10th Sept 2022

DNA Storage & Security

→ DNA

- Four letter alphabet
  - Adenine (A)
  - Cytosine (C)
  - Guanine (G)
  - Thymine (T)

DNA Synthesizer - write
DNA Sequencer - read.

Steps:
1. Encoding
2. Synthesis
3. Storage
4. Retrieval
5. Sequencing
6. Decoding

Repetition Codes:

Code $C = \{000, 111\}$

$0 \rightarrow 000$
$1 \rightarrow 111$

min Hamming dist = d

we can correct $\left\lfloor \frac{d-1}{2} \right\rfloor$ errors.

& detect $d-1$ errors.

→ Z_n is a field if n is prime

$GF(4) = \{a+\alpha b \mid \alpha, b \in Z_2, \alpha^2+\alpha+1=0\}$

$= \{0, 1, \alpha, \alpha^2\}$

$= \{0, 1, \alpha, 1+\alpha\}$

$= \{00, 01, 10, 11\}$

→ $GF(p^m) = GF(q)$

→ $Z_2^n = \{(x_1, x_2, \ldots x_n) \mid x_i \in Z_2\}$

→ $F^n = \{(x_1, x_2, \ldots x_n) \mid x_i \in F\}$

$C = \{000, 111\}$ is a subspace of Z_2^3

→ Basis of $Z_2^3 = \{001, 010, 100\}$

$\dim Z_2^3 = 3$

→ $F_q^n = GF(q) = \{(x_1, x_2, \ldots x_n) \mid x_i \in F_q\}$

$C \leq F_q^n$

k-dim subspace

$\mathbb{Z}_2^n$ is a vector space of dim n

Linear code: $C \underset{K\text{-dim}}{\leq} \mathbb{Z}_2^n$

Eg: $C \underset{1\text{-dim}}{\leq} \mathbb{Z}_2^3 = \{000, 111\}$

$B = \{\bar{v}_1, \bar{v}_2, \ldots \bar{v}_k\}$

(1) $\sum_{i=1}^{n} \lambda_i v_i = 0 \Rightarrow \lambda_i = 0$ Linearly independent

(2) $\langle \bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k \rangle = C$

$$G = \begin{bmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_k \end{bmatrix}_{k \times n}$$

Generator matrix of code C.

$(x_1, x_2, \ldots x_k)\, G_{k \times n} = (y_1, y_2, \ldots y_n)_{1 \times n}$

block of K bits | Generator matrix | encoded bits

$C = [n, k, d]_2$

$d = \min \{ d_H(\bar{x}, \bar{y}) \mid \bar{x}, \bar{y} \in C \}$

$\bar{x} \neq \bar{y}$

Eg

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

$C = \{000, 110, 101, 011\}$

$C = [3, 2, 2]_2$

$$[0\ 1] \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = [1\ 0\ 1]$$

$F_q \rightarrow$ Finite field with q elements

$q = p^m$ p is prime.

$F_q^n = \{ (x_1, x_2, \cdots x_n) \mid x_i \in F_q \}$

$C \underset{k\text{-dim}}{\leq} F_q^n \qquad \dim C = k$

Dual Code

$$C^{\perp} = \{ \bar{y} \in F_q^n \mid \langle \bar{x}, \bar{y} \rangle = 0 : \forall \bar{x} \in C \}$$

$$\boxed{\begin{aligned} \dim C^{\perp} &= n - k \\ &= n - \dim C \end{aligned}}$$

(1) C is self orthogonal if $C \subseteq C^{\perp}$

(2) C is self dual if $C = C^{\perp}$

Eg $$C^{\perp} = \{ \bar{y} \in \mathbb{Z}_2^n \mid \langle \bar{x}, \bar{y} \rangle = 0 \ \forall \bar{x} \in C \}$$

$n = 3$ & $C = \{000, 111\}$

$$C^{\perp} = \{000, 110, 101, 011\}$$

$$\boxed{\dim C + \dim C^{\perp} = \dim \mathbb{Z}_2^3}$$

$$1 + 2 = 3$$

In simple cases, $C \cap C^{\perp} = 0$ vector.

Standard Generator matrix

$C = [n, k, d]_2$

$G_{k \times n} = [I_k \mid A]$

$H \rightarrow$ parity check matrix

$H = [-A^t \mid I_{n-k}]$ $\quad \boxed{\text{rank } H = n-k}$

& $\boxed{GH^t = 0}$

$\langle G \rangle = C = \{\bar{x} \mid Hx^t = 0\}$ = Null space of H

Eg $G = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} = [I_2 \mid A]$

$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad A^t = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

$H = [-A^t \mid I_{n-k}] = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$

$$C = \{ \bar{x} \in Z_2^n \mid Hx^t = 0 \}$$

$$Z_2^n = \{ (x_1, x_2, \dots x_n) \mid x_i \in Z_2 \}$$

$C \leq Z_2^n$ (subset)

→ $C : (n, M, d)$ (Non-linear codes)

$d \downarrow$ min d_H

$M \downarrow$ # of codewords

$n \downarrow$ length

→ $C \underset{k\text{-dim}}{\subseteq} Z_2^n$ $\quad C = [n, k, d]$ Linear Codes

$G_{k \times n} = [I_k \mid A]$ std. gen. matrix

$|C| = 2^k$ Cardinality

$\text{rank}(G) = k$

$\text{rank}(H) = n - k$

$H_{(n-k) \times n}$ P.C.M

$G H^t = 0$

$$C = \{ \bar{x} \in Z_2^n \mid Hx^t = 0 \}$$

$H = [-A^t \mid I_{n-k}]$

Page No.:
Date: / /

Array Decoding

x : sent

y : received

$$\boxed{Syn(y) = Hy^t}$$ How to decode?

