Linear Algebra

Course 11

Chapter 4. Introduction to Coding Theory

Part II

Contents

Matrix representation

- A binary n-digit word $a_0 a_1 \dots a_{n-1}$ may be identified with a matrix $\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} \in M_{n,1}(\mathbb{Z}_2)$.
- For an (n, k)-code, we see the 2^k possible messages as the elements of the vector space $\mathbb{Z}_2^k$ over $\mathbb{Z}_2$, and the 2^n possible received words as the elements of the vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$.

Definition

- An *encoder* of an (n, k)-code is an injective function $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ (or equivalently, $\gamma : M_{k,1}(\mathbb{Z}_2) \to M_{n,1}(\mathbb{Z}_2)$).
- An (n, k)-code is called *linear* if its encoder is a linear map.

Theorem

Any (n, k)-code generated by a polynomial of degree $n - k$ is linear.

E.g. *Reed-Solomon code*, used for CD's, DVD's, Blu-ray discs etc.

Generator matrix

Definition

Consider a linear (n, k)-code with encoder $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$. Let E, E' be the canonical bases of the $\mathbb{Z}_2$-vector spaces $\mathbb{Z}_2^k$ and $\mathbb{Z}_2^n$ respectively. Then the matrix

$$G = [\gamma]_{EE'}$$

is called the *generator matrix* of the code.

A message $m \in \mathbb{Z}_2^k$ encodes as $\gamma(m)$.

But for $m \in \mathbb{Z}_2^k$, we have $[\gamma(m)]_{E'} = [\gamma]_{EE'} \cdot [m]_E$.

Hence a message $m \in M_{k,1}(\mathbb{Z}_2)$ encodes as $G \cdot [m]_E$.

Generator matrix - cont.

Use the above notation.

Theorem

(i) The code words of the (n, k)-code are the vectors in the subspace $\operatorname{Im}\gamma$ of $\mathbb{Z}_2^n$. Hence a binary (n, k)-code means a k-dimensional subspace of the vector space $\mathbb{Z}_2^n$.
(ii) The columns of G form a basis of this subspace, and so a vector is a code vector if and only if it is a unique linear combination of the columns of G.

Remark. A code word contains the message digits on the last k positions. Hence the generator matrix G of an (n, k)-code is always of the form

$$G = \begin{pmatrix} P \\ I_k \end{pmatrix} \in M_{n,k}(\mathbb{Z}_2),$$

where $P \in M_{n-k,k}(\mathbb{Z}_2)$ and $I_k \in M_k(\mathbb{Z}_2)$ is the identity matrix.

Parity check matrix

Definition

With the above notation, the matrix

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} \in M_{n-k,n}(\mathbb{Z}_2)$$

is called the *parity check matrix* of the code.

Theorem

Consider a linear (n, k)-code with parity check matrix $H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} \in M_{n-k,n}(\mathbb{Z}_2)$. Then a received vector $u \in \mathbb{Z}_2^n$ (or $u \in M_{n,1}(\mathbb{Z}_2)$) is a code vector if and only if $H \cdot [u]_{E'} = [0]_{E'}$.

Matrix representation - example

Example. *Determine the generator matrix and the parity check matrix of the* $(6,3)$*-code generated by the polynomial* $p = 1 + X + X^3 \in \mathbb{Z}_2[X]$*, and characterize the code vectors.*

Solution. Note that $n = 6$ and $k = 3$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, i.e. $\gamma : \mathbb{Z}_2^3 \to \mathbb{Z}_2^6$. The encoding of v is $\gamma(v)$.

- The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2^3$ and $\mathbb{Z}_2^6$ respectively. We have

$$\begin{aligned} e_1 = (1,0,0) &\rightsquigarrow 100 \rightsquigarrow m = 1 \rightsquigarrow m \cdot X^{n-k} = X^3 \\ &\rightsquigarrow r = m \cdot X^{n-k} \bmod p = X^3 \bmod p = 1 + X \\ &\rightsquigarrow v = r + m \cdot X^{n-k} = 1 + X + X^3 \\ &\rightsquigarrow \boxed{110}\boxed{100} \rightsquigarrow (1,1,0,1,0,0) = \gamma(e_1). \end{aligned}$$

Similarly, $e_2 = (0,1,0) \rightsquigarrow (0,1,1,0,1,0) = \gamma(e_2)$ and $e_3 = (0,0,1) \rightsquigarrow (1,1,1,0,0,1) = \gamma(e_3)$.

