Estimation: chapter 6

Best Linear Unbiased Estimators

Lifang Feng
lffeng@ustb.edu.cn

北京科技大学
University of Science and Technology Beijing

Spring 2022

Finding estimators so far

1. CRLB - may give you the MVUE
2. Linear models - MVUE and its statistics explicitly!
3. Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem - may give you the MVUE if you can find sufficient and complete statistics

MVUE still may be tough to find

All assume we know p(x;θ)

Finding estimators so far

1. CRLB - may give you the MVUE
2. Linear models - MVUE and its statistics explicitly!
3. Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem - may give you the MVUE if you can find sufficient and complete statistics

MVUE still may be tough to find

All assume we know p(x;θ)

- only need first and second order moments of $p(\mathbf{x};\theta)$ = fairly practical!

Best Linear Unbiased Estimator

- simplify finding an estimator by constraining the class of estimators under consideration to the class of **linear estimators**, i.e.

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

- The vector **a** is a vector of constants, whose values we will design to meet certain criteria.
- Note that there is no reason to believe that a linear estimator will produce either an efficient estimator (meeting the CRLB), an MVUE, or an estimator that is optimal in any sense.

We are trading optimality for practicality!!

BLUE vs. MVUE

- when does the BLUE become the MVUE?

  - $x[n] = A + w[n]$, $w[n] \sim \mathcal{N}(0, \sigma^2)$
  - $x[n] = w[n]$, $w[n] \sim \mathcal{U}(0, \beta)$
  - $x[n] = w[n]$, $w[n] \sim \mathcal{N}(0, \sigma^2)$

BLUE assumptions

We need only assume the following:

- $x[n] = s[n]\theta + w[n], \quad n = 0, 1, \cdots N-1$ for $s[n]$ known. This ensures we can find an unbiased estimator! $\mathbf{x} = \mathbf{s}\theta + \mathbf{w}$
- $\mathbf{C} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T]$ is known

Thus, we need only the first and second order moments of $\mathbf{x}$ and not the whole pdf! Now, we wish to find the "best" linear unbiased estimator, where "best" means *minimum variance*. Taking into account the above, our linear model assumes:

- $\hat{\theta} = \mathbf{a}^T\mathbf{x}$
- $E[\hat{\theta}] = \theta \Rightarrow \mathbf{a}^T\mathbf{s} = 1$
- $var(\hat{\theta}) = \mathbf{a}^T\mathbf{C}\mathbf{a}$

LINEAR MODEL without Gaussian noise!

How to pick **a**?

BLUE - scalar

$$\hat{\theta} = \mathbf{a}^T \mathbf{x}$$

$$\mathbf{a}_{\mathrm{OPT}} = \frac{\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}$$

$$var(\hat{\theta}) = \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}$$

Examples

- $x[n] = A + w[n]$, $w[n]$ not Gaussian, but independent, identically distributed of zero mean and variance σ^2
- $x[n] = A + w[n]$, $w[n]$ not Gaussian, but independent, zero mean and variance σ_n^2

BLUE - vector

- Gauss-Markov theorem for BLUEs:

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, θ is a $p \times 1$ vector of parameters to be estimated and $\mathbf{w}$ is a $N \times 1$ noise vector with zero mean and covariance $\mathbf{C}$ (the PDF of $\mathbf{w}$ is otherwise arbitrary), then the BLUE of θ is

$$\hat{\theta} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

and the minimum variance of θ_i is

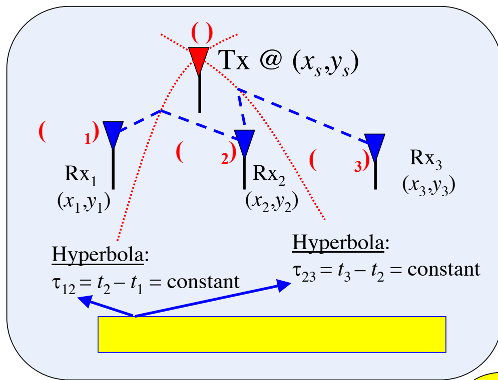
$$var(\theta_i) = \left[(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\right]_{ii}.$$

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

$$\mathbf{C}_{\hat{\theta}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}.$$

Example 6.3: source localization

Ex. 4.3: TDOA-Based Emitter Location



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)

Taken from http://www.ws.binghamton.edu/fowler/fowler%20personal%20page/EE522.htm

http://en.wikipedia.org/wiki/Multilateration

TOA Measurement Model

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

There are measurement errors

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

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)

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

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

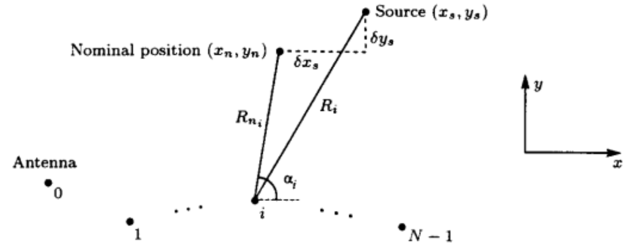


Figure 6.3 Source localization geometry

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

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

known known known

Three unknown parameters to estimate: T_o, δy_s, δy_s

http://en.wikipedia.org/wiki/Multilateration

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

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

$\boxed{\text{In matrix form: } \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}}$

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

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

$$\mathbf{w} = \underbrace{\begin{bmatrix} -1 & 1 & 0 & 0 & \ldots & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & -1 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_{N-1} \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

$$\mathbf{C} = E[\mathbf{A}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\mathbf{A}^T] = \sigma^2\mathbf{A}\mathbf{A}^T$$

Apply BLUE to TDOA Linearized Model

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

Dependence on σ^2 cancels out!!!

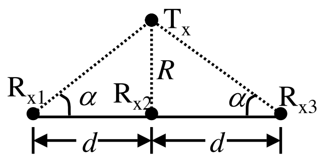
$$\boxed{\begin{aligned}\mathbf{C}_{\hat{\boldsymbol{\theta}}} &= \left(\mathbf{H}^T\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}\right)^{-1} \\ &= \sigma^2\left(\mathbf{H}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{H}\right)^{-1}\end{aligned}}$$

Describes how large the location error is

<u>**Things we can now do**</u>:

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



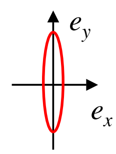
$$\mathbf{H} = \frac{1}{c}\begin{bmatrix} -\cos\alpha & 1-\sin\alpha \\ -\cos\alpha & -(1-\sin\alpha) \end{bmatrix}$$

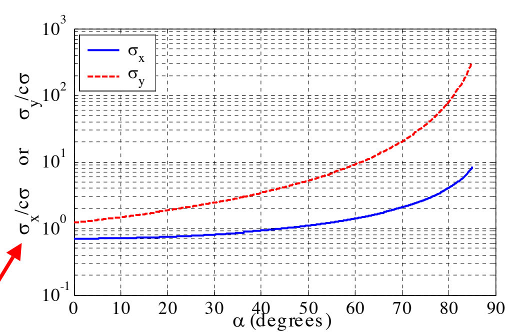
$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

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

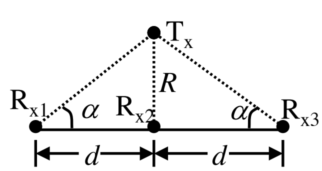
Diagonal Error Cov ⇒ Aligned Error Ellipse

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**