Coding Theory: Coset Decoding

Alex Andonian and Milan Brankovic

Math 309

April 8, 2016

Introduction and Motivation

Coding theory originated in the 1940s as a way to develop error-correcting for the reliable transmission of information across noisy channels. Coding theory has many practical applications such as how to best encode and decode messages sent over communication channels. In this project, we investigate how coded messages can be decoded using cosets. Specifically we examine what it means to decode a received binary word of length n and how this can be accomplished using cosets. Furthermore, our discussion of coset decoding leads us to topics such as the *coset leader* and *syndrome* of a word x. We are particularly interested in this topic because it is about learning efficient ways of sending and understanding received information, which is ubiquitous in today's modern world.

Elementary Definitions and Theorems

Definition A *binary word* $\mathbf{a} = a_1 a_2 \cdots a_n$ is a string of 0s and 1s, such as 0010111, 1010011, etc., and is considered to be the most basic way of transmitting information. The number of 0s and 1s in any binary word is called its *length*.

Definition If a word $\mathbf{a} = a_1 a_2 \cdots a_n$ is sent, but a word $\mathbf{b} = b_1 b_2 \cdots b_n$ is received (where a_i and the b_j are 0s or 1s), then the *error pattern* is the word $\mathbf{e} = e_1 e_2 \cdots e_n$ where

$$e_i = \begin{cases} 0 & \text{if } a_i = b_i \\ 1 & \text{if } a_i \neq b_i \end{cases}$$

With this in mind, we can define an operation of *adding words* in the following way: If $\mathbf{a}$ and b are both words of binary words of length n, we add them by adding corresponding digits according to the following rules

$$0 + 0 = 0 \qquad 1 + 1 = 0 \qquad 0 + 1 = 1 \qquad 1 + 0 = 1.$$

Thus, the addition of $\mathbf{a}$ and $\mathbf{b}$ is

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

The symbol $\mathbb{B}^n$ designates the set of all the binary words of length n, and it can be shown that $\mathbb{B}^n$ forms a group.

Definition A *code* is just a subset of $\mathbb{B}^n$. Notice that there are 2^n possible words of length n, but only a subset of these words called *codewords* need to be selected to make up a code, and only codewords are transmitted.

Definition The *weight* of a binary word is just the number of 1s in the word.

Definition The *distance* between two binary words is the number of positions in which the two words differ. With this, we can define the *minimum distance* in a code to be the smallest distance among all the distances between pairs of codewords.

Definition In practice, every codeword contained in a code is constructed with a certain number of *information positions*, with the remaining positions, typically labeled as *redundancy positions*, satisfying *parity-check equations*.

Definition A $m \times n$ matrix $\mathbf{G}$ is a *generator matrix* for the code C if C is the group generated by the rows of $\mathbf{G}$.

Definition *Encoding* a message is the process of adding the parity-check digits to the end of the message. If $\mathbf{x}$ is a message, then $E(\mathbf{x})$ denotes the encoded word. Furthermore, if G is the generator matrix of a code and $\mathbf{x}$ is a message, then $E(\mathbf{x})$ is equal to the product $\mathbf{xG}$.

Definition The *parity-check matrix* of a code C is a matrix that describes the linear relations that the components of a codeword must satisfy It can be used to decide whether a particular vector is a codeword and is also used in decoding algorithms. Specifically, the rows of a parity check matrix are the coefficients of the parity check equations. That is, they show how linear combinations of certain digits of each codeword equal zero.

Definition To *decode* a received word $\mathbf{x}$ means to find the codeword $\mathbf{a}$ closest to $\mathbf{x}$, that is, the codeword $\mathbf{a}$ such that the distance $d(\mathbf{a}, \mathbf{x})$ is a minimum.

Definition A *coset leader* is a word of minimum weight in any particular coset, that is, a word with the lowest amount of non-zero entries. Sometimes there are several words of equal minimum weight in a coset, and in that case, any one of those words may be chosen to be the coset leader.

Definition Let C be a code and let H be the parity-check matrix of C. If $\mathbf{x}$ is any word, $\mathbf{Hx}$ is called the *syndrome* of $\mathbf{x}$ and is denoted $\text{syn}(\mathbf{x})$.

