Kalman Filter and Localization

Tsang-Kai Chang

Oct. 23 and 25, 2018

Summary

- Multivariate Gaussian random vector
- Kalman filter
- Properties of Kalman filter
- Extended Kalman filter (EKF)
- Localization by EKF and its problem

Multivariate Gaussian

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

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

- We write $X \sim \mathcal{N}(\mu, \Sigma)$.
- Σ is a positive definite matrix.

Definition

A symmetric matrix A is positive definite if $x^\mathsf{T} A x > 0$ for all non-zero vector x. Simply write $A > 0$.

Multivariate Gaussian - Sum

- For two independent r.v.s $X_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $X_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$, we have

$$X = X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$$

$$\begin{aligned}
\mathsf{E}[X] &= \mathsf{E}[X_1 + X_2] = \mathsf{E}[X_1] + \mathsf{E}[X_2] \\
\mathsf{Cov}(X) &= \mathsf{E}(X - \mu)(X - \mu)^\mathsf{T} \\
&= \mathsf{E}(X_1 + X_2 - \mu_1 - \mu_2)(X_1 + X_2 - \mu_1 - \mu_2)^\mathsf{T} \\
&= \\
&= \\
&= \Sigma_1 + \Sigma_2
\end{aligned}$$

Multivariate Gaussian - Affine Transformation

- If $Y = c + BX$ is an affine transformation of $X \sim \mathcal{N}(\mu, \Sigma)$, where c is a constant vector and B is a constant matrix, then

$$Y \sim \mathcal{N}(c + B\mu, B\Sigma B^{\mathsf{T}}).$$

Multivariate Gaussian - Conditional Distribution

- Suppose that

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right),$$

  then

$$X_1 | X_2 \sim \mathcal{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

- Direct write down the formula for conditional probability then the answer shows up. [1]

[1] http://cs229.stanford.edu/section/more_on_gaussians.pdf

Multivariate Gaussian - Conditional Distribution

$$
\begin{aligned}
p(x_1|x_2) &= \frac{p(x_1, x_2; \mu, \Sigma)}{\int_{x_1 \in \mathbf{R}^m} p(x_1, x_2; \mu, \Sigma) dx_1} \\
&= \frac{1}{Z} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \right) \\
&= \frac{1}{Z} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \right) \\
&= \frac{1}{Z} \exp\Bigg(-\frac{1}{2} \Big[(x_1 - \mu_1)^\mathsf{T} V_{11} (x_1 - \mu_1) + (x_1 - \mu_1)^\mathsf{T} V_{12} (x_2 - \mu_2) \\
&\qquad + (x_2 - \mu_2)^\mathsf{T} V_{21} (x_1 - \mu_1) + (x_2 - \mu_2)^\mathsf{T} V_{22} (x_2 - \mu_2) \Big] \Bigg)
\end{aligned}
$$

Multivariate Gaussian - Conditional Distribution

- completion of squares

$$\frac{1}{2}z^\mathsf{T}Az+b^\mathsf{T}z+c=\frac{1}{2}(z+A^{-1}b)^\mathsf{T}A(z+A^{-1}b)^\mathsf{T}+c-\frac{1}{2}b^\mathsf{T}A^{-1}b$$

- matrix inversion

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} V_{11} & -V_{11}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}V_{11} & (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{bmatrix}$$

where $V_{11}=(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$

Multivariate Gaussian - Conditional Distribution

$$
\begin{aligned}
p(x_1|x_2) = \frac{1}{Z}\exp\Bigg(-\frac{1}{2}\Big[&(x_1-\mu_1)^\mathsf{T} V_{11}(x_1-\mu_1) + (x_1-\mu_1)^\mathsf{T} V_{12}(x_2-\mu_2) \\
&+ (x_2-\mu_2)^\mathsf{T} V_{21}(x_1-\mu_1) + (x_2-\mu_2)^\mathsf{T} V_{22}(x_2-\mu_2)\Big]\Bigg) \\
= \frac{1}{Z'}\exp\Bigg(-\frac{1}{2}\Big[&(x_1-\mu_1+V_{11}^{-1}V_{12}(x_2-\mu_2))^\mathsf{T} V_{11} \\
&(x_1-\mu_1+V_{11}^{-1}V_{12}(x_2-\mu_2))\Big]\Bigg)
\end{aligned}
$$

Kalman Filter

- state estimation for dynamic system
- optimal if noises are Gaussians
- recursive formula
- While KF is studied by researchers from different fields, the way to explain it somehow differs.

