Introduction to Error Correcting Codes

Amir Shpilka
Arazim ©

January 6, 2016

In this lesson:

- Codes with linear encoding and decoding which are able to fix $\Omega(1)$ percent of errors.
- Codes such as the ones described which reach the Shannon bound.
- Complexity problems.

1 Linear complexity codes

Another way to build codes from graphs We will take the left side to have k vertices and denote each value $x_1, \ldots x_k$ and the right side has $n-k$ vertices with the value at $y_j = \sum_{j \sim i} x_i$. Each vertice on the left side has a degree of d and each one on the left c. The generating matrix for this code is $\left(\frac{I}{B}\right) \in M_{k \times n}$ where B is the matrix for the graph. If c is constant then the encoding can be done in $c \cdot n$ time. minimal value in the matrix is $d+1$.

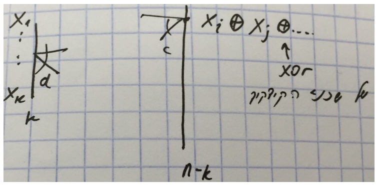
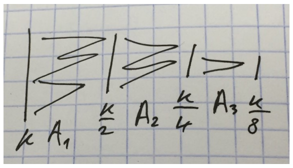
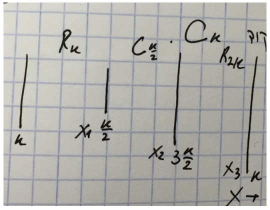


Figure 1: How to build

Another idea



Another idea for building a graph is taking a $\log k$ layered graph, with the i-th one having a size of $c = \frac{k}{2^i}$. Then we can create a code with a size of $2k$ by taking the first matrix as $\left(\begin{smallmatrix} I \\ A_1 \end{smallmatrix}\right) \in M_{k \times \frac{3k}{2}}$, the second $\left(\begin{smallmatrix} I \\ A_2 \end{smallmatrix}\right) \in M_{\frac{3k}{2} \times \frac{7k}{8}}$, etc.



Our construction: We will show a recursive construction of a code with a rate of $\frac{1}{4}$. Asume tat we have a family of $(d, 2d)$-regular graphs ($d \geq 8$) with $|L| = k$ for all k, say, a power of 2. In addition we shall assume that the graphs are $\left(\delta, \frac{7}{8}d\right)$ expanding. We will define R_k as the code with a graph as defined and for the right side $|R| = \frac{k}{2}$ where each value on the right is the sum of its neighbors. Now we will recursively define a code C_k as the encoding $x \mapsto (x, x_1, x_2, x_3)$.

- We encode x using R_k and arrive at a code with a length of $\frac{k}{2}$.

- We encode x_1 and using using $C_{k/2}$ and arrive at x_2 which has a length of $\frac{3k}{2}$.
- We encode (x_1, x_2) using R_{2k} and reach x_3 which has a length of k.

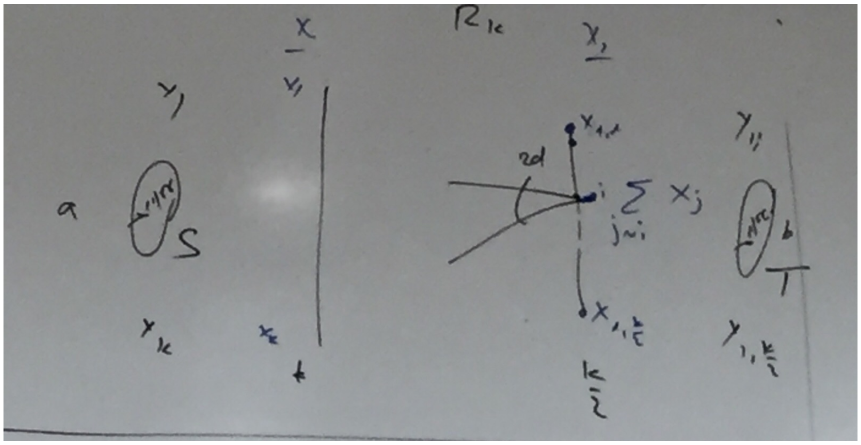
Analyzing the complexity of this code:

Encoding

$$T(k) = 2d \cdot \frac{k}{2} + T(k) + 2d \cdot k = T(\frac{k}{2}) + 3d \cdot k \Rightarrow T(k) = 6dk$$

Decoding Idea: We will prove that when we perform an iterative FLIP on R_k, if the number of original errors is not too large, then the number of errors in a certain part of the message is "Not too large". meaning that after FLIP on R_{2k} we have that y_1, y_2' in which there are not too many errors.
Using recursion, we will fix all of the errors in y_1, y_2 in a code $C_{k/2}$ meaning that we will get exactly x_1. Finally, we will perform the FLIP algorithm and return exactly $x \Rightarrow$ linear decoding time.

