HW 2

Ping-Jung Liu

1、(a) $\Sigma_{ij} = E[(x_i-\mu_i)(x_j-\mu_j)] = E[(x_j-\mu_j)(x_i-\mu_i)] = \Sigma_{ji}$

$\Rightarrow$ Symmetric

(b) Let $v \in R^n$

$$\Rightarrow v^T \Sigma v = v^T \cdot E[(x-\mu)(x-\mu)^T] \cdot v = E[\underbrace{v^T(x-\mu)}_{\text{scalar}}\underbrace{(x-\mu)^T v}_{\text{scalar}}]$$

$$= E[(v^T(x-\mu))^2]$$

$\Rightarrow E[(v^T(x-\mu))^2] \geq 0$ (square)

$\Rightarrow v^T \Sigma v \geq 0 \Rightarrow \Sigma$ is positive semi-definite

(c) Mean:

Start from the definition of multivariate Gaussian and take the log-likelihood of the product of m distribution.

$$\Rightarrow -\tfrac{m}{2}\log - \frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$\Rightarrow \partial/\partial\mu = \sum_{i=1}^{m}(x_i-\mu)^T\Sigma^{-1} = 0$$

$$\Rightarrow \mu \approx \frac{1}{m}\sum_{i=1}^{m} x_i \quad \text{(mean of sample)}$$

Covariance:

$$-\frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$= \frac{m}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$\Rightarrow \partial/\partial\Sigma^{-1} = \frac{m}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^T(x_i-\mu) = 0$$ (I admit this part is a little weird; transpose ends up in the wrong place.)

$\Rightarrow$ Theoretically, the result should be

$$\Sigma \approx \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T \quad (\mu \text{ here is the sample mean})$$

(d) Need to find A and B such that

$x = Ay + B$

B is simply μ. $AA^T = \Sigma$

$\Rightarrow A$ could be obtained by taking the eigenvalue decomposition of Σ.

(e) $\eta = \Sigma^{-1}\mu$, $\Lambda = \Sigma^{-1}$

$$N(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \cdot e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}x^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu + x^T\Sigma^{-1}\mu\right)}$$

$$= \underbrace{\frac{1}{(2\pi)^{d/2}|\Lambda^{-1}|^{1/2}} \cdot e^{\left(-\frac{1}{2}x^T\Lambda x - \frac{1}{2}\eta^T\Lambda^{-1}\eta + x^T\eta\right)}}_{N(x;\eta,\Lambda)}$$

2. (a) Since x and v are both gaussians

$$\Rightarrow \hat{z} = H\hat{x}^- \ , \quad \Sigma_z = H\Sigma^- H^T + R$$

(b) $P(z)\cdot P(x|z) = P(x, z) = P(x)\cdot P(z|x)$

Given x, z is entirely determined by v

$$\Rightarrow P(z|x) = P(v) \quad \Rightarrow P(z)\cdot P(x|z) = P(x)\cdot P(v)$$

$$\Rightarrow P(x|z) = \frac{P(x)\cdot P(v)}{P(z)}$$

(c) From (b) we know $P(x|z) \propto P(x)\cdot P(v)$

$$\Rightarrow P(x|z) \propto \exp\left\{-\frac{1}{2}(x-\hat{x}^-)^T \Sigma^{-1} (x-\hat{x}^-)\right\} \cdot \underbrace{\exp\left\{-\frac{1}{2} v^T R^{-1} v\right\}}_{\exp\left\{-\frac{1}{2}(z-Hx)^T R^{-1}(z-Hx)\right\}}$$

$$\Rightarrow P(x|z) \propto \exp\left\{-\frac{1}{2}(x-\hat{x}^-)^T \Sigma^{-1}(x-\hat{x}^-) - \frac{1}{2}(z-Hx)^T R^{-1}(z-Hx)\right\}$$

(d) $p(x|z) \propto \exp\left\{-\frac{1}{2}(x-\hat{x}^-)^T {\Sigma^-}^{-1}(x-\hat{x}^-) - \frac{1}{2}(z-Hx)^T R^{-1}(z-Hx)\right\}$

