Chapter 7

Expander Graphs

In this chapter, we first define expander graphs and see some of their properties. Then, we study a deterministic combinatorial construction of expander graphs, called the zig-zag products. Then, we discuss various interesting and important applications of expander graphs. Most of the material in this chapter is extracted from the excellent survey by Hoory, Linial, and Wigderson [HLW06].

There are several possible ways to define regular expander graphs.

1. Combinatorically, expander graphs are graphs with very good "connectivity", e.g. graphs with good edge expansion or vertex expansion.

2. Probabilistically, expander graphs are graphs in which random walks mix rapidly.

3. Algebraically, expander graphs are graphs with a large spectral gap $\alpha_1 - \alpha_2$.

We have already seen in chapter 4 and chapter 6 that these definitions are closely related. Cheeger's inequality in Theorem 4.3 states that a graph has a large spectral gap if and only if its edge expansion is large. The spectral analysis in Corollary 6.17 and Problem 6.21 show that lazy random walks mix quickly if and only if the spectral gap is large.

Note that complete graphs are the best expander graphs in each of the above definitions, but we are interested in sparse expander graphs with linear number of edges, that is, d-regular expander graphs with constant d. In constructions of expander graphs, the spectral definition is the most convenient, and we will use the following stronger spectral defintion that also bounds the last eigenvalue.

Definition 7.1 (Spectral Expanders)**.** *Let G be a d-regular graph and let $d = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq -d$ be the spectrum of its adjacency matrix. We say that G is an (n, d, ϵ)-graph if it has n vertices, is d-regular, and with $\max\{\alpha_2, |\alpha_n|\} \leq \epsilon d$. The quantity $\alpha := \max\{\alpha_2, |\alpha_n|\}$ is called the spectral radius of the graph.*

The smaller is the spectral radius, the stronger the graph is as a spectral expander. Probabilistically, the spectral radius is small if and only if the non-lazy random walks mix rapidly, as shown in Theorem 6.16. Combinatorically, $|\alpha_n|$ is small if and only if there is no nearly bipartite component, as shown in Theorem 5.4.

7.1 Properties of Expander Graphs

We collect more combinatorial and probabilistic properties of a spectral expander in this section.

Expander Mixing Lemma

A well-known and useful property of expander graphs is that it behaves as a random d-regular graph. Consider the number of edges between two subsets S, T of vertices.

Definition 7.2 (Induced Edges)**.** *Given a graph $G = (V, E)$ and $S, T \subseteq V$, define $E(S,T) := \{(u,v) \mid u \in S, v \in T, uv \in E\}$ be the set of ordered pairs where $u \in S$ and $v \in T$. Note that an edge with $u \in S \cap T$ and $v \in S \cap T$ is counted twice, as both (u,v) and (v,u) are in $E(S,T)$.*

In a random graph where every pair of vertices has an edge with probability $\frac{d}{n}$, we expect that $|E(S,T)|$ is close to $\frac{d}{n}|S||T|$. The expander mixing lemma says that in a spectral expander $|E(S,T)|$ is close to this expectation.

Theorem 7.3 (Expander Mixing Lemma)**.** *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. If the spectral radius of G is α, then for every $S \subseteq V$ and $T \subseteq V$,*

$$\left| \left|E(S,T)\right| - \frac{d|S||T|}{n} \right| \leq \alpha\sqrt{|S||T|}.$$

Proof. First, we write $|E(S,T)|$ as an algebraic expression. Let χ_S and χ_T be the characteristic vectors of S and T, such that $\chi_S(i) = 1$ if $i \in S$ and $\chi_S(i) = 0$ if $i \notin S$. Notice that $|E(S,T)| = \chi_S^T A \chi_T$, where A is the adjacency matrix of G.

Then, we use eigen-decompositions of χ_S and χ_T to relate $|E(S,T)|$ to the eigenvalues of A. Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors of A. Recall that $\alpha_1 = d$ and $v_1 = \frac{1}{\sqrt{n}}\vec{1}$. Write $\chi_S = \sum_{i=1}^n a_i v_i$ and $\chi_T = \sum_{i=1}^n b_i v_i$ as linear combination of the eigenvectors. So, $a_1 = \langle \chi_S, v_1 \rangle = \frac{|S|}{\sqrt{n}}$ and $b_1 = \langle \chi_T, v_1 \rangle = \frac{|T|}{\sqrt{n}}$. Then, by orthonormality of $v_1, \ldots, v_n$,

$$\left|E(S,T)\right| = \chi_S^T A \chi_T = \sum_{i=1}^n \alpha_i a_i b_i = \frac{d|S||T|}{n} + \sum_{i=2}^n \alpha_i a_i b_i.$$

Therefore, by the definition of spectral radius and an application of the Cauchy-Schwarz inequality,

$$\left| \left|E(S,T)\right| - \frac{d|S||T|}{n} \right| \leq \left| \sum_{i=2}^n \alpha_i a_i b_i \right| \leq \alpha \sum_{i=2}^n |a_i||b_i| \leq \alpha \|\vec{a}\|_2 \|\vec{b}\|_2 = \alpha \|\chi_S\|_2 \|\chi_T\|_2 = \alpha\sqrt{|S||T|},$$

where $\vec{a} = (a_1, \ldots, a_n)$ and $\vec{b} = (b_1, \ldots, b_n)$. □

The following is a consequence of the expander mixing lemma.

Exercise 7.4 (Maximum Independent Set of Spectral Expanders)**.** *Let $G = (V, E)$ be a d-regular graph with $V = [n]$ with spectral radius α. Show that the size of a maximum independent set is at most $\frac{\alpha n}{d}$. Conclude that an (n, d, ϵ)-graph has chromatic number at least $\frac{1}{\epsilon}$.*

Converse of Expander Mixing Lemma

Interestingly, Bilu and Linial [BL06] proved a converse of the expander mixing lemma, showing that it comes close in characterizing the spectral radius of a graph.

Theorem 7.5 (Converse of Expander Mixing Lemma [BL06]). *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. Suppose that for any subsets $S, T \subseteq V$ with $S \cap T \neq \emptyset$, it holds that*

$$\left| |E(S,T)| - \frac{d|S||T|}{n} \right| \leq \alpha \sqrt{|S||T|}.$$

Then all but the largest eigenvalue of $A(G)$ are bounded in absolute value by $O\big(\alpha(1 + \log \frac{d}{\alpha})\big)$.

The proof of Theorem 7.5 is based on the following linear algebraic lemma.

Lemma 7.6 (Bounding Spectral Radius [BL06]). *Let A be an $n \times n$ real symmetric matrix such that the ℓ_1-norm of each row of A is at most d, and all diagonal entries of A are with absolute value $O\big(\alpha \log(\frac{d}{\alpha}) + 1\big)$. Suppose that for any two vectors $u, v \in \{0,1\}^n$ with $\mathrm{supp}(u) \cap \mathrm{supp}(v) = \emptyset$, it holds that $|u^T A v| \leq \alpha \|u\|_2 \|v\|_2$. Then the spectral radius of A is $O\big(\alpha \log(\frac{d}{\alpha}) + 1\big)$.*

The proof of Lemma 7.6 is based on linear programming duality and is not quite intuitive. It would be very interesting if there is a proof of Theorem 7.5 which is of a similar style of Trevisan's proof of Cheeger's inequality in Theorem 4.3.

