Oded (January 29, 2023): On the analysis of the Zig-Zag product[1]

Given a (big) D-regular graph $G = (V, E)$, and a (small) d-regular graph $H = ([D], F)$, their Zig-Zag product, denoted $G \textcircled{z} H$, consists of the vertex set $V \times [D]$, which is partitioned to D-vertex clouds such that the cloud that corresponds to vertex $v \in V$ is the set of vertices $C_v = \{(v, i) : i \in [D]\}$, and edges that correspond to certain 3-step walks (as detailed next).

Actually, it is instructive to first consider the graph, denoted $G \textcircled{r} H$, in which copies of H are placed on the clouds (i.e., for every $v \in V$ and $\{i, j\} \in F$ we place the intra-cloud edge $\{(v, i), (v, j)\}$), and edges of G connect the corresponding clouds by using corresponding edges; that is, if $\{u, v\} \in E$ is the i^{th} (resp., j^{th}) edge incident at u (resp., at v), then we place the inter-cloud edge $\{(u, i), (v, j)\}$. Note that each vertex in $G \textcircled{r} H$ has d intra-cloud edges and a single inter-cloud edge. Now, the edges of $G \textcircled{z} H$ correspond to 3-step walks in $G \textcircled{r} H$ that start with an intra-cloud edge, then take the (only available) inter-cloud edge, and lastly take some intra-cloud edge; that is, such a generic walk has the form $(v, i) \to (v, j) \to (w, k) \to (w, \ell)$, where $\{i, j\}, \{k, \ell\} \in F$ and $\{(v, j), (w, k)\}$ is an inter-cloud edge in $G \textcircled{r} H$ (i.e., $\{v, w\} \in E$ is the j^{th} edge incident at v and the k^{th} edge incident at w).

We shall assume that both G and H are connected and are not bipartite. In that case it is clear that the graph $G \textcircled{r} H$ is also connected and non-bipartite, and it can be shown that also $G \textcircled{z} H$ has these properties. Showing the latter is simpler when assuming that H has self-loops on each vertex.[2]

Our focus is on upper-bounding the convergence rate of random walks on $G \textcircled{z} H$ (aka *second eigenvalue of the corresponding matrix*) in terms of the corresponding rates of the graphs G and H. (Recall that we refer to the eigenvalues of the corresponding normalized adjacency matrices, where the normalization consists of dividing each entry by the degree of the (regular) graph.) