March 24, 2025 Error-Correcting Codes

Want to transmit or store bits. These can be corrupted (today flipped). How communicate reliably?

Idea 1: parity bit. To send $x_1, \dots, x_m$, append $x_{m+1} = x_1 \oplus \cdots \oplus x_m$ (mod 2)

If no errors, $x_1 \oplus \cdots \oplus x_m \oplus x_{m+1} = 0$.

one error, $= 1$ so know something went wrong

two errors ?

$\mathbb{F}_2 = \{0,1\}$, write $+$ for $+$ mod 2.

To locate and correct error, need more parity bits. Send $x_{i,j}$ $1 \le i,j \le k$

$$\begin{array}{cccc}
x_{1,1} & & x_{1,k} & x_{1,k+1} \\
 & & & \\
x_{k,1} & & x_{k,k} & x_{k,k+1} \\
x_{k+1,1} & & x_{k+1,k} & x_{k+1,k+1} \leftarrow \text{parity} \\
 & & & \uparrow \\
 & & & \text{parity}
\end{array}$$

set parity bits so each row & col sums to zero

If 1 error, will get sum 1 in that row and column

Ex $k=2$

$$\begin{array}{cccc}
1 & 0 & 1 & \\
0 & \cancel{0} & 0 & 1 \\
1 & 0 & 1 & \\
 & 1 & &
\end{array}$$

4 message bits
5 parity bits
1 error correction

Goal: be as reliable and efficient as possible.

Hamming Code 4 message bits 3 parity, 1 error. Message bits x_3, x_5, x_6, x_7

Parity bits x_1, x_2, x_4 chosen so

$$x_1 \oplus x_3 \oplus x_5 \oplus x_7 = 0$$
$$x_2 \oplus x_3 \oplus x_6 \oplus x_7 = 0$$
$$x_4 + x_5 + x_6 + x_7 = 0$$

in matrix form $\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $\quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$

A bit flip gives sum = column of flip
all cols different → can find it.
In fact, is binary rep of index

Asymptotics: Code is a map $C: \{0,1\}^m \to \{0,1\}^n$ $\quad n > m$ $\quad C(b)$ is a codeword

Rate $r = \frac{m}{n}$. Hamming dist $(C_1, C_2) = |C_1 - C_2|$, where $|x| = \#1\text{s in } x$

Minimum distance $= \min\limits_{C_1 \neq C_2} \text{dist}(C_1, C_2) \triangleq d$. Can correct up to $d/2$ errors.

If $\text{dist}(C_1, r) < \frac{d}{2}$, then for all $C_2 \neq C_1$, $\text{dist}(C_2, r) > \frac{d}{2}$.

<u>proof</u> $d \leq \text{dist}(C_1, C_2) \leq \text{dist}(C_1, r) + \text{dist}(r, C_2) < \frac{d}{2} + \text{dist}(r, C_2) \to \text{dist}(r, C_2) > d/2$

Minimum relative distance $\delta = d/n$

Asymptotically good codes: δ, r fixed, n grows — so a family of codes.

Can use Random Linear Codes. Linear code: $C(b) = Gb$ for some $n \times m$ matrix $G \in \mathbb{F}_2^{n \times m}$

Set of codewords is a vector space over $\mathbb{F}_2^n$. $G(b_1 + b_2) = Gb_1 + Gb_2 = G(b_1 - b_2)$ (mod 2)

min-dist $= \min_{b \neq 0} |G(b)|$. $C_G = \{C(b) : b \in \mathbb{F}_2^m\}$

$Gb = \begin{matrix} g_1 b \\ \vdots \\ g_n b \end{matrix}$ where $G = \begin{pmatrix} -g_1- \\ \vdots \\ -g_n- \end{pmatrix}$ If $b \neq 0$ and g_i is iid random, $g_i b$ is uniform in $\mathbb{F}_2$

