Model Answers to Exercise 7 of November 1, 2017

http://www.isi.ee.ethz.ch/teaching/courses/it1.html

Problem 1  
On the Achievable Rate

The second channel, $W_2$, can be considered to consist of two parallel channels $W_1$ as shown in the figure below, i.e., one use of $W_2$ corresponds to two independent uses of $W_1$.

![Diagram of Channels](image)

- $x_1$ is encoded into $y_1$ via $W_1(y_1|x_1)$.
- $x_2$ is encoded into $y_2$ via $W_1(y_2|x_2)$.
- $y_1$ and $y_2$ are combined using $W_2(y_1, y_2|x_1, x_2)$.

a) Let $n$ and $R$ be fixed. Let $f: \{1, \ldots, 2^{2nR}\} \to \mathcal{X}^{2n}$ and $\phi: \mathcal{Y}^{2n} \to \{1, \ldots, 2^{2nR}\}$ be a length-2n rate-$R$ encoder and decoder pair for channel $W_1$. We construct a length-$n$ rate-$2R$ encoder and decoder pair $g: \{1, \ldots, 2^{n2R}\} \to (\mathcal{X} \times \mathcal{X})^n$ and $\psi: (\mathcal{Y} \times \mathcal{Y})^n \to \{1, \ldots, 2^{n2R}\}$ for channel $W_2$ by interleaving:

- Note that both encoder and decoder pairs have the same message set since $2^{2nR} = 2^{n2R}$.
- To encode a message $m$, use $f$ to obtain the codeword $(x_1, \ldots, x_{2n}) = f(m)$, and define $g(m)$ to be

$$
\begin{bmatrix}
[x_1] \\
[x_2] \\
[x_3] \\
\vdots \\
[x_{2n}] \\
\end{bmatrix}
$$

(The top entries correspond to the first input of $W_2$, and the bottom entries correspond to the second input of $W_2$.)

- To decode a sequence $y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{2n} \end{bmatrix}$, use $\phi$, i.e., define $\psi(y)$ to be $\phi(y_1, \ldots, y_{2n})$. (The top entries correspond to the first output of $W_2$, and the bottom entries correspond to the second output of $W_2$.)

Since one use of $W_2$ corresponds to two independent uses of $W_1$, the encoder and decoder pair $(g, \psi)$ has the same probability of error as the encoder and decoder pair $(f, \phi)$.

If $R$ is achievable on $W_1$, there must exist a sequence of encoder and decoder pairs for $W_1$ whose probability of error tends to zero as $n$ tends to infinity. In this case, the above construction yields a sequence of encoder and decoder pairs for $W_2$ whose probability of error tends to zero as $n$ tends to infinity, which shows that $2R$ is achievable on $W_2$. 

© Amos Lapidoth, 2017
b) We use the same idea as in Part a). For a given length-$n$ rate-$R$ encoder on $W_2$, we construct a length-$2n$ rate-$R/2$ encoder for $W_1$ that alternates between emitting the first input for $W_2$ and the second input for $W_2$. Similarly, for a given length-$n$ rate-$R$ decoder on $W_2$, we construct a length-$2n$ rate-$R/2$ decoder for $W_1$ that splits the received sequence into two sequences and invokes the decoder for $W_2$. The encoder and decoder pair for $W_1$ then has the same probability of error as the encoder and decoder pair for $W_2$.

If $R$ is achievable on $W_2$, there must exist a sequence of encoder and decoder pairs for $W_2$ whose probability of error tends to zero as $n$ tends to infinity. Using the above construction and ignoring that it only yields encoder and decoder pairs for $W_1$ with even blocklength, we see that $R/2$ is achievable on $W_1$.

