Handout 12 — Principles of Digital Communications
Solutions to Graded HW 1 — Apr. 1, 2020

PROBLEM 1. Consider the following hypothesis testing problem:

$$H_0 : f_{Y|H}(y|0) = \exp(-y)$$
$$H_1 : f_{Y|H}(y|1) = 2\exp(-2y)$$
$$H_2 : f_{Y|H}(y|2) = 2y\exp(-y^2),$$

where $y \geq 0$. We want to decide whether $H = 0$ or $H \neq 0$. To this end, we will design an estimator

$$\hat{H}_\alpha(y) = \begin{cases} 0 & y \geq \alpha \\ 1 & y < \alpha. \end{cases}$$

The estimator is evaluated using the following metrics:

$$p_{\text{det}} := P_{\hat{H}_\alpha(Y)|H}(0|0)$$
$$p_{\text{fp}} := \max\{P_{\hat{H}_\alpha(Y)|H}(0|1), P_{\hat{H}_\alpha(Y)|H}(0|2)\}.$$

A good estimator will have high probability of detection p_{det} and low probability of false positive p_{fp}.

a. Calculate the following probabilities : $P_{\hat{H}_\alpha(Y)|H}(0|0)$, $P_{\hat{H}_\alpha(Y)|H}(0|1)$, and $P_{\hat{H}_\alpha(Y)|H}(0|2)$. [Hint : $\frac{d}{dx}\exp(\lambda x) = \lambda\exp(\lambda x)$ and $\frac{d}{dx}\exp(\lambda x^2) = 2\lambda x\exp(\lambda x^2)$.]

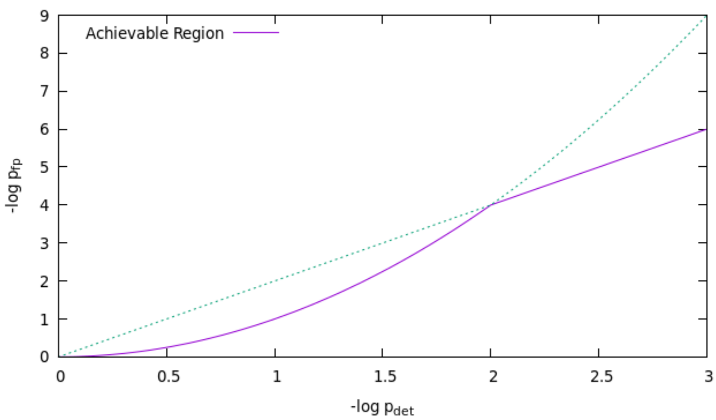
b. Sketch a plot of all points $(-\ln p_{\text{det}}, -\ln p_{\text{fp}})$, $0 \leq -\ln p_{\text{det}} \leq 3$, that can be achieved using $\hat{H}_\alpha$ when we vary α.

SOLUTION 1.

a. Due to the form of $\hat{H}_\alpha(Y)$, we have $P(\hat{H}_\alpha(Y) = 0|H) = P(Y \geq \alpha|H)$. This is equal to

$$P(\hat{H}_\alpha(Y) = 0|H = 0) = \int_\alpha^\infty \exp(-y)dy = \exp(-\alpha)$$
$$P(\hat{H}_\alpha(Y) = 0|H = 1) = \int_\alpha^\infty 2\exp(-2y)dy = \exp(-2\alpha)$$
$$P(\hat{H}_\alpha(Y) = 0|H = 2) = \int_\alpha^\infty 2y\exp(-y^2)dy = \exp(-\alpha^2)$$

b. The achievable region is given as follows



Problem 2. Consider a binary hypothesis test with observation $Y = \begin{bmatrix} Y_1 & Y_2 & Y_3 & Y_4 \end{bmatrix}^T$. Under hypothesis $H = i$, the observation is given by $Y = (-1)^i \mu + Z$ where

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} \sim \mathcal{N}\left(0, \sigma^2 \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}\right)$$

and $\mu = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$.

