Linear Codes and Syndrome Decoding

These notes are intended to be used as supplementary reading to Sections 16.7–9 of Grimaldi’s Discrete and Combinatorial Mathematics. The proofs of the theorems are left as exercises.

In the following, an \((n, m)\) block code will be the image \(C = E(\mathbb{Z}_2^m)\) of an encoding function \(E : \mathbb{Z}_2^m \to \mathbb{Z}_2^n\) where we assume, except in Section 5, that the first \(m\) bits of the code word \(E(w) \in \mathbb{Z}_2^n\) agrees with the first \(m\) bits of the message word \(w \in \mathbb{Z}_2^m\).

1. Generator Matrices and Linear Codes

Let \(G\) be an \(m \times n\) matrix over \(\mathbb{Z}_2\). We will assume that \(G\) has the form \(G = [I_m | A]\) where \(I_m\) is the \(m \times m\) identity matrix, and \(A\) is an \(m \times (n - m)\) matrix over \(\mathbb{Z}_2\). We will consider the elements of \(\mathbb{Z}_2^k, k \geq 1\), as row \(k\)-vectors (in other words, \(1 \times k\) matrices) over \(\mathbb{Z}_2\).

We define an encoding function \(E_G : \mathbb{Z}_2^m \to \mathbb{Z}_2^n\) by \(E_G(w) = wG\) for \(w \in \mathbb{Z}_2^m\). \(G\) is called the generator matrix of the code \(C = E(\mathbb{Z}_2^m) \subset \mathbb{Z}_2^n\), and \(C\) is called the code generated by \(G\).

**Example 1:** Let 
\[
G = \begin{bmatrix}
1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 1
\end{bmatrix}.
\]

Then, for instance,
\[
E(110) = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} G
= \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix}
1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 1
\end{bmatrix}
= \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}
= 110101.
\]

The code consists of the words obtained by adding any number of rows of \(G\). Hence, \(C = \{000000, 100110, 010011, 001101, 110101, 101011, 011110, 111000\}\).

**Example 2:** Let \(E : \mathbb{Z}_2^5 \to \mathbb{Z}_2^6\) be the \((6, 2)\) triple repetition code defined by \(E(w) = www\) for all \(w \in \mathbb{Z}_2^5\). Then 
\[
G = \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}
\]
is the generator matrix for this code.

**Example 3:** Let \(E : \mathbb{Z}_2^5 \to \mathbb{Z}_2^6\) be the \((6, 5)\) parity-check code defined by \(E(w) = wp_w\) for all \(w \in \mathbb{Z}_2^5\) where \(p_w\) is 1 if \(w\) has an odd number of 1’s, and 0 if \(w\) has an even number of 1’s. Then 
\[
G = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}
\]
is the generator matrix for this code.
**Definition:** A code $C$ is called a *linear code* if $c_1 + c_2 \in C$ for all $c_1, c_2 \in C$.

All of the codes in the examples above are linear codes. More generally we have the following theorem.

**Theorem 1:** Let $C = E(\mathbb{Z}_2^m) \subset \mathbb{Z}_2^n$ be a code with the encoding function $E : \mathbb{Z}_2^m \rightarrow \mathbb{Z}_2^n$. Then the following conditions are equivalent:

1. $C$ is a linear code.
2. $C$ is generated by a generator matrix.
3. $E(w_1 + w_2) = E(w_1) + E(w_2)$ for all $w_1, w_2 \in \mathbb{Z}_2^m$.

---

### 2. Parity-Check Matrices and Error Correction

For a generator matrix $G = [I_m|A]$, we define the *parity-check matrix* $H$ by $H = [A^t|I_{n-m}]$ where $A^t$ is the transpose of $A$, and $I_{n-m}$ is the $(n - m) \times (n - m)$ identity matrix. Note that $H$ is an $(n - m) \times m$ matrix over $\mathbb{Z}_2$.

**Example 4:** For $G$ of Example 1,

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

is the parity-check matrix.

**Example 5:** For the $(6, 2)$ triple repetition code the parity check matrix is

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

**Example 6:** For the $(6, 5)$ parity-check code the parity check matrix is

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

**Example 7:** Consider the code of Examples 1 and 4. Note that for the code word $x = 110101$, we have

$$Hx^t = H \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}^t = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$
On the other hand for \( x = 001110 \), which is not a code word, we have

\[
H r^t = H [0 \ 0 \ 1 \ 1 \ 1 \ 0]^t = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\]

What is observed in Example 7 is true in general.

