Expander Codes

These notes are based on the Coding theory lecture notes of Sudan [Sud13, Lecture 14-15] and Guruswami [Gur14].

1 Expanders

Given a bipartite graph $G = (L, R, E)$ where $|L| = n$ and $|R| = m$ with $m < n$, for any subset $S \subseteq L$ of the left vertices define the neighbourhoods (standard, odd and unique as follows)

$$\begin{aligned}\Gamma(S) &= \{v \in R | \exists u \in S, (u,v) \in E\},\\ \Gamma^{\text{oddd}}(S) &= \{v \in R | \#|\{u \in S \mid (u,v) \in E\}| = \text{odd}\},\\ \Gamma(S)^{+} &= \{v \in R | \#|\{u \in S \mid (u,v) \in E\}| = 1\}.\end{aligned}$$

We say the graph $G = (L, R, E)$ is (d, D)-bounded if the degree of every left vertex is at most d and the degree of every right vertex is at most D.

We say that the graph is (γ, δ)-expander if for all subsets $S \subseteq L, |S| \leq \delta n$, we have $|\Gamma(S)| \geq \gamma|S|$.

Similarly, we that the graph is $(\widetilde{\gamma}, \delta)$-unique expander if for all subsets $S \subseteq L, |S| \leq \delta n$, we have $|\Gamma^{+}(S)| \geq \widetilde{\gamma}|S|$.

The following is an easy claim to prove.

Claim 1.1. *Let $\gamma > d/2$. If G is (d, D)-bounded and a (γ, δ)-expander, then G is a $(2\gamma - d, \delta)$-unique expander.*

We will assume (w/o proof) the explicit construction of expanders with expansion parameter $d(1 - \varepsilon)$ for every $\varepsilon \in (0, 1)$.

Theorem 1.2 (Capalbo-Reingold-Vadhan-Wigderson)**.** *For all $\varepsilon, \beta \in (0, 1)$, there exist $d \in \mathbb{Z}^{\geq 0}$ and $\delta \in (0, 1)$ such that for all sufficiently large n, there exists explicit constructions of $(d, \lfloor d/\beta \rfloor)$-bounded $(d(1 - \varepsilon), \delta)$-expanders $G = (L, R, E)$ with $|L| = n$ and $|R| = \lfloor \beta n \rfloor$.*

2 Expander Codes

We can define the corresponding code $C(G)$ based on the graph $G = (L, R, E)$.

$$C(G) = \left\{ y \in \{0,1\}^L \middle| \forall v \in R, \sum_{u \in \Gamma(v)} y_u = 0 \right\}.$$

Clearly, $C(G)$ is a $[n, \geq n - m]_2$ code. The following claim (due to Sipser-Spielman) shows that this code has linear distance when G is (d, D)-bounded $(d(1 - \varepsilon), \delta)$-expander and $\varepsilon < 1/2$.

Lemma 2.1. *Let $\varepsilon \in (0, \frac{1}{2})$. Let G be a (d, D)-bounded $(d(1 - \varepsilon), \delta)$-expander on (n, m) vertices. Then $C(G)$ has distance at least $2\delta(1 - \varepsilon)n$.*

Proof. Let $c \in \{0,1\}^L$ be a non-zero codeword of minimum weight. Let $S = \{u \in L | c_u = 1\}$. It suffices if we show that $|S| \geq 2\delta(1-\varepsilon)n$. For contradiction, assume the contrary.

Since G is a $(d(1-\varepsilon), \delta)$-expander and (d, D)-bounded, by Claim 1.1 G is also a $(d(1-2\varepsilon)$-unique expander.

To begin with, suppose $|S| \leq \delta n$, then

$$|\Gamma^{\text{odd}}(S)| \geq |\Gamma^+(S)| \geq d(1-2\varepsilon)|S| > 0.$$

Observe that every constraint corresponding to $\Gamma^{\text{odd}}(S) \neq \emptyset$ is violated by c contradicting that c is a codeword. Hence, $|S| > \delta n$.

We now have $\delta n < |S| < 2\delta(1-\varepsilon)n$. Let $Q \subseteq S$ be any subset of size exactly δn. We now have

$$|\Gamma^{\text{odd}}(S)| \geq |\Gamma^+(S)| \geq |\Gamma^+(Q)| - |\Gamma(S \setminus Q) > d(1-2\varepsilon)\delta n - d[2\delta(1-\varepsilon)n - \delta n] = 0.$$

Arguing as before, we have that c is not a codeword. Hence, $|S| \geq 2\delta(1-\varepsilon)n$, thus proving the lemma. □

