High Dimensional Expanders

Lecture 6

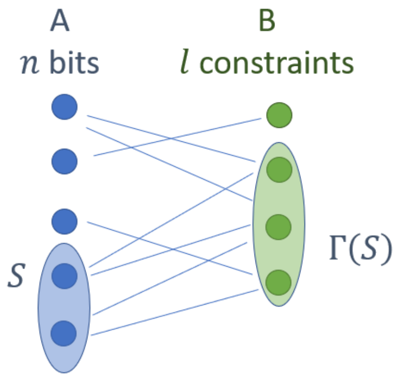
Lecture by: Irit Dinur, Weizmann Institute of Science *Scribe: Dafna Sadeh*

In this lecture we define a variant of expansion that focuses on small set, called Small Set Expander and show a randomized construction of such graphs. We demonstrate how to use this expander as a constraint graph in order to build an error correcting code with a constant relative distance. Finally, we show a novel construction by Vadhan of an HDX Cayley graph related to this code.

1 From expander to error correcting code

A bipartite graph (A, B, E) is (d, e) regular if for each $a \in A$ $deg(a) = d$ and for each $b \in B$ $deg(b) = e$.

Definition 1.1. *Small Set Expander. A (d, e) regular bipartite graph (A, B, E) is an (α, h)-Small Set Expander (SSE) if $\forall S \subseteq A \quad s.t \quad |S| \le \alpha|A|, \quad |\Gamma(S)| \ge hd|S|$ where $\Gamma(S) = \{b \in B | \exists a \in S. \quad (a, b) \in E\}$.*



Using the probabilistic method we construct a SSE.

Lemma 1.2. *randomized construction of SSE. There exist $d, n_0 \in \mathbb{N}$ s.t for any $n > n_0$ a random bipartite graph on $|A| = n$ and $|B| = \frac{3}{4}n = m$ chosen by letting each $u \in A$ choose d neighbors independently is a $(\alpha = \frac{1}{100d}, h = \frac{3}{4})$ SSE with probability > 0.*

Proof. Fix any $S \subset A$ of size $|S| = s \le \alpha n = \frac{n}{100d}$ and $T \subset B$ of size $|T| = t = hd|S| = \frac{3}{4}d|S|$. The probability of a "*Bad* event" that $\Gamma(S) \subseteq T$ is $\le \left(\frac{t}{m}\right)^{ds}$. We sum over the probability of all possible bad events and get that:

$$\sum_{s=1}^{\alpha n} \binom{n}{s} \Pr(Bad(S, T) \text{ for } |S| = s \text{ and } |T| = hds) < 1,$$

Which means that Pr(The graph is SSE) > 0.

$$
\begin{aligned}
\sum_{s=1}^{\alpha n}\binom{n}{s}\binom{m}{t}(\frac{t}{m})^{ds} &\overset{(1)}{\leq} \sum_{s=1}^{\alpha n}(\frac{ne}{s})^{s}(\frac{me}{t})^{t}\left(\frac{t}{m}\right)^{ds} \overset{(2)}{=} \\
&\sum_{s=1}^{\alpha n} e^{s+\frac{3}{4}ds}(\frac{n}{s})^{s}(\frac{m}{t})^{\frac{3}{4}ds}(\frac{t}{m})^{ds} \overset{(3)}{=} \\
&\sum_{s=1}^{\alpha n} e^{s+\frac{3}{4}ds}(\frac{n}{s})^{s}(\frac{t}{m})^{\frac{ds}{4}} \overset{(4)}{=} \\
&\sum_{s=1}^{\alpha n} e^{ds(\frac{1}{d}+\frac{3}{4})}(\frac{n}{s})^{s}(\frac{ds}{n})^{\frac{ds}{4}} \overset{(5)}{=} \\
&\sum_{s=1}^{\alpha n} e^{ds(\frac{1}{d}+\frac{3}{4})}(\frac{n}{s})^{s}(\frac{ds}{n})^{s}(\frac{ds}{n})^{\frac{ds}{4}-s} \overset{(6)}{=} \\
&\sum_{s=1}^{\alpha n} e^{ds(\frac{1}{d}+\frac{3}{4}+\frac{\ln d}{d})}\left(\left(\frac{ds}{n}\right)^{\frac{1}{4}-\frac{1}{d}}\right)^{ds} \overset{(7)}{\leq} \\
&\sum_{s=1}^{\alpha n} e^{ds(\frac{1}{d}+\frac{3}{4}+\frac{\ln d}{d})}\left(\left(\frac{1}{100}\right)^{\frac{1}{4}-\frac{1}{d}}\right)^{ds} \overset{(8)}{\leq} \\
&\sum_{s=1}^{\alpha n} \frac{1}{4}^{ds} < \frac{1}{3} < 1
\end{aligned}
\tag{1.1}
$$

