An innovations sequence and its applications. Let $[Y_1 \ Y_2 \ Y_3 \ X]^\top$ be a zero-mean random vector with covariance matrix

$$
\begin{bmatrix}
1 & 0.5 & 0.5 & 0 \\
0.5 & 1 & 0.5 & 0.25 \\
0.5 & 0.5 & 1 & 0.25 \\
0 & 0.25 & 0.25 & 1
\end{bmatrix}
$$

a. Let $\tilde{Y} = [\tilde{Y}_1 \ \tilde{Y}_2 \ \tilde{Y}_3]^\top$ be the innovations sequence of $Y = [Y_1 \ Y_2 \ Y_3]^\top$. Find the matrix $A$ such that

$$
\tilde{Y} = AY.
$$

b. Find the covariance matrix of $\tilde{Y}$ and the cross-covariance matrix of $X$ and $\tilde{Y}$.

c. Find the constants $a, b$, and $c$ that minimize $E[(X - a\tilde{Y}_1 - b\tilde{Y}_2 - c\tilde{Y}_3)^2]$.

**Solution** (10 points)

a. The causal best MSE linear estimates are

$$
\hat{Y}_1 = 0 \\
\hat{Y}_2(Y_1) = \frac{\text{Cov}(Y_1, Y_2)}{\text{Var}(Y_1)}Y_1 = \frac{1}{2}Y_1 \\
\hat{Y}_3(Y_1, Y_2) = \Sigma_{Y_3Y_{12}}\Sigma^{-1}_{Y_{12}}Y_{12} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\
-\frac{1}{2} & 1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\
Y_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y_1 \\
Y_2 \end{bmatrix},
$$

where $Y_{12}$ is shorthand for the vector $[Y_1 \ Y_2]^\top$. The innovations are given as

$$
\tilde{Y}_i = Y_i - \hat{Y}_i(Y_1, \ldots, Y_{i-1}),
$$
and therefore
\[
\begin{bmatrix}
\tilde{Y}_1 \\
\tilde{Y}_2 \\
\tilde{Y}_3
\end{bmatrix}
=\begin{bmatrix}
1 & 0 & 0 \\
-\frac{1}{3} & 1 & 0 \\
-\frac{1}{3} & -\frac{1}{3} & 1
\end{bmatrix}
\begin{bmatrix}
Y_1 \\
Y_2 \\
Y_3
\end{bmatrix}.
\]

b. Using the matrix \( A \) from part (a), we have
\[
\Sigma_{\tilde{Y}} = A \Sigma_Y A^T
\]
\[
= \begin{bmatrix}
1 & 0 & 0 \\
0 & \frac{3}{4} & 0 \\
0 & 0 & \frac{2}{3}
\end{bmatrix},
\]
\[
\Sigma_{XY} = E[XY^T]
\]
\[
= E[XY^T]A^T
\]
\[
= \Sigma_{XY} A^T
\]
\[
= \begin{bmatrix}
0 & \frac{1}{4} & \frac{1}{6}
\end{bmatrix}.
\]

The covariance matrix of \( \tilde{Y} \) is diagonal, since the innovations are uncorrelated among each other by definition.

Remark: The matrix \( A \) is related to the lower-triangular whitening matrix \( W \) of \( \Sigma_Y \). Let us compare the two:
\[
A = \begin{bmatrix}
1 & 0 & 0 \\
-\frac{1}{3} & 1 & 0 \\
-\frac{1}{3} & -\frac{1}{3} & 1
\end{bmatrix},
\]
\[
\Rightarrow A \Sigma_Y A^T = \begin{bmatrix}
1 & 0 & 0 \\
0 & \frac{3}{4} & 0 \\
0 & 0 & \frac{2}{3}
\end{bmatrix},
\]
\[
W = \begin{bmatrix}
1 & 0 & 0 \\
-1/\sqrt{3} & 2/\sqrt{3} & 0 \\
-1/\sqrt{6} & -1/\sqrt{6} & \sqrt{3}/2
\end{bmatrix},
\]
\[
\Rightarrow W \Sigma_Y W^T = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}.
\]

Here, the whitening matrix is computed as \( W = L^{-1} \), where \( L \) is from the Cholesky decomposition \( \Sigma_Y = LL^T \). Observe that both transformations \( A \) and \( W \) are lower triangular and diagonalize the covariance matrix \( \Sigma_Y \). Among all such transformations, \( W \) is chosen such that the diagonal elements of \( W \Sigma_Y W^T \) are equal to one, whereas \( A \) is chosen such that its own diagonal elements are equal to one.

