

PROBLEM 1. (7 points)

Suppose  $Z = (Z_1, Z_2)$  is uniformly distributed on the unit disc  $\{(x, y) : x^2 + y^2 \leq 1\}$ . In a binary hypothesis problem the observation  $Y$  is given by

$$Y = \begin{cases} Z & \text{if } H = 0, \\ \beta Z & \text{if } H = 1, \end{cases}$$

where  $\beta > 1$  is a known constant. Let  $p_0 = \Pr(H = 0)$  and  $p_1 = 1 - p_0 = \Pr(H = 1)$ .

(a) (3 pts) Find the MAP decision rule  $\hat{H}_{\text{MAP}}(y)$ .

*Hint:* If  $X$  is uniformly distributed on a set  $A \subset \mathbb{R}^2$ , then  $f_X(x) = \frac{1}{\text{Area}(A)} \mathbb{1}\{x \in A\}$ .

*Solution:* We have

$$\begin{aligned} f_{Y|H}(y|0) &= \frac{1}{\pi}, \text{ if } y_1^2 + y_2^2 \leq 1, \\ f_{Y|H}(y|1) &= \frac{1}{\pi\beta^2}, \text{ if } y_1^2 + y_2^2 \leq \beta^2. \end{aligned}$$

Hence, the MAP rule is

– if  $1 < y_1^2 + y_2^2 \leq \beta^2$ , then  $\hat{H}_{\text{MAP}}(y) = 1$ , since

$$\begin{aligned} p_H(0)f_{Y|H}(y|0) &\stackrel{\hat{H}=0}{\geq} p_H(1)f_{Y|H}(y|1) \\ &\iff p_0 \cdot 0 \stackrel{\hat{H}=0}{\geq} p_1 \frac{1}{\pi\beta^2}. \end{aligned}$$

– if  $0 \leq y_1^2 + y_2^2 \leq 1$ , then decide  $\hat{H}_{\text{MAP}}(y) = 0$  if  $\beta^2 \frac{p_0}{p_1} \geq 1$  and 1 else, since

$$\begin{aligned} p_0 \cdot \frac{1}{\pi} &\stackrel{\hat{H}=0}{\geq} p_1 \frac{1}{\pi\beta^2} \\ \frac{p_0}{p_1} &\stackrel{\hat{H}=0}{\geq} \frac{1}{\beta^2}. \end{aligned}$$

(b) (2 pts) Are there values of  $p_0$  for which the MAP rule does not depend on  $y$ ? If so, find them.

*Solution:* In the answer to part (a), we see that the MAP rule is to always decide 1 if  $y_1^2 + y_2^2 > 1$ . Hence, the MAP rule will not depend on  $y$  if it also decides 1 for  $y_1^2 + y_2^2 \leq 1$ , which happens when  $\beta^2 \frac{p_0}{1-p_0} < 1$ , or equivalently,  $p_0 < \frac{1}{1+\beta^2}$ .

(c) (2 pts) Assume  $p_0 = 1/2$ . Find  $\Pr(\text{error}|H = 0)$  and  $\Pr(\text{error}|H = 1)$ .

*Solution:* When  $p_0 = 1/2$ , the MAP rule is to decide  $\hat{H}_{\text{MAP}}(y) = 0$  if  $y_1^2 + y_2^2 \leq 1$  and 1 else. When  $H = 0$ , we necessarily have  $y_1^2 + y_2^2 \leq 1$ , and there is no error, i.e.,  $\Pr(\text{error}|H = 0) = 0$ . When  $H = 1$ , we make an error if the decision is 0, i.e., if we have  $y_1^2 + y_2^2 \leq 1$ . Hence, we have that  $\Pr(\text{error}|H = 1)$  is equal to the probability that  $\beta Z$  lies in the unit disc. Since  $\beta Z$  is uniformly distributed on the disc of radius  $\beta$ , this is equal to  $\frac{1}{\beta^2}$ .

*Remarks:* In this problem,  $Y_1^2 + Y_2^2$  is a sufficient statistic. If the hypotheses are a priori equally likely, then the MAP rule is to decide 0 if  $Y$  lies in the smaller disc and 1 otherwise. In part (b), we see that if our prior belief about  $H$  is sufficiently biased towards 1, then we never decide 0 — the “evidence” is not strong enough to overcome our initial bias towards  $H = 1$ , regardless of the observation  $Y$ .

