

EE6617: Tutorial 6

6.1 Show that if an efficient estimator exists, the maximum likelihood method will produce it. Assume a scalar parameter case. (Hint: use the CRLB theorem.)

6.2 Consider $\mathbf{H} \in \mathbb{R}^{N \times p}$ $\text{rank}(\mathbf{H}) = p$. Let the singular value decomposition (SVD) of $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where $\mathbf{U} \in \mathbb{R}^{N \times N}$ satisfies $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_N$, $\mathbf{V} \in \mathbb{R}^{p \times p}$ satisfies $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_p$, and $\mathbf{\Sigma} = \begin{bmatrix} \bar{\mathbf{\Sigma}} \\ \mathbf{0}_{(N-p) \times p} \end{bmatrix}$ where $\bar{\mathbf{\Sigma}} = \text{diag}(\sigma_1, \dots, \sigma_p)$. Show that $\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T = \bar{\mathbf{U}}\bar{\mathbf{U}}^T$ where $[\mathbf{U}]_{1:p} = \bar{\mathbf{U}}$. Also, check that if we set $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T = \bar{\mathbf{U}}\bar{\mathbf{U}}^T$, $(\mathbf{P})^k = \bar{\mathbf{U}}\bar{\mathbf{U}}^T = \mathbf{P}$ for $k \geq 0$. The matrix with $\mathbf{P}^k = \mathbf{P}$ property is called the idempotent matrix.

6.3 Assume the observation $\mathbf{x} = \mathbf{s} + \mathbf{w} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ where $\mathbf{H} \in \mathbb{R}^{N \times p}$. We want to design LSE of $\mathbf{s}$. When $p = N$ what is the LSE $\hat{\mathbf{s}}$?

6.4 Suppose linearly constrained LS problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmin}} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \quad \text{s.t.} \quad \mathbf{A}\boldsymbol{\theta} = \mathbf{b}$$

where $\mathbf{A} \in \mathbb{R}^{r \times p}$, $r \leq p \leq N$. We already know from the class that

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{uc} - (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T \left(\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T \right)^{-1} \left(\mathbf{A}\hat{\boldsymbol{\theta}}_{uc} - \mathbf{b} \right)$$

where $(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}\mathbf{x} = \hat{\boldsymbol{\theta}}_{uc}$. Alternatively, show this result by solving an equivalent problem

$$\hat{\tilde{\mathbf{x}}} = \underset{\tilde{\mathbf{x}}}{\text{argmin}} \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} \quad \text{s.t.} \quad \mathbf{A}\boldsymbol{\theta} = \mathbf{b}. \tag{1}$$

where $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{H}\boldsymbol{\theta}$.

Sol. 6.1 If an efficient estimator exists, the CRLB guarantees that the decomposition is true, i.e.,

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

Notice that this condition is iff condition. To find MLE, the critical point is evaluated such that

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} &= 0 \\ \Rightarrow I(\theta) (\hat{\alpha}(\mathbf{x}) - \theta) &= 0 \\ \Rightarrow \hat{\theta} &= \hat{\alpha}(\mathbf{x}) \end{aligned}$$

and $\hat{\theta} = \hat{\alpha}(\mathbf{x})$ is the efficient estimator. This is both the efficient MVU estimator and MLE.

Sol. 6.2 Denote $\bar{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_p)$.

$$\begin{aligned} \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T &= \mathbf{U} \Sigma \mathbf{V}^T (\mathbf{V} \Sigma^T \mathbf{U}^T \mathbf{U} \Sigma \mathbf{V}^T)^{-1} \mathbf{V} \Sigma^T \mathbf{U}^T \\ &= \mathbf{U} \Sigma \mathbf{V}^T \mathbf{V} (\Sigma^T \Sigma)^{-1} \mathbf{V}^T \mathbf{V} \Sigma^T \mathbf{U}^T \\ &= \mathbf{U} \Sigma (\Sigma^T \Sigma)^{-1} \Sigma^T \mathbf{U}^T \\ &= \bar{\mathbf{U}} \bar{\Sigma} (\bar{\Sigma}^T \bar{\Sigma})^{-1} \bar{\Sigma}^T \bar{\mathbf{U}}^T \\ &= \bar{\mathbf{U}} \bar{\Sigma} \bar{\Sigma}^{-1} \bar{\Sigma}^{-T} \bar{\Sigma}^T \bar{\mathbf{U}}^T \\ &= \bar{\mathbf{U}} \bar{\mathbf{U}}^T. \end{aligned}$$

It is straightforward to show that $\mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \bar{\mathbf{U}} \bar{\mathbf{U}}^T \bar{\mathbf{U}} \bar{\mathbf{U}}^T = \bar{\mathbf{U}} \bar{\mathbf{U}}^T$.

