1. Exercise 10.2 in Gallager.

Solution:

(a) Since $Y$ and $U$ are both linear combinations of $X, Z$, they are jointly Gaussian by Definition 3.3.1. Since $E[YU] = 0$, i.e., $Y$ and $U$ are uncorrelated, and since they are jointly Gaussian, they are independent.

(b) Since $Y$ and $U$ are independent and are each $N(0, 2)$, the joint density is

$$f_{Y,U} = \frac{1}{4\pi} \exp \left( -\frac{1}{4}y^2 - \frac{1}{4}u^2 \right).$$

Since the covariance matrix of $Y$ and $U$ is $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$, this is the same as (3.22).

(c) From first principles, $\hat{x}(y) = E[X | Y = y] = \frac{y}{2}$.

(d) From (c),

$$\xi = X - \hat{X}(Y) = X - (X + Z)/2 = (X - Z)/2 = -U/2.$$

(e) Since $Y$ and $U$ are statistically independent,

$$f_{U|Y}(u | y) = f_U(u) = \frac{1}{\sqrt{4\pi}} \exp \left( -u^2/4 \right).$$

Since $\xi = U/2$,

$$f_{\xi|Y}(\xi | y) = f_{\xi}(\xi) = \frac{1}{\sqrt{\pi}} \exp \left( -\xi^2/2 \right).$$

(f) The circles below are equiprobability contours of $X$ and $Z$. Since $y^2 + u^2 = 2(x^2 + z^2)$, they are also equiprobable contours of $Y$ and $U$.

The point of the problem, in associating the estimation error with $u/2$, is to give a graphical explanation of why the estimation error is independent of the estimate. $Y$ and $U$ are independent since the $y$ axis and the $u$ axis are at 45 degree rotations from the the $x$ and $y$ axes, and thus 90 degree from each other.
2. Consider the estimation problem:

\[ Y_1 = X + Z_1 \]
\[ Y_2 = X + Z_2 \]

where \( X, Z_1, Z_2 \) are mutually independent Gaussian random variables. \( X \) has mean \( \bar{X} \) and variance \( \sigma_0^2 \), while \( Z_1, Z_2 \) have mean 0 and variance 1.

(a) Compute the conditional distribution of \( X \) given \( Y_1 = y_1 \).

(b) Compute the conditional distribution of \((X, Z_1)\) given \( Y_2 = y_2 \).

(c) Using (a) and (b) compute the conditional distribution of \( X \) given \( Y_1 = y_1 \) and \( Y_2 = y_2 \) in a successive fashion by first conditioning on \( Y_1 = y_1 \) and then on \( Y_2 = y_2 \).

(d) Check that your answer in part (c) is the same as what you would have got by computing the conditional distribution directly.

**Solution:**

The covariance of \([X, Y_1, Y_2]^\top\) is given by

\[
\text{Cov} \begin{bmatrix} X \\ Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} K_X & K_X & K_X \\ K_X & K_X + K_{Z_1} & K_X \\ K_X & K_X & K_X + K_{Z_2} \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 + 1 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 & \sigma_0^2 + 1 \end{bmatrix},
\]

and the mean components are

\[
\mathbb{E} \begin{bmatrix} X \\ Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \bar{X} \\ \bar{X} \\ \bar{X} \end{bmatrix}.
\]

(a) The conditional distribution of \( X \) given \( Y_1 = y_1 \) is

\[
X \mid Y_1 = y_1 \sim \mathcal{N} \left( \bar{X} + K_{X,Y_1} K_{Y_1}^{-1} (y_1 - \bar{Y}), K_{X,Y_1} K_{Y_1}^{-1} K_{X,Y_1}^\top \right)
\]

\[
= \mathcal{N} \left( \bar{X} + \frac{\sigma_0^2}{\sigma_0^2 + 1} (y_1 - \bar{X}), \frac{\sigma_0^2}{\sigma_0^2 + 1} \right).
\]

(b) The covariance between \([X, Z_1]^\top\) and \( Y_2 \) satisfies

\[
\text{Cov} \begin{bmatrix} X \\ Z_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} K_X & 0 & K_X \\ 0 & K_{Z_1} & 0 \\ K_X & 0 & K_X + K_{Z_2} \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & 0 & \sigma_0^2 \\ 0 & 1 & 0 \\ \sigma_0^2 & 0 & \sigma_0^2 + 1 \end{bmatrix}.
\]

