extended Kalman filter (EKF)

Multivariate Gaussian Random Vector

Multivariate Gaussian

- A random vector $\mathbf{x} \in \mathbb{R}^n$ is Gaussian if its pdf is given by

$$p_{\mathbf{x}}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)\right).$$

- We write $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$.
- Σ is a positive definite matrix.
- $\sqrt{(x-\mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)}$ is known as the Mahalanobis distance.

Multivariate Gaussian - Affine Transformation

Proposition (Affine transformation)

If $\mathbf{x}$ *is multivariate Gaussian with distribution* $N(\mu, \Sigma)$*, then* $A\mathbf{x} + b$ *has distribution* $N(A\mu + b, A\Sigma A^{\mathsf{T}})$*.*

Multivariate Gaussian - Independence

Proposition (Independence)

Given a Gaussian random vector $[\mathbf{x}^\mathsf{T}, \mathbf{y}^\mathsf{T}]^\mathsf{T}$, $\mathbf{x}$ *and* $\mathbf{y}$ *are independent if and only if the covariance matrix is block diagonal.*

Multivariate Gaussian - Addition

Proposition (Addition)

If $\mathbf{x} \sim N(\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$ *and* $\mathbf{y} \sim N(\mu_{\mathbf{y}}, \Sigma_{\mathbf{y}})$ *are two independent multivariate Gaussian random vectors with the same dimension, then* $\mathbf{x} + \mathbf{y} \sim N(\mu_{\mathbf{x}} + \mu_{\mathbf{y}}, \Sigma_{\mathbf{x}} + \Sigma_{\mathbf{y}})$*.*

- We can use the previous two properties to see this.

Multivariate Gaussian - Condition

Proposition (Conditioning)

If $\mathbf{x} \in \mathbb{R}^m$ *and* $\mathbf{y} \in \mathbb{R}^n$ *are jointly Gaussian*

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{y}} \end{bmatrix} \right),$$

then the conditional distribution of $\mathbf{x}$ *given* $\mathbf{y} = y$ *is still Gaussian with distribution*

$$N\left(\mu_{\mathbf{x}} + \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}(y - \mu_{\mathbf{y}}), \Sigma_{\mathbf{x}} - \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}\Sigma_{\mathbf{yx}}\right).$$

Multivariate Gaussian - Condition

- undergraduate probability:

$$p_{\mathbf{x}|\mathbf{y}}(x|y) = \frac{p(x, y; \mu, \Sigma)}{\int_{x \in \mathbb{R}^m} p(x, y; \mu, \Sigma) dx}$$

- I present another interpretation based on conditional expectation (still undergraduate probability).

Multivariate Gaussian - Condition

Proposition (Conditional expectation)

If $\mathbf{x} \in \mathbb{R}^m$ *and* $\mathbf{y} \in \mathbb{R}^n$ *are jointly Gaussian*

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{y}} \end{bmatrix} \right),$$

then

$$\mathsf{E}[\mathbf{x}|\mathbf{y}] = \mu_{\mathbf{x}} + \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} (\mathbf{y} - \mu_{\mathbf{y}}).$$

- $\mathsf{E}[\mathbf{x}|\mathbf{y}]$ is a random vector; moreover, a function of $\mathbf{y}$

Multivariate Gaussian - Condition

- Consider a simpler case

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{y}} \end{bmatrix} \right).$$

- Define $\mathbf{x} = \hat{\mathbf{x}} + \tilde{\mathbf{x}}$ with

$$\begin{aligned} \hat{\mathbf{x}} &= \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} \mathbf{y}, \\ \tilde{\mathbf{x}} &= \mathbf{x} - \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} \mathbf{y}. \end{aligned}$$

- Since

$$\begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} I_m & -\Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} \\ 0 & I_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix},$$

$$\begin{aligned} \text{Cov}\left(\begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{y} \end{bmatrix} \right) &= \begin{bmatrix} I_m & -\Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} \\ 0 & I_n \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{y}} \end{bmatrix} \begin{bmatrix} I_m & -\Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} \\ 0 & I_n \end{bmatrix}^{\mathsf{T}} \\ &= \begin{bmatrix} \Sigma_{\mathbf{x}} - \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} \Sigma_{\mathbf{yx}} & 0 \\ 0 & \Sigma_{\mathbf{y}} \end{bmatrix}. \end{aligned}$$

Multivariate Gaussian - Condition

- $\tilde{\mathbf{x}}$ and $\mathbf{y}$ are independent.

