Homework Set #6

Due 19 November 2008, before 12:00 noon, INR 031/032/038

Problem 1 (MORE ON HAMMING CODES)

We have seen the $(7, 4)$ Hamming code in the class. This family of code exist for different parameters. We will study them in this problem.

Let $r \geq 2$ be an arbitrary integer. Consider all non-zero binary vectors of length r and list them as the column of a matrix, called $\mathbf{H}_r$.

(a) What is the size of matrix $\mathbf{H}_r$?

(b) Let t be the minimum number of columns of $\mathbf{H}_r$ which are linearly dependent, i.e., there exist columns indicated by $h_{i_1}, h_{i_2}, \ldots, h_{i_t}$ and binary numbers $c_1, c_2, \ldots, c_t$ such that

$$\sum_{j=1}^{t} c_j h_{i_j} = 0.$$

Prove that $t = 3$ for any $r \in \mathbb{N}$.

(c) Given the fact that $\mathbf{H}_r$ is a full-rank matrix, find the size of the right kernel of $\mathbf{H}_r$, that is the number of vectors $\mathbf{x}$ which satisfy $\mathbf{H}_r \cdot \mathbf{x} = \mathbf{0}$.

(d) Define the code $\mathcal{C}_r$ as the set of set of vectors $\mathbf{x}$ is part (c). Show that this code is linear. Find the codeword length, dimension, minimum distance and the rate of the code.

We are given T coins out of which one is fake, *i.e.*, whose weight is different than the others. We also have a set of detectors which are able to tell us whether the fake coin is in a set (chosen by us) or not (See Problem 2 of Homework 3 for more details).

(e) Let $T = 2^k$. How many times should we use a detector to find the fake coin?

(f) Assume we have as many number of detectors as you wish, but only one of them, which we don't know, is broken, i.e., its answer might be true or false. How many times should we use the detectors to find the fake coin? What is your strategy to find the fake coin?

Problem 2 (NOISY TYPEWRITER)

You are using a typewriter (for younger audience: a machine people used for text editing before computers existed). To simplify things, we will assume that you are typing in English and only using letters and a space symbol, which we denote as "$-$". Hence, the input and output alphabets are $\mathcal{X} = \mathcal{Y} = \{a, b, \ldots, z, -\}$. Since this machine is very old, it produces errors: in particular if you type the symbol x_i, it will produce one of the nearby symbols $x_{i-1}, x_i,$ or x_{i+1} with equal probability, in a circular fashion (for example, typing "a" may result in "$-$", "a", or "b").

(a) What is the capacity of the noisy typewriter channel?

(b) The typewriter has a property that a specific output symbol can be observed given that a specific input symbol has been used (e.g. if the output of the typewriter is "b", you can be sure that the input is not "m"). You can use this to carefully restrict your input to a subset of the alphabet. What is the maximum number of input symbols that you can choose such that you can perfectly reconstruct the input symbol upon observing the output symbol?

(c) Assume now that all input symbols are equally likely. Devise a coding strategy that ensures that you can reconstruct the input given the output and compare this to the channel capacity. What can you conclude?

Problem 3 (REED-SOLOMON CODES)

A matrix in the following form has a special name, Vandermonde matrix.

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{k-1} & x_2^{k-1} & x_3^{k-1} & \cdots & x_n^{k-1} \end{pmatrix}$$

Let this Vandermonde matrix be a generator matrix for a linear code $\mathcal{C}$ and let x_i's be distinct elements of a finite field $\mathcal{X}$ with $n \leq |\mathcal{X}|$ and $k \leq n$.

(a) Is the generator matrix in the systematic form?

(b) What is the information sequence $(u_0, u_1, \cdots u_{k-1})$ encoded to?

(c) What is the minimum distance of $\mathcal{C}$?

(d) How many codewords of Hamming weight d are there in the codebook?

Problem 4

Consider a binary symmetric channel with cross probability p. Therefore, a binary sequence $\mathbf{x}$ is transmitted and the received sequence would be $\mathbf{y} = \mathbf{x} + \mathbf{z}$, where the elements of z are chosen independently at random according to the distribution $\Pr(z_i = 1) = 1 - \Pr(z_i = 0) = p$.

(a) What is the capacity of this channel?

(b) What is the typical Hamming weight of the noise sequences, that is the most probable w such that $w_H(\mathbf{z}) = w$?
Hint: Find $Pr(w_H(\mathbf{z}) = i)$ as a function of i, and determine when the function is increasing and decreasing.

(c) Assume that we use a t-ball bounded distance decoder which acts as follows. Given received $\mathbf{y}$, the decoder looks for codewords $\mathbf{x}$ within distance t of $\mathbf{y}$. If there is a unique such codeword, then the decoder maps $\mathbf{y}$ to $\mathbf{x}$. Otherwise, it declares an error. We want to design a code such that using the code and a t-ball bounded distance decoder we can correct any error of the typical weight obtained in part (b). Find the minimum required distance of the code.

(d) Consider codes $\mathcal{C}_n$ of increasing blocklength n and fixed rate r, *i.e.*, the code has $2^{\lfloor nr \rfloor}$ codewords. Let $d(\mathcal{C}_n)$ denote the minimum distance of the code and define $\delta = d(\mathcal{C}_n)/n$ as the normalized distance of the code. Recall the Gilbert-Varshamov bound studied in the class. Use the approximation $\sum_{i=1}^{d-1} \binom{n}{i} \simeq 2^{nH_2(d/n)}$ for large enough n, where $H_2(\cdot)$ is the binary entropy function. Rewrite the bound in terms of the rate r and the normalized minimum distance δ of the code.

(e) Compare the capacity of channel with the GV bound for $p = 1/4$.

(f) How much error can a capacity achieving code with a bounded distance decoder tolerate for arbitrary p? What is your conclusion about the role of minimum distance in performance of a code used to communicate over a noisy channel?