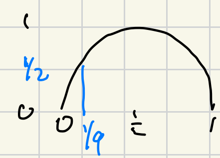
If entries of G iid, Gb is uniform in $\mathbb{F}_2^n$

Lem 1 $\Pr_G[C_G \text{ has min dist} > d] \geq 1 - \frac{2^m}{2^n} \sum_{i=0}^{d} \binom{n}{i}$

Could use Hoeffding

proof $\Pr_G[\text{min dist of } C_G \leq d] \leq \sum_{b \in \mathbb{F}_2^m - 0} \Pr[|C_G(b)| \leq d] = (2^m - 1)\left(\sum_{i=0}^{d} \binom{n}{i}\right) 2^{-n} \leq \frac{2^m}{2^n} \sum_{i=0}^{d} \binom{n}{i}$

Fact: $\sum_{i=0}^{d} \binom{n}{i} \leq 2^{nH(d/n)}$ where $H(x) = \frac{1}{x}\lg_2 x + \frac{1}{1-x} \lg_2(1-x)$



→ Thm 1 For δ, r s.t. $H(\delta) < 1 - r$, a random lin code of rate r prob has min rel dist $\geq \delta$

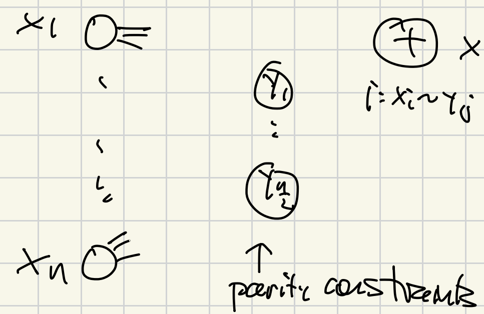
Holds when $\frac{2^m}{2^n} 2^{nH(\delta)}$ is small $\Leftrightarrow$ $m - n + H(\delta) < 0$ $m = rn$, so $H(\delta) + r - 1 < 0$

So, random codes probably have good minimum distance. Decoding? seems hard.

Gallager. Low-Density Parity-Check codes. Random bipartite graph. n nodes on left $\frac{n}{2}$ on right

Say, 3-regular on left, 6-regular on right

Related to expanders, next lecture.



$$\bigoplus_{i : x_i \sim y_j} x_i = 0$$

Generalized hypercube. Given $G = \begin{matrix} -g_1- \\ \vdots \\ -g_n- \end{matrix}$ $\quad g_i \in \mathbb{F}_2^m \quad n > m$

A graph with vertex set $\mathbb{F}_2^m$. $(a, b) \in E$ if $b = a + g_i$ some i

<u>Thm</u> Eigvals of Laplacian are $2|Gx|$ for $x \in \mathbb{F}_2^m$.

<u>proof</u> For $x \in \mathbb{F}_2^m$, let $\psi_x(a) = (-1)^{x^T a}$. To see is an eigvec, note $\psi_x(a+g) = (-1)^{x^T a + x^T g} = \psi_x(a)\psi_x(g)$

$$(L\psi_x)(a) = \sum_i \psi_x(a) - \psi_x(a+g_i) = n\psi_x(a) - \sum_i \psi_x(a)\psi_x(g_i) = \psi_x(a)\left(n - \sum_i \psi_x(g_i)\right)$$

So, it is an eigenvector. And, $\sum_i \psi_x(g_i) = \sum_i \begin{cases} 1 & \text{if } x^T g_i = 0 \\ -1 & \text{if } x^T g_i = 1 \end{cases} = n - 2|Gx|$

so, eigval is $2|Gx|$

If min rel dist $\geq \delta$, all eigvals $\geq 2\delta n$. So, is an $(1-2\delta)$-expander.

If max rel-dist $\leq 1-\delta$, all eigvals $\leq 2(1-\delta)n$

$\delta \to 0$ as $r \to 0$. Only defect is not constant degree.