6.5 Consider the time-series model

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}, \qquad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}, \qquad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}),$$

where $\mathbf{w}, \mathbf{v}$ are i.i.d. Gaussian noise variables. Further, assume that $p(\mathbf{x}_0) = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$.

a. What is the form of $p(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T)$? Justify your answer (you do not have to explicitly compute the joint distribution).

Solution.

$p(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T) = p(\mathbf{x}_0) * p(\mathbf{x}_1, \ldots, \mathbf{x}_T | \mathbf{x}_0)$

$= p(\mathbf{x}_0) * p(\mathbf{x}_1 | \mathbf{x}_0) * p(\mathbf{x}_2, \ldots, \mathbf{x}_T | \mathbf{x}_0, \mathbf{x}_1)$

$= p(\mathbf{x}_0) * p(\mathbf{x}_1 | \mathbf{x}_0) * \ldots * p(\mathbf{x}_{T-1} | \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{T-2}) * p(\mathbf{x}_T | \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{T-1})$

$p(\mathbf{x}_0)$ is a Gaussian distribution. Due to properties of linearity, multiplying it by a scalar ($\mathbf{A}$) and adding another Gaussian variable ($\mathbf{w}$) to it results in another Gaussian variable. Therefore, $p(\mathbf{x}_i)$ is a Gaussian distribution for all i.

The product of these Guassian distributions will give us another Gaussian distribution.

Therefore, $p(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T)$ is a Gaussian distribution.

b. Assume that $p(\mathbf{x}_t | \mathbf{y}_1, \ldots \mathbf{y}_t) = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$.

1. Compute $p(\mathbf{x}_{t+1} | \mathbf{y}_1, \ldots \mathbf{y}_t)$.

Solution.

$$p(\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t) = p(\mathbf{A}\mathbf{x}_t + \mathbf{w}|\mathbf{y}_1,\dots\mathbf{y}_t)$$

$\mathbf{A}\mathbf{x}_t + \mathbf{w}$ is a transformation of the random Gaussian variable $\mathbf{x}_t$, and is itself a Gaussian variable.

Using 6.50, the mean of this distribution is

$$\mathbb{E}[\mathbf{A}\mathbf{x}_t + \mathbf{w}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$= \mathbb{E}[\mathbf{A}\mathbf{x}_t|\mathbf{y}_1,\dots\mathbf{y}_t] + \mathbb{E}[\mathbf{w}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$= \mathbf{A}\mathbb{E}[\mathbf{x}_t|\mathbf{y}_1,\dots\mathbf{y}_t] + \mathbb{E}[\mathbf{w}]$$

$$\text{since } \mathbf{w} \text{ is independent of } \mathbf{y}_1,\dots\mathbf{y}_t \text{ and } \mathbf{x}_t$$

$$= \mathbf{A}\boldsymbol{\mu}_t + \mathbf{0}$$

$$= \mathbf{A}\boldsymbol{\mu}_t$$

Using 6.51,

$$\mathbb{V}[\mathbf{A}\mathbf{x}_t + \mathbf{w}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$= \mathbb{V}[\mathbf{A}\mathbf{x}_t|\mathbf{y}_1,\dots\mathbf{y}_t] + \mathbb{V}[\mathbf{w}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$+Cov[\mathbf{A}\mathbf{x}_t|\mathbf{y}_1,\dots\mathbf{y}_t, \mathbf{w}|\mathbf{y}_1,\dots\mathbf{y}_t] + Cov[\mathbf{w}|\mathbf{y}_1,\dots\mathbf{y}_t, \mathbf{A}\mathbf{x}_t|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$= \mathbf{A}\mathbb{V}[\mathbf{x}_t|\mathbf{y}_1,\dots\mathbf{y}_t]\mathbf{A}^T + \mathbb{V}[\mathbf{w}] + \mathbf{0} + \mathbf{0}$$

$$\text{since } \mathbf{w} \text{ is independent of } \mathbf{y}_1,\dots\mathbf{y}_t \text{ and } \mathbf{x}_t$$

$$= \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T + \mathbf{Q}$$

since $\mathbf{w}$ is independent of $\mathbf{y}_1, \ldots \mathbf{y}_t$ and $\mathbf{x}_t$

$$\implies p(\mathbf{x}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t) = \mathcal{N}(\mathbf{A}\boldsymbol{\mu}_t, \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T + \mathbf{Q})$$

2. Compute $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t)$.

   Solution.