Kalman Filter

We consider a linear dynamic system:

- time evolution (process) model:

$$x_{t+1} = Fx_t + w_t \tag{1}$$

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

- observation (measurement) model:

$$y_t = Hx_t + v_t \tag{2}$$

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

We use $\hat{x}_t$ with Gaussian distribution $N(\bar{x}_t, \Sigma_t)$ to estimate x_t.

Time Update

Begin with the time evolution model

$$\hat{x}_{t+1} = F\hat{x}_t + w_t,$$

the distribution of $\hat{x}_{t+1}$ can be characterized by its mean

$$\begin{aligned}\bar{x}_{t+1} &= \mathsf{E}[F\hat{x}_t + w_t] \\ &= F\,\mathsf{E}[\hat{x}_t] + \mathsf{E}[w_t] \\ &= F\bar{x}_t\end{aligned}$$

and covariance

$$\begin{aligned}\Sigma_{t+1} &= \mathsf{E}\left[(x_{t+1} - \bar{x}_{t+1})(x_{t+1} - \bar{x}_{t+1})^{\mathsf{T}}\right] \\ &= \mathsf{E}\left[(F\tilde{x}_t + w_t)(F\tilde{x}_t + w_t)^{\mathsf{T}}\right] \\ &= \mathsf{E}\left[F\tilde{x}_t\tilde{x}_t^{\mathsf{T}}F^{\mathsf{T}}\right] + \mathsf{E}\left[w_t w_t^{\mathsf{T}}\right] \\ &= F\Sigma_t F^{\mathsf{T}} + Q\end{aligned}$$

Observation Update

$$y_t = Hx_t + v_t$$

- Condition on $\hat{x}_{t^-}$, the information up to time t,

$$y_t|\hat{x}_{t^-} = Hx_t|\hat{x}_{t^-} + v_t|\hat{x}_{t^-} = Hx_t|\hat{x}_{t^-} + v_t$$

- Gaussian

$$\begin{bmatrix} x_t|\hat{x}_{t^-} \\ y_t|\hat{x}_{t^-} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{x}_{t^-} \\ H\bar{x}_{t^-} \end{bmatrix}, \begin{bmatrix} \Sigma_{t^-} & \Sigma_{t^-}H^\mathsf{T} \\ H\Sigma_{t^-} & H\Sigma_{t^-}H^\mathsf{T} + R \end{bmatrix} \right)$$

Observation Update

- Define $\hat{x}_{t^+} = x_t | y_t = (x_t | \hat{x}_{t^-}) | (y_t | \hat{x}_{t^-})$, we have

$$\bar{x}_{t^+} = \bar{x}_{t^-} + \Sigma_{t^-} H^\mathsf{T} \left(H \Sigma_{t^-} H^\mathsf{T} + R \right)^{-1} (y_t - H \bar{x}_{t^-})$$

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

- Or you may see other notation

$$\bar{x}_{t^+} = \bar{x}_{t^-} + K_t (y_t - H \bar{x}_{t^-})$$

$$\Sigma_{t^+} = (I - K_t H) \Sigma_{t^-}$$

- $(y_t - H \bar{x}_{t^-})$ is called innovation
- $K_t = \Sigma_{t^-} H^\mathsf{T} \left(H \Sigma_{t^-} H^\mathsf{T} + R \right)^{-1}$ is defined as Kalman gain
- but I don't think this is a good formula

Kalman Filter

- time update

$$\bar{x}_{t+1} = F\bar{x}_t$$
$$\Sigma_{t+1} = F\Sigma_t F^\mathsf{T} + Q$$

- observation update

$$\bar{x}_{t^+} = \bar{x}_{t^-} + \Sigma_{t^-} H^\mathsf{T} \left(H\Sigma_{t^-} H^\mathsf{T} + R \right)^{-1} (y_t - H\bar{x}_{t^-})$$
$$\Sigma_{t^+} = \Sigma_{t^-} - \Sigma_{t^-} H^\mathsf{T} \left(H\Sigma_{t^-} H^\mathsf{T} + R \right)^{-1} H\Sigma_{t^-}$$

notation remark

- time update: $t+1|t \leftarrow t|t$
- observation update: $t|t \leftarrow t|t-1$