The first inequality is due do Stirling approximation, the 4th equality is by the definitions of t and m and the 7th and 8th inequality are for large enough n_0 and d. □

Remarks:

- If (A, B, E) is a random (d, e)-regular graph then the analysis is more subtle, but it still true. In our analysis the B side isn't necessarily regular. This makes the probabilities independent and easier to analyze.
- If we want to analyze the spectrum we can use Cheeger's inequality.
- Friedman proved a near-optimal spectral bound: $\lambda_2 \leq 2\sqrt{d-1}+\epsilon$

Claim 1.3. Unique Neighbor Claim. If a (d, e) regular bipartite graph (A, B, E), is an (α, h)-small set expander for $h > 1/2$, then every set $S \subset A$ of size $|S| \leq \alpha|A|$ has a unique neighbor i.e $\exists v \in B$ s.t v is adjacent to exactly one element in S.

Proof. Fix S of size $|S| \leq \alpha|A|$ and let $T = \Gamma(S)$. Assume by contradiction that every $v \in T$ has at least two neighbors in S, then:

$$d|S| < 2hd|S| \leq 2|T| \leq E(S, T) = d|S|.$$

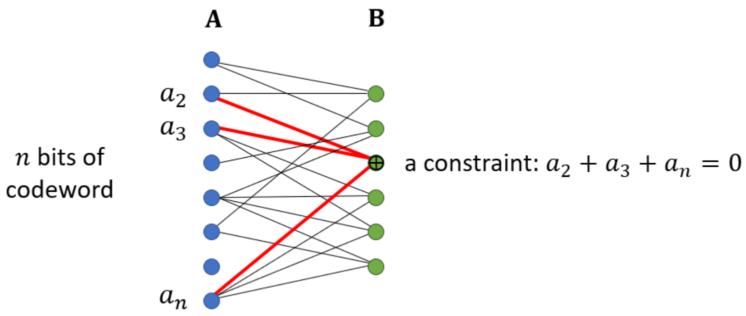
□

We use (d, e) regular bipartite graphs $F = (A, B, E)$ as a constraint graphs to build a linear code. Assuming $|A| = n$ and $|B| = m$ we define:

$$C_F = \{w \in \{0,1\}^n \mid \forall b \in B. \sum_{i\sim b} w_i = 0 \ mod\ 2\}.$$

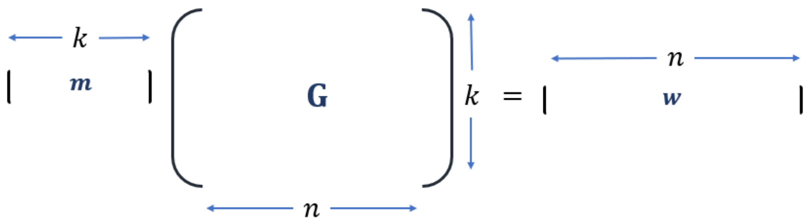
Each constraint is over e bits. C_F is a linear sub-space of dimension $dim(C_F) \leq n - m = |A| - |B|$.

Claim 1.4. from SSE to code. Assume (d, e) regular bipartite graph (A, B, E) is an (α, h)-small set expander for $h > 1/2$. Then C_F is a linear code of relative distance $\geq \alpha$.

