Problem 1

"Double"-Gaussian Channel

a) The capacity of this channel is given by

\[ C = \frac{1}{2} \log \left( 1 + \frac{1}{N} \right). \]

The corresponding plot is shown in Figure 1.

Figure 1: Channel capacity \( C \) (in bits per channel use) versus variance of the noise \( N \).

b) i) Since the total power is divided equally between the two channels,

\[ C = \frac{1}{2} \log \left( 1 + \frac{E_s}{N_1} \right) + \frac{1}{2} \log \left( 1 + \frac{E_s}{N_2} \right). \]

For scheme A we therefore get

\[ C_A = 2 \cdot \frac{1}{2} \log \left( 1 + \frac{E_s}{2N} \right) = \log \left( 1 + \frac{E_s}{2N} \right). \]
As it is seen in Figure 1, \( f(x) = \log \left( 1 + \frac{E_s}{2x} \right) \) is a convex function of \( x \). Therefore, using Jensen’s inequality,

\[
C = \frac{1}{2} \log \left( 1 + \frac{E_s}{2N_1} \right) + \frac{1}{2} \log \left( 1 + \frac{E_s}{2N_2} \right) \\
= \frac{1}{2} f(N_1) + \frac{1}{2} f(N_2) \\
\geq f \left( \frac{1}{2} N_1 + \frac{1}{2} N_2 \right) \\
= \log \left( 1 + \frac{E_s}{2(N_1/2 + N_2/2)} \right) \\
= \log \left( 1 + \frac{E_s}{2N} \right) = C_A.
\]

Thus, scheme B is better.

ii) Note first that \( f(x) = \log \left( 1 + \frac{x}{N} \right) \) is a concave function of \( x \). Therefore, with \( \bar{\lambda} = 1 - \lambda \),

\[
C_A = \frac{1}{2} \log \left( 1 + \frac{\lambda E_s}{N} \right) + \frac{1}{2} \log \left( 1 + \frac{\bar{\lambda} E_s}{N} \right) \\
= \frac{1}{2} f(\lambda E_s) + \frac{1}{2} f(\bar{\lambda} E_s) \\
\leq f \left( \frac{1}{2} \lambda E_s + \frac{1}{2} \bar{\lambda} E_s \right) \\
= f \left( \frac{1}{2} E_s \right) \\
= \log \left( 1 + \frac{E_s}{2N} \right) \\
= \frac{1}{2} \log \left( 1 + \frac{E_s}{2N} \right) + \frac{1}{2} \log \left( 1 + \frac{E_s}{2N} \right),
\]

which means that for scheme A, an equal input power \( \frac{E_s}{2} \) achieves the highest capacity. Now, since scheme B is better even under equal input power (see b-i)), if we optimize over input power we are going to find that scheme B is still better.

Problem 2

Additive Noise Channel

a) The probability density of the random variables \( Z_k \) for all \( k \) is given as follows:

\[
f_Z(z) = \frac{1}{10} \delta(z) + \frac{9}{10} f_G(z), \quad z \in \mathbb{R},
\]

where \( f_G(z) \) is the probability density of the Gaussian distribution \( \mathcal{N}(0, \sigma^2) \). Hence,

\[
h(Z) = -\int_{-\infty}^{\infty} f_Z(z) \log f_Z(z) \, dz \\
= -\frac{1}{10} \int_{-\infty}^{\infty} \delta(z) \log \left( \frac{1}{10} \delta(z) + \frac{9}{10} f_G(z) \right) \, dz \\
- \frac{9}{10} \int_{-\infty}^{\infty} f_G(z) \log \left( \frac{1}{10} \delta(z) + \frac{9}{10} f_G(z) \right) \, dz
\]
\[-\frac{1}{10} \log \left( \frac{1}{10} \delta(0) + \frac{9}{10} f_G(0) \right) - \frac{9}{10} \int_{-\infty}^{\infty} f_G(z) \log \left( \frac{1}{10} \delta(z) + \frac{9}{10} f_G(z) \right) dz \]
\[-\infty - \frac{9}{10} \int_{-\infty}^{\infty} f_G(z) \log \left( \frac{1}{10} \delta(z) + \frac{9}{10} f_G(z) \right) dz.\]

Since
\[\log \left( \frac{1}{10} \delta(z) + \frac{9}{10} f_G(z) \right) \geq \log \left( \frac{9}{10} f_G(z) \right), \quad z \in \mathbb{R},\]
we get
\[\int_{-\infty}^{\infty} f_G(z) \log \left( \frac{1}{10} \delta(z) + \frac{9}{10} f_G(z) \right) dz \geq \int_{-\infty}^{\infty} f_G(z) \log \left( \frac{9}{10} f_G(z) \right) dz = \log \left( \frac{9}{10} \right) - \frac{1}{2} \log(2\pi e\sigma^2),\]
and thus
\[h(Z) \leq -\infty - \frac{9}{10} \log \left( \frac{9}{10} \right) - \frac{9}{20} \log(2\pi e\sigma^2) = -\infty,\]
i.e., \(h(Z) = -\infty\). This is actually clear as \(Z_k\) takes the fixed value 0 with a nonzero probability.

