STO8 LECTURE 3 2/13/2008

Note Title 2/13/2008

TODAY

- Converse Coding Theorem for BSC
- Error-Correcting Codes :
  - Parameters of Interest
  - Greedy Codes : Gilbert bound
  - Linear Codes ; Dual ; Varshamov bound.

Recall Coding Theorem

For BSC(p), let $C = C(p) = 1 - H(p)$

then $\forall R < C$, suff. large n, $\exists E, D$

mapping $k = Rn$ bits to n bits & back s.t.

Decoding error Prob. $\leq \exp(-n)$.

Converse if $R > C$ then $\forall$ suff. large n,

$\forall$ E, D mapping $k = Rn$ bits to n bits,

Prob. [Decoding Error] $\geq 1 - \exp(-n)$.

Proof: Fix R, n, E, D

uniform

Transmit $E(m)$ for random m

Receive $E(m) + Y$ for random Y.

BSC(p)

Bad Events

E1 : Too few error $wt(Y) \leq (p - \frac{\epsilon}{2}) n$

E2: $\exists$ x s.t. E2(x) where

$E2(x)$:

(i) $\mathrm{Wt}\left(E(D(x)) - x\right) \geq (p - \frac{\epsilon}{2})n$

(ii) $m = D(x)$

(iii) $Y = E(D(x))$

(A) $\Pr[E1] \longrightarrow \exp(-n)$ [Chernoff]

(B) $\Pr[E2(x)]$

$$= \underset{\uparrow \; \Pr[(ii)]}{2^{-k}} \cdot \underset{\uparrow \; \Pr[(iii) \text{ given } (i)]}{2^{-H(p)\cdot n}}$$

$$\Pr[E2] \leq \sum_x \Pr[E2(x)]$$

$$= 2^n \cdot 2^{-k} \cdot 2^{-H(p)\cdot n} \longrightarrow \exp(-n)$$

(C) Neither E1 nor E2 $\Rightarrow$ Decoding failure:

Proof: Assume $\neg E1$, $\neg E2$,

& Correct Decoding.

Let $x = E(m) + y$

Correct Decoding $\Rightarrow$ $D(x) = m$

$\neg E1 \Rightarrow wt(y) \geq (p - \frac{\epsilon}{2})n$

$$\Downarrow$$

$$wt(x - E(m)) \geq (p - \frac{\epsilon}{2})n$$

$$\Downarrow$$

E2(x) is true $\Leftarrow$
(i) – $wt(x - E(D(x))) \geq (p - \frac{\epsilon}{2})n$
(ii) $D(x) = m$
(iii) $y = x - E(D(x))$

Contradiction! $\boxtimes$

Notes on Shannon Theory

1. Theory much broader (than just BSC)
   ... but no crisp characterization of when it holds

2. Even in cases where it holds, not clear how to compute C.
   Example: DELETION channel (drops bit with prob. p)

3. Proofs Not Constructive.

   Rest of Course: Effort to seek Constructive Versions.

Contrast between Hamming & Shannon

- Surprisingly disjoint emphasis.

  Shannon: E, D but not what makes pair work.

  Hamming: C ($= \text{image}(E)$) but no mention of E, D themselves

- Different Error Models:

  Shannon: random errors

  Hamming: bounded # worst-case errors.

  (Why?)

We'll use Hamming theory: Why?

bit more "constructive"

- Can at least "prove" C is not good (in all cases)

- Can use "math" to prove C is good.

Codes & Parameters

Code C over alphabet Σ is an

$(n, k, d)_q$ code if

(i) $q = |\Sigma|$

(ii) $C \subseteq \Sigma^n$

(iii) $|C| \geq q^k$

(iv) $$\Delta(C) = \min_{\substack{x \neq y \\ x, y \in C}} \{\Delta(x,y)\} \geq d$$

