Error-Correcting Codes

G. Eric Moorhouse, UW Math

Corrected copies of transparencies for this seminar series should soon be available at

`http://math.uwyo.edu/~moorhous/quantum/`

References

F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes,* North-Holland, 1977.

J.H. van Lint, *Introduction to Coding Theory,* 2nd ed., Springer-Verlag, 1992.

Goal of Coding Theory

Digital information in the real world is subject to errors, i.e. alteration due to unreliability of storage media and interference in communications channels.

The goal of coding theory is to represent digital information in a form which allows for the recovery of the original data from its corrupted form, if the number of errors is not too large.

This requires that some redundancy be incorporated into the stored information.

Key People

Richard Hamming (1915–1998), pioneer in computer design and error-correcting codes.

Claude Shannon (1916–?), founder of Information Theory, researcher at Bell Telephones 1941-1972.

Both Hamming and Shannon were involved in the Manhattan Project.

Alphabet and Words

Information is stored and transmitted as a stream of *letters* from a chosen *alphabet* F.

Most popular is the *binary alphabet* $F = \{0, 1\}$.

More generally, $F = \{0, 1, 2, \ldots, p-1\}$ with addition and multiplication mod p (where p is a prime) is popular because F is a field. In this case

$$F^n = \{(a_1, a_2, \ldots, a_n) : a_i \in F\}$$

is an n-dimensional vector space over F.

A *word* of length n is a string of n characters from the alphabet F. If $|F| = q$ then there are q^n words of length n. These are identified with the vectors of F^n.

A *code of length* n is a subset $\mathcal{C} \subseteq F^n$. Elements of $\mathcal{C}$ are *codewords.* If $F = \{0, 1\}$ then $\mathcal{C}$ is a *binary code.*

Example 1: Parity Check Codes

The following binary code $\mathcal{C}_1 = \{00000, 00011, \ldots, 11110\}$ of length 5 is formed by appending a *parity check bit* to the end of each message word.

Message word	Codeword
0000	00000
0001	00011
0010	00101
0011	00110
0100	01001
0101	01010
0110	01100
0111	01111
1000	10001
1001	10010
1010	10100
1011	10111
1100	11000
1101	11011
1110	11101
1111	11110

Using the code $\mathcal{C}_1$, we can detect up to one bit error during transmission, but we cannot correct any errors.

Example 2: 3-Repetition Codes

The following binary code $\mathcal{C}_2$ of length 12 is formed by repeating each message word three times.

Message word	Codeword
0000	0000 0000 0000
0001	0001 0001 0001
0010	0010 0010 0010
0011	0011 0011 0011
0100	0100 0100 0100
0101	0101 0101 0101
0110	0110 0110 0110
0111	0111 0111 0111
1000	1000 1000 1000
1001	1001 1001 1001
1010	1010 1010 1010
1011	1011 1011 1011
1100	1100 1100 1100
1101	1101 1101 1101
1110	1110 1110 1110
1111	1111 1111 1111

Using this code we can correct up to one bit error during transmission.

This gain comes at a price: $\mathcal{C}_2$ has information rate $\frac{4}{12} = \frac{1}{3}$, lower than the information rate of $\mathcal{C}_1$ which is $\frac{4}{5}$.

The information rate of a binary code is the ratio of the number of significant bits of information in each word, to the total length of each word.

More generally for an alphabet of size $|F| = q$, the *information rate* of a code $\mathcal{C}$ of length n over F is

$$\frac{\log_q |\mathcal{C}|}{n}.$$

We seek codes with

(i) high information rate, *and*

(ii) high error-correcting capability.

The goal (ii) requires that codewords be 'far apart' from each other.

Hamming Distance

The *Hamming distance* between two words $x, y \in F^n$, denoted $d(x, y)$, is the number of coordinate positions in which they differ.

E.g. $d(10010, 00111) = 3$.

The *minimum distance* of a code $\mathcal{C} \subseteq F^n$ is the minimum of $d(x, y)$ for all $x \neq y$ in the code $\mathcal{C}$.

