

Perfect Codes

We shall examine the problem of determining how many words a linear code $C(n, k, d)$ can have. This problem is far from resolved in general, though it has been settled for certain values of n and d . We can however find some bounds on the size of a code with these given parameters.

Theorem 1. (*The Hamming Bound*). Consider the block code C of length n and distance d , where $d = 2t + 1$ or $d = 2t + 2$, then

$$|C| \leq \frac{2^n}{\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{t}}$$

Example 1. What is the largest possible dimension of a $C(7, k, 3)$ linear code?

The Hamming bounds gives:

$$|C| = 2^k \leq \frac{2^7}{\binom{7}{0} + \binom{7}{1}} = \frac{128}{1 + 7} = \frac{128}{8} = 16.$$

Thus $k \leq 4$.

Perfect Codes. A code C of odd distance $d = 2t + 1$ is called *perfect*, if C attains the Hamming bound. The code in the above example is a perfect code. It is not difficult to prove that the only perfect code of length n and distance $d = 1$ is \mathbb{K}^n and the dimension of a perfect code of length n and distance $d = n$ is 1. These codes are called *trivial perfect codes*.

Next, we give an improvement over the Hamming bound for linear codes.

Theorem 2. (*Gilbert-Varshamov Bound*). There exists a linear code of length n , dimension k , and distance d , if

$$\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{d-2} < 2^{n-k}.$$

Corollary 2.1. If $n \neq 1$ and $d \neq 1$, then there exists a linear code $C(n, k, d)$ such that

$$|C| = 2^k \geq \frac{2^{n-1}}{\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{d-2}}$$

Example 2. Find a upper bound and a lower bound of the dimension of a linear code $C(9, k, 5)$; then construct a generator.

By using Theorem 1, we find an upper bound on k :

$$|C| = 2^k \geq \frac{2^9}{\binom{9}{0} + \binom{9}{1} + \binom{9}{2}} = \frac{512}{1 + 9 + 36} = \frac{512}{46} \approx 11.13$$

Hence $k \leq 3$. Now we use Corollary 2.1, and we find:

$$|C| = 2^k \geq \frac{2^8}{\binom{9-1}{0} + \binom{9-1}{1} + \binom{9-1}{2} + \binom{9-1}{3}} = \frac{2^8}{\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \binom{8}{3}} = \frac{2^8}{1 + 8 + 28 + 56} = \frac{256}{93} \approx 2.75$$

Thus $2 \leq k \leq 3$. The matrix $G = \begin{pmatrix} 101111000 \\ 010001111 \end{pmatrix}$ is a generator for $C(9, 2, 5)$ linear code.

Remark. The above example shows that the actual dimension of a linear code may be much lower than its Hamming bound. Although $2.75 < 2^3 = 8 < 11.13$, but there is no way to construct a $C(9, 3, 5)$ linear code.

Gilbert-Varshamov bound, according to Wikipedia is defined as:

$$\text{http} : // \text{en.wikipedia.org/wiki/Gilbert-Varshamov_bound}$$

Let $A_q(n, d)$ denote the maximum possible size of a q -ary code C with length n and minimum Hamming weight d (a q -ary code is a code over the field \mathbb{F}_q of q elements). Then:

$$A_q(n, d) \geq \frac{q^n}{\sum_{j=0}^{d-1} \binom{n}{j} (q-1)^j}$$

For the q a prime power, one can improve the bound to $A_q(n, d) \geq q^k$ where k is the greatest integer for which

$$A_q(n, d) \geq \frac{q^n}{\sum_{j=0}^{d-2} \binom{n-1}{j} (q-1)^j}.$$

So, according to Wikipedia, for $q = 2$, $n = 9$, and $d = 5$, the Gilbert-Varshamov bound:

$$A_2(9, 5) \geq \frac{2^9}{\sum_{j=0}^{5-1} \binom{9}{j} (2-1)^j} = \frac{2^9}{\binom{9}{0} + \binom{9}{1} + \binom{9}{2} + \binom{9}{3} + \binom{9}{4}} = \frac{512}{256} = 2$$

and
$$A_2(9, 5) \geq \frac{2^9}{\sum_{j=0}^{5-2} \binom{9-1}{j} (2-1)^j} = \frac{2^9}{\binom{9-1}{0} + \binom{9-1}{1} + \binom{9-1}{2} + \binom{9-1}{3}} = \frac{512}{93} \approx 5.5$$

Thus $k = 3$ which is a contradiction with the fact that there is no way to construct a $C(9, 3, 5)$ linear code. Hence 2^n in the second formula, should be replaced by 2^{n-1} .

