Problem 1 \hspace{1cm} \textit{Rate Distortion with Two Distortion Functions}

a) We first outline the proof of the direct part. Fix a conditional distribution \( P_{X|X} \) that achieves the minimum in

\[
\rho(D_1, D_2) \triangleq \min_{P_{X|X}(\cdot|\cdot)} \mathbb{I}(X; \hat{X}).
\]

Compute the induced marginal distribution on \( \hat{X} \),

\[
P_{\hat{X}}(\hat{x}) = \sum_x P_X(x) P_{X|\hat{X}}(\hat{x}|x), \quad \hat{x} \in \hat{X}.
\]

Generate a codebook of size \( |2^n R| \) at random by drawing each symbol of each codeword IID according to the marginal \( P_{\hat{X}} \). To describe a source sequence \( x^n \) the encoder tries to find a codeword \( \hat{x}^n(i) \) such that \( (x^n, \hat{x}^n(i)) \in T_1(n)(P_{X,\hat{X}}) \) and then sends the smallest such index \( i \). When observing \( i \), the reconstructor produces \( \hat{x}^n(i) \). For a fixed codebook \( C_n \) let \( \mathcal{E}(C_n) \) denote the set of source sequences for which no jointly typical codeword can be found. For a fixed codebook, the expected distortions can then be written as

\[
\mathbb{E}[d_1(X^n, \hat{X}^n)] = \sum_{x^n \in X^n} P_{X^n}(x^n) d_1(x^n, g(f(x^n)))
\]

\[
= \sum_{x^n \in \mathcal{E}(C_n)} P_{X^n}(x^n) d_1(x^n, g(f(x^n))) + \sum_{x^n \in \mathcal{E}(C_n)} P_{X^n}(x^n) d_1(x^n, g(f(x^n)))
\]

\[
\leq d_{1, \max} P_{X^n}(\mathcal{E}(C_n)) + (1 + \epsilon) D_1.
\]

Similarly,

\[
\mathbb{E}[d_2(X^n, \hat{X}^n)] = \sum_{x^n \in X^n} P_{X^n}(x^n) d_2(x^n, g(f(x^n)))
\]

\[
= \sum_{x^n \in \mathcal{E}(C_n)} P_{X^n}(x^n) d_2(x^n, g(f(x^n))) + \sum_{x^n \in \mathcal{E}(C_n)} P_{X^n}(x^n) d_2(x^n, g(f(x^n)))
\]

\[
\leq d_{2, \max} P_{X^n}(\mathcal{E}(C_n)) + (1 + \epsilon) D_2.
\]

It thus suffices to show that if \( R > \rho(D_1, D_2) \), then there exists a sequence of codebooks with \( P_{X^n}(\mathcal{E}(C_n)) \to 0 \). To this end, we will show that \( P_{X^n}(\mathcal{E}(C_n)) \) averaged over all realizations of the codebook is small:

\[
\sum_{C_n} P(C_n) P_{X^n}(\mathcal{E}(C_n)) = \sum_{C_n} P(C_n) \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) 1\{x^n \in \mathcal{E}(C_n)\}
\]

© Amos Lapidoth, 2015/2016
\[
\sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) \sum_{C_n} P(C_n) I\{x^n \in \mathcal{E}(C_n)\} = \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) \Pr[x^n \in \mathcal{E}(C_n)] = \sum_{x^n \in \mathcal{T}_e(n)(P_X)} P_{X^n}(x^n) \Pr[x^n \in \mathcal{E}(C_n)] + \sum_{x^n \not\in \mathcal{T}_e(n)(P_X)} P_{X^n}(x^n) \Pr[x^n \in \mathcal{E}(C_n)] \leq \sum_{x^n \in \mathcal{T}_e(n)(P_X)} P_{X^n}(x^n) \Pr[x^n \in \mathcal{E}(C_n)] + \Pr[X^n \notin \mathcal{T}_e(n)(P_X)].
\]

