Homework 2

Solutions

Part I: Reading Assignment

1. Read textbook Sec. 3.2 $\sim$ Sec. 3.5, Sec. 4.1, and Sec. 4.2, which you should have learned before in college level Probability course.
2. Read Chapter 2 of Gallager's note.

Part II: Problem Assignment

1. (a) Let $\mathbf{t} = [t_1, t_2, t_3]^T$. The joint MGF of $\mathbf{w}$ is given by

$$\begin{aligned}\theta_w(\mathbf{t}) &= \exp\left(\mathbf{t}^T\mathbf{m}_w + \frac{1}{2}\mathbf{t}^T\mathbf{K}_w\mathbf{t}\right) \\ &= \exp\left(t_1 + 2t_2 + \frac{3}{2}t_1^2 + \frac{3}{2}t_2^2 + \frac{1}{2}t_3^2 - t_1t_2 + t_1t_3\right).\end{aligned}$$

(b) For any real a and b,

$$a(X - \alpha Y) + bY = aX + (b - a\alpha)Y,$$

which, by definition, is a Gaussian random variable since X and Y are jointly Gaussian. Thus, $(X - \alpha Y)$ and Y are jointly Gaussian.
With that $(X - \alpha Y)$ and Y are jointly Gaussian, what we need to do is to find α such that

$$E[(X - \alpha Y)Y] = E[X - \alpha Y]E[Y]. \tag{1}$$

Thus, we need to know $E[XY]$ and $E[Y^2]$.
The covariance matrix tells us that $E[(X - 1)(Y - 2)] = -1$ and $E[(Y - 2)^2] = 3$, from which we can obtain $E[XY] = 1$ and $E[Y^2] = 7$. Substituting the above results into (1) yields

$$\underline{\alpha = -\frac{1}{3}}.$$

(c) From the covariance matrix we know that Y and Z are independent. Thus, $E[Y^3|Z] = E[Y^3]$. With $E[Y] = 2$, we have $E[(Y - 2)^3] = 0$, expanding which gives

$$\begin{aligned}E[(Y - 2)^3] &= E[Y^3] - 6E[Y^2] + 12E[Y] - 8 \\ &= E[Y^3] - 6 \cdot 7 + 12 \cdot 2 - 8 \\ &= 0,\end{aligned}$$

where $E[Y^2] = 7$ can be obtained from $E[(Y - 2)^2] = 3$. Therefore, we have

$$E[Y^3] = 26.$$

(d) Since X and Y are jointly Gaussian, we know $X + 2Y$ is a Gaussian random variable. We need its mean and variance for the density function. The mean is $E[X] + 2E[Y] = 5$, and the variance is $E[(S - E[X] - 2E[Y])^2] = E[(X - E[X] + 2(Y - E[Y]))^2] = 11$.

(e) Although X and Y are not zero mean in the current case, the decorrelation procedure is identical to that in the problem on page 273 of the textbook.

2. (a) This problem is in essence equivalent to showing that $\mathbf{x}$ and $\mathbf{z}$ are collectively jointly Gaussian. Since $\mathbf{x}$ and $\mathbf{z}$ are independent, $\boldsymbol{\alpha}^T\mathbf{x}$ and $\boldsymbol{\beta}^T\mathbf{y}$ are also independent for any real vector $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Thus, we know

$$\boldsymbol{\alpha}^T\mathbf{x} + \boldsymbol{\beta}^T\mathbf{y}$$

is also a Gaussian random variable. This proves that $\mathbf{x}$ and $\mathbf{z}$ are collectively jointly Gaussian.
Turning back to see whether $\mathbf{u}$ is a Gaussian random vector (jointly Gaussian). We express

$$\mathbf{u} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{H} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix}, \tag{2}$$

where $\mathbf{I}_n$ and $\mathbf{I}_m$ are $n \times n$ and $m \times m$ identity matrix respectively. We see that $\mathbf{u}$ is a linear transformation of jointly Gaussian $\mathbf{x}$ and $\mathbf{z}$. We therefore can conclude that $\mathbf{u}$ is a random vector with jointly Gaussian elements.

