LINEAR CODES

If the set of codes, C, in $V = \mathbb{Z}_2^n$ is a subspace of V, then C is called a linear code. where, if $\dim C = r$, it in called an $[n, r]$ code

C can be defined by a basis and C is all linear combinations of the basis. The coefficents are 0's and 1's, so if $\dim C = r$ then $|C| = 2^r$.

To construct a linear code, start with an $r \times n$ matrix of full row rank, G Then $G: \mathbb{Z}_2^r \to \mathbb{Z}_2^n$ by $wG \subseteq \mathbb{Z}_2^n$ for $w \in \mathbb{Z}_2^r$. Since G has full row rank, the rows of G

are linearly independent and form a basis for $C \subseteq \mathbb{Z}_2^n$ and $\dim C = r$. The minimum distance between vectors in $\mathbb{Z}_2^r$ is 1 so that one can not correct errors. The minimum distance in C is hopefully more. In fact how one takes G determines this minimal distance

Ex.1. $G = \begin{pmatrix} 1111 & 0000 & 111 \\ 0000 & 1111 & 111 \end{pmatrix}$

$n = 11$ $r = 2$ and the rows are linearly independent, so G has full row rank. C is all linear combinations of the rows of G. So, beside the rows of G, (11111111000) and (00000000000) make up C. $\dim C = 2$. It is seen that $d = 7$, so $t = 3$ and 3 errors can be corrected

3 Since G has full row rank, it has a right inverse, U, so $wGU = w$ and U takes elements in C back into $\mathbb{Z}_2^r$.

Since G is $r \times n$ with full row rank, the null space of G has dimension $n-r$. To find it, solve $Gx = 0$. The solutions are the null space. Take a basis for the null space and put them in a matrix H^t as its columns. Then $GH^t = 0$ and $HG^t = 0$. In particular, if $c \in C$, $c = wG$ for some $w \Leftrightarrow Hc^t = H(wG)^t = HG^t w^t = 0$. So c is a code word iff $Hc^t = 0$. H is used to check if a vector is a code word. We will also use it to correct errors. This very important matrix is called the parity check matrix for the code.

4 In constructing these codes, we can start with G and compute H. Or we can start with H and compute G. To see this let

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

G is found from the null space of H

Let $Hx = 0$ $\quad x^T = (x_1\, x_2\, x_3\, x_4\, x_5\, x_6\, x_7)$

Then

$$x_4 + x_5 + x_6 + x_7 = 0$$
$$x_2 + x_3 + x_6 + x_7 = 0$$
$$x_1 + x_3 + x_5 + x_7 = 0$$

Row reduction gives

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$\to$ $x_1 = x_3 + x_5 + x_7$

$x_2 = x_3 + x_6 + x_7$

$x_4 = x_5 + x_6 + x_7$

Remember, all computation is mod 2.

Linearly independent solutions

$(1\,1\,1\,0\,0\,0\,1)$

$(0\,1\,0\,1\,0\,1\,0)$

$(1\,0\,0\,1\,1\,0\,0)$

$(1\,1\,1\,0\,0\,0\,0)$

These are the rows in G

G is $4 \times 7 \to \dim C = 4$

generated by the rows of G

5 G is 4×7 and H is 3×7

Always the number of rows in G + the number of rows in H = the number of columns in G (and also H)

As always, we are interested in the number of errors that a code is guaranteed to correct. As we know, for any C, we computed the desired distance d which is the minimum of all $d(x,y)$, $x, y \in C$. Here the code is linear so both $x-y$ and $0 \in C$. Also $d(x,y) = d(x-y, 0)$. The second term is the length of $x-y$. Hence, in linear codes, d = min of the lengths of the none zero vectors in C. In our example above, only three vectors need to be checked and we find $d=7$. Hence $t=3$

A useful way to compute d and t is given next

Theorem. Let C be a code with parity check matrix $H = [h_1, \ldots h_n]$ where the h_j are the colomns of H. Then d = minimum number of colomns that give a dependent set

Proof. In general $c \in C$ iff $Hc^t = 0$ iff $h_1 c_1 + \ldots + h_n c_n = 0$ ~~each~~ $c_i = 0$ or 1. Then $(c_1, \ldots c_n) \in C$ the shortest $(c_1, \ldots c_n) \in C$ has d ~~colomns~~ of the c_i equal to one, the rest 0. That means d of the c_i in $h_1 c_1 + \ldots + h_n c_n = 0$ are 1 and those related colums are linearly dependent. The argument works in the other direction as well so the result holds

Ex 2 Continued

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

The first 3 columns are linearly dependent and no two columns are, so $d = 3$ and $t = 1$

