Lecture: Decoding linear codes (Contd.)

From Lecture 3, recall that the cosets of $\mathcal{L}$ in $\mathbb{F}_2^n$ partition $\mathbb{F}_2^n$; furthermore, each coset has size exactly $|\mathcal{L}| = 2^k$. There are hence exactly $2^n/2^k = 2^{n-k}$ cosets of $\mathcal{L}$ in $\mathbb{F}_2^n$.

Example: Consider the (systematic) linear code $\mathcal{L}$ generated by

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

Clearly, we have $\dim(\mathcal{L}) = 2 = k$. The cosets of $\mathcal{L}$ in $\mathbb{F}_2^n$ can be written as follows:

$\mathcal{L}\ (= 0 + \mathcal{L})$	00000	01001	10101	11100
$00001 + \mathcal{L}$	00001	01000	10100	11101 ← Two candidate coset leaders!
$00010 + \mathcal{L}$	00010	01011	10111	11110
$00011 + \mathcal{L}$	00011	01010	10110	11111
$00100 + \mathcal{L}$	00100	01101	10001	11000
$00101 + \mathcal{L}$	00101	01100	10000	11001
$00110 + \mathcal{L}$	00110	01111	10011	11010
$00111 + \mathcal{L}$	00111	01110	10010	11011

Standard array table for $\mathcal{L}$

A coset leader of a coset is any vector of minimum Hamming weight in that coset (see above).

Given such a standard array, the MDD picks the coset that is $y+\mathcal{C}$ and simply returns $y - \hat{e}_{\mathcal{C}}$, where $\hat{e}_{\mathcal{C}}$ is the (now fixed) coset leader of $y+\mathcal{C}$.

Example: In the standard array above, suppose that $y = 01011$ was received. This corresponds to the coset $00010 + \mathcal{C}$. The MDD thus returns $01011 + \underbrace{00010}_{\text{coset leader}} = 01001 \in \mathcal{C}$.

Remark: By properties of cosets, it is clear that $\text{MDD}(y)$ always lies in $\mathcal{C}$.

The complexity of MDD is precisely the complexity of storing and searching the standard array, which is $\Theta(2^n)$. [This is bad (high complexity)!]

Returning to our motivation of decoding over the BSC(p), it follows that the probability of correct decoding by the MDD is precisely the probability that the "true" error vector is one of the coset leaders, i.e.,

$$\Pr[\text{MDD}(y) = \underline{c} \mid \underline{c} \text{ transmitted}] = \Pr[y + \underline{c} = \text{coset leader} \mid \underline{c} \text{ transmitted}]$$

$$= \sum_{w=0}^{n} N(w) \cdot p^w (1-p)^{n-w}, \qquad (1)$$

where $N(w)$ = # coset leaders of Hamming weight w.

HW: Justify the equality in (1) above and compute this expression for the standard array above.

HW: Suppose that $\mathcal{C}$ is an $[n, k, d]_2$ code. Argue that in any coset with a sequence of weight $\left\lfloor \frac{d-1}{2} \right\rfloor$ or less, this sequence is the unique coset leader.

Reducing the decoding complexity via syndromes

The standard array allows one to perform ML (or MAP) decoding over a BSC(p) ($p < 1/2$). However, such a procedure is computationally expensive.

In what follows, we argue that one can reduce the complexity of MDD to $\Theta(2^{n-k})$, via a clever use of linearity of the code $\mathcal{C}$.

Note that given y, we are interested in finding the coset leader of $y + \mathcal{C}$. Can we somehow succinctly store (coset identifier, coset leader)?

Def. 1: For a fixed parity-check matrix H of $\mathcal{C}$, the syndrome of $y \in \mathbb{F}^n$ is the vector $\underline{s} = Hy^T$.

Proposition 1: Two vectors $y_1, y_2 \in \mathbb{F}^n$ are in the same coset of $\mathcal{C}$ iff they have the same syndrome.

Proof: y_1, y_2 in the same coset of $\mathcal{C}$ $\Leftrightarrow$ $y_1 - y_2 \in \mathcal{C}$

$\Leftrightarrow$ $H(y_1 - y_2)^T = 0$

$\Leftrightarrow$ $Hy_1^T = Hy_2^T$. ▨

The above proposition gives rise to the following version of MDD with reduced complexity, called "syndrome decoding".

Step 0 (offline): Store the table of (syndrome, coset leader).

Step 1: Compute the syndrome $\underline{s} = Hy^T$, given y.

Step 2: Obtain the coset leader $\underline{e}$ corresponding to $\underline{s}$ (from Step 0)

Step 3: Return $\hat{\underline{c}} = y - \underline{e}$.

HW: Find the syndromes corresponding to each coset in the standard array above.