Chapter 6
Best Linear Unbiased Estimate (BLUE)

Motivation for BLUE

Except for Linear Model case, the optimal MVU estimator might:

1. not even exist
2. be difficult or impossible to find

⇒ **Resort to a *sub-optimal* estimate**

BLUE is one such sub-optimal estimate

Idea for BLUE:

1. Restrict estimate to be linear in data **x**
2. Restrict estimate to be unbiased
3. Find the best one (i.e. with minimum variance)

Advantage of BLUE: Needs only $\underbrace{1^{st} \text{ and } 2^{nd} \text{ moments of PDF}}_{\textbf{Mean \& Covariance}}$

Disadvantages of BLUE:

1. Sub-optimal (in general)
2. Sometimes totally inappropriate (see bottom of p. 134)

6.3 Definition of BLUE (scalar case)

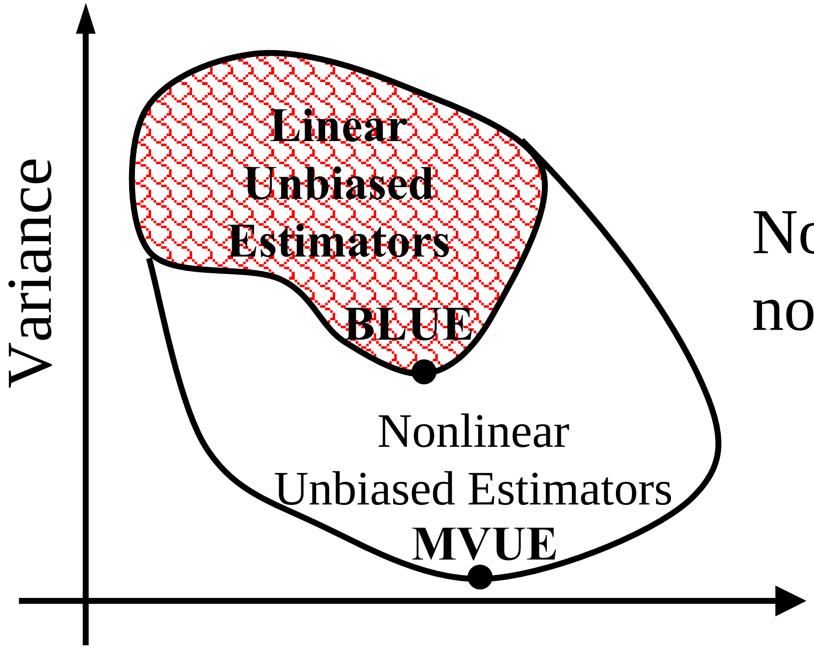
Observed Data: $\mathbf{x} = \left[x[0]\ x[1]\ .\ .\ .\ x[N-1]\ \right]^{\mathrm{T}}$

PDF: $p(\mathbf{x};\theta)$ depends on unknown θ

BLUE constrained to be linear in data: $$\hat{\theta}_{BLU} = \sum_{n-0}^{N-1} a_n x[n] = \mathbf{a}^T \mathbf{x}$$

Choose a's to give:

1. unbiased estimator
2. then minimize variance



Note: This is not Fig. 6.1

6.4 Finding The BLUE (Scalar Case)

1. Constrain to be Linear: $\hat{\theta} = \sum_{n-0}^{N-1} a_n x[n]$

2. Constrain to be Unbiased: $E\{\hat{\theta}\} = \theta$

$$\Downarrow$$

Using linear constraint

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

Q: When can we meet both of these constraints?

A: Only for certain observation models (e.g., linear observations)

Finding BLUE for Scalar Linear Observations

Consider scalar-parameter linear observation:

$$x[n] = \theta s[n] + w[n] \quad \Rightarrow \quad E\{x[n]\} = \theta s[n]$$

Then for the unbiased condition we need: $E\{\hat{\theta}\} = \theta \sum_{n-0}^{N-1} \underbrace{a_n s[n]}_{\Downarrow} = \theta$

Tells how to choose weights to use in the BLUE estimator form

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

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

Now… given that these constraints are met…

We need to minimize the variance!!