For each of the following functions, indicate if it is a sufficient statistic, and briefly explain why. [Hint : Do not apply Fisher-Neyman factorization. Observe instead that $\mathbb{E}[(Z_1 + Z_4)^2] = \mathbb{E}[(Z_2 + Z_3)^2] = 0$]

a. $T_1(Y) = Y_1 + Y_2 + Y_3 + Y_4$

b. $T_2(Y) = Y_1 + Y_2 - \frac{Y_3 + Y_4}{2}$

c. $T_3(Y) = Y_1 - Y_4 + \frac{Y_2 - Y_3}{2}$

d. $T_4(Y) = \langle \begin{bmatrix} p & 1-p & -p & p-1 \end{bmatrix}, Y \rangle$ for some $p \in \mathbb{R}$

Solution 2. It is not possible to apply Fisher-Neyman easily in this case since the distribution of Z is a degenerate Gaussian, indeed Z lies in a 2 dimensional space. Observe that $Z_1 = -Z_4$ and $Z_2 = -Z_3$. So that $T_1(Y) = 4(-1)^H$, $T_2(Y) = (-1)^H + \frac{3}{2}(Z_1 + Z_2)$, $T_3(Y) = 2Z_1 + Z_2$ and $T_4(Y) = Z_1 + Z_2$.

It is pretty clear that T_3 and T_4 pure noise and independent of H and so are not sufficient statistics, and T_1 is a noiseless function of H and so is sufficient. The decision based on any sufficient statistic should lead to the same error probability which in this case is 0 when we use T_1. If we were to decide using T_2 we would have some non zero probability of error since we have a noise. So T_2 is not a sufficient statistic.

PROBLEM 3. Assume that $H \in \{0,1,2,3,4\}$. For each $H = i$, the transmitter transmits codeword μ_i and the receiver observes Y where

$$Y = \mu_i + Z \qquad\qquad Z \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

The codewords are:

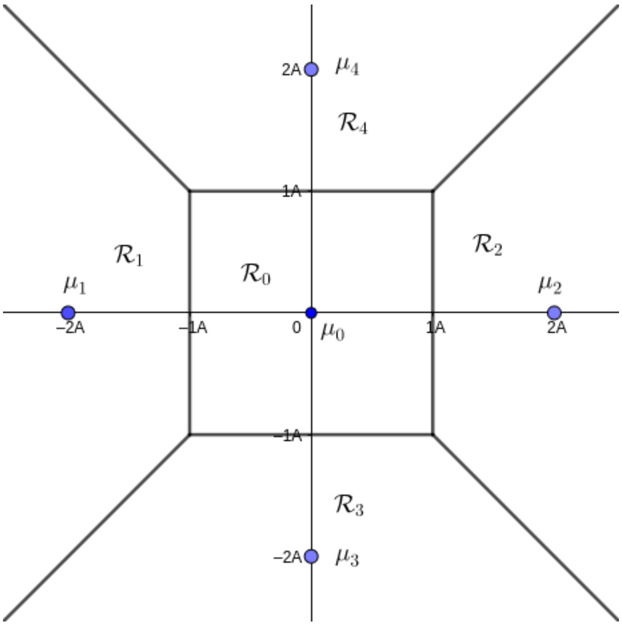
$$\mu_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad -\mu_1 = \mu_2 = \begin{bmatrix} 2A \\ 0 \end{bmatrix} \qquad -\mu_3 = \mu_4 = \begin{bmatrix} 0 \\ 2A \end{bmatrix}.$$

a. Sketch the decision regions assuming that all messages $P_H(i)$ are equally likely.

b. Calculate $P(Error|H = 0)$ for the decision regions you found in part (a).

c. Calculate $P(Error|H = 1)$ for the decision regions you found in part (a).

