For part i) I referenced Essential Aspects of Bayesian Data Imputation by Holt and Nguyen
Holt, William, and Duy Nguyen. "Essential Aspects of Bayesian Data Imputation." Marist College, 2022.

2.5.i)

2.5 - 2.7 $\quad x \sim \mathcal{N}(\bar{x}, P_{xx}),\ y \sim \mathcal{N}(\bar{y}, P_{yy})$

2.5i Let (x, y) be jointly Gaussian random vectors s.t.

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{pmatrix}\right), P_{yy} \succ 0$$

i) show that $p(x|y) \sim \mathcal{N}(\hat{x}, P_{x|y})$
where $\hat{x} = \bar{x} + P_{xy}P_{yy}^{-1}(y - \bar{y})$ and $P_{x|y} = P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}$

$$p(x|y) \sim \mathcal{N}\left(\bar{x} + P_{xy}P_{yy}^{-1}(y-\bar{y}), P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}\right)$$

show $\mathbb{E}(p(x|y)) = \bar{x} + P_{xy}P_{yy}^{-1}(y - \bar{y}) = \hat{x}$

$$= \mathbb{E}\left(\frac{P(x \cap y)}{P(y)}\right)$$

$$P_{xx} - P_{xy}P_{yy}^{-1}P_{yx} = \mathbb{E}\left((x - \hat{x})(x - \hat{x})^T\right) = P_{x|y}$$

$$\mathbb{E}\left(\left((x - \bar{x}) - P_{xy}P_{yy}^{-1}(y - \bar{y})\right)\left((x - \bar{x}) - P_{xy}P_{yy}^{-1}(y - \bar{y})\right)^T\right)$$

$$P_{xx} = \mathbb{E}\left((x - \bar{x})(x - \bar{x})^T\right)$$

Proof: $p(x,y|\hat{x}, P_{x|y}) = \frac{1}{(2\pi)^{(d_1+d_2)/2}|P_{x|y}|^{1/2}} \exp\left(-\frac{1}{2}\begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}^T \begin{pmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{pmatrix}^{-1} \begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}\right)$

Note the following identity is directly obtained

Let $P_{x|y}/P_{yy} = P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}$

$$\begin{pmatrix} I & -P_{xy}P_{yy}^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{pmatrix}\begin{pmatrix} I & 0 \\ -P_{yy}^{-1}P_{yx} & I \end{pmatrix} = \begin{pmatrix} P_{xy}/P_{yy} & 0 \\ 0 & P_{yy} \end{pmatrix} \quad (1)$$

Taking determinant of both sides of (1) we get

$$|P_{x|y}| = \begin{vmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{vmatrix} = |P_{x|y}/P_{yy}||P_{yy}| \qquad (2)$$

inverting both sides of (1).

$$\begin{pmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -P_{yy}^{-1}P_{yx} & I \end{pmatrix} \begin{pmatrix} (P_{x|y}/P_{yy})^{-1} & 0 \\ 0 & P_{yy}^{-1} \end{pmatrix} \begin{pmatrix} I & -P_{xy}P_{yy}^{-1} \\ 0 & I \end{pmatrix}$$

Resulting from this

$$p(x,y|\hat{x},P_{x|y}) \propto \exp\left(-\frac{1}{2}\begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}^T \begin{pmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{pmatrix}^{-1} \begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}\right)$$

$$= \exp\left(-\frac{1}{2}(x-\bar{x}-P_{xy}P_{yy}^{-1}(y-\bar{y}))^T(P_{x|y}/P_{yy})^{-1}(x-\bar{x}-P_{xy}P_{yy}^{-1}(y-\bar{y}))\right)$$

$$\cdot \exp\left(-\frac{1}{2}(y-\bar{y})^T P_{yy}^{-1}(y-\bar{y})\right)$$

From (2) it can be seen

$$(2\pi)^{(d_1+d_2)/2}|P_{x|y}|^{1/2} = (2\pi)^{(d_1+d_2)/2}(|P_{x|y}/P_{yy}||P_{yy}|)^{1/2}$$

$$= (2\pi)^{d_1/2}|P_{x|y}/P_{yy}|^{1/2}(2\pi)^{d_2/2}|P_{yy}|^{1/2}$$

Hence

$$p(x,y|\hat{x},P_{x|y}) = \frac{1}{(2\pi)^{d_1/2}|P_{x|y}/P_{yy}|^{1/2}}$$

$$\cdot \exp\left(-\frac{1}{2}(x-\bar{x}-P_{xy}P_{yy}^{-1}(y-\bar{y}))^T(P_{x|y}/P_{yy})^{-1}(x-\hat{x}-P_{xy}P_{yy}^{-1}(y-\bar{y}))\right)$$

$$\cdot \frac{1}{(2\pi)^{d_2/2}|P_{yy}|^{1/2}}\exp\left(-\frac{1}{2}(y-\bar{y})^T P_{yy}^{-1}(y-\bar{y})\right)$$

$$= p(x|y)\,p(y)$$

Therefore

$$p(x|y) = \frac{1}{(2\pi)^{d_1/2}|P_{x|y}/P_{yy}|^{1/2}}$$

$$\cdot \exp\left(-\frac{1}{2}(x-\bar{x}-P_{xy}P_{yy}^{-1}(y-\bar{y}))^T(P_{x|y}/P_{yy})^{-1}(x-\bar{x}-P_{xy}P_{yy}^{-1}(y-\bar{y}))\right)$$

$$\sim \mathcal{N}(\hat{x}, P_{x|y})$$

where

$$\hat{x} = x + P_{xy}P_{yy}^{-1}(y-\bar{y})$$

$$P_{x|y} = P_{x|y}/P_{yy} = P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}$$

