ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 35 Information Theory and Coding

Solutions to Final Exam Feb. 05, 2018

PROBLEM 1.

(a) Observe that

$$\begin{aligned} Z_n &= \mathbb{1}\{N \geq n\} \\ &= 1 - \mathbb{1}\{N \leq n-1\}, \end{aligned}$$

and, as the value of $\mathbb{1}\{N \leq n-1\}$ is determined by U^{n-1}, so is the value of Z_n. Therefore $H(Z_n|U^{n-1}) = 0$.

(b) We have shown in the previous question that Z_n only depends on $U_1, \cdots, U_{n-1}$. Given that $U_1, U_2, \cdots$ are i.i.d. random variables and Z_n is a function of $U_1, \cdots, U_{n-1}$, we see that Z_n is independent of U_n.

(c) Assume $\mathcal{V}$ is not a prefix-free collection. Then there exists two words $\mathbf{v}_1 = v_{11}v_{12}\cdots v_{1m}$ and $\mathbf{v}_2 = v_{21}v_{22}\cdots v_{2n}$, with $m < n$ and $\mathbf{v}_1$ is prefix of $\mathbf{v}_2$. Thus observing $U^m = \mathbf{v}_1$ we cannot determine if $N = m$ or $N = n$. This is a contradiction to the property that the event $\{N = m\}$ is determined by U^m.

(d) We have

$$\begin{aligned} E\left[g\left(U_1, U_2, \cdots\right)\right] &= E\left[\sum_{n=1}^{N} f\left(U_n\right)\right] \\ &\overset{(a)}{=} E\left[\sum_{n=1}^{\infty} f\left(U_n\right) \mathbb{1}\{N \geq n\}\right] \\ &\overset{(b)}{=} \sum_{n=1}^{\infty} E\left[f\left(U_n\right) Z_n\right] \\ &\overset{(c)}{=} \sum_{n=1}^{\infty} E\left[f\left(U_n\right)\right] E\left[Z_n\right] \\ &\overset{(d)}{=} E\left[f(U_1)\right] \sum_{n=1}^{\infty} E[Z_n] \\ &= E\left[f(U_1)\right] E\left[\sum_{i=1}^{\infty} Z_n\right] \\ &\overset{(e)}{=} E\left[f\left(U_1\right)\right] E\left[N\right], \end{aligned}$$

where in (a) we replaced the finite random sum by an infinite sum while multiplying by $\mathbb{1}\{N \geq n\}$ to make sure that for indices bigger than N we are multiplying by 0. (b) is due to the fact that the bounds of the sum are not random anymore and hence is obtained by linearity of the expectation. (c) is due to the independence of Z_n and U_n, thus of Z_n and $f\left(U_n\right)$. (d) is due to the fact that the U_n's are i.i.d and hence $E\left[f\left(U_n\right)\right] = E\left[f\left(U_1\right)\right]$. (e) is due to the fact that N is a non-negative random variable thus $N = \sum_{n=1}^{\infty} \mathbb{1}\{N \geq n\} = \sum_{n=1}^{\infty} Z_n$, as per the hint.

(e) Let's take $g\left(U_1, U_2, \cdots\right) = \log p_V(V)$ in our result in (d). Then,

$$g\left(U_1, U_2, \cdots\right) = \log p_V(V) = \log p_{U^N}\left(U^N\right) = \log \Pi_{i=1}^{N} p_U\left(U_i\right) = \sum_{i=1}^{N} \log p_U\left(U_i\right) = \sum_{i=1}^{N} f(U_i),$$

where $f(U_i) = \log p_U(U_i)$.

$$\begin{aligned}
H(V) &= -E\left[\log p_V(V)\right] \\
&= -E\left[g\left(U_1, U_2, \cdots\right)\right] \\
&\overset{(a)}{=} -E\left[N\right] E\left[f(U_1)\right] \\
&\overset{(b)}{=} -E\left[\log p_U(U_1)\right] E\left[N\right] \\
&\overset{(c)}{=} H(U_1) E\left[N\right],
\end{aligned}$$

where (a) is obtained by applying the result in question (d). (b) is obtained by replacing $f(.)$ by its expression and (c) is obtained using the definition of entropy.

Problem 2.

(a)

$$\Pr(U = u|V = ?) = \frac{\Pr(V = ?|U = u)\, p_U(u)}{\Pr(V = ?)} = \frac{p_U(u)p}{p} = p_U(u)$$

(b)

$$\begin{aligned} I(U;V) &= H(U) - H(U|V) \\ &= H(U) - \Pr(V = ?)H(U|V = ?) - \Pr(V \neq ?)H(U|V \neq ?) \\ &\overset{(a)}{=} H(U) - p\sum_{u=1}^{K} \Pr(U = u|V = ?) \log \frac{1}{\Pr(U = u|V = ?)} \\ &\overset{(b)}{=} H(U) - p\sum_{u=1}^{K} p_U(u) \log \frac{1}{p_U(u)} = H(U) - pH(U) = (1-p)H(U), \end{aligned}$$

where (a) is obtained by noticing that if $V \neq ?$ then $V = U$ and $H(U|V \neq ?) = 0$ and (b) is obtained since $\Pr(U = u|V = ?) = p_U(u)$.

(c) Let C_p be the capacity of this channel. Then,

$$C_p = \max_{p_U} I(U,V) = \max_{p_U}(1-p)H(U) = (1-p)\max_{p_U} H(U) = (1-p)\log K,$$

with the maximum achieved when U is uniformly distributed over $\{1, \cdots, K\}$.

(d)

$$\begin{aligned} I(X;Z) &= H(X) - H(X|Z) \\ &= H(X) - \Pr(Z = ?)H(X|Z = ?) - \Pr(Z \neq ?)H(X|Z \neq ?) \\ &\overset{(a)}{=} H(X) - pH(X|Z = ?) - (1-p)H(X|Y) \\ &\overset{(b)}{=} (1-p)(H(X) - H(X|Y)) + p(H(X) - H(X|Z = ?)) \\ &\overset{(c)}{=} (1-p)I(X;Y) + p(H(X) - H(X)) = (1-p)I(X;Y) \end{aligned}$$

where

- (a) is due to the fact $\Pr(Z = ?) = p$ and when we don't have erasures $Z = Y$.
- (b) $H(X) = pH(X) + (1-p)H(X)$.
- (c) First notice that $X - Y - Z$ forms a Markov chain, so $\Pr(Z = ?|X = x, Y = y) = \Pr(Z = ?|Y = y)$. Then

$$\begin{aligned} \Pr(X = x|Z = ?) &= \frac{p_X(x)\Pr(Z = ?|X = x)}{\Pr(Z = ?)} \\ &= \frac{p_X(x)\sum_y \Pr(Y = y|X = x)\Pr(Z = ?|Y = y)}{p} \\ &= \frac{p_X(x)p\sum_y \Pr(Y = y|X = x)}{p} = p_X(x). \end{aligned}$$

So $H(X|Z = ?) = H(X)$.

(e) Let C_{tot} be the capacity of the channel from X to Z. Then,

$$C_{tot} = \max_{p_X} I(X;Z) = \max_{p_X}(1-p)I(X;Y) = (1-p)\max_{p_X} I(X;Y) = (1-p)C.$$

Problem 3.

(a) In this exercise we assume all the vectors are column vectors. We know that $(X_1, \cdots, X_n, Y_1, \cdots, Y_m$ are jointly Gaussian random variables if and only if any linear combination of these variables is normally distributed. This means that any linear combination of $\mathrm{X} = (X_1, X_2, \cdots, X_n)$ is normally distributed and thus X is an n Gaussian random vector. Similarly, the vector $\mathrm{Y} = (Y_1, \cdots, Y_m)$ is an m dimensional random vector.

Moreover, we can write $(X_1, \cdots, X_n, Y_1, \cdots, Y_m) = (\mathrm{X}, \mathrm{Y})$. So its covariance matrix is

$$K = E\left(\begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \end{bmatrix} \begin{bmatrix} \mathrm{X}^T & \mathrm{Y}^T \end{bmatrix}\right) = \begin{bmatrix} E\left(\mathrm{XX}^T\right) & E\left(\mathrm{XY}^T\right) \\ E\left(\mathrm{YX}^T\right) & E\left(\mathrm{YY}^T\right) \end{bmatrix} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}.$$

So $K_{11} = E\left(\mathrm{XX}^T\right)$ and $K_{22} = E\left(\mathrm{YY}^T\right)$. Thus the vector $\mathrm{X} = (X_1, \cdots, X_n)$ is normally distributed with covariance matrix K_{11} and the vector $\mathrm{Y} = (Y_1, \cdots, Y_m)$ is normally distributed with covariance matrix K_{22}.

