Lecture 17: Perfect Codes and Gilbert-Varshamov Bound

Maximality of Hamming code

Lemma

Let C be a code with distance 3, then:

$$|C| \leqslant \frac{2^n}{n+1}$$

- Codes that meet this bound: Perfect codes
- Hamming code is a perfect code

Hamming Bound

Lemma

Let C be a code with distance d, then:

$$|C| \leqslant \frac{2^n}{\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i}}$$

Proof: Think about $B(c, \lfloor \frac{d-1}{2} \rfloor)$, for any $c \in C$

Theorem (Tietavainen and van Lint)

There following are all the binary perfect codes:

- *Hamming code*
- *The* $[23, 12, 7]_2$ *Golay code*
- *Trivial codes* ($\{0\}$, $\{1^n, 0^n\}$ *for odd* n, $\{0,1\}^n$)

Dual Code

Definition (Dual Code)

For a linear code C, define

$$C^{\perp} := \{z \colon z \in \mathbb{F}_2^n, \forall c \in C \text{ we have } z^{\intercal} c = 0\}$$

- $(C^{\intercal})^{\intercal} = C$
- If H is the parity check matrix for C then H is the generator matrix for $C^{\intercal}$
- If $C^{\intercal} \subseteq C$ then C is called self-orthogonal
- If $C^{\intercal} = C$ then C is called self-dual

- Dual code of (generalized) Hamming code is Simplex code (it is $[2^r - 1, r]_2$ code)
- Add an all zero column of the parity check matrix of (generalized) Hamming code. The code generated by it is: Hadamard code (it is $[2^r, r]_2$ code). Its distance is 2^{r-1}.

Volume of a Ball

- Let $\text{Ball}_q(n, \ell)$ be the set of all elements in $\mathbb{F}_q^n$ with weight $\leqslant \ell$

Definition (Volume)

Size of $\text{Ball}_q(n, \ell)$ is:

$$\text{Vol}_q(n, \ell) := \sum_{j=0}^{\ell} \binom{n}{j} (q-1)^j$$

Definition (Largest Code)

The largest q-ary code of block length n and distance d is defined to have $A_q(n, d)$ codewords

Gilbert-Varshamov Bound

Lemma (Gilbert-Varshamov Bound)

$$A_q(n,d) \geqslant \frac{q^n}{\mathrm{Vol}_q(n,d-1)}$$

- For sets A, B, we define $A + B = \{a + b \colon a \in A, b \in B\}$
- Let $C = \emptyset$
- Greedily add to C any $c \in \mathbb{F}_q^n$ not covered in $C + \mathrm{Ball}_q(n, d-1)$
- If $|C| < A_q(n,d)$ then $|C + \mathrm{Ball}_q(n,d-1)| < q^n$ and there exists such c

Lemma (Gilbert-Varshamov Bound)

There exists a linear $[n, k]_q$ code C such that

$$k \geqslant \left\lfloor \log_q \frac{q^n}{\mathrm{Vol}_q(n, d-1)} \right\rfloor$$

- Suppose $C = \langle v_1, \dots, v_{k-1} \rangle$
- Define $S = C + \mathrm{Ball}_q(n, d-1)$
- If $\mathbb{F}_q^n \setminus S$ is non-empty, then choose v_k from it
- Note that $v_k \notin S$
- We want to claim:

Claim

For any $v \in C$ and $\alpha \in \mathbb{F}_q$, the codeword $v + \alpha v_k$ is not in S

- If the claim is true then we are done

Proof of the claim:

- Suppose there exists $v \in C, \alpha \in \mathbb{F}_q$ such that $v + \alpha v_k \in S$
- So, there exists $v' \in C$ such that: $\Delta(v + \alpha v_k, v') < d$
- Implies, $\Delta(\alpha v_k, (v' - v)) < d$
- Let $v'' = \alpha^{-1}(v' - v)$ and $v'' \in C$
- So, $v_k \in \{v''\} + \mathrm{Ball}_q(n, d-1) \subseteq S$, a contradiction

Entropy Function

Definition (Entropy Function)

$$h_q(x) = x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x)$$

- For $q = 2$, the binary entropy function $h(x) = -x\log x - (1-x)\log(1-x)$

Lemma

$$(h_q(p) - o(1))n \leqslant \log_q \mathrm{Vol}_q(n, pn) \leqslant h_q(p)n$$

Asymptotic GV Bound for Linear Codes

Theorem (Asymptotic GV Bound)

For every prime power q, $p \in (0,1)$ and $\varepsilon \in (0, 1 - h_q(p))$, there exists n_0 such that for all $n \geqslant n_0$ there exists an $[n,k,d]_q$ code where $d = pn$ and $k = (1 - h_q(p) - \varepsilon)n$. In fact, a random generator matrix $G \in \mathbb{F}_q^{k \times n}$ corresponds to such a code, except with probability $\exp(-\Omega(n))$.

Proof of Full Row Rank:

- Probability that the i-th row is in the span of previous $(i-1)$ rows: $q^{i-1}/q^n < q^{-(n-k)}$
- Probability that all rows are linearly independent (by union bound) $\leqslant kq^{-(n-k)} = \exp(-\Omega(n))$

Asymptotic GV Bound for Linear Codes (continued)

Proof of high distance:

- Linear Code has low distance if and only if there exists a low weight codeword
- For $S \subseteq [k]$, let $G_S := \oplus_{i \in S} G_i$, where G_i is the i-th row of the matrix
- Fix S and consider the random variable G_S
- Note that it is a uniform variable over $\mathbb{F}_q^n$ and the probability that G_S has weight $\leqslant \ell$ is $\mathrm{Vol}_q(n, \ell)/q^n$
- Therefore we have: $\Pr_G[G_S \in \mathrm{Ball}_q(n, d-1)] \leqslant q^{-(1-h_q(p))n}$
- Now, $\Pr_G[\exists S \colon G_S \in \mathrm{Ball}_q(n, d-1)] \leqslant q^k \cdot q^{-(1-h_q(p))n} \leqslant q^{-\varepsilon n}$

More Perspective

- Prove: Previous theorem also holds true for linear codes defined by choosing a random parity check matrix
- Prove: Previous theorem also holds true for random generator matrices in systematic form
- Think: Can we beat the GV bound using explicit constructions?