Discrete Structures (Monsoon 2021)

Ashok Kumar Das

Associate Professor
IEEE Senior Member
Center for Security, Theory and Algorithmic Research
International Institute of Information Technology, Hyderabad (IIIT Hyderabad)

E-mail: *ashok.das@iiit.ac.in*
URL: http://www.iiit.ac.in/people/faculty/ashokkdas
https://sites.google.com/site/iitkgpakdas/

Coding Theory (Group Codes)

Error Detection

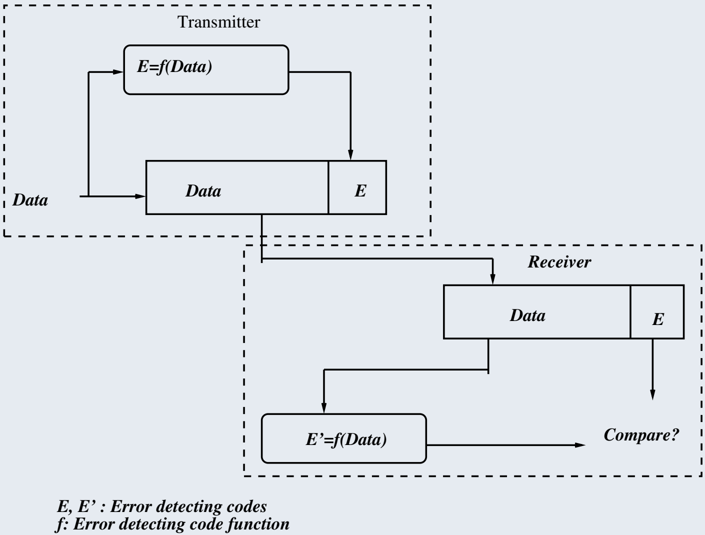


Figure: Error detection

Error Detection

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

- For a given frame of bits, additional bits that constitute an error-detecting code are added by the transmitter. This code is calculated as a function of the other transmitted bits.
- The receiver performs the same calculation and compares the two results. A detected error occurs if and only if there is a mismatch.

Group Codes

Definition

Let x and y be binary n-tuples, i.e., $x = \langle x_1, x_2, \ldots, x_n \rangle$ and $y = \langle y_1, y_2, \ldots, y_n \rangle$, where $x_i, y_i \in \{0, 1\}$. The Hamming distance between x and y denoted as $H(x, y)$ is the number of co-ordinates (components) in which they differ.

- Example: The Hamming distance between $\langle 1, 0, 1 \rangle$ and $\langle 1, 1, 0 \rangle$ is $H(\langle 1, 0, 1 \rangle, \langle 1, 1, 0 \rangle) = 2$.
- The Hamming distance between two n-tuples is equal to the number of independent single errors needed to change one n-tuple into the other.

Group Codes

Properties

- $H(x, y) \geq 0$, $\forall x, y \in C$, where C is the set of code words which are n-tuples $c_i = \langle c_{i,1}, c_{i,2}, \ldots, c_{i,n} \rangle$, $c_{i,j} \in \{0, 1\}$.
- $H(x, y) = 0$ if and only if $x = y$.
- $H(x, y) = H(y, x)$, $\forall x, y \in C$.
- $H(x, z) \leq H(x, y) + H(y, z)$, $\forall x, y, z \in C$.

Definition

The minimum distance (or minimum Hamming distance) of an n-coordinate code, C is $H_c = min_{c_i, c_j \in C} H(c_i, c_j)$.

Group Codes

Theorem

A code C can detect all combinations of d or fewer errors if and only if its minimum distance is at least $(d+1)$.
In other words,
C can detect $\leq d$ errors
if and only if
$H_c = \text{minimum distance of } C = min_{c_i, c_j \in C} H(c_i, c_j) \geq (d+1)$.

Group Codes

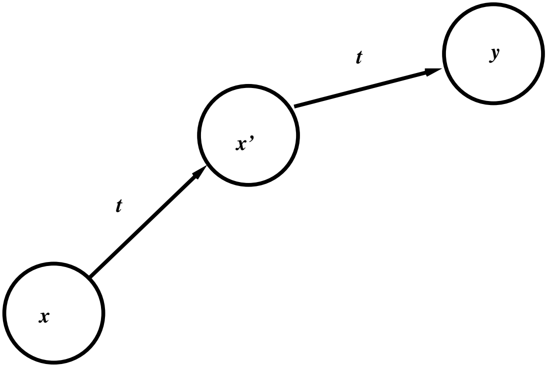
Theorem

A code C can correct every combination of t or fewer errors if and only if its minimum distance is at least $(2t+1)$.

Proof. Let C be a code of n-tuple code words c_i, where $c_i = \langle c_{i,1}, c_{i,2}, \ldots, c_{i,n} \rangle$, $c_{i,j} \in \{0, 1\}$.
The Hamming distance $H(x, y)$ between two n-tuple code words x and y, where $x, y \in C$, is $H(x, y) =$ number of coordinates in which they differ.
The minimum Hamming distance is given by $H_c = min_{c_i, c_j \in C} H(c_i, c_j)$.
$(\Rightarrow)$: Given C can correct $\leq t$ errors.
RTP: $H_c = 2t + 1$, that is, $\forall x, y \in C, H(x, y) \geq (2t + 1)$.
If possible, let $\exists x, y \in C$ such that $H(x, y) = 2t$.
Let $l_1, l_2, \ldots, l_{2t}$ be the coordinates (positions) where x and y differ.
Select $l_1, l_2, \ldots, l_t$ and change x to another n-tuple x' by changing x in these positions. Therefore, $H(x, x') = t$.

Group Codes

Proof (Continued . . .)



Group Codes

Proof (Continued . . .) But, then from the property of Hamming distance, we have:

$$\begin{aligned} H(x, y) &\leq H(x, x') + H(x', y) \\ &= t + t \\ H(x, y) &\leq 2t. \end{aligned}$$

There exists some n-tuple x' that satisfies $H(x, x') = t$ and $H(x', y) \leq t$.
This is a contradiction. Hence, $H_c = 2t + 1$, that is, $\forall x, y \in C, H(x, y) \geq (2t + 1)$.

Group Codes

Proof (Continued . . .)

$(\Leftarrow)$: Given $H_c = 2t + 1$, that is, $\forall x, y \in C$,

$$H(x, y) \geq 2t + 1. \quad (1)$$

Let x' be a received n-tuple that is corrupted by NOT more than t errors and x be a code word. x' has thus changed from x by t or fewer errors. Hence,

$$H(x, x') \leq t. \quad (2)$$

From the properties of Hamming distance, we have

$$\begin{aligned} H(x, y) &\leq H(x, x') + H(x', y) \\ H(x', y) &\geq H(x, y) - H(x, x') \\ &\geq t + 1, \text{ using Eqns. (1) and (2).} \end{aligned}$$

Therefore, every code word y is farther than x' than is x, and x can be correctly decoded. $\square$

Group Codes

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

Definition

A *group code* is a code from which n-tuple code words forms a group with respect to the operation $\oplus$ (modulo-2 or bitwise XOR), where $x \oplus y = \langle x_1 \oplus y_1, x_2 \oplus y_2, \ldots, x_n \oplus y_n \rangle$.

Definition

The weight of a code word x, denoted by $w(x)$, is the number of its coordinates (or components) that are 1s, that is, $w(x) =$ number of 1s in x.

Example: $w(\langle 1, 1, 1, 1 \rangle) = 4$
$w(\langle 1, 1, 0, 0 \rangle) = 2$.
We denote the n-tuple $\langle 0, 0, \ldots, 0 \rangle$ by 0.
Note that $w(x) = H(x, 0)$,
$H(x, y) = H(x \oplus y, 0) = w(x \oplus y)$.

Group Codes

Lemma

The minimum distance of a group code, C is equal to the minimum weight of its non-zero code words.

Definition (Null Space)

Let H be an $r \times n$ binary matrix. Then the set of binary n-tuples x that satisfies $x.H^t = 0$ is called the *null space* of H, $N(H)$. In other words,

$$N(H) = \{x | x.H^t = 0, x \in C\},$$

where C is the group code and H^t the transposition of the matrix H.

Group Codes

Theorem

The null space $N(H)$ of an $r \times n$ binary matrix H is a group under $\oplus$, component-wise addition modulo-2 (XOR).

Proof. Let H be an $r \times n$ binary matrix (parity-check matrix) and C a group code of n-tuples code words. Then the *null space* of H, $N(H)$ is

$$N(H) = \{x | x.H^t = 0, x \in C\},$$

where C is the group code and H^t the transposition of the matrix H.
RTP: $\langle N(H), \oplus \rangle$ is a group.

- Closure: Let $x, y \in N(H)$. Then, $x.H^t = 0$ and $y.H^t = 0$. Therefore, $x.H^t \oplus y.H^t = 0 \Rightarrow (x \oplus y).H^t = 0. \Rightarrow (x \oplus y) \in N(H)$. Hence, closure axiom holds.

Proof (Continued . . .). .

