ECE531 Screencast 3.1: MVU Parameter Estimation for Linear Models

D. Richard Brown III

Worcester Polytechnic Institute
Introduction

A special case that covers a wide variety of common problems is when the observations and unknown parameters are related by a linear model. The most general form of this model is

\[ Y = H\theta + s + W \]

where

- \( H \in \mathbb{R}^{N \times p} \) is a known observation matrix with linearly independent columns, i.e. \( \text{rank}(H) = p \).
- \( \theta \in \mathbb{R}^{p \times 1} \) is a vector of unknown parameters
- \( s \in \mathbb{R}^{N} \) is a vector of known signal samples
- \( W \in \mathbb{R}^{N} \) is a Gaussian noise vector with distribution \( W \sim \mathcal{N}(0, C) \), where \( C = \text{E}[WW^\top] \) is known.

We are interested in finding MVU estimators of \( \theta \) in this context.
Example 1

Our standard problem of estimating a constant in zero-mean white Gaussian noise fits this model. Recall the observation model

\[ Y_k = \theta + W_k \text{ for } k = 0, \ldots, N - 1 \]

with \( W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \). We can put this into matrix/vector form as follows:

\[
\begin{bmatrix}
Y_0 \\
\vdots \\
Y_{N-1}
\end{bmatrix} =
\begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix} \theta +
\begin{bmatrix}
0 \\
\vdots \\
0
\end{bmatrix} +
\begin{bmatrix}
W_0 \\
\vdots \\
W_{N-1}
\end{bmatrix}
\]

The noise covariance \( C = \text{E}[WW^\top] \). In this case, since the noise is i.i.d., we have \( C = \sigma^2 I \).
Example 2

Many interesting problems also fit this model. For example, consider a binary communication system with an intersymbol interference channel. Suppose $M$ known symbols $x[0], \ldots, x[M-1]$ are sent through the length-$L$ unknown multipath channel $h$ and we observe

$$
Y_k = \sum_{\ell=0}^{L-1} h_\ell x_{k-\ell} + W_k
$$

for $k \in \{0, \ldots, L + M - 2\}$ and $W_k \sim \mathcal{N}(0, \sigma^2)$. We wish to estimate the channel coefficients $\theta = [h_0, \ldots, h_{L-1}]$. Letting $N = L + M - 2$, we can put this into matrix/vector form as follows:

$$
\begin{bmatrix}
Y_0 \\
Y_1 \\
\vdots \\
Y_{N-2} \\
Y_{N-1}
\end{bmatrix} =
\begin{bmatrix}
x_0 & 0 & \cdots & 0 \\
x_1 & x_0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & x_{L-1} & x_{L-2} \\
0 & \cdots & 0 & x_{L-1}
\end{bmatrix}
\begin{bmatrix}
h_0 \\
h_1 \\
\vdots \\
h_{L-2} \\
h_{L-1}
\end{bmatrix} +
\begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix} +
\begin{bmatrix}
W_0 \\
W_1 \\
\vdots \\
W_{N-2} \\
W_{N-1}
\end{bmatrix}
$$
MVU Estimator for the General Linear Model

Given observations related to parameters by the linear model

\[ Y = H\theta + s + W \]

with \( W \sim \mathcal{N}(0, C) \), the MVU estimator is

\[ \hat{\theta}(Y) = (H^\top C^{-1} H)^{-1} H^\top C^{-1} (Y - s). \]

This estimator is efficient, hence

\[ \text{cov}(\hat{\theta}(Y)) = \left( H^\top C^{-1} H \right)^{-1} \]
Intuition

The previous result is based on transforming the observation as

\[ Z = D(Y - s) \]

where \( C^{-1} = D^\top D \) and \( D \in \mathbb{R}^{N \times N} \). Since \( C \) is real, symmetric, and positive definite, so is \( C^{-1} \). This is called the Cholesky factorization (see Matlab function `chol`) and exists for all positive definite matrices.

We can write the transformed observation model as

\[
Z = D(Y - s) \\
= D(H\theta + W) \\
= DH\theta + W'
\]

where \( W' \) is a Gaussian random vector. The mean \( \mathbb{E}[W'] = DE[W] = 0 \) and the covariance

\[
\text{cov}(W') = \mathbb{E}[DWW^\top D^\top] = DCD^\top = D(D^\top D)^{-1}D^\top = DD^{-1}(D^\top)^{-1}D^\top = I.
\]

In other words, \( W'_k \sim \text{i.i.d. } \mathcal{N}(0, 1) \).
Proof Sketch

The crux of the proof is in the **attainability conditions** of the CRLB. Recall that, under the conditions of the theorem in Chapter 3, the CRLB is attainable if and only if you can write

\[ \nabla_{\theta} \ln p_Y(y; \theta) = I(\theta)(g(y) - \theta) \]

for some function \( g(y) \) that is not a function of \( \theta \). When this is true, \( g(y) \) is the MVU estimator and it achieves the CRLB (it is efficient).

You will need some multivariable calculus to compute \( \nabla_{\theta} \ln p_Y(y; \theta) \). In the case with i.i.d. white Gaussian noise \( W_k \sim \mathcal{N}(0, 1) \) and \( s = 0 \), the result is

\[ \nabla_{\theta} \ln p_Y(y; \theta) = H^\top H \underbrace{I(\theta)}_{\text{\( H^\top H \)}} \begin{bmatrix} (H^\top H)^{-1}H^\top y - \theta \\ g(y) \end{bmatrix} \]

hence the MVU estimator is \( \hat{\theta}(y) = (H^\top H)^{-1}H^\top y \) and it achieves the CRLB, i.e. \( \text{cov}(\hat{\theta}(Y)) = I^{-1}(\theta) = (H^\top H)^{-1} \).