PROBLEM 2. (12 points)

Suppose  $Z = [Z_1, Z_2, Z_3]^T \sim \mathcal{N}(0, K)$ , with

$$K = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

(a) (2 pts) Show that  $Z_1 + Z_2 + Z_3 = 0$  with probability 1.

*Hint:*  $E[X^2] = 0$  implies that  $X = 0$  with probability 1.

*Solution:* Guided by the hint, we compute

$$\begin{aligned} \mathbb{E}[(Z_1^2 + Z_2^2 + Z_3^2)] &= \mathbb{E}[Z_1^2 + Z_2^2 + Z_3^2 + 2Z_1Z_2 + 2Z_2Z_3 + 2Z_3Z_1] \\ &= 2 + 2 + 2 + 2(-1) + 2(-1) + 2(-1) = 0. \end{aligned}$$

Hence, we have that  $Z_1 + Z_2 + Z_3 = 0$  with probability 1.

(b) (3 pts) Let  $A = \begin{bmatrix} 1 & 0 & 0 \\ 1/\sqrt{3} & 2/\sqrt{3} & 0 \\ 1 & 1 & 1 \end{bmatrix}$ , and let  $U = AZ$ . What is the covariance matrix of  $U$ ?

*Hint:* Show that  $U_1$  and  $U_2$  are independent, and use (a).

*Solution:* First note that  $U$  is a zero mean Gaussian vector. Also observe that  $U_3 = Z_1 + Z_2 + Z_3$ , which, by part (a), is equal to 0 with probability 1. Finally, we also have that  $U_1$  and  $U_2$  are independent, since  $\mathbb{E}[U_1U_2] = \mathbb{E}\left[Z_1 \frac{1}{\sqrt{3}}(Z_1 + 2Z_2)\right] = \frac{1}{\sqrt{3}}[2 + 2(-1)] = 0$ . To complete the covariance, all we require is the variance of  $U_2$ , which is given by  $\frac{1}{3}\mathbb{E}[(Z_1 + 2Z_2)^2] = \frac{1}{3}[2 + 4(2) + 4(-1)] = 2$ . Hence, the covariance matrix of  $U$  is given by

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let  $c_1 = [1, 2, 3]^T$  and  $c_2 = [5, 1, 0]^T$  be the codewords of a communication system with two equally likely messages, and suppose  $Y = c_i + Z$  (with  $Z$  as above) be the receiver’s observation if message  $i$  is sent.

- (c) (2 pts) Let  $\tilde{Y} = AY$ . (Note that  $A$  is an invertible matrix, so  $\tilde{Y}$  is equivalent to  $Y$ .) Show that  $(\tilde{Y}_1, \tilde{Y}_2)$  is a sufficient statistic.

*Solution:* We have  $\tilde{Y} = Ac_i + AZ = Ac_i + U$ . Define  $\tilde{c}_i = Ac_i$ , then we have  $\tilde{c}_1 = [1, 5/\sqrt{3}, 6]^T$  and  $\tilde{c}_2 = [5, 7/\sqrt{3}, 6]^T$ . Also note that, since  $U_3$  is 0 mean and 0 variance it is 0, with probability 1. Hence,  $\tilde{Y}_3 = \tilde{c}_{i,3} + U_3 = 6$ , irrespective of  $i$ , which implies that it is irrelevant and  $(\tilde{Y}_1, \tilde{Y}_2)$  is a sufficient statistic.

- (d) (3 pts) Find the probability of error of the MAP decision rule for the communication system above.

*Solution:* The error probability is simply given by  $Q\left(\frac{d}{2\sigma}\right)$ , where  $\sigma^2 = \text{Var}(U_1) = \text{Var}(U_2) = 2$  and  $d = \|\tilde{c}_1 - \tilde{c}_2\| = \sqrt{\frac{52}{3}}$ .

Suppose we replace  $c_1$  and  $c_2$  above with  $c_1 = [0, 0, 1]^T$  and  $c_2 = [0, 0, -1]^T$ . The observation  $Y$  is still  $c_i + Z$ , and  $\tilde{Y} = AY$ .

