UDC: 512.642/.643:004.422.61'415
Conference Paper

Using classes of equivalence in the detecting and correcting errors in the wordcode

Azir Jusufi [1], Xhevair Beqiri[1], Bukurie Imeri-Jusufi[2]

[1]*Department of Mathematics, Faculty of Natural Sciences and Mathematics, University of Tetova*
[2] *Departament of applied economic, Faculty of technical sciences University "Mother Teresa"- Skopje*
[*]*Corresponding Author: e-mail: azir.jusufi@unite.edu.mk*

Abstract

Developments in recent decades of the sector in digital communication have created a close relationship between mathematics and computer engineering. An important class of codes are linear codes in the vector space V (n, q) , where GF (q), is a finite field of order q. Let C be a linear [n, k] code. A generator matrix of C is a k × n matrix such that the rows are basis vectors of C [n, k]. If we specify V(n,q) as a vector space on then the linear binary code C[n,k] is nothing else, but a subspace of the vector space V(n,q). When transferring the wordcodes through different types of obstacle channels, errors may occur, which we need to detect and correct. With classes a+C, where a is a vector from V(n, q), we construct the standard group, which we will use to detect and correct the errors in the wordcode. If the word code c is sent, but the received vector is r, we define the error vector e = r − c. The error vector which will be corrected are precisely the coset leaders, irrespective of which wordcode is transmitted. By choosing a minimum weight vector in each coset as coset leader, we ensure that standard array decoding is a nearest neighbour decoding scheme.

Keywords: Vector space, generator matrix code, wordcode, detection, class.

1. Vector Spaces and General Linear Codes

In order to develop the theory for general linear codes, we need some definitions from abstract and linear algebra. The binary alphabet, along with modulo 2 addition and multiplication, is the smallest example of a field.

Defnition 1.1. Let F be a field. A set V of elements called vectors is a *vector space* if for any u,v,w $\in V$ and for any $c,d \in F$:
1. $u + v \in V$
2. $cv \in V$
3. $u + v = v + u$.
4. $u + (v + w) = (u + v) + w$
5. $\exists\ 0 \in V$ such that $u + 0 = u$.
6. $\forall\ u \in V$, $\exists\ -u \in V$, $u + (-u) = 0$.
7. $c(u + v) = cu = cv$.
8. $(c + d)u = cu + du$
9. $c(du) = (cd)u$.
10. $1u = u$.

We often say that F is the field of *scalars*. An important example in coding theory is the vector space of all binary n-tuples. Here, the vectors are the binary n-tuples and the field is $GF(2)=\{0,1\}$. We say that 0 and 1 are the scalars. Note V (n,q) denotes the vectorial space of vectors with length n on the field GF {q}.

Example 1.1. Let's mark it V(3,2)={000,100,010,001,110,101,011,111} a vector space over GF {2}.

We will use the vector subspaces more often than the above formal definition of a vector space.

Defnition 1.2. Let V be a vector space. A subset U of V is called a *subspace* of V if U is itself a vector space under the same operations as V.

Theorem 1.1 ([1].p.148) Let V be a vector space and F be a field. A subset U of V is a subspace if:

1. $\forall$ u,v $\in U$, u + v $\in U$;
2. $\forall c \in F$, u $\in U$, $\forall c$u $\in U$.

The vectors $u_1, u_2, \ldots, u_n$ of a vector space V are said to be *linearly dependent*, if there exist a *finite* number scalars $c_1, c_2, \ldots, c_n$, not all zero, such that $c_1u_1 + c_2u_2 + \ldots + c_nu_n = 0$. where 0 denotes the zero vector.

The vectors $u_1, u_2, \ldots, u_n$ of a vector space V are said to be *linearly independent* if the equation $c_1u_1 + c_2u_2 + \ldots + c_nu_n = 0$ can only be satisfied $c_i = 0$ for i=1,2,…,n.

Definition 1.3. A *basis* is a (minimal cardinality) set of linearly independent vectors that span the vector space. The *dimension* of a vector space V is number of vectors in any basis for V .

Definition 1.4. Let V be a vector space over a field F and let W be a subset of V . If W is a vector space over F, then W is a *vector subspace* of V .

Theorem 1.2. ([2], p.196) If W is a subset of a vector space V over a field F, then W is a vector subspace of V if $a\boldsymbol{u} + b\boldsymbol{v} \in W$ for all $a,b \in F$ and all $u,v \in W$.

Proof. Suppose that W is a vector subspace of V , a vector space over the field F. Then, W is itself a vector space over F. Therefore, by the definition of vector spaces v +u $\in W$ for any v, u $\in W$ and $av \in W$ for any $a \in F$. This directly implies that $a\boldsymbol{u} + b\boldsymbol{v} \in W$ for all $a,b \in F$ and all $u,v \in W$..

