Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi
Sayı: 16 Eylül 2008 ISSN - 1302 - 3055

ERROR CORRECTION IN LINEAR CODES WITH COMPUTER

Aydın SEÇER* & Ayten ÖZKAN* & Mustafa BAYRAM**

*Ataturk University, Faculty of Arts and Sciences, Department of Mathematics, Erzurum/Turkey
e-mail: asecer@atauni.edu.tr,

**Yildiz Tecnical University, Faculty of Arts and Sciences, Department of Mathematics, 34210-Davutpasa-Istanbul/Turkey
e-mail: msbayram@yildiz.edu.tr

Geliş tarihi: 12.06.2008 Kabul tarihi: 25.07.2008

ABSTRACT

Within the context of this study, we have improved a computer program on the syndrome decoding method for correcting codewords received incorrect. This program generates code from given generator matrix and calculates the hamming distance which appertains to this code, also, it finds Slepian(1960) standart array. It gives a list of the decoding table of code and an error pattern relating to be received incorrect codewords. We have used Maple Computer Algebra for calculations [5]. The Algorithm, we have given, reproduces different results from given generator matrix. In here, we have chosen a test problem whose code word's lengthy is 4. But, it is possible for algorithm that it can make calculations for longer codes.

Key Words: *Linear Code, Sydrome Decoding, Coset, Coset Leader, MAPLE.*
AMS (2000) Clasification: 94B05.

1. INTRODUCTION

Coding theory is the branch of mathematics concerned with accurate and efficient transfer of data across noisy channels as the theory of message sent. A transmission channel is the physical medium through which the information is transmitted, such as telephone lines, or atmosphere in the case of wireless comunication. Undesirable disturbance (noise) can occur across the comunication channel, causing the received information to be different from the original information sent [6]. Another prominent example is the ubiquitous CD: a scratch on a music CD has no effect on sound quality, completely unlike phonograph records. Coding theory deals with detection and correction of transmission errors caused by noise in the channel. The primary goal of coding theory is efficient encoding of information, easy transmission of encoded messages, fast decoding of received information and correction of errors introduced in the channel [2]. Coding theory is used all the time: in reading CDs, receiving transmissions from satellites, or in cell phones[3].
The purpose of our study is to implement error detection and correction into computer applications in Linear Codes which some of the most important error correcting codes of Coding Theory, using computer programming language like *MAPLE Computer Algebra System*[5].

Linear codes are error correcting codes whose codewords form a vector space. A vector space is a collection of vectors, here codewords, which is closed under vector addition and scalar multiplications, in other words contains all the possible linear combinations of its codewords. If the vector space of all codewords is n dimensional and the subspace formed by codewords of a code k dimensional then the code is described as an (n,k) - linear code (the dimension of a vector space is given by the number of vectors in the basis of the vector space; a basis for a vector space or subspace is a minimum collection of linearly independents vectors which generate the entire vector space of subspace; linearly independent vectors are vectors such that none of them can be expressed as a sum of multiples of the others) [2].

There are two ways of describing a linear code C. The first is using a generator matrix G which has as its rows a set of basis vectors of the linear subspace C. For every linear code there is an equivalent code which has

generator matrix of the form $G=\left[I_k \mid A\right]$, where I_k is the $k\times k$ identity matrix and *A* is a $k\times n-k$ matrix. The second description of a linear code *C* consists in specifying not vectors in *C* but rather the vectors orthogonal to *C*. The orthogonal complement of *C* is a subspace and in fact is another code called the dual code of *C* denoted by $C^{\perp}$. If *H* is generator matrix for $C^{\perp}$ then *H* is a $(n-k)\times n$ matrix and it is called a parity check matrix for *C*. In general the rows of the parity check matrix for a code *C* are orthogonal to all codewords of *C* ($G\cdot H^T=0$) and any matrix *H* is a parity check matrix if the rows of *H* generate the dual code of *C*. Therefore, a code *C* is defined by such a parity check matrix *H* following way:

$$C=\left\{x\in V(n,q)\mid x\cdot H^T=0\right\}. \tag{1.1}$$

The parity check matrix can be obtained from the null space of generator matrix and the generator matrix can be obtained by finding the null space of parity check matrix.

Let *C* be a code over *q*-ary alphabet (alphabet consisting of *q* code symbols, for example a binary alphabet consist of two code symbols such as 0 and 1). If *C* has *M* codewords and minimum distance (distance between codewords is number of distinct code symbols) *d*, it is called an (n,M,d)-code [3].

