Linear Algebra

Course 10

Chapter 4. Introduction to Coding Theory

Part I

Contents

Coding theory

Starting points:

- Shannon 1948: Information Theory
- Hamming 1950: Error-Correcting Codes

Main classes of codes:

- source coding: data compression
- channel coding: error-correcting codes

Probabilities of errors

Suppose that we have a communication channel whose probability of a correct transmission is p. The probability of t errors in a message of length m is

$$C_m^t p^{m-t}(1-p)^t.$$

For instance, for $p = 0.99$ and $m = 50$, we have the following table:

t	Probability of t errors
0	0.605
1	0.3056
2	0.0756
3	0.0122
4	0.00145

These probabilities decrease if m is small enough, more precisely when $m < \frac{p}{1-p}$. Hence we should not expect too many errors during a transmission. But still they happen, and should be detected and corrected.

A first example

EAN-13 International Article Number

It is a sequence of 13 digits $a_1, a_2, \ldots, a_{13}$ that identifies a product. Digit a_{13} is a check digit that is computed as

$$a_{13} = 10 - (a_1 + 3a_2 + a_3 + 3a_4 + \cdots + a_{11} + 3a_{12}) \bmod 10.$$

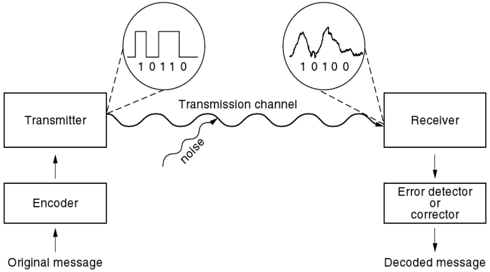
Digits are written in binary; black bars for 1, white bars for 0.

In particular:

- ISBN (International Standard Book Number)
- UPC (Universal Product Code) etc.

Error-correcting (detecting) codes

General scheme:

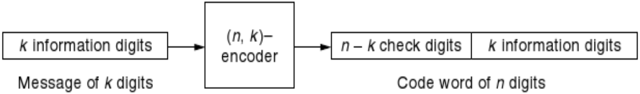


Different codes are suitable for different applications:

- satellite and space transmissions
- credit cards
- CD's, DVD's, Blu-ray discs etc.

The coding problem

- We discuss *binary codes*. In general: codes over finite fields.
- We consider *symmetric channels*: the probability of 1 being changed into 0 is the same as that of 0 being changed into 1.
- It is assumed that the number of errors is less than the number of correctly transmitted bits.
- We talk about (n, k)-*codes*:



There are 2^k possible messages, and so 2^k code words.
There are 2^n possible words received.

Aim

Find the right balance between k and $n - k$.

Two simple codes - The $(3, 2)$-parity check code

- The check digit is the sum modulo 2 of the message digits.
- Encoding:

Message	Code word
00	000
01	101
10	110
11	011

  How many errors can this code detect/correct?

- Decoding:

Received words	101	111	100	000	110
Parity check	passes	fails	fails	passes	passes
Decoded words	01	-	-	00	10

Two simple codes - The $(3, 1)$-repeating code

- The two check digits repeat the message digit.
- Encoding:

Message	Code word
0	000
1	111

  How many errors can this code detect/correct?
- Decoding:

Received words	111	010	011	000
Decoded words	1	0	1	0

Hamming distance

Definition

The *Hamming distance* between two words of the same length is the number of positions in which they difer.

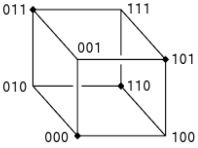
Notation $d(u, v)$.
Example: $d(101, 100) = 1$, $d(110, 001) = 3$, $d(101, 011) = 2$.

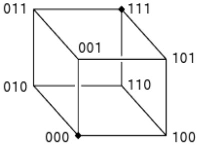
Theorem

The Hamming distance has the following properties hold for every $u, v, w \in \mathbb{Z}_2^n$:
(1) $d(u, v) = d(v, u)$.
(2) $d(u, v) + d(v, w) \geq d(u, w)$.
(3) $d(u, v) \geq 0$ with equality if and only if $u = v$.

Hamming distance - cont.