* → Generate H such that no two vectors are multiple of each other

Z_2^m

Eg $m=2$ $\quad Z_2^2 = \{00, 01, 10, 11\}$ $\quad H = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}_{2\times 3}$

$C = [7, 4, 3]_2$ (n, k, d)

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}_{3\times 7}$$

$$C = \{x \in Z_2^3 \mid Hx^t = 0\}$$

$H_m : [2^m - 1, 2^m - 1 - m, 3]_2$

Hamming code

→ x ∈ C sent, y is received.

$\bar{y} = \bar{x} + \bar{e}$

$Syn(\bar{y}) = Syn(\bar{x} + \bar{e})$

$= H\bar{x}^t + H\bar{e}^t$

$= 0 + He^t$

$= He^t$

$Syn(\bar{y}) = \sum_{i=1}^{n} e_i H_i = H_a + H_b + H_c$

↓

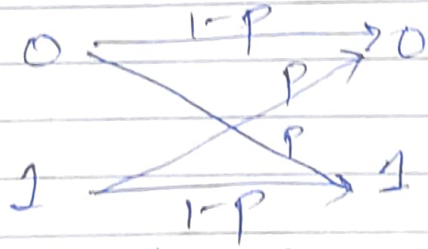
find col in H & flip that bit in y.

∴ 1 bit error can be corrected.

* $C = [000, 111]$ $C = [3, 1, 3]$

$G = [111]$

BSC (Binary Symmetric Channel)



Received $v^r = 110$ $p = 0.05$

$P(110 \text{ rec} \mid 000 \text{ sent}) = p^2(1-p)$ greater probability that codeword is sent.

$P(110 \text{ rec} \mid 111 \text{ sent}) = (1-p)^2 p$

* Cyclic code

= Linear code
- Any cyclic shift of codeword is again a codeword.

$C_0, C_1, \dots C_{n-1} \in C$

$\Rightarrow C_{n-1}, C_0, C_1, \dots, C_{n-2} \in C$

$C = \{000, 101, 011, 110\}$

$C_0, C_1, \dots C_{n-1} \rightsquigarrow C_0 + C_1 x + \dots C_{n-1} x^{n-1}$

$\underbrace{\quad\quad}_{C(x)}$

$xC(x) = C_0 x + C_1 x^2 + \dots + C_{n-1} x^n$

Here $x^n = 1$

Eg V_2^3

000	$\rightarrow 0$
100	1
010	x
001	x^2
110	$1+x$
011	$x+x^2$
101	$1+x^2$
111	$1+x+x^2$

$f(x) = x^2 + x + 1$

$Z_2[x]/(x^2+x+1) = GF(4)$.

If polynomial is irreducible, then this is a field.

Eg $Z_2[x]/(x^3-1)$ $\quad f(x) = x^3 - 1$

$= R_3$ (Ring)

$Z_2[x]/(x^n-1) = R_n$ (Polynomial Ring)

Eg: $C = \{000, 110, 101, 011\} = \{0, 1+x, 1+x^2, x+x^2\}$. C ideal R_3.

Ideal if (1) $a+b, a-b \in I$ $\forall a, b \in I$

(2) $r \cdot a \in I$ $\forall r \in R$ & $a \in I$

$\rightarrow R_3 = Z_2[x]/(x^3-1)$

$I = \{0, 1+x, 1+x^2, x+x^2\}$

$I = \langle 1+x \rangle$ $\quad \therefore r(1+x) = I \quad r \in R_3$

↳ Principal ideal if $\exists g \in I$ s.t $I = \langle g \rangle = \{rg \mid r \in R\}$

→ C is a cyclic code in $R_n = Z_2[x]/(x^3-1)$ (x^n-1) gen

(1) ∃ a unique monic poly g(x) of smallest deg in C

(2) $C = \langle g(x) \rangle$

(3) g(x) is a factor of $x^n - 1$.

$$g(x) = g_0 + g_1 x + \cdots + g_r x^r$$

$$G = \begin{bmatrix} g_0 & g_1 & g_2 & \cdots & g_r & 0 & \cdots & 0 \\ 0 & g_0 & g_1 & \cdots & & g_r & & \\ & & & & & & & \\ 0 & \cdots & & & & g_0 & g_1 & \cdots g_r \end{bmatrix}$$

This will not be a standard generator matrix

Page No.:
Date: / /

* How to find all binary cyclic codes of length 3?

$x^3 - 1 = (x+1)(x^2+x+1)$

*

Gen Poly	Code in R_3	Code in Z_2^3
1	All of R_3	all of Z_2^3
$x+1$	$\{0, 1+x, 1+x^2, x+x^2\}$	$\{000, 110, 011, 101\}$
x^2+x+1	$\{0, 1+x+x^2\}$	$\{000, 111\}$
x^3-1	$\{0\}$	$\{000\}$

*

n	Factorization of $x^n - 1$	No of cyclic codes
1	$1+x$	2
2	$(1+x)^2$	3
3	$(1+x)(1+x+x^2)$	4
4	$(1+x)^4$	5
5	$(1+x)(1+x+x^2+x^3+x^4)$	4
6	$(1+x)^2(1+x+x^2)^2$	9
7	$(1+x)(1+x^2+x^3)(1+x+x^3)$	8
8	$(1+x)^8$	8
9	$(1+x)(1+x+x^2)(1+x^3+x^6)$	8
10	$(1+x)^2(1+x+x^2+x^3+x^4)^2$	9

Hamming code can be constructed for out of