- We now have

$$\begin{aligned}
\mathsf{E}[\mathbf{x}|\mathbf{y}] &= \mathsf{E}[\hat{\mathbf{x}} + \tilde{\mathbf{x}}|\mathbf{y}] \\
&= \mathsf{E}[\hat{\mathbf{x}}|\mathbf{y}] + \mathsf{E}[\tilde{\mathbf{x}}|\mathbf{y}] \\
&= \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}\mathbf{y} + \mathsf{E}[\tilde{\mathbf{x}}] \\
&= \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}\mathbf{y}.
\end{aligned}$$

Multivariate Gaussian - Condition

- For the general case that

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{y}} \end{bmatrix} \right)$$

- We choose

$$\begin{aligned} \hat{\mathbf{x}} &= \mu_{\mathbf{x}} + \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} (\mathbf{y} - \mu_{\mathbf{y}}), \\ \tilde{\mathbf{x}} &= \mathbf{x} - \hat{\mathbf{x}} \\ &= \mathbf{x} - \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{y}}^{-1} \mathbf{y} \end{aligned}$$

- Since
  1. $\tilde{\mathbf{x}}$ and $\mathbf{y}$ are independent
  2. $\mathsf{E}[\tilde{\mathbf{x}}] = 0$

  the proposition is proved.

Multivariate Gaussian - Condition

- The conditional PDF of $\mathbf{x}$ can then be easily determined.

Proposition (Conditioning)

If $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ are jointly Gaussian

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{y}} \end{bmatrix} \right),$$

then the conditional distribution of $\mathbf{x}$ given $\mathbf{y} = y$ is still Gaussian with distribution

$$N\left(\mu_{\mathbf{x}} + \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}(y - \mu_{\mathbf{y}}), \Sigma_{\mathbf{x}} - \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{y}}^{-1}\Sigma_{\mathbf{yx}}\right).$$

Summary

- The properties of sum and conditioning are actually the skeleton of the Kalman filter.

Kalman Filter and its Properties

Kalman Filter

We consider a linear dynamic system:

- time evolution (process) model:

$$\mathbf{s}_{t+1} = F\mathbf{s}_t + \mathbf{w}_t \in \mathbb{R}^n, \tag{1}$$

  - $\mathbf{s}_t$: the state
  - $\mathbf{w}_t$: process noise, independent zero-mean Gaussian with $\mathsf{E}[\mathbf{w}_t\mathbf{w}_t^\mathsf{T}] = Q > 0$

- observation (measurement) model:

$$\mathbf{o}_t = H\mathbf{s}_t + \mathbf{v}_t \in \mathbb{R}^m, \tag{2}$$

  - $\mathbf{o}_t$: the observed output
  - $\mathbf{v}_t$: measurement noise, independent zero-mean Gaussian with $\mathsf{E}[\mathbf{v}_t\mathbf{v}_t^\mathsf{T}] = R > 0$

Conditioning is the optimal estimator

- Given $\mathbf{o}_{1:t} = o_{1:t}$, what is the best estimator of $\mathbf{s}$ and $\mathbf{s}_{t+1}$?

Theorem

The σ-algebra generated by $\mathbf{z}$ is denoted by $\sigma(\mathbf{z})$, which contains all Borel-measurable functions of $\mathbf{z}$. Then

$$\arg\min_{\mathbf{y}\in\sigma(\mathbf{z})} \mathsf{E}[||\mathbf{x}-\mathbf{y}||^2] = \mathsf{E}[\mathbf{x}|\mathbf{y}]. \tag{3}$$

- The Kalman filter is an efficient algorithm to compute the conditional distributions of $\mathbf{s}_{t+1}$ given $\mathbf{o}_{1:t} = o_{1:t}$ and $\mathbf{o}_{1:t+1} = o_{1:t+1}$, recursively.

Time update

- Based on the time dynamic model,

$$\mathbf{s}_{t+1|t} = F\mathbf{s}_{t|t} + \mathbf{w}_t.$$

- If $\mathbf{s}_{t|t}$ follows the distribution $N(\bar{s}_{t|t}, \Sigma_{t|t})$, $\mathbf{s}_{t+1|t}$ is also Gaussian with mean and covariance

$$\begin{aligned}\bar{s}_{t+1|t} &= F\bar{s}_{t|t},\\ \Sigma_{t+1|t} &= F\Sigma_{t|t}F^{\mathsf{T}} + Q.\end{aligned}$$

Observation update

- From the observation model and condition each random variable on the information up to now:

$$\begin{aligned}\mathbf{o}_{t+1}|\mathbf{o}_{1:t} &= H\mathbf{s}_{t+1}|\mathbf{o}_{1:t} + \mathbf{v}_{t+1}|\mathbf{o}_{1:t} \\ &= H\mathbf{s}_{t+1}|\mathbf{o}_{1:t} + \mathbf{v}_{t+1}.\end{aligned}$$

