2WC09 Coding & Crypto 1

Tuesday 10 sep 2013

Christiane Peters, Technical University of Denmark, email: cppet (at) dtu.dk

---

- C: **code**
- over an **alphabet** Q where $|Q|=q$
- q**-ary code**
- $|C|=M$ **size** of the code
- send codeword **c** (information + redundancy). Receiver gets **c'** and will try to decode

- consider **hard-decision decoding**, i.e., no erasures, every entry in **c** is a valid symbol from the alphabet

- **Maximum-likelihood decoding**: decode **y** to the nearest codeword

- Def: Distance: d between $x=(x_1,\ldots,x_n)$ and $y=(y_1,\ldots,y_n)$

$d(x,y)=\#\{1\leq i\leq n: x_i\neq y_i\}$

=number of symbols that change when going from **x** to **y**

- Error-correcting capability of a code C with min dist d is $e=(d-1)/2$

- If no more than e errors are introduced one can recover codeword with maximum-likelihood decoding correctly

- Example: **repetition code** (1,1,1,1,1), (0,0,0,0,0) of length $n=5$ has minimum distance $d=2$. Can correct 2 errors.

- **Def: Minimum distance**

The minimum distance of a code $C\neq 0$ is

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

- **Hamming bound** last week

- **covering radius**: max distance of an arbitrary word in the ambient space to a codeword

$\rho = \max\{ d(x,C) \mid x \text{ in } Q^n \}$

- every word x in Q^n is at distance at most ρ to some codeword, say c, it is also inside at least one of the spheres of radius ρ around the codewords

- So spheres of radius ρ around all codewords cover Q^n

- Spheres of radius $e = (d-1)/2$ around codewords have no overlap

- So $e \leq \rho$

- Def: code is called **perfect** if $e = \rho$

- All spheres around codewords have no overlap

- For odd length n the repetition code is perfect

$B_{[(n-1)/2]}(0,...0)$ contains all elements of $\{0,1\}^n$ with $\leq [(n-1)/2]$ 1's and

$B_{[(n-1)/2]}(1,...1)$ contains all the other elements in $\{0,1\}^n$

- e.g.

n=5 $B_2(0,...0)$ contains (0,..0) and all words with one or two 1's

(5 choose 2)+(5 choose 1)+ (5 choose 0)=16 words

$B_2(1,...1)$ contains (1,...,1) and all words with three or more 1's

(5 choose 3)+ (5 choose 4)+ (5 choose 5)=16 words

Partition of the set $\{0,1\}^n$

- There are $(n \text{ choose } i)(q-1)^i$ elements at distance i

- Thm 2.1.5 (didn't call it like this but showed the formula)

- Thm 2.1.7 **Sphere-Packing Bound**

Thm. ***Singleton-bound***

Let ***C*** be a ***q***-ary ***(n,M,d)*** code. Then

$$M \leq q^{n-d+1}$$

Proof.

Erase in all codewords the last ***d-1*** coordinates. Because all words have minimum distance ***d*** in ***C*** the new words will still be distinct. Their length is ***n-(d-1)=n-d+1***. However, there are only q^{n-d+1} possible words of lengt n-d+1 over an alphabet of size ***q***

- Codes with parameters (n, q^{n-d+1}, d) are called **maximum-distance separable** or simply **MDS codes**

Thm 2.1.9 Gilbert-Varshamov bound

Proof.

Let ***C'*** be a code with ***M'*** elements and minimum distance ***>d***.

If M'*sum i=0^(d-1)<q^n

then the spheres of radius ***d-1*** do not cover the whole space Q^n and we can find at least one element ***x*** in Q^n which is not in ***C'*** with ***d(x,C)≥d***. Build a new code ***C=C' join {x}*** with ***M'+1*** codewords and minimum distance ***d***.

QED

- Note this was a purely theoretical result. It took years to find actual examples

Linear Codes

- Want alphabets to be finite fields, i.e. $Q=F_q$
- note: book uses notation $F_q=GF(q)$ "Galois Field", both are commonly used
- Def. A linear code is a vector space $C \text{ subset } (F_q)^n$
- As vector space C has a dimension, denoted by k, so C has size $M=q^k$
- For a linear code we sometimes write **(n,k,d)-code** to say it has length n, dimension k and minimum distance d

- Hamming weight of a vector x in $(F_q)^n$ is the number of non-zero coordinates of c

- We have $w(x)=d(x,0)$ for any element x in $(F_q)^n$

- Thm 2.2.3 The minimum distance of a linear code equals the minimum weight of a nonzero codeword in C.

Proof. min distance : $d=min\{d(x,y) \text{ for all } x\neq y \text{ in } C\}$

$d(x,y)=d(x-y,0)=w(x-y)$

Since C is linear $x-y$ is a codeword in C.

- Recall linear algebra
- A linear code C is a vector space of dimension k.

So we can write down a basis of C consisting of k linearly independent vectors in C

- Write down as a kxn matrix (k basis elements, each of length n)
- Consider a vector m in $(F_q)^k$. Then $m \cdot G$ is just taking linear combinations of the rows of G

- Such a matrix gives an embedding of $(F_q)^k$ into $(F_q)^n$
- messages are mapped to codewords.

- Def. A generator matrix ***G*** for ***C*** is a ***k x n*** matrix such that

$$C=\{m\cdot G \mid m \text{ in } (F_q)^k\}$$

- note ***G*** has **full rank** by construction

- Examples:

- **Repetition code** of length ***n***

F_q ***into*** $(F_q)^n$ ***, c --> (c,c,...,c)***

has ***G=(1,1,...,1)***; codewords are generated as ***0·G*** and ***1·G***

- The **binary even-weight code** (n,n-1,2) has generator matrix

(1 0 0 1)

(0 1 0 1)

( )

(0 0 1 1)

(obviously independent, exactly two 1's per row, taking linear combinations of G's rows produces words of even weight)

- Any generator matrix can be written in ***systematic form*** (or ***standard form***) $G=(I_k|Q)$ for some ***kx(n-k)*** matrix ***Q***

mapping m to ***m·G=(m,******)***. If ***G*** is in systematic form, we call the first ***k*** symbols of ***m·G*** **information symbols** and the remaining ***r=n-k*** **redundancy symbols**

- very important for code-based cryptography!

- Translation of bounds

- linear Sphere-packing bound

- linear Singleton bound $k\leq n-d+1$

- linear GV bound