

EE6617: Assignment 2

Due data: 09 July 2015, 7:00pm, during the class

100 points in total and the last problem is for additional (or bonus) points

2.1 In Tutorial 3, we have looked at the problem estimating two parameters A and σ^2 by using CRLB theorem for the observation $x[n] = A + w[n]$, $n = 0, 1, \dots, N-1$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. Unfortunately, CRLB factorization does not exist for this estimation problem. In this assignment, we want to check whether or not an MVU estimator for $\boldsymbol{\theta} = [A \ \sigma^2]^T$ can be designed by using Neyman-Fisher factorization and RBLs theorem.

- (a) Use Neyman-Fisher factorization and find sufficient statistics for $\boldsymbol{\theta} = [A \ \sigma^2]^T$.
- (b) Since $\{x[n]\}_{n=0}^{N-1}$ are Gaussian distributed, we assume that $\mathbf{T}(\mathbf{x})$ is complete ($T_1(\mathbf{x})$ is still Gaussian and $T_2(\mathbf{x})$ is chi-square distributed). Compute $E[\mathbf{T}(\mathbf{x})]$.
- (c) Design the estimator by RBLs theorem such that $\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{T}(\mathbf{x}))$ and $E[\mathbf{g}(\mathbf{T}(\mathbf{x}))] = [A \ \sigma^2]^T$. First try to evaluate $E[\frac{1}{N}T_1(\mathbf{x})]$ and $E[\frac{1}{N}T_2(\mathbf{x})]$. You may realize $E[\frac{1}{N}T_2(\mathbf{x})]$ is biased, and removing the bias term could be done by $E[\frac{1}{N}T_2(\mathbf{x}) - (\frac{1}{N}T_1(\mathbf{x}))^2]$ and scaling the result.

2.2 Assume that $x[n]$ is the result of Bernoulli trial (a coin toss) with

$$\begin{aligned} p(x[n] = 1) &= \theta \\ p(x[n] = 0) &= 1 - \theta \end{aligned}$$

and the N i.i.d. observations have been made. Assume the Neyman-Fisher factorization theorem holds for discrete random variables and find a sufficient statistics for θ . Notice $p(x[n]; \theta) = \theta^{x[n]}(1 - \theta)^{1-x[n]}$. Then, assuming completeness, find the MVU estimator of θ .

2.3 Assume that $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $E[\mathbf{w}] = \mathbf{0}$ and $E[\mathbf{w}\mathbf{w}^T] = \mathbf{C}$. Let $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{q \times 1}$ be the BLUE for $\boldsymbol{\theta} \in \mathbb{R}^{q \times 1}$. Prove that if we wish to estimate

$$\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\theta} + \mathbf{b}$$

where $\mathbf{B} \in \mathbb{R}^{q \times q}$ is invertible and $\mathbf{b} \in \mathbb{R}^{q \times 1}$ is a known vector, the BLUE of $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}} = \mathbf{B}\hat{\boldsymbol{\theta}} + \mathbf{b}.$$

Hint: Try to express $\boldsymbol{\theta}$ in $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ in terms of $\boldsymbol{\alpha}$.

2.4 Observe N samples written in a vector form as $\mathbf{x} = \mathbf{h}\theta + \mathbf{w}$ where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and $\mathbf{h}$ is an $N \times 1$ vector. We know both $\mathbf{h}$ and $\mathbf{C}$.

- (a) Derive the MVU estimator for θ . For full credit use either CRLB theorem or a sufficient statistic argument.
- (b) Derive the variance of the estimator in (a). You can compute this by using CRLB theorem.
- (c) Suppose that we control $\mathbf{h}$, but we are forced to have $\mathbf{h}^T \mathbf{h} = 1$. Also suppose that $\mathbf{C}$ has a factorization of the form

$$\mathbf{C} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$$

where $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_N] \in \mathbb{R}^{N \times N}$ is an unitary matrix such that $\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I}$ and $\mathbf{\Lambda}$ is a diagonal matrix with diagonal entries $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Design the $\mathbf{h}$ that minimize the variance in (b) and give the corresponding minimum variance.