Notes: - Four basic Parameters

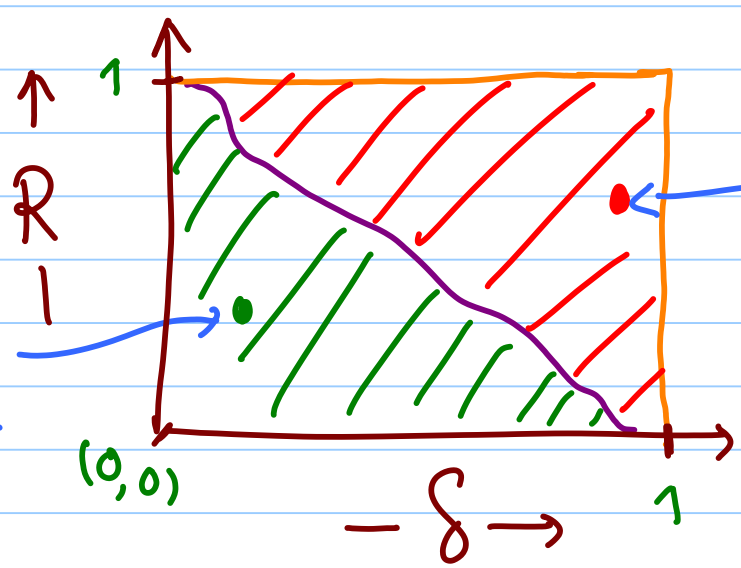
- Two too many for 2-d plots.

Asymptotics

- Fix q (today $q=2$)
- Study $R \triangleq \frac{k}{n}$ vs. $\delta = \frac{d}{n}$

  as $n \longrightarrow \infty$

- Need to fill this 2-d plot



Such codes do exist

Such codes don't exist

- Prior to Shannon: Probably didn't think $R>0$ & $\delta>0$ possible ...

GREEDY CODE

(Achieves $R, \delta > 0$)

Fix δ ; take large n ; let $d = \delta n$.

$C \leftarrow \phi$; $S = \{0,1\}^n$

while $S \neq \phi$ do

Pick $x \in S$ arbitrarily ;

$C \leftarrow C \cup \{x\}$;

$S \leftarrow S - \text{Ball}(x, d-1)$;

endwhile

Output C ;

Claim : $\Delta(C) \geq d$ (Obvious)

Claim : $|C| \geq \dfrac{2^n}{|\text{Ball}(\bar{0}, d-1)|} \approx 2^{n(1-H(\delta))}$

Theorem: $\exists$ Codes with $R \approx 1 - H(\delta)$

Problem Set 2:

- Prove random code does not achieve this
- Prove random code + deletion does achieve this.

Linear Codes

- $C \subseteq \{0,1\}^n$ is linear if $\forall x, y \in C$, $x + y \in C$.

FACT: C linear

$\Longleftrightarrow \exists G$ $k \times n$ matrix s.t.

generator

$$C = \{ x \cdot G \mid x \in \{0,1\}^k \}$$

$\Longleftrightarrow \exists H$ $n \times (n-k)$ s.t.

Parity check

$$C = \{ y \in \{0,1\}^n \mid yH = 0 \}$$

Note: Definition extends to $\Sigma = \mathbb{F}_q$ (finite field on q elements)

Notation: Linear $(n,k,d)_q$ denoted $[n,k,d]_q$.

FACT: H is parity check of $[n,k,d]_q$ code iff every subset of $d-1$ rows linearly independent.

Varshamov's Greedy (Linear) Code:

- Add rows to H greedily;
- Initially H empty.
- While $\exists\, v \in \{0,1\}^{n-k}$ s.t.

$v \neq h_1 + \dots h_\ell$ for any (every) $h_1 \dots h_\ell \in H$, $\ell \le d-2$

$H \leftarrow H \cup \{v\}$

- Output H;

Claim: $C = \{y \mid yH = 0\}$ has $\Delta(C) \geq d$.