- We have

$$\begin{bmatrix}\mathbf{s}_{t+1}|\mathbf{o}_{1:t} \\ \mathbf{o}_{t+1}|\mathbf{o}_{1:t}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\bar{s}_{t+1|t} \\ H\bar{s}_{t+1|t}\end{bmatrix}, \begin{bmatrix}\Sigma_{t+1|t} & \Sigma_{t+1|t}H^{\mathsf{T}} \\ H\Sigma_{t+1|t} & H\Sigma_{t+1|t}H^{\mathsf{T}} + R\end{bmatrix}\right).$$

- With $\mathbf{s}_{t+1|t+1} = (\mathbf{s}_{t+1}|\mathbf{o}_{1:t})|(\mathbf{o}_{t+1}|\mathbf{o}_{1:t})$.

$$\bar{s}_{t+1|t+1} = \bar{s}_{t+1|t} + \Sigma_{t+1|t}H^{\mathsf{T}}\left(H\Sigma_{t+1|t}H^{\mathsf{T}} + R\right)^{-1}\left(o_{t+1} - H\bar{s}_{t+1|1:t}\right),$$

$$\Sigma_{t+1|t+1} = \Sigma_{t+1|t} - \Sigma_{t+1|t}H^{\mathsf{T}}\left(H\Sigma_{t+1|t}H^{\mathsf{T}} + R\right)^{-1}H\Sigma_{t+1|t}.$$

Observation update

- Another common expression is:

$$\bar{s}_{t+1|t+1} = \bar{s}_{t+1|t} + K_{t+1}(o_{t+1} - H\bar{s}_{t+1|t}),$$
$$\Sigma_{t+1|t+1} = (I - K_{t+1}H)\Sigma_{t+1|t},$$

- the Kalman gain $K_{t+1} = \Sigma_{t+1|t}H^{\mathsf{T}}\left(H\Sigma_{t+1|t}H^{\mathsf{T}} + R\right)^{-1}$

- I like this expression better, from matrix inversion lemma

$$\Sigma_{t+1|t+1}^{-1} = \Sigma_{t+1|t}^{-1} + H^{\mathsf{T}}R^{-1}H.$$

Systems with Constant Inputs

- Consider

$$\mathbf{s}_{t+1} = F\mathbf{s}_t + Gu_t + \mathbf{w}_t,$$
$$\mathbf{o}_t = H\mathbf{s}_t + J\lambda_t + \mathbf{v}_t,$$

  u_t and λ are not random vectors, but are parameters in the system models.

- time update:

$$\bar{s}_{t+1|t} = F\bar{s}_{t|t} + Gu_t,$$
$$\Sigma_{t+1|t} = F\Sigma_{t|t}F^{\mathsf{T}} + Q.$$

- observation update:

$$\bar{s}_{t+1|t+1} = \bar{s}_{t+1|t} + \Sigma_{t+1|t}H^{\mathsf{T}}\left(H\Sigma_{t+1|t}H^{\mathsf{T}} + R\right)^{-1}\left(o_{t+1} - H\bar{s}_{t+1|1:t} - J\lambda_t\right),$$
$$\Sigma_{t+1|t+1} = \Sigma_{t+1|t} - \Sigma_{t+1|t}H^{\mathsf{T}}\left(H\Sigma_{t+1|t}H^{\mathsf{T}} + R\right)^{-1}H\Sigma_{t+1|t}.$$

Summary

- time update:

$$\bar{s}_{t+1|t} = F\bar{s}_{t|t},$$
$$\Sigma_{t+1|t} = F\Sigma_{t|t}F^{\mathsf{T}} + Q.$$

- observation update:

$$\bar{s}_{t+1|t+1} = \bar{s}_{t+1|t} + \Sigma_{t+1|t}H^{\mathsf{T}}\left(H\Sigma_{t+1|t}H^{\mathsf{T}} + R\right)^{-1}\left(o_{t+1} - H\bar{s}_{t+1|1:t}\right),$$
$$\Sigma^{-1}_{t+1|t+1} = \Sigma^{-1}_{t+1|t} + H^{\mathsf{T}}R^{-1}H.$$

time update	observation update
interval	instance
summation	condition
Σ increase	Σ decrease

Extended Kalman Filter (EKF)

Extended Kalman Filter

- We now consider non-linear model:

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{w}_t),$$
$$\mathbf{o}_t = h(\mathbf{s}_t, \mathbf{v}_t).$$

- "The EKF is simply an *ad hoc* state estimator that only approximates the optimality of Bayes' rule by linearization." [Welch and Bishop, 2011]

- We can approximate these nonlinear models by separating how the mean propagates and how the uncertainty evolves linearly.

