Chapter 31 Coding Theory

- Use algebraic techniques to protect data transmission affected by noise (human error, imperfect channels, interference etc.)
- Suppose a $(0,1)$ sequence of length of $x_1 \cdots x_{500}$ is transmitted.
- If there is 1% probability that x_i is transmitted incorrectly for each x_i, then the probability of correct transmission is $(0.99)^{500} \sim .0066$.
- If each x_i is transmitted as $x_i x_i x_i$, and the message is decoded by the maximum likelihood scheme, then the probability for x_i to be wrongly decoded is:

$$3(0.01)^2(0.99)^2 + (0.01)^3 \sim 0.000298 < .0003.$$

- Thus, the probability of correct transmission for each x_i is larger than $.9997$, and the probability of correct transmission of $x_1 \cdots x_{500}$ is larger than $(.9997)^{500} \sim .86$.
- But repeating many times is not an efficient scheme, so we use algebraic techniques.

Hamming (7,4) Code

Example Encode (x_1, x_2, x_3, x_4) by $(x_1, x_2, x_3, x_4)G$ with

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

Then we have 2^4 code words in 2^7 such that every pair code words differ in at least 3 digits.

This can be checked by examining the list of nonzero code words in p. 528. Here we need only compute $(x - y)H$ using $\mathbb{Z}_2$ arithmetic for any $x - y \neq 0$.

Linear Code

Definition An (n, k) linear code over a finite field is a k-dimensional subspace V in $\mathbb{F}^n$ such that the elements in V are the code words. When $\mathbb{F}$ is $\mathbb{Z}_2$, we have a binary code.

Remark The Hamming (7,4) code is a binary code.

Example The set $\{0000, 0101, 1010, 1111\}$ is a (4,2) binary code.

Example The set $\{0000, 0121, 0212, 1022, 1110, 1201, 2011, 2102, 2220\}$ is a (4,2) linear code over $\mathbb{Z}_3$.

Hamming Distance, Hamming Weight

Definition The Hamming weight $wt(u)$ of $u \in \mathbb{F}^n$ is the number of nonzero entries in $u \in \mathbb{F}^n$. The Hamming distance $d(u, v)$ of $u, v \in \mathbb{F}^n$ is the number of positions in which they differ so that $d(u, v) = wt(u - v)$.

Theorem 31.1 The Hamming distance is a metric (a distance function) in $\mathbb{F}^n$.

Proof. (1) $d(u, v) \geq 0$ with equality if and only if $u - v = 0$.

(2) $d(u, v) = d(v, u)$.

(3) For u, v, w,

$$d(u, w) = wt(u - w) \leq wt(u - v) + wt(v - w) = d(u, v) + d(v, w).$$

To see that $wt(u - w) \leq wt(u - v) + wt(v - w)$, note that if u_i, w_i are different then u_i, v_i or v_i, w_i are different. $\square$

Theorem 31.2 Suppose the Hamming weight of a linear code is at least $2t + 1$. Then it can correct any t or fewer errors. Alternatively, it can be used to detect 2t or few errors.

Proof. We use the nearest neighbor decoding.

Suppose v is a received word. Decode it as the nearest code word.

If there is more than one, do not decode. [There are too many errors.]

Suppose u is transmitted and v is received with no more than t errors so that $d(u, v) \leq t$.

Let w be a code word other than u.

Then $2t + 1 \leq d(u, w) \leq d(u, v) + d(v, w) \leq t + d(v, w)$

so that $d(v, w) > t$. So, u is the unique correct code word nearest v.

Clearly, if u is transmitted as v with fewer than $2t$ error, then it cannot be another code word.

So, one can detect that are errors in the transmission. $\square$

Remark We cannot use it to do both.

For the Hamming (7,3) code, when there are two errors, one may decode it assuming one error occurs, or assume two errors and refuse to decode.

Systematic code.

We encode a code word $(a_1 \cdots a_k)$ as $(a_1 \cdots a_k)G$, where $G = \begin{bmatrix} I_k \mid A \end{bmatrix}$.

The first k digits are the message digits.