Exercises. Problem 1. Find an upper bound for the size or dimension of a linear code for given values of n and d .

$$(a) n = 8, d = 3 \quad (b) n = 7, d = 3 \quad (c) n = 10, d = 5$$

$$(d) n = 15, d = 3 \quad (e) n = 15, d = 5 \quad (f) n = 23, d = 7$$

Problem 2. Verify the Hamming bound for the linear code C with the given generator matrix.

$$G_1 = \begin{bmatrix} 11111000000000 \\ 00000111110000 \\ 00000111111111 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 100111 \\ 010101 \\ 001011 \end{bmatrix}, \quad \text{and} \quad G_3 = \begin{bmatrix} 1000111 \\ 0100110 \\ 0010101 \\ 0001011 \end{bmatrix}.$$

Theorem 3. If C is a non-trivial perfect code of length n and distance $d = 2t + 1$, then

California State University, East Bay

- (i) either $n = 23$ and $d = 7$, or $n = 2^r - 1$ for some $r \geq 2$ and $d = 3$;
(ii) C will correct all error patterns of weight less than or equal to t , and no other error patterns.

Extended Codes. If $C(n, k)$ is a linear code of an odd length n , then the $(n + 1, k)$ code obtained by adding an overall parity-check digit (usually the last digit) is called the *extended C*. This process is called *extending C*. This extension is performed only when C has some odd weight vectors and then the extended code has only even weight vectors. If the original code C has a $k \times n$ generator matrix $G = [I_k \ B]$, the the extended code C^* has $k \times (n + 1)$ generator matrix

$$G^* = [I_k \ B, b],$$

where the last column b of G^* is appended so that each row of G^* has even weight. A parity-check matrix of C^* can be constructed from G^* and one of the two algorithms for finding a parity check matrix. But there is an easier way if we are given a parity-check matrix H for the original code C . We construct H^* as follows:

$$H^* = \begin{bmatrix} H & J \\ \theta & 1 \end{bmatrix},$$

where J is the $n \times 1$ column of all ones. Since H has rank $n - k$, the last row of H^* ensures that H^* has rank $n - k + 1$. Moreover, since the i th entry of GJ is the weight of the i th row of G , we have

$$G^* H^* = [G, b] \begin{bmatrix} H & J \\ \theta & 1 \end{bmatrix} = [GH, GJ + b] = [\theta, 0] = \theta.$$

Note that for a given C , extending G at a different column gives a *equivalent* code.

Punctured Codes. If $C(n, k)$ is a linear code of an even length n , then the $(n - 1, k)$ code obtained by removing a column of a generator matrix of C is called the *punctured C*. This process is called *puncturing C*. Note that for a given C , it is possible that puncturing at different columns give *inequivalent* codes. The puncture code has length $n - 1$ and dimension k or $k - 1$. If C has weight d , then the puncture code usually has weight $d - 1$, but could conceivably have weight d .

Doubly Even Codes. A self-dual code C generated by the matrix $G = [I \ B]$ is called *doubly even* if the weight of every row of B is divisible by 4. A doubly even $(n, n/2)$ code exists if and only if, n is divisible by 8.

♣ **Hamming Codes.** A code of length $n = 2^r - 1$, $r \geq 2$, having parity-check matrix H whose rows consist of all nonzero vectors of length r is called a *Hamming code* of length $2^r - 1$. Hamming codes are *perfect single-error correcting codes* (see Example 1.)

One possibility for a parity-check matrix H and a generator matrix G for a Hamming code of length 7 ($r = 3$) is

$$H = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Here is an *SDA* for a Hamming code:

Coset leader		Syndrome \mathbf{uH}
θ		θ
I_n		H

Here is a generator and a parity-check matrix of an Extended Hamming code C^* :

$$G^* = \left(\begin{array}{ccccccc|c} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{array} \right) \quad \text{and} \quad H^* = \left(\begin{array}{cccc|c} 1 & 1 & 1 & & 1 \\ 1 & 1 & 0 & & 1 \\ 1 & 0 & 1 & & 1 \\ 0 & 1 & 1 & & 1 \\ 1 & 0 & 0 & & 1 \\ 0 & 1 & 0 & & 1 \\ 0 & 0 & 1 & & 1 \\ - & - & - & - & - \\ 0 & 0 & 0 & & 1 \end{array} \right).$$

Note that $G^*G^{*t} = Z_4$, thus G^{*t} is also a parity check-matrix of the extended Hamming code $C^*(8,4)$ which is self-dual and doubly even. The fact that the generator matrix of any extended hamming code of length 8 is obtained by the product of some elementary matrices by G^* ; we conclude that any extended hamming code of length 8 is a self dual code.