Given that **C** is the covariance matrix of **x** we have:

$$\operatorname{var}\left\{\hat{\theta}_{BLU}\right\} = \operatorname{var}\left\{\mathbf{a}^T \mathbf{x}\right\} = \mathbf{a}^T \mathbf{C} \mathbf{a}$$

Like var{aX} =a^2 var{X}

Goal: minimize $\mathbf{a}^T\mathbf{C}\mathbf{a}$ subject to $\mathbf{a}^T\mathbf{s} = 1$

$\Rightarrow$ *Constrained optimization*

Appendix 6A: Use Lagrangian Multipliers:

Minimize $J = \mathbf{a}^T\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{s} - 1)$

$$\text{Set}: \quad \frac{\partial J}{\partial a} = 0 \quad \Rightarrow \quad \underbrace{\mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}}_{\mathbf{a}^T\mathbf{s}=1}$$

$$\Rightarrow \quad \mathbf{a}^T\mathbf{s} = -\frac{\lambda}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = 1 \quad \Rightarrow \quad -\frac{\lambda}{2} = \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}$$

$$\boxed{\mathbf{a} = \frac{\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}$$

$$\boxed{\hat{\theta}_{BLUE} = \mathbf{a}^T\mathbf{x} = \frac{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}} \qquad \boxed{\operatorname{var}(\hat{\theta}) = \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}$$

Appendix 6A shows that this achieves a global minimum

Applicability of BLUE

We just derived the BLUE under the following:

1. Linear observations but with no constraint on the noise PDF
2. No knowledge of the noise PDF other than its mean and cov!!

What does this tell us???
BLUE is applicable to linear observations
But… noise need not be Gaussian!!!
(as was assumed in Ch. 4 Linear Model)
And all we need are the 1st and 2nd moments of the PDF!!!

But… we'll see in the Example that we
can often linearize a nonlinear model!!!

6.5 Vector Parameter Case: Gauss-Markov Thm

Gauss-Markov Theorem:

If data can be modeled as having linear observations in noise:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

Known Matrix

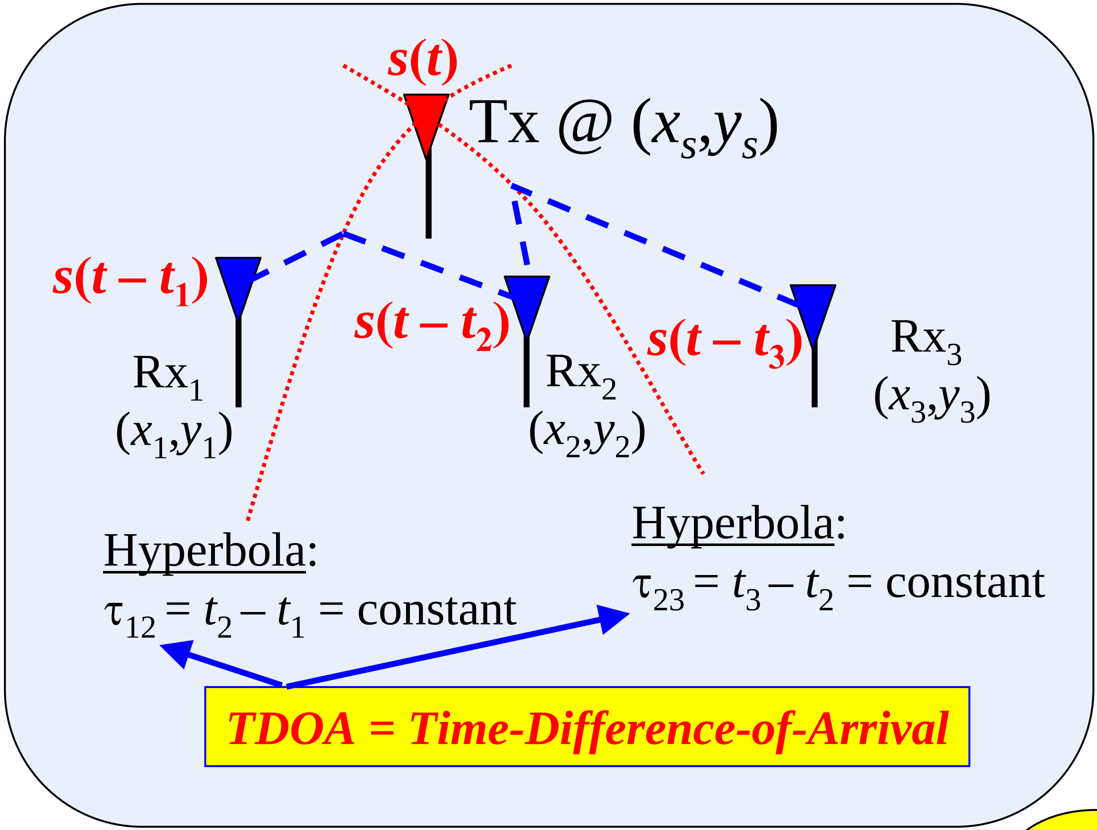
Known Mean & Cov (PDF is otherwise arbitrary & unknown)