Definition 1.1. The minimum distance of a code is defined following way;

$$d(C)=\min\left\{d(x,y)\middle| x,y\in C, x\neq y\right\} \tag{1.2}$$

Definition 1.2. Let F_q^n denote the vector space of all *n*-tuples over finite field F_q. The weight $W(x)$ of a vector $x\in F_q^n$ is the number of nonzero coordinates in *x* [4].

Theorem 1. (i) A code *C* can detect up to *s* errors in any codewords if $d(C)\geq s+1$.

(ii) A code *C* can correct up to *t* errors in any codewords if $d(C)\geq 2t+1$ [3].

Corollary 1. If a code *C* has minimum distance *d*, then *C* can detect up to $d-1$ errors in any codewords and correct up to $\left[\frac{(d-1)}{2}\right]$ errors in any codewords [3].

Theorem 2. If $G=\left[I_k \mid A\right]$ is a generator matrix for for the $\left[n,k\right]$ code *C* in standard form, then

$$H=\left[-A^T \mid I_{n-k}\right] \tag{1.3}$$

is a parity check matrix for *C* [4].

Theorem 3. Two $k\times n$ matrices generate equivalent linear $\left[n,k,d\right]$-codes over $GF(q)$ (Galois Field) if one of matrix can be obtained from the other by a sequence of operation of following types [3];

(i) Permutation of rows,
(ii) Multiplication of a row by a non-zero scalar,
(iii) Addition of a scalar multiple of one row to another.
(iv) Permutation of columns,
(v) Multiplication of any column by a non-zero scalar [6].

Definition 1.3. Suppose that *C* is an $\left[n,k\right]$-code over $GF(q)$ and that a is any vector in $V(n,q)$. Then the set $a+C$ defined by

$$a+C=\left\{a+x\mid x\in C\right\} \tag{1.4}$$

is called a *coset* of *C* [3].

Lemma 2. Suppose that $a+C$ is a coset of *C* and that $b \in a+C$. Then,

$$b+C = a+C \tag{1.5}$$

[3].

Theorem 4. (Lagrange) Suppose *C* is an $[n,k]$-code over $GF(q)$. Then

(i) Every vector of $V(q,n)$ is in some coset of *C*,

(ii) Every cosets contains exactly q^k vector,

(iii) Two cosets either are disjoint or coincide [3]

Definition 1.4. (Coset Leader) The vector having minimum weight in a coset is called the *coset leader*. (If there is more than one vector with the minimum weigh, we choose one at random and call it the coset leader).

Therorem 4 shows that $V(n,q)$ is partitioned into disjoint cosets of *C*:

$$V(n,q) = (0+C) \cup (a_1+C) \cup ... \cup (a_s+C) \tag{1.6}$$

where $s = q^{n-k} - 1$, and, by Lemma 2 we may take $0, a_1, a_2, a_3, ..., a_s$, to be the coset leaders. A (Slepian 1960) *standart array* for an $[n,k]$-code *C* is a $q^{n-k} \times q^k$ array of all the vectors in $V(n,q)$ in which the first row consists of code *C* with 0 on the extreme left and the other rows are cosets $a_i + C$ each arranged in corresponding order with the coset leader on the left [7]

2. ENCODING WITH A LINEAR CODE

Let *C* be an $[n,k]$-code over $GF(q)$ with generator matrix *G*. *C* contains q^k codewords and so can be used to communicative any one of q^k distinct messages. We identify these messages with the q^k *k*-tuples of $V(k,q)$ and we *encode* a message vector $u = u_1u_2u_3...u_k$ simply by multiplying it on the right by *G*. If the rows of *G* are $r_1, r_2, ..., r_k$, then

$$u \cdot G = \sum_{i=1}^{k} u_i r_i \tag{2.1}$$

and so $u \cdot G$ is indeed a codeword of *C*, being a linear combination of the rows of generator matrix. Note that the encoding function $u \to u \cdot G$ maps the vector space $V(k,q)$ on to a *k*-dimensional subspace of $V(n,q)$.

The encoding rule is even simpler if *G* is standard form. Suppose $G = [I_k \mid A]$, where $A = [a_{ij}]$ is a $k \times (n-k)$ matrix. Then message vector *u* is encoded as

$$x = u \cdot G = x_1x_2x_3...x_kx_{k+1}...x_n, \tag{2.2}$$

where $x_i = u_i$, $1 \le i \le k$, are message digits and

$$x_{k+i} = \sum_{j=1}^{k} a_{ji}u_j, \quad 1 \le i \le n-k \tag{2.3}$$

are the *check digits*. The check digits represent *redundancy* which has been added to the message to give protection against noise [3].

