Delft University of Technology
Faculty of Electrical Engineering, Mathematics, and Computer Science
Circuits and Systems Group

ET 4386 Estimation and Detection - Answers

27 January 2017, 9:00–12:00

This exam is a closed book exam. One double sided self-handwritten A4 formula cheat sheet is allowed.

This exam has four questions (40 points in total).

Question 1 (10 points)

(2 p) (a)

$$p(\mathbf{x};\lambda) = \begin{cases} \lambda^N \exp\left(-\lambda \sum_{n=0}^{N-1} x[n]\right) & x[n] > 0 \\ 0 & x[n] < 0. \end{cases}$$

$$\frac{\partial \log p(\mathbf{x};\lambda)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{n=0}^{N-01} x[n]$$

$$\frac{\partial^2 \log p(\mathbf{x};\lambda)}{\partial \lambda^2} = -\frac{N}{\lambda^2} = -I(\lambda)$$

CRLB: $\text{var}(\hat{\lambda}) \geq 1/I(\lambda) = \frac{\lambda^2}{N}$

(2 p) (b) An unbiased etimator that attains the CRLB bound can be found if and only if

$$\frac{\partial \log p(\mathbf{x};\lambda)}{\partial \lambda} = I(\lambda)(g(\mathbf{x}) - \lambda).$$

In this case, it is impossible to write $\frac{\partial \log p(\mathbf{x};\lambda)}{\partial \lambda}$ in this form and thus there is no unbiased etimator that attains the CRLB bound.

(2 p) (c)

$$\frac{\partial \log p(\mathbf{x};\lambda)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{n=0}^{N-1} x[n] = 0.$$

From this it follows that $\hat{\lambda}_{MLE} = \frac{N}{\sum_{n=0}^{N-1} x[n]}$.

(2 p) (d) The MAP is given by

$$\arg\max_{\lambda} p(A|\mathbf{x}) = \arg\max_{\lambda} \log(p(\mathbf{x}|A)) + \log(p(A)).$$

Taking the derivative to λ then leads to $\hat{\lambda}_{MAP} = \frac{N}{\sum_{n=0}^{N-1} x[n] + a}$.

(2 p) (e) If $a \to 0$, the variance goes to $\text{var}(\lambda) \to \infty$. In that case, the prior is non-informative and the Bayesian estimator becomes more similar to the classical estimator. Similarly, if N increases, the MAP estimatot will rely more on the data instead of on the prior. This can be shown by rewriting the MAP estimator as $\hat{\lambda}_{MAP} = \frac{1}{\frac{\sum_{n=0}^{N-1} x[n]}{N} + \frac{a}{N}}$ which converges to $\frac{1}{\frac{\sum_{n=0}^{N-1} x[n]}{N}}$ for large N.

Question 2 (10 points)

(2,5 p) **(a)** The Neyman-Pearson detector $T(\mathbf{x})$ is given by $T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s}A$. As $\mathbf{x}$ is Gaussian (under both hypotheses), $T(\mathbf{x})$ is also Gaussian as it is a sum of scaled Gaussian random variables. The first moments and variances under both hypotheses are:

$$E(T(\mathbf{x}), \mathcal{H}_0) = E(\mathbf{w}^T\mathbf{C}^{-1}\mathbf{s}A) = 0$$

$$E(T(\mathbf{x}), \mathcal{H}_1) = E(\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s}A) = A\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = As^2[0] + A2s^2[1]$$

$$var(T(\mathbf{x}), \mathcal{H}_0) = E((\mathbf{w}^T\mathbf{C}^{-1}\mathbf{s}A)^2) = A^2\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = A^2s^2[0] + A^22s^2[1]$$

$$var(T(\mathbf{x}), \mathcal{H}_1) = E((\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s}A - E(\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s}A))^2) = A^2\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = A^2s^2[0] + A^22s^2[1]$$

Under NP, the optimal threshold is given by $\gamma' = Q^{-1}(P_{fa})\sqrt{A^2\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}} = Q^{-1}(P_{fa}\sqrt{A^2s^2[0] + A^22s^2[1]})$. The P_D is then given by $P_D = Q(\frac{\gamma' - A^2\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}{\sqrt{A^2\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}) = Q(Q^{-1}(P_{fa}) - A\sqrt{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}})$.

