ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 24 Information Theory and Coding

Homework 10 December 11, 2007

PROBLEM 1. Consider a binary block code with M codewords and blocklength n. That is, each codeword is sequence of n bits. Suppose this code can correct up to (and including) e errors.

(a) For a codeword x consider the set S of all binary sequences for which the decoder decides x. Show that

$$|S| \geq \sum_{i=0}^{e} \binom{n}{i}$$

(b) Show that the number of codewords M satisfies

$$M \leq 2^n \Big/ \sum_{i=0}^{e} \binom{n}{i}$$

This is known as the "sphere packing bound."

(c) In the class we showed that Hamming codes can correct up to 1 error, and for blocklength $2^m - 1$ they contained 2^{2^m-m-1} codewords. Show that Hamming codes satisfy the sphere packing bound with equality. Thus, for these blocklengths they are the highest rate single error correcting codes.

PROBLEM 2.

(a) Consider the following method of constructing a binary block code of minimum distance d and block length n:

(i) Start with the list of all 2^n binary sequences as potential codewords, and an empty list of chosen codewords.

(ii) While the list of potential codewords is not empty, pick any of its members and add it to the list of chosen codewords. Remove it and all sequences that are distance less than $d-1$ or less from it from the list of potential codewords.

(iii) The constructed code is the set of chosen codewords.

Argue that this procedure will yield a code of minimum distance at least d.

(b) Show that at each iteration of step (ii) one codeword is added to the list of chosen codewords, and at most $\sum_{i=0}^{d-1} \binom{n}{i}$ sequences are deleted from the list of potential codewords. Conclude that the number of chosen codewords M satisfies

$$M \geq 2^n \Big/ \sum_{i=0}^{d-1} \binom{n}{i}$$

This is known as the "Gilbert-Varshamov bound" (after Ed Gilbert and R.R. Varshamov who independently discovered it).

Problem 3. Consider appending an overall parity check to the codewords of Hamming code: Each codewords 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 and odd number of 1's. For example 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 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' found above to D 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 for the 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 columns are the x_i's gives us the generator matrix.

Problem 5.

(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 code word, averaged over all codewords in a linear binary code of blocklength N, can be at most $N/2$.

PROBLEM 6. The weight of a binary sequence of length N is the number of 1's in the sequence. The Hamming distance between two binary sequences of length N is the weight of their modulo 2 sum. Let $\mathbf{x}_1$ be an arbitrary codeword in a linear binary code of block length N and let $\mathbf{x}_0$ be the all-zero codeword. Show that for each $n \leq N$, the number of codewords at distance n from $\mathbf{x}_1$ is the same as the number of code words at distance n from $\mathbf{x}_0$. [Hint: show that it is smaller and larger.]