Another example of finding the MVUE using sufficient statistic

- Consider tossing a biased coin for $N$ times.
- Suppose the coin is biased such that the probability of a head is $\theta$ and that of a tail is $1 - \theta$.
- The coin is tossed $N$ times resulting in a sequence $x[n] \in \{0, 1\}$, for $n = 0, 1, \ldots, N - 1$. Here, one denotes a head and zero a tail.
Another example of finding the MVUE using sufficient statistic

- The probability of the sequence is given by the Bernoulli distribution:

\[ p(x; \theta) = \theta \sum x[n] (1 - \theta)^{N - \sum x[n]}, \]

Note that \( \sum x[n] \) is simply the number of heads in the sequence.
Another example of finding the MVUE using sufficient statistic

- The task is to find the MVUE for $\theta$. The Neyman-Fisher factorization is again simple:

$$p(x; \theta) = \theta \sum x[n] (1 - \theta)^{N - \sum x[n]} \cdot \frac{1}{g(T(x), \theta) h(x)}$$

where $T(x) = \sum x[n]$ is the sufficient statistic.

- Now it remains to find a function $f$ of $T(x)$ that makes it unbiased.
Another example of finding the MVUE using sufficient statistic

- Let’s see the bias of $T(x)$ first:

$$
E[T(x)] = E\left[ \sum_{n=0}^{N-1} x[n] \right] = \sum_{n=0}^{N-1} E[x[n]] = \sum_{n=0}^{N-1} \theta = N\theta.
$$

- Thus, division by $N$ makes an unbiased estimator:

$$
\hat{\theta} = \frac{\sum_{n=0}^{N-1} x[n]}{N}.
$$

Unbiasedness check:

$$
E[\hat{\theta}] = \frac{\sum_{n=0}^{N-1} E[x[n]]}{N} = \frac{N\theta}{N} = \theta.
$$
Best linear unbiased estimator (BLUE)
Best linear unbiased estimator (BLUE)

- This time we’ll look into a simple subclass of all estimators: linear estimators.
- If we don’t know the pdf of the data or we are not willing to model it we cannot apply the CRLB or RBLS theorem.
- Even if the pdf is known the CRLB and RBLS theorem may not produce the MVU estimator.
- Then, we might resort to some sub-optimal estimator.
- The BLUE restricts the estimator to be linear in data.
- The BLUE can be determined only with the knowledge of the first two moments (instead of the complete PDF).
BLUE - definition

• The BLUE restricts the estimator to be linear in data

\[ \hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]. \]

• To determine the BLUE we must determine \( a_n \) so that the estimator is unbiased and has minimum variance.

• The BLUE will be optimal (the MVU estimator) if the MVU estimator is linear in data, otherwise it is sub-optimal.

• If the MVU estimator is nonlinear in data, then the BLUE is sub-optimal and the difference in performance can be substantial.
**BLUE - definition**

- For some estimation problems the use of the BLUE can be totally inappropriate.
  - For example in the estimation of the power of WGN, all linear estimators will be biased, hence BLUE does not exist.
    - However, if we can use transformed data: \( y[n] = x^2[n] \), the BL estimator becomes viable
      \[
      \hat{\sigma}^2 = \sum_{n=0}^{N-1} a_n x^2[n]
      \]
  - Sometimes their performance is low compared to the MVUE; for example in the estimation of the mean of an uniform distribution, whose BLUE is the sample mean.
Finding the BLUE

• We constrain $\hat{\theta}$ to be linear and unbiased and then find the $a_n$’s to minimize the variance.
• The unbiasedness constraint says that:

$$E(\hat{\theta}) = \sum_{n=0}^{N-1} a_n E(x[n]) = \theta$$  \hspace{1cm} (1)

• The problem is to minimize the variance $\text{var}(\hat{\theta})$:

$$\text{var}(\hat{\theta}) = E \left[ \left( \sum_{n=0}^{N-1} a_n x[n] - E\left( \sum_{n=0}^{N-1} a_n x[n] \right) \right)^2 \right].$$
Finding the BLUE

