**Problem 1.**

(a) Observe that

\[ Z_n = 1 \{ N \geq n \} = 1 - 1 \{ N \leq n - 1 \}, \]

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

(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 \( v_1 = v_{11}v_{12} \cdots v_{1m} \) and \( v_2 = v_{21}v_{22} \cdots v_{2n} \), with \( m < n \) and \( v_1 \) is prefix of \( v_2 \). Thus observing \( U_m = v_1 \) 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

\[
E [g(U_1, U_2, \cdots)] = E \left[ \sum_{n=1}^{N} f(U_n) \right] = \sum_{n=1}^{\infty} f(U_n) \mathbb{1} \{ N \geq n \} = \sum_{n=1}^{\infty} E[f(U_n)Z_n] = \sum_{n=1}^{\infty} E[f(U_n)] E[Z_n] = E[f(U_1)] \sum_{n=1}^{\infty} Z_n = E[f(U_1)] E[N],
\]

where in (a) we replaced the finite random sum by an infinite sum while multiplying by \( 1 \{ N \geq n \} \) 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 \( f(U_n) \). (d) is due to the fact that the \( U_n \)'s are i.i.d and hence \( E[f(U_n)] = E[f(U_1)] \). (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 $g(U_1, U_2, \cdots) = \log p_V(V)$ in our result in (d). Then,

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

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

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

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) \[
I(U; V) = H(U) - H(U|V) = H(U) - \Pr(V = ?) H(U|V = ?) - \Pr(V \neq ?) H(U|V \neq ?)
\]
\[
\equiv \[a\] H(U) - p \sum_{u=1}^{K} \Pr(U = u | V = ?) \log \frac{1}{\Pr(U = u | V = ?)}
\]
\[
\equiv \[b\] H(U) - p \sum_{u=1}^{K} p_U(u) \log \frac{1}{p_U(u)} = H(U) - p H(U) = (1 - p) H(U),
\]

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 \(\{1, \ldots, K\}\).

(d) \[
I(X; Z) = H(X) - H(X|Z)
\]
\[
= H(X) - \Pr(Z = ?) H(X|Z = ?) - \Pr(Z \neq ?) H(X|Z \neq ?)
\]
\[
\equiv \[a\] H(X) - p H(X|Z = ?) - (1 - p) H(X|Y)
\]
\[
\equiv \[b\] (1 - p) (H(X) - H(X|Y)) + p (H(X) - H(X|Z = ?))
\]
\[
\equiv \[c\] (1 - p) I(X; Y) + p (H(X) - H(X)) = (1 - p) I(X; Y)
\]

where

- (a) is due to the fact \(\Pr(Z = ?) = p\) and when we don’t have erasures \(Z = Y\).
- (b) \(H(X) = p H(X) + (1 - p) H(X)\).
- (c) First notice that \(X - Y - Z\) forms a Markov chain, so \(\Pr(Z = ? | X = x, Y = y) = \Pr(Z = ? | Y = y)\). Then
\[
\Pr(X = x | Z = ?) = \frac{p_X(x) \Pr(Z = ? | X = x)}{\Pr(Z = ?)} = \frac{p_X(x) \sum_y \Pr(Y = y | X = x) \Pr(Z = ? | Y = y)}{p} = \frac{p_X(x) p \sum_y \Pr(Y = y | X = x)}{p} = p_X(x).
\]

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 \(X = (X_1, X_2, \cdots , X_n)\) is normally distributed and thus \(X\) is an \(n\) Gaussian random vector. Similarly, the vector \(Y = (Y_1, \cdots , Y_m)\) is an \(m\) dimensional random vector.

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

\[
K = E \begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X^T \\ Y^T \end{pmatrix}
\]

\[
= \begin{bmatrix} E(XX^T) & E(XY^T) \\ E(YX^T) & E(YY^T) \end{bmatrix}
\]

\[
= \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}
\]

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

Hence, using the results derived in class we get

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

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

and

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

(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

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

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

and

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

Moreover, we know that

\[
h(X_1, \cdots , X_n, Y_1, \cdots , Y_m) \leq h(X_1, \cdots , X_n) + h(Y_1, \cdots , Y_m).
\]

So,

\[
\frac{1}{2} \ln \left( (2\pi e)^{n+m} \det (A) \right) \leq \frac{1}{2} \ln \left( (2\pi e)^n \det (A_{11}) \right) + \frac{1}{2} \ln \left( (2\pi e)^m \det (A_{22}) \right)
\]

\[
(2\pi e)^{n+m} \det (A) \leq (2\pi e)^n \det (A_{11}) \times (2\pi e)^m \det (A_{22})
\]

\[
\det (A) \leq \det (A_{11}) \det (A_{22}).
\]
Problem 4. In this exercise, unless stated otherwise, all vectors are column vectors.