- Associativity: Since $((x \oplus y) \oplus z).H^t = (x \oplus (y \oplus z)).H^t$, $\forall x, y, z \in N(H)$, we have $(x \oplus y) \oplus z = x \oplus (y \oplus z)$. Associativity under $\oplus$ holds.
- Existence of Identity: We have: $(0 \oplus x).H^t = (x \oplus 0).H^t = x.H^t$, $\forall x \in N(H)$. Thus, $0 \oplus x = x = x \oplus 0$, $\forall x \in N(H)$. This implies that $0 = \langle 0, 0, \ldots, 0 \rangle$ is the identity in $N(H)$.
- Existence of Inverse: It is noted that
  $(x \oplus x).H^t = 0.H^t = 0$
  $\Rightarrow x \oplus x = 0$, $\forall x \in N(H)$.
  It shows that every element $x \in N(H)$ is its own inverse.

As a result, $N(H)$ forms a group under $\oplus$. □

Corollary

$\langle N(H), \oplus \rangle$ is an abelian (commutative) group.

Group Codes

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

Theorem

Let $c_1, c_2, \ldots, c_d$ be d distinct columns of the parity check $r \times n$ matrix H. Then the r-tuple sum $c_1 \oplus c_2 \oplus \cdots \oplus c_d$ is 0 if and only if the null space of H, $N(H)$ has a code word of weight d.

Theorem

H is a parity-check matrix for a code of minimum weight at least 3 if and only if
(i) no column of H is all 0s; and
(ii) no two columns are identical.
(iii) there exists three columns, whose sum is 0, that is, $\exists C_i, C_j, C_k$ such that $C_i \oplus C_j \oplus C_k = 0$.

Error detection/correction capability

Theorem

Let H be an $r \times n$ binary parity-check matrix of the form $[P|I_r]$, where I_r is an $r \times r$ identity matrix, and P an arbitrary $r \times (n-r)$ matrix. Then the code defined by H has 2^{n-r} code words. H is called the canonical parity-check matrix.

Error detection/correction capability of $N(H)$, the null space of a parity-check matrix H of a code, C
$=$ minimum weight of C
$=$ minimum number of columns, d of H that sum to 0
$= d$.

Code generation by parity checks

Let $H = [P|I_r]$ be a canonical parity-check matrix, where I_r is an $r \times r$ identity matrix, and P an arbitrary $r \times (n-r)$ matrix.

Let $k = n - r$.

Let

$$H = \left(\begin{array}{cccc|cccc} h_{11} & h_{12} & \cdots & h_{1k} & 1 & 0 & \cdots & 0 \\ h_{21} & h_{22} & \cdots & h_{2k} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{r1} & h_{r2} & \cdots & h_{rk} & 0 & 0 & \cdots & 1 \\ & & P & & & I_r & & \end{array} \right).$$

Encoding Procedure:

- Given a k-tuple message $x = \langle x_1, x_2, \ldots, x_k \rangle$, we need to compute the corresponding n-tuple code word (frame = message + error code) $y = \langle y_1, y_2, \ldots, y_k, y_{k+1}, \ldots, y_n \rangle$, where $k = n - r$, that is, $n = k + r$.
- Set $y_i \leftarrow x_i$, for all $1 \leq i \leq k$.

Code generation by parity checks

- Compute y_{k+i} for $1 \leq i \leq r$ as the modulo-2 sum:

$$
\begin{aligned}
y_1 h_{11} \oplus y_2 h_{12} \oplus \cdots & \\
\oplus y_k h_{1k} \oplus y_{k+1} h_{1,k+1} &= 0, \text{ since } h_{1,k+1} = 1 \\
\Rightarrow y_{k+1} &= y_1 h_{11} \oplus y_2 h_{12} \oplus \cdots \oplus y_k h_{1k}. \\
\text{Similarly,} & \\
y_{k+2} &= y_1 h_{21} \oplus y_2 h_{22} \oplus \cdots \oplus y_k h_{2k}. \\
\text{In general,} & \\
y_{k+i} &= \bigoplus_{j=1}^{k} y_j h_{i,j}.
\end{aligned}
$$

Code generation by parity checks

Decoding Procedure:

- Let C be a group code with individual code words c_i.
- Assume that the true code word is the n-tuple x, but the observed n-tuple is x', which is x after it has been corrupted by errors.
- Note that Hamming code is a single-error correcting code since H generates a code of minimum weight at least 3.
- Let ϵ be the error n-tuple that satisfies

$$\begin{aligned} x' &= x \oplus \epsilon \\ \Rightarrow x &= x' \oplus \epsilon. \end{aligned}$$

- We now show that the problem of finding ϵ reduces the problem of finding the coset to which x' belongs.