Vertex Expansion

Cheeger's inequality proves that a d-regular spectral expander has large edge expansion. One could lower bound the vertex expansion of a d-regular graph through edge expansion, but with a factor d loss. Tanner's theorem proves a much stronger lower bound than that followed from edge expansion.

Definition 7.7 (Vertex Boundary). *Let $G = (V, E)$ be an undirected graph. For $S \subseteq V$, the open vertex boundary of S is defined as $\partial(S) := \{v \in V - S \mid \exists u \in S \text{ with } uv \in E\}$, and the closed vertex boundary of S is defined as $\partial[S] := S \cup \partial(S)$.*

Definition 7.8 (Vertex Expansion). *Let $G = (V, E)$ be an undirected graph. The vertex expansion of a subset $S \subseteq V$ and of a graph are defined as*

$$\psi(S) := \frac{|\partial(S)|}{|S|} \quad \text{and} \quad \psi(G) := \min_{S : |S| \leq |V|/2} \psi(S).$$

Theorem 7.9 (Tanner's Theorem). *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. Suppose the spectral radius of G is at most ϵd for some $0 < \epsilon < 1$. Then, for any $0 < \delta \leq 1/2$, for any subset $S \subseteq V$ with $|S| = \delta n$,*

$$\psi(S) \geq \frac{1}{\delta(1 - \epsilon^2) + \epsilon^2} - 1.$$

Proof. The key is to consider the quantity $\|A\chi_S\|_2^2$, where A is the adjacency matrix and χ_S is the characteristic vector of $S \subseteq V$. For a vertex $v \in V$, let $\deg_S(v) := |\{u \in S \mid uv \in E\}|$ be the number of neighbors of v in S. On one hand,

$$\|A\chi_S\|_2^2 = \sum_{v \in V} \deg_S(v)^2 = \sum_{v \in \partial[S]} \deg_S(v)^2 \geq \frac{\big(\sum_{v \in \partial[S]} \deg_S(v)\big)^2}{|\partial[S]|} = \frac{(d|S|)^2}{|\partial[S]|}.$$

On the other hand, we write $\chi_S = \sum_{i=1}^n c_i v_i$ as a linear combination of the orthonormal eigenvectors of A, with $v_1 = \vec{1}/\sqrt{n}$ and $c_1 = \langle \chi_S, v_1 \rangle = |S|/\sqrt{n}$. Then

$$\|A\chi_S\|_2^2 = \Big\|\sum_{i=1}^n c_i \alpha_i v_i\Big\|_2^2 = \sum_{i=1}^n c_i^2 \alpha_i^2 \le \frac{d^2|S|^2}{n} + (\epsilon d)^2\big(\|\chi_S\|^2 - c_1^2\big) = d^2|S|\big(\delta + \epsilon^2(1-\delta)\big),$$

where the second equality is by orthonormality of $v_1, \dots, v_n$, the inequality is by the assumption of the spectral radius and $\sum_{i=1}^n c_i^2 = \|\chi_S\|^2$, and the final equality is by plugging in $|S| = \delta n$. Combining the inequalities yields the theorem. □

Note that when $\delta \ll \epsilon^2$, Tanner's theorem gives $\psi(S) \gtrsim 1/\epsilon^2$, which implies that $|\partial(S)|$ is much larger than $|S|$ when $|S|$ is small enough. Check that a straightforward application of Cheeger's inequality only gives $\psi(S) \ge \frac{1}{2}(1-\epsilon)$ for $|S| \le |V|/2$.

Alon-Boppana Bound

How small can the spectral radius be? There are graphs, called Ramanujan graphs, with spectral radius $2\sqrt{d-1}$. This is essentially tight, as the following theorem by Alon and Boppana showed.

Theorem 7.10 (Alon-Boppana Bound)**.** *Let $G = (V, E)$ be a d-regular graph and α_2 be the second largest eigenvalue of its adjacency matrix. Then*

$$\alpha_2 \ge 2\sqrt{d-1} - \frac{2\sqrt{d-1} - 1}{\lfloor \mathrm{diag}(G)/2 \rfloor},$$

where $\mathrm{diag}(G)$ *denotes the diameter of the graph G.*

Note that the theorem implies that if we have an infinite family of d-regular expander graphs each has spectral radius at most α, then $\alpha \ge 2\sqrt{d-1}$ as the diameter goes to infinity as the size of the graph grows.

There are two different proofs of this result. One is by the trace method, which computes the number of closed walks that starts and ends at some given vertex. Another is by constructing a function with small Rayleigh quotient. These two methods can also be used to solve Problem 3.10, which is closely related to the spectral radius of Ramanujan graphs as we will see later in the course.

We will not prove Theorem 7.10 and refer the reader to [HLW06] or Trevisan's blog posts for proofs. We just present an easy proof that the spectral radius is at least $\sqrt{d}\big(1 - o(1)\big)$, using a very simple trace argument.

Claim 7.11 (Easy Lower Bound on Spectral Radius)**.** *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. Then its spectral radius α is at least $\sqrt{d}\sqrt{\frac{n-d}{n-1}}$.*

Proof. Note that $\mathrm{Tr}(A^2) \ge nd$, as each edge uv contributes one length-two walk from u to u and one length-two walk from v to v. On the other hand, by Fact 2.35, $\mathrm{Tr}(A^2) = \sum_{i=1}^n \alpha_i^2 \le d^2 + (n-1)\alpha^2$. Combining the two inequalities gives the claim. □

Random Walks on Expander Graphs

We know from Theorem 6.16 that random walks on a (n, d, ϵ)-graph converge to the uniform distribution in $O(\log n/(1-\epsilon))$ steps. Interestingly, random walks on expander graphs not only give good randomness properties for the final vertex in the walk, but also for the sequence of vertices traversed in the walk. In some applications, the sequence of vertices of a walk can be used to replace a sequence of indpendent uniform random variables.

The following result is not of the most general form, but it will be enough for the application of probability amplication that we will see in section 7.3. See [HLW06, Vad12] for more general statements. To get an intuition, it is instructive to compare the probability bound below with the probability bound when each X_i is an independent uniform random sample.