Then the BLUE is: $\boxed{\hat{\boldsymbol{\theta}}_{BLUE} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}}$

and its covariance is: $\boxed{\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}}$

Note: If noise *is* Gaussian then BLUE *is* MVUE

Ex. 4.3: TDOA-Based Emitter Location



Hyperbola:
$\tau_{12} = t_2 - t_1$ = constant

Hyperbola:
$\tau_{23} = t_3 - t_2$ = constant

TDOA = Time-Difference-of-Arrival

Assume that the i^{th} Rx can measure its TOA: t_i

We won't worry about "how" they do that. Also… there are TDOA systems that never actually estimate TOAs!

Then… from the set of TOAs… compute TDOAs

Then… from the set of TDOAs… estimate location (x_s,y_s)

TOA Measurement Model

Assume measurements of TOAs at N receivers (only 3 shown above):

There are measurement errors

$$t_0, t_1, \dots, t_{N-1}$$

TOA measurement model:

T_o = Time the signal emitted

R_i = Range from Tx to Rx$_i$

c = Speed of Propagation (for EM: $c = 3\text{x}10^8$ m/s)

$$\boxed{t_i = T_o + R_i/c + \varepsilon_i \qquad i = 0, 1, \dots, N\text{-}1}$$

Measurement Noise ⇒ zero-mean, variance σ^2, independent (but **PDF unknown**)
(variance determined from estimator used to estimate t_i's)

Now use: $R_i = [\,(x_s - x_i)^2 + (y_s - y_i)^2\,]^{1/2}$

$$\boxed{t_i = f(x_s, y_s) = T_o + \frac{1}{c}\sqrt{(x_s - x_i)^2 + (y_s - y_i)^2} + \varepsilon_i}$$

Nonlinear Model

Linearization of TOA Model

So… we linearize the model so we can apply BLUE:

Assume some rough estimate is available (x_n, y_n)

$$x_s = x_n + \delta x_s \qquad y_s = y_n + \delta y_s$$

know estimate know estimate

$$\Rightarrow \quad \theta = [\delta x \ \ \delta y]^T$$

Now use truncated Taylor series to linearize $R_i\,(x_n, y_n)$:

$$R_i \approx R_{n_i} + \underbrace{\frac{x_n - x_i}{R_{n_i}}}_{\triangleq A_i}\delta x_s + \underbrace{\frac{y_n - y_i}{R_{n_i}}}_{\triangleq B_i}\delta y_s$$

Known

$$\text{Apply to TOA:}\quad \tilde{t}_i = t_i - \frac{R_{n_i}}{c} = \boxed{T_o} + \frac{A_i}{c}\boxed{\delta x_s} + \frac{B_i}{c}\boxed{\delta y_s} + \varepsilon_i$$

known known known

Three unknown parameters to estimate: T_o, δy_s, δy_s

TOA Model vs. TDOA Model

Two options now:

1. Use TOA to estimate 3 parameters: T_o, δy_s, δy_s
2. Use TDOA to estimate 2 parameters: δy_s, δy_s

Generally the fewer parameters the better…

Everything else being the same.

