Math 2000 Project 3: Error-Detecting and Error-Correcting Codes[*]

Subhadip Chowdhury

- **Purpose:** To learn a method for detecting and correcting errors made in the transmission of encoded messages is constructed.
- **Prerequisite:** Abstract vector spaces and the concepts of null space, rank, and dimension.
- **Resources:** Read textbook section 4.(3-6) first.

When a message is transmitted, it has the potential to get scrambled by noise. This is certainly true of voice messages, and is also true of the *digital* messages that are sent to and from computers. Now even sound and video are being transmitted in this manner. A digital message is a sequence of **0**'s and **1**'s which encodes a given message. More data will be added to a given binary message that will help to detect if an error has been made in the transmission of the message; adding such data is called an *error-detecting code*. More data may also be added to the original message so that errors made in transmission may be detected, and also to figure out what the original message was from the possibly corrupt message that was received. This type of code is an *error-correcting code*.

A common type of error-detecting code is called a *parity check*. For example, consider the message **1101**. Add a **0** or **1** to the end of this message so that the resulting message has an even number of 1's. The message **1101** would thus be encoded as **11011**. If the original message were **1001**, it would be encoded as **10010**, since the original message already had an even number of **1**'s. Now consider receiving the message **10101**. Since the number of **1**'s in this message is odd, an error has been made in transmission. However, it is not known how many errors happened in transmission or which digit(s) were effected. Thus a parity check scheme detects errors, but does not locate them for correction.

Example 1. The United States Postal Service uses a code to express the zip code on a letter as a series of long and short bars. The digits are coded as in figure 1.

Zip codes are encoded and placed on the envelope. A long bar begins and ends each code. An additional parity check digit is encoded. This digit, when added to those in the five-digit zip code, produces a number which is a multiple of ten. If the six encoded digits do not add to a multiple of ten, then an error in transmission must have occurred. Thus the zip codes **29733** and **28209** become figure 2.

Since $\mathbf{2+9+7+3+3=24}$, and $\mathbf{24+6=30}$, a **6** was added to the code for **29733**; likewise a **9** was added to the code for **28209**, since $\mathbf{2+8+2+0+9+9=30}$.

[*] Adapted from Pearson Education website.

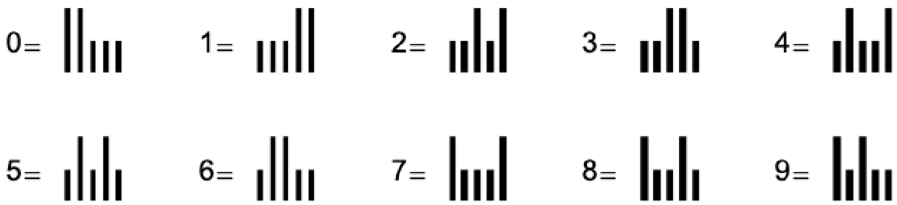


Figure 1

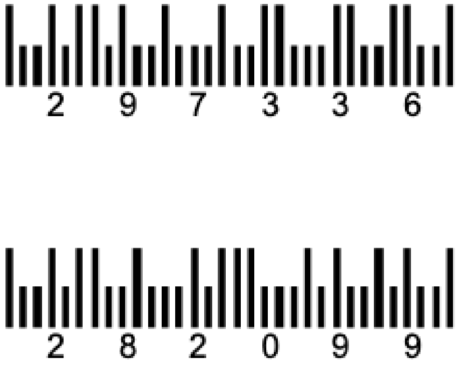


Figure 2

In order to discuss error-correcting codes, attention will be restricted to digital sequences: messages of $\mathbf{0}$'s and $\mathbf{1}$'s. Define the set $\mathbb{Z}_2$ to be the set $\{\mathbf{0}, \mathbf{1}\}$. It will first be useful to do arithmetic on $\mathbb{Z}_2$. Addition and multiplication for $\mathbf{0}$ and $\mathbf{1}$ are given in the following tables:

+	0	1
0	0	1
1	1	0

and

×	0	1
0	0	0
1	0	1

One may check that these operations have the familiar properties of addition and multiplication of real numbers. One peculiarity is the fact that since $\mathbf{1}+\mathbf{1}=\mathbf{0}$, $\mathbf{1}=-\mathbf{1}$. That is, $\mathbf{1}$ is its own additive inverse, and thus subtraction is exactly the same as addition in $\mathbb{Z}_2$.

