EE/Ma 127a Error-Correcting Codes - Midterm Exam

Ling Li, ling@cs.caltech.edu

October 26, 2000

M.1 Hamming bound

The Hamming bound says for code $(n, k = 2t + 1)$,

$$2^k \cdot V(n,t) \leq 2^n,$$

where $V(n,t) = 1 + n + \binom{n}{2} + \cdots + \binom{n}{t}$. Since redundancy $r = n - k$, we have

$$r \geq \log_2 V(n,t).$$

(a) $n = 23$, $d = 7$. $t = 3$, $V(n,t) = 1 + 23 + \binom{23}{2} + \binom{23}{3} = 2048$, so $r \geq 11$.

(b) $n = 2^m - 1$, $d = 3$. $t = 1$, $V(n,t) = 1 + (2^m - 1) = 2^m$, so $r \geq m$.

(c) $n = 2m + 1$, $d = 2m + 1$. $t = m$,

$$\begin{aligned} V(n,t) &= \sum_{i=0}^{m} \binom{2m+1}{i} = \sum_{i=0}^{m} \frac{1}{2}\left[\binom{2m+1}{i} + \binom{2m+1}{2m+1-i}\right] \\ &= \frac{1}{2}\sum_{i=0}^{2m+1}\binom{2m+1}{i} = 2^{2m}. \end{aligned}$$

So $r \geq 2m$.

(d) In case (b), when $r = m$, the code is Hamming code $(2^m - 1, 2^m - m - 1)$. The parity-check matrix for this code is a $m \times (2^m - 1)$ matrix, whose columns are all distinct nonzero vectors of length m.
In case (c), when $r = 2m$, the code is $(2m + 1, 1)$, which has only 2 codewords. One example is the repetition code.
In case (a), $r = 11$ is really a close lower bound for r. I have no idea about how to design a code $(23, 12)$ with $d = 7$. However, we can find a cyclic code for $(23, 12)$ with generator polynomial $g(x) = x^{11} + x^9 + x^7 + x^6 + x^5 + x + 1$. The weight of $g(x)$ is 7, but I don't know whether the minimum codeword distance is 7.

M.2 Encoding and decoding for $(7, 4)$

One parity-check matrix for the $(7, 4)$ Hamming code is

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

One corresponding generator matrix is

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

It is easy to verify that $HG^T = 0$.

Encoding: the codeword is generated by

$$(x_1, x_2, x_3, x_4)G$$

or

$$(x_5, x_6, x_7) = (x_1, x_2, x_3, x_4) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Decoding:

- When codeword $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ is received, calculate the syndrome

$$\mathbf{s} = (s_1, s_2, s_3) = \mathbf{x}H^T.$$

- If the syndrome $\mathbf{s}$ is $(0, 0, 0)$, accept the codeword $\mathbf{x}$.
- Otherwise, an error is detected. For one-bit error, the syndrome indicates an error in x_i, where $i = 2^2 s_1 + 2^1 s_2 + 2^0 s_3 = 4s_1 + 2s_2 + s_3$.[1]
- We can correct this one-bit error, or just request a re-transmission.

M.3 Self-dual codes

For a "self-dual" code (n, k), its dual code $(n, n-k)$ is itself. Thus $k = n - k$, i.e., $k = n/2$. In this problem, $n = 4$, thus $k = 2$.

Let $A(z) = \sum_{i=0}^4 A_i z^i$ denote the weight enumerator for one such self-dual code $(4, 2)$. The MacWilliams identities says

$$A(z) = \frac{1}{2^2} \sum_{i=0}^{4} A_i (1-z)^i (1+z)^{4-i},$$

which gives

$$\begin{cases} 3A_0 - A_1 - A_2 - A_3 - A_4 = 0 \\ 2A_0 - A_1 - A_3 - 2A_4 = 0 \\ A_0 - A_2 + A_4 = 0 \\ A_0 - A_1 + A_2 - A_3 - 3A_4 = 0 \end{cases}$$

[1]For only one error, in x_i, we have

$$\mathbf{s} = \mathbf{x}H^T = h_i,$$

where h_i is the ith column of H. Since h_i is designed in a 'binary' order, namely, $i = 2^2 s_1 + 2^1 s_2 + 2^0 s_3$, we can directly give the error position by simple calculation.