♣ **Extended Golay Code.** This code was used in the Voyager spacecraft program which, in the early 1980's, brought us those marvellous close-up photographs of Jupiter and Saturn.

let $x = 11011100010$ and $P = (p_{ij})$ be the 11×11 full-cycle permutation matrix, i.e. the only non-zero entries are:

$$p_{2,1} = p_{3,2} = p_{i+1,i} = \cdots \cdots = p_{n,n-1} = p_{1,n} = 1.$$

Then define the matrix B as follows:

$$B = \left[\begin{array}{c|c} x & 1 \\ xP & 1 \\ xP^2 & 1 \\ \vdots & \vdots \\ xP^{10} & 1 \\ \hline J & 0 \end{array} \right] = \begin{bmatrix} 11011100010 & 1 \\ 10111000101 & 1 \\ 01110001011 & 1 \\ 11100010110 & 1 \\ 11000101101 & 1 \\ 10001011011 & 1 \\ 00010110111 & 1 \\ 00101101110 & 1 \\ 01011011100 & 1 \\ 10110111000 & 1 \\ 01101110001 & 1 \\ \hline 11111111111 & 0 \end{bmatrix}$$

We now list several important facts about the extended Golay code C_{24} with generator $G = [I_{12} \ B]$ with a parity-check matrix $H = \begin{bmatrix} B \\ I_{12} \end{bmatrix}$.

1. C_{24} is a $(24, 12, 8)$ code with $2^{12} = 4096$ codewords.
2. C_{24} is a self-dual doubly even code.
3. C_{24} is also generated by the matrix $G_1 = [B \ I_{12}]$ with a parity-check matrix $H_1 = \begin{bmatrix} I_{12} \\ B \end{bmatrix}$.
4. C_{24} is a three-error correcting code.
5. All codewords in C_{24} have even weights.

Here is the weight distribution:

Weight	0	4	8	12	16	20	24
Words	1	0	759	2576	759	0	1

♠ Algorithms for Decoding the Extended Golay Code for IMLD. We shall now explain an algorithm for *IMLD* for C_{24} code. Throughout this section, $w = [w_1, w_2]$ denotes the word received, $v = [v_1, v_2]$ the closest codeword to w and $u = [u_1, u_2]$ the error pattern $v + w$. We denote by b_i the i th row of the matrix B and e_i the i th row of the 12×12 identity matrix I_{12} . Our aim is to determine the coset leader, u of the coset containing w without having to refer to the *SDA* of C_{24} .

Since we are assuming that $wt(u) \leq 3$, either $wt(u_1) \leq 1$ or $wt(u_2) \leq 1$. Let s_1 be the syndrome of $w = v + u$ using the parity check matrix $H_1 = \begin{bmatrix} I_{12} \\ B \end{bmatrix}$ and s_2 be the syndrome of $w = v + u$ using the parity check matrix $H = \begin{bmatrix} B \\ I_{12} \end{bmatrix}$. Using the fact that $B^2 = I_{12}$, we have:

$$s_1 = wH_1 = [u_1, u_2]H_1 = u_1 + u_2B \quad \text{and} \quad s_2 = wH = u_1B + u_2 = (u_1 + u_2B)B = s_1B.$$

- (a) If $wt(u_2) = 0$, then s_1 consists of a word of weight at most 3.

(b) If $wt(u_2) = 1$, then s_1 consists of a row of B with at most 2 of its digits changed.

Similarly, if $wt(u_1) \leq 1$, then the syndrome s_2 consists of either a word of weight at most 3 or a row of B with at most 2 of its digits changed.

Algorithm IMLD for the Extended Golay Code.

Step 1. Compute the syndrome $s_1 = wH_1 = [w_1, w_2B]$

Step 2. If $wt(s_1) \leq 3$, then $u = [s_1, \theta]$.

Step 3. If $wt(s_1 + b_i) \leq 2$ for some row b_i of B , then $u = [s_1 + b_i, e_i]$. To obtain $s_1 + b_i$, we choose the row of the matrix $(1, 1, 1, \dots, 1, 1)^t s_1 + B$ with minimum weight.

Step 4. Compute the second syndrome $s_2 = s_1B$.

Step 5. If $wt(s_2) \leq 3$, then $u = [\theta, s_2]$.

Step 6. If $wt(s_2 + b_i) \leq 2$ for some row b_i of B , then $u = [e_i, s_2 + b_i]$. Use the same procedure as in Step 3 to obtain the desired $s_2 + b_i$.

