

The Cosets Code

Aubrey Neumann *

Summer 2010

1 Introduction

This paper will propose a new binary linear code, determine its fundamental properties and explore the effectiveness single iteration voting decoding. While this knowledge is useful in terms of transmission and decoding, even those with minimal background in codes may find value in the study of the code as a mathematical object.

We commence by introducing the basics of binary linear codes, beginning with the definition.

A binary linear code is a vector space in mod 2 formed by the left null space of an arbitrary binary matrix or *check matrix*.

We call each element of the vector space a *codeword* and each coordinate of the codeword a *bit*. The check matrix may be seen as giving a series of *check equations* describing how certain bits of any given codeword must add to 0.

Also, note that in mod 2 addition and subtraction are equivalent. [1]

*Lawrence University. This work was supported by The Lawrence University Distinctiveness Fund, The Lawrence University Chester Hill, Jr. Memorial Fund, and Professional Development funds under Professor Alan E. Parks.

1.1 Properties of Codes and Their Codewords

The *length* of a code is equal to the number of bits in one of its codewords. The *dimension* of a code is equal to k where 2^k is the number of codewords belonging to the code. Furthermore, the check matrix for this code will have *length – dimension* linearly independent rows.

The *weight* of a vector is the number of non-zero bits. The *distance* between two vectors is the number of bits in which they differ. Note, in binary the distance between two vectors is equal to the weight of their sum vector.

The *minimum distance* of a code is the distance between the two closest codewords in that code. This is equal to the minimum weight of a non-zero codeword.

Often times a code of length n , dimension k , and minimum distance d is called an $[n, k, d]$ code. [1]

2 The Question

We wish to determine the length, dimension and minimum distance of the *cosets code* defined below.¹Also, we wish to determine the probability of bit error after a single iteration of voting decoding.

2.1 Defining the Cosets Code

Let m, n be positive integers. Let $G = \mathbb{Z}_m^n$. Note that while G is a group unto itself, we will instead consider the group ring \mathbb{Z}_2G as a vector space in which to define this code. So if $x \in \mathbb{Z}_2G$, then

$$x = \sum_{g \in G} x_g \cdot g \quad \text{where} \quad x_g \in \mathbb{Z}_2$$

¹The cosets code was introduced to me by Professor Alan E. Parks under the auspice of a summer researchship.

In other words, if the coordinate $x_g = 1$ then g is present in x . The elements present in x form a subset of G . If $x, y \in \mathbb{Z}_2G$, then define the dot product on \mathbb{Z}_2G

$$x \circ y = \sum_{g \in G} x_g \cdot y_g \quad \text{in mod } 2$$

Let $e_j \in G$ be the j -th unit basis element, so that

$$e_j[i] = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Let $G_j = G/\langle e_j \rangle$, the cosets of $\langle e_j \rangle$. Note that while each $H \in G_j$ is a coset of G , H is also a single element of \mathbb{Z}_2G . We then define the code as the set of $x \in \mathbb{Z}_2G$ such that

$$x \circ H = 0 \quad \text{for all } H \in G_j \text{ and } 1 \leq j \leq n \quad (1)$$

The equations (1) are the check equations of the code.

2.2 Notation

Before we go further, it is necessary to introduce some new notation. We note that each element $a \in G$ is composed of n coordinates. Thus, we choose to denote each $a = (a_1, a_2, a_3, \dots, a_n)$. In a similar manner, for each $H \in G_j$ we write $H = H_j$ and note that

$$H_j = \sum_{a_i \in \mathbb{Z}_m} (a_1, \dots, a_n) \in \mathbb{Z}_2G$$

This notation implies that the values of a_i for $1 \leq i \leq n$ and $i \neq j$ are fixed but arbitrary. We write $H_{j,i}(a_i)$ to describe an H_j coset with a constant, a_i , in the i -th position and $i \neq j$.

Also, we define the set $(\mathbb{Z}_m)' = \{x \in \mathbb{Z}_m | x \neq 0\}$.

3 Length

The computation of length is trivial. From our definition of the group ring \mathbb{Z}_2G , the number of bits any given codeword is equal to $|G|$. There are m values for the n coordinates of any given $a \in G$. Thus, $|G| = m^n$ and the length of the cosets code is m^n .

4 Dimension

By definition, the dimension of this code is equal to the number of independent H subtracted from the length. Thus, we must compute the number of independent H .