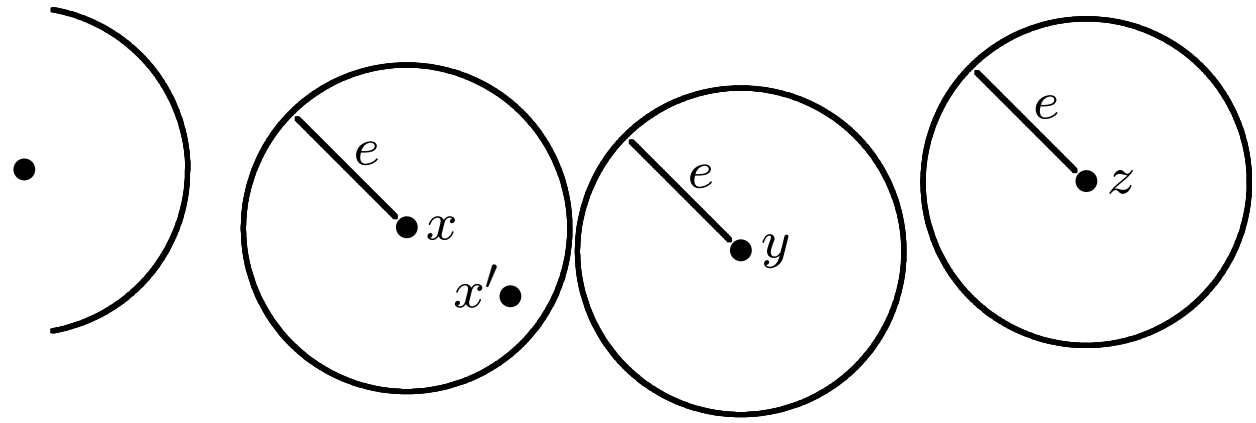
Theorem. *A code $\mathcal{C}$ corrects e errors if and only if the minimum distance of C is at least $2e+1$.*

Proof. Suppose $\mathcal{C}$ has minimum distance at least $2e+1$. If a codeword $x \in \mathcal{C}$ suffers at most e bit errors, the corrupted word x' satisfies $d(x', x) \leq e$. And x is the *only* codeword

having distance $\leq e$ from x' since for every codeword $y \neq x$,

$$2e+1 \leq d(y,x) \leq d(y,x')+d(x',x) \leq d(y,x')+e$$

by the triangle inequality, so $d(x',y) \geq e+1$. The word x' is unambiguously decoded as $x \in \mathcal{C}$. The converse is clear. □



Balls of radius e centered at codewords

Relationship with Sphere Packing

Finding a large code with minimum distance e is the same as packing as many balls of radius e as possible in F^n.

Example 3:
The Binary Hamming Code of Length 7

The following binary code $\mathcal{C}_3$ of length 7 has minimum distance 3 and so corrects one bit error.

Message word	Codeword
0000	0000000
0001	1010101
0010	0110011
0011	1100110
0100	0001111
0101	1011010
0110	0111100
0111	1101001
1000	1111111
1001	0101010
1010	1001100
1011	0011001
1100	1110000
1101	0100101
1110	1000011
1111	0010110

Encoding Using a Generator Matrix

In the binary Hamming code $\mathcal{C}_3$, the codeword for the message $x = (x_1, x_2, x_3, x_4) \in F^4$ is the matrix product xG where

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

For example, the codeword for 1010 is

$$(1, 0, 1, 0)G = (1, 0, 0, 1, 1, 0, 0).$$

Decoding Using a Check Matrix

A binary word $w \in F^7$ is a codeword in $\mathcal{C}_3$ if and only if $Hw^\top = 0$ where

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

If $w \notin \mathcal{C}_3$ then $Hw^\top$ is the binary representation of $i \in \{1, 2, \ldots, 7\}$ and by switching the ith bit of w we obtain the unique codeword at distance 1 from w.

For example, $w = 0011011$ gives

$$Hw^\top = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

so to decode w we switch the 6th bit of w, giving 0011001 as the unique codeword at distance 1 from w.

Syndromes

Note that the vector $Hw^\top$, called the *syndrome* of w, does not depend on the original message word, but only on the bit error incurred.

Linear Codes

A code $\mathcal{C} \subseteq F^n$ is *linear* if the alphabet F is a field and $\mathcal{C}$ is a subspace of F^n.

An $[n, k, d]$ q-ary code is a k-dimensional subspace of F^n with minimum distance d, where $q = |F|$. In this case $|\mathcal{C}| = q^k$ and so the information rate of $\mathcal{C}$ is

$$\frac{\log_q |\mathcal{C}|}{n} = \frac{k}{n}.$$

$\mathcal{C}_1$ is a $[5, 4, 2]$ binary code.

