Question #1 (15 points) –

Commutativity of Discrete Linear Shift Invariant Operators

Show that two LSI operators, denoted by the matrices $H^{(1)}_{LSI}$ and $H^{(2)}_{LSI}$, commute, i.e., prove that $H^{(1)}_{LSI} H^{(2)}_{LSI} = H^{(2)}_{LSI} H^{(1)}_{LSI}$.

**Solution:**

The matrices $H^{(1)}_{LSI}$ and $H^{(2)}_{LSI}$ correspond to LSI operators, hence are circulant.

Note that since the operators are defined as matrices, this question considers the discrete settings (i.e., related signals and systems are discrete).

Since a circulant matrix is diagonalized by the DFT matrix, $F$, we can write

$$FH^{(1)}_{LSI} F^* = \Lambda^{(1)} \rightarrow H^{(1)}_{LSI} = F^* \Lambda^{(1)} F$$

$$FH^{(2)}_{LSI} F^* = \Lambda^{(2)} \rightarrow H^{(2)}_{LSI} = F^* \Lambda^{(2)} F$$

Accordingly,

$$H^{(1)}_{LSI} H^{(2)}_{LSI} = F^* \Lambda^{(1)} FF^* \Lambda^{(2)} F = F^* \Lambda^{(1)} \Lambda^{(2)} F$$

where the second equality relies on $FF^* = I$, which holds since DFT matrix is unitary.

Since $\Lambda^{(1)}$ and $\Lambda^{(2)}$ are diagonal matrices, their product $\Lambda^{(1)} \Lambda^{(2)}$ is formed by the scalar products of their corresponding diagonal components. Since scalar multiplication is commutative, the product of two diagonal matrices is also commutative, hence

$$\Lambda^{(1)} \Lambda^{(2)} = \Lambda^{(2)} \Lambda^{(1)}$$

Then, we can write that

$$H^{(1)}_{LSI} H^{(2)}_{LSI} = F^* \Lambda^{(1)} \Lambda^{(2)} F = F^* \Lambda^{(2)} \Lambda^{(1)} F = F^* \Lambda^{(2)} F \Lambda^{(1)} F = H^{(2)}_{LSI} H^{(1)}_{LSI}$$

Proving the commutativity of two discrete LSI operators.
**Question #2 (20 points) – LSI Systems and Derivatives**

We define the derivative operator applied on some (differentiable) signal as

\[ D\{\varphi(t)\} = \frac{d}{dt}\varphi(t) \]

Prove that LSI systems are invariant to derivation.

Specifically, for an LSI system \( \mathcal{H} \) the following holds:

\[ \mathcal{H}\{D\{\varphi_{in}(t)\}\} = D\{\mathcal{H}\{\varphi_{in}(t)\}\} \]

where \( \varphi_{in}(t) \) is the input signal to the entire processing procedure.

**Solution:**

We start by showing that derivation is a linear shift-invariant operator.

Linearity is due to the linearity of derivation (prove it by yourself).

Shift invariance:

Let us define the shifted input signal as

\[ \varphi_{in, \text{shift}}(t) = T_{t_0}\{\varphi_{in}(t)\} = \varphi_{in}(t - t_0) \]

And the derivative of the non-shifted input as

\[ \varphi_{in}'(t) = D\{\varphi_{in}(t)\} = \frac{d}{dt}\varphi_{in}(t) \]

Then the derivative of the shifted signal is

\[ D\{T_{t_0}\{\varphi_{in}(t)\}\} = D\{\varphi_{in, \text{shift}}(t)\} = \frac{d}{dt}\varphi_{in, \text{shift}}(t) = \frac{d}{dt}\varphi_{in}(t - t_0) \]

\[ = \frac{d}{dt}\{t - t_0\} \cdot \varphi_{in}'(t - t_0) = \varphi_{in}'(t - t_0) \]

and the shifted derivative result is just

\[ T_{t_0}\{D\{\varphi_{in}(t)\}\} = T_{t_0}\{\varphi_{in}'(t)\} = \varphi_{in}'(t - t_0) \]

Hence we got that

\[ D\{T_{t_0}\{\varphi_{in}(t)\}\} = T_{t_0}\{D\{\varphi_{in}(t)\}\} \]

thus derivation is shift invariant.

