Today: List-Decoding: Combinatorics

So far:
Rate vs. Errors in Shannon Model (random errors)

- upper bound
- lower bound
- algorithmic

\[ y = 1 - H(x) \]

all equal
Rate vs. Errors in Hamming Model

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

Hamming's bound: $c = \frac{d-1}{2}$
List-Decoding

Relaxed notion of recovery... for adversarial errors.

- $E : \Sigma^k \rightarrow \Sigma^n$ is $(p, L)$-list-decodable if $\exists$ list-Decoder $D : \Sigma^n \rightarrow (\Sigma^k)^L$ such that $\forall m \in \Sigma^k, \eta \in \text{Ball}(\hat{0}, pn)$

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

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

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

  $\forall y \in \Sigma^n$, \quad $|\text{Ball}(y, pn) \cap C| \leq L$
Notes:

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

- Is this reasonable? Depends...
  1. If channel is "probabilistic", then typical list size = 1.
  2. Can disambiguate with second channel.
  3. Can add some cryptography to protect against any computationally bounded channel.
List-decoding Radius ($p$) vs. Rate

Upper Bound: Shannon's converse

\[ R \leq 1 - H(p) \]

**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)

\[ R \geq 1 - H(p) - \epsilon \quad \forall \epsilon > 0 \]

[Zyablov, Pinsker, Blinovskii]
Proof 1:

- Pick $C \subseteq \mathcal{S}_n^{2^k}$, $|C| = 2^k$ at random.

- $\Pr \left[ i^n \text{ codeword in } \text{Ball}(y, pn) \right]$
  \[ \leq \mu = 2^{(p-1)n} \rightarrow 0 \]

- $\Pr \left[ \exists L \text{ codewords in } \text{Ball}(y, pn) \right]$
  \[ \leq \binom{2^k}{L} \cdot \mu^L \leq \left( 2^k \cdot 2^{(p-1)n} \right)^L \]

- $\Pr \left[ \exists y, L \text{ codewords in } \text{Ball}(y, pn) \right]$
  \[ \leq 2^n \left( 2^k \cdot 2^{(p-1)n} \right)^L \leq 2^n \cdot 2^{-cL n} \rightarrow 0 \text{ if } L \gg \frac{1}{\varepsilon} \]

$\bowtie$
Linear Codes?

- Still same bounds ....
- Pick C linear at random.
- \[ \Pr \left[ \exists \text{ L codewords in Ball}(y, \rho n) \right] \leq \Pr \left[ \exists \log L \text{ independent codewords in Ball}(y, \rho n) \right] \]

... works if \( L \geq 2^{\frac{1}{\varepsilon}} \).

Can we do better?

Greedy Construction ....
- **Pick** $b_1, ..., b_k \in \mathbb{R}^l$ **greedily** so follows:
  - Let $C_i = \text{span \{b_1, ..., b_i\}}$
  - Let $\mathcal{N}_i(y) = |\text{Ball}(y, \rho_n) \cap C_i|$
  - Let $\Phi_i = \mathbb{E}_y \left[ 2^{\mathcal{N}_i(y)} \right]$
  - Pick $b_{i+1}$ to minimize $\Phi_i$ given $b_1, ..., b_i$
  - Stop when $\Phi_k = 2$

--- **Analysis**:

Initially: $\Phi_0$ = ?

$\mathcal{N}_i(y) = 0$ if $y \notin \text{Ball}(0, \rho_n)$

$= 1$ if $y \in \text{Ball}(0, \rho_n)$

$\Phi_0 = 1 + 2^{\frac{H(p) \cdot n}{2n}} = (1+m)$
Finally: $\Phi \leq 2 \Rightarrow n_i(y) \leq 2^{n+1} + y$

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

**How many steps?**

**Claim:** $\mathbb{E} \left[ \Phi_i \right] \leq \Phi_i^2$

**Proof:**

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

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

$\Rightarrow \mathbb{E} \left[ \Phi_i \right] = \frac{1}{2^n} \cdot \frac{1}{2^n} \cdot \frac{2^{n_i(y)} \cdot 2^{n_i(y+b_{i+1})}}{y} \leq \frac{1}{y} \cdot 2^{n_i(y)} \cdot 2^{n_i(y+b_{i+1})} \leq \frac{1}{4^n} \cdot 2^{n_i(y)} \cdot 2^{n_i(y+b_{i+1})} \leq \Phi_i^2 \quad \blacksquare$
Thus after $k$ steps, $I_k \leq \Phi_0^{2^k}$

if $k$ suff. small

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

Conclusions:

$R = 1 - H(p)$

Negative

Existential

Constructive?

Algorithmic?

Only ideas for constructive: Try same idea, but what is their reduction?
Johnson Bound:

Gives list-decoding bounds on codes of known distance.

Thm: \[ C \text{ is } [n,k,d]_q \text{ code} \]

\[ \Rightarrow C \text{ is } (1 - \sqrt{1 - \frac{d}{n}}, \text{poly}) - \text{list decode} \]

Proof: [Dai-kumar Rathakrishnan]:

Say \( C_1, \ldots, C_L \) within distance \( p \) of \( y \)

\[
\begin{align*}
&10 & \text{if } (C_i)_j = y_j \\
&20 & i \rightarrow 3 \\
&0 & \vdots \\
&0 & \vdots
\end{align*}
\]
Draw bipartite graph

Left vertices = \{1, \ldots, L\}

Right vertices = \{1, \ldots, n\}

\[ i \leftrightarrow j \iff (C_i)_j = 1 \]

Conditions on graph:

- No \( K_{2, n-d+1} \) since \( C \) has distance \( d \).

- Every vertex on left has degree \( \geq (1-p)n \)

- "ZARANKIEWICZ" technique \( L \leq \ldots \ldots \) (details below)
- Pick random $i_1, i_2$ distinct in $\{1, \ldots, L\}$. 

A lower bound Expected # common neighbors.

- Let right degrees $= q_1, \ldots, q_n$.

Let $\overline{q} = \frac{\sum q_i}{n} > (1-p)\overline{q}$.

- $P_r \left[ j \text{ adjacent to } i_1, i_2 \right]$

\[ \frac{q_j}{\binom{L}{2}} \]

- $E \left[ \text{# common neighbors} \right] \leq \sum_j \frac{r_j}{\binom{L}{2}}$

\[ \geq n \cdot \frac{\overline{q}}{\binom{L}{2}} \]
\[ n \left( \frac{\hat{e}}{2} \right) \leq (n-d+1) \]

\[ \Rightarrow \quad n \left( \frac{(1-p)L}{2} \right) \leq (n-d+1) \]

\[ \Rightarrow \quad n (1-p)^2 \leq (n-d+1) \]

\[ \Rightarrow \quad (1-p)^2 \leq \left( 1 - \frac{d}{n} \right) \]

\[ p \leq 1 - \sqrt{1 - \frac{d}{n}} \]
Example

- RS code with degree \( k = \frac{n}{100} \)

- \( d = \frac{99}{100} n \)

- Unique decoding radius = \( \frac{d}{2} = \frac{99}{200} n \)
  \( \approx 0.495 n \)

- List decoding radius \( \geq 0.9 n \)

Algorithm? Next lecture.

Note: no dependence on \( q \) in lower bound.

\( q \) - any version:
**Johnson Bound:**

If $C = (n, k, d)_{q}$ code with

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

then $C$ is $(\frac{q-1}{q}, 1 - \sqrt{\varepsilon})$, poly-list decodable.

*Proof:* See Lecture 4 Notes.