Recitation 6 – Estimation of Random Vector from a Random Vector

Definition of the problem:

- $\underline{X}$ – the estimated vector
- $\underline{Y}$ – the measurements vector
- $\underline{\hat{X}} = g(\underline{Y})$ – the estimator
- $\underline{e} = \underline{X} - \underline{\hat{X}}$ – the estimation error
- $d(\underline{X}, \underline{\hat{X}}) = \sum_{i=1}^{n} d(X_i, \hat{X}_i)$ – an additive distortion measure

Important examples for additive distortion measure:

1. Optimal estimator in the sense of minimum probability of error:

$$\text{Minimize } \sum_{i=1}^{n} P(\hat{X}_i \neq X_i) \Leftrightarrow \text{ minimize } P(\hat{X}_i \neq X_i) \qquad i = 1, \ldots, n$$

2. Optimal estimator in the sense of mean squared error (MSE):

$$\text{Minimize } E[\|\underline{e}\|^2] = E\left[\sum_{i=1}^{n} e_i^2\right] = \sum_{i=1}^{n} E[e_i^2] \Leftrightarrow minimize\ E[e_i^2] \quad i = 1, \ldots, n$$

3. Optimal **linear** estimator in the sense of mean squared error (MSE).

Reminder – MSE estimation of a random vector from another random vector:

The estimator $\underline{\hat{X}} = g(\underline{Y})$ is optimal in the sense of MSE (minimizes the MSE) iff the estimation error is orthogonal to any function of the measurements:

$$E[\underline{e}\, h(\underline{Y})] = \underline{0} \qquad \forall h(\underline{y})$$

The only solution is the conditional expectation estimator:

$$\underline{\hat{X}} = E[\underline{X} \mid \underline{Y}]$$

Notice that for all i, the estimator of X_i is the conditional expectation of X_i given $\underline{Y}$:

$$\hat{X}_i = E[X_i \mid \underline{Y}]$$

That is to say, each one of the elements of $\underline{X}$ is optimally estimated separately.

Covariance matrix of the estimation error:

$$C_{\underline{e}} = E\left\{(\underline{X} - \underline{\hat{X}}_{MMSE})(\underline{X} - \underline{\hat{X}}_{MMSE})^T\right\} = C_{\underline{X}} - C_{\underline{\hat{X}}_{MMSE}}$$

On the diagonal of this matrix appear the **error variances** of the estimators $\hat{X}_i$ (namely, the MSEs are scalars).

Reminder – optimal linear estimation in the sense of MSE:

* An estimator with the form $\hat{X} = A\underline{Y} + \underline{b}$ is desired, which minimizes the mean squared error (MSE).
* An estimator is optimal iff its error is orthogonal to all **linear** functions of the measurements.
* If the measurements vector $\underline{Y}$ has an invertible covariance matrix, the optimal linear estimator is:

$$\underline{\hat{X}}_{BLE} = \underline{\eta}_X + C_{\underline{X}\underline{Y}} C_{\underline{Y}}^{-1} (\underline{Y} - \underline{\eta}_Y)$$

* The covariance matrix of the estimation error is given by:

$$C_{\underline{e}} = E\left\{ (\underline{X} - \underline{\hat{X}}_{BLE})(\underline{X} - \underline{\hat{X}}_{BLE})^T \right\} = C_{\underline{X}} - C_{\underline{X}\underline{Y}} C_{\underline{Y}}^{-1} C_{\underline{Y}\underline{X}}$$

An interesting case is the passing of random vector $\underline{X}$ through a noisy linear system, i.e. $\underline{Y} = H\underline{X} + \underline{N}$, where $\underline{X}$ and $\underline{N}$ are uncorrelated random vectors with expectation zero and covariance matrices $C_{\underline{X}}$ and $C_{\underline{N}}$, respectively.

