A Note on the Applications of Coding Theory

Mulatu Lemma, Blessing Enya and Keith Lord
Department of Mathematics
Savannah State University
Savannah, GA 31404
USA

Abstract: Communication and disc records are common in the society but sometimes there are errors in information transmission from one source to another. However, transmission has been made easier thanks to the discovery of the study of error-control code known as coding theory. Several methods like binomial coefficient and linear codes like Hamming codes play a role in ensuring the transmission of information through a noiseless channel. Even though binomial coefficient is a mathematical formula, it helps error-control smaller codes. Linear codes are easier and quick to use, but require a prime element. An example of linear codes is the Hammond codes and they are perfect because they attain the highest rates for codes within a minimum length distance of 3. Other examples of codes include repetition codes, cyclic codes, parity check and sum-O codes, and generalized Reed-Solomon code.

Introduction: Mobile phones are an excellent source of communication and they come in handy when trying to browse for something on the internet. CDs and DVDs are used to record songs, books, and movies, which gives a chance to listen to tracks or watch movies on them with clarity. However, what happens "behind the scene" of this communication process? How does a person in one country get information about another country through the internet? What is the code behind recording things on CDs? Well, information being transmitted across a noiseless channel appear as strings of 0 (zero) and 1 (one). Sometimes, the 0 (zero) might be sent as 0 (zero), but could be received as 1 (one). Thanks to years of research, the discovery of a process known as error control coding was created. The error-control code helps spot and modify the errors found during transmission. The study of error-control codes is known as coding theory, and it is used for protecting informational transmission through a noise channel. It is also known as algebraic coding

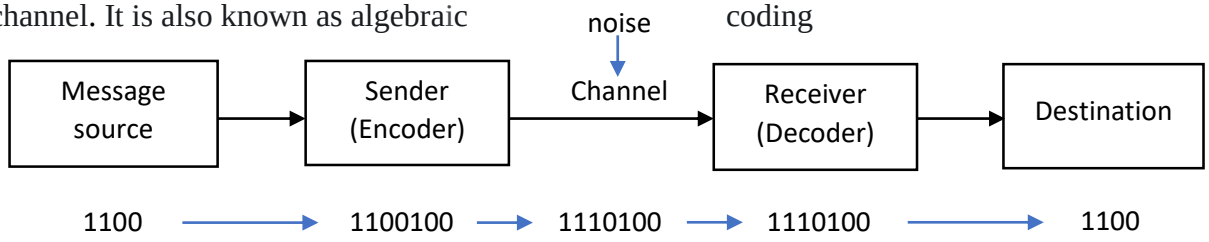

Fig. 1: Coding Theory channel

theory, and relates to courses like discrete mathematics and number theory. Coding theory is used in TV s, internet, fax machines, and even phones to correct code errors, because transmission can sometimes be noisy. Without this theory, there would be no satellite communication on TV, no mobile messaging or calling, and no space travel. Fig. 1 shows the process of coding theory and how information are being transmitted across

the world. Message source is usually the person or object sending the information. The sender (encoder) is the means through which message is being accessed for errors and compatibility. Both must go through a channel, and this is where the coding theory comes in. A channel is the technology or substance that the information is being transmitted through, and a good example would be telephone lines. After going through a channel containing error-control codes, the information is decoded, the receiver acquires this message and it gets to its destination. This new method of coding was discovered in 1948 by Claude Shannon while he was working at Bell laboratories in the United States. It has since then been explored, explained, and expanded by Richard Hamming and John Leech. This paper will focus on some of the different categories of coding theory such as binomial coefficient, linear codes and generating matrices, and Hamming's codes.

