Coding Theory

Josh Brakensiek*

MOP 2017
Black group

1 Definitions (read as needed)

- An *alphabet* is any finite set Σ of characters. For example, $\Sigma = \{0, 1\}$, $\Sigma = \mathbb{F}_q$ (q prime power), or $\Sigma = \{M, O, P\}$. A *code* of length n is a subset $C \subseteq \Sigma^n$. Any element of C is a *codeword*. A code C is *linear* C, if Σ is a vector space, and C is a linear subspace of Σ^n.
- A *metric* is a function $d : \Sigma^n \times \Sigma^n \to [0, \infty)$ such that (1) $d(x, y) = 0$ if and only if $x = y$, (2) $d(x, y) = d(y, x)$ for all x and y, and (3) $d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in \Sigma^n$.
- The most common metric is the *Hamming distance* (d_{Ham}): the minimum number of letters needed to be changed to go from one string to another. For example, $d_{\text{Ham}}(0011, 0101) = 2$ and $d_{\text{Ham}}(\text{MOP}, \text{COW}) = 2$.
- The *distance* of a code is C with respect to a metric d is the minimum distance between two distinct code words.
- A *channel* is any "black box" which takes in codewords and outputs modified codewords For example the *one-bit deletion channel* takes a string and deletes exactly one bit. Thus, 01010 can become any of 1010, 0010, 0110, 0100, 0101. Often these are randomized and/or adversarial.

2 Constructing Classic Codes

1. (Hamming code) Let $n = 2^k - 1$ for some positive integer k. Find a linear code $C \leq \mathbb{F}_2^n$ with $|C| = 2^{n-k}$ and Hamming distance at least 3.
2. (Dual/Hadamard code) Find another linear code $C \leq \mathbb{F}_2^n$ with $|C| = 2^k$ (same n and k as the previous problem) but with Hamming distance at least $n/2$.
3. (Reed-Solomon) Let q be a prime power, and let n, k be positive integers such that $q \geq n \geq k$ (commonly $q = n$). Find a k-dimensional subset $C \leq \mathbb{F}_q^n$ such that $d_{\text{Ham}}(s, t) \geq n - k + 1$ for all $s \neq t \in C$. (Hint: consider polynomials in $\mathbb{F}_q[x]$.)
4. (BCH) Find a linear binary code (subspace of $\mathbb{F}_2^n$) of dimension at least $n-(n-k+1)\log_2(n+1)$ with Hamming distance at least $n - k + 1$. (Hint: modify the construction from the previous problem.)

*Contact: first dot last at gmail. Also special thanks to Venkatesan Guruswami, Sidhanth Mohanty and Ray Li.

3 Bounds on the Size of Codes

5. (Hamming/Gilbert-Varshamov) Construct $C \subseteq \Sigma^n$ with Hamming distance $d \leq n$ such that

$$|C| \geq \frac{|\Sigma|^n}{\sum_{i=0}^{d-1} (|\Sigma| - 1)^i \binom{n}{i}}.$$

6. (Singleton/Hamming) If $C \subseteq \Sigma^n$ has Hamming distance $d \leq n$ then

$$|C| \leq \min\left(|\Sigma|^{n-d+1}, \frac{|\Sigma|^n}{\sum_{i=0}^{\lfloor (d-1)/2 \rfloor} (|\Sigma| - 1)^i \binom{n}{i}}\right).$$

   (As a corollary, the Reed-Solomon code is optimal for its size and distance.)

7. (Shannon) The MAA has started a new binary string transfer service. The only catch is that with probability $p \in (0, 1/2)$, each bit will be uniformly at random flipped from 0 to 1 or 1 to 0. If Becky wants to send to Po-Shen an n-bit message using the service, what is (asymptotically) the minimum number of bits that Becky must transfer in order for Po-Shen to recover her original message with probability .999? (Assume both have unlimited computing power.)