Problem 3

Remember that the Z-channel looks as shown in Figure 1.

\begin{figure}[h]
\centering
\begin{tikzpicture}
  \node (0) at (0,0) {0};
  \node (1) at (0,-1) {1};
  \node (2) at (0,1) {1};
  \node (3) at (1,0) {0};
  \draw (0) -- (1) node[midway,above] {$\frac{1}{2}$};
  \draw (0) -- (2) node[midway,above] {$\frac{1}{2}$};
  \draw (1) -- (3) node[midway,above] {1};
\end{tikzpicture}
\caption{Z-Channel.}
\end{figure}

First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $p = \Pr[X = 1]$:

\begin{align*}
H(Y|X = 0) &= 0; \\
H(Y|X = 1) &= H_b\left(\frac{1}{2}\right) = 1 \text{ bit;} \\
\implies H(Y|X) &= \Pr[X = 0] \cdot 0 + \Pr[X = 1] \cdot 1 = p \text{ bits;} \\
\Pr[Y = 0] &= \Pr[X = 0] \cdot 1 + \Pr[X = 1] \cdot \frac{1}{2} = 1 - p + p \cdot \frac{1}{2} = 1 - \frac{1}{2} p; \\
\Pr[Y = 1] &= \Pr[X = 0] \cdot 0 + \Pr[X = 1] \cdot \frac{1}{2} = \frac{1}{2} p; \\
\implies H(Y) &= H_b\left(\frac{p}{2}\right); \\
\implies I(X;Y) &= H(Y) - H(Y|X) = H_b\left(\frac{p}{2}\right) - p \text{ bits.}
\end{align*}

Since $I(X;Y) = 0$ if $p = 0$ or $p = 1$, the maximum mutual information is obtained for some value of $p$ such that $0 < p < 1$. Using calculus, we determine that

$$\frac{d}{dp} I(X;Y) = \frac{1}{2} \log_2 \frac{2 - p}{p} - 1,$$

which is equal to zero for $p = \frac{2}{5}$. (It is reasonable that $\Pr[X = 1] < \frac{1}{2}$ because $X = 1$ is the noisy input to the channel.) So the capacity of the Z-channel is

$$H_b\left(\frac{1}{5}\right) - \frac{2}{5} \approx 0.722 - 0.4 = 0.322 \text{ bits per channel use.}$$
Problem 4  

Independent Parallel Channels

The capacity of the parallel channel can be computed as

$$C = \max_Q I(X_1, X_2; Y_1, Y_2),$$

where the maximum is over all input distributions on the alphabet $\mathcal{X}_1 \times \mathcal{X}_2$. Because the channels are independent, we have

$$H(Y_1, Y_2|X_1, X_2) \overset{\text{(i)}}{=} H(Y_1|X_1) + H(Y_2|X_2).$$

(Since

$$\Pr[Y_1 = y_1, Y_2 = y_2|X_1 = x_1, X_2 = x_2] = \Pr[Y_1 = y_1|X_1 = x_1] \cdot \Pr[Y_2 = y_2|X_2 = x_2],$$

we have

$$H(Y_1, Y_2|X_1 = x_1, X_2 = x_2) = H(Y_1|X_1 = x_1) + H(Y_2|X_2 = x_2),$$

which implies (i).) Therefore,

$$I(X_1, X_2; Y_1, Y_2) = H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2)$$
$$= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2)$$
$$= H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2)$$
$$\overset{\text{(ii)}}{\leq} H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2)$$
$$= I(X_1; Y_1) + I(X_2; Y_2),$$

where (ii) holds because conditioning does not increase entropy. Thus,

$$C = \max_Q I(X_1, X_2; Y_1, Y_2)$$
$$\leq \max_Q \{I(X_1; Y_1) + I(X_2; Y_2)\}$$
$$\overset{\text{(iii)}}{\leq} C_1 + C_2,$$

where (iii) holds because $I(X_1; Y_1) \leq C_1$ and $I(X_2; Y_2) \leq C_2$ by the definition of $C_1$ and $C_2$. Choosing $Q(x_1, x_2) = Q_1(x_1)Q_2(x_2)$, where $Q_1$ and $Q_2$ are capacity-achieving input distributions for $W_1$ and $W_2$, respectively, leads to

$$I(X_1, X_2; Y_1, Y_2) \overset{\text{(iv)}}{=} I(X_1; Y_1) + I(X_2; Y_2)$$
$$\overset{\text{(v)}}{=} C_1 + C_2,$$