3 Decoding Expander Codes

We will now see a decoding algorithm that corrects all errors if fraction of errors is at most $\delta(1-2\varepsilon)$ provided $\varepsilon < 1/4$ (i.e., expansion parameter is greater than $3d/4$). Given a received word $r \in \{0,1\}^L$, let $c \in C(G)$ be the closest codeword in C to y (assuming $\Delta(r, C(G)) \leq \delta(1-2\varepsilon)n$. A word $x \in \{0,1\}^L$, we can label each right vertex $v \in R$ "sat" or "unsat" if the corresponding constraint is satisfied. The decoding algorithm proceeds along the following simple belief propogation idea: while there exists a vertex $u \in L$ such that it has more unsat neighbours than sat neighbours, flip the value of the bit x_u.
DecodingFLIP Algorithm

- Input: received word r
  1. Set $i \leftarrow 0$, $x^{(0)} \leftarrow r$ and label the right vertices based on $x^{0)}$.
  2. While there exists a vertex $u \in L$ with more unsat neighbours than sat neighbours,
     - Flip x_u. Formally, set $x_u^{(i+1)} = 1 - x_u^{(i+1)}$ and $x_{u'}^{(i+1)} = x_{u'}^{(i+1)}$ for all $u' \neq u$. Relabel the right vertices suitably and increment i.
  3. Output $x^{(i)}$.

We first observe that the number of unsat right vertices strictly reduces with each iteration of the while loop. Since the initial number of unsat vertices is at most m, the algorithm terminates in at most m iterations of the while loops.

Let $S^{(i)} = \{u \in L | x_u^{(i)} \neq c_u\}$, ie., $S^{(i)}$ is the set of locations where the current word $x^{(i)}$ disagrees with the codeword c, closest to the original word $x^{(0)} = R$. By assumption, we have $|S^{(0)}| < \delta(1-2\varepsilon)n$.

We now make the following two claims

Claim 3.1. *Let $\varepsilon < 1/4$. If $0 < |S^{(i)}| \leq \delta n$, then there exists a $u \in L$ such that u has more unsat neighbours than sat neighbours.*

Proof. It is easy to observe that every unique neighbour of $S^{(i)}$ is an unsat vertex. Since $|S^{(i)}| \leq \delta n$, we have that $|\Gamma^+(S^{(i)})| \geq d(1-2\varepsilon)|S^{(i)}| > d|S^{(i)}|/2$. We thus have that among the potential $d|S^{(i)}|$ neighbours of $S^{(i)}$, at least $d|S^{(i)}|/2$ are unsat. In particular, there exists a vertex which has more unsat neighbours than sat neighbours. □

The above claim shows that as long as $|S^{(i)}| < \delta n$ (which is true at the very beginning $i = 0$), the algorithm enters the while loop iteration unless $S^{(i)} = \emptyset$ in which case the current word $x^{(i)}$ is a codeword (we still do not know if it is the codeword c that we wish the algorithm to output). The following claim shows that if the initial number of disagreements is less than $\delta(1-2\varepsilon)n$, then this is indeed the case (ie., for each iteration of the while loop $|S^{(i)}| < \delta n$.

Claim 3.2. *If $|S^{(0)}| < \delta(1-2\varepsilon)n$, then at the beginning of each while loop we have $|S^{(i)}| < \delta n$.*

Proof. The number of unsat vertices at the very beginning is at most $d|S^{(0)}| < \delta(1-2\varepsilon)dn$. At each iteration, the size of $S^{(i)}$ changes by exactly one (it either increases of decreases by one). Suppose by induction, we have that $S^{(i)}$ is of size less than δn up to iteration i. Observe that this is true at the very beginning (ie., $i = 0$). At the next iteration, we have $|S^{(i+1)}| \leq \delta n$. Hence, $|\Gamma^+(S^{(i+1)})| \geq d(1-2\varepsilon)|S^{(i+1)}|$. However, every vertex in $\Gamma^+(S^{(i+1)})$ is unsat and as we noticed before, the number of unsat vertices only decreases with each iteration. Hence, $d(1-2\varepsilon)|S^{(i+1)}| < \delta(1-2\varepsilon)dn$ which implies that $|S^{(i+1)}| < \delta n$. □

These two claims show that the size of $S^{(i)}$ is always less than δn, in which case the while condition is satisified unless $S^{(i)}$ is empty. However, we do know that the algorithm terminates, hence it must be the case that after the final iteration we have $S^{(i)}$ is empty, in which case the output word $x^{(i)}$ is exactly the closest codeword c to the input word $x^{(0)}$.

References

[Gur14] VENKATESAN GURUSWAMI. *15-859Y: Coding theory*, 2014. (A course on coding theory at CMU, Fall 2014).

[Sud13] MADHU SUDAN. *6.440: Essential coding theory*, 2013. (A course on coding theory at MIT, Spring 2013).