Exam (January 11, 2016) Solution

Last modified January 21, 2016

Problem 1 - Entropy, Source Coding, and Channel Capacity [15 points]

Part I

The entropy of X is

$$H(X) = \sum_{i=1}^{16} p_i \log_2 \left(\frac{1}{p_i}\right) = \log_2(16) = 4 \text{ bits}.$$

The entropy of Y is

$$\begin{aligned} H(Y) &= \sum_{k=1}^{\infty} 2^{-k} \log_2(2^k) \\ &= \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k \\ &= \sum_{k=0}^{\infty} k \left(\frac{1}{2}\right)^k \\ &= \frac{0.5}{(1-0.5)^2} \\ &= 2 \text{ bits}. \end{aligned}$$

Since X and Y are independent, the joint entropy of X and Y is

$$H(X,Y) = H(X) + H(Y) = 2 + 4 = 6 \text{ bits}.$$

Part II

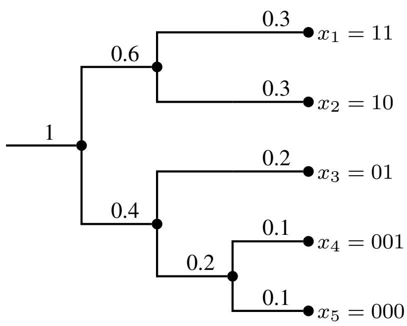
1. The source entropy is

$$H(X) = \sum_{i=1}^{5} p_i \log_2 \left(\frac{1}{p_i}\right) = 2.171 \text{ bits}.$$

2. One possibility is

symbol	codeword	probability
x_1	11	0.3
x_2	10	0.3
x_3	01	0.2
x_4	001	0.1
x_5	000	0.1

where the corresponding tree is depicted below.



3. The average codeword length is

$$\bar{L} = 2 \cdot 0.3 \cdot 2 + 0.2 \cdot 2 + 2 \cdot 0.1 \cdot 3 = 2.2 \text{ bits}.$$

Therefore, the efficiency is

$$\eta = \frac{H(P)}{\bar{L}} = \frac{2.171}{2.2} = 98.7\%.$$

4. The average codeword length is equal to the entropy of P' if and only if the codeword lengths satisfy

$$l_i = \log_2\left(\frac{1}{p_i'}\right) \ \forall i.$$

Solving for p_i' gives

$$p_i' = \frac{1}{2^{l_i}} = 2^{-l_i}.$$

Using the codeword lengths from question 2, P' becomes

$$P' = (2^{-2}, 2^{-2}, 2^{-2}, 2^{-3}, 2^{-3}) = (0.25, 0.25, 0.25, 0.125, 0.125).$$

Part III

1. The entropy of the source is

$$H(X) = -0.5\log_2(0.5) - 0.5\log_2(0.5) = 1.$$

The probability distribution of the output is

$$\begin{aligned}
P(Y=0) &= P(Y=0|X=0)P(X=0) + P(Y=0|X=1)P(X=1)\\
&= 0.5(1-\varepsilon) + 0.5\varepsilon\\
&= 0.5,\\
P(Y=2) &= P(Y=2|X=0)P(X=0) + P(Y=2|X=1)P(X=1)\\
&= 0.5\varepsilon + 0.5(1-\varepsilon)\\
&= 0.5.
\end{aligned}$$

Therefore, the entropy of the output distribution is

$$H(Y) = -P(Y=0)\log_2(P(Y=0)) - P(Y=2)\log_2(P(Y=2)) = 1.$$

2. The product rule gives

$$P(x,y) = P(y|x)P(x).$$

Therefore, the joint probability distribution of the source and the output is

$$\begin{aligned}
P(X=0,Y=0) &= P(Y=0|X=0)P(X=0) = 0.5(1-\varepsilon)\\
P(X=1,Y=0) &= P(Y=0|X=1)P(X=1) = 0.5\varepsilon\\
P(X=0,Y=2) &= P(Y=2|X=0)P(X=0) = 0.5\varepsilon\\
P(X=1,Y=2) &= P(Y=2|X=1)P(X=1) = 0.5(1-\varepsilon).
\end{aligned}$$

The joint entropy is then computed as

$$\begin{aligned}
H(X,Y) = &-P(X=0,Y=0)\log_2(P(X=0,Y=0)) - P(X=0,Y=2)\log_2(P(X=0,Y=2))\\
&- P(X=1,Y=0)\log_2(P(X=1,Y=0)) - P(X=1,Y=2)\log_2(P(X=1,Y=2))\\
= &-0.5(1-\varepsilon)\log_2(0.5(1-\varepsilon)) - 0.5\varepsilon\log_2(0.5\varepsilon) - 0.5\varepsilon\log_2(0.5\varepsilon) - 0.5(1-\varepsilon)\log_2(0.5(1-\varepsilon))\\
= &-(1-\varepsilon)\log_2(1-\varepsilon) - (1-\varepsilon)\log_2(0.5) - \varepsilon\log_2(\varepsilon) - \varepsilon\log_2(0.5)\\
= &\ 1 + H_b(\varepsilon).
\end{aligned}$$