(b) From (2), the covariance matrix $\mathbf{K}_u$ is

$$\mathbf{K}_u = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{H} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{K}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_z \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{H}^T \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix}, \tag{3}$$

where the matrix

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{H} & \mathbf{I}_n \end{bmatrix}$$

is nonsingular, no matter whether $\mathbf{H}$ is square, full rank or anything else. Therefore, in order for $\mathbf{K}_u$ to be invertible (nonsingular), we need the condition that both $\mathbf{K}_x$ and $\mathbf{K}_z$ are invertible.

(c) From (a), we know that $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian. Thus, the conditional density is also a Gaussian density with $\mathcal{N}\Big(\mathbf{m}_{x|y}, \mathbf{K}_{x|y}\Big)$ where

$$\begin{aligned} \mathbf{m}_{x|y} &= E\Big[\mathbf{x}|\mathbf{y} = \boldsymbol{y}\Big] \\ &= \mathbf{m}_x + \mathbf{K}_{\mathbf{xy}}\mathbf{K}_{\mathbf{y}}{}^{-1}(\boldsymbol{y} - \mathbf{m}_y) \\ &= \mathbf{K}_{xy}\mathbf{K}_y^{-1}\boldsymbol{y} \qquad (\text{since } \mathbf{m}_x = \mathbf{m}_y = \mathbf{0}) \\ &= \mathbf{K}_x\mathbf{H}^T\big(\mathbf{H}\mathbf{K}_x\mathbf{H}^T + \mathbf{K}_z\big)^{-1}\boldsymbol{y}. \end{aligned}$$

and

$$\begin{aligned} \mathbf{K}_{x|y} &= \mathbf{K}_x - \mathbf{K}_{xy}\mathbf{K}_y^{-1}\mathbf{K}_{yx} \\ &= \mathbf{K}_{xy}\mathbf{K}_y^{-1}\mathbf{y} \\ &= \mathbf{K}_x - \mathbf{K}_x\mathbf{H}^T\big(\mathbf{H}\mathbf{K}_x\mathbf{H}^T + \mathbf{K}_z\big)^{-1}\mathbf{H}\mathbf{K}_x. \end{aligned}$$

3. (10 points)
For a circularly symmetric complex Gaussian random variable $X = X_r + jX_i$, we know that its mean and pseudo-covariance are both zero. Zero mean suggests that X_r and X_i both have zero mean too. And, the pseudo-covariance gives

$$E[XX^T] = E[X_r^2 - X_i^2] + j2E[X_rX_i],$$

which is zero only when its real part and imaginary part are both zero, suggesting

$$E[X_r^2] = E[X_i^2] \quad \text{and} \quad E[X_rX_i] = 0.$$

This says the real part and the imaginary part have equal 2nd moment, and are uncorrelated. Since X_r and X_i are jointly Gaussian, uncorrelatedness implies independence. Thus, we have shown that the real part and the imaginary part are i.i.d.

4. Problem 2.13 in Gallager's note.

(a) For $\mathbf{y} = \mathbf{A}\mathbf{w}$, we have

$$E[\mathbf{w}\mathbf{y}^T] = E[\mathbf{w}\mathbf{w}^T]\mathbf{A}^T = \mathbf{A}^T.$$

So, we need $\mathbf{A}^T = \mathbf{K}$. That is $\mathbf{A} = \mathbf{K}^T$.

(b) $E[\mathbf{w}\mathbf{z}^T] = E[\mathbf{w}\mathbf{w}^T]\mathbf{A}^T = \mathbf{K_z}\mathbf{B}^T$. Thus, we need $\mathbf{B} = \mathbf{K}^T\mathbf{K_z}{}^{-1}$.

(c) From part (b), we know that we are looking for $\mathbf{B}$ such that $\mathbf{B}\mathbf{K}_\mathbf{z} = \mathbf{K}^T$. Assuming that $\mathbf{B}^T = [\mathbf{b}_1|\cdots|\mathbf{b}_m]$ and $\mathbf{K} = [\mathbf{k}_1|\cdots|\mathbf{k}_m]$, let us rewrite ${\mathbf{K}_\mathbf{z}}^T\mathbf{B}^T = \mathbf{K}$ in the following form

$$\mathbf{K}_\mathbf{z}[\mathbf{b}_1|\cdots|\mathbf{b}_m] = [\mathbf{k}_1|\cdots|\mathbf{k}_m]$$