SOLUTION 3. a. The decision region is shown on the following figure.



b. This correspond to having Z_1 and Z_2 not in $\mathcal{R}_0$:

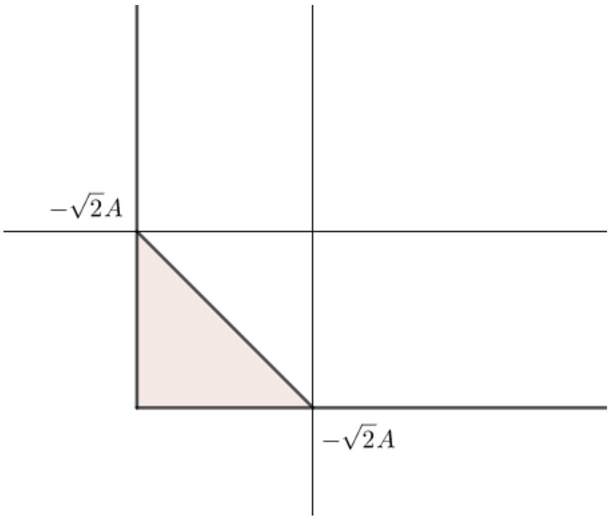
$$\begin{aligned} \mathbb{P}(Error|H = 0) &= 1 - (1 - 2Q(A))^2 \\ &= 4Q(A) - 4Q(A)^2 \end{aligned}$$

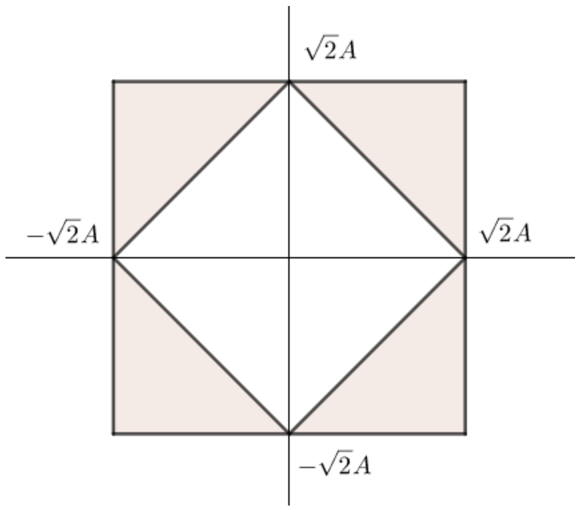
c. This is a bit more tricky, after applying a rotation, it is equivalent to compute the probability that Z is in the upper right part of this plot

Which in turn is $1 - Q(-\sqrt{2}A)^2 + V = 1 - (1 - Q(\sqrt{2}A))^2 + V$ where V is the probability that (Z_1, Z_2) is in the shaded triangle.

In order to compute V one can observe that it appears 4 times on the following figure The probability of the big square is $(1 - 2Q(\sqrt{2}A))^2$ and the probability of the small one is $(1 - 2Q(A))^2$. This means that $4V + (1 - 2Q(A))^2 = (1 - 2Q(\sqrt{2}A))^2$ and so $V = \frac{(1-2Q(\sqrt{2}A))^2-(1-2Q(A))^2}{4}$ and

$$\begin{aligned} \mathbb{P}(Error|H = 1) &= 1 - (1 - Q(\sqrt{2}A))^2 + \frac{(1 - 2Q(\sqrt{2}A))^2 - (1 - 2Q(A))^2}{4} \\ &= Q(\sqrt{2}A) + Q(A) - Q(A)^2 \end{aligned}$$





PROBLEM 4. Consider a transmission scheme with $H \in \{0, 1, \ldots, m-1\}$ uniformly and when $H = i$, the transmitter transmits codeword $\mu_i \in \mathbb{R}^n$. The receiver observes

$$Y \sim \mathcal{N}(\mu_i, C)$$

where C is a positive definite matrix in $\mathbb{R}^{n \times n}$.