Messages can now be expressed as column vectors of elements of $\mathbb{Z}_2$. The messages **1001** and **1101** would be expressed as

$$\begin{pmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{0} \\ \mathbf{1} \end{pmatrix}$$

Assume that each message is $\boldsymbol{n}$ digits long; the set of all possible messages of length $\boldsymbol{n}$ digits will be called $\mathbb{Z}_2^n$. In other words, $\mathbb{Z}_2^n$ is the set of all vectors with $\boldsymbol{n}$ elements taken from $\mathbb{Z}_2$. The set $\mathbb{Z}_2^4$ contains the following sixteen vectors:

$$\begin{pmatrix}0\\0\\0\\0\end{pmatrix},\begin{pmatrix}0\\0\\0\\1\end{pmatrix},\begin{pmatrix}0\\0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\\1\end{pmatrix},\begin{pmatrix}0\\1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\\1\end{pmatrix},\begin{pmatrix}0\\1\\1\\0\end{pmatrix},\begin{pmatrix}0\\1\\1\\1\end{pmatrix},$$

$$\begin{pmatrix}1\\0\\0\\0\end{pmatrix},\begin{pmatrix}1\\0\\0\\1\end{pmatrix},\begin{pmatrix}1\\0\\1\\0\end{pmatrix},\begin{pmatrix}1\\0\\1\\1\end{pmatrix},\begin{pmatrix}1\\1\\0\\0\end{pmatrix},\begin{pmatrix}1\\1\\0\\1\end{pmatrix},\begin{pmatrix}1\\1\\1\\0\end{pmatrix},\begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

These vectors can be added just as in $\mathbb{R}^n$; these vectors may also be multiplied by scalars taken from $\mathbb{Z}_2$.

Example 2.

$$\begin{pmatrix}0\\1\\0\\0\end{pmatrix}+\begin{pmatrix}1\\1\\0\\1\end{pmatrix}=\begin{pmatrix}1\\0\\0\\1\end{pmatrix} \qquad \text{and} \qquad 1\cdot\begin{pmatrix}1\\0\\0\\1\end{pmatrix}=\begin{pmatrix}1\\0\\0\\1\end{pmatrix}$$

In fact, if $\mathbb{Z}_2$ are the scalars, and the operations of vector addition and scalar multiplication as given in the last examples are used, then $\mathbb{Z}_2^n$ is a vector space: to make clear that $\mathbb{Z}_2$ are the scalars, $\mathbb{Z}_2^n$ is called a vector space over $\mathbb{Z}_2$. The material in Sections 4.2 to 4.6 on matrices of real numbers also applies to matrices whose entries are taken from $\mathbb{Z}_2$, except that all arithmetic is done in $\mathbb{Z}_2$.

Example 3. To find a basis for the column space, a basis for the null space, and the rank of $A=\begin{pmatrix}1&1&0&0\\1&0&1&1\\0&1&1&1\end{pmatrix}$, first row reduce A using $\mathbb{Z}_2$ arithmetic (remember that $1+1=0$):

$$\begin{pmatrix}1&1&0&0\\1&0&1&1\\0&1&1&1\end{pmatrix}\sim\begin{pmatrix}1&1&0&0\\0&1&1&1\\0&1&1&1\end{pmatrix}\sim\begin{pmatrix}1&1&0&0\\0&1&1&1\\0&0&0&0\end{pmatrix}\sim\begin{pmatrix}1&0&1&1\\0&1&1&1\\0&0&0&0\end{pmatrix}$$

A basis for $\mathbf{Col}\,A$ is the pivot columns in A:

$$\left\{\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}1\\0\\1\end{pmatrix}\right\}$$

Thus $\mathbf{rank}\,A=2$. To find a basis for $\mathbf{Nul}\,A$, solve $A\vec{x}=0$ and get the equations

$$x_1=-1x_3-1x_4$$

and

$$x_2=-1x_3-1x_4$$

Since $1=-1$,

$$x_1=x_3+x_4$$

and

$$x_2 = x_3 + x_4$$

so a basis for **Nul** A would be

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Notice that these results differ from those which would be calculated if A were treated as a matrix of real numbers; you may confirm that in that case $\mathbf{rank}\, A = 3$. Here are all of the members of **Nul** A:

$$\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

Note that the number of vectors in **Nul** A is $4 = 2^2$, which is 2 raised to the dimension of **Nul** A. This is true for any subspace of $\mathbb{Z}_2^n$.

