Bayesian Filtering

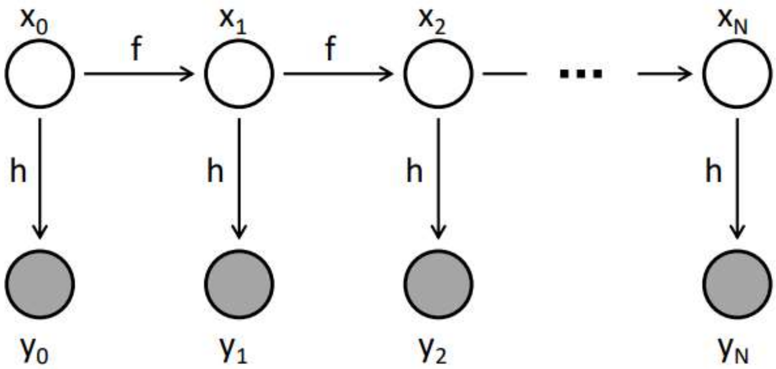
1. Dynamic Systems

시간에 따라 상태가 변하는 시스템 모델은 다음과 같이 표현할 수 있다.

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

여기서 x_t는 t에서의 상태, y_t는 t에서의 관측값, u_t, w_t는 노이즈이다.

또한 이를 그림으로 보면 다음과 같다.



해당 동적 시스템에서 Bayesian Filtering의 과정은 다음과 같이 prediction과 measurement update 과정으로 구성된다.

1) $P(X_t|y_{0:t})$로부터 $P(X_{t+1}|y_{0:t})$를 예측한다. (Prediction)

2) 예측값으로부터 다음 관측값인 y_{t+1}을 얻는다.

3) 얻은 관측값을 이용하여 $P(X_{t+1}|y_{0:t+1})$을 계산한다. (measurement update)

다음 과정이 재귀적으로 일어날 수 있다.

우선 Prediction 과정의 경우, 총합의 법칙과 바로 이전 상태에만 의존성을 지니는 마르코프 가정에 따라 다음과 같이 계산된다.

$$\begin{aligned} P(X_{t+1}|y_{0:t}) &= \int P(X_{t+1}|x_t, y_{0:t})P(x_t|y_{0:t})dx_t \\ &= \int P(X_{t+1}|x_t)P(x_t|y_{0:t})dx_t \end{aligned}$$

또한 Measurement Update 과정의 경우에는 베이지안 룰에 따라 다음과 같이 계산된다.

$$\begin{aligned} P(X_{t+1}|y_{0:t+1}) &= \frac{P(y_{t+1}|X_{t+1}, y_{0:t})P(X_{t+1}|y_{0:t})}{\int P(y_{t+1}|x_{t+1}, y_{0:t})P(x_{t+1}|y_{0:t})dx_{t+1}} \\ &= \frac{P(y_{t+1}|X_{t+1})P(X_{t+1}|y_{0:t})}{\int P(y_{t+1}|x_{t+1})P(x_{t+1}|y_{0:t})dx_{t+1}} \end{aligned}$$

2. Kalman Filtering

다음과 같은 linear dynamic model을 생각해보자.

$$x_{t+1} = Ax_t + Gw_t$$
$$y_t = Cx_t + v_t$$

여기서 초기 상태 분포는 $x_0 \sim N(\mu_0, \Sigma_0)$ (여기서 $\mu_0 = 0$ 가정) 과 같고 노이즈의 분포도 $w_t \sim N(0, Q), v_t \sim N(0, R)$이며 w_t, v_t, x_0은 서로 독립이다.

우선 x_t의 분포를 살펴보면 다음과 같다.

$$\begin{aligned} E(x_t) &= 0 \\ var(x_{t+1}) &= \Sigma_{t+1} = E(x_{t+1}x_{t+1}^T) \\ &= E((Ax_t + Gw_t)(Ax_t + Gw_t)^T) \\ &= AE(x_t x_t^T)A^T + \geq (w_t w_t^T)G^T \\ &= A\Sigma_t A^T + GQG^T \end{aligned}$$

또한 여기서 기본적으로 주어진 변수를 살펴보면 다음과 같다.

$\hat{x}_{t|t} = E(x_t|y_{0:t})$: $y_{0:t}$가 주어졌을 때 t에서의 상태의 평균

$P_{t|t} = E(\tilde{x}_{t|t}\tilde{x}_{t|t}^T|y_{0:t})$: $y_{0:t}$가 주어졌을 때 t에서의 상태의 공분산

$\tilde{x}_{t|t} = x_t - \hat{x}_{t|t}$: $y_{0:t}$가 주어졌을 때 t에서의 상태의 오차

이제 주된 과정인 Prediction과 Measurement update를 살펴보자.