The optimal linear estimator given $\underline{Y}$ in this case is:

$$\underline{\hat{X}}_{BLE} = \underline{\eta}_X + C_{XY} C_Y^{-1} (\underline{Y} - \underline{\eta}_Y) = C_{XY} C_Y^{-1} \underline{Y} = C_X H^T \left(H C_X H^T + C_N \right)^{-1} \underline{Y}$$

The covariance matrix of the estimation error will be:

$$C_{ee}^{MMSE} = C_X - C_X H^T (H C_X H^T + C_N)^{-1} H C_X$$

<u>Problem 1:</u>

Given is the Gaussian random vector $\underline{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}^T$, which is distributed as:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix} \right)$$

The above vector passes through a noisy channel, and the following measurements are taken:

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}}_{H} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix}$$

where $\underline{N} = \begin{bmatrix} N_1 & N_2 \end{bmatrix}^T$ is a noise vector independent of $\underline{X}$ that satisfies:

$$\begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

1. Prove that the six variables $(X_1, X_2, N_1, N_2, Y_1, Y_2)$ are jointly Gaussian.

2. Find the optimal estimator in the sense of MMSE of $\underline{X}$ from $\underline{Y}$.

3. Find the optimal estimator in the sense of MMSE of X_1 from $\underline{Y}$.

4. To what do the coefficients of the estimator from section 2 approach in both the limits $\gamma \to 0$ (low signal to noise ratio) and $\gamma \to \infty$ (high signal to noise ratio)?

Solution:

1. The two vectors $\begin{bmatrix} X_1 & X_2 \end{bmatrix}$, $\begin{bmatrix} N_1 & N_2 \end{bmatrix}$ are independent Gaussian vectors, thus also their combination $\begin{bmatrix} X_1 & X_2 & N_1 & N_2 \end{bmatrix}$ is a Gaussian random vector. The vector $(X_1, X_2, N_1, N_2, Y_1, Y_2)$ is a linear transformation of the above vector, and is, therefore, also a Gaussian random vector.

2. Based on section 1, all the elements of the vectors $\underline{X}, \underline{Y}$ are jointly Gaussian, and, therefore, the optimal MMSE estimator of $\underline{X}$ from $\underline{Y}$ is also the optimal linear estimator –moreover, the estimation error vector $\underline{e} = \underline{X} - \underline{\hat{X}}$ is **<u>independent</u>** on the random vector $\underline{Y}$.
   Let us calculate the optimal linear estimator (in the sense of MMSE) of $\underline{X}$ from $\underline{Y}$. In our case:
$$\underline{\eta}_Y = E\left\{\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}\right\} = H \cdot E\left\{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right\} + E\left\{\begin{bmatrix} N_1 \\ N_2 \end{bmatrix}\right\} = \underline{0}$$
   In order to find the covariance matrix of $\underline{Y}$, let us define $\underline{R} = H \cdot \underline{X}$; namely, $\underline{Y} = \underline{R} + \underline{N}$. It is clear that $\underline{R}$ and $\underline{N}$ are independent, therefore:
$$C_Y = C_R + C_N = HC_X H^T + C_N$$
   The cross-covariance matrix between $\underline{X}$ and $\underline{Y}$ is:
$$C_{XY} = E\left\{\underline{X}\,\underline{Y}^T\right\} - \underbrace{E\left\{\underline{X}\right\}E\left\{\underline{Y}^T\right\}}_{0} = E\left\{\underline{X}\left(H \cdot \underline{X} + \underline{N}\right)^T\right\} = E\left\{\underline{X}\,\underline{X}^T H^T\right\} + E\left\{\underline{X}\,\underline{N}^T\right\} =$$
$$= \underbrace{E\left\{\underline{X}\,\underline{X}^T\right\}}_{C_X} H^T + \underbrace{E\left\{\underline{X}\right\}E\left\{\underline{N}^T\right\}}_{0} = C_X H^T$$
   We get the estimator:
$$\underline{\hat{X}}_{BLE} = \underline{\eta}_X + C_{XY}C_Y^{-1}(\underline{Y} - \underline{\eta}_Y) = C_{XY}C_Y^{-1}\underline{Y} = C_X H^T\left(HC_X H^T + C_N\right)^{-1}\underline{Y}$$
   where:
$$HC_X H^T + C_N = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \gamma\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5\gamma+1 & 4\gamma \\ 4\gamma & 5\gamma+1 \end{bmatrix}$$
   Let us substitute and we get:

$$\underline{\hat{X}}_{BLE} = C_X H^T \left(HC_X H^T + C_N \right)^{-1} \underline{Y} = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5\gamma+1 & 4\gamma \\ 4\gamma & 5\gamma+1 \end{bmatrix}^{-1} \underline{Y} =$$

