Problem 1. Let \( \{f_i : \mathbb{R} \to \mathbb{R}\}_{1 \leq i \leq n} \) be a set of convex functions on \( \mathbb{R} \) and \( c_i \geq 0 \) for all \( i \in \{1, 2, \ldots, n\} \).

(a) Show that the function \( f : x \mapsto \sum_{i=1}^{n} c_i f_i(x) \) is convex.

(b) Show that the function \( g : (x_1, x_2, \ldots, x_n) \mapsto \sum_{i=1}^{n} c_i f_i(x_i) \) is convex.

Problem 2. Let \( \{f_i(x)\}_{i \in I} \) be a set of convex real-valued functions defined over a convex domain \( D \). Assuming that \( f(x) = \sup_{i \in I} f_i(x) \) is finite for all \( x \in D \), show that \( f(x) \) is convex.

Problem 3. Let \( f : U \to \mathbb{R} \) be a convex function on \( U \) and assume that there exists \( a, b \in \mathbb{R} \) such that \( a \leq f(x) \leq b \) for all \( x \in U \). Let \( h \) be an increasing convex function defined on the interval \( [a, b] \). Show that the function \( g = h \circ f \) is convex on \( U \).

Problem 4. A function \( f(v) \) is defined on a convex region \( R \) of a vector space. Show that \( f(v) \) is convex iff the function \( f(\lambda v_1 + (1 - \lambda)v_2) \) is a convex function of \( \lambda \), \( 0 \leq \lambda \leq 1 \), for all \( v_1, v_2 \in R \).

Problem 5.

(a) Show that \( I(U;V) \geq I(U;V|T) \) if \( T, U, V \) form a Markov chain, i.e., conditional on \( U \), the random variables \( T \) and \( V \) are independent.

Fix a conditional probability distribution \( p(y|x) \), and suppose \( p_1(x) \) and \( p_2(x) \) are two probability distributions on \( X \).

For \( k \in \{1, 2\} \), let \( I_k \) denote the mutual information between \( X \) and \( Y \) when the distribution of \( X \) is \( p_k(\cdot) \).

For \( 0 \leq \lambda \leq 1 \), let \( W \) be a random variable, taking values in \( \{1, 2\} \), with
\[
\Pr(W = 1) = \lambda, \quad \Pr(W = 2) = 1 - \lambda.
\]

Define
\[
p_{W,X,Y}(w, x, y) = \begin{cases} 
\lambda p_1(x)p(y|x) & \text{if } w = 1 \\
(1 - \lambda)p_2(x)p(y|x) & \text{if } w = 2.
\end{cases}
\]

(b) Express \( I(X;Y|W) \) in terms of \( I_1, I_2 \) and \( \lambda \).

(c) Express \( p(x) \) in terms of \( p_1(x), p_2(x) \) and \( \lambda \).

(d) Using (a), (b) and (c) show that, for every fixed conditional distribution \( p_{Y|X} \), the mutual information \( I(X;Y) \) is a concave \( \cap \) function of \( p_X \).

Problem 6. Suppose \( Z \) is uniformly distributed on \( [-1, 1] \), and \( X \) is a random variable, independent of \( Z \), constrained to take values in \( [-1, 1] \). What distribution for \( X \) maximizes the entropy of \( X + Z \)? What distribution of \( X \) maximizes the entropy of \( XZ \)?
Problem 7. Show that among all non-negative random variables with mean \( \lambda \) the exponential random variable has the largest differential entropy. Hint: let \( p(x) = e^{-x/\lambda}/\lambda \) be the density of the exponential random variable and let \( q(x) \) be some other density with mean \( \lambda \). Consider \( D(q\|p) \) and mimic the proof in class for the maximal entropy of the Gaussian.

Problem 8. Consider an additive noise channel with input \( x \in \mathbb{R} \), and output

\[
Y = x + Z
\]

where \( Z \) is a real random variable independent of the input \( x \), has zero mean and variance equal to \( \sigma^2 \).

In this problem we prove in two different ways that the Gaussian channel has the smallest capacity among all additive noise channels of a given noise variance. Let \( \mathcal{N}_\sigma \) denote the Gaussian density with zero mean and variance \( \sigma^2 \).

**First Method**

Let \( X \) be a Gaussian random variable with zero-mean and variance \( P \). Let \( \mathcal{N}_P \) denote its density \( \mathcal{N}_P(x) = \frac{1}{\sqrt{2\pi P}} e^{-x^2/2P} \).

(a) Show that \( I(X;Y) = H(X) - H(X - \alpha Y | Y) \) for any \( \alpha \in \mathbb{R} \).

(b) Show that \( H(X - \alpha Y) \leq \frac{1}{2} \log 2\pi e E((X - \alpha Y)^2) \) for any \( \alpha \in \mathbb{R} \).

(c) Deduce from (a) and (b) that

\[
I(X;Y) \geq H(X) - \frac{1}{2} \log 2\pi e E((X - \alpha Y)^2)
\]

for any \( \alpha \in \mathbb{R} \).

(d) Show that \( E((X - \alpha Y)^2) \geq \frac{\sigma^2 P}{\sigma^2 + P} \) with equality if and only if \( \alpha = P / (P + \sigma^2) \).

(e) Deduce from (c) and (d) that

\[
I(X;Y) \geq \frac{1}{2} \log \left( 1 + \frac{P}{\sigma^2} \right)
\]

and conclude that the Gaussian channel has the smallest capacity among all additive noise channels of a given noise variance.

**Second Method**

(a) Denote the input probability density by \( p_X \). Verify that

\[
I(X;Y) = \int \int p_X(x)p_Z(y-x) \ln \frac{p_Z(y-x)}{p_Y(y)} \, dx \, dy \text{ nats.}
\]

where \( p_Y \) is the density of the output when the input has density \( p_X \).

(b) Now set \( p_X = \mathcal{N}_P \). Verify that

\[
\frac{1}{2} \ln(1 + P/\sigma^2) = \int \int p_X(x)p_Z(y-x) \ln \frac{\mathcal{N}_\sigma^2(y-x)}{\mathcal{N}_P+\sigma^2(y)} \, dx \, dy.
\]

(c) Still with \( p_X = \mathcal{N}_P \), show that

\[
\frac{1}{2} \ln(1 + P/\sigma^2) - I(X;Y) \leq 0.
\]

[Hint: use (a) and (b) and \( \ln t \leq t - 1 \).]

(d) Show that an additive noise channel with noise variance \( \sigma^2 \) and input power \( P \) has capacity at least \( \frac{1}{2} \log_2(1 + P/\sigma^2) \) bits per channel use. Conclude that the Gaussian channel has the smallest capacity among all additive noise channels of a given noise variance.