4 Further Challenges

8. (Locally recoverable codes: Goplan, Huang, Simitci, Yekhanin) Assume $t \ll k < n$. A code $C \subseteq \{0, 1\}^n$ of size 2^k has the property that single bit flips are t-locally recoverable: for any $i \in \{1, \ldots, n\}$ there exists $S_i \subseteq \{1, \ldots, n\}$ with $|S_i| \leq t$, such that the ith bit is a function of the bits at the indices of S_i. (For example, if $i = 1$ and $S_i = \{2, 3\}$, then one should be able to figure out what the first bit is by only looking and that second and third bits.)

   Prove that the hamming distance of C is at most $n - k - \lceil \frac{k}{t} \rceil + 2$.

9. (Varshamov-Tenegolts) Alice wants to pick a code $A \subseteq \{0, 1\}^n$ as large as possible for the one-bit deletion channel (see definitions). That is, for any distinct $a_1, a_2 \in A$, an adversary (Eve) cannot delete one bit from a_1 and one bit from a_2 to get the same string.

   Show that Alice can have $|A| = \Theta(2^n/n)$, but not better. (Hint: consider $x_1 + 2x_2 + \cdots + nx_n = 0 \mod (n + 1)$.)

10. (ϵ-balanced) Let n, k be a positive integers, an ϵ-*balanced code* is a linear binary code of dimension k, Hamming distance $(1/2 - \epsilon)n$, and every codeword has Hamming weight in the range $((1/2 - \epsilon)n, (1/2 + \epsilon)n)$

    (a) Show that if and ϵ-balanced code exists with parameters n and k, there exists $S \subseteq \{0, 1\}^k$ of size n which is ϵ-*biased*: for any $v \in \{0, 1\}^k$,

$$\left|\Pr_{s \in S}[v \cdot s] - \frac{1}{2}\right| < \epsilon,$$

    where $v \cdot s = v_1 s_1 + \cdots + v_n s_n \mod 2$.

(b) Show that there exists an infinite family of ϵ-balanced codes with $n \leq O(\frac{k}{\epsilon^2})$.

(c) (Alon, Goldreich, Håstad, Peralta) Explicitly construct an infinite family of ϵ-balanced codes with $n \leq O(\frac{k^2}{\epsilon^2})$.

(d) (Ta-Shma: *Very* hard) Explicitly construct an infinite family of ϵ-balanced codes with $n \leq O(\frac{k}{\epsilon^{2+o(1)}})$.

5 Open Problems

11. (Chee, Kiah, Ling, Nguyen, Vu, Zhang) A permutation $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is *short* if for all $i \in \{1, \ldots, n\}$, $|i - \pi(i)| \leq 1$. (Note that the identity permutation is short.) A permutation π acts on a binary string of length $s \in \{0,1\}^n$ so that s_i is sent to the $\pi(i)$th position (denoted by $\pi(s)$).

    Find the largest possible subset $A \subseteq \{0,1\}^n$ such for any two distinct $a_1, a_2 \in A$ there do not exist short permutations π_1, π_2 such that $\pi_1(a_1) = \pi_2(a_2)$. The state of the art is

$$\Omega(2^{.643n}) \leq |A| \leq O(2^{2n/3}).$$

12. (Guruswami) The 2-bit deletion channel takes any n-bit string and outputs any $(n-2)$-bit string with 2 characters deleted.

    (a) (Not open) There exists a code for the 2-bit deletion channel with $\Omega(2^n/n^{10})$ codewords.

    (b) Find an *explicit*[1] example of such a code from part (a).

6 Further Reading

- Lecture notes from CMU.

  `https://www.cs.cmu.edu/~venkatg/teaching/codingtheory-au14/`

- Whole book on coding theory.

  `http://www.cse.buffalo.edu/faculty/atri/courses/coding-theory/book/`

[1]Explicit can take on a variety of meanings. Here it means "can be described on the back of an envelope using math." This problem *is* solved if explicit means "solvable by a Turing machine in poly(n) time" and $\Omega(2^n/n^{10})$ is replaced with $\Omega(2^n/n^{10^{10}})$.