Binomial coefficients: These are essential in error-control for smaller error-control coding and they are written mathematically as

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

where.

$$n! = n * (n-1) * (n-2) * (n-3) * \ldots * 3 * 2 * 1$$

In the equation, *k* represents unordered selections of distinct objects and *n* represent unique objects. In coding, binomial coefficient represents the amount of *k* from *n.* For example,

$$\frac{6!}{4!\,(6-4)!}$$

$$= \frac{6!}{4!\,(2)!}$$

$$= \frac{6 * 5 * 4 * 3 * 2 * 1}{(4 * 3 * 2 * 1)(2 * 1)}$$

$$= \frac{6 * 5 * \cancel{4 * 3 * 2 * 1}}{(\cancel{4 * 3 * 2 * 1})(2 * 1)}$$

$$= \frac{6 * 5}{2 * 1}$$

$$= \frac{30}{2}$$

$$\binom{6}{4} = 15$$

Therefore, if there are a set of 6 distinct objects, there are 15 unordered selections of 4

objects, which means that if the objects are *a*, *b*, C, *d*, *e*, *f*, then the unordered selections are:

$$(\mathbf{a},\mathbf{b}),(\mathbf{a},\mathbf{c}),(\mathbf{a},\mathbf{d}),(\mathbf{a},\mathbf{e}),(\mathbf{a},\mathbf{f}),(\mathbf{b}.\mathbf{c}),(\mathbf{b},\mathbf{d}),(\mathbf{b},\mathbf{e}),(\mathbf{b},\mathbf{f}),(\mathbf{c},\mathbf{d}),(\mathbf{c},\mathbf{e}),(\mathbf{c},\mathbf{f}),(\mathbf{d},\mathbf{e}),(\mathbf{d},\mathbf{f}),(\mathbf{e},\mathbf{f})$$

Linear Codes: Group codes $\boldsymbol{C}$ with length $\boldsymbol{n}$ are linear subspaces of the vector space $\boldsymbol{F_q^n}$, where $\boldsymbol{F_q}$ is the finite field with $\boldsymbol{q}$ prime elements, and the dimension of the subspace is $\boldsymbol{k}$. These codes are also known as q-ary $[\boldsymbol{n}, \boldsymbol{q^k}]$-*code*. If $\boldsymbol{q = 2}$, the code is a binary code, and if $\boldsymbol{q = 3}$, the code is a ternary code. These field with additions and multiplications is known as the Galois field $\boldsymbol{GF(q)}$. There are vectors in $\boldsymbol{C}$, known as codewords, and as easy to use and quick the linear codes are, there are certain criteria that must be met, such all codewords should be listed to specify a non-linear code, the size of a code is the number of codewords present in $\boldsymbol{C}$ to equals $\boldsymbol{q^k}$, and the sum of two codewords is another codeword. It is, however, possible to layout these vectors in matrix forms known as generator matrix containing dimension $\boldsymbol{n^*k}$.

Definition: A subset $\boldsymbol{C \subseteq V(n,q)}$ is a linear code if $\boldsymbol{u + v \in C}$ for all $u, v \in \boldsymbol{C}$, and $\boldsymbol{au \in C}$ for all $\boldsymbol{u \in C}$, $\boldsymbol{a \in GF(q)}$.

Generator matrix: Let $\boldsymbol{C \subseteq F_q^n}$ be a linear code of dimension $\boldsymbol{k}$. A matrix $\boldsymbol{G \in F_q^{n*k}}$ is said to be a generator matrix for $\boldsymbol{C}$ if its $\boldsymbol{k}$ columns span $\boldsymbol{C}$. The generator matrix $\boldsymbol{G}$ provides a way to encode a message $\boldsymbol{x \in F_q^k}$ as the codeword $\boldsymbol{Gx \in C \subseteq F_q^n}$.

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

this [4,8]-code has

$$\mathbf{2^k = 8}.$$

where

$$\mathbf{8 = 2^3}$$

r

then

$$\mathbf{2^k = 2^3}$$

$$\mathbf{\cancel{2}^k = \cancel{2}^3}$$

$$\mathbf{k = 3}$$

and the generator matrix is

$$\begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

The [4,8]- code can be recreated through a construction of the linear combination of the rows of the previously found generator matrix:

$$[\mathbf{xyz}] * \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{x}[\mathbf{0100}] + \mathbf{y}[\mathbf{0011}] + \mathbf{z}[\mathbf{1100}]$$

When $x = 1, y = 0 \ \ and \ z = 1$, the seventh codeword can be generated:

$$[\mathbf{0100}] + [\mathbf{1100}] = [\mathbf{0+1}, \mathbf{1+1}, \mathbf{0+0}, \mathbf{0+0}] = [\mathbf{1000}]$$