Step 7. If u is not yet determined, then request retransmission.

Note. Since the algorithm is designed for \mathcal{IMLD} and G corrects all the error patterns of weight at most 3, there is only one $i = 1, 2, \dots, 12$ such that the inequalities:

$$\mathbf{wt}(s_1) \leq \mathbf{3}, \mathbf{wt}(s_1 + b_i) \leq \mathbf{2}, \mathbf{wt}(s_2) \leq \mathbf{3}, \text{ and } \mathbf{wt}(s_2 + b_i) \leq \mathbf{2}$$

in Step 2, Step 3, Step 5, and Step 6 hold respectively.

Examples. 1. Decode $w = [1011\ 1110\ 1111, 0100\ 1001\ 0010]$. The syndrome

$$s_1 = wH_1 = 1011\ 1110\ 1111 + 0011\ 1110\ 1110 = 1000\ 0000\ 0001$$

which has weight 2. Since $wt(s_1) \leq 3$, we conclude that

$$u = [s_1, \theta] = [1000\ 0000\ 0001, 0000\ 0000\ 0000]$$

is the error pattern and

$$v = w + u = [0011\ 1110\ 1110, 0100\ 1001\ 0010]$$

was the codeword sent. Since $G = [I_{12}\ B]$, the message sent is

$$\mathbf{m} = \mathbf{v}_1 = \mathbf{0011\ 1110\ 1110}.$$

2. Decode $w = [0010\ 0100\ 1101, 1010\ 0010\ 1000]$. The syndrome

$$s_1 = wH_1 = 0010\ 0100\ 1101 + 1110\ 0000\ 0100 = 1100\ 0100\ 1001$$

which has weight 5. Proceeding to step 3 of the algorithm, we find

$$s_1 + b_5 = 0000\ 0001\ 0010$$

Since $wt(s_1 + b_5) \leq 2$, we conclude that

$$u = [s_1 + b_5, e_5] = [0000\ 0001\ 0010, 0000\ 1000\ 0000]$$

is the error pattern and

$$v = w + u = [0010\ 0101\ 1111, 1010\ 1010\ 1000]$$

was the codeword sent.

3. Decode $w = [0001\ 1100\ 0111, 0110\ 1101\ 0000]$. The syndrome

$$s_1 = wH_1 = 0001\ 1100\ 0111 + 1010\ 1010\ 1101 = 1011\ 0110\ 1010$$

which has weight 5. Proceeding to step 3 of the algorithm, we find

$$s_1 + b_i \geq 3, \quad \text{for } i = 1, 2, \dots, 12$$

We continue to step 4; the second syndrome is

$$s_2 = s_1B = 1110\ 0111\ 1101$$

Forging ahead to step 5, we find

$$s_2 + b_4 = 0000\ 0101\ 0000.$$

Since $wt(s_2 + b_4) \leq 2$, we conclude that

$$u = [e_4, s_2 + b_4] = [0001\ 0000\ 0000, 0000\ 0101\ 0000]$$

is the error pattern and

$$v = w + u = [0000\ 1100\ 0111, 0110\ 1000\ 0000]$$

was the codeword sent.

4. Decode $w = [1111\ 1100\ 0000, 1110\ 0011\ 1000]$. Compute the syndrome:

$$s_1 = 1000\ 1001\ 0010$$

Since the weight of $s_1 \geq 3$, we compute the weight of $s_1 + b_i$, for $i = 1, 2, \dots, 12$

$$wt = [7\ 5\ 7\ 9\ 5\ 3\ 9\ 7\ 7\ 7\ 7\ 7].$$

All the weight are greater than 2, we need to find the second syndrome

$$s_2 = s_1 B = [0101 1010 0000]$$

Since the weight of s_2 is greater than 2, we compute the weight of $s_2 + b_i$, for $i = 1, 2, \dots, 12$.

$$wt = [5 7 7 7 9 7 7 9 3 7 5 7].$$

All the weight are greater than 2, this time we request **retransmission**.

♣ **Golay Code**. By puncturing C_{24} (we shall remove the last digit of each codeword in C_{24}), we obtain the *Golay code* $C_{23} = C(23, 12, 7)$.