This can be decomposed into a set of linear equations

$$\mathbf{K}_\mathbf{z}\mathbf{b}_i = \mathbf{k}_i \tag{4}$$

for $i = 1\cdots m$. Since now $\mathbf{K}_\mathbf{z}$ is singular, we cannot uniquely determine $\mathbf{b}_i$. Assume the covariance matrix $\mathbf{K}_\mathbf{z}$ has rank r, it can be decomposed to $\mathbf{K}_\mathbf{z} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^T$ where $\Lambda = \text{diag}\left(\lambda_1\cdots\lambda_r, 0, \cdots, 0\right)$. Substituting this into (4) gives

$$\boldsymbol{\Lambda}\mathbf{E}^T\mathbf{b}_i = \mathbf{E}^T\mathbf{k}_i,$$

of which the lower $(m-r)$ equations equal to zero, i.e. $\mathbf{e}_j^T\mathbf{k}_i = 0$, for $j \geq r+1$ where $\mathbf{e}_j$ is the jth column vector of the matrix $\mathbf{E}$. In other words, each column of $\mathbf{K}$ is orthogonal to all the eigenvectors that have a zero eigenvalue. This is equivalent to saying that each column of $\mathbf{K}$ belongs the space $V = span\{\mathbf{e}_1, \cdots, \mathbf{e}_r\}$.

In conclusion, for $\mathbf{B}$ to satisfy the desired cross-correlation matrix, each column of $\mathbf{K}$ must belongs to the space spanned by the eigenvectors that correspond to the non-zero eigenvalues of $\mathbf{K}_\mathbf{z}$. ■

5. This problem again reflects the importance of Gaussian density, where with only mean and variance of a random variable, we can model the random variable as Gaussian in the sense that the Gaussian density maximizes the entropy of the random variable.

In this problem, the only things we have are:

$$\int_{-\infty}^{\infty} p(x)dx = 1 \tag{5}$$

$$\int_{-\infty}^{\infty} xp(x)dx = \mu \tag{6}$$

$$\int_{-\infty}^{\infty} x^2p(x)dx = \mu^2 + \sigma^2. \tag{7}$$

We want to find a $p(x)$ that maximizes the entropy

$$H[X] \triangleq -\int_{-\infty}^{\infty} p(x)\ln p(x)dx,$$

while satisfying the above three conditions. This is a typical constrained optimization problem, and we can resort to Lagrange Multiplier technique to find a solution.

The cost function for the constrained optimization problem here is

$$Q(p(x)) = \int_{-\infty}^{\infty} \left[-p(x)\ln p(x) + \lambda_1 p(x) + \lambda_2 xp(x) + \lambda_3 x^2 p(x)\right] dx,$$

where λ_1, λ_2, andλ_3 are the Lagrange multipliers. Taking the derivative with respect to $p(x)$ and letting the result equal to zero give

$$\ln p(x) = -1 + \lambda_1 + \lambda_2 x + \lambda_3 x^2.$$

So, we have

$$p(x) = K \cdot e^{\lambda_2 x + \lambda_3 x^2}, \tag{8}$$

where $K = e^{-1+\lambda_1}$. We will solve for these 3 Lagrange multipliers using (1), (2), and (3). Actually, we don't need to solve for λ_1, λ_2 and λ_3 explicitly. Based on (1) and (4), we know

$$K \cdot \int_{-\infty}^{\infty} e^{\lambda_2 x + \lambda_3 x^2} dx = 1,$$

where the exponent is quadratic with respect to x. Merely from this observation, we can conclude that $p(x)$ is a Gaussian density with proper coefficient K. Then, from (2) and (3), we know $p(x)$ is Gaussian with mean μ and variance σ^2. ■

Extra Problems

This problem is very important. It states that, when the jointly Gaussian random vectors $\mathbf{y}$ and $\mathbf{z}$ are dependent, conditioning on $\mathbf{z}$ can always be replaced by conditioning on another Gaussian random vector $\hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ and $\mathbf{y}$ are statistically independent. We will use this property later this semester when discussing the recursive minimum mean squared estimator, or the so-called ***Kalman filter***.