**Theorem 2:** Let \( C \) be the code generated by \( G \). Then for \( x \in \mathbb{Z}_2^n \), \( x \in C \) if and only if \( H x^t = 0 \).

**Remark:** We will follow the usual convention and denote zero vectors of any size by \( \mathbf{0} \), and assume that the size is clear from the context.

Suppose that the code word \( c \in C \) is transmitted, and \( r \in \mathbb{Z}_2^n \) is received where \( r = c + e \) and \( e \in \mathbb{Z}_2^n \) is the error pattern. Then \( H r^t = H e^t \) by Theorem 2.

**Example 8:** Again consider the code of Examples 1, 4 and 7. If the received word is \( r = 001110 \), then as we computed in Example 7, \( H r^t = [0 \ 1 \ 1]^t \). If \( e \in \mathbb{Z}_2^n \) is the error pattern, then by the observation above, \( H e^t = H r^t = [0 \ 1 \ 1]^t \).

\([0 \ 1 \ 1]^t \) is the second column of \( H \). Therefore the error pattern \( e = 010000 \) is a possibility. It has weight one. \([0 \ 1 \ 1]^t \) is also the sum of the fifth and the sixth columns, and the sum of the first, the fourth and the sixth columns. These correspond to the error patterns \( e = 000011 \) and \( e = 100101 \) of weights two and three, respectively.

In a binary symmetric channel, for a given received word, the error pattern with the lowest weight, or equivalently, the code word closest to the received word, is the most likely one. So we will assume that \( e = 010000 \) is the error pattern, correct the received word as \( c = r + e = 001110 + 010000 = 011110 \), and decode it as \( w = 011 \).

On the other hand, if \( r = 111111 \), then \( H r^t = [1 \ 1 \ 1]^t \). So \( r \) is not a code word, and since \( H r^t \) is not equal to a column of \( H \), an error of weight one is not possible. There are three possibilities for error patterns of weight two: \( e = 100001, 010100, 001010 \).

3. **Syndromes and Error Correction**

**Definition:** Let \( C \subset \mathbb{Z}_2^n \) be a linear code with the parity-check matrix \( H \). For a received word \( r \in \mathbb{Z}_2^n \), \( H r^t \in \mathbb{Z}_2^m \) is called the **syndrome** of \( r \). The set of all received words with the same syndrome is called a **coset** of the code \( C \).

Having the same syndrome is an equivalence relation, and the cosets are the equivalence classes of this equivalence relation.

We will denote the set \( \{ x + c : c \in C \} \) by \( x + C \). Note that \( x \in x + C \).
Theorem 3: Each coset of the linear code $C \subset \mathbb{Z}_2^n$ is of the form $x + C$ for some $x \in \mathbb{Z}_2^n$. In particular, each coset has $|C| = 2^m$ elements, and there are $2^{n-m}$ cosets.

Definition: An element of a coset whose weight is smaller than or equal to the weights of all other elements in that coset is called a coset leader for the coset.

Example 9: The following is the coset table for the code of Examples 1, 4, 7.

<table>
<thead>
<tr>
<th></th>
<th>000000</th>
<th>100110</th>
<th>010011</th>
<th>001101</th>
<th>110101</th>
<th>101011</th>
<th>011110</th>
<th>111000</th>
</tr>
</thead>
<tbody>
<tr>
<td>100000</td>
<td>001110</td>
<td>110011</td>
<td>101101</td>
<td>010101</td>
<td>001011</td>
<td>111110</td>
<td>011000</td>
<td></td>
</tr>
<tr>
<td>010000</td>
<td>110110</td>
<td>000011</td>
<td>011101</td>
<td>100101</td>
<td>111011</td>
<td>001110</td>
<td>101000</td>
<td></td>
</tr>
<tr>
<td>001000</td>
<td>101110</td>
<td>011011</td>
<td>000101</td>
<td>111011</td>
<td>100011</td>
<td>010110</td>
<td>110000</td>
<td></td>
</tr>
<tr>
<td>000100</td>
<td>100010</td>
<td>010111</td>
<td>001001</td>
<td>110011</td>
<td>101101</td>
<td>010101</td>
<td>111100</td>
<td></td>
</tr>
<tr>
<td>000010</td>
<td>100100</td>
<td>010001</td>
<td>001101</td>
<td>110101</td>
<td>101001</td>
<td>011000</td>
<td>111010</td>
<td></td>
</tr>
<tr>
<td>000001</td>
<td>101011</td>
<td>000110</td>
<td>011001</td>
<td>100110</td>
<td>100110</td>
<td>011110</td>
<td>111010</td>
<td></td>
</tr>
<tr>
<td>010100</td>
<td>110010</td>
<td>000111</td>
<td>011001</td>
<td>100101</td>
<td>101111</td>
<td>001010</td>
<td>101100</td>
<td></td>
</tr>
</tbody>
</table>

