CS 229r Information Theory in Computer Science Feb 10, 2020

Lecture 5

Instructor: Madhu Sudan *Scribe: Matthew Hase-Liu*

Today, we look at the list-decoding bound, some musings on the limitations and tightness of our bounds, and the beginnings of "algebraic" coding theory, which we'll go into much more depth next time.

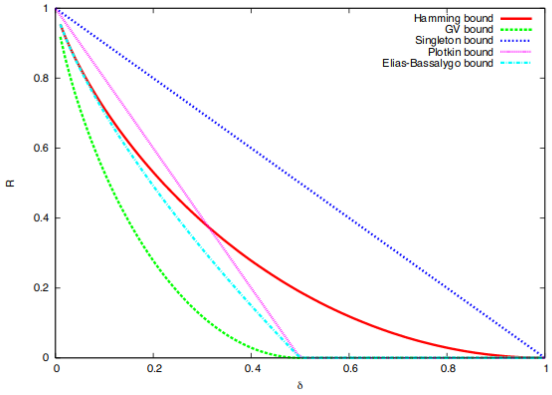
As a reminder, don't forget to sign up for office hours this and next week (in groups of two)! Also, Chi-Ning's office hours will be streamed.

1 What we've done so far

We've proved two upper bounds (respectively the Hamming bound and Plotkin bound) in previous classes: $R \leq 1 - H(\delta/2)$ by packing balls of radius $d/2$, which looks like a convex curve, and $R \leq 1 - 2\delta$ by considering an embedding of $\{0,1\}^n$ into the Euclidean space $\mathbb{R}^n$. In particular, the codes corresponding to stuff above these curves are unattainable.

On the other hand, we've also shown a lower bound (i.e. codes that are attainable) by greedy and random methods (the Gilbert-Varshamov bounds): $R \geq 1 - H(\delta)$.

2 The Elias-Bassalygo bound



The best (upper) bound we'll get in this course is called the Elias-Bassalygo bound:

$$R \leq 1 - H\left(\frac{1}{2}(1 - \sqrt{1 - 2\delta})\right)$$

.

The approach combines techniques from both the Hamming and Plotkin bounds. In particular, recall that a Hamming code of distance d corrects approximately $\delta/2$ fraction of errors uniquely. We similarly want to say that we can actually correct $\frac{1}{2}(1 - \sqrt{1 - 2\delta})$ fraction of errors with **small lists**, which are lists of size at most n^2.

Recall that in the Hamming case, saying that we can uniquely decode $\delta/2$ fraction of errors means that for any two codewords, balls of radius $d/2$ around them will not intersect.

We modify this by saying that if we draw balls of radius $\frac{1}{2}(1 - \sqrt{1 - 2\delta})$ centered at codewords in our space, although there may be intersections, there won't be any points that land in more than n^2 balls. To prove this, we'll consider an embedding into Euclidean space again.

Returning to the Hamming case, assuming the above is true, the analysis above yields the bound $2^n \geq 2^k \text{vol}(n, d/2)$, from which we derive the Hamming bound.

Exercise 1. *Assuming the unproven claim above, show that* $R \leq 1 - H\left(\frac{1}{2}(1 - \sqrt{1 - 2\delta})\right)$ *by essentially copying the same idea from the Hamming case.*

As a hint, you should get the bound $n^2 2^n \geq 2^k \mathrm{vol}\left(n, \frac{n}{2}(1-\sqrt{1-2\delta})\right)$. Note that the left side is a worse bound, while the right side is better. Balancing this tension is precisely what gives us the improved bound.

We now prove the following lemma, which will fill in the details of our proof outline above.

Lemma 2. *Let C be a code with relative distance δ, $\tau = \frac{1}{2}(1-\sqrt{1-2\delta})$, $t = \tau n$, and $d = \delta n$. Then, for all $w \in \{0,1\}^n$, there are at most n^2 codewords $c_1, \ldots, c_{n^2} \in C$ so that, for each i, w is contained in a ball of radius t centered at v_i.*

Proof. We proceed by contradiction and assume τ isn't fixed (this will be made clearer in a bit). Suppose C, then, is not list-decodable from errors with list size n^2. Then, there is some vector $w \in \{0,1\}^n$, along with codewords $c_1, \ldots, c_{n^2+1}$ so that $\Delta(c_i, c_j) \geq d$ and $\Delta(c_i, w) \leq t$. We want to show that, for the "correct" value of τ (i.e. we will ultimately show that $\tau = \frac{1}{2}(1-\sqrt{1-2\delta})$ is the "correct" value), this is, in fact, impossible.

Like in the Plotkin bound argument, we consider the following embedding in to Euclidean space: take $0 \mapsto 1$ and $1 \mapsto -1$. Renormalizing and rephrasing everything, we currently have the following setup: vectors $v, u_1, \ldots, u_{n^2+1} \in \frac{1}{\sqrt{n}}\{-1,1\}^n$ so that $\langle u_i, u_i \rangle = 1 = \langle v, v \rangle$, $\langle u_i, u_j \rangle \leq 1 - 2\delta$, and $\langle u_i, v \rangle \geq 1 - 2\tau$.

Intuitively, we want to reorient the origin so that the shifted vectors that are close to v have large angles between each other (at least 90 degrees, for instance). Doing so forces the inner products of different shifted vectors to be at most 0, from which we can conclude, recalling our analysis of the Plotkin bound, that there can be at most $2n$ vectors (which is clearly less than $n^2 + 1$ for $n \geq 2$), which yields a contradiction.