The goal of this exercise is to characterise the decoding regions and give an expression for the probability of error of this scheme, in order to achieve that we transform the problem to the one of the AWGN channel that you already know.

a. A positive definite matrix C can always be rewritten as $C = U^T \Sigma U$ where Σ is a diagonal matrix containing positive elements and U is a unitary matrix (hence $UU^T = U^T U = I$). Show that under hypothesis $H = i$, we have $UY \sim \mathcal{N}(U\mu_i, \Sigma)$.

b. Let us write $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$, let $\Sigma^{-\frac{1}{2}} = \begin{bmatrix} \sigma_1^{-1} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^{-1} \end{bmatrix}$. let $\tilde{Y} := \Sigma^{-\frac{1}{2}} UY$. Give the distribution of $\tilde{Y}$.

c. Show that $\left[\langle \tilde{Y}, \Sigma^{-\frac{1}{2}} U\mu_0 \rangle, \ldots, \langle \tilde{Y}, \Sigma^{-\frac{1}{2}} U\mu_{m-1} \rangle\right]^T$ is a sufficient statistic.

d. Give the decoding region in the $\tilde{Y}$ space. Deduce the decoding regions in the Y space.

e. For the special case $m = 2$, give the probability of error as a function of $\sum_i \frac{(\nu_{0,i}-\nu_{1,i})^2}{\sigma_i^2}$ where $\nu_0 = U\mu_0$ and $\nu_1 = U\mu_1$ and $\nu_{k,i}$ denote the i'th coordinate of the vector ν_k.

f. Now, consider if C is not positive definite but positive semi-definite, i.e., some σ_i^2 is equal to 0. Then we cannot construct the matrix $\Sigma^{-\frac{1}{2}}$ since it involves dividing by zero. Give an optimal decision rule to determine H from UY and its probability of error. [Hint: use part (e) to gain intuition.]

Solution 4.

a. Under $H = i$, UY have mean $U\mu_i$ and have variance $UCU^T = UU^T\Sigma UU^T = \Sigma$, furthermore it is a Gaussian vector.

b. Under $H = i$, $\Sigma^{-\frac{1}{2}}UY$ have mean $\Sigma^{-\frac{1}{2}}U\mu_i$ and have variance $\Sigma^{-\frac{1}{2}}UCU^T\Sigma^{-\frac{T}{2}} = \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{T}{2}} = I$, furthermore it is a Gaussian vector so $\Sigma^{-\frac{1}{2}}UY \sim \mathcal{N}(\Sigma^{-\frac{1}{2}}U\mu_i, I)$.

c. There is a one to one function between Y and $\tilde{Y}$ hence $\tilde{Y}$ is a sufficient statistic. If we do estimation only based on $\tilde{Y}$, we know that $\left[\langle \tilde{Y}, \Sigma^{-\frac{1}{2}}U\mu_0\rangle, \ldots, \langle \tilde{Y}, \Sigma^{-\frac{1}{2}}U\mu_{m-1}\rangle\right]^T$ is a sufficient statistic, this can be seen by only considering estimation based on $\tilde{Y}$ and applying usual methods for A.W.G.N. channels. We have transformed our colored noise into a white one, this is called whitening.

d. The regions are obvious for $\tilde{Y}$ since it is an A.W.G.N scenario. To get the regions for Y it is sufficient to pass the regions for $\tilde{Y}$ through the inverse of the linear transformation represented by the matrix $\Sigma^{-\frac{1}{2}}U$, this can be computed to be the matrix $U^T\Sigma^{\frac{1}{2}}$ with

$$\Sigma^{\frac{1}{2}} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix}.$$

e. We can use the whitened signal $\tilde{Y}$ to get $\mathbb{P}(E) = Q\left(\frac{\left\|\Sigma^{-\frac{1}{2}}U\mu_0 - \Sigma^{-\frac{1}{2}}U\mu_1\right\|}{2}\right) = Q\left(\frac{1}{2}\sqrt{\sum_i \frac{(\nu_{0,i}-\nu_{1,i})^2}{\sigma_i^2}}\right)$

