PROBLEM 1. 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 σ^2.

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

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

$$I(X;Y) = \iint p_X(x)p_Z(y-x)\ln\frac{p_Z(y-x)}{p_Y(y)}\,dxdy \quad \text{nats.}$$

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

$$\tfrac{1}{2}\ln\big(1+P/\sigma^2\big) = \iint p_X(x)p_Z(y-x)\ln\frac{\mathcal{N}_{\sigma^2}(y-x)}{\mathcal{N}_{P+\sigma^2}(y)}\,dxdy.$$

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

$$\tfrac{1}{2}\ln(1+P/\sigma^2) - I(X;Y) \le 0.$$

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

(d) Show that an additive noise channel with noise variance σ^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.

PROBLEM 2. A discrete memoryless channel has three input symbols: $\{-1,0,1\}$, and two output symbols: $\{1,-1\}$. The transition probabilities are

$$p(-1|-1) = p(1|1) = 1, \quad p(1|0) = p(-1|0) = 0.5.$$

Find the capacity of this channel with cost constraint β, if the cost function is $b(x) = x^2$.

PROBLEM 3. Consider a vector Gaussian channel described as follows:

$$Y_1 = x + Z_1$$
$$Y_2 = Z_2$$

where x is the input to the channel constrained in power to P; Z_1 and Z_2 are jointly Gaussian random variables with $E[Z_1] = E[Z_2] = 0$, $E[Z_1^2] = E[Z_2^2] = \sigma^2$ and $E[Z_1Z_2] = \rho\sigma^2$, with $\rho \in [-1,1]$, and independent of the channel input.

(a) Consider a receiver that discards Y_2 and decodes the message based only on Y_1. What rates are achievable with such a receiver?

(b) Consider a receiver that forms $Y = Y_1 - \rho Y_2$, and decodes the message based only on Y. What rates are achievable with such a receiver?

(c) Find the capacity of the channel and compare it to the part (b).

PROBLEM 4. In this problem we will show that a binary linear code contains 2^k codewords for some k. Suppose C is a binary linear code of block length n, that is, C is a non-empty set of binary sequences of length n with the property that if x and y are in C so is their modulo 2 sum. Consider the following algorithm.

(i) Initialize D to be the set that contains only the all-zero sequence.

(ii) If C does not contain any element not in D stop. Otherwise C contains an element x not in D. Form $D' = \{x + y : y \in D\}$.

(iii) Augment D to $D \cup D'$ where D' is found above, and go to step (ii).

(a) Show that the all-zero sequence is in C so that at the end of step (i) $D \subset C$. Note that initially $|D| = 1$ which is a power of 2.

(b) Show that if D is a linear subset of C and there is an x that is in C but not in D, then D' formed in (ii) is a subset of C. [The phrase "A is a linear subset of B" means that A is a subset of B, and that if $x \in A$ and $y \in A$ then $x + y \in A$.]

(c) Under the assumptions of (b) show that D' is disjoint from D.

(d) Again under the assumptions of (b) show that D' has the same number of elements as D.

(e) Still under the assumptions of (b) show that $D \cup D'$ is a linear subset of C.

(f) Using parts (b), (c), (d) and (e) show that if at the beginning of step (ii) D is a linear subset of C, then at the end of step (iii) D is still a linear subset of C and it has twice as many elements as in the beginning. Conclude that when the algorithm terminates $D = C$ and the number of elements in D is a power of 2.

Note that the above algorithm also gives a generator matrix G for the code: Let $x_1, \ldots, x_k$ be the codewords that are picked at the successive stages of step (ii) of the algorithm. It then follows that each codeword in C can be written as a (unique) linear combination of these x_i's. Taking G as the matrix whose rows are the x_i's gives us the generator matrix.

PROBLEM 5. Consider appending an overall parity check to the codewords of Hamming code: Each codeword of a Hamming code is extended by 1 bit which is 0 if the codeword contains an even number of 1's and 1 if the codeword contains an odd number of 1's. For example, for the (7,4,3) Hamming code discussed in class, the codeword 0000000 becomes 00000000, the codeword 1110000 becomes 11100001, the codeword 1111111 becomes 11111111, etc. Show that this new code has minimum distance 4, can correct 1 error, and can detect 2 errors. This class of $(2^m, 2^m - m - 1, 4)$ codes are known as the "extended Hamming codes."

PROBLEM 6.

(a) Show that in a binary linear code, either all codewords contain an even number of 1's or half the codewords contain an odd number of 1's and half an even number.

(b) Let $x_{m,n}$ be the nth digit in the mth codeword of a binary linear code. Show that for any given n, either half or all of the $x_{m,n}$ are zero. If all of the $x_{m,n}$ are zero for a given n, explain how the code could be improved.

(c) Show that the average number of ones per codeword, averaged over all codewords in a linear binary code of blocklength N, can be at most $N/2$.