(2,5 p) **(b)** We know that $P_D = Q(Q^{-1}(P_{fa}) - A\sqrt{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}})$. From this it follows that $P_D > 10P_{fa} = Q(Q^{-1}(P_{fa}) - A\sqrt{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}})$. From this we then obtain $A \geq \frac{Q^{-1}(p_{fa}) - Q^{-1}(10p_{fa})}{\sqrt{3}}$

(2,5 p) **(c)** The detection performance depends on the "SNR": $A^2\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = (s^2[0] + 2s^2[1])$.
By inspection: Given that $s^2[0] + s^2[1] = 2$, the SNR $(s^2[0] + 2s^2[1])$ can be written as $SNR = (2 + s^2[1])$ We know from the constraint $s^2[0] + s^2[1] = 2$, that $s^2[1]$ has a minimum value of 0 and a maximum of 2. The maximum SNR is obtained by putting all energy in $s[1]$, that is, $s[n] = \sqrt{2}$, and setting $s[0] = 0$. This is logical, as the noise variance at $n = 0$ is the highest, while the noise is uncorrelated.
Using eigenvectors of the correlation matrix: The eigenvectors of $\mathbf{C}$ are given by $[1, 0]^T$ and $[0, 1]^T$. Maximizing $\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$ while constraining $\mathbf{s}^T\mathbf{s} = 2$ can be done by maximizing

$$\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} + \lambda(2 - \mathbf{s}^T\mathbf{s}).$$

Taking the derivative with respect to $\mathbf{s}$ and setting this to zero we obtain

$$\mathbf{C}^{-1}\mathbf{s} = \lambda\mathbf{s}.$$

This implies that $\mathbf{s}$ is an eigenvector of $\mathbf{C}^{-1}$. As we want to maximize $\mathbf{s}\mathbf{C}^{-1}\mathbf{s}$, we have

$$\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = \lambda\mathbf{s}^T\mathbf{s} = \lambda 2.$$

From this it follows that we should chose $\mathbf{s}$ as the eigenvector of $\mathbf{C}^{-1}$ whose eigenvalue λ is maximum (or equivalently, as the eigenvector of $\mathbf{C}$ whose eigenvalue is minimum). In this particular case that corresponds to setting $\mathbf{s}^T = [s[0], s[1]] = [0, \sqrt{2}]$.

(2,5 p) (d) Maximizing $\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$ while constraining $\mathbf{s}^T\mathbf{s} = 2$ can be done by maximizing

$$\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} + \lambda(2 - \mathbf{s}^T\mathbf{s}).$$

Taking the derivative with respect to $\mathbf{s}$ and setting this to zero we obtain

$$\mathbf{C}^{-1}\mathbf{s} = \lambda\mathbf{s}.$$

This implies that $\mathbf{s}$ is an eigenvector of $\mathbf{C}^{-1}$. As we want to maximize $\mathbf{s}\mathbf{C}^{-1}\mathbf{s}$, we have

$$\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = \lambda\mathbf{s}^T\mathbf{s}.$$

From this it follows that we should chose $\mathbf{s}$ as the eigenvector of $\mathbf{C}^{-1}$ whose eigenvalue λ is maximum (or equivalently, as the eigenvector of $\mathbf{C}$ whose eigenvalue is minimum).

Question 3 (10 points)

(a) With N independent observations, due to the additive property, the Fisher Information will be

$$I(\rho) = \frac{N}{2} \text{tr} \left[\left(\mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} \right)^2 \right].$$

The Cramér-Rao lower bound (CRLB) is the inverse of the Fisher information.

The inverse of

$$\mathbf{C}(\rho) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

is

$$\mathbf{C}^{-1}(\rho) = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \quad \text{and} \quad \frac{\partial \mathbf{C}(\rho)}{\partial \theta} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

So we can compute

$$\mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} = \frac{1}{1-\rho^2} \begin{bmatrix} -\rho & 1 \\ 1 & -\rho \end{bmatrix},$$

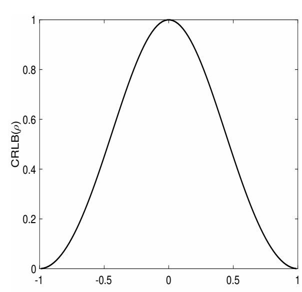
which on squaring results in

$$\left(\mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} \right)^2 = \frac{1}{(1-\rho^2)^2} \begin{bmatrix} 1+\rho^2 & * \\ * & 1+\rho^2 \end{bmatrix},$$

Therefore, the Cramér-Rao lower bound will be

$$\text{var}(\hat{\rho}) \geq \frac{(1-\rho^2)^2}{N(1+\rho^2)}.$$

(b) The CRLB for $\rho \in [-1, 1]$ is shown below.



For $\rho \to \pm 1$, CRLB goes to zero as it is easier to determine if random variables are coherent. On the other hand, CRLB is equal to one (larger variance) if the random variables are completely uncorrelated.