Clearly, \( Z_1 \) and \( Y_2 \) are independent, i.e. observing \( Y_2 \) does not provide any information about \( Z_1 \). Consequently, using the result of Part (a) we get

\[
\begin{bmatrix} X \\ Z_1 \end{bmatrix} \mid Y_2 = y_2 \sim \mathcal{N} \left( \begin{bmatrix} \bar{X} \\ 0 \end{bmatrix} + \begin{bmatrix} \sigma_0^2 \\ \sigma_0^2 + 1 \end{bmatrix} \frac{1}{\sigma_0^2 + 1} y_2 \right), \begin{bmatrix} \sigma_0^2 \\ \sigma_0^2 + 1 \end{bmatrix}
\]

\[
= \mathcal{N} \left( \begin{bmatrix} \frac{1}{\sigma_0^2 + 1} \bar{X} + \frac{\sigma_0^2}{\sigma_0^2 + 1} y_2 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sigma_0^2}{\sigma_0^2 + 1} \\ 1 \end{bmatrix} \right)
\]

(c) Conditional on \( Y_1 = y_1 \), we know from part (a) that

\[
X \mid Y_1 = y_1 \sim \mathcal{N} \left( \bar{X} + \frac{\sigma_0^2}{\sigma_0^2 + 1} (y_1 - \bar{X}), \frac{\sigma_0^2}{\sigma_0^2 + 1} \right)
\]
Let \( X_{y_1} := X \mid Y = y_1 \). Note that \( Z_2 \)’s are independent of \( X, Y_1 \) and \( X_{y_1} \). Conditional on \( Y_2 = y_2 \), we can compute the distribution similar to part (a) but with \( \mathbb{E} [X_{y_1}] \) in place of \( \overline{X} \), \( \sigma^2_{X_{y_1}} \) in place of \( \sigma^2_0 \):

\[
X_{y_1} \mid Y_2 = y_2 \sim \mathcal{N} \left( \overline{X} + \frac{\sigma^2_0}{\sigma^2_0 + 1} (y_1 - X) + \frac{\sigma^2_0}{\sigma^2_0 + 1} (y_2 - X - \frac{\sigma^2_0}{\sigma^2_0 + 1} (y_1 - X), \frac{\sigma^2_0}{\sigma^2_0 + 1} + 1 \right)
\]

\[
= \mathcal{N} \left( \overline{X} + \frac{\sigma^2_0}{\sigma^2_0 + 1} (y_1 - X) + \frac{\sigma^2_0}{2\sigma^2_0 + 1} (y_2 - X - \frac{\sigma^2_0}{\sigma^2_0 + 1} (y_1 - X), \frac{\sigma^2_0}{2\sigma^2_0 + 1} \right)
\]

(d) We know that covariance between \( X \) and \( [Y_1, Y_2] \) satisfies

\[
\text{Cov} \begin{bmatrix} X \\ Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \sigma^2_0 & \frac{\sigma^2_0}{\sigma^2_0 + 1} & \frac{\sigma^2_0}{\sigma^2_0 + 1} \\ \frac{\sigma^2_0}{\sigma^2_0 + 1} & \sigma^2_0 + 1 & \frac{\sigma^2_0}{\sigma^2_0 + 1} \\ \frac{\sigma^2_0}{\sigma^2_0 + 1} & \frac{\sigma^2_0}{\sigma^2_0 + 1} & \sigma^2_0 + 1 \end{bmatrix}.
\]