The second term vanishes as \(n \to \infty\) by the law of large numbers. As for the first term, if \(x^n \in \mathcal{T}_e(n)(P_X)\), then the probability that a sequence \(\hat{X}^n\) drawn IID from \(P_X\) is strongly jointly typical with \(x^n\) is at least \(2^{-n(I(X;\hat{X})+\epsilon_1)}\) where \(\epsilon_1 \to 0\) as \(\epsilon \to 0\). Consequently, for \(x^n \in \mathcal{T}_e(n)(P_X)\),

\[
\Pr[x^n \in \mathcal{E}(C_n)] \leq \left(1 - 2^{-n(I(X;\hat{X})+\epsilon_1)}\right)2^{nR} \leq \exp\left(2^{n(R-I(X;\hat{X})-\epsilon_1)}\right),
\]

where we used the basic inequality \(1 - \xi \leq e^{-\xi}\). Thus, if \(R > I(X;\hat{X}) + \epsilon_1\), then

\[
\sum_{C_n} P(C_n)P_{X^n}(\mathcal{E}(C_n)) \to 0, \quad n \to \infty.
\]

This proves the existence of a sequence of codes \(\{C_n\}_{n=1}^\infty\) of rate \(R\) arbitrarily close to \(I(X;\hat{X})\) with \(P_{X^n}(\mathcal{E}(C_n)) \to 0\) as \(n \to \infty\).

To prove the converse part, we first show that \(\rho(D_1, D_2)\) is convex in the pair \((D_1, D_2)\). Indeed, if \(P_{X|X}^{(1)}\) achieves \(\rho(D_1^{(1)}, D_2^{(1)})\) and \(P_{X|X}^{(2)}\) achieves \(\rho(D_1^{(2)}, D_2^{(2)})\), then \(\lambda P_{X|X}^{(1)} + \lambda P_{X|X}^{(2)}\) must satisfy

\[
E[d_1(X, \hat{X})] \leq \lambda D_1^{(1)} + \lambda D_1^{(2)},
E[d_2(X, \hat{X})] \leq \lambda D_2^{(1)} + \lambda D_2^{(2)}.
\]

The convexity can then be verified as follows:

\[
\rho\left(\lambda D_1^{(1)} + \lambda D_1^{(2)}, \lambda D_2^{(1)} + \lambda D_2^{(2)}\right) \leq I\left(P_X, \lambda P_{X|X}^{(1)} + \lambda P_{X|X}^{(2)}\right) \leq \lambda I\left(P_X, P_{X|X}^{(1)}\right) + \lambda I\left(P_X, P_{X|X}^{(2)}\right) = \lambda \rho\left(D_1^{(1)}, D_2^{(1)}\right) + \lambda \rho\left(D_1^{(2)}, D_2^{(2)}\right).
\]

Suppose now that there exists a code that achieves

\[
E[d_1(X^n, \hat{X}^n)] \leq D_1, \quad E[d_2(X^n, \hat{X}^n)] \leq D_2.
\]

We have

\[
nR \geq I(X^n; \hat{X}^n)
\]
\[ = H(X^n) - H(X^n | \hat{X}^n) \]
\[ = \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i | \hat{X}^n, X^{i-1}) \]
\[ \geq \sum_{i=1}^{n} I(X_i; \hat{X}_i) \]
\[ \geq n \sum_{i=1}^{n} \frac{1}{n} \rho \left( E[d_1(X_i, \hat{X}_i)], E[d_2(X_i, \hat{X}_i)] \right) \]
\[ \geq n \rho(D_1, D_2), \]

where the last line follows from the convexity of \( \rho(D_1, D_2) \). This concludes the proof of the converse part of the theorem.

b) This result follows from the first part by defining \( \hat{X} = \hat{X}_1 \times \hat{X}_2, \hat{X} = (\hat{X}_1, \hat{X}_2) \) and setting \( d_1(x, \hat{x}) = d_1(x, \hat{x}_1) \) and \( d_2(x, \hat{x}) = d_2(x, \hat{x}_2) \). The two reconstructors \( g_1^{(n)} \) and \( g_2^{(n)} \) can be viewed as the coordinates of a single reconstructor \( g^{(n)} : \{1, \ldots, 2^{nR}\} \rightarrow X_1 \times \hat{X}_2 \). The result from the first part can then be applied to show that \( (R, D_1, D_2) \) is achievable if, and only if,
\[ R \geq \min_{P_{X|X}(\cdot)} \min_{E[d_1(X, \hat{X})] \leq D_1, E[d_2(X, \hat{X})] \leq D_2} I(X; \hat{X}) = \min_{P_{X_1,X_2|X}(\cdot, \cdot)} \min_{E[d_1(X, \hat{X}_1)] \leq D_1, E[d_2(X, \hat{X}_2)] \leq D_2} I(X; \hat{X}_1, \hat{X}_2) \]

