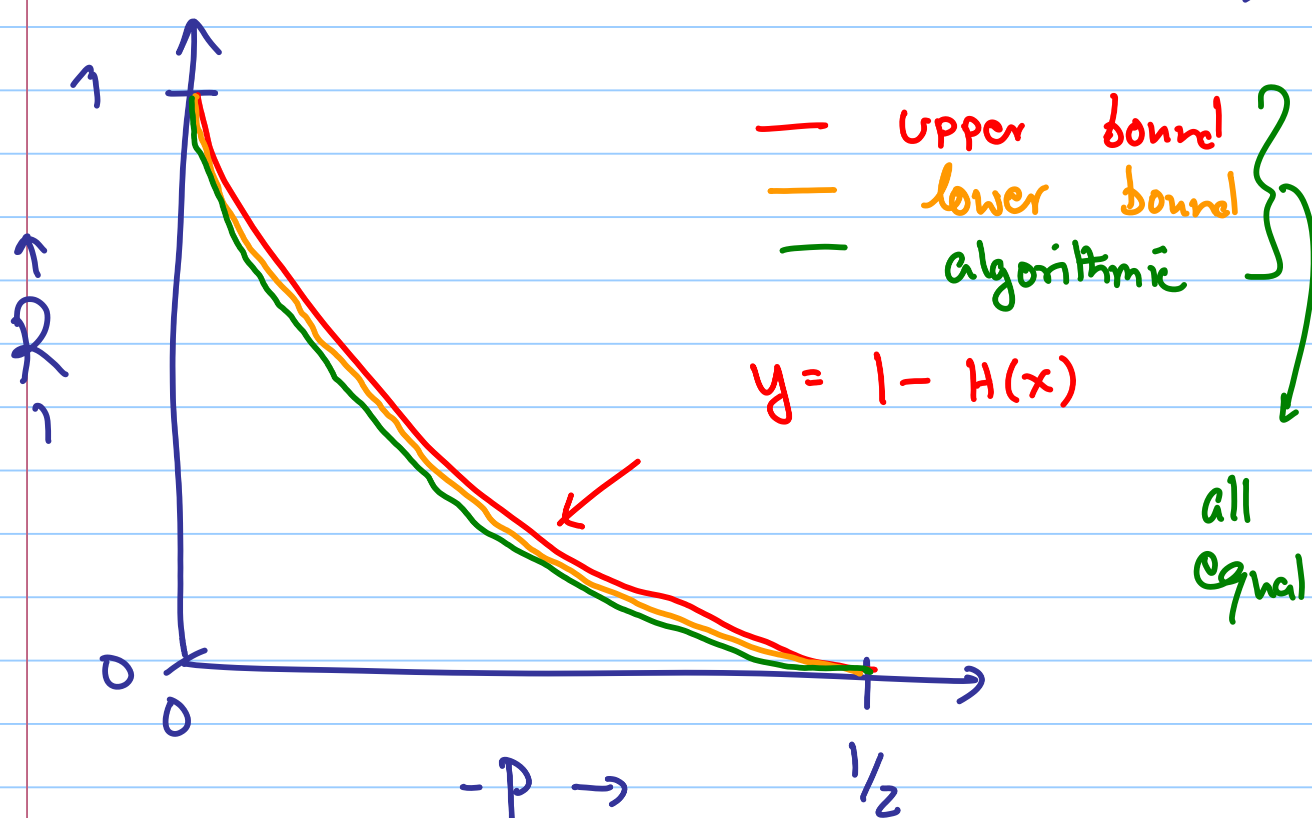
ST08 LECTURE 12 3/17/2008

Note Title 3/16/2008

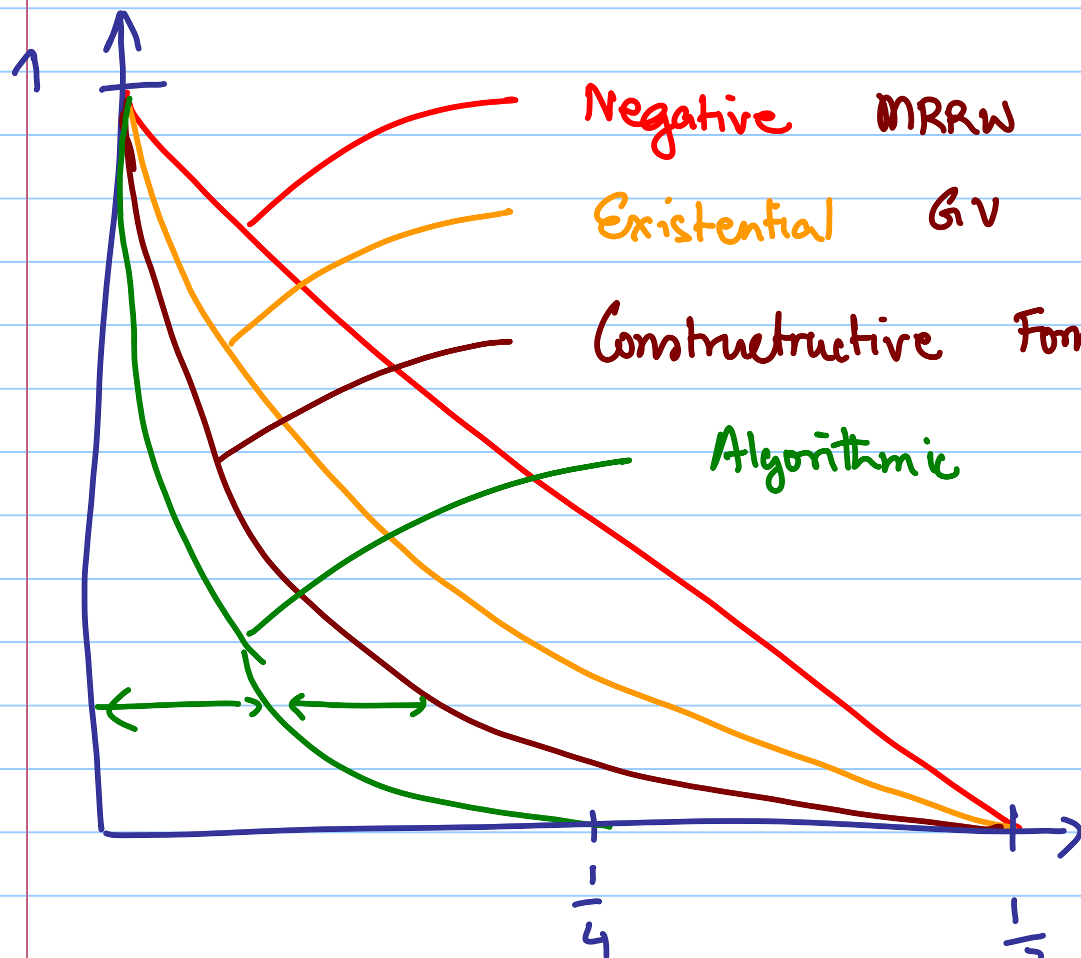
Today: List-Decoding: Combinatorics

So far:

Rate vs. Errors in Shannon Model
(random errors)



Rate vs. Errors in Hamming Model



Why such a gap between distance ( ) & error-correction ( ) ?

Hamming's bane: $e = \frac{d-1}{2}$

List-Decoding

Relaxed notion of recovery ... for adversarial errors.

- $E : \Sigma^k \to \Sigma^n$ is (p, L)-list-decodable if $\exists$ List-Decoder $D : \Sigma^n \to (\Sigma^k)^L$ s.t. $\forall\ m \in \Sigma^k,\ \eta \in \mathrm{Ball}(\bar{0}, pn)$

$$m \in D(E(m) + \eta)$$

(i.e. list-decoder outputs L messages & includes message if # errors $\le pn$)

- (Equivalent) $C \subseteq \Sigma^n$ $\quad |C| = |\Sigma|^k$

$\forall\ y \in \Sigma^n, \quad |\mathrm{Ball}(y, pn) \cap C| \le L$

Notes:

- Definition is combinatorial, not computational. (in particular "D" is not required to be efficient).