Discussion

How about the system?

$$x_{t+1} = Fx_t + Gu_t + w_t$$

$$\begin{aligned}
\bar{x}_{t+1} &= \mathsf{E}[F\hat{x}_t + Gu_t + w_t] \\
&= F\,\mathsf{E}[\hat{x}_t] + Gu_t + \mathsf{E}[w_t] \\
&= F\bar{x}_t + Gu_t \\
\Sigma_{t+1} &= F\Sigma_t F^{\mathsf{T}} + Q
\end{aligned}$$

Joseph's Form

- Numerical rounding problem
- no longer PSD

$$\Sigma_{t^+} = [I - K_t H]\Sigma_{t^-}[I - K_t H]^\mathsf{T} + K_t R K_t^\mathsf{T} \tag{3}$$

Woodbury Matrix Identity

- Woodbury matrix identity

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1}$$

- Then we can rewrite the observation update as

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

  Since $H^\mathsf{T} R^{-1} H > 0$, $\Sigma_{t^+} < \Sigma_{t^-}$.

- If $H^\mathsf{T} R^{-1} H$ is nonsingular, $\Sigma_{t^+} < (H^\mathsf{T} R^{-1} H)^{-1}$.

Summary

- Multivariate Gaussian random vector
- Kalman filter
- Properties of Kalman filter
- Extended Kalman filter (EKF)
- Localization by EKF and its problem

time update	observation update
interval	instance
summation	condition
$\Sigma_{t+1} > \Sigma_t$	$\Sigma_{t^+} < \Sigma_{t^-}$

Toy Example

Let's consider a toy example with $x_t \in \mathbf{R}^2$

$$F = Q = \Sigma_0 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

- Case 1:

$$H_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad R_1 = 1$$

- Case 2:

$$H_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For each time instance, time update is executed then observation update.
What is your observation?

Riccati Recursion

- For each time instance with one time and one observation update, the covariance update is given by

$$\begin{aligned}\Sigma_{t+1} &= F\left(\Sigma_t^{-1} + H^{\mathsf{T}} R^{-1} H\right)^{-1} F^{\mathsf{T}} + Q \\ &= F\Sigma_t F^{\mathsf{T}} - F\Sigma_t H^{\mathsf{T}}\left(R + H^{\mathsf{T}}\Sigma_t H\right)^{-1} H\Sigma F^{\mathsf{T}} + Q.\end{aligned}$$

- The update of Σ_t follows discrete-time Riccati recursion.
- The convergence of Σ_t is essential for the estimation performance.

Riccati Recursion

Lemma

Given $(F, Q^{1/2})$ *stabilizable and* (F, H) *detectable and* $\Sigma_0 \geq 0$*, then*

$$\lim_{t \to \infty} \Sigma_t = \Sigma$$

exponentially fast, where Σ *is the solution of the discrete Riccati equation*

$$\Sigma = F\Sigma F^{\mathsf{T}} + Q^{\mathsf{T}} - F\Sigma H^{\mathsf{T}} \left(R + H^{\mathsf{T}}\Sigma H\right)^{-1} H\Sigma F^{\mathsf{T}}.$$

- S. W. Chan, G. C. Goodwin, and K. S. Sin, "Convergence properties of the Riccati difference equation in optimal filtering of nonstabilizable systems," *IEEE Trans. on Automatic Control*, vol. AC-29, no. 2, 1984.

Extended Kalman Filter

- With *non-linear* stochastic difference equations for system model

$$x_{t+1} = f(x_t, u_t, w_t)$$
$$y_t = h(x_t, v_t)$$

- Linear approximation on noise

$$\tilde{x}_{t+1} \approx F_t \tilde{x}_t + W_t w_t$$

$$F_{ij} = \frac{\partial f_i}{\partial x_j}(x_t, u_t, 0), \quad W_{ij} = \frac{\partial f_i}{\partial w_j}(x_t, u_t, 0)$$

$$\tilde{y}_t \approx H_t \tilde{x}_t + V_t v_t$$
$$H_{ij} = \frac{\partial h_i}{\partial x_j}(x_t, 0), \quad V_{ij} = \frac{\partial h_i}{\partial v_j}(x_t, 0)$$

