# E4: Linear Models

If the generative model is linear, the MVU estimator will become easy to obtain.

**Thm 4.1** MVU Estimator for the Linear Model

If the data observed can be modeled as $\mathbf{x} = H\theta + w$ where $x$ is an $N \times 1$ vector of observations, $H$ is a known $N \times p$ observation matrix $(N > p)$ with rank $p$, $\theta$ is a $p \times 1$ vector of parameters to be estimated, and $w$ is an $N \times 1$ noise vector with pdf $\mathcal{N}(0, \sigma^2 I)$, then the MVU estimator is $\hat{\theta} = (H^T H)^{-1} H^T x$ and $C_{\hat{\theta}} = \sigma^2 (H^T H)^{-1}$. For the linear model, the MVU estimator is efficient.

~~Proof~~: Clearly $E[\hat{\theta}] = \theta$, so $\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2 (H^T H)^{-1})$.

Proof: $x \sim \mathcal{N}(H\theta, \sigma^2 I)$.

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

$$= \frac{-1}{2\sigma^2}\frac{\partial}{\partial \theta}\left[x^T x - 2x^T H\theta + \theta^T H^T H \theta\right]$$

$$= \frac{1}{\sigma^2}\left[H^T x - H^T H \theta\right]$$

Assuming $(H^T H)^{-1}$ exists.

$$= \frac{H^T H}{\sigma^2}\left[(H^T H)^{-1} H^T x - \theta\right] \Rightarrow I(\theta) = \frac{H^T H}{\sigma^2}$$

From CRLB theorem: $\hat{\theta} = (H^T H)^{-1} H^T x$

$C_{\hat{\theta}} = I^{-1}(\theta) = \sigma^2 (H^T H)^{-1}$

Ex] Curve fitting (polynomial models or other linear-in-parameter models)

Measurements $\{x(t_0), x(t_1), \ldots, x(t_{N-1})\}$ are taken for a signal with generative model

$$x(t) = \theta_1 + \theta_2 t + \theta_3 t^2 + w(t)$$

$x = [x(t_0), \ldots, x(t_{N-1})]^T$, $\theta = [\theta_1, \theta_2, \theta_3]^T$ and $w = [w(t_0), \ldots, w(t_{N-1})]^T$ yields $x = H\theta + w$.

Here $H = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{N-1} & t_{N-1}^2 \end{bmatrix}$. This could be easily generalized to the case of fitting an order-$(p-1)$ polynomial (with $\theta = [\theta_1 \ldots \theta_p]^T$ and $H$ with $p$ columns).

$H = \begin{bmatrix} 1 & t_0 & \cdots & t_0^{p-1} \\ \vdots & \vdots & & \vdots \\ 1 & t_{N-1} & \cdots & t_{N-1}^{p-1} \end{bmatrix}$ is a Vandermonde matrix.

$$\hat{s}(t) = \sum_{i=1}^{p} \hat{\theta}_i t^{i-1}$$

In general, a linear-in parameters type model is $x(t) = \sum_{i=1}^{p} \theta_i b_i(t) + w(t)$ where $b_i(t)$ are basis functions. Then $H = \begin{bmatrix} b_1(t_0) & \cdots & b_p(t_0) \\ \vdots & & \vdots \\ b_1(t_{N-1}) & \cdots & b_p(t_{N-1}) \end{bmatrix}$.

Ex] Fourier Analysis

Consider $x[n] = \sum_{k=1}^{M} a_k \cos\left(\frac{2\pi k n}{N}\right) + \sum_{k=1}^{M} b_k \sin\left(\frac{2\pi k n}{N}\right) + w[n]$

for $n = 0, 1, \ldots, (N-1)$. $w[n]$ is WGN. The frequencies are $f_k = k/N$.

The parameters $\{a_k, b_k\}_{k=1}^{M}$ are to be estimated.