- Is this reasonable? Depends....
  - (i) If channel is "probabilistic", then typical list size = 1.
  - (ii) Can disambiguate with second channel.
  - (iii) Can add some cryptography to protect against any computationally bounded channel.

List-decoding Radius (p) vs. Rate

Upper Bound: Shannon's converse

$\Rightarrow \quad R \leq 1 - H(p)$

Pf: if (p, poly) code exists, can correct from p fraction error with error

$1 - \frac{1}{\text{poly}}$ (while Shannon $\Rightarrow 1 - \frac{1}{\exp}$)

Lower Bounds (non-constructive)

$R \geq 1 - H(p) - \epsilon \qquad \forall \epsilon > 0$

[Zyablov, Pinsker, Blinovskii]

Proof 1:

- Pick $C \subseteq \{0,1\}^n$, $|C| = 2^k$ at random.
- $\Pr[\, i^{th}$ codeword in $\mathrm{Ball}(y, pn)\,]$

$$\leq \mu = 2^{(H(p)-1)\cdot n} \quad (\to 0)$$

- $\Pr[\, \exists\, L$ codewords in $\mathrm{Ball}(y, pn)\,]$

$$\leq \binom{2^k}{L} \cdot \mu^L \leq \left(2^k \cdot 2^{(H(p)-1)n}\right)^L$$

- $\Pr[\, \exists\, y,\ L$ codewords in $\mathrm{Ball}(y, pn)\,]$

$$\leq 2^n \cdot \left(2^k \cdot 2^{(H(p)-1)n}\right)^L$$

$$\leq 2^n\, 2^{-\epsilon L n} \longrightarrow 0 \text{ if } L \gg \frac{1}{\epsilon}$$

$\boxtimes$

Linear Codes?

- Still same bounds ....
- Pick C linear at random.
- $\Pr[\exists\ L$ codewords in $Ball(y, pn)]$

$$\leq \Pr[\exists \log L \text{ independent codewords in } Ball(y, pn)]$$

....

works if $L \geq 2^{\frac{1}{\epsilon}}$.

Can we do better?

Greedy Construction ....

- Pick $b_1, \ldots b_k \in \{0,1\}^\ell$ greedily as follows:
    - Let $C_i = \text{span}\{b_1, \ldots b_i\}$
    - Let $n_i(y) = |\text{Ball}(y, pn) \cap C_i|$
    - Let $\Phi_i = E_y\left[2^{n_i(y)}\right]$
    - Pick b_{i+1} to minimize Φ_{i+1} given $b_1 \ldots b_i$.
    - Stop when $\Phi_k = 2$.
- Analysis:

Initially: $\Phi_0 = ?$

$$n_i(y) = 0 \quad \text{if} \quad y \notin \text{Ball}(0, pn)$$
$$= 1 \quad \text{if} \quad y \in \text{Ball}(0, pn)$$

$$\Phi_0 = 1 + \frac{2^{H(p)\cdot n}}{2^n} = (1+\mu)$$

Finally: $\Phi_k \le 2 \Rightarrow n_i(y) \le 2^{n+1}$

$\forall y$

$\Rightarrow C$ is $(p, n+1)$ list-decodable.

How many steps?

