6.440: Essential Coding Theory	February 19, 2013
Lecture 4	
Prof. Madhu Sudan	*Scribe: Matthew Coudron*

1 Overview: Limits on Rates of Codes

1. Singleton Bound (Pigeon-Hole Principle)
2. Hamming Bound (Balls/Packing)
3. Plotkin Bound (Geometric Argument)

2 Quick Review

The expression $[n, k, d]_q$ denotes the set of linear codes over $\mathbb{F}_q$ (or some alphabet Σ of size q) of length n, dimension k, and distance d.

The rate of a code C is defined by: $\text{Rate}(C) \equiv \frac{k}{n}$.

The relative distance of C is defined by: $\delta(C) \equiv \frac{d}{n}$.

We define the q-ary Entropy function $H_q(\delta)$ as: $H_q(\delta) \equiv -\delta \log_q(\delta) - (1-\delta)log_q(1-\delta) + \delta log_q(q-1)$.

We know that there exist codes with rate R and relative distance δ for every pair R, δ such that $R \leq 1 - H_q(\delta)$.

The goal of this lecture is to explore known bounds on error correcting codes.

3 Singleton Bound

Consider the map $\Pi : \Sigma^n \rightarrow \Sigma^{k-1}$ defined by $\Pi(a_1, ..., a_n) = (a_1, ..., a_{k-1})$.

Given a code $C \subset \Sigma^n$ with $|C| > |\Sigma|^{k-1}$ it follows by the Pigeonhole Principle that $\exists x \neq y \in C$ such that $\Pi(x) = \Pi(y)$ (this follows because the image of Π contains at most $|\Sigma|^{k-1}$ elements).

This pair x, y are then identical on the first $k-1$ coordinates, so they can only differ on other $n-k+1$ coordinates, and thus $\Delta(x, y) \leq n-k+1$.

It follows that $\Delta(C) \leq n-k+1$, and thus $\delta = \frac{\Delta(c)}{n} \leq \frac{n-k+1}{n} \leq 1 - R + \frac{1}{n}$.

Alternatively, writing $\Delta(C) = d$, we may express the bound as $k \leq n-d+1$.

This reasoning gives what is known as the Singleton Bound.

4 Reed-Solomon Codes

Here we give a brief description of a class of codes, called Reed-Solomon codes, which demonstrates that the Singleton bound is tight. In particular Reed-Solomon codes allow us to conclude that no bound can improve on the Singleton bound without taking q (the alphabet size) into account.

A Reed-Solomon code over $\mathbb{F}_q$ ($q \geq n$) is specified by a set $\{\alpha_1, ..., \alpha_n\}$ of n distinct elements in $\mathbb{F}_q$ and a parameter k. A message $m = (m_0, ..., m_{k-1}) \in \mathbb{F}_q^k$ corresponds to the following polynomial:

$m(x) = \sum_{i=0}^{k-1} m_i x^i$

A message can be encoded as follows:

$\text{Encoding}(m) \equiv (m(\alpha_1), ..., m(\alpha_n)) \in \mathbb{F}_q^n$

This code has dimension k by definition. Since any non-zero polynomial of degree $k-1$ can have at most $k-1$ distinct roots, it follows that distinct codewords can agree in at most $k-1$ distinct positions. Thus, distinct codewords must differ in at least $n-(k-1) = n-k+1$ positions. Therefore, the code has distance $n-k+1$. These parameters saturate the Singleton bound exactly, thus demonstrating that it is a tight bound.

5 Hamming Bound/Sphere Packing Bound

Consider a $(n, k, d)_q$ code C. Define $t \equiv \lfloor \frac{d-1}{2} \rfloor$, and imagine a ball of radius t about every codeword in C. No two such balls can intersect since an intersection would imply that the corresponding codewords are separated by a distance less than d (a contradiction of the definition of d). Consequently, the sum of the volumes of all of these balls must be less than the volume of the entire codeword space. Letting $V_q(t)$ denote the volume of a ball of radius t (about any point), we have established the following:

$$q^n \geq q^k \cdot V_q(t)$$

.

A simple calculation gives $V_q(t) = \sum_{i=0}^{t} \binom{n}{i}(q-1)^i$, and so we have

$$q^n \geq q^k \cdot V_q(t) = q^k \sum_{i=0}^{t} \binom{n}{i}(q-1)^i$$

This relationship is known as the Hamming Bound, or the Sphere Packing Bound.

Note that $\log_q(V_q(t))$ is approximately $H_q(\frac{t}{n})n$ so that, by taking logarithms of the above expression, we get the approximate inequality

$$n \geq k + H_q(\frac{t}{n})n$$

and dividing by n gives

$$1 \geq \frac{k}{n} + H_q(\frac{t}{n}) = R + H_q(\frac{t}{n})$$

This is an approximate statement of the Hamming Bound which can be made precise for large t and n.

Comment: A class of codes called BCH codes give a way to pack balls into $\mathbb{F}_q^n$ very efficiently for constant distances. These codes show that, for $q = 2$ and constant distances, the Hamming bound is essentially tight.

6 Plotkin Bound

Theorem 1. *Plotkin Bound*

1. *If* $C \subset \{0,1\}^n$ *and* $\Delta(C) \geq \frac{n}{2}$ *then* $|C| \leq 2n \to \delta \geq \frac{1}{2} \to R \leq 0$
2. $R \leq 1 - \frac{q}{q-1}\delta = 1 - \delta - \frac{\delta}{q-1}$. *In particular, for* $q = 2$, $R \leq 1 - 2\delta$.

Proof. For part 1: Let $C = \{c_1, ..., c_m\} \subset \mathbb{F}_2^n$ be our code, so $\Delta(C) \geq \frac{n}{2}$ by assumption. Define the map $T : \mathbb{F}_2^n \to \mathbb{R}^n$ by applying the following map coordinatewise:

$0 \to 1$

$1 \to -1$

For $x, y \in \mathbb{F}_2^n$ it is easy to show that $||T(x) - T(y)||_2^2 = 4d(x,y)$, and $||T(x)||_2^2 = n$. A direct calculation shows that for $i \neq j \in [m]$,

$$\langle T(c_i), T(c_j) \rangle = n - 2d(c_i, c_j) \leq n - 2\Delta(C) \leq n - 2\frac{n}{2} = 0$$

We now normalize all of the vectors $T(c_i)$ (which doesn't change the sign of their inner product), and apply the part 2 of the following interesting mathematical fact.

Lemma 2. *If* $v_1, ..., v_m \in \mathbb{R}^n$ *are unit vectors such that:*

1. $\langle v_i, v_j \rangle < 0 \ \forall i \neq j$ *then* $m \leq n + 1$
2. $\langle v_i, v_j \rangle \leq 0 \ \forall i \neq j$ *then* $m \leq 2n$

It follows that we must have $m = |C| \leq 2n$, from which we see that $\delta = \frac{\Delta(C)}{|C|} \geq \frac{\frac{n}{2}}{n} = \frac{1}{2}$, and $R \leq 0$.

□