Solutions to Homework Set #2

3.9. (a) The operational definition of the capacity cost function is

\[ C(B) = \sup \{ R : (R, B) \text{ is achievable} \}. \]

Since the supremum is taken over a bigger set as \( B \) increases, \( C(B) \) is nondecreasing for \( B \geq 0 \).

To prove the concavity, let \( B_1 \) and \( B_2 \) be two cost constraints. Suppose that \( R_1 \) is achievable under \( B_1 \) and \( R_2 \) is achievable under \( B_2 \) i.e. \( R_1 \leq C(B_1) \) and \( R_2 \leq C(B_2) \). Let \( k = |\alpha n|, k' = n - k \) and \( \alpha \in [0,1] \). We can construct a code by using a \((2^{kR_1}, k)\) code for the first \( k \) transmissions and a \((2^{k'R_2}, k')\) code for the rest of \( k' \) transmissions. Hence, the resulting code can achieve rate \( \alpha R_1 + \alpha R_2 \) with cost constraint \( E(b(X)) \leq \alpha B_1 + \alpha B_2 \). Therefore, \( C(B) \) is concave for \( B \geq 0 \).

(b) The information capacity–cost function is defined as

\[ C(B) = \max_{p(x):E(b(X)) \leq B} I(X; Y). \]

The monotonicity is trivial. To prove the concavity, let \( B_1 \) and \( B_2 \) be two cost constraints, and let \( p_1(x) \) and \( p_2(x) \) be two probability distributions that attain \( I_{p_1}(X; Y) = C(B_1) \) and \( I_{p_2}(X; Y) = C(B_2) \), respectively. Let \( p(x) = \alpha p_1(x) + \bar{\alpha} p_2(x) \) for \( \alpha \in [0,1] \). Then, \( E_p(b(X)) = \alpha E_{p_1}(b(X)) + \bar{\alpha} E_{p_2}(b(X)) \leq \alpha B_1 + \bar{\alpha} B_2 \), which implies that

\[ I_p(X; Y) \leq C(B). \]

Now by the concavity of the mutual information \( I(X; Y) \) in \( p(x) \) for fixed \( p(y|x) \), we have

\[ I_p(X; Y) \geq \alpha I_{p_1}(X; Y) + \bar{\alpha} I_{p_2}(X; Y) = \alpha C(B_1) + \bar{\alpha} C(B_2). \]

Combining the two bounds on \( I_p(X; Y) \) establishes the concavity of \( C(B) \).

The continuity of \( C(B) \) on \((0, \infty)\) follows immediately from the concavity. For the continuity at \( B = 0 \), observe that the set \( \{(R, B) : R \leq I(X; Y), B \geq E(b(X)) \text{ for some } p(x) \} \) is closed.

3.12. It follows immediately from the operational definition of capacity. Alternatively, note that the output becomes \( Y' = agX + aZ \), where \( a \neq 0 \). Thus,

\[ I(X; Y') = I(X; agX + aZ) \]

\[ = h(agX + aZ) - h(aZ) \]

\[ = \frac{1}{2} \log(2\pi e(a^2 g^2 P)) - \frac{1}{2} \log(2\pi e(a^2)) \]

\[ = \frac{1}{2} \log(g^2 P). \]

3.14. (a) Consider the random codebook generation as in the standard achievability proof for the DMC. For the decoding, let \( \mathcal{A} = \{ m : (x^n(m), y^n) \in \mathcal{T}_e(n) \} \) and declare \( \mathcal{L} = \mathcal{A} \) if \( |\mathcal{A}| \leq 2^nL \) (and take an arbitrary \( \mathcal{L} \) otherwise). Suppose \( M = 1 \). Then the probability of error is bounded as

\[ P_e(n) = P\{ M \notin \mathcal{L}(Y^n) | M = 1 \} \]

\[ \leq P\{ (X^n(1), Y^n) \notin \mathcal{T}_e(n) | M = 1 \} + P\{ |\mathcal{A}| > 2^nL | M = 1 \}. \]
By the LLN, the first term tends to zero as \( n \to \infty \). By the Markov inequality and the joint typicality lemma, the second term is bounded as

\[
P\{|A| > 2^{nL}|M = 1\} \leq \frac{\mathbb{E}(|A||M = 1)}{2^{nL}L}
\leq \frac{1}{2^{nL}} \left(1 + \sum_{m=2}^{2^{nR}} P\{(X^n(m), Y^n) \in T^{(n)}_\epsilon\}\right)
\leq 2^{-nL} \log(1 + 2^{nR}L - I(X; Y) + \delta(\epsilon)),
\]

which tends to zero as \( n \to \infty \) if \( R < L + I(X; Y) - \delta(\epsilon) \).

