Lecture 8 : Operations on Codes and Weight Enumerators

In this lecture, we shall discuss how one can generate new linear codes old. Several exercises ahead!

① Puncturing codes

Def 1: Let $\mathcal{C}$ be an $[n,k,d]_q$ code. The punctured code $\mathcal{C}^{*,i}$ is obtained by deleting symbol i in every codeword.

- If G is a generator matrix for $\mathcal{C}$, then a generator matrix $G^{*,i}$, for $\mathcal{C}^{*,i}$, is obtained by deleting the i^{th} column in G and removing any possible repeated rows.

Theorem 1: For an $[n,k,d]_q$ code $\mathcal{C}$,

(a) If $d>1$, then $\mathcal{C}^{*,i}$ is an $[n-1,k,d^{*,i}]$ code, where $d^{*,i}=d-1$, if $\mathcal{C}$ has a min. wt. codeword with a non-zero i^{th} coordinate, and $d^{*,i}=d$, otherwise.

(b) If $d=1$, then $\mathcal{C}^{*,i}$ is an $[n-1,k^{*,i},d^{*,i}]$ code, where $k^{*,i}=k$ and $d^{*,i}=1$, if $\mathcal{C}$ has no codeword of wt. 1 with non-zero entry in coordinate i, and $k^{*,i}=k-1$ (for $k>1$) and $d^{*,i}\geq 1$, otherwise.

Proof: (a) If $d>1$ and $\mathcal{C}$ has a min wt. codeword with a non-zero i^{th} co-ord., then after puncturing, $\exists$ codeword in $\mathcal{C}^{*,i}$ with wt. $d-1$.

If $\mathcal{C}$ has no min. wt. codeword with a non-zero i^{th}-co-ord., then, all codewords in $\mathcal{C}^{*,i}$ have weight $\geq d$, & $\exists$ codeword of wt. exactly d.

Moreover, since $d>1$, no two rows in $G^{*,i}$ are identical, thereby ensuring that $|\mathcal{C}^{*,i}| = q^k$.

(b) Exercise. $\boxtimes$

HW: Let $\mathcal{C}$ be the $[4,2,1]$ binary code with generator matrix

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

Consider the code $\mathcal{C}^{*,1}$. What are its parameters?

② Extending codes

Def 2: If $\mathcal{C}$ is an $[n,k,d]_q$ code, the extended code $\hat{\mathcal{C}}$ is

$$\hat{\mathcal{C}} = \left\{ (c_1, c_2, \ldots, c_{n+1}) : (c_1, \ldots, c_n) \in \mathcal{C} \text{ and } \sum_{i=1}^{n+1} c_i = 0 \right\}.$$

Theorem 2: If $\mathcal{C}$ is a binary code, then $\hat{\mathcal{C}}$ is an $[n+1, k, \hat{d}]$ code, where

$$d = \begin{cases} d, & \text{if } \hat{d} \text{ is even,} \\ d+1, & \text{o.w.} \end{cases}$$

Proof: Exercise.

HW: (i) Obtain a generator matrix $\hat{G}$ for $\hat{\mathcal{C}}$ using a given generator matrix G for $\mathcal{C}$.

(ii) Argue that

$$\hat{H} = \left[\begin{array}{ccc|c} 1 & 1 \cdots & 1 & 1 \\ \hline & & & 0 \\ & H & & \vdots \\ & & & 0 \end{array}\right]$$

is a parity-check matrix for $\hat{\mathcal{C}}$, where H is a parity-check matrix for $\mathcal{C}$.

(iii) Consider the code $\mathcal{C}$ with generator matrix

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

Provide a generator matrix for the code obtained by first puncturing the second co-ordinate & then via an extension on the last coordinate.

③ Shortening codes:

Let $T \subseteq [n]$. Given an $[n, k, d]_q$ code $\mathcal{C}$, consider the set $\mathcal{C}(T)$ of codewords that are $\underline{0}$ in the locations in T.

Def 3: The code $\mathcal{C}_T$ that is the code $\mathcal{C}$ shortened on T is obtained by puncturing $\mathcal{C}(T)$ in the coordinates in T.