And hence

\[
X \mid \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}
\]

\[
= \mathcal{N} \left( \overline{X} + [\sigma^2_0, \sigma^2_0] K^{-1} Y_1 \begin{bmatrix} y_1 - \overline{X} \\ y_2 - \overline{X} \end{bmatrix}, K_X - [\sigma^2_0, \sigma^2_0] K^{-1} Y_1 \begin{bmatrix} \sigma^2_0 \\ \sigma^2_0 \end{bmatrix} ^\top \right)
\]

\[
= \mathcal{N} \left( \overline{X} + 1 \frac{2\sigma^2_0 + 1}{\sigma^2_0} \begin{bmatrix} \sigma^2_0 + 1 \\ \sigma^2_0 \end{bmatrix} \begin{bmatrix} y_1 - \overline{X} \\ y_2 - \overline{X} \end{bmatrix}, K_X - 1 \frac{1}{2\sigma^2_0 + 1} \begin{bmatrix} \sigma^2_0, \sigma^2_0 \end{bmatrix} \begin{bmatrix} \sigma^2_0 + 1 \\ -\sigma^2_0 \end{bmatrix} \right)
\]

\[
= \mathcal{N} \left( \overline{X} + \frac{2\sigma^2_0 + 1}{\sigma^2_0} \begin{bmatrix} \sigma^2_0 + 1 \\ \sigma^2_0 \end{bmatrix} \begin{bmatrix} y_1 - \overline{X} \\ y_2 - \overline{X} \end{bmatrix}, \frac{\sigma^2_0}{\sigma^2_0 + 1} \right),
\]

which coincides with our computation in Part (c).

3. Consider the MMSE estimation problem with a \( n \)-dimensional observation:

\[
Y = hX + Z,
\]

where \( X \sim \mathcal{N}(0, 1) \) and \( Z \sim \mathcal{N}(0, \sigma^2 I) \) and independent of \( X \).

(a) Show that \( V = h^T Y \) is a sufficient statistic for the estimation problem. Recall that a sufficient statistic \( V = v(Y) \) is such that \( X, V, Y \) forms a Markov chain.

(b) Does \( V \) remain a sufficient statistic if \( X \) is non-Gaussian?

Solution:

(a) Let \( Q \) be an orthonormal matrix whose first column is \( \frac{h}{|h|} \). Further let \( \tilde{Y} = Q^T Y \). Then only the first element of \( \tilde{Y} \) depends on \( X \). Given \( h^T Y \), the first element of \( \tilde{Y} \) is determined. Then given \( h^T Y \), \( \tilde{Y} \) is independent of \( X \). Because \( \tilde{Y} \) is an orthonormal transformation of \( Y \), given \( h^T Y \), \( Y \) is also independent of \( X \).
(b) Yes. As long as \( Z \sim \mathcal{N}(0, \sigma^2 I) \), the first element of \( \tilde{Y} \) is independent of the rest of the elements. Then given \( h^T Y \), the first element of \( \tilde{Y} \) becomes a number and the rest elements of \( \tilde{Y} \) are still i.i.d. Gaussian. Then the argument follows.

4. Exercise 10.3 in Gallager.

Solution:

Exercise 10.3: a) Let \( X, Z_1, Z_2, \ldots, Z_n \) be independent zero-mean Gaussian rv’s with variances \( \sigma_X^2, \sigma_{Z_1}^2, \ldots, \sigma_{Z_n}^2 \) respectively. Let \( Y_j = h_j X + Z_j \) for \( j \geq 1 \) and let \( Y = (Y_1, \ldots, Y_n)^T \). Use (10.9) to show that the MMSE estimate of \( X \) conditional on \( Y = y = (y_1, \ldots, y_n)^T \), is given by
\[
\hat{x}(y) = \frac{1}{\sigma_Y^2} \sum_{j=1}^n h_j y_j \frac{1}{1 + \sum_{i=1}^n h_i^2/\sigma_{Z_i}^2}.
\]
(A.213)

Hint: Let the row vector \( g^T \) be \([K_X Y][K_Y^{-1}] \) and multiply \( g^T \) by \( K_Y \) to solve for \( g^T \).

Solution: From (10.9), we see that \( \hat{x}(y) = g^T y \) where \( g^T = [K_X Y][K_Y^{-1}] \). Since \( Y = h X + Z \), we see that \( [K_X Y] = h \sigma_X^2 h^T + [K_Z] \) and \( [K_X Y] = \sigma_Y^2 h^T \). Thus we want to solve the vector equation \( g^T h \sigma_X^2 h^T + g^T [K_Z] = \sigma_Y^2 h^T \). Since \( g^T h \) is a scalar, we can rewrite this as \( (1 - g^T h) \sigma_X^2 h^T = g^T [K_Z] \). The \( j \)th component of this equation is
\[
g_j = \frac{(1 - g^T h) \sigma_X^2 h_j}{\sigma_{Z_j}^2}.
\]
(A.214)