4.1 Identifying Independent Cosets

To identify the independent cosets we define a set \mathbb{H} of all cosets H of the form $H = H_{j,i}(a_i)$ with $a_i \neq 0$ for all $1 \leq i < j$. We wish to show that \mathbb{H} forms a basis for the set of all H .

Proposition 1. \mathbb{H} spans all $H \in G_j$ for $1 \leq j \leq n$.

Proof. This proof requires the use of complete induction, so we begin by proving the base cases. First, we let $j = 1$. Since there exists no i such that $1 \leq i < j$, all $H_1 \in \mathbb{H}$. Thus, the proposition holds for $j = 1$.

Next, let $j = 2$. While this case is not necessary to the induction proof, it provides a concrete example of what later occurs in more abstract cases. Consider an arbitrary coset $H_2 \in G_2$. If $H_2 \in \mathbb{H}$, we are done. If not, then $a_1 = 0$ and we get

$$H_{2,1}(0) = \sum_{a_2 \in \mathbb{Z}_m} H_{1,2}(a_2) + \sum_{a_1 \in (\mathbb{Z}_m)'} H_{2,1}(a_1)$$

Note that all coordinates a_l with $l \neq 1, 2$ are held constant throughout the equation.²All $H_1 \in \mathbb{H}$ as do those H_2 with $a_1 \neq 0$, so our arbitrary coset is dependent on cosets in \mathbb{H} . Thus, the proposition holds for $j = 2$.

Now, let $j = p$ and assume the proposition holds for all cases less than p . Consider an arbitrary coset $H_p \in G_p$. If $H_p \in \mathbb{H}$, we are done. If not, then there exists a least k such that $a_k = 0$ and $1 \leq k < p$ and we get

$$H_{p,k}(0) = \sum_{a_p \in \mathbb{Z}_m} H_{k,p}(a_p) + \sum_{a_k \in (\mathbb{Z}_m)'} H_{p,k}(a_k)$$

Again note that all coordinates a_l with $l \neq k, p$ are held constant throughout the equation. By definition, $k < p$ so the proposition holds for all H_k . If a_k was the only $a_i = 0$, then all the H_p with $a_k \neq 0$ are also in \mathbb{H} , and we are done. If not, we choose a next least k' such that $a_{k'} = 0$ and $k < k' < p$ and repeat the processes for $H_{p,k,k'}(a_k, 0)$ for $a_k \in (\mathbb{Z}_m)'$.

Since there are a finite number of $i < p$, after only a finite number of repetitions it will be evident that our arbitrary coset is dependent on cosets of the proposed form. Thus, the proposition holds for $j = p$, and according to the principal of mathematical induction, the proposition is true for all j . \square

While we have now proved that all $H \notin \mathbb{H}$ are dependent upon $H \in \mathbb{H}$, we have yet to show that all $H \in \mathbb{H}$ are independent of each other. We do so with this next proposition.

Proposition 2. \mathbb{H} is an independent set.

Proof. Suppose otherwise, that some of the cosets in \mathbb{H} are dependent. Then, $0 = \sum_{h \in \mathbb{H}} (c_h h)$ with at least one constant $c_h \neq 0$. We choose a coset $h \in G_j$ where j is the least such that $c_h = 1$. Let y be the element present in h with $a_j = 0$. There must exist some $h' \in \mathbb{H}$ with y present in $h' \neq h$ and $c_{h'} = 1$ for the sum to be 0. Let $h' \in G_k$. Given our choice of j , there exist no h' with $k < j$. Since cosets are disjoint, there are also no h' with $k = j$ and y present. Similarly no $h' \in \mathbb{H}$ with $k > j$ will have y present. Thus, we have a contradiction as no such h' exists.

²For an example, see Appendix A

Our only assumption was the dependence of the cosets, therefore the cosets of \mathbb{H} must be independent. \square

Together *Propositions 1* and *2* do indeed prove that \mathbb{H} forms a basis for the set of all H .

4.2 Counting Independent Cosets

Now that we've identified them, we must determine the number of independent cosets or the order of \mathbb{H} .

