Lecture 9: ABNNR and AEL expander codes

Topics in Error-Correcting Codes (Fall 2022)
University of Toronto
Swastik Kopparty
Scribe: Haohua Tang and Lawrence Li

1 Bipartite Expander Graph

First we introduce the definitions of Bipartite Expander Graph. We assume these are d-regular.

Definition 1. *An (α, β) matrix expander bipartite graph is a bipartite graph $(V = L + R, E)$ such that*

$$\forall S \subseteq L, |S| \leq \alpha n \implies |\Gamma(S)| \geq \beta |S|$$

Definition 2. *A λ spectral absolute bipartite expander graph is a bipartite graph $(V = L + R, E)$ such that λ is the second largest absolute value of singular values of the adjacency matrix.*

Lemma 3. *Expander Mixing Lemma for Bipartite Expander:*

$$\forall S \subseteq L, T \subseteq R, \left| e(S,T) - \frac{d}{n}|S||T| \right| \leq \lambda \sqrt{|S||T|}$$

2 ABNNR Codes

2.1 Encoding

The Alon-Brooks-Naor-Naor-Roth(ABNNR) codes can be constructed basing on a given code. Given code $C \subseteq \Sigma^n$ with dimension k and distance δn, we take an (α, β) expander, in which there are n vertices on each side. Let c be a code word in C. For each vertex l_i on the left side, we associate it with c_i, and we send its value to the right side via its edges. That is, for each vertex r_i on the right side, we associate it with tuple of $(c_{\Gamma(r_i)_1}, c_{\Gamma(r_i)_2}, ...c_{\Gamma(r_i)_d})$, where neighbours in Γ are in increasing order.

The operation above gives a code word $\tilde{c} \in \tilde{\Sigma}^n$, where $\tilde{\Sigma} = \Sigma^d$, and thus we can obtain a code $\tilde{C} \subseteq \tilde{\Sigma}^n$ by performing it on every c in C.

Claim 4. *If $\delta \leq \alpha$, then $\tilde{C}$ has distance at least $\beta\delta n$.*

Proof. Consider $\tilde{c}_1, \tilde{c}_2 \in \tilde{C}$. Let $S = \{i \in L \ s.t. \ (c_1)_i \neq (c_2)_i\}$, where c_1, c_2 are the code words in C that from which we obtain $\tilde{c}_1, \tilde{c}_2$ by the ABNNR construction. Then $|S| \geq \delta n$. This implies $\exists S' \subseteq S, |S'| = \delta n$. By our assumption of $\delta \leq \alpha$, using the expansion property we can obtain $|\Gamma(S')| \geq \beta |S'|$.

By how we construct the ABNNR code, for all $j \in R$ $s.t.$ $j \in \Gamma(S)$, $(\tilde{c}_1)_j \neq (\tilde{c}_2)_j$. Thus we can see

$$\begin{aligned}\Delta(\tilde{c}_1, \tilde{c}_2) &\geq |\Gamma(S)| \\ &\geq |\Gamma(S')| \\ &\geq \beta|S'| \\ &= \beta\delta n\end{aligned}$$

□

2.2 Decoding

Given $y \in \tilde{\Sigma}^n$ $s.t.$ $\exists \tilde{c} \in \tilde{C}, \Delta(y, \tilde{c}) \leq \gamma n$ for some $\gamma \leq \frac{\alpha}{2}$, we create $r \in \Sigma^n$ by majority:

$$r_i = \begin{cases} b & \text{if } (y_j)_i = b \text{ for more than } d/2 \text{ neighbours } j \in R \text{ of i} \\ \text{JUNK} & \text{otherwise} \end{cases} \tag{1}$$

Claim 5. $\Delta(r, c) \leq \frac{\gamma}{\beta - \frac{d}{2}} n$

Proof. Given $T \subseteq R, |T| \leq \gamma n$, define $S = \{i \in L, |\Gamma(S) \cap T| \geq \frac{d}{2}\}$. First we will bound the size of S.

Since the graph is d-regular, we know that $d|T| \leq \frac{d}{2}|S|$, which implies $|S| \leq 2|T| \leq 2\gamma n \leq \alpha n$. By the expander property, $|\Gamma(S)| \geq \beta|S|$.