Since derivation is an LSI operator it can be represented by a convolution:

\[ D\{\varphi_{in}(t)\} = d(t) * \varphi_{in}(t) = \int_{-\infty}^{\infty} \varphi_{in}(\tau)d(t - \tau)d\tau \]

where \( d(t) \) is the derivation kernel function. It can be shown that \( d(t) = \delta'(t) \) i.e., it is the derivative of the (Dirac’s) delta function. Nevertheless, the solution presented here relies on the above result that derivation is an LSI operator (system).

Now we continue the proof by showing that two LSI operators commute, i.e., their application order can be switched.

Consider two LSI systems \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) which their independent operation on an input signal \( \varphi_{in}(t) \) can be written in the convolution forms
We will prove here that
\[ \mathcal{H}_2 \{ \mathcal{H}_1 \{ \varphi_{\text{in}}(t) \} \} = \mathcal{H}_1 \{ \mathcal{H}_2 \{ \varphi_{\text{in}}(t) \} \} \]

Let us denote the output signals for each of the systems as
\[ \varphi_{\text{out},1}(t) = \mathcal{H}_1 \{ \varphi_{\text{in}}(t) \} \]
\[ \varphi_{\text{out},2}(t) = \mathcal{H}_2 \{ \varphi_{\text{in}}(t) \} \]

Let us develop the left side of the equation
\[ \mathcal{H}_2 \{ \mathcal{H}_1 \{ \varphi_{\text{in}}(t) \} \} = \mathcal{H}_2 \{ \varphi_{\text{out},1}(t) \} = \int_{-\infty}^{\infty} \varphi_{\text{out},1}(\tau) h_2(t - \tau) \, d\tau \]
\[ = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \varphi_{\text{in}}(\tau - z) h_1(z) \, dz \right) h_2(t - \tau) \, d\tau \]
\[ = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \varphi_{\text{in}}(\tau - z) h_2(t - \tau) \, d\tau \right) h_1(z) \, dz \]
\[ = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \varphi_{\text{in}}(\tilde{\tau}) h_2(t - \tilde{\tau} - z) \, d\tilde{\tau} \right) h_1(z) \, dz \]
\[ = \int_{-\infty}^{\infty} \varphi_{\text{out},2}(t) h_1(z) \, dz = \int_{-\infty}^{\infty} \varphi_{\text{out},2}(t - z) h_1(z) \, dz \]
\[ = \mathcal{H}_1 \{ \varphi_{\text{out},2}(t) \} = \mathcal{H}_1 \{ \mathcal{H}_2 \{ \varphi_{\text{in}}(t) \} \} \]

Hence, we proved that the application order of two LSI systems can be changed. In this question we are given an LSI system \( \mathcal{H}\{\cdot\} \) and a derivation operator \( D\{\cdot\} \), which we proved to be an LSI operator. By our second result that two LSI systems commute, we conclude that
\[ \mathcal{H} \{ D\{\varphi_{\text{in}}(t)\} \} = D \{ \mathcal{H}\{\varphi_{\text{in}}(t)\} \} \]
Question #3 (25 points) – The Discrete Cosine Transform

The \( N \times N \) DCT matrix, \( \mathbf{U} \), is comprised from the elements defined as

\[
\mathbf{U}_{n,k} = \begin{cases} 
\frac{1}{\sqrt{N}} & \text{for } k = 1, \ 1 \leq n \leq N \\
\frac{2}{N} \cos \left( \frac{\pi (2n-1)(k-1)}{2N} \right) & \text{for } 2 \leq k \leq N, \ 1 \leq n \leq N
\end{cases}
\]

where \( u_{n,k} \) is the value of the matrix component at the \( k^{th} \) column and \( n^{th} \) row.

a. Prove that the DCT matrix \( \mathbf{U} \) is unitary.

Solution:

Let us show that the columns of \( \mathbf{U} \), denoted here as \( \mathbf{u}_1, \ldots, \mathbf{u}_N \), are orthonormal (the following inner products calculations consider the fact that \( \mathbf{U} \) is real-valued).