We begin by determining the number of $H_p \in \mathbb{H}$. Each $a_i \neq 0$, but may take on any other value. So there are $m - 1$ values of the $p - 1$ coordinates a_i . Since we do not wish to double-count cosets, we will only consider 1 value of a_p . Finally, there are m values for the remaining $n - p$ terms. Thus, we get

$$(m - 1)^{p-1} m^{n-p} = \text{the number of } H_p \in \mathbb{H}$$

To determine the number of independent cosets in p quotient groups, it is a simple matter of using summation notation.

$$\sum_{i=1}^p (m - 1)^{i-1} m^{n-i} = \text{the number of } H \in \mathbb{H} \text{ from } p \text{ quotient groups}$$

Substituting n for p in the previous equation, we get

$$\sum_{i=1}^n (m - 1)^{i-1} m^{n-i} = \text{the order of } \mathbb{H}$$

However, we note the geometric series above and choose to simplify it. Thus, we get

$$\begin{aligned} \sum_{i=1}^n (m - 1)^{i-1} m^{n-i} &= (m^{n-1}) \sum_{i=0}^{n-1} \left(\frac{m - 1}{m} \right)^i \\ &= (m^{n-1}) \left(\frac{\left(\frac{m-1}{m} \right)^n - 1}{\frac{m-1}{m} - 1} \right) \\ &= m^n - (m - 1)^n \end{aligned}$$

So, we have

$$m^n - (m - 1)^n = \text{the number of independent } H \quad (2)$$

4.3 Conclusion

Now that we have both the length and the number of independent H we may subtract one from the other to get

$$m^n - (m^n - (m - 1)^n) = (m - 1)^n = \text{the dimension}$$

Thus, there are exactly $2^{(m-1)^n}$ codewords for the proposed code.

5 Minimum Distance

As previously stated, in binary the minimum distance of a code is equal to the minimum weight of a non-zero code word.

Proposition 3. *For the cosets code, the minimum weight of a non-zero code-word is 2^n .*

Proof. Let x be a codeword of minimum weight. We first want to show that x must at least have weight 2^n . In other words, 2^n elements of G must be present in x . We also note that for the dot product to be zero, there must be an even number of elements from each coset H present in x . Since x is non-zero, there must be at least one element $(a_1, a_2, a_3, \dots, a_n)$ present in x . This is a single element in an H_1 coset. Thus, a second element $(a'_1, a_2, a_3, \dots, a_n) \in H_1$ must also be present in x .

Each of these elements also exist in an H_2 coset. However,

$$(a_1, a_2, a_3, \dots, a_n) \in H_{2,1}(a_1) \quad \text{while} \quad (a'_1, a_2, a_3, \dots, a_n) \in H_{2,1}(a'_1)$$

Thus, a third and fourth element

$$(a_1, a'_2, a_3, \dots, a_n) \in H_{2,1}(a_1) \quad \text{and} \quad (a'_1, a''_2, a_3, \dots, a_n) \in H_{2,1}(a'_1)$$

must also be present in x .

With the added consideration of a new quotient group G_3 , we see that every two identified elements b, c present in x have $b_i \neq c_i$ for some $1 \leq i < 3$. So, either $b_1 \neq c_1$ or $b_2 \neq c_2$. Thus, the elements present in x exist in disjoint H_3 cosets and at least twice as many elements must be present in x to create even pairings. Repeating this doubling process for each of the n quotient groups gives us a minimum of 2^n elements present in x .

Note that if the prime coordinates of each element are equal to each other (ie $a'_2 = a''_2$) then the elements added to insure an even number of elements in H_k cosets will exist in pairs in H_i cosets for $1 \leq i < k$. Therefore, any x created in this manner would weigh exactly 2^n . \square

Thus, the minimum distance of the code is 2^n . The code can detect 2^{n-1} errors and correct $2^{n-1} - 1$ errors.

6 Decoding the Cosets Code

We wish to determine the probability that an arbitrary bit x will be decoded incorrectly after one iteration of voting decoding.

In voting decoding, each bit of a codeword is checked by taking the dot-product of the codeword itself with each check vector containing the bit. If the majority of the dotproducts are 1 then the "vote" is in favor of change and the bit is changed. Otherwise, the bit is not changed.[2]

6.1 A Single Iteration

After a single iteration, there are two ways for x to be in error-the voting decoding may fail to correct an error or it may cause new error. Both cases occur when the majority of check vectors involving the x bit have an odd number of errors in the remaining 1 bits(those non- x bits with value 1).

