E6: Best Linear Unbiased Estimators (BLUE)

In practice finding the MVU estimator, even if it exists, may be hard or not possible. In that case, we may resort to a suboptimal estimator by restricting the search. In many cases, the suboptimal solution may exhibit a variance that is acceptable (i.e. meets the specs).

We will now consider the class of linear estimators that are unbiased and has minimum variance.

Defn: BLUE are linear functions of data that are unbiased estimators of the parameters with minimum variance.

Given $\{x[0], x[1], \ldots, x[N-1]\}$ with pdf $p(\mathbf{x};\theta)$, consider a linear estimator of the form:

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

DC Level in WGN: $\hat{\theta}_{MVU} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ (here $a_n = \frac{1}{N}\ \forall n$)

The MVU estimator was linear in data in this case. So here $\hat{\theta}_{BLUE} = \hat{\theta}_{MVU}$.

Mean of $\mathcal{U}[0,\theta]$: $\hat{\theta}_{MVU} = \frac{N+1}{2N} \max_n x[n]$.

In this case the MVU estimator is nonlinearly dependent on the data. $\hat{\theta}_{BLUE}$ for this problem will be suboptimal.

Power of WGN: $\widehat{\sigma^2}_{MVU} = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n]$

In this case an estimate of the form
$$\widehat{\sigma^2} = \sum_{n=0}^{N-1} a_n x[n]$$
would be completely inappropriate and a very poor suboptimal solution. Note that
$$E[\widehat{\sigma^2}] = \sum_n a_n E[x[n]] = 0$$
Here, we cannot even make the linear estimator unbiased, let alone minimum variance.

If we define $y[n] = x^2[n]$, and use this transformed data: $\widehat{\sigma^2} = \sum_n a_n y[n]$, then
$E[\widehat{\sigma^2}] = \sum_n a_n \sigma^2 = \sigma^2$ is possible with $\sum_n a_n = 1$.

Then within this constraint, we can seek the a vector that minimizes variance.

Finding the BLUE

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] \Rightarrow E[\hat{\theta}] = \sum_n a_n E[x[n]] = \theta$$

is required $\forall\theta$. [redacted]

$$Var(\hat{\theta}) = E\left[\left(a^T x - a^T E[x]\right)^2\right] = E\left[\left(a^T (x - E[x])\right)^2\right]$$

$$= E\left\{ a^T (x - E[x])(x - E[x])^T a \right\} = a^T C a$$

where $C = E\left\{ (x - E[x])(x - E[x])^T \right\} = Cov(x)$.

For $\hat{\theta}$ to be unbiased we must have $E[x[n]] = s[n]\theta$

so that $\left(\sum_n a_n s[n]\right)\theta = \theta \quad \forall\theta \Leftrightarrow \sum_n a_n s[n] = 1$

becomes the unbiased-ness constraint.

For a general ~~error~~ $x[n]$, let $\mu_x[n] = E[x[n]]$

and write $x[n] = \mu_x[n] + (x[n] - E[x[n]])$

$$= \mu_x[n] + w[n]$$

where $E[w[n]] = 0$ and $var(w[n]) = var(x[n])$ etc.

(i.e. $w[n]$ has the same pdf as $x[n]$ except it is 0-mean).

Let $\mu_x[n] = \theta s[n]$: $x[n] = \theta s[n] + w[n]$. We see that BLUE could be applicable to the estimation of θ, the amplitude of a signal in additive noise.

The constrained optimization problem to find the BLUE.

$$\min_{a} a^T C a \quad \text{s.t.} \quad \sum_{n=0}^{N-1} a_n E[x[n]] = \theta$$

$$\left(\text{or} \sum_{n=0}^{N-1} a_n s[n] = 1 \iff a^T s = 1\right)$$

where $s = [s[0], s[1], --, s[N-1]]^T$ is known.

The Lagrangian function becomes $\mathcal{L} = a^T C a + \lambda (a^T s - 1)$

and $\frac{\partial \mathcal{L}}{\partial a^T} = 2Ca + \lambda s = 0$ yields $a_{opt} = -\frac{\lambda_{opt}}{2} C^{-1} s$.

$\frac{\partial \mathcal{L}}{\partial \lambda} = a_{opt}^T s - 1 = 0$ yields $-\frac{\lambda}{2} s^T C^{-1} s = 1 \Rightarrow -\frac{\lambda_{opt}}{2} = \frac{1}{s^T C^{-1} s}$.

Then $a_{opt} = \frac{C^{-1} s}{s^T C^{-1} s}$ is the only stationary point of the Lagrangian (with the λ_{opt} specified) and since $\mathcal{L}$ has a positive definite Hessian, this stationary point must be a minimizer (in fact the global minimum).

$$\mathrm{var}(\hat{\theta}) = a_{opt}^T C a_{opt} = \frac{s^T C^{-1} C C^{-1} s}{(s^T C^{-1} s)^2} = \frac{1}{s^T C^{-1} s}$$

The BLUE is $\hat{\theta} = a_{opt}^T x = \frac{s^T C^{-1} x}{s^T C^{-1} s}$.

