6.440 Essential Coding Theory Feb 19, 2008

Lecture 4

Lecturer: Madhu Sudan *Scribe: Ning Xie*

Today we are going to discuss limitations of codes. More specifically, we will see rate upper bounds of codes, including Singleton bound, Hamming bound, Plotkin bound, Elias bound and Johnson bound.

1 Review of last lecture

Let $C \subseteq \Sigma^n$ be an error correcting code. We say C is an $(n,k,d)_q$ code if $|\Sigma| = q$, $|C| \geq q^k$ and $\Delta(C) \geq d$, where $\Delta(C)$ denotes the minimum distance of C. We write C as $[n,k,d]_q$ code if furthermore the code is a linear subspace over $\mathbb{F}_q$ (i.e., C is a linear code). Define the rate of code C as $R := \frac{k}{n}$ and relative distance as $\delta := \frac{d}{n}$. Usually we fix q and study the asymptotic behaviors of R and δ as $n \to \infty$.

Recall last time we gave an existence result, namely the Gilbert-Varshamov(GV) bound constructed by greedy codes (Varshamov bound corresponds to greedy linear codes). For $q = 2$, GV bound gives codes with $k \geq n - \log_2 \mathrm{Vol}(n, d-2)$. Asymptotically this shows the existence of codes with $R \geq 1 - H(\delta)$, which is similar to Shannon's result. Today we are going to see some upper bound results, that is, code beyond certain bounds does not exist.

2 Singleton bound

Theorem 1 (Singleton bound) *For any code with any alphabet size q, $R + \delta \leq 1$.*

Proof Let $C \subseteq \Sigma^n$ be a code with $|C| \geq |\Sigma|^k$. The main idea is to project the code C on to the first $k-1$ coordinates. Namely, define a projection map $\pi : \Sigma^n \to \Sigma^{k-1}$ such that $(x_1, \cdots, x_n) \longmapsto (x_1, \cdots, x_{k-1})$. Let $\pi(C) = \{\pi(x) | x \in C\}$. Since $\pi(C) \subseteq \Sigma^{k-1}$ so $|\pi(C)| \leq |\Sigma|^{k-1} < |\Sigma|^{k-1} \leq |C|$. It follows by the Pigeonhole Principle that there exist two distinct codewords x and y in C such that $\pi(x) = \pi(y)$. This is to say that the first $k-1$ bits of x and y all agree, thus the distance between them is at most $n-k+1$. Since d is defined to be the minimum distance between any pair of codewords, we conclude that $d \leq n-k+1$. Asymptotically, this yields $\delta \leq 1 - R$ or $R + \delta \leq 1$. ■

Note that Singleton bound holds for any q and when q is large it can be met exactly by some codes.

3 Hamming bound

Now we fix $q = 2$ and discuss the relation between Shannon and Hamming's results. Recall that codes of relative distance δ can correct $\delta/2$ fraction of errors (with probability one if the number of error bits is small). Note that this error correcting ability is independent of the underlying error models. We may call this as error correcting in the Hamming sense. According to GV bound, there exist codes of rate $R \geq 1 - H(\delta)$ that are able to correct $\delta/2$ fraction of errors. On the other hand, Shannon's result shows that $\delta/2$ of random, independent errors can be corrected by codes with rate at least $1 - H(\delta/2)$. We will call this as error correcting in the Shannon sense. The converse is also true, namely, $\delta/2$ fraction of random and independent errors requires $R \leq 1 - H(\delta/2)$. Since Shannon's error model is a restricted one, this upper bound result must also hold in the Hamming sense: for all binary codes, $R \leq 1 - H(\delta/2)$. Indeed, this was proved by Hamming and is known as the Hamming bound (also known as Packing bound). Since the result is weaker, the proof of Hamming bound is more compact than that of Shannon's.

Theorem 2 (Hamming bound) $R + H(\delta/2) \leq 1$.

Proof Since the code C has minimum distance d, all Hamming balls centered at codewords of radius $\frac{d-1}{2}$ must be disjoint. Since the volume of the union of all these balls is at most 2^n and there are (roughly) at most 2^k codewords, this gives $2^k\mathrm{Vol}(n, \frac{d-1}{2}) \le 2^n$. Asymptotically, when $n \to \infty$, $R + H(\frac{\delta}{2}) \le 1$. ■

4 Plotkin bound

After seeing these two bounds, a natural question to ask is: is the x-intercepts of the bounds correct? For example, is there a code that can achieve $\delta = 0.75$? The following Plotkin bound excludes such possibility.