3. The joint entropy, $H(X,Y)$, can be rewritten as

$$H(X,Y) = H(X|Y) + H(X).$$

Therefore, the mutual information can be computed, using $H(X)$, $H(Y)$, and $H(X,Y)$, as

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \\ &= 1 + 1 - 1 - H_b(\varepsilon) \\ &= 1 - H_b(\varepsilon). \end{aligned}$$

4. For $\varepsilon = 0$ or $\varepsilon = 1$, $H_b(\varepsilon) = 0$. In that case, $I(X;Y)$ is maximal and the mutual information is 1 bit.

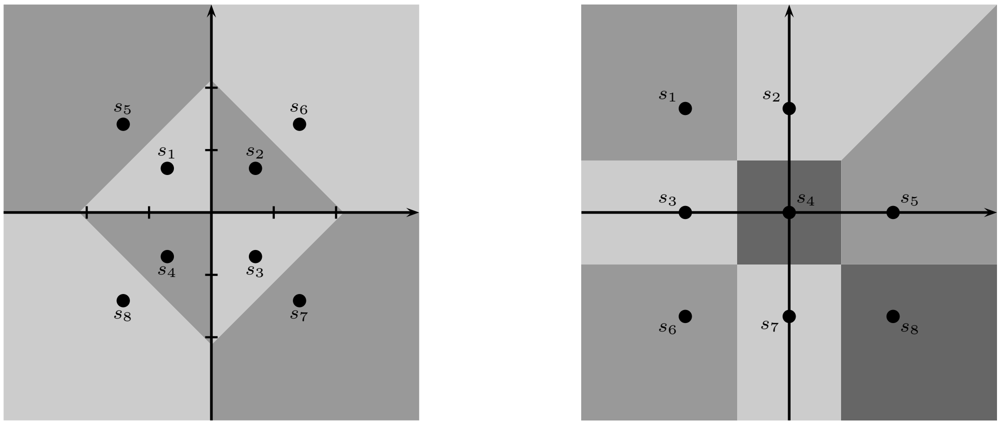
5. For $\varepsilon = 0.5$, $H_b(\varepsilon) = 1$. In that case, $I(X;Y)$ is minimal and the mutual information is 0.

Problem 2 - Signal Constellations and Detection [15 points]

1. For Ω_1, we have 4 points with energy a^2 and 4 points with energy $(2a)^2 = 4a^2$. The average comes out to $E_s = 20a^2/8$ and therefore $a = \sqrt{2/5}$.

   For Ω_2, we have 4 points with energy b^2, 3 points with energy $b^2 + b^2 = 2b^2$, and 1 points with energy 0. The average comes out to $E_s = 10b^2/8$ and therefore $b = \sqrt{4/5}$.

2. The maximum likelihood decision regions are depicted below.



(a) Ω_1 *(b)* Ω_2

3. In general, the nearest neighbor approximation is given by

$$P_s(e) \approx A_d Q\left(\sqrt{\frac{d_{E,\min}^2}{2N_0}}\right),$$

   where A_d is the average number of points at minimum distance $d_{E,\min}$.

   For Ω_1, we have $d_{E,\min} = a = \sqrt{2/5}$ and $A_d = 1$. Therefore

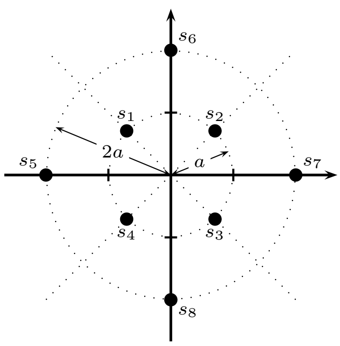
$$P_s(e) \approx Q\left(\sqrt{\frac{1}{5N_0}}\right).$$

   For Ω_2, we have $d_{E,\min} = b = \sqrt{4/5}$ and $A_d = (5 \cdot 2 + 2 \cdot 3 + 4)/8 = 5/2$. Therefore

$$P_s(e) \approx \frac{5}{2} Q\left(\sqrt{\frac{2}{5N_0}}\right).$$

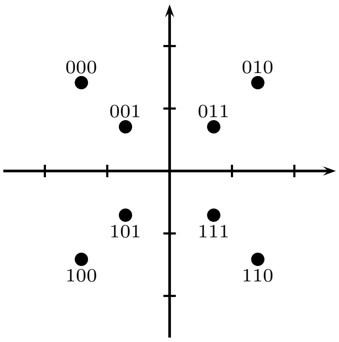
4. Ω_2 performs better because it has a larger minimum distance than Ω_1.

5. The constellation Ω_3 is depicted below.



Ω_3 has $d_{E,\min} = \sqrt{2}a$, which is larger than the minimum distance of Ω_1. Therefore Ω_3 is more power efficient than Ω_1.

6. For Ω_1, a Gray mapping can be found and one example would be



For Ω_2, no Gray mapping is possible, since s_4 has 4 neighbors at minimum distance.

7. Since θ is uniformly distributed between $-\pi/6$ and $\pi/6$, the phase noise will cause each point in 8-PSK to cross over to the ML decision regions of its neighbors, with nonzero probability. Hence, the symbol error probability will always be nonzero, regardless of the signal-to-noise ratio.

   For Ω_1, the angular spacing between the points is $\pi/2$, and therefore the phase noise will not cause the points to cross over to the ML decision regions of their neighbors. Thus, the symbol error probability will go to zero as the signal-to-noise ratio goes to infinity.

8. The maximum likelihood decision rule can be simplified as

$$\begin{aligned}
\operatorname*{argmax}_{s\in\mathcal{X}} p(r|s) &= \operatorname*{argmax}_{s\in\mathcal{X}} \log p(r|s) \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} -|r-s|^2 \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} \left(-|r|^2 + 2\mathrm{Re}\{rs^*\} - |s|^2\right) \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} \mathrm{Re}\{rs^*\} && (1) \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} \mathrm{Re}\{|r|\exp(j\arg\{r\})|s|\exp(-j\arg\{s\})\} \\
&= \operatorname*{argmax}_{s\in\mathcal{X}} \cos(\arg\{r\} - \arg\{s\}) && (2)
\end{aligned}$$

where (1) and (2) follow from the fact that $|r|$ is constant with respect to maximization over s, and $|s|$ is constant because PSK constellations have constant amplitude.

Problem 3 - Linear Block Codes [15 points]

1. In order to find a systematic parity-check matrix, we are allowed to apply elementary row operations to the matrix $\boldsymbol{H}$. With this, we find that

$$\mathbf{H}_s = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

and the corresponding generator matrix is given by

$$\mathbf{G}_s = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

2. The codewords can be obtained through $\boldsymbol{x} = \boldsymbol{u}\mathbf{G}_s$, where $\boldsymbol{u}$ are all possible 3-bit information words. Thus:

$\boldsymbol{u}$	$\boldsymbol{x}$
000	000000
001	101001
010	110010
011	011011
100	011100
101	110101
110	101110
111	000111

The weight spectrum is $A_0 = 1$, $A_3 = 4$, $A_4 = 3$, and $A_1 = A_2 = A_5 = A_6 = 0$.

3. The code parameters are $(6, 3, 3)$. The code rate is $R_c = 1/2$.

4. The syndrome tables based on $\boldsymbol{H}$ and $\boldsymbol{H}_s$ are given by

syndrome	error vector
000	000000
001	001000
010	000001
011	100000
100	000010
101	001010
110	000100
111	010000

and

syndrome	error vector
000	000000
001	001000
010	010000
011	000100
100	100000
101	000001
110	000010
111	100100

respectively. The 6th entry in the left table and the last entry in the right table are not unique.

5. The syndromes are given by (110) for $\boldsymbol{H}$ and (011) for $\boldsymbol{H}_s$. Using the above syndrome tables, the estimated error vector in both cases is (000100) and the ML codeword is thus (011100).

6. The generator matrix for the dual code $\mathcal{C}_\perp$ is given by the parity-check matrix of the code $\mathcal{C}$. With this, we find that the dual code comprises the codewords

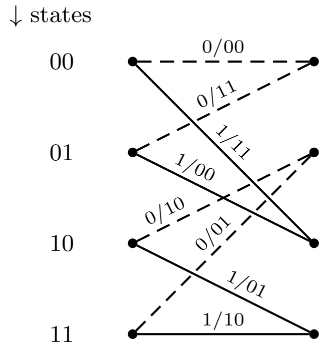
$\boldsymbol{u}$	$\boldsymbol{x}$
000	000000
001	001101
010	010110
011	011011
100	100011
101	101110
110	110101
111	111000

7. Both codes have the same distance spectrum. Moreover, they can both *certainly* correct $\lfloor (d_{\mathrm{H,min}} - 1)/2 \rfloor = 1$ error and detect up to (and including) 2 errors.

