

The zig-zag product

Neil Olver

1 Motivation

The expander constructions based on algebraic methods can give expanders that are both explicit (i.e. we can quickly construct the graph, or even obtain neighborhood information without constructing the entire graph, and *Ramanujan*, meaning that the spectral gap is essentially as large as possible. It also follows from this spectral bound that the edge expansion of Ramanujan graphs is essentially as large as possible.

So why do we need some other expander construction? One reason is the consideration of vertex expansion. Define

$$\Psi_V(G, k) := \min_{S \subset V: |S| \leq k} \frac{|\Gamma(S) \setminus S|}{|S|}.$$

This allows us to ask more fine-tuned questions about the vertex expansion, since we can consider the expansion for sets of different sizes. It can be shown that a random d -regular graph has very good vertex expansion on sufficiently small sets: in particular, for any $\delta > 0$, there is an $\epsilon > 0$ s.t.

$$\Psi_V(G, \epsilon n) \geq d - 2 - \delta. \tag{1}$$

However, the spectral gap is insufficient to get this. Kahale [3] gave a construction of Ramanujan graphs s.t. $\Psi_V(G, \epsilon n) \approx d/2$, no matter how small $\epsilon > 0$; see pg. 477 in [2]. So spectral properties alone are no longer enough to give us what we need.

Expanders satisfying Eq. (1) are known as “lossless expanders”. Such expanders, and variants of them, have a number of uses in computer science; constructions of good linear codes, data structures, the design of routing networks, and in computational complexity. See pg. 515 of [2] for a long list of references. Note that these applications typically require explicit constructions, and so the fact that random graphs are lossless expanders with high probability is insufficient.

There has also been recent work [1] which gives an expander construction based on similar methods to the zig-zag product, and for which they are able to give a completely combinatorial proof of the expansion properties, without recourse to the spectral gap.

2 The construction

Recall the following notation:

Definition 1. An (n, d) -graph is a d -regular graph on n vertices.

Definition 2. An (n, d, α) -graph G is an (n, d) -graph with $\lambda_2(G)/d \leq \alpha$.

The scaling down by d will be particularly convenient, because we will mostly be working with the transition matrix \hat{A} of the random walk on G ; if the adjacency matrix of G is A , this is just $\hat{A} = A/d$.

To construct the zig-zag product, we first define the *replacement product*. Throughout, G will represent an (n, m) -graph, and H an (m, d) -graph.

Roughly speaking, $G \boxtimes H$ replaces each vertex of G with a copy of H (and keeps all the edges of H in all the copies), and distributes the edges of G so that an edge $e = (u, v)$ in G becomes an edge between some vertex in the “cloud” from u and some vertex in the cloud from v . In addition, these G edges are distributed

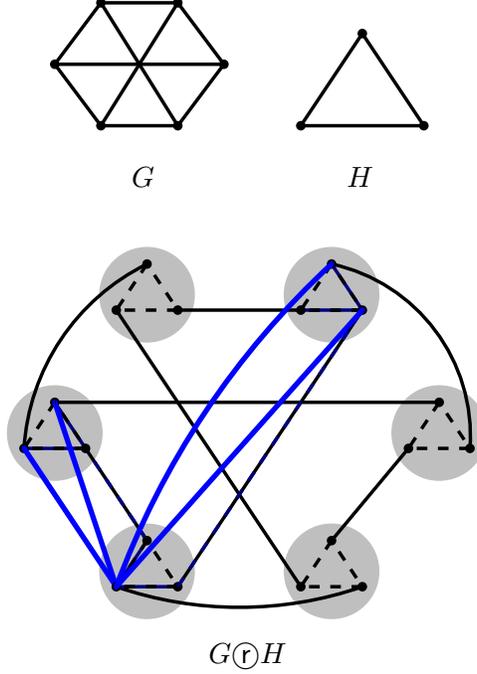


Figure 1: The replacement product of two graphs, and (in colour) the edges in the zigzag product from one particular vertex.

so that they form a *matching* in $G \oplus H$, i.e. every vertex in $G \oplus H$ is adjacent to one such edge. All of this ensures that $G \oplus H$ is $d + 1$ -regular.