The first column is clearly normalized since

\[
\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \sum_{n=1}^{N} (u_{n,1})^2 = \sum_{n=1}^{N} \frac{1}{N} = 1
\]

The \( k^{th} \) column \((k = 2, \ldots, N)\) energy is

\[
\langle \mathbf{u}_k, \mathbf{u}_k \rangle = \sum_{n=1}^{N} (u_{n,k})^2 = \sum_{n=1}^{N} \left( \frac{2}{N} \cos \left( \frac{\pi (2n-1)(k-1)}{2N} \right) \right)^2
\]

\[
= \frac{2}{N} \sum_{n=1}^{N} \cos^2 \left( \frac{\pi (2n-1)(k-1)}{2N} \right)
\]

\[
= \frac{2}{N} \sum_{n=1}^{N} \frac{1}{2} \left( 1 + \cos \left( \frac{2 \pi (2n-1)(k-1)}{2N} \right) \right)
\]

\[
= \frac{1}{N} \sum_{n=1}^{N} \left( 1 + \cos \left( \frac{\pi (2n-1)(k-1)}{N} \right) \right)
\]

\[
= 1 + \frac{1}{N} \sum_{n=1}^{N} \cos \left( \frac{\pi (2n-1)(k-1)}{N} \right)
\]

\[
= 1 + \frac{1}{N} \sum_{n=1}^{N} \left( \frac{1}{2} W_N^{(n-\frac{1}{2})(k-1)} + \frac{1}{2} W_N^{-(n-\frac{1}{2})(k-1)} \right)
\]

\[
= 1 + \frac{1}{N} \left( \frac{1}{2} W_N^{-\frac{1}{2}(k-1)} \sum_{n=1}^{N} W_n^{(n-\frac{1}{2})} + \frac{1}{2} W_N^{\frac{1}{2}(k-1)} \sum_{n=1}^{N} W_n^{-(n-\frac{1}{2})} \right)
\]

\[
= 1 + \frac{1}{N} \left( \frac{1}{2} W_N^{-\frac{1}{2}(k-1)} \cdot 0 + \frac{1}{2} W_N^{\frac{1}{2}(k-1)} \cdot 0 \right)
\]

\[
= 1
\]

and the last result implies that the \( k^{th} \) column \((k = 2, \ldots, N)\) of \( \mathbf{U} \) is normalized.
Note the above sums became zeros due to the fact that $k > 1$ and that the sum of the roots of unity of order $N$ (or their nonzero powers) is zero.

Orthogonality proofs:
The inner product of the first column with another column ($k = 2, \ldots, N$) is

$$
\langle u_1, u_k \rangle = \sum_{n=1}^{N} u_{n,1} u_{n,k} = \sum_{n=1}^{N} \frac{1}{\sqrt{2}} \cdot \sqrt{\frac{2}{N}} \cos \left( \frac{\pi (2n - 1)(k - 1)}{2N} \right)
$$

$$
= \sqrt{\frac{1}{N}} \sum_{n=1}^{N} \cos \left( \frac{\pi (2n - 1)(k - 1)}{2N} \right) = (*)
$$

Again, noting that using $W_N = e^{-\frac{2\pi i}{N}}$

$$
\cos \left( \frac{\pi (2n - 1)(k - 1)}{2N} \right) = \frac{1}{2} W_N \frac{1}{2} (n^{-1} z)^{(k-1)} + \frac{1}{2} W_N^{-1} (n^{-1} z)^{(k-1)}
$$

$$
= \frac{1}{2} W_N^{\frac{1}{4} (k-1)} W_N^{\frac{1}{4} (k-1)} + \frac{1}{2} W_N^{\frac{1}{4} (k-1)} W_N^{-\frac{1}{4} (k-1)}
$$

then, returning to the computation of $\langle u_1, u_k \rangle$ shows

$$
(*) = \frac{1}{2} \sqrt{\frac{1}{N}} W_N^{-\frac{1}{4} (k-1)} \sum_{n=1}^{N} W_N^{\frac{1}{2} (k-1)} W_N^{\frac{1}{4} (k-1)} + \frac{1}{2} \sqrt{\frac{1}{N}} W_N^{\frac{1}{4} (k-1)} \sum_{n=1}^{N} W_N^{-\frac{1}{2} (k-1)} = 0
$$

where the last equality is due to the fact that $k > 1$ and that the sum of the roots of unity of order $N$ (or their nonzero powers) is zero.