Proposition 1. *If W is a subspace of $\mathbb{Z}_2^n$ with $\mathbf{dim}\, W = k$, then the number of vectors in W is equal to 2^k.*

Assume that the messages are each 4 digits long. A self-correcting code for these messages will now be created. A more sophisticated version of the parity check is done; three numbers will be added to the end of each 4 digit message. Thus the encoded messages will be elements of $\mathbb{Z}_2^7$. To begin, consider the matrix

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

Notice that the columns in H, which will be called $\vec{\mathbf{h}}_1, \ldots, \vec{\mathbf{h}}_7$, happen to be all of the non-zero members of $\mathbb{Z}_2^3$. A basis for the null space of H may be found as above:

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

For reasons which will be made clear later, it will be better to have a different basis for **Nul** H. This new basis will be created by making a matrix whose rows are the vectors in the old basis (so that the row space of the new matrix is **Nul** H), row reducing this matrix (which doesn't change the row space), then using the non-zero rows of the resulting matrix as a basis for **Nul** H (which is still the row space). This process is allowable by Theorem 13 in Section 4.6.

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Thus the following set of vectors is also a basis for **Nul** H:

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

Since the dimension of **Nul** H is 4, by proposition 1, **Nul** H contains 16 vectors. Of course, $\mathbb{Z}_2^4$ also contains 16 vectors, so *each vector in* $\mathbb{Z}_2^4$ *can be encoded using a different vector in* **Nul** H. For that reason **Nul** H is called the **Hamming (7, 4) code**. To encode the vectors in $\mathbb{Z}_2^4$, form a matrix A whose columns are the basis elements for **Nul** H; the matrix A will be the encoding matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

Example 4. To encode the message **1101**, compute

$$A \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Notice that since the first four rows of A are the identity matrix, multiplication by A merely adds three digits to the original message. *This is the reason that the alternative basis for* **Nul** H *is desirable - the encoding process merely appends data onto each encoded vector.*

The matrix H itself was chosen because its null space has some very interesting properties which allow for the detection and correction of *single errors* in transmitted messages. Assume at this point that any transmitted message has at most one error in transmission. If the probability of an error in transmission is small, then this is a reasonable assumption.

Consider the standard basis vectors called $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \ldots, \vec{\mathbf{e}}_7 \in \mathbb{Z}_2^7$.

$$\vec{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{\mathbf{e}}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\vec{\mathbf{e}}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{\mathbf{e}}_5 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{\mathbf{e}}_6 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{\mathbf{e}}_7 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Notice that adding one of these vectors to an encoded message vector $\vec{\mathbf{x}}$ is equivalent to making a single error in the transmission of $\vec{\mathbf{x}}$. Notice also that the vectors $\vec{\mathbf{e}}_1, \ldots, \vec{\mathbf{e}}_7$ are not in the null space of H since $H\vec{\mathbf{e}}_i = \vec{\mathbf{h}}_i \neq \mathbf{0}$. In fact, there is the following theorem.

Theorem 2. *If H is the matrix given above, and if $\vec{\mathbf{x}}$ is in* **Nul** H*, then $x + \vec{\mathbf{e}}_i$ is not in* **Nul** H.

Proof. Since $\vec{\mathbf{x}}$ is in **Nul** H, $H\vec{\mathbf{x}} = \vec{\mathbf{0}}$. By the above note, $H\vec{\mathbf{e}}_i = \vec{\mathbf{h}}_i \neq \mathbf{0}$. Thus

$$H(\vec{\mathbf{x}} + \vec{\mathbf{e}}_i) = H\vec{\mathbf{x}} + H\vec{\mathbf{e}}_i = \vec{\mathbf{0}} + \vec{\mathbf{h}}_i = \vec{\mathbf{h}}_i \neq \vec{\mathbf{0}}$$

and $\vec{\mathbf{x}} + \vec{\mathbf{e}}_i$ is not in **Nul** H. □

This result means that if a *single* error is made in the transmission of a message $\vec{\mathbf{x}}$, then that error may be detected by checking to see whether the received message lies in **Nul** H.

Example 5. If the message **0100101** is received, check that

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

Since the message vector is in **Nul** H, no single transmission error has happened. If a single error had happened, the theorem says that the resulting message vector would not be in **Nul** H.

