Expander Graphs and their Applications*
Lecture 8: The Zig Zag Product

Notes taken by Eyal Bigman

May 2002

Summary: In this lecture we shall introduce a new kind of graph product called the *Zig Zag Product*. We shall analyze the expansion properties of the zig zag product of expanding graphs and use them to create an explicit recursive construction of a family of good expanders. Furthermore we will connect the entropy of a random walk on the graph and its expansion.

1 Introduction

We begin with some standard definitions: Let $G =< V, E >$ be a d-regular graph ($|V| = n$). The adjacency matrix $A(G) = (a_{ij})_{i,j=1}^n$ of the graph G is a symmetric $n \times n$ non-negative integer matrix such that $a_{ij} = k$ iff there are k edges between vertices i and j. It follows from regularity that the sum of every row is d. The matrix $\hat{G} = \frac{1}{d}A$ is the *normalized adjacency matrix*, it is the transition matrix of a random walk on G. Thus if $p \in \Re^n$ is a probability distribution on the vertices at time t then $\hat{G}p$ is the distribution at time $t+1$.

G is a (n, d, α)-expander if α is an upper bound on the second eigenvalue of the normalized adjacency matrix $\lambda_2 \leq \alpha$ ($\lambda_1 = 1$). It follows from the spectral gap theorem that $(1-\alpha)d/2 \leq h(G)$ where $h(G)$ is the expansion of G. Thus G is a better expander when α is close to 0.

The *square* $G^2 = (V, E')$ is a graph on the same vertices and $(u, w) \in E'$ iff there is a path of length 2 in G from u to w. If A is the adjacency matrix of G then A^2 is the adjacency matrix of G^2. It is easy to see that G^2 is a (n, d^2, α^2)-expander.

The zig zag product will be an unsymmetric binary function, that given an m-regular graph of size n and a d-regular graph of size m, it yields a d^2-regular graph of size mn. After we define the zig zag product we will prove:

Theorem 1 (The Zig Zag Theorem). *Let G be a (n, m, α)-expander and H be a (m, d, β)-expander then $G \text{ⓩ} H$ is a $(nm, d^2, f(\alpha, \beta))$-expander where $f(\alpha, \beta) = \alpha + \beta + \beta^2$.*

We will present a proof that uses only elementary linear algebra. With some more algebra this bound is improved in [1] to $f(\alpha + \beta) = \alpha + \beta$, and if $\alpha, \beta < 1$, it can be shown $f(\alpha + \beta) < 1$.

Before we proceed with the definition let us show how it can be used for an explicit construction of a family of expanders with constant degree. For d constant let H be a $(d^4, d, \frac{1}{4})$-expander, there is a probabilistic proof that such an expander exists and since d is constant we can find such a graph by an exhaustive search in constant time, there are also efficient constructions of such graphs - see [1]. Define recursively:

$$G_1 = H^2$$
$$G_{n+1} = (G_n)^2 \text{ⓩ} H$$

Proposition 2. *G_n is a $(d^{4n}, d^2, \frac{1}{2})$-expander for all n*

Proof. By induction. The case $n = 1$ follows from the definition. If we assume the proposition for n then G_n^2 is a $(d^{4n}, d^4, \frac{1}{4})$-expander and from theorem 1 it follows that G_{n+1} is a $(d^{4(n+1)}, d^2, \frac{1}{2})$ graph $\square$

* Lecture Notes for a course given by Nati Linial and Avi Wigderson at the Hebrew University, Israel.

2 The Construction

Let G be a (n, m, α)-expander and H be a (m, d, β)-expander. For every vertex $v \in V(G)$ let $e_v^1, \ldots e_v^m$ be the edges connected to the vertex. Also, we regard the vertices of H as the integers $1, \ldots, m$, denoted by $[m]$. To obtain $G ⓩ H$, we replace every vertex v with a cloud of m vertices $(v, 1), \ldots (v, m)$ one for every edge connected. The vertices within a cloud are connected by "miniedges" so that every cloud forms a mini copy of H. The edges of G are augmented on both sides by these miniedges to form the edges of $G ⓩ H$. More formally:

Definition 3. $G ⓩ H = < V(G) \times [m], E' >$, where $((v, i), (u, j)) \in E'$ iff there are some $k, l \in [m]$ such that $(i, k), (l, j) \in V(E)$ and $e_v^k = e_u^l$.

It is essential of course for the sake of well definedness that the degree of G equals the size of H.

Proof of the Zig Zag theorem. It is easy to see that $G ⓩ H$ is a graph of size mn and degree d^2. The expansion constant of $G ⓩ H$ is a bound on the second eigenvalue of the transition matrix of the random walk on the graph. Each step of the random walk on an edge of $G ⓩ H$ can be regarded as a random step on a miniedge within a cloud, a deterministic step on an edge connecting two clouds and another random step within a cloud. Thus, the transition from (v, i) to (u, j) consists of a random step from (v, i) to (v, k) for some $k \in [m]$, a deterministic step from (v, k) to (u, l) and another random step from (u, l) to (u, j).