This proves that $u_1$ is orthogonal to $u_k$ for $k = 2, \ldots, N$.

The inner product of the $k^{th}$ column ($k = 2, \ldots, N$) with the $l^{th}$ column ($l \neq k, \ l = 2, \ldots, N$) is
\[ (u_k, u_l) = \sum_{n=1}^{N} u_{n,k}u_{n,l} = \sum_{n=1}^{N} 2 \frac{2}{N} \cos \left( \frac{(2n-1)(k-1)}{2N} \right) \cos \left( \frac{(2n-1)(l-1)}{2N} \right) \]
\[ = \frac{1}{N} \sum_{n=1}^{N} \left[ \cos \left( \frac{(2n-1)(k-l)}{2N} \right) + \cos \left( \frac{(2n-1)(k+l-2)}{2N} \right) \right] \]
\[ = \frac{1}{2N} \sum_{n=1}^{N} W_n^{1(\frac{n-1)(k-l)}{2}} + \frac{1}{2N} \sum_{n=1}^{N} W_n^{2\frac{n-1}{N}(k+l-2)} + \frac{1}{2N} \sum_{n=1}^{N} W_n^{-\frac{n-1}{2}(k+l-2)} \]
\[ = 0 \]

where the last equality to zero is due to the fact that \( k - l \neq 0 \) and \( k + l - 2 \neq 0 \) (since \( k, l > 1 \)) and that the sum of the roots of unity of order \( N \) (or their nonzero powers) is zero.

This proves that the \( k^{th} \) and the \( l^{th} \) columns \( (l \neq k, k = 2, \ldots, N, l = 2, \ldots, N) \) are orthogonal.

To conclude this section, we proved that the columns of the DCT matrix are orthonormal, hence, the DCT matrix is unitary.

b. Is the DCT matrix the real part of the DFT matrix? Explain.

**Solution:**

The \((n, k)\) component of the DFT matrix is

\[ [DFT]_{n,k} = \frac{1}{\sqrt{N}} (W^*)^{(n-1)(k-1)} = \frac{1}{\sqrt{N}} e^{i2\pi(n-1)(k-1)/N} \]

and its real part is

\[ Re([DFT]_{n,k}) = \frac{1}{\sqrt{N}} \cos \left( \frac{2\pi(n-1)(k-1)}{N} \right) \]

that clearly differs from the \((n, k)\) component of the DCT matrix defined in the question (for \( k > 1 \)) as

\[ [DCT]_{n,k} = \sqrt{\frac{2}{N}} \cos \left( \frac{\pi(2n-1)(k-1)}{2N} \right) \]
c. Consider a class of discrete random signals that is described by zero mean and the autocorrelation matrix:

\[
R = \begin{bmatrix}
1 - \alpha & -\alpha & 0 \\
-\alpha & 1 & -\alpha \\
0 & -\alpha & 1 - \alpha
\end{bmatrix}
\]

where \(0 < \alpha < 1\) is some scalar value.

What is the PCA for this class? i.e., what is the unitary matrix that diagonalizes \(R\)?

**Solution:**

The PCA matrix, denoted here as \(V\), diagonalizes the autocorrelation matrix \(R\), i.e.,

\[
V^* RV = \Lambda
\]

where \(\Lambda\) is a diagonal matrix.

In this case where the matrix \(R\) is not circulant we need to explicitly calculate its eigenvectors (that of course form the diagonalizing matrix \(V\)):

\[
det (R - \lambda I) = det \begin{bmatrix}
1 - \alpha - \lambda & -\alpha & 0 \\
-\alpha & 1 - \lambda & -\alpha \\
0 & -\alpha & 1 - \alpha - \lambda
\end{bmatrix}
\]

\[
= (1 - \alpha - \lambda)(1 - \lambda)(1 - \alpha - \lambda) - \alpha^2(1 - \alpha - \lambda)
\]

\[
= (1 - \alpha - \lambda)(1 - \lambda - 2\alpha)(1 - \lambda + \alpha)
\]

Then, \(det (R - \lambda I) = 0\) gives

\[
(1 - \alpha - \lambda)(1 - \lambda - 2\alpha)(1 - \lambda + \alpha) = 0
\]

and the eigenvalues are:

\[
\Rightarrow \lambda_1 = 1 + \alpha \, , \, \lambda_2 = 1 - \alpha \, , \, \lambda_3 = 1 - 2\alpha \quad \Rightarrow \quad \Lambda = \begin{bmatrix}
1 + \alpha & 0 \\
0 & 1 - \alpha
\end{bmatrix}
\]

(In the following calculations \(u_1, u_2, u_3\) are variables for the purpose of the specific computations.)

