# Lecture 14: Linear Codes: Examples and Properties

# Space

- Given a field $(\mathbb{F}, +, \cdot)$
- We consider the set of all $n$-tuples with entries in $\mathbb{F}$
- That is, we will consider the set $\mathbb{F}^n$
- The total number of elements in the set if $|\mathbb{F}|^n$

# Basic Terminology

- A code $\mathcal{C} \subseteq \mathbb{F}^n$
- A codeword $c = (c_1, \dots, c_n) \in \mathcal{C}$ such that all $c_1, \dots, c_n \in \mathbb{F}$
- Block length is $n$
- Weight of a codeword: $\mathrm{wt}(c) = \left|\{i : c_i \neq 0\}\right|$
- (Hamming) Distance $d_H(c, c') = \mathrm{wt}(c - c')$
- Distance of a code: $d(\mathcal{C}) = \min_{\substack{c, c' \in \mathcal{C} \\ c \neq c'}} \mathrm{wt}(c - c')$
- $(N, K, d)$-code: $|\mathcal{C}| = K$, $|\mathbb{F}|^n = N$ and $d(\mathcal{C}) = d$

# Intuition

- Given fixed $\mathbb{F}$ and $n$
- We want to maximize $|\mathcal{C}|$ and $d(\mathcal{C})$
  - $|\mathcal{C}|$ determines how much information can be transmitted over the channel, and
  - $d(\mathcal{C})$ determines the robustness of the encoding (because, to force the maximum likelihood decoding algorithm to output an incorrect codeword, the channel needs to introduce at least $\lceil d(\mathcal{C})/2 \rceil$ errors)
- We will see later that these two parameters are conflicting and there is a trade-off of these two parameters

- Linear Code: If $\mathcal{C}$ is a vector subspace of $\mathbb{F}^n$
- Suppose $(c_1, \ldots, c_k)$ is a basis of the vector subspace $\mathcal{C}$
- $[n, k, d]_{\mathbb{F}}$ code: A code $\mathcal{C}$ that is a vector subspace of $\mathbb{F}^n$, of dimension $k$, and $d(\mathcal{C}) = d$
- The generator matrix $G$ of a code $\mathcal{C}$ is defined to a matrix in $\mathbb{F}^{k\times n}$ as defined below.

$$G = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

- Note that $\begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_k \end{pmatrix} \cdot G$ generates all codewords in $\mathcal{C}$, where $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$

# Distance of a Linear Code

## Claim

$$d(\mathcal{C}) = \min_{c \in \mathcal{C}} \mathrm{wt}(c)$$

## Proof.

- Let $d(\mathcal{C})$ is realized by the distance between the codewords $c$ and $c'$
- Note that $c - c'$ is also a codeword (because $\mathcal{C}$ is a vector space)
- Note that $\mathrm{wt}(c - c') = d(\mathcal{C})$
- If there exists $\widetilde{c}$ such that $\mathrm{wt}(\widetilde{c}) < d(\mathcal{C})$ then $d_H(0, \widetilde{c}) < d(\mathcal{C})$ (which is a contradiction)
- Therefore, we have the claim

□

# Example:Repetition code

- The repetition code $\{0^n, 1^n\}$ has generator matrix

$$\begin{pmatrix} \overbrace{1 \quad 1 \quad \cdots \quad 1}^{n\text{-times}} \\ \end{pmatrix}$$

- It is an $[n, 1, n]_2$ code

# Equivalent Code

- Let $G$ be the generator matrix of a code $\mathcal{C}$
- Let $G'$ be the generator matrix obtained by replacing the row $G_{i,*}$ in the matrix $G$ by the row $\alpha G_{i,*}$, for $\alpha \in \mathbb{F}^*$, then $G'$ generates the same code as $G$
- Let $G'$ be the generator matrix obtained by replacing the row $G_{i,*}$ in the matrix $G$ by $G_{i,*} + \alpha G_{j,*}$, for $i \neq j$ and $\alpha \in \mathbb{F}^*$, then $G'$ generates the same code as $G$
- We write $G \equiv G'$