Theorem 0.1. *Let H be the parity-check matrix of a code C in $\mathbb{B}^n$. A word $\mathbf{x}$ in $\mathbb{B}^n$ is a codeword if and only if $\mathbf{Hx} = 0$.*

K. Coding Theory: Coset Decoding

1 Let C_1 be the code consisting of the following binary words of length 5:

00000
00111
01001
01110
10011
10100
11010
11101

(a) List the elements in each of the cosets of C_1.

In general, when creating cosets, it is helpful to leave the information positions unchanged: only change the numbers in parity-check positions. So one can create every coset by adding every word that has nonzero elements only in parity-check positions to the code C. The number of unique cosets that are created is equal to the number of words that have nonzero elements only in parity-check positions.

Using this method, we found four cosets:

$$C_1 = C_1 + (00000)$$
$$C_2 = C_1 + (00001)$$
$$C_3 = C_1 + (00010)$$
$$C_4 = C_1 + (00011)$$

(b) Find the coset leader in each coset.(There may be more than one word of minimum weight in a coset; choose one of them as the coset leader.)

As a general strategy for finding the coset leader, for the nth coset $C_n = C_1 + a_n$, we find e_n by adding a_n to each element of C_n and then take the coset leader to be the result with the fewest non-zero elements.

Now, for this particular set, we see that elements of each coset are shown below:

C_1	C_2	C_3	C_4
$\boxed{00000}$	$\boxed{00001}$	$\boxed{00010}$	00011
00111	00110	00101	$\boxed{00100}$
01001	01000	01011	01010
01110	01111	01100	01101
10011	10010	10001	10000
10100	10101	10110	10111
11010	11011	11000	11001
11101	11100	11111	11110

It is not challenging to verify that the boxed word in each column above serve as the coset leader of coset represented in that column.

(c) Use the procedure described above to decode the following words $\mathbf{x}$: 11100, 01101, 11011, 00011.
To summarize the procedure used to decode a word $\mathbf{x}$, we first examine the coset $C + \mathbf{x}$ to find the coset leader $\mathbf{e}$, and then add $\mathbf{e}$ to $\mathbf{x}$ to produce the decoded word $\mathbf{e} + \mathbf{x}$.

<u>11100:</u>

$$C + 11100 = \begin{cases} 11100 \\ 11011 \\ 10101 \\ 10010 \\ 01111 \\ 01000 \\ 00110 \\ \boxed{00001 = \mathbf{e}} \quad \text{(one coset leader)} \end{cases}$$

Therefore, the decoded word is $\mathbf{e} + \mathbf{x} = 11101$.
<u>01101:</u>

$$C + 01101 = \begin{cases} 01101 \\ 01010 \\ \boxed{00100 = \mathbf{e}} \quad \text{(one coset leader)} \\ 00011 \\ 11110 \\ 11001 \\ 10111 \\ 10000 \end{cases}$$

Therefore, the decoded word is $\mathbf{e} + \mathbf{x} = 11101$.
<u>00011:</u>

$$C + 00011 = \begin{cases} 01101 \\ 01010 \\ \boxed{00100 = \mathbf{e}} \quad \text{(one coset leader)} \\ 00011 \\ 11110 \\ 11001 \\ 10111 \\ 10000 \end{cases}$$

Therefore, the decoded word is $\mathbf{e} + \mathbf{x} = 00111$.

2 Let C_1 be the following Hamming code in $\mathbb{B}^7$: the first four positions are information positions, and the parity-check equations are $a_5 = a_2 + a_3 + a_4$, $a_6 = a_1 + a_3 + a_4$, and $a_7 = a_1 + a_2 + a_4$. List the elements in each of the cosets of C_1 and find a coset leader in each coset. Then use coset decoding to decode the following words: $\mathbf{x} : 1100001, 0111011, 1001011$.

