DECODING

Proposition (1) (# of errors a code can detect) *Let C be an $[n,k,d]$-code. Then C can detect $\leq s$ errors if $d \geq s+1$.*

Proof: Assume $d \geq s+1$. If a codeword $c \in C$ is sent and s or fewer errors occur then the received message r cannot be a codeword because if $r \in C$ then we must have $d_H(c,r) \geq s+1$. $\square$

Proposition (2) (# of errors a code can correct) *Let C be an $[n,k,d]$-code. Then C can correct $\leq t$ errors if $d \geq 2t+1$.*

Proof: Suppose that $d \geq 2t+1$. Assume that the codeword c is sent and the received word r has $\leq t$ errors, i,e., $d_H(c,r) \leq t$. We will show that if $c_1 \in C$ is any other codeword then $d_H(c_1,r) \geq t+1$. Assume that $d_H(c_1,r) \leq t$. It follows from the definition of d (the minimal Hamming distance between distinct codewords in C) and the triangle inequality for Hamming distance that

$$2t+1 \leq d \leq d_H(c,c_1) \leq d_H(c,r) + d_H(r,c_1) \leq 2t.$$

This is a contradiction, so we must have $d_H(c_1,r) > t$. It follows that c is the unique codeword $x \in C$ which satisfies $d_H(x,r) \leq t$, and we can correct the error by replacing r with the codeword with closest Hamming distance to r. $\square$

SYNDROMES

Let $G = \big(I_k,\ P\big)$ be a generator for an $[n,k]$-bilinear code (which we denote by C) where I_k is the $k \times k$ identity matrix and and P is a $k \times (n-k)$ matrix with entries in $\mathcal{F}_2$. Let

$$H = \big(-P^T,\ I_{n-k}\big)$$

be the check matrix for C where P^T denotes the transpose of the matrix P.

Definition (Syndrome) *Let $u \in \mathbb{F}_2^n$. Then the syndrome of u (denoted $S(u)$) is defined to be* $\boxed{S(u) := u \cdot H^T.}$

Remark: The most important property of syndromes is that $S(u) = 0$ if and only if $u \in C$.

Examples of Syndromes: Let

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

be the generator and check matrix for a [5,3]-code.

Consider the following vectors in $\mathbb{F}_2^5$.

$$u_1 = 01010, \qquad u_2 = 10010, \qquad u_3 = 11100.$$

To compute the syndromes of u_1, u_2, u_3 we first write down the transpose of the matrix H given by $H^T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then

$$S(u_1) = u_1 \cdot H^T = (0,1), \qquad S(u_2) = u_2 \cdot H^T = (0,0), \qquad S(u_3) = u_3 \cdot H^T = (0,1).$$

<u>We see that only u_2 is a codeword.</u>

COSETS OF THE VECTOR SPACE $\mathbb{F}_2^n$

Definition (Coset of $\mathbb{F}_2^n$) *Let C be an $[n,k]$-code. Let $u \in \mathbb{F}_2^n$. Then the set of vectors $u + \mathbb{C} := \{u + c_1, u + c_2, \quad \ldots, \quad u + c_N\}$ is called a coset of $\mathbb{F}_2^n$.*

Remark: Since C is an $[n,k]$-code it is clear that $N = 2^k$.

Remark: The trivial coset is C itself when $u = 000\ldots 0$.

Examples of cosets: Let $C = \{00000, 10010, 01011, 00100, 11001, 01111, 10101, 11101\}$ be the [5,3]=code with generator matrix $G = \begin{pmatrix} 1&0&0&1&0 \\ 0&1&0&1&1 \\ 0&0&1&0&0 \end{pmatrix}$.

- *When $u = 00000$, then C is the trivial coset.*
- *When $u = 10000$ then $u + C = \{10000, 00010, 11011, 10100, 01001, 11111, 00101, 01101\}$.*
- *When $u = 01000$ then $u + C = \{01000, 11010, 00011, 01100, 10001, 00111, 11101, 10101\}$.*

Table of all the distinct cosets of the above code C:

Each coset must have the same number of elements as the code C. In the above, each coset has 8 vectors. Since $\mathbb{F}_2^5$ has 32 vectors it follows that there must be exactly 4 cosets. We have found already found 3 of the cosets. To determine the remaining one we note that 00001 is not in any of these 3 cosets, so we can take $u = 00001$ to determine the fourth coset. We now list all 4 cosets.

- $00000 + C = \{00000, 10010, 01011, 00100, 11001, 01111, 10101, 11101\}$
- $10000 + C = \{10000, 00010, 11011, 10100, 01001, 11111, 00101, 01101\}$.
- $01000 + C = \{01000, 11010, 00011, 01100, 10001, 00111, 11101, 10101\}$.
- $00001 + C = \{00001, 10011, 01010, 00101, 11000, 01110, 10100, 11100\}$

Definition (Coset Leader) *The coset leader of a coset is an element of the coset with minimal Hamming weight (it has the fewest number of ones).*

Remark: In the example above the vectors 00000, 10000, 01000, 00001 are all coset leaders.

Remark: Coset leaders do not have to be unique. Note that in the coset $10000 + C$ we could also choose 00010 to be a coset leader.

How to construct all the cosets:

Step 1: The trivial coset is just the code C itself.

Step 2: Choose a vector $u_1 \in \mathbb{F}_2^n$ which is not in C and has smallest Hamming weight. Construct the coset $u_1 + C$. Note: *the vector u_1 may not be unique.*

Step 2: Choose a vector $u_2 \in \mathbb{F}_2^n$ which is not in C or $u_1 + C$ with minimal Hamming weight. Construct the coset $u_2 + C$.

