

## Binary Linear Codes

♠ **Generator and Parity-Check Matrices.** A subset  $C$  of  $\mathbb{K}^n$  is called a *linear code*, if  $C$  is a subspace of  $\mathbb{K}^n$  (i.e.,  $C$  is closed under addition). A linear code of dimension  $k$  contains precisely  $2^k$  codewords. The distance of a linear code is the minimum weight of any nonzero codeword. If  $C$  is a block code (not necessarily linear) of length  $n$  and if  $P$  is an  $n \times n$  permutation matrix, the block code  $C' = \{v * P : v \in C\}$  is said to be *equivalent* to  $C$ . A  $k \times n$  matrix  $G$  is a *generator matrix* for some linear code  $C$ , if the rows of  $G$  are linearly independent; that is if the rank of  $G$  equals  $k$ . A linear code generated by an  $k \times n$  generator matrix  $G$  is called a  $(n, k)$  code. An  $(n, k)$  code with distance  $d$  is said to be an  $(n, k, d)$  code. If  $G_1$  is row equivalent to  $G$ , then  $G_1$  also generates the same linear code  $C$ . If  $G_2$  is column equivalent to  $G$ , then the linear code  $C_2$  generated by  $G_2$  is equivalent to  $C$ .

Consider the  $3 \times 5$  generator matrix

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

of rank 3. By using some row operations, we obtain another generator matrix

$$G_1 = \begin{pmatrix} 1 & 0 & 0 & \vdots & 1 & 0 \\ 0 & 1 & 0 & \vdots & 1 & 0 \\ 0 & 0 & 1 & \vdots & 1 & 0 \end{pmatrix} = [I_3 \quad B]$$

for the same linear code. Note that  $G_1$  is in *reduced row echelon form* ( $\mathcal{RREF}$ ). This linear code has an information rate of  $3/5$  (i.e.,  $G$  and  $G_1$  accept all the messages in  $\mathbb{K}^3$  and change them into words of length 5). A generator matrix in the form  $G = [I_3 \quad B]$  is said to be in *standard form*, and the code  $C$  generated by  $G$  is called a *systematic code*. Not all linear codes have a generator matrix in standard form. For example, the linear code  $C = \{000, 100, 001, 101\}$  has six generator matrices

$$G_1 = \begin{pmatrix} 100 \\ 001 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 001 \\ 100 \end{pmatrix}, \quad G_3 = \begin{pmatrix} 100 \\ 101 \end{pmatrix}, \quad G_4 = \begin{pmatrix} 001 \\ 101 \end{pmatrix}, \quad G_5 = \begin{pmatrix} 101 \\ 100 \end{pmatrix}, \quad \text{and} \quad G_6 = \begin{pmatrix} 101 \\ 001 \end{pmatrix}.$$

None of these matrices are in standard form. Note that the matrix  $G' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$  in standard form generates the code  $C' = \{000, 100, 010, 110\}$  which is equivalent to  $C$ . If  $G$  is in  $\mathcal{RREF}$ , then any column of  $G$  which is equal to the vector  $e_i$  is called a *leading column*. If  $\mathbf{m} \in \mathbb{K}^k$  is the message and  $v = \mathbf{m}G \in \mathbb{K}^n$  is the codeword of a systematic code, then the first  $k$  digits of  $v$  which represent the message  $\mathbf{m}$  are called *information digits*, while the last  $n - k$  digits are called *redundancy* or *parity-check* digits. If  $C$  is not a systematic code, then to recover the message from a codeword we select the digits corresponding to the leading columns  $e_1, e_2, \dots, e_k$ . For example, if  $G = \begin{pmatrix} 001 \\ 100 \end{pmatrix} = [e_2 \quad \theta \quad e_1]$  and  $v = 001$ , then we recover the message  $\mathbf{m} = 10$  from the last digit and the first digit of  $v$  respectively.

Let  $S$  be a subset of  $\mathbb{K}^n$ . The set of all vectors orthogonal to  $S$  is denoted by  $S^\perp$  and called the *orthogonal complement* of  $S$ . It can readily be shown that  $S^\perp$  is a linear code. If  $C = \langle S \rangle$ , then  $C^\perp = S^\perp$  which is also a linear code is called the *dual code* of  $C$ .

A matrix  $H$  is called a *parity-check matrix* for a linear code  $C$  of length  $n$  generated by the matrix  $G$ , if the columns of  $H$  form a basis for the dual code  $C^\perp$ . If  $v$  is a word in  $C$ , then  $vH = \theta$ . A code  $C$  is called *self-dual* if  $C = C^\perp$ . In this case  $n$  must be even and  $C$  must be an  $(n, n/2)$  code. If  $G$  is a generator matrix of a self-dual code, then  $H = G^t$ . Both the generator matrices

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad G_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

generate self-dual codes but only  $G_1$  is in  $\mathcal{RR}\mathcal{EF}$ . If  $G = [I \ B]$  is a generator of a self-dual code, then  $B^2 = I$ .