Definition 3. Let G be an (n, m) -graph and H an (m, d) -graph, with the vertices of H labelled from 1 to m . In addition, let each $v \in G$ be given an associated numbering of its neighbours, v^1, v^2, \dots, v^m . Then the replacement product $G \oplus H$ has vertex set $V(G) \times V(H)$, and edge set

$$\{(v, i), (v, j) : (i, j) \in E(H)\} \cup \{(v, i), (w, j) : v, w \in V(G), i, j \in V(H), w = v^i \text{ and } v = w^j\}.$$

This is only well-defined when the neighbour ordering is provided in addition to G and H ; however, everything we do will be true for any ordering, and so we do not specify it explicitly. Figure 1 gives an example of the zigzag product of two graphs (ignore the coloured edges for now).

We refer to the set $\{(v, i) : i \in H\}$ as the *cloud representing* v , for any $v \in G$. I'll also call edges between clouds “ G -edges”, although this is not standard.

Now we are ready to define the zig-zag product. Basically, each edge in $G \otimes H$ corresponds with a length-3 path in $G \oplus H$, where the first step is an edge within some cloud, the second step is a G -edge, and the third edge is another cloud step. Figure 1 shows two of the edges of $G \otimes H$, shown in blue. Since from any vertex there d choices for the first step, no choice in the second step, and d choices on the third step, this gives a d^2 -regular graph.

Definition 4. Let G be an (n, m) -graph and H an (m, d) -graph, with $V(H) = [m]$, where $[m]$ denotes the set $\{1, 2, \dots, m\}$. Let E_r be the edge set of $G \oplus H$. Then $G \otimes H$ has vertex set $V(G) \times [m]$, and edge set

$$E = \{(v, i), (w, l) : \exists j, k \in [m] \text{ where } ((v, i), (v, j)), ((v, j), (w, k)) \text{ and } ((w, k), (w, l)) \text{ are all in } E_r\}.$$

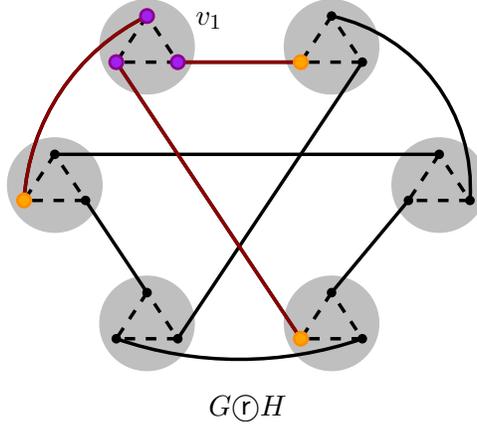


Figure 2: The effect of the first two steps within a single move in $G \otimes H$, starting from a uniform distribution over the cloud v_1 .

3 Some intuition

The following section will be very hand-wavy: the primary goal is to give some intuition as to why this particular construction, and also some intuition behind the proof that we'll see in the next section. The intuition here can be formalized a bit (*entropy* is the key concept) and tied more closely to the proof, but I'm aiming here to keep the probability requisites to a minimum.

Let \hat{A} be the transition matrix for the random walk on G . Expansion is tied up with rapid mixing of the random walk:

Lemma 5. For an (n, d, α) -graph G with transition matrix \hat{A} , and any starting distribution \mathbf{p} ,

$$\|\hat{A}^t \mathbf{p} - \pi\|_2 \leq \|\mathbf{p} - \pi\|_2 \alpha^t \leq \alpha^t \quad \text{for all } t \in \mathbb{N}^+,$$

where π is the uniform distribution.

Proof. We only need to show this for $t = 1$. But $\mathbf{p} - \pi$ is orthogonal to π , which is proportional to $\mathbf{1}$. Hence $\|A(\mathbf{p} - \pi)\|_2 \leq \alpha \|\mathbf{p} - \pi\|_2$. \square

So expanders are sparse graphs that mix quickly; at each step, the distribution spreads out quite a bit (gets closer to uniform). We would like to see why $G \otimes H$ should have this property.