Now suppose that $a\boldsymbol{u} + b\boldsymbol{v} \in W$ for all $a,b \in F$ and all $u,v \in W$, then, setting $a=b=1 \in F$ gives the first requirement of a vector space and setting $a = 1 \in F$ and $b= 0 \in F$ gives the second requirement of a vector space.

2. Binary Linear Codes

Definition 2.1. A *binary linear code C* of length n is a set of binary *n*-tuples such that the component wise modulo 2 sum of any two wordcodes is contained in *C*.

Example 2.1.

Let'sV(3,2)={000,100,010,001,110,101,011,111} a vector space over GF {2}.

The set C(3,2)={000,100,001,101} is a subset of the V(3,2),

we see that 000+000=000 ∈ C, 000+100=100 ∈ C, 000+001=001 ∈ C, 000+101=101 ∈ C,

100+100=000 ∈ C, 100+001=101 ∈ C, 100+101=001 ∈ C, 001+001=000 ∈ C, 001+101=100 ∈ C, 101+101=000 ∈ C.

so C is the binary linear code. Also C completes the conditions of the subspaces.

Definition 2.2. A *linear code C* of length *n* over *GF*(*q*) is a subspace of the vector space *V* (*n, q*).

Definition 2.3. The *Hamming weight* *w*(c) of a wordcode c is the number of nonzero components in the wordcode.

Example 2.2. *w*(0000) = 0, *w*(111) = 3, *w*(1011001) = 4.

Definition 2.3. The *Hamming distance* between two wordcodes *d*(x*;* y) is the number of places in which the wordcodes x and y differenc. In other words, *d*(x*;* y) is the Hamming weight of the vector x -y, representing the component-wise difference of the vectors x and y.

Example 2.3. *d*(0001*;* 1110) = 4,

Definition 2.4. The *minimum (Hamming) distance* of a code *C* is the minimum distance between any two wordcodes in the code: $d(C) = \min \{d(x; y) \mid x \neq y;\ x, y \in C\}$.

To easily calculate the distance of code C, we get

Theorem 2.1.([6], p.6) The minimum distance, *d*(*C*), of a linear code *C* is equal to *w**(*c*), the weight of the lowest-weight nonzero wordcode.

Proof. There exist wordcodes x and y in *C* such that *d*(*C*) = *d*(x y). By the definition of Hamming distance, we can rewrite this as *d*(*C*) = *w*(x-y). Note that x-y is a wordcode in *C* by the linearity of *C*. Therefore, since *w**(*c*) is the weight of the lowest weight wordcode, we have *w**(*c*) ≤ *w*(x-y) =*d*(*C*).

On the other hand, there exists some wordcode c ∈ *C* such that *w**(*c*) = *w*(c). By the definition of the weight of a wordcode, we can write *w*(c) = *d*(c,0). Since c and 0 are both wordcodes, the distance between them must be greater than or equal to the minimum distance of the code: *d*(c, 0) ≥ *d*(*C*). Stringing together these (in) equalities, shows that *w**(*C*) ≥ *d*(*C*).

Since we have now shown that both *d*(*C*) ≥ *w**(*c*) and *d*(*C*) ≤ *w**(*c*), we conclude that *d*(*C*) =*w**(*c*).

Example 2.4. The minimum distance, *d*(*C*), of a linear code *C*={000,100,011,111} is *d*(*C*) =*w**(*100*)=1.

3. Generator Matrices. Equation check matrices

Definition 3.1. A *kxn*-matrix *G* whose rows form a basis for an [*n; k*] linear code *C* is called a *generator matrix* of the code *C*.

Example 3.1. Suppose we need a [3,2] binary linear code *C*, that is, we need a 2-dimensional vector subspace of *V* (3*;* 2), the vector space of binary 3-tuples. In order to defne a 2-dimensional vector subspace of *V* (3*;* 2), we need 2 linearly independent basis vectors. Arbitrarily, let's choose these vectors to be 101 and 010. Let *C* be spanned by these vectors. Then the generator matrix is

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

At first we see if the rows are *linearly independent* :

$\alpha(101) + \beta(010) = 0 \Rightarrow \alpha = \beta = 0$

Taking all the linear combinations of the rows in *G*, we generate the code $C = \{000,101,010,111\}$.

If *generator matrix G is* $G = [I_k \mid P]$ *then it is called in* ***standard form.***

Example 3.2. $G = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = [I_2 \mid P] \Rightarrow P = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Definition 3.2. Let *C* be an [*n; k*] linear code. A *equation check matrix* for *C* is an (*n* - *k*) *x n* matrix *H* such that c $\in C$ if and only if c$H^T = 0$.

