An introduction to error correcting codes

Mathematics sin fronteras

María Chara

Universidad Nacional del Litoral

May 26, 2021

FIQ UNL. FACULTAD DE INGENIERÍA QUÍMICA

Mariner 9, 1971

Error-Correction Codes achievements [edit]

To control for errors in the reception of the grayscale image data sent by Mariner 9 (caused by a low signal-to-noise ratio), the data had to be encoded before transmission using a so-called forward error-correcting code (FEC). Without FEC, noise would have made up roughly a quarter of a received image, while the FEC encoded the data in a redundant way which allowed for the reconstruction of most of the sent image data at reception.

Since the flown hardware was constrained with regards to weight, power consumption, storage, and computing power, some considerations had to be put into choosing an FEC, and it was decided to use a Hadamard code for Mariner 9. Each image pixel was represented as a six-bit binary value, which had 64 possible grayscale levels. Because of limitations of the transmitter, the maximum useful data length was about 30 bits. Instead of using a repetition code, a [32, 6, 16] Hadamard code was used, which is also a 1st-order Reed-Muller code. Errors of up to seven bits per each 32-bit word could be corrected using this scheme.[10][11] Compared to a five-repetition code, the error correcting properties of this Hadamard code were much better, yet its data rate was comparable. The efficient decoding algorithm was an important factor in the decision to use this code. The circuitry used was called the "Green Machine", which employed the fast Fourier transform, increasing the decoding speed by a factor of three.[12]

Figure: https://en.wikipedia.org/wiki/Mariner_9

Parity check matrix and the dual code

- For an $[n, k]$ code $\mathcal{C}$ over $\mathbb{F}_q$ and generator matrix G:

$$\begin{array}{ccc} \mathbb{F}_q^k & \longrightarrow & \mathbb{F}_q^n \\ u & \rightarrow & c = uG \end{array}$$

Parity check matrix and the dual code

- For an $[n, k]$ code $\mathcal{C}$ over $\mathbb{F}_q$ and generator matrix G:

$$\begin{array}{ccc} \mathbb{F}_q^k & \longrightarrow & \mathbb{F}_q^n \\ u & \rightarrow & c = uG \end{array}$$

- A matrix H of size $(n-k) \times n$ which is a generator matrix for $\mathcal{C}^\perp$ is called parity check matrix for C.

Parity check matrix and the dual code

- For an $[n, k]$ code $\mathcal{C}$ over $\mathbb{F}_q$ and generator matrix G:

$$\begin{array}{ccc} \mathbb{F}_q^k & \longrightarrow & \mathbb{F}_q^n \\ u & \rightarrow & c = uG \end{array}$$

- A matrix H of size $(n-k) \times n$ which is a generator matrix for $\mathcal{C}^{\perp}$ is called parity check matrix for C.

- H 'checks' whether a vector $y \in \mathbb{F}_q^n$ is a codeword or not:

$$y \in \mathcal{C} \qquad \Longleftrightarrow \qquad yH^T = 0 \qquad \Longleftrightarrow \qquad S(y) = 0$$

Parity check matrix and the dual code

- For an $[n, k]$ code $\mathcal{C}$ over $\mathbb{F}_q$ and generator matrix G:

$$\begin{array}{ccc} \mathbb{F}_q^k & \longrightarrow & \mathbb{F}_q^n \\ u & \rightarrow & c = uG \end{array}$$

- A matrix H of size $(n-k) \times n$ which is a generator matrix for $\mathcal{C}^{\perp}$ is called parity check matrix for C.

- H 'checks' whether a vector $y \in \mathbb{F}_q^n$ is a codeword or not:

$$y \in \mathcal{C} \qquad \Longleftrightarrow \qquad yH^T = 0 \qquad \Longleftrightarrow \qquad S(y) = 0$$

- If $y \notin \mathcal{C}$ then y is corrected as $y - e$ where $S(y) = S(e)$.

Parity check matrix and the dual code

- For an $[n, k]$ code $\mathcal{C}$ over $\mathbb{F}_q$ and generator matrix G:

$$\begin{array}{ccc} \mathbb{F}_q^k & \longrightarrow & \mathbb{F}_q^n \\ u & \rightarrow & c = uG \end{array}$$

- A matrix H of size $(n-k) \times n$ which is a generator matrix for $\mathcal{C}^\perp$ is called parity check matrix for C.