- Using vector notation, this becomes easier to read. Let’s denote \( \mathbf{a} = [a_0, a_1, \ldots, a_{N-1}]^T \):

\[
\text{var}(\hat{\theta}) = \mathbb{E}\left[ (\mathbf{a}^T \mathbf{x} - \mathbf{a}^T \mathbb{E}(\mathbf{x}))^2 \right] \\
= \mathbb{E}\left[ (\mathbf{a}^T (\mathbf{x} - \mathbb{E}(\mathbf{x})))^2 \right] \\
= \mathbb{E}\left[ \mathbf{a}^T (\mathbf{x} - \mathbb{E}(\mathbf{x}))(\mathbf{x} - \mathbb{E}(\mathbf{x}))^T \mathbf{a} \right] \\
= \mathbf{a}^T \mathbf{C} \mathbf{a}. \tag{2}
\]

- To satisfy the unbiasedness constraint, \( \mathbb{E}(\mathbf{x}[n]) \) must be linear in \( \theta \): \( \mathbb{E}(\mathbf{x}[n]) = s[n] \theta \), where \( s[n] \) is a known signal.
Finding the BLUE

• What does this representation mean: what kind of problems is the BLUE good for?
• Note, that if we write $x[n]$ as

$$ x[n] = E(x[n]) + (x[n] - E(x[n])) , $$

we can view the random term $(x[n] - E(x[n]))$ as noise $w[n]$: 

$$ x[n] = s[n]\theta + w[n] $$

• Thus, the BLUE is applicable to problems, which can be modeled as above.
• *But isn’t this exactly the linear model?*
Finding the BLUE

- No, the linear model assumes Gaussian noise. Here we only assume the knowledge of the first two moments of the data: \( s[n] \) and \( C \).

1. If your noise is Gaussian, use \( \hat{\theta} = (H^T H)^{-1} H^T x \).

2. If not (and you know the first 2 moments instead), use the BLUE estimator we’re about to derive.

3. If you are unable to assume even the mean and covariance, use Least Squares instead.

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>MVUE guaranteed?</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Model</td>
<td>Yes</td>
<td>( \hat{\theta} = (H^T H)^{-1} H^T x )</td>
</tr>
<tr>
<td>BLUE</td>
<td>No</td>
<td>( \hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} x )</td>
</tr>
<tr>
<td>(Weighted) Least Squares</td>
<td>No</td>
<td>( \hat{\theta} = (H^T W H)^{-1} H^T W x )</td>
</tr>
</tbody>
</table>
Finding the BLUE

- The linearity of $E(x[n])$ gives the unbiasedness constraint the following form

\[
\sum_{n=0}^{N-1} a_n E(x[n]) = \theta
\]

\[\Leftrightarrow \sum_{n=0}^{N-1} a_n s[n] \theta = \theta\]

\[\Leftrightarrow \sum_{n=0}^{N-1} a_n s[n] = 1\]
Finding the BLUE

which we can write more briefly as

$$\mathbf{a}^T \mathbf{s} = 1.$$ 

• Now we arrive at a constrained linear programming problem:

$$\text{Minimize } \text{var}(\hat{\theta}) = \mathbf{a}^T \mathbf{C} \mathbf{a} \text{ subject to the constraint } \mathbf{a}^T \mathbf{s} = 1.$$ 

• This problem can be solved using Lagrange multipliers as follows.
The method of Lagrangian multipliers adds the constraint as a penalty term into the objective function and after solving sets the penalty to zero. In our case the Lagrangian function to be minimized becomes

\[ J(a) = a^T Ca + \lambda(a^T s - 1) \]

Note that when the constraint is satisfied, this is equal to the objective function.
Finding the BLUE

- Differentiating with respect to $a$ gives (see the differentiation rules given on slide 4 of lecture 3):

$$\frac{\partial J}{\partial a} = 2Ca + \lambda s$$

Setting this equal to zero gives the optimum $a_{opt}$:

$$2Ca_{opt} + \lambda s = 0, \quad \text{or}$$

$$a_{opt} = -\frac{\lambda}{2C^{-1}}s$$
Finding the BLUE

• Next, we get rid of $\lambda$ by choosing the particular value of $\lambda$ that actually forces the constraint to hold. Our constraint requires that $a^T s = 1$, which has to hold also for the above optimum $a_{opt}$. Thus,

$$a_{opt}^T s = s^T a_{opt} = -\frac{\lambda}{2} s^T C^{-1} s = 1.$$

• Solving for $\lambda$ gives

$$\lambda = -\frac{2}{s^T C^{-1} s}$$
Finding the BLUE

- Substituting this into the equation of $a_{opt}$ gives the optimal coefficient vector:

$$a_{opt} = -\frac{\lambda}{2} C^{-1} s = \frac{C^{-1} s}{s^T C^{-1} s}$$

- Since

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] = a^T x,$$

we get the BLUE as

$$\hat{\theta}_{\text{BLUE}} = a_{opt}^T x = \frac{s^T C^{-1} x}{s^T C^{-1} s}$$
Finding the BLUE

