1. **Noise cancellation**  A classic problem in statistical signal processing involves estimating a weak signal (e.g., the heart beat of a fetus) in the presence of a strong interference (the heart beat of its mother) by making two observations—one with the weak signal present and one without (by placing one microphone on the mother’s belly and another close to her heart). The observations can then be combined to estimate the weak signal by “canceling out” the interference. The following is a simple version of this application.

Let the weak signal $X$ be a random variable with mean $\mu$ and variance $P$. Let the observations be $Y_1 = X + Z_1$ and $Y_2 = Z_1 + Z_2$, where $Z_1$ is the strong interference and $Z_2$ is measurement noise. Assume that $Z_1$ and $Z_2$ are zero mean with variances $N_1$ and $N_2$, respectively. Further assume that $X$, $Z_1$, and $Z_2$ are uncorrelated. Find the LLSE estimate of $X$ given $Y_1$ and $Y_2$ and the corresponding MSE. Interpret the results.

**Solution:** This is a vector MSE linear estimation problem. Since $Z_1$ and $Z_2$ are zero mean, $\mu_{Y_1} = \mu_X + \mu_{Z_1} = \mu$ and $\mu_{Y_2} = \mu_{Z_1} + \mu_{Z_2} = 0$. We first normalize the random variables by subtracting their means to get

$$X' = X - \mu \quad \text{and} \quad Y' = \begin{bmatrix} Y_1 - \mu \\ Y_2 \end{bmatrix}.$$ 

To find the best linear estimate $\hat{X}'$ of $X'$, we first find

$$\Sigma_Y = \begin{bmatrix} P + N_1 \\ N_1 \\ N_1 + N_2 \end{bmatrix} \quad \text{and} \quad \Sigma_{YX} = \begin{bmatrix} P \\ 0 \end{bmatrix}.$$ 

Therefore

$$\hat{X}' = \Sigma_{YX}^{-1} \Sigma_Y^{-1} Y'$$

$$= \begin{bmatrix} P & 0 \end{bmatrix} \frac{1}{P(N_1 + N_2) + N_1 N_2} \begin{bmatrix} N_1 + N_2 & -N_1 \\ -N_1 & P + N_1 \end{bmatrix} \begin{bmatrix} Y_1 - \mu \\ Y_2 \end{bmatrix}$$

$$= \frac{P}{P(N_1 + N_2) + N_1 N_2} \begin{bmatrix} N_1 + N_2 & -N_1 \end{bmatrix} \begin{bmatrix} Y_1 - \mu \\ Y_2 \end{bmatrix}.$$ 

The best linear MSE estimate is $\hat{X} = \hat{X}' + \mu$. Thus

$$\hat{X} = \frac{P(N_1 + N_2)(Y_1 - \mu) - PN_1 Y_2}{P(N_1 + N_2) + N_1 N_2} + \mu$$

$$= \frac{P(N_1 + N_2)Y_1 - PN_1 Y_2 + N_1 N_2 \mu}{P(N_1 + N_2) + N_1 N_2}.$$ 

The MSE can be calculated by

$$\text{MSE} = \sigma_{X'}^2 - \Sigma_{X'}^{-1} \Sigma_{YX}^{-1} \Sigma_{YX}$$

$$= \frac{P}{P(N_1 + N_2) + N_1 N_2} \begin{bmatrix} N_1 + N_2 & -N_1 \end{bmatrix} \begin{bmatrix} P \\ 0 \end{bmatrix}$$

$$= \frac{P^2(N_1 + N_2)}{P(N_1 + N_2) + N_1 N_2} = \frac{PN_1 N_2}{P(N_1 + N_2) + N_1 N_2}.$$
Note that if either \( N_1 \) or \( N_2 \) go to 0, the MSE also goes to 0. This is because the estimator will then use the measurement with zero noise variance (that is, the one with no noise) to perfectly reconstruct \( X \).