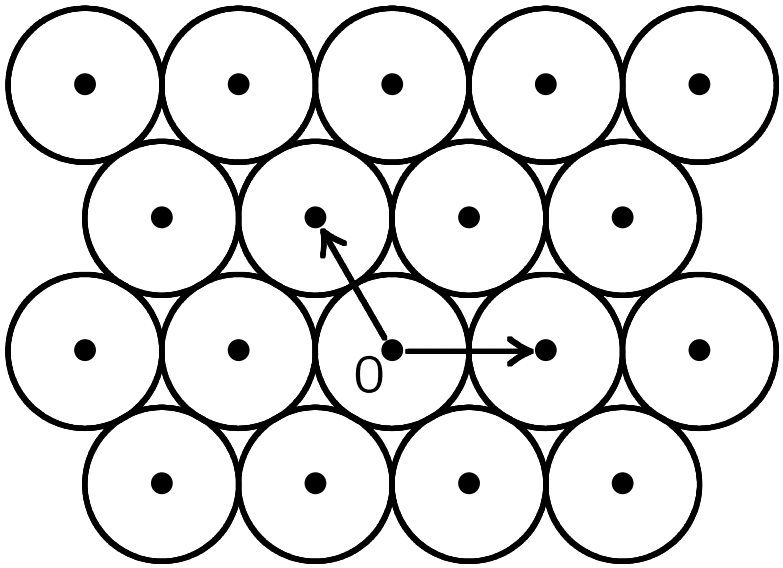
$\mathcal{C}_2$ is a $[12, 4, 3]$ binary code.

$\mathcal{C}_3$ is a $[7, 4, 3]$ binary code.

Some Good Reasons for Using Linear Codes

1. Linearity reduces the encoding and decoding processes to easily automated linear algebra.

2. Many of the best codes (i.e. highest information rate for a given length and minimum distance) are linear. We suspect this to be true by analogy with dense sphere-packings.



Centres of balls consist of all $\mathbb{Z}$-linear combinations of the two vectors shown.

Hamming Weight

The *Hamming weight* of a word $w \in F^n$ is $d(w, 0)$, i.e. the number of nonzero coordinates in the vector w.

In any linear code $\mathcal{C}$, $d(x, y) = d(x-y, 0)$ where $x-y \in \mathcal{C}$ so the minimum distance of $\mathcal{C}$ is simply equal to the *minimum weight* of $\mathcal{C}$, i.e. the minimum of $d(w, 0)$ for all $w \neq 0$ in $\mathcal{C}$.

Shannon's Theorem

Shannon showed that codes exist with probability of decoding errors as small as desired, *and* high information rate (depending on the channel).

Consider binary codes for which each bit transmitted has probability $p < \frac{1}{2}$ of error, and errors in different bits are statistically independent.

Fix a desired information rate R with $0 < R < 1-H(p)$ where

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

(the *entropy function.*)

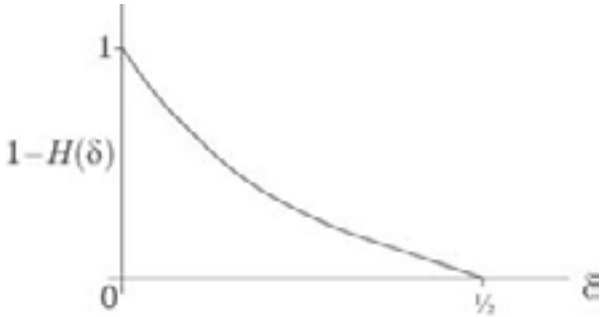
Theorem (Shannon, 1948)**.** *For all $\varepsilon > 0$ there exists a code $\mathcal{C}$ with information rate at least R such that the probability of incorrectly decoding a typical codeword is less than ε.*

The Gilbert-Varshamov Bound

Fix $\delta < \frac{1}{2}$.

For large n, there exist binary codes of length n, minimum distance at least δn, and infor-

mation rate as close as desired to $1 - H(\delta)$.



Recent Improvement

Tsfasman, Vlādut and Zink (1982) used algebraic curves over finite fields to obtain codes which (for $q \geq 49$) do better than the Gilbert-Varshamov bound (i.e. have asymptotically higher information rate for the same length and minimum distance).

$\mathbb{Z}_4$-Linear Codes

Coding theorists have long been puzzled by the fact that for certain n and d, the best binary codes of length n and minimum distance d (i.e. highest information rate) are not linear.

Calderbank, Hammons, Kumar, Sloane and Solé (c. 1995) showed that such codes are $\mathbb{Z}_4$-linear ($\mathbb{Z}_4 = \{0, 1, 2, 3\}$ with addition and multiplication mod 4; this is *not* a field!)