Theorem 7.12 (Concentration Property of Random Walks on Spectral Expanders)**.** *Let $G = (V, E)$ be a d-regular graph with spectral radius ϵd for some $\epsilon \leq 1/10$. Let $B \subseteq V$ with $|B| \leq \frac{1}{100}|V|$. Let X_0 be a uniform random vertex, and $X_1, \ldots, X_t$ be the vertices produced by t steps of a random walk. Let $S = \{i \mid X_i \in B\}$ be the set of times when the random walk is in B. Then*

$$\Pr\left(|S| > \frac{t}{2}\right) \leq \left(\frac{2}{\sqrt{5}}\right)^{t+1}.$$

Proof. We first set up the matrix formulation of the problem. Let $n = |V|$. The initial distribution $p_0 = \vec{1}/n$ of X_0 is the uniform distribution. Let χ_B and $\chi_{\overline{B}}$ be the characteristic vectors of B and $\overline{B}$ respectively, where $\overline{B} = V - B$. Let I_B be the diagonal matrix with a 1 in the i-th diagonal entry if $i \in B$ and zero otherwise, and similarly $I_{\overline{B}}$. Let p be a probability vector, i.e. p is non-negative and the sum of its entries is at most one. Then $I_B \cdot p$ is the probability vector that is the restriction of p on B, such that $(I_B \cdot p)(i) = p(i)$ if $i \in B$ and $(I_B \cdot p)(i) = 0$ if $i \notin B$. Check that the probability the random walk is in B at precisely the time steps in S is

$$p_S := \vec{1}^T (I_{Z_t}\mathcal{A})(I_{Z_{t-1}}\mathcal{A})(I_{Z_{t-2}}\mathcal{A}) \ldots (I_{Z_2}\mathcal{A})(I_{Z_1}\mathcal{A})p_0,$$

where $Z_i = B$ if $i \in S$ and $Z_i = \overline{B}$ if $i \notin S$, and $\mathcal{A}$ is the normalized adjacency matrix which is the probability transition matrix of the random walks. We will prove that $p_S \leq (\frac{1}{5})^{|S|}$. The theorem will then follow by a union bound as

$$\Pr\left(|S| > \frac{t}{2}\right) \leq \sum_{S:|S|>t/2} p_S \leq \sum_{S:|S|>t/2} \left(\frac{1}{5}\right)^{|S|} \leq \sum_{S:|S|>t/2} \left(\frac{1}{5}\right)^{\frac{t+1}{2}} \leq 2^{t+1}\left(\frac{1}{5}\right)^{\frac{t+1}{2}} = \left(\frac{2}{\sqrt{5}}\right)^{t+1}.$$

To prove $p_S \leq (\frac{1}{5})^{|S|}$, we use the concept of operator norm in Definition 2.17. Check that $\|I_B\|_{\text{op}} = \|I_{\overline{B}}\|_{\text{op}} = \|\mathcal{A}\|_{\text{op}} = 1$. We will prove that $\|I_B\mathcal{A}\|_{\text{op}} \leq \frac{1}{5}$, and this would imply that $p_S \leq (\frac{1}{5})^{|S|}$ because

$$\begin{aligned}
p_S &= \vec{1}^T (I_{Z_t}\mathcal{A})(I_{Z_{t-1}}\mathcal{A})(I_{Z_{t-2}}\mathcal{A}) \ldots (I_{Z_2}\mathcal{A})(I_{Z_1}\mathcal{A})p_0 \\
&\leq \|\vec{1}\|_2 \cdot \left\|(I_{Z_t}\mathcal{A})(I_{Z_{t-1}}\mathcal{A})(I_{Z_{t-2}}\mathcal{A}) \ldots (I_{Z_2}\mathcal{A})(I_{Z_1}\mathcal{A})p_0\right\|_2 && \text{by Cauchy-Schwarz} \\
&\leq \|\vec{1}\|_2 \cdot \left(\prod_{i=1}^{t} \|I_{Z_i}\mathcal{A}\|_{\text{op}}\right) \cdot \|p_0\|_2 && \text{by Fact 2.19} \\
&\leq \|\vec{1}\|_2 \cdot \left(\frac{1}{5}\right)^{|S|} \cdot \|p_0\|_2 && \text{as } \|I_B\mathcal{A}\|_{\text{op}} \leq \frac{1}{5} \text{ and } \|I_{\overline{B}}\mathcal{A}\|_{\text{op}} \leq 1 \\
&= \left(\frac{1}{5}\right)^{|S|} && \text{as } \|1\|_2 = \sqrt{n} \text{ and } \|p_0\|_2 = \frac{1}{\sqrt{n}}.
\end{aligned}$$

It remains to prove that $\|I_B\mathcal{A}\|_{\text{op}} \le \frac{1}{5}$. Let x be any nonzero vector. Write $x = c_1v_1 + \ldots + c_nv_n$, where $v_1, \ldots, v_n$ are the orthonormal eigenvectors of $\mathcal{A}$ with eigenvalues $\alpha_1 \ge \ldots \ge \alpha_n$. Then

$$\|I_B\mathcal{A}x\|_2^2 = \|I_B\mathcal{A}(c_1v_1 + \ldots + c_nv_n)\|_2^2 = \left\|I_B\sum_{i=1}^n c_i\alpha_iv_i\right\|_2^2 \le 2\|I_Bc_1\alpha_1v_1\|_2^2 + 2\left\|I_B\sum_{i=2}^n c_i\alpha_iv_i\right\|_2^2,$$

where the inequality is by $\|x+y\|_2^2 \le 2\|x\|_2^2 + 2\|y\|_2^2$. Recall that $\alpha_1 = 1$, $v_1 = \vec{1}/\sqrt{n}$ and $c_1 = \langle x, v_1\rangle = \frac{1}{\sqrt{n}}\cdot\sum_{i=1}^n x(i)$. So, the first term on the RHS is

$$2\|I_Bc_1\alpha_1v_1\|_2^2 = 2\left\|\frac{1}{n}\Big(\sum_{i=1}^n x(i)\Big)I_B\vec{1}\right\|_2^2 = 2|B|\left(\frac{\sum_{i=1}^n x(i)}{n}\right)^2 \le 2|B|\cdot\frac{\|x\|_2^2}{n} \le \frac{1}{50}\|x\|_2^2,$$

where the first inequality is by Cauchy-Schwarz and the second inequality is by the assumption that $|B| \le \frac{n}{100}$. The second term on the RHS is

$$2\left\|I_B\sum_{i=2}^n c_i\alpha_iv_i\right\|_2^2 \le 2\|I_B\|_{\text{op}}^2\cdot\left\|\sum_{i=2}^n c_i\alpha_iv_i\right\|_2^2 = 2\sum_{i=2}^n c_i^2\alpha_i^2 \le 2\epsilon^2\sum_{i=2}^n c_i^2 \le 2\epsilon^2\|x\|_2^2 \le \frac{1}{50}\|x\|_2^2,$$

where the equality is by orthnormality of $v_1, \ldots, v_n$, the second inequality is by the assumption that the spectral radius of the adjacency matrix A is at most ϵd and so the eigenvalues of $\mathcal{A} = A/d$ satisfies $\max_{2\le i\le n}\{|\alpha_i|\} \le \epsilon$, and the last inequality is by the assumption that $\epsilon \le \frac{1}{10}$. Combining the two terms,

$$\|I_B\mathcal{A}x\|_2^2 \le \frac{1}{25}\|x\|_2^2 \quad\Longrightarrow\quad \|I_B\mathcal{A}\|_{\text{op}} \le \frac{1}{5}.$$

□

See [Gil98] for a well-known Chernoff bound for spectral expanders, and [GLSS18] for a recent generalization to the matrix setting. See [CPT21] for an interesting recent paper showing that many functions are fooled by expander random walks, in that they cannot distinguish independent random samples from those obtained by expander random walks.

7.2 Constructions of Expander Graphs

It can be shown that a random d-regular graph is an expander graph with high probability using the combinatorial definitions, by standard techniques using Chernoff bound and union bound. It is a good problem to work out the details; see [HLW06, Vad12] for the precise statements and proofs.