(c) To find the least squares estimator, we will compute the derivative of

$$J(\rho) = \text{tr}\left\{[\mathbf{S} - \mathbf{C}(\rho)]^T[\mathbf{S} - \mathbf{C}(\rho)]\right\} = \text{tr}\left[\mathbf{S}^T\mathbf{S} + \mathbf{C}^T(\rho)\mathbf{C}(\rho) - 2\mathbf{S}^T\mathbf{C}(\rho)\right],$$

towards ρ and set it to zero. So we get

$$\begin{aligned}\frac{\partial J(\rho)}{\partial \rho} &= 2\text{tr}\left\{\mathbf{C}^T(\rho)\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right\} - 2\text{tr}\left\{\mathbf{S}^T(\rho)\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right\} \\ &= 4\rho - 2\text{tr}\left\{\mathbf{S}^T(\rho)\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right\} = 0,\end{aligned}$$

Observing that

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}[n]\mathbf{x}^T[n] = \frac{1}{N}\begin{bmatrix} \sum_{n=1}^N x_1^2[n] & \sum_{n=1}^N x_1[n]x_2[n] \\ \sum_{n=1}^N x_1[n]x_2[n] & \sum_{n=1}^N x_2^2[n] \end{bmatrix},$$

we can compute the LSE of ρ as

$$\hat{\rho} = \frac{1}{N}\sum_{n=1}^{N} x_1[n]x_2[n].$$

(d) The estimator for the correlation coefficient will be non-linear, e.g., see the LSE in the previous question. So BLUE is not appropriate. However, by transforming the observed data and defining $y[n] = x_1[n]x_2[n]$, one may find the BLUE as

$$\hat{\rho} = \mathbf{a}^T\mathbf{y},$$

where $\mathbf{y} = [y[0], \cdots, y[N-1]]^T$ and $\mathbf{a}$ contains the combining weights to be determined.

Question 4 (10 points)

(a) We decide on $\mathcal{H}_1$ if

$$p(\mathbf{x}|\mathcal{H}_1) > p(\mathbf{x}|\mathcal{H}_0)$$

or if

$$\frac{1}{2\pi}(\det\mathbf{C})^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} > \frac{1}{2\pi}(\det\mathbf{C})^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right\}.$$

Taking logarithm on both sides, the minimum P_e will be

$$\mathbf{x}^T\mathbf{C}^{-1}\boldsymbol{\mu} > \frac{1}{2}\boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}.$$

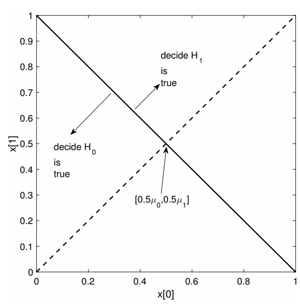
(b) For $\rho = 0$, the minimum P_e simplifies to

$$x[0]\mu_0 + x[1]\mu_1 > \frac{1}{2}\mu_0^2 + \mu_1^2$$

or

$$x[1] > -\frac{\mu_0}{\mu_1}x[0] + \frac{1}{2}\frac{\mu_0^2+\mu_1^2}{\mu_1}.$$

The decision boundary and decision regions is show in the picture below, where the solid line is the decision boundary and the dashed line is the line segment joining the origin and $\boldsymbol{\mu} = [\mu_0, \mu_1]^T$.



The decision boundary has a slope of -1 (i.e., perpendicular to the line segment joining the origin and $\boldsymbol{\mu} = [\mu_0, \mu_1]^T$) and intersects the line segment joining the origin and $\boldsymbol{\mu} = [\mu_0, \mu_1]^T$ at $[\mu_0/2, \mu_1/2]$.

(c) For an arbitrary value of the correlation coefficient, the minimum P_e detector will be

$$x[1] > -x[0]\frac{\mu_0 - \rho\mu_1}{\mu_1 - \rho\mu_0} + \frac{1}{2}\frac{\mu_0^2 + \mu_1^2 - 2\rho\mu_0\mu_1}{\mu_1 - \rho\mu_0}.$$

For $\boldsymbol{\mu} = [1, 1]^T$, the above detection rule simplifies to

$$x[1] > -x[0]\frac{1-\rho}{1-\rho} + \frac{1-\rho}{1-\rho},$$

or equivalently to

$$x[1] > -x[0] + 1.$$

Therefore, for $\boldsymbol{\mu} = [1, 1]^T$ the detector is independent of ρ. This is because, both the equicorrelated samples have common mean (or in other words identical distributions). However, the decision boundary line is still a perpendicular bisector of the line segment from **0** to **1**. This will not be true for $\mu_0 \neq \mu_1$.