An alternative way to construct \( A \) is to first compute \( W \) and then scale each row.
such that the diagonal becomes one.

c. The best MSE linear estimator of $X$ given $\tilde{Y}$ is

$$\hat{X}(\tilde{Y}) = \Sigma_{X\tilde{Y}} \Sigma_{\tilde{Y}}^{-1} \tilde{Y}$$

$$= \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \tilde{Y}_3 \end{bmatrix},$$

where we have used the results from part (b). The needed constants are thus $a = 0$, $b = \frac{1}{3}$, and $c = \frac{1}{4}$.

2. Cellphone. We aim to design a cellphone which is able to denoise a signal modeled as $Y_1 = V + Z$, where $V$ is a random variable representing the user’s voice and $Z$ is a random variable representing background noise. An extra microphone measures the background. However this measurement also includes some distorted voice signal. This is taken into account by modeling it as $Y_2 = Z + U$, where $U$ is a random variable representing the distortion. Assume that $V$, $Z$ and $U$ are all zero mean, both $(V, Z)$ and $(U, Z)$ are uncorrelated and $\text{Corr}(U, V) = \rho$. We also know that $\text{Var}(V) = P$, $\text{Var}(Z) = N$ and $\text{Var}(U) = Q$.

We decide to obtain a linear estimate of $V$ from $Y_1$ and $Y_2$.

Figure 1: Illustration of the system
a. What is the innovation sequence $\tilde{Y}_1$ and $\tilde{Y}_2$ corresponding to $Y_1$ and $Y_2$?

b. What is the linear MMSE estimate of $V$ given the measurements expressed as a function of $\tilde{Y}_1$, $\tilde{Y}_2$, $P$, $Q$ and $N$?

c. What is the corresponding MSE in terms of $P$, $Q$ and $N$?

**Solution** (15 points)

a. The innovation sequence is equal to

$$
\tilde{Y}_1 = Y_1 \\
\tilde{Y}_2 = Y_2 - \tilde{Y}_2(Y_1) = Y_2 - \frac{\text{Cov}(Y_1,Y_2)}{\text{Var}(Y_1)} Y_1 = Y_2 - \frac{N + \rho\sqrt{PQ}}{N + P} Y_1.
$$

b. The innovation sequence is uncorrelated by construction, so we can compute the linear MMSE estimates of $V$ given $\tilde{Y}_1$ and $\tilde{Y}_2$ separately and then add them up. It will be helpful to note that

$$
\tilde{Y}_2 = \frac{P - \rho\sqrt{PQ}}{N + P} Z + U - \frac{N + \rho\sqrt{PQ}}{N + P} V.
$$

Using this we obtain

$$
\hat{V}(\tilde{Y}_1) = \frac{\text{Cov}(\tilde{Y}_1,V)}{\text{Var}(\tilde{Y}_1)} \tilde{Y}_1 = \frac{\text{Cov}(Y_1,V)}{\text{Var}(Y_1)} \tilde{Y}_1 = \frac{P}{N + P} \tilde{Y}_1,
$$

$$
\hat{V}(\tilde{Y}_2) = \frac{\text{Cov}(\tilde{Y}_2,V)}{\text{Var}(\tilde{Y}_2)} \tilde{Y}_2
$$

$$
= \frac{\rho - \frac{(N + \rho\sqrt{PQ})P}{N + P}}{\left(\frac{P - \rho\sqrt{PQ}}{N + P}\right)^2 N + Q + \left(\frac{N + \rho\sqrt{PQ}}{N + P}\right)^2 P - 2\frac{N + \rho\sqrt{PQ}}{N + P} \rho\sqrt{PQ}} \tilde{Y}_2
$$

$$
= \frac{\rho(N + P)^2 - (N + \rho\sqrt{PQ})P(N + P)}{N(P - \rho\sqrt{PQ})^2 + Q(N + P)^2 + (N + \rho\sqrt{PQ})^2 P - 2(N + \rho\sqrt{PQ})(N + P)\rho\sqrt{PQ}} \tilde{Y}_2.
$$