(a) Let $\mathbf{s} \triangleq [\mathbf{y}^T\mathbf{z}^T]^T$ be the $(m+r)\times 1$ vector collecting $\mathbf{y}$ and $\mathbf{z}$. In topic 3, we learn that the conditional mean

$$E[\mathbf{x}|\mathbf{y},\mathbf{z}] = E[\mathbf{x}|\mathbf{s}] = \mathbf{m_x} + \mathbf{K_{xs}}\mathbf{K_s}^{-1}(\mathbf{s}-\mathbf{m_s}), \tag{9}$$

where $\mathbf{m_x}$ and $\mathbf{m_s}$ are the mean vector of $\mathbf{x}$ and $\mathbf{s}$, respectively. In the following, the notation $\mathbf{m}_*$ refers to the mean vector of $*$.

We first carry out $\mathbf{K_{xs}}$ and $\mathbf{K_s}^{-1}$. The cross-covariance matrix between $\mathbf{x}$ and $\mathbf{s}$ is an $n\times(m+r)$ matrix and can be written in a block matrix form as

$$\mathbf{K_{xs}} = E\left[(\mathbf{x}-\mathbf{m_x})(\mathbf{s}-\mathbf{m_s})^T\right] = [\mathbf{K_{xy}} \mid \mathbf{K_{xz}}],$$

where $\mathbf{K_{xy}}$ and $\mathbf{K_{xz}}$ are the cross-covariance matrix of [$\mathbf{x}$ and $\mathbf{y}$, and [$\mathbf{x}$ and $\mathbf{z}$, respectively.

Since $\mathbf{y}$ and $\mathbf{z}$ are independent, their cross-covariance matrix $\mathbf{K_{yz}}$ is a zero matrix. Therefore, the covariance matrix of $\mathbf{s}$ is

$$\mathbf{K_s} = E\left[\begin{bmatrix}\mathbf{y}-\mathbf{m_y}\\ \mathbf{z}-\mathbf{m_z}\end{bmatrix}\cdot[(\mathbf{y}-\mathbf{m_y})^T,(\mathbf{z}-\mathbf{m_z})^T]\right] = \left[\begin{array}{c|c}\mathbf{K_y} & \mathbf{0}\\ \hline \mathbf{0} & \mathbf{K_z}\end{array}\right].$$

It is clear that

$$\mathbf{K_s}^{-1} = \left[\begin{array}{c|c}\mathbf{K_y}^{-1} & \mathbf{0}\\ \hline \mathbf{0} & \mathbf{K_z}^{-1}\end{array}\right].$$

This can be seen by directly expanding the matrix multiplication $\mathbf{K_s}\mathbf{K_s}^{-1}$, and show the result is an identity matrix. With straightforward manipulations, we find

$$\begin{aligned}\mathbf{K_{xs}}\mathbf{K_s}^{-1}(\mathbf{s}-\mathbf{m_s}) &= [\mathbf{K_{xy}}\mathbf{K_y}^{-1},\ \mathbf{K_{xz}}\mathbf{K_z}^{-1}]\cdot\begin{bmatrix}\mathbf{y}-\mathbf{m_y}\\ \mathbf{z}-\mathbf{m_z}\end{bmatrix}\\ &= \mathbf{K_{xy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y}) + \mathbf{K_{xz}}\mathbf{K_z}^{-1}(\mathbf{z}-\mathbf{m_z})\\ &= E[\mathbf{x}|\mathbf{y}] - \mathbf{m_x} + E[\mathbf{x}|\mathbf{z}] - \mathbf{m_x}. \end{aligned}\tag{10}$$

Substituting (10) into (9) yields

$$E[\mathbf{x}|\mathbf{y},\mathbf{z}] = E[\mathbf{x}|\mathbf{y}] + E[\mathbf{x}|\mathbf{z}] - \mathbf{m_x},$$

when $\mathbf{y}$ and $\mathbf{z}$ are statistically independent.