Finding eigenvector #1:

\[
\begin{bmatrix}
1 - \alpha & -\alpha & 0 \\
-\alpha & 1 & -\alpha \\
0 & -\alpha & 1 - \alpha
\end{bmatrix} \begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix} = (1 + \alpha) \begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
\]

\[
(1 - \alpha)u_1 - \alpha u_2 = (1 + \alpha)u_1 \rightarrow u_1 = -\frac{1}{2}u_2 \\
-\alpha u_1 + u_2 - \alpha u_3 = (1 + \alpha)u_2 \
\]
\(-\alpha u_2 + (1 - \alpha) u_1 = (1 + \alpha) u_1, \ \rightarrow \ u_1 = -\frac{1}{2} u_2 = u_3 \ \Rightarrow \ u = \begin{bmatrix} \beta \\ -2\beta \\ \beta \end{bmatrix} \Rightarrow u = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}\)

Noting that the values of \(\beta\) were determined such that the vector is normalized.

Finding eigenvector #2:
\[
\begin{bmatrix} 1 - \alpha & -\alpha & 0 \\ -\alpha & 1 & -\alpha \\ 0 & -\alpha & 1 - \alpha \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = (1 - \alpha) \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}
\]

\((1 - \alpha) u_1 - \alpha u_2 = (1 - \alpha) u_1, \ \rightarrow u_2 = 0\)

\(-\alpha u_1 + u_2 = (1 - \alpha) u_2, \ \rightarrow u_1 = -u_3 \ \Rightarrow u = \begin{bmatrix} \beta \\ 0 \\ -\beta \end{bmatrix} \Rightarrow u = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}\)

Finding eigenvector #3:
\[
\begin{bmatrix} 1 - \alpha & -\alpha & 0 \\ -\alpha & 1 & -\alpha \\ 0 & -\alpha & 1 - 2\alpha \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = (1 - 2\alpha) \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}
\]

\((1 - \alpha) u_1 - \alpha u_2 = (1 - 2\alpha) u_1, \ \rightarrow u_1 = u_2\)

\(-\alpha u_1 + u_2 = (1 - 2\alpha) u_2, \ \rightarrow u_3 = u_2 \Rightarrow u = \begin{bmatrix} \beta \\ \beta \\ \beta \end{bmatrix} \Rightarrow u = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}\)

We have got the 3 \times 3 PCA matrix:
\[
V = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}
\]
d. Is the PCA matrix from subsection c related to the $3 \times 3$ DCT matrix?

Solution:

By the definition given in the question to the DCT matrix we see that the $3 \times 3$ matrix is

$$[DCT] = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

By changing the order of the columns in the $3 \times 3$ DCT matrix we get the PCA matrix that we found above.
**Question #4 (15 points)**

A Gaussian random variable $X \sim \mathcal{N}(\mu_x, \sigma^2_x)$ is corrupted by an additive Gaussian noise sample $N \sim \mathcal{N}(0, \sigma^2_n)$, which is statistically independent of $X$, resulting in the measured noisy variable is $Y = X + N$.

What is the probability density function of the random variable $y$? Provide a full mathematical proof.

Hint: recall that the probability density function (PDF) of a sum of two independent random variables is the convolution of their PDFs.

**Solution:**

$$X \sim \mathcal{N}(\mu_x, \sigma^2_x) \quad \rightarrow \quad p_X(x) = \frac{1}{\sqrt{2\pi\sigma_x}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$$

$$N \sim \mathcal{N}(0, \sigma^2_n) \quad \rightarrow \quad p_N(n) = \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{n^2}{2\sigma_n^2}}$$