Similar to the steps in textbook TBF

$\Rightarrow$ Let $J_2 = \frac{1}{2}(x-\hat{x}^-)^T {\Sigma^-}^{-1}(x-\hat{x}^-) + \frac{1}{2}(z-Hx)^T R^{-1}(z-Hx)\}$

$\Rightarrow \partial J/\partial x = {\Sigma^-}^{-1}(x-\hat{x}^-) - H^T R^{-1}(z-Hx) = 0$

$\partial J/\partial x^2 = {\Sigma^-}^{-1} + H^T R^{-1} H$

$\Downarrow$

$\Sigma^+ = ({\Sigma^-}^{-1} + H^T R^{-1} H)^{-1}$

Replace x with the new mean $\hat{x}^+$

$\Rightarrow \cancel{{\Sigma^-}^{-1}(x-\hat{x}^-) - H^T R^{-1}(z-H\hat{x}^+)}\quad {\Sigma^-}^{-1}(\hat{x}^+-\hat{x}^-) = H^T R^{-1}(z-H\hat{x}^+)$

$\Rightarrow H^T R^{-1}(z-H\hat{x}^+) = H^T R^{-1}(z-H\hat{x}^+ + H\hat{x}^- - H\hat{x}^-)$

$= H^T R^{-1}(z-H\hat{x}^-) - H^T R^{-1} H(\hat{x}^+-\hat{x}^-) = {\Sigma^-}^{-1}(\hat{x}^+-\hat{x}^-)$

$\Rightarrow H^T R^{-1}(z-H\hat{x}^-) = \underbrace{H^T R^{-1} H + {\Sigma^-}^{-1}}_{{\Sigma^+}^{-1}}(\hat{x}^+-\hat{x}^-)$

$\Rightarrow \hat{x}^+ = \hat{x}^- + \Sigma^+ H^T R^{-1}(z-H\hat{x}^-)$

$\Sigma^+ = ({\Sigma^-}^{-1} + H^T R^{-1} H)^{-1}$

$p(x|z) \propto \exp\left\{-\frac{1}{2}(x-\hat{x}^+)^T {\Sigma^+}^{-1}(x-\hat{x}^+)\right\}$

(e)

$\Rightarrow$ If we include $p(z)$ from part (a), we will obtain the update equations for kalman filter.
Σ^+ here is the innovation covariance.

3. (a) $\begin{bmatrix} x_{t-1} \\ x_t \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mu_{t-1} \\ \mu_t \end{bmatrix}, \begin{bmatrix} \Sigma_{t-1} & \Sigma_{t-1,t} \\ \Sigma_{t,t-1} & \Sigma_t \end{bmatrix} \right)$

$\Sigma_{t-1} = \sigma^2_{t-1}$

$\mu_{t-1} = \mu_{t-1}$

$\mu_t = A\mu_{t-1} + Bu_t$

$$\Sigma_t = E\left[(x_t - \mu_t)^2 \right]$$

$$= E\left[\left((Ax_{t-1} + Bu_t + v_{t-1}) - (A\mu_{t-1} + Bu_t) \right)^2 \right]$$

$$= E\left[\left((Ax_{t-1} - A\mu_{t-1}) + v_{t-1} \right)^2 \right]$$

~~Since~~ We know $E[vx] = 0$

$$= E\left[(Ax_{t-1} - A\mu_{t-1})^2 \right] + E\left[v_{t-1}^2 \right]$$

$$= A^2 \sigma_{t-1}^2 + R_{t-1}^2$$

$$\Sigma_{t-1,t} = E\left[(x_{t-1} - \mu_{t-1})(x_t - \mu_t) \right]$$

$$= E\left[(x_{t-1} - \mu_{t-1})(Ax_{t-1} - A\mu_{t-1} + v_{t-1}) \right]$$

$$= A \cdot \sigma_{t-1}^2$$

$$= \Sigma_{t,t-1} \text{ for scaler case}$$

$$\Rightarrow \begin{bmatrix} x_{t-1} \\ x_t \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mu_{t-1} \\ \cancel{\mu_t} \end{bmatrix}, \begin{bmatrix} \sigma^2_{t-1} & A \cdot \sigma_{t-1}^2 \\ A \cdot \sigma_{t-1}^2 & A^2 \sigma_{t-1}^2 + R_{t-1}^2 \end{bmatrix} \right)$$