(b) This part is somewhat involved, and its result is very important, as I mentioned. Intuitively, since $\mathbf{y}$ and $\mathbf{z}$ are dependent, knowing both $\mathbf{y}$ and $\mathbf{z}$ is not necessary. We can work on statistically independent $\mathbf{y}$ and $\hat{\mathbf{z}}$, and apply part (1) to find $E[\mathbf{x}|\mathbf{y},\hat{\mathbf{z}}]$. The proof in this part basically includes 3 things:

— expand $E[\mathbf{x}|\mathbf{y},\mathbf{z}]$

— expand $E[\mathbf{x}|\mathbf{y},\hat{\mathbf{z}}]$, and

— compare the above expanded results

Again, let $\mathbf{s} \triangleq [\mathbf{y}^T\mathbf{z}^T]^T$. We know

$$E[\mathbf{x}|\mathbf{y},\mathbf{z}] = \mathbf{m_x} + \underbrace{[\mathbf{K_{xy}} \mid \mathbf{K_{xz}}]}_{=\mathbf{K_{xs}}}\underbrace{\begin{bmatrix}\mathbf{K_y} & \mathbf{K_{yz}}\\ \mathbf{K_{zy}} & \mathbf{K_z}\end{bmatrix}^{-1}}_{=\mathbf{K_s}^{-1}}\begin{bmatrix}\mathbf{y}-\mathbf{m_y}\\ \mathbf{z}-\mathbf{m_z}\end{bmatrix}. \tag{11}$$

The inverse of $\mathbf{K_s}$ can be further carried out as

$$\mathbf{K_s}^{-1} = \begin{bmatrix}\mathbf{K_y} & \mathbf{K_{yz}}\\ \mathbf{K_{zy}} & \mathbf{K_z}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^T & \mathbf{C}\end{bmatrix},$$

where

$$\begin{aligned}
\mathbf{A} &= \mathbf{K_y}^{-1} + \mathbf{K_y}^{-1}\mathbf{K_{yz}}\mathbf{C}\mathbf{K_{zy}}\mathbf{K_y}^{-1} \\
\mathbf{B} &= -\mathbf{K_y}^{-1}\mathbf{K_{yz}}\mathbf{C} \\
\mathbf{C} &= \left(\mathbf{K_z} - \mathbf{K_{zy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}\right)^{-1}.
\end{aligned}$$

By plugging all these into (11), we have

$$\begin{aligned}
E[\mathbf{x}|\mathbf{y},\mathbf{z}] &= \mathbf{m_x} + \Big(\mathbf{K_{xy}}\mathbf{K_y}^{-1} + \mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}\mathbf{C}\mathbf{K_{zy}}\mathbf{K_y}^{-1} - \mathbf{K_{xz}}\mathbf{C}\mathbf{K_{zy}}\mathbf{K_y}^{-1}\Big)(\mathbf{y}-\mathbf{m_y}) \\
&\quad + \Big(\mathbf{K_{xz}}\mathbf{C} - \mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}\mathbf{C}\Big)(\mathbf{z}-\mathbf{m_z}) \\
&= \mathbf{m_x} + \Big(\mathbf{K_{xy}}\mathbf{K_y}^{-1} + \left(\mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}} - \mathbf{K_{xz}}\right)\mathbf{C}\mathbf{K_{zy}}\mathbf{K_y}^{-1}\Big)(\mathbf{y}-\mathbf{m_y}) \\
&\quad + \Big(\mathbf{K_{xz}} - \mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}\Big)\mathbf{C}(\mathbf{z}-\mathbf{m_z}). \qquad (12)
\end{aligned}$$

On the other hand, the conditional expectation conditioned on $\mathbf{y}$ and $\hat{\mathbf{z}}$ can be given by

$$\begin{aligned}
E[\mathbf{x}|\mathbf{y},\hat{\mathbf{z}}] &= \mathbf{m_x} + [\mathbf{K_{xy}} \mid \mathbf{K_{x\hat{z}}}]\begin{bmatrix} \mathbf{K_y} & \mathbf{0} \\ \mathbf{0} & \mathbf{K_{\hat{z}}} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{y}-\mathbf{m_y} \\ \hat{\mathbf{z}}-\mathbf{m_{\hat{z}}} \end{bmatrix} \\
&= \mathbf{m_x} + \mathbf{K_{xy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y}) + \mathbf{K_{x\hat{z}}}\mathbf{K_{\hat{z}}}^{-1}\left(\hat{\mathbf{z}}-\mathbf{m_{\hat{z}}}\right), \qquad (13)
\end{aligned}$$

where $\mathbf{m_{\hat{z}}} = E[\hat{\mathbf{z}}] = \mathbf{0}$ and we have used the property that $\mathbf{y}$ and $\hat{\mathbf{z}}$ are independent, i.e.