Perhaps surprisingly, while almost every d-regular graph is an expander graph, it is very difficult to come up with a deterministic construction of expander graphs. One possible explanation is that random graphs have high descriptive complexity, while in deterministic constructions the infinite family of expander graphs can be described in a succinct way.

There are explicit constructions of d-regular expander graphs, most of them are algebraic constructions.

- A family of 8-regular graphs G_m for every integer m. The vertex set is $V = \mathbb{Z}_m \times \mathbb{Z}_m$. The neighbors of vertex (x, y) is $(x, y\pm x), (x\pm y, y), (x, y+1\pm x), (x+1\pm y, y)$, where all additions are mod m. Note that this family is very explicit, meaning that the neighbors of a vertex can

be computed in $O(\log m)$ time, which is very useful for some applications such as probability amplification as we will see. This construction is due to Margulis whose proof did not give any explicit bound. Gabber and Galil proved that its spectral radius is at most $5\sqrt{2} < 8$. Their proof uses Fourier analysis; see [HLW06].

- A family of 3-regular p-vertex graph for every prime number p. The vertex set is $\mathbb{Z}_p$, and a vertex is connected to $x+1, x-1$ and its multiplicative inverse x^{-1} (for vertex 0 its inverse is 0), where the additions are mod p. The proof uses some deep results in number theory.
- The main source of explicit deterministic construction is from Cayley graphs, which are graphs defined by groups. Some of the stronger expanders, the Ramanujan graphs with spectral radius $2\sqrt{d-1}$, are from Cayley graphs and the proofs require sophisticated mathematical tools.

In the second part of the course, we will see a new way to show the existence of "bipartite" Ramanujan graphs using combinatorial and probabilistic methods, through interlacing family of polynomials.

In the following, we will study a combinatorial construction of expander graphs, known as the zig-zag product, whose proof is more elementary and intuitive, although the bound is not as sharp and the construction is not as explicit.

Combinatorial Constructions

The general idea of the combinatorial constructions is to construct bigger expander graphs from smaller expander graphs.

The base case could simply be a constant size complete graph. Let G be an (n, k, ϵ_1)-graph and H be an (k, d, ϵ_2)-graph. A natural product of G and H is to replace each vertex v in G by a copy of H, so that each edge incident on v is incident on a different vertex of H. This is called the replacement product of G and H.

Definition 7.13 (Replacement Product)**.** *Let G be a k-regular graph on n vertices and H be a d-regular graph on k vertices. The replacement product $G \textcircled{r} H$ is a graph where the vertex set is the Cartesian product $[n] \times [k]$ of the vertex set of G and H, and two vertices (u, i) and (v, j) have an edge if and only if (1) $u = v$ and $ij \in E(H)$ or (2) $vu \in E(G)$ and v is the i-th neighbor of u in G and u is the j-th neighbor of v in G.*

Intuitively, $G \textcircled{r} H$ is a combinatorial expander if G and H are combinatorial expanders. Consider a set $S \subseteq V(G \textcircled{r} H)$. If S has either large of small intersection with each "cloud" (copy of H), then S should have large expansion because of the large expansion of G as S is basically a set of vertices in G. If S has medium intersections with many clouds, then S should have large expansion because of the large expansion of H as there are many crossing edges within each such cloud. However, it is not clear how to make this intuition precise, as there seems to be no clean way to decompose a subset's contribution into its contribution from G and its contribution from H. In a way, the spectral proof that we are going to see soon can be thought of as a linear algebraic approach to carry out this idea in a more general setting.

Zig-Zag Product

The actual construction by Reingold, Vadhan and Wigderson [RVW02] that we will analyze is slightly more complicated.

Definition 7.14 (Zig-Zag Product). *Let G be a k-regular graph on n vertices and H be a d-regular graph on k vertices. The zig-zag product $G \textcircled{z} H$ is a graph with the same vertex set $[n] \times [k]$ as the replacement product, and two vertices (u, i) and (v, j) have an edge if and only if $u \neq v$ and there exists $a \in [k]$ such that (u, i)–(u, a), (u, a)–(v, b), and (v, b)–(v, j) are all edges in the replacement product $G \textcircled{r} H$, where (u, a)–(v, b) is the unique edge incident on (u, a) with $v \neq u$ (i.e. the unique edge incident on (u, a) that leaves the cloud of u in the replacement product).*

In words, each edge in the zig-zag product $G \textcircled{z} H$ corresponds to a length three walk in the replacement product $G \textcircled{r} H$, where the first step is within a cloud, the second step is the unique way to leave a cloud, and the third step is within the other cloud.

The intuition that the zig-zag product is a spectral expander comes from random walks. Edge edge in $G \textcircled{z} H$ corresponds to a random step in H, a deterministic step in G, and a random step in H. We should think of the first two steps as going to random neighboring cloud, and the third step corresponds to moving to a random neighbor within the neighboring cloud. Since both G and H are spectral expanders and thus have the fast mixing property, after not many steps of random walks, we won't know which cloud we are in and the location within the cloud, and so $G \textcircled{z} H$ also has the fast mixing property and hence a spectral expander.

Theorem 7.15 (Zig-Zag Theorem). *Let G be an (n, k, ϵ_1)-graph and H be an (k, d, ϵ_2)-graph. Then $G \textcircled{z} H$ is an $(nk, d^2, \epsilon_1 + \epsilon_2 + \epsilon_2^2)$-graph.*

We will prove the theorem in the next subsection. Let us first see how zig-zag product can be used to construct bigger and bigger constant degree expander graphs. The idea is to combine with the following standard operation that decreases the spectral radius.

Definition 7.16 (Graph Power). *Let G be a graph with adjacency matrix A. The k-th power G^k is the graph with the same vertex set as G and with (weighted) adjacency matrix A^k.*

In words, the number of parallel edges between u and v in G^k is equal to the number of length k walks between u and v in G. Note that the spectral radius of G^k has improved significantly, but the degree of G^k has also improved significantly.

Exercise 7.17 (Spectral Radius of Graph Power). *If G is an (n, d, ϵ)-graph, then G^k is an (n, d^k, ϵ^k)-graph.*

The idea of the combinatorial construction is to use graph power to decrease the spectral radius, and then use zig-zag product to decrease the degree while not increasing the spectral radius too much.

Theorem 7.18 (Expanders from Zig-Zag Product). *For large enough constant d, there is an infinite family of d^2-regular with spectral radius at most $\frac{1}{4}d^2$.*

Proof. Let H be a $(d^4, d, 1/16)$-graph. We can prove its existence by a probabilistic argument when d is a large enough constant. Since d is a constant, one can find it by an exhaustive search in constant time.

Using the building block H, we inductively define G_i by $G_1 = H^2$ and $G_{i+1} = G_i^2 \textcircled{z} H$. We claim that G_i is a $(d^{4i}, d^2, 1/4)$-graph for all $i \geq 1$. The base case is clearly true by Exercise 7.17. Assume G_i is a $(d^{4i}, d^2, 1/4)$-graph. Then G_i^2 is a $(d^{4i}, d^4, 1/16)$-graph by Exercise 7.17. And $G_i^2 \textcircled{z} H$ is a $(d^{4(i+1)}, d^2, 1/4)$-graph by Theorem 7.15. □

Proof of the Zig-Zag Theorem

Check that $G \textcircled{z} H$ has nk vertices and is d^2-regular. We bound the spectral radius of $G \textcircled{z} H$ in the rest of this subsection.