1) Prediction

$$\begin{aligned} \hat{x}_{t|t} &:= \mathbb{E}(x_t|y_0,\ldots,y_t) \\ P_{t|t} &= \mathbb{E}(\tilde{x}_{t|t}\tilde{x}_{t|t}^T|y_0,\ldots,y_t) \\ \tilde{x}_{t|t} &= x_t - \hat{x}_{t|t} \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \hat{x}_{t+1|t} &:= \mathbb{E}(x_{t+1}|y_0,\ldots,y_t) \\ P_{t+1|t} &= \mathbb{E}\left(\tilde{x}_{t+1|t}\tilde{x}_{t+1|t}^T|y_0,\ldots,y_t\right) \\ \tilde{x}_{t+1|t} &= x_{t+1} - \hat{x}_{t+1|t} \end{aligned}$$

$$\begin{aligned} \hat{x}_{t+1|t} &= E(x_{t+1}|y_{0:t}) \\ &= E(Ax_t + Gw_t|y_{0:t}) \\ &= A\hat{x}_{t|t} \\ P_{t+1|t} &= E(\tilde{x}_{t+1|t}\tilde{x}_{t+1|t}^T|y_{0:t}) \\ &= E[(Ax_t + Gw_t - A\hat{x}_{t|t})(Ax_t + Gw_t - A\hat{x}_{t|t})^T|y_{0:t}) \\ &= E[(A(x_t - \hat{x}_{t|t}) + Gw_t)(A(x_t - \hat{x}_{t|t}) + Gw_t)^T|y_{0:t}) \\ &= AP_{t|t}A^T + GQG^T \end{aligned}$$

2) Measurement Update

구하고자 하는 것은 다음과 같이 정리된다.

$$\begin{aligned} \hat{x}_{t+1|t} &:= \mathbb{E}(x_{t+1}|y_0,\ldots,y_t) \\ P_{t+1|t} &= \mathbb{E}\left(\tilde{x}_{t+1|t}\tilde{x}_{t+1|t}^T|y_0,\ldots,y_t\right) \\ \tilde{x}_{t+1|t} &= x_{t+1} - \hat{x}_{t+1|t} \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \hat{x}_{t+1|t+1} &:= \mathbb{E}(x_{t+1}|y_0,\ldots,y_{t+1}) \\ P_{t+1|t+1} &= \mathbb{E}(\tilde{x}_{t+1|t+1}\tilde{x}_{t+1|t+1}^T|y_0,\ldots,y_{t+1}) \\ \tilde{x}_{t+1|t+1} &= x_{t+1} - \hat{x}_{t+1|t+1} \end{aligned}$$

우선,

$$P(x_{t+1}, y_{t+1}|y_{0:t}) \sim N(\mu, \Sigma),$$
$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}$$

일 때, 다음의 Multivariate Gaussian에서의 조건부 확률 정리를 활용하면,

Theorem 10.2 (Conditional PDF of Multivariate Gaussian) *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are jointly Gaussian, where* $\mathbf{x}$ *is* $k \times 1$ *and* $\mathbf{y}$ *is* $l \times 1$, *with mean vector* $[E(\mathbf{x})^T \; E(\mathbf{y})^T]^T$ *and partitioned covariance matrix*

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix} = \begin{bmatrix} k \times k & k \times l \\ l \times k & l \times l \end{bmatrix} \tag{10.23}$$

so that

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{\frac{k+l}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}\left(\begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}\right)^T \mathbf{C}^{-1}\left(\begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}\right)\right],$$

then the conditional PDF $p(\mathbf{y}|\mathbf{x})$ *is also Gaussian and*

$$E(\mathbf{y}|\mathbf{x}) = E(\mathbf{y}) + \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \tag{10.24}$$
$$\mathbf{C}_{y|x} = \mathbf{C}_{yy} - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}. \tag{10.25}$$

구하고자 하는 $y_{0:t+1}$일 때 x_{t+1}의 확률 분포는 다음과 같이 정리된다.

$$P(x_{t+1}|y_{0:t+1}) \sim N(\mu_{x|y}, \Sigma_{x|y}),$$
$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$$
$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$$

