

EE6617: Assignment 1

Due data: 25 June 2015, 7:00pm, during the class

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

1. Suppose $\mathbf{A} = \mathbf{A}^T$, i.e., $\mathbf{A}$ is symmetric. Show that the eigenvalue decomposition of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \quad (1)$$

where $\mathbf{V} \in \mathbb{R}^{n \times n}$ is orthonormal, i.e., $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$ and $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$ is the diagonal matrix.

2. (a) Show that if $\mathbf{A}$ is symmetric and positive definite, the eigenvalues of $\mathbf{A}$ are positive.
(b) Show, for $\mathbf{X} \in \mathbb{R}^{m \times n}$, that $\dim(\mathcal{R}(\mathbf{X})) = \text{rank}(\mathbf{X})$.
3. Suppose $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{S}^n$, i.e., symmetric. The ratio between the norm of $\|\mathbf{x}\|^2$ and the quadratic form $\mathbf{x}^T \mathbf{A} \mathbf{x}$ is called Rayleigh quotient

$$\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}.$$

The eigenvalues of $\mathbf{A}$ are ordered in decreasing order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Show that the minimum of Rayleigh quotient is the least eigenvalue λ_n of $\mathbf{A}$, i.e.,

$$\lambda_n = \min_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \quad (2)$$

and this is achieved by the eigenvector $\mathbf{x} = \mathbf{v}_n$ associated with λ_n .

4. Suppose the observations $x[n] = A + Bn + w[n]$, $n = 0, \dots, N-1$, $A \in \mathbb{R}$, $B \in \mathbb{R}$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. We want to estimate A and B . This problem is equivalent to estimate a line $\hat{A} + \hat{B}n$ to best fit to $x[n]$.
- (a) Find the likelihood function $p(\mathbf{x}; \boldsymbol{\theta})$ with $\boldsymbol{\theta} = [A \ B]^T$.
- (b) Compute the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$.
- (c) Find CRLBs for $\hat{A}$ and $\hat{B}$. Given the CRLBs, state which parameter is easier to be estimated than the other?
5. Suppose you want to estimate $\boldsymbol{\alpha} \in \mathbb{R}^{r \times 1}$ which is a linear function of the parameters $\boldsymbol{\theta} \in \mathbb{R}^{p \times 1}$,

$$\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta} + \mathbf{b},$$

where $\mathbf{A} \in \mathbb{R}^{r \times p}$ and $\mathbf{b} \in \mathbb{R}^{r \times 1}$. Let's assume that after some investigation, you realize that you are able to find an estimator for $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$. Show that if $\hat{\boldsymbol{\theta}}$ is [efficient](#), i.e., $\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta})$ and $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$, its linear mapping $\hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{b}$ is also [efficient](#). You will find following formula to be useful

$$\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} (\mathbf{g}(\boldsymbol{\theta}))^T = \frac{\partial}{\partial \boldsymbol{\theta}} (\boldsymbol{\theta}^T \mathbf{A}^T + \mathbf{b}^T) = \mathbf{I} \mathbf{A}^T = \mathbf{A}^T.$$

6. In the problem **3.3** of Tutorial 3, given the observation $x[n] = A + w[n]$, $n = 1, \dots, N - 1$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, we have computed CRLB of $\boldsymbol{\theta} = [A \ \sigma^2]^T$.

The ratio between the signal power A^2 and noise power σ^2 is called signal-to-noise-ratio (SNR) and one can write SNR as a function of the parameters $\boldsymbol{\theta} = [A \ \sigma^2]^T$,

$$SNR = \alpha = g(\boldsymbol{\theta}) = \frac{A^2}{\sigma^2}.$$

Estimating the SNR in a communications system is a very important step to adapt the signal transmission and reception (e.g., modulation and demodulation) strategies to the operating SNR.

Find CRLB of $\hat{\alpha}$, i.e, calculate the bound of $\text{var}(\hat{\alpha})$. You would like to use the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ computed in Tutorial 3 and apply bounds in CRLB-Multivariate function of $\boldsymbol{\theta}$.

7. For square and non-singular matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, show that $\det(\mathbf{A}^T) = \det(\mathbf{A})$.

Sol. 1 Suppose the eigenvalue decomposition, $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$. Since $\mathbf{A}$ is symmetric,

$$\mathbf{A} = \mathbf{A}^T \Rightarrow \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} = (\mathbf{V}^{-1})^T \mathbf{\Lambda} \mathbf{V}^T$$

This implies $\mathbf{V}^{-1} = \mathbf{V}^T$ and $\mathbf{V} = (\mathbf{V}^{-1})^T$. The $\mathbf{V}^{-1} = \mathbf{V}^T$ implies $\mathbf{V} = (\mathbf{V}^{-1})^T$ or vice versa. This can be checked by $(\mathbf{V}^{-1})^T = (\mathbf{V}^T)^T \Rightarrow (\mathbf{V}^{-1})^T = \mathbf{V}$.

