State Estimation

1. Problem of Estimation.

① $P(x \mid z, u)$

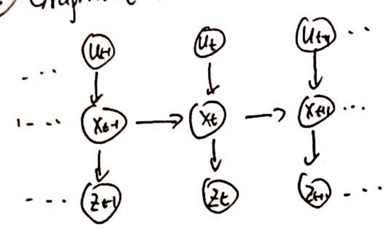
x: state. like position of robots.

u: control commands. like go left 1m.

z: observation. like the robot sees a tree 1m in front of him

② State estimation: we know robot commands $u_{1:t}$. we know what the robot observes: $z_{1:t}$. we want to know where it is x_t.

③ Graphical Model.



<1>. x_t's are hidden.

<2>. u_t and x_{t-1} affects x_t.

arrow "$\rightarrow$" means "affect"

2. Bayes Filter.

denote $bel(x_t) = P(x_t \mid z_{1:t}, u_{1:t})$.

Bayes Rule: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)} = \eta\, P(B|A)P(A)$ — we don't care about $P(B)$.

$= \eta\, P(z_t \mid x_t, z_{1:t-1}, u_{1:t})\, P(x_t \mid z_{1:t-1}, u_{1:t})$.

If x_t is observed. z_t is independent of $z_{1:t-1}, u_{1:t}$.

$= \eta\, P(z_t|x_t)\, P(x_t \mid z_{1:t-1}, u_{1:t})$

Law of total probability. $P(x_1) = \int P(x_1, x_2)\,dx_2 = \int P(x_1|x_2)P(x_2)\,dx_2$

$= \eta\, P(z_t|x_t) \int_{x_{t-1}} P(x_t \mid x_{t-1}, z_{1:t-1}, u_{1:t})\, P(x_{t-1} \mid z_{1:t-1}, u_{1:t})\,dx_{t-1}$

Given x_{t-1}, we only need u_t.

$= \eta\, P(z_t|x_t) \int_{x_{t-1}} P(x_t \mid x_{t-1}, u_t)\, P(x_{t-1} \mid z_{1:t-1}, u_{1:t})\,dx_{t-1}$

"u_t does not affect x_{t-1}" → An assumption!

$= \eta\, P(z_t|x_t) \int_{x_{t-1}} P(x_t \mid x_{t-1}, u_t)\, bel(x_{t-1})\,dx_{t-1}$.

Bayes Filter. $bel(x_{t-1}) \rightarrow bel(x_t)$.

$\overline{bel}(x_t) = \int P(x_t \mid u_t, x_{t-1})\, bel(x_{t-1})\, dx_{t-1}$ Prediction

$bel(x_t) = \eta\, P(z_t|x_t)\, \overline{bel}(x_t)$. Correction

Bayes Filter Says: if we know $P(x_t|u_t, x_{t-1})$ and $P(z_t|x_t)$, we can use $bel(x_{t-1})$ to predict $bel(x_t)$.

motion model: $P(x_t \mid u_t, x_{t-1})$ we know x_{t-1} and commands u_t, what can we say about x_t

observation model: $P(z_t|x_t)$ we know x_t, what's probability we observe z_t?

now the questions come. How do we get motion and observation model?

3. Kalman Filter.

Kalman Filter provides a model for $\begin{cases} P(x_t|u_t, x_{t-1}) \\ P(z_t|x_t) \end{cases}$

① The Kalman Filter Model.

$x_t = A_t x_{t-1} + B_t u_t + \varepsilon_t$

$z_t = C_t x_t + \delta_t$.

Assumption: everything is Gaussian and of a linear transform

② Components:

A_t: How state changes without command

B_t: How commands change state from $t-1$ to t.

C_t: How to map state to an observation.

ε_t and δ_t are Gaussian noise. $\varepsilon_t \sim N(0, R_t)$

$\delta_t \sim N(0, Q_t)$

③ In Review of MVG. we have seen how to use affine transformation of Gaussian to solve such linear systems!

<1>. $p(x_t | u_t, x_{t-1})$ now u_t and x_{t-1} are observed.

so x_t is a transformation of Gaussian ε_t

$\Rightarrow p(x_t | u_t, x_{t-1}) \sim N(A_t x_{t-1} + B_t u_t, R_t)$.

<2> $p(z_t | x_t) = N(C_t x_t, Q_t)$.

now we can plugin in these two terms in Bayes Filter and solve for $bel(x_t)$

④ Let's use the trick similar to that used previously.

$\overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1})\, p(x_{t-1} | u_{1:t-1}, z_{1:t-1})\, dx_{t-1}$..

note $p(x_t | u_t, x_{t-1}) = p(x_t | x_{t-1}, u_{1:t}, z_{1:t-1})$
$p(x_{t-1} | u_{1:t-1}, z_{1:t-1}) = p(x_{t-1} | u_{1:t}, z_{1:t-1})$ } Inverse of our assumption

So we are doing:

$\tilde{p}(x_t | x_{t-1})$, $\tilde{p}(x_{t-1}) \overset{f}{\Rightarrow} \tilde{p}(x_t)$ Recall Affine transformation!