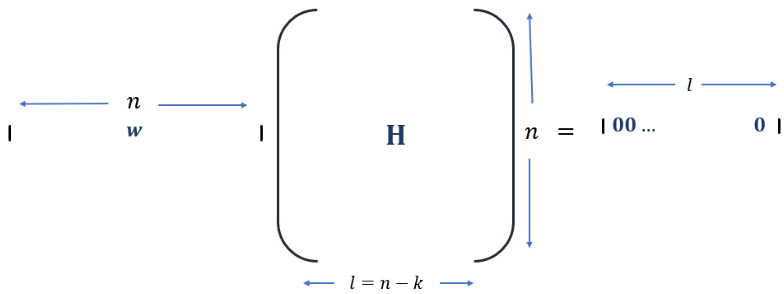


Proof. Let $w \neq w' \in C_F$ and define $S = \{i \in [n] | w_i \neq w'_i\}$. Assume by contradiction that $|S| < \alpha n$, then S has a unique neighbor v that adjacent to exactly one element in S. Since $\sum_{i \sim v}(w \oplus w')_i = 1$, we get that $\sum_{i \sim v} w_i = \sum_{i \sim v} w'_i + 1$ which means that the constraint defined by v is unsatisfied either by w or by w'. A contradiction to $w, w' \in C$. $\square$

Low Density Parity Check Codes (LDPC) [Gallager 63] The idea of building of a code from a set of sparse constraints is well studied. This codes are called LDPC for low density parity check. Linear codes can be given by a $k \times n$ generating matrix G over $\mathbf{F}_2$ as $C = \{m^T G | m \in \mathbf{F}_2^k\} \subseteq \mathbf{F}_2^n$.



An alternative way to give a code is using the linear relations that the coordinates of a code word must satisfy. Formally, define H, an $n \times l$, $(l = n - k)$ parity check matrix over $\mathbf{F}_2$, $C = \{w \in \mathbf{F}_2^n | w^T H = 0\} = \textit{left\_}\ker(H)$.



H is a "parity check" in that every column of H is a parity check constraint. Observe that the rows of G are the basis of the words that satisfy the H-constraints, thus, $\text{span}(col\,(H)) = (rows\,(G))^{\perp}$. Each column of H is a linear constraint on the words in C. We say that H is an LDPC if it has a "few" 1's it is.

Q: Can G be sparse?

A: No. Since each row of G is a code word, thus must have $\geq \delta n$ 1's in a code of relative distance δ.

Q: Can H be sparse?

A: Yes. We need to find a sparse basis for $(Rows(G))^{\perp}$

H can be viewed in a combinatorial way as a constraint graph.

2 Construction of Cayley HDX - Salil Vadhan Nov' 18

Let $F = (A, B, E)$ be a right 3-regular bipartite graph, $A = \{a_1, a_2, ..., a_n\}$, $|B| = 0.99n$. We use F as a constraint graph for the code $C_F = \{w \in \{0,1\}^n | \sum_{i \sim v} w_i = 0 \quad \forall v \in B\}$ - "the left kernel of F". Let $G_{k \times n}$ for $k \geq 0.01n$, be a generating matrix for C_F. Since each $a_i \in A$ represents a bit in location i in a code word, there is a matching between A to the columns of G. Let $S = \{s_1, s_2, ..., s_n\}$ be the columns of G.