Matrix Formulation: The first step is to write down the walk matrix Z of the zig-zag product $G \textcircled{z} H$. Let $W(H)$ be the $k \times k$ walk matrix of H, which is simply $\frac{1}{k}A(H)$ where $A(H)$ is the adjacency matrix of H. Let W be the $nk \times nk$ matrix with n copies of W_H on the diagonal, which is the transition matrix of one step of random walk within the clouds in $G \textcircled{z} H$. The steps between clouds are deterministic: the walk moves from a vertex (u, i) to a unique vertex (v, j) with $v \neq u$. The transition matrix for this deterministic step is thus a permutation matrix P with $P_{(u,i),(v,j)} = 1$ for each inter-cloud edge and zero otherwise. It follows from the definition of the zig-zag product that

$$Z = WPW.$$

So the random walk matrix of $G \textcircled{z} H$ has a very nice form, which should be the reason for the definition of zig-zag product in Definition 7.14.

The graph $G \textcircled{z} H$ is a regular graph, and so $\vec{1}_{nk}$ is an eigenvector of Z with eigenvalue 1. To prove the zig-zag product theorem, we will prove that for all $f \perp \vec{1}_{nk}$ the Rayleigh quotient

$$R_Z(f) = \frac{|f^T Z f|}{\|f\|_2^2} \leq \epsilon_1 + \epsilon_2 + \epsilon_2^2,$$

and this will imply that the spectral radius of Z is at most $\epsilon_1+\epsilon_2+\epsilon_2^2$ by the optimization formulation of the second eigenvalue in Lemma 2.11 and an analogous formulation for the last eigenvalue.

Vector Decomposition For any $f \perp \vec{1}_{nk}$, we decompose f to two vectors to apply the results in G and in H. This is where the power of linear algebra comes from, as in the larger domain $\mathbb{R}^{nk}$ there is a natural way to decompose the vector, while in the combinatorial setting it is not clear how to decompose a set of vertices in $G \textcircled{z} H$ into a set of vertices in G and a set of vertices in H to apply the expansion properties of G and of H as we discussed before.

Define f_G as the average of f on clouds, such that $f_G(u, i) = \frac{1}{k}\sum_{j=1}^{k} f(u, j)$ for all $(u, i) \in V(G \textcircled{z} H)$, so that two vertices in the same cloud have the same value in f_G. Define $f_H = f - f_G$. Note that f_H sums to zero in each cloud, such that $\sum_{j=1}^{k} f_H(u, j) = 0$ for each $u \in G$. Using triangle inequality,

$$|f^T Z f| = |f^T WPWf| = |(f_G+f_H)^T WPW(f_G+f_H)| \leq |f_G^T WPWf_G|+2|f_G^T WPWf_H|+|f_H^T WPWf_H|.$$

Since $W(H) \cdot \vec{1}_k = \vec{1}_k$ as H is a regular graph, it follows that $Wf_G = f_G$ as vertices in the same cloud have the same value in f_G. Therefore,

$$|f^T Z f| \leq |f_G^T P f_G| + 2|f_G^T PWf_H| + |f_H^T WPWf_H|.$$

We will use the spectral expansion of G to prove $|f_G^T P f_G| \leq \epsilon_1 \|f_G\|_2^2$ in Claim 7.21, the spectral expansion of H to prove $|f_H^T WPWf_H| \leq \epsilon_2^2 \|f_H\|_2^2$ in Claim 7.19, and a simple argument to bound $2|f_G^T PWf_H| \leq 2\epsilon_2 \|f_G\|_2 \|f_H\|_2$ in Claim 7.20. Assuming these claims, then

$$\begin{aligned} |f^T Z f| &\leq \epsilon_1 \|f_G\|_2^2 + 2\epsilon_2 \|f_G\|_2 \|f_H\|_2 + \epsilon_2^2 \|f_H\|_2^2 \\ &\leq \epsilon_1 \|f_G\|_2^2 + \epsilon_2 \big(\|f_G\|_2^2 + \|f_H\|_2^2\big) + \epsilon_2^2 \|f_H\|_2^2 \\ &\leq (\epsilon_1 + \epsilon_2 + \epsilon_2^2)\|f\|_2^2, \end{aligned}$$

where the last inequality holds because $f_G \perp f_H$ and so $\|f\|_2^2 = \|f_G\|_2^2 + \|f_H\|_2^2$ and also $\|f_G\|_2 \leq \|f\|_2$ and $\|f_H\|_2 \leq \|f\|_2$. This will complete the proof of Theorem 7.15 and so it remains to prove the three claims.

Spectral Expansion The following claim uses the spectral expansion of H and that f_H sums to zero in each cloud.

Claim 7.19 (Quadratic Term of H)**.** $|f_H^T W P W f_H| \leq \epsilon_2^2 \|f_H\|_2^2$

Proof. As the spectral radius of $W(H)$ is ϵ_2, we claim that $\|W(H) \cdot x\|_2 \leq \epsilon_2 \|x\|_2$ for any $x \perp \vec{1}_k$. To see this, let $x = \sum_{i=1}^k c_i v_i$ where $v_1, \ldots, v_n$ is an orthonormal basis of eigenvectors of $W(H)$ with eigenvalues $\alpha_1, \ldots, \alpha_k$. Note that $c_1 = 0$ as $v_1 = \vec{1}/\sqrt{k}$ and $x \perp \vec{1}$. Then

$$\|W(H) \cdot x\|_2^2 = \left\| W(H) \cdot \left(\sum_{i=2}^k c_i v_i \right) \right\|_2^2 = \left\| \sum_{i=2}^k c_i \alpha_i v_i \right\|_2^2 = \sum_{i=2}^k c_i^2 \alpha_i^2 \leq \epsilon_2^2 \sum_{i=2}^k c_i^2 \leq \epsilon_2^2 \|x\|_2^2,$$

where the first inequality is by the spectral radius of $W(H)$. This implies that $\|W f_H\|_2 \leq \epsilon_2 \|f_H\|$ as the sum of the entries in each cloud is zero in f_H as we argued earlier. Therefore,

$$|f_H^T W P W f_H| \leq \|W f_H\|_2 \cdot \|P W f_H\|_2 = \|W f_H\|_2^2 \leq \epsilon_2^2 \|f_H\|_2^2,$$

where the first inequality is by Cauchy-Schwarz and the equality is because P is a permutation matrix. □

The second claim is straightforward.

Claim 7.20 (Cross Term)**.** $|f_G^T P W f_H| \leq \epsilon_2 \|f_G\|_2 \|f_H\|_2$.

Proof. By Cauchy-Schwarz,

$$|f_G^T P W f_H| \leq \|f_G\|_2 \cdot \|P W f_H\|_2 = \|f_G\|_2 \cdot \|W f_H\|_2 \leq \epsilon_2 \|f_G\|_2 \|f_H\|_2,$$

where the last inequality was established in the proof of Claim 7.19. □

The final claim uses the spectral expansion of G and that $f \perp \vec{1}_{nk}$.

Claim 7.21 (Quadratic Term of G)**.** $|f_G^T P f_G| \leq \epsilon_1 \|f_G\|_2^2$.