Let \hat{B} be the 12×11 matrix obtained from the matrix B by deleting the last column. The 12×23 matrix $G = [I_{12} \hat{B}]$ is the generator of the Golay code $C^*(24) = C_{23}$. Golay Code is a perfect code and according to Theorem 3, it corrects all error patterns of weight 3 or less, and no others. Therefore every received word w of distance at most 3 from a codeword in C_{23} may be corrected. In all that follows, we assume that $wt(u) = wt(v+w) \leq 3$. So if we append the digit 0 or 1 to w forming $\hat{w} = w0$ or $\hat{w} = w1$ respectively so that the resulting word has even weight. We then use the algorithm for decoding the Extended Golay code C_{24} ; once \hat{v} in C_{24} found, we remove the last digit.

Algorithm IMLD for the Golay Code.

Step 1. Form $\hat{w} = w0$ or $\hat{w} = w1$, whichever makes \hat{w} an even word.

Step 2. Decode \hat{w} using the algorithm for decoding the Extended Golay code.

Step 3. Remove the last digit of \hat{v} .

Example. Decode $w = [0010 0100 1001 , 1111 1110 000]$. Since w has odd weight, form

$$\hat{w} = w1 = [0010 0100 1001 , 1111 1110 0001].$$

The syndrome

$$\hat{s}_1 = \hat{w}H_1 = 0010 0100 1001 + 0101 0000 1001 = 0111 0100 0000$$

Since the weight of $s_1 \geq 3$, we compute the weight of $s_1 + b_i$, for $i = 1, 2, \dots, 12$

$$wt = [5 7 5 7 7 11 7 7 7 5 5 7].$$

All the weight are greater than 2, so we need to find the second syndrome

$$s_2 = s_1 B = [1010 0000 0110]$$

Since the weight of s_2 is greater than 2, we compute the weight of $s_2 + b_i$, for $i = 1, 2, \dots, 12$.

$$wt = [7 \ 5 \ 5 \ 5 \ 7 \ 5 \ 7 \ 7 \ 11 \ 7 \ 7 \ 7].$$

All the weight are greater than 2, this time we request **retransmission**.

♡ **Matlab.** Suppose the word $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2]$ is received.

By using $x = 11011100010$ and an 11×11 full-cycle permutation matrix P , we define B .

$$\gg x = [1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0]; P = [\text{zeros}(1,10) \ 1 ; \text{eye}(10) \ \text{zeros}(10,1)];$$

$$\gg \text{for } i = 1 : 11, B1(i, 1 : 11) = [x * P^{i-1}]; \text{end};$$

$$\gg J1 = \text{ones}(11,1); B = [B1, J1; J1', 0]$$

Using B , we define a generator matrix G and a parity check matrix H .

$$\gg I2 = \text{eye}(12); G = [I2, B]; H = [I2; B]; J2 = \text{ones}(12,1);$$

Step 1. Compute the syndrome using the vector w :

$$\gg s1 = \text{rem}(w * H, 2)$$

Step 2. Find the weight of the syndrome s_1 using the vector $J2$:

$$\gg wts1 = s1 * J2$$

If the weight of s_1 is less than or equal to 3, then the error pattern is:

$$\gg u = [s1, \text{zeros}(1,12)]$$

The codeword sent was:

$$\gg v = \text{rem}(w + u, 2)$$

Step 3. Compute the matrix $RB1$

$$\gg R1 = J2 * s1; RB1 = \text{rem}(R1 + B, 2)$$

Find the weight of each row of $RB1$:

$$\gg wt = (RB1 * J2)'$$

If the weight of a row r (say 5) is less than or equal to 2, then the error pattern is:

$$\gg r = 5; u = [\text{rem}(RB(r, 1 : 12), 2), I2(r, 1 : 12)]$$

and the codeword sent was:

$$\gg v = \text{rem}(w + u, 2)$$

Step 4. Compute the second syndrome $s_2 = s_1 B$.

$$\gg s2 = \text{rem}(s1 * B, 2)$$

Step 5. Find $wt(s_2)$,

$$\gg wts2 = (s2 * J2)'$$

If $wt(s_2) \leq 3$, then $u = [\theta, s_2]$.

$$\gg u = [\text{zeros}(1, 12), s_2]$$

and the codeword sent was:

$$\gg v = \text{rem}(w + u, 2)$$

Step 6. Compute the matrix $RB2$ and find the weight of each row:

$$\gg R2 = J2 * s2 ; RB2 = \text{rem}(R2 + B, 2) , wt = (RB2 * J2)'$$

If the weight of a row r (say 8) is less than or equal to 2, then the error pattern is:

$$\gg r = 8; u = [I2(r, 1 : 12) , \text{rem}(RB2(r, 1 : 12), 2)]$$

and the codeword sent was:

$$\gg v = \text{rem}(w + u, 2)$$

Step 7. Request retransmission.