Lemma 2. Suppose that $a+C$ is a coset of *C* and that $b \in a+C$. Then,

$$b+C = a+C \tag{1.5}$$

[3].

Theorem 4. (Lagrange) Suppose *C* is an $[n,k]$-code over $GF(q)$. Then

(i) Every vector of $V(q,n)$ is in some coset of *C*,

(ii) Every cosets contains exactly q^k vector,

(iii) Two cosets either are disjoint or coincide [3]

Definition 1.4. (Coset Leader) The vector having minimum weight in a coset is called the *coset leader*. (If there is more than one vector with the minimum weigh, we choose one at random and call it the coset leader).

Therorem 4 shows that $V(n,q)$ is partitioned into disjoint cosets of *C*:

$$V(n,q) = (0+C) \cup (a_1+C) \cup ... \cup (a_s+C) \tag{1.6}$$

where $s = q^{n-k} - 1$, and, by Lemma 2 we may take $0, a_1, a_2, a_3, ..., a_s$, to be the coset leaders. A (Slepian 1960) *standart array* for an $[n,k]$-code *C* is a $q^{n-k} \times q^k$ array of all the vectors in $V(n,q)$ in which the first row consists of code *C* with 0 on the extreme left and the other rows are cosets $a_i + C$ each arranged in corresponding order with the coset leader on the left [7]

2. ENCODING WITH A LINEAR CODE

Let *C* be an $[n,k]$-code over $GF(q)$ with generator matrix *G*. *C* contains q^k codewords and so can be used to communicative any one of q^k distinct messages. We identify these messages with the q^k *k*-tuples of $V(k,q)$ and we *encode* a message vector $u = u_1u_2u_3...u_k$ simply by multiplying it on the right by *G*. If the rows of *G* are $r_1, r_2, ..., r_k$, then

$$u \cdot G = \sum_{i=1}^{k} u_i r_i \tag{2.1}$$

and so $u \cdot G$ is indeed a codeword of *C*, being a linear combination of the rows of generator matrix. Note that the encoding function $u \to u \cdot G$ maps the vector space $V(k,q)$ on to a *k*-dimensional subspace of $V(n,q)$.

The encoding rule is even simpler if *G* is standard form. Suppose $G = [I_k \mid A]$, where $A = [a_{ij}]$ is a $k \times (n-k)$ matrix. Then message vector *u* is encoded as

$$x = u \cdot G = x_1x_2x_3...x_kx_{k+1}...x_n, \tag{2.2}$$

where $x_i = u_i$, $1 \le i \le k$, are message digits and

$$x_{k+i} = \sum_{j=1}^{k} a_{ji}u_j, \quad 1 \le i \le n-k \tag{2.3}$$

are the *check digits*. The check digits represent *redundancy* which has been added to the message to give protection against noise [3].

Example 1. Let *C* be the binary $[7,4]$-code with generator matrix

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

If we apply a sequence of operations in Theorem 3 to this matrix we can get the standard form of *G* as

$$G=\left[\begin{array}{cccc|ccc}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{array}\right]$$

Thus, that $u=(u_1u_2u_3u_4)$ message vector is encoded as

$$u\cdot G=(u_1,u_2,u_3,u_4,u_1+u_2+u_3,u_2+u_3+u_4,u_1+u_2+u_4) \tag{2.4}$$

$$\boxed{\textit{Message source}} \xrightarrow[u=u_1u_2u_3\ldots u_k]{\text{Message vector}} \boxed{\begin{array}{l}\textit{Encoder}\\ u\to uG=x\end{array}} \xrightarrow[x=x_1x_2x_3\ldots x_n]{\text{Codeword}} \boxed{\begin{array}{c}\downarrow\text{Noise}\\ \textit{Channel}\end{array}} \to \ldots.$$

Figure 2.1. Linear Encoding.

3. DECODING WITH A LINEAR CODE

Suppose the codeword $x=x_1x_2\ldots x_n$ is sent through the channel and that the received vector is $y=y_1y_2\ldots y_n$. We define the *error vector e* to be

$$e=y-x=e_1e_2\ldots e_n \tag{2.5}$$

The decoder must decide from *y* which error vector *e* has occured. An elegant nearest neighbour decoding scheme for linear codes, devised by Slepian (1960), uses the fact that a linear code is a subgroup of the additive group $V(n,q)$ [3].