Together with

$$\sum_{i=0}^{4} A_i = 2^k = 4,$$

we have

$$A_0 = 1, A_1 = 2 - A_3 - 2A_4, A_2 = 1 + A_4.$$

Since $A_i \geq 0$, we have $A_4 \leq 1$, and

- If $A_4 = 1$, then $A_1 = A_3 = 0$, $A_2 = 2$, thus

$$A(z) = 1 + 2z^2 + z^4.$$

- If $A_4 = 0$, then $A_0 = A_2 = 1$, $A_1 = 2 - A_3$. For $A_3 = 0, 1, 2$, we get

$$A(z) = 1 + z^2 + 2z^3, \text{ or } 1 + z + z^2 + z^3, \text{ or } 1 + 2z + z^2.$$

However, $C^{\perp} = C$ means if $\mathbf{x}_1, \mathbf{x}_2$ are different codewords of C,

$$\mathbf{x}_1 \cdot \mathbf{x}_2 = \mathbf{x}_1 \mathbf{x}_2^T = 0. \tag{1}$$

Here '$\cdot$' is the inner-product of vectors. If the weight of $\mathbf{x}_1$ is odd and $\mathbf{x}_2 \neq 0$, we have $\mathbf{x}_1 + \mathbf{x}_2 \neq \mathbf{x}_1$ is also a codeword of C, and

$$\mathbf{x}_1 \cdot (\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{x}_1 \cdot \mathbf{x}_1 + \mathbf{x}_1 \cdot \mathbf{x}_2 = \mathbf{x}_1 \cdot \mathbf{x}_1 \neq 0.$$

This contradicts the property (1) of C. Thus A_1 and A_3 should both be 0.

So the only possible weight enumerator is

$$A(z) = 1 + 2z^2 + z^4.$$

One example of C is $\{0000, 0011, 1100, 1111\}$.

M.4 One error and one erasure

Suppose (n, k) is such a code and H is one parity-check matrix for this code. Let h_i denote the ith column of H, $i = 1..n$, and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote the codeword before sent.

- Correct one error. This requires $h_i \neq 0$, and for $i \neq j$, $h_i \neq h_j$.
- Correct one erasure when no other error happens. If the jth bit of the codeword is erased, that is, we do not know x_j, we can calculate

$$s_j = \sum_{i=1, i \neq j}^{n} x_i h_i.$$

If the erased bit x_j is 0, $s_j = H\mathbf{x}^T = 0$; otherwise $s_j = 0 - h_j = h_j$. So correcting one erasure requires $h_j \neq 0$ for $j = 1..n$.

- Correct one error and one erasure. If x_i is inverted and x_j is erased, the received codeword $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is
$$y_i = 1 + x_i, y_j = ?, y_l = x_l \text{ for all } l \neq i, j.$$
Thus
$$s_j = \sum_{l=1, l\neq j}^{n} y_l h_l = \sum_{l=1}^{n} x_l h_l + h_i + x_j h_j = h_i + x_j h_j.$$
  - To distinguish from only one erasure, we need
$$h_i \neq 0 \text{ and } h_i \neq h_j$$
for all $i \neq j$.
  - To distinguish from different one error at i' and one erasure at j,[2] we need
$$h_i \neq h_{i'} \text{ and } h_i + h_j \neq h_{i'} \text{ and } h_j \neq 0$$
for all i, i', j are different from each other.

So, in order to correct simultaneously one error and/or one erasure, any 3 columns of H should be linear independent. This is discussed in homeworks 1.5 and 2.1. For $k = 4$, $n \geq 8$. We can design an H for code $(8, 4)$
$$H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$
which is directly extended from $(7, 4)$.

M.5 Number of codes

(a) Using Gaussian binomial coefficient, the number of $(7, 4)$ binary linear codes is
$$\begin{bmatrix} 7 \\ 4 \end{bmatrix}_2 = \prod_{i=0}^{4-1} \frac{2^7 - 2^i}{2^4 - 2^i} = 11811.$$

(b) $d \geq 3$ requires that the columns of the parity-check matrix H are distinct nonzero vectors of length 3. Since $n = 7 = 2^3 - 1$, any nonzero vector of length 3 is one column of H. Thus the number of different H is 7!. However, for one code there are
$$\prod_{i=0}^{r-1} (2^r - 2^i) = (2^3 - 1)(2^3 - 2)(2^3 - 2^2) = 7 \times 6 \times 4$$
different parity-check matrices. So the number of different codes such that $d \geq 3$ is
$$\frac{7!}{7 \times 6 \times 4} = 30.$$

[2] The position of erasure cann't change. However, the position of one error may change, and the erased bit x_j may be 0 or 1.