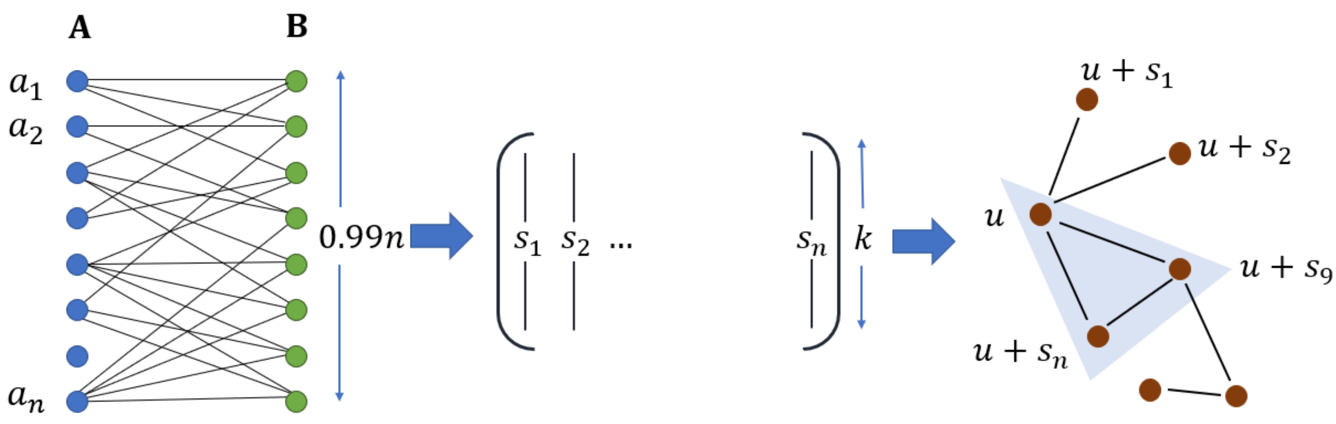
We look at the Cayley graph $Cay(\{0,1\}^k, S)$. Let $X = X(0), X(1), X(2)$ be the "clique complex" of this graph.

$X(0) = \{0,1\}^k$, $|X(0)| \geq 2^{0.01n}$

$X(1) = \{(u, v) | u = v + s \text{ for some } s \in columns(G)\}$

$X(2) = \{(u_1, u_2, u_3) | \text{ all 3 edges } (u_1, u_2), (u_1, u_3) \text{ and } (u_2, u_3) \text{ belongs to } X(1)\}$

Observe that $(X(0), X(1))$ is the Cayley graph $Cay(\{0,1\}^k, S)$, thus it's a $\lambda_2 \leq 1 - \delta$ expander (we have proved this in Lecture 5), with low degree - $\forall u \in X(0) \ \deg(u) = n$, logarithmic in $|X(0)|$.



Constraint graph F — *Generating matrix of a code G* — *Clique complex*

Fix some $u \in \{0,1\}^k$. We look on the vertices of the link $X_u(0) = \{u + s_1, u + s_2, ..., u + s_n\}$ and match them to A, by the bijection $h : X_u(0) \to A$ given by $h(u + s_i) = a_i$.

Claim 2.1. $u + s_i, u + s_j \in X_u(0)$ are connected by an edge in $X(1)$ (and in $X_u(1)$) iff $dist_F(h(u + s_i), h(u + s_j)) = dist_F(a_i, a_j) = 2$.

Proof. ($\Leftarrow$): $dist_F(a_i, a_j) = 2$ means that $\exists$ a constraint $x \in B$ neighboring both a_i and a_j. Let a_k be the third neighbor of x, which means that for any codeword w of C, $w_i + w_j + w_k = 0$ and thus $s_i + s_j + s_k = 0$. We get that $u + s_i = u + s_j + s_k$, therefore, by definition, $u + s_i$ and $u + s_j$ are connected by an edge in $X(1)$.
($\Rightarrow$) Assume $u + s_i, u + s_j \in X_u(0)$ are connected by an edge in $X(1)$, this means that exists some $s_k \in S$ such that $u + s_i = u + s_j + s_k \Rightarrow s_i + s_j + s_k = 0$. We gets that in every code word $w \in C_F$, $w_i + w_j + w_k = 0$. Assuming the dependencies between the constraints in F don't create new linear constraints of three bits, we gets that there exists $x \in B$ adjacent to a_i, a_j and $a_k \Rightarrow dist_F(a_i, a_j) = 2$. □

The meaning of this claim is that the link of every vertex represents a two steps walk in a random graph and thus also an expander.