   $\mathbf{y}_{t+1}$ is a linear transformation of $\mathbf{x}_{t+1}$, so $p(\mathbf{y}_{t+1}$ is a Gaussian distribution, and from section 6.5.1, we know that marginals and conditionals of Gaussians are Gaussians, so $p(\mathbf{y}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t)$ is also a Gaussian distribution.

$$\mathbb{E}[\mathbf{y}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t] = \mathbb{E}[\mathbf{C}\mathbf{x}_{t+1} + \mathbf{v}|\mathbf{y}_1, \ldots \mathbf{y}_t]$$

$$\implies \mathbb{E}[\mathbf{y}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t] = \mathbb{E}[\mathbf{C}\mathbf{x}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t] + \mathbb{E}[\mathbf{v}|\mathbf{y}_1, \ldots \mathbf{y}_t]$$

$$\implies \mathbb{E}[\mathbf{y}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t] = \mathbf{C}\mathbb{E}[\mathbf{x}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t] + \mathbb{E}[\mathbf{v}] \qquad \text{since } \mathbf{v} \text{ is independent of } \mathbf{y}_1, \ldots \mathbf{y}_t$$

$$\implies \mathbb{E}[\mathbf{y}_{t+1}|\mathbf{y}_1, \ldots \mathbf{y}_t] = \mathbf{C}\mathbf{A}\mathbf{u}_t + \mathbf{0} = \mathbf{C}\mathbf{A}\mathbf{u}_t$$

$$\mathbb{V}[\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t] = \mathbb{V}[\mathbf{C}\mathbf{x}_{t+1} + \mathbf{v}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$\implies \mathbb{V}[\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t] = \mathbb{V}[\mathbf{C}\mathbf{x}_{t+1} + \mathbf{v}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$\implies \mathbb{V}[\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t] = \mathbb{V}[\mathbf{C}\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t] + \mathbb{V}[\mathbf{v}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$+Cov[\mathbf{C}\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t, \mathbf{v}|\mathbf{y}_1,\dots\mathbf{y}_t] + Cov[\mathbf{v}|\mathbf{y}_1,\dots\mathbf{y}_t, \mathbf{C}\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t]$$

$$\implies \mathbb{V}[\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t] = \mathbf{C}\mathbb{V}[\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t]\mathbf{C}^T + \mathbb{V}[\mathbf{v}] + \mathbf{0} + \mathbf{0}$$

since $\mathbf{v}$ is independent of all $\mathbf{x}_i s$ and all $\mathbf{y}_i s$.

$$\implies \mathbb{V}[\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t] = \mathbf{C}(\mathbf{A}\mathbf{\Sigma}_t\mathbf{A}^T + \mathbf{Q})\mathbf{C}^T + \mathbf{R}$$

From 6.64,

$$p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t)$$

$$=$$

$$\mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t} \\ \boldsymbol{\mu}_{\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t} & \mathbf{\Sigma}_{\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t,\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t} \\ \mathbf{\Sigma}_{\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t,\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t} & \mathbf{\Sigma}_{\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t} \end{bmatrix}\right)$$

$$\mathbf{\Sigma}_{\mathbf{x}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t,\mathbf{y}_{t+1}|\mathbf{y}_1,\dots\mathbf{y}_t}$$

$$= \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)(\mathbf{y}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)^T] - \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)]\mathbb{E}[(\mathbf{y}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)]^T$$

$$= \boldsymbol{\Sigma}^T_{\mathbf{y}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t,\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t}$$

$$\mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)(\mathbf{y}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)^T] = \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{x}_t)(\mathbf{C}\mathbf{x}_{t+1} + \mathbf{v}|\mathbf{y}_1,\ldots\mathbf{y}_t)^T]$$

$$= \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{x}_t)(\mathbf{C}\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)^T] + \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{x}_t)(\mathbf{v}|\mathbf{y}_1,\ldots\mathbf{y}_t)^T]$$

