Introduction to Error Correcting Codes

Amir Shpilka
Arazim ©

December 30, 2015

1 Combinatoric construction of codes

1.1 Low-density parity check codes

Definition 1. We will say that a matrix H is d-sparse if for every row there are at most d ones.

If v is not a codeword, $Hv \neq 0$ and for example, $(Hv)_i \neq 0$ then "the 1's in the i-th row of H indicate an error".

Definition 2. A family C_n of codes with $n \to \infty$ and $C_n \subseteq \{0,1\}^n$ is LDPC if there is a $d > 0$ such that for every parity check matrix is $d - sparse$.

1. How good can an LDPC be?
2. Can we reach the GV bound with a LDPC code?
3. In what way can we correct errors? What more o we need to know in order to do that?

Theorem 1. *You can reach the GV bound with LDPC codes.*

Proof. Probabilistic method. □

We can create a two-sided graph, denoting the right side as R and the left side as L. We can think of the right side as a collection of linear constraints, and this gives us a parallelization between a parity check matrix and two-sided graphs.
We say that $\bar{x} \in \{0,1\}^n$ is a codeword if for every vertice $i \in R$ $\sum_{j \sim i} x_j = 0$

Definition 3. A two-sided graph (L, R) is called (d, c) regular if the degree of every veritce in R is c and every one in L is d.

Note 1. $|L| \cdot d = |E| = |R| \cdot c$.

Definition 4. A two-sided graph is (δ, γ)-expanding if for all $S \subseteq L$ such that $|S| \leq \delta \cdot |L|$ we have $\left|\Gamma(S)\right| \geq \gamma \cdot |S|$ where $\Gamma(S)$ is defined as the neighbors of S.

Note 2. If the graph (d, c) is (δ, γ)-expanding, then $d \geq \gamma$.

Theorem 2. *For all $0 < \alpha < 1$ there exist graphs with $|L| = n$, $|R| = \alpha \cdot n$ that are (d, c)-regular, $c = d/\alpha$ and they are (δ, γ) expanding for $\gamma = d - 1 - \varepsilon$ and $\delta = \mathcal{O}_{\alpha,\varepsilon}(1)$.*

Let G be a two-sided graph with $|L| = n$, assume that G is (d, c)-regular and (δ, γ)-expanding for $\gamma \geq \left(\frac{3}{4} + \varepsilon\right) \cdot d$.
Let $C \subseteq \{0,1\}^n$ be the code that is defined by the graph (parity check on the veritices of R).

Claim 1. There is an efficient algorithm for correcting $\frac{1}{2}(1+4\varepsilon)\delta n$ errors.
Note 3. The dimension of the code if at least:

$$dim \geq |L| - |R| = n - n\frac{d}{c} = n \cdot \left(1 - \frac{d}{c}\right)$$

Belief propogation - A vertice $i \in L$ will change its value if more than half of the parity checks with its neighbors fail.
Claim 2. Under the assumption that the graph is (d, c) regular and (δ, γ)-expandingm with $\gamma > \frac{d}{2}$ we have that the minimal distance in C is at least $\delta \cdot n$ and in particular, the minimal distance s larger than $\frac{2g}{d} \cdot \delta \cdot n$

Proof. Let $s \subseteq [n]$ be a set such that the vector 1_s is a codeword with minimal weight. We will say that $j \in R$ is a unique neighbor of S if j has a single neighbor in S. Denoting with $\Gamma_1(S)$ the set of unique-neighbors of S mad mptice that if $\Gamma_1(s) \neq \emptyset$ then 1_s is not a codeword. □

Claim 3. If $|S| < \delta n$ then $\left|\Gamma_1(S)\right| \geq (2r - d) \cdot |S|$ and in particular, if $r > \frac{d}{2}$ then $\left|\Gamma_1(S)\right| > 0$

Proof. In $E(S, \Gamma(S))$ we have that

$$\left|\Gamma_1(S)\right| + 2 \cdot \left|\Gamma(S) \backslash \Gamma_1(S)\right| \leq E(S, \Gamma(S)) = d \cdot |S| \Rightarrow 2\left|\Gamma(S)\right| - \left|\Gamma_1(S)\right|$$

$$\left|\Gamma_1(S)\right| \geq 2\left|\Gamma(S)\right| - d \cdot |S| \geq (2r - d) \cdot |S|$$

Where the last inequality occurs if $\left|\Gamma(s)\right| \geq \gamma \cdot |S|$ and $|S| < \delta n$ □

We will prove the stronger claim for the minimal distance. We have seen that

$$2\gamma\delta n - d|S| \leq 2\Gamma(S) - d|S| \leq \left|\Gamma_1(S)\right|$$

If 1_s is a codeword then $\left|\Gamma_1(s)\right| = 0 \Rightarrow$ *something*

Flip algorithm As long as the is a vertice $i \in L$ ofr which

$$\# \{j \sim i : \text{The equation on } j \text{ doesnt hold}\} > \frac{d}{2}$$

We will flip the value of the i-th coordinate.
Claim 4. IF we have arrived at a word with a number of errors that is smaller than $\frac{\delta}{2d} \cdot n$ then the FLIP algorithm runs in linear time and fixes all of the errors.