Extended Kalman Filter

- time update

$$\bar{x}_{t+1} = f(\bar{x}_t, u_t, 0)$$
$$\Sigma_{t+1} = F_t \Sigma_t F_t^{\mathsf{T}} + W_t Q W_t^{\mathsf{T}}$$

- observation update

$$\bar{x}_{t^+} = \bar{x}_{t^-} + \Sigma_{t^-} H_t^{\mathsf{T}} \left(H_t \Sigma_{t^-} H_t^{\mathsf{T}} + R \right)^{-1} (y_t - h(\bar{x}_{t^-}, 0))$$
$$\Sigma_{t^+} = \Sigma_{t^-} - \Sigma_{t^-} H_t^{\mathsf{T}} \left(H_t \Sigma_{t^-} H_t^{\mathsf{T}} + R \right)^{-1} H_t \Sigma_{t^-}$$

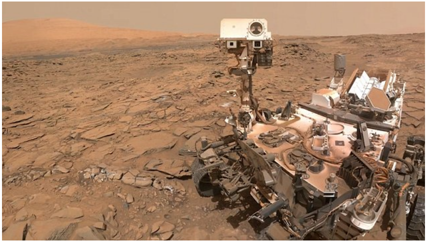
Extended Kalman Filter

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

- It is relative easy to find consistent estimate, but it is not easy to stay with Gaussian assumption.

Localization

- Goal: try to know the spatial state of a mobile agent by all the available information



[Credit: NASA, 2018]

System Model

- time evolution model

$$\begin{aligned}\theta_{t+1} &= \theta_t + (\omega_t + w_{\omega,t})\Delta t \\ x_{t+1} &= x_t + (v_t + w_{v,t})\cos\theta_t \Delta t \\ y_{t+1} &= y_t + (v_t + w_{v,t})\sin\theta_t \Delta t\end{aligned}$$

  - orientation θ_t
  - position x_t, y_t
  - translational velocity v_t
  - angular velocity ω_t

System Model

- observation model
  (assume there is a landmark with position known)

$$\begin{bmatrix} d_t \\ \phi_t \end{bmatrix} = \begin{bmatrix} \sqrt{(x_t - x_l)^2 + (y_t - y_l)^2} \\ \tan^{-1}\left(\frac{y_t - y_l}{x_t - x_l}\right) - \theta_t \end{bmatrix} + \begin{bmatrix} v_{d,t} \\ v_{\phi,t} \end{bmatrix}$$

  - distance d_t
  - bearing ϕ_t
  - landmark position (x_l, y_l)

[Credit: UTIAS dataset, 2011]

Localization by EKF

- spatial state s_t

$$s_t = \begin{bmatrix} \theta_t \\ x_t \\ y_t \end{bmatrix}$$

- spatial state

$$\hat{s}_t = \begin{bmatrix} \hat{\theta}_t \\ \hat{x}_t \\ \hat{y}_t \end{bmatrix} = \mathcal{N}(\bar{s}_t, \Sigma_t)$$

Time update

- time update

$$\begin{bmatrix} \bar{\theta}_{t+1} \\ \bar{x}_{t+1} \\ \bar{y}_{t+1} \end{bmatrix} = \begin{bmatrix} \bar{\theta}_t + \omega_t \Delta t \\ \bar{x}_t + v_t \cos \bar{\theta}_t \Delta t \\ \bar{y}_t + v_t \sin \bar{\theta}_t \Delta t \end{bmatrix}$$

- noise propagation by linearization

$$\begin{aligned} \begin{bmatrix} \tilde{\theta}_{t+1} \\ \tilde{x}_{t+1} \\ \tilde{y}_{t+1} \end{bmatrix} &\approx \begin{bmatrix} 1 & 0 & 0 \\ -v_t \sin \bar{\theta}_t \Delta t & 1 & 0 \\ v_t \cos \bar{\theta}_t \Delta t & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\theta}_t \\ \tilde{x}_t \\ \tilde{y}_t \end{bmatrix} + \begin{bmatrix} \Delta t & 0 \\ 0 & \cos \bar{\theta}_t \Delta t \\ 0 & \sin \bar{\theta}_t \Delta t \end{bmatrix} \begin{bmatrix} w_{\omega,t} \\ w_{v,t} \end{bmatrix} \\ &= F_t \begin{bmatrix} \tilde{\theta}_t \\ \tilde{x}_t \\ \tilde{y}_t \end{bmatrix} + W_t \begin{bmatrix} w_{\omega,t} \\ w_{v,t} \end{bmatrix} \end{aligned}$$