The final estimate is

$$
\hat{V}(\tilde{Y}_1,\tilde{Y}_2) = \frac{P}{N + P} \tilde{Y}_1 - \frac{\rho(N + P)^2 - (N + \rho\sqrt{PQ})P(N + P)}{N(P - \rho\sqrt{PQ})^2 + Q(N + P)^2 + (N + \rho\sqrt{PQ})^2 P - 2(N + \rho\sqrt{PQ})(N + P)\rho\sqrt{PQ}} \tilde{Y}_2.
$$
c. The MSE is
\[
MSE = \text{Var}(V) - \frac{\text{Cov}^2(Y_1, V)}{\text{Var}(Y_1)} - \frac{\text{Cov}^2(Y_2, V)}{\text{Var}(Y_2)}
\]
\[
= P - \frac{P^2}{N + P} - \frac{\left(\rho - \frac{(N + \rho\sqrt{PQ})P}{N + P}\right)^2}{N + Q + \frac{(N + \rho\sqrt{PQ})^2}{N + P}P - 2 \frac{N + \rho\sqrt{PQ}}{N + P} \rho \sqrt{PQ}}
\]
\[
= P - \frac{P^2}{N + P} - \frac{(\rho(N + P) - NP + \rho P \sqrt{PQ})^2}{N(P - \rho \sqrt{PQ})^2 + Q(N + P)^2 + (N + \rho \sqrt{PQ})^2P - 2(N + \rho \sqrt{PQ})(N + P)\rho \sqrt{PQ}}
\]

3. Vector Kalman filter experiment. Consider the state space model
\[
X_{i+1} = \begin{bmatrix} 1 & 1 \\ 0 & 0.9 \end{bmatrix} X_i + U_i, \quad \text{for } i = 1, \ldots, n.
\]
The first and second component of \(X_i\) are the one-dimensional position and velocity of a moving object. In each step, the position and velocity evolve according to Newton's laws of physics. Due to friction, the velocity is dampened by a factor of 0.9 in each time step. The state is disturbed by independent noise vectors \(U_i\), distributed according to
\[
U_i \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 10 \end{bmatrix}\right), \quad \text{for } i = 1, \ldots, n.
\]
The initial state is independent of the noise vectors \(U_i\) and distributed according to
\[
X_1 \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1000 & 0 \\ 0 & 0 \end{bmatrix}\right)
\]
The observations are
\[
Y_i = X_i + V_i, \quad \text{for } i = 1, \ldots, n + 1,
\]
where the noise \(V_i\) is distributed according to
\[
V_i \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1000 & 0 \\ 0 & 1 \end{bmatrix}\right), \quad \text{for } i = 1, \ldots, n + 1,
\]
independent of the initial state \(X_1\) and the state noise vectors \(U_i\).
Download the posted MATLAB file \texttt{vectorKalman.m} and complete the code to compute the Kalman prediction filter \(\hat{X}_{i+1|j}\) from the observations \(Y^j\), for \(i = 0, \ldots, n+1\). For a single realization with time horizon \(n = 100\), plot the true state and its prediction over the time index \(i\). In a separate figure, plot the prediction error over \(i\). Hand in your MATLAB code and the plots.

\textbf{Solution} (10 points)
function [Xhatp,SigmaXhatp] = vectorKalmanPredictor(A,P1,Q,N,Y) % inputs:
% A(sxsxn)-systemmatrixfortimes1:n
% P1 (s x s) - covariance matrix of state X_1
% Q (s x s x n) - covariance matrix of state noise at times 1:n
% N (s x s x n) - covariance matrix of observation noise at times 1:n
% Y (s x n) - observations for times 1:n % outputs:
% Xhatp (s x n+1) - predicted state.
% hatXp(:,i+1) is Xhat(i+1|i)
% hatXp(:,1) is Xhat(1|0)
% hatXp(:,2) is Xhat(2|1), etc.
% SigmaXhatp (s x s x n) - MSE matrix of the prediction
% SigmaXhatp(:,i+1) is Sigma(i+1|i)
% s = size(A,1);
% n = size(A,3);
% Xhatp = zeros(s,n+1); SigmaXhatp = zeros(s,s,n+1); K = zeros(s,s,n);
% for i=1:n
% K(:,:,i) = A(:,:,i)*SigmaXhatp(:,:,i)*inv(SigmaXhatp(:,:,i)+N(:,:,i));
% SigmaXhatp(:,:,i+1) = A(:,:,i)*SigmaXhatp(:,:,i)*...
% (eye(s) - inv(SigmaXhatp(:,:,i)+N(:,:,i))* SigmaXhatp(:,:,i))*...
% A(:,:,i) + Q(:,:,i);
% Xhatp(:,i+1) = A(:,:,i)*Xhatp(:,i) + K(:,:,i)*(Y(:,i) - Xhatp(:,i));
% end
end

The plots generated by the code are shown in Figures 2 and 2.