• Furthermore, the variance of $\hat{\theta}$ is

$$\text{var}(\hat{\theta}) = \frac{1}{s^T C^{-1} s}$$
Example: DC Level in White Noise

- DC level in White Noise (note: not necessarily Gaussian):
  \[ x[n] = A + w[n], \quad n = 0, 1, \ldots, N - 1 \]

- Now \( s = 1 \) and \( C = \sigma^2 I \). Thus, the BLUE is given by
  \[
  \hat{A} = \frac{s^T C^{-1} x}{s^T C^{-1} s} = \frac{1^T \frac{1}{\sigma^2} I x}{1^T \frac{1}{\sigma^2} I 1} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]
  \]

- Moreover,
  \[
  \text{var}(\hat{A}) = \frac{1}{s^T C^{-1} s} = \frac{1}{1^T \frac{1}{\sigma^2} 1} = \frac{\sigma^2}{N}
  \]
Example: DC Level in Uncorrelated Noise

- Here, uncorrelated noise means that the samples are uncorrelated, but may have different distributions. Otherwise the model is the same:

\[ x[n] = A + w[n], \quad n = 0, 1, \ldots, N - 1 \]

We only need to know the variances (\( \text{var}(w[n]) = \sigma_n^2 \)) and the mean (\( s = 1 \)).
Example: DC Level in Uncorrelated Noise

- The covariance matrix is now diagonal:

\[
C = \begin{pmatrix}
\sigma_0^2 & 0 & \cdots & 0 \\
0 & \sigma_1^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_{N-1}^2
\end{pmatrix}
\]
Example: DC Level in Uncorrelated Noise

- The inverse becomes

\[
C^{-1} = \begin{pmatrix}
\frac{1}{\sigma_0^2} & 0 & \cdots & 0 \\
0 & \frac{1}{\sigma_1^2} & \cdots & 0 \\
0 & 0 & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{\sigma_{N-1}^2}
\end{pmatrix}
\]

- Again \( s = 1 \), so the BLUE is given by

\[
\hat{A} = \frac{s^T C^{-1} x}{s^T C^{-1} s} = \frac{1^T C^{-1} x}{1^T C^{-1} 1}
\]
Example: DC Level in Uncorrelated Noise

- The numerator is now

\[ 1^T C^{-1} x = \left( \frac{1}{\sigma_0^2}, \frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_{N-1}^2} \right) \begin{pmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{pmatrix} = \sum_{n=0}^{N-1} \frac{x[n]}{\sigma_n^2} \]

- The denominator acts simply as a scaling factor to normalize the gain:

\[ 1^T C^{-1} 1 = \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}. \]
Example: DC Level in Uncorrelated Noise

- Thus, the BLUE for uncorrelated noise is

$$\hat{A} = \frac{\sum_{n=0}^{N-1} \frac{x[n]}{\sigma_n^2}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}$$

- In other words, more reliable samples with smaller variance are given more weight.
- Finally, the estimator variance is

$$\text{var}(\hat{A}) = \frac{1}{1^T C^{-1} 1} = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}$$
BLUE for Vector Parameter Case

- The BLUE can be extended also to the vector parameter case. Let us first extend the definitions of linearity and unbiasedness there.