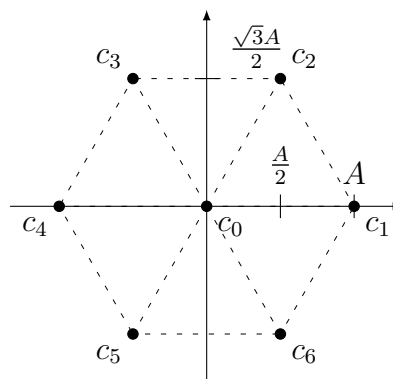
- (e) (2 pts) What is the probability of error of the MAP decision rule for this new system? Is  $(\tilde{Y}_1, \tilde{Y}_2)$  still a sufficient statistic? (Explain).

*Solution:* In this new system, observe that  $\tilde{c}_1 = [0, 0, 1]^T$  and  $\tilde{c}_2 = [0, 0, -1]^T$ . Hence,  $\tilde{Y}_3 = 1$  when  $i = 1$  and  $-1$  when  $i = 2$  (since  $U_3 = 0$ ). Thus, the error probability is 0. Further,  $(\tilde{Y}_1, \tilde{Y}_2)$  is no longer a sufficient statistic, since we cannot get zero error probability without looking at  $\tilde{Y}_3$ . In fact, in this problem, we actually have that not only is  $(\tilde{Y}_1, \tilde{Y}_2)$  not a sufficient statistic, but it is also irrelevant.

*Remarks:* Though  $Z$  is a vector in  $\mathbb{R}^3$ , it only has dimension 2, since it lives in a two-dimensional plane given by  $Z_1 + Z_2 + Z_3 = 0$ . This suggests that there is a direction (perpendicular to the plane) where there is “no noise”. In part (e), we exploit this by choosing codewords that differ along this noiseless direction, allowing us to achieve zero error probability. In part (c) and (d), however, our choice of codewords was suboptimal — by choosing codewords that belonged to the same two-dimensional subspace as the noise, we could not exploit the existence of a noiseless dimension.

### PROBLEM 3. (11 points)

Consider the constellation with seven codewords  $\{c_i\}_{i=0}^6$  as given in the diagram. Assume that the seven messages are equally likely, and let  $e_{\text{hex}}(A)$  be the error probability of the MAP decoder that observes  $Y = c_i + Z$ , where  $Z \sim \mathcal{N}(0, I_2)$ .

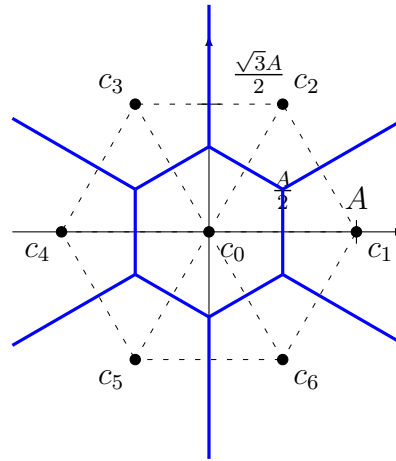


- (a) (3 pts) Show that  $e_{\text{hex}}(A)$  is upper bounded by  $\Pr(Z_1^2 + Z_2^2 \geq A^2/4)$ .