Clearly, $E[\hat{\theta}] = \frac{s^T C^{-1} E[x]}{s^T C^{-1} s} = \frac{s^T C^{-1} \theta s}{s^T C^{-1} s} = \theta$ (unbiased).

☆ To find (the) BLUE, we need the knowledge of
1) s, the scaled mean of $p(x;\theta)$,
2) C, the covariance of $p(x;\theta)$.

Ex] DC Level in White Noise

$x[n] = A + w[n] \quad n = 0, 1, \ldots, N-1$, $w[n] \sim$ white noise with var σ^2.

Here $w[n]$ need not be Gaussian, just white with any finite-variance distribution. (e.g. Cauchy will not work).

Since $w[n]$ is not necessarily Gaussian, "white" does not imply iid; the covariance is $C = \sigma^2 I$. The vector measurement model is

$$x = A\mathbf{1} + w$$

so $s = \mathbf{1}$ and the BLUE is $\hat{A}_{BLUE} = \dfrac{\mathbf{1}^T(\frac{1}{\sigma^2}I)x}{\mathbf{1}^T(\frac{1}{\sigma^2}I)\mathbf{1}}$

$$\Rightarrow \hat{A}_{BLUE} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] = \bar{x}.$$

$$\text{Var}(\hat{A}_{BLUE}) = \frac{1}{\mathbf{1}^T(\frac{1}{\sigma^2}I)\mathbf{1}} = \frac{\sigma^2}{N}.$$

The BLUE estimator is independent of the pdf of the data and is the MVU estimator when the pdf is Gaussian.

Ex] DC Level in Uncorrelated Noise

Let $w[n]$ be 0-mean uncorrelated noise with $\text{var}(w[n]) = \sigma_n^2$ (so $C = \text{diag}(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{N-1}^2)$).

$$\hat{A}_{BLUE} = \frac{1^T C^{-1} x}{1^T C^{-1} 1} = \sum_{n=0}^{N-1} \gamma_n \, x[n] \quad \text{where } \gamma_n = \frac{\sigma_n^{-2}}{\sum_{k=0}^{N-1} \sigma_k^{-2}}.$$

$$Var(\hat{A}_{BLUE}) = \left(\sum_{n=0}^{N-1} \sigma_n^{-2} \right)^{-1} \quad \left(\text{from } \frac{1}{1^T C^{-1} 1} \right)$$

Note that γ_n is larger when σ_n^2 is smaller. The BLUE attempts to weight samples inversely proportional to their variances. The denominator in the expression for γ_n is the normalization needed to make the estimator unbiased.

In general, the $C^{-1}x$ term in the numerator prewhitens the data prior to averaging. The denominator is normalization.

Extension to a Vector Parameter

If $\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}$, then $\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} \, x[n]$ is the elementwise linear estimator. In vector form:

$$\hat{\theta} = A x$$

where A is $p \times N$. $\hat{\theta}$ needs to be unbiased:

$$E[\hat{\theta}_i] = \sum_{n=0}^{N-1} a_{in} \, E\big[x[n]\big] = \theta_i \quad \forall i \in \{1, \dots, p\}.$$

In vector form $E[\hat{\theta}] = A\, E[x] = \theta$ is required $\forall \theta$.
If in a given problem $E[x] = H\theta$ for a known $N \times p$ matrix H, then the unbiased-ness constraint reduces to $AH = I$.

Let $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_p^T \end{bmatrix}$ and $H = [h_1 \cdots h_p]$,

where $a_i = [a_{i0} \cdots a_{i(N-1)}]^T$, so that $\hat{\theta}_i = a_i^T x$.
To have an unbiased linear estimator we need $a_i^T h_j = \delta_{ij}$ $\forall i,j \in \{1, \ldots, p\}$ (from $AH = I$).

The variances are $\mathrm{var}(\hat{\theta}_i) = a_i^T C a_i$ $\forall i$, where $C = \mathrm{Cov}(x)$. The vector valued BLUE is derived considering one a_i at a time. For $i = 1, \ldots, p$:

$$\min_{a_i} a_i^T C a_i \quad \text{s.t.} \quad a_i^T h_j = \delta_{ij} \quad j = 1, \ldots, p.$$

The Lagrangian is $\mathcal{L}_i = a_i^T C a_i + \sum_{j=1}^{p} \lambda_j^{(i)} (a_i^T h_j - \delta_{ij})$.

$$\frac{\partial \mathcal{L}_i}{\partial a_i^T} = 2Ca_i + \sum_{j=1}^{p} \lambda_j^{(i)} h_j .$$ Let $\lambda_i = [\lambda_1^{(i)} \cdots \lambda_p^{(i)}]^T$.

