Coding and Information Theory
Chapter 7:
Linear Codes - A

Xuejun Liang

Chapter 7: Linear Codes

1. Matrix Description of Linear Codes
2. Equivalence of Linear Codes
3. Minimum Distance of Linear Codes
4. The Hamming Codes
5. The Golay Codes
6. The Standard Array
7. Syndrome Decoding

Key content in this chapter

- Will study linear codes in greater detail by applying elementary linear algebra and matrix theory
  - including an even simpler method for calculating the minimum distance.
- Theoretical background required includes
  - Topics such as linear independence, dimension, and row and column operations
  - Linear space on a finite field

7.1 Matrix Description of Linear Codes

- Given a linear code $C \subseteq V = F^n$ and let dim(C) = k. A **generator matrix G for C is** defined as a $k \times n$ matrix, in which the row vectors are a basis of C.

- Example 7.1
  - The repetition code R_n over F has a single basis vector $\mathbf{u_1}$ = 11 . . . 1, so it has a generator matrix G = (11 … 1)

Example 7.2

The parity-check code P_n over F has basis $\mathbf{u_1}, \ldots, \mathbf{u_{n-1}}$ where each $\mathbf{u}_i = \mathbf{e_i} - \mathbf{e_n}$ in terms of the standard basis vectors $\mathbf{e_1}, \ldots, \mathbf{e_n}$ of V, so it has a generator matrix G

$$G = \begin{pmatrix} 1 & & & & -1 \\ & 1 & & & -1 \\ & & \ddots & & \vdots \\ & & & 1 & -1 \end{pmatrix}$$

We have proved $\mathbf{u_1}, \ldots, \mathbf{u_{n-1}}$ are linearly independent in Example 6.4

$$\boldsymbol{e_1} = (10 \ldots 00) \qquad \boldsymbol{u_1} = \boldsymbol{e_1} - \boldsymbol{e_n} = (10 \ldots 0-1)$$

$$\boldsymbol{e_2} = (01 \ldots 00) \qquad \boldsymbol{u_2} = \boldsymbol{e_2} - \boldsymbol{e_n} = (01 \ldots 0-1)$$

$$\boldsymbol{e_n} = (00 \ldots 01) \qquad \boldsymbol{u_{n-1}} = \boldsymbol{e_{n-1}} - \boldsymbol{e_n} = (00 \ldots 1-1)$$

$$\boldsymbol{a} = (a_1 a_2 \ldots a_{n-1} a_n) \in P_n \Rightarrow a_1 + a_2 + \cdots + a_{n-1} + a_n = 0$$

$$\Rightarrow a_n = -(a_1 + a_2 + \cdots + a_{n-1})$$

$$\Rightarrow a_1 \boldsymbol{u_1} + a_2 \boldsymbol{u_2} + \cdots + a_{n-1} \boldsymbol{u_{n-1}} = (a_1 a_2 \ldots a_{n-1} \, -(a_1 + a_2 + \cdots + a_{n-1})) = (a_1 a_2 \ldots a_{n-1} a_n) = \boldsymbol{a}$$

Example 7.3

A basis $\mathbf{u_1}$ = 1110000, $\mathbf{u_2}$ = 1001100, $\mathbf{u_3}$ = 0101010, $\mathbf{u_4}$ = 1101001 for the binary Hamming code H_7 was given in Example 6.5. So, this code has a generator matrix G.

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

Recall: How to construct the code for **a** = $a_1a_2a_3a_4$

Let the code word u = $u_1u_2u_3u_4u_5u_6u_7$

Bits $u_3 = a_1$, $u_5 = a_2$, $u_6 = a_3$, and $u_7 = a_4$

Bits u_1, u_2, u_4 for checking, determined by

$$u_4 + u_5 + u_6 + u_7 = 0$$
$$u_2 + u_3 + u_6 + u_7 = 0$$
$$u_1 + u_3 + u_5 + u_7 = 0$$

$$a_1\begin{pmatrix}1\\1\\1\\0\\0\\0\\0\end{pmatrix} + a_2\begin{pmatrix}1\\0\\0\\1\\1\\0\\0\end{pmatrix} + a_3\begin{pmatrix}0\\1\\0\\1\\0\\1\\0\end{pmatrix} + a_4\begin{pmatrix}1\\1\\0\\1\\0\\0\\1\end{pmatrix} = \begin{pmatrix}a_1 + a_2 + a_4\\a_1 + a_3 + a_4\\a_1\\a_2 + a_3 + a_4\\a_2\\a_3\\a_4\end{pmatrix} = \begin{pmatrix}u_1\\u_2\\u_3\\u_4\\u_5\\u_6\\u_7\end{pmatrix} = \boldsymbol{u}$$

Encoding of Source

- Given a linear code $C \subseteq V = F^n$ and let dim(C) = k.
- Then the k-dimensional vector space $A = F^k$ can be regarded as a source
- Encoding of source $A = F^k$ is a linear isomorphism $A \to C \subseteq V = F^n$ ($\boldsymbol{a} \in A \mapsto \boldsymbol{u} \in C$) given by the matrix G

$$\boldsymbol{u} = \boldsymbol{a}G$$

  - $\boldsymbol{a} = a_1 \dots a_k$ is a word
  - $\boldsymbol{u} = u_1 \dots u_n$ is a code-word
  - Thus encoding is multiplication by a fixed matrix
- Example 7.4
  - The repetition code R_n has k = 1, so A = F^1 = F. Each $\boldsymbol{a}$ = a ∈ A is encoded as $\boldsymbol{u}$ = $\boldsymbol{a}G$ = $a \dots a$ ∈ R_n.

