Problem 1. **Forrest Gump**

Forrest Gump is running across the United States, and we would like to track his progress. Assume that on day \( n \in \mathbb{N} \) he runs \( X(n) \) miles, and the amount he runs each day is determined by the amount he ran on the previous day with some random noise in the following manner: \( X(n) = \alpha X(n-1) + V(n) \). Unfortunately, the measurements of the distance he traveled on each day are also subject to some noise. Assume that \( Y(n) \) gives the measured number of miles Forrest Gump traveled on day \( n \) and that \( Y(n) = \beta X(n) + W(n) \). For this problem, assume that \( X(0) \sim \mathcal{N}(0, \sigma_X^2), W(n) \sim \mathcal{N}(0, \sigma_W^2), V(n) \sim \mathcal{N}(0, \sigma_V^2) \) are independent.

1. Suppose that you observe \( Y(0) \). Find the MMSE of \( X(0) \) given this observation.

2. Express both \( \mathbb{E}[Y(n) | Y(0), \ldots, Y(n-1)] \) and \( \mathbb{E}[X(n) | Y(0), \ldots, Y(n-1)] \) in terms of \( \hat{X}(n-1) \), where \( \hat{X}(n-1) \) is the MMSE of \( X(n-1) \) given the observations \( Y(0), Y(1), \ldots, Y(n-1) \).

3. Show that:

\[
\hat{X}(n) = \alpha \hat{X}(n-1) + k_n[Y(n) - \alpha \beta \hat{X}(n-1)]
\]

where

\[
k_n = \frac{\text{cov}(X(n), Y(n))}{\text{var} Y(n)}
\]

and \( \bar{Y}(n) = Y(n) - L[Y(n) | Y(0), Y(1), \ldots, Y(n-1)] \).

Solution 1. 1. We can see that \( X(0), Y(0) \) are jointly Gaussian random variables, so:

\[
\mathbb{E}[X(0) | Y(0)] = L[X(0) | Y(0)] \\
= \mathbb{E}[X(0)] + \frac{\text{cov}(X(0), Y(0))}{\text{var} Y(0)}(Y(0) - \mathbb{E}[Y(0)])
\]

We find:

\[
\text{cov}(X(0), Y(0)) = \mathbb{E}[X(0)Y(0)] = \mathbb{E}[\beta X(0)^2 + X(0)W(0)] = \beta \sigma_X^2
\]

Additionally,

\[
\text{var} Y(0) = \mathbb{E}[(\beta X(0) + W(0))^2] = \beta^2 \sigma_X^2 + \sigma_W^2.
\]

Thus, we have

\[
\mathbb{E}[X(0) | Y(0)] = \frac{\beta \sigma_X^2}{\beta^2 \sigma_X^2 + \sigma_W^2} Y(0).
\]

**Geometric Solution**

Consider Figure [I].
2. We have the following:

\[
\mathbb{E}[Y(n) \mid Y^{(n-1)}] = \mathbb{E}[\beta X(n) + W(n) \mid Y^{(n-1)}]
\]
\[
= \beta \mathbb{E}[X(n) \mid Y^{(n-1)}]
\]
\[
= \beta \mathbb{E}[\alpha X(n-1) + V(n) \mid Y^{(n-1)}]
\]
\[
= \alpha \beta \hat{X}(n-1)
\]

Likewise, we see that:

\[
\mathbb{E}[X(n) \mid Y^{(n-1)}] = \mathbb{E}[\alpha X(n-1) + V(n) \mid Y^{(n-1)}] = \alpha \hat{X}(n-1)
\]

3. We are interested in the quantity \( \hat{X}(n) = \mathbb{E}[X(n) \mid Y^{(n)}] \). In other words, we would like the best estimator after \( n \) measurements have been taken. In an online setting, many of these measurements come in sequential fashion, so we would ideally like to have an estimate at time \( n-1 \) and simply update the estimate when a new measurement comes at time \( n \). Now, we can see that \( \hat{X}(n) = \mathbb{E}[X(n) \mid Y^{(n)}] = L[X(n) \mid Y^{(n)}] \). How does this play into updating our observation? We can equivalently write this as \( L[X(n) \mid Y(n), Y^{(n-1)}] \).

We now state the following Theorem (8.1 in the book):

**Theorem 1.** If \( X, Y, Z \) are 0 mean random variables such that \( Y \) and \( Z \) are orthogonal, \( L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z] \).

The proof is given in the book and we will not repeat it here, but we will give the corresponding diagram which gives the intuition in Figure 3.