# Similar Code

- Suppose $G'$ is a generator matrix obtained by swapping two columns of the generator matrix $G$
- Then the code generated by $G'$ is similar to the code generated by $G$
- We write $G \sim G'$
- If $G$ is the generator matrix of an $[n, k, d]_{\mathbb{F}}$ code, then $G'$ also generates an $[n, k, d]_{\mathbb{F}}$ code (the codewords can be bijectively mapped where the mapping swaps the $i$-th and the $j$-th coordinate of the codeword)

# Systematic Form

- Let $G$ be a generator matrix of an $[n, k, d]_{\mathbb{F}}$ code
- Then there exists $P \in \mathbb{F}^{k\times(n-k)}$ such that

$$G \sim \left[ I_{k\times k} | \, P \right] ,$$

where $I_{k\times k}$ is the identity matrix of dimension $k \times k$

- Proof Outline: Row rank = Column rank, swap $k$ independent columns of $G$ into its first $k$ columns, and perform Gaussian Elimination

- Define the matrix $H$ in $\mathbb{F}^{(n-k)\times n}$ as follows

$$H = \left[ -P^\top \middle| I_{(n-k)\times(n-k)} \right]$$

**Claim**

*The inner product of any row $G_{i,*}$ and any row $H_{j,*}$ is always 0.*

**Proof.**

- Note that $G_{i,*} = (\delta_i, P_{i,*})$
- Note that $H_{i,*} = \left(-P^\top_{*,j}, \delta_j\right)$
- Their inner product is $-P_{i,j} + P_{i,j} = 0$

□

**Claim**

*Every codeword in the code generated by $G$ is orthogonal to every codeword in the code generated by $H$*

**Proof.**

- A codeword in the code generated by $G$ looks like $\sum_{i=1}^{k} \alpha_i \cdot G_{i,*}$, for $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$
- A codeword in the code generated by $H$ looks like $\sum_{j=1}^{n-k} \beta_j \cdot H_{j,*}$, for $\beta_1, \ldots, \beta_{n-k} \in \mathbb{F}$
- Now, we have

$$\left\langle \sum_{i=1}^{k} \alpha_i G_{i,*}, \sum_{j=1}^{n-k} \beta_j H_{j,*} \right\rangle = \sum_{i=1}^{k} \sum_{j=1}^{n-k} \alpha_i \beta_j \langle G_{i,*}, H_{j,*} \rangle$$
$$= \sum_{i=1}^{k} \sum_{j=1}^{n-k} \alpha_i \beta_j \cdot 0 = 0 \qquad \square$$

- Let $\mathcal{C}$ be the code generated by $G$
- We denote the code generated by $H$ as $\mathcal{C}^\perp$ (dual of $\mathcal{C}$)
- Show that $\left(\mathcal{C}^\perp\right)^\perp = \mathcal{C}$
- We represent the $d(\mathcal{C}^\perp)$ by $d^\perp$
- Let $t(H)$ represent the minimum number of columns of $H$ that can be (non-trivially) linearly combined to yield the 0 column

**Claim**

$$d(\mathcal{C}) = t(H)$$

Proof is left as an exercise

# Hadamard Code

- The columns of the generator matrix $G$ has all binary strings of length $r$
- For example, for $r = 3$, we have

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

- This generates a $[2^r, r, 2^{r-1}]_2$ code (Prove this)

# Punctured Hadamard Code

- From the generator matrix of the Hadamard Code, we remove all those columns that have a 0 as their top-most entry
- For example, for $r = 3$, we have

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

- This generates a $[2^{r-1}, r, 2^{r-2}]_2$ code (Prove this)

# Simplex Code

- From the generator matrix of the Hadamard Code, we remove the all-0 column
- For example, for $r = 3$, we have

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

- This generates a $[2^r - 1, r, 2^{r-1}]_2$ code (Prove this)

**Claim**

*Let $G$ be the generator matrix of the Simplex Code. We have $t(G) = 3$.*

Prove this.

# Hamming Code

- Hamming Code is the dual of the Simplex code
- Therefore, it is a $[2^r - 1, 2^r - r - 1, 3]_2$ code (Prove this)
- Write down the generator matrix for $r = 3$ case