$$= \gamma \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5\gamma+1 & 4\gamma \\ 4\gamma & 5\gamma+1 \end{bmatrix}^{-1} \underline{Y} = \frac{\gamma}{(5\gamma+1)^2-(4\gamma)^2} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5\gamma+1 & -4\gamma \\ -4\gamma & 5\gamma+1 \end{bmatrix} \underline{Y} =$$

$$= \frac{\gamma}{9\gamma^2+10\gamma+1} \begin{bmatrix} 6\gamma+2 & -3\gamma+1 \\ -3\gamma+1 & 6\gamma+2 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \frac{\gamma}{9\gamma^2+10\gamma+1} \begin{bmatrix} (6\gamma+2)Y_1+(-3\gamma+1)Y_2 \\ (-3\gamma+1)Y_1+(6\gamma+2)Y_2 \end{bmatrix}$$

3. Let $\underline{\hat{X}}_{BLE} = \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \end{bmatrix}$ be the estimator from section 2. Then $\hat{X}_1$ is the optimal linear estimator of X_1 from $\underline{Y}$ and, similarly, $\hat{X}_2$ is the optimal linear estimator of X_2 from $\underline{Y}$.

4. From section 2, the coefficients of the estimator are the elements of the matrix:

$$A = \frac{1}{9\gamma^2+10\gamma+1} \begin{bmatrix} 6\gamma^2+2\gamma & -3\gamma^2+\gamma \\ -3\gamma^2+\gamma & 6\gamma^2+2\gamma \end{bmatrix}.$$

   Let us take a look at the limit of the matrix A in both of the following cases:

   1. $\gamma \to 0$ (low signal to noise ratio):

   $$A \to \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

   If the noise is very strong compared to $\underline{X}$, then the optimal estimator completely ignores the measurements and is just equal to zero.

   2. $\gamma \to \infty$ (high signal to noise ratio):

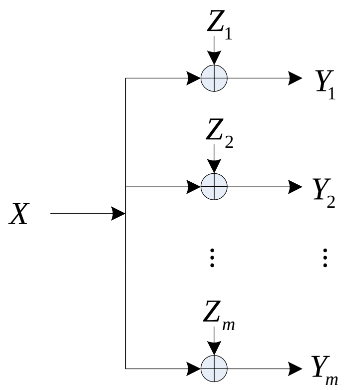
   $$A \to \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} = H^{-1}$$

   Namely, the optimal estimator is the inverse matrix of the channel.

Problem 2:

Consider the following communication system:

The input is $X \in \{-1,+1\}$ with equal probability, and the outputs are $Y_i = X + Z_i$, $i = 1,\ldots,m$. The noises $Z_1,\ldots,Z_m$ are i.i.d Gaussian noises (independent of X), with expected value zero and variance σ^2.



1. Prove that the <u>optimal</u> estimator of X from **only a specific** Y_i, in the sense of minimum probability of error, is given by $\hat{X}_i = sign(Y_i) = \begin{cases} +1 & Y_i \geq 0 \\ -1 & Y_i < 0 \end{cases}$. What is the estimation error?

2. Prove that the <u>optimal</u> estimator of X from **all** the measurements, in the sense of minimum probability of error, is given by $\hat{X} = sign\left(\sum_{i=1}^{m} Y_i\right)$.

3. Find the optimal estimator (in the sense of minimum probability of error) of X from $\underline{\hat{X}} = \left[\hat{X}_1 \ \hat{X}_2 \ \cdots \ \hat{X}_m\right]^T$ (using the definition from section 1).

4. For $m = 2$, an engineer suggested to estimate $\hat{X}_{eng} = \hat{X}_1$ (an estimator that ignores the value of $\hat{X}_2$). Will the probability of error when using this estimator be better/worse/equal to that of the optimal estimator $\hat{X}$ (from section 3)?

<u>Solution:</u>

1. The optimal estimator in the sense of minimum probability of error is given by MAP rule.

   Given the value of X, Y is distributed normally with expectation X and variance σ^2. We will use Bayes' theorem and get:

$$\begin{aligned}\hat{X}_{MAP} &= \arg\max_{x\in\{-1,+1\}} P_{X|Y_i}(x \mid y_i) = \arg\max_{x\in\{-1,+1\}} \frac{f_{Y_i|X}(y_i \mid x) P_X(x)}{f_{Y_i}(y_i)} \\ &= \arg\max_{x\in\{-1,+1\}} f_{Y_i|X}(y_i \mid x) = \arg\max_{x\in\{-1,+1\}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-x)^2}{2\sigma^2}} \\ &= \arg\min_{x\in\{-1,+1\}} |y_i - x|\end{aligned}$$

   That is to say – given y_i, $\hat{X}_{MAP}$ is the estimator that brings $|y_i - x|$ to a minimum, namely, **the value** x **that is closest to** y_i. For positive values of y_i, this value is $+1$, and for negative values of y_i, it is -1, and, all in all, $sign(y_i)$. In other words, the optimal estimator is a **slicer**.

