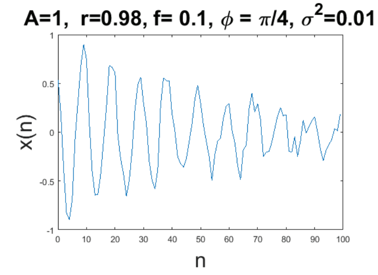
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Estimation theory – Lecture 2

- Ch. 4 Linear models
- Ch. 5 General MVU estimation

Magnus Jansson

Recall the example



A damped sinusoid signal $s(n)$ is observed in noise:

$$x(n) = s(n) + w(n)$$

for $n = 0, 1, \ldots, N-1$. Here

$$s(n) = Ar^n \cos(2\pi f n + \phi)$$

and $w(n)$ is zero mean white Gaussian noise (WGN) with variance σ^2.

Problem: Estimate unknown parameters $\boldsymbol{\theta} = [A \; r \; f \; \phi \; \sigma^2]$, or subsets thereof, from $\mathbf{x} = [x(0) \; x(1) \; \ldots x(N-1)]$.

Example: Amplitude and phase estimation

Assume now that we know $f = f_0$ and that $r = 1$. Hence, we observe

$$x(n) = A\cos(2\pi f_0 n + \phi) + w(n)$$

for $n = 0, 1, \ldots, N-1$, and want to estimate A and ϕ.

- The signal model is linear in A but nonlinear in ϕ.

Example cont'd

A possible approach: Rewrite model into linear form

$$x(n) = a\cos(2\pi f_0 n) + b\sin(2\pi f_0 n) + w(n)$$

where $A = \sqrt{a^2 + b^2}$, $a = A\cos(\phi)$ and $b = -A\sin(\phi)$.

- The signal model is now linear in a and b.
- Simple estimation problem!?

The linear Gaussian model

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}; \quad \mathbf{w} \in \mathcal{N}(0, \sigma^2 \mathbf{I})$$

In our example, we have

$$\mathbf{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} \cos(2\pi f_0 0) & \sin(2\pi f_0 0) \\ \cos(2\pi f_0 1) & \sin(2\pi f_0 1) \\ \vdots & \vdots \\ \cos(2\pi f_0(N-1)) & \sin(2\pi f_0(N-1)) \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(N-1) \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} a \\ b \end{bmatrix}$$

Linear Gaussian model cont'd

Probability density function

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right\}$$

Score function

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \ln\ p(\mathbf{x}; \boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \left[-\ln(2\pi\sigma^2)^{N/2} - \frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right] \\ &= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\theta}} \left[\mathbf{x}^T \mathbf{x} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{x} + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} \right] \\ &= -\frac{1}{2\sigma^2} \left[-2\mathbf{H}^T \mathbf{x} + 2\mathbf{H}^T \mathbf{H} \boldsymbol{\theta} \right] \\ &= \frac{\mathbf{H}^T \mathbf{H}}{\sigma^2} \left[(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} - \boldsymbol{\theta} \right] \end{aligned}$$

Linear Gaussian model cont'd

Cf. the CRB identity condition:

$$\boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln\ p(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta}) (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}).$$

Hence, the estimator

$$\hat{\boldsymbol{\theta}} = g(\mathbf{x}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$$

is unbiased and its error covariance matrix is

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1}$$

which is equal to the CRLB and hence $\hat{\boldsymbol{\theta}}$ is MVU.

Linear Gaussian model cont'd

Note also that

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} \\ &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) \\ &= \boldsymbol{\theta} + (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \end{aligned}$$

Hence, $\hat{\boldsymbol{\theta}} = \mathcal{N}(\boldsymbol{\theta}, \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1})$.

Example cont'd

Assuming the frequency f_0 is a Fourier frequency ($f_0 = k/N$ for some $k = 0, 1, \ldots, N-1$) we have

$$\mathbf{H}^T\mathbf{H} = \frac{N}{2}\mathbf{I}$$

Then

$$\hat{\boldsymbol{\theta}} = \frac{2}{N}\mathbf{H}^T\mathbf{x}$$
$$\mathbf{C}_{\hat{\theta}} = \frac{2\sigma^2}{N}\mathbf{I}$$

Example cont'd

Relatively easy to estimate a and b. What about A and ϕ? Cf. Ch. 3 (transformation of variables in CRB and of estimates) and for this example, see the paper:

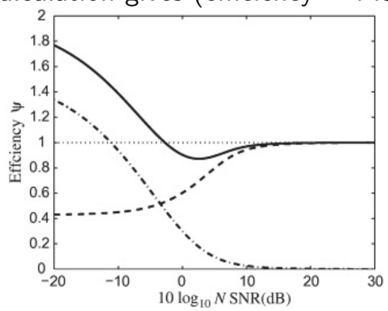
Peter Händel, Amplitude estimation using ieee-std-1057 three-parameter sine wave fit: Statistical distribution, bias and variance, Measurement, vol. 43, no. 6, pp. 766-770, 2010.

$$\mathrm{E}\{\hat{A}\} = \mathrm{E}\{\sqrt{\hat{a}^2 + \hat{b}^2}\} \approx A + \frac{\sigma^2}{AN}$$
$$\mathrm{var}\{\hat{A}\} \approx \frac{2\sigma^2}{N} - \frac{\sigma^4}{A^2N^2}$$
$$\mathrm{mse}\,\hat{A} \approx \frac{2\sigma^2}{N},$$

which is also the CRB of $\hat{A}$.

