

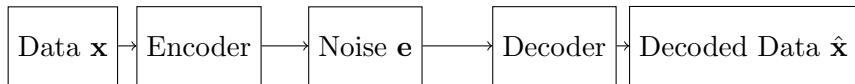
Coding Theory: Linear Codes and Distances

A Generic Communication Channel

Transmitter

Channel

Receiver



$$\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{F}_q^k \mapsto \mathbf{c} = \underbrace{(c_1, \dots, c_n)}_{\text{codeword, } n \geq k} = (x_1, \dots, x_k) \underbrace{[\mathbf{I}_k | A]}_{\text{systematic generator matrix}} \in \mathbb{F}_q^n$$

Generator matrix

$$\mathcal{C} = \{\mathbf{c} = \mathbf{x}G, \mathbf{x} \in \mathbb{F}_q^k\}$$

$$\mathcal{C}^\perp = \{\mathbf{c} = \mathbf{x}H, \mathbf{x} \in \mathbb{F}_q^{n-k}\}$$

Parity check matrix

$$\mathcal{C} = \{\mathbf{v} \in \mathbb{F}_q^n, H\mathbf{v}^T = \mathbf{0}\}$$

$$\mathcal{C}^\perp = \{\mathbf{v} \in \mathbb{F}_q^n, G\mathbf{v}^T = \mathbf{0}\}$$

Hamming distance

For two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$, their Hamming distance $d(\mathbf{x}, \mathbf{y})$ is the number of coordinates in which they differ.



(source: wikipedia)

Hamming Distance

■ Axioms of distance

$d(\mathbf{x}, \mathbf{y}) \geq 0$: d counts a number of coordinates, it varies between 0 and n for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$.

$d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$: $d(\mathbf{x}, \mathbf{y}) = 0$ means \mathbf{x} and \mathbf{y} differ in zero coordinate.

$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$: d counts the differences between \mathbf{x} and \mathbf{y} (which is the same as between \mathbf{y} and \mathbf{x}).

$d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$: suppose \mathbf{x} and \mathbf{z} differ in one coordinate; it cannot be that \mathbf{y} agrees with both \mathbf{x} and \mathbf{z} on this coordinate. If \mathbf{y} agrees with either \mathbf{x} or \mathbf{z} , this contributes to 1 to $d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$, if \mathbf{y} disagrees with both, it contributes to 2.

Weight of a vector

For $\mathbf{x} \in \mathbb{F}_q^n$, its weight $wt(\mathbf{x})$ counts the number of nonzero coordinates of \mathbf{x} .

- $wt((1, 0, 1, 0, 0)) = 2$
- $d(\mathbf{x}, \mathbf{y}) = wt(\mathbf{x} - \mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$. Indeed, the vector $\mathbf{x} - \mathbf{y}$ will have zero coordinates exactly where \mathbf{x} and \mathbf{y} agree on their coordinates.

Hamming distance of \mathcal{C}

For an (n, k) linear code \mathcal{C} over \mathbb{F}_q , its minimum distance $d_H(\mathcal{C})$ is the minimum weight of the nonzero codewords of \mathcal{C} .

- This works only for linear codes.
- If $d_H(\mathcal{C}) = d$, the notation (n, k, d) is sometimes used.

Hamming Distance

■ Examples

The $(n, 1)$ repetition code

- The dimension is $k = 1$ and the length is n .
- $(x_1) \mapsto (x_1, \dots, x_1)$.

Hamming Distance

■ Examples

The $(n, 1)$ repetition code

- The dimension is $k = 1$ and the length is n .
- $(x_1) \mapsto (x_1, \dots, x_1)$.
- Its minimum distance is $d_H(\mathcal{C}) = n$: indeed, there is only one codeword which is not zero, namely $(1, \dots, 1)$ whose weight is n .

Hamming Distance

■ Examples

The $(n, n - 1)$ single parity check code

- The dimension is $k = n - 1$ and the length is n .
- $(x_1, \dots, x_k) \mapsto (x_1, \dots, x_k, \sum_{i=1}^k x_i)$.

Hamming Distance

■ Examples

The $(n, n - 1)$ single parity check code

- The dimension is $k = n - 1$ and the length is n .
- $(x_1, \dots, x_k) \mapsto (x_1, \dots, x_k, \sum_{i=1}^k x_i)$.
- Its minimum distance is $d_H(\mathcal{C}) = 2$: indeed, the codeword $(1, 0, \dots, 0, 1)$ has weight 2. It is not possible to have a codeword with weight 1, because if there is a single data symbol which is not zero, then the parity symbol is also not zero. And if there are at least two data symbols, then the weight is at least 2.