   > <u>Note:</u> this estimator is optimal only when the distribution of the input is uniform, i.e. the a-priori distribution of X is $P(X=1) = P(X=-1) = \frac{1}{2}$.

   The probability of error is given by:

$$
\begin{aligned}
P_e &= P(\hat{X} \neq X) \\
&= \frac{1}{2} P\{Y_i > 0 \mid X = -1\} + \frac{1}{2} P\{Y_i < 0 \mid X = 1\} \\
&= \frac{1}{2} P\{Y_i - X > 1 \mid X = -1\} + \frac{1}{2} P\{Y_i - X < -1 \mid X = 1\} \\
&= \frac{1}{2} \underbrace{P\{Z_i > 1\}}_{Q\left(\frac{1}{\sigma}\right)} + \frac{1}{2} \underbrace{P\{Z_i < -1\}}_{Q\left(\frac{1}{\sigma}\right)} = Q\left(\frac{1}{\sigma}\right)
\end{aligned}
$$

2. Here, too, the optimal estimator is given by MAP rule. Given X, the measurements are independent and Gaussian:

$$
f_{\underline{Y}|X}\left(\underline{y} \mid x\right) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - x)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-m/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - x)^2\right] =
$$

$$
= \left(2\pi\sigma^2\right)^{-m/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{m} \left(y_i^2 - 2xy_i + x^2\right)\right]
$$

Let us use Bayes' law once more and get that the optimal estimator is:

$$
\begin{aligned}
\hat{X}_{MAP} &= \arg\max_{x \in \{-1,+1\}} P_{X|\underline{Y}}\left(x \mid \underline{y}\right) = \arg\max_{x \in \{-1,+1\}} f_{\underline{Y}|X}\left(\underline{y} \mid x\right) = \\
&= \arg\max_{x \in \{-1,+1\}} \left(2\pi\sigma^2\right)^{-m/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{m} \left(y_i^2 - 2xy_i + x^2\right)\right] = \\
&= \arg\min_{x \in \{-1,+1\}} \sum_{i=1}^{m} \left(y_i^2 - 2xy_i + x^2\right) = \arg\max_{x \in \{-1,+1\}} x \sum_{i=1}^{m} y_i
\end{aligned}
$$

Now, if $\sum_{i=1}^{m} y_i > 0$, then the maximum is achieved for $x = +1$, whereas if $\sum_{i=1}^{m} y_i < 0$, then the maximum is achieved for $x = -1$. Therefore, the MAP estimator is:

$$
\hat{X}_{MAP} = sign\left(\sum_{i=1}^{m} y_i\right)
$$

3. Now the measurements we have are the estimators' outputs $\hat{X}_1, \ldots, \hat{X}_m$. From these measurements, we would like to estimate the value of X, in the sense of minimum probability of error.

The optimal decision rule is, as usual, the MAP rule:

$$
\hat{X}_{MAP} = \arg\max_{x \in \{-1,+1\}} P_{X|\underline{\hat{X}}}\left(x \mid \underline{\hat{x}}\right) = \arg\max_{x \in \{-1,+1\}} P_{\underline{\hat{X}}|X}\left(\underline{\hat{x}} \mid x\right)
$$

Now, notice that **given the value of** X, the estimators $\hat{X}_1, \ldots, \hat{X}_m$ are mutually independent, thus:

$$\hat{X}_{MAP} = \arg\max_{x\in\{-1,+1\}} \prod_{i=1}^{m} P_{\hat{X}_i|X}(\hat{x}_i \mid x)$$

As seen in section 1 – given the value of X, $\hat{X}_i$ differs from X with probability $Q\left(\frac{1}{\sigma}\right)$, and is equal to X with probability $1-Q\left(\frac{1}{\sigma}\right)$. Since $Q\left(\frac{1}{\sigma}\right) < 1-Q\left(\frac{1}{\sigma}\right)$, **the product will be maximal iff x is chosen to be that which the greatest amount of values of $\hat{x}_i$ are equal to it** – namely, the estimator operates on the principle "the majority determines":

$$\hat{X}_{MAP} = \begin{cases} 1 & \sum_{i=1}^{m} \hat{X}_i > 0 \\ 1\ or\ -1 & \sum_{i=1}^{m} \hat{X}_i = 0 \\ -1 & \sum_{i=1}^{m} \hat{X}_i < 0 \end{cases}$$

If there is a "tie", then 1 or -1 can be chosen arbitrarily – in both cases we will get the same probability of error.

