CIE6020/MAT3350

Selected Topics in Information Theory

Lecture 16: Linear Codes

4 April 2019

The Chinese University of Hong Kong, Shenzhen

Linear Codes

Linear Codes

- Suppose that $\mathcal{A}$ is the input alphabet of a channel.
- A block *error correcting code* $\mathcal{C}$ is a subset of $\mathcal{A}^n$, where n is called the *block length*.
- Most practical channel codes are linear codes, where $\mathcal{A}$ is a finite field.
- A code $\mathcal{C} \subset \mathcal{A}^n$ is *linear* if it is closed under linear combinations, in other words,

$$\alpha \mathbf{x} + \alpha' \mathbf{x}' \in \mathcal{C}, \quad \forall \mathbf{x}, \mathbf{x}' \in \mathcal{C},\ \forall \alpha, \alpha' \in \mathcal{A}.$$

- A linear code $\mathcal{C}$ is a subspace of $\mathcal{A}^n$.
- A linear code with length n and dimension k is said to be an (n, k) code.

Generator Matrix

- For an (n, k) code $\mathcal{C}$, a $k \times n$ matrix G, whose rows form a basis of $\mathcal{C}$, is called a generator matrix for $\mathcal{C}$.
- $\mathcal{C} = \langle G \rangle = \{uG : u \in \mathcal{A}^k\}$.
- A generator matrix G of $\mathcal{C}$ is said to be *systematic* if $G = [I \ P]$, where I is a $k \times k$ identity matrix.

Dual Code and Parity-Check Matrix

- The *dual code* $\mathcal{C}^\perp$ of a linear code $\mathcal{C}$ is defined by

$$\mathcal{C}^\perp = \{\mathbf{v} \in \mathcal{A}^n : \mathbf{v} \cdot \mathbf{x}^\top = 0, \forall \mathbf{x} \in \mathcal{C}\} = \{\mathbf{v} : G\mathbf{v}^\top = \mathbf{0}\}.$$

- The dimension of $\mathcal{C}^\perp$ is $n - k$.
- A generator matrix H of the dual code $\mathcal{C}^\perp$ is also called a *parity-check matrix* of the original code $\mathcal{C}$.
- We can write

$$\mathcal{C} = \{\mathbf{x} : H\mathbf{x}^\top = \mathbf{0}\}.$$

Why Linear Codes?

- The description of linear codes is simple.
- Encoding complexity $O(n^2)$, and even simpler if there exists a sparse generator matrix.
- Linear codes achieve the capacity.

Examples of Linear Codes

- Hamming codes (1950)
- Reed-Solomon codes (early 1950s)
- BCH codes (1959)
- Convolutional codes (1955)
- Turbo codes (1993)
- LDPC (1962, 1997)
- Fountain codes (1998)
- Polar codes (2006)

Hamming Distance

- Let $\mathbb{A}$ be an alphabet of q elements.
- The *Hamming distance* of two vector $\mathbf{x}, \mathbf{y} \in \mathbb{A}^n$, denoted by $d(\mathbf{x}, \mathbf{y})$, is the number of coordinates i with different values.
- The Hamming distance is a metric since
    1. $d(\mathbf{x}, \mathbf{y}) \geq 0$, with equality iff $\mathbf{x} = \mathbf{y}$.
    2. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.
    3. $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{y}, \mathbf{z})$.

Minimum Distance Decoding

- Consider a memoryless BSC with cross-over probability $\epsilon \leq 1/2$.
- The *maximum likelihood* (ML) decoding rule for received vector $\mathbf{y}$ reads

$$\begin{aligned}
\hat{\mathbf{x}} &= \underset{\mathbf{x}:H\mathbf{x}^\top=0}{\operatorname{argmax}} W_n(\mathbf{y}|\mathbf{x}) \\
&= \underset{\mathbf{x}:H\mathbf{x}^\top=0}{\operatorname{argmax}} \prod_{i=1}^{n} W(y_i|x_i) \\
&= \underset{\mathbf{x}:H\mathbf{x}^\top=0}{\operatorname{argmax}} \epsilon^{d(\mathbf{x},\mathbf{y})}(1-\epsilon)^{n-d(\mathbf{x},\mathbf{y})} \\
&= \underset{\mathbf{x}:H\mathbf{x}^\top=0}{\operatorname{argmin}} d(\mathbf{x},\mathbf{y}).
\end{aligned}$$

Syndrome Decoding

- Let $\mathbf{s} = H\mathbf{y}^\top$, which is called the syndrome. We further have

$$\hat{\mathbf{x}} = \underset{\mathbf{x}: H\mathbf{x}^\top = 0}{\operatorname{argmin}} \, w(\mathbf{x} - \mathbf{y}) \\ = \mathbf{y} - \underset{\mathbf{e}: H\mathbf{e}^\top = \mathbf{s}}{\operatorname{argmin}} \, w(\mathbf{e})$$

ML decision problem
Is there $\mathbf{e} \in \{0, 1\}^n$ such that $w(\mathbf{e}) \leq c$ and $H\mathbf{e}^\top = \mathbf{s}$?

Theorem
The ML decision problem for BSC is NP-complete.

Hat Problem

- A number N of players are each wearing a hat, which may be of blue or red colours.
- Players can see the colors of all other players' hats, but not that of their own.
- Without any communication, some of the players must guess the color of their hat. Not all players are required to guess.
- All players who guess must decide at the same predetermined time, i.e., they don't know other's guess.
- Players win if at least one player guesses and all of those who guess do so correctly.
- How can the players maximise their chance of winning?