Claim: $\mathbb{E}_{b_{i+1}}\left[\Phi_{i+1}\right] \le \Phi_i^2$

Proof: $\Phi_{i+1} = \frac{1}{2^n} \sum_y 2^{n_i(y) + n_i(y+b_{i+1})}$

$$= \frac{1}{2^n} \sum_y 2^{n_i(y)} \cdot 2^{n_i(y+b_{i+1})}$$

$$\mathbb{E}_{b_{i+1}}\left[\Phi_{i+1}\right] = \frac{1}{2^n} \cdot \frac{1}{2^n} \cdot \sum_{b_{i+1}} \sum_y 2^{n_i(y)} \cdot 2^{n_i(y+b_{i+1})}$$

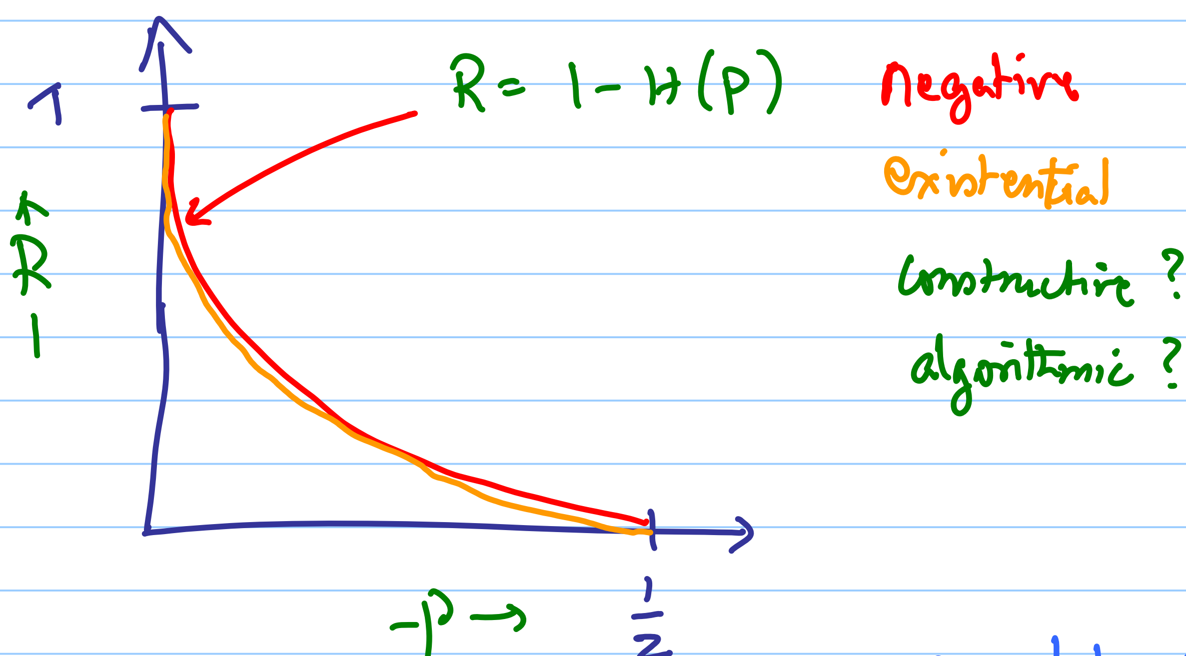
$$= \frac{1}{4^n} \sum_z 2^{n_i(z)} \cdot \sum_y 2^{n_i(y)} = \Phi_i^2 \quad \blacksquare$$

Thus after k steps. $\Phi_k \leq \Phi_0^{2^k}$

if k suff. small

$$\Phi_k \leq 1 + 2^k \cdot \mu$$

Conclusions:



Only ideas for constructive: Try same codes, but what is their radius?

Johnson Bound:

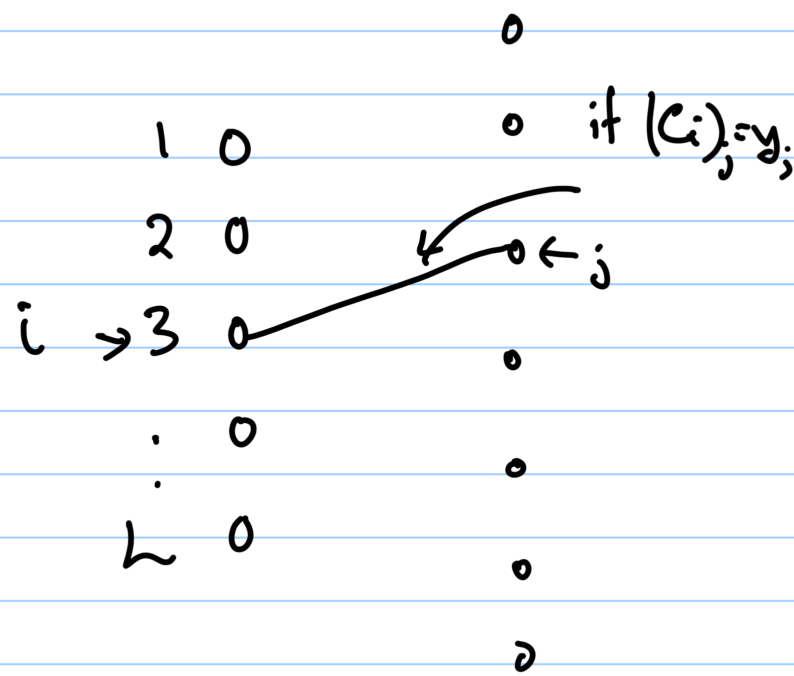
Gives list-decoding bounds on codes of known distance.

Thm: C is $[n,k,d]_q$ code

$\Rightarrow C$ is $\left(1-\sqrt{1-\frac{d}{n}}\,,\ \text{poly}\right)$- list decode

Proof: [Jaikumar Radhakrishnan] :

Say $c_1, \dots c_L$ within distance pn of y



- Draw bipartite graph
- Left vertices = $\{1, \ldots, L\}$
- Right vertices = $\{1, \ldots n\}$
- $i \leftrightarrow j \iff (c_i)_j = y_j$

Condition on graph:

- No $K_{2, n-d+1}$ $\boxed{t = n-d+1}$

  since C has distance d.

- Every vertex on left has degree $\geq (1-p)n$
- "ZARANKEWICZ" technique $\Rightarrow L \leq \ldots \ldots$

  (details below)

- Pick random i_1, i_2 distinct in $\{1, .. L\}$
  & lower bound Expected # common ngbrs.

- Let right degrees $= q_1 ... q_n$

  & let $\bar{q} = \frac{\sum q_i}{n}$; $\bar{q} > (1-p)L$

- $$\Pr_{i_1, i_2}\left[j \text{ adjacent to } i_1, i_2 \right] = \frac{\binom{q_j}{2}}{\binom{L}{2}}$$

- $$\mathbb{E}_{i_1, i_2}\left[\# \text{ common neighbours} \right] = \sum_j \frac{\binom{q_j}{2}}{\binom{L}{2}}$$

  $$\geq n \cdot \frac{\binom{\bar{q}}{2}}{\binom{L}{2}}$$

$$\frac{n\binom{\bar{q}}{2}}{\binom{L}{2}} \le (n-d+1)$$

$$\Rightarrow \quad \frac{n\binom{(1-p)L}{2}}{\binom{L}{2}} \le (n-d+1)$$

$$\approx \quad \Rightarrow \quad n(1-p)^2 \le (n-d+1)$$

$$\Rightarrow \quad (1-p)^2 \le \left(1-\frac{d}{n}\right) \qquad \boxtimes$$

$$p \le 1-\sqrt{1-\frac{d}{n}}$$

Example

- RS Code with degree $k = \frac{n}{100}$
- $d = \frac{99}{100} n$
- Unique decoding radius $= \frac{d}{2} = \frac{99}{200} n$

  $\approx .495 n$

- List decoding radius $\geq .9 n$

  Algorithm? Next Lecture.

Note: no dependence on q in lower bound.

q-ary version:

Johnson Bound:

if $C = (n,k,d)_q$ code with

$$d = \left(\frac{q-1}{q}\right)(1-\epsilon)\, n$$

then C is $\left(\frac{(q-1)}{q}\cdot(1-\sqrt{\epsilon}),\ \text{poly}\right)$ - list decodable.

Proof: See Lecture 4 Notes.