

ERROR CORRECTING CODES CONTINUED

Recap

→ What are ECCs?

→ Rate of ECC

→ Minimum Distance

(Examples: Parity Code
Repetition Code)

→ Linear ECC & its properties

Error detection easy → due to parity check matrix

Code $C = \text{rowspan}(G)$

$$\text{Enc}(x) = xG$$

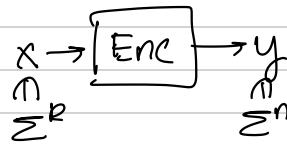
Generator
Matrix

Dual Code $C^\perp = \text{Nullspace}(G)$

$$z \in C \Leftrightarrow Hz = 0$$

$$\text{Enc}^\perp(x) = xH$$

Parity Check
Matrix



up to t errors
 $y \rightsquigarrow y'$
 Goal → to recover y
 & $\therefore x$.

Today

Hamming Code

1950

Reed Solomon
codes

1960

Hadamard
codes

1970

Applications

- Bar codes, QR codes
- CDs, DVDs
- Satellite comm.
- Broadcaster services etc...

Application

used in 1971 to
transmit photos of Mars
back to Earth by NASA
space probe Mariner 9

- Optimal tradeoff between rate & min-distance.

Linear Code Facts

- ① Min-distance $d(C) =$ minimum Hamming weight of a non-zero codeword. → # non-zero entries.

Pf:

let $y \neq y', y, y' \in C$
 then $\Delta(y, y') = \text{HammingWeight}(y - y')$ (by linearity)

hamming weight is distance from 0 notation for Hamming weight.

$$\therefore d(C) = \min_{y \neq y' \in C} \{\Delta(y, y')\} = \min_{y \neq 0 \in C} \{wt(y)\}$$

- ② $d(C) =$ min # cols of H which are linearly dependent.

Pf:

$$d(C) = \min \{wt(y) : y \in C, y \neq 0\}$$

$$\text{but } y \in C \Leftrightarrow Hy = 0$$

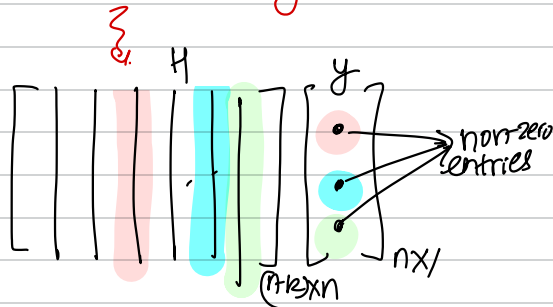
$$\therefore d(C) = \min \{wt(y) : Hy = 0, y \neq 0\}$$

So a min wt y means you find the sparsest linear combination of cols of H adding to 0.

this denotes linear combination of columns corresponding to non-zero entries in y .

i.e. minimum d such that every $d-1$ column set of H is linearly independent.

$$\sum_{j=1}^n y_j H_j = 0 \quad y_j \neq 0$$



I

HAMMING CODE (q=2)

Start with goal: High-rate linear code (i.e. lots of strings in codeword space) with some "decent" min-distance.

→ Goal is high-rate $\Rightarrow$ large k
 $\rightarrow$ smaller $n-k$
(shorter matrix H)

→ For min-distance $\rightarrow$ we want higher min-distance
 $\Rightarrow$ By property ② above, want the min number of linearly dependent columns in H to be large

We want min-distance ≥ 3 (so that you can at least correct 1 error)

for this we want to maximize rate.

①

$\Leftrightarrow$ How can you guarantee that min # linearly dependent columns in H is at least 3?

(Recall: we are in $\mathbb{F}_2$)

Ans: Make sure H has at least 1 non-zero column and doesn't have two identical columns.

You don't want a linearly dependent set of size 2

$\Rightarrow$ make sure you don't have 2 identical columns ($\because$ in $\mathbb{F}_2$)

$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ is linearly dependent by itself.

if you have 1 non-zero column & ≥ 2 linearly dependent columns.

