Today

Existence of asymptotically good codes.

- The Gilbert proof.
- The Varshamov proof.
- The Gilbert-Varshamov bound.
- Wozencraft's Ensemble.

Plan for the First Part of Course

- Will give some atomic "constructions" of codes.
- Then give some composition results - and that will give explicit constructions.
- But today: Exponential time/Randomized polytime constructions. Why?
  - Prove such codes exist.
  - Even better, randomized polynomial construction.
  - Gives target for deterministic (explicit) constructions.
  - Used in deterministic constructions.

The Gilbert Construction

- Exponential time, deterministic construction.
- Greedy algorithm:
  1. Initially $S \leftarrow \{0,1\}^n$, $C \leftarrow \emptyset$.
  2. While $S \neq \emptyset$ do
     (a) Pick any element $x \in S$;
     (b) $C \leftarrow C + \{x\}$, $S \leftarrow S - \mathcal{B}(x, d-1)$.
- Gives code of minimum distance d.
- Not linear.
- How many codewords? At least $2^n/\mathrm{Vol}_2(d-1,n)$.
- Will analyze quantitative results shortly.

The Varshamov Result

- Polynomial time, randomized construction, working with positive probability.
- Algorithm:
  1. Pick a $k \times n$ matrix G at random.
  2. Let $C = \{\mathbf{x} \cdot G | \mathbf{x}\}$.
- Claim: w.h.p. C has 2^k distinct elements. Furthermore, their pairwise distance is at least d provided $2^k - 1 < 2^n/\mathrm{Vol}_2(d-1,n)$.
- Proof:
  1. Suffices to verify that for every non-zero vector $\mathbf{x}$, $\mathbf{x} \cdot G$ is not in $\mathcal{B}(\mathbf{0}, d-1)$

2. Fix $\mathbf{x}$. $\mathbf{x}G$ is a random vector. Falls in $\mathcal{B}(\mathbf{0}, d-1)$ w.p. $\mathrm{Vol}_2(d-1, n)/2^n$.
3. By union bound prob. exists $\mathbf{x}$ such that $\mathbf{x}G \in \mathcal{B}(\mathbf{0}, d-1)$ is at most $(2^k - 1)\mathrm{Vol}_2(d-1, n)/2^n$. If this quantity is less than 1, then such a code exists. If much less, then found with high probability.

Gilbert-Varshamov bounds

- Famed phrase in coding theory.
- Non-asymptotic version: There exist $(n, k, d)_2$ codes with $2^k \geq 2^n/\mathrm{Vol}_2(d-1, n)$.
- Asymptotic version: For every R, δ such that
$$R < 1 - H(\delta)$$
there exists a family of codes with rate R and relative distance δ.

Reflections on the G-V bound

- Asymptotically good families exist!
- Striking similarity to Shannon's result. Coincidence? Shannon's result implicitly proved the GV bound (earlier than GV).
- Terminology: When "dealing with errors in inf. transmission" talk of the Shannon bound; When "combinatorics of minimum distance" we talk of GV bound.
- In other words: $\sup_{\mathcal{C}}\{R(\mathcal{C})\} \geq 1 - H(\delta)$.
- Contrast with Hamming (Volume) bound: $\sup_{\mathcal{C}}\{R(\mathcal{C})\} \leq 1 - H(\delta/2)$.

- Actually far from each other. Which one is right?

Is the GV bound tight?

- Reigning belief in coding community: GV bound is tight. ("Almost every code meets the GV bound, except the ones we know".)
- Applies only to:
  - Asymptotically good families of codes:
    - Hamming codes beat GV.
    - Hadamard codes beat GV.
    - BCH codes beat GV.
    - RS codes beat GV.
  - $q = 2$.
    - GV construction extends to $q > 2$. Roughly says for fixed $R, \delta > 0$ as $q \to \infty$. $\sup_{\mathcal{C}}\{R(\mathcal{C})\} \geq 1 - \delta - O(1/\log q)$
    - But there exist algebraic codes s.t. $R \geq 1 - \delta - O(1/\sqrt{q})$

- "Almost every code meets the GV bound, except the ones we know, which are better."?
- If counterexample exists, where could it be?
  - Try $\delta \to 0$. Codes with $R \geq 1 - \frac{1}{2}\delta \log \delta^{-1} - o(\delta)$ beat the bound. (Or any $\alpha < 1$.)
  - Some possibilities in the range $\delta \to \frac{1}{2}$.

More non-explicit constructions

- More randomness: Completely random code.
- Less randomness: Pick G to be Toeplitz.
- Wozencraft: Nice, with a slightly more explicit feel.

Wozencraft's Ensemble

- Basic idea: "Pack" $\{0,1\}^n$ with linear codes $C_1, \ldots, C_t$.
  - $C_i \cap C_j = \{\mathbf{0}\}$.
  - $\cup_i C_i = \{0,1\}^n$.
- $t \geq \mathrm{Vol}_2(d-1, n)$ implies some C_i has distance d. More strongly, $\epsilon t \geq \mathrm{Vol}_2(d-1, n)$ implies more than $1 - \epsilon$ fraction of C_i's have distance d.
  - Proof: Every point in $\mathcal{B}(\mathbf{0}, d-1)$ rules out one code C_i. But we have more codes than points in $\mathcal{B}(\mathbf{0}, d-1)$.
- $t =$?: Conditions imply $t = (2^n - 1)/(2^k - 1)$. So exist codes satisfying $2^n - 1 \geq$

$(2^k - 1) \cdot \mathrm{Vol}_2(d-1, n)$ provided we can "pack".

Wozencraft Packing

- Say $n = ck$.
  - View elements of $\mathbb{F}_2^k$ as elements of $\mathbb{F}_{2^k}$.
  - So message is one field element.
  - Encoding is c field elements.
  - Codes described by c field elements $\alpha_1, \ldots, \alpha_c$, not all zero, with first non-zero element being 1.
  - Claim: This Packs $\{0,1\}^n$.

Conclusion

- Can get very low-randomness constructions. Down to $O(n)$ randomness.
- Nothing explicit yet.
- GV bound very natural - comes up in so many ways.
- So it is right?
- When it fails, no intuition as to why it fails. No natural proofs of existence of Hamming codes, BCH codes, RS codes. etc.
- Next lecture: Explicit constructions of asymptotically good codes.