Matrix representation - example

- Hence $G = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} P \\ I_3 \end{pmatrix} = \begin{pmatrix} P \\ I_k \end{pmatrix}$.
- The parity check matrix is

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_3 & P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

- $(u_1, u_2, u_3, u_4, u_5, u_6) \in \mathbb{Z}_2^6$ is a code word $\Leftrightarrow H \cdot [u]_{E'} = [0]_{E'}$

$$\Leftrightarrow \begin{cases} u_1 + u_4 + u_6 = 0 \\ u_2 + u_4 + u_5 + u_6 = 0 \\ u_3 + u_5 + u_6 = 0 \end{cases} \Leftrightarrow \begin{cases} u_1 = u_4 + u_6 \\ u_2 = u_4 + u_5 + u_6 \\ u_3 = u_5 + u_6 \end{cases}.$$

Error-correcting and decoding

A naive method:

- Given a received word, compute all Hamming distances to the code words.
  (Recall that the Hamming distance between two words of the same length is the number of positions in which they difer.)
- The code word closest to the received word will be assumed to be the most likely transmitted word.

Not practical!

Intermezzo: quotient vector spaces

Theorem

Let U be a K-vector space and let V be a subspace of U. For $u \in U$ we denote $u + V = \{u + v \mid v \in V\}$. Then

$$U/V = \{u + V \mid u \in U\}$$

is a vector space over K with respect to the addition and the scalar multiplication given by

$$\begin{aligned} (u_1 + V) + (u_2 + V) &= (u_1 + u_2) + V, \quad \forall u_1, u_2 \in U, \\ k \cdot (u + V) &= (k \cdot u) + V, \quad \forall k \in K, \forall u \in U. \end{aligned}$$

Then U/V is called a quotient vector space*, and $u + V$ $(u \in U)$ is called a* coset*.*

Coset leaders

Consider an (n, k)-code with encoding function $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ and denote $V = \operatorname{Im} \gamma$ (the subspace of code vectors).

- Start with a code vector $v \in V = \operatorname{Im} \gamma \leq \mathbb{Z}_2^n$, and assume that an error $e \in \mathbb{Z}_2^n$ occurs during transmission.
- Then the received vector is $u = v + e \in \mathbb{Z}_2^n$. The receiver determines the most likely transmitted vector by finding the most likely error pattern (called the *coset leader*)

$$e = u - v = u + v \in u + V.$$

- The coset leader will usually be the coset containing the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's in the error pattern are bunched together as much as possible.

Syndromes

Theorem

Consider a linear (n, k)-code with generator and parity check matrices $G \in M_{n,k}(\mathbb{Z}_2)$ and $H \in M_{n-k,n}(\mathbb{Z}_2)$ respectively. Let

$$\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n \text{ and } \eta : \mathbb{Z}_2^n \to \mathbb{Z}_2^{n-k}$$

be the $\mathbb{Z}_2$-linear maps corresponding to G and H respectively. Then $V = \operatorname{Im}\gamma = \operatorname{Ker}\eta$.

If E' and E'' are the canonical bases of the $\mathbb{Z}_2$-vector spaces $\mathbb{Z}_2^n$ and $\mathbb{Z}_2^{n-k}$ respectively, then $[\eta(u)]_{E''} = [\eta]_{E'E''} \cdot [u]_{E'} = H \cdot [u]_{E'}$.

Definition

With the above notation, the vector $\eta(u) \in \operatorname{Im}\eta \leq \mathbb{Z}_2^{n-k}$ (or $H \cdot u \in M_{n-k,1}(\mathbb{Z}_2)$) is called the *syndrome* of u.

The number of syndromes for an (n, k)-code is 2^{n-k}.

A general method for decoding

1. Calculate the syndrome of the received word.
2. Find the coset leader of the coset corresponding to the syndrome.
3. Subtract the coset leader from the received word to obtain the most likely transmitted word.
4. Drop the check digits to obtain the most likely message.

Decoding - examples

Example 1. Construct a table of coset leaders and syndromes for the $(6, 3)$-code with parity check matrix

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

and then decode the received words 011100 and 100011.