Proof. The main point is to see that the LHS is equal to a corresponding quadratic form of the walk matrix of G. To see this, we "contract" each cloud to a single vertex. Define $g : V(G) \to \mathbb{R}$ as $g(v) = \sqrt{k} \cdot f_G(v, i)$. Note that $\|g\|_2^2 = \|f_G\|_2^2$. Note also that $f_G^T P f_G = g^T W(G) g$, where $W(G)$ is the random walk matrix of G, as each edge (u, i)-(v, j) in $G \textcircled{z} H$ contributes $f_G(u, i) \cdot f_G(v, j)$ to $f_G^T P f_G$ while the corresponding edge $uv \in G$ contributes $\left(\sqrt{k} f_G(u, i)\right)\left(\frac{1}{k}\right)\left(\sqrt{k} f_G(v, j)\right) = f_G(u, i) \cdot f_G(v, j)$ to $g^T W g$. Therefore,

$$\frac{f_G^T P f_G}{\|f_G\|_2^2} = \frac{g^T W g}{\|g\|^2}.$$

Since $f \perp \vec{1}$, it follows that $f_G \perp \vec{1}$ and thus $g \perp \vec{1}$. As G is an (n, k, ϵ_1)-graph, we conclude that

$$\frac{f_G^T P f_G}{\|f_G\|_2^2} = \frac{g^T W g}{\|g\|^2} \leq \epsilon_1.$$

□

This concludes the proof of Theorem 7.15. The idea of decomposing a vector into different components is useful in many proofs. We will use it again when we study high dimensional expanders in the third part of the course.

7.3 Applications of Expander Graphs

We discuss some of the many interesting applications of expander graphs in this section, with more details on expander codes as they are the basics of the recent breakthroughs in designing asymptotically good codes that are locally testable [DEL+21, PK21].

Probability Amplification

Suppose we have a randomized algorithm with error probability $1/100$ requiring n random bits. To decrease the failure probability, a standard way is to run the randomized algorithm independently k times, and then take the majority answer as the output. By a standard Chernoff bound argument, this decreases the failure probability to δ^k for some small constant δ. The number of random bits used is kn.

We show how to achieve exponentially small error probability while using only $n + ck$ bits where c is a constant. First, we see the above analysis in a slightly different perspective. Let V be the set of all n-bit strings. The randomized algorithm has error probability at most $1/100$ is equivalent in saying that among the 2^n n-bit strings, at most $2^n/100$ of them are "bad" strings. Denote this set of bad strings by $B \subseteq V$. The standard algorithm of taking the majority answer would fail if and only if we choose more than $k/2$ random strings from B, which is highly unlikely as $|B| \leq \frac{1}{100}|V|$. We can interpret the standard algorithm as doing a random walk of length k on the complete graph on V, and use the corresponding bit strings of the vertices $X_1, \ldots, X_k$ on this walk.

The idea is to replace a random walk on the complete graph on V by a random walk on a constant degree expander graph on V. Construct a d-regular expander graph G with 2^n vertices with spectral radius ϵd where d is a constant and $\epsilon \leq 1/100$. This can be done, say, by taking a large enough constant power of a Margulis expander. In the first step of the random walk, we use an n-bit random string, with error probability at most $1/100$. In the subsequent steps, instead of using n random bits to find the next n-bit string, we just choose a random neighbor of the current string in G and use the corresponding string in this random neighbor. Since G is a d-regular graph, we just need to use $\lceil \log_2 d \rceil$ random bits to choose a random neighbor in each subsequent step. Thus, the total number of bits used is $n + (k-1) \cdot \lceil \log_2 d \rceil$. Note that it is important that the neighbors of a Margulis expander can be computed quickly, so that we can find out the corresponding strings in this random walk quickly.

What is the error probability of this expander walk algorithm? This is exactly what Theorem 7.12 is formulated for, which shows that the error probability of taking the majority answer of a random walk of length k on a spectral expander with $\epsilon \leq 1/100$ is at most $(2/\sqrt{5})^k$.

This is just one example of using expander graphs in derandomization; see [HLW06, Vad12, AB06] for many more. The expander mixing lemma in Theorem 7.3 is very useful in derandomization.

Constructing Efficient Objects

We can think of a d-regular expander graph as a very efficient, as it only has a linear number of edges and it achieves very high connectivity. It should not be surprising that expander graphs are

useful in constructing efficient networks.

One interesting example is the construction of *superconcentrators*, which are directed graphs with n input nodes and n output nodes, satisfying the strong connectivity property that for any $k \leq n$ there are k vertex disjoint paths between any k input nodes and any k output nodes. For instance, the complete bipartite graph $K_{n,n}$ satisfies this property, but it has $\Theta(n^2)$ edges. Valiant conjectured that there is no superconcentrator with $O(n)$ edges, in an attempt to prove circuit lower bound. Later, he found a recursive construction of superconcentrator with $O(n)$ edges using expander graphs as building blocks. See [HLW06] for details.

Superconcentrators and expander graphs can be used to design efficient algorithms as well. One application is in designing fast algorithms for computing matrix rank [CKL13], where an expander graph or a superconcentrator is used to "compress" a rectangular matrix $A \in \mathbb{F}^{m \times n}$ with $n \gg m$ into a square matrix $B \in \mathbb{F}^{m \times m}$ in linear time such that $\operatorname{rank}(A) = \operatorname{rank}(B)$ with high probability.

A famous classical example of using expander graphs is to construct optimal sorting networks [AKS83], with $O(n \log n)$ edges and depth $O(\log n)$.

Undirected Connectivity in Log-Space

A striking application of the zig-zag product in Definition 7.14 is to solve the s-t connectivity problem in an undirected graph in logarithmic space. If we are allow to use randomized algorithms, then there is a very simple algorithm to solve the s-t connectivity problem in log-space, simply running a random walk for $O(n^3)$ steps would do, as it is well-known that the expected cover time for any undirected graph is at most $O(n^3)$. There is a deterministic algorithm by Savitch that solves the more general problem of s-t connectivity in *directed* graphs in $O(\log^2 n)$ space, by recursively guessing the midpoint of a directed s-t path. It has been a long standing and important open problem whether directed s-t connectivity can be solved in log-space. If such an algorithm exists, then this would imply that $NL = L$, the complexity classes of non-deterministic log-space problems and deterministic log-space problems are the same.

Reingold [Rei08] discovered a determinstic $O(\log n)$ space algorithm for s-t connectivity in undirected graphs using zig-zag products. Suppose the input graph G is a d-regular expander graph for a constant d. Then it can be shown that G has diameter $O(\log n)$. Then one can enumerate all paths of length $O(\log n)$ in $O(\log n)$ space, since each neighbor can be described in $\lceil \log_2 d \rceil$ space as we have seen in the probability amplification application above. Reingold's idea is to transform any graph G into a d-regular expander graph H such that s, t are connected in G if and only if s, t are connected in H. First, one can reduce G into a d-regular graph with constant d by replacing each vertex of high degree by a constant degree expander graph (and adding self-loops to each low degree vertex), similar to what was done in the replacement product in Definition 7.13. To improve the expansion, one can construct the graph $(G \textcircled{z} C)^8$, where C is a $(d, d^{1/16}, 1/2)$-graph. Using a variant of the zig-zag theorem in Theorem 7.15, it is possible to prove that the spectral gap doubles in the resulting graph. Then, one just needs to repeat this construction $O(\log n)$ times to get a graph H with constant spectral gap, as the initial spectral gap is at least $\Omega(1/n^2)$ for any connected undirected graph. Note that the size of H is at most a polynomial factor larger than the size of G, and s, t are connected in G if and only if s, t are connected in H.