Figure 2: Vector Kalman filter experiment: True and predicted state.
4. **The filtering version of the Kalman filter** We have studied the derivation of scalar Kalman filter to predict the next state in class. Now we are interested in estimating the current state. Derive the update equations and the error:

\[
\hat{X}_{i+1|i+1} = a_i (1 - k_i) \hat{X}_{i|i} + k_i Y_{i+1}
\]

where

\[
k_i = \frac{a_i^2 \sigma_{ii}^2 + Q_i}{a_i^2 \sigma_{ii}^2 + Q_i + N_{i+1}}
\]

and

\[
\sigma_{i+1|i+1} = (1 - k_i)(a_i^2 \sigma_{ii}^2 + Q_i).
\]

**Solution** (15 points)

***SOLUTIONS NOTE:*** Depending on how the Kalman filter is derived during class, it is possible that \( Q_i \) may be replaced by \( Q_{i+1} \) in the derivations/outcomes.

Assume we know the estimate \( \hat{X}_{i|i} \) and its MSE \( \sigma_{ii}^2 \). Now we receive a new observation \( Y_{i+1} \) and would like to compute the updated estimate \( \hat{X}_{i+1|i+1} \) and its MSE \( \sigma_{i+1|i+1}^2 \).

First, we consider the estimate (prediction) of the state at time \( i + 1 \) without new
observations. We have
\[
\hat{X}_{i+1|i} = a_i \hat{X}_{i|i},
\]
\[
\sigma_{i+1|i}^2 = a_i^2 \sigma_{i|i}^2 + Q_i.
\]
When we observe \(Y_{i+1}\), we are only interested in its innovation \(\tilde{Y}_{i+1}\), i.e., the part that was not predictable from the previous observations. It is given as
\[
\tilde{Y}_{i+1} = Y_{i+1} - \hat{Y}_{i+1|i}.
\]
The predictable part \(\hat{Y}_{i+1|i}\), in turn, is simply
\[
\hat{Y}_{i+1|i} = \hat{X}_{i+1|i} = a_i \hat{X}_{i|i},
\]
as computed above. The new state estimate is
\[
\hat{X}_{i+1|i+1} = \hat{X}_{i+1|i} + k_i \tilde{Y}_{i+1},
\]
because \(X_{i+1|i}\) is a function of \(Y_i\) and therefore orthogonal to \(\tilde{Y}_{i+1}\). Substituting, we obtain
\[
\hat{X}_{i+1|i+1} = a_i \hat{X}_{i|i} + k_i (Y_{i+1} - a_i \hat{X}_{i|i})
= a_i (1 - k_i) \hat{X}_{i|i} + k_i Y_{i+1}.
\] (1)
The coefficient \(k_i\) is computed as
\[
k_i = \frac{\text{Cov}(X_{i+1}, \tilde{Y}_{i+1})}{\text{Var}(\tilde{Y}_{i+1})},
\]
where
\[
\text{Cov}(X_{i+1}, \tilde{Y}_{i+1}) = \text{Cov}(a_i X_i + U_i, Y_{i+1} - a_i \hat{X}_{i|i})
= \text{Cov}(a_i X_i + U_i, a_i X_i + U_i + V_{i+1} - a_i \hat{X}_{i|i})
= a_i^2 \text{Cov}(X_i, X_i - \hat{X}_{i|i}) + Q_i
= a_i^2 \sigma_{i|i}^2 + Q_i,
\]
and
\[
\text{Var}(\tilde{Y}_{i+1}) = \text{Var}(X_{i+1} + V_{i+1} - a_i \hat{X}_{i|i})
= \text{Var}(aX_i + U_i + V_{i+1} - a_i \hat{X}_{i|i})
= Q_i + N_{i+1} + a_i^2 \sigma_{i|i}^2.
\]
Substituting, we have
\[
k_i = \frac{a_i^2 \sigma_{i|i}^2 + Q_i}{a_i^2 \sigma_{i|i}^2 + Q_i + N_{i+1}}.
\] (2)
Finally, the updated MSE is
\[
\sigma_{i+1|i+1}^2 = \sigma_{i+1|i}^2 - \frac{\text{Cov}(X_{i+1}, \tilde{Y}_{i+1})^2}{\text{Var}(Y_{i+1})}
\]
\[
= \sigma_{i|i}^2 + Q_i - \frac{(a_i^2 \sigma_{i|i}^2 + Q_i)^2}{a_i^2 \sigma_{i|i}^2 + Q_i + N_{i+1}}
\]
\[
= (1 - k_i)(a_i^2 \sigma_{i|i}^2 + Q_i).
\]
Equations (1), (2), and (3) were what we set out to show.