Note that here we have $2M$ basis functions in the form of cosines and sines with harmonic frequencies.

$$\theta = [a_1 \ a_2 \ \cdots \ a_M \ b_1 \ b_2 \ \cdots \ b_M]^T$$

$$H = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ \cos\left(\frac{2\pi}{N}\right) & \cdots & \cos\left(\frac{2\pi M}{N}\right) & \sin\left(\frac{2\pi}{N}\right) & \cdots & \sin\left(\frac{2\pi M}{N}\right) \\ \vdots & & \vdots & \vdots & & \vdots \\ \cos\left(\frac{2\pi(N-1)}{N}\right) & \cdots & \cos\left(\frac{2\pi M(N-1)}{N}\right) & \sin\left(\frac{2\pi(N-1)}{N}\right) & \cdots & \sin\left(\frac{2\pi M(N-1)}{N}\right) \end{bmatrix}$$

$H \in \mathbb{R}^{N \times 2M}$ $(p = 2M)$. To have $N > p$, we need $M < N/2$.

Let $H = [h_1 h_2 \cdots h_{2M}]$ where $h_i$ is the $i^{th}$ column. It can be shown that $h_i^T h_j = 0$ for $i \neq j$. Then

$$H^T H = \begin{bmatrix} h_1^T \\ \vdots \\ h_{2M}^T \end{bmatrix} [h_1 \cdots h_{2M}] = \begin{array}{c} \scriptstyle j^{th}\text{ col} \\ \begin{bmatrix} & \vdots & \\ \cdots & h_i^T h_j & \cdots \\ & \vdots & \end{bmatrix} \end{array} i^{th}\text{ row}$$

Clearly $H^T H$ is a diagonal matrix. The orthogonality of columns seen here also arise in DFT. Specifically

$$H^T H = \frac{N}{2} I \quad \left(\text{because } h_i^T h_i = \frac{N}{2}\right).$$

Then $\hat{\theta} = (H^T H)^{-1} H^T x = \frac{2}{N} H^T x$, or explicitly

$$\left.\begin{aligned} \hat{a}_k &= \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi k n}{N}\right) \\ \hat{b}_k &= \frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi k n}{N}\right) \end{aligned}\right\} \text{DFT coefficients.}$$

Since the linear model is unbiased $E[\hat{a}_k] = a_k$ and $E[\hat{b}_k] = b_k$. $C_{\hat{\theta}} = \sigma^2 (H^T H)^{-1} = \sigma^2 \left(\frac{N}{2} I\right)^{-1} = \frac{2\sigma^2}{N} I$.

Ex | System Identification

$x[n] = h[n] * u[n] + w[n]$ where $h[n]$ is FIR.

$\mathcal{H}(z) = \sum_{k=0}^{p-1} h[k] z^{-k}$, $u[n]$ is known, $w[n]$ is WGN.

$$x[n] = \sum_{k=0}^{p-1} h[k]\, u[n-k] + w[n] \qquad n = 0, 1, \ldots, (N-1)$$

$$= [u[n]\ u[n-1]\ \cdots\ u[n-p+1]] \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[p-1] \end{bmatrix} + w[n]$$

In vector-matrix form, letting $x = [x[0], \ldots, x[N-1]]^T$, etc.

$$x = \begin{bmatrix} u[0] & 0 & \cdots & 0 \\ u[1] & u[0] & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ u[N-1] & u[N-2] & \cdots & u[N-p] \end{bmatrix} \underbrace{\begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[p-1] \end{bmatrix}}_{\theta} + \begin{bmatrix} w[0] \\ \vdots \\ w[N-1] \end{bmatrix}$$

$= H\theta + w$

The MVU estimator is $\hat{\theta} = (H^T H)^{-1} H^T x$ and $C_{\hat{\theta}} = \sigma^2 (H^T H)^{-1}$.

A key question is how to choose the probing signal $u[n]$.