Example cont'd

A more exact calculation gives (efficiency = MSE/CRB):



(solid line: MSE/CRB, dash-dot: squared bias/CRB, dashed: var/CRB)

Estimator MSE lower than CRB at some SNRs, i.e., super efficient!

Linear Gaussian model with colored noise

What if the noise is correlated $\mathrm{E}\{w(n)w(n-k)\} \neq 0$?

New data model:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}; \quad \mathbf{w} \in \mathcal{N}(0, \mathbf{C})$$

Colored noise cont'd

Probability density function

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det\{\mathbf{C}\}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\right\}$$

Score function

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}} \ln\ p(\mathbf{x};\boldsymbol{\theta}) &= \frac{\partial}{\partial\boldsymbol{\theta}}\left[-\frac{1}{2}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\right] \\ &= -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\theta}}\left[\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} - 2\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta}\right] \\ &= -\frac{1}{2}\left[-2\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} + 2\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta}\right] \\ &= \mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\left[(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} - \boldsymbol{\theta}\right]\end{aligned}$$

Colored noise cont'd

The CRB identity condition gives that

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

is unbiased and its error covariance matrix is

$$\mathbf{C}_{\hat{\theta}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$$

which is equal to the CRLB and hence $\hat{\boldsymbol{\theta}}$ is MVU.
Note also

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{H}\boldsymbol{\theta}+\mathbf{w}) \\ &= \boldsymbol{\theta} + (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{w}\end{aligned}$$

Hence, $\hat{\boldsymbol{\theta}} = \mathcal{N}(\boldsymbol{\theta}, (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1})$.

Sufficient statistics (Ch. 5)

Probability density function for linear Gaussian model:

$$\begin{aligned}p(\mathbf{x};\boldsymbol{\theta}) &= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x}^T\mathbf{x} - 2\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{x} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta})\right\} \\ &= \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x}^T\mathbf{x})\right\}}_{=:h(\mathbf{x})}\underbrace{\exp\left\{-2\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{x} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}\right\}}_{=:g\left(T(\mathbf{x}),\theta\right)}\end{aligned}$$

The part of the PDF containing information about $\boldsymbol{\theta}$ only depends on $\mathbf{x}$ via the function $T(\mathbf{x}) = \mathbf{H}^T\mathbf{x}$.
$T(\mathbf{x})$ is a *sufficient statistic* for $\boldsymbol{\theta}$!

Neyman-Fisher factorization theorem

- General model: $p(\mathbf{x};\boldsymbol{\theta})$
- $T(\mathbf{x})$ is a *sufficient statistic* iff $p(\mathbf{x};\boldsymbol{\theta}) = g(T(\mathbf{x}),\theta)h(\mathbf{x})$

The conditional PDF of $\mathbf{x}$ given $T(\mathbf{x})$ should not depend on $\boldsymbol{\theta}$.

- $T(\mathbf{x})$ is a *minimal* sufficient statistic if it is of minimal dimension
- $T(\mathbf{x})$ is a *complete* sufficient statistic if there is exactly one unbiased function of $T(\mathbf{x})$ for all $\boldsymbol{\theta}$

Rao-Blackwell-Lehman-Sheffe (RBLS) theorem

- Unbiased estimator: $\check{\theta} = \check{g}(\mathbf{x})$
- New unbiased estimator: $\hat{\theta} = \mathrm{E}(\check{\theta}|T(\mathbf{x})) = \hat{g}(T(\mathbf{x}))$

The new estimator

1. is a valid estimator (not function of $\boldsymbol{\theta}$)
2. unbiased
3. is better, i.e., $\mathrm{var}(\hat{\theta}) \leq \mathrm{var}(\check{\theta})$
4. is the MVU estimator if $T(\mathbf{x})$ is complete.

Provides a method for obtaining the MVU estimator when the CRLB can not be computed or no efficient estimator exists

Completeness

Check that

$$\mathrm{E}\{\mathbf{v}(\mathbf{T}(\mathbf{x}))\} = \int \mathbf{v}(\mathbf{T})p(\mathbf{T};\boldsymbol{\theta})d\mathbf{T} = 0 \quad \forall\boldsymbol{\theta}$$

is only true for $\mathbf{v}(\mathbf{T}) = 0$ for all $\mathbf{T}$.

Completeness depends on PDF; true for exponential family of PDFs.

NF/RBLS : Procedure to find MVU

1. Find sufficient statistic $T(\mathbf{x})$ for θ (NF thm)
2. Check if $T(\mathbf{x})$ is complete
3. Find one arbitrary unbiased estimator $\check{\theta} = \check{g}(\mathbf{x})$ and compute: $\hat{\theta} = \mathrm{E}(\check{\theta}|T(\mathbf{x})) = \hat{g}(T(\mathbf{x}))$
4. Alternatively, try to find/guess a function $g(T(\mathbf{x}))$ that is an unbiased estimator of θ. Because of uniqueness (completeness) this is the MVU.