4. SYNDROME DECODING

Suppose *H* is a parity check matrix of an $[n,k]$-code *C*. Then for any vector $y\in V(n,q)$, the $1\times(n-k)$ row vector

$$S(y)=yH^T \tag{2.6}$$

is called *syndrome* of *y* [6]. If the rows of *H* are $h_1,h_2,h_3,\ldots,h_{n-k}$, then

$$S(y)=(yh_1,yh_2,yh_3,\ldots,yh_{n-k}) \tag{2.7}$$

and

$$S(y)=0\Leftrightarrow y\in C. \tag{3.10}$$

Lemma 4.1. Two vectors *u* and *v* are in the sme coset of *C* if and only if they have the same syndromes [2].

Corollary 4.1. There is one-to-one correspondence between cosets and syndromes [2].

COROLLARIES

We have given main theorems and definitions of the coding theory by now. We explained how can be encoded and corrected linear codes using generator matrix and standatd array. We have calculated procedures to encode and correct codewords which are received incorrect. In some situation doing this calculation without using computer may be impossible because of the lenght of code. Thus, we have improved a computer program which calculate this processes as quickly and efficiently. At the same time our program can encode and correct longer code. We have used Maple Computer Algebra System for all calculations.

Now we are giving a test problem for hand calculation. Then we will use it in our algorithm.

TEST PROBLEM

Let *C* be a $[n,k]$-code with the generator matrix as following;

$$G=\begin{bmatrix}1&0&1&1\\0&1&0&1\end{bmatrix}_{2\times 4},k=2,n=4 \tag{4.1}$$

and so by theorem 3 a parity check matrix is

$$H=\begin{bmatrix}1&0&1&0\\1&1&0&1\end{bmatrix}. \tag{4.2}$$

Hence the Syndromes of coset leaders are

$$\left.\begin{aligned}S(0000)=(0000)H^T=(00),S(1000)=(1000)H^T=(11)\\S(0100)=(0100)H^T=(01),S(0010)=(0010)H^T=(10)\end{aligned}\right\}. \tag{4.3}$$

The standart array become:

Table 4.1. Standart array

Coset Leaders				Syndomes
0000	1011	0101	1110	00
1000	0011	1101	0110	11
0100	1111	0001	1010	01
0010	1011	0111	1100	10

The decoding algorithm is now: when a vector is *y* received, calculate $S(y)=yH^T$ and locate $S(y)$ in syndromes column of the array. Locate *y* in the corresponding row and and decode as the codeword at the top of the column containing *y*. If 1111 is received, $S(1111)=01$, and so 1111 occurs in third row of the array. This is called a *syndrome look-up table*[7]. The syndrome look-up table for this code is

Syndromes (*y*)		Coset Leaders $f(y)$
00		0000
11		1000
01		0100
10		0010

The decoding procedure: (We also used this procedure in our program.)

Step 1: For a received vector *y* calculate $S(y)=yH^T$.

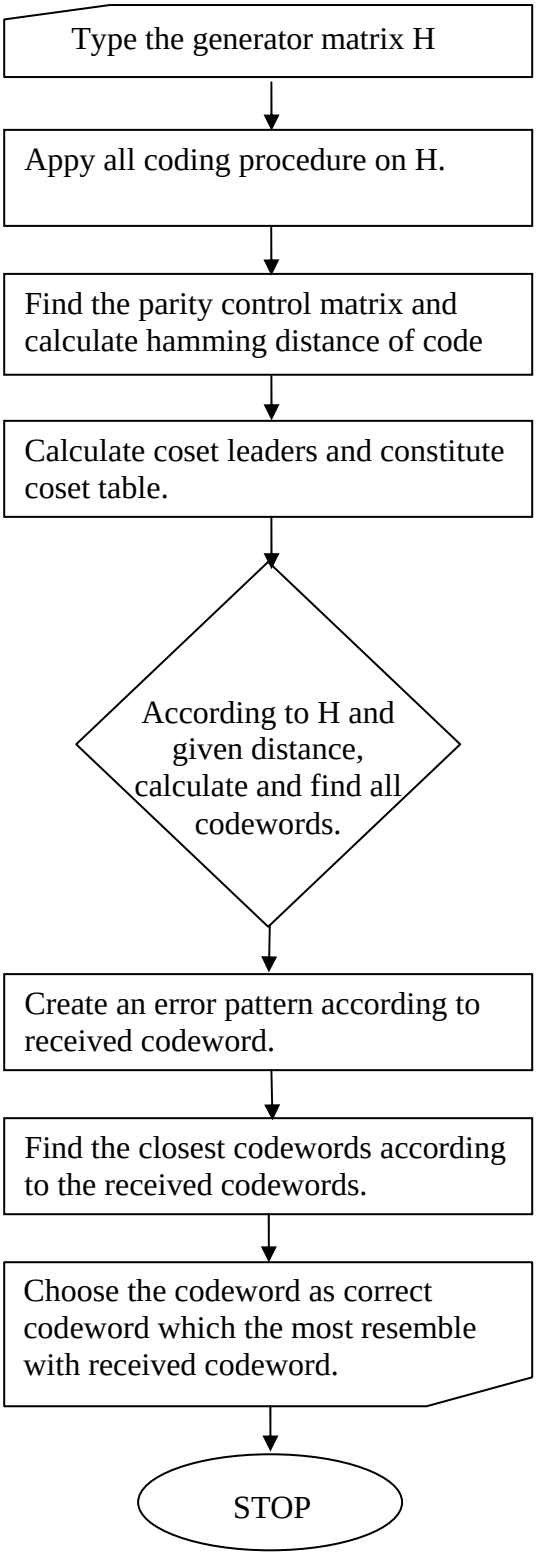
Step 2: Let $z=S(y)$, and locate *z* in the first column of the look-up table.