Further we note that from conditioning that reduces entropy we have
\[h(Y) = h(X + Z) \geq h(X + Z|Z) = h(X|Z) = h(X). \quad (1)\]
Hence,
\[C = \max_{f_X} I(X; Y) \]
\[= \max_{f_X} \left\{ h(Y) - h(Y|X) \right\} \]
\[= \max_{f_X} \left\{ h(Y) - h(X + Z|X) \right\} \]
\[= \max_{f_X} \left\{ h(Y) - h(Z|X) \right\} \]
\[= \max_{f_X} h(Y) - h(Z) \]
\[\geq \max_{f_X} h(X) - h(Z) \quad \text{(by (1))} \]
\[= \infty, \quad >-\infty \]
i.e., the capacity is infinite: \(C = \infty\).

b) In order for a coding scheme to achieve capacity, we must show that for any \(\epsilon > 0\), there exists an \(n_0\) such that for any \(n > n_0\), we can send an arbitrary number of bits using \(n\) channel uses with probability of error less than \(\epsilon\). We propose two different schemes:

**Coding scheme 1:** The encoder maps the message to a rational number with absolute value less than \(\sqrt{E}\) and transmits that number repeatedly. The decoder looks at the received sequence to see if there is a rational number. If yes, declare it to be the input. Otherwise, declare an error.
Since the output will be irrational with probability one if $Z$ is Gaussian, the probability of error is $(0.9)^n$. For any $\epsilon > 0$, we can find an $n_0$ such that $(0.9)^n < \epsilon$, for $n > n_0$.

Note that there is an infinite number of rational numbers between $-\sqrt{E_s}$ and $\sqrt{E_s}$, so we can indeed send an infinite number of messages.

**Coding scheme 2:** The encoder maps the message to a real number with absolute value less than $\sqrt{E_s}$ and transmits the number repeatedly. The decoder looks at the received sequence to see if there are any repeated numbers. If yes, declare them to be the input. Otherwise, declare an error.

With probability one, the only repeated received symbols will be the noise free copies of the input. Thus, the probability of error for this scheme is the probability that $Z = 0$ for none or only one out of the $n$ transmissions: $(0.9)^n + \binom{n}{1}(0.1)(0.9)^{n-1}$. Again, we can choose $n_0$ large enough to ensure probability of error less than any $\epsilon > 0$, and there is an infinite number of real numbers between $-\sqrt{E_s}$ and $\sqrt{E_s}$.

---

**Problem 3**

**Exponential Noise Channel**

We shall use the following capacity formula without proving it:

$$C = \max_{E[X] \leq \lambda} I(X; Y).$$

We can rewrite $I(X; Y)$ as

$$I(X; Y) = h(Y) - h(Y|X)$$
$$= h(Y) - h(Z|X)$$
$$= h(Y) - h(Z).$$

Recall that, for a nonnegative random variable $W$ satisfying $E[W] = \nu$, the entropy-maximizing distribution is the exponential distribution with mean $\nu$:

$$f(w) = \frac{1}{\nu} e^{-\frac{w}{\nu}}, \quad w \geq 0,$$

and that the corresponding differential entropy is

$$h(W) = \log e\nu.$$

(See Example 12.2.5 in Cover & Thomas.)

Thus, we can bound $I(X; Y)$ under the condition $E[X] \leq \lambda$ as

$$I(X; Y) = h(Y) - h(Z)$$
$$= h(Y) - \log e\mu$$
$$\leq \log e E[Y] - \log e\mu$$
$$= \log e(E[X] + E[Z]) - \log e \mu$$
$$\leq \log e(\lambda + \mu) - \log e \mu$$
$$= \log \left(1 + \frac{\lambda}{\mu}\right). \quad (2)$$

Therefore we obtain

$$C \leq \log \left(1 + \frac{\lambda}{\mu}\right).$$
We next show that equality in (2) can be achieved. To this end, we choose \( X \) to be equal to 0 with probability \( \frac{\mu}{\mu + \lambda} \) and to be exponentially distributed with mean \( \mu + \lambda \) otherwise. In other words, we choose \( X \) in such a way that
\[
\Pr[X \geq x] = \frac{\lambda}{\mu + \lambda} e^{-\frac{x}{\mu + \lambda}}.
\]
It can be easily verified that the distribution on \( Y \) induced by the above choice of \( X \) is exponential with mean \( \lambda + \mu \). Hence, this choice achieves equality in (2) and therefore we have
\[
C = \log \left( 1 + \frac{\lambda}{\mu} \right).
\]

**Remark:** A more rigorous derivation of this result uses Fano’s Inequality to prove the converse and uses a joint-typicality decoder to prove the achievability. It is done in a similar way to the derivation of the capacity of the Gaussian channel.