c) This follows by noting that if \( P_{X|X} \) achieves
\[ \min_{P_{X|X}(\cdot)} \min_{E[d_1(X, \hat{X})] \leq D_1, E[d_2(X, \hat{X})] \leq D_2} I(X; \hat{X}), \]
then
\[ P_{X_1,X_2|X}(\hat{x}_1, \hat{x}_2|x) = P_{X|X}(\hat{x}_1|x) I(\hat{x}_2 = \hat{x}_1) \]
satisfies \( E[d_1(X, \hat{X}_1)] \leq D_1 \) and \( E[d_2(X, \hat{X}_2)] \leq D_2 \), so that the minimum over all permissible \( P_{X_1,X_2|X}(\cdot, \cdot) \) can only be smaller.

**Problem 2**

*Source–Channel Separation with Feedback*

Consider a combined source-channel coding scheme which achieves expected average distortion \( D \). Such a scheme, together with the source distribution and the channel law, induces a distribution on \((U^k, X^n, Y^n, \hat{U}^k)\) satisfying that
\[ \frac{1}{k} \sum_{i=1}^{k} E[d(U_i, \hat{U}_i)] \leq D. \]

To lower-bound \( I(U^k; \hat{U}^k) \), we proceed similarly as in the proof of the converse part of the Rate-Distortion Theorem:
\[ I(U^k; \hat{U}^k) = H(U^k) - H(U^k | \hat{U}^k) \]
\[ = \sum_{i=1}^{k} H(U_i) - \sum_{i=1}^{k} H(U_i | \hat{U}^k, U^{i-1}) \]
\[ \geq \sum_{i=1}^{k} H(U_i) - \sum_{i=1}^{k} H(U_i | \hat{U}_i) \]
\[ = \sum_{i=1}^{k} I(U_i ; \hat{U}_i) \]
\[ \geq \sum_{i=1}^{k} R\left( \mathbb{E}\left[ d(U_i, \hat{U}_i) \right] \right) \]
\[ \geq kR(D), \]

where the last inequality follows from the convexity of \( R(D) \).

We then upper-bound \( I(U^k ; \hat{U}^k) \) starting with the data processing inequality:

\[ I(U^k ; \hat{U}^k) \leq I(U^k ; Y^n) \]
\[ = H(Y^n) - H(Y^n | U^k) \]
\[ = H(Y^n) - \sum_{i=1}^{n} H(Y_i | Y^{i-1}, U^k) \]
\[ = H(Y^n) - \sum_{i=1}^{n} H(Y_i | Y^{i-1}, U^k, X_i) \]
\[ = H(Y^n) - \sum_{i=1}^{n} H(Y_i | X_i) \]
\[ \leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i | X_i) \]
\[ = \sum_{i=1}^{n} I(X_i ; Y_i) \]
\[ \leq nC. \]

Thus we obtain
\[ kR(D) \leq nC, \]

which concludes the proof.

**Problem 3**

*Source-Channel Separation and Average Bit-Error Probability*

a) The desired distortion function is

\[ d(u, \hat{u}) = \begin{cases} 0 & \text{if } u = \hat{u}, \\ 1 & \text{if } u \neq \hat{u}. \end{cases} \]

b) By Problem 2 we can achieve expected average distortion

\[ \frac{1}{k} \sum_{i=1}^{k} \mathbb{E}[d(U_i, \hat{U}_i)] \leq D \]

only if

\[ R(D) \leq \frac{n}{k}C. \]
Using the distortion function of Part a) and recalling that the rate distortion function for a \( \text{Ber}(1/2) \) source with Hamming distortion is given by

\[
R(D) = \begin{cases} 
1 - H_b(D) & \text{if } 0 \leq D \leq \frac{1}{2}, \\
0 & \text{otherwise},
\end{cases}
\]

we obtain the desired result that

\[
\frac{1}{k} \sum_{i=1}^{k} \Pr[U_i \neq \hat{U}_i] \geq H_b^{-1}\left(1 - \frac{n}{k} C\right) > 0.
\]