2. **Singular Covariance Matrix.** From the book we know the general optimal linear estimator of \( X \) given vector \( Y \) is given by:

\[
\hat{X}_{\text{lin}}(Y) = K_{XY}K_Y^{-1}(Y - \bar{Y}) + \bar{X},
\]

where \( \bar{Y} \) and \( \bar{X} \) are the mean of \( Y \) and \( X \).

However, this formula would not work if the covariance matrix \( K_Y \) is not invertible. In this problem, you are required to investigate this situation and obtain the corresponding solutions.

(a) Prove that if the covariance matrix of a \( d \)-dimensional random vector \( Y \) is not invertible, then there must exist at least one component of \( Y \) which is expressible as a linear combination of the others.

(b) Suppose \( X \) is a zero-mean scalar random variable, \( Y \) is a zero-mean random vector with covariance matrix

\[
K_Y = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix},
\]

and the covariance between \( X \) and \( Y \) is

\[
K_{XY} = [3, 4, 7].
\]

Derive the optimal linear estimator of \( X \) based on \( Y \).

**Solution:**

(a) As shown in class, each covariance matrix has a unique Cholesky decomposition. In other words, there exists a lower-triangular matrix \( L \) such that

\[
K_Y = LL^T.
\]

If \( K_Y \) is not invertible, then there must be at least one zero component in the diagonal of \( L \). Indeed, if all of the diagonal components of \( L \) are non-zero, then \( L \) is full rank, which implies \( K_Y \) is invertible.

Assume the \( i \)-th diagonal component of \( L \) is zero. Now we show that \( Y_i \) could be expressed as a linear combination of \( Y_1, Y_2, \ldots, Y_{i-1} \). Indeed, if we denote the first \( i - 1 \) components of \( Y \) as an \( i - 1 \)-dimensional vector \( Y_{[1:i-1]} \), then the Cholesky decomposition shows that there exists an \( d \)-dimensional vector \( U \) with zero mean and identity covariance such that

\[
Y_{[1:i-1]} = L_{[1:i-1]}U_{[1:i-1]}.
\]

Equivalently,

\[
U_{[1:i-1]} = L_{[1:i-1]}^{-1}Y_{[1:i-1]}.
\]

Since we have assumed the \( i \)-th diagonal component of \( L \) is zero, we know that

\[
Y_i = \sum_{j=1}^{i-1} L_{ij}U_j.
\]

Equivalently, we have

\[
Y_i = \{L_{i1}, L_{i2}, \ldots, L_{i,i-1}\}U_{[1:i-1]} = \{L_{i1}, L_{i2}, \ldots, L_{i,i-1}\}L_{[1:i-1]}^{-1}Y_{[1:i-1]}.
\]
Thus, we have shown that \( Y_i \) could be expressed as linear combination of other components in \( \mathbf{Y} \).

Conversely, if there exists a component \( Y_i \) such that it can be expressed as a linear combination of other components, then there must exist a non-zero \( d \)-dimensional vector \( a \) such that

\[
a^T \mathbf{Y} = 0.
\]  

We have

\[
\mathbb{E}[a^T \mathbf{Y}]^2 = \mathbb{E}[\mathbf{Y}^T a] = a^T \mathbb{E}[\mathbf{Y} \mathbf{Y}^T] a = a^T \mathbf{K}_Y a = 0.
\]  

The existence of a non-zero vector \( a \) such that \( a^T \mathbf{K}_Y a \) implies that \( \mathbf{K}_Y \) is not invertible.

(b) We compute the Cholesky decomposition of \( \mathbf{K}_Y \):

\[
L = \begin{bmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
2 & 1 & 0
\end{bmatrix}
\]  

Since it is singular, we know that there must exist a component of \( \mathbf{Y} \) which can be expressed as a linear combination of the other components. Using arguments in the solution of part (a), we have

\[
\hat{Y}_3 = Y_1 + Y_2.
\]

Thus, we abandon \( Y_3 \) and use \( \mathbf{Y}' = (Y_1, Y_2)^T \) to construct the optimal linear estimator of \( X \) given \( Y \).