We have the elements of $C$ in the first row. Each row consists of the elements of a coset of $C$. An element in the first column is a coset leader for the coset in that row. The coset leader is unique for the cosets in the rows 1-7. The coset in the eighth row has three coset leaders of weight two: 010100, 100001 and 001010. Here we chose the first one when making this table.

Using this table we correct the errors as follows: If $r = 001100$ is received, then we find it in this table. It is the fourth element in the seventh row. The coset leader of this coset is $e = 000001$. So we correct the received word as $c = r + e = 001100 + 000001 = 001101$.

Since the cosets are in a one-to-one correspondence with the syndromes, we can also do this using a smaller table with two columns in which the first column lists the syndrome corresponding to a coset, and the second column lists the coset leader of the coset.

<table>
<thead>
<tr>
<th>$Hr^t$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>000000</td>
</tr>
<tr>
<td>110</td>
<td>100000</td>
</tr>
<tr>
<td>011</td>
<td>010000</td>
</tr>
<tr>
<td>101</td>
<td>001000</td>
</tr>
<tr>
<td>100</td>
<td>000100</td>
</tr>
<tr>
<td>010</td>
<td>000010</td>
</tr>
<tr>
<td>001</td>
<td>000001</td>
</tr>
<tr>
<td>111</td>
<td>010100</td>
</tr>
</tbody>
</table>

Now for the received word $r = 001100$, we simply compute $Hr^t = [0 \ 0 \ 1]^t$, find 001 in the first row, and read the coset leader $e = 000001$ in the second row. \[\square\]

We can prepare the syndrome table for a linear code as follows: Starting with weight zero, and in the order of increasing weight compute the syndrome of each $x \in \mathbb{Z}_2^n$. (We order the elements of the same weight in some way.) If the syndrome we compute is not in the table, we add it to the first column of the table, and add $x$ to the second column as the corresponding coset leader; if the syndrome we compute is already in the table, we move on to the next $x$. We stop when there are $2^{n-m}$ syndromes in the table.
Example 10: We apply this algorithm to get the following syndrome table for the (6, 2) triple repetition code.

<table>
<thead>
<tr>
<th>$Hr^t$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>000000</td>
</tr>
<tr>
<td>1010</td>
<td>100000</td>
</tr>
<tr>
<td>0101</td>
<td>010000</td>
</tr>
<tr>
<td>1000</td>
<td>001000</td>
</tr>
<tr>
<td>0100</td>
<td>000100</td>
</tr>
<tr>
<td>0010</td>
<td>000010</td>
</tr>
<tr>
<td>0001</td>
<td>000001</td>
</tr>
<tr>
<td>1111</td>
<td>110000</td>
</tr>
<tr>
<td>1110</td>
<td>100100</td>
</tr>
<tr>
<td>1011</td>
<td>100001</td>
</tr>
<tr>
<td>1101</td>
<td>011000</td>
</tr>
<tr>
<td>0111</td>
<td>010010</td>
</tr>
<tr>
<td>1100</td>
<td>001100</td>
</tr>
<tr>
<td>1001</td>
<td>001001</td>
</tr>
<tr>
<td>0110</td>
<td>000110</td>
</tr>
<tr>
<td>0011</td>
<td>000011</td>
</tr>
</tbody>
</table>

□

Suppose $r \in \mathbb{Z}_2^n$ is the received word, and let $e \in \mathbb{Z}_2^n$ be the coset leader of the coset $r + C$. Then $\text{wt}(e) = \min\{\text{wt}(r + c) : c \in C\}$. If we apply the syndrome method, then we correct $r$ as $c' = r + e$. Since $d(c', r) = \text{wt}(c' + r) = \text{wt}(e) \leq \text{wt}(r + c) = d(c, r)$ for all $c \in C$, $c'$ is one of the closest code words to $r$ in $\mathbb{Z}_2^n$. In other words, the syndrome error correction is the nearest neighbor error correction for a linear code.