Step 3: Assuming the cosets C, $u_1 + C$, $\ldots u_k + C$ are found. Choose a vector $u_{k+1} \in \mathbb{F}_2^n$ which is not in any of C, $u_1 + C$ $\ldots$, $u_k + C$ with minimal Hamming. Construct the coset $u_{k+1} + C$.

Step 3: Keep repeating the above procedure until all the cosets are found.

Theorem: *Let C be an $[n,k]$-code. Let $u \neq u'$ be vectors in $\mathbb{F}_2^n$. Then either $u + C = u' + C$ or the two cosets $u + C$, $u' + C$ have no elements in common.*

Proof: Assume some vector $u + c$ in the coset $u + C$ equals some vector $u' + c'$ in the coset $u' + C'$. This implies that $u + c + C = u' + c' + C$. But $u + c + C = u + C$ and $u' + c' + C = u' + C$. So the two cosets are the same. $\square$

Remark: The above theorem guarantees that the method to construct all the cosets on the previous page has to work.

An immediate corollary of the above theorem is the following.

Corollary: *Two vectors $u, v \in \mathbb{F}_2^n$ belong to the same coset if and only if they have the same syndrome, i.e., $S(u) = S(v)$.*

Proof: It follows from the above theorem that two vectors $u, v \in \mathbb{F}_2^n$ belong to the same coset if and only if $u - v \in C$. This implies that

$$0 = S(u - v) = (u - v) \cdot H^T = u \cdot H^T - v \cdot H^T = S(u) - S(v). \quad \square$$

Definition (Syndrome of a Coset) The syndrome of a coset $u + C$ is defined to be $S(u)$.

Remark: This definition is well defined since every vector in the coset has the same syndrome. In particular, even if we have a different coset leader the syndrome will be the same.

SYNDROME OR NEAREST NEIGHBOR DECODING

Let C be an $[n,k]$ bilinear code. Assume that a codeword $c \in C$ is transmitted across a noisy channel and received as $r \in \mathbb{F}_2^n$.

Syndrome Decoding Protocol:

- **Precomputation:** *Make a table of coset leaders and their syndromes.*
- **Step (1)** *Compute the syndrome $S(r)$ of the received vector r.*
- **Step (2)***Find the coset leader u with the same syndrome as S(r).*
- **Step (3)** *Decode r as $r - u$.*

EXAMPLE (Syndrome Decoding in the Hamming [7,4]-code)

The Hamming [7,4]-code (let's denote it as $\boldsymbol{\mathcal{C}}$) has 16 codewords

$$\begin{aligned}\boldsymbol{\mathcal{C}} = \big\{&0000000, 1000110, 0100101, 0010011,\\ &0001111, 1100011, 1010101, 1001001,\\ &0110110, 0101010, 0011100, 1110000,\\ &1011010, 1101100, 0111001, 1111111\big\}.\end{aligned}$$

We see that every non-zero codeword has at least 3 one's in it. This tells us that $d_H\big(\boldsymbol{\mathcal{C}}\big) = 3$. It then follows from Propositions 1, 2 that $\boldsymbol{\mathcal{C}}$ can detect up to 2 errors and correct exactly one error.

Step (1) The first step in syndrome decoding is to make a table of coset leaders and their syndromes. Note that there will be exactly 8 cosets.

Coset Leader	Syndrome
0000000	000
0000001	001
0000010	010
0000100	100
0001000	111
0010000	011
0100000	101
1000000	110

Assume the sender transmits the codeword c= 1100011 and it is received as r = 0100011.

Step (2) We compute $S(r) = r \cdot H^T = (0,1,0,0,0,1,1) \cdot \begin{pmatrix} 1&1&0 \\ 1&0&1 \\ 0&1&1 \\ 1&1&1 \\ 1&0&0 \\ 0&1&0 \\ 0&0&1 \end{pmatrix} = (1,1,0)$. We see that the coset leader $u = 1000000$ has the same syndrome 110.

Step (3) We decode r as $r - u = 0100011 - 1000000 = 1100011$.

Assume the sender transmits the codeword c= 0100101 and it is received as r = 0110101.

Step (2) We compute $S(r) = r \cdot H^T = (0,1,1,0,1,0,1) \cdot \begin{pmatrix} 1&1&0 \\ 1&0&1 \\ 0&1&1 \\ 1&1&1 \\ 1&0&0 \\ 0&1&0 \\ 0&0&1 \end{pmatrix} = (0,1,1)$. We see that the coset leader $u = 0010000$ has the same syndrome 011.

Step (3) We decode r as $r - u = 0110101 - 0010000 = 0100101$.

DUAL CODES

Definition (Dot Product) *The dot product of two vectors* $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n)$ *which are in* $\mathbb{F}_2^n$ *is defined to be:* $u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$.

Definition (Dual Code) *Let* C *be an* $[n, k]$ *bilinear code. The dual code (denoted* $C^{\perp}$*) is defined as the set of all vectors* $u \in \mathbb{F}_2^N$ *satisfying* $u \cdot c = 0$ *for all* $c \in C$.

Proposition *Let* C *be an* $[n, k]$ *bilinear code with generating matrix* $G = (I_k, P)$ *and check matrix* $H = \left(-P^T, I_{n-k}\right)$. *Then the dual code* $C^{\perp}$ *has generating matrix* H *and check matrix* G.

Proof: See page 415 in Introduction to Cryptography with Coding Theory.