So consider some distribution $\mathbf{p} := \mathbf{p}(t)$ on $G \otimes H$. For any fixed cloud v , if we look at \mathbf{p} restricted to this cloud, and scale up accordingly, we obtain a distribution $\mathbf{p}^{(v)}$, indexed by H . We also have a distribution $\bar{\mathbf{p}}$, indexed by G , which describes how the probability is distributed among the clouds: $\bar{p}(v) := \sum_i p(v, i)$.

This distribution $\mathbf{p}^{(v)}$ may be very concentrated over the cloud v , or it may be quite spread out. If the $\mathbf{p}^{(v)}$'s are mostly quite spread out, and \mathbf{p} is quite spread out between the clouds (the total weight $\sum_{i \in H} p(v, i)$ in each cloud is about the same), then \mathbf{p} is overall quite spread out. So if \mathbf{p} is not close to uniform, either (i) lots of the $\mathbf{p}^{(v)}$'s are far from uniform¹, or (ii) the weight is unevenly distributed between the clouds.

Consider what happens when we take a random step in $G \otimes H$. We think in terms of the construction; this is equivalent to taking 3 steps in $G \uplus H$, where the first and third steps are along cloud edges, and the middle step is along a G -edge. If (i) occurs, then the very first step will cause things to spread out a lot. If (i) does not occur however, the first random step doesn't really do anything; we made a random move, but we

¹More accurately, if we measure how spread out each $\mathbf{p}^{(v)}$ is, and take an average over all the clouds weighted by \bar{p} , this is small; to do this properly, one would consider the *conditional entropy*.

were already uniform from the perspective of H , so this extra randomness is “lost”. To make things clearer, imagine that \mathbf{p} is exactly uniform over cloud v_1 (v_1 some fixed node in $V(G)$, and zero outside this cloud). See Figure 2—suppose all the purple nodes have probability $1/3$, and the rest zero. So after the first step, the distribution is completely unchanged (still $1/3$ each for all nodes in cloud v_1). But what happens in the second step? Since the position in cloud v_1 is uniformly random, the G -edge we pick is exactly a uniformly random one from the set of G -edges adjacent to the cloud (the red edges). Now there is some subtlety here; this second step was completely deterministic, and so cannot increase the amount of randomness in the system (formally, it cannot increase the entropy). What in fact it does is shift the randomness into a different place; there is now lots of uncertainty in terms of which cloud we moved to, *but* the marginal distributions within these clouds are now very far away from uniform. In the picture, each orange node has probability $1/3$ after the first two steps; for each cloud w adjacent to v_1 , the distribution $\mathbf{p}^{(w)}(t+1)$ has all its weight on a single node. The deterministic second step has sucked out randomness from the marginal distributions $\mathbf{p}^{(v)}$, and put it into the \bar{p} . This means that when we take the final, random, step within a cloud, the situation is far from uniform, and this step does indeed add a lot of extra randomness into the system.

4 Analyzing the expansion

We prove the following:

Theorem 6. (*Zig-zag theorem, weakened version; Reingold-Vadhan-Wigderson 2002*). *If G is an (n, m, α) -graph and H is an (m, d, β) -graph, then $G \otimes H$ is an $(nm, d^2, \varphi(\alpha, \beta))$ graph, with $\varphi(\alpha, \beta) = \alpha + \beta + \beta^2$.*

In fact, the above is true with

$$\varphi(\alpha, \beta) = \min(\alpha + \beta, 1 - (1 - \beta^2)(1 - \alpha)/2).$$

(So in particular, if $\alpha, \beta < 1$, then $\varphi(\alpha, \beta) < 1$.) We will only prove the weaker statement however (and it will be sufficient for the construction of a family of expanders that we will do in the next section).

Proof. Motivated by our entropy intuition, we will think in terms of random walks on $G \otimes H$. In particular, we will think of a random step in $G \otimes H$ as 3 consecutive steps in $G \oplus H$, which motivates the following factorization.