그런데 이 때,

$$\begin{aligned} \mu_y = \hat{y}_{t+1|t} &= E(y_{t+1}|y_{0:t}) \\ &= E(Cx_{t+1} + v_{t+1}|y_{0:t}) = C\hat{x}_{t+1|t} \end{aligned}$$

$$\begin{aligned} \Sigma_{yy} &= E[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})^T|y_{0:t}] \\ &= E[(Cx_{t+1} + v_{t+1} - C\hat{x}_{t+1|t})(Cx_{t+1} + v_{t+1} - C\hat{x}_{t+1|t})^T|y_{0:t}] \\ &= E[(C(x_{t+1} - \hat{x}_{t+1|t}) + v_{t+1})(C(x_{t+1} - \hat{x}_{t+1|t}) + v_{t+1})^T|y_{0:t}] \\ &= CP_{t+1|t}C^T + R \\ \Sigma_{yx} &= E[(y_{t+1} - \hat{y}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T|y_{0:t}] \\ &= E[(Cx_{t+1} + v_{t+1} - C\hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T|y_{0:t}] \\ &= CP_{t+1|t} \end{aligned}$$

이를 이용하여 정리하면,

$$P(x_{t+1}|y_{0:t+1}) \sim N(\hat{x}_{t+1|t+1}, P_{t+1|t+1}),$$
$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}(y_{t+1} - C\hat{x}_{t+1|t})$$
$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}CP_{t+1|t}$$

와 같이 정리된다.

Kalman Filtering의 과정을 정리하면 다음과 같다.

- Dynamic update

$$\begin{aligned}\hat{x}_{t+1|t} &= A\hat{x}_{t|t}\\ P_{t+1|t} &= AP_{t|t}A^T + GQG^T\end{aligned}$$

- Measurement update

$$\begin{aligned}\hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + K_{t+1}(y_{t+1} - C\hat{x}_{t+1|t})\\ P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}CP_{t+1|t}\end{aligned}$$

- Kalman gain

$$K_{t+1} = P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}$$

- Dynamic model: $x_{t+1} = Ax_t + Gw_t$
- Measurement model: $y_t = Cx_t + v_t$
- Initial state distribution: $x_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$
- Noises: $w_t \sim \mathcal{N}(0, Q)$, $v_t \sim \mathcal{N}(0, R)$. w_t, v_t, and x_t are independent.

Ex) particle이 random forces와 damping 아래 평면을 움직인다.

particle의 상태는 $x = [x^1, \dot{x}^1, x^2, \dot{x}^2]^T$이며 (x^1, x^2)는 particle의 위치, $(\dot{x}^1, \dot{x}^2)$는 particle의 속도를 의미한다.

각 상태는 다음 조건을 만족한다.

$$\begin{aligned}x^i_{t+1} &= x^i_t + \dot{x}^i_t\\ \dot{x}^i_{t+1} &= 0.98\dot{x}^i_t + w^i_t\end{aligned}$$

이에 따라 dynamic model과 measurement model을 구할 수 있다.

$$x_{t+1} = \begin{bmatrix} x^1 \\ \dot{x}^1 \\ x^2 \\ \dot{x}^2 \end{bmatrix}_{t+1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0.98 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0.98 \end{bmatrix} \begin{bmatrix} x^1 \\ \dot{x}^1 \\ x^2 \\ \dot{x}^2 \end{bmatrix}_t + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w^1 \\ w^2 \end{bmatrix}_t = Ax_t + Gw_t$$

$$y_t = \begin{bmatrix} y^1 \\ y^2 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x^1 \\ \dot{x}^1 \\ x^2 \\ \dot{x}^2 \end{bmatrix}_t + v_t = Cx_t + v_t$$

3. Other Filtering Algorithms

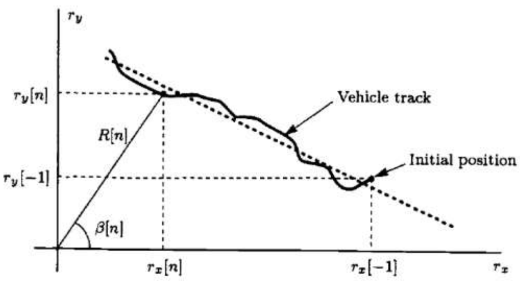
1) Extended Kalman Filters

dynamic 또는 measurement 식이 nonlinear한 경우에는 앞서 구한 식들을 사용할 수 없다. 따라서 이런 경우에는 nonlinear function들을 linearizing함으로써 이전의 식들을 적용할 수 있도록 만든다.

nonlinear한 경우의 예시에는 다음이 있다.

Vehicle tracking example

Vehicle's position: (r_x, r_y)



Nonlinear measurement model

$$\hat{R}[n] = R[n] + w_R[n]$$
$$\hat{\beta}[n] = \beta[n] + w_\beta[n]$$
$$R[n] = \sqrt{r_x^2[n] + r_y^2[n]} \quad \text{Range}$$
$$\beta[n] = \arctan \frac{r_y[n]}{r_x[n]}. \quad \text{Bearing (angle)}$$

우선 다음과 같은 nonlinear system을 가정하자.