The dimension of a shortened code depends on $|T|$ and its relationship with $d_{min}(\mathcal{C}^{\perp})$. We shall not delve into greater detail on this; for more, see [Huffman-Vera-Pless, Thm. 1.5.7].

HW*: Let $\mathcal{C}$ be an $[n,k,d]_q$ code. Argue that any $n-d+1$ coordinates of $\mathcal{C}$ contains an information set.

[Hint: Consider a generator matrix G of $\mathcal{C}$ and pick any s coordinates of $\mathcal{C}$, giving rise to the submatrix G_s. If these coordinates do not contain an information set, then $\exists$ a lin. combination of rows of G_s giving rise to $\underline{0}$. Obtain an upper bound on d via s.].

HW: Given codes $\mathcal{C}_1, \mathcal{C}_2$, with parameters $[n_1,k_1,d_1]_q$ and $[n_2,k_2,d_2]_q$, respectively, their direct sum is defined as:

$$\mathcal{C}_1 \oplus \mathcal{C}_2 = \{(\underline{c}_1, \underline{c}_2) : \underline{c}_1 \in \mathcal{C}_1,\ \underline{c}_2 \in \mathcal{C}_2\}.$$

Obtain generator and parity-check matrices for $\mathcal{C}_1 \oplus \mathcal{C}_2$, given that G_i, H_i, respectively, are generator & parity-check matrices for $\mathcal{C}_i$, $i \in \{1,2\}$.

What is the minimum distance of $\mathcal{C}_1 \oplus \mathcal{C}_2$?

HW: Let $\mathcal{C}$ be the binary code with generator matrix

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

Via row operations, argue that $\mathcal{C}$ is a direct sum of two binary codes; find generator matrices for these codes.

Weight Distribution of Codes

Def 4: Let $\mathcal{C}$ be an $[n,k,d]_q$ linear code. We define, for $w \in [0:n]$,

$$A_w \triangleq \left| \{ \underline{c} \in \mathcal{C} : wt(\underline{c}) = w \} \right|,$$

to be the "weight enumerator" of $\mathcal{C}$ at the weight w.

The collection $(A_0, A_1, \ldots, A_n)$ is called the "weight distribution" of $\mathcal{C}$.
Note that $A_0 = 1$ and $A_i = 0$, for $i \in [1:d-1]$.

Weight enumerators of important code families are extensively studied in the literature.

Motivation for studying weight enumerators:

① Error detection: Suppose that a codeword of an $[n,k,d]_q$ linear code $\mathcal{C}$ is transmitted over a BSC(p). We define the "undetected error probability" as the average (over codewords) probability that the received sequence is a codeword that is different from the transmitted codeword, i.e.,

$$P_{ud} \triangleq \frac{1}{|\mathcal{C}|} \cdot \sum_{\underline{c} \in \mathcal{C}} \Pr_{\underline{Y}}\left[\underline{Y} = \underline{c}', \text{ for some } \underline{c}' \neq \underline{c}\right]$$

$$= \frac{1}{|\mathcal{C}|} \cdot \sum_{\underline{c} \in \mathcal{C}} \cdot \sum_{\underline{y}} p^{d(\underline{y},\underline{c})}(1-p)^{n-d(\underline{y},\underline{c})} \cdot \mathbb{1}\{\underline{y} \in \mathcal{C}, \underline{y} \neq \underline{c}\}$$

$$= \frac{1}{|\mathcal{C}|} \cdot \sum_{\underline{c} \in \mathcal{C}} \sum_{w=d}^{n} A_w \cdot p^w (1-p)^{n-w}$$

$$= \sum_{w=1}^{n} A_w\, p^w (1-p)^{n-w}.$$

② ML error estimation: Consider the setting where a codeword $\underline{c} \in \mathcal{C}$ is transmitted over a BSC(p) and decoded via the MAP/ML decoder.

FACT: The probability of error under ML decoding can be bounded as:

$$P_{e,ML} \leq \sum_{w=1}^{n} A_w \left(Z(p)\right)^w,$$

where $Z(p) \triangleq 2\sqrt{p(1-p)}$.