2.5 We consider the estimation of the noise variance σ^2 in the model $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{w}$ is zero mean with covariance matrix $\sigma^2 \mathbf{I}$. In the problem 2.1, we have considered the case when the noise is Gaussian and the PDF is known. In this problem, however only those $E[\mathbf{w}] = \mathbf{0}$ and $E[\mathbf{w}\mathbf{w}^T] = \sigma^2 \mathbf{I}$ are known and the PDF is unknown.

We already studied when $\mathbf{H} \in \mathbb{R}^{N \times p}$ with $\text{rank}(\mathbf{H}) = p$, the LSE of $\mathbf{H}\boldsymbol{\theta}$ is $\mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$. Based on the LSE, we propose a noise variance estimator

$$\hat{\sigma}^2 = \frac{1}{N} \left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}} \right)^T \left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}} \right)$$

where $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$. Is this estimator is unbiased? The identities $E[\text{tr}(\mathbf{A}\mathbf{B})] = \text{tr}(E[\mathbf{B}\mathbf{A}]) = E[\text{tr}(\mathbf{B}\mathbf{A})] = \text{tr}(E[\mathbf{B}\mathbf{A}])$ will be found useful. If the estimator is biased, can you propose one that is unbiased?

2.6 Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\text{rank}(\mathbf{A}) = r$. Then, SVD of A is given

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ with $\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I}_m$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ and $\mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T = \mathbf{I}_n$, and $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ with $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_p)$ such that $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r \geq 0$ and $\sigma_{r+1} = \cdots = \sigma_p = 0$. Denote $\bar{\mathbf{\Sigma}} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, $\bar{\mathbf{U}} = [\mathbf{U}]_{1:r}$, and $\bar{\mathbf{V}} = [\mathbf{V}]_{1:r}$.

(a) show that

$$\sigma_1 = \max_{\mathbf{y}} \frac{\|\mathbf{A}\mathbf{y}\|_2}{\|\mathbf{y}\|_2}.$$

(b) show that

$$\sigma_1 = \max_{\mathbf{x}} \frac{\|\mathbf{x}^T \mathbf{A}\|_2}{\|\mathbf{x}\|_2}.$$

(c) show that

$$\sigma_1 = \max_{\mathbf{x}, \mathbf{y}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}.$$

2.7 Given a symmetric $\mathbf{A} \in \mathbb{S}^n$, show that $\mathbf{A}$ is negative definite if all the eigenvalues (or singular values) are negative.

Sol. 2.1 (a) By Neyman-Fisher factorization,

$$\begin{aligned} p(\mathbf{x}; \boldsymbol{\theta}) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right) \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x[n]^2 - 2A \sum_{n=0}^{N-1} x[n] + NA^2 \right) \right). \end{aligned}$$

Set $T_1(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ and $T_2(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n]$. Then,

$$\begin{aligned} g(\mathbf{T}(\mathbf{x}); \boldsymbol{\theta}) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left(-\frac{1}{2\sigma^2} (T_2(\mathbf{x}) - 2AT_1(\mathbf{x}) + NA^2) \right) \\ h(\mathbf{x}) &= 1 \end{aligned}$$

and by Neyman-Fisher factorization

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} T_1(\mathbf{x}) = \sum_{n=0}^{N-1} x[n] \\ T_2(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}$$

is the sufficient statistics.

(b) $E[T_1(\mathbf{x})] = NA$ and

$$E[T_2(\mathbf{x})] = E \left[\sum_{n=0}^{N-1} x^2[n] \right] = \sum_{n=0}^{N-1} E[x^2[n]] = \sum_{n=0}^{N-1} (\sigma^2 + A^2) = N(\sigma^2 + A^2).$$

(c) First try to remove the dependency on N by dividing $T_1(\mathbf{x})$ and $T_2(\mathbf{x})$ by N , i.e.,

$$\begin{aligned} E \left[\frac{1}{N} T_1(\mathbf{x}) \right] &= A \\ E \left[\frac{1}{N} T_2(\mathbf{x}) \right] &= \sigma^2 + A^2. \end{aligned}$$

Now, for $\frac{1}{N} T_2(\mathbf{x})$, if the effect of A^2 is removed, then we are done. Try to subtract out $\left(\frac{1}{N} T_1(\mathbf{x})\right)^2$ to get rid of A^2 .