Theorem 3 (Plokin bound) *For $q = 2$, $R + 2\delta \le 1$. In fact, we have*

(i) (n, k, d) code implies $(n-1, k-1, d)$ code;

(ii) (n, k, d) code with $d > \frac{n}{2}$ implies $2^k \le n + 1$ (asymptotically, if $\delta > \frac{1}{2}$ then $R = 0$).

Proof For part (i), let us list all the codewords in C as a $2^k \times n$ matrix. If we partition C into two subcodes according to the values in the last column, then since there are only two possible values (0 and 1) as the last bit for each codeword, one of two subcodes will have size at least $n/2$. Moreover, since the minimum distance of each subcode is at least d and all codewords in a subcode agree on last bit, every pair of codewords in the subcode must disagree on at least d positions in the first $n-1$ bits. If we pick the larger subcode and delete the last bits of each codeword, we get an $(n-1, k-1, d)$ code. Note that for general q, this proof shows that $(n, k, d)_q$ codes implies $(n-1, k-1, d)_q$ code.

The proof of part (ii) requires a very useful idea in coding theory: embedding the Hamming space into the Euclidean space. In doing this, Hamming distances in Hamming space are mapped to ℓ_2-distance in Euclidean space and we can use linear algebra tools, in particular dimension argument, to prove coding bounds. Consider the mapping $\{0,1\} \to \mathbb{R}$ such that $0 \mapsto 1$ and $1 \mapsto -1$. For Hamming cube, this gives $\{0,1\}^n \mapsto \{-1,1\}^n \subseteq \mathbb{R}^n$. Let $x, y \in \{0,1\}^n$ be two codewords and let $\tilde{x}$ and $\tilde{y}$ be their images in Euclidean space under this mapping. There is a nice relation between the Hamming distance between x and y and the Euclidean distance (defined by inner product) between $\tilde{x}$ and $\tilde{y}$. Namely, it is easy to check that, if $\Delta(x,y) = d$ then $\langle \tilde{x}, \tilde{y} \rangle = n - 2d$. Suppose the binary code C has m codewords in it, $C = \{x_1, x_2, \ldots, x_m\} \subseteq \{0,1\}^n$. What we need to show is that, if $d > \frac{n}{2}$ then $m \le n+1$. After applying the mapping, we have m n-dimensional real vectors $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_m$. Since $d > \frac{n}{2}$, for any two distinct vectors $\tilde{x}_i$ and $\tilde{x}_j$, $\langle \tilde{x}_i, \tilde{x}_j \rangle = n - 2\Delta(\tilde{x}_i, \tilde{x}_j) \le n - 2d < 0$. Intuitively, we can not put more than $n+1$ vectors in n dimensional space such that any pair-wise angle between two vectors is great than 90°. This fact is proved in the following geometric lemma.

Lemma 4 *If there exist m n-dimensional vectors $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_m$ such all the pariwise angles between these vectors are larger than 90°, then $m \ge n+1$. Note that the bound is tight by n-dimensional simplex.*

Proof The Lemma can be proved by induction but we are going to prove it by linear algebra argument. Suppose for the purpose of contradiction $m \ge n+2$. Since the vectors are in n-dimensional space, $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{n+1}$ are linearly dependent. That is, without loss of generality, there exist $\lambda_1, \ldots, \lambda_\ell, -\lambda_{\ell+1}, \ldots, -\lambda_t$ with $t \le n+1$ and $\lambda_i > 0$ for every i such that

$$\sum_{i=1}^{\ell} \lambda_i \tilde{x}_i - \sum_{j=\ell+1}^{t} \lambda_j \tilde{x}_j = 0.$$

Now we consider the following two possibilities.

- case 1: $t > \ell > 0$

  Let $z = \sum_{i=1}^{\ell} \lambda_i \tilde{x}_i = \sum_{j=\ell+1}^{t} \lambda_j \tilde{x}_j$. The we have

$$\begin{aligned} 0 \le \langle z, z \rangle &= \langle \sum_{i=1}^{\ell} \lambda_i \tilde{x}_i, \sum_{j=\ell+1}^{t} \lambda_j \tilde{x}_j \rangle \\ &= \sum_{i,j} \lambda_i \lambda_j \langle \tilde{x}_i, \tilde{x}_j \rangle < 0. \end{aligned}$$

- case 2: $t = \ell$

  Then we have
$$0 = \langle \tilde{x}_{n+2}, 0 \rangle = \langle \tilde{x}_{n+2}, \sum_{i=1}^{\ell} \lambda_i \tilde{x}_i \rangle = \sum_{i=1}^{\ell} \langle \tilde{x}_{n+2}, \tilde{x}_i \rangle < 0.$$

So we reach a contradiction for either case and so we prove that $m \leq n+1$. [1] ■

This complete the proof of part (ii) of the theorem.