Hence, using the results derived in class we get

$$h(X_1, \cdots, X_n) = \frac{1}{2} \ln\left((2\pi e)^n \det\left(K_{11}\right)\right),$$

$$h(Y_1, \cdots, Y_m) = \frac{1}{2} \ln\left((2\pi e)^m \det\left(K_{22}\right)\right)$$

and

$$h(X_1, \cdots, X_n, Y_1, \cdots, Y_m) = \frac{1}{2} \ln\left((2\pi e)^{n+m} \det\left(K\right)\right).$$

(b) Let A_{11} be an $n \times n$ matrix and A_{22} be an $m \times m$ matrix. So A becomes an $(n+m) \times (n+m)$ matrix. Since A is a positive definite matrix then there exists an $n+m$ dimensional Gaussian random vector which covariance matrix is A. Let's denote this vector as $(X_1, \cdots, X_n, Y_1, \cdots, Y_m)$. From question (a) we know that

$$h(X_1, \cdots, X_n) = \frac{1}{2} \ln\left((2\pi e)^n \det\left(A_{11}\right)\right),$$

$$h(Y_1, \cdots, Y_m) = \frac{1}{2} \ln\left((2\pi e)^m \det\left(A_{22}\right)\right)$$

and

$$h(X_1, \cdots, X_n, Y_1, \cdots, Y_m) = \frac{1}{2} \ln\left((2\pi e)^{n+m} \det\left(A\right)\right).$$

Moreover, we know that

$$h\left(X_1, \cdots, X_n, Y_1, \cdots, Y_m\right) \leq h\left(X_1, \cdots, X_n\right) + h\left(Y_1, \cdots, Y_m\right).$$

So,

$$\begin{aligned} \frac{1}{2} \ln\left((2\pi e)^{n+m} \det\left(A\right)\right) &\leq \frac{1}{2} \ln\left((2\pi e)^n \det\left(A_{11}\right)\right) + \frac{1}{2} \ln\left((2\pi e)^m \det\left(A_{22}\right)\right) \\ (2\pi e)^{n+m} \det\left(A\right) &\leq (2\pi e)^n \det\left(A_{11}\right) \times (2\pi e)^m \det\left(A_{22}\right) \\ \det(A) &\leq \det(A_{11}) \det(A_{22}). \end{aligned}$$

PROBLEM 4. In this exercise, unless stated otherwise, all vectors are column vectors.

(a) Let X_1 and X_2 be two codewords in $\mathcal{C}$. Then X_1 and X_2 have their columns and their rows belong to $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Since $\mathcal{C}_1$ and $\mathcal{C}_2$ are linear codes and the rows and columns of $X_1 + X_2$ correspond to the sum of the rows and columns of X_1 and X_2, then the rows of $X_1 + X_2$ belong to $\mathcal{C}_2$ and its columns to $\mathcal{C}_1$. Therefore, $X_1 + X_2 \in \mathcal{C}$. This proves the linearity of $\mathcal{C}$.

(b) Since $X \in \mathcal{C}$, then its columns $(\mathbf{v}_1, \cdots, \mathbf{v}_{n_2})$ are $\mathcal{C}_1$ codewords. This means that there exist $\mathbf{u}_1, \cdots, \mathbf{u}_{n_2} \in \mathbb{F}_2^{k_1}$ such that $\mathbf{v}_j = G_1\mathbf{u}_j = \begin{bmatrix} \mathbf{u}_j \\ A_1\mathbf{u}_j \end{bmatrix}$ for $j = 1, \cdots, n_2$. So the codeword can be written as

$$X = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n_2} \\ A_1\mathbf{u}_1 & A_1\mathbf{u}_2 & \cdots & A_1\mathbf{u}_{n_2} \end{bmatrix}. \tag{1}$$