Let \hat{A}, \hat{B} be the transition matrices for G and H respectively, and let Z be the transition matrix for $G \otimes H$. We can factor $Z = \tilde{B}P\tilde{B}$, where P is a permutation matrix defined by a set of swaps (one for each G -edge), and $\tilde{B} = \hat{B} \otimes I_n$. This is simply a restatement of the fact that a random walk in $G \otimes H$ corresponds to a 3-step walk in $G \oplus H$, the first two being random steps in the clouds (giving \tilde{B}), and the second a completely deterministic walk along a G -edge, which defines P .

The Rayleigh formula tells us that

$$\lambda_2(Z) = \sup_{f \neq 0, f \perp \mathbf{1}_{mn}} \frac{|f^T Z f|}{\|f\|^2}.$$

So our goal is to show that

$$|f^T Z f| \leq \varphi(\alpha, \beta) \|f\|^2. \tag{2}$$

Define f^{\parallel} by

$$f^{\parallel}(v, i) = \frac{1}{m} \sum_{j \in [m]} f(x, j),$$

and let $f^{\perp} = f - f^{\parallel}$. (Comparing with the intuition: f^{\parallel} is constant over each cloud, and is just telling us how much weight each cloud has.) Expand out the LHS of (2):

$$\begin{aligned} |f^T Z f| &= |f^T \tilde{B}P\tilde{B} f| \\ &\leq |f^{\parallel} \tilde{B}P\tilde{B} f^{\parallel}| + 2|f^{\parallel} \tilde{B}P\tilde{B} f^{\perp}| + |f^{\perp} \tilde{B}P\tilde{B} f^{\perp}| \end{aligned}$$

Now we know that $\tilde{B}f^\parallel = f^\parallel$, because $\tilde{B} = \hat{B} \otimes I_n$ (\tilde{B} makes a step within clouds, but f^\parallel is uniform over every cloud). We also have that

$$\|\tilde{B}f^\perp\| = \left\| \sum_{v \in G} \hat{B}f^\perp|_v \right\| \leq \sum_{v \in G} \beta \|f^\perp|_v\| = \beta \|f^\perp\|.$$

So

$$|f^{\perp T} \tilde{B}P\tilde{B}f^\perp| \leq \beta^2 |f^{\perp T} P f^\perp| \leq \|f^\perp\|^2.$$

This last step follows by the Rayleigh quotient again using that eigenvalues of P all have magnitude 1.

Let $g(v) := \sqrt{m}f^\parallel(v, i)$ (the scaling is so that $\|f^\parallel\| = \|g\|$). Then $f^{\parallel T} P f^\parallel = g^T \hat{A}g$ (again going back to the intuition, this just says that if the distribution is uniform over a cloud, the middle step is uniform over the edges adjacent to that cloud). Since $f^\parallel \perp \mathbf{1}_{mn}$, $g \perp \mathbf{1}_{mn}$ and so $g^T \hat{A}g \leq \alpha \|g\|^2$. Hence

$$|f^\parallel P f^\parallel| \leq \alpha \|f^\parallel\|^2.$$

The following is a standard result, but the short proof is included:

Claim 7. *Let M be any $n \times n$ **stochastic matrix**, i.e. a square matrix having all entries nonnegative, and all rows summing to exactly 1. Then M is a contraction wrt the 2-norm: $\|M\mathbf{x}\| \leq \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^n$.*

Proof.

$$\begin{aligned} \|Mx\|^2 &= \sum_i (Mx)_i^2 = \sum_i \left(\sum_j M_{ij}x_j \right)^2 \\ &\leq \sum_i \sum_j M_{ij}x_j^2 \quad (\text{by Jensen's inequality, since } \sum_j M_{ij} = 1) \\ &= \sum_j \left(\sum_i M_{ij} \right) x_j^2 \\ &= \sum_j x_j^2 = \|x\|^2 \end{aligned}$$