$s[n] = a(s[n-1]) + Bu[n]$
$x[n] = h(s[n]) + w[n],$
$u[n] \sim N(0, Q) \quad w[n] \sim N(0, C[n])$

여기서 $a()$와 $h()$는 nonlinear vector valued function이다.
각 nonlinear function들에 대해 First-order Taylor expansion을 전개하면

$$a(s[n-1]) \approx a(\hat{s}[n-1|n-1]) + \left.\frac{\partial a}{\partial s[n-1]}\right|_{s[n-1] = \hat{s}[n-1|n-1]} (s[n-1] - \hat{s}[n-1|n-1])$$
$$h(s[n]) \approx h(\hat{s}[n|n-1]) + \left.\frac{\partial h}{\partial s[n]}\right|_{s[n] = \hat{s}[n|n-1]} (s[n] - \hat{s}[n|n-1])$$

여기서 첫번째 미분 항들을 다음과 같이 정의할 수 있다.

$$A[n-1] = \left.\frac{\partial a}{\partial s[n-1]}\right|_{s[n-1] = \hat{s}[n-1|n-1]}$$
$$H[n] = \left.\frac{\partial h}{\partial s[n]}\right|_{s[n] = \hat{s}[n|n-1]}$$

따라서 nonlinear system을 다음과 같이 linearize할 수 있다.

$s[n] = A[n-1]s[n-1] + Bu[n] + (a(\hat{s}[n-1|n-1]) - A[n-1]\hat{s}[n-1|n-1])$
$x[n] = H[n]s[n] + w[n] + (h(\hat{s}[n|n-1]) - H[n]\hat{s}[n|n-1])$

위 linearized system을 바탕으로 앞서 구한 식에 대입하면 다음과 같이 구할 수 있다.
Prediction :

$\hat{s}[n|n-1] = a(\hat{s}[n-1|n-1])$

Minimum Prediction MSE Matrix :

$M[n|n-1] = A[n-1]M[n-1|n-1]A^T[n-1] + BQB^T$

Kalman Gain Matrix :

$K[n] = M[n|n-1]H^T[n](C[n] + H[n]M[n|n-1]H^T[n])^{-1}$

Correction :

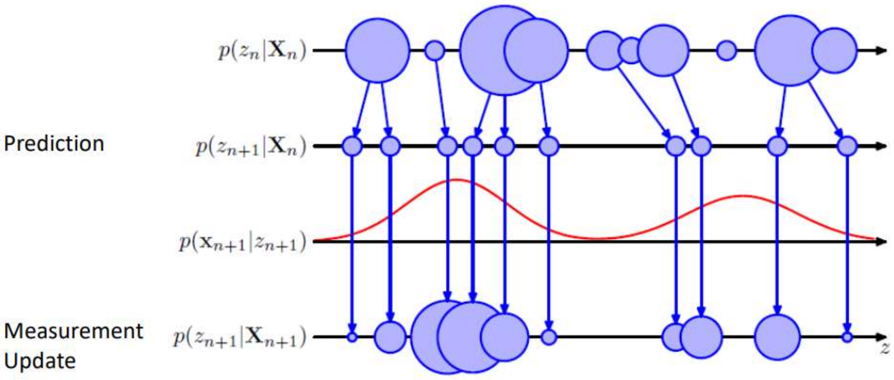
$\hat{s}[n|n] = \hat{s}[n|n-1] + K[n](x[n] - h(\hat{s}[n|n-1]))$

Minimum MSE Matrix :

$M[n|n] = (I - K[n]H[n])M[n|n-1]$

2) Particle Filters

Linear Gaussian을 따르지 않는 dynamical system에 대해 사용할 수 있는 필터이다.



동작 알고리즘은 다음과 같다.

[1] 정해진 구간에 particle을 랜덤으로 분포시킨다.

[2] 각 particle들이 각자의 위치에서 variance * gaussian random[-1, 1]만큼 움직인다고 하고 움직인 이후의 위치를 추정한다.

[3] 예측된 particle들의 위치에 따라 예측된 위치와 센서로 측정된 위치의 차이를 이용해 정규 분포 함수에 따라 각 particle들의 weight을 업데이트한다.

[4] [3]에서 업데이트 된 particle들의 weight별로 weight이 작은 particle은 줄이고, weight이 큰 particle은 늘린다. [4] 과정 전후의 particle의 개수는 같아야 한다.

[5] [2]-[4]를 반복한다.