f. We iterate over the values of σ_i^2 to reduce the problem to a scenario with full rank matrix. Let $A = \{i : \sigma_i^2 = 0\}$ and let $\{B_j\}$ be a partition of $\{0, \ldots, m-1\}$ such that for $k, \ell \in B_j$ and all $i \in A$ we have $\nu_{k,i} = \nu_{\ell,i}$ and if k and ℓ are in different B_j then their is a $i \in A$ such that $\nu_{k,i} \neq \nu_{\ell,i}$. Let U_A, respectively U_{A^c} be the sub-matrices of U corresponding to rows of indices in A and respectively the rest, let also Σ_{A^c} be the non zero eigenvalues in the same order as the columns of U_A. Then upon receiving observation Y, We are interested in the two statistics U_AY and $\Sigma_{A^c}^{-\frac{1}{2}}U_{A^c}$, the first one is a function of H with no additive noise, it will permit us to decide which B_j we are in and this will have the effect of changing the prior to $P_{H'}(i) = \frac{1\{i \in B_j\}p_H(i)}{\sum_{k \in B_j} p_H(k)}$. The vector $\Sigma_{A^c}^{-\frac{1}{2}}U_{A^c}$ is a standard A.W.G.N. scenario with non singular co-variance matrix, we can apply part d to it.

PROBLEM 5. Consider the set of message $H \in \{0, 1, \ldots, m-1\}$ with prior $\mathbb{P}(H = k) = p_k$. We wish to transmit over an AWGN channel where the noise $N(t)$ have power spectral density $\frac{N_0}{2}$. If $H = k$, the transmitter sends $A\psi_k(t)$, with $A \geq 0$ and $\psi_k(t) = \sqrt{2}\sin\left(2^{k+1}\pi t\right)$ for $t \in [0, 1]$.

a. Determine the optimal receiver structure.

b. Consider another transmission scheme for H. In this scheme, the transmitter sends $A\hat{\psi}_k(t)$, where $\hat{\psi}_k(t) = \mathbb{1}\{t \in [k, k+1)\}$. Determine the optimal receiver structure. Does this transmission scheme has lower error probability compared to the transmission scheme using ψ_k?

c. Back to the transmitter in part a. Consider the following sub-optimal decoder,

$$\hat{H}_\lambda(R) = \begin{cases} i & \langle R, \psi_i \rangle \geq \lambda \text{ and there is no } j \neq i \text{ such that } \langle R, \psi_j \rangle \geq \lambda \\ ? & \text{otherwise} \end{cases}$$

where R is the received signal. Calculate $P\left(\hat{H}_\lambda(R) \neq i \,\middle|\, H = i\right)$.

d. Take $p_k = 1/m$, $A = \sqrt{C \ln m}$, and $\lambda = (1-\epsilon)A$. Show that the following upper bound holds

$$P(\hat{H}_\lambda(R) \neq i | H = i) \leq Q\left(\epsilon\sqrt{\frac{2C\ln m}{N_0}}\right) + (m-1)Q\left((1-\epsilon)\sqrt{\frac{2C\ln m}{N_0}}\right).$$

e. Show that if

$$\frac{(1-\epsilon)^2 C}{N_0} > 1$$

then $P(\hat{H}_\lambda(R) \neq i | H = i) \to 0$ as $m \to \infty$. [Hint : Use the fact that $Q(x) \leq \exp\left(-\frac{x^2}{2}\right)$.

SOLUTION 5. a. To understand the optimal receiver structure, we have to find the optimal decision rule. Before that, we observe that $\psi_k(t)$ form an orthonormal basis

$$\langle \psi_i, \psi_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Hence, the MAP rule corresponds to,

$$\begin{aligned} \hat{H}_{MAP}(R) &= \text{argmax}_{i \in H} f(R|i) p_i \\ &= \text{argmax}_{i \in H} \exp\left(-\frac{(\langle R, \psi_i \rangle - A)^2}{N_0}\right) p_i \\ &= \text{argmax}_{i \in H} - \frac{(\langle R, \psi_i \rangle - A)^2}{N_0} + \log p_i \end{aligned}$$

Notice that we need $\langle R, \psi_i \rangle$ to make the decision, hence the optimal receiver structure is achieved by first compute the inner product $\langle R, \psi_i \rangle$ and the perform the maximization in the third equality to obtain the MAP rule.