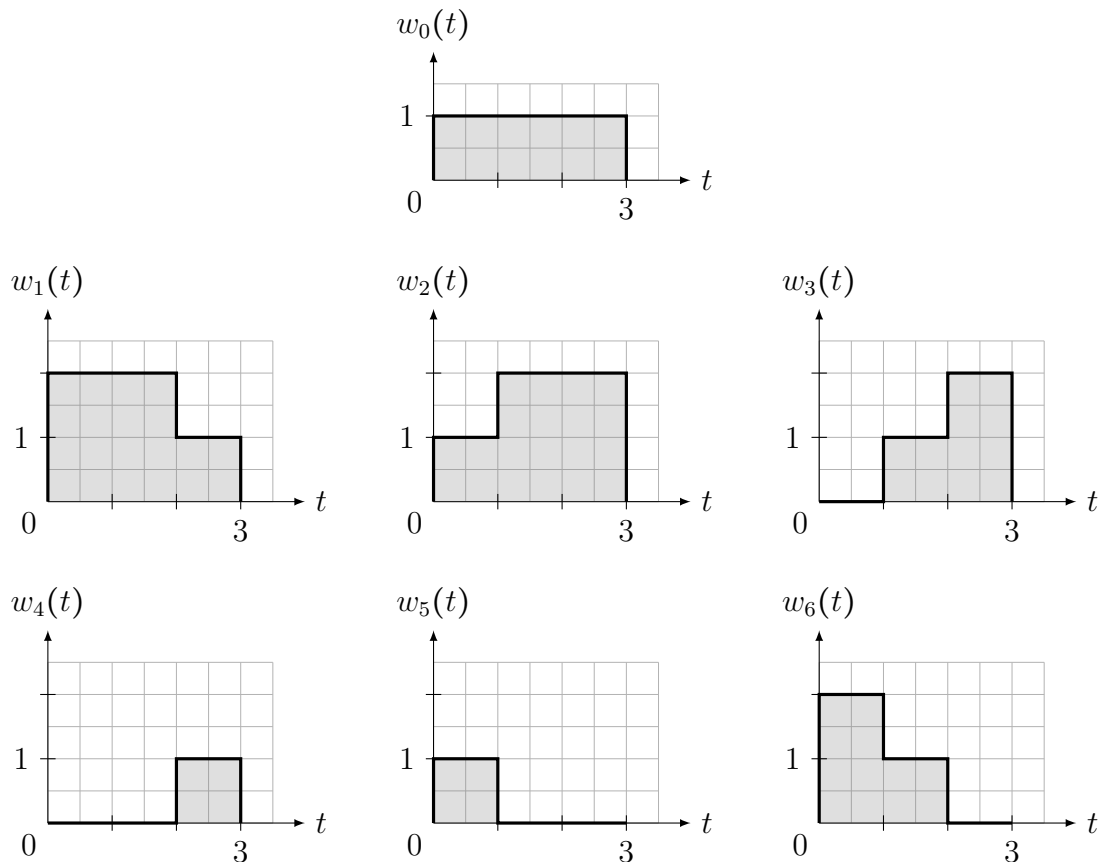
*Hint:* You may find it helpful to draw the decision regions.

*Solution:*

The MAP decision regions are marked in the figure to the right. Observe that a correct decision is necessarily made when  $(Z_1, Z_2)$  lies in the circle of radius  $\frac{A}{2}$  centered at the origin. Hence, the probability of being correct is at least  $\Pr(Z_1^2 + Z_2^2 < A^2/4)$ , which implies that the probability of error is at most  $\Pr(Z_1^2 + Z_2^2 \geq A^2/4)$ .

