Developing New Error Correcting Codes

Anubhab Das

September 22 2022

Faculty Consultant: Prof. Harm Derksen

1 Abstract

In this project, I will be studying in the fields of Information and Coding Theory to learn about different kinds of error correction to try and develop some error-correcting codes. Specifically, Reed Solomon Error correction is something that is of particular interest to me. I will be studying various coding theory concepts like the Noisy Coding Theorem, Minimum Distance Decoding and aspects of Algebra that will be relevant like finite fields, so that I can better understand error-correction more rigorously. On my own I plan trying to possibly develop new error-correcting codes by coalescing different methods I learn about (Reed-Solomon, Quadratic-residue codes, BCH codes etc).

2 Motivation

I am a double major in both Computer Science and Mathematics, so I feel that this topic is natural intersection of the two majors. I have always been interested in Cryptography and the mechanics behind concepts like Hashing Algorithms and Error Correction is a very natural adjacent topic. According to my consultant the prerequisite knowledge in Algebra and Probability theory is minor enough to the point where it is very accessible to me, which means it will not be out of my depth. While I am familiar with basic error correction and hamming codes, I plan on delving deeper into Shannon's famous paper and even the development behind Reed Solomon's code, so there will also be plenty to learn. I anticipate that most of my learning will go into learning the Algebra and the proper aspects of Coding Theory as I feel I have an adequate background in probability theory.

3 Approach

In order to lay the foundation of the project I am planning on learning general aspects about Codes (ie Error Detection and Correction, Minimum Distance

Decoding, different families of Codes, etc). I will also try to learn the required Algebra that I am not accustomed to that will be relevant like what exactly a finite field is and how it is prevalent when it comes to discussing codes. From then on I will delve into the mechanics of Reed-Solomon encoding and decoding among other coding schemes, so that I can develop my own. A strong sign of success I think would be able to explain my coding scheme to the class, so that everyone at least has a somewhat basic understanding. Moreover, I think it would be possible to write a computer program that serves as a demonstration on a much larger scale (ie much longer messages that could not physically be written on a board).

4 Terminology

- *Field* - A Field F is a set under an additive operation and a multiplicative operation. The set must be associative, commutative, distributive, have an identity element and have inverses for each element. This must be true for both operations except for that the additive identity does not have a multiplicative inverse. These properties for the field axioms.
- *Finite Field* - A set that satisfies the field axioms with a finite number of elements.
- *Entropy* - A calculation that measures the average level of uncertainty of a Random Variable's outcome. Explicitly the way to calculate the entropy a discrete random variable X is given by the equation:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

- *Hamming Distance* - A way to measure the distance between different code words (can allow us to construct a geometric object that represents a coding scheme). Explicitly we can express Hamming Distance between two codes $w = (w_1, ..., w_n)$ and $v = (v_1, ..., v_n)$ as

$$d(w, v) = |\{i \ | w_i \neq v_i\}|.$$

- *Block Codes* - A large family of error correcting codes that encodes data into blocks.
- *Linear Codes* - Error-correcting codes where any linear combination of codes is also a code. More explicitly if $w = (w_1, ..., w_n)$ and $v = (v_1, ..., v_n)$ are codes then $aw + bv$ is a code for constants a and b.
- *Cyclic Codes* - A Block code where circular shifts of codes also create codes ie if $w = (w_1, ..., w_n)$ is a code, so is $w' = (w_n, w_1, ..., w_{n-1})$.

5 Bibliography

- Roman, Steven. *Coding and Information Theory.* Springer, 2011. This is the main book I plan on using for learning about Coding Theory.
- Shannon, Claude Elwood, and Warren Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, 1999. This is Shannon's original paper that I think will be a useful resource, when it comes to building a foundation of knowledge.
- Dummit, David Steven, and Richard M. Foote. *Abstract Algebra.* Wiley, 2004. This will serve as my general Algebra research throughout the project. It is a book that I am pretty familiar with.
- Hill, Raymond. *A First Course in Coding Theory.* Clarendon Press, 2003. This is a fairly introductory text for learning about Coding Theory that is not too technical and may work better for a general overview in some instrances.
- Ling, San, and Chaoping Xing. *Coding Theory: A First Course.* Univ. Press, 2010. This is another introductory text that is less rigorous and may be useful if I am more interested in learning ideas.