Chapter 2: Error Detection and Correction Codes

Lecture 6

Binary Hamming error-correcting codes

Let us construct a code $\mathcal{C}$ that

- is binary with code alphabet $\{0, 1\}$
- has fixed length n codewords $\mathbf{x} = x_1 \cdots x_n$
- is single-error correcting
- provides user-friendly error-correcting
- uses m independent linear parity checks

$$\sum_{j=1}^{n} a_{ij} x_j \equiv 0 \pmod 2 \quad \text{where} \quad i = 1, \ldots, m \quad \text{and} \quad a_{ij} \in \{0, 1\}$$

We need to choose

- $\mathcal{C}$
- an encoding scheme
- a correcting scheme
- a decoding scheme

The m parity checks

$$\sum_{j=1}^{n} a_{ij}x_j \equiv 0 \pmod 2 \quad \text{where} \quad i = 1, \ldots, m \quad \text{and} \quad a_{ij} \in \{0, 1\}$$

can be expressed as $H\mathbf{x}^T = \mathbf{0}$ (in $\mathbb{Z}_2$)
where H is the $m \times n$ parity check matrix with entries a_{ij}.

Let $\mathcal{C}$ be the null space of H:

$$\mathbf{x} \in \mathcal{C} \quad \text{if and only if} \quad H\mathbf{x}^T = \mathbf{0}$$

Let $\mathcal{C} = \{\mathbf{x} \in \mathbb{Z}_2^n \ : \ H\mathbf{x}^T = \mathbf{0}\}$ be the null space of H.

Define the syndrome $S(\mathbf{y}) = H\mathbf{y}^T$.

- $S(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x} \in \mathcal{C}$
- $S(\mathbf{y})$ tells us when $\mathbf{y}$ has an error
- In fact, we can get $S(\mathbf{y})$ to tell us where $\mathbf{y}$ has an error!

Let $\mathbf{x}$ be a codeword of $\mathcal{C} = \{\mathbf{x} \in \mathbb{Z}_2^n \ : \ H\mathbf{x}^T = \mathbf{0}\}$.
Consider a word $\mathbf{y}$ with a single error ($\mathbf{x} \rightsquigarrow \mathbf{y}$), in position i.
Then $\mathbf{y}^T = \mathbf{x}^T + \mathbf{e}_i$, so

$$S(\mathbf{y}) = H\mathbf{y}^T = H(\mathbf{x}^T + \mathbf{e}_i) = H\mathbf{x}^T + H\mathbf{e}_i = \mathbf{0} + H\mathbf{e}_i = H\mathbf{e}_i$$

Now, $H\mathbf{e}_i$ is the ith column of H, so we can make error-correcting easy by defining the ith column of H to be the binary expression for i.
Then $S(\mathbf{y}) = H\mathbf{e}_i$ tells us the position i of the (single) error:

If $S(\mathbf{y})^T = 0000$, then there are no errors
If $S(\mathbf{y})^T = 0001$, then the error is in position 1
If $S(\mathbf{y})^T = 0010$, then the error is in position 2
If $S(\mathbf{y})^T = 0011$, then the error is in position 3
etc.

Codes defined in this way are called binary Hamming-type codes.

Example

The binary Hamming-type code for $n = 7$, $m = 3$ has parity check matrix

$$H = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

This is the parity check matrix for the binary Hamming (7,4) code.
The number 4 refers to $k = n - m = 7 - 3$.

Note that we need $2^m - 1 \geq n$, or $2^m \geq n + 1$.
If $2^m \geq n + 1$, then the code is the binary Hamming (n, k) code.

Error-correcting with binary Hamming-type codes

The binary Hamming-type code for $n = 5$, $m = 3$ has parity check matrix

$$H = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

The word $\mathbf{y} = 00111$ has a single error.
To find this error, we calculate the syndrome $S(\mathbf{y})$:

$$S(\mathbf{y}) = H\mathbf{y}^T = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

This corresponds to the binary number 010, namely 2.
We therefore correct bit number 2 in $\mathbf{y}$, and get the codeword 01111.

Encoding/decoding with binary Hamming-type codes

$$H \; = \; \begin{array}{c} \begin{array}{ccccc} \cancel{1} & \cancel{2} & 3 & \cancel{4} & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} \qquad \begin{array}{l} n = 5 \\ m = 3 \end{array}$$

We see that the $m = 3$ columns 1, 2, and 4 are leading
whereas columns 3 and 5 are non-leading.
Therefore when solving $H\mathbf{x}^T = \mathbf{0}$ for $\mathbf{x} = x_1 \cdots x_5$,
x_3 and x_5 are free parametric variables and together determine x_1, x_2, x_4.

We can use x_3 and x_5 as information bits and x_1, x_2, x_4 as check bits.

The binary Hamming-type codes are systematic, with

- $k = n - m = 2$ information bits,
- $m = 3$ check bits (in columns $1, 2, 4, \ldots, 2^{m-1}$ in general), and
- $2^k = 4$ codewords.

Encoding/decoding with binary Hamming-type codes

$$
H = \begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
\end{array}
\qquad n = 5 \qquad m = 3
$$

To encode a message $\mathbf{w} = w_1 \cdots w_k$ where $k = n - m$:

- Substitute $\mathbf{w}$ into the k parametric variables of $\mathbf{x}$
- Solve $H\mathbf{x}^T = \mathbf{0}$ to find the m check (leading) variables
- $\mathbf{x}$ is the resulting codeword.

To decode a codeword $\mathbf{x} = x_1 \cdots x_n$:

- Extract the sequence of parametric variable values.
- This gives the decoded message.

Example

$$H = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} \qquad \begin{array}{l} n = 5 \\ m = 3 \end{array}$$

To encode $\mathbf{w} = 01$, set $x_3 = 0$ and $x_5 = 1$ in $\mathbf{x} = x_1 \cdots x_5$.
Now solve $H\mathbf{x}^T = \mathbf{0}$:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \\ x_4 \\ 1 \end{pmatrix} = \mathbf{0} \quad \text{or} \quad \begin{array}{ccccccc} x_1 & & & + & 1 & = & 0 \\ & x_2 & & & & = & 0 \\ & & x_4 & + & 1 & = & 0 \end{array}$$

We see that $x_1 = x_4 = 1$ and $x_2 = 0$.
The encoded message is therefore $\mathbf{x} = 10011$.

To decode $\mathbf{x} = 10011$, just extract the non-leading entries: 01

Binary Hamming (n, k) codes

- Binary
- Block codes with codeword length n
- Systematic
- k information bits
- $m = n - k$ check bits (in positions $1, 2, 4, \ldots, 2^{m-1}$)
- 2^k codewords
- $2^m = n + 1$ (not true for all Binary Hamming-type codes)
- To encode, write message as information bits of $\mathbf{x}^T$ & solve $H\mathbf{x}^T = \mathbf{0}$
- To correct, calculate syndrome $S(\mathbf{x})$ to find error position
- To decode, extract message from information bits of $\mathbf{x}^T$

Binary Hamming (n, k) codes

- Parameters:

m	$k = 2^m - m - 1$	$n = k + m$	$R = \frac{k}{n}$
3	4	7	0.57
4	11	15	0.73
5	26	31	0.84
6	57	63	0.90
7	120	127	0.94
8	247	255	0.97
9	502	511	0.98
10	1013	1023	0.99

Exercise

For the Hamming (7,4) code,

(a) encode 1001

(b) correct and decode 0110001