Sol. 2 (a) Recall the definition of the positive definite, i.e., a matrix $\mathbf{A}$ is called positive definite if $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for $\mathbf{x} \neq 0$. Now, we apply the condition $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ to the symmetric matrix $\mathbf{A}$. The eigenvalue decomposition of $\mathbf{A}$ is $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ and we pick the i th eigenvector associated with the i th eigenvalue of $\mathbf{A}$ as $\mathbf{x} = \mathbf{v}_i$. Then, the condition returns $\mathbf{v}_i^T \mathbf{A} \mathbf{v}_i = \lambda_i > 0$ because of the positive definiteness. This implies the i th eigenvalue λ_i of the positive definite matrix is positive. Repeating this $\forall i$ concludes the proof.

(b)

$$\begin{aligned} \text{rank}(\mathbf{X}) &= \text{Number of independent columns of } (\mathbf{x}_1, \dots, \mathbf{x}_n) \\ &= \text{Number of bases of } \mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_n) \\ &= \text{Number of bases of } \mathcal{R}(\mathbf{X}) \\ &= \dim(\mathcal{R}(\mathbf{X})). \end{aligned}$$

Sol. 3 Given the eigenvalue decomposition $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ where $\mathbf{V}\mathbf{V}^T = \mathbf{I}$, the Rayleigh quotient is rewritten by

$$\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \frac{\mathbf{x}^T \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \mathbf{x}}{\mathbf{x}^T \mathbf{V} \mathbf{V}^T \mathbf{x}} = \frac{\mathbf{y}^T \mathbf{\Lambda} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \quad (3)$$

where $\mathbf{y} = \mathbf{V}^T \mathbf{x}$. The (3) is

$$\frac{\mathbf{y}^T \mathbf{\Lambda} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \frac{y_1^2 \lambda_1 + y_2^2 \lambda_2 + \dots + y_n^2 \lambda_n}{y_1^2 + y_2^2 + \dots + y_n^2} \geq \frac{y_1^2 \lambda_n + y_2^2 \lambda_n + \dots + y_n^2 \lambda_n}{y_1^2 + y_2^2 + \dots + y_n^2} = \lambda_n. \quad (4)$$

Notice that we can not change λ_i . Thus, only way to achieve (4) is to design $\mathbf{x}$ such that $y_1 = \dots = y_{n-1} = 0$ and $y_n \neq 0$. Notice that $\mathbf{y} = \mathbf{V}^T \mathbf{x}$. If $\mathbf{x} = \mathbf{v}_n$, $\mathbf{y} = [0 \ 0 \ \dots \ 0 \ 1]^T$ and we have

$$\frac{\mathbf{v}_n^T \mathbf{A} \mathbf{v}_n}{\mathbf{v}_n^T \mathbf{v}_n} = \lambda_n \frac{\mathbf{v}_n^T \mathbf{v}_n}{\mathbf{v}_n^T \mathbf{v}_n} = \lambda_n.$$

Sol. 4 (a) For $\boldsymbol{\theta} = [A \ B]$,

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

(b) The Fisher information matrix for $\boldsymbol{\theta} = [A \ B]$ is defined by

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} -E \left[\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial^2 A} \right] & -E \left[\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial A \partial B} \right] \\ -E \left[\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial B \partial A} \right] & -E \left[\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial^2 B} \right] \end{bmatrix}$$

$$\begin{aligned}\frac{\partial \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial A} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - (A + Bn)) \\ \frac{\partial \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial B} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n(x[n] - (A + Bn))\end{aligned}$$

Since $E\left[\frac{\partial \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial A}\right] = 0$ and $E\left[\frac{\partial \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial B}\right] = 0$, the regularity condition holds. Now, we further compute

$$\begin{aligned}-\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial^2 A} &= \frac{N}{\sigma^2} \\ -\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial^2 B} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^2 = \frac{N(N-1)(2N-1)}{6\sigma^2} \\ -\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial A \partial B} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n = \frac{N(N-1)}{2\sigma^2}\end{aligned}\tag{5}$$

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & \frac{N(N-1)}{2\sigma^2} \\ \frac{N(N-1)}{2\sigma^2} & \frac{N(N-1)(2N-1)}{6\sigma^2} \end{bmatrix}.$$

- (c) Based on the formula $\mathbf{A}^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix}$.
Now, the inverse of $\mathbf{I}(\boldsymbol{\theta})$ is

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{2(2N-1)\sigma^2}{N(N+1)} & -\frac{6\sigma^2}{N(N+1)} \\ -\frac{6\sigma^2}{N(N+1)} & \frac{12\sigma^2}{N(N^2-1)} \end{bmatrix}.$$

Thus,

$$\begin{aligned}\text{var}(\hat{A}) &\geq \frac{2(2N-1)\sigma^2}{N(N+1)} \\ \text{var}(\hat{B}) &\geq \frac{12\sigma^2}{N(N^2-1)}.\end{aligned}$$

Comparing two bounds, we have $\frac{12\sigma^2}{N(N^2-1)} = O\left(\frac{1}{N^3}\right) \leq \frac{2(2N-1)\sigma^2}{N(N+1)} = O\left(\frac{1}{N}\right)$, implying it is much harder to estimate A than B . Why? This is because as n grows, B dominates the mean of $x[n]$.