$$\begin{aligned}
\mathrm{Cov}(\mathbf{y},\hat{\mathbf{z}}) &= E[(\mathbf{y}-\mathbf{m_y})\hat{\mathbf{z}}^T] = E[\mathbf{y}\hat{\mathbf{z}}^T] \\
&= E[\mathbf{y}\mathbf{z}^T] - E[\mathbf{y}E[\mathbf{z}^T|\mathbf{y}]] \\
&= E[\mathbf{y}\mathbf{z}^T] - E[E[\mathbf{y}\mathbf{z}^T|\mathbf{y}]] \\
&= E[\mathbf{y}\mathbf{z}^T] - E[\mathbf{y}\mathbf{z}^T] = \mathbf{0}.
\end{aligned}$$

Now, it remains to find $\mathbf{K_{x\hat{z}}}$ and $\mathbf{K_{\hat{z}}}^{-1}$ in (13). It is straightforward to show that

$$\begin{aligned}
&\mathbf{K_{x\hat{z}}} = E\left[\big(\mathbf{x}-\mathbf{m_x}\big)\big(\mathbf{z}-\mathbf{m_z}-\mathbf{K_{zy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y})\big)^T\right] = \mathbf{K_{xz}} - \mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}} \\
&\mathbf{K_{\hat{z}}} = E\left[\big(\mathbf{z}-\mathbf{m_z}-\mathbf{K_{zy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y})\big)\big(\mathbf{z}-\mathbf{m_z}-\mathbf{K_{zy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y})\big)^T\right] \\
&\quad = \mathbf{K_z} - \mathbf{K_{zy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}.
\end{aligned}$$

We see that $\mathbf{K_{\hat{z}}}^{-1} = \mathbf{C}$. Again, plugging these results into (13) yields

$$\begin{aligned}
E[\mathbf{x}|\mathbf{y},\hat{\mathbf{z}}] &= \mathbf{m_x} + \mathbf{K_{xy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y}) + \left(\mathbf{K_{xz}} - \mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}\right)\mathbf{C}\left(\mathbf{z}-E[\mathbf{z}|\mathbf{y}]\right) \\
&= \mathbf{m_x} + \mathbf{K_{xy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y}) + \left(\mathbf{K_{xz}} - \mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}\right)\mathbf{C} \\
&\quad \times \left(\mathbf{z}-\mathbf{m_z}-\mathbf{K_{zy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y})\right) \\
&= \mathbf{m_x} + \Big(\mathbf{K_{xy}}\mathbf{K_y}^{-1} + \left(\mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}} - \mathbf{K_{xz}}\right)\mathbf{C}\mathbf{K_{zy}}\mathbf{K_y}^{-1}\Big)(\mathbf{y}-\mathbf{m_y}) \\
&\quad + \Big(\mathbf{K_{xz}} - \mathbf{K_{xy}}\mathbf{K_y}^{-1}\mathbf{K_{yz}}\Big)\mathbf{C}(\mathbf{z}-\mathbf{m_z}),
\end{aligned}$$

which is exactly the same as the result in (12).

(Another approach:)

Another easier way to prove the statement given in the problem is using the concept of linear transformation. Since $\mathbf{y}$ and $\mathbf{z}$ are jointly Gaussian, we know

$$\hat{\mathbf{z}} = \mathbf{z} - E[\mathbf{z}|\mathbf{y}] = \mathbf{z} - \mathbf{m_z} - \mathbf{K_{zy}}\mathbf{K_y}^{-1}(\mathbf{y}-\mathbf{m_y}) \qquad (14)$$

Let $\hat{\mathbf{s}} \triangleq [\mathbf{y}^T\hat{\mathbf{z}}^T]^T$. Then, it follows from (14) that $\hat{\mathbf{s}}$ can be written as

$$\hat{\mathbf{s}} = \begin{bmatrix} \mathbf{y} \\ \hat{\mathbf{z}} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{K_{zy}}\mathbf{K_y}^{-1} & \mathbf{I} \end{bmatrix}}_{\triangleq \mathbf{A}}\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{0} \\ -\mathbf{m_z} + \mathbf{K_{zy}}\mathbf{K_y}^{-1}\mathbf{m}_y \end{bmatrix}}_{\triangleq \mathbf{b}}.$$

