Can You Hear Me Now?

An Introduction to Coding Theory

William J. Turner

Department of Mathematics & Computer Science
Wabash College
Crawfordsville, IN 47933

19 October 2004

Outline

Coding Theory vs. Cryptography

Coding Theory

Efficient and reliable communication in an uncooperative (possibly hostile) environment.

- Encode and decode
- *Clear* communication

Cryptography

Disguising messages so only certain people can see through the disguise.

- Encrypt and decrypt
- *Hidden* communication

Applications of Coding Theory

Example (Digital communication)

- Email, internet, intranet
- Radio, satellite
- Photographs from deep space
- Remote control of unmanned drones

Example (Products)

- Store scanners (bar codes)
- International Standard Book Number (ISBN)

Error Detecting Codes vs. Error Correcting Codes

Error Detecting Codes

Detect when an error occurse in transmission.

Error Correcting Codes

Detect and *correct* errors in transmission.

Repetition Code

Idea

Send each letter or word multiple times.

Example (Send each letter of message twice)

- Receive message 2234.
- Detect error in second digit: 3 or 4?

Example (Send each letter of message three times)

- Receive message 222343.
- Correct error in second letter: 3.

Parity Check

Idea

- Augment binary messages with an extra bit to make an even number of 1s.
- If you receive a message with an odd number of bits, you know an error occurred in transmission.

Examples (Seven Bits + Parity Check)

1. $22 \longrightarrow (0,0,1,0,1,1,0) \longrightarrow (0,0,1,0,1,1,0,1)$
2. $23 \longrightarrow (0,0,1,0,1,1,1) \longrightarrow (0,0,1,0,1,1,1,0)$

International Standard Book Number (ISBN)

Idea

Ten-digit number (codeword) assigned by publisher:

$$x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10}$$

- x_1: language
- $x_2 x_3$: publisher
- $x_4 x_5 \cdots x_9$: book (assigned by publisher)
- x_{10}: assigned so $\sum_{i=1}^{10} i\ x_i \equiv 0 \pmod{11}$
- $x_{10} = \sum_{i=1}^{9} i\ x_i \bmod 11$

Possible to

- Detect and correct error in one digit.
- Detect transposition of two digits.

International Standard Book Number (ISBN)

Example (*Accelerated C++* by Koenig and Moo)

- Published in English: $x_1 = 0$
- Published by Addison-Wesley: $x_2 x_3 = 20$
- $x_4 x_5 \cdots x_9 = 170353$
- $x_{10} = ??$

$$\sum_{i=1}^{9} i\ x_i = 131$$

$$\sum_{i=1}^{9} i\ x_i \bmod 11 = 10$$

$$x_{10} = X$$

ISBN: $020170353X$

Vector Space

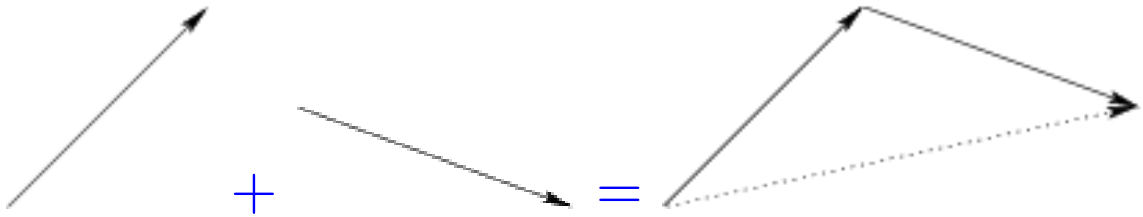
Idea

- A *vector* has a magnitude and direction.
- Adding two vectors together creates a new vector that might have a new direction.
- Multiplying a vector by a *scalar* changes its magnitude but not its direction.
- *Vector space* is set of all such combinations of vectors.

Vector Space

Example ($\mathbb{R}^2$ over the field $\mathbb{R}$)

- $\mathbb{R}$ is field of real numbers.
- $\mathbb{R}^2$ is set of all pairs of real numbers.
  I.e., all points on two-dimensional plane.



Vector Space

Example ($\mathbb{Z}_2^n$ over the field $\mathbb{Z}_2$)

- $\mathbb{Z}_2$ is field of integers modulo 2.

$$1 + 1 = 2 \bmod 2 = 0$$

- $\mathbb{Z}_2^n$ is set of all n-tuples of 0s and 1s.

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{Z}_2^7$$

Linear Codes

Definition (Linear Code)

A subspace of a vector space.