- If the parameter to be estimated is a $p \times 1$ vector $\theta = [\theta_0, \theta_1, \ldots, \theta_{N-1}]$, then the linear estimator has the form

\[
\theta_i = \sum_{n=0}^{N-1} a_{in} x[n], \quad i = 1, 2, \ldots, p.
\]

- In matrix form this becomes

\[
\hat{\theta} = Ax
\]
BLUE for Vector Parameter Case

- The unbiasedness requirement is now:

\[ E(\hat{\theta}_i) = \sum_{n=0}^{N-1} a_{in} E(x[n]) = \theta_i, \quad i = 1, 2, \ldots, p \]

or in matrix form

\[ E(\hat{\theta}) = AE(x) = \theta \]

- Just like the scalar case, also here this must imply that the model is linear in data:

\[ E(x) = H\theta \]
BLUE for Vector Parameter Case

- Substituting this form into the unbiasedness constraint above gives

\[ AE(x) = \theta \]
\[ \Leftrightarrow \quad AH\theta = \theta \]
\[ \Leftrightarrow \quad AH = I \]

- Thus, a linear estimator is unbiased if its coefficient matrix \( A \) satisfies \( AH = I \).
BLUE for Vector Parameter Case

• Similar derivation as in the scalar case gives the vector BLUE (see Appendix 6B):

\[ \hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} x \]

and the covariance matrix of \( \hat{\theta} \) is \( C_{\hat{\theta}} = (H^T C^{-1} H)^{-1} \).

• Formally, this can be stated as the following theorem.

• Note, that though many details are the same as in Chapter 4 (Linear models), the difference is in that we do not assume Gaussianity (or any other distribution either).

• On the other hand, the following theorem does not always result in MVUE, but in a suboptimal solution.
Theorem

Gauss-Markov theorem: If the data are of the general linear model form \( \mathbf{x} = \mathbf{H}\theta + \mathbf{w} \) where \( \mathbf{H} \) is a known \( N \times p \) matrix, \( \theta \) is a \( p \times 1 \) vector of parameters to be estimated, and \( \mathbf{w} \) is a \( p \times 1 \) noise vector with zero mean and covariance \( \mathbf{C} \) then the BLUE of \( \theta \) is

\[
\hat{\theta} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}
\]

the variance of \( \hat{\theta}_i \) is

\[
\text{var}(\hat{\theta}_i) = [(\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}]_{ii}
\]

and the covariance matrix of \( \hat{\theta} \) is \( \mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \).
Example: Source Localization

- Consider the problem of localizing an aircraft based on a signal it is sending.
- The signal is received by $N$ antennas at known locations $(x_0, y_0), (x_1, y_1), \ldots, (x_{N-1}, y_{N-1})$. 
Example: Source Localization

Aircraft at position 
\((x_s, y_s)\)

\[\text{Distance } d_0 \]
\[\text{Distance } d_1 \]
\[\text{Distance } d_{N-1} \]

Antenna 0 at 
\((x_0, y_0)\)

Antenna 1 at 
\((x_1, y_1)\)

Antenna \(N-1\) at 
\((x_{N-1}, y_{N-1})\)
Example: Source Localization

- A signal emitted at time $T_0$ is received at antenna $k$ at time

$$t_k = T_0 + \frac{d_k}{c} + \epsilon_k, \quad k = 0, 1, 2 \ldots, N - 1,$$

where $c$ is the propagation speed.

- Due to the Pythagorean theorem, $d_k$ is related to the locations by

$$d_k = \sqrt{(x_s - x_k)^2 + (y_s - y_k)^2},$$
Example: Source Localization

resulting in the model

\[ t_0 = T_0 + \frac{\sqrt{(x_s - x_0)^2 + (y_s - y_0)^2}}{c} + \epsilon_0 \]

\[ t_1 = T_0 + \frac{\sqrt{(x_s - x_1)^2 + (y_s - y_1)^2}}{c} + \epsilon_1 \]

\[ \vdots \]

\[ t_{N-1} = T_0 + \frac{\sqrt{(x_s - x_{N-1})^2 + (y_s - y_{N-1})^2}}{c} + \epsilon_{N-1} \]

- Unfortunately the model is nonlinear, and a BLUE can not be applied.
Example: Source Localization

- Instead, the model for the location change can be linearized, and can be used for a tracking application.
- Define the location change vector as

\[ \theta = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} x_s - x_{\text{old}} \\ y_s - y_{\text{old}} \end{pmatrix}, \]

where \((x_{\text{old}}, y_{\text{old}})\) is the previous location estimate.
Example: Source Localization

Aircraft at position \((x_s, y_s)\)

Old location \((x_{\text{old}}, y_{\text{new}})\)

Antenna 0 at \((x_0, y_0)\)
Antenna 1 at \((x_1, y_1)\)
Antenna \(N-1\) at \((x_{N-1}, y_{N-1})\)
Example: Source Localization