Hamming Distance

■ Examples

The (7, 4) Hamming code

- It has dimension $k = 4$, length $n = 7$ and alphabet \mathbb{F}_2 .
- Encoding given by the generator matrix:

$$G = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array} \right],$$

so

$$\mathbf{x} \mapsto (x_1, x_2, x_3, x_4, x_2 + x_3 + x_4, x_1 + x_3 + x_4, x_1 + x_2 + x_4).$$

Hamming Distance

■ Examples

The (7, 4) Hamming code

- $\mathbf{x} \mapsto (x_1, x_2, x_3, x_4, x_2 + x_3 + x_4, x_1 + x_3 + x_4, x_1 + x_2 + x_4)$.
- If $x_1 = 1$ and $x_2 = x_3 = x_4 = 0$, then we get a weight of 3. If we have two data symbols that are not zero, say x_1 and x_2 , the only way to get a smaller weight would be that all parities are 0, which is not possible since x_1 and x_2 appear at different positions. Thus $d_H(\mathcal{C}) = 3$.

Hamming Distance

■ Examples

The $(4, 2)$ tetracode

- It has dimension $k = 2$, length $n = 4$ and alphabet \mathbb{F}_3 .
- Encoding given by the generator matrix:

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{bmatrix}$$

so a codeword is of the form $(x_1, x_2, x_1 + x_2, x_1 - x_2)$.

Hamming Distance

■ Examples

The $(4, 2)$ tetracode

- It has dimension $k = 2$, length $n = 4$ and alphabet \mathbb{F}_3 .
- Encoding given by the generator matrix:

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{bmatrix}$$

so a codeword is of the form $(x_1, x_2, x_1 + x_2, x_1 - x_2)$.

- If say $x_1 = 1$ and $x_2 = 0$, then we get a weight of 3. To get a smaller weight, we could try to have both x_1, x_2 non-zero, but then the parities should both be 0, which is not possible. So $d_H(\mathcal{C}) = 3$.

Hamming Distance

■ Examples

n	k	\mathbb{F}_q	name	d_H
n	1	\mathbb{F}_q	repetition	n
n	$n - 1$	\mathbb{F}_q	single parity check	2
7	4	\mathbb{F}_2	Hamming	3
4	2	\mathbb{F}_3	tetracode	3

Distance and Erasures

An (n, k) linear code \mathcal{C} over \mathbb{F}_q with minimum Hamming distance $d_H(\mathcal{C}) = d$ can recover from $d - 1$ erasures.

Any two codewords differ in d coordinates.

If up to $d - 1$ coordinates are missing, there is still at least one left to recognize the codeword.

Erasure Recovery

■ Examples

The $(4, 3, 2)$ single parity check code

Over \mathbb{F}_2 , codewords are

$$\begin{aligned} &(0, 0, 0, 0), (1, 0, 0, 1), (0, 1, 0, 1), (1, 1, 0, 0), \\ &(0, 0, 1, 1), (1, 0, 1, 0), (0, 1, 1, 0), (1, 1, 1, 1). \end{aligned}$$

If we receive $(0, *, 1, 0)$,

Erasure Recovery

■ Examples

The $(4, 3, 2)$ single parity check code

Over \mathbb{F}_2 , codewords are

$$\begin{aligned} &(0, 0, 0, 0), (1, 0, 0, 1), (0, 1, 0, 1), (1, 1, 0, 0), \\ &(0, 0, 1, 1), (1, 0, 1, 0), (0, 1, 1, 0), (1, 1, 1, 1). \end{aligned}$$

If we receive $(0, *, 1, 0)$, we must have sent $(0, 1, 1, 0)$.

If we receive $(0, *, 1, *)$,

Erasure Recovery

■ Examples

The $(4, 3, 2)$ single parity check code

Over \mathbb{F}_2 , codewords are

$$\begin{aligned} &(0, 0, 0, 0), (1, 0, 0, 1), (0, 1, 0, 1), (1, 1, 0, 0), \\ &(0, 0, 1, 1), (1, 0, 1, 0), (0, 1, 1, 0), (1, 1, 1, 1). \end{aligned}$$

If we receive $(0, *, 1, 0)$, we must have sent $(0, 1, 1, 0)$.

If we receive $(0, *, 1, *)$, we could have sent $(0, 1, 1, 0)$ or $(0, 0, 1, 1)$.

Erasure Recovery

■ Examples

The $(4, 2, 3)$ tetracode

Over \mathbb{F}_3 , codewords are of the form

$(x_1, x_2, x_1 + x_2, x_1 - x_2)$:

$$\begin{aligned} &(0, 0, 0, 0), (0, 1, 1, 2), (0, 2, 2, 1), (1, 0, 1, 1), \\ &(1, 1, 2, 0), (1, 2, 0, 2), (2, 0, 2, 2), (2, 1, 0, 1), (2, 2, 1, 0). \end{aligned}$$

If we receive $(0, *, *, 0)$,

Erasure Recovery

■ Examples

The $(4, 2, 3)$ tetracode

Over \mathbb{F}_3 , codewords are of the form

$(x_1, x_2, x_1 + x_2, x_1 - x_2)$:

$$(0, 0, 0, 0), (0, 1, 1, 2), (0, 2, 2, 1), (1, 0, 1, 1), \\ (1, 1, 2, 0), (1, 2, 0, 2), (2, 0, 2, 2), (2, 1, 0, 1), (2, 2, 1, 0).$$

If we receive $(0, *, *, 0)$, we must have sent $(0, 0, 0, 0)$.