Since $Y = X + N$, the probability density function (pdf) of the random variable $Y$ is the convolution of the pdfs of $X$ and $N$:

$$p_Y(y) = (p_X * p_N)(y) = \int_{-\infty}^{\infty} p_X(\tau)p_N(y-\tau)d\tau$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_x}} e^{-\frac{(\tau-\mu_x)^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{(y-\tau)^2}{2\sigma_n^2}} d\tau = \frac{1}{2\pi\sigma_x\sigma_n} \int_{-\infty}^{\infty} e^{-\frac{\sigma_n^2(\tau-\mu_x)^2+\sigma_x^2(y-\tau)^2}{2\sigma_x^2\sigma_n^2}} d\tau = (*)$$

Let us develop the expression in the exponent:
\[
\frac{\sigma_n^2 (\tau - \mu_x)^2 + \sigma_n^2 (y - \tau)^2}{2\sigma_n^2} = \frac{\sigma_n^2 (\tau^2 - 2\tau \mu_x + \mu_x^2) + \sigma_n^2 (y^2 - 2\tau y + \tau^2)}{2\sigma_n^2}
\]

\[
= \frac{(\sigma_x^2 + \sigma_n^2) \tau^2 - 2(\mu_x \sigma_n^2 + y \sigma_n^2) \tau + \mu_n^2 \sigma_n^2 + y^2 \sigma_n^2}{2\sigma_x^2 \sigma_n^2}
\]

\[
= \frac{\tau^2 - 2\mu_x \sigma_n^2 + y \sigma_n^2 \tau + \mu_n^2 \sigma_n^2 + y^2 \sigma_n^2}{2\sigma_x^2 \sigma_n^2}
\]

\[
= \frac{(\sigma_x^2 + \sigma_n^2) \tau^2 - 2\mu_x \sigma_n^2 + y \sigma_n^2 \tau + (\mu_n^2 \sigma_n^2 + y \sigma_n^2) - (\mu_n^2 \sigma_n^2 + y \sigma_n^2)^2 + \frac{\mu_n^2 \sigma_n^2 + y^2 \sigma_n^2}{\sigma_x^2 + \sigma_n^2}}{2\sigma_x^2 \sigma_n^2}
\]

\[
= \frac{(\tau - \mu_x \sigma_n^2 + y \sigma_n^2)^2}{2\sigma_x^2 \sigma_n^2} - \frac{(\mu_n^2 \sigma_n^2 + y \sigma_n^2)^2}{2\sigma_x^2 \sigma_n^2} + \frac{\mu_n^2 \sigma_n^2 + y \sigma_n^2}{\sigma_x^2 + \sigma_n^2}
\]

\[
= \frac{2 \sigma_x^2 \sigma_n^2}{\sigma_x^2 + \sigma_n^2} + \frac{(\mu_n^2 \sigma_n^2 + y \sigma_n^2)^2}{2\sigma_x^2 \sigma_n^2} + \frac{\mu_n^2 \sigma_n^2 + y \sigma_n^2}{\sigma_x^2 + \sigma_n^2}
\]

\[
= \frac{(\tau - \mu_x \sigma_n^2 + y \sigma_n^2)^2}{2\sigma_x^2 \sigma_n^2} + \frac{\mu_n^2 \sigma_n^2 + y \sigma_n^2}{\sigma_x^2 + \sigma_n^2}
\]

\[
= \frac{2 \sigma_x^2 \sigma_n^2}{\sigma_x^2 + \sigma_n^2} + \frac{(\mu_n^2 \sigma_n^2 + y \sigma_n^2)^2}{2\sigma_x^2 \sigma_n^2} + \frac{\mu_n^2 \sigma_n^2 + y \sigma_n^2}{\sigma_x^2 + \sigma_n^2}
\]

\[
= \frac{(\tau - \mu_x \sigma_n^2 + y \sigma_n^2)^2}{2\sigma_x^2 \sigma_n^2}
\]