$$\begin{aligned} E \left[\left(\frac{1}{N} T_1(\mathbf{x}) \right)^2 \right] &= E \left[\frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} x[n]x[m] \right] \\ &= \frac{A^2 + \sigma^2}{N} + \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0, m \neq n}^{N-1} E[x[n]x[m]] \\ &= \frac{A^2 + \sigma^2}{N} + \frac{1}{N^2} N(N-1)A^2 \\ &= A^2 + \sigma^2/N \end{aligned}$$

and

$$E \left[\frac{1}{N} T_2(\mathbf{x}) - \left(\frac{1}{N} T_1(\mathbf{x}) \right)^2 \right] = \sigma^2 \frac{N-1}{N} \quad (1)$$

Thus, now we design the MVU estimator

$$\begin{aligned}
\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{T}(\mathbf{x})) &= \begin{bmatrix} \frac{1}{N} T_1(\mathbf{x}) \\ \frac{N}{N-1} \left(\frac{1}{N} T_2(\mathbf{x}) - \left(\frac{1}{N} T_1(\mathbf{x}) \right)^2 \right) \end{bmatrix} \\
&= \begin{bmatrix} \bar{\mathbf{x}} \\ \frac{1}{N-1} \left(\sum_{n=0}^{N-1} x^2[n] - N \bar{\mathbf{x}}^2 \right) \end{bmatrix} \\
&= \begin{bmatrix} \bar{\mathbf{x}} \\ \frac{1}{N-1} \left(\sum_{n=0}^{N-1} (x[n] - \bar{\mathbf{x}})^2 \right) \end{bmatrix}.
\end{aligned}$$

The $\bar{\mathbf{x}} = \sum_{n=0}^{N-1} x[n]$ is the sample mean and $\bar{\sigma}^2 = \frac{1}{N-1} \left(\sum_{n=0}^{N-1} (x[n] - \bar{\mathbf{x}})^2 \right)$ is the sample variance. It is well known that $\bar{\mathbf{x}}$ and $\bar{\sigma}^2$ are independent and that

$$\begin{aligned}
\bar{\mathbf{x}} &\sim \mathcal{N} \left(A, \frac{\sigma^2}{N} \right) \\
\frac{(N-1)\bar{\sigma}^2}{\sigma^2} &\sim \chi_{N-1}^2.
\end{aligned}$$

Sol. 2.2 The PDF can be equivalently written as

$$p(x[n]; \theta) = \theta^{x[n]} (1 - \theta)^{1-x[n]}.$$

Since $x[n]$ is i.i.d., the PDF of $\mathbf{x}$ is

$$\begin{aligned}
p(\mathbf{x}; \theta) &= \prod_{n=0}^{N-1} \theta^{x[n]} (1 - \theta)^{1-x[n]} \\
&= \underbrace{\theta^{\sum_{n=0}^{N-1} x[n]} (1 - \theta)^{N - \sum_{n=0}^{N-1} x[n]}}_{g(T(\mathbf{x}); \theta)} \underbrace{1}_{h(\mathbf{x})}
\end{aligned}$$

$\Rightarrow T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ is sufficient statistics.

$\Rightarrow$ Assume completeness and apply RBLs theorem

$\Rightarrow$ Consider an estimator $g(T(\mathbf{x}))$ and compute $E[T(\mathbf{x})] = \sum_{n=0}^{N-1} E[x[n]] = N\theta$. In order to have $E[g(T(\mathbf{x}))] = \theta$, the estimator should be designed such that $\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$.

Sol. 2.3 Given $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\theta} + \mathbf{b}$, one can manipulate $\boldsymbol{\theta} = \mathbf{B}^{-1}(\boldsymbol{\alpha} - \mathbf{b})$. Then,

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} = \mathbf{H}\mathbf{B}^{-1}\boldsymbol{\alpha} + \mathbf{w} - \mathbf{H}\mathbf{B}^{-1}\mathbf{b}. \quad (2)$$

Now, (2) is the linear model with non-zero mean noise. Similar to the “linear MVU estimator”, we could equalize the biased noise to zero mean by

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{H}\mathbf{B}^{-1}\mathbf{b} = \mathbf{H}\mathbf{B}^{-1}\boldsymbol{\alpha} + \mathbf{w}. \quad (3)$$