Code generation by parity checks

Decoding Procedure (Continued...):

- For each c_i, let us find the error vector ϵ_i that satisfies $x' = c_i \oplus \epsilon_i$, that is, $\epsilon_i = c_i \oplus x'$.
- The error vectors ϵ_is form the set $E = C \oplus x'$. Because C is a subgroup of the group, $G = \langle \{ \text{ all } n\text{-tuples } \}, \oplus \rangle$, $C \oplus x'$ is a coset (right) of the group G.
- Thus, we wish to find ϵ, the n-tuple of least weight in the coset that contains x' (by the Maximum Likelihood method). This ϵ is called the "coset leader" for that coset.
- In summary,
  (i) Determine the coset to which the observed n-tuple x' belongs;
  (ii) Find the coset leader ϵ for that coset; and
  (iii) Decode x' as the n-tuple $x = x' \oplus \epsilon$.

Code generation by parity checks

Definition

For any observed n-tuple x', the *syndrome* of x' is the r-tuple $x'.H^t$, where r is the number of parity-check bits.

Theorem

Two n-tuples are in the same coset if and only if they have the same syndrome.

Code generation by parity checks

Problem:

Given the following 4×9 parity-check matrix H.

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

(a) Does its null space $N(H)$ have single-error correcting capability? Justify your answer.
(b) Encode the message tuple (1 1 0 1 0).
(c) Find the error, if any, in the tuple (0 1 0 1 1 1 0 0 1) and hence show that its syndrome is same as that of error tuple.

Code generation by parity checks

Solution:

Here $r = 4, n = 9, k = n - r = 5$.
(a) $N(H)$, the null space of H has single-error correcting capability, because H satisfies the following properties:
(i) No column of H is all 0's;
(ii) No two columns of H are identical;
(iii) at least three columns sum is 0, i.e., minimum weight is at least 3, since $\exists$

$$c_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, c_4 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, c_9 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ such that } c_1 \oplus c_4 \oplus c_9 = 0.$$

Code generation by parity checks

Solution (Continued...):

b) Here the message tuple is (1 1 0 1 0) = $\langle x_1, x_2, x_3, x_4, x_5 \rangle$. H is of the form $[P|I_r]$, where P is an 4×5 matrix and I_4 is the identity matrix. Let the encoded message tuple be $y = \langle y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9 \rangle$.

Set $y_1 = x_1 = 1$;
$y_2 = x_2 = 1$;
$y_3 = x_3 = 0$;
$y_4 = x_4 = 1$;
$y_5 = x_5 = 0$.

The parity-check equations are given by

$y_1 \oplus y_2 \oplus y_4 \oplus y_6 = 0 \Rightarrow y_6 = 1$;
$y_1 \oplus y_4 \oplus y_5 \oplus y_7 = 0 \Rightarrow y_7 = 0$;
$y_2 \oplus y_3 \oplus y_5 \oplus y_8 = 0 \Rightarrow y_8 = 1$;
$y_3 \oplus y_4 \oplus y_9 = 0 \Rightarrow y_9 = 1$.

Hence, the encoded message is $\langle$ 1 1 0 1 0 1 0 1 1 $\rangle$.

Code generation by parity checks

Solution (Continued...):

(c) The observed received tuple is $x' = \langle\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 0\ 1 \rangle$. The error syndrome is $x'.H^t = \langle\ 1\ 0\ 0\ 0\ \rangle$. Thus, there is a single error at $(1\,0\,0\,0)_2 =$ 8-th position of x'. Hence, the decoded tuple is $x = x' \oplus \epsilon = \langle\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ \rangle$, by simply flipping the 8-th bit position of x'. $\square$

Code generation by parity checks

Problem: Let H be an $r \times (2^r - 1)$ parity-check matrix for a Hamming code for which the i-th column is the binary representation of the integer i. Let H' be created from H by appending a row of all 1s. Show that the null space of H' is a group code with minimum distance 4.

Solution: Here H has the following form

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

where i-th column of H is the binary representation of the integer i.

Code generation by parity checks

Solution (Continued...): Now, H' will have the following form

$$H' = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 1 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 1 & 1 & 1 & 1 & \cdots & 1 \end{pmatrix},$$

where the last row of H is appended with all 1s.

Code generation by parity checks

Solution (Continued...): $N(H')$ is a group code with minimum distance 4, since

- No column of H' is all 0s;
- No two columns are identical;
- There does not exist three columns of H', whose sum is 0; and
- There exists four columns C_2, C_3, C_4, C_5 such that $C_2 \oplus C_3 \oplus C_4 \oplus C_5 = 0$.

End of this lecture