A technical difficulty in carrying out this approach is to compute a neighbor of a vertex in H in log-space. The hope is that there are only $O(\log n)$ recursion levels for the zig-zag construction, and in each level we only need constant space, as there are only three steps and the degree is constant. Reingold proved that this can indeed be done; see [Rei08, Vad12] for details.

Hardness Amplification

Random walks on expander graphs can also be used for hardness amplifications, to take instances that are hard to approximate and construct instances that are even harder to approximate. See for example Chapter 22 of [AB06] for a simple application of expander random walks in proving hardness of approximating maximum independent sets.

Dinur [Din07] found an amazing proof of the very important PCP theorem using expander random walks. Her proof was inspired by Reingold's result, which involves many iterations of "powering" and "degree reduction", that makes the underlying constraint satisfaction problem harder and harder to approximate. See [AB06] for a good exposition of the PCP theorem. This is a great project topic especially for those who are interested in complexity theory.

Expander Codes

A main motivation for early developments in expander graphs is from coding theory.

A code $C \subseteq \{0,1\}^n$ of length n is a subset of n-bit strings, where each string in C is called a codeword. To design a good error correcting code, we would like to choose codewords that are far from each other so as to correct more errors, but at the same time choose as many codewords as possible so as to maximize the information rate. This can be thought of as a sphere packing problem, where the objective is to fit in as many disjoint spheres of a certain radius as possible in $\mathbb{F}_2^n$.

Definition 7.22 (Distance of Code). *Given $C \subseteq \{0,1\}^n$, the distance of C is defined as $\operatorname{dist}(C) := \min_{x \neq y \in C} d_H(x,y)$, where $d_H(x,y)$ is the Hamming distance between two codewords x and y. The relative distance of C is defined as $\operatorname{dist}(C)/n$.*

Definition 7.23 (Rate of Code). *Given $C \subseteq \{0,1\}^n$, the rate of C is defined as $\log|C|/n$, where $\log|C|$ can be thought of as the number of bits of information sent.*

Definition 7.24 (Asymptotically Good Code). *A family $C_n \in \{0,1\}^n$ of codes is asymptotically good if there are constants $r > 0$ and $\delta > 0$ such that for all n both the relative distance of C_n is at least δ and the rate of C_n is at least r.*

The existence of an asymptotically good code can be proved a standard probabilistic method. For the codes to be useful in practice, we would also like that encoding and decoding can be done in polynomial time in n (and ideally linear time in n), but this makes the problem much more challenging.

A common class of codes is the class of linear codes, where C is a linear subspace of $\mathbb{F}_2^n$. Linear codes have the advantage that they can be described by a basis and so encoding can be done in $O(n^2)$ time. Also, a simple but useful property of linear codes is that the minimum distance of the code is equal to the minimum ℓ_1-norm of a non-zero codeword, because $d_H(x,y) = \|x - y\|_1$ and $x - y$ is a codeword. The natural decoding strategy is to find the nearest codeword of a received word, but this is an NP-complete problem even for linear codes.

Low Density Parity Check Codes The idea of constructing codes from graphs was first suggested by Gallager, who uses sparse bipartite graphs to design low-density parity check codes (LDPC codes).

Let A be a parity check matrix for code C, such that $C = \{x \mid Ax = 0\}$ where $A \in \{0,1\}^{m \times n}$ with $m < n$. Each row i of A is a parity-check constraint, requiring $\sum_{j=1}^{n} A_{ij} \cdot x(j) = 0$ where the addition is mod 2. Note that the rate of this code is $1 - m/n$, so we want m/n to be bounded away from 1.

The matrix A can be viewed as a bipartite graph $G = (L, R; E)$ with $L = [n]$ and $R = [m]$ between the variables and the constraints, where there is a vertex in L for each variable and a vertex in R for each constraint, and variable i and constraint j has an edge if and only if $A_{ij} = 1$. We will see that good expansion of G yields good LDPC codes.

Definition 7.25 (Left Small-Set Vertex Expansion)**.** *Let $G = (L, R; E)$ be a bipartite graph with $|L| = n$ and $|R| = m$ and $m < n$. For any $0 < \delta < 1$, define the left δ-small-set vertex expansion of G as*

$$\psi_\delta^L(G) := \min_{S \subseteq L : |S| \leq \delta n} \frac{|\partial(S)|}{|S|},$$

where $\partial(S)$ is the vertex boundary in Definition 7.7.

Note that $\psi_\delta^L(G) \leq k$ for any k-left-regular bipartite graph G and any δ. Kahale proved that a Ramanujan graph satisfies $\psi_\delta^L(G) \approx \frac{1}{2}k$ for some constant $\delta > 0$ and this bound cannot be improved. In the following, we will need a stronger requirement that $\psi_\delta^L(G) \geq \frac{3}{4}k$, which is satisfied in a random k-left-regular bipartite graph with high probability. Capalbo, Reingold, Vadhan, Wigderson gave deterministic constructions of these "lossless expanders" satisfying $\psi_\delta^L(G) \geq 0.99k$ for some $\delta > 0$ and $m/n < 0.99$ using some variant of the zig-zag product.

First we see that the relative distance of a lossless expander code is a constant. The proof uses the unique neighbor property of a lossless expander.

Theorem 7.26 (Distance of Expander Code [SS96])**.** *Let $G = (L, R; E)$ be a left k-regular bipartite graph with $\psi_\delta^L(G) > \frac{1}{2}k$. Then the parity check code $C(G)$ defined by G has relative distance greater than δ.*

Proof. Let $S \subseteq L$ be a subset of left vertices with $|S| \leq \delta n$. Then $|\partial(S)| > \frac{k}{2}|S|$ by the left small-set vertex expansion assumption of G. A simple counting argument shows that that there exists a vertex $v \in \partial(S) \subseteq R$ with only one neighbor in S. Let us call such a vertex a unique neighbor of S.

To lower bound the minimum distance, recall that it is equivalent to lower bounding the ℓ_1-norm/support-size of a codeword $x \in \{0,1\}^n$. Let S be the support of x. If $|S| \leq \delta n$, by the previous paragraph, there exists a unique neighbor $v \in R$ of S. This implies that the parity constraint on v is not satisfied by x, and thus x is not a codeword of the parity check code defined by G. Therefore, any codeword of this parity check code must have support size greater than δn, and thus the minimum distance of this code is greater than δn. □

The key feature of the LPDC codes defined by expander graphs is that there is a surprisingly simple and efficient decoding algorithm.

Algorithm 4 Flip Algorithm for Expander Code

Require: A parity check matrix $A \in \{0,1\}^{m \times n}$ and a bit string $x \in \{0,1\}^n$.

