Error Correcting Codes: Combinatorics, Algorithms and Applications (Spring 2010)

Lecture 25: List Decoding from Random Errors

March 22, 2010

Lecturer: Atri Rudra *Scribe: Kushal Suryamohan*

1 Overview

We have previously seen in Hamming theory that for worst case errors $\geq \delta/2$ errors cannot be handled. However, such bad examples are rare where the error is $\geq \delta/2$. So for random errors, the list size should be small. List coding sort of achieves a trade-off between the rate and the fraction of errors that can be corrected in a code.

2 List Decoding from Random Errors

Theorem 2.1. *((p, L)-list-decodability)*
Let us fix some numbers δ (distance). ($0 < \delta < 1 - 1/q$). $\varepsilon > 0$, $q \geq 2^{\Omega(1/\varepsilon)}$.
Then the following is true for $n \geq \Omega(1/\varepsilon)$).
Let C be a q-ary code, $C \subseteq \{0, 1, ...q-1\}^n$ with relative distance δ and block length n.
Let $S \subseteq [n]$ such that $|S| = (1-\rho)n$, where $(0 < \rho \leq \delta - \varepsilon)$.
Then, for all $c_1 \in C$ and all but an exponentially small fraction ($q^{-\Omega(\varepsilon n)}$) of error patterns, $\bar{e} \in \{0, 1...q-1\}^n$ *such that*

$$\mathbf{e_s} = \bar{\mathbf{0}}$$

with

$$wt(\mathbf{e}) = \rho n.$$

Proof: We see that everything outside of S has errors and is non-zero. So, the transmitted codeword $\bar{c_1}$ is the only codeword within *hamming distance* = ρn *of* $\mathbf{c_1} + \mathbf{e_1}$.

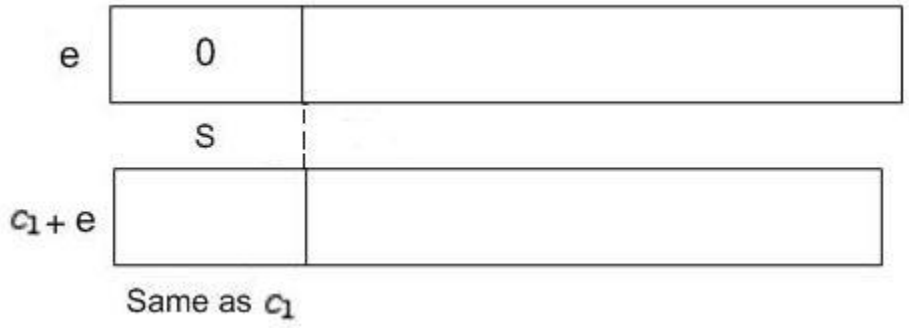
Let us for convenience define $\mathcal{E}_s$ to be the number of possible such error patterns $\mathbf{e}$ such that $\mathbf{e_s} = 0$ and $wt(\mathbf{e}) = \rho n$.
$|\mathcal{E}_s| = (q-1)^{\rho n}$ (as every error position has$(q-1)$ choices and there are ρn such positions).
Call an error pattern $\mathbf{e} \in \mathcal{E}_s$ as *bad* if there exists another codeword $\mathbf{c_2} \neq \mathbf{c_1}$ such that

$$\triangle(\mathbf{c_2}, \mathbf{c_2} + \mathbf{e}) \leq \rho n.$$

Now, we need to show that the number of bad error patterns

$$\mathbf{e_S} \leq q^{-\Omega(\varepsilon n)} |\mathcal{E}_s|.$$

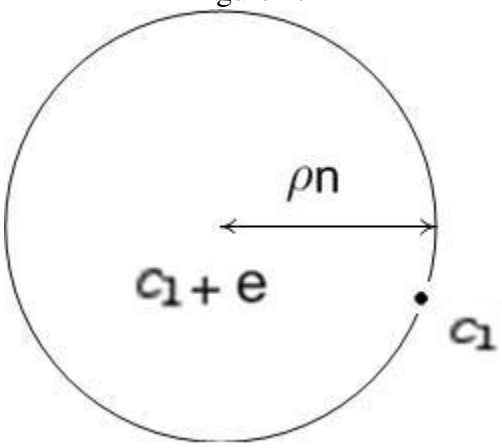
Figure 1:



Claim 1. *The Singleton bound claims that for any $(n, k, d)_q$ code, $k \leq n - d + 1$. So if we fix $(1 - \delta)n + 1$ positions in a codeword out of a possible n (in C), then at most one codeword can agree with the fixed values.*

Let us associate with $\mathbf{e}$, $c(\mathbf{e})$, which is the closest other codeword which lies within hamming distance ρn.
Note that for every bad error pattern there is at least one other codeword within hamming distance ρn.

Figure 2:



Definition: Let A be the set of positions where $c(\bar{e})$ agrees with $c_1 + e$, $A \subseteq [n]$ with $|A| \triangleq \alpha n$
Where

$$\alpha \geq 1 - \rho \geq 1 - \delta + \varepsilon,$$

(since the $c(\mathbf{e}) and c_1 + \mathbf{e}$ agree in at least $1 - \rho$ positions).

$$|A_2| = (\alpha - \beta)^n \leq \rho n$$

$$A_1 = A \bigcap S$$

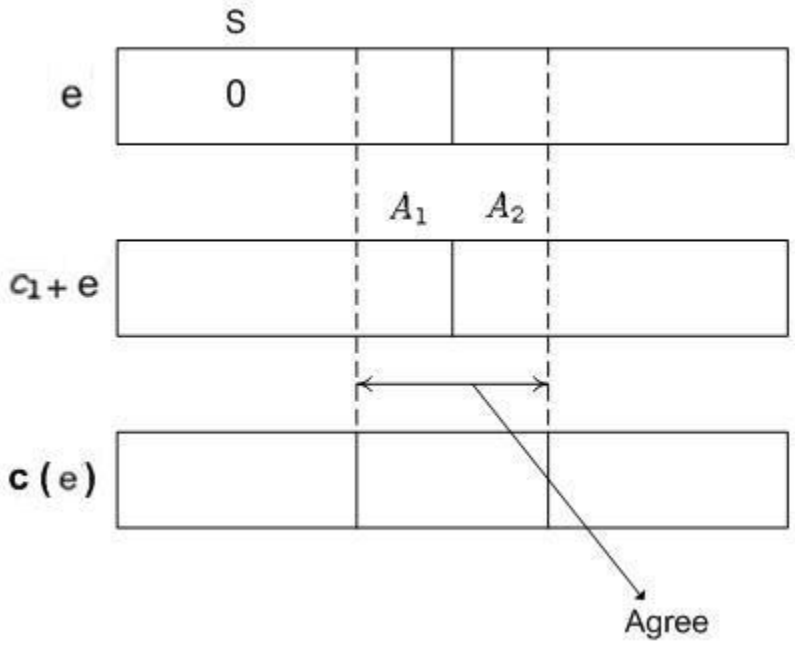
$$A_2 = A \setminus A_1$$

$$|A_1| = \beta n$$

$$A = A_1 \bigcup A_2$$

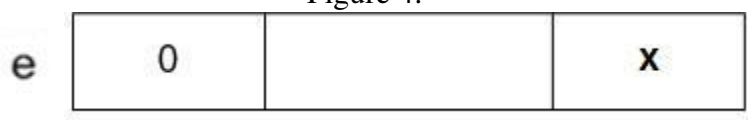
(*note that since* $A_1 \subseteq A$, βn *is at most* αn, $(\alpha - \beta) \leq \beta \leq \alpha$)

Figure 3:



What we've seen here is that every bad error pattern $\mathbf{e}$ corresponds to another $c(\mathbf{e})$ and associate it to the pair (A_1, A_2).
So, we fix (A_1, A_2) and then count the number of bad $\bar{e}$'s that map to (A_1, A_2). (Later on we will aggregate this count over all the 2^n choices of (A_1, A_2).