Exercise 1: Note that in both of Examples 9 and 10, every $e \in \mathbb{Z}_2^n$ of weight one occurs as a coset leader in the syndrome table, but this is not true for weight two elements. There are $\binom{6}{2} = 15$ such elements, but only 9 of them occurs as coset leaders in Example 10, and only one in Example 9. If $t$ is the largest integer such that every $e \in \mathbb{Z}_2^n$ with $\text{wt}(e) \leq t$ is the unique coset leader of its coset, what can you say about $t$?

Exercise 2: Check that the remaining $15 - 9 = 6$ weight two elements in Example 10 lie in cosets whose coset leader has weight one. Therefore every coset of the (6, 2) triple repetition code has a unique coset leader. But in Example 9, the last coset in the table has three coset leaders as we observed in Example 9. What can you say about the codes every coset of which has a unique coset leader?

4. Parity-Check Matrix and Minimum Distance

Let $C \subset \mathbb{Z}_2^n$ be a linear code with minimum distance $d$ and the parity-check matrix $H$.

Theorem 4: For a linear code $C$, $d = \min\{\text{wt}(c) : c \in C, c \neq 0\}$.

Theorem 5: Suppose that there are $d'$ columns in $H$ whose sum is zero, but for all $0 \leq i \leq d' - 1$, the sum of any $i$ columns of $H$ is different from zero. Then $d' = d$. 

5. Hamming Codes

Let \( r \) be a positive integer. Let \( H \) be an \( r \times (2^r - 1) \) matrix whose columns consist of all possible nonzero elements of \( \mathbb{Z}_2^r \). The corresponding generator matrix \( G \) is \((2^r - 1 - r) \times (2^r - 1)\). The code generated by \( G \) is called the \( (2^r - 1, 2^r - 1 - r) \) Hamming code.

Example 11: Let \( r = 3 \).

\[
H = \begin{bmatrix}
0 & 1 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}
\]

is the parity-check matrix, and

\[
G = \begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 1
\end{bmatrix}
\]

is the generator matrix of the \((7, 4)\) Hamming code. By Theorem 5, this code has minimum distance 3. Therefore it corrects errors of weight one. Moreover, it has \(2^7-4 = 8\) cosets, and each coset has a unique coset leader of weight less than or equal to one. \(\square\)

Example 12: So far we followed the convention that the first \( m \) bits of each code word is the corresponding message word. This is equivalent to the conditions that \( G \) has the form \([I_m | A]\), and \( H \) has the form \([A^t | I_{n-m}]\). Now we will forgo this condition.

Consider the matrices

\[
H' = \begin{bmatrix}
0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}
\quad \text{and} \quad
G' = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}
\]

\(E_{G'} : \mathbb{Z}_2^3 \to \mathbb{Z}_2^7\) gives an encoding such that

\[
E(w_1 w_2 w_3 w_4) = (w_1 + w_2 + w_4)(w_1 + w_3 + w_4)w_1(w_2 + w_3 + w_4)w_2w_3w_4.
\]

Therefore the message word bits occur as the third, the fifth, the sixth, and the seventh bits of the code word. \(H'\) is the parity-check matrix for this code.

Note that for \(0 \leq i \leq 7\), \(i\)th column of \(H'\) has the digits of the binary expansion of \(i\) from top to bottom. For instance, \(3 = 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = (011)_2\), and the third column is \([0 \ 1 \ 1]^t\).

The advantage of using this set-up is in the error correction. Suppose \(r = 1011100\) is received. Then the syndrome is \(H'r^t = [0 \ 1 \ 1]^t\). This tells us directly that the corresponding weight one error pattern has its only 1 in the 3rd position. That is, \(e = 0010000\). \(\square\)