Example 6. If the message **1001000** is received, check that

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

Thus (assuming that at most one error in transmission has been made) a single transmission error has happened.

So the **Hamming (7, 4) code** is an error-detecting code. The following theorem will show that it is also an error-correcting code.

Theorem 3. *If H is the matrix given above, and if $H\vec{\mathbf{x}} = \vec{\mathbf{h}}_i$, then $\vec{\mathbf{x}} + \vec{\mathbf{e}}_i$ is in $\mathbf{Nul}\,H$, and $\vec{\mathbf{x}} + \vec{\mathbf{e}}_j$ is not in $\mathbf{Nul}\,H$ for $j \neq i$.*

Proof. Suppose that $H\vec{\mathbf{x}} = \vec{\mathbf{h}}_i$. Then

$$H(\vec{\mathbf{x}} + \vec{\mathbf{e}}_i) = \vec{\mathbf{h}}_i + \vec{\mathbf{h}}_i = 0$$

Likewise if $i \neq j$,

$$H(\vec{\mathbf{x}} + \vec{\mathbf{e}}_j) = \vec{\mathbf{h}}_i + \vec{\mathbf{h}}_j \neq 0$$

□

Suppose a message $\vec{\mathbf{x}}$ is received that has had a single error happen in transmission. By Theorem 2, $H\vec{\mathbf{x}} \neq \mathbf{0}$, so $H\vec{\mathbf{x}} = \vec{\mathbf{h}}_i$ for some i. The result in Theorem 3 implies that the single error in transmission must have occurred to the i^{th} digit; changing this digit (by adding $\vec{\mathbf{e}}_i$ to $\vec{\mathbf{x}}$) will produce a vector in $\mathbf{Nul}\,H$, and thus a properly encoded vector. Changing any other digit in $\vec{\mathbf{x}}$ will not produce a vector in $\mathbf{Nul}\,H$.

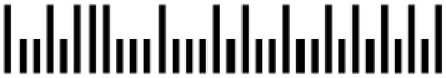
Example 7. The message **1001000** was in error by a previous example. In fact,

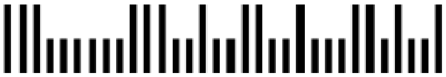
$$H\begin{pmatrix}1\\0\\0\\1\\0\\0\\0\end{pmatrix}=\begin{pmatrix}0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\1&0&1&0&1&0&1\end{pmatrix}\begin{pmatrix}1\\0\\0\\1\\0\\0\\0\end{pmatrix}=\begin{pmatrix}1\\0\\1\end{pmatrix}=\mathbf{h}_5$$

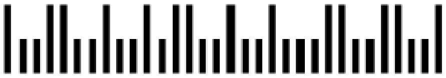
By Theorem 3, the single error in transmission must have occurred at the fifth digit. Thus the true message which was sent is **1001100**.

Exercise 1

The following United States Postal Service codes were found on envelopes; determine whether an error was made in transmission.

(a) 

(b) 

(c) 

Exercise 2

Consider the following vectors in $\mathbb{Z}_2^4$.

$$\vec{\mathbf{a}} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{\mathbf{b}} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{\mathbf{c}} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

(a) Compute $\vec{\mathbf{a}} + \vec{\mathbf{b}}$ and $\vec{\mathbf{c}} - \vec{\mathbf{b}} + \vec{\mathbf{a}}$.

(b) Is the set $\{\vec{\mathbf{a}}, \vec{\mathbf{b}}, \vec{\mathbf{c}}\}$ linearly independent or linearly dependent?

Exercise 3

Find a basis for the column space, a basis for the null space, and the rank of

$$\boldsymbol{B} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

where the entries in $\boldsymbol{B}$ are considered to be in $\mathbb{Z}_2$.

Exercise 4

Encode the following messages using the Hamming (7,4) code.

(a) **1001**

(b) **0011**

(c) **0101**

Exercise 5

Each of the following messages has been received, and each had been encoded using the Hamming (7,4) code. During transmission at most one element in the vector was changed. Either determine that no error was made in transmission, or find the error made in transmission and correct it.

(a) **0101101**

(b) **1000011**

(c) **0010111**

(d) **0101010**

(e) **0111100**

(f) **1001101**

(g) **1010010**

(h) **1110111**