Then $$\frac{\partial \mathcal{L}_i}{\partial a_i^T} = 2Ca_i + H\lambda_i = 0 \Rightarrow a_i = -\frac{1}{2} C^{-1} H \lambda_i .$$

Let $e_i = [0 \cdots 0 \overset{i^{th}\text{ entry}}{1} 0 \cdots 0]^T$ be $p \times 1$. Then $H^T a_i = e_i$ is the set of constraints. From the a_i expression above we get

$$H^T a_i = -\frac{1}{2} H^T C^{-1} H \lambda_i = e_i \xrightarrow[\text{exists}]{\text{Assume } (H^T C^{-1} H)^{-1}} -\frac{1}{2}\lambda_i = (H^T C^{-1} H)^{-1} e_i$$

$$\therefore\ a_{i,opt} = C^{-1} H (H^T C^{-1} H)^{-1} e_i$$

with $\mathrm{var}(\hat{\theta}_i) = a_{i,opt}^T C a_{i,opt} = \cdots = e_i^T (H^T C^{-1} H)^{-1} e_i$.

Concatenating each $a_{i,opt}$ into A_{opt}, we get

$$\hat{\theta}_{BLUE} = (H^T C^{-1} H)^{-1} H^T C^{-1} x$$ (The e_i's disappear by forming an identity matrix.)

The covariance of $\hat{\theta}_{BLUE}$ is, dropping "BLUE".

$$C_{\hat{\theta}_{BLUE}} = E\left[(\hat{\theta} - E[\hat{\theta}])(\hat{\theta} - E[\hat{\theta}])^T\right]$$

where
$$\begin{aligned}\hat{\theta} - E[\hat{\theta}] &= (H^T C^{-1} H)^{-1} H^T C^{-1} (H\theta + w) - E[\hat{\theta}] \\ &= (H^T C^{-1} H)^{-1} H^T C^{-1} w\end{aligned}$$

$$\begin{aligned}\therefore\ C_{\hat{\theta}_{BLUE}} &= (H^T C^{-1} H)^{-1} H^T C^{-1} C C^{-1} H (H^T C^{-1} H)^{-1} \\ &= (H^T C^{-1} H)^{-1}\end{aligned}$$

and $\mathrm{var}(\hat{\theta}_i) = \left[(H^T C^{-1} H)^{-1}\right]_{ii}$.

If $x = H\theta + w$ is jointly Gaussian, then $\hat{\theta}_{BLUE}$ is also $\hat{\theta}_{MVU}$.

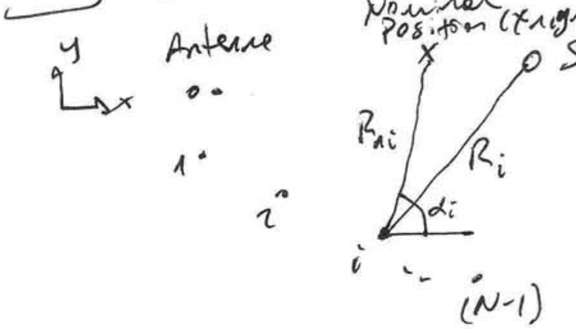
Thm 6.1 Gauss-Markov Theorem

If the data are of the general linear model form $x = H\theta + w$, where H is a known $N \times p$ matrix, θ is a $p \times 1$ vector of parameters, and w is a $N \times 1$ noise vector with zero-mean and covariance C, then the BLUE of θ is $\hat{\theta}_{BLUE} = (H^T C^{-1} H)^{-1} H^T C^{-1} x$ and the minimum variances are $var(\hat{\theta}_i) = [(H^T C^{-1} H)^{-1}]_{ii}$ $i = 1, \dots, p$.

In addition, the covariance matrix of $\hat{\theta}_{BLUE}$ is

$$C_{\hat{\theta}_{BLUE}} = (H^T C^{-1} H)^{-1}.$$

Ex Source Localization



Suppose time of arrival measurements are available for a source emission at t_0 (signal travels in space and arrives at antennas with different delays. Use this to localize the source.

For a signal emitted by the source at $t = T_0$, the time-of-arrival measurements are $t_i = T_0 + R_i/c + \epsilon_i$ $i = 0, \dots, (N-1)$.
(This assumes clocks are synchronized.)

$R_i^2 = (x_s - x_i)^2 + (y_s - y_i)^2$, ϵ_i is noise with 0 mean and variance σ^2, uncorrelated with each other.