②

High rate $\Rightarrow k$ larger $\Rightarrow$ want more codewords.
 $\rightarrow$ short & wide matrix

So, what we do to get Hamming code's parity check matrix H is take short matrix and fill it with distinct non-zero columns.

Defn Hamming Code *Nice video by 3Blue1Brown)

Defined by having parity check matrix

$$H = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 1 \end{bmatrix} \begin{matrix} \uparrow \\ r = n-k \\ \downarrow \end{matrix}$$

$\xleftarrow{\quad n = 2^r - 1 \quad} \xrightarrow{\quad}$

Columns are all possible non-zero strings of some fixed length r ordered lexicographically.

∴ It is a $\left[\underset{n}{2^r - 1}, \underset{k \approx n - \log_2(n+1)}{2^r - 1 - r}, 3 \right]_2$ -code.

$$\underline{\text{Rate}} = \frac{2^r - 1 - r}{2^r - 1} \approx 1 - \frac{\log_2(n+1)}{n} \quad (\text{very good})$$

Min-distance = 3 $\Rightarrow$ can correct up to 1 error

Nice property of Hamming code:

If y is a codeword, and it is corrupted at 1 position, say the i -th position.

Then, the received string is $z = y + e_i$

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \begin{matrix} \text{ith} \\ \text{pos} \end{matrix}$$

Now, $H z = H(y + e_i) = H e_i \quad (\because H y = 0)$

$\hookrightarrow$ this is the i th column of H . (which is i written in binary)

So HZ tells us the corrupted position i .
 $\Rightarrow$ can just decode by flipping the i th bit.

II HADAMARD CODE ($q=2$)

want to now flip the goal & get best min-distance.

Take dual of Hamming code, i.e. it is generator matrix for this code.

$$\text{Enc}(x) = \left[\underbrace{\quad x \quad}_{r} \right] \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \dots & 1 \end{bmatrix}_{r \times 2^r}$$

equivalently

add 0-column.

What is this encoding really doing?

Defn

$\text{Enc}: \mathbb{F}_2^r \rightarrow \mathbb{F}_2^{2^r}$ is defined as
 (very bad rate)

$$\text{Enc}(x) := (x \cdot a)_{a \in \mathbb{F}_2^r}$$

equivalently

Defn: For $x \in \mathbb{F}_2^r$, define $L_x: \mathbb{F}_2^r \rightarrow \mathbb{F}_2$ by
 $L_x(a) := x_1 a_1 + \dots + x_r a_r \rightarrow$ linear r -variate poly over $\mathbb{F}_2$.

Now $\text{Enc}(x) :=$ "truth table of L_x "

Remark (Why named Hadamard code?): replace $\mathbb{F}_2$ with $\{+1, -1\}$, and write each of the 2^r codewords in a row (each of length 2^r) $\Rightarrow$ you get H_{2^r} , the Hadamard matrix.

Proposition: Hadamard is a $[2^r, r, d]_2$ -code for $d = 2^{r-1}$ (n, k, min dist)
 i.e. it is an $[n, \log_2 n, \frac{n}{2}]_2$ -code.

Proof: (Using Schwartz-Zippel with $d=1$)

$L_x(a) := x_1 a_1 + \dots + x_r a_r$
 non-zero polynomial of degree 1 over $\mathbb{F}_2$

$$\Pr_{a_1, a_2, \dots, a_r \in \mathbb{F}_2} [L_x(a_1, \dots, a_r) = 0] \leq \frac{1}{2}$$

∴ for 2 different messages x & x' ($x \neq x'$)
 $L_x - L_{x'}$ is a non-zero polynomial of degree 1

$$\Rightarrow \Pr_{a_1, \dots, a_r \in \mathbb{F}_2} [(L_x - L_{x'})(a_1, \dots, a_r) \neq 0] > \frac{1}{2}$$

$\Rightarrow \text{Enc}(x)$ & $\text{Enc}(x')$ differ in at least $\frac{2^r}{2}$ locations

$$\circ \circ d = 2^{r-1}$$

Remark: Hadamard code has great min distance.
 Is it the optimal min distance? $\rightarrow$ It is known for binary code if $d \geq (\frac{1}{2} + \epsilon)n$ then it will only have $O(1)$ codewords ($\epsilon \rightarrow 0$), i.e. really bad rate.
 But, Hadamard's rate is bad ($\frac{\log_2 n}{n}$).