Setting the last result back into the convolution expression gives us

\[
(*) = \frac{1}{2\pi \sigma_x \sigma_n} \int_{-\infty}^{\infty} e^{2 \sigma_x \sigma_n (\frac{\tau - \mu_x^2 + y \sigma_n^2}{\sigma_x^2 + \sigma_n^2})^2} \frac{(y - \mu_x)^2}{2(\sigma_x^2 + \sigma_n^2)} e^{2 \sigma_x \sigma_n (\frac{y - \mu_x^2 + y \sigma_n^2}{\sigma_x^2 + \sigma_n^2})^2} \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_n^2)}} \int_{-\infty}^{\infty} e^{-\frac{(y - \mu_x)^2}{2(\sigma_x^2 + \sigma_n^2)}} dt = (**)
\]

\[
(*) = \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_n^2)}} \int_{-\infty}^{\infty} e^{-\frac{(y - \mu_x)^2}{2(\sigma_x^2 + \sigma_n^2)}} \int_{-\infty}^{\infty} e^{-\frac{(y - \mu_x^2 + y \sigma_n^2)^2}{2\sigma_x^2 \sigma_n^2}} \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_n^2)}} dt = (**)
\]
Noting that \[ \int_{-\infty}^{\infty} e^{-\frac{(\tau - \mu_x \sigma_n^2 + y \sigma_x^2)^2}{\sigma_x^2 + \sigma_n^2}} \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2} \ d\tau = 1 \] since this is an integration over a Gaussian distribution with mean \( \frac{\mu_x \sigma_n^2 + y \sigma_x^2}{\sigma_x^2 + \sigma_n^2} \) and variance \( \frac{\sigma_x^2 \sigma_n^2}{\sigma_x^2 + \sigma_n^2} \). Then, we got from (**) that

\[ p_Y(y) = \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_n^2)}} e^{-\frac{(y - \mu_y)^2}{2(\sigma_x^2 + \sigma_n^2)}} \]

meaning that \( Y \) is a Gaussian random variable: \( Y \sim \mathcal{N}(\mu_x, \sigma_x^2 + \sigma_n^2) \).
**Question #5 (15 points)**

The signal vector $\mathbf{x}$ is corrupted by an additive noise vector $\mathbf{n}$, so the measured noisy signal is $\mathbf{y} = \mathbf{x} + \mathbf{n}$.

The observed noisy signal, $\mathbf{y}$, is linearly filtered to obtain the estimate:

$$\hat{\mathbf{x}} = \mathbf{H}\mathbf{y}$$

where $\mathbf{H}$ is the filter matrix.

Prove that for the optimal linear filter (in the minimal expected MSE sense), the error is orthogonal to the noisy measurements, i.e., show that $E[\mathbf{y}^T \mathbf{e}] = 0$,

where $\mathbf{e} \triangleq \mathbf{x} - \hat{\mathbf{x}}$ is the error signal.

**Solution:**

The error signal is

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - \mathbf{H}\mathbf{y}$$

The optimal linear filter minimizes the expected squared error, i.e.,

$$\min_{\mathbf{H}} E[\|\mathbf{e}\|_2^2]$$

and since $\|\mathbf{e}\|_2^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{x} - \mathbf{H}\mathbf{y})^T (\mathbf{x} - \mathbf{H}\mathbf{y})$, the optimization can be written as

$$\min_{\mathbf{H}} E[(\mathbf{x} - \mathbf{H}\mathbf{y})^T (\mathbf{x} - \mathbf{H}\mathbf{y})]$$

The optimal $\mathbf{H}$ can be determined by

$$\frac{\partial}{\partial \mathbf{H}} E[(\mathbf{x} - \mathbf{H}\mathbf{y})^T (\mathbf{x} - \mathbf{H}\mathbf{y})] = 0$$

Let us develop the optimization cost function:

$$E[(\mathbf{x} - \mathbf{H}\mathbf{y})^T (\mathbf{x} - \mathbf{H}\mathbf{y})] = E[\text{trace}((\mathbf{x} - \mathbf{H}\mathbf{y})^T (\mathbf{x} - \mathbf{H}\mathbf{y}))]$$

$$= E[\text{trace}((\mathbf{x} - \mathbf{H}\mathbf{y})(\mathbf{x} - \mathbf{H}\mathbf{y})^T)]$$

$$= E[\text{trace}(\mathbf{x}\mathbf{x}^T)] - E[\text{trace}(\mathbf{x}\mathbf{y}^T\mathbf{H}^T)] - E[\text{trace}(\mathbf{H}\mathbf{y}\mathbf{x}^T)] + E[\text{trace}(\mathbf{H}\mathbf{y}^T\mathbf{H}^T)]$$

$$= E[\text{trace}(\mathbf{x}\mathbf{x}^T)] - 2E[\text{trace}(\mathbf{H}\mathbf{y}\mathbf{x}^T)] + E[\text{trace}(\mathbf{H}\mathbf{y}^T\mathbf{H}^T)]$$