$$= \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{x}_t)(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)^T\mathbf{C}^T] + \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{x}_t)(\mathbf{v})^T]$$

$$= \mathbb{E}[(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{x}_t)(\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t)^T]\mathbf{C}^T + \mathbf{0}$$

$$= (\boldsymbol{\Sigma}_t + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T)\mathbf{C}^T$$

$$\boldsymbol{\Sigma}_{\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t,\mathbf{y}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t} = (\boldsymbol{\Sigma}_t + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T)\mathbf{C}^T - (\mathbf{A}\boldsymbol{\mu}_t)(\mathbf{C}\mathbf{A}\mathbf{u}_t)^T$$

$$= (\boldsymbol{\Sigma}_t + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T - \mathbf{A}\boldsymbol{\mu}_t\mathbf{u}_t^T\mathbf{A}^T)\mathbf{C}^T$$

$$\boldsymbol{\Sigma}_{\mathbf{y}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t,\mathbf{x}_{t+1}|\mathbf{y}_1,\ldots\mathbf{y}_t} = ((\boldsymbol{\Sigma}_t + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T - \mathbf{A}\boldsymbol{\mu}_t\mathbf{u}_t^T\mathbf{A}^T)\mathbf{C}^T)^T = \mathbf{C}(\boldsymbol{\Sigma}_t^T + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T - \mathbf{A}\boldsymbol{\mu}_t\mathbf{u}_t^T\mathbf{A}^T)$$

Substituting in the expression for $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}|\mathbf{y}_1, \dots \mathbf{y}_t)$,

$$p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}|\mathbf{y}_1, \dots \mathbf{y}_t) =$$
$$\mathcal{N}\left(\begin{bmatrix} \mathbf{A}\mathbf{u}_t \\ \mathbf{C}\mathbf{A}\mathbf{u}_t \end{bmatrix}, \begin{bmatrix} \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T + \mathbf{Q} & (\boldsymbol{\Sigma}_t + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T - \mathbf{A}\boldsymbol{\mu}_t\mathbf{u}_t^T\mathbf{A}^T)\mathbf{C}^T \\ \mathbf{C}(\boldsymbol{\Sigma}_t^T + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T - \mathbf{A}\boldsymbol{\mu}_t\mathbf{u}_t^T\mathbf{A}^T) & \mathbf{C}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T + \mathbf{Q})\mathbf{C}^T + \mathbf{R} \end{bmatrix}\right)$$

3. At time $t+1$, we observe the value $\mathbf{y}_{t+1} = \hat{\mathbf{y}}$. Compute the conditional distribution $p(\mathbf{x}_{t+1}|\mathbf{y}_1, \dots \mathbf{y}_{t+1})$.

Solution.

Using product rule:

$$p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}|\mathbf{y}_1, \dots \mathbf{y}_t) = p(\mathbf{x}_{t+1}|\mathbf{y}_1, \dots \mathbf{y}_{t+1}) * p(\mathbf{y}_{t+1})$$

$$\implies p(\mathbf{x}_{t+1}|\mathbf{y}_1, \dots \mathbf{y}_{t+1}) = \frac{p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}|\mathbf{y}_1, \dots \mathbf{y}_t)}{p(\mathbf{y}_{t+1})}$$
$$= \frac{1}{p(\hat{\mathbf{y}})}\mathcal{N}\left(\begin{bmatrix} \mathbf{A}\mathbf{u}_t \\ \mathbf{C}\mathbf{A}\mathbf{u}_t \end{bmatrix}, \begin{bmatrix} \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T + \mathbf{Q} & (\boldsymbol{\Sigma}_t + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T - \mathbf{A}\boldsymbol{\mu}_t\mathbf{u}_t^T\mathbf{A}^T)\mathbf{C}^T \\ \mathbf{C}(\boldsymbol{\Sigma}_t^T + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^T - \mathbf{A}\boldsymbol{\mu}_t\mathbf{u}_t^T\mathbf{A}^T) & \mathbf{C}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T + \mathbf{Q})\mathbf{C}^T + \mathbf{R} \end{bmatrix}\right)$$