When $x = 1, y = 1 \ \ and \ z = 0$, the fifth codeword can be generated:

$$[\mathbf{0100}] + [\mathbf{0011}] = [\mathbf{0+0}, \mathbf{1+0}, \mathbf{0+1}, \mathbf{0+1}] = [\mathbf{0111}]$$

When $x = 0, y = 1 \ \ and \ z = 1,$ the last codeword can be generated:

$$[\mathbf{0011}] + [\mathbf{1100}] = [\mathbf{0+1}, \mathbf{0+1}, \mathbf{1+0}, \mathbf{1+0}] = [\mathbf{1111}]$$

Theorem: For any subset $\boldsymbol{S}$ of a linear space, $\boldsymbol{S}^{\wedge}$ SA is a linear space containing the zero word, all words in ***S***, and the sum of two or more words in ***S***. Therefore, repeating the process of addition by alternating ***0***s and ***1***s for ***x, y***, and ***z***, there will be a creation of the entire matrix.

Hamming codes: These are a class of linear error-correcting codes, and because they are perfect, they attain the highest rates for codes within a minimum length distance of ***3***. A binary Hamming code of length $n = 2^f - 1 (f \geq 2)$ has parity check matrix $\boldsymbol{H}$ whose columns consist of all nonzero vectors of length **f**. This gives an $[\boldsymbol{n}, \boldsymbol{k}, \boldsymbol{d}]$ linear code where $n = 2^f - 1$ is the block length, $k = 2^f - 1 - f$ is the message length, and $d = 3$ is the distance.

For example, since $f \geq 2$, then

Where $f = 3$

$$n = 2^3 - 1$$

$$n = 8 - 1$$

$$n = 7$$

}

and

$$k = 2^3 - 1 - 3$$

$$k = 8 - 1 - 3$$

$$k = 7 - 3$$

$$k = 4$$

Therefore, the Hamming code parity check matrix [7,4,3] is perfect and it would look like

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

and

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

where $\boldsymbol{H}$ is the Hamming code and $\boldsymbol{G}$ is the generator matrix.

Definition (q-ary Hamming Code): Given an integer

$$\boldsymbol{r \geq 2}$$

let

$$\boldsymbol{n = \frac{q^r - 1}{q - 1}}$$

The q-ary Hamming code Ham ***(r, q)*** is a linear $[\boldsymbol{n, n - r}]$ code in $\boldsymbol{F_q^n}$, whose parity-check

matrix $\boldsymbol{H}$ has the property that the columns of $\boldsymbol{H}$ are made up of precisely one nonzero vector
from each vector subspace of dimension $\boldsymbol{1}$ of $\boldsymbol{F_q^n}$. Other examples of codes include repetition codes, cyclic codes, parity check and ***sum-0*** codes, and generalized Reed-Solomon code. Repetition codes exist for any length $\boldsymbol{n}$ and any alphabet $\boldsymbol{A}$. The most fundamental case is that of binary repetition codes, those with alphabet $\boldsymbol{A} = \{0,1\}$. The parity check code of length $\boldsymbol{n}$ is composed of all binary (alphabet $\boldsymbol{A} = \{\mathbf{0}, \mathbf{1}\}$) $\boldsymbol{n}$-tuples that contain an even number of ***1'***s.

References

Barg, A. (1993). At the Dawn of the Theory of Codes. *The Mathematical Intelligencer, 20-26.*

Chen, H. (2011). Hamming Codes. *Coding Theory, 47-52.*

Fiedler, 1. (2004). Hamming Codes. *Iowa State University Math Newsletter, 1-8.*

Grossman, 1. (2008). Coding Theory: Introductions to Linear Codes and Applications. *Insight: RIVIER ACADEMIC JOURNAL, 1-17.*

Hall, 1. (2010). Linear Codes. *Notes on Coding Theory, 4-56.*

Tonchev, V. (2009). Linear Codes. *An Introduction to Coding Theory: Lecture Notes, 2-26.*

Weisstein, E. (2017, February 11). *Coding Theory.* Retrieved from MathWorld: *http://mathworld.* wolfram.com/CodingTheory. html