Alternatively, we can partition \( 2^{nR} \) messages together into \( 2^{n(R-L)} \) equal-size groups and map each group into a single codeword. The encoder sends the group index \( k \in [1 : 2^{n(R-L)}] \) by transmitting \( x^n(k) \) for \( m \in [(k-1)2^{nL} + 1 : k2^{nL}] \). The decoder finds the correct group index \( \hat{k} \) and simply forms the list of messages associated with \( \hat{k} \), i.e., \( L = [(\hat{k} - 1)2^{nL} + 1 : \hat{k}2^{nL}] \).

Finally, by the channel coding theorem for the standard DMC, the group index can be recovered if \( R - L < C \), which completes the proof of achievability.

(b) Note that we have a new definition of error in this problem. An error occurs if \( M \notin L(Y^n) \).

Define an error random variable \( E \) which takes value 1 if there is an error, and 0 otherwise. Then,

\[
H(M|Y^n) = H(M|Y^n) + H(E|M, Y^n)
= H(E|Y^n) + H(M|E, Y^n)
\leq H(E) + H(M|E, Y^n)
\leq 1 + P\{E = 0\}H(M|E = 0, Y^n) + P\{E = 1\}H(M|E = 1, Y^n)
\leq 1 + (1 - P_e^n) \log 2^{nL} + P_e^n \log(2^{nR} - 1)
\leq 1 + nL + nR \epsilon_e^n = nL + n \epsilon_e^n,
\]

This implies that given any sequence of \( (2^{nR}, 2^{nL}, n) \) codes with \( P_e^n \to 0 \) as \( n \to \infty \), we have

\[
nR = H(M)
\leq I(M; Y^n) + H(M|Y^n)
\leq I(M; Y^n) + nL + n \epsilon_e^n
\leq nC + nL + n \epsilon_e^n,
\]

where (a) follows by the proof of the converse for DMC. Therefore, \( R \leq C + L \).

10.1. (a) The optimal rate is \( R^* = H(X|Y) \).

(b) Since both the encoder and the decoder know the side information, the encoder only needs to describe the sequences \( X^n \) such that \( X^n \in T^{(n)}_\epsilon(X|y^n) \) for each observed \( y^n \) sequence, which requires \( n(H(X|Y) + \delta(\epsilon)) \) bits. An error occurs only if \( (X^n, Y^n) \notin T^{(n)}_\epsilon(X, Y) \). By the LLN, the probability of this event tends to zero as \( n \to \infty \).
(c) Consider

\[ nR \geq H(M) \]
\[ \geq I(M; X^n|Y^n) \]
\[ = H(X^n|Y^n) - H(X^n|M, Y^n) \]
\[ \geq \sum_{i=1}^{n} H(X_i|Y^n, X^{i-1}) - n\epsilon_n \]
\[ = \sum_{i=1}^{n} H(X_i|Y_i) - n\epsilon_n \]
\[ = nH(X|Y) - n\epsilon_n, \]

where (a) follows by Fano’s inequality.

(d) For distributed lossless source coding, suppose that a genie provides side information \( X_2^n \) to both encoder 1 and the decoder. From part (c), the rate required in this setting is \( R_1 \geq H(X_1|X_2) \), which establishes an upper bound on \( R_1 \) for the case without genie. Similarly, by introducing a genie providing side information \( X_1^n \) to both encoder 2 and the decoder, we have \( R_2 \geq H(X_2|X_1) \).

10.6. (a) By identifying \( X_1 = (X, Y) \) and \( X_2 = Y \) in the distributed lossless source coding setting, the optimal rate region is the set of rate pairs \((R_1, R_2)\) such that

\[ R_1 \geq H(X, Y|Y), \]
\[ R_2 \geq H(Y|X, Y), \]
\[ R_1 + R_2 \geq H(X, Y), \]

or equivalently,

\[ R_1 \geq H(X|Y), \]
\[ R_1 + R_2 \geq H(X, Y). \]

(b) The optimal rate region is the set of rate pairs \((R_1, R_2)\) such that

\[ R_1 + R_2 \geq H(Y). \]

This can be achieved by ignoring \( X^n \) at encoder 1. For the converse, consider

\[ n(R_1 + R_2) \geq H(M_1, M_2) \]
\[ \geq I(Y^n; M_1, M_2) \]
\[ \geq H(Y^n) - n\epsilon_n \]
\[ = nH(Y) - n\epsilon_n, \]

where (a) follows by Fano’s inequality.

10.8. The optimal rate is

\[ R^* = \max\{H(X|Y_1), H(X|Y_2)\}. \]

This can be achieved using Slepian–Wolf coding (recall that random binning does not depend on the
side information or its distribution). For the converse, consider

\[ nR \geq H(M) \]
\[ \geq I(X^n; M|Y^n) \]
\[ \geq H(X^n|Y^n) - n\epsilon_n \]
\[ \geq \sum_{i=1}^{n} H(X_i|Y_{ji}) - n\epsilon_n \]
\[ = nH(X|Y_j) - n\epsilon_n \]

for \( j = 1, 2 \).