Coding Theory

CO 331

Prof. Alfred John Menezes

Preface

Disclaimer Much of the information on this set of notes is transcribed directly/indirectly from the lectures of CO 331 during Winter 2020 as well as other related resources. I do not make any warranties about the completeness, reliability and accuracy of this set of notes. Use at your own risk.

For any questions, send me an email via https://notes.sibeliusp.com/contact/.

You can find my notes for other courses on https://notes.sibeliusp.com/.

Sibelius Peng

Contents

P	reface	1
0	Pre	3
1	Introduction & Fundamentals	5

0

Pre

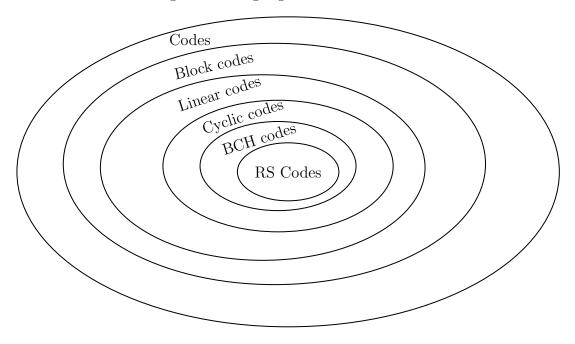
Example: Replication code

```
codewords
source msgs
                       0
      0
                       1
      1
\# of errors/codeword that be detected: 0
# errors/codeword that can be corrected: 0
Rate: 1
                   codewords
source msgs
                       00
      0
      1
                       11
\# of errors/codeword that be detected: 1
# errors/codeword that can be corrected: 0
Rate: 1/2
                   codewords
source msgs
      0
                      000
                      111
# of errors/codeword that be detected: 2
# errors/codeword that can be corrected: 1 (nearest neighbour decoding)
Rate: 1/3
source msgs
                   codewords
      0
                     00000
                     11111
\# of errors/codeword that be detected: 4
# errors/codeword that can be corrected: 2 (nearest neighbour decoding)
Rate: 1/5
```

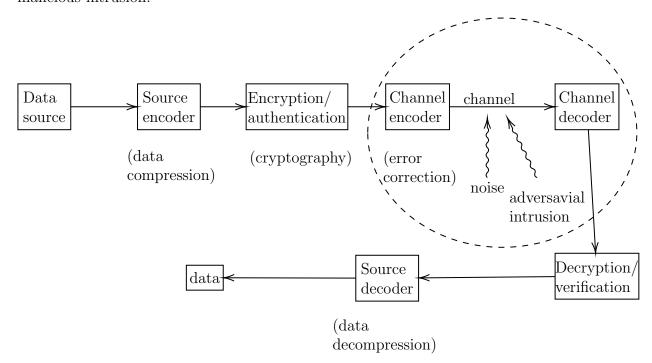
CHAPTER 0. PRE 4

Goal of Coding Theory Design codes so that:

- 1. High information rate
- 2. High error-correcting capability
- 3. Efficient encoding & decoding algorithms



The big picture In its broadest sense, coding deals with the reliable, efficient, secure transmission of data over channels that are subject to inadvertent noise and malicious intrusion.



mid: Feb 26th

Introduction & Fundamentals

alphabet, word, length...

An alphabet A is a finite set of $q \ge 2$ symbols. E.g. $A = \{0, 1\}$.

A word is a finite sequence of symbols from A. (tuples or vectors)

The *length* of a word is the number of symbols in it.

A code C over A is a finite set of words over A (of size ≥ 2).

A codeword is a word in C.

A block code is a code where all codewords have the same length.

A block code C of length n containing M codewords over A is a subset $C \subseteq A^n$, with |C| = M. This is denoted by [n, M].

Example:

 $A = \{0, 1\}$. $C = \{00000, 11100, 00111, 10101\}$ is a [5, 4]-code over $\{0, 1\}$.

Messages		Codeword
00	\rightarrow	00000
10	\rightarrow	11100
01	\rightarrow	00111
11	\rightarrow	10101

Encoding 1-1 map

The channel encoder transmits only codewords. But, what's received by the channel decoder might not be codeword.

Example:

Suppose the channel decoder receives r = 11001. What should it do?

Example: q = 2 (Binary symmetric channel, BSC)

q=3

Assumptions about the communications channel

- 1) The channel only transmits symbols from A.
- 2) No symbols are deleted, added, or transposed.
- 3) (Errors are "random") Suppose the symbol transmitted are X_1, X_2, X_3, \ldots Suppose the symbols received and Y_1, Y_2, Y_3, \ldots Then for all $i \geq 1$, and all $i \leq j, k \leq q$,

$$Pr(Y_i = a_j | X_i = a_k) = \begin{cases} 1 - p, & \text{if } j = k \\ \frac{p}{q - 1}, & \text{if } j \neq k \end{cases}$$

where p = symbol error prob.

Notes about BSC

- (i) If p = 0, the channel is perfect.
- (ii) If $p = \frac{1}{2}$, the channel is useless.
- (iii) If $1 \ge p > \frac{1}{2}$, then simply flip all bits that are received.
- (iv) WLOG, we will assume that 0 .
- (v) Analogously, for a q-ary channel, we can assume that 0 . (Optional exercise)

Hamming distance

If $x, y \in A^n$, the Hamming distance d(x, y) is the # of coordinate positions in which x & y differ.

The distance of a code C is

$$d(C) = \min\{d(x,y) \in C, x \neq y\}$$

Example:

$$d(10111, 01010) = 4$$

Theorem 1.1

d is a metric. For all $x, y, z \in A^n$

- (i) $d(x, y) \ge 0$, and d(x, y) = 0 iff x = y.
- (ii) d(x,y) = d(y,x)

(iii)
$$\triangle$$
 inequality $d(x, z) \le d(x, y) + d(y, z)$

rate

The rate of an [n, M]-code C over A with |A| = q is

$$R = \frac{\log_q M}{n}.$$

If the source messages are all k-tuples over A,

$$R = \frac{\log_q(q^k)}{n} = \frac{k}{n}.$$

Example:

$$C = \{00000, 11100, 00111, 10101\} \qquad A = \{0, 1\}$$
 Here $R = \frac{2}{5}$ and $d(C) = 2.$

Index

Α	R
alphabet, word, length 5	rate
Н	
Hamming distance	