Exercise sheet 2
Solutions

Exercise 1. Assuming that the two directions of motion are independent, the state equations for the different components of $\boldsymbol{\theta}_k$ can be expressed as follows:

$$\begin{aligned}\boldsymbol{x}_k^{(i)} &= \boldsymbol{x}_{k-1}^{(i)} + \Delta \dot{\boldsymbol{x}}_{k-1}^{(i)} + \boldsymbol{u}_{k,i} \\ \dot{\boldsymbol{x}}_k^{(i)} &= \dot{\boldsymbol{x}}_{k-1}^{(i)} + \boldsymbol{u}_{k,i+2},\end{aligned}$$

for any $i \in \{1, 2\}$, which can be expressed more concisely in a matrix form (following the same approach for the noise) as $\boldsymbol{\theta}_k = F\boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k$ with $\boldsymbol{u}_k \sim \mathrm{N}(\cdot\,; 0, U)$ and with

$$F = \begin{pmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad U = \begin{pmatrix} \Delta^4/4 & 0 & \Delta^3/2 & 0 \\ 0 & \Delta^4/4 & 0 & \Delta^3/2 \\ \Delta^3/2 & 0 & \Delta^2 & 0 \\ 0 & \Delta^3/2 & 0 & \Delta^2 \end{pmatrix}.$$

Similarly, assuming that the position is observed, the observation equation is $\boldsymbol{y}_k = H\boldsymbol{\theta}_k + \boldsymbol{v}_k$ with the observation matrix H defined as

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

and the noise can be assumed to be of the form $\boldsymbol{v}_k \sim \mathrm{N}(\cdot\,; 0, \sigma^2 I_2)$.

Exercise 2. 1. The state and observation equations are $\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k$ and $\boldsymbol{y}_k = \boldsymbol{\theta}_k + \boldsymbol{v}_k$ with $(\boldsymbol{u}_k)_k$ and $(\boldsymbol{v}_k)_k$ i.i.d. sequences of random variables distributed according to $\mathrm{N}(\cdot\,; 0, U)$ and $\mathrm{N}(\cdot\,; 0, V)$ respectively.

2. The prediction of the Kalman filter becomes $m_k = \hat{m}_{k-1}$ and $P_k = \hat{P}_{k-1} + U$, the innovation, Kalman gain and covariance of the innovation become

$$z_k = y_k - m_k, \qquad K_k = \frac{P_k}{S_k}, \qquad \text{and} \qquad S_k = P_k + V$$

so that the update step can be simplified to

$$\begin{aligned}\hat{m}_k &= m_k + K_k z_k = \frac{V}{P_k + V} m_k + \frac{P_k}{P_k + V} y_k \\ \hat{P}_k &= (1 - K_k) P_k = \frac{V P_k}{P_k + V}\end{aligned}$$

This is indeed what was found in Section 1.3.1 in the case of an unknown mean and a known variance.

3. The predictive distribution of $\boldsymbol{y}_k$ given $\boldsymbol{y}_{0:k-1} = y_{0:k-1}$ is normal (as a linear combination of normally-distributed random variables) with mean m_k and covariance $P_k + V$. The latter is the same as the covariance of the innovation, which is not surprising since the (random) innovation $\boldsymbol{z}_k$ is simply $\boldsymbol{y}_k$ shifted by a constant ($-m_k$), so they have the same covariance matrix.

Exercise 3. 1. From the previous exercise, we easily compute that

$$\begin{aligned}\hat{P}_k &= \frac{V P_k}{P_k + V} = \frac{V \hat{P}_{k-1} + V U}{\hat{P}_{k-1} + U + V} \\ K_k &= \frac{P_k}{P_k + V} = \frac{\hat{P}_{k-1} + U}{\hat{P}_{k-1} + U + V}\end{aligned}$$

Since it holds that $\hat{P}_{k-1} = VK_{k-1}$, it follows that

$$K_k = \frac{VK_{k-1} + U}{VK_{k-1} + U + V}.$$

These equations can be expressed more concisely as

$$\begin{aligned} \hat{P}_k^{-1} &= (P_k + V)^{-1} + V^{-1} \\ K_k^{-1} &= \Big(K_{k-1} + \frac{U}{V}\Big)^{-1} + 1 \end{aligned}$$