Suppose that a nominal position (such as a previous localization result) is available. Then, assuming that the current position is close to the nominal position:

$$R_i \approx R_{ni} + \frac{x_n - x_i}{R_{ni}} \delta x_s + \frac{y_n - y_i}{R_{ni}} \delta y_s$$

is the first order Taylor series approximation where $\delta x_s = x_s - x_n$ and $\delta y_s = y_s - y_n$. Then

$$t_i \approx T_0 + \frac{R_{ni}}{c} + \frac{x_n - x_i}{R_{ni} c} \delta x_s + \frac{y_n - y_i}{R_{ni} c} \delta y_s + \epsilon_i$$

where c is the speed of propagation for the signal, assuming a uniform transmission medium.

Noticing that $\cos\alpha_i = \frac{x_n - x_i}{R_{ni}}$ (from the geometry), and $\sin\alpha_i = \frac{y_n - y_i}{R_{ni}}$, we have

$$t_i \approx T_0 + \frac{R_{ni}}{c} + \frac{\cos\alpha_i}{c} \delta x_s + \frac{\sin\alpha_i}{c} \delta y_s + \epsilon_i$$

Here R_{ni}/c is a known constant. Let $\tau_i = t_i - \frac{R_{ni}}{c}$.

$$\therefore \tau_i = T_0 + \frac{\cos\alpha_i}{c} \delta x_s + \frac{\sin\alpha_i}{c} \delta y_s + \epsilon_i$$

where $T_0, \delta x_s, \delta y_s$ are unknown. Consider differences of arrival times between antennas:

$$\left.\begin{aligned} \xi_1 &= \tau_1 - \tau_0 \\ \xi_2 &= \tau_2 - \tau_1 \\ &\vdots \\ \xi_{N-1} &= \tau_{N-1} - \tau_{N-2} \end{aligned}\right\} \quad \begin{aligned} \xi_i = {} & \tfrac{1}{c}(\cos\alpha_i - \cos\alpha_{i-1})\,\delta x_s \\ & + \tfrac{1}{c}(\sin\alpha_i - \sin\alpha_{i-1})\,\delta y_s \\ & + (\epsilon_i - \epsilon_{i-1}), \quad i = 1, \ldots, N-1 \end{aligned}$$

In standard form, this can be written using

$$\theta = [\delta x_s, \delta y_s]^T \qquad w = \begin{bmatrix} \epsilon_1 - \epsilon_0 \\ \vdots \\ \epsilon_{N-1} - \epsilon_{N-2} \end{bmatrix}$$

$$H = \frac{1}{c}\begin{bmatrix} (\cos\alpha_1 - \cos\alpha_0) & (\sin\alpha_1 - \sin\alpha_0) \\ \vdots & \vdots \\ (\cos\alpha_{N-1} - \cos\alpha_{N-2}) & (\sin\alpha_{N-1} - \sin\alpha_{N-2}) \end{bmatrix} \qquad C = \sigma^2 A A^T$$

We notice that $w = \underbrace{\begin{bmatrix} -1 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{bmatrix}}_{A \text{ is } (N-1)\times N} \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_{N-1} \end{bmatrix}$

$Cov(\epsilon) = \sigma^2 I$

$\Rightarrow Cov(w) = \sigma^2 A A^T \triangleq C$

and $E[w] = A\,E[\epsilon] = 0$.

Suggested Problems:
1, 3, 4, 5, 8
12, 15, 16

Using the Gauss-Markov theorem:

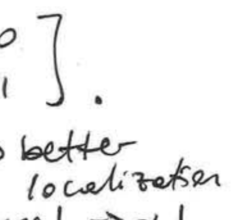
$$\hat{\theta}_{BLUE} = (H^T C^{-1} H)^{-1} H^T C^{-1} z$$

$$= \left[H^T (AA^T)^{-1} H \right]^{-1} H^T (AA^T)^{-1} z$$

and $var(\hat{\theta}_i) = \sigma^2 \left[(H^T (AA^T)^{-1} H)^{-1} \right]_{ii}$

$\hat{\theta}_{BLUE,i}$

or $C_{\hat{\theta}_{BLUE}} = \sigma^2 \left[H^T (AA^T)^{-1} H \right]^{-1}$



Special Case: 3 antennas in a linear array with distance of d between antennas.

$$\Rightarrow H = \frac{1}{c}\begin{bmatrix} -\cos\alpha & 1-\sin\alpha \\ -\cos\alpha & -(1-\sin\alpha) \end{bmatrix} \text{ and } A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

$$\Rightarrow C_{\hat{\theta}_{BLUE}} = \sigma^2 c^2 \begin{bmatrix} (2\cos^2\alpha)^{-1} & 0 \\ 0 & \frac{3}{2}(1-\sin\alpha)^{-2} \end{bmatrix}$$

$\alpha \downarrow \Rightarrow$ better localization

$d \uparrow$ or Range$\downarrow \Rightarrow \alpha\downarrow$.