Chapter 7

Basic Algebraic Coding Theory

7.1 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

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

One practical reason to use linear codes is that it is easy for encoding. To record all the code words in a hard drive is not possible!

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

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

Syndrome Decoding

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

$$\begin{aligned}
\hat{\mathbf{x}} &= \operatorname*{argmin}_{\mathbf{x}: H\mathbf{x}^\top = 0} w(\mathbf{x} - \mathbf{y}) \\
&= \mathbf{y} - \operatorname*{argmin}_{\mathbf{e}: H\mathbf{e}^\top = \mathbf{s}} w(\mathbf{e})
\end{aligned}$$

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 7.1 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?

7.2 Minimum Distance

Minimum Distance

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

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

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

Error Detection

- Decoder: return the correct codeword or announce errors.
- Example: CRC
- A code is c-error detecting if the code can detect correctly for any c or less than c errors.

Theorem 7.3 A code is c-error detecting iff $d_{\min} \geq c + 1$.

Erasure Correction

- A code is c-error correcting for erasure if the code can decode correctly for any c or less than c erasures.

Theorem 7.4 A code is c-error correcting for erasure iff $d_{\min} \geq c + 1$.

Intractability of Computing Minimum Distance

Theorem 7.5 The problem of computing the minimum distance of a binary linear code is NP-hard, and the corresponding decision problem is NP-complete.

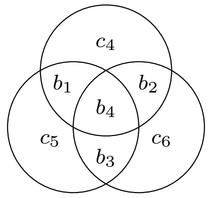
7.3 Hamming Codes

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

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

- $\text{rank}(H) = 3$.
- $\dim(C) = 4$.
- 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 7.6 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 non-zero codeword is 3.
3. A binary vector of length n is either a codeword, or one flip away from a unique codeword.

Proof. 1. H is full rank. 2. Any two columns of H are linearly independents, but not for some set of three columns of H. 3. Check that $2^k + 2^k n = 2^n$. ∎

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 7.7 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.

7.4 Reed-Solomon Codes

Applications of Reed-Solomon Codes

- Burst error protection: in many scenarios, couple bits are treated as a symbol.
- Communications
- Storage
- Bar code

Reed-Solomon Codes

- The alphabet is the finite field $\mathbb{F}$ with q elements, where $q \geq n$.
- Let $\alpha_1, \ldots, \alpha_n$ be n distinct elements of $\mathbb{F}$.
- Encoding:
  - For a message $\mathbf{m} = (m_0, \ldots, m_{k-1})$, define polynomial

$$p_{\mathbf{m}}(x) = m_0 + m_1 x + \cdots + m_{k-1} x^{k-1}.$$

  - $\mathbf{m} \mapsto (p_{\mathbf{m}}(\alpha_1), \ldots, p_{\mathbf{m}}(\alpha_n))$.
- Generator matrix:

$$G = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \cdots & \alpha_n^{k-1} \end{bmatrix}$$

How to generate a systematic Reed-Solomon code?

Decoding of Reed-Solomon Codes

- The Reed-Solomon code with above parameters is a $(n, k, n-k+1)$ code.
- Decoding algorithms:
  - Syndrome decoding (E.g. Berlekamp-Massey algorithm)
  - List decoding (Sudan and Guruswami's algorithms)
  - Soft decoding (Kötter and Vardy)

Welch-Berlekamp Algorithm

- Decoding problem:
  - Given: n pairs of field elements (α_i, r_i), $i = 1, \ldots, n$, and a parameter k.
  - Task: Find a polynomial $p(x)$ of degree less than k such that $p(\alpha_i) = r_i$ for at least $(n+k)/2$ values of $i \in \{1, \ldots, n\}$.
- Error polynomial $E(x)$
  - $p(\alpha_i) \neq r_i$ implies $E(\alpha_i) = 0$.
  - Given E, p can be computed efficiently.
  - Such an E exists: $E(x) = \prod_{i: r_i \neq p(\alpha_i)} (x - \alpha_i)$.
  - E has degree equal to the number t of errors and the most significant coefficient is 1.
- Key equation: $r_i E(\alpha_i) = E(\alpha_i) p(\alpha_i)$ for $i = 1, \ldots, n$.

Welch-Berlekamp Algorithm

- Let $Q(x) = E(x)p(x)$, which has degree $k-1+t$.
- Take the unknown coefficients of Q and E as variables and solve the linear system

$$r_i E(\alpha_i) = Q(\alpha_i), i = 1, \ldots, n.$$

- Try $t = 0, 1, \ldots, (n-k)/2$.
- A solution exists, but may not be unique.

Suppose (E, Q) and (E', Q') are both solutions of the linear system. We have for $i = 1, \ldots, n$

$$r_i E(\alpha_i) = Q(\alpha_i), \quad r_i E'(\alpha_i) = Q'(\alpha_i),$$

and hence

$$r_i E(\alpha_i) Q'(\alpha_i) = Q(\alpha_i) Q'(\alpha_i) = r_i E'(\alpha_i) Q(\alpha_i).$$

When $r_i \neq 0$, we obtain

$$E(\alpha_i) Q'(\alpha_i) = E'(\alpha_i) Q(\alpha_i).$$

When $r_i = 0$, we have

$$E(\alpha_i) Q'(\alpha_i) = E'(\alpha_i) Q(\alpha_i) = 0.$$

So for n values, $E(x)Q'(x)$ and $E'(x)Q(x)$ are the same.