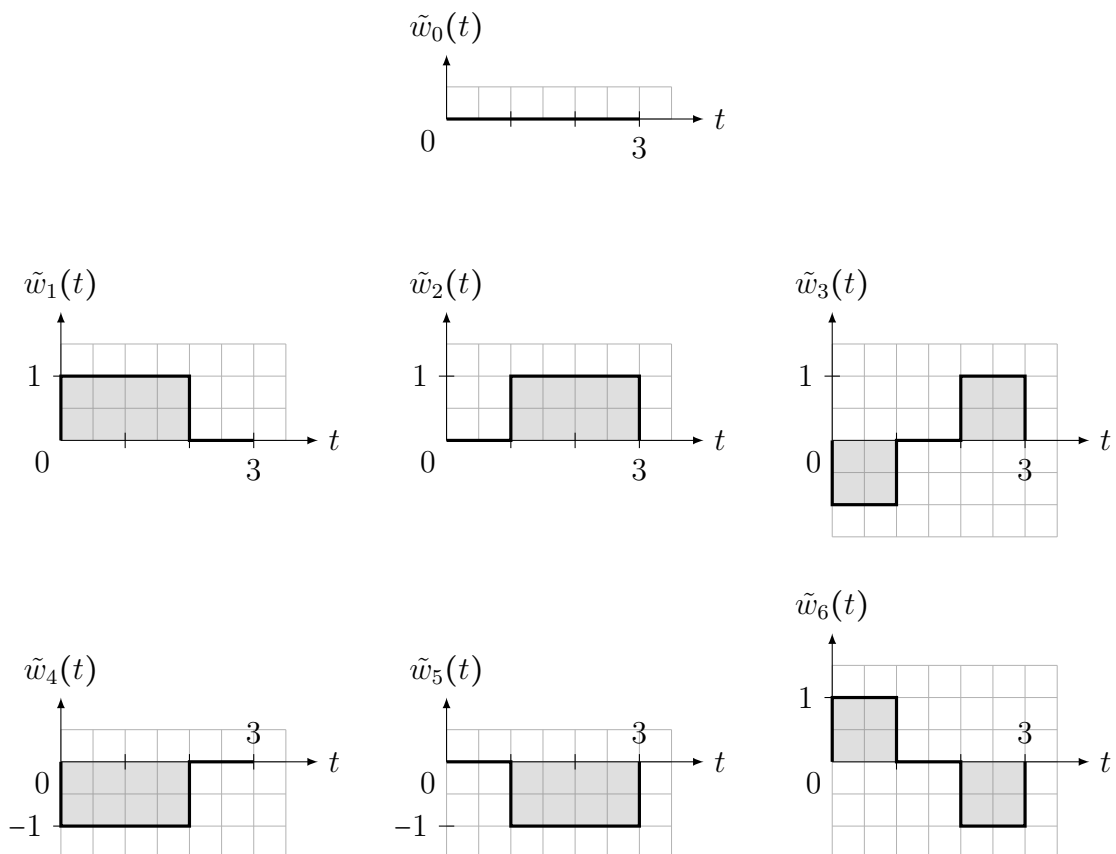


Now consider the waveforms  $\{w_i\}_{i=0}^6$  as shown below.



- (b) (2 pts) Assume that all messages are equally likely. Find a translation of this waveform set to minimize the average energy. Let the new waveforms be  $\{\tilde{w}_i\}_{i=0}^6$ .

*Solution:* To translate the waveforms to obtain the minimum average energy, we subtract the average of the waveforms from each. Note that the average is given exactly by  $w_0$ , hence the new waveforms are as given below.



- (c) (2 pts) Show that  $\tilde{w}_1 + \tilde{w}_4 = \tilde{w}_2 + \tilde{w}_5 = \tilde{w}_3 + \tilde{w}_6 = 0$ , and that  $\|\tilde{w}_1\| = \|\tilde{w}_2\| = \|\tilde{w}_3\|$ .

*Solution:* We see clearly from the figure that  $\tilde{w}_1 + \tilde{w}_4 = \tilde{w}_2 + \tilde{w}_5 = \tilde{w}_3 + \tilde{w}_6 = 0$ . In addition, we also have  $\|\tilde{w}_1\| = \|\tilde{w}_2\| = \|\tilde{w}_3\| = \sqrt{2}$ .

- (d) (2 pts) Find the inner products  $\langle \tilde{w}_1, \tilde{w}_2 \rangle$ ,  $\langle \tilde{w}_1, \tilde{w}_3 \rangle$ , and  $\langle \tilde{w}_2, \tilde{w}_3 \rangle$ .

*Solution:* By a direct computation, we obtain  $\langle \tilde{w}_1, \tilde{w}_2 \rangle = 1$ ,  $\langle \tilde{w}_1, \tilde{w}_3 \rangle = -1$ , and  $\langle \tilde{w}_2, \tilde{w}_3 \rangle = 1$ .

- (e) (2 pts) Consider a communication system which uses the waveforms  $\{w_i\}_{i=0}^6$  to communicate over a white Gaussian noise channel with intensity  $\frac{N_0}{2}$ . Express the optimal error probability of this system *in terms of*  $e_{\text{hex}}(\cdot)$ .

*Hint:* No lengthy computations needed.