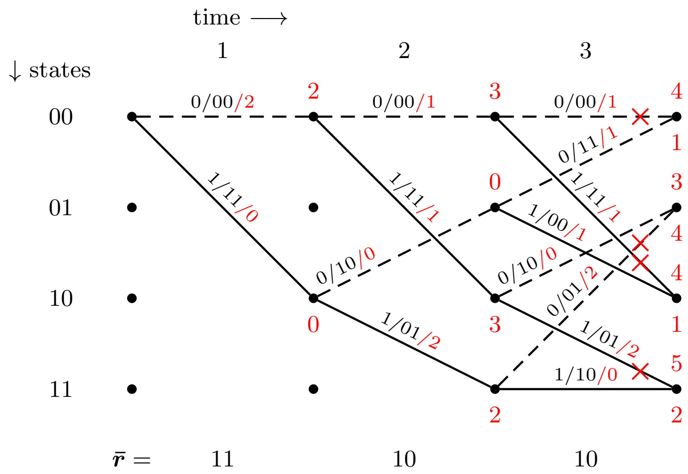
   To explain this, note that we have $\boldsymbol{P} = \boldsymbol{P}^T$. This means that the first three columns of $\mathbf{G}_{\mathrm{s}}$ correspond exactly to the last three columns of $\mathbf{H}_{\mathrm{s}}$. This implies that the codewords in $\mathcal{C}_{\perp}$ can be found by simply swapping the first three and last three bits of the codewords in $\mathcal{C}$.

Problem 4 - Viterbi Algorithm [15 points]

1. One trellis section should look like this:

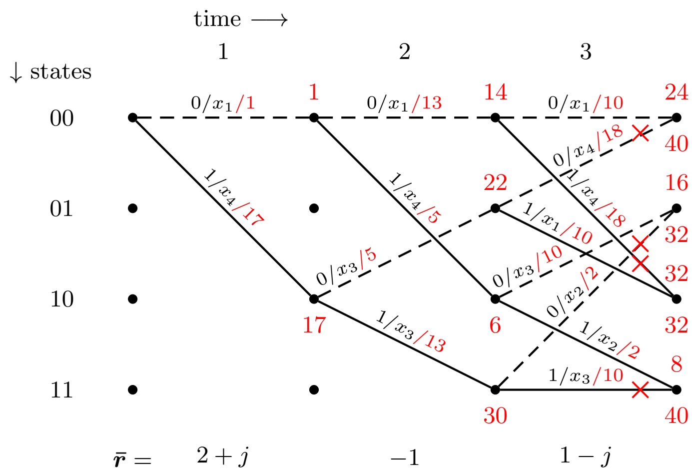


2. In the figure below we show the complete trellis including all branch metrics, state metrics, and survivor paths:



Thus, we can chose $\hat{\boldsymbol{u}}_{\mathrm{ML}} = (101)$ and $\hat{\boldsymbol{c}}_{\mathrm{ML}} = (111000)$.

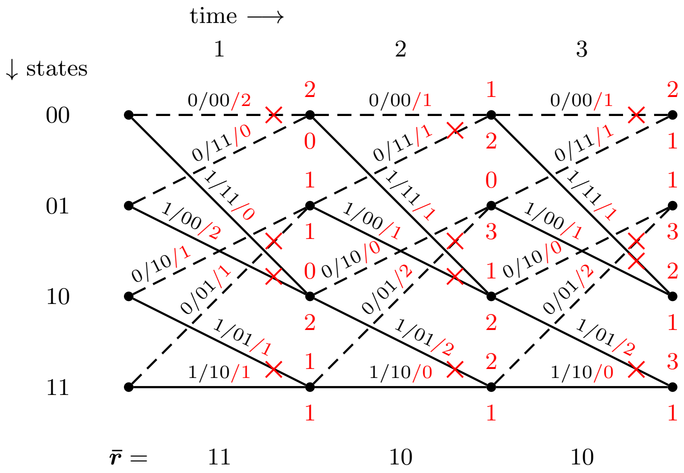
3. In the figure below we show the complete trellis including all branch metrics, state metrics, and survivor paths:



Thus, we have $\hat{\boldsymbol{u}}_{\mathrm{ML}} = (011)$ and $\hat{\boldsymbol{c}}_{\mathrm{ML}} = (001101)$.

Note that the distances are squared Euclidean distances. For example, we have $d_E^2(x_1, 2+j) = 0^2+1^2 = 1$ and $d_E^2(x_4, 2+j) = 4^2+1^2 = 17$ for the first trellis section.

4. In the figure below we show the complete trellis including all branch metrics, state metrics, and survivor paths:



We can chose any of the surviving paths with metric 1. For example, we can declare $\hat{\boldsymbol{u}}_{\mathrm{ML}} = (101)$ and $\hat{\boldsymbol{c}}_{\mathrm{ML}} = (111000)$.