• The change in distance can be approximated by

\[ d_k \approx d_{old} + \frac{x_{old} - x_k}{d_{old}} \Delta x + \frac{y_{old} - y_k}{d_{old}} \Delta y, \]

which is a first order Taylor expansion around \((x_{old}, y_{old})\). There’s no error if the plane is moving linearly towards or away from the antenna.
Example: Source Localization

- Substituting this into the model

\[ t_k = T_0 + \frac{d_k}{c} + \epsilon_k, \quad k = 0, 1, 2 \ldots, N - 1, \]

gives

\[ t_k \approx T_0 + \frac{d_{\text{old}}}{c} + \frac{x_{\text{old}} - x_k}{d_{\text{old}} c} \Delta x + \frac{y_{\text{old}} - y_k}{d_{\text{old}} c} \Delta y + \epsilon_k, \]

which is a linear model of the arrival times with unknown parameters \( T_0, \Delta x \) and \( \Delta y \).
The model can be further simplified using the following observations:

\[
\frac{x_{\text{old}} - x_k}{d_{\text{old}}} = \cos \alpha_k
\]

\[
\frac{y_{\text{old}} - y_k}{d_{\text{old}}} = \sin \alpha_k
\]

and by introducing a new variable

\[
\tau_k = t_k + \frac{d_{\text{old}}}{c}.
\]
Example: Source Localization

- These result in the model

\[ \tau_k \approx T_0 + \frac{\cos \alpha_k}{c} \Delta x + \frac{\sin \alpha_k}{c} \Delta y + \epsilon_k, \]

where the unknowns are \( T_0, \Delta x \) and \( \Delta y \).

- The estimation of \( T_0 \) is often impractical, as it would require synchronization between the plane and the antennas.

- Therefore, it is customary to model the time difference of arrival (TDOA), where the observable is the time difference between antennas:

\[ \xi_k = \tau_k - \tau_{k-1}, \quad \text{for } k = 1, 2, \ldots, N - 1. \]

---

Note that this model assumes that the angle \( \alpha_k \) between the antenna and the aircraft stays constant. This can also be taken into account, but let’s keep it less complicated for the moment.
Example: Source Localization

• This results in the final model

\[ \xi_k = \frac{1}{c} (\cos \alpha_k - \cos \alpha_{k-1}) \Delta x + \frac{1}{c} (\sin \alpha_k - \sin \alpha_{k-1}) \Delta y + \epsilon_k - \epsilon_{k-1}. \]

• In matrix form this becomes

\[ \xi = H\theta + w \]

or
Example: Source Localization

Now, the BLUE solution requires the knowledge of the covariance matrix. Since our simplified model assumes that $H$ and $\theta$ are constant, the covariance matrix is that of $w$. 

\[
\begin{pmatrix}
\xi_1 \\
\xi_2 \\
\vdots \\
\xi_{N-1}
\end{pmatrix}
= \frac{1}{c}
\begin{pmatrix}
\cos \alpha_1 - \cos \alpha_0 & \sin \alpha_1 - \sin \alpha_0 \\
\cos \alpha_2 - \cos \alpha_1 & \sin \alpha_2 - \sin \alpha_1 \\
\vdots & \vdots \\
\cos \alpha_{N-1} - \cos \alpha_{N-2} & \sin \alpha_{N-1} - \sin \alpha_{N-2}
\end{pmatrix}
\cdot
\begin{pmatrix}
\Delta x \\
\Delta y
\end{pmatrix}
+ \begin{pmatrix}
\epsilon_1 - \epsilon_0 \\
\epsilon_2 - \epsilon_1 \\
\vdots \\
\epsilon_{N-1} - \epsilon_{N-2}
\end{pmatrix}.
\]
Example: Source Localization

The vector $w$ can be simplified:

$$w = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_{N-1} \end{pmatrix}.$$
Example: Source Localization

- Therefore, the covariance matrix becomes

\[ C = E[\mathbf{A} \cdot \mathbf{e} \cdot \mathbf{e}^T \cdot \mathbf{A}^T] = \sigma^2 \mathbf{A} \mathbf{A}^T. \]

Using the Gauss-Markov theorem we obtain the BLUE

\[ \hat{\theta} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \xi, \]

\[ = (\mathbf{H}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{A} \mathbf{A}^T)^{-1} \xi. \]