Singleton Bound

Theorem 7.8 For a block code $\mathcal{C} \subset \mathcal{A}^n$ satisfies

$$|\mathcal{C}| \leq q^{n-d_{\min}+1}.$$

- Codes that achieve the Singleton bound is also called maximum distance separable (MDS) codes.
- Reed-Solomon codes are MDS.

Proof. Generate a matrix M of $|\mathcal{C}|$ rows and n columns. Remove any $d_{\min} - 1$ columns from the matrix. In the remaining part M', all the rows are different. Otherwise, suppose two rows are the same, and then the two rows in M are two codewords of distance at most $d_{\min} - 1$. ∎

MDS conjecture

- There exist linear MDS codes over $\mathbb{F}_q$ of length $n = q + 1$.
- (Bush 1952) If $k \geq q + 1$, then for any MDS codes $n \leq k + 1$.
- (MDS conjecture, Segre 1955) If $k \leq q$ then for any MDS codes $n \leq q + 1$, unless $q = 2^h$ and $k = 3$ or $k = q - 1$, in which case $n \leq q + 2$.

$$G = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n & 0 \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 & \vdots \\ \vdots & \vdots & \ddots & \vdots & 0 \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \cdots & \alpha_n^{k-1} & 1 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ 0 & 0 & \ddots & \vdots & 1 \\ 0 & 0 & \cdots & 1 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & 0 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_q & 0 & 1 \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_q^2 & 1 & 0 \end{bmatrix}$$

7.5 Greedy algorithms

Gilbert-Varshamov Bound (Sphere-Covering Bound)

Theorem 7.9 There exists a code $\mathcal{C} \subset \mathcal{A}^n$ such that

$$|\mathcal{C}| \geq \frac{q^n}{\sum_{i=0}^{d_{\min}-1} \binom{n}{i}(q-1)^i}.$$

Theorem 7.10 There exists a linear code $\mathcal{C} \subset \mathcal{A}^n$ with dimension k such that

$$k \geq n - \log_q \sum_{i=0}^{d_{\min}-1} \binom{n}{i}(q-1)^i.$$

For any $\mathbf{x} \in \mathcal{A}^n$, let

$$\mathcal{B}(\mathbf{x}) = \{\mathbf{y} \in \mathcal{A}^n : d(\mathbf{x}, \mathbf{y}) \leq d_{\min} - 1\}.$$

We have

$$|\mathcal{B}(\mathbf{x})| = \sum_{i=0}^{d_{\min}-1} \binom{n}{i}(q-1)^i.$$

Proof of Theorem 7.9. There exists a code $\mathcal{C}$ of minimum distance $d_{\min}$ such that

$$\mathcal{A}^n \subset \cup_{\mathbf{x}\in\mathcal{C}} \mathcal{B}(\mathbf{x}),$$

since otherwise, we can add certain $\mathbf{x} \in \mathcal{A}^n \setminus (\cup_{\mathbf{x}\in\mathcal{C}} \mathcal{B}(\mathbf{x}))$ to $\mathcal{C}$ without changing the minimum distance. Hence,

$$|\mathcal{A}^n| \leq |\cup_{\mathbf{x}\in\mathcal{C}} \mathcal{B}(\mathbf{x})| \leq |\mathcal{C}||\mathcal{B}(\mathbf{x})|,$$

which leads to the theorem. ■

Proof of Theorem 7.10. For $\mathcal{B}, \mathcal{C} \in \mathcal{A}^n$, define

$$\mathcal{B} \oplus \mathcal{C} = \{b + c : b \in \mathcal{B}, c \in \mathcal{C}\}.$$

There exists a code $\mathcal{C}$ of minimum distance $d_{\min}$ such that

$$\mathcal{A}^n \subset \mathcal{C} \oplus \mathcal{B}(\mathbf{0}),$$

where $\mathbf{0}$ is the all zero vector in $\mathcal{A}$. Otherwise, i.e., there exists $\mathbf{x} \in \mathcal{A}^n \setminus (\mathcal{C} \oplus \mathcal{B}(\mathbf{0}))$, we claim that $\mathcal{C}\oplus < \mathbf{x} >$ is a linear code of the same minimum distance. For certain $\mathbf{c} \in \mathcal{C}$ and $\alpha \neq 0$, if the weight of $\mathbf{c} + \alpha\mathbf{x}$ is less than $d_{\min}$, i.e., $\mathbf{c} + \alpha\mathbf{x} \in \mathcal{B}(\mathbf{0})$, then $\mathbf{x} \in \mathcal{C} \oplus \mathcal{B}(\mathbf{0})$, a contradiction. Hence,

$$|\mathcal{A}^n| \leq |\mathcal{C} \oplus \mathcal{B}(\mathbf{0})| \leq |\mathcal{C}||\mathcal{B}(\mathbf{0})|,$$

which leads to the theorem. ■

Asymptotic Gilbert-Varshamov Bound

- Let $\delta = d/n$.
- For a fixed rate r, $0 < r < 1$,

$$\delta^*(r) = \limsup_{n\to\infty} \max\{d(C)/n : C \in \mathcal{C}(n, 2^{\lfloor nr \rfloor})\}.$$

Theorem 7.11 $h(\delta^*(r)) \geq 1 - r$.

Proof. Using G-V bound and $\binom{n}{n\delta} \approx 2^{nh(\delta)}$. ■