Figure 4:



Towards this end, first we do the following:

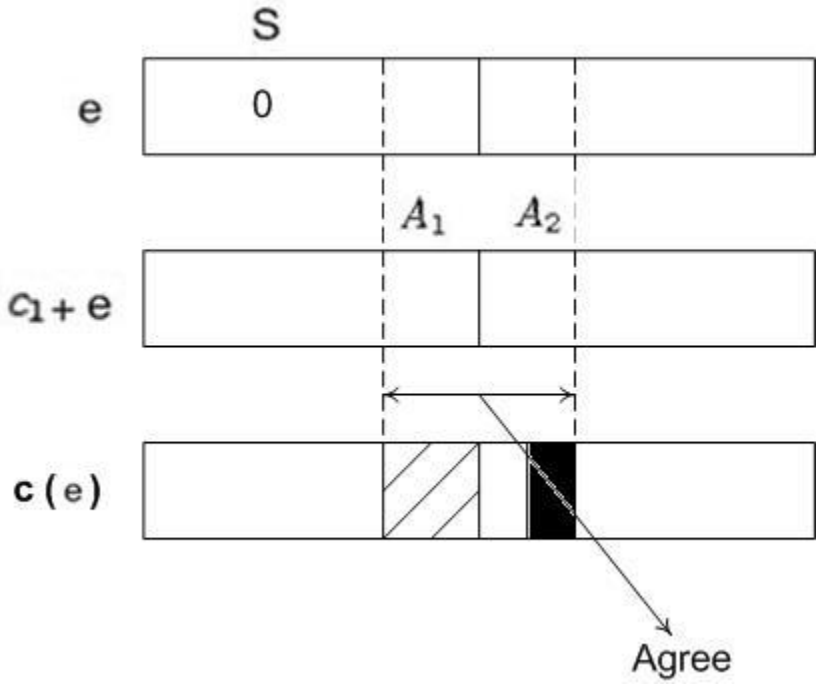
- Allow $\bar{e}$ to have arbitrary values in $[n] \setminus (S \bigcup A_2)$
  (Note that the number of possible

$$\bar{\mathbf{e}}_{[n]\setminus(S \bigcup A_2)} = (q-1)^{n-|S|-A_2|}, q^{n-(1-\rho)n-(\alpha-\beta)n})$$

- Next we fix x such that

$$\bar{e}_{[n]\setminus(S \bigcup A_2)} = \bar{x}$$

Figure 5:



By fixing $c(\bar{e})$, we fix $(1-\delta)n+1$ positions and hence we fix $\bar{e}$.
By fixing $(1-\delta)n+1-|A_1|$ positions in A_2, by our claim, then $c(\bar{e})$ is fixed and hence $\bar{e}$is fixed.
$\therefore$ The number of choices for $\bar{e_2} = (q-1)^{1-\delta n+1-\beta n}$
Thus, the number of possible bad error patterns $\bar{\mathbf{e}}$ that map (A_1, A_2) is upper bounded by

$$(q-1)^{n-(1-\rho)n-\alpha n+\beta n+(1-\delta)n+1-\beta n}$$

(Since $-\alpha$ is at most $1-\delta+\varepsilon$), the above is at most $(q-1)^{\rho n-\varepsilon n+1} \leq (q-1)^{-\varepsilon n+1}|\mathcal{E}_s|$.
(Recall the last inequality $|\mathcal{E}_s| = (q-1)^{\rho n}$)

Finally, summing up over all choices of A, we get,The total number of bad

$$2^n.(q-1)^{-\varepsilon n+1}.|\varepsilon_s| \leq q^{\frac{n}{\log_2 q}-\frac{\varepsilon n}{2}+\frac{1}{2}}.|\varepsilon_s|$$

$$(q-1 \leq \sqrt{q}) \leq q^{-\Omega(\varepsilon n)}.|\mathcal{E}_s|$$

For large enough n, $\frac{n+1/2}{\log_2 q} < \varepsilon n$ which implies that the fraction of bad error patterns is $q^{-\varepsilon n/4}$, which is exponentially small. This completes the proof.

Remark 2.2. *Theorem is not true for $q = 2^{o(\frac{1}{\varepsilon})}$. See the paper by Rudra and Uurtamo for the details*

3 References

- Two Theorems in List Decoding, Atri Rudra and Steve Uurtamo, ECCC Technical Report TR10-007, 2010
- Atri Rudra's lecture notes for Spring 2010(http://www.cse.buffalo.edu/ atri/courses/coding-theory/)