$N(A_t x_{t-1} + B_t u_t, R_t)$, $N(\mu_{t-1}, \Sigma_{t-1}) \overset{f}{\Rightarrow} N(\bar{\mu}_t, \bar{\Sigma}_t)$

where. $\bar{\mu}_t = A_t \mu_{t-1} + B_t u_t$

$\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^T + R_t$.

similarly we do for $bel(x_t) = \eta\, p(z_t | x_t)\, \overline{bel}(x_t)$

$bel(x_t) = p(x_t | z_{1:t}, u_{1:t}) = p(x_t | z_t, z_{1:t-1}, u_{1:t})$

$p(z_t | x_t) = p(z_t | x_t, z_{1:t-1}, u_{1:t})$

$\overline{bel}(x_t) = \int_{x_{t-1}} p(x_t | x_{t-1}, z_{1:t-1}, u_{1:t})\, p(x_{t-1} | u_{1:t}, z_{1:t-1})\, dx_{t-1}$

$= \int p(x_t, x_{t-1} | z_{1:t-1}, u_{1:t})\, dx_{t-1}$

$= p(x_t | z_{1:t-1}, u_{1:t})$.

again: $\tilde{p}(x_t)$, $\tilde{p}(z_t | x_t) \overset{f}{\Rightarrow} \tilde{p}(x_t | z_t)$.

Use Affine transform:

$K_t = \bar{\Sigma}_t C_t^T (C_t \bar{\Sigma}_t C_t^T + Q_t)^{-1}$

$\mu_t = \bar{\mu}_t + K_t (z_t - C_t \bar{\mu}_t)$

$\Sigma_t = (I - K_t C_t) \bar{\Sigma}_t$

⑤ Finally we call this Kalman Filter Algos

bel$(x_{t-1}) \longrightarrow$ bel(x_t)

$\uparrow$

$(\mu_{t-1}, \Sigma_{t-1}) + (u_t, z_t) \longrightarrow (\mu_t, \Sigma_t)$.

⑥ perfect Sensor.

A perfect sensor has no noise. $Q_t = 0$

$\Rightarrow K_t = C_t^{-1} \Rightarrow \mu_t = C_t^{-1} z_t$.

x_t is only affected by observation.

⑦ Extremely noisy sensor.

$Q_t \to \infty \Rightarrow K_t = 0 \Rightarrow \mu_t = A_t \mu_{t-1} + B_t u_t$.

not affected by observation

⑧ kalman Gain: K_t. It knows when to discard noise and when there's no noise. It is clever to take noise into consideration.

4. Extended Kalman Filter.

what if we don't Assume a linear transform?

① Extended model:

$x_t = g(u_t, x_{t-1}) + \varepsilon_t$ Now. non-linear transfom of Gaussian may not be Gaussian.

$z_t = h(x_t) + \delta_t$.

② We can use Taylor expansion to linearize the model.

$g(u_t, x_{t-1}) \approx g(u_t, \mu_{t-1}) + \underbrace{\frac{\partial g(u_t, \mu_{t-1})}{\partial x_{t-1}}}_{G_t} (x_{t-1} - \mu_{t-1})$

$h(x_t) = h(\bar{\mu}_t) + \underbrace{\frac{\partial h(\bar{\mu}_t)}{\partial x_t}}_{H_t} (x_t - \bar{\mu}_t)$.

H_t Jacobian matrix.

③ plug in!

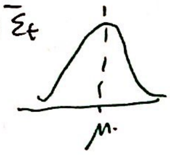
$\bar{\mu}_t = g(u_t, \mu_{t-1})$ $\bar{\Sigma} = G_t \Sigma_{t-1} G_t^T + R_t$

$K_t = \bar{\Sigma}_t H_t^T (H_t \bar{\Sigma}_t H_t^T + Q_t)^{-1}$

$\mu_t = \bar{\mu}_t + K_t (z_t - h(\bar{\mu}_t))$

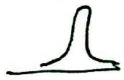
$\Sigma_t = (I - K_t H_t) \bar{\Sigma}_t$

why expand at μ_{t-1}?



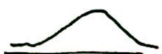
μ_{t-1} is where most points are!

$\Rightarrow$ if we have slim gaussian



we have good result.

if we have fat Gaussian



not as good.

5. Uncented Kalman Filter.

Instead of using linearization, why not use uncent transfer?

① sigma point: $\chi_t^{[i]}$. $[\mu_{t-1}, \mu_{t-1} \mp (\gamma \sqrt{\Sigma_{t-1}})_i]$

② prediction step: $\bar{\mu}_t$, $\bar{\Sigma}_t$.