Claim 1. Let R_k be the code we built using a $(d, 2d)$ graph, $d \geq 8$ and $(\delta, \frac{7}{8}d$ expanding. Let (x, x_1) be a codeword (the length of x is k) and the length of x_1 is $\frac{k}{2}$. We will assume that $(\bar{y}, \bar{y_1}$ is a codeword such that $dist(\bar{y_1}, \bar{x_1}) \leq b$, $dist(\bar{y}, \bar{x}) < a$ such that $a \cdot (d+1) + b \leq \delta \cdot k$. Thus, the iterative FLIP algorithm (a code in L changes its value if it is connected to more errors than not) will return $(\bar{y}, \bar{y_1}$ such that $dist(\bar{y}, \bar{x}) \leq \frac{1}{2}b$



Proof. We will denote

$$S = \{i : x_i \neq y_i\} \qquad T = \left\{j : x_{1,j} \neq y_{1,j}\right\} \qquad S' = \left\{x_i \neq y_i'\right\}$$

Clearly, $|S| \leq a$ and $|T| \leq b$. The number of stages in the algorithm is smaller or equal to the number of vertices in the right side whose value is different from the values of its neighbors (we denote these as unsatisfied values).
The codes which have a disagreement are a subset of $\Gamma(S) \cup T$. Therefore the total number of original disagreements is

$$\#\left\{\text{original disagreements}\right\} \leq \left|\Gamma(S)\right| + |T| \leq d \cdot |S| + |T| \leq d \cdot a + b$$

Therefore at the end of the algorithm ,te number of erros in the left side is

$$\left|S'\right| \leq \overbrace{a}^{\substack{\text{Orignal} \\ \text{errors}}} + \overbrace{(d \cdot a + b)}^{\#\{\text{stages}\}} \leq \delta \cdot k$$

And since $\left|S'\right| \leq \delta k$. According to a lemma that we proved:

$$\left|\Gamma_1(S')\right| \geq (2 \cdot \frac{7}{8}d - d) \cdot \left|S'\right|$$

We will define

$$u = \left\{ k : j \in \Gamma(S') \wedge y_{1,j} \neq \sum_{i \sim j} y_i' \right\}$$

Since for every vertice in L has more satisfied neighbors than not, $|u| \leq \frac{d}{2} - \left|S'\right|$
On the other hand every unique neighbor of S which does't have an error, is not satisified. In particular, $u \geq \Gamma_1(S') \backslash T$

$$\left|\Gamma_1(S')\right| = |T| \leq \left|\Gamma_1(S') \backslash T\right| \leq |u| \leq \frac{d}{2}\left|S'\right| \Rightarrow \left|S'\right| \leq \frac{4}{d}b \leq \frac{1}{2}b$$

□

1. FLIP on R_{2k} $\left(\left((\bar{y_1}, \bar{y_2}), \bar{y_3}\right)\right)$ will return $\left((\bar{y_1}', \bar{y_2}'), \bar{y_3}'\right)$.
2. We will fix $C_{\frac{k}{2}}$ $\left(\bar{y_1}', \bar{y_2}'\right)$ We will get $(\bar{y_1}'', \bar{y_2}'') = (\bar{x_1}, \bar{x_2}$.
3. We will perform a FLIP on $(\bar{y}, \bar{y_1}'')$ and get $\bar{x}$

Claim 2. If the percentage fo original errors is $\leq \varepsilon$ thenthe algorithm we described bixes C_k.

Proof. The starting number of erros is $\leq \varepsilon \cdot 4k$ (since $dist(\bar{x}\bar{x_1}\bar{x_2}\bar{x_3}, \bar{y}\bar{y_1}\bar{y_2}\bar{y_3}) \leq \varepsilon \cdot 4k$) and in particular,

$$dist(\bar{y_3}, \bar{x_3}) \leq \overbrace{\varepsilon \cdot 4k}^{b}, \qquad dist\left((\bar{y_1}, \bar{y_2}), (\bar{x_1}, \bar{x_2})\right) \leq \overbrace{\varepsilon \cdot 4k}^{a}$$

In R_{2k} we need

$$\varepsilon \leq \frac{\delta}{2(d+2)} \Leftarrow \varepsilon \cdot 4k(d+1) \rightarrow \varepsilon \cdot 4k \leq \delta \cdot (2k)$$

According to the error reduction clain, we have that $dist((\bar{y_1}', \bar{y_2}'), (\bar{x_1}, \bar{x_2})) \leq \varepsilon 4 \cdot \frac{k}{2} = \varepsilon 2k \Rightarrow$ By induction, stage 2 will <u>fix</u> $\bar{y_1}', \bar{y_2}'$ and return $\bar{x_1}, \bar{x_2}. \Rightarrow$ if $4\varepsilon k(d+1) \leq \delta k$ then FLIP will return 0 errors (by the error reduction claim). □

2 Test

The test will consist of questions from the homework or similar to them. In addition, there may be proofs from class.