MA 552, Spring 2011

Topic 1. Hamming codes

(1) *Do not submit anything for this problem in written. This problem will be stated again later in this problem set in more appropriate terms. It is however possible to solve this problem without any "science".*

(a) Alice and Bob play a game. Alice has a secret integer number from 0 to 10, Bob wants to guess it. Bob can ask yes/no questions, and Alice is allowed to lie once (at a moment of her choice). Show that Bob can succeed in 7 questions.

(b) Suppose you need to transmit 4 binary bits of data over a not entirely reliable channel. More specifically, there is a chance that 1 bit in a transmission will be corrupted (the possibility that more than 1 bit can be corrupted is neglected). Show that you can transmit a message that consists of 7 bits so that the receiving party will be able to recover the original message regardless of where the corruption occurred and whether it occurred.

These two problems are described by the same mathematical setting. Consider a set $W_N = \{0, 1\}^N$, the set of all possible sequences of 0's and 1's of length N. A subset $C \subseteq W_N$ is called a *code*. We say that code C *corrects a single error* if any two elements $c, c' \in C$ differ at least in 3 places. Elements of C are interpreted as messages.

The term "single error correcting code" is explained by the following consideration: if after transmission one (or none) digit of a message $c \in C$ gets altered, the result still differs from any other $c' \in C$ at least in 2 places, and differs from original c at most in 1 place, so if the receiving party knows that only elements of C are possible messages, they can recover the original message (thus "correcting a single error" in a transmission).

The number $|C|$ of elements in C is the number of different messages that can be transmitted using C. Naturally, the closer $|C|$ is to $|W_N| = 2^N$ the more efficient communication is. For example, one can suggest the following code C (called "triple repetition code"): all sequences of $0, 1$ where each digit is repeated 3 times (assume here that N is a multiple of 3). If one digit in a message $c \in C$ gets altered, it is still easy to recover c. However, this is an example of highly inefficient code: $|C| = 2^{N/3}$, so "useful information" is only $N/3$ bits long, compared to N bits in the whole message.

Problems (1a) and (1b) suggest to find $C \subseteq W_7$ with $|C| \geq 11$ and $|C| \geq 16$, respectively.

Search for efficient error correcting (or detecting) codes is an established and well-developed branch of coding theory. Nearly every "real-life" massive data transmission (cell phones, satellite broadcasting, computer networks, data compression, read/write operations with any kind of memory in a computer, data storage) includes some kind of error control. The particular code introduced in the problems below, despite being one of the earliest and simplest examples, is actually used in certain types of RAM.

While generally finding a "good" $|C|$ is a combinatorial question, significant progress has been made employing linear algebra techniques.

For the rest of this topic, fix field F to be $\mathbb{Z}_2$, the field on two elements $0, 1$ with usual arithmetic operations modulo 2. (By the way, does not hurt to check that it really is a field.)

(2) Let V be a vector space over F of dimension N. Prove that V has precisely 2^N elements.

—— *see next page* ——

Let $c, c' \in W_N$. Let $c = (c_1, c_2, \ldots, c_N)$, $c' = (c'_1, c'_2, \ldots, c'_N)$. We say that *Hamming distance* between c and c' is the number

$$d_H(c, c') = |\{1 \le i \le N \;:\; c_i \ne c'_i\}|,$$

that is, $d_H(c, c')$ is the number of places where c and c' differ.

Then our condition on elements of C can be reformulated as follows: we are looking for $C \subseteq W_N$ such that

$$\min_{c, c' \in C, c \ne c'} d_H(c, c') \ge 3.$$

The value $d_H(C) = \min_{c, c' \in C, c \ne c'} d_H(c, c')$ is called the *minimum Hamming distance* of code C.

From now on, we treat W_N as an N-dimensional vector space over F, with coordinate-wise operations.

(3) Let $c, c' \in W_N$. Prove that $d_H(c, c') = d_H(0, c - c')$.