This shows that the weighting factors \( g_j \) in \( \hat{x}(y) \) depend on \( j \) only through \( h_j/\sigma_{Z_j} \), which is reasonable. We still must determine the unknown constant \( 1 - g^T h \). To do this, multiply (A.214) by \( h_j \) and sum over \( j \), getting
\[
g^T h = (1 - g^T h) \sum_j \frac{\sigma_X^2 h_j^2}{\sigma_{Z_j}^2}.
\]
Solving for \( g^T h \), from this,
\[
g^T h = \frac{\sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}, \quad 1 - g^T h = \frac{1}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}.
\]
(A.215)

Substituting the expression for \( 1 - g^T h \) into (A.214) yields (A.213).

b) Let \( \xi = X - \tilde{X}(Y) \) and show that (10.29) is valid, i.e., that
\[
1/\sigma_\xi^2 = 1/\sigma_X^2 + \sum_{i=1}^n h_i^2/\sigma_{Z_i}^2.
\]

Solution: Using (10.6) in one dimension, \( \sigma_\xi^2 = \mathbb{E}[\xi X] = \sigma_X^2 - \mathbb{E} \left[ \tilde{X}(Y) X \right] \). Since \( \tilde{X}(Y) = \sum_j g_j Y_j \) from (A.213), and since \( \mathbb{E}[Y_i X] = \sigma_X^2 h_i \), we have
\[
\sigma_\xi^2 = \sigma_X^2 - \sum_{i=1}^n g_i \mathbb{E}[Y_i X] = \sigma_X^2 \left( 1 - \sum_{i=1}^n g_i h_i \right)
\]
\[
= \sigma_X^2(1 - g^T h) = \frac{\sigma_X^2}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2} = \frac{1}{1/\sigma_X^2 + \sum_j h_j^2/\sigma_{Z_j}^2},
\]
where we have used (A.215). This is equivalent to (10.29).

c) Show that (10.28), i.e., \( \hat{x}(y) = \sigma_X^2 \sum_{j=1}^n h_j g_j / \sigma_{Z_j} \), is valid.

Solution: Substitute the expression for \( \sigma_\xi^2 \) in (b) into (A.213).
(d) Show that the expression in (10.29) is equivalent to the iterative expression in (10.25).

**Solution:** First, we show that (10.29) implies (10.27). We use $\xi_n$ to refer to the error for $n$ observations and $\xi_{n-1}$ for the error using the first $n-1$ of those observations. Using (10.29),

$$
\frac{1}{\sigma_{\xi}^2} = \frac{1}{\sigma_x^2} + \sum_{i=1}^{n} \frac{h_i^2}{\sigma_{z_i}^2} = \frac{1}{\sigma_x^2} + \sum_{i=1}^{n-1} \frac{h_i^2}{\sigma_{z_i}^2} + \frac{h_n^2}{\sigma_{z_n}^2}
$$

$$
= \frac{1}{\sigma_{\xi_{n-1}}^2} + \frac{h_n^2}{\sigma_{z_n}^2},
$$

(A.216)

which is (10.27). This holds for all $n$, so (10.27) for all $n$ also implies (10.29).

(e) Show that the expression in (10.28) is equivalent to the iterative expression in (10.25).

**Solution:** Breaking (10.28) into the first $n-1$ terms followed by the term for $n$, we get

$$
\hat{x}(y_1^n) = \sigma_{\xi}^2 \sum_{j=1}^{n-1} \frac{h_j y_j}{\sigma_{z_j}^2} + \sigma_{\xi}^2 \frac{h_n y_n}{\sigma_{z_n}^2} = \sigma_{\xi}^2 \hat{x}(y_1^{n-1}) + \sigma_{\xi}^2 \frac{h_n y_n}{\sigma_{z_n}^2}.
$$

(A.217)

where we used (10.28) for $\hat{x}(y_1^{n-1})$. We can solve for $\sigma_{\xi}^2 / \sigma_{\xi_{n-1}}^2$ by multiplying (A.216) by $\sigma_{\xi}^2$, getting

$$
\frac{\sigma_{\xi}^2}{\sigma_{\xi_{n-1}}^2} = 1 - \frac{\sigma_{\xi}^2 h_n^2}{\sigma_{z_n}^2}.
$$