Extended Kalman Filter

- time update

$$\begin{aligned}
\bar{s}_{t+1} &\approx f(\bar{s}_t, 0), \\
\delta \mathbf{s}_{t+1} &\approx \frac{\partial}{\partial s_t} f(s_t, w_t)\Big|_{s_t=\bar{s}_t, w_t=0} \delta \mathbf{s}_t + \frac{\partial}{\partial w_t} f(s_t, w_t)\Big|_{s_t=\bar{s}_t, w_t=0} \mathbf{w}_t \\
&= F_{s,t} \delta \mathbf{s}_t + F_{w,t} \mathbf{w}_t.
\end{aligned}$$

- observation update

$$\begin{aligned}
\bar{o}_{t+1} &\approx h(\bar{s}_{t+1}, 0), \\
\delta \mathbf{o}_{t+1} &\approx \frac{\partial}{\partial s_{t+1}} h(s_{t+1}, v_{t+1})\Big|_{s_{t+1}=\bar{s}_{t+1}, v_{t+1}=0} \delta \mathbf{s}_t \\
&\quad + \frac{\partial}{\partial s_{t+1}} h(s_{t+1}, v_{t+1})\Big|_{s_{t+1}=\bar{s}_{t+1}, v_{t+1}=0} \mathbf{v}_{t+1} \\
&= H_{s,t+1} \delta \mathbf{s}_{t+1} + H_{v,t+1} \mathbf{v}_{t+1}.
\end{aligned}$$

A simple 2D model

- unicycle model f:

$$\mathbf{s}_{t+1} = \begin{bmatrix} \mathbf{s}_{\theta,t+1} \\ \mathbf{s}_{x,t+1} \\ \mathbf{s}_{y,t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_{\theta,t} + \Delta t(u_{\omega,t} + \mathbf{w}_{\omega,t}) \\ \mathbf{s}_{x,t} + \Delta t \cos(\mathbf{s}_{\theta,t})(u_{v,t} + \mathbf{w}_{v,t}) \\ \mathbf{s}_{y,t} + \Delta t \sin(\mathbf{s}_{\theta,t})(u_{v,t} + \mathbf{w}_{v,t}) \end{bmatrix}$$

- bearing-and-range model h:

$$\mathbf{o}_t = \begin{bmatrix} \mathbf{o}_{\phi,t} \\ \mathbf{o}_{r,t} \end{bmatrix} = \begin{bmatrix} \tan^{-1}\left(\frac{\lambda_y - \mathbf{s}_{y,t}}{\lambda_x - \mathbf{s}_{x,t}}\right) - \mathbf{s}_{\theta,t} \\ \sqrt{(\lambda_x - \mathbf{s}_{x,t})^2 + (\lambda_y - \mathbf{s}_{y,t})^2} \end{bmatrix} + \mathbf{v}_t.$$

Extended Kalman Filter

- time update

$$\bar{s}_{t+1|t} = f(\bar{s}_t, 0),$$
$$\Sigma_{t+1|t} = F_{s,t}\Sigma_{t|t}F_{s,t}^{\mathsf{T}} + F_{w,t}QF_{w,t}^{\mathsf{T}}.$$

- observation update

$$\bar{s}_{t+1|t+1} = \bar{s}_{t+1|t} + \Sigma_{t+1|t}H_s^{\mathsf{T}}\left(H_s\Sigma_{t+1|t}H_s^{\mathsf{T}} + H_vRH_v^{\mathsf{T}}\right)^{-1}\left(o_{t+1} - \bar{o}_{t+1}\right),$$
$$\Sigma_{t+1|t+1} = \Sigma_{t+1|t} - \Sigma_{t+1|t}H_s^{\mathsf{T}}\left(H_s\Sigma_{t+1|t}H_s^{\mathsf{T}} + H_vRH_v^{\mathsf{T}}\right)^{-1}H_s\Sigma_{t+1|t}.$$

Neuroscience

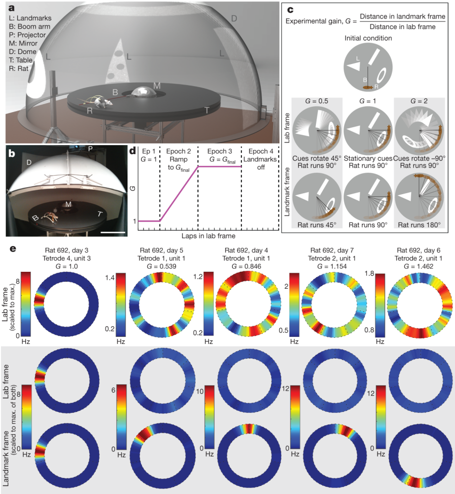

Figure: Recalibration of path integration in hippocampal place cells. [Jayakumar, 2019]

Summary

- the theory of Kalman filter
- the implementation of EKF