Step 3: Decode *y* as $y-f(z)$.

We use now this algorithm for decoding test problem: If we $y=1111$, then $S(y)=01$ and we decode as $1111-0100=1011$.

Now let us give an algorithm which corrects all incorrect codewords.

4.3. Decoding Algorithm



4.4. The results of our program

We have given only main procedures follows

H2:=mat(["1010","1101"]); (Generator matrix)

$$H2 := \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

C2:=KOD(H2);

HAMMINGMESAFESI(C2); (Hamming distance of code 2)

KODADONUSTUR(C2); (Code Convertor)

$$\{[1, 1, 0, 1, 1], [0, 0, 1, 1, 1], [1, 0, 1, 0, 1], [1, 1, 1, 0, 0]\}$$

KOSETTABLOSU1(H2); (Coset table)

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

KOSETTABLOSU2(H2);

```
Syndrome     Coset Leader
  11              1000
  10              0010
  00              0000
  01              0001
```

COZUMTABLOSU1(C2); (decoding table of C2)

```
Code: C2={0101, 0000, 1011, 1110}
```

Table 4.5. Decoding Table, (*) 4.1. Test problem, (**) 4.2. Test Problem.

Received Codewords	**Error Pattern**	**Corrected Codewords**
1101 (**)	**[1101, 0110, 1000, 0011]**	**0101**
0000	[0000, 1011, 0101, 1110]	0000
0111 (*)	**[0111, 1100, 0010, 1001]**	**0101**
0001	[0001, 1010, 0100, 1111]	0000
1011	[1011, 0000, 1110, 0101]	1011
0110	[0110, 1101, 0011, 1000]	1110
0011	[0011, 1000, 0110, 1101]	1011
1010	[1010, 0001, 1111, 0100]	1011
0101	**[0101, 1110, 0000, 1011]**	**0101**
1111	[1111, 0100, 1010, 0001]	1011
1100	[1100, 0111, 1001, 0010]	1110
1001	[1001, 0010, 1100, 0111]	1011
1110	[1110, 0101, 1011, 0000]	1110
1000	[1000, 0011, 1101, 0110]	0000
0100	[0100, 1111, 0001, 1010]	0000
0010	[0010, 1001, 0111, 1100]	0000

5. CONCLUSION

In this study we have seen results generated by our algorithm that makes computation easy and efficient. On the other hand the algorithm can also calculates all coding procedures and give correct results for more longer linear binary codes.

REFERENCES

[1] Bose R. C., Ray-Chaudhuri D. K., 1960, On a class of error-correcting binary group codes, *Info and Control* 3, 68-79.

[2] Hill R., 1986, A first course in coding theory, Clarendon Press, Oxford.

[3] Huffman W. C. and Pless V., 2003, Fundamentals of Error Correcting Codes, Cambridge University Press.

[4] Hamming R. W., 1950, Error detecting and error-correcting codes. *Bell Syst. T.J.* 29, 147-60.

[5] Monagan M. B., Geddes K. O., Heal K. M., Labahn G., Vorkoetter S. M., McCarron J. and DeMarco P., 2003, Maple 9 Advanced Programming Guide. *Maplesoft, a division of Waterloo Maple Inc.*

[6] Shannon C. E., 1948, A mathematical theory of communication. *Bell Syst. T.J.* 27, 379-423.

[7] Slepian D., 1960, Some further theory of group codes. *Bell Syst. Tech.J.* 39, 1219-52.