We have

\[
\hat{X}_{\text{lin}}(\mathbf{Y}) = \hat{X}_{\text{lin}}(\mathbf{Y}') = \mathbf{K}_{\mathbf{X}\mathbf{Y}'} \mathbf{K}_{\mathbf{Y}'}^{-1} \mathbf{Y}',
\]

since we have assumed \( \mathbb{E}X = 0, \mathbb{E}Y = 0 \).

We have

\[
\mathbf{K}_{\mathbf{X}\mathbf{Y}'} = [3, 4],
\]  

and

\[
\mathbf{K}_{\mathbf{Y}'}^{-1} = \begin{bmatrix}
2 & -1 \\
-1 & 1
\end{bmatrix}.
\]  

Hence we have

\[
\hat{X}_{\text{lin}}(\mathbf{Y}) = [2, 1] \mathbf{Y}' = 2Y_1 + Y_2.
\]

3. Additive Non-White Gaussian Noise Channel

Let \( Y_i = X + Z_i \) for \( i = 1, 2, \ldots, n \) be \( n \) observations of a signal \( X \sim \mathcal{N}(0, P) \). The additive noise random variables \( Z_1, Z_2, \ldots, Z_n \) are zero mean jointly Gaussian random variables that are independent of \( X \) and have correlation \( \mathbb{E}(Z_i Z_j) = N \cdot 2^{-|i-j|} \) for \( 1 \leq i, j \leq n \). Find the best MSE estimate of \( X \) given \( Y_1, Y_2, \ldots, Y_n \), and its MSE. Hint: the coefficients for the best estimate are of the form \( \mathbf{h}^T = [a \ b \ b \ \cdots \ b \ b \ a] \).

Solution: The best estimate of \( X \) is of the form

\[
\hat{X} = \sum_{i=1}^{n} h_i Y_i.
\]

We apply the orthogonality condition \( \mathbb{E}(XY_j) = \mathbb{E}(\hat{X}Y_j) \) for \( 1 \leq j \leq n \):

\[
P = \sum_{i=1}^{n} h_i \mathbb{E}(Y_i Y_j) = \sum_{i=1}^{n} h_i \mathbb{E}((X + Z_i)(X + Z_j)) = \sum_{i=1}^{n} h_i (P + N \cdot 2^{-|i-j|}).
\]
There are $n$ equations with $n$ unknowns:

\[
\begin{bmatrix}
P
P
\vdots
P
\end{bmatrix} =
\begin{bmatrix}
P + N & P + N/2 & \cdots & P + N/2^{n-2} & P + N/2^{n-1} \\
P + N/2 & P + N & \cdots & P + N/2^{n-3} & P + N/2^{n-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
P + N/2^{n-2} & P + N/2^{n-3} & \cdots & P + N & P + N/2 \\
P + N/2^{n-1} & P + N/2^{n-2} & \cdots & P + N/2 & P + N
\end{bmatrix}
\begin{bmatrix}
h_1 \\
h_2 \\
\vdots \\
h_{n-1} \\
h_n
\end{bmatrix}.
\]

By the hint, there are only 2 degrees of freedom given, $a$ and $b$. Solving this equation using the first 2 rows of the matrix, we obtain

\[
\begin{bmatrix}
h_1 \\
h_2 \\
\vdots \\
h_{n-1} \\
h_n
\end{bmatrix} = \begin{bmatrix}
P \\
2 \\
1 \\
\vdots \\
2
\end{bmatrix}
\begin{bmatrix}
3N + (n + 2)P
\end{bmatrix}.
\]

The minimum mean square error is

\[
\text{MSE} = E(X - \hat{X})X = P - P \sum_{i=1}^{n} h_i Y_i
\]

\[
= P \left(1 - \frac{(n + 2)P}{3N + (n + 2)P}\right) = \frac{3PN}{3N + (n + 2)P}.
\]