*Solution:* First, observe that since  $\{w_i\}_{i=0}^6$  and  $\{\tilde{w}_i\}_{i=0}^6$  are simply translations of each other, the error probability of the system that uses  $\{w_i\}_{i=0}^6$  will be the same as that of a system that uses  $\{\tilde{w}_i\}_{i=0}^6$ . Also observe that codewords  $\{c_i\}_{i=0}^6$  in the hexagonal constellation satisfy (i)  $c_0 = c_1 + c_4 = c_2 + c_5 = c_3 + c_6 = 0$ ,  $\|c_1\|^2 = \|c_2\|^2 = \|c_3\|^2$ , and  $\langle c_1, c_2 \rangle = \langle c_2, c_3 \rangle = -\langle c_1, c_3 \rangle = \|c_1\|^2/2$ . We thus see that  $\langle c_i, c_k \rangle = (A^2/2)\langle \tilde{w}_i, \tilde{w}_k \rangle$ . Hence, there is an orthonormal basis  $\{\psi_1, \psi_2\}$  such that the waveforms  $\{\tilde{w}_i\}_{i=0}^6$  are generated by the codewords  $\{c_i\}_{i=0}^6$  using this basis, with scaling  $\sqrt{2}/A$ . Let  $R = \tilde{w}_i + N$  be the received signal, then computing the sufficient statistic  $Y = (\langle R, \psi_1 \rangle, \langle R, \psi_2 \rangle)$ , we see that  $Y = \frac{\sqrt{2}}{A}c_i + Z$ , where  $Z \sim \mathcal{N}(0, \frac{N_0}{2}I_2)$ . Hence, defining  $\tilde{Y} = \sqrt{\frac{2}{N_0}}Y$ , we have  $\tilde{Y} = \frac{2}{A\sqrt{N_0}}c_i + Z$ , where  $Z \sim \mathcal{N}(0, I_2)$ . Thus, the error probability of this system is the same as that using such a hexagonal constellation with  $A = \frac{2}{\sqrt{N_0}}$ , in terms of  $e_{\text{hex}}$ , is given by  $e_{\text{hex}}\left(\frac{2}{\sqrt{N_0}}\right)$ .

*Remarks:* This is a variant of the triangular and tetrahedral constellations that have been considered in previous midterms.

PROBLEM 4. (13 points)

Let  $e_1, \dots, e_n$  denote the standard basis for  $\mathbb{R}^n$ , i.e.,  $e_1 = [1, 0, \dots, 0]^T$ ,  $e_2 = [0, 1, 0, \dots, 0]^T$ ,  $\dots$ ,  $e_n = [0, \dots, 0, 1]^T$ .

We have a communication system with  $m = 2n$  codewords  $c_1, \dots, c_{2n}$  in  $\mathbb{R}^n$ , with  $c_i = Ae_i$  for  $i = 1, \dots, n$ , and  $c_i = -c_{i-n}$  for  $i = n + 1, \dots, 2n$ . Here  $A > 0$  is a positive constant.

The  $m$  messages are equally likely, and the receiver's observation is given by  $Y = c_i + Z$  if message  $i$  is sent, where  $Z \sim \mathcal{N}(0, \sigma^2 I_n)$ .

- (a) (2 pts) Find the average energy  $\mathcal{E}$  and the average energy per bit  $\mathcal{E}_b$  of the signal constellation above.

*Solution:* For every  $i$ ,  $\|c_i\|_2^2 = A^2$ . Therefore  $\mathcal{E} = \frac{1}{2n} \sum_{i=1}^{2n} \|c_i\|_2^2 = A^2$ . Consequently,  $\mathcal{E}_b = A^2 / \log_2(2n)$ .

- (b) (2 pts) Consider the decision method that first computes  $i_0 = \arg \max_{i=1, \dots, n} |y_i|$ , and sets  $\hat{H}(y) = \begin{cases} i_0 & \text{if } y_{i_0} > 0, \\ n + i_0 & \text{else.} \end{cases}$

Is this rule optimal? (Explain your answer.)