That is, with $\mathbf{y}$ and $\mathbf{z}$ being jointly Gaussian, the vector $\hat{\mathbf{s}}$ is in fact a linear transformation of the vector $\mathbf{s}$ with

$$\hat{\mathbf{s}} = \mathbf{A}\mathbf{s} + \mathbf{b}, \tag{15}$$

where $\mathbf{A}$ and $\mathbf{b}$ are defined in the previous equation. Since $\mathbf{x}$ and $\mathbf{s}$ are jointly Gaussian, it can be easily shown that $\mathbf{x}$ and $\hat{\mathbf{s}}$ are also jointly Gaussian. Then, the conditional mean can be expressed by

$$E[\mathbf{x}|\hat{\mathbf{s}}] = \mathbf{m_x} + \mathbf{K_{x\hat{s}}}\mathbf{K_{\hat{s}}}^{-1}(\hat{\mathbf{s}} - \mathbf{m_{\hat{s}}}). \tag{16}$$

What remains is to find the expression of $\mathbf{K_{x\hat{s}}}$ and $\mathbf{K_{\hat{s}}}$. With the equality (15), we can find explicit relationships between $\mathbf{K_{x\hat{s}}}$ and $\mathbf{K_{xs}}$, as well as $\mathbf{K_{\hat{s}}}$ and $\mathbf{K_s}$ as follows:

$$\begin{aligned}
\mathbf{K_{x\hat{s}}} &= E\left[(\mathbf{x} - \mathbf{m_x})(\hat{\mathbf{s}} - \mathbf{m_{\hat{s}}})^T\right] \\
&= E\left[\big(\mathbf{x} - \mathbf{m_x}\big)\big(\mathbf{A}\mathbf{s} + \mathbf{b} - (\mathbf{A}\mathbf{m_s} + \mathbf{b})\big)^T\right] \\
&= E\left[\big(\mathbf{x} - \mathbf{m_x}\big)\big(\mathbf{s} - \mathbf{m_s}\big)^{\mathbf{T}}\right] \cdot \mathbf{A}^T \\
&= \mathbf{K_{xs}} \cdot \mathbf{A}^T
\end{aligned}$$

$$\begin{aligned}
\mathbf{K_{\hat{s}}} &= E\left[(\hat{\mathbf{s}} - \mathbf{m_{\hat{s}}})(\hat{\mathbf{s}} - \mathbf{m_{\hat{s}}})^T\right] \\
&= \mathbf{A} \cdot E\left[(\mathbf{s} - \mathbf{m_s})(\mathbf{s} - \mathbf{m_s})^T\right] \cdot \mathbf{A}^T \\
&= \mathbf{A}\mathbf{K_s}\mathbf{A}^T.
\end{aligned}$$

Substituting these two into (16) gives rise to

$$\begin{aligned}
E[\mathbf{x}|\hat{\mathbf{s}}] &= \mathbf{m_x} + \mathbf{K_{x\hat{s}}}\mathbf{K_{\hat{s}}}^{-1}(\hat{\mathbf{s}} - \mathbf{m_{\hat{s}}}) \\
&= \mathbf{m_x} + \mathbf{K_{xs}} \cdot \mathbf{A}^T \left(\mathbf{A}\mathbf{K_s}\mathbf{A}^T\right)^{-1} \cdot (\mathbf{A}\mathbf{s} - \mathbf{A}\mathbf{m_s}) \\
&= \mathbf{m_x} + \mathbf{K_{xs}}\mathbf{K_s}^{-1}(\mathbf{s} - \mathbf{m_s}),
\end{aligned}$$

which is exactly the result of $E[\mathbf{x}|\mathbf{s}]$ under the condition that $\mathbf{A}^{-1}$ exists. We can justify $\mathbf{A}$ is indeed nonsingular by checking its determinant $\det(\mathbf{A}) = 1$.
This implies that, since $\mathbf{A}$ is nonsingular, the linear transformation $\mathbf{A}\mathbf{s} + \mathbf{b}$ does not lose any information that $\mathbf{s}$ originally provides, nor does it add in anything new. Therefore, conditioning on $\mathbf{A}\mathbf{s} + \mathbf{b}$ is equivalent to conditioning on $\mathbf{s}$. ■