G = (11 … 1)

Example 7.5

- If $C = P_n$ then $k = n - 1$, so $A = F^{n-1}$.
- Each **a** = $a_1 \ldots a_{n-1} \in A$ is encoded as

  u = **a**G = $a_1 \ldots a_{n-1}a_n$

  where $a_n = -(a_l + \ldots + a_{n-1})$, so $\sum_i a_i = 0$

$$G = \begin{pmatrix} 1 & & & -1 \\ & 1 & & -1 \\ & & \ddots & \vdots \\ & & 1 & -1 \end{pmatrix}$$

$$(a_1 a_2 \ldots a_{n-1})G = (a_1 a_2 \ldots a_{n-1} \; -(a_1 + a_2 + \cdots + a_{n-1})) \\ = (a_1 a_2 \ldots a_{n-1} a_n) = \boldsymbol{a}$$

$$a_1 \boldsymbol{u_1} + a_2 \boldsymbol{u_2} + \cdots + a_{n-1} \boldsymbol{u_{n-1}} = (a_1 a_2 \ldots a_{n-1} \; -(a_1 + a_2 + \cdots + a_{n-1})) \\ = (a_1 a_2 \ldots a_{n-1} a_n) = \boldsymbol{a}$$

Example 7.6

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

- If $C = H_7$ then n = 7 and k = 4, so A = F_2^4.
- Each **a** = a_1 ... a_4 ∈ A is encoded as

  $$\mathbf{u} = \mathbf{a}G \in H_7.$$

- For example, **a** = 0110, then **u** = **a**G = (1100110)

$$(0\ 1\ 1\ 0) \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} = (1\ 1\ 0\ 0\ 1\ 1\ 0)$$

Whether a Vector is a Code Word ?

- Given a linear code $C \subseteq V = F^n$ and let dim(C) = k.
- Want to determine whether a vector $v \in V$ is in C
- C consists of all solutions of a set of n - k simultaneous linear equations.
- Example 7.7
  - The repetition code R_n consists of the vectors $v = v_1 \ldots v_n \in V$ satisfying $v_1 = \cdots = v_n$, which can be regarded as a set of n - k = n - 1 simultaneous linear equations $v_i - v_n = 0$ (i = 1, .. . , n - 1).

Two More Examples

- Example 7.8
  - The parity-check code P_n (which has n - k = 1) is the subspace of V defined by the single linear equation $v_1 + \cdots + v_n = 0$.
- Example 7.9
  - The Hamming code H_7 consists of the vectors $v = v_1 \ldots v_7 \in V = F_2^7$ satisfying

$$v_4 + v_5 + v_6 + v_7 = 0,$$
$$v_2 + v_3 + v_6 + v_7 = 0,$$
$$v_1 + v_3 + v_5 + v_7 = 0.$$

Parity-Check Matrix H for C

- These equations are called parity-check equations
- Their matrix H of coefficients is called a parity-check matrix for C

- Lemma 7.10
  - Let C be a linear code, contained in V, with parity-check matrix H, and let $v \in V$. Then $v \in C$ if and only if

$$vH^T = 0,$$

where H^T denotes the transpose of the matrix H.

Compute parity-check matrix H for C

- Example 7.11: The repetition code R_n.

$v_i - v_n = 0$ (i = 1, .. . , n - 1).

$$H = \begin{pmatrix} 1 & & & & -1 \\ & 1 & & & -1 \\ & & \ddots & & \vdots \\ & & & 1 & -1 \end{pmatrix}$$

Compute parity-check matrix H for C

- Example 7.12: The parity-check code P_n.

$$v_1 + \cdots + v_n = 0$$

$$H = (1 \quad 1 \quad \ldots \quad 1)$$

Compute parity-check matrix H for C

- Example 7.13: The Hamming code H_7.

$$v_4 + v_5 + v_6 + v_7 = 0,$$

$$v_2 + v_3 + v_6 + v_7 = 0,$$

$$v_1 + v_3 + v_5 + v_7 = 0.$$

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

Dual Code of C

- H can be viewed as the matrix of a linear transformation $h: V \to W = F^{n-k}$
  - $\boldsymbol{v} \mapsto h(\boldsymbol{v}) = \boldsymbol{v}H^T$
- We have
  - $C = \ker(h) = \{\boldsymbol{v}: h(\boldsymbol{v}) = 0\}$
  - $im(h) = \{h(\boldsymbol{v}): v \in V\}$
  - $\dim(V) = \dim(\ker(h)) + \dim(im(h))$
  - H has rank $n–k$.
- So, n-k rows of H forms a basis of a linear space $D \subseteq V$ of dimension n-k. This linear code, with generator matrix H, called the dual code of C.