4. Let us write the optimal estimators for $m = 2$:

$\hat{X}_1$	$\hat{X}_2$	$\hat{X}$
1	1	1
1	−1	?
−1	1	?
−1	−1	−1

As mentioned in section 3, for a "tie" both values of the estimator (± 1) result in the same probability of error so it does not matter which value is chosen and $\hat{X}$ will still be optimal. One of the legitimate choices is:

$\hat{X}_1$	$\hat{X}_2$	$\hat{X}$
1	1	1
1	−1	1
−1	1	−1
−1	−1	−1

This is exactly the estimator suggested by the engineer thus the probability of error will be the same in both cases (in other words, the estimator that the engineer suggested is optimal in the sense of minimum probability of error).

Extra Questions

Problem 3:

1. Prove or refute: given three random variables X, Y, Z, when Z and Y satisfy the relation $Z = g(Y)$, where $g(\cdot)$ is a deterministic function, let us denote:
   a. $\hat{Y}_{MMSE}$ - the optimal estimator in the sense of MMSE of Y given X.
   b. $\hat{Z}_{MMSE}$ - the optimal estimator in the sense of MMSE of Z given X.

   then the following necessarily holds: $\hat{Z}_{MMSE} = g(\hat{Y}_{MMSE})$.

2. Does the relation from section 1 necessarily exist when $Z = g(Y)$ is a linear function?

3. Repeat sections 1 and 2 for the optimal **linear** estimator (in the sense of minimizing MSE).

Solution:

1. The claim is not true.
   Counter-example: let us define X, Y to be two standard Gaussian random variables independent of each other, and let us define $Z = Y^2$:
$$\hat{Y}_{MMSE} = E[Y \mid X] = E[Y] = 0$$
$$\hat{Z}_{MMSE} = E[Z \mid X] = E[Y^2 \mid X] = E[Y^2] = 1 \neq 0^2$$

2. If g is linear then the claim is true:
$$Z = aY + b$$
$$\Rightarrow \hat{Z}_{MMSE} = E[aY + b \mid X] = aE[Y \mid X] + b = a\hat{Y}_{MMSE} + b$$

   Another way of seeing this is through the perpendicularity principle:
   Let us assume we know that $\hat{Y}_{MMSE}$ is the optimal MMSE estimator of Y from X. We are interested in testing the estimator $\hat{Z} = a\hat{Y}_{MMSE} + b$ as a "candidate" for the optimal MMSE estimator of Z from X. For this cause, it is sufficient to show that the estimation error, $Z - \hat{Z}$, is orthogonal to all functions of the measurement:
$$Z = aY + b$$
$$\hat{Z} = a\hat{Y}_{MMSE} + b$$
$$\Rightarrow E\left[\left(Z - \hat{Z}\right) h(X)\right] = E\left[\left(aY + b - a\hat{Y}_{MMSE} - b\right) h(X)\right] = a \underbrace{E\left[\left(Y - \hat{Y}_{MMSE}\right) h(X)\right]}_{=0} = 0$$

We got that the estimation error $Z-\hat{Z}$ is orthogonal to all function of the measurement X, namely $\hat{Z}$ is indeed the optimal MMSE estimator of Z from X:

$$\hat{Z}_{MMSE}=\hat{Z}=a\hat{Y}_{MMSE}+b$$

3. Section 1 is **not true** also for the optimal linear estimator – this can be shown using the **same** counter-example!