Substituting this into (A.217) yields

$$
\hat{x}(y_1^n) = \hat{x}(y_1^{n-1}) + \sigma_{\xi}^2 \frac{h_n y_n h_n^2}{\sigma_{z_n}^2}.
$$

Finally, if we invert (A.216), we get

$$
\sigma_{\xi}^2 = \frac{\sigma_{\xi_{n-1}}^2 \sigma_{z_n}^2}{h_n^2 \sigma_{z_n}^2 + \sigma_{z_n}^2}.
$$

Substituting this into (A.217), we get (10.27).

5. Consider the vector dynamical system:

$$
\begin{align*}
X_1 & \sim \mathcal{N}(0, K_1) \\
X_{n+1} &= A X_n + B W_{n+1} \\
Y_n &= C X_n + D Z_n
\end{align*}
$$

where $W_1, \ldots, Z_1, \ldots$, are independent and $W_n \sim \mathcal{N}(0, K_w)$ and $Z_n \sim \mathcal{N}(0, K_z)$.

(a) Reformulate the scalar system:

$$
\begin{align*}
X & \sim \mathcal{N}(0, 1) \\
X_{n+1} &= X_n + 0.5X_{n-1} + 0.2X_{n-2} + W_{n+1} \\
Y_n &= X_n + 0.4X_{n-1} + Z_n
\end{align*}
$$

as a vector dynamical system. Here, $W_n$'s and $Z_n$'s are independent and $\mathcal{N}(0, 1)$ distributed.

(b) For the general vector dynamical system, derive the recursion for the Kalman filter estimates and for the covariance matrices of the errors. Assuming that the error covariance matrices approach a
limit as \( n \to \infty \), characterize the limit in terms of the solution of an equation.

**Solution:**

(a) By setting \( \tilde{X}_n = \begin{bmatrix} X_n \\ X_{n-1} \\ X_{n-2} \end{bmatrix} \), one get the system dynamics in the vector form as follows

\[
\tilde{X}_{n+1} = \begin{bmatrix} X_{n+1} \\ X_n \\ X_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0.2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_n \\ X_{n-1} \\ X_{n-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} W_{n+1}.
\]

\[
Y_n = \begin{bmatrix} 1 & 0.4 & 0 \end{bmatrix} \begin{bmatrix} X_n \\ X_{n-1} \\ X_{n-2} \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} Z_n.
\]

(b) See detailed derivation of the vector Kalman filter in Chapter 10.4.3-10.4.4 of Gallager’s book. The only difference is that this problem is equivalent to

\[
X_{n+1} = AX_n + \tilde{W}_{n+1}
\]

\[
Y_n = CX_n + \tilde{Z}_n
\]

where \( \tilde{W}_n \sim \mathcal{N}(0, BK_w B^\top) \) and \( \tilde{Z}_n \sim \mathcal{N}(0, DK_z D^\top) \). Therefore, the resulting Kalman filter will be the recursive equations in (10.79)-(10.83) with \( A_n, H_n, K_{W_n}, K_{Z_n} \) replaced by \( A, C, BK_w B^\top, DK_z D^\top \), respectively.

In particular, from (10.81) and (10.83) we show that

\[
K_{\zeta_{n+1}} = AK_{\zeta_n} A^\top + BK_w B^\top
\]

\[
= A \left( K_{\zeta_n} - K_{\zeta_n} C^\top (CK_{\zeta_n} C^\top + DK_z D^\top)^{-1} C K_{\zeta_n} \right) A^\top + BK_w B^\top
\]

\[
= AK_{\zeta_n} A^\top + BK_w B^\top - AK_{\zeta_n} C^\top (CK_{\zeta_n} C^\top + DK_z D^\top)^{-1} C K_{\zeta_n} A^\top,
\]

which is a Riccati recursion. If this converges, then the error covariance matrix will be the solution \( K \) to the following equation

\[
K = AK A^\top + BK_w B^\top - AK C^\top (CK C^\top + DK_z D^\top)^{-1} CK A^\top,
\]

which is the well-known discrete time algebraic Riccati equation (DARE).