5. Absolute value random walk. Let \(X_n\) be a random walk defined by \(X_0 = 0, X_n = \sum_{i=1}^n Z_i, n \geq 1\), where \(\{Z_i\}\) is an i.i.d. process with \(P\{Z_1 = -1\} = P\{Z_1 = +1\} = \frac{1}{2}\). Define the absolute value random process \(Y_n = |X_n|\).

a. Find \(P\{Y_n = k\}\).

b. Find \(P\{\max\{Y_i : 1 \leq i < 20\} = 10 | Y_{20} = 0\}\).

**Solution** (10 points)

a. If \(k \geq 0\) then
\[
P\{Y_n = k\} = P\{X_n = +k \text{ or } X_n = -k\}.
\]
If \(k > 0\) then \(P\{Y_n = k\} = 2P\{X_n = k\}, \) while \(P\{Y_n = 0\} = P\{X_n = 0\}\). Thus
\[
P\{Y_n = k\} = \begin{cases} 
\left(\frac{n}{(n+k)/2}\right) \left(\frac{1}{2}\right)^{(n-1)} & k > 0, n - k \text{ is even, } n - k \geq 0 \\
\left(\frac{n}{(n/2)}\right) \left(\frac{1}{2}\right)^n & k = 0, n \text{ is even, } n \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

b. If \(Y_{20} = |X_{20}| = 0\) then there are only two sample paths with \(\max_{1 \leq i < 20} |X_i| = 10\). These two paths are shown in Figure 4. Since the total number of sample paths is \(\binom{20}{10}\) and all paths are equally likely,
\[
P\left\{\max_{1 \leq i < 20} Y_i = 10 | Y_{20} = 0\right\} = \frac{2}{\binom{20}{10}} = \frac{2}{184756} = \frac{1}{92378}.
\]

6. Poisson process branching. Let \(N(t)\) be a Poisson process with rate \(\lambda\). We split \(N(t)\) into two counting subprocesses \(N_1(t)\) and \(N_2(t)\) such that \(N(t) = N_1(t) + N_2(t)\) as follows: each event is randomly and independently assigned to process \(N_1(t)\) with probability \(p\), otherwise it is assigned to \(N_2(t)\). Prove that \(N_1(t)\) is a Poisson process with rate \(p\lambda\) and \(N_2(t)\) is a Poisson process with rate \((1-p)\lambda\).

**Solution** (10 points)

*Proof.* We calculate the joint PMF for \(N_1(t), N_2(t)\) for arbitrary \(t\). Note that when conditioned on the number of arrivals \(N(t)\), the joint PMF for \(N_1(t), N_2(t)\) is binomial
Figure 4: Sample paths for problem 5

i.e.

\[ P\{N_1(t) = m, N_2(t) = k | N(t) = m + k\} = \binom{m+k}{m} p^m (1-p)^k \]

because each event is randomly and independently assigned to one of the processes.

Hence,

\[ P\{N_1(t) = m, N_2(t) = k \} = P\{N_1(t) = m, N_2(t) = k | N(t) = m + k\} P\{N(t) = m + k\} \]

\[ = \binom{m+k}{m} p^m (1-p)^k \frac{(\lambda t)^m k e^{-\lambda t}}{(m+k)!} \]

\[ = \frac{p^m (1-p)^k \lambda t^{m+k} e^{-\lambda t}}{m! k!} \]

\[ = \frac{(p \lambda t)^m e^{-p \lambda t}}{m!} \frac{[(1-p) \lambda t]^k e^{-(1-p) \lambda t}}{k!} \]

\[ = P\{N_1(t) = m\} P\{N_2(t) = k\}. \]

Thus, we conclude that \( N_1(t) \) is a Poisson process with rate \( p \lambda \) and \( N_2(t) \) is a Poisson process with rate \( (1-p) \lambda \).