(a) Given the parity-check equations, let us first find the generator matrix $\mathbf{G_1}$, which will allow us to generate all the words in the code.

$$\mathbf{G_1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Since $\mathbf{G_1}$ is the generator matrix of the code C_1, we can generate the code words of C_1 by taking all the linear combinations of the rows of $\mathbf{G_1}$. That is , $C_1 = \mathbf{xG_1}$ for all $\mathbf{x} \in \mathbb{B}^4$. By this method, we have that C_1 contains the following 16 words:

0000000	1000011
0001111	1001100
0010110	1010101
0011001	1011010
0100101	1100110
0101010	1101001
0110011	1110000
0111100	1111111

Next, to find the nth coset of C_1 (there are 8 in total), we simply take $C_n = C_1 + (0,0,0,0|\mathbf{a_n})$ for all $\mathbf{a_n} \in \mathbb{B}^3$. In other words, we just add every word with nonzero elements only in the parity-check positions to the code C_1. Thus we obtain the following cosets:

C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
0000000	0000001	0000010	0000011	0000100	0000101	0000110	0000111
0001111	0001110	0001101	0001100	0001011	0001010	0001001	0001000
0010110	0010111	0010100	0010101	0010010	0010011	0010000	0010001
0011001	0011000	0011011	0011010	0011101	0011100	0011111	0011110
0100101	0100100	0100111	0100110	0100001	0100000	0100011	0100010
0101010	0101011	0101000	0101001	0101110	0101111	0101100	0101101
0110011	0110010	0110001	0110000	0110111	0110110	0110101	0110100
0111100	0111101	0111110	0111111	0111000	0111001	0111010	0111011
1000011	1000010	1000001	1000000	1000111	1000110	1000101	1000100
1001100	1001101	1001110	1001111	1001000	1001001	1001010	1001011
1010101	1010100	1010111	1010110	1010001	1010000	1010011	1010010
1011010	1011011	1011000	1011001	1011110	1011111	1011100	1011101
1100110	1100111	1100100	1100101	1100010	1100011	1100000	1100001
1101001	1101000	1101011	1101010	1101101	1101100	1101111	1101110
1110000	1110001	1110010	1110011	1110100	1110101	1110110	1110111
1111111	1111110	1111101	1111100	1111011	1111010	1111001	1111000

The word $1100001 + 0000000 = 1100001$ is an element of C_8. Therefore, $C_1 + 1100001 = C_8$. The leader of C_8 is 0001000 so the word 1100001 will be decoded to $0001000 + 1100001 = 1101001$.

The word $0111011 + 0000000 = 0111011$ is an element of C_8. Therefore, $C_1 + 0111011 = C_8$. The leader of C_8 is 0001000 so the word 0111011 will be decoded to $0001000 + 0111011 = 0110011$.

The word $1001011 + 0000000 = 1001011$ is an element of C_8. Therefore, $C_1 + 1001011 = C_8$. The leader of C_8 is 0001000 so the word 1001011 will be decoded to $0001000 + 1001011 = 1000011$.

3 Let C be a code and let $\mathbf{H}$ be the parity-check matrix of C. Prove that $\mathbf{x}$ and $\mathbf{y}$ are in the same coset of C if and only if $\mathbf{Hx} = \mathbf{Hy}$.

Proof. ($\Rightarrow$) To prove the forward direction, let $\mathbf{x}$ and $\mathbf{y}$ be elements of the same coset with the leader $\mathbf{e}$. The closest word to $\mathbf{x}$ is $\mathbf{a} = \mathbf{x} + \mathbf{e}$ and closest to $\mathbf{y}$ is $\mathbf{b} = \mathbf{y} + \mathbf{e}$. Since $\mathbf{a}$ and $\mathbf{b}$ are codewords of C, we can then express $\mathbf{x} = \mathbf{a} + \mathbf{e}$ and $\mathbf{y} = \mathbf{b} + \mathbf{e}$. Therefore, we have that

$$\mathbf{Hx} = \mathbf{H}(\mathbf{a} + \mathbf{e}) = \mathbf{Ha} + \mathbf{He} = 0 + \mathbf{He} \quad \text{and} \quad \mathbf{Hy} = \mathbf{H}(\mathbf{b} + \mathbf{e}) = \mathbf{Hb} + \mathbf{He} = 0 + \mathbf{He}.$$

So $\mathbf{Hx} = \mathbf{Hy} = \mathbf{He}$ when $\mathbf{x}$ and $\mathbf{y}$ are in the same coset.