But… here "everything else" is not the same:

Options 1 & 2 have different noise models

(Option 1 has independent noise)

(Option 2 has correlated noise)

In practice… we'd explore both options and see which is best.

Conversion to TDOA Model

N–1 TDOAs rather than *N* TOAs

TDOAs: $\tau_i = \tilde{t}_i - \tilde{t}_{i-1}, \quad i = 1, 2, \ldots, N-1$

$$= \underbrace{\frac{A_i - A_{i-1}}{c}}_{\text{known}} \delta x_s + \underbrace{\frac{B_i - B_{i-1}}{c}}_{\text{known}} \delta y_s + \underbrace{\varepsilon_i - \varepsilon_{i-1}}_{\text{correlated noise}}$$

In matrix form: **x = Hθ + w**

$$\mathbf{x} = \begin{bmatrix} \tau_1 & \tau_2 & \cdots & \tau_{N-1} \end{bmatrix}^T \qquad \boldsymbol{\theta} = \begin{bmatrix} \delta x_s & \delta y_s \end{bmatrix}^T$$

$$\mathbf{H} = \frac{1}{c} \begin{bmatrix} (A_1 - A_0) & \vdots & (B_1 - B_0) \\ (A_2 - A_1) & \vdots & (B_2 - B_1) \\ \vdots & \vdots & \vdots \\ (A_{N-1} - A_{N-2}) & \vdots & (B_{N-1} - B_{N-2}) \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} \varepsilon_1 - \varepsilon_0 \\ \varepsilon_2 - \varepsilon_1 \\ \vdots \\ \varepsilon_{N-1} - \varepsilon_{N-2} \end{bmatrix} = \mathbf{A}\boldsymbol{\varepsilon}$$

See book for structure of matrix **A**

$$\mathbf{C}_\mathbf{w} = \mathrm{cov}\{\mathbf{w}\} = \sigma^2 \mathbf{A}\mathbf{A}^T$$

Apply BLUE to TDOA Linearized Model

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_{BLUE} &= \left(\mathbf{H}^T\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{x} \\
&= \left(\mathbf{H}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{x}
\end{aligned}
$$

Dependence on σ² cancels out!!!

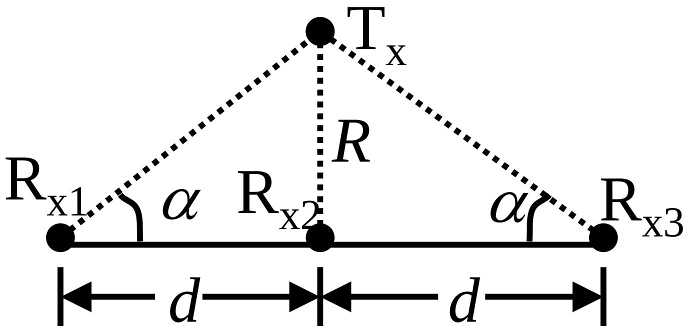
$$
\begin{aligned}
\mathbf{C}_{\hat{\boldsymbol{\theta}}} &= \left(\mathbf{H}^T\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}\right)^{-1} \\
&= \sigma^2\left(\mathbf{H}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{H}\right)^{-1}
\end{aligned}
$$

Describes how large the location error is

Things we can now do:

1. Explore estimation error cov for different Tx/Rx geometries
   - Plot error ellipses
2. Analytically explore simple geometries to find trends
   - See next chart (more details in book)