1: Let $x^{(0)} := x$ and $t = 0$.
2: **while** there is an unsatisfied parity check constraint **do**
3: Find a bit i such that flipping it decreases the number of unsatisfied parity constraints. That is, an $i \in [n]$ such that $\left\|A\left(x^{(t)} + \chi_i\right)\right\|_1 < \left\|Ax^{(t)}\right\|_1$, where χ_i is the characteristic vector of i and the addition is under arithmetic mod 2. Set $x^{(t+1)} := x^{(t)} + \chi_i$ and $t \leftarrow t + 1$.
4: **end while**
5: **return** $x^{(t)}$.

The analysis of the flip algorithm uses a stronger assumption about the left small-set vertex expansion than than in Theorem 7.26.

Theorem 7.27 (Efficient Decoding of Expander Code [SS96])**.** *Let $G = (L, R; E)$ be a left k-regular bipartite graph with $L = [n]$ and $R = [m]$ and $\psi_\delta^L(G) > \frac{3}{4}k$. Let x be an n-bit string whose distance from a codeword y is at most $\frac{1}{2}\delta n$. Then Algorithm 4 will return y in at most m iterations.*

Proof. Let $\Delta^{(t)} := \{i \in [n] \mid x^{(t)}(i) \neq y(i)\}$ be the set of errors at the t-th iteration. The plan is to argue that as long as $\mathrm{dist}_H\left(x^{(t)}, y\right) = \left|\Delta^{(t)}\right| \leq \delta n$, there exists a bit i such that flipping it decreases the number of unsatisfied constraints, and also argue that $\mathrm{dist}_H\left(x^{(t)}, y\right) \leq \delta n$ for all t if $\mathrm{dist}_H\left(x^{(0)}, y\right) \leq \frac{1}{2}\delta n$. These would imply that after at most $\tau \leq m$ iterations, there will be no unsatisfied constraints and so $x^{(\tau)}$ is a codeword, and thus $x^{(\tau)}$ must be equal to y as $\mathrm{dist}_H(x^{(\tau)}, y) \leq \delta n$ while the distance between y and other codewords is strictly bigger than δn.

For ease of notation, let $\Delta := \Delta^{(t)}$ be the set of error at some iteration t. Assume that $0 < |\Delta| \leq \delta n$, we would like to argue that there is a bit i that flipping it decreases the number of unsatisfied constraints. Partition $\partial(\Delta)$ into the set of satisfied neighbors $\partial_+(\Delta)$ of Δ and the set of unsatisfied neighbors $\partial_-(\Delta)$ of Δ. On one hand, since $|\Delta| \leq \delta n$, by the left small-set vertex expansion of Δ,

$$|\partial_+(\Delta)| + |\partial_-(\Delta)| = |\partial(\Delta)| > \frac{3}{4}k|\Delta|.$$

On the other hand, when we consider the $k|\Delta|$ number of edges between Δ and $\partial(\Delta)$, observe that each vertex in $\partial_+(\Delta)$ has at least two such edges while each vertex in $\partial_-(\Delta)$ has at least one such edge, and so

$$2|\partial_+(\Delta)| + |\partial_-(\Delta)| \leq k|\Delta|.$$

Combining these two inequalities gives that $|\partial_-(\Delta)| > \frac{1}{2}k|\Delta|$. This implies that there must exist a vertex $i \in \Delta$ with strictly more unsatisfied neighbors than satisfied neighbors. Therefore, as long as $|\Delta| \leq \delta n$, there must exist a bit i such that flipping it decreases the number of unsatisfied constraints.

To complete the proof, we argue that $|\Delta| \leq \delta n$ in any iteration. Suppose this is not true, then since $|\Delta|$ changes by one in each iteration, there is an (earliest) iteration such that $|\Delta| = \delta n$. Then, by the argument in the previous paragraph, there are strictly more than $\frac{1}{2}k|\Delta| = \frac{1}{2}k\delta n$ unsatisfied constraints in that iteration. However, since $|\Delta^{(0)}| \leq \frac{1}{2}\delta n$, the number of unsatisfied constraints in the beginning is at most $\frac{1}{2}k\delta n$. This contradicts with the previous paragraph that the number of unsatisfied constraints is decreasing when $|\Delta| \leq \delta n$. □

Spielman showed that it is possible to use expander codes to obtain asymptotically good codes that are linear time encodable and decodable!

Tanner Codes Tanner code is a generalization of LDPC code in which the "base code" can be more general than just checking parity. Let $C_0 \subseteq \{0,1\}^k$ be the base code. Let $G = (V, E)$ be a k-regular graph with $V = [n]$ and $E = [m]$. The Tanner code is defined as $C(G) := \{y \in \{0,1\}^m \mid y_{|\delta(i)} \in C_0 \ \forall i \in [n]\}$, where $y_{|\delta(i)}$ is the vector y restricted on the k edges in $\delta(i)$ for a vertex $i \in V$. That is, each bit $y(j)$ of a codeword is on an edge $j \in E$ of G, and a binary string y is a codeword if $y_{|\delta(i)}$ is a codeword of the base code C_0 for every vertex $i \in V$ of G.

The advantage of using Tanner code is that we could use a stronger base code with larger minimum distance, rather than just the parity check code with minimum distance only two. With a base code C_0 of minimum distance d_0, the requirement on the vertex expansion of G can be relaxed to k/d_0 to achieve the same distance as that of the corresponding LDPC code. In particular, because of Tanner's theorem in Theorem 7.9, one can simply use a spectral expander as G to design asymptotically good codes that are linear time encodable and decodable, without using lossless expanders. The decoding algorithm is still an iterative "fixing" algorithm where we replace an invalid codeword on a vertex by its nearest codeword. The analysis has a similar flavor that if the decoding algorithm fails, then one argues that there must be a "denser" subgraph than what is allowed by the expander mixing lemma.

The recent breakthroughs [DEL+21, PK21] in designing asymptotically good codes that are also locally testable is a generalization of Tanner codes on 2-dimensional expanders (where graphs are 1-dimensional expanders). Hope we will have some time to discuss it in the third part of the course when we study high-dimensional expanders.

7.4 References

[AB06] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2006. 65, 67

[AKS83] Miklós Ajtai, János Komlós, and Endre Szemerédi. An o(n log n) sorting network. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25-27 April, 1983, Boston, Massachusetts, USA*, pages 1–9. ACM, 1983. 66

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Comb.*, 26(5):495–519, 2006. 57

[CKL13] Ho Yee Cheung, Tsz Chiu Kwok, and Lap Chi Lau. Fast matrix rank algorithms and applications. *J. ACM*, 60(5):31:1–31:25, 2013. 66

[CPT21] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: a fourier-analytic approach. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1643–1655. ACM, 2021. 60

[DEL+21] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. *Electron. Colloquium Comput. Complex.*, page 151, 2021. 65, 70

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007. 67

[Gil98] David Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998. 60

[GLSS18] Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava. A matrix expander Chernoff bound. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1102–1114. ACM, 2018. 60

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 24, 27, 55, 58, 59, 60, 61, 65, 66

[PK21] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. *CoRR*, abs/2111.03654, 2021. 65, 70

[Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):17:1–17:24, 2008. 66

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1):pp. 157–187, 2002. 61

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inf. Theory*, 42(6):1710–1722, 1996. 68, 69

[Vad12] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. 59, 60, 65, 66