III

REED-SOLOMON CODES

generalization of Hadamard codes to univariate higher deg polynomials

Goal: I want great rate & great min-distance
(willing to increase alphabet size q)

For $k < n$, $q \geq n$ & $S \subseteq \mathbb{F}_q$ with $|S| = n$

$$\text{Enc}: \mathbb{F}_q^k \rightarrow \mathbb{F}_q^n$$

large alphabet size

$$\text{Enc}(m = (m_0, m_1, \dots, m_{k-1})) := (P_m(a))_{a \in S}$$

think of this like
coefficient representation
of a $(k-1)$ -degree
polynomial, $P_m(x)$

$$P_m(x) = m_0 + m_1x + \dots + m_{k-1}x^{k-1}$$

$$x \in \mathbb{F}_q[x]$$

all evaluations of
the polynomial P_m
on n points.

$$\left(\begin{array}{l} (k-1)\text{-degree univariate poly} \\ \text{in coefficient representation} \end{array} \right) \xrightarrow{\text{Enc}} \left(\begin{array}{l} n > k \text{ evaluations} \\ \text{of the polynomial} \end{array} \right)$$

Remark

① usually $q = n$ & $S = \mathbb{F}_q$

② Question: Is this a linear code? why?

Yes, in fact for $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$

$$\text{Enc}(m) = (m_0 \ m_1 \ \dots \ m_{k-1}) \begin{pmatrix} 1 & 1 & \dots & 1 \\ \alpha_1 & \alpha_2 & & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & & \alpha_n^{k-1} \end{pmatrix}$$

Vandermonde

Proposition: Distance of the Reed Solomon code is at least $n - (k-1)$.

Pf: Exercise (use fundamental thm of Algebra)

• RS is $[[n, k, n-k+1]]_q$ -code.

In fact, assuming $q \geq n$, the rate-distance trade off is optimal, by the following bound:

Singleton Bound: For $[[n, k, d]]_q$ -code, $k \leq n - d + 1$
(i.e. $d \leq n - k + 1$)

notation: not-necessarily linear code

Proof: Exercise (Hint: use Pigeonhole principle)

→ Example: For $k = \frac{n}{2}$, $\text{rate} = \frac{1}{2}$, $\text{min-distance} = \frac{n+1}{2}$
(But you need $q \geq n$)

Can we get best of all worlds?

Fix q , the alphabet size you want.

Let $\mathcal{C} = (\mathcal{C}_n)$ be a sequence of $[[n, k(n), d(n)]]_q$ codes.

Eg: Hamming: $[[n, n - \log_2(n+1), 3]]_2$
Hadamard: $[[n, \log_2 n, n/2]]_2$

Reed Solomon: $[n, k, n-k+1]_q$
 not fixed $q \leq n$.

Asymptotic Rate: $R(C) = \lim_{n \rightarrow \infty} \frac{k(n)}{n}$

Asymptotic Relative Distance : $g(C) = \lim_{n \rightarrow \infty} \frac{d(n)}{n}$

Eg: $R(\text{Hamming}) = 1$ $S(\text{Hamming}) = 0$

$$R(\text{Hadamard}) = 0 \quad \mathcal{S}(\text{Hadamard}) = 1/2$$

$$R(\text{Reed Solomon}) = R_0 \quad f(\text{Reed Solomon}) = 1 - R_0$$

(for any R_0)
(But q not fixed, for $q \geq n$ it's optimal)

Definition Dream Goal: Asymptotically good code family

$$C: \text{fixed } q, \quad R(C) \geq R_0 > 0, \quad \delta(C) \geq \delta_0 > 0$$

Theorem (Feasibility) $\rightarrow$ Do good codes even exist?