*Solution:* Let us first focus on the decision boundary between  $c_1, c_2$ . Since messages are equally likely we need to look at the likelihood ratio and see when it is 0. That is,

$$\begin{aligned} \log \frac{f_{Y|H}(x^n|1)}{f_{Y|H}(x^n|2)} &= \log \frac{\frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{(x_1-A)^2}{2\sigma^2}} e^{-\frac{x_2^2}{2\sigma^2}} \dots e^{-\frac{x_n^2}{2\sigma^2}}}{\frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{(x_1)^2}{2\sigma^2}} e^{-\frac{(x_2-A)^2}{2\sigma^2}} \dots e^{-\frac{x_n^2}{2\sigma^2}}} \\ &= \log \frac{e^{-\frac{(x_1-A)^2}{2\sigma^2}} e^{-\frac{x_2^2}{2\sigma^2}}}{e^{-\frac{(x_1)^2}{2\sigma^2}} e^{-\frac{(x_2-A)^2}{2\sigma^2}}} \\ &= \frac{A(x_1 - x_2)}{\sigma^2}. \end{aligned}$$

Since  $A > 0$ , we choose  $c_1$  over  $c_2$  whenever  $Y_1 > Y_2$ . Similarly if we look at the decision boundary between  $c_1$  and  $c_{n+1}$ , we see that, we prefer  $c_1$  over  $c_{n+1}$  whenever  $Y_1 > 0$ , this is because,

$$\begin{aligned} \log \frac{f_{Y|H}(x^n|1)}{f_{Y|H}(x^n|n+1)} &= \log \frac{e^{-\frac{(x_1-A)^2}{2\sigma^2}}}{e^{-\frac{(x_1+A)^2}{2\sigma^2}}} \\ &= \frac{2Ax_1}{\sigma^2}. \end{aligned}$$

We can generalize this idea to any likelihood ratio. Therefore, the MAP decoder will find the index  $i$  such that  $|Y_i|$  is the largest, and choose the message  $i$  if  $Y_i > 0$  and  $n + i$ , otherwise.

- (c) (2 pts) Upper bound the probability of error using the union bound.

*Hint:* For each codeword  $c_i$ , the codeword  $-c_i$  is at distance  $2A$  from it; what is the distance between  $c_i$  and the other  $2n - 2$  codewords?

*Solution:* Because the noise is white Gaussian, we are essentially doing minimum distance decoding.

$$\begin{aligned}\mathbb{P}(\text{error}|H = i) &= \mathbb{P}(\exists j \neq i \ \|Y - c_j\|_2 \leq \|Y - c_i\|_2 \mid H = i) \\ &\leq \sum_{j \neq i} \mathbb{P}(\|Y - c_j\|_2 \leq \|Y - c_i\|_2 \mid H = i)\end{aligned}$$

Note that, for every  $j$ , the term  $\mathbb{P}(\|Y - c_j\|_2 \leq \|Y - c_i\|_2 \mid H = i)$  is equal to  $Q\left(\frac{\|c_i - c_j\|_2}{2\sigma}\right)$ . In our case there is one  $j$  such that  $\|c_i - c_j\|_2 = 2A$  and  $2n - 2$   $j$ 's such that  $\|c_j - c_i\|_2 = \sqrt{2}A$ . Therefore,

$$\mathbb{P}(\text{error}|H = i) \leq Q\left(\frac{A}{\sigma}\right) + (2n - 2)Q\left(\frac{A}{\sqrt{2}\sigma}\right).$$

Since this value does not depend on  $i$ , we have,

$$\mathbb{P}(\text{error}) \leq Q\left(\frac{A}{\sigma}\right) + (2n - 2)Q\left(\frac{A}{\sqrt{2}\sigma}\right).$$

- (d) (2 pts) Show that if  $\mathcal{E}_b/\sigma^2 > 4 \ln 2$ , then the probability of error of this communication system approaches zero as  $n$  gets large.

*Hint:* Use (c), that  $Q(\sqrt{x}) \leq \exp(-x/2)$ , and note that  $(m - 1) \exp(-\alpha \log_2(m))$  tends to zero as  $m$  gets large if  $\alpha > \ln 2$ .