where we used above the linearity of the trace and expectation operators, and the trace property that $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$ for two matrices $\mathbf{A}$ and $\mathbf{B}$ (of sizes $M \times N$ and $N \times M$, respectively).
We use the obtained expression for determining the following derivative:

\[
\frac{\partial}{\partial H} E\{(x - Hy)^T(x - Hy)\} = -2 \frac{\partial}{\partial H} E\{\text{trace}(Hy^T)\} + \frac{\partial}{\partial H} E\{\text{trace}(Hy^TH^T)\}
\]

\[
= -2E\left( \frac{\partial}{\partial H} \text{trace}(Hy^T) \right) + E\left( \frac{\partial}{\partial H} \text{trace}(Hy^T H^T) \right)
\]

\[
= -2E\{yx^T\} + 2E\{Hy^T\}
\]

\[
= -2E\{yx^T\} + 2E\{Hy^T\}
\]

where we used above the derivation formulas, defined for two matrices \(A\) (of size \(M \times N\)) and \(B\) as:

\[
\frac{\partial}{\partial A} \text{trace}\{AB\} = \frac{\partial}{\partial A} \text{trace}\{BA\} = B \quad (B \text{ is of size } N \times M)
\]

\[
\frac{\partial}{\partial A} \text{trace}\{ABA^T\} = A(B + B^T) \quad (B \text{ is of size } N \times N)
\]

Demanding equality to zero for the derivative expression yields

\[
E\{Hy^T\} - E\{yx^T\} = 0
\]

implying also that

\[
\text{trace}\{E\{Hy^T\} - E\{yx^T\}\} = 0
\]

then we further develop the last expression

\[
\text{trace}\{E\{Hy^T\} - E\{yx^T\}\} = \text{trace}\{E\{Hy^T\}\} - \text{trace}\{E\{yx^T\}\}
\]

\[
= E\{\text{trace}(Hy^T)\} - E\{\text{trace}(yx^T)\}
\]

\[
= E\{\text{trace}(y^THy)\} - E\{\text{trace}(x^Ty)\}
\]

\[
= E\{y^THy\} - E\{x^Ty\}
\]

\[
= E\{y^THy\} - E\{y^Tx\}
\]

\[
= E\{y^T(\text{Hy} - x)\}
\]

and by demanding equality to zero we get

\[
E\{y^T(\text{Hy} - x)\} = 0
\]

which also means

\[
E\{y^T(x - Hy)\} = 0
\]

and

\[
E\{y^Te\} = 0
\]

as needed to prove.
**Question #6  (10 points)**

A unitary transform $\mathbf{U}$ is applied on the random vector $\mathbf{x}$, resulting in $\mathbf{y} = \mathbf{Ux}$.

The autocorrelation matrices of $\mathbf{x}$ and $\mathbf{y}$ are denoted as $\mathbf{R}_x$ and $\mathbf{R}_y$, respectively.

a. Prove that $\text{Trace}[\mathbf{R}_y] = \text{Trace}[\mathbf{R}_x]$.

b. What is the meaning of the above equality?

**Solution:**

Here the direct transform of $\mathbf{x}$ is defined as $\mathbf{y} = \mathbf{Ux}$

(Note that the direct transform is often defined as $\mathbf{y} = \mathbf{U^*x}$, however, we address here the specified settings).

a. The autocorrelation matrix of the transformed signal is

$$\mathbf{R}_y = E[\mathbf{yy}^*] = E[\mathbf{Uxx}^*\mathbf{U}^*] = \mathbf{UE}[\mathbf{xx}^*]\mathbf{U}^* = \mathbf{UR}_x\mathbf{U}^*$$

Then,

$$\text{Trace}[\mathbf{R}_y] = \text{Trace}[\mathbf{UR}_x\mathbf{U}^*] = \text{Trace}[\mathbf{U^*UR}_x] = \text{Trace}[\mathbf{R}_x]$$

due to the trace properties and since $\mathbf{U}$ is unitary.

b. The trace of an autocorrelation matrix is the sum of the variances of all the vector components, which is the expected energy of the signal. The above result shows that the energy of the signal is preserved under unitary transformation.