$$x_t = A x_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, Q)$$

$$y_t = H x_t + \eta_t \qquad \eta_t \sim \mathcal{N}(0, R)$$

$$t = 1, 2, \ldots, T \qquad x_1 \sim \mathcal{N}(\mu_0, Q_0)$$

Prediction :

$$P(x_t \mid y_{1:t-1})$$

Filtering :

$$P(x_t \mid y_{1:t})$$

Smoothing :

$$P(x_t \mid y_{1:T})$$

$$P(x_1) \simeq \mathcal{N}(\mu_0, Q_0)$$

$$P(y_1 \mid x_1) = \mathcal{N}(H x_1, R)$$

$$P(x_1 \mid y_1) = \frac{P(y_1 \mid x_1) \cdot P(x_1)}{\underbrace{\int P(y_1 \mid x_1) \cdot P(x_1)}_{= P(y_1)}}$$

(Bayes Rule).

At $t = 1$

$$p(x_1) = \mathcal{N}(x_1 ; \mu_0, Q_0)$$

$$p(y_1 \mid x_1) = \mathcal{N}(y_1 ; H x_1, R)$$

We know, that

Bayes Rule gaussian

If $p(x) = \mathcal{N}(x ; \mu, \Sigma)$

$p(y/x) = \mathcal{N}(y ; Hx + d, T)$

then

$$p(x \mid y) = \mathcal{N}(\mu_{x|y}, \Sigma_{x|y})$$

where,

$$\Sigma_{x|y} = (\Sigma^{-1} + H^T \Gamma^{-1} H)^{-1}$$

$$\mu_{x|y} = (\Sigma^{-1} + H^T \Gamma^{-1} H)^{-1} \left[H^T \Gamma^{-1} (y - d) + \Sigma^{-1} \mu \right]$$

Plugging this into the problem above:

$$p(x_1 | y_1) = \mathcal{N}(\mu_{1|1}, P_{1|1}), \quad \text{where}$$

$$P_{1|1} = (Q_0^{-1} + C^T R^{-1} C)^{-1}$$

$$\mu_{1|1} = (Q_0^{-1} + C^T R^{-1} C)^{-1} [C^T R^{-1} y_1 + Q_0^{-1} \mu_0]$$

By the Woodbury Matrix Identity:

$$(A + B D^{-1} C)^{-1} = A^{-1} - A^{-1} B (D + C A^{-1} B)^{-1} C A^{-1}$$

Therefore,

$$P_{1|1} = (Q_0^{-1} + C^T R^{-1} C)^{-1} = Q_0 - \underbrace{Q_0 C^T (R + C Q_0 C^T)^{-1}}_{= K} C Q_0.$$

$$= (I - KC) Q_0.$$

Now,

$$(Q_0^{-1} + C^T R^{-1} C)^{-1} C^T R^{-1}$$

$$= (I - KC) Q_0 C^T R^{-1}$$

$$= Q_0 C^T R^{-1} - Q_0 C^T (R + C Q_0 C^T)^{-1} C Q_0 C^T R^{-1}$$

$$= Q_0 C^T \underbrace{\left(I - (R + C Q_0 C^T)^{-1} C Q_0 C^T\right) R^{-1}}_{(I + R^{-1} \cdot C Q_0 C^T)^{-1} R^{-1}}$$

Now,

$$(I + AB)^{-1} A = A(I + BA)^{-1}$$

$$= Q_0 C^T R^{-1} (I + C Q_0 C^T R^{-1})^{-1}$$

$$= Q_0 C^T R^{-1} (R R^{-1} + C Q_0 C^T R^{-1})^{-1}$$

$$= Q_0 C^T R^{-1} \left[(R + C Q_0 C^T) R^{-1} \right]^{-1}$$

$$= Q_0 C^T R^{-1} \left[R (R + C Q_0 C^T)^{-1} \right]$$

$$= Q_0 C^T (R + C Q_0 C^T)^{-1} = K_1$$

Therefore,

$$P_{1|1} = (I - KC) Q_0$$

$$\mu_{1|1} = \underbrace{(Q_0^{-1} + C^T R^{-1} C)^{-1} C^T R^{-1}}_{= K} y_1 + \underbrace{(Q_0^{-1} + C^T R^{-1} C)^{-1}}_{(I - KC) Q_0} Q_0^{-1} \mu_0$$

$$\mu_{1|1} = K_1 y + (I - KC) \cancelto{I}{Q_0 Q_0^{-1}} \mu_0$$

$$= \mu_0 + K_1 (y_1 - \underbrace{C \mu_0}_{\text{predicted mean of } y_1})$$

↳ predicted mean of y_1