We know that $e(S, T) \geq \frac{d}{2}|S|$ by the definition of S, which means $e(S, R \setminus T) \leq \frac{d}{2}|S|$. Thus we have $\Gamma(S) \leq |T| + \frac{d}{2}|S|$.

Combining these we have

$$|S| \leq \frac{|T|}{\beta - \frac{d}{2}} \leq \frac{\gamma}{\beta - \frac{d}{2}} n$$

We let $T \subseteq R$ to be the set that indicates errors in y, then $|T| \leq \gamma n$. Observe that $S \subseteq L$ indicates where errors may occur after the decoding algorithm, equivalently, for all $i \notin S$, r_i must be correct. Thus we can conclude that

$$\Delta(r, c) \leq |S| \leq \frac{\gamma}{\beta - \frac{d}{2}} n$$

□

By this claim, we can use a decoder of C on r to retrieve the original string.

3 Alon-Edmonds-Luby(AEL) Codes

AEL codes are similar in spirit to concatenation codes, combined with ideas from ABNNR codes. We first begin with a small code $C_0 \subseteq \Sigma_0^d$, with $|C_0| = |\Sigma|$ and encoding function Enc: $\Sigma \to C_0$. Next, take a λ-absolute spectral bipartite expander.

AEL codes allow us to get ϵ-close to the singleton bound with alphabet size $O(1)$.

We begin with a small code C_0:

1. $C_0 \subseteq \Sigma_0^d$, with $|C_0| = |\Sigma|$.
2. Encoding function Enc: $\Sigma \to C_0$.
3. Dimension $k = (1-\epsilon)n$.
4. Rate R_0.
5. Distance δ.

And a Reed-Solomon code:

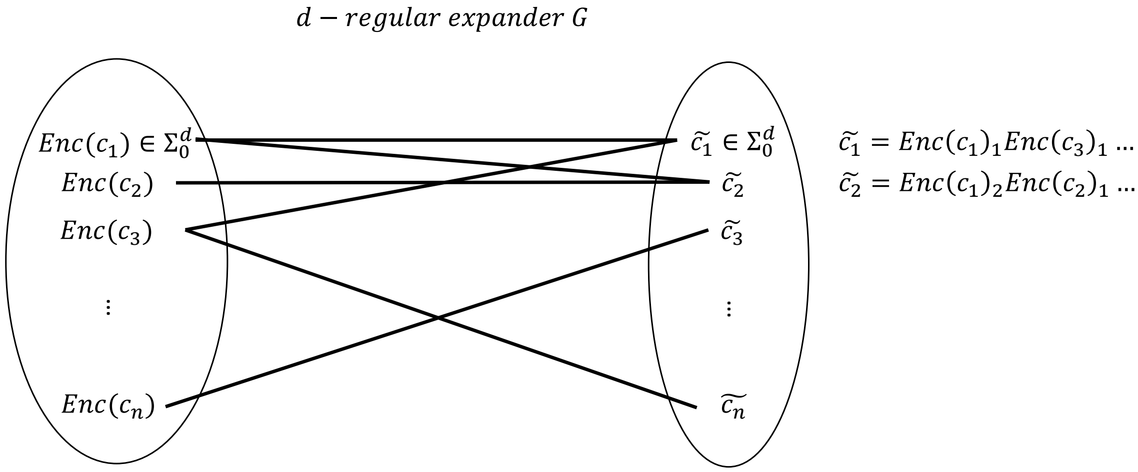
1. $C \subseteq \Sigma^n$, with $|\Sigma| = n$.
2. Rate $1-\epsilon$.
3. Distance ϵ.

And a d-regular λ-absolute bipartite expander G.

We end with a code $\tilde{C}$:

1. $\tilde{C} \subseteq \tilde{\Sigma}^n$, with $\tilde{\Sigma} = \Sigma_0^d$.
2. New encoding function from the reed solomon code to the new code Enc: $\Sigma^n \to \tilde{C}$. If $\tilde{c} = c_1 c_2 ... c_n, c_i \in \Sigma$, then $\text{Enc}(\tilde{c}) = \tilde{c_1}\tilde{c_2}...\tilde{c_n}$, where $\tilde{c_i} = \{(\text{Enc}(c_j))_k, i \text{ is the } k\text{-th neighbour of } j \text{ in } G\}$ for $c_i \in \Sigma$.
3. Rate $R_0 - \epsilon$.
4. Distance $\delta - \epsilon$.