Apply TDOA Result to Simple Geometry

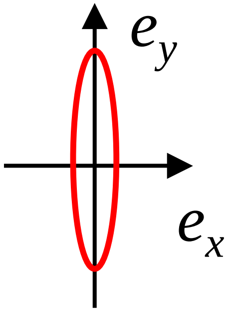


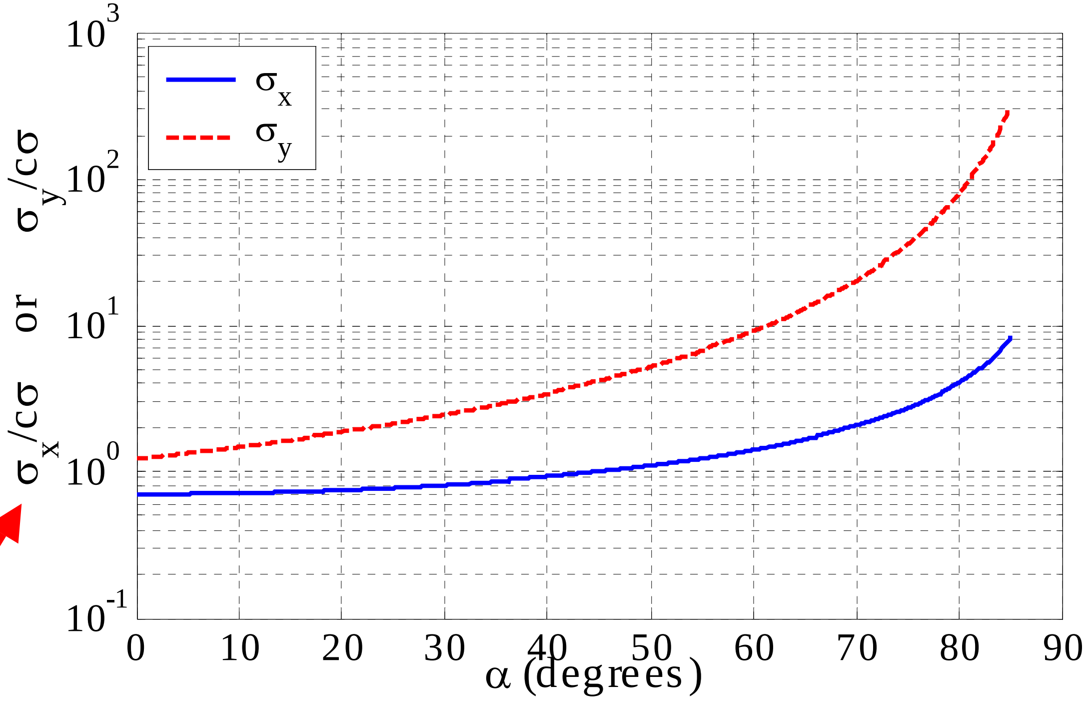
Then can show:

$$\boxed{\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2 c^2 \begin{bmatrix} \dfrac{1}{2\cos^2\alpha} & 0 \\ 0 & \dfrac{3/2}{(1-\sin\alpha)^2} \end{bmatrix}}$$

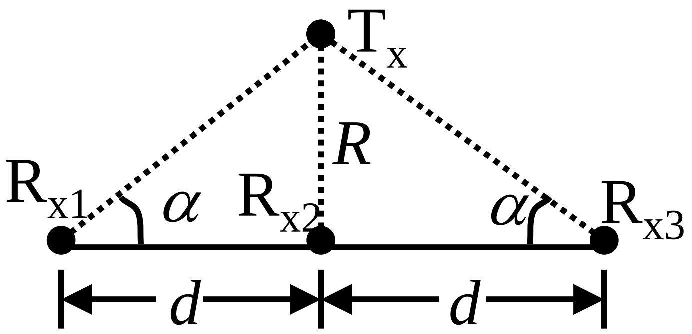
Diagonal Error Cov ⇒ Aligned Error Ellipse

And… y-error always bigger than x-error





- Used Std. Dev. to show units of X & Y
- Normalized by cσ… get actual values by multiplying by your specific cσ value



- **For Fixed Range *R*: Increasing Rx Spacing *d* Improves Accuracy**
- **For Fixed Spacing *d*: Decreasing Range *R* Improves Accuracy**