The transition matrix of the random walk on $G ⓩ H$ will therefore be the product of the transition matrices of these three steps, i.e. $\widehat{G ⓩ H} = \tilde{H}\tilde{G}\tilde{H}$, where $\tilde{H} = \hat{H} \otimes I_n$ is the transition matrix of a random step in each cloud, and

$$\tilde{G}_{(v,k),(u,l)} = \begin{cases} 1 & \text{if } e_v^k = e_u^l \\ 0 & \text{otherwise} \end{cases}$$

is a matrix of transpositions. It is easy to see that $\| \tilde{H} \|, \| \tilde{G} \| \leq 1$.

$G ⓩ H$ is a regular graph thus the constant vector 1_{mn} is an eigenvector, it follows from Rayleigh's theorem that $\lambda_2 = max_{f \perp 1_{mn}} \frac{|f \widehat{G ⓩ H} f|}{\|f\|^2}$. Therefore it suffices to show $\frac{|f \widehat{G ⓩ H} f|}{\|f\|^2} \leq \alpha + \beta + \beta^2$ for every $f \perp 1_{mn}$.

We can write any vector f as $f = f^{\|} + \tilde{H} f^{\perp}$, where $f^{\|}$ is a vector that is constant within each cloud, and $\tilde{H} f^{\perp}$ sums up to zero within each cloud. Since 1_m is an eigenvector of H with eigenvalue one, we get that $\tilde{H} f^{\|} = f^{\|}$.

Therefore, for every $f \perp 1_{mn}$:

$$\begin{aligned} |f \widehat{G ⓩ H} f| &= |f \tilde{H}\tilde{G}\tilde{H} f| \\ &= |f^{\|} \tilde{H}\tilde{G}\tilde{H} f^{\|}| + 2|f^{\|} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| + |f^{\perp} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| \\ &= |f^{\|} \tilde{G} f^{\|}| + 2|f^{\|} \tilde{G}\tilde{H} f^{\perp}| + |f^{\perp} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| \end{aligned}$$

Now, since $f^{\|} \perp 1_{mn}$, we know that $\| \tilde{G} f^{\|} \| \leq \alpha \| f^{\|} \|$. Also, since $f^{\perp}$ sums up to zero in each cloud, we have $\| \tilde{H} f^{\perp} \| \leq \beta \| f^{\perp} \|$. Therefore:

$$|f^{\|} \tilde{G} f^{\|}| \leq \alpha \| f^{\|} \|^2$$

$$|f^{\|} \tilde{G}\tilde{H} f^{\perp}| \leq \| \tilde{G} \| \cdot \| f^{\|} \| \cdot \| \tilde{H} f^{\perp} \| \leq \beta \| f^{\|} \| \cdot \| f^{\perp} \|$$

$$|f^{\perp} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| \leq \| \tilde{G} \| \cdot \| \tilde{H} f^{\perp} \|^2 \leq \beta^2 \| f^{\perp} \|^2$$

The inequalities follow from Cauchy Schwartz and the definition of the operator norm.

$$\begin{aligned} |f \widehat{G ⓩ H} f >| &\leq \alpha \| f^{\|} \|^2 + 2\beta \| f^{\|} \| \cdot \| f^{\perp} \| + \beta^2 \| f^{\perp} \|^2 \\ &\leq \alpha \| f \|^2 + 2\beta \| f^{\|} \| \cdot \| f^{\perp} \| + \beta^2 \| f \|^2 \\ &= \alpha \| f \|^2 + \beta \cdot (\| f^{\|} \|^2 + \| f^{\perp} \|^2 - (\| f^{\|} \| - \| f^{\perp} \|)^2) + \beta^2 \| f \|^2 \\ &= (\alpha + \beta + \beta^2) \cdot \| f \|^2 - \beta \cdot (\| f^{\|} \| - \| f^{\perp} \|)^2 \\ &\leq (\alpha + \beta + \beta^2) \cdot \| f \|^2 \end{aligned}$$

□

3 Entropy Analysis

There are several different definitions of entropy for distribution p, the classical definitions $H(p) = -\sum_{i=1}^{n} p_i \log(p_i)$, the H_2 entropy $H_2(p) = -\log(\| p \|_2)$ and the min-entropy $H_\infty(p) = -\log(\| p \|_\infty)$. For any transition matrix A with expansion constant α and distribution $p = \frac{1}{n}1_n + f$

$$\| Ap \|^2 = \| \frac{1}{n}1_n + Af \|^2 \leq \| \frac{1}{n}1_n \|^2 + \| Af \|^2 \leq ((1-\lambda^2) + \alpha^2\lambda^2) \| p \|^2$$

where $\lambda = \frac{\|f\|}{\|p\|} \leq 1$. Hence

$$H_2(Ap) = H_2(p) + \log((1-\lambda^2) + \alpha^2\lambda^2) = H_2(p) + \log(1-(1-\alpha^2)\lambda^2) = H_2(p) - \Delta_E$$

