Department of
Mathematical Sciences
College of Art and Sciences

Algebraic Coding Theory and Applications

Jillian Gaietto, Senior, Pure Mathematics, Kent State University
Advisor: Hai Q. Dinh, Ph.D., Professor of Mathematics, Kent State University

CELEBRATE CENTENNIAL KENT STATE UNIVERSITY 1910·2010

Abstract

When communicating across a channel, it is inevitable that such pathways of communication be "noisy", thus there is always some sort of interference across the channel. This results in messages not always being received as they were sent. In order to solve these problems, coding theory developed and is used both to detect and correct errors. It is used for data compression, error correction, cryptography and network coding. In error correction, a concentration on algebraic coding theory lies with linear codes, including cyclic and constacyclic codes. In this poster presentation, we will discuss the history of coding theory, going in depth with cyclic and constacyclic codes, as well as discussing applications and current problems being resolved using algebraic coding theory.

History

- Claude Shannon, American mathematician, electrical engineer, and cryptographer wrote a paper titled "A Mathematical Theory of Communication" in 1948.
- Not related to what you say but what you could say.
- Focused on the best way to encode information that a sender wants to transmit.
- Introduced the term "bits" to reference a binary digit.

$$H = -\sum p(x)\log p(x)$$

- H=Shannon Entropy, measure of information in a message in bits; p(x)=Probability of a certain symbol, x, turning up; log(p(x))=Number of bits needed to represent x
- Used probability theory to prove
- ~10 years later Cyclic codes discovered
- ~10 years later Negacyclic codes discovered
- Constacyclic codes discovered

Definitions

- *Coding Theory* is the study of methods for efficient and accurate transfer of information from one place to another; finding noise and correcting errors
- *Code* is a set of codewords. A *block code* is a set of codewords of the same length
- *Codewords* are the words belonging to a given code; made up of digits
- A *channel* is the physical medium through which the information is transmitted. A *binary channel* only sends digits of 0 or 1.
- *Noise* is the undesirable disturbances which may cause information received to differ from that which was sent
- *Length* is the number of digits in a codeword
- A binary channel is *symmetric* if 0 and 1 are transmitted in equal accuracy.
- The *information rate* is a number designed to measure the proportion of each codeword that is carrying the message $-\frac{1}{n}\log_2|C|$ code c of length n
- The *Hamming weight* is the number of times the digit 1 occurs in a codeword v, denoted wt(v).
- The *Hamming distance* is the number of positions in which w and v disagree, denoted d(v,w)
- *Parity Digit* is an added digit that follows a certain algorithm to reduce errors
- A code C of distance d is an *error correcting code* if it detects all error patterns of weight less than or equal to (d-1) and there is at least one error pattern of weight d which C will not detect.
- A *cyclic code* is a block code where the circular shifts of each codeword gives another word that belongs to the code, error-correcting

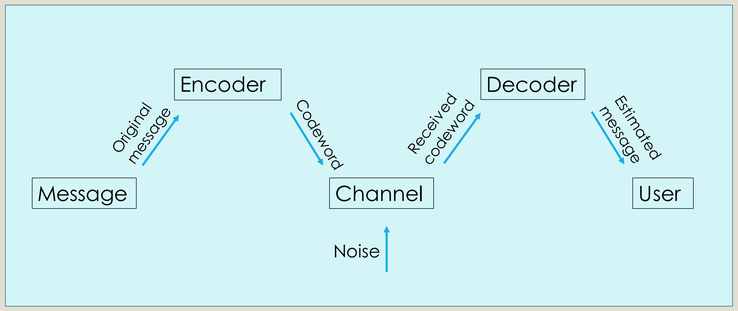
$$\tau(x_0, x_1, \dots, x_{n-1}) = (x_{n-1}, x_0, x_1, \dots, x_{n-2})$$

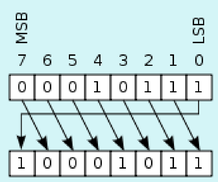
- A *constacyclic code* is a cyclic code where the circular shifts of each codeword gives another word that belongs to the code with the first symbol being a λ–tuple

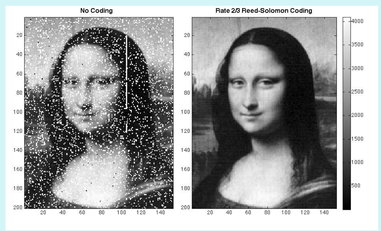
$$\tau_\lambda(x_0, x_1, \dots, x_{n-1}) = (\lambda x_{n-1}, x_0, x_1, \dots, x_{n-2})$$

- A *negacyclic code* is a cyclic code where the circular shifts of each codeword gives another word that belongs to the code with the first symbol changing sign

$$v(x_0, x_1, \dots, x_{n-1}) = (-x_{n-1}, x_0, x_1, \dots, x_{n-2})$$







Compact Disc

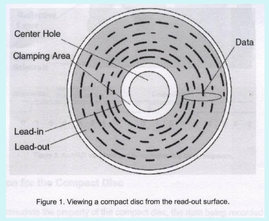


Figure 1. Viewing a compact disc from the read-out surface.

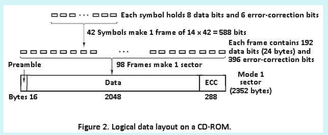
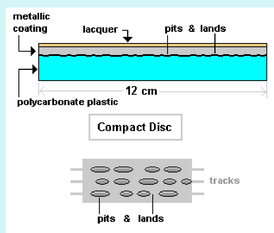


Figure 2. Logical data layout on a CD-ROM.