7. Markov processes. Let \( \{X_n\} \) be a discrete-time continuous-valued Markov random process, that is,

\[ f(x_{n+1}|x_1, x_2, \ldots, x_n) = f(x_{n+1}|x_n) \]

for every \( n \geq 1 \) and for all sequences \( (x_1, x_2, \ldots, x_{n+1}) \).

a. Show that \( f(x_1, \ldots, x_n) = f(x_1)f(x_2|x_1) \cdots f(x_n|x_{n-1}) = f(x_n)f(x_{n-1}|x_n) \cdots f(x_1|x_2) \).

b. Show that \( f(x_n|x_1, x_2, \ldots, x_k) = f(x_n|x_k) \) for every \( k \) such that \( 1 \leq k < n \).

c. Show that \( f(x_{n+1}, x_{n-1}|x_n) = f(x_{n+1}|x_n)f(x_{n-1}|x_n) \), that is, the past and the future are independent given the present.

**Solution** (10 points)
a. We are given that \( f(x_{n+1}|x_1, x_2, \ldots, x_n) = f(x_{n+1}|x_n) \). From the chain rule, in general,
\[
f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_1, x_2) \cdots f(x_n|x_1, x_2, \ldots, x_{n-1}).
\]
Thus, by the definition of Markovity,
\[
f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_2) \cdots f(x_n|x_{n-1}). \quad (4)
\]
Similarly, applying the chain rule in reverse we get
\[
f(x_1, x_2, \ldots, x_n) = f(x_n)f(x_{n-1}|x_n)f(x_{n-2}|x_{n-1}, x_n) \cdots f(x_1|x_2, x_3, \ldots, x_n).
\]
Next,
\[
f(x_i|x_{i+1}, \ldots, x_n) = \frac{f(x_i, x_{i+1}, \ldots, x_n)}{f(x_{i+1}, \ldots, x_n)} = \frac{f(x_i)f(x_{i+1}|x_i)}{f(x_{i+1})} = f(x_i|x_{i+1}), \quad (5)
\]
where the second equality follows from (4). Therefore
\[
f(x_1, x_2, \ldots, x_n) = f(x_n)f(x_{n-1}|x_n)f(x_{n-2}|x_{n-1}, x_n) \cdots f(x_1|x_2, x_3, \ldots, x_n)
= f(x_n)f(x_{n-1}|x_n)f(x_{n-2}|x_{n-1}) \cdots f(x_1|x_2),
\]
where the second line follows from (5).

b. First consider
\[
f(x_n|x_1, \ldots, x_k) = \frac{f(x_1, \ldots, x_k, x_n)}{f(x_1, \ldots, x_k)}
= \frac{f(x_n)f(x_k|x_n)f(x_{k-1}|x_k, x_n) \cdots f(x_1|x_2, \ldots, x_k, x_n)}{f(x_k)f(x_{k-1}|x_k) \cdots f(x_1|x_2)}, \quad (6)
\]
where the denominator in the second line follows from part (a). Next consider
\[
f(x_{k-1}, x_k, \ldots, x_n) = f(x_k, x_n)f(x_{k-1}|x_k, x_n)f(x_{k-1}x_{k+2}, \ldots, x_{n-1}|x_{k-1}, x_k, x_n)
= f(x_n)f(x_{n-1}|x_n) \cdots f(x_{k-1}|x_k),
\]
where the second line follows from (5). Integrating both sides over \( x_{k+1}, \ldots, x_{n-1} \) (i.e., using the Law of Total Probability), we get
\[
f(x_k, x_n)f(x_{k-1}|x_k, x_n) = f(x_k, x_n)f(x_{k-1}|x_k).
\]
Finally, substituting into (6), we get
\[
f(x_n|x_1, \ldots, x_k) = \frac{f(x_n)f(x_k|x_n)f(x_{k-1}|x_k) \cdots f(x_1|x_2)}{f(x_k)f(x_{k-1}|x_k) \cdots f(x_1|x_2)}
= \frac{f(x_n)f(x_k|x_n)}{f(x_k)}
= f(x_n|x_k).
\]
c. By the chain rule for conditional densities,
\[
f(x_{n+1}, x_{n-1}|x_n) = f(x_{n+1}|x_n)f(x_{n-1}|x_{n+1}, x_n) = f(x_{n+1}|x_n)f(x_{n-1}|x_n),
\]
where the second equality follows from [5].