Give (3), the BLUE for α is

$$\begin{aligned}
\hat{\alpha} &= (\mathbf{B}^{-T} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H} \mathbf{B}^{-1})^{-1} \mathbf{B}^{-T} \mathbf{H}^T \mathbf{C}^{-1} \tilde{\mathbf{x}} \\
&= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{B}^T \mathbf{B}^{-T} \mathbf{H}^T \mathbf{C}^{-1} \tilde{\mathbf{x}} \\
&= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \tilde{\mathbf{x}} \\
&= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} + \mathbf{H} \mathbf{B}^{-1} \mathbf{b}) \\
&= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} + \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H} \mathbf{B}^{-1} \mathbf{b} \\
&= \mathbf{B} \hat{\theta} + \mathbf{b}.
\end{aligned}$$

Sol. 2.4 (a) For the PDF $p(\mathbf{x}; \theta)$,

$$\begin{aligned}
p(\mathbf{x}; \theta) &= \frac{1}{(2\pi)^{N/2} \sqrt{\det(\mathbf{C})}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{h}\theta)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{h}\theta) \right) \\
\ln p(\mathbf{x}; \theta) &= -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln \det(\mathbf{C}) - \frac{1}{2} (\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} + \theta^2 \mathbf{h}^T \mathbf{C}^{-1} \mathbf{h} - 2\theta \mathbf{h}^T \mathbf{C}^{-1} \mathbf{x}) \\
\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} &= \mathbf{h}^T \mathbf{C}^{-1} \mathbf{x} - \theta (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h}) \\
&= (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h}) ((\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1} \mathbf{h}^T \mathbf{C}^{-1} \mathbf{x} - \theta).
\end{aligned}$$

Thus, by CRLB theorem

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

(b) By CRLB,

$$I(\theta) = \mathbf{h}^T \mathbf{C}^{-1} \mathbf{h}.$$

So,

$$\text{var}(\hat{\theta}) = I^{-1}(\theta) = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1}$$

(c) The variance is $\text{var}(\hat{\theta}) = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1}$ and this is minimized when $(\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})$ is maximized. Notice that $\mathbf{C}^{-1} = \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^T$ and by Rayleigh quotient theorem, $(\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})$ is maximized when

$$\mathbf{h} = \mathbf{u}_N$$

and it achieves the lower bound

$$\text{var}(\hat{\theta}) \geq (\lambda_N^{-1})^{-1} = \lambda_N.$$

Sol. 2.5

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{N} \mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x} \\
&= \frac{1}{N} \mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x}.
\end{aligned}$$

$$\mathbf{x} = \mathbf{H}\theta + \mathbf{w}, \quad E[\mathbf{w}] = \mathbf{0}, \quad E[\mathbf{w}\mathbf{w}^T] = \sigma^2 \mathbf{I}.$$

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{N} \theta^T \mathbf{H}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{H} \theta + \frac{1}{N} \mathbf{w}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{H} \theta \\
&\quad + \frac{1}{N} \theta^T \mathbf{H}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{w} + \frac{1}{N} \mathbf{w}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{w}
\end{aligned}$$

$$E[\hat{\sigma}^2] = \frac{1}{N} \boldsymbol{\theta}^T \mathbf{H}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{H} \boldsymbol{\theta} + \frac{\sigma^2}{N} N - \frac{1}{N} \text{tr} (E [\mathbf{w}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}])$$

Since $\text{tr} (E [\mathbf{w}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}]) = \text{tr} (E [\mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T]) = \text{tr} (\mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E [\mathbf{w} \mathbf{w}^T]) = \sigma^2 \text{tr} (\mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) = \sigma^2 \text{tr} ((\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H}) = \sigma^2 \text{tr} (\mathbf{I}_p) = \sigma^2 p$.

$$\begin{aligned} E[\hat{\sigma}^2] &= \frac{1}{N} \boldsymbol{\theta}^T \mathbf{H}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{H} \boldsymbol{\theta} + \frac{\sigma^2}{N} N - \frac{\sigma^2}{N} p \\ &= \frac{1}{N} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} - \frac{1}{N} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} + \frac{\sigma^2}{N} (N - p) \\ &= \frac{\sigma^2}{N} (N - p) \end{aligned}$$

Thus, the estimator is biased unless $N = p$.