Similarly, denoting by $(\mathbf{r}_1, \cdots, \mathbf{r}_{n_1})$ the rows of X, there exist $\mathbf{w}_1, \cdots, \mathbf{w}_{n_1} \in \mathbb{F}_2^{k_2}$ such that $\mathbf{r}_l = G_2\mathbf{w}_l = \begin{bmatrix} \mathbf{w}_l \\ A_2\mathbf{w}_l \end{bmatrix}$ for $l = 1, \cdots, n_1$. So X can be also written as

$$X = \begin{bmatrix} \mathbf{r}_1^T \\ \vdots \\ \mathbf{r}_{n_1}^T \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1^T & \mathbf{w}_1^T A_2^T \\ \mathbf{w}_2^T & \mathbf{w}_2^T A_2^T \\ \vdots & \vdots \\ \mathbf{w}_{n_1}^T & \mathbf{w}_{n_1}^T A_2^T \end{bmatrix}. \tag{2}$$

This means that

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{k_2} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_{k_1}^T \end{bmatrix}.$$

From (2) we deduce that $X_{12} = UA_2^T$ and from (1) we deduce that $X_{21} = A_1U$. To find X_{22} we can either notice from (1) that $X_{22} = A_1X_{12} = A_1UA_2^T$ or notice from (2) that $X_{22} = X_{21}A_2^T = A_1UA_2^T$. Hence,

$$X = \begin{bmatrix} U & UA_2^T \\ A_1U & A_1UA_2^T \end{bmatrix}.$$

This shows that X is determined solely by U. Therefore the number of codewords in $\mathcal{C}$ is equal to the number of possible values of U. Since there are $2^{k_1k_2}$ possible values of U then $\mathcal{C}$ has $2^{k_1k_2}$ codewords.

(c) If $U_{rs} = 1$ this means that the s^{th} column of X is a non-zero codeword of the linear code $\mathcal{C}_1$. Therefore, its weight $w_1 \geq d_1$.

(d) If $X_{ts} = 1$ this means that the t^{th} row of X is a non-zero codeword of the linear code $\mathcal{C}_2$. Therefore, its weight $w_2 \geq d_2$.

(e) To show that X with $X_{rs} = c_r^{(1)}c_s^{(2)}$ is a codeword of $\mathcal{C}$ with $c^{(i)} \in \mathcal{C}_i$, we need to show that its columns belong to $\mathcal{C}_1$ and its rows belong to $\mathcal{C}_2$. The s^{th} column of X can be written as

$$\begin{bmatrix} c_1^{(1)} \\ c_2^{(1)} \\ \vdots \\ c_{n_1}^{(1)} \end{bmatrix} c_s^{(2)}.$$

If $c_s^{(2)} = 0$ then the s^{th} column of X is the all zero vector and thus a codeword of $\mathcal{C}_1$. If $c_s^{(2)} = 1$ then the s^{th} column of X is the $c^{(1)}$ vector and thus also a codeword of $\mathcal{C}_1$. This shows that all columns of X are codewords of $\mathcal{C}_1$. Similarly, the r^{th} row of X can be written as

$$\begin{bmatrix} c_1^{(2)} & \cdots & c_{n_2}^{(2)} \end{bmatrix} c_r^{(1)}.$$

If $c_r^{(1)} = 0$, then the r^{th} row is the all-zero vector. If $c_r^{(1)} = 1$, then the r^{th} row is the codeword $c^{(2)}$. In both cases the r^{th} is a codeword of $\mathcal{C}_2$. Hence all rows are codeword of $\mathcal{C}_2$. This shows that X is a codeword of $\mathcal{C}$.

(f) The minimum distance d of a codeword X of $\mathcal{C}$ is the smallest number of 1 possible in X. From part (c) we know that the number of 1's per column $w_1 \geq d_1$ and from part (d) we know that the number of 1's per row $w_2 \geq d_2$. Thus $d \geq d_1 d_2$. Moreover, take $c^{(1)}$ and $c^{(2)}$ to be minimum weight codewords in $\mathcal{C}_1$ and $\mathcal{C}_2$ so they are of weight d_1 and d_2, then the codeword X such that $X_{rs} = c_r^{(1)} c_s^{(2)}$ will have exactly d_2 columns with each column having d_1 1's. This means that the weight of such X is $d_1 d_2$. Therefore $d \leq d_1 d_2$. Therefore $d = d_1 d_2$.