We now consider error correction. Suppose c is sent and r is received. Then $r = c + e$ and e is the error. Note that $Hr = Hc + He = He$. Since we can compute Hr, if we can find e from $Hr = He$, we can correct the error. Since $r = c + e$, r and e are in the same coset of C in $V = \mathbb{Z}_2^n$. This coset $= \{r + x \mid x \in C\}$. If r can be corrected, then the length of e is $\leq t$. If there is another $* \in$ the coset of length $\leq t$, then $* = e + c_1$ for

some $c_1 \in C$. The length of c_1 is greater than equal to d and equals the length of $x+e$ $\leq$ length x + length $e \leq 2t$. Hence $d \leq 2t$ a contradiction Hence there is only one vector in the coset that is shorter or equal in length to t. That is logically the error so we need to find the shortest vector in the coset So far, in summary, the vector r is received and code C is known. To compute the coset $r+C = \{r+c; c \in C\}$ will lead to a shortest vector, the error as long as r has no more than t errors. The problem is this could be a long process if C is large

Instead, we compute Hr. For any vector, x, in the coset $r+C$, $Hx = H(r+c) = Hr + Hc = Hr$. Thus, Hr gives the coset to consider. Hr is called the syndrome of r. For a code that is used a lot, one lists the syndromes of all cosets $r+C$ where r has less than or equal to t errors. The shortest vector in that coset is found and listed next to the syndrome. Thus computing the syndrome and checking the table of syndromes and shortest vectors, will find e.

Ex 3 $G = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$

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

The last 3 columns in H are linearly dependent. Thus $d = 3$, $t = 1$. Vectors with one or zero errors are listed in the first column of the following table along with their syndromes. For example, if $e = (0\,0\,1\,0\,0)$, $He = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ is the syndrome.

TABLE

Syndrome	Coset leader (shortest vector)
$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$(0\,0\,0\,0\,0$

Next Page

Coset Leader (Shortest Vector)	Syndrome
$(0\,0\,0\,0\,0)$	$(0\,0\,0)^t$
$(1\,0\,0\,0\,0)$	$(1\,1\,1)^t$
$(0\,1\,0\,0\,0)$	$(1\,0\,0)^t$
$(0\,0\,1\,0\,0)$	$(0\,1\,1)^t$
$(0\,0\,0\,1\,0)$	$(0\,1\,0)^t$
$(0\,0\,0\,0\,1)$	$(0\,0\,1)^t$

Suppose $r = (0\,0\,0\,1\,1)$ is received

$Hr = (0\,1\,1)$ Hence $e = (0\,0\,1\,0\,0)$

and $c = r + e = (0\,0\,1\,1\,1)$

Suppose $r = (0\,1\,0\,0\,1)$

$Hr = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ This is not a syndrome. r's coset does not contain on of the vectors in the first column of the table and the error can not be corrected

Several things to note

In the last example, if $t=2$, then syndromes for all vectors with exactly 2 ones would be added to the table. There would be $\binom{5}{2}=10$ of them.

In computing the table, we only need H and t. Then correcting r, we compute Hr to get the syndrome for r and look up the error in the table

HAMMING CODES

We have considered the parity check matrix

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

with

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

The first 3 columns of H are linearly dependent. Hence $d=3$, $t=1$. The rows in G generate a $[7,4]$ code. H is quite special. The columns of H represent the numbers from 1 to 7 in binary. The sixth colum is $1 \cdot 2^2 + 1 \cdot 2 + 0 \cdot 1 = 6$. This is a small example of a class of codes, called Hamming codes. Here H is 3 by 7 and G is 4 by 7. Letting $n=3$, H is n by 2^n-1 and G is 2^n-n-1 by 2^n-1.

For each n there is such a code

$n=4$ gives H is 4 by 15 and G is 11 by 15. Here

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

and C has dimension 11 which equals the number of rows in G

Hamming codes are 1 error correcting since the first 3 columns in H are always linearly dependent

Thus the only errors correctable have 1 one and the rest 0

Considering the 7x4 Hamming code

Let $e = (0\,0\,0\,0\,1\,0\,0)$

$He = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ which is 5 in binary and is also the position of the 1 in e. So to correct r, compute

Hr = the column whose position gives the 1 in the error

Ex $r = (1\,1\,0\,1\,1\,0\,0)$

$Hr = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ which represents 4

so, $e = (0\,0\,0\,1\,0\,0\,0)$ and

$c = (1\,1\,0\,0\,1\,0\,0)$

Thm. Hamming codes are perfect

Pick n $\quad H$ is $2^n - 1$ by n

G is $2^n - 1$, $2^n - 1 - n$

$|C| = 2^{2^n - 1 - n}$, $\binom{2^n-1}{0} + \binom{2^n-1}{1}$

The product $= 2^{2^n-1-n}(1 + 2^n - 1)$

$= 2^{2^n - 1 - n + n} = 2^{2^n - 1}$ = number of elements in V

Thus every vector is in a ball of radius $2^n - 1$ around a code word