I.e., a subset of a vector space that is itself a vector space.

Example ($C_{7,4}$)

The set linear combinations (span) over $\mathbb{Z}_2$ of

$$u_1 = (1,0,0,0,0,1,1)$$
$$u_2 = (0,1,0,0,1,0,1)$$
$$u_3 = (0,0,1,0,1,1,0)$$
$$u_4 = (0,0,0,1,1,1,1)$$

Every vector in $C_{7,4}$ is a unique linear combination of $\{u_1, u_2, u_3, u_4\}$ because these four vectors are *linearly independent*.

$\{u_1, u_2, u_3, u_4\}$ forms a *basis* for $C_{7,4}$.

Encoding a Message

Idea

Use word's bits as scalars for linear combination of basis vectors.

Example ($C_{7,4}$)

- Want to encode the word $x = (1, 0, 1, 1)$.
- Encode x as $w = 1u_1 + 0u_2 + 1u_3 + 1u_4 = (1, 0, 1, 1, 0, 1, 0)$.

Encoding a Message

Definition (Generator Matrix)

- A matrix formed from the basis vectors of a linear code.
- Encode word by multiplying with generator matrix.

Example ($C_{7,4}$)

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

$$w = xG = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Decoding a Message

Idea

Solve the linear system

$$xG = w$$

or

$$G^T x^T = w^T$$

Example ($C_{7,4}$)

Suppose received message $w = (0, 1, 1, 0, 0, 1, 1)$.

$$w = xG \Longrightarrow x = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}$$

Detecting and Correcting Errors

Idea

- Use a matrix H whose (left) null space is the linear code.
  I.e., $wH = 0$ for every codeword w.
- If $wH \neq 0$, an error must have occurred.
- wH may tell us where error occurred.

Detecting Errors

Example ($C_{7,4}$)

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

- Suppose received message $w = (0, 1, 1, 1, 0, 1, 1)$.
- $wH = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \neq 0$ so an error must have occurred.

Correcting Errors

Example ($C_{7,4}$)

- $wH = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \neq 0$
- wH is second row of $H \Longrightarrow$ error is second entry of w.
- Correct $w_c = (0, 0, 1, 1, 0, 1, 1)$.
- Decode $x = (0, 0, 1, 1)$.

Detecting and Correcting Errors

Definition (Hamming Distance)

The distance between two codewords is the number of bits in which they differ.

Example ($C_{7,4}$)

$$d(u_1, u_2) = 4$$

$$d(u_1, u_4) = 3$$

Detecting and Correcting Errors

Idea

Suppose $\min_{u,v \in C} \{d(u,v)\} = d$.

- Can detect at most $d-1$ errors.
  - Change d bits $\Longrightarrow$ may get another codeword.
- Can correct at most $d/2 - 1$ errors.
  - Change $d/2$ bits $\Longrightarrow$ may be closer to different codeword.

Detecting and Correcting Errors

Example ($C_{7,4}$)

$$\min_{u,v \in C_{7,4}} \{d(u,v)\} = 3$$

- Can detect at most 2 errors.
  - Start with word $x = (1,0,0,0)$.
  - Encode as $w = (1,0,0,0,0,1,1)$.
  - Three errors during transmission to receive $\bar{w} = (0,0,0,1,1,1,1)$.
  - Decode message $\bar{x} = (0,0,0,1)$.
- Can correct at most 1 errors.
  - Start with word $x = (1,0,0,0)$.
  - Encode as $w = (1,0,0,0,0,1,1)$.
  - Two errors during transmission to receive $\bar{w} = (1,1,1,0,0,1,1)$.
  - "Correct" message as $\bar{w}_c = (0,1,1,0,0,1,1)$.
  - Decode message $\bar{x} = (0,1,1,0)$.

Detecting and Correcting More Errors

Idea

Must increase distance between codewords.

- Want to correct e errors.
- Must be able to move distance e from a codeword and still remain at least distance $e+1$ from any other codeword.
- Consider *sphere* of radius e around each codeword.
- Spheres of any pair of codewords cannot intersect.
- Increase $e \Longrightarrow$ increase size of spheres
  - Fewer spheres can fit in same box.
  - Need bigger box to keep same number of spheres.

Sphere packing problem

References

Raymond Hill (1986).
A First Course in Coding Theory.
Oxford University Press, Oxford.

Richard Hill (1996).
Elementary Linear Algebra with Applications.
Harcourt, Orlando, 3rd edition.