If we receive $(*, *, 1, *)$,

Erasure Recovery

■ Examples

The $(4, 2, 3)$ tetracode

Over \mathbb{F}_3 , codewords are of the form

$(x_1, x_2, x_1 + x_2, x_1 - x_2)$:

$$(0, 0, 0, 0), (0, 1, 1, 2), (0, 2, 2, 1), (1, 0, 1, 1), \\ (1, 1, 2, 0), (1, 2, 0, 2), (2, 0, 2, 2), (2, 1, 0, 1), (2, 2, 1, 0).$$

If we receive $(0, *, *, 0)$, we must have sent $(0, 0, 0, 0)$.

If we receive $(*, *, 1, *)$, we could have sent $(0, 1, 1, 2)$ or $(1, 0, 1, 1)$ or $(2, 2, 1, 0)$.

Parity check matrix and Hamming distance

If $\mathbf{c} \in \mathcal{C}$, an (n, k) code, the columns of the parity check matrix H corresponding to the nonzero coordinates of \mathbf{c} are linearly dependent.

$$\mathcal{C} = \{\mathbf{v} \in \mathbb{F}_q^n, H\mathbf{v}^T = \mathbf{0}\}$$

If $\mathbf{c} \in \mathcal{C}$, $H\mathbf{c}^T = \mathbf{0}$.

If \mathbf{h}_i are columns of H , then

$$c_1\mathbf{h}_1 + c_2\mathbf{h}_2 + \dots + \mathbf{h}_nc_n = \mathbf{0}.$$

The nonzero c_i create a linear dependency among their columns.

Parity check matrix and Hamming distance

Conversely, if a linear dependency with nonzero coefficients exists among w columns of H , then there is a codeword of weight w whose nonzero coordinates correspond to these columns.

If \mathbf{h}_i are columns of H and

$$v_1\mathbf{h}_1 + v_2\mathbf{h}_2 + \dots + \mathbf{h}_n v_n = \mathbf{0}$$

for exactly w nonzero v_i .

Then for $\mathbf{v} = (v_1, \dots, v_n)$,
 $H\mathbf{v}^T = \mathbf{0}$.

Since

$\mathcal{C} = \{\mathbf{v} \in \mathbb{F}_q^n, H\mathbf{v}^T = \mathbf{0}\}$
we have that $\mathbf{v} \in \mathcal{C}$.

Parity check matrix and Hamming distance

A linear code \mathcal{C} has minimum Hamming distance d if and only if its parity check matrix H has a set of d linearly dependent columns but no set of $d - 1$ linearly dependent columns.

\mathcal{C} has minimum Hamming distance d if and only if there is a codeword \mathbf{c} of weight d but no codeword of weight less than d . Therefore there are linearly dependent columns of H corresponding to the nonzero coordinates of \mathbf{c} , and no $d - 1$ columns that are linearly dependent.

Hamming Distance

■ Examples

The $(n, 1)$ repetition code

$d_H(\mathcal{C}) = d$ if and only if H has a set of d linearly dependent columns but no set of $d - 1$ linearly dependent columns.

- parity check matrix:

$$\begin{bmatrix} -1 & & \\ \vdots & \mathbf{I}_{n-1} & \\ -1 & & \end{bmatrix}$$

- We have n linearly dependent columns since the first column is obtained as minus the sum of the others, but any set of $n - 1$ columns is linearly independent, thus $d_H(\mathcal{C}) = n$.

Hamming Distance

■ Examples

The (7, 4) Hamming code

$d_H(\mathcal{C}) = d$ if and only if H has a set of d linearly dependent columns but no set of $d - 1$ linearly dependent columns.



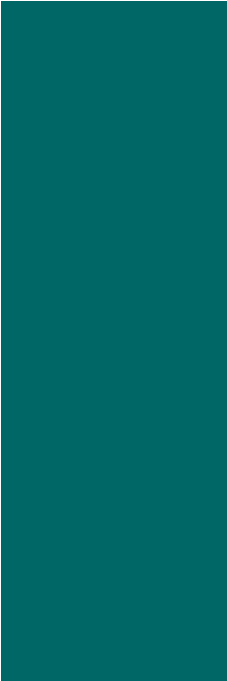
$$H = \left[\begin{array}{cccc|ccc} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array} \right]$$

- Columns 4,5,6,7 are linearly dependent, but so are columns 1,6,7. All two columns are linearly independent, so $d_H(\mathcal{C}) = 3$.

Hamming Distance

■ Examples

n	k	\mathbb{F}_q	name	d_H	recovers from
n	1	\mathbb{F}_q	repetition	n ✓	$n - 1$ erasures
n	$n - 1$	\mathbb{F}_q	single parity check	2	1 erasure
7	4	\mathbb{F}_2	Hamming	3 ✓	2 erasures
4	2	\mathbb{F}_3	tetracode	3	2 erasures



Weight of a vector

Hamming distance

Connection to erasure recovery

Two ways of computing the Hamming distance