Claim: Get code of length n with 2^k codewords provided

$$2^{n-k} > |\mathrm{Ball}(n-1, d-2)|$$

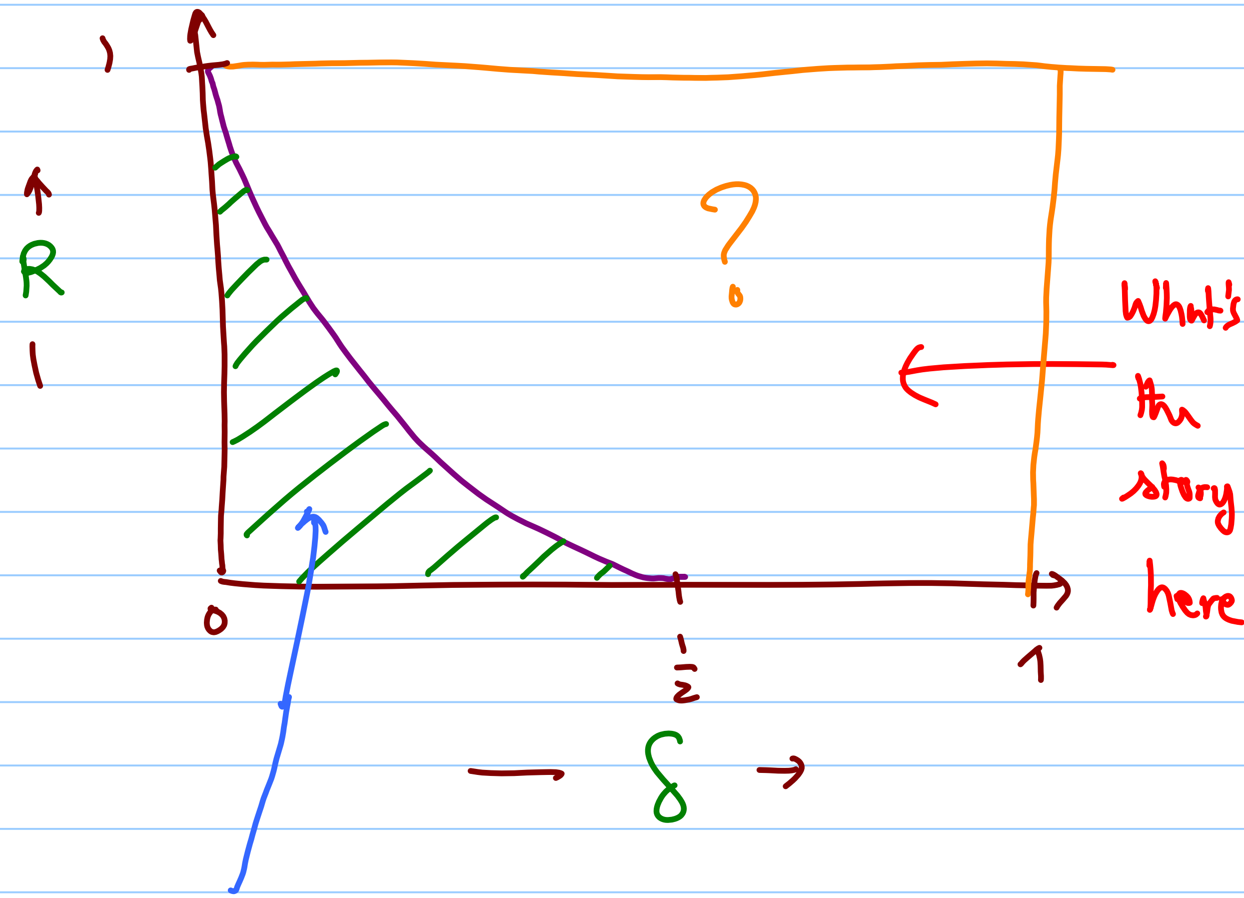
$$\left(\text{i.e., } 2^k < \frac{2^n}{|\mathrm{Ball}(n-1, d-2)|} \text{ achievable} \right)$$