Proof. At every stage, the number of equations not satisfied goes down, therefore the number of stages $\leq$ number of unsatisfied equations. Thus, at the end we have at most

$$\overbrace{\frac{\delta}{2}}^{\#\text{stages}} n + \overbrace{\frac{\delta}{2d}}^{\#\text{errors}} \cdot n < \delta n$$

Errors
Claim 5. At the end of the alorithm, all of the equations are satisfied. In particular, at the end of the algorithm we have a codeword. According to the calculation, the distance from the original codeword $< \delta n \leq min - dist$ and this must be the original codeword.

proof of claim. Let S be the set of errors at a certain stage. we have shown that $|S| < \delta n$ and therefore

$$\left|\Gamma_1(S)\right| \geq (2\gamma - d)|S| \overset{\gamma > \frac{3}{4d}}{>} \frac{d}{2}|S|$$

$\Rightarrow$ there is a vertice in S with more than $\frac{d}{2}$ unique neighbors and they are all unsatisfied. □

paragraph about the running time of the algorithm. □

Parallel FLIP algorithm At every stage, every vertice that is connected to mode than $\frac{d}{2}$ unsatisfied equations, changes the value of the word written in them.

Claim 6. If $\gamma \geq \left(\frac{3}{4}+\varepsilon\right)d$, the number of errors is $\leq \frac{1}{2}(1+4\varepsilon)$ then the number of stages that the algorthm performs is $\mathcal{O}(\log n)$ and at the end we arrive at the original codeword.

Proof. We will show that at each stage, the number of errors grows smaller by a factor of $(1-4\varepsilon)$. We will look at the first stage. Denoting S' as the set of errors at the end of the stage and S as the errors at the beginning. □

Claim 7.

$$|S \cup S| < \delta \cdot n$$

proof of claim. If the union is larger than δn then let $S'' \subset S'$ such that $\left|S \cup S''\right| = \delta n$. We will defien $S''_{in} = S'' \cap S$ and $S''_{out} \backslash S$

$$\left(\frac{3}{4}+\varepsilon\right) \cdot d \cdot \delta n \leq \gamma \cdot \delta n \leq \left|\Gamma\left(S \cup S''\right)\right| = \left|\Gamma(S)\right| + \left(\left|\Gamma\left(S''_{out}\right)\right| - \left|\Gamma\left(S''_{out}\right) \cap \Gamma\left(S\right)\right|\right)$$

$$\leq \left|\Gamma(S)\right| + \frac{d}{2}\left|S''_{out}\right| \leq d \cdot |S| + \frac{d}{2}\left|S''_{out}\right| \leq \frac{d}{2}\left(|S| + \left|S''_{out}\right|\right) = \frac{d}{2}|S| + \frac{d}{2}\delta n$$

$$\left(\frac{3}{4}+\varepsilon\right)\delta dn \leq \frac{d}{2}|S| + \frac{d}{2}\delta n$$

$$\frac{1}{2}\left(1+4\varepsilon\right)\delta n = \left(\frac{1}{2}+2\varepsilon\right)\delta n \leq |S|$$

And this is a contradiction to the assumption that the number of errors is smaller than $\frac{1}{2}(1+4\varepsilon)\cdot\delta n$ □

Claim 8.

$$\left|S'\right| \leq (1-4\varepsilon)\cdot|S|$$

Proof.

$$\left(\frac{3}{4}+\varepsilon\right) d \cdot \left(|S| + \left|S'_{out}\right|\right) \leq \gamma\left|S \cup S'\right| \leq \Gamma\left(S \cup S'\right) \leq d \cdot \left|S \backslash S'_{in}\right| + \frac{d}{2}\left|S'_{in}\right| + \frac{d}{4}\left|S'_{in}\right|$$

And after moving sides, we arrive at

$$\frac{1}{4}\left|S'\right| \leq \frac{1}{4}\left|S'_{in}\right| + \left(\frac{1}{4}+\varepsilon\right)\left|S'_{out}\right| \leq \left(\frac{1}{4}-\varepsilon\right)|S|$$

□