(a) Let $X_1$ and $X_2$ be two codewords in $C$. Then $X_1$ and $X_2$ have their columns and their rows belong to $C_1$ and $C_2$ respectively. Since $C_1$ and $C_2$ 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 $C_2$ and its columns to $C_1$. Therefore, $X_1 + X_2 \in C$. This proves the linearity of $C$.

(b) Since $X \in C$, then its columns $(v_1, \cdots, v_{n_2})$ are $C_1$ codewords. This means that there exist $u_1, \cdots, u_{n_2} \in F_2^{k_1}$ such that $v_j = G_1 u_j = \begin{bmatrix} u_j \\ A_1 u_j \end{bmatrix}$ for $j = 1, \cdots, n_2$. So the codeword can be written as

$$X = \begin{bmatrix} u_1 & u_2 & \cdots & u_{n_2} \\ A_1 u_1 & A_1 u_2 & \cdots & A_1 u_{n_2} \end{bmatrix}.$$ (1)

Similarly, denoting by $(r_1, \cdots, r_{n_1})$ the rows of $X$, there exist $w_1, \cdots, w_{n_1} \in F_2^{k_2}$ such that $r_l = G_2 w_l = \begin{bmatrix} w_l \\ A_2 w_l \end{bmatrix}$ for $l = 1, \cdots, n_1$. So $X$ can be also written as

$$X = \begin{bmatrix} r_1^T \\ \vdots \\ r_{n_1}^T \end{bmatrix} = \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_{n_1}^T \end{bmatrix} = \begin{bmatrix} w_1^T A_2^T \\ w_2^T A_2^T \\ \vdots \\ w_{n_1}^T A_2^T \end{bmatrix}.$$ (2)

This means that

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_{n_2} \end{bmatrix}^T = \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_{n_1}^T \end{bmatrix} = \begin{bmatrix} w_1^T \\ w_2^T \\ w_{n_1}^T \end{bmatrix}.$$ (3)

From (2) we deduce that $X_{12} = U A_2^T$ and from (1) we deduce that $X_{21} = A_1 U$. To find $X_{22}$ we can either notice from (1) that $X_{22} = A_1 X_{12} = A_1 U A_2^T$ or notice from (2) that $X_{22} = X_{21} A_2^T = A_1 U A_2^T$. Hence,

$$X = \begin{bmatrix} U & U A_2^T \\ A_1 U & A_1 U A_2^T \end{bmatrix}.$$ (4)

This shows that $X$ is determined solely by $U$. Therefore the number of codewords in $C$ is equal to the number of possible values of $U$. Since there are $2^{k_1 k_2}$ possible values of $U$ then $C$ has $2^{k_1 k_2}$ codewords.

(c) If $U_{rs} = 1$ this means that the $s^{th}$ column of $X$ is a non-zero codeword of the linear code $C_1$. 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 $C_2$. 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 $C$ with $c_r^{(1)} \in C_1$, we need to show that its columns belong to $C_1$ and its rows belong to $C_2$. The $s^{th}$ column of $X$ can be written as

$$\begin{bmatrix} c_r^{(1)} \\ c_r^{(2)} \\ \vdots \\ c_r^{(n_1)} \end{bmatrix}.$$
If $c_s^{(2)} = 0$ then the $s^{th}$ column of $X$ is the all zero vector and thus a codeword of $C_1$. If $c_s^{(2)} = 1$ then the $s^{th}$ column of $X$ is the $c^{(1)}$ vector and thus also a codeword of $C_1$. This shows that all columns of $X$ are codewords of $C_1$. Similarly, the $r^{th}$ row of $X$ can be written as

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

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 $C_2$. Hence all rows are codeword of $C_2$. This shows that $X$ is a codeword of $C$.

(f) The minimum distance $d$ of a codeword $X$ of $C$ 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 $d \geq d_1d_2$. Moreover, take $c^{(1)}$ and $c^{(2)}$ to be minimum weight codewords in $C_1$ and $C_2$ 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 $d_1d_2$. Therefore $d \leq d_1d_2$. Therefore $d = d_1d_2$. 