[Gilbert-Varshamov Bound]

$$\forall q, \forall 0 \leq \varepsilon_0 \leq 1 - \frac{1}{q}, \exists C \text{ with } \mathcal{S}(C) = \varepsilon_0$$

$$\star R(\mathcal{C}) = 1 - h_q(\mathcal{E}_0)$$

↳ "q-ary entropy function"
q-ary strings in a radius- ϵn Hamming ball is $\Theta(q^{h_q(\frac{\epsilon}{2})n})$.

(It is a greedy construction)

Also, random linear codes attain this GV Bound!

What about efficiently encodable & decodable asymptotically optimal codes? $\rightarrow$ They exist!

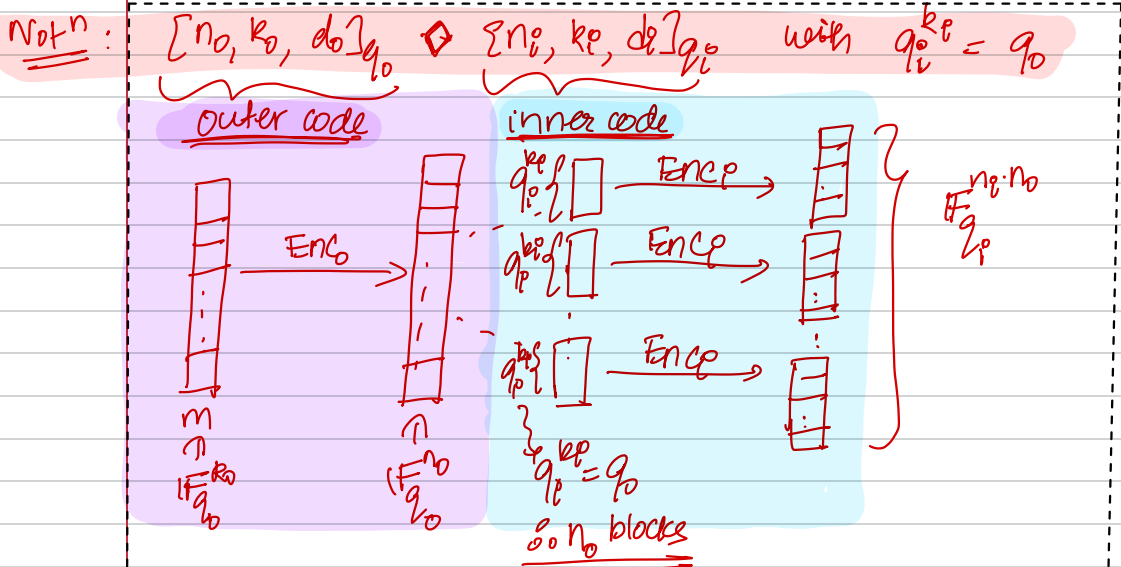
Theorem [Justesen, 1972]

For $q=2$, $\forall$ poly(n)-time encodable/decodable asymptotically good linear codes:

$$\forall 0 < R_0 < 1 \text{ \& } \delta_0 \approx 0.11(1-R_0)$$

* This theorem use concatenation codes. (defined below)

Concatenation codes



This is a $[n_0 \cdot n_i, k_0 \cdot k_i, d_0 \cdot d_i]_{q_0}$ -code

→ Show this → Exercise.

Example [Forney, 1966]: Reed Solomon $\diamond$ Hadamard

$([q, k, q-k+1]_q \diamond [q, \log_2 q, \frac{q}{2}]_2)$

for $q=2^r$

This is a $[q^2, R \log_2 q, q(q-k+1)/2]_2$ -code

If $R = \epsilon \cdot q$ ($0 < \epsilon < 1$), then $R = O(\frac{\log_2 q}{q})$, $\delta \approx \frac{1}{2}$

(Even $R(\text{ReedSolomon} \diamond \text{Hadamard}) = 0$, it's better than Hadamard)

Remark: Justesen's Code is a concatenation code with Reed Solomon as the outer code & "Wozencraft Ensemble".

→ CHALK AND TALK TOPIC #6 (Group of 2)

===== We continue after Midsem =====

I will announce more Chalk & Talk topics on moodle, sign up for it → Presentation will be before End-sems in April.

GOOD LUCK FOR THE MID-SEMS