Note: Matches Hamming for $d = 3$

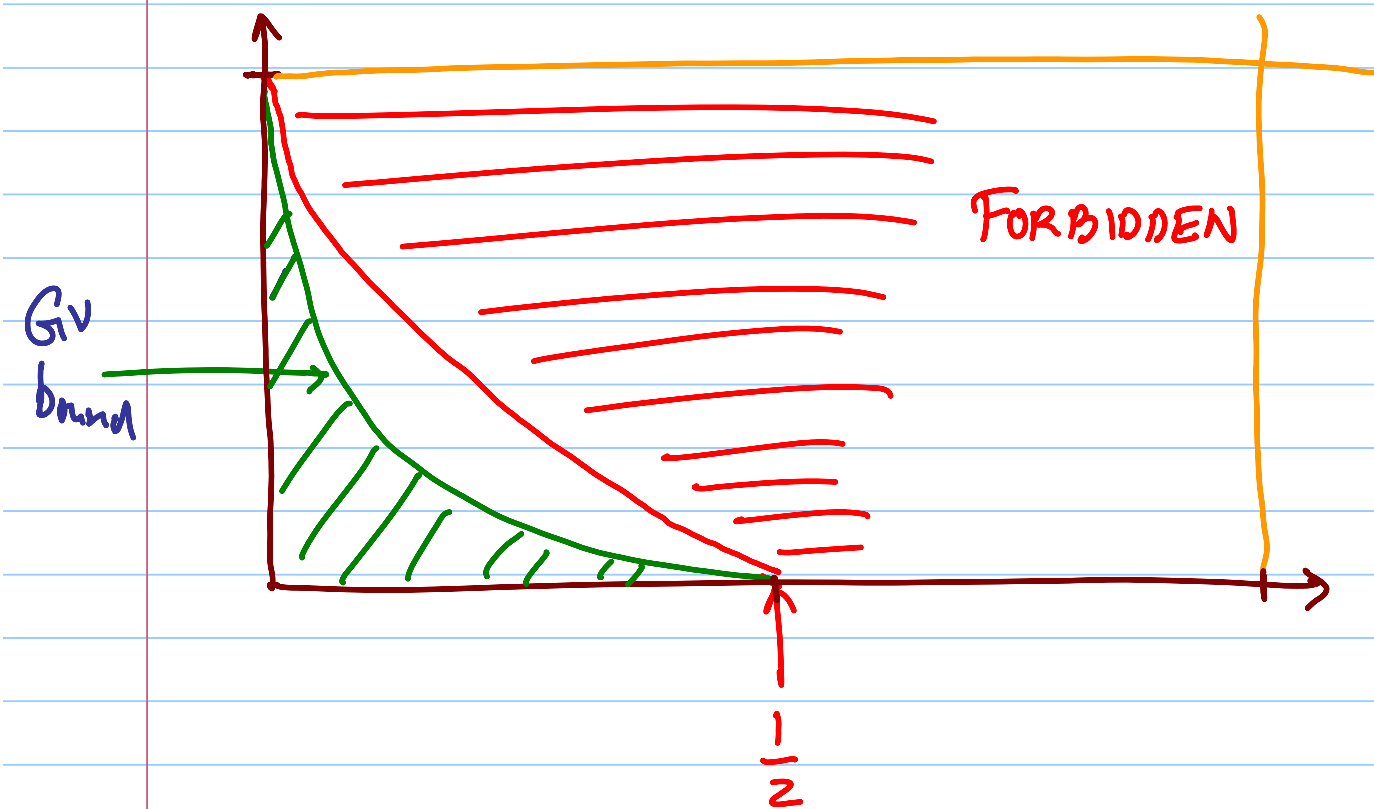
[Gilbert] Greedy $|C| = \Omega\left(\frac{2^n}{n^2}\right)$

[Varshamov] Greedy $|C| = \Omega\left(\frac{2^n}{n}\right)$

Gilbert-Varshamov Plot



Next Few Lectures



GV bound: $R \geq 1 - H(\delta)$

Best known for binary codes

Some conjecture optimal.

(Non-asymptotic Improvements)

GV : $2^k \geq \dfrac{2^n}{\mathrm{Vol}(n, d-2)}$ [$\mathrm{Vol}(n,d) \triangleq |\mathrm{Ball}(0,d)|$]

[BCH] $2^k \geq \Omega\left(\dfrac{2^n}{n^{(d-1)/2}}\right)$ (Will see later)

[Jiang-Vardy] : $2^k \geq \Omega(d)\, \dfrac{2^n}{\mathrm{Vol}(n, d-2)}$

Proof Sketch:

- Let G = graph with vertices $\{0,1\}^n$ & edge (x,y) if $\Delta(x,y) \leq d-1$
- $C \subseteq \{0,1\}^n$ has $\Delta(C) \geq d$ if C independent set of G

- $D \triangleq$ degree of vertices of G

  $= \mathrm{Vol}(n, d-1)$

- Turan's thm: G has ind. set. of size

  $|V(G)|/(D+1) \Rightarrow$ GV bound.

- But # triangles in G small

  $\# \Delta\text{'s} = 2^n \cdot D^{2-\epsilon}$ (for $\epsilon > 0$)

- [Ajtai Komlos Szemeredi] ...

  $\Rightarrow \# \Delta\text{'s} = \Omega\left(\frac{\log D}{D} \cdot 2^n\right)$

  ... $\boxtimes$