(i) transform points: $\bar{\chi}_t^* = g(u, \chi_{t-1})$

(ii) estimate. $\bar{\mu}_t = \sum_{i=0}^{2n} w_m^{[i]} \bar{\chi}_t^{*[i]}$

don't forget noise.

$\bar{\Sigma}_t = \underbrace{\sum_{i=0}^{2n} w_c^{[i]} (\bar{\chi}_t^{*[i]} - \bar{\mu}_t)(\bar{\chi}_t^{*[i]} - \bar{\mu}_t)^T}_{G_t \Sigma_{t-1} G_t^T \text{ transformed covariance}} + R_t$.

③ Correction step.

(i) compute transformed points $\bar{Z}_t = h(\bar{\chi}_t)$

(ii) $\hat{z}_t = \sum_{i=0}^{2n} w_m^{[i]} \bar{Z}_t^{[i]}$

$$S_t = \sum_{i=0}^{2n} w_c^{[i]} (\bar{Z}_t^{[i]} - \hat{z}_t)(\bar{Z}_t^{[i]} - \hat{z}_t)^T + Q_t$$

$$z_t = h(x_t) + \delta_t$$

(iii) Kalman Gain.

$$K_t = \underbrace{\bar{\Sigma}_t H_t^T}_{\bar{\Sigma}_t^{x,z}} (\underbrace{H_t \bar{\Sigma}_t H_t^T + Q_t}_{S_t})^{-1}$$

$H_t \bar{\Sigma}_t H_t^T$: the transformed covariance under $h(\,)$.

$\bar{\Sigma}_t H_t^T$: partially transformed covariance

$$\bar{\Sigma}_t H_t^T = \sum_{i=0}^{2n} w_c^{[i]} \underbrace{(\bar{\chi}_t^{[i]} - \bar{\mu}_t)}_{\text{untransformed variable}} \underbrace{(\bar{Z}_t^{[i]} - \hat{z}_t)^T}_{\text{transformed variable}}$$

$$K_t = \bar{\Sigma}_t^{x,z} S_t^{-1}$$

An intuition.

$H_t \bar{\Sigma}_t H_t^T$ let $\bar{\Sigma}_t = LL^T$

$\Rightarrow H_t \bar{\Sigma}_t H_t^T = (HL)(HL)^T$. → transformation of variable.

now $\bar{\Sigma}_t H_t^T = L\,(HL)^T$ → transformed

L untransformed

(iv) $\Sigma_t = (I - K_t H_t)\bar{\Sigma}_t$

$= \bar{\Sigma}_t - K_t (\bar{\Sigma}_t^{x,z})^T$

$= \bar{\Sigma}_t - K_t S_t K_t^T$

Remark:

$x_{t-1} \overset{g}{\sim} x_t \overset{h}{\sim} z_t \to x_{t+1}$

we have $\chi^{[i]}$, then compute transformed covariance under g.

plus noise we get $\bar{\Sigma}_t$.

Then S_t is covariance under transformation $h(\,) + \delta_t$

But S_t is not Σ_t!

Kalman filter bridges $\bar{\Sigma}_t$ and Σ_t. Thus we need to compute Kalman Gain and use S_t to get Σ_t.

$$\mu_t = \bar{\mu}_t + K_t(z_t - \hat{z}_t)$$

④ Compare UKF and EKF.

(i) For slim banana:  we expect EKF to be good!

(ii) For large variance, UKF is good.

6. Information Filter: Kalman Filter in Information Space.

① idea: transform μ, Σ into ξ, Ω. That's it!

② prediction: $\bar{\Omega}_t = (A_t \Omega_{t-1}^{-1} A_t^T + R_t)^{-1}$ expensive

$\bar{\xi}_t = \bar{\Omega}_t (A_t \Omega_{t-1}^{-1} \xi_{t-1} + B_t u_t)$.

③ correction. $\begin{cases} \Omega_t = C_t^T Q_t^{-1} C_t + \bar{\Omega}_t \\ \xi_t = C_t^T Q_t^{-1} z_t + \bar{\xi}_t \end{cases}$ cheap.

note for KF, prediction is cheap, correction is expensive

7. Extended Information Filter.

① Since g( ) h( ) are defined in momentum space, we need to go back to moment for some term → extra computation.

② prediction.

$$\bar{\xi}_t = \bar{\Omega}_t \, g(u_t, \underbrace{\Omega_{t-1}^{-1} \xi_{t-1}}_{\mu_{t-1}})$$

③ correction.

$$\Omega_t = \bar{\Omega}_t + H_t^T Q_t^{-1} H_t$$

$$\xi_t = \bar{\xi}_t + H_t^T Q_t^{-1} (z_t - h(\bar{\mu}_t) + H_t \bar{\mu}_t)$$

$$\bar{\mu}_t = g(u_t, \mu_{t-1})$$

$$= g(u_t, \Omega_{t-1}^{-1} \xi_{t-1})$$