The generator matrix can be used to construct an equation control matrix for a code and vice versa. If the generator matrix G is in the standard form $G = [I_k \mid P]$,then the equation control matrix is $H = [P^T \mid I_{n-k}]$. This is a conclusion of the fact that the equation control matrix by C is a generator matrix of the dual code $C^{\perp}$ of C .

Example 3.3. The equation control matrix of $G = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = [I_2 \mid P] \Rightarrow P = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is $H = [P^T \mid I_{2-1}] = [101]$.

4. Classes of equivalence. The detecting and correcting errors in the word code

The defining equivalence of a equation check matrix (c $\in C$ $\Leftrightarrow$ c$H^T = 0$) is used in the first step of the decoding of linear codes.

We mark it: c- sent vector, r- received vector , e- error vector.

Note that c + e = r (mod2)

Upon receiving a vector *r*, the product rH^T is computed. If the result is 0, we assume that no errors have occurred and decode the received vector as *r*.

But if $rH^T \neq 0$?

Definition 4.1. Let *C* is an [*n; k*] linear code over *GF*(*q*) and that *a* is any vector in *V* (*n, q*). The set a + *C* = {a + c | c $\in C$ } is called a classe of equivalence of *C*.

$a \in V(n,q)$ is the class representative (leader) and usually takes the minimum weight.

Example 4.1. Let *C* be the binary [4*;* 2] linear code with generator matrix $G = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$.

Then C={0000,1001,0111,1110} and the classes of equivalence of *C* are

- 0000 + *C* = {0000, 1001, 0111, 1110}
- 1000 + *C* = {1000, 0001, 1111, 0110}
- 0100 + *C* = {0100, 1101, 0011, 1010}
- 0010 + *C* = {0010, 1011, 0101, 1100}

if we find it 0001 + *C* = {0001, 1000, 0110, 1111}=1000 + *C* .

Since two equivalence classes are either disjoint or matching, we conclude that the only classes are 0000+C, 1000+C, 0100+C and 0010+C.

In order to carry out the above decoding algorithm, we make a standard array by listing the classes' elements of C in rows, where the first entry in each row is the class leader.

Now, we make a standard array by listing the classes' elements of C in rows, where the first entry in each row is the class leader:

0000, 1001, 0111, 1110
1000, 0001, 1111, 0110
0100, 1101, 0011, 1010
0010, 1011, 0101, 1100

Definition 4.2. Expression S (c) = cH^T, we call syndrome of vector c .
Theorem 4.1. ([5], p.30) For a binary code, the syndrome of a received wordcode is a vector that is equal to the sum of the columns of H that correspond to the positions where errors occurred.
Proof. Let r be a received wordcode with syndrome $S(r) = rH^T$. Since r = c+e, where c is a valid wordcode and e is an error vector, we have

$$S(r) = rH^T = (c + e)H^T = cHT + eH^T = 0 + eH^T = eH^T$$

The equation control matrix of

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

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

Next, we count the syndrome of the row leader in the standard array table and add them to the end of the respective row:

- [0000] H^T =[00]
- [1000] H^T =[01]
- [0100] H^T =[11]
- [0010] H^T =[10]

We take:

0000, 1001, 0111, 1110, **00**
1000, 0001, 1111, 0110, **01**
0100, 1101, 0011, 1010, **11**
0010, 1011, 0101, 1100, **10**

The syndrome decoding algorithm works as follows. When a vector r is received, calculate the syndrome $S(r) = rH^T$. Locate the syndrome in the table, which requires looking through only one column. Decode r as r-e where e is the class leader in the row containing $S(r)$. This is equivalent to decoding r as the wordcode at the top of the column containing r.

Example 4.2. Suppose 1011 is received vector, and compute $S(1011) = [1110] \cdot \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = [10]$

We find this syndrome in the fourth row of the syndrome column. The class leader for this row is 0010, and so we decode 1011 as 1011 -0010 = 1001.

References

[1]. Bashkim Gazidede, "Algjebra 2", SH.B Ilar, Tiranë, 2006
[2]. David Cherney, Tom Denton, Rohit Thomas and Andrew Waldron, "Linear Algebra" First Edition. Davis California, 2013.
[3]. R. Hill, A flrst course in coding theory, Oxford University Press, New York, flrst ed., 1986.
[4]. Steven J. Leon, "Linear Algebra with Applications", Eighth Edition, University of Massachusetts, Dartmouth
[5]. Sarah Spence Adams, "Introduction to Algebraic Coding Theory" , 2008
[6]. Yehuda Lindell, "Introduction to Coding Theory Lecture Notes", Department of Computer Science Bar-Ilan University, Israel January 25, 2010
[7]. E. R.: Algebraic Coding Theory. New York: McGraw-Hill, 1968