- H 'checks' whether a vector $y \in \mathbb{F}_q^n$ is a codeword or not:

$$y \in \mathcal{C} \iff yH^T = 0 \iff S(y) = 0$$

- If $y \notin \mathcal{C}$ then y is corrected as $y - e$ where $S(y) = S(e)$.

- To perform error correction we need a table with all the coset leaders and their syndromes.

Example

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \qquad H = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathcal{C} = \{(000000), (100011), (010101), (001110), \\ (110110), (011011), (101101), (111000)\}.$$

Example

$$\mathcal{C} = \{(000000), (100011), (010101), (001110), (110110), (011011), (101101), (111000)\}.$$

$e_0 = (000000)\ S(e_0) = (000)$	$e_4 = (000100)\ S(e_4) = (100)$
$e_1 = (100000)\ S(e_1) = (011)$	$e_5 = (000010)\ S(e_5) = (010)$
$e_2 = (010000)\ S(e_2) = (101)$	$e_6 = (000001)\ S(e_6) = (001)$
$e_3 = (001000)\ S(e_3) = (110)$	$e_7 = (100100)\ S(e_7) = (111)$

$$H^T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- $x = (010101) \rightarrow y = (010001)$
- $x = (010101) \rightarrow y = (011101)$
- $x = (011011) \rightarrow y = (011111)$
- $x = (110110) \rightarrow y = (110010)$
- 2 errors: $x = (110110) \rightarrow y = (110101)$

Hadamard Code

Hadamard Code

A Hadamard code is an $[2^r, r, 2^{r-1}]$ linear code over a binary alphabet which can correct many errors. The generator matrix G can be constructed in the following way: the i^{th} column of G is the binary representation of i for $i = 0, \ldots, 2^r - 1$.

Hadamard Code

A Hadamard code is an $[2^r, r, 2^{r-1}]$ linear code over a binary alphabet which can correct many errors. The generator matrix G can be constructed in the following way: the i^{th} column of G is the binary representation of i for $i = 0, \ldots, 2^r - 1$.

$$23 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

Hadamard Code

A Hadamard code is an $[2^r, r, 2^{r-1}]$ linear code over a binary alphabet which can correct many errors. The generator matrix G can be constructed in the following way: the i^{th} column of G is the binary representation of i for $i = 0, \ldots, 2^r - 1$.

$$23 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

- Example: the $[8, 3, 4]$ Hadamard code has generator matrix:

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

Coordinate recovery

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ g_0 & g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{pmatrix}$$

$\vec{x}$ $\quad$ $\vec{c} = \vec{x} \cdot G$ $\quad$ $g_7 = g_1 + g_6$

$000 \rightarrow (00000000)$
$001 \rightarrow (01010101)$
$010 \rightarrow (00110011)$
$100 \rightarrow (00001111)$
$101 \rightarrow (01011010)$
$011 \rightarrow (01100110)$
$110 \rightarrow (00111100)$
$111 \rightarrow (01101001)$

Punctured Hadamard code

Punctured Hadamard code

The punctured Hadamard code is a $[2^{r-1}, r, 2^{r-2}]$ linear code over the binary alphabet. The generator matrix for this code is obtained from the generator matrix of a Hadamard code without the columns starting with 0.

Punctured Hadamard code

The punctured Hadamard code is a $[2^{r-1}, r, 2^{r-2}]$ linear code over the binary alphabet. The generator matrix for this code is obtained from the generator matrix of a Hadamard code without the columns starting with 0.

- Example: $r = 3$:

$$G_{[8,3,4]} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad \rightarrow \quad G_{[4,3,2]} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Punctured Hadamard code

The punctured Hadamard code is a $[2^{r-1}, r, 2^{r-2}]$ linear code over the binary alphabet. The generator matrix for this code is obtained from the generator matrix of a Hadamard code without the columns starting with 0.

- Example: $r = 3$:

$$G_{[8,3,4]} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad \rightarrow \quad G_{[4,3,2]} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$\mathcal{C} = \{(0000), (0101), (0011), (1111), (0110), (1010), (1100), (1001)\}$$

Código de de Reed-Muller de primer orden

Reed-Muller codes are among the oldest known codes (1954).