**Theorem 1.** Let  $H$  be a parity-check matrix for a linear code  $C$  generated by the  $k \times n$  matrix  $G$ . Then

- (i) the rows of  $G$  are linearly independent;
- (ii) the columns of  $H$  are linearly independent;
- (iii)  $GH = Z_k$ , where  $Z_k$  is the  $k \times k$  zero matrix;
- (iv) by permuting columns of  $H$ , we obtain another parity-check matrix corresponding to  $G$ ,
- (v)  $\dim(C) = \text{rank}(G)$ ,  $\dim(C^\perp) = \text{rank}(H)$ , and  $\dim(C) + \dim(C^\perp) = n$ ;
- (vi)  $H^t$  is a generator matrix for  $C^\perp$  with  $G^t$  its parity-check matrix;
- (vii) if  $C$  is self-dual with  $G = [I_k \ B]$  its generator, then  $G_1 = [B \ I_k]$  also generates  $C$ ;
- (viii)  $C$  has distance  $d$  if and only if any set of  $d - 1$  rows of  $H$  is linearly independent, and at least one set of  $d$  rows of  $H$  is linearly dependent.

**♣ Algorithms for Finding Generator and Parity-Check Matrices.**

**Example 1.** Let  $S = \{01100, 01010, 11100, 00110\}$  be a subset of  $\mathbb{K}^5$  generating the linear code  $C$ . By using the words in  $S$ , we define the matrix

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

After some row operations and deleting a row, we obtain the generator matrix

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

in  $\mathcal{RR}\mathcal{EF}$ . Let  $B$  be the matrix formed by the last two rows of  $G$ , the parity-check matrix of  $C$  is the following matrix

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

**Example 2.** Let  $S = \{1010010101, 0001010001, 0000100100, 0000001001, 0000000011\}$  be a linearly independent set generating  $C$ . The generator matrix

$$G = \begin{matrix} & e_1 & & e_2 & e_3 & & e_4 & & e_5 \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{matrix}$$

is in  $\mathcal{RRE}\mathcal{F}$  but not in standard form.

We permute the columns of  $G$  into order 1, 4, 5, 7, 9, 2, 3, 6, 8, 10 to form the matrix

$$G_1 = G * P = \begin{pmatrix} 1010010101 \\ 0001010001 \\ 0000100100 \\ 0000001001 \\ 0000000011 \end{pmatrix} \begin{pmatrix} 1000000000 \\ 0000010000 \\ 0000001000 \\ 0100000000 \\ 0010000000 \\ 0000000100 \\ 0001000000 \\ 0000000010 \\ 0000100000 \\ 0000000001 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \vdots & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & \vdots & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & \vdots & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & \vdots & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & \vdots & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Then we form the matrix  $H_1$  and finally rearrange the rows of  $H_1$  into their natural order to form the parity-check matrix  $H$ .

$$H_1 = \begin{bmatrix} B \\ I_5 \end{bmatrix} = \begin{pmatrix} 01111 \\ 00101 \\ 00010 \\ 00001 \\ 00001 \\ 10000 \\ 01000 \\ 00100 \\ 00010 \\ 00001 \end{pmatrix} \begin{matrix} 1 \\ 4 \\ 5 \\ 7 \\ 9 \\ 2 \\ 3 \\ 6 \\ 8 \\ 10 \end{matrix} ; H = P * H_1 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix}$$

The columns of  $H$  form a basis for  $C^\perp$ .

♡ **Matlab.** To permute the columns of  $G$  into order 1, 4, 5, 7, 9, 2, 3, 6, 8, 10 to form the matrix  $G_1$ , first we define the permutation matrix  $P$  as follows:

```
>> P = eye(10) ; T = P ; < Return key >
>> P(:, 2) = T(:, 4) ; P(:, 3) = T(:, 5) ; P(:, 4) = T(:, 7) ; P(:, 5) = T(:, 9) ; P(:, 6) =
T(:, 2) ; P(:, 7) = T(:, 3) ; P(:, 8) = T(:, 6) ; P(:, 9) = T(:, 8) ; < Return key >
>> G1 = G * P < Return key >
```

Finally  $H$  is obtained as follows:

```
>> B = G1(:, 6 : 10) ; H1 = [B ; eye(5)] ; H = P * H1 < Return key >
```

♠ **Maximum Likelihood Decoding (MLD) for Linear Codes.** We will describe a procedure for either  $\mathcal{CMCD}$  or  $\mathcal{TMCD}$  for a linear codes.

If  $C \in \mathbb{K}^n$  is a linear code of dimension  $k$ , and if  $u \in \mathbb{K}^n$ , we define the *coset of  $C$  determined by  $u$*  denoted  $\hat{u}$  as follows:

$$\hat{u} = C + u = \{v + u : v \in C\}.$$

There are as many as  $2^{n-k}$  distinct cosets of  $C$  in  $\mathbb{K}^n$  of order  $2^k$ , where every word in  $\mathbb{K}^n$  is contained in one of the cosets.