Solution.

- We have an (n, k)-code, where $n = 6$ and $k = 3$.
- The number of syndromes is $2^{n-k} = 2^3 = 8$.
- We write down all possible syndromes in a table, and then we determine their corresponding coset leaders.

Decoding - examples

syndrome	coset leader
000	
001	
010	
011	
100	
101	
110	
111	

syndrome	coset leader
000	000000
001	001000
010	010000
011	000010
100	100000
101	000110
110	000100
111	000001

Decoding - examples

- The coset leaders (the most likely errors) are chosen such that they contain the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's are bunched together as much as possible.
- We first consider the coset leader with all bits 0, then coset leaders having only one bit 1, then two consecutive bits 1, then two bits 1 not necessarily consecutive etc., until we find all correspondences with the syndromes.
- We use the general matrix equality:

$$[\text{syndrome}] = H \cdot [\text{vector}].$$

Decoding - examples

- The syndrome of $u = 000000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 100000$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 010000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 001000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.
- The syndrome of $u = 000100$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 000010$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.
- The syndrome of $u = 000001$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Decoding - examples

- For the last syndrome, namely 101, we try with 110000, 011000, 001100, 000110 or 000011. The correct one is $u = 000110$, because $H \cdot [u] = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.
- To decode $u_1 = 011100$, compute its syndrome $H \cdot [u_1] = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.
  Its corresponding coset leader is $e_1 = 000110$.
  The most likely code vector is $v_1 = u_1 + e_1 = 011010$.
  Hence the most likely message is 010.
- To decode $u_2 = 100011$, compute its syndrome $H \cdot [u_2] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
  Its corresponding coset leader is $e_2 = 000000$.
  The most likely code vector is $v_2 = u_2 + e_2 = 100011$.
  Hence the most likely message is 011.

Decoding - examples

Example 2. Construct a table of coset leaders and syndromes for the $(7, 3)$-code generated by the polynomial

$$p = 1 + X + X^4 \in \mathbb{Z}_2[X].$$

Solution. Note that $n = 7$ and $k = 3$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, i.e. $\gamma : \mathbb{Z}_2^3 \to \mathbb{Z}_2^7$. The encoding of v is $\gamma(v)$.

- The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2^3$ and $\mathbb{Z}_2^7$ respectively. We have

$$\begin{aligned} e_1 = (1,0,0) &\rightsquigarrow 100 \rightsquigarrow m = 1 \rightsquigarrow m \cdot X^{n-k} = X^4 \\ &\rightsquigarrow r = m \cdot X^{n-k} \bmod p = X^4 \bmod p = 1 + X \\ &\rightsquigarrow v = r + m \cdot X^{n-k} = 1 + X + X^4 \\ &\rightsquigarrow \boxed{1100}\boxed{100} \rightsquigarrow (1,1,0,0,1,0,0) = \gamma(e_1). \end{aligned}$$

Similarly, $e_2 = (0,1,0) \rightsquigarrow (0,1,1,0,0,1,0) = \gamma(e_2)$ and $e_3 = (0,0,1) \rightsquigarrow (0,0,1,1,0,0,1) = \gamma(e_3)$.

Decoding - examples

- Hence $G = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} P \\ I_3 \end{pmatrix} = \begin{pmatrix} P \\ I_k \end{pmatrix}$.
- The parity check matrix is

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_4 & P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Decoding - examples

- The number of syndromes is $2^{n-k} = 2^4 = 16$.
- We write down all possible syndromes in a table, and then we determine their corresponding coset leaders.
- The coset leaders (the most likely errors) are chosen such that they contain the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's are bunched together as much as possible.
- We first consider the coset leader with all bits 0, then coset leaders having only one bit 1, then two consecutive bits 1, then two bits 1 not necessarily consecutive etc., until we find all correspondences with the syndromes.
- We use the general matrix equality:

$$[\text{syndrome}] = H \cdot [\text{vector}].$$

Decoding - examples

After computations, we obtain the following table:

syndrome	coset leader
0000	0000000
0001	0001000
0010	0010000
0011	0000001
0100	0100000
0101	0000011
0110	0000010
0111	0100001

syndrome	coset leader
1000	1000000
1001	1001000
1010	0000110
1011	1000001
1100	0000100
1101	0001100
1110	0010100
1111	0000101