where (iv) holds because $X_1$ and $X_2$ are independent and because the channels are independent, so $Y_1$ and $Y_2$ independent, which implies that (ii) holds with equality; and (v) holds because $X_1$ and $X_2$ are distributed according to $Q_1$ and $Q_2$, respectively, which are assumed to be capacity-achieving input distributions. Combining (1) and (2), we conclude that

$$C = C_1 + C_2.$$  

Note: The above result can be generalized to $n$ independent parallel channels. By induction,

$$C = \sum_{k=1}^{n} C_k.$$  

© Amos Lapidoth, 2017
a) Let \( S \in \{1, \ldots, \nu\} \) be the chance variable representing the selected channel, i.e., \( S \sim s \) where \( s \) is the probability vector \((s_1, \ldots, s_\nu)\).

Applying the chain rule twice to \( H(X, S) \) leads to
\[
H(X) + H(S|X) = H(X, S) = H(S) + H(X|S),
\]
where the underbraced term is zero because \( X \) determines \( S \). Applying the chain rule twice to \( H(X, S|Y) \) leads to
\[
H(X|Y) + H(S|X,Y) = H(X, S|Y) = H(S|Y) + H(X|Y, S),
\]
where the underbraced terms are zero because \( Y \) determines \( S \). Therefore, we have
\[
I(X; Y) = H(X) - H(X|Y) = H(S) + H(X|S) - H(X|Y, S) = H(S) + \sum_{i=1}^{\nu} s_i I(X; Y|S = i).
\]

By the law of total probability, we rewrite the input distribution of the sum channel as
\[
P_X(x) = \sum_{i=1}^{\nu} s_i P_{X|S}(x|i).
\]

Plugging these expression into the formula for the capacity leads to
\[
C = \max_{P_X} I(X; Y) = \max_{s} \max_{P_X|S} \left( H(S) + \sum_{i=1}^{\nu} s_i I(X; Y|S = i) \right) = \max_{s} \left( H(S) + \sum_{i=1}^{\nu} s_i C_i \right),
\]
where the last equality holds because \( I(X; Y|S = i) \leq C_i \) (by the definition of the capacity) and because \( I(X; Y|S = i) = C_i \) if and only if, conditional on \( S = i \), \( X \) is chosen to be a capacity-achieving input distribution for the \( i \)-th channel. Therefore, \( C \) is equal to the entropy of the channel selection plus a weighted average of the channel capacities, and it remains to determine the optimal channel selection.

We proceed with
\[
H(S) + \sum_{i=1}^{\nu} s_i C_i = \sum_{i=1}^{\nu} s_i \log \frac{1}{s_i} + \sum_{i=1}^{\nu} s_i \log 2^{C_i} = -\sum_{i=1}^{\nu} s_i \log \frac{s_i}{2^{C_i}} = -\sum_{i=1}^{\nu} s_i \log \frac{s_i/\alpha}{2^{C_i}/\alpha} \quad \left( \text{where } \alpha = \sum_{i=1}^{\nu} 2^{C_i} \right)
\]

© Amos Lapidoth, 2017 4
\[
\sum_{i=1}^{\nu} s_i \log \alpha - \sum_{i=1}^{\nu} s_i \log \frac{s_i}{t_i} = \log \alpha - D(s\|t) \geq 0
\]

and conclude that

\[
C = \max_s \left( H(S) + \sum_{i=1}^{\nu} s_i C_i \right) = \log \alpha = \log \sum_{i=1}^{\nu} 2^{C_i},
\]

where the maximum is achieved for \( s = t \), i.e., \( s_i = \frac{2^{C_i}}{\sum_{j=1}^{\nu} 2^{C_j}} \) for all \( i \in \{1, \ldots, \nu\} \).

b) The capacity of the BSC is \( 1 - H_b(\epsilon) \) bits and the capacity of the other channel is zero. Thus by Part a) the capacity of the sum channel is

\[
C = \log \left( 2^{1-H_b(\epsilon)} + 2^0 \right) = \log \left( 1 + 2^{1-H_b(\epsilon)} \right).
\]