The asymptotic Plotkin bound can be obtained by combining the two parts of the theorem and applying the first part recursively. ■

For general q, Plotkin bounds gives $R + \frac{q}{q-1}\delta \leq 1$ and usually there are codes meeting this bound.

5 Elias-Bassalygo bound

Next we are going to see a single bound that is better than both Hamming and Plotkin bounds, the so-called Elias-Bassalygo bound. The main idea is try to pack more codewords in the Hamming cube with limited overlap. More specifically, we fit the Hamming sphere with larger balls of radius τ around each codeword such that no point in the Hamming cube is covered by more than L balls. Then we have
$$2^k \cdot 2^{H(\tau)} \leq \text{sum of the volume of all balls} \leq L \cdot 2^n.$$
If we set L to a polynomial of n, then when $n \to \infty$ after taking logarithm, the $\log L$ term vanishes and we get $k + H(\tau) \leq 1$.

The radius τ defined in this way is called *list decoding radius.* To this end, we introduce a new notion of error correcting, which may be called error correcting in the Elia's sense (apart from Shannon and Hamming). In this model, we use an encoding algorithm E to encode message m. After transmitting through a noisy channel, we use a list decoding algorithm D to output a list of candidate messages $\{m_1, m_2, \ldots, m_L\}$. We say the algorithm successfully recovers the errors if $m \in \{m_1, m_2, \ldots, m_L\}$, where the size of the list L is a polynomial in n. A code C is called list decodable with list decoding radius τ if for any point x in the Hamming cube there are at most a polynomially many codewords in C that are within distance τn from x. Clearly the usual notion of decoding is a special case of list-decoding with $L = 1$.

To prove Elias-Bassalygo bound we need the following Johnson bound whose proof can be found in the appendix of the lecture note.

Theorem 5 (Johonson bound) *Every binary code of relative distance δ has list decoding radius*
$$\tau \geq \frac{1}{2}(1 - \sqrt{1-2\delta}).$$

Using Johnson bound, we thus proved the following theorem:

Theorem 6 (Elias-Bassalygo bound)
$$R + H(\frac{1}{2}(1 - \sqrt{1-2\delta})) \leq 1.$$

The Elias-Bassalygo bound is clearly better than Hamming bound and it is also better than Plotkin bound. Now we look at the behavior of Elias-Bassalygo bound at the neighborhood of $\delta = 0$ and $\delta = 1/2$.

Since $\sqrt{1-x} \leq 1 - \frac{x}{2}$, so $\frac{1}{2}(1 - \sqrt{1-2\delta}) \geq \frac{\delta}{2}$. Therefore, when $\delta \to 0$, $\frac{1}{2}(1 - \sqrt{1-2\delta}) \to \frac{\delta}{2}$.

What about $\delta \to \frac{1}{2}$? Let $\delta = \frac{1}{2} - \epsilon$, then $\frac{1}{2}(1 - \sqrt{1-2\delta}) = \frac{1}{2}(1 - \sqrt{2\epsilon})$. Since $H(\frac{1}{2} - \alpha) \approx \Theta(\alpha^2)$, when $\alpha = o(1)$, it follows that $H(\frac{1}{2}(1 - \sqrt{1-2\delta})) \approx \Theta(\epsilon)$, when $\epsilon \to 0$. That is $R \leq \Theta(\epsilon)$. Is this tight? Recall Shannon's result shows that random codes can achieve $R \geq 1 - H(\delta)$. When $\delta = \frac{1}{2} - \epsilon$ and $\epsilon \to 0$, Shannon's bound gives $R \geq O(\epsilon^2)$. So which bound is the correct one? It turns out another bound, called LP bound gives better upper bound $R \leq \tilde{O}(\epsilon^2)$. But that is outside the scope of this course.

[1] Many proofs of this flavor can be found in the beautiful book of Babai and Frankl "Linear Algebra Methods in Combinatorics".