   Section 2 is **true** also for the optimal linear estimator – we will, again, prove it using the perpendicularity principle, only that this time we will limit h to be a **linear** function, i.e. $h(X)=\alpha X+\beta$:

$$Z=aY+b$$

$$\hat{Z}=a\hat{Y}_{BLE}+b$$

$$\Rightarrow E\left[\left(Z-\hat{Z}\right)h(X)\right]=E\left[\left(aY+b-a\hat{Y}_{BLE}-b\right)h(X)\right]=a\underbrace{E\left[\left(Y-\hat{Y}_{BLE}\right)h(X)\right]}_{=0}=0$$

   We got that the estimation error $Z-\hat{Z}$ is orthogonal to all **linear** function of the measurement X, namely $\hat{Z}$ is indeed the optimal linear MMSE estimator of Z from X:

$$\hat{Z}_{BLE}=\hat{Z}=a\hat{Y}_{BLE}+b$$

   Note: the above expression can easily be generalized to the vector case.

Problem 4:

Given is a sequence of i.i.d random variables: $\{A_n\}_{n=1}^{\infty}$, having expected value μ_A.

Let us define its partial summation sequence S_n as follows:

$$S_n=\sum_{k=1}^{n}A_k$$

1. Find the optimal estimator (in the sense of MMSE) of A_1 given the measurements $\{S_m\}_{m=n}^{\infty}$.

2. Find the optimal estimator (in the sense of MMSE) of S_n given $\{S_k\}_{k=1}^{m}$, where $m\le n$.

3. Repeat sections 1 and 2 for the optimal **linear** estimator (in the sense of MSE).

Solution:

1. First, let us point out that if the measurements $\{S_m\}_{m=n}^{\infty}$ are given, then, in fact, the measurements $\{S_n,A_{n+1},A_{n+2},\cdots\}$ are given. That is to say:

$$\{S_n = s_n, S_{n+1} = s_{n+1}, ...\} \quad \Leftrightarrow \quad \{S_n = s_n, A_{n+1} = s_{n+1} - s_n, A_{n+2} = s_{n+2} - s_{n+1}, ...\}$$

Now, notice that all the random variables $\{A_{n+1}, A_{n+2}, \cdots\}$ are mutually independent of the pair of random variables A_1, S_n, so the optimal estimator in the sense of MMSE is:

$$\hat{A}_1 = E\{A_1 \mid S_n, S_{n+1}, ...\} = E\{A_1 \mid S_n, A_{n+1}, A_{n+2}, ...\} = E\{A_1 \mid S_n\}$$

In order to calculate this expectation, we will use symmetry: if $E[A_1 \mid S_n] = \hat{A}_1$, then, since the random variables $\{A_1, \cdots, A_n\}$ are i.i.d, for reasons of symmetry it holds also that:

$$\begin{gathered} E[A_1 \mid S_n] = \hat{A}_1 \\ E[A_2 \mid S_n] = \hat{A}_1 \\ \vdots \\ E[A_n \mid S_n] = \hat{A}_1 \end{gathered}$$

Let us add all these equations to get:

$$\begin{aligned} & E[A_1 \mid S_n] + E[A_2 \mid S_n] + \cdots + E[A_n \mid S_n] = n \cdot \hat{A}_1 \\ & \Rightarrow E[A_1 + A_2 + \cdots + A_n \mid S_n] = n \cdot \hat{A}_1 \\ & \Rightarrow E[S_n \mid S_n] = n \cdot \hat{A}_1 \\ & \Rightarrow S_n = n \cdot \hat{A}_1 \\ & \Rightarrow \hat{A}_1 = \frac{S_n}{n} \end{aligned}$$

2. Here, too, the MMSE estimator is the conditional expectation estimator:

$$E\{S_n \mid S_m, S_{m-1}, S_{m-2}, ...S_1\} = E\left\{S_m + \sum_{k=m+1}^{n} A_k \mid S_m, S_{m-1}, S_{m-2}, ...S_1\right\} =$$

$$= S_m + E\left\{\sum_{k=m+1}^{n} A_k \mid S_m, S_{m-1}, S_{m-2}, ...S_1\right\} = S_m + \sum_{k=m+1}^{n} E\{A_k \mid S_m, S_{m-1}, S_{m-2}, ...S_1\} =$$

$$= S_m + \sum_{k=m+1}^{n} E\{A_k\} = S_m + (n-m) \cdot \mu_A$$

3. As recalled, when we say "linear estimator", we actually mean an affine estimator, i.e. an estimator of the form $\hat{X} = aY + b$.

   Both of the estimators we found in the previous sections are of this form, and, thus, these are also the optimal linear MMSE estimators.