Sol. 5 We first show the unbiasedness of $\hat{\boldsymbol{\alpha}}$ by $E[\hat{\boldsymbol{\alpha}}] = \mathbf{A}\boldsymbol{\theta} + \mathbf{b} = \boldsymbol{\alpha}$. Next, in order to show $\mathbf{C}_{\hat{\boldsymbol{\alpha}}} = \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \mathbf{I}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)$, we first evaluate the covariance matrix of $\hat{\boldsymbol{\alpha}}$,

$$\begin{aligned}\mathbf{C}_{\hat{\boldsymbol{\alpha}}} = E[(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T] &= E\left[\mathbf{A}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T\right] = \mathbf{A}E\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\right] \mathbf{A}^T \\ &= \mathbf{A}\mathbf{C}_{\hat{\boldsymbol{\theta}}}\mathbf{A}^T.\end{aligned}$$

From the formula $\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{A}^T$, we have

$$\left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{I}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{A} \mathbf{C}_{\hat{\boldsymbol{\theta}}} \mathbf{A}^T,$$

verifying $\mathbf{C}_{\hat{\boldsymbol{\alpha}}} = \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{I}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{A} \mathbf{C}_{\hat{\boldsymbol{\theta}}} \mathbf{A}^T$. Thus, the linear mapping $\hat{\boldsymbol{\alpha}} = \mathbf{A} \hat{\boldsymbol{\theta}} + \mathbf{b}$ is efficient.

Sol. 6 We want to estimate $SNR = g(\boldsymbol{\theta}) = \frac{A^2}{\sigma^2}$ where $\boldsymbol{\theta} = [A \ \sigma^2]^T$. The CRLB for $\hat{\alpha}$ is

$$\text{var}(\hat{\alpha}) \geq \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{I}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)$$

and since we know $\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$, we need to compute $\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$.

$$\begin{aligned} \frac{\partial g(\boldsymbol{\theta})}{\partial A} &= \frac{2A}{\sigma^2} \\ \frac{\partial g(\boldsymbol{\theta})}{\partial \sigma^2} &= -\frac{A^2}{\sigma^4} \Rightarrow \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{2A}{\sigma^2} \\ -\frac{A^2}{\sigma^4} \end{bmatrix} \end{aligned}$$

Thus,

$$\begin{aligned} \text{var}(\hat{\alpha}) &\geq \begin{bmatrix} \frac{2A}{\sigma^2} & -\frac{A^2}{\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix} \begin{bmatrix} \frac{2A}{\sigma^2} \\ -\frac{A^2}{\sigma^4} \end{bmatrix} \\ &= \frac{4A^2\sigma^2 + 2A^4}{N\sigma^4} = \frac{4\alpha + 2\alpha^2}{N}. \end{aligned}$$

Sol. 7 The definition of determinant is $\det(\mathbf{A}) = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma_i}$. Let the i th and j th element of $\mathbf{A}^T$ as $\tilde{a}_{ij}$. Then,

$$\det(\mathbf{A}^T) = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \prod_{i=1}^n \tilde{a}_{i,\sigma_i}.$$

Given a permutation σ that maps $i \xrightarrow{\sigma} j$, there always exists inverse map $i \xleftarrow{\sigma^{-1}} j$, because the map σ is one-to-one. Thus, for any permutation σ , we have

$$\begin{aligned} \prod_{i=1}^n \tilde{a}_{i,\sigma_i} &= \prod_{i=1}^n a_{\sigma_i,i} \\ &= \prod_{j=1}^n a_{j,\sigma_j^{-1}}. \end{aligned}$$

Notice that $\text{sgn}(\sigma) = \text{sgn}(\sigma^{-1})$. Thus,

$$\begin{aligned}
\det(\mathbf{A}^T) &= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \prod_{i=1}^n \tilde{a}_{i, \sigma_i} \\
&= \sum_{\sigma^{-1} \in \mathcal{S}_n} \text{sgn}(\sigma^{-1}) \prod_{j=1}^n a_{j, \sigma_j^{-1}} \\
&= \sum_{\varsigma \in \mathcal{S}_n} \text{sgn}(\varsigma) \prod_{j=1}^n a_{j, \varsigma_j}
\end{aligned}$$

where the sum is computed over all permutations $\sigma^{-1} \in \mathcal{S}_n$ of the set $\{1, \dots, n\}$, whose sum is same as taking with respect to all permutations $\varsigma \in \mathcal{S}_n$.