(4) Suppose C is a linear subspace of W_N. Prove that

$$\min_{c, c' \in C, c \ne c'} d_H(c, c') = \min_{c \in C, c \ne 0} d_H(0, c).$$

Now, instead of looking for an arbitrary $C \subseteq W_N$, we restrict ourselves to those C which are linear subspaces. The problem above explains that in such case, it is much easier to control the minimum Hamming distance of C: one only needs to find what is the minimum number of 1's in a non-zero element of C. (On the other hand, we may miss out on some "good" non-linear subsets C. Let's not worry about it right now.)

(5) In problem 1a, Bob needs to come up with a $C \subseteq W_7$ such that $|C| \ge 11$. What is the minimal possible dimension of a linear subset C of W_7 such that $|C| \ge 11$? What is the minimal possible dimension of a linear subset C of W_N such that $|C| \ge n$?

Next idea is as follows. Since we want to deal with linear subsets of W_N, why don't we describe C by the corresponding linear system.

(6) Let $H \in F^{m \times N}$ be an $m \times N$ matrix over $F = \mathbb{Z}_2$, let X be a column of variables $(x_1, \ldots, x_N)$. Denote $C_H \subseteq W_N$ to be the linear space of solutions of the system

$$HX = 0.$$

(a) In terms of columns of matrix H, describe when $HX = 0$ has a solution c with $d_H(0, c) = 1$. That is, answer the following question: suppose for some X of the form

$$X = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

we have

$$H \cdot \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

What can one say about columns of H?

—— *see next page* ——

(b) In terms of columns of matrix H, describe when $HX = 0$ has a solution c with $d_H(0, c) = 2$.

(7) If columns of C have none of two properties obtained in the previous problem, then any non-zero solution c of the system $HX = 0$ is at least distance 3 from zero: $d_H(0, c) \geq 3$. For a fixed m, what is the largest possible N? (Hint: answer is $N = 2^m - 1$.) A matrix H with the largest possible N is denoted H_m. (Particular choice of H_m does not matter at this point.)

(8) Write matrix H_3.

(9) What is the rank of H_m? (Hint: the answer is $\operatorname{rank} H = m$.)

(10) Find $\dim C_{H_m}$. Find m such that $N = 7$ and $\dim C_{H_m} = 4$.

Now for each m we have found a particular code $C = C_{H_m}$ and found $\dim C$. C_{H_m} is called $(\dim C_{H_m}, N)$ *Hamming code.* For example, solution to problem 1 is a $(4, 7)$ Hamming code. (Which corresponds to m that you found in the problem above.)

Compare N to $\dim C$. One can think of the number $N - \dim C$ in the following way: in a message $c \in C$ of length N there are $\dim C$ "useful" bits, and remaining $N - \dim C$ bits are redundant "checksums". The fewer bits are wasted on checksums, the better. For example, in triple repetition code this difference equals $2N/3$. In the Hamming codes this difference is a very small number compared to N.

In the remaining few problems we explain two things: 1) how the receiving party recovers original message c, given the result of transmission c'; 2) how the sending party encodes "useful" message consisting of $\dim C$ bits into a message $c \in C \subseteq W_N$.

For the former question, the idea is as follows: since we have a matrix H_m whose null space is C, why don't we use the same matrix to decide which bit is wrong (if any) in c'.

Suppose $c \in C$ is the original message and c' is the received message such that $c' \in W_N$ and $d_H(c, c') \leq 1$. Write $c' = c + e$. Then

$$H_m c' = H_m(c + e) = H_m c + H_m e = H_m e.$$

Note that e is a string of zeros, except for a single 1 in i-th place.

(11) Let e be as described above, that is $e \in W_n$ and e is a string of zeros, except for a single 1 in i-th place. Express $H_m e$ through elements of H_m. (Hint: which columns of H_m matter when computing $H_m e$?) How one finds i given the column $H_m e$?

(12) Suppose $c \in C$ and $c' \in W_N$ is such that $d_H(c, c') \leq 1$. Find what $H_m c'$ looks like if $d_H(c, c') = 0$ (that is, $c = c'$), and if $d_H(c, c') = 1$ and (that is, a single error occurred in i-th place). How to recover c', given $H_m c'$?

The problem above suggests the following procedure: the receiving party, having received a message c', multiplies $H_m c'$. Then, using results of the problem, judges whether any bit was altered in transmission, and, if any, which one. This allows to tell which $c \in C$ was the original message. (Note that if there are more than 1 error in transmission, this procedure will give an incorrect answer.)

—— *see next page* ——

Now, to the latter question. There is a simple (and also not very practical) answer: make a table of correspondence between $(\dim C)$-bit messages and C. There is a better way, though. As implied in problem 7, a particular choice of a specific matrix H_m so far did not matter. Now, let's choose H_m of the form

$$H_m = (A \mid I_m) = \begin{pmatrix} * & * & * & * & 1 & 0 & \cdots & 0 \\ * & * & * & * & 0 & 1 & & \vdots \\ * & * & * & * & \vdots & & \ddots & 0 \\ * & * & * & * & 0 & \cdots & 0 & 1 \end{pmatrix}$$

where and I_m is $m \times m$ identity matrix.

(13) Write such a matrix H_3. (Again, there are multiple possible matrices A. Pick one.)

Consider i-th row of H_m. It looks like

$$(h_{i1}\ h_{i2}\ \dots\ h_{i,N-m}\ 0\ \dots\ 0\ 1\ 0\ \dots 0)$$

where the 1 is in $(N - m + i)$-th place. The corresponding equation is

$$(h_{i1}x_1 + h_{i2}x_2 + \dots + h_{i,N-m}x_{N-m}) + x_{N-m+i} = 0,$$

or (over the field on two elements)

$$h_{i1}x_1 + h_{i2}x_2 + \dots + h_{i,N-m}x_{N-m} = x_{N-m+i}.$$

That is, bits $x_1, \dots, x_{N-m}$ can be chosen arbitrary and bits x_{N-m+i} are expressed through them (note: matrix H_m is now in row-reduced form!).

This allows to interpret bits 1 through $N-m$ as "useful information" and bits $N - m + 1$ through N as "parity checks" (bit number $N - m + i$ checks parity of sum of bits in places which have nonzero values of $h_{i1}, \dots, h_{i,N-m}$).

Hence the encoding procedure: given a matrix H_m of the form specified above, the sending party takes "useful information" $(x_1, \dots, x_{N-m})$, computes m sums

$$h_{i1}x_1 + h_{i2}x_2 + \dots + h_{i,N-m}x_{N-m}, \quad i \le 1 \le m,$$

and uses them as bits $x_{N-m+1}, \dots, x_N$ in the transmission $(x_1, \dots, x_N)$.

(14) For the matrix H_3 that you picked, encode messages 0101 and 0111. Enjoy seeing how 1 bit difference transformed into 3 bit difference.

(15) Assuming that Alice and Bob are both fluent with binary numerical system, solve problem 1a. (Hint: start by assuming that the secret number is $b_3b_2b_1b_0$ in binary notation; for example, 5 is represented as $5 = 0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$, that is, by the binary digit string 0101. Each Bob's question corresponds to a bit in W_7. That is, each question that Bob asks is of the form "If $b_3b_2b_1b_0$ is your number in binary notation, is $b_{\text{something}} = 0$?" or "... is $b_{\text{something}} + \dots b_{\text{something}} = 0$?".)

A final remark about Hamming codes. Let's forget for a second about them and go back to the original question. Suppose you have a transmitted message of length n and suppose you know that a single error might have occurred. How you find and correct the error without retransmitting the whole thing? Well, one way is to split the message in first and second halves and ask to transmit the *sums* of bits in the first half and second half. If either of these two numbers does not coincide with corresponding sums in your received message, you immediately narrow your search down to a half of the message. Repeat this $\log n$ times and find the error.

Basing on this consideration, one may actually solve the initial problem without any linear algebra. However, the description of the code will be rather clumsy. Linear algebra over $\mathbb{Z}_2$ provides an extremely fortunate toolset for this particular problem. One can put nails into a wall with one's forehead (if it is strong enough), but using hammer makes it so much easier.