Lecture 02: More bounds on codes and Shannon's theorem

Error-Correcting Codes (Spring 2016)
Rutgers University
Swastik Kopparty
Scribes: Harsha Tirumala and Malihe Alikhani

The lecture began with a revision of the definitions of rate of a code, minimum distance of a code and the relative minimum distance.
For the remainder of this lecture C is used to denote the code and its minimum distance is represented by d.

1 Encoding and Decoding functions

Let $C \subseteq \{0,1\}^n$ and $|C| = 2^k$. Then C can be used to encode k-bit messages into n-bit codewords. Choose a bijection $E : \{0,1\}^k \to C$ which represents an Encoding function. The codeword corresponding to a message $m \in \{0,1\}^k$ is given by $E(m) \in \{0,1\}^n$

The message m to be transmitted is converted to the codeword $E(m)$ using the encoding function E and is sent for receiving. However, unless the transmission is ideal (which is rare) it will be a modified $E(m)$ that will be received. Let the error be represented by z. So the received codeword $x = E(m) + z$.

Retrieving the original message from the received codeword x is non-trivial (unless there is a guarantee of no errors). A Decoding map $D : \{0,1\}^n \to \{0,1\}^k$ is used to perform this task.

$$D(x) = \{E^{-1}(c) : c \in C\}$$

where $c \in C$ is the codeword closest to x

The output $D(x)$ is considered to be the message that was intended to be sent. Now, given some meessage m and a code C with minimum distance d:

$$\Delta(x, E(m)) < \frac{d}{2} \Rightarrow D(x) = m.$$

2 Bounds for $d = \Theta(1)$

2.1 Greedy Existence bound

This bound gives the existence of a code C with minimum distance d of a certain size (greedy construction can yield such a C hence the name) :

$$\exists C \quad |C| \geq \frac{2^n}{|B_n(d-1)|}$$

$|B_n(d-1)|$ denotes the volume of a ball of radius d-1 in $\{0,1\}^n$.

2.2 Impossibility bound

For a code C with minimum distance d, it is necessary that the balls of radius $\lfloor \frac{d-1}{2} \rfloor$ around its codewords do not intersect each other. This gives the following upper limit on the size of C:

$$\forall C \quad |C| \leq \frac{2^n}{|B_n(\frac{d-1}{2})|}$$

3 Linear Codes

Consider a linear code $C \subseteq \mathbb{F}_2^n$ which is a linear subspace in $\mathbb{F}_2^n$. Let $|C| = 2^k$. There are two succinct representations of such linear codes C :

1. Specify a basis G of k vectors ($G \subset \{0,1\}^n$) for C. This is the generator matrix G.
2. Specify C by giving a parity check matrix H of $(n-k)$ vectors $H \subset \{0,1\}^n$. H is a basis for the nullspace $C^\perp$.

4 Bounds for $d = \Theta(n)$

Recall that the relative minimum distance δ of a code $C \subseteq \{0,1\}^n$ with minimum distance d is given by $\delta = \frac{d}{n}$. Here $\delta = \Theta(1)$

Let R denote the rate of code C. Then the following bounds hold:

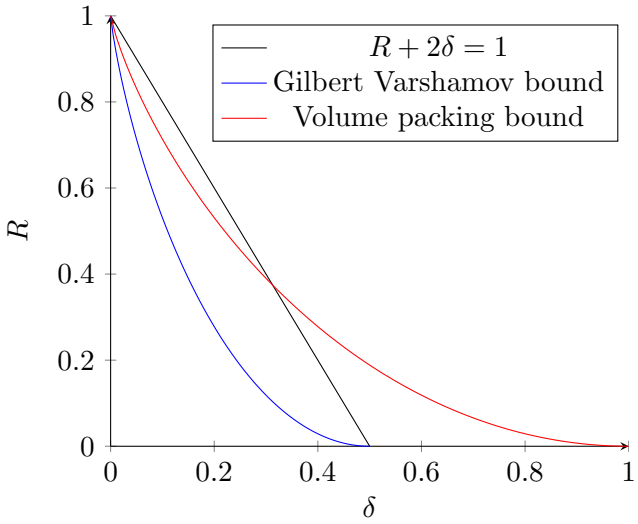
4.1 Gilbert Varshamov bound

When $\delta < \frac{1}{2}$:

$$\exists C \quad R \geq 1 - H(\delta) - o(1)$$

4.2 Impossibility bound

$$\forall C \quad R \leq 1 - H(\frac{\delta}{2}) + o(1)$$



Our knowledge so far :

- Codes definitely exist in the region under the Gilbert Varshamov bound
- Codes definitely don't exist in the region outside the Volume packing bound
- Codes in the region enclosed between them - to be further investigated !

4.3 Open questions on the plot-kin bound

- What is the best trade-off for R, δ ?
- Give an explicit code construction with $R \geq 1 - H(\delta) - o(1)$
- Provide efficient decoding functions for the above codes

4.4 What we do know

- Explicit codes with $R, \delta > 0$

- Decoding algorithms for these codes (from $\frac{\delta}{2}$ fraction errors)
- Random linear codes have $R \geq 1 - H(\delta)$

5 Random Linear Codes

As the name suggests, random linear codes are linear codes built randomly. But, what is random in this construction? Recall that a linear code can be specified by either giving a generator matrix G or a parity check matrix H. Here, the basis for G is chosen randomly. We try to build a random linear code C (of dimension k) with minimum distance d and analyze the success of such a process.

Constructing a random linear code C

1. Pick k vectors $\{v_1, v_2, v_3.....v_k\} \subset \mathbb{F}_2^n$ independently and uniformly at random . These vectors $\{v_1, v_2, v_3.....v_k\}$ form the random basis.
2. Let $C = span(v_1, v_2, v_3, ...v_k)$ be the code that spans this basis.

The constraints on C

The above two steps construct a linear code C. The following analysis reveals the probabilities associated with the two constraints on C :

1. Dimension of $C = k$
2. Minimum distance of code $C \geq d$

5.1 Is C a code of k dimensions?

Since the choice of vectors $v_1, v_2, v_3.....v_k$ was random, it is possible that they do not form a basis (i.e. $v_1, v_2, v_3.....v_k$ are linearly dependent). However, analysis below reveals that they indeed form a basis with a very high probability :

Consider a setting where $v_1, v_2, v_3.....v_k$ are picked iteratively in order from $\mathbb{F}_2^n$. If $v_1, v_2, v_3, ...v_k$ are linearly dependent $\exists$ an $i \in \{1, 2, ...k\}$ such that v_i lies in the span of the set $\{v_1, v_2, ...v_{i-1}\}$. Let $Pr(x)$ denote the probability of an event x. Then :

$$Pr(v_i \in span(v_1, v_2, ...v_{i-1})) \leq \frac{2^{i-1}}{2^n}$$

$$\Rightarrow Pr(\exists i \leq k : v_i \in span(v_1, v_2, ...v_{i-1})) \leq \Sigma_{i=1}^{k} \frac{2^{i-1}}{2^n} \leq \frac{2^{Rn}}{2^n} = 2^{-n(1-R)}$$

This probability is exponentially small in n. So with very high probability the set $\{v_1, v_2, ...v_k\}$ forms a basis i.e. C is k-Dimensional.

5.2 Is the minimum distance of $C \geq d$?

Recall the property of linear codes presented as proposition 15 in lecture 1 that :

minimum distance of $C \geq d \iff$ Every non zero element of C has atleast d non-zero elements

We show that a codeword of hamming weight $< \delta n$ exists with a very small probability which binds the probability that C has minimum distance $< d$ from the above bi-implication.

We have constructed the random basis for C above (with a high probability).For each $w \in \mathbb{F}_2^k$ there is a unique codeword $C_w = \Sigma_{i=1}^n w_i v_i$

Claim 1. *For any $w \neq 0, C_w$ is distributed uniformly at random.*

Proof. The proof of this claim is left as an exercise. □

Observation 2. *for any fixed $u \in \mathbb{F}_2^n$ and uniformly random $x \in \mathbb{F}_2^n : u + x$ is uniformly random in $\mathbb{F}_2^n$.*

So for any fixed $w \neq 0$ $Pr[wt.(C_w) < \delta n] \leq \frac{|B_n(\delta n - 1)|}{2^n}$

$$\Rightarrow Pr[\exists w \neq 0 wt.(C_w) < \delta n] \leq \frac{|B_n(\delta n - 1)|}{2^n}(2^{Rn} - 1) \approx \frac{|B_n(\delta n - 1)|}{2^n}(2^{Rn})$$

$$\Rightarrow Pr[\exists w \neq 0 wt.(C_w) < \delta n] \leq 2^{-n(R+H(\delta)-1)}$$

So if $R + H(\delta) < 1$ then this Probability is $2^{-\Omega(n)}$. This analysis reveals that the probability of existence of a codeword of hamming weight less than d is exponentially small ;and hence that the minimum distance of the constructed code C is atleast d with a very high probabiliy.

From 5.1 and 5.2 we can conclude that the random linear code C has the desired k dimensions as well as a minimum distance d with very high probability.