**Problem 4**

**DMC with a Cost Constraint**

We first show the achievability part. We will show that, for any \( \delta > 0 \) and
\[
R < \sup_{E[b(X)] \leq \beta - \delta} I(X; Y),
\]
there exists a sequence of \( (n, R) \) codes satisfying, for every \( m \in \{1, \ldots, 2^{nR}\} \),
\[
\frac{1}{n} \sum_{i=1}^{n} b(x_i(m)) \leq \beta, \quad m \in \{1, \ldots, 2^{nR}\},
\]
such that as \( n \) tends to infinity, the maximum error probability tends to zero. To this end, for \( R \) satisfying (3) and for any \( n \), we randomly generate an \( (n, R) \) codebook where the codewords are drawn independently, and where the symbols in each codeword are drawn IID according to \( p^* \), where \( p^* \) is the distribution that achieves the supremum on the RHS of (3).

When a sequence \( Y^n \) is received, the decoder looks in the codebook for a codeword that is jointly typical with \( Y^n \), and satisfies the cost constraint (4). If there is only one such codeword, the decoder outputs the message corresponding to this codeword; if there is more than one or no such codeword, the decoder declares an error.

We now analyze the average error probability averaged over all such randomly generated codebooks. We list and analyze the possible error types as follows:

- The transmitted codeword does not satisfy the cost constraint (4), and is thus always rejected by the decoder. By the law of large numbers, the probability of this type of errors tends to zero as \( n \) tends to infinity.
- The transmitted codeword and the received sequence are not jointly typical. Also by AEP, the probability of this type of errors tends to zero as \( n \) tends to infinity.
- There exists a codeword which is not transmitted but which is jointly typical with \( Y^n \). The probability of this event tends to zero as \( n \) tends to infinity because \( R \) satisfies (3).

We have thus shown that, for any positive \( \epsilon \), for all large enough \( n \), the average error probability averaged over the above codebooks is smaller than \( \epsilon \). It then follows that there must be at least one codebook whose average error probability is smaller than \( \epsilon \). We then throw away half of the codewords in this codebook with higher error probabilities, then the maximum error probability of
the remaining codewords must be smaller than $2\epsilon$. Furthermore, all the remaining codewords must satisfy the cost constraint (4), since those codewords that do not satisfy (4) have error probability one. Note that by throwing away half of the codewords we lose $\frac{1}{n}$ bit of rate, which tends to zero as $n$ tends to infinity. This concludes the proof of the achievability part.

We next prove the converse part. We will show that if there exists a sequence of codebooks of rate $R$ whose average error probability tends to zero as the block-length $n$ tends to infinity satisfying the weaker cost constraint
\begin{equation}
\frac{1}{2n} \sum_{m=1}^{2^n} \frac{1}{n} \sum_{i=1}^{n} b(x_i(m)) \leq \beta, \tag{5}
\end{equation}
then
\begin{equation}
R \leq \sup_{E[b(X)] \leq \beta} I(X; Y). \tag{6}
\end{equation}
To this end, we bound $nR$ as follows:
\[
nR \leq I(X^n; Y^n) + n\epsilon_n
\]
\[
= h(Y^n) - h(Y^n|X^n) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Y_i|X_i) + n\epsilon_n
\]
\[
= \sum_{i=1}^{n} I(X_i; Y_i) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^{n} \sup_{E[b(X_i)] \leq E[b(X_i)]} I(X_i; Y_i) + n\epsilon_n, \tag{7}
\]
where $E[b(X_i)]$ is the cost of the $i$th symbol averaged over the codebook, and where $\epsilon_n$ tends to zero as $n$ tends to infinity. Here, the first inequality follows by Fano’s inequality; and the second equality follows because the channel is memoryless. We next notice that the RHS of (6) is concave in $\beta$. To see this, let $p_1$ and $p_2$ achieve the supremum on the RHS of (6) when $\beta = \beta_1$ and when $\beta = \beta_2$, respectively, and let $p^* = \lambda p_1 + (1 - \lambda)p_2$, $\lambda \in (0, 1)$. Then $p^*$ satisfies
\[
E_{p^*}[b(X)] = \lambda E_{p_1}[b(X)] + (1 - \lambda) E_{p_2}[b(X)] \leq \lambda \beta_1 + (1 - \lambda) \beta_2.
\]
On the other hand, because mutual information is concave in the input distribution, we have
\[
I(X; Y)|_{p^*} \geq \lambda I(X; Y)|_{p_1} + (1 - \lambda) I(X; Y)|_{p_2},
\]
which establishes the desired concavity. We use this concavity and (6) to continue (7) as
\[
nR \leq \sum_{i=1}^{n} \sup_{E[b(X)] \leq E[b(X_i)]} I(X_i; Y_i) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^{n} \sup_{E[b(X)] \leq \beta} I(X_i; Y_i) + n\epsilon_n.
\]
Since $\epsilon_n$ tends to zero as $n$ tends to infinity, we have proved (6). This concludes the proof of the converse and hence also the whole proof.

© Amos Lapidoth, 2014/2015 6