We are not quite done, however, as \( Y(n) \) and \( Y^{(n-1)} \) are not necessarily orthogonal. To deal with this, we note that \( Y(n) - L[Y(n) \mid Y(n-1)] \) and \( Y(n-1) \) are orthogonal. Additionally, we can see that any linear combination of \( Y(n), Y^{(n-1)} \) is a linear combination of

\[
(Y(n) - L[Y(n) \mid Y^{(n-1)}], Y^{(n-1)})
\]

(going forward, we let \( \tilde{Y}(n) = Y(n) - L[Y(n) \mid Y^{(n-1)}] \)). Thus, we have:

\[
L[X(n) \mid Y(n), Y^{(n-1)}] = L[X(n) \mid Y^{(n-1)}, \tilde{Y}(n)]
\]
Additionally, by the theorem above, we have:
\[ L[ X(n) \mid Y(n-1), \tilde{Y}(n)] = L[ X(n) \mid Y(n-1)] + L[ X(n) \mid \tilde{Y}(n)] \]

We tackle these parts separately. Note that we determined \( L[ X(n) \mid Y(n-1)] \) in Part (b). Now, we look at \( L[ X(n) \mid \tilde{Y}(n)] \). Recall that from part b, we have \( L[Y(n) \mid Y(n-1)] = \alpha \beta \hat{X}(n-1) \). Thus, we have the following:

\[ L[X(n) \mid Y(n-1)] + L[X(n) \mid \tilde{Y}(n)] \]
\[ = \alpha \hat{X}(n-1) + k_n(Y(n) - L[Y(n) \mid Y(n-1)]) \]
\[ = \alpha \hat{X}(n-1) + k_n(Y(n) - \alpha \beta \hat{X}(n-1)) \]

where (2) follows from the known slope for the LLSE.

Congratulations! You have just derived the recursive structure of the scalar Kalman filter. We refer to \( k_n \) as the gain of the filter, and it turns out that this can be precomputed. Thus, as you have shown, the Kalman filter can be used to compute the MMSE at time \( n \) as a linear function of the MMSE at time \( n-1 \) and the newest observation.

**Problem 2. Hidden Markov Models**

A hidden Markov model (HMM) is a Markov chain \( \{X_n\}_{n=0}^{\infty} \) in which the states are considered “hidden” or “latent”. In other words, we do not directly observe \( \{X_n\}_{n=0}^{\infty} \). Instead, we observe \( \{Y_n\}_{n=0}^{\infty} \), where \( Q(x, y) \) is the probability that state \( x \) will emit observation \( y \). \( \pi_0 \) is the initial distribution for the Markov chain, and \( P \) is the transition matrix.

1. What is \( \Pr(X_0 = x_0, Y_0 = y_0, \ldots, X_n = x_n, Y_n = y_n) \), where \( n \) is a positive integer, \( x_0, \ldots, x_n \) are hidden states, and \( y_0, \ldots, y_n \) are observations?

2. What is \( \Pr(X_0 = x_0 \mid Y_0 = y_0) \)?

3. We observe \( (y_0, \ldots, y_n) \) and we would like to find the most likely sequence of hidden states \( (x_0, \ldots, x_n) \) which gave rise to the observations. Let

\[ U(x_m, m) = \max_{x_{m+1}, \ldots, x_n \in \mathcal{X}} \Pr(X_{m+1:n} = x_{m+1:n}, Y_{m+1:n} = y_{m+1:n} \mid X_m = x_m) \]
denote the largest probability for a sequence of hidden states beginning at state $x_m$ at time $m \in \mathbb{N}$, along with the observations $(y_{m+1}, \ldots, y_n)$. Develop a recursion for $U(x_m, m)$ in terms of $U(x_{m+1}, m+1)$, $x_{m+1} \in \mathcal{X}$.

**Solution 2.**

1. The probability is

$$\pi_0(x_0)Q(x_0, y_0) \prod_{i=1}^{n} P(x_{i-1}, x_i)Q(x_i, y_i).$$

2. This is a simple application of Bayes rule.

$$\Pr(X_0 = x_0 \mid Y_0 = y_0) = \frac{\Pr(X_0 = x_0, Y_0 = y_0)}{\Pr(Y_0 = y_0)} = \frac{\pi_0(x_0)Q(x_0, y_0)}{\sum_{x \in \mathcal{X}} \pi_0(x)Q(x, y_0)}.$$

3. The probability of transitioning to $x_{m+1}$ is $P(x_m, x_{m+1})$. The probability of emission is $Q(x_{m+1}, y_{m+1})$. Once we are in state $x_{m+1}$, the most likely sequence of hidden states for the observations $(y_0, \ldots, y_n)$, beginning at $x_{m+1}$ at time $m+1$, is $U(x_{m+1}, m+1)$. Hence,

$$U(x_m, m) = \max_{x_{m+1} \in \mathcal{X}} P(x_m, x_{m+1})Q(x_{m+1}, y_{m+1})U(x_{m+1}, m+1). \quad (4)$$

To avoid numerical issues, we often work with the logarithms of the above quantities instead.

Note also that the recursion should be solved backwards for efficiency. If we simply try to solve for $U(x_0, m)$, we would have to evaluate all $|\mathcal{X}|^{n+1}$ possible paths, which is computationally prohibitive. Instead, if we solve the equations backwards using (4), then each step requires taking the maximum over $|\mathcal{X}|$ possibilities, so the algorithm will terminate in at most $O(n|\mathcal{X}|)$ steps.