ECE 788: Network Information Theory
Assignment 4 (due on Oct. 5)

1. We want to prove that the rate-distortion function $R(D)$ is convex in D, i.e., that

$$R(\lambda D_1 + (1-\lambda)D_2) \leq \lambda R(D_1) + (1-\lambda)R(D_2),$$

with $D = \lambda D_1 + (1-\lambda)D_2$ and $0 \leq \lambda \leq 1$ (check that this condition coincides with the geometric interpretation of convexity).
1.a. Argue that proving the above is the same as proving that we can find a compression/ decompression scheme that achieves distortion $D = \lambda D_1 + (1-\lambda)D_2$ with rate $\lambda R(D_1) + (1-\lambda)R(D_2)$ (Hint: The rate-distortion function $R(D)$ is the *minimum* rate required to compress to distortion D, i.e., it is the performance of the best strategy).
So now let us find such scheme. We start with the two schemes that achieve $R(D_1)$ and $R(D_2)$, respectively. We refer to them as scheme 1 and scheme 2, respectively. We show that by *time-sharing* between the two schemes, we can achieve $D = \lambda D_1 + (1-\lambda)D_2$ with rate $\lambda R(D_1) + (1-\lambda)R(D_2)$.
1.b. Consider scheme 1 (that achieves $R(D_1)$). Assume that we use such scheme for the first λn symbols (rounding λn to the closest integer is in principle necessary) of the source vector X^n, while for the remaining $(1-\lambda)n$ symbols we accept distortion $d_{\max}$. Show that the average distortion of this scheme is $\lambda D_1 + (1-\lambda)d_{\max}$ and that the rate is $\lambda R(D_1)$ bits/ symbol (Hint: Use the definition of distortion $\frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i)$ and break the summation into the two sets of symbols. Note that $\frac{1}{\lambda n}\sum_{i=1}^{\lambda n} d(x_i, \hat{x}_i) = D_1$ by definition of scheme 1)
1.c. Now consider a strategy that applies scheme 1 for the first λn symbols and scheme 2 for the remaining $(1-\lambda)n$ symbols. What is the distortion achieved by this scheme? What is the rate? This should conclude the proof!

2. Consider a bynary symmetric source (BSS), i.e., $\mathcal{X} = \{0, 1\}$ and $P_X(0) = 1/2$ and Hamming distortion ($d^n(x^n, \hat{x}^n) = 1/n \sum_{i=1}^{n} d(x_i, \hat{x}_i)$ with $d(x_i, \hat{x}_i) = 1\{x_i \neq \hat{x}_i\}$). Focusing on the multiple description problem, we want to achieve central distortion $D_{12} = 0$ with $R_1 = 0.5$ and $R_2 = 0.5$. Propose a simple scheme that satisfies these conditions and evaluate the side distortions D_1 and D_2 (Hint: Think about the compression scheme used for the binary erasure source. You should find that $D_1 = D_2 = 0.25$).

3. In the proof of achievability for the multiple description problem, the key step is the evaluation of a lower bound on $E[S^n]$ and an upper bound on $var(S^n)$ (recall the discussion in class), where $S^n = \sum_{w_1,w_2} S^n_{w_1,w_2}$ and $S^n_{w_1,w_2} = 1\{\mathcal{E}_{w_1,w_2}\}$ with

$$\mathcal{E}_{w_1,w_2} = \{(x^n, \hat{X}_1^n(w_1), \hat{X}_2^n(w_2), \hat{X}_{12}^n(w_1, w_2)) \in T_\epsilon^n(P_{X\hat{X}_1\hat{X}_2\hat{X}_{12}})\}$$

for $x^n \in T_\epsilon^n(P_X)$. Let us calculate a lower bound on $E[S^n]$. Notice that the expectation is taken with respect to the codebook distribution according to the random coding approach.
3.a. Argue that

$$E[S^n] = 2^{n(R_1+R_2)} \Pr[\mathcal{E}_{w_1,w_2}].$$

(Hint: Use the symmetry of the codebook construction)

3.b. Show that

$$\Pr[\mathcal{E}_{w_1,w_2}] = \sum_{T_\epsilon^n(P_{X\hat{X}_1\hat{X}_2\hat{X}_{12}}|x^n)} P^n_{\hat{X}_1}(\hat{x}_1^n)P^n_{\hat{X}_2}(\hat{x}_2^n)P^n_{\hat{X}_{12}|\hat{X}_1\hat{X}_2}(\hat{x}_{12}^n|\hat{x}_1^n,\hat{x}_2^n),$$

where as usual we defined $P^n_{\hat{X}_1}(\hat{x}_1^n) = \prod_{i=1}^n P_{\hat{X}_1}(\hat{x}_{1i})$ and similarly for the other probabilities. (Hint: Recall how the codebook was created)

3.c. Show that neglecting terms of the order of ϵ or terms that go to zero for n large we have the following bound

$$\Pr[\mathcal{E}_{w_1,w_2}] \geq 2^{n(H(\hat{X}_1\hat{X}_2\hat{X}_{12}|X)-H(\hat{X}_1)-H(\hat{X}_2)-H(\hat{X}_{12}|\hat{X}_1,\hat{X}_1))}.$$

What theorems should you use to prove the above? (Hint: Bound each term in $\Pr[\mathcal{E}_{w_1,w_2}]$ separately using the fact that the sum is over typical sequences only)