2.5.ii)

ii) Derive Kalman filter formula (2.52) and (2.53)

Specifically, show if $p(x_{t-1}|y_{0:t-1}) \sim \mathcal{N}(\hat{x}_{t-1|t-1}, P_{t-1|t-1})$
then $p(x_t|y_{0:t-1}) \sim \mathcal{N}(\hat{x}_{t|t-1}, P_{t|t-1})$ where $\hat{x}_{t|t-1}$ and $P_{t|t-1}$ are given by (2.52). Then using part i) show that $p(x_t|y_{0:t}) \sim \mathcal{N}(\hat{x}_{t|t}, P_{t|t})$ where $\hat{x}_{t|t}$ and $P_{t|t}$ are given by (2.53)

(2.52) Prediction:

$$\hat{x}_{t|t-1} = A_t \hat{x}_{t-1|t-1} \qquad (2.52a)$$
$$P_{t|t-1} = A_t P_{t-1|t-1} A_t^T + W_t \qquad (2.52b)$$

(2.53) Update:

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + P_{t|t-1} C_t^T (C_t P_{t|t-1} C_t^T + V_t)^{-1} (y_t - C_t \hat{x}_{t|t-1}) \qquad (2.53a)$$
$$P_{t|t} = P_{t|t-1} - P_{t|t-1} C_t^T (C_t P_{t|t-1} C_t^T + V_t)^{-1} C_t P_{t|t-1} \qquad (2.53b)$$

Show if $p(x_{t-1}|y_{0:t-1}) \sim \mathcal{N}(\hat{x}_{t-1|t-1}, P_{t-1|t-1})$
then $p(x_t|y_{0:t-1}) \sim \mathcal{N}(\hat{x}_{t|t-1}, P_{t|t-1})$ where $\hat{x}_{t|t-1}$ from (2.52a)
$P_{t|t-1}$ from (2.52b)

$w_t \sim \mathcal{N}(0, W_t)$

Linear model $x_{t+1} = A_t x_t + w_t$ so it is a linear transform

so $p(A_t x_{t-1|t-1}) \sim \mathcal{N}(A_t x_{t-1|t-1}, A_t P_{t-1|t-1} A_t^T + W_t)$

$x_{t|t-1} = A_t x_{t-1|y_{0:t-1}}$

$\Rightarrow p(x_t|y_{0:t-1}) \sim \mathcal{N}(A_t \hat{x}_{t-1|t-1}, A_t P_{t-1|t-1} A_t^T + W_t)$ ■

covariance adds for gaussian

⊙ Now using part i) show that $p(x_t|y_{0:t}) \sim \mathcal{N}(\hat{x}_{t|t}, P_{t|t})$ where $\hat{x}_{t|t}$ & $P_{t|t}$ given by (2.53)

From i) $\hat{x} = \bar{x} + P_{xy}P_{yy}^{-1}(y-\bar{y})$

$P_{x|y} = P_{x|y}/P_{yy} = P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}$

Want $$\hat{x}_{t|t} = \underbrace{\hat{x}_{t|t-1}}_{\bar{x}} + \underbrace{P_{t|t-1}C_t^T}_{P_{xy}}\underbrace{(C_tP_{t|t-1}C_t^T+V_t)^{-1}}_{P_{yy}^{-1}}\underbrace{(y_t - C_t\hat{x}_{t|t-1})}_{y-\bar{y}}$$

$$\underbrace{P_{t|t}}_{P_{xx}} = P_{t|t-1} - \underbrace{P_{t|t-1}^T C_t^T}_{P_{xy}}\underbrace{(C_tP_{t|t-1}C_t^T+V_t)^{-1}}_{P_{yy}^{-1}}\underbrace{C_tP_{t|t-1}}_{P_{yx}}$$

⇓ directly $p(x_t|y_{0:t}) \sim \mathcal{N}(\hat{x}_{t|t}, P_{t|t})$

Expected value of y is $C_t\hat{x}_{t|t-1}$

because $y_t = C_tx_t + v_t$, $v_t \sim \mathcal{N}(0, V_t)$

$\Rightarrow \mathbb{E}(y_t) = \mathbb{E}(C_tx_t) = C_t\mathbb{E}(x_t) = C_t\hat{x}_{t|t-1}$

$\mathbb{E}((y_t - C_t\hat{x}_{t|t-1})(y_t - C_t\hat{x}_{t|t-1})^T) = C_tP_{t|t-1}C_t^T + V_t$

$= P_{yy}$

How to show $P_{yx} = C_tP_{t|t-1}$?

$= \mathbb{E}((y_t - C_t\hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^T)$

$= \mathbb{E}((C_tx_t - C_t\hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^T)$

$= \mathbb{E}(C_t(x_t - \hat{x}_{t|t-1})(x_t\ \hat{x}_{t|t-1})^T)$

$= C_t\mathbb{E}((x_t - \hat{x}_{t|t-1})(x_t\ \hat{x}_{t|t-1})^T)$

$= C_tP_{t|t-1}$ ■

KALMAN FILTER DERIVED!!