Minimum Distance

Minimum Distance

- The minimum distance of a code $\mathcal{C}$ is

$$d_{\mathsf{min}} \triangleq \min_{\mathbf{x} \neq \mathbf{y} \in \mathcal{C}} d(\mathbf{x}, \mathbf{y}).$$

Hamming Weight

- The *Hamming weight* of vector $\mathbf{z} \in \mathcal{A}^n$, denoted by $w(\mathbf{z})$, is the number of non-zero components in $\mathbf{z}$.
- Suppose $\mathcal{A}$ is a finite field.
- For $\mathbf{x}, \mathbf{y} \in A^n$, $d(\mathbf{x}, \mathbf{y}) = w(\mathbf{x} - \mathbf{y})$.
- For a linear code $d_{\min} = \min_{\mathbf{x} \neq \mathbf{0} \in \mathcal{C}} w(\mathbf{x})$.

Error Correction

- A code is t-error correcting if there exists a decoding algorithm such that the code can be decoded correctly for any t or less than t errors.

Theorem
A code is t-error correcting iff $d_{min} \geq 2t + 1$.

Error Detection

- A code is t-error detecting if there exists a decoding algorithm such that the code can be decoded correctly when there is no errors and an error message is generated for any c, $0 < c \leq t$, errors.

Theorem
A code is t-error detecting iff $d_{min} \geq t + 1$.

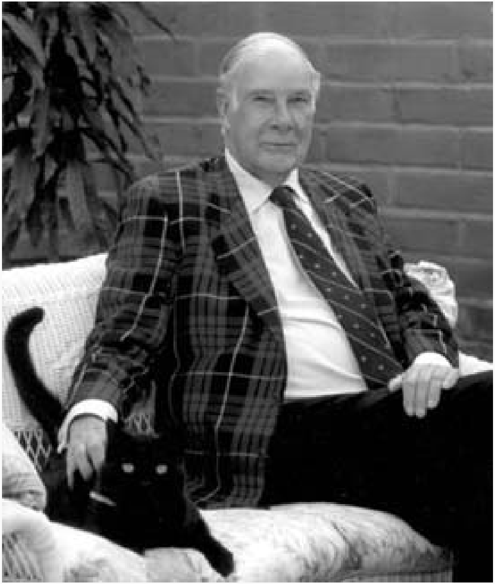
Erasure Correction

- A code is t-erasure correcting if the code can be decoded correctly for any t or less than t erasures.

Theorem
A code is t-erasure correcting iff $d_{min} \geq t + 1$.

Hamming Codes

Richard Hamming (1915 -1998)



All storage devices make errors!

1. magnetic tape
2. hard disk, floppy disk
3. optical disk
4. flash memory
5. distributed storage
6. cloud storage

Error Models

- Bit-flip errors.
- Erasure is also common in storage devices.
- More sophisticated error models can be obtained by investigating the underlying physical phenomenons of a particular storage devices.

Hamming's quesiton

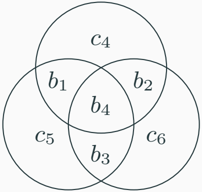
If there exists only one bit flip, how to correct it?

Repetition codes:

- Repeat each bit three times
- Majority vote

$(7, 4)$ **Hamming Code**

- Encode each block of 4 bits to a 7-bit codeword.



- Generator matrix

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

- Encoding: $\mathbf{c} = [b_1 b_2 b_3 b_4]G$.

$(7, 4)$ Hamming Code

- Parity check matrix

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

- $\mathrm{rank}(H) = 3$.
- $\mathrm{rank}(C) = 4$.
- The minimum (Hamming) weight of a codeword is 3.

General Hamming Codes

- Let m be a nonnegative integer, and $n = 2^m - 1$.
- Let H be an $m \times n$ binary matrix whose columns are formed by all the nonzero m-tuples.

Theorem
The code $\mathcal{C}$ with H as the parity-check matrix has the following properties:

1. *The dimension of $\mathcal{C}$ is $k = 2^m - m - 1$.*
2. *The minimum weight of a codeword is 3.*
3. *A binary vector of length 2^n is either a codeword, or one flip away from a unique codeword.*

Syndrome Decoding for Hamming Codes

- Transmit $\mathbf{x} \in \mathcal{C}$.
- Receive $\mathbf{y} = \mathbf{x} + \mathbf{e}_i$.
- Calculate $H\mathbf{y}^\top = H\mathbf{x}^\top + H\mathbf{e}_i^\top = h_i$.
- So $H\mathbf{y}^\top$ tells the position of the error.

Hamming Bound (Sphere-Packing Bound)

Theorem
For a block code $\mathcal{C} \subset \mathbb{A}^n$ *satisfies*

$$|\mathcal{C}| \leq \frac{q^n}{\sum_{i=0}^{t} \binom{n}{i} (q-1)^i}$$

where $t = \lfloor (d_{min} - 1)/2 \rfloor$.

Hamming Bound (Sphere-Packing Bound)

Theorem
For a block code $\mathcal{C} \subset \mathbb{A}^n$ *satisfies*

$$|\mathcal{C}| \leq \frac{q^n}{\sum_{i=0}^{t} \binom{n}{i}(q-1)^i}$$

where $t = \lfloor (d_{min} - 1)/2 \rfloor$.

Binary Hamming codes achieve the Hamming bound.