Making this more formal, we let the new origin be a scaling of v, i.e. say αv, for some $\alpha \in \mathbb{R}$. Then, we can write out the shifted vectors as $\tilde{v_i} = v_i - \alpha v$. The inner product between the shifted vectors $\tilde{v_i}, \tilde{v_j}$ is then given by $\langle \tilde{v_i}, \tilde{v_j} \rangle = \langle v_i, v_j \rangle - \alpha \langle v_i, v \rangle - \alpha \langle v_j, v \rangle + \alpha^2 \langle v_i, v_j \rangle \leq 1 - 2\delta - 2\alpha(1-2\tau) + \alpha^2$.

We can minimize the quantity on the right by taking α to be $1 - 2\tau$. By manipulating the terms, it's clear that taking $\tau \geq \frac{1}{2}(1-\sqrt{1-2\delta})$ then forces all the inner products to be nonpositive, and then we're done. ✪

As might be expected, n^2 isn't anything special in the proof above. In fact, we can take the length of the list to be any polynomial in n:

Exercise 3. *Show that everything above works out if we take the list size to be polynomial in n.*

In the following section, we analyze how the Elias-Bassalygo bound compares to the other bounds from earlier.

3 Comparison of the Elias-Bassalygo bound with other bounds

First, we note that because $H(\cdot)$ is monotone in $\delta/2 \leq \tau \leq \delta$, it follows that the Elias-Bassalygo bound lies in between the Gilbert-Varshamov and Hamming bounds. There is, in fact, another bound called the McEliece-Rodemich-Rumsey-Welch (also JPL or LP) bound [1] that is slightly better than the Elias-Bassalygo bound, but we will not cover in this course.

We note that as $\delta \to 1/2$, we have $\tau \to 1/2$, and similarly as $\delta \to 0$, we have $\tau \to 0$. Near 0, we have $H(\delta) \to \delta \log(1/\delta)$. Both the Elias-Bassalygo bound and the Hamming bound essentially tell us that, in this area, $R \leq 1 - \frac{\delta}{2}\log\frac{2}{\delta} \approx 1 - \frac{\delta}{2}\log\frac{1}{\delta}$. From the Gilbert-Varshamov bound (recall the greedy and random constructions), we also have $R \geq 1 - \delta \log \frac{1}{\delta}$.

In the other extreme, the Elias-Bassalygo bound comes close to the Gilbert-Varshamov bound. If we write $\delta = 1/2 - \varepsilon$, the Gilbert-Varshamov bound gives $R \geq \Omega(\varepsilon^2)$, whereas list-decoding gives us $R \leq 1 - H\left(\frac{1}{2}(1-\sqrt{1-2\delta})\right) = O(\varepsilon)$. It turns out that best we can do comes from the LP bound, which gives us $R = O\left(\varepsilon^2 \log^2 \frac{1}{\varepsilon}\right)$—the point is that the right answer should look something like ε^2 instead of ε.

At this point we took a quick break and did the following exercise as a class:

Exercise 4. *Introduce yourself. What is your name, grade, and concentration? And why are you taking this class?*

4 Beginnings of the algebraic theory

So far, we haven't seen too many concrete things that are useful for explicit, algorithmic purposes—perhaps the only thing we've encountered is Gilbert's greedy construction.

When talking about the algebraic theory, we start to care about the structures that can be put on $[n]$. Some examples include $\mathbb{F}_q, \mathbb{F}_q^m$, and subsets of $\mathbb{F}_q^m$, where $\mathbb{F}_q$ is the[1] finite field of q elements.

Let $S \subset \mathbb{F}_q^m$. We can consider codes C that are linear subspaces of the set of functions $\{f : S \to \mathbb{F}_q\}$, by which we mean that $f, g \in C, \alpha \in \mathbb{F}_q \implies \alpha f + g \in C$.

As we'll see in the future, we want C to comprise functions who have relatively few zeroes in S. In the simplest case, we can let $S = \mathbb{F}_q$ and take polynomial functions of bounded degree. This leads to the construction of Reed-Solomon codes.

Example 1 (Reed-Solomon codes)**.** Let $C_k = \{\text{polynomials of deg at most } k-1 \text{ with coefficients in } \mathbb{F}_q\}$. We then take $\alpha_1, \ldots, \alpha_n$ to be n distinct elements in $\mathbb{F}_q$, and define a function $E : \mathbb{F}_q^k \to \mathbb{F}_q^n$ that takes as input some element $m \in \mathbb{F}_q^k$ that corresponds to the coefficients of a polynomial of degree at most $k-1$ (note that such a polynomial has at most k nonzero coefficients) in, say, increasing order (represent this polynomial as f), and then output $[f(\alpha_1), \ldots, f(\alpha_n)]$.

Noting that the number of roots of a nonzero polynomial of degree $k-1$ is at most $k-1$, we immediately get $d \geq n - k + 1$.

This means the Reed-Solomon code reaches the Singleton bound!

Exercise 5. *In the proof above, we used the fact that the number of roots of a nonzero polynomial of degree k is at most k. Why is this true? This is a little subtle, because the polynomial $x^2 - 1$, for instance, actually has more than two roots modulo 8!*

References

[1] J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch. "New Upper Bounds on the Rate of a Code via the Delsarte-MacWilliams Inequalities." IEEE Trans. Inform. Theory, vol.IT-23, pp. 157–166, Mar. 1997.

[1] We can say **the** finite field of a certain size because all finite fields of the same size are isomorphic up to isomorphism! In particular, the finite field of size $q = p^n$, with p prime, is the splitting field of the polynomial $x^q - x$ over $\mathbb{F}_p$.