1. **Strong converse for source coding.** Given a sequence of \( (2^nR, n) \) lossless source codes with \( R < H(X) \), show that \( P_e^{(n)} \to 1 \) as \( n \to \infty \). (Hint: A \( (2^nR, n) \) code can represent only \( 2^nR \) points in \( X^n \). Using typicality, show that if \( R < H(X) \), the probability of these \( 2^nR \) points converges to zero, no matter how we choose them.)

2. **Rate-distortion function.** Consider the lossy source coding for a DMS \( X \) with distortion measure \( d \).
   
   (a) Using the operational definition, show that the rate–distortion function \( R(D) \) is nonincreasing and convex for \( D \geq 0 \).
   
   (b) Show that the information rate–distortion function \( \hat{R}(D) \) is nonincreasing, convex, and continuous for \( D \geq 0 \).

3. **Bounds on the rate distortion function for squared error distortion.** Let \( X \) be an arbitrary memoryless (stationary) source with variance \( P \), and let \( d(x, \hat{x}) = (x - \hat{x})^2 \). Show that the rate distortion function is bounded as
   
   \[
   \frac{1}{2} \log(2\pi eD) \leq R(D) \leq \frac{1}{2} \log \left( \frac{P}{D} \right)
   \]
   
   with equality if \( X \) is a WGN(\( P \)) source. (The lower bound is referred to as the Shannon lower bound.) Is the Gaussian source harder or easier to describe than other sources with the same variance? (Hint: For the upper bound, consider \( \hat{X} = (P - D)X/P + Z \), where \( Z \sim N(0, D(P - D)/P) \) is independent of \( X \).)

4. **Lossy source coding from a noisy observation.** Let \((X, Y) \sim p(x, y)\) be a 2-DMS and let \( d(x, \hat{x}) \) be a distortion measure. Consider the following lossy source coding problem, in which the encoder has access to a noisy version \( Y \) of the source \( X \) instead of \( X \) itself.
   
   Unlike the regular lossy source coding setup, the encoder maps each \( y^n \) sequence to an index \( m \in [1 : 2^nR] \). Otherwise, the definitions of \( (2^nR, n) \) codes, achievability, and rate distortion function are the same as before.
   
   Let \( D_{\min} = \min_{\hat{x}(y)} \mathbb{E}[d(X, \hat{x}(Y))] \). Show that the rate distortion function for this setting is
   
   \[
   R(D) = \min_{p(\hat{x}|y): \mathbb{E}[d(X, \hat{X})] \leq D} I(Y; \hat{X}) \quad \text{for} \ D \geq D_{\min}.
   \]
   
   (Hint: Define a new distortion measure \( d'(y, \hat{x}) = \mathbb{E}(d(X, \hat{x})|Y = y) \), and show that
   
   \[
   \mathbb{E}[d(X^n, \hat{x}^n(m(Y^n)))] = \mathbb{E}[d'(Y^n, \hat{x}^n(m(Y^n)))].
   \]

5. **To code or not to code.** Consider a WGN(\( P \)) source \( U \) and an AWGN channel with output \( Y = gX + Z \), where \( Z \) in WGN(1), and average power constraint \( P \) on the input \( X \). We wish to send the source over the channel at rate \( r = 1 \) symbol/transmission with the smallest possible squared error distortion.
   
   (a) Find the minimum mean squared error distortion achieved using separate source and channel coding.
   
   (b) Find the squared error distortion achieved when the sender at time \( i \) simply sends \( X_i = U_i \), i.e., performs no coding, and the receiver uses the best linear MSE estimate \( \hat{U}_i \) of \( U_i \) given \( Y_i \). Compare this to the distortion in part (a) and comment on the results.
6. Capacity of multiple access channels.

(a) Consider the binary multiplier MAC in Example 4.1. We established the capacity region using time division between two individual capacities. Show that the capacity region can be also expressed as the union of \( \mathcal{C}(X_1, X_2) \) sets (with no time-sharing/convexification), specify the set of pmfs \( p(x_1)p(x_2) \) on \( (X_1, X_2) \) that achieve the boundary of the region.

(b) Find the capacity region for the modulo-2 additive MAC, where \( X_1, X_2, \) and \( Y \) are binary, and \( Y = X_1 \oplus X_2 \). Again show that the capacity region can be expressed as the union of \( \mathcal{C}(X_1, X_2) \) sets and therefore time-sharing is not necessary.