Let $e_i \triangleq [0 \cdots 0\ \underset{i\text{th entry}}{1}\ 0 \cdots 0]^T$. $\operatorname{var}(\hat{\theta}_i) = e_i^T C_{\hat{\theta}} e_i$.

Let $C_{\hat{\theta}}^{-1} = D^T D$ where $D$ is an invertible $p \times p$ matrix.

Note that $(e_i^T D^T D^{-T} e_i)^2 = 1$. Let $\xi_1 = D e_i$, $\xi_2 = D^{-T} e_i$

and apply the Cauchy Schwarz inequality. Since $\xi_1^T \xi_2 = 1$,

$$1 \le (e_i^T D^T D e_i)(e_i^T D^{-1} D^{-T} e_i) = (e_i^T C_{\hat{\theta}}^{-1} e_i)(e_i^T C_{\hat{\theta}} e_i)$$

$\therefore \operatorname{var}(\hat{\theta}_i) \ge \dfrac{1}{e_i^T C_{\hat{\theta}}^{-1} e_i} = \dfrac{\sigma^2}{[H^T H]_{ii}}$. Equality holds iff

$\xi_1 = c\,\xi_2$ or $D e_i = c_i D^{-T} e_i \quad i = 1, 2, \ldots, p$ for some $\{c_i\}_{i=1}^{p}$.

$D^T D = C_{\hat{\theta}}^{-1} = \frac{H^T H}{\sigma^2}$, so $\frac{H^T H}{\sigma^2} e_i = c_i e_i$. In matrix form:

$$H^T H = \sigma^2 \begin{bmatrix} c_1 & & & 0 \\ & c_2 & & \\ & & \ddots & \\ 0 & & & c_p \end{bmatrix}$$

$\therefore$ In order to minimize the variance of the MVU estimator, $u[n]$ must be chosen to make $H^T H$ diagonal. Since $[H]_{ij} = u[i-j]$, $[H^T H]_{ij} = \sum_{n=1}^{N} u[n-i]u[n-j]$

$i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, p$. For large $N$, we get

$$[H^T H]_{ij} \approx \sum_{n=0}^{N-1-|i-j|} u[n]u[n+|i-j|]$$

which is the correlation function of a deterministic sequence. With this approximation $H^T H$ becomes a Toeplitz, symmetric, autocorrelation matrix.

$$H^T H \approx N \begin{bmatrix} r_{uu}[0] & r_{uu}[1] & \cdots & r_{uu}[p-1] \\ r_{uu}[1] & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{uu}[1] \\ r_{uu}[p-1] & \cdots & r_{uu}[1] & r_{uu}[0] \end{bmatrix}$$

where $r_{uu}[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} u[n]u[n+k]$

For $H^T H$ to be diagonal, we require $r_{uu}[k] = 0$ for $k \neq 0$ which is approximately realized if we use a pseudorandom noise (PRN) sequence. In that case, we get $H^T H \approx N r_{uu}[0] I$ and $\mathrm{var}(\hat{h}[i]) \approx \frac{1}{(N r_{uu}[0]/\sigma^2)}$ $i = 0, 1, \ldots, p-1$.

Finally, in this case $\hat{\theta} = (H^T H)^{-1} H^T x$ becomes

$$\hat{h}[i] \approx \frac{1}{N r_{uu}[0]} \sum_{n=0}^{N-1} u[n-i]\, x[n] = \frac{\frac{1}{N}\sum_{n=0}^{N-1-i} u[n]\, x[n+i]}{r_{uu}[0]}$$

The numerator is the crosscorrelation $r_{ux}[i]$, so if PRN input is used $\hat{h}[i] = \frac{r_{ux}[i]}{r_{uu}[0]}$ $i = 0, \ldots, p-1$,

where $r_{ux}[i] = \frac{1}{N}\sum_{n=0}^{N-1-i} u[n]\, x[n+i]$, $r_{uu}[0] = \frac{1}{N}\sum_{n=0}^{N-1} u^2[n]$.