Example Let

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

$$000000, 001111, 010101, 100110, 110011, 101001, 011010, 111100.$$

All nonzero code words have weight at least 3.

So, it will correct single error, or detect up to 2 errors.

Example Example Messages are: $00, 01, 02, 10, 11, 12, 20, 21, 22$. Let

$$G = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 2 & 2 \end{bmatrix}.$$

Code words:

$$0000, 0122, 0211, 1021, 1110, 1202, 2012, 2101, 2220.$$

All code words have weights at least 3.

So, it will correct single error, or detect up to 2 errors.

Parity-Check Matrix Decoding

Suppose $G = [I_k \mid A]$ is the generator matrix. Let

$$H = \begin{bmatrix} -A \\ I_{n-k} \end{bmatrix}$$

be the parity check matrix.

- If w is received, computer wH.
- If $wH = 0$, then assume no error.
- If wH equals s times the ith row of H, then decode w as $w - se_i$, where e_i is the ith row of I_n.

  If there are more than one such instance, do not decode.

  If the code is binary, we simply change the ith position of w.
- If the last two cases do not happen, assume more than two errors occur, and do not decode.

Use the Hamming (7,4) code with

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

If $v = 0000110$ is received, then $vH = 110$ which is the first row of H.

So, we decode it as 1000110 so that the original message is 1000.

If $w = 1011111$ is received, then $wH = 101$, which is the second row of H. So, we decode it as 1111111 so that the original message is 1111.

If $u = 1001101$ is sent, and $z = 1001011$ is received so that $zH = 110$ is the first row. We will wrongly decode it as 0001.

Orthogonality relation

Lemma Let C be a systematic (n, k) linear code over $\mathbb{F}$ with a standard generator matrix G and parity matrix H. Then $v \in C$ if and only if $vH = 0$.

Proof. If $v \in C$, then $v = uG$ so that $vH = uGH = uO = 0$.

If $vH = 0$, then $v \in Ker(T)$ where $T : \mathbb{F}^n \to \mathbb{F}^{n-k}$ defined by $x \mapsto xH$.

We will show that the row space of G is $Ker(T)$ so that $vH = 0$ ensures that $v = uG$ for some u.

Note that row space of G has dimension k, range space of T has $n - k$. So, it suffices to prove that $C \subseteq Ker(T)$.

It is clear because $GH = 0$. $\square$

Theorem 31.3 Parity-check matrix decoding will correct any single error if and only if the rows of the parity-check matrix are nonzero and no one row is a scalar multiple of any other rows.

Proof.

Coset Decoding

Example Consider the (6,3) binary linear code.

$C = \{000000, 100110, 010101, 001001, 1110011, 101101, 011110, 111000\}.$

One can construct the 8 cosets, and use the element with minimum weight as the coset leaders for each of them.

In the above example, the coset leaders (listed as the first column) are:

000000, 100000, 010000, 001000, 000100,000010,000001,100001.

One can decode a received word as the code word in the vertical column containing the received word.

Theorem 31.4 The coset decoding is the same as minimum distance decoding.

proof. Let w be a received word.

If v is the coset leader of the coset containing w, then $w + C = v + C$.

If w is decoded as c, and c' is another code word, then

$$d(w, c') = wt(w - c') \geq wt(v) = wt(w - c) = d(w, c).$$

Thus, w is decoded as c, which has a minimum distance to w among all code words. □

Same coset - same syndrome

Definition If an (n, k) linear code over $\mathbb{F}$ has parity-check matrix H, then, for any vecotr $u \in \mathbb{F}^n$, the vector uH is call the dyndrome of u.

Theorem 31.5 Let C be an (n,k) linear code over $\mathbb{F}$ with a parity-check matrix H. Then, two vectors of $\mathbb{F}^n$ are in the same coset of C if and only if they have the same syndrome.

Proof. Two vectors u and v are in the same coset of C if and only if
$u - v \in C$.
By the orthogonality lemma, u and v are in the same coset if and only if
$0 = (u - v)H = uH - vH$. □

Syndrome decoding for a received word w.

1. Compute wH, the syndrome.
2. Find the coset leader v such that $wH = vH$.
3. Decode the vector sent was $w - v$.