We construct the code as follows: For each element $c = c_1 c_2 ... c_n \in C$ in the Reed-Solomon code, we let the element $c' = \text{Enc}(c)$ be an element in the new code $\tilde{C}$.



Rate of $\tilde{C}$: Let the rate of the new code be R. We have:

$$(|\tilde{\Sigma}|^{Rn}) = |\tilde{C}| = |\Sigma|^{(1-\epsilon)n}$$
$$(|\Sigma_0|^d)^{Rn} = |\Sigma|^{(1-\epsilon)n}$$

We know that $|\Sigma_0|^{R_0 d} = |\Sigma|$, so we have:

$$|\Sigma_0|^{Rd} = |\Sigma|^{1-\epsilon}$$
$$\implies R = R_0(1-\epsilon)$$

Distance of $\tilde{C}$: Take codewords $\tilde{w_1}, \tilde{w_2} \in \Sigma_0^{dn}$, and let w_1 and w_2 be such that $\text{Enc}(w_1) = \tilde{w_1}$, and $\text{Enc}(w_2) = \tilde{w_2}$. Let S be the indices where w_1 and w_2 differ, $S = \{i \in [n] | (w_1)_i \neq (w_2)_i\}$, and let T be the indices where $\tilde{w_1}$ and $\tilde{w_2}$ differ, $T = \{j \in [n] | (\tilde{w_1})_j \neq (\tilde{w_2})_j\}$. Our objective is to show that $|T|$ is large. Since w_1 and w_2 are words from a Reed-Solomon code of distance ϵ, we have the guarantee that $|S| \geq \epsilon n$.

Let H be the set of i, j pairs such that $(\text{Enc}(w_1)_i)_j \neq (\text{Enc}(w_2)_i)_j$. Notice that for any vertex in S, we have that $(w_1)_i \neq (w_2)_i$. Since Enc is an error correcting code of distance δ, we have that $\text{Enc}((w_1)_i)$ and $\text{Enc}((w_2)_i)$ differ on at least δd coordinates. Hence H has at least δd edges incident on every vertex of S.

We can now obtain a bound on the size of T:

$$\begin{aligned}
e(S,T) &\geq \delta d|S| \\
\delta d|S| \leq e(S,T) &\leq \frac{d}{n}|S||T| + \lambda\sqrt{|S||T|} \quad \text{by the expander mixing lemma.} \\
\implies |T| &\geq n\delta - n(\frac{\lambda}{d})\frac{\sqrt{|S||T|}}{|S|} \\
|T| &\geq n\delta - n\frac{\lambda}{d}\sqrt{\delta\epsilon^{-1}} \quad \text{since } S \geq \epsilon n, T \leq n\delta \\
|T| &\geq n\delta - n\epsilon \quad \text{pick } \lambda \leq d\epsilon^{1.5}\delta^{-0.5}
\end{aligned}$$

This implies that the distance is at least $\delta - \epsilon$.

This produces a code that is ϵ-close to the singleton bound with alphabet size $O(1)$.

This implies the following theorem:

Theorem 6. *Let $p \in [0, \frac{1}{2})$. Define $C_p = 1 - H(p)$, where $H(p) = -p \log p - (1-p)\log(1-p)$ is the entropy of p. If $R < C_p$, then there exists codes in $\{0,1\}^n$ of length n, rate R, such that given a corrupted codeward r prodoced by taking a true codeward, flipping each bit with probability p independently, then:*

$$Pr[\text{ the nearest codeward to } r \text{ is the original codeward }] = 1 - exp(-n)$$

Conversely, if $r > C_p$, any procedure will be wrong with probability $1 - exp(-n)$.

Furthermore, this code is explicit with a polynomial time decoding algorithm.

Theorem 7. *There exists an explicit code and poly time decoding algorithm for the above.*