CS 496 Lecture 3: Expander Codes

Sidhanth Mohanty

September 23, 2025

Today, we will wrap up the analysis of majority dynamics from the previous lecture (see notes for Lecture 2 for a full proof), and then we will see an application of expander graphs to the construction of good error-correcting codes. The material in this lecture is based on the work of Sipser and Spielman [SS02].

1 Error-correcting codes 101

The setup here is that Alice wants to send a k-bit message to Bob, but the channel she is sending the message over undergoes some amount of corruption by an adversary: i.e., the adversary has some budget of bitflips they are allowed to perform on the message. Therefore, Alice must introduce some amount of redundancy in her messages. The question is: how can Alice encode her message so as to not add too much redundancy, and yet have her encoded message be resilient to adversarial errors.

This is formalized via the following notion.

Definition 1.1. A *code* $\mathcal{C}$ is a subset of $\{0,1\}^n$. We say the *rate* of $\mathcal{C}$ is $r(\mathcal{C}) := \frac{\log|\mathcal{C}|}{n}$, and the *distance* of $\mathcal{C}$ is $\Delta(\mathcal{C}) := \min_{x,y\in\mathcal{C}} \text{dist}(x,y)$, where dist measures the Hamming distance.

Remark 1.2. Alice can pick some function Enc that maps the bijectively maps the space $\{0,1\}^k$ of possible messages to some $\mathcal{C}$ of size 2^k, and if she wishes to send x to Bob, she sends him the bits of $\text{Enc}(x)$. If the number of corruptions is $< \Delta(\mathcal{C})/2$, then Bob can recover Alice's message by rounding his corrupted word y to the closest codeword y^*, which has to equal $\text{Enc}(x)$, and then computing $\text{Enc}^{-1}(y^*)$.

To get a sense of what rate and distance are acceptable to us, we define the notion of a *good code*.

Definition 1.3 (Good code)**.** We say that a family of codes $\mathcal{C}_n$ is a family of *good codes* if there are absolute constants $r_0 > 0$ and $\Delta_0 > 0$ such that $r(\mathcal{C}_n) \geqslant r_0$ and $\Delta(\mathcal{C}_n) \geqslant \Delta_0$. I.e. the rate and distance do not depend on n.

One of the first things you are told about codes is that a random code is a good code! However, it is impossible to succinctly describe a random linear code, or give an efficient algorithm for encoding and decoding.

It turns out that a random *linear* code is also a good code—i.e. the code $\mathcal{C}$ is a random k-dimensional subspace of $\mathbb{F}_2^n$. The proof of this fact will be relegated to homework—the distance of random linear codes is related to the expansion of an associated Cayley graph over $\mathbb{F}_2^n$!

A random linear code is a step up from a fully random code, because for one, it is efficiently storeable, and there is an efficient encoding algorithm too: consider a $n \times k$ matrix A whose columns span the code, and simply encode a message x with Ax. However, it is completely unclear how one efficiently decodes a corrupted codeword from a random linear code, and it is even possibly a hard problem.

Today, we will see a construction of a good code based on expander graphs!

2 Codes from expanders

Fix a d-regular λ-spectral expander $G = (V, E)$ on $|V| = n$ vertices (so $|E| = m = \frac{dn}{2}$), where $\lambda < \varepsilon$ is a sufficiently small constant. Let $C_0 \subseteq \{0,1\}^d$ be a linear "small code" of rate $r_0 > 1/2$ and (absolute) distance at least $2\varepsilon d$.

Definition of the code. We define a "big code" $C \subseteq \{0,1\}^E$ whose coordinates are indexed by the edges of G. For $x \in \{0,1\}^E$ and a vertex $v \in V$, fix an arbitrary ordering of the edges incident to v as $(v,1), \ldots, (v,d)$, and write

$$x|_{N(v)} = \big(x_{(v,1)}, \ldots, x_{(v,d)}\big) \in \{0,1\}^d.$$

Then $x \in C$ iff for every $v \in V$ we have the local constraint $x|_{N(v)} \in C_0$.

Linearity and rate. Since C_0 is a linear subspace, there is a parity-check matrix $A \in \{0,1\}^{(1-r_0)d \times d}$ with $C_0 = \{z \in \{0,1\}^d : Az = 0\}$. Imposing $Az = 0$ independently at each vertex gives $(1 - r_0)d$ linear constraints per vertex, for a total of $(1 - r_0)dn$ linear equations over $\mathbb{F}_2$. Because $m = \frac{dn}{2}$ variables live on edges while constraints live on vertices, standard counting (and independence of local checks up to edge sharing) yields

$$\dim(C) \geqslant m - (1 - r_0)dn = dn\big(r_0 - \tfrac{1}{2}\big),$$

so

$$r(C) = \frac{\dim(C)}{m} \geqslant 2r_0 - 1.$$

In particular, C has good rate whenever $r_0 > 1/2$.

Distance via expansion. A basic fact about linear codes is that the distance is equal to the minimum Hamming weight codeword in the space.

Let $x \in C$ be a nonzero codeword, and let $F \subseteq E$ be the support of x. Let $S := V(F)$ be the set of vertices incident to at least one edge of F. Because each local view $x|_{\Gamma(v)}$ is a codeword of C_0 of distance $\geqslant 2\varepsilon d$, we have, for every $v \in S$,

$$\deg_F(v) \in \{0\} \cup [\,2\varepsilon d, d\,],$$

and in particular $\deg_F(v) \geqslant 2\varepsilon d$ for all $v \in S$. Summing over $v \in S$ and using $\sum_{v \in S} \deg_F(v) = 2|F|$,

$$2|F| \geqslant (2\varepsilon d)\,|S| \qquad \Rightarrow \qquad |S| \leqslant \frac{|F|}{\varepsilon d}.$$

Every edge in F has both endpoints in S, so $|F| \leqslant e(S)$. By the expander mixing lemma and $\lambda < \varepsilon$,

$$|F| \leqslant e(S) \leqslant \frac{d}{2n}|S|^2 + \frac{\lambda d}{2}|S| \leqslant \frac{d}{2n}\left(\frac{|F|}{\varepsilon d}\right)^2 + \frac{\varepsilon d}{2}\left(\frac{|F|}{\varepsilon d}\right) = \frac{|F|^2}{2\varepsilon^2 nd} + \frac{|F|}{2}.$$

Rearranging gives

$$\frac{|F|}{2} \leqslant \frac{|F|^2}{2\varepsilon^2 nd} \qquad \Rightarrow \qquad |F| \geqslant \varepsilon^2 nd = 2\varepsilon^2 m.$$

Hence every nonzero codeword has Hamming weight at least $2\varepsilon^2$ times the blocklength m, and so the (relative) distance of C satisfies

$$\delta(C) \geqslant 2\varepsilon^2.$$

References

[SS02] Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 2002. 1