**Theorem 2.** Let  $C$  be a linear code. Then

- (i)  $\hat{\theta} = C$ ,
- (ii) if  $v \in \hat{u} = C + u$ , then  $\hat{v} = \hat{u}$ ,
- (iii)  $u + v \in C$  if and only if  $u$  and  $v$  are in the same coset.

The parity-check matrix and cosets of the code play fundamental roles in the decoding process.

Let  $C$  be a linear code. Assume the codeword  $v$  in  $C$  is transmitted and the word  $w$  is received, resulting in the *error pattern*  $u = v + w$ . Then  $w + u = v$  is in  $C$ , so **the error pattern  $u$  and the received word  $w$  are in the same coset of  $C$** . Since error patterns of small weight are the most likely to occur, we choose a word  $u$  of least weight in the coset  $\hat{u}$  (which must contain  $w$ ) and conclude that  $v = w + u$  was the word sent.

Let  $C \in \mathbb{K}^n$  be a linear code of dimension  $k$  and let  $H$  be a parity-check matrix. For any word  $w \in \mathbb{K}^n$ , the *syndrome of  $w$*  is the word  $s(w) = wH$  in  $\mathbb{K}^{n-k}$ .

**Theorem 3.** Let  $H$  be a parity-check matrix for a linear code  $C$ . Then

- (i)  $wH = \theta$  if and only if  $w$  is a codeword in  $C$ .
- (ii)  $w_1H = w_2H$  if and only if  $w_1$  and  $w_2$  lie in the same coset of  $C$ .
- (iii) If  $u$  is the error pattern in a received word  $w$ , then  $uH$  is the sum of the rows of  $H$  that correspond to the positions in which errors occurred in transmission.

A table which matches each syndrome with its coset leader, is called a *standard decoding array*, or *SDA*. To construct an *SDA*, first list all the cosets for the code, and choose from each coset word of least weight as coset leader  $u$ . Then find a parity-check matrix for the code and, for each coset leader  $u$ , calculate its syndrome  $uH$ .

**Example.** Here is the list of all the cosets of  $C = \{0000, 1011, 0101, 1110\}$  generated by the

generator matrix  $G = \begin{pmatrix} 1011 \\ 0101 \end{pmatrix}$  with a parity-check matrix  $H = \begin{pmatrix} 11 \\ 01 \\ 10 \\ 01 \end{pmatrix}$ :

$$\begin{aligned} \widehat{0000} &= \{0000, 1011, 0101, 1110\} \\ \widehat{1000} &= \{1000, 0011, 1101, 0110\} \\ \widehat{0100} &= \{0100, 1111, 0001, 1010\} \\ \widehat{0010} &= \{0010, 1001, 0111, 1100\} \end{aligned}$$

Here is the *SDA* for the code:

<u>Coset leader <math>u</math></u>	<u>Syndrome <math>uH</math></u>
0000	00
1000	11
0100	01
0010 or 0001	10*

The syndrome with a \* indicates a retransmission in the case of  $\mathcal{IMCD}$ . Notice that the set of error patterns that can be corrected using  $\mathcal{IMCD}$  is equal to the set of unique coset leaders.

If  $w = 1101$  is received, then the syndrome of  $w$  is  $s(w) = wH = 11$ . Notice that the word of least weight in the coset  $\hat{w}$  is  $u = 1000$  and the syndrome of  $u$  is  $s(u) = uH = 11 = wH$ . Furthermore,  $\mathcal{CMCD}$  concludes  $v = w + u = 1101 + 1000 = 0101$  was sent, so there was an error in the first digit. Notice also that  $s(w) = 11$  picks up the first row of  $H$  corresponding to the location of the most likely error; also the coset leader in the  $\mathcal{SDA}$  is 1000. The calculations

$$\begin{aligned} d(0000, 1101) &= 3 & d(0101, 1101) &= 1 \\ d(1011, 1101) &= 2 & d(1110, 1101) &= 2 \end{aligned}$$

give the distances between  $w$  and each codeword in  $C$ , show that indeed  $v = 0101$  is the closest word in  $C$  to  $w$ .

For  $w = 1111$  received, however, the same calculations

$$\begin{aligned} d(0000, 1111) &= 4 & d(0101, 1111) &= 2 \\ d(1011, 1111) &= 1 & d(1110, 1111) &= 1 \end{aligned}$$

reveal a tie for the closest word in  $C$  to  $w$ . This is not surprising, since there was a choice for a coset leader for the syndrome  $1111H = 01$ . In the case of  $\mathcal{CLMD}$ , we arbitrary choose a coset leader, which in effect arbitrary selects one codeword in  $C$  closest to  $w$ . Using  $\mathcal{IMCD}$ , we ask for retransmission.