b. The most important observation here is that the geometry of transmission scheme in (a) and the transmission scheme in (b) is equivalent. In other words, we have that

$$\langle \psi_i, \psi_j \rangle = \langle \hat{\psi}_i, \hat{\psi}_j \rangle.$$

As the projection AWGN is a multi-variate gaussian, it does not change under the transformation that preserves inner product. Hence, we can conclude that transmission scheme in (a) and in (b) is equivalent, therfore its optimal receiver structure is equivalent (except we take the inner product with respect to $\hat{\psi}_i$ instead of ψ_i) and its error probability is also equal.

c. Let us calculate the correct detection probability,

$$\begin{aligned} P\left(\hat{H}_\lambda(R) = i \,\middle|\, H = i\right) &= P\left(\{\langle R, \psi_i \rangle \geq \lambda\} \bigcap_{j \neq i} \{\langle R, \psi_j \rangle < \lambda\} \,\middle|\, H = i\right) \\ &= P\left(\{\langle R, \psi_i \rangle \geq \lambda\} | H = i\right) \prod_{j \neq i} P\left(\{\langle R, \psi_j \rangle < \lambda\} | H = i\right) \end{aligned}$$

The second equality is justified because $\langle R, \psi_i \rangle$ is independent of $\langle R, \psi_j \rangle$ if $\langle \psi_j, \psi_j \rangle = 0$ under AWGN. Observe that $H = 1$, then $\langle R, \psi_i \rangle \sim \mathcal{N}(A, N_0/2)$ and $\langle R, \psi_j \rangle \sim \mathcal{N}(0, N_0/2)$ for all $j \neq i$, therefore

$$g(A, \lambda) := P\left(\{\langle R, \psi_i \rangle \geq \lambda\} | H = i\right) = \begin{cases} 1 - Q\left(\frac{A - \lambda}{\sqrt{N_0/2}}\right) & A \geq \lambda \\ Q\left(\frac{\lambda - A}{\sqrt{N_0/2}}\right) & A < \lambda \end{cases}$$

and for $j \neq i$

$$h(\lambda) := P\left(\{\langle R, \psi_j \rangle < \lambda\} | H = i\right) = \begin{cases} 1 - Q\left(\frac{\lambda}{\sqrt{N_0/2}}\right) & \lambda \geq 0 \\ Q\left(\frac{-\lambda}{\sqrt{N_0/2}}\right) & \lambda < 0. \end{cases}$$

Therefore we obtained

$$P\left(\hat{H}_\lambda(R) \neq i \,\middle|\, H = i\right) = 1 - g(A, \lambda) h(\lambda)^{m-1}$$

d. This bound is produced by the union bound on error probability. In this case, the error probability is given as follows,

$$\begin{aligned} P\left(Error | H = i\right) &= P\left(\{\langle R, \psi_i \rangle < \lambda\} \bigcup_{j \neq i} \{\langle R, \psi_j \rangle \geq \lambda\} \,\middle|\, H = i\right) \\ &\leq P\left(\{\langle R, \psi_i \rangle < \lambda\} | H = i\right) + \sum_{j \neq i} P\left(\{\langle R, \psi_j \rangle \geq \lambda | H = i\}\right) \\ &= Q\left(\epsilon \sqrt{\frac{2C \ln m}{N_0}}\right) + (m-1) Q\left((1-\epsilon) \sqrt{\frac{2C \ln m}{N_0}}\right). \end{aligned}$$

e. From (d.) we can upper bound using the hint to obtain:

$$P(\hat{H}_\lambda(R) \neq i | H = i) \leq \exp\left(-\epsilon^2 \frac{2C \ln m}{N_0}\right) + \exp\left(-(1-\epsilon)^2 \frac{2C}{N_0} \ln m + \ln(m-1)\right).$$

The first term always goes to 0 as $m \to \infty$ as the exponent goes to $-\infty$. While the second term's exponent also goes to $-\infty$ under the assumption.