($\Leftarrow$) Now, to prove the reverse direction, suppose that $\mathbf{Hx} = \mathbf{Hy}$. This implies that $\mathbf{H}(\mathbf{x} + \mathbf{y}) = 0$ which implies that $\mathbf{x} + \mathbf{y}$ is a codeword of C. So let $\mathbf{a}$ be the codeword of C that is equal to $\mathbf{x} + \mathbf{y}$. Then we can write $\mathbf{x}$ as $\mathbf{x} = \mathbf{a} + \mathbf{y}$ and $\mathbf{y}$ as $\mathbf{y} = 0 + \mathbf{y}$ where 0 is the identity element of the code C. So $\mathbf{x}$ and $\mathbf{y}$ are both elements of the same coset $C + \mathbf{y}$. So if $\mathbf{Hy} = \mathbf{Hx}$, then $\mathbf{x}$ and $\mathbf{y}$ are in the same coset. □

4 Let code C have q cosets, and let the coset leaders be $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_q}$. Explain why the following is true: To decode a received word $\mathbf{x}$, compare $\text{syn}(\mathbf{x})$ with $\text{syn}(\mathbf{e_1})$,..., $\text{syn}(\mathbf{e_q})$ and find the coset leader $\mathbf{e_i}$ such that $\text{syn}(\mathbf{x}) = \text{syn}(\mathbf{e_i})$. Then $\mathbf{x}$ is to be decoded to $\mathbf{x} + \mathbf{e_i}$.

If $\text{syn}(\mathbf{x}) = \text{syn}(\mathbf{e_i})$, then by what we have proven in problem three, $\mathbf{x}$ must be in the same coset as $\mathbf{e_i}$. Thus, we can find the codeword closest to $\mathbf{x}$ by finding $\mathbf{x} + \mathbf{e_i}$. It is explained in exercise K why we can do this.

5 Find the syndromes of the coset leaders in part 2. Then use the method of part 4 to decode the words, $\mathbf{x} = 1100001$ and $\mathbf{x} = 1001011$.

The coset leaders for part two are:

$$\mathbf{e_1} = 0000000$$
$$\mathbf{e_2} = 0000001$$
$$\mathbf{e_3} = 0000010$$
$$\mathbf{e_4} = 1000000$$
$$\mathbf{e_5} = 0000100$$
$$\mathbf{e_6} = 0100000$$
$$\mathbf{e_7} = 0010000$$
$$\mathbf{e_8} = 0001000$$

The parity-check matrix for part 2 is equal to:

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

We can find the syndrome of a coset leader by multiplying the parity-check matrix by the coset leader:

$$\text{syn}(\mathbf{e_1}) = 000$$
$$\text{syn}(\mathbf{e_2}) = 001$$
$$\text{syn}(\mathbf{e_3}) = 010$$
$$\text{syn}(\mathbf{e_4}) = 011$$
$$\text{syn}(\mathbf{e_5}) = 100$$
$$\text{syn}(\mathbf{e_6}) = 101$$
$$\text{syn}(\mathbf{e_7}) = 110$$
$$\text{syn}(\mathbf{e_8}) = 111$$

The syndrome of the word $\mathbf{x} = 1100001$ is $\mathbf{Hx} = 111$ which implies that $\mathbf{x} = 1100001$ is an element of a coset with the leader $\mathbf{e_8} = 0001000$. Thus, the word 1100001 should be decoded to $\mathbf{e_8} + \mathbf{x} = 1101001$.

The syndrome of the word $\mathbf{x} = 1001011$ is $\mathbf{Hx} = 111$ which implies that $\mathbf{x} = 1001011$ is an element of a coset with the leader $\mathbf{e_8} = 0001000$. Thus, the word 1001011 should be decoded to $\mathbf{e_8} + \mathbf{x} = 1000011$.