$A\mu_{t-1} + Bu_t$

Margrize out x_{t-1}

$$\Rightarrow \mathcal{N}\left(A\mu_{t-1} + Bu_t, \; A^2 \sigma_{t-1}^2 + R_{t-1}^2 \right)$$

(b) $\begin{bmatrix} x_t \\ z_t \end{bmatrix} = N\left(\begin{bmatrix} \mu_t \\ \bar{z}_t \end{bmatrix}, \begin{bmatrix} \Sigma_{xt} & \Sigma_{xt,zt} \\ \Sigma_{zt,xt} & \Sigma_{zt} \end{bmatrix} \right)$

$\Sigma_{xt} = \sigma_t^2$

$\mu_t = \mu_t$

$\bar{z}_t = H\mu_t$

$$\Sigma_{zt} = E\left[(z_t - \bar{z}_t)^2\right]$$

$$= E\left[((Hx_t + w_t) - H\mu_t)^2\right]$$

$$= H^2 \cdot E\left[(x_t - \mu_t)^2\right] + E\left[w_t^2\right]$$

$$= H^2 \cdot \sigma_t^2 + Q_t$$

$$\Sigma_{xt,zt} = E\left[(x - \mu_t)(z - \bar{z}_t)\right]$$

$$= E\left[(x - \mu_t)(Hx_t - H\mu_t + w_t)\right]$$

$$= H \cdot \sigma_t^2$$

$$\Rightarrow \begin{bmatrix} x_t \\ z_t \end{bmatrix} = N\left(\begin{bmatrix} \mu_t \\ H\mu_t \end{bmatrix}, \begin{bmatrix} \sigma_t^2 & H\sigma_t^2 \\ H\sigma_t^2 & H^2 \cdot \sigma_t^2 + Q_t \end{bmatrix} \right)$$

Now, condition x_t on z_t

$$\Rightarrow N\left(\mu_t + \frac{H\sigma_t^2}{H^2\sigma_t^2 + Q_t}(z - H\mu_t), \; \sigma_t^2 - \frac{H^2\sigma_t^4}{H^2\sigma_t^2 + Q_t} \right)$$

(c)

Prediction step:

$$\Sigma_{t-1} = \Sigma_{t-1} \quad , \mu_{t-1} = \mu_{t-1}, \quad \mu_t = A\mu_{t-1} + Bu_t$$

$$\Rightarrow \Sigma_{t-1,t} = E\left[(x_{t-1} - \mu_{t-1})(x_t - \mu_t)^T \right]$$

$$= \Sigma_{t-1} \cdot A^T$$

I skipped some steps that are extremely similar to (a), (b).

$$\Sigma_{t,t-1} = A\Sigma_{t-1}$$

$$\Sigma_t = E\left[(x_t - \mu_t)(x_t - \mu_t)^T \right]$$

$$= E\left[(A(x_{t-1} - \mu_{t-1}) + v_{t-1})(\quad '' \quad)^T \right]$$

$$= A\Sigma_{t-1}A^T + R_{t-1}$$

$$\Rightarrow \begin{bmatrix} x_{t-1} \\ x_t \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{t-1} \\ A\mu_{t-1} + Bu_t \end{bmatrix}, \begin{bmatrix} \Sigma_{t-1} & \Sigma_{t-1}A^T \\ A\Sigma_{t-1} & A\Sigma_{t-1}A^T + R_{t-1} \end{bmatrix} \right)$$