An unbiased version could be

$$\hat{\sigma}_{ub}^2 = \frac{N}{N - p} \hat{\sigma}^2 \quad (4)$$

Sol. 2.6 (a) Given $\mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$,

$$\max_{\mathbf{y}} \frac{\|\mathbf{A} \mathbf{y}\|_2}{\|\mathbf{y}\|_2} = \max_{\tilde{\mathbf{y}} = \frac{\mathbf{y}}{\|\mathbf{y}\|_2}} \sqrt{\tilde{\mathbf{y}}^T \mathbf{A}^T \mathbf{A} \tilde{\mathbf{y}}} = \sqrt{\max_{\tilde{\mathbf{y}}} \tilde{\mathbf{y}}^T \mathbf{V} \boldsymbol{\Sigma}^T \boldsymbol{\Sigma} \mathbf{V}^T \tilde{\mathbf{y}}} = \sqrt{\sigma_1^2}$$

where the last equality follows from the Rayleigh quotient theorem.

(b) Similar to (a),

$$\max_{\mathbf{x}} \frac{\|\mathbf{x}^T \mathbf{A}\|_2}{\|\mathbf{x}\|_2} = \max_{\tilde{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}} \sqrt{\tilde{\mathbf{x}}^T \mathbf{A} \mathbf{A}^T \tilde{\mathbf{x}}} = \sqrt{\max_{\tilde{\mathbf{x}}} \tilde{\mathbf{x}}^T \mathbf{U} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^T \mathbf{U}^T \tilde{\mathbf{x}}} = \sqrt{\sigma_1^2} = \sigma_1.$$

(c) This can be shown by Cauchy-Schwartz inequality $\mathbf{a}^T \mathbf{b} \leq \|\mathbf{a}\|_2 \|\mathbf{b}\|_2$ where the equality is achieved if and only if $\mathbf{a} = c \mathbf{b}$ for arbitrary $c \geq 0$. Consider $\mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T = \bar{\mathbf{U}} \bar{\boldsymbol{\Sigma}} \bar{\mathbf{V}}^T$.

$$\begin{aligned} \max_{\mathbf{x}, \mathbf{y}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} &= \max_{\mathbf{x}, \mathbf{y}} \frac{\mathbf{x}^T \bar{\mathbf{U}} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{V}}^T \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \\ &= \max_{\tilde{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|_2, \tilde{\mathbf{y}} = \mathbf{y}/\|\mathbf{y}\|_2} \tilde{\mathbf{x}}^T \bar{\mathbf{U}} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{V}}^T \tilde{\mathbf{y}} \\ &= \max_{\tilde{\mathbf{x}}} (\bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{U}}^T \tilde{\mathbf{x}})^T \max_{\tilde{\mathbf{y}}} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{V}}^T \tilde{\mathbf{y}} \\ &\leq \sqrt{\max_{\tilde{\mathbf{x}}} \tilde{\mathbf{x}}^T \bar{\mathbf{U}} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{U}}^T \tilde{\mathbf{x}}} \sqrt{\max_{\tilde{\mathbf{y}}} \tilde{\mathbf{y}}^T \bar{\mathbf{V}} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{V}}^T \tilde{\mathbf{y}}} \\ &= \sigma_1^{1/2} \sigma_1^{1/2} = \sigma_1. \end{aligned} \quad (5)$$

The inequality holds as the equality if and only if

$$\bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{U}}^T \tilde{\mathbf{x}} = c \bar{\boldsymbol{\Sigma}}^{1/2} \bar{\mathbf{V}}^T \tilde{\mathbf{y}}, \quad c \geq 0$$

and this holds if $\tilde{\mathbf{x}} = \mathbf{u}_1$ and $\tilde{\mathbf{y}} = \mathbf{v}_1$ with $c = 1$.

Sol. 2.7 Given $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ with $\text{rank}(\mathbf{A}) = r \leq n$, one can write

$$\mathbf{A} = \sum_{i=1}^r \lambda_i \mathbf{v}_i \mathbf{v}_i^T.$$

To check the positive definiteness of $\mathbf{A}$, compute for $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^r \lambda_i \mathbf{x}^T \mathbf{v}_i \mathbf{v}_i^T \mathbf{x} = \sum_{i=1}^r \lambda_i (\mathbf{x}^T \mathbf{v}_i)^2.$$

Since $(\mathbf{x}^T \mathbf{v}_i)^2 \geq 0$, it is obvious that $\mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0$ if $\lambda_i \leq 0$, $i = 1, \dots, r$. This conclude the proof.