6 Impossibility results on Codes with high δ

Fact 3. *There do not exist codes with $R > 0, \delta \geq \frac{1}{2}$*

The following lemma will help in bounding sizes of codes with $\delta \geq \frac{1}{2}$:

Lemma 4. *If $C \subseteq \{0,1\}^n$ such that $\delta \geq \frac{1}{2}$ then $|C| \leq 2n$.*

Proof. Let $C = c_1, c_2, c_3, ...c_K \subseteq \{0,1\}^n$ be the set of codewords . Since $\delta \geq \frac{1}{2}$ we have :
$\forall i, j \in \{1, 2, ...n\}$ and $i \neq j$ $\Delta(c_i, c_j) \geq \frac{n}{2}$
Consider the following map $f : \{0,1\}^n \to \{+1, -1\}^n \in \mathbb{R}^n$ with $f(0) \to +1 and f(1) \to -1$.
Let $f(c_i) = x, f(c_j) = y$ then :

$$\Delta(x,y) = \frac{||f(x) - f(y)||^2}{4} = \frac{||f(x)||^2 + ||f(y)||^2 - 2 < f(x), f(y) >}{4}$$

$$\Rightarrow \Delta(x,y) = \frac{2n - 2f(x), f(y)}{4} = \frac{n}{2} - \frac{< f(x), f(y) >}{2}$$

where $< f(x), f(y) >$ is the inner product of the vectors $f(x), f(y)$

Observation 5. *Let $v_1, v_2, ...v_K$ be given by $v_i = \frac{f(c_i)}{\sqrt[2]{n}}$. Then the following facts hold:*

1. $||v_i|| = 1 \ \forall i \in \{1, 2, 3, ...K\}$
2. $< v_i, v_j > \leq 0 \quad (\because \Delta(c_i, c_j) \leq 0 \forall$ *i,j and* $i \neq j$ *)*

To simplify calculations, rotate all vectors v_i such that $v_1 = (1, 0, 0, 0...0)$. Note that rotation does not change the value of the inner products. From the above fact on inner products we can conclude that all other vectors have first coordinate ≤ 0. Now, delete the first coordinate from these remaining $K - 1$ vectors. These pairs also have negative inner product ; but since all of them had negative first coordinate their negative inner product must come from these remaining $n - 1$ dimensions. Note that a vector with negative first dimension and remaining $0's$ is a possibility so set this aside (since its inner product with the other $K - 2$ vectors $= 0$).

Now there are atmost $K - 2$ vectors in $n - 1$ dimensions with all inner products among them negative. This is similar to the initial situation of K vectors in n dimensions with all inner products negative. So the above argument repeats and for loss of every 1 dimension atmost 2 vectors can be added. So $K \leq 2n$ i.e. $|C| \leq 2n for \delta \geq \frac{1}{2}$.

Rate $R = \frac{\log |C|}{n} \leq \frac{\log 2n}{n} \to 0$. This proves fact 3.

The following interesting claim was made with the proof left as an exercise:

Claim 6. *If $v_1, v_2, ...v_K \in \mathbb{R}^n$ be unit vectors with $< v_i, v_j > \leq -\epsilon$ then $K \leq 1 + \frac{2}{\epsilon}$*

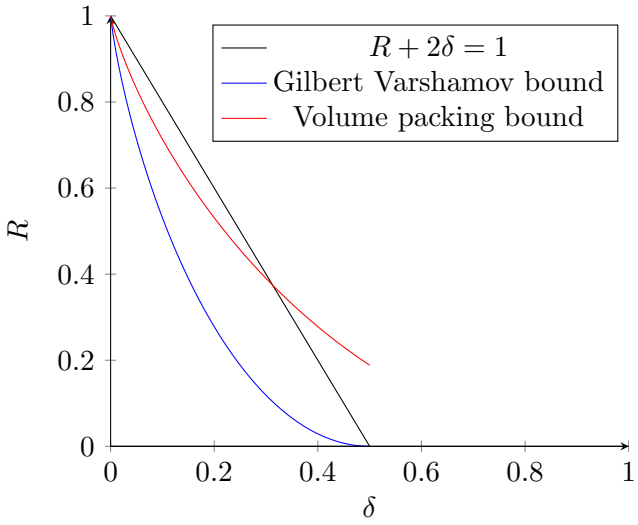
Lemma 7. $\forall C \quad R + 2\delta \leq 1$

Proof. Take a code C with K codewords. Replace the first $n - 2d$ coordinates of all the codewords with the most popular setting of these coordinates. Pigeon hole principle guarantees that there exists a set of size atleast $\frac{K}{2^{n-2d}}$ with the same first $n - 2d$ coordinates.