- In an (n, k)-code, the 2^n received words can be thought of as placed at the vertices of an n-dimensional cube with unit sides.
- The Hamming distance between two words is the shortest distance between their corresponding vertices along the edges of the n-cube.
- The 2^k code words form a subset of the 2^n vertices, and the code has better error-correcting and error-detecting capabilities the farther apart these code words are.
- Cube representations of the $(3, 2)$-parity check and $(3, 1)$-repeating codes:





Error detection/correction capabilities

Theorem

A code detects all sets of t or fewer errors $\iff$ the minimum Hamming distance between code words is at least $t+1$.

Theorem

A code corrects all sets of t or fewer errors $\iff$ the minimum Hamming distance between code words is at least $2t+1$.

Code	Minimum distance between words	No. of detectable errors	No. of correctable errors	Information rate
(n, k)-code	d	d-1	$\leq \frac{d-1}{2}$	$\frac{k}{n}$
$(3, 2)$-parity check code	2	1	0	$\frac{2}{3}$
$(3, 1)$-repeating code	3	2	1	$\frac{1}{3}$

Polynomial representation

- A binary n-digit word $a_0 a_1 \dots a_{n-1}$ may be identified with a polynomial $a_0 + a_1 X + \cdots + a_{n-1} X^{n-1} \in \mathbb{Z}_2[X]$.

Definition

Let $p \in \mathbb{Z}_2[X]$ be of degree $n-k$. The *polynomial code generated by* p is an (n,k)-code whose code words are those polynomials of degree less than n which are divisible by p. Then the polynomial p is called the *generator* of the code.

- A message of length k is represented by a polynomial $m \in \mathbb{Z}_2[X]$ of degree less than k.
- Since the message is stored in the right hand side of a word, the message digits are carried by the higher-order coefficients of a polynomial. So we consider $m \cdot X^{n-k}$.

Polynomial representation - cont.

- To encode the message polynomial m we first use the Division Algorithm to find unique $q, r \in \mathbb{Z}_2[X]$ such that

$$m \cdot X^{n-k} = q \cdot p + r\,, \quad degree(r) < degree(p) = n - k.$$

Then the code polynomial is

$$v = r + m \cdot X^{n-k}.$$

The check digits of the message are carried by r.

Theorem

With the above notation, the code polynomial v is divisible by p.

Proof. We have $v = r + m \cdot X^{n-k} = r + q \cdot p + r = q \cdot p$, because $r \in \mathbb{Z}_2[X]$, and so $r + r = 0$.

Polynomial representation - examples

Example 1. *Let* $p = 1 + X^2 + X^3 + X^4 \in \mathbb{Z}_2[X]$ *be the generator polynomial of a* $(7,3)$*-code. Let us encode the message 101.*

Solution. Note that $n = 7$ and $k = 3$.

$$\begin{aligned}
\text{message } 101 &\rightsquigarrow m = 1 \cdot 1 + 0 \cdot X + 1 \cdot X^2 = 1 + X^2 \\
&\rightsquigarrow mX^{n-k} = (1 + X^2) \cdot X^4 = X^4 + X^6 \\
&\rightsquigarrow r = mX^{n-k} \bmod p = (X^4 + X^6) \bmod p = 1 + X \\
&\rightsquigarrow v = r + mX^{n-k} = 1 + X + X^4 + X^6 \\
&\rightsquigarrow \text{code word } \boxed{1100}\boxed{101}
\end{aligned}$$

Example 2. *If the generator polynomial of a* $(6,3)$*-code is* $p = 1 + X + X^3 \in \mathbb{Z}_2[X]$, *test whether the following received words contain detectable errors: 100011, 100110.*

Solution. We check if the received words are code words, that is, their associated polynomials are divisible by p [...].

Polynomial representation - examples

Example 3. *Write down all the code words for the* $(6, 3)$*-code generated by the polynomial* $p = 1 + X + X^3 \in \mathbb{Z}_2[X]$.

Solution. Note that $n = 6$, $k = 3$, and we have $2^k = 8$ code words. We obtain the following table:

message	code word
000	000000
001	111001
010	011010
011	100011
100	110100
101	001101
110	101110
111	010111

E.g.: $110 \rightsquigarrow m = 1 + X \rightsquigarrow mX^{n-k} = X^3 + X^4$

$$\rightsquigarrow r = mX^{n-k} \bmod p = (X^3 + X^4) \bmod p = 1 + X^2$$

$$\rightsquigarrow v = r + mX^{n-k} = 1 + X^2 + X^3 + X^4 \rightsquigarrow \boxed{101}\boxed{110}$$