Deep Space

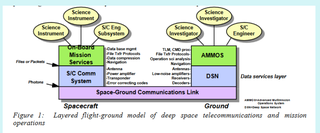
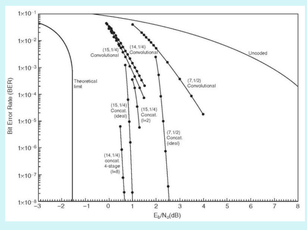


Figure 1: Layered flight-ground model of deep space telecommunications and mission operations

Tab. 1 The distance and delay between the Earth and other planets in solar system

Planets	Minimum distance ($\times 10^6$ km)	Minimum delay (min)	Maximum distance ($\times 10^6$ km)	Maximum delay (min)
Venus	39.6	2.20	261.0	14.50
Mars	59.6	3.311	401.3	22.294
Mercur	101.1	5.617	221.9	12.378
Jupiter	593.7	32.983	968.0	53.778
Saturn	1199.7	86.661	1659.1	92.172
Uranus	2591.9	143.994	3155.1	175.283
Neptune	4304.9	239.161	4694.1	260.783
Pluto	4297.9	238.772	7535.1	418.617



Compact Disc (CD)

- Reed-Solomon Codes with binary digits represented on the disc as pits and lands, first instituted in 1982
- Code is so strong most playback error comes from tracking errors causing the laser to jump tracks
- Reed-Solomon codes discovered in 1960 by Irving Reed and Gustave Solomon.
- Can detect and correct multiple errors, including burst errors.
- Can correct a burst error of up to 4000 bad bits, or a physical defect of 2.47 mm long through parity digits and interleaving.
- Interpolation can conceal errors up to 13,700 bits of 8.5 mm long
- Two layers of Reed-Solomon code separated by a 28-way convolutional interleaver, Cross-Interleaved Reed-Solomon Code (CIRC)
  - High random error correctability
  - Long burst error correctability
  - In case exceeded, interpolation provide concealment approximation
  - Very high efficiency
  - Simple decoder strategy with reasonable sized memory
- Codewords consist of all function tables of polynomials of degree less than k over the finite field with n elements (n is prime)
- Interpret k given symbols as the first segment of the function table. Remaining n-k symbols be generated by evaluating polynomial at points
- Since n transmitted symbols from an overdetermined system that specifies polynomial of degree less than k, Interpolation can recover original message
- Adds a parity digits to every three
- 1st Circle: relatively weak Reed-Solomon (32,28), can correct up to 2 bit errors in 32 bit block and flags erasures with more than 2 bit
- 2nd Circle: Reed-Solomon (28,24) can correct up to 4 erasures per block
- CIRC interleaves audio frames through disc over several consecutive frame
- A physical frame contains information from many audio frames. This adds 64 bits of error correction data to each frame. 8 bits of subcode added to each frame

Deep Space Communication

- *Deep space communication* is communication between earth stations and remote spacecraft, other planets, or space beyond Earth's gravitational field.
  - Most missions never return to earth, failed reception and consequent retransmission not an option
  - Communication sporadic and ultra long distances
  - Long delay, weak received signal, and variable distances according with orbits
  - Asymmetrical uplink and downlink capacities
  - Limited mass, power source, and volume
  - Intensity of electromagnetic radiation decreases according to $\frac{1}{r^2}$ as you leave Earth
- Channel coding major solution to deep space issues
- Traditionally used concatenation of convolutional code and Reed-Solomon codes.
  - Convolutional code have greater simplicity of implementation over a block code of equal power
  - Infinite but fundamentally don't offer more protection against noise than the equivalent block code
  - Encoder usually a simple circuit with memory and logic while decoder in software or firmware

$$c(x) = (c_1(x), c_2(x), \dots, c_n(x)) = (m(x)g_1(x), m(x)g_2(x), \dots, m(x)g_n(x))$$

- By adding specific types of redundancy, can recover data perfectly with high probability, even under huge amounts of noise
- *Low-Density Parity-Check (LDPC)* are on a matrix containing only a few ones in each row and column
- LDPC decoded on parity check matrix which grows larger as the code rate is decreased, low rate LDPC more complex
- *Turbo codes* are constructed by applying 2 or more simple to decode encoding rules to different permutations of the same information sequence, achieve data rates more near Shannon limit (theoretical max)
- Turbo codes decoded on trellises with one trellis per information bit corresponding to several code symbols
- LDPC now international standard while Turbo codes are used for extremely long transmissions(outer planets or outside solar system)

Resources: http://www.usna.edu/Users/math/wdj/_files/documents/reed-sol.htm ; http://www.cs.cmu.edu/~guyb/realworld/reedsolomon/reed_solomon_ ; https://www.theguardian.com/science/2014/jun/22/shannon-information ; https://www.technologyreview.com/s/420369/code-quest/ ; Cyclic Codes & Reed Solomon Codes by John MacLaaren Walsh, ECES 421, Spring, 2013 ; Introduction to Coding Theory, Massoud Malek, California State University ; http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.2977&rep=rep1&type=pdf ; http://www.nasa.gov/mission_pages/LRO/news/mona-lisa.html ; http://scienceandtechnology.jpl.nasa.gov/research/research-topics-list/communications-computing-software/deep-space-communications ; Concatenated Turbo/LDPC Codes for Deep Space Communications: Performance and Impementation, Carol Condo,2013 ; Performance Study of a Deep Space Communications system with LDPC under Solar Scintillation, Qi Li,2012 ; A Survey of Deep Space Comunication, Zhang Gengxin,2011; Coding Theory and Cryptography, The Essentials