Distances of these codewords from each other must come from the last $2d$ coordinates because they match on the first $n-2d$ coordinates. Remove the first $n-2d$ coordinates of these codewords. So now we have atleast $\frac{K}{2^{n-2d}}$ vectors in $\{0,1\}^{2d}$ with distance d (so $\delta = \frac{1}{2}$. From lemma 4 we have the following bound :

$$\frac{K}{2^{n-2d}} \leq 2(2d) \Rightarrow K \leq 4d\{2^{n-2d}\}$$

$$\Rightarrow RateR = \frac{\log K}{n} \leq \frac{n-2d}{n} + \frac{\log 4d}{n} \leq 1 - 2\delta + o(1)$$



Conclusions :

- Codes with $\delta > \frac{1}{2}$ do not exist (asypmtotic sense)

7 Random Errors

Assume that you have a channel and you send a bit. Each bit is filliped with probability p independently.

$$m \in \{0,1\}^k, \Pr(\text{m goes correctly}) = (1-p)^k$$

Now send m, L times. Decode by taking majority of each bit.

$$P(\text{m decodes wrong}) \leq \sum_i \Pr(i^{th} \text{ bit decodes wrong}) \leq ke^{-\Omega(L)}$$

If $L = \Omega(logk)$ we successfully transmitted with probability $1 - O(1)$. Therefore to send a k bit message, this skim transmits $klogk$ bits.

8 Shannon Theorem

8.1 Theorem

(i) There exists findings E and D that,

$$E : \{0,1\}^{R_n} \to \{0,1\}^n$$

and

$$D : \{0,1\}^n \to \{0,1\}^{R_n}$$

such that for all m

$$\Pr(D(E(m) + z) \neq m) \leq \exp(-n)$$

when $z \sim \{0,1\}^n$ and $\Pr(z_i = 1) = p$.

(ii) For all $R > C$, for all E and D, there exists m such that,

$$\Pr(D(E(m) + z) = m) \leq \exp(-n)$$

8.2 Proof

Assume $R < C \quad$ and $\quad R = 1 - H(p) - \epsilon$. Let $k = 2^{R_n}$ and pick $x_1, \ldots, x_k$ unifying at random in $\{0,1\}^n$. Let E be a bijection $\{0,1\}^n \to \{x_1, \ldots x_k\}$. Let D be the following:

$$D(y) = \begin{cases} E^{-1}(x_i) & \text{if there exists exactly one i such that } \triangle\,(x_i, y) \leq n(p + \frac{\epsilon}{2}) \\ ERROR & otherwise \end{cases}$$

The goal is that after some modification with high probability E and D are such that all messages m can be transmitted correctly with probability at least $-\exp(-n)$.

We have two possible sources of error. The first one is when $wt > p + \frac{\epsilon}{2}$ meaning $E(m) + 2$ is too far from $E(m)$. The second is when some other code-words $E(m)$ lies in $B(E(m) + z, n(p + \frac{\epsilon}{2}))$.

The probability of the first bad event is at most $e^{-\Omega(n\epsilon^2)}$ and the probability of the second bad event is conditioned on the choice of $m, E(m)$ and z.

What is left is the random choice of $E(m')$ for all $m \neq m'$.

$$\Pr(\text{bad event}) = \Pr\begin{cases}\text{Pick } k-1 \text{ points uniformally from } \{0,1\}^n \\ \text{One of the points in } B(E(m)+z, n(p+\frac{\epsilon}{2}))\end{cases}$$

$$\leq (k-1)\frac{|B(n(p+\frac{\epsilon}{2}))|}{2^n} \leq 2^{n(R+H(p+\frac{\epsilon}{2})-1)} \leq 2^{-\Omega(n)}$$

where $R = 1 - H(p+\frac{\epsilon}{2}) - 1)$.

$$E_E[E_m[\Pr_z[D(E(m)+z) \neq m]]] \leq \exp(-n)$$

so there exists E such that

$$E_m[\Pr_z[D(E(m)+z) \neq m]] \leq \exp(-n).$$

Therefore

$$\Pr_m[\Pr_z[D(E(m)+z) \neq m] \geq 2\exp(-n)] \leq \frac{1}{2}.$$

Hence for at least $\frac{1}{2}k$ m's in $\{0,1\}^{R_n}$, $\Pr_z[D(E(m)+z) \neq m] < 2\exp(-n)$ take the E and D restricted to those m's. So that was the forward direction (the existence). The reverse is:

Take any E and D that,

$$E : \{0,1\}^{R_n} \to \{0,1\}^n$$

and

$$D : \{0,1\}^n \to \{0,1\}^{R_n}.$$

Let S_y be the set of strings. $S_m = \{x \in \{0,1\}^n s.t D(x) = m\}$, $\sum_m |S_m| = 2^n$.

$$\exists m \quad \text{with} |S_m| \leq \frac{2^n}{2^{R_n}} = 2^{n(1-R_n)} < 2^{nH(p-\epsilon)}.$$

$E(m)+z$ is to smeared out for $\Pr E(m)+z \in S_m$ to be big. The largest that $\Pr E(m)+z \in S_m$ can be is $\Pr(wt(z) \leq n(p-\epsilon) < e^{n\epsilon^2}$ for most m.