$0 < 1-\alpha^2 \leq 1$ implies $0 \leq 1-(1-\alpha^2)\lambda^2 < 1$ hence $\Delta_E = -\log((1-\lambda^2)+\alpha^2\lambda^2) > 0$ as long as $\lambda > 0$. This shows that the entropy increases by at least Δ_E as long as there is a positive nonuniform component. It follows that for better expanders (α smaller) the H_2 entropy grows faster. It can be shown that the H_2 and the H_∞ are correlated, therefore the same increase is true for the min-entropy.

We will show next that the classical entropy also grows when $\lambda > 0$ but we shall make very different considerations. It is currently unknown how fast the entropy grows and how the growth rate relates to the expansion constant. We note that for all the definitions of entropy, the increase of entropy for nonuniform distributions is essentially the second law of thermodynamics.

For random variables X and Y with some joint distribution we have the standard entropy equation $H(X,Y) = H(X)+H(Y|X)$ thus the entropy of the joint distribution of X and Y is the sum of the entropy of X and the conditioned entropy of $Y|X$. In the case of the zig zag graph the set of vertices is $V \times [m]$, the random variables we will analyze will be the projections X and Y to the first and second coordinates. We will see that a random step on the zig zag graph will increase both the entropy of X and the conditioned entropy $H(Y|X)$ and thus increase the entropy of the joint distribution.

As we mentioned before a random step consists of a random step within a cloud (zig) a deterministic step between clouds and another random step within a cloud (zag). Since the zig and the zag steps are within a cloud, they only affect the second coordinate. They are random steps therefore they increase the conditioned entropy of $Y|X$ as long as that is less than maximal. For a distribution which is uniform on every cloud the zig and zag steps have no effect.

We can think of X and Y as basins of entropy, taking either a zig or a zag step is like pouring entropy from an external source into the basin. How much entropy can be poured in? Well that depends on the capacity of the random variable (maximum entropy) and the amount of entropy present in the basin, i.e. how far is the distribution on every cloud from uniform.

What about the deterministic step between the clouds? Well this step is deterministic and therefore it does not change the total amount of entropy in the system, but that does not mean it cannot change the division of the entropy between X and Y.

Let us think of the extreme case $p = e_v \otimes \frac{1}{m}1_m$ where the entropy of X is zero and the entropy of Y is maximal, i.e. the distribution is concentrated uniformly on the vertices of the cloud v (obviously p remains unchanged by the zig step). In this case $H(Y|X) = \log(m)$ and $H(X,Y) = H(X) + H(Y|X) = 0 + \log(m)$. After the deterministic step there is equal probability to be in any one of v's neighboring clouds, but within these clouds the distribution is concentrated on the vertex that is actually connected to a vertex in v. The entropy in these clouds is 0. The entropy of clouds not connected to v remains zero, and the entropy in v is zero. Thus we see that the entropy of X goes from zero to $\log(m)$ and the entropy of Y plummets to zero. We still have the same entropy in the system $H(X,Y) = H(X)+H(Y|X) = \log(m)+0$, but the division between X and Y changes, all the entropy of Y is passed to X. In the general case the same thing happens, entropy is exchanged between X and Y. Since deterministic steps are reversible, not necessarily is the entropy transfered from Y to X, or from a variable with high entropy to a variable with low entropy. Nevertheless, it can be shown that if the entropy of $Y|X$ is maximal and the joint distribution is not uniform then the deterministic step reduces the entropy of $Y|X$. Thus we see that the deterministic step does not change the entropy in the systems but pours entropy from one vessel to another, and in the case that Y is full and X is not, some entropy is poured from Y to X.

We can see now the effect of a random step, first entropy is poured into Y, if it is not full then some entropy is added to the joint distribution. Next entropy is exchanged between X and Y, necessarily making room for more entropy in

Y if it is full and X is not. Finally yet more entropy is poured into Y and again if it is not full already then the entropy of the joint distribution is increased.

Thus we see that the total amount of entropy in the system necessarily increases unless the two basins are full already. Why do we have to pour in entropy from an external source twice? Because in cases like p above, whatever entropy was poured while Y was full will be wasted and the entropy of the whole system will remain the same. Only the second step will make room in Y for more entropy, thus if we avoid the zag step the entropy in the system will not increase. On the other hand, since the deterministic step is reversible it could also be the case that Y will be filled from X and thus the entropy poured in the zag step will be wasted. It can be shown (or follows immediatly form reversability) that in such a case Y is not full from the start. Thus if we don't pour entropy on Y at the beginig it may not be possible later. Therefore both the zig and the zag steps are essential in order to assure that the entropy will actually increase.

References

[1] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1):157–187, 2002.