(c) The capacity regions for the above two examples and for the Gaussian MAC can all be expressed as union of \( \mathcal{C}(X_1, X_2) \) sets and no time sharing is necessary. Is time sharing ever necessary? Find the capacity of the push-to-talk MAC with binary inputs and output, given by \( p(0,0) = p(1,0) = 1 \) and \( p(0,1) = 0.5 \). Why is this channel called “push-to-talk”? Show that the capacity region cannot be completely expressed as the union of \( \mathcal{C}(X_1, X_2) \) sets and that time sharing (convexification) is necessary.

7. Maximum vs. average probability of error. We proved the channel coding theorem for a discrete memoryless channel under the average probability of error criterion. It is straightforward to show that the capacity is the same if we instead deal with the maximum probability of error. If we have an \((2^R, n)\) code with \( P_e^{(n)} \leq \epsilon \), then at least half its codewords have \( \lambda_m \leq 2\epsilon \) (Markov inequality) and this half has a rate \( R - 1/n \to R \). Such an argument cannot be used to show that the capacity of an arbitrary MAC under the maximum probability of error criterion is the same as that with the average probability of error criterion.

(a) Argue that simply discarding half of the codeword pairs with the highest probability of error does not work in general.

(b) How about throwing out the worst half of each sender’s codewords? Show that this does not work either. (Hint: Provide a simple example of a set of probabilities \( p_{ij} \in [0, 1], (i, j) \in [1 : n] \times [1 : n] \), with \( 1/n^2 \sum_{i,j} p_{ij} \leq \epsilon \), for some \( 0 < \epsilon < 1/4 \), such that there are no subsets \( I, J \subseteq [1 : n] \) with cardinalities \( |I|, |J| \geq n/2 \) satisfying \( p_{ij} \leq 4\epsilon \) for all \((i, j) \in I \times J \).

Remark: A much stronger statement holds. Dueck (1978) provided an example of a MAC for which the capacity region with maximum probability of error is strictly smaller than that with average probability of error. This gives yet another example in which a result from single-user information theory does not necessarily carry over to the multiple-user case.

8. Convex closure of the union of sets. Let \( \mathcal{R}_1 = \{(r_1, r_2) : r_1 \leq I_1, r_2 \leq I_2, r_1 + r_2 \leq I_{12}\} \) for some \( I_1, I_2, I_{12} \geq 0 \) and \( \mathcal{R}_1' = \{(r_1, r_2) : r_1 \leq I_1', r_2 \leq I_2', r_1 + r_2 \leq I_{12}'\} \) for some \( I_1', I_2', I_{12}' \geq 0 \). Let \( \mathcal{R}_2 \) be the convex closure of the union of \( \mathcal{R}_1 \) and \( \mathcal{R}_1' \), and \( \mathcal{R}_3 = \{(r_1, r_2) : r_1 \leq \alpha I_1 + \alpha I_1', r_2 \leq \alpha I_2 + \alpha I_2', r_1 + r_2 \leq \alpha I_{12} + \alpha I_{12}' \} \) for some \( \alpha \in [0, 1] \).

(a) Show that \( \mathcal{R}_3 \subseteq \mathcal{R}_2 \).

(b) Provide a counterexample showing that the inclusion can be strict.

(c) Under what condition on \( I_1, I_2, I_{12}, I_1', I_2', I_{12}' \) are \( \mathcal{R}_2 = \mathcal{R}_3 \)?

(Hint: Consider the case \( I_1 = 2, I_2 = 5, I_{12} = 6, I_1' = 1, I_2' = 4, \) and \( I_{12}' = 7 \).

9. Cooperative capacity of a MAC. Consider a DM-MAC \((X_1 \times X_2, P(y|x_1, x_2), \mathcal{Y})\). Assume that both senders have access to both messages \( m_1 \in [2^{nR_1}] \) and \( m_2 \in [2^{nR_2}] \), thus the codewords \( x_1^m(m_1, m_2) \) and \( x_2^m(m_1, m_2) \) can depend on both messages.

(a) Find the capacity region.

(b) Evaluate the region for the AWGN-MAC with noise power 1 and power constraints \( P_1 \) and \( P_2 \).