*Solution:*

Using part (c), we have,

$$\begin{aligned}\mathbb{P}(\text{error}) &\leq Q\left(\frac{A}{\sigma}\right) + (2n - 2)Q\left(\frac{A}{\sqrt{2}\sigma}\right) \\ &\leq e^{-\frac{A^2}{2\sigma^2}} + (2n - 2)e^{-\frac{A^2}{4\sigma^2}}\end{aligned}\tag{*}$$

The question suggests that if the energy per bit is above a certain threshold, the error will decay to 0. Note that the energy per bit,  $\mathcal{E}_b = A^2/\log_2(2n)$ . Therefore, if the energy per bit is above a certain threshold, then  $A$  will grow to infinity as  $n$  gets larger and larger. Therefore, the first term in (\*) will decay to 0 anyways.

The second term is

$$\begin{aligned}(2n - 2)e^{-\frac{A^2}{4\sigma^2}} &= e^{\ln(2n-2)} e^{-\frac{A^2}{4\sigma^2}} \\ &= e^{\ln(2n-2)} e^{-\frac{\mathcal{E}_b \log_2(2n)}{4\sigma^2}} \\ &= e^{(\ln 2)(1+\log_2(n-1))} e^{-\frac{\mathcal{E}_b(1+\log_2(n))}{4\sigma^2}}\end{aligned}$$

Note that, the first multiplicative term grows exponentially with  $\log_2 n$  and the second decays exponentially with  $\log_2 n$ . The multiplication overall will decay if the decaying one dominates. That is, if

$$\frac{\mathcal{E}_b}{4\sigma^2} > \ln 2,$$

the error will decay to 0.

(e) (2 pts) Show that for  $i = 1, \dots, n$ ,

$$\ln \frac{p_{H|Y}(i|y)}{p_{H|Y}(n+i|y)} = 2Ay_i/\sigma^2.$$

*Solution:* This is shown already in part (b), by using Bayes' rule and the fact that the messages are equally likely.

(f) (3 pts) Suppose  $T = t(Y)$  where  $t(\cdot)$  is a deterministic function. Suppose that there exist  $y$  and  $\tilde{y}$ , with  $y \neq \tilde{y}$  and  $t(y) = t(\tilde{y})$ . Can  $T$  be a sufficient statistic? Explain.

*Hint:* Use (e).

*Solution:*  $T = t(Y)$  is a sufficient statistic if  $P_{H|T(Y)}(h|t) = P_{H|T(Y),Y}(h|t, y)$  for every  $h, t$  and  $y$  such that  $t(y) = t$ . Since  $t(\cdot)$  is a deterministic function,  $P_{H|T(Y),Y}(h|t, y) = P_{H|Y}(h|y)$  for all such  $y$ . That is, for every  $h$  and  $t$ ,  $P_{H|T(Y)}(h|t) = P_{H|Y}(h|y)$  for all  $y$  such that  $t(y) = t$ .

Suppose now that  $H$  and  $Y$  are sampled and you are given the value of  $t(Y) = T$ . From this you can calculate  $P_{H|T(Y)}(\cdot|T)$ , if  $t(Y)$  is a sufficient statistic, as we have shown, you also know  $P_{H|Y}$  for every  $Y$  that could have been sampled. Using part e we see that for every  $y$  that could have been sampled, you can calculate the quantity,

$$\ln \frac{P_{H|Y}(i|y)}{P_{H|Y}(n+i|y)} = 2Ay_i/\sigma^2,$$

and you can do this for every  $i$ . Therefore, knowing the value of  $T$  you know the value of  $Y_i$  for every  $i$ . That is  $Y$  must be a function of  $T$ . We have reached this conclusion assuming that  $T = t(Y)$  is a sufficient statistic. However, this will lead to a contradiction with the assumption of the question that two different values of  $y$  is mapped to the same  $t$ . Therefore,  $T = t(Y)$  with the given property cannot be a sufficient statistic.

*Remarks:* The computation in part (d) shows that if the signal-to-noise ratio (energy per bit to noise variance) is large enough, then the error probability decays to zero as  $n$  gets large. In part (f), we show that no  $T = t(Y)$  with  $t(\cdot)$  that is not injective can be a sufficient statistic. Of course,  $Y$  itself is always a sufficient statistic. This question shows something interesting, that  $Y$  itself is a *minimal sufficient statistic* here. That is, there is no sufficient statistic which is a function of  $Y$  but does not allow you to recover  $Y$ . [Compare this against other sufficient statistics that you have seen before — is it always possible to recover  $Y$  from  $T$ ?]