Sol. 6.3 The LSE of $\boldsymbol{\theta}$ is

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

When $N = p$, $\hat{\boldsymbol{\theta}} = \mathbf{H}^{-1} \mathbf{x}$. Thus, the LSE of $\mathbf{s}$ is $\hat{\mathbf{s}} = \mathbf{H} \mathbf{H}^{-1} \mathbf{x} = \mathbf{x}$.

Sol. 6.4 Given $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{H} \boldsymbol{\theta}$, we express the $\boldsymbol{\theta} = \mathbf{H}^\dagger (\mathbf{x} - \tilde{\mathbf{x}})$ where $\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$. Then, the constraint $\mathbf{A} \boldsymbol{\theta} = \mathbf{b}$ is

$$\begin{aligned} \mathbf{A} \mathbf{H}^\dagger (\mathbf{x} - \tilde{\mathbf{x}}) &= \mathbf{b} \\ \Rightarrow \mathbf{A} \mathbf{H}^\dagger \tilde{\mathbf{x}} &= \mathbf{A} \mathbf{H}^\dagger \mathbf{x} - \mathbf{b} \\ \Rightarrow \tilde{\mathbf{A}} \tilde{\mathbf{x}} &= \tilde{\mathbf{A}} \mathbf{x} - \mathbf{b} \\ \Rightarrow \tilde{\mathbf{x}} &= \tilde{\mathbf{A}}^\dagger (\tilde{\mathbf{A}} \mathbf{x} - \mathbf{b}) \end{aligned} \tag{2}$$

where $\tilde{\mathbf{A}}^\dagger = \mathbf{H}^{\dagger T} \mathbf{A}^T (\mathbf{A} \mathbf{H}^\dagger \mathbf{H}^{\dagger T} \mathbf{A}^T)^{-1}$. Then, by substituting $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{H} \boldsymbol{\theta}$ in (2)

$$\begin{aligned} \mathbf{x} - \mathbf{H} \boldsymbol{\theta} &= \tilde{\mathbf{A}}^\dagger (\tilde{\mathbf{A}} \mathbf{x} - \mathbf{b}) \\ \Rightarrow \mathbf{H} \boldsymbol{\theta} &= \mathbf{x} - \tilde{\mathbf{A}}^\dagger (\tilde{\mathbf{A}} \mathbf{x} - \mathbf{b}) \\ \Rightarrow \hat{\boldsymbol{\theta}} &= \mathbf{H}^\dagger \mathbf{x} - \mathbf{H}^\dagger \tilde{\mathbf{A}}^\dagger (\tilde{\mathbf{A}} \mathbf{x} - \mathbf{b}). \end{aligned}$$

Now, plugging in $\tilde{\mathbf{A}}^\dagger = \mathbf{H}^{\dagger T} \mathbf{A}^T \left(\mathbf{A} \mathbf{H}^\dagger \mathbf{H}^{\dagger T} \mathbf{A}^T \right)^{-1}$ and $\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}$, we have

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \mathbf{H}^\dagger \mathbf{x} - \mathbf{H}^\dagger \mathbf{H}^{\dagger T} \mathbf{A}^T \left(\mathbf{A} \mathbf{H}^\dagger \mathbf{H}^{\dagger T} \mathbf{A}^T \right)^{-1} \left(\tilde{\mathbf{A}} \mathbf{x} - \mathbf{b} \right) \\ &= \hat{\boldsymbol{\theta}}_{uc} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T \left(\mathbf{A} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T \right)^{-1} \left(\mathbf{A} \hat{\boldsymbol{\theta}}_{uc} - \mathbf{b} \right). \end{aligned}$$