Definition (Recursiva)

The (first order) Reed-Muller codes $\mathcal{R}(1, m)$ are binary codes recursively defined, for all $m \geq 1$, by:

- $\mathcal{R}(1, 1) = \{00, 01, 10, 11\} = \mathbb{Z}_2^2$;
- para $m \geq 1$,

$$\mathcal{R}(1, m) = \{(u, u), (u, u + \mathbf{1}) : u \in \mathcal{R}(1, m - 1) \text{ and } \mathbf{1} = \text{ all 1 vector}\}$$

Proposition

For $m \geq 1$ the Reed-Muller code $\mathcal{R}(1, m)$ is an $[2^m, m+1, 2^{m-1}]$ binary linear code in which every codeword, except all 0 and all 1, has weight 2^{m-1}.

We have that:

- Since the minimum distance is $d = 2^{m-1}$, the code can correct up to $2^{m-2} - 1$ errors.
- The all zero word and the all one word (of weight 2^m) are in the code.
- All other codewords have half of its coordinates 0 and half 1.

Error correction

- Convert all the codewords (and the received vector) to ± 1 vectors by turning the 0's into -1's.
- Take the dot product of the received vector with each of the codewords in turn.
- As soon as the result is $2^{m-1} + 2$ or greater, decode as that codeword.

Error correction

$$\mathcal{R}(1,3) = \{(00000000), (00001111), (01010101), (01011010),$$
$$(10101010), (10100101), (11111111), (11110000)$$
$$(00110011), (00111100), (01100110), (01101001)$$
$$(10011001), (10010110), (11001100), (11000011)\}$$

$$d(c_i, c_j) = w(c_i - c_j) = \begin{cases} 0 & \text{if } i = j, \\ 8 & \text{if all coordinates are different,} \\ 4 & \text{in other case.} \end{cases}$$

$$c_i \cdot c_j = \begin{cases} 8 & \text{if } i = j, \\ -8 & \text{f all coordinates are different,} \\ 0 & \text{in other case.} \end{cases}$$

$$c \quad \rightarrow \quad y = c + e \quad \rightarrow \quad y \cdot c_i = \begin{cases} \geq 6 & \text{if } c_i = c, \\ \leq 2 & \text{in other case.} \end{cases}$$

Reed-Muller code

- For $0 \leq r \leq m$ the Reed-Muller denoted by $\mathcal{R}(r, m)$ is an $[n, k, d]$ binary code with

$$n = 2^m, \qquad k = \sum_{i=0}^{r} \binom{m}{i}, \qquad d = 2^{m-r}.$$

- The generator matrix for $\mathcal{R}(r, m)$ is

$$G_{r,m} = \begin{pmatrix} G_{r,m-1} & G_{r,m-1} \\ 0 & G_{r-1,m-1} \end{pmatrix} \qquad \text{si } 0 < r < m$$

$$G_{0,m} = \underbrace{(1\ 1\ \cdots 1)}_{m+1} \qquad G_{m,m} = \begin{pmatrix} G_{m-1,m} \\ 0 \cdots 0\ 1 \end{pmatrix}$$

Challenge

Write the generator matrix of the

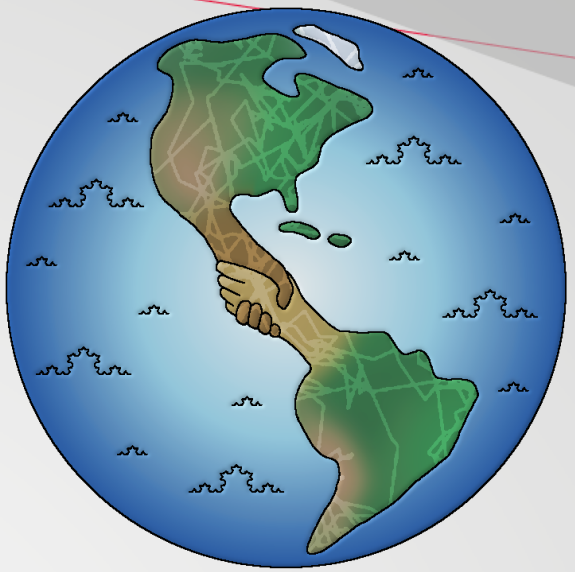
$$[32, 6, 16]$$

code used by Mariner 9.

Questions?

María Chara
charamaria@gmail.com

Saraí Hernández-Torres
sarai.h@campus.technion.ac.il



FIQ

UNL . FACULTAD DE
INGENIERÍA QUÍMICA