- covariance update

$$\Sigma_{t+1} = F_t \Sigma_t F_t^{\mathsf{T}} + W_t Q W_t^{\mathsf{T}}$$

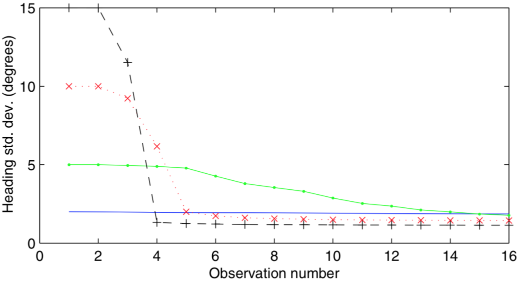
Observation update

- observation noise approximation

$$
\begin{aligned}
\begin{bmatrix} \tilde{d}_t \\ \tilde{\phi}_t \end{bmatrix} &\approx \begin{bmatrix} \frac{\bar{x}_t - x_l}{d_t} & \frac{\bar{y}_t - y_l}{d_t} & 0 \\ -\frac{\bar{y}_t - y_l}{d_t^2} & \frac{\bar{x}_t - x_l}{d_t^2} & -1 \end{bmatrix} \begin{bmatrix} \tilde{\theta}_t \\ \tilde{x}_t \\ \tilde{y}_t \end{bmatrix} + \begin{bmatrix} v_{d,t} \\ v_{\phi_t} \end{bmatrix} \\
&= H_t \begin{bmatrix} \tilde{\theta}_t \\ \tilde{x}_t \\ \tilde{y}_t \end{bmatrix} + \begin{bmatrix} v_{d,t} \\ v_{\phi_t} \end{bmatrix}
\end{aligned}
$$

How good is EKF localization?

- linearization inconsistency
  - excessive or spurious information gain
  - peculiar update characteristics of the state mean



[Bailey, et al., 2006]

Explanation

- Consider the simplified case:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \omega_t \Delta t$$
$$\hat{x}_{t+1} = \hat{x}_t + v_t \cos \hat{\theta}_t \Delta t$$
$$\hat{y}_{t+1} = \hat{y}_t + v_t \sin \hat{\theta}_t \Delta t$$

- $\hat{s}_t$ is estimated by some distribution, says Gaussian. Will $\hat{s}_{t+1}$ remain Gaussian?

- If not, will the mean remain valid at least?

von Mises Distribution

- The von Mises distribution $vM(\mu, \kappa), \kappa > 0$ has PDF

$$g(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \quad 0 \leq \theta < 2\pi. \tag{4}$$

where I_p is the modified Bessel function of the first kind and order p, which can be defined by

$$I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) e^{\kappa \cos \theta} d\theta. \tag{5}$$

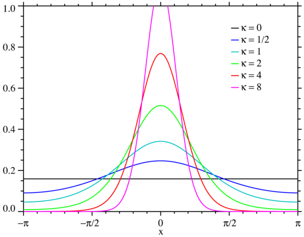


Figure: PDFs of von Mises distribution. [from Wiki]

Explanation

- Assume that

$$\begin{aligned}\hat{\theta}_t &\sim vM(\bar{\theta}_t, \kappa_t)\\ \hat{x}_t &\sim \mathcal{N}(\bar{x}_t, \sigma_{x,t})\\ \hat{y}_t &\sim \mathcal{N}(\bar{y}_t, \sigma_{y,t})\end{aligned}$$

- We have

$$\begin{aligned}\mathsf{E}[\hat{x}_{t+1}] &= \mathsf{E}[\hat{x}_t + v_t \cos(\hat{\theta}_t)\Delta t]\\ &= \bar{x}_t + v_t\,\mathsf{E}[\cos(\hat{\theta}_t)]\Delta t\\ &= \bar{x}_t + v_t \frac{I_1(\kappa_t)}{I_0(\kappa_t)} \cos(\bar{\theta}_t)\Delta t\\ &\neq \bar{x}_t + v_t \cos(\bar{\theta}_t)\Delta t\end{aligned}$$

Reference

- G. Welch and G. Bishop, "An Introduction to the Kalman Filter," SIGGRAPH, 2011.