Marginalize out x_{T-1}

$$\Rightarrow N\left(A\mu_{t-1} + Bu_t, \; A\Sigma_{t-1}A^T + R_{t-1} \right)$$

Update step:

$$\bar{\Sigma}_{x_t} = \Sigma_{x_t}, \quad \bar{\mu}_t = \mu_t, \quad \bar{z}_t = H\mu_t$$

$$\Rightarrow \Sigma_{x_t, z_t} = E\left[(x_t - \mu_t)(z_t - \bar{z}_t)^T\right]$$

$$= \Sigma_{x_t} \cdot H^T$$

$$\Sigma_{z_t, x_t} = H \cdot \Sigma_{x_t}$$

$$\Sigma_{z_t} = E\left[(z_t - \bar{z}_t)(z_t - \bar{z}_t)^T\right]$$

$$= E\left[(H(x_t - \mu_t) + w_t)(\quad " \quad)^T\right]$$

$$= H\Sigma_{x_t}H^T + Q_t$$

$$\Rightarrow \begin{bmatrix} x_t \\ z_t \end{bmatrix} = N\left(\begin{bmatrix} \mu_t \\ H\mu_t \end{bmatrix}, \begin{bmatrix} \Sigma_{x_t} & \Sigma_{x_t}H^T \\ H\Sigma_{x_t} & H\Sigma_{x_t}H^T + Q_t \end{bmatrix}\right)$$

Condition x_t on z_t

$$= N\Big(\mu_t + \Sigma_{x_t}H^T(H\Sigma_{x_t}H^T + Q_t)^{-1}(z - H\mu_t),$$
$$\Sigma_{x_t} - \Sigma_{x_t}H^T(H\Sigma_{x_t}H^T + Q_t)^{-1}H\Sigma_{x_t}\Big)$$

(d) From (c), we know

$$\mu_t = \bar{\mu}_t + \bar{\Sigma}_t H^T (H\bar{\Sigma}_t H^T + Q)^{-1}(z_t - H\bar{\mu}_t)$$
$$\Sigma_t = \bar{\Sigma}_t - \bar{\Sigma}_t H^T (H\bar{\Sigma}_t H^T + Q)^{-1} H\bar{\Sigma}_t$$

$$\Rightarrow \bar{\mu}_{t+1} = A\left(\bar{\mu}_t + \bar{\Sigma}_t H^T (H\bar{\Sigma}_t H^T + Q)^{-1}(z_t - H\bar{\mu}_t)\right) + B u_{t+1}$$

$$\bar{\Sigma}_{t+1} = A\left(\bar{\Sigma}_t - \bar{\Sigma}_t H^T (H\bar{\Sigma}_t H^T + Q)^{-1} H\bar{\Sigma}_t\right) A^T + R$$

There is no steady-state formulation for the means because of z_s and u_s in the equation.

For $\bar{\Sigma}$, it can be found by evaluating $\bar{\Sigma}$ with the following

$$\bar{\Sigma} = A\left(\bar{\Sigma} - \bar{\Sigma} H^T (H\bar{\Sigma} H^T + Q)^{-1} H\bar{\Sigma}\right) A^T + R$$

(e) Since A is a linear operator, the expression of $\bar{\Sigma}_{t+1}$ is simply of the form $A(\ )A^T + R$.

4. (a)

$$F = \begin{bmatrix} 1 & 0 & -dt \cdot \sin(\theta_{t-1}) \\ 0 & 1 & dt \cdot \cos(\theta_{t-1}) \\ 0 & 0 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} 2x_t & 2y_t & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(b) $\mu_T = \begin{bmatrix} 2.61 & -4.481 & 1.48 \end{bmatrix}$

$$\Sigma_T = \begin{bmatrix} 3.24e{-02} & 1.89e{-02} & -4.26e{-13} \\ 1.89e{-02} & 1.1e{-02} & -2.485e^{-13} \\ -4.26e{-13} & -2.485e{-13} & 1.74e{-08} \end{bmatrix}$$

If I ~~changed~~ increased the coefficients of R and Q to 1E-1 and 1E-3 respectively, the initial few estimations would get extremely close to the ground truth.

If I changed those coefficients to 5E-6 and 5E-8 instead, the error plot would show that the errors fell outside the intervals more than half of the time.