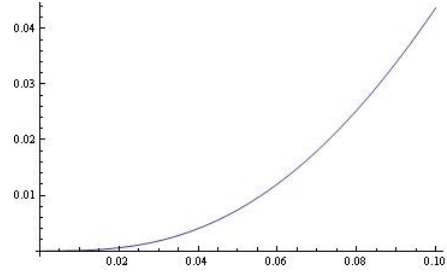
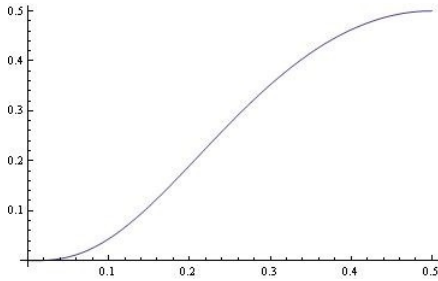
Therefore, we find the probability $g(p)$ that there will be an odd number of errors as a function of the probability of error, p .

$$g(p) = \sum_{y=0}^{\lfloor (m-2)/2 \rfloor} \binom{m-1}{2y+1} (p)^{2y+1} (1-p)^{m-1-(2y+1)}$$

We then find the probability $f(p)$ that the majority of the n check vectors with x present will have this quality. Note in the case where there are an equal number of odd error check vectors and even error check vectors, x won't be corrected if it is in error but neither will it be changed if it is correct. This accounts for the slight difference in summation parameters below.

$$f(p) = \sum_{x=\lfloor \frac{n}{2} + 1 \rfloor}^n \binom{n}{x} g(p)^x (1-g(p))^{n-x} (1-p) + \sum_{x=\lceil \frac{n}{2} \rceil}^n \binom{n}{x} g(p)^x (1-g(p))^{n-x} p$$

We then obtain graphs of the probability x will be decoded incorrectly when $m = 3$ and $n = 5$ as a function of p . We note that $f(p) < p$ for all $p < .21$ which is an extremely high error rate. We also calculate $h(.001) = 7.95214 \times 10^{-8}$ as a future reference point.



7 Summary

The cosets code is an $[m^n, (m-1)^n, 2^n]$ code, meaning it consists of $2^{(m-1)^n}$ codewords of length m^n and, in ideal decoding, the code can detect 2^{n-1} errors and correct $2^{n-1} - 1$ errors.

Voting decoding seems to be a promising algorithm for decoding the cosets code. Even after a single iteration the likelihood of error in a bit of the vector decreases dramatically.

A Clarifying Calculations

The following is an example of the equation seen in *Proposition 1* in section *4.1 Identifying Independent Cosets*.

Let $n = 2$ and $m = 3$. Then, there is only one coset $H_{2,1}(0)$ namely the one in which $(0,0)$, $(0,1)$ and $(0,2)$ are all present. Thus, the equation

$$H_{2,1}(0) = \sum_{a_2 \in \mathbb{Z}_m} H_{1,2}(a_2) + \sum_{a_1 \in (\mathbb{Z}_m)'} H_{2,1}(a_1)$$

may also be written

$$\begin{aligned} (0,0) + (0,1) + (0,2) &= (0,0) + (1,0) + (2,0) \\ &\quad + (0,1) + (1,1) + (2,1) \\ &\quad + (0,2) + (1,2) + (2,2) \\ &\quad + (1,0) + (1,1) + (1,2) \\ &\quad + (2,0) + (2,1) + (2,2) \end{aligned}$$

Note the addition above is being done in \mathbb{Z}_2G so the sum of the two $(1,0)$ elements is a zero vector, and they cancel each other. If we identify all the elements on the right side that cancel, the truth of the equation becomes evident.

$$\begin{aligned} (0,0) + (0,1) + (0,2) &= (0,0) + \cancel{(1,0)} + \cancel{(2,0)} \\ &\quad + (0,1) + \cancel{(1,1)} + \cancel{(2,1)} \\ &\quad + (0,2) + \cancel{(1,2)} + \cancel{(2,2)} \\ &\quad + \cancel{(1,0)} + \cancel{(1,1)} + \cancel{(1,2)} \\ &\quad + \cancel{(2,0)} + \cancel{(2,1)} + \cancel{(2,2)} \end{aligned}$$

References

- [1] F.J. MacWilliams and N.J.A. Sloane, The Theory of Error-Correcting Codes, *North-Holland Mathematical Library*, (Amsterdam, The Netherlands, 2006).
- [2] R.G. Gallager, Information Theory and Reliable Communication, (Wiley, New York, 1968).