2. Denoting R the ratio U/V, it holds that

$$\begin{aligned} K_k - K_{k-1} &= K_k K_{k-1}(K_{k-1}^{-1} - K_k^{-1}) \\ &= K_k K_{k-1}\big((K_{k-2} + R)^{-1} - (K_{k-1} + R)^{-1}\big) \\ &= \frac{K_k K_{k-1}(K_{k-1} - K_{k-2})}{(K_{k-1} + R)(K_{k-2} + R)} \end{aligned}$$

so that

$$\begin{aligned} \frac{K_k - K_{k-1}}{K_{k-1} - K_{k-2}} &= \frac{K_k K_{k-1}}{(K_{k-1} + R)(K_{k-2} + R)} \\ &= K_k K_{k-1}(K_k^{-1} - 1)(K_{k-1}^{-1} - 1) \\ &= (1 - K_k)(1 - K_{k-1}) \in (0, 1) \end{aligned}$$

since $K_k \in (0, 1)$ for any $k \geq 0$. The lower bound implies that the sequence is monotonic and the upper bound implies that the increment becomes smaller with k, so that the sequence as a unique limit K.

Exercise 4. The distribution of $\boldsymbol{\theta}_{k+1}$ given $\boldsymbol{y}_{0:k-1} = y_{0:k-1}$ is also normal with mean

$$\mathbb{E}(\boldsymbol{\theta}_{k+1} \,|\, \boldsymbol{y}_{0:k-1} = y_{0:k-1}) = \mathbb{E}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k-1} = y_{0:k-1}) = \hat{m}_{k-1}$$

and variance

$$\mathrm{var}(\boldsymbol{\theta}_{k+1} \,|\, \boldsymbol{y}_{0:k-1} = y_{0:k-1}) = \mathrm{var}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k-1} = y_{0:k-1}) + U = \hat{P}_{k-1} + 2U$$

In these conditions, the Kalman gain at time step $k + 1$ without update at time k becomes

$$\tilde{K}_{k+1} = \frac{\hat{P}_{k-1} + 2U}{\hat{P}_{k-1} + 2U + V},$$

This alternative Kalman gain is greater than what would have been K_k if y_k had been available, so that the influence of y_{k+1} will be greater. This is to be expected since there is more uncertainty when observations are missing.

Exercise 5. To answer this question, it is convenient to rewrite the state $\boldsymbol{\theta}_k$ as

$$\boldsymbol{\theta}_{k+i} = \boldsymbol{\theta}_k + \sum_{j=1}^{i} \boldsymbol{u}_{k+j}$$

for any $i > 0$, so that

$$\boldsymbol{y}_{k+i} = \boldsymbol{\theta}_{k+i} + \boldsymbol{v}_{k+i} = \boldsymbol{\theta}_k + \boldsymbol{v}_{k+i} + \sum_{j=1}^{i} \boldsymbol{u}_{k+j}$$

The aggregate future sales $A_{k,\delta}$ can then be expressed as

$$\begin{aligned}
A_{k,\delta} &= \sum_{i=1}^{\delta} \boldsymbol{y}_{k+i} \\
&= \sum_{i=1}^{\delta} \Big(\boldsymbol{\theta}_k + \boldsymbol{v}_{k+i} + \sum_{j=1}^{i} \boldsymbol{u}_{k+j} \Big) \\
&= \delta \boldsymbol{\theta}_k + \sum_{i=1}^{\delta} \boldsymbol{v}_{k+i} + \sum_{i=1}^{\delta} \sum_{j=1}^{i} \boldsymbol{u}_{k+j} \\
&= \delta \boldsymbol{\theta}_k + \sum_{i=1}^{\delta} \boldsymbol{v}_{k+i} + \sum_{i=1}^{\delta} (\delta - i + 1) \boldsymbol{u}_{k+i}.
\end{aligned}$$

Since $A_{k,\delta}$ is defined as a linear combination of normally-distributed random variables, it also follows a normal distribution with mean

$$\mathbb{E}(A_{k,\delta} \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) = \delta \mathbb{E}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) = \delta \hat{m}_k$$

and variance

$$\begin{aligned}
\operatorname{var}(A_{k,\delta} \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) &= \delta^2 \operatorname{var}(\boldsymbol{\theta}_k \,|\, \boldsymbol{y}_{0:k} = y_{0:k}) + \sum_{i=1}^{\delta} \operatorname{var}(\boldsymbol{v}_{k+i}) + \sum_{i=1}^{\delta} (\delta - i + 1)^2 \operatorname{var}(\boldsymbol{u}_{k+i}) \\
&= \delta^2 \hat{P}_k + \delta V + U \sum_{i=1}^{\delta} i^2 \\
&= \delta^2 \hat{P}_k + \delta V + \frac{1}{6} \delta (\delta + 1)(2\delta + 1) U.
\end{aligned}$$