□

The matrices \tilde{B} and P are both stochastic; this is trivial for P , and for \tilde{B} it is simply because all transition matrices are stochastic (hence the name!). So by the above claim,

$$|f^{\parallel T} \tilde{B}P\tilde{B}f^\perp| \leq \|f^\parallel\|^T \tilde{B} \cdot \|P\tilde{B}f^\perp\| \leq \|f^\parallel\| \|f^\perp\|.$$

From all this, we get

$$\begin{aligned} |f^T Z f| &\leq \alpha \|f^\parallel\|^2 + 2\beta \|f^\parallel\| \cdot \|f^\perp\| + \beta^2 \|f^\perp\|^2 \\ &= F^T \begin{pmatrix} \alpha & \beta \\ \beta & \beta^2 \end{pmatrix} F, \end{aligned}$$

where $F = \begin{pmatrix} \|f^\parallel\| \\ \|f^\perp\| \end{pmatrix}$

$$\leq \lambda_{max} \|F\|^2.$$

Here, λ_{max} is the largest e-value of the above matrix, which can be checked to be less than $\alpha + \beta + \beta^2$.

□

5 Using the zig-zag product to construct large expanders

While $G \circledast H$ has good expansion if G and H do, it's not quite *as* good. To handle this, we will use the following iterative construction, starting from a small expander H .

Let H be a $(d^4, d, 1/4)$ -graph. Such graphs do exist: this can be shown using the probabilistic method (Christophe's lecture). Since d is "small", we can find them just by an exhaustive search. Now define

$$G_1 = H^2, \quad G_{n+1} = (G_n)^2 \circledast H \quad \forall n \geq 1.$$

The point is that the squaring operation improves the expansion, since for an (n, d, α) -graph G , G^2 is an (n, d^2, α^2) graph. The problem is it pumps up the degree; but the zig-zag product takes its degree from the second operand H , and so that's fine.

Theorem 8. *For each $n \geq 1$, G_n is a $(d^{4n}, d^4, 1/2)$ -graph, and hence $\{G_n : n \in \mathbb{N}\}$ is an expander family.*

Proof. We proceed by induction. The claim is clearly true for $n = 1$.

So assume G_n is a $(d^{4n}, d^4, 1/2)$ -graph. So $(G_n)^2$ is a $(d^{4n}, d^8, 1/2)$ -graph—this follows by noting that $\hat{A}((G_n)^2) = \hat{A}(G_n)^2$, and hence the eigenvalues are simply squared. So now applying the (weakened) zig-zag theorem, we get that G_{n+1} is a $(d^{4n}, d^4, \varphi(1/2, 1/4))$ -graph, where

$$\varphi(1/2, 1/4) \leq 1/2 + 1/4 + (1/4)^2 < 1/2,$$

as required. □

We also get the following (I assume here the full zig-zag theorem, in particular that $\varphi(\alpha, \beta) < 1$ if $\alpha, \beta < 1$):

Corollary 9. *If G and H are an (n, m, α) and (m, d, β) graph respectively, then $G \circledast H$ is an $(nm, d + 1, \sqrt[3]{\varphi(\alpha, \beta)})$ -graph. In particular, it is an expander if G and H are.*

Proof. Observe that $(G \circledast H)^3 \supseteq G \circledast H$, since edges of $G \circledast H$ correspond to some subset of all 3-step walks in $G \circledast H$. Hence $\lambda_2((G \circledast H)^3)/d \leq \varphi(\alpha, \beta)$ (the extra edges can only help). Hence $G \circledast H$ is an $(nm, d + 1, \sqrt[3]{\varphi(\alpha, \beta)})$ -graph. □

We can in turn use this corollary to construct degree 3 expanders from a $d > 3$ expander G : just consider $G \circledast C_d$, where C_d is a cycle of length d .

References

- [1] N. Alon, O. Schwartz, and A. Shapira. An elementary construction of constant-degree expanders. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 454–458, 2007.
- [2] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43:439–561, 2006.
- [3] N. Kahale. Eigenvalues and expansion of regular graphs. *J. ACM*, 42(5):1091–1106, 1995.