## Extension to the Linear Model with Nonwhite Noise

In the general case noise is not white: $w \sim \mathcal{N}(0, C)$, where $C$ is not a scaled identity matrix. We assume $C > 0$. Then $C^{-1} = D^T D$ exists with invertible square $D$.

Since $E[(Dw)(Dw)^T] = DCD^T = D D^{-1} D^{-T} D^T = I$, $Dw$ is white. We define $x' = Dx$. Then

$$x' = Dx = DH\theta + Dw = H'\theta + w' \text{ where } w' \sim \mathcal{N}(0, I).$$

The usual linear model with white noise applies.

$$\hat{\theta} = (H'^T H')^{-1} H'^T x' = (H^T D^T D H)^{-1} H^T D^T D x$$

$$= (H^T C^{-1} H)^{-1} H^T C^{-1} x$$

$$C_{\hat{\theta}} = (H'^T H')^{-1} = (H^T C^{-1} H)^{-1}$$

Clearly $C = \sigma^2 I$ reduces this general result to the previous case we examined.

Ex) DC Level in Colored Noise

$x[n] = A + w[n]$ $n = 0, 1, \ldots, N-1$ where $w[n]$ is colored Gaussian noise such that for $w = [w[0], \ldots, w[N-1]]^T$ the covariance matrix is $C$. Here $H = \mathbf{1} = [1 \cdots 1]^T$ and the MVU estimator is

$$\hat{A} = (H^T C^{-1} H)^{-1} H^T C^{-1} x = \frac{\mathbf{1}^T C^{-1} x}{\mathbf{1}^T C^{-1} \mathbf{1}}$$

with $\operatorname{var}(\hat{A}) = (H^T C^{-1} H)^{-1} = \dfrac{1}{\mathbf{1}^T C^{-1} \mathbf{1}}$.

If $C = \sigma^2 I$ ($w[n]$ is WGN), then the MVU estimator is the sample average with variance $\sigma^2/N$. Let $C^{-1} = D^T D$:

$$\hat{A} = \frac{\mathbf{1}^T D^T D x}{\mathbf{1}^T D^T D x} = \frac{(D\mathbf{1})^T x'}{\mathbf{1}^T D^T D \mathbf{1}} = \sum_{n=0}^{N-1} d_n x'[n]$$

where $d_n = [D\mathbf{1}]_n / (\mathbf{1}^T D^T D \mathbf{1})$. So $\hat{A}$ is a weighted average of prewhitened data.

Extension to a Linear Model with Known Signal Components

Let $x = H\theta + s + w$ where $s$ is a known signal. Define $x' = x - s = H\theta + w$ brings the problem to standard form. Then $\hat{\theta} = (H^T H)^{-1} H^T (x - s)$ with $C_{\hat{\theta}} = \sigma^2 (H^T H)^{-1}$.

Ex) DC Level and Exponential Signal in White Noise

$x[n] = A + r^n + w[n]$ for $n = 0, 1, \ldots, N-1$ where $r$ is known, $w[n]$ is WGN, $A$ is to be estimated.

$x = A\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + s + w$ where $s = [1 \; r \; \cdots \; r^{N-1}]^T$.

The MVU estimator is $\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - r^n)$

with $\mathrm{var}(\hat{A}) = \sigma^2/N$.

Thm 4.2 MVU Estimator for General Linear Model.

If $x = H\theta + s + w$, $x \sim N\times 1$ vector, $H \sim N\times p$ is known $(N>p)$ with rank $p$, $\theta \sim p\times 1$, $s \sim N\times 1$ and known, $w \sim N\times 1$ with pdf $\mathcal{N}(0, C)$, then the MVU estimator is

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

and the covariance is $C_{\hat{\theta}} = (H^T C^{-1} H)^{-1}$. This MVU estimator is efficient.

Suggested Problems: 1, 3, 6, 8, 12, 13, 14