Discrete Mathematics — Lent 2009
MA 210

Notes for lectures 19 and 20

4 Coding Theory

4.4 Some examples of binary codes

4.4.1 Parity check codes

Let C_e be the set of all sequences in $\{0,1\}^n$ with even weigh. Hence C_e contains all $0,1$-sequences of length n with an even number of 1's. We have seen before that the number of $0,1$-sequences of length n with an even number of 1's is the same as the number of $0,1$-sequences of length n with an odd number of 1's. Since the total number of $0,1$-sequences of length n is 2^n, the number of $0,1$-sequences of length n with an even number of 1's is $2^n/2 = 2^{n-1}$. Hence we have that $|C_e| = 2^{n-1}$. You should convince yourself that the minimum distance of C_e is $\delta(C_e) = 2$. Hence C_e can detect one error.

Another way to get the code C_e is as follows: To every $0,1$-sequence of length $n-1$ $\bar{x}^- \in \{0,1\}^{n-1}$ we add an extra n-thbit as follows: if the number of 1's in $\bar{x}^-$ is even, then the extra bit is 0. If the number of 1's in $\bar{x}^-$ is odd, then the extra bit is 1. In other words, the extra bit makes the total number of 1's even. Let C^+ be the collection of all words that results in this way, by varying $\bar{x}^-$ over all sequebces in $\{0,1\}^n$. Then $C^+ = C_e$.

The extra bit we added is sometimes called the *parity check bit*. The codes C_e are called *parity check codes*.

For $n = 4$, we get the parity check code

$$C_e = \{0000, 0011, 0101, 1001, 1100, 1010, 0110, 1111\}.$$

4.4.2 d-repetition code

Let k and d be positive integers, and set $n = k \cdot d$. Then the *d-repetition code of length n* is obtained as follows: For every $\bar{y} \in \{0,1\}^k$, form $\bar{x}$ by putting d copies of $\bar{y}$ in a row. Let C be the collection of all all words $\bar{x}$ that can be formed in this way.

It is easy to see that the d-repetition code C of length n has minimum distance $\delta(C) = d$.

For $k = 2$ and $d = 3$, we get $n = 6$ and

$$C = \{000000, 010101, 101010, 111111\}.$$

© London School of Economics, 2009

4.5 Preliminaries for linear codes

From now on we no longer consider $\{0,1\}$ as just a set with two element, but as a set with additional algebraic structure. We define addition in $\{0,1\}$ modulo 2:

$$0+0=0,\ 1+0=1,\ 0+1=1,\ 1+1=0.$$

For two words $\bar{x} = x_1x_2\ldots x_n \in \{0,1\}^n$ and $\bar{y} = y_1y_2\ldots y_n \in \{0,1\}^n$, we define the sum $\bar{z} = \bar{x} + \bar{y}$ as $\bar{z} = z_1z_2\ldots z_n$, where $z_i = x_i + y + i$ for $i = 1,2,\ldots,n$.

With addition defined above, the set $\{0,1\}^n$ becomes an abelian group with identity $\bar{\mathbf{0}}$ in which every element is its own inverse (why?).

4.6 Linear codes

A code $C \subseteq \{0,1\}^n$ is called a *linear binary code* if for all $\bar{x}, \bar{y} \in C$ also $\bar{x} + \bar{y} \in C$.

Fact 4.8. *If C is a linear code, then $\bar{\mathbf{0}} \in C$.*

It follows from Fact 4.8 and from the definition of the linear code that every linear code $C \subseteq \{0,1\}^n$ is a subgroup of $\{0,1\}^n$. By Lagrange's Theorem, the size of C must divide the size of $\{0,1\}^n$. Since the size of $\{0,1\}^n$ is 2^n, the size of C must be 2^k for some k, $0 \le k \le n$. We call this k the *dimension* of C.

Theorem 4.9. *Let C be a linear code. Then the minimum distance of C is the minimum weight of a non-zero codeword in C; hence*

$$\delta(C) = \min\{w(\bar{x})\ :\ \bar{x} \in C,\ \bar{x} \neq \bar{\mathbf{0}}\}.$$

4.7 Construction of linear codes

Let H be a binary matrix with k rows and n columns. For a word $\bar{x} \in \{0,1\}^n$, we write $\bar{x}'$ for the word $\bar{x}$ considered as a *column vector*. Then $H\bar{x}'$ is a column vector with k entries.

Theorem 4.10. *Let H be a binary matrix with n columns. Then the set*

$$C = \{\bar{x} \in \{0,1\}^n\ :\ H\bar{x}' = \bar{\mathbf{0}}'\}$$

is a linear code.

If C is a code constructed as above, then the matrix H is called the *parity-check matrix* or just the *check matrix* of C.

For example, let H be the matrix

$$H = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Then the code C contains every word $\bar{x} = x_1x_2x_3$ satisfying

$$H\bar{x}' = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \bar{\mathbf{0}}'.$$

This gives $x_1 + x_3 = 0$ and $x_2 = 0$, which is the same as $x_1 = x_3$ and $x_2 = 0$. We conclude that $C = \{000, 101\}$.

4.8 Correcting errors in linear codes

Theorem 4.11. *Let H be a binary matrix in which no column consists entirely of zeros, and in which no two columns are the same. Then the code C with H as the chceck matrix can correct one error.*

Suppose that C is a code with minimum distance at least 3. Hence we know we must be able to correct if at most one error occurs. So suppose that a codeword $\bar{\boldsymbol{x}} \in C$ is transmitted and a word $\bar{\boldsymbol{y}}$ with at most one error is received. How can we determine whether $\bar{\boldsymbol{y}}$ is a correct word or, if not, what the error is?

Remember that the receiver only knows $\bar{\boldsymbol{y}}$. So, in order to check whether $\bar{\boldsymbol{y}}$ is a codeword, the receiver can compare $\bar{\boldsymbol{y}}$ with all the codewords in C. If there is no match, then the receiver should chceck again and find a codeword that differs from $\bar{\boldsymbol{y}}$ in exactly one bit.

Now, this can be a rather tedious process, especially when the code is large. However, for linear codes we can do much better. Let C be a linear code determined by the check matrix H with minimum distance at least 3. Suppose that a codeword $\bar{\boldsymbol{x}}$ is sent and an error occurs in the i-th bit. Hence, the word $\bar{\boldsymbol{y}}$ that is received satisfies

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{x}} + \bar{\boldsymbol{e}}_i,$$

where $\bar{\boldsymbol{e}}_i$ is the word with all bits equal to 0, except the i-th bit, which is 1. Then we can calculate

$$H\bar{\boldsymbol{y}}' = H(\bar{\boldsymbol{x}} + \bar{\boldsymbol{e}}_i) = H\bar{\boldsymbol{x}}' + H\bar{\boldsymbol{e}}_i'.$$

Since $\bar{\boldsymbol{x}}$ is a codeword, we have $H\bar{\boldsymbol{x}} = \bar{\boldsymbol{0}}'$. So,

$$H\bar{\boldsymbol{y}}' = H\bar{\boldsymbol{e}}_i'.$$

It is easy to see that $H\bar{\boldsymbol{e}}_i'$ is equal to the i-th column of H.

So, we have the following procedure to detect and correct single errors in a linear code C with $\delta(C) \geq 3$ and with the check matrix H:

1. Compute $H\bar{\boldsymbol{y}}'$, where $\bar{\boldsymbol{y}}'$ is the received word.
2. If $H\bar{\boldsymbol{y}}' = \bar{\boldsymbol{0}}'$, then $\bar{\boldsymbol{y}}$ is a codeword.
3. If $H\bar{\boldsymbol{y}}' \neq \bar{\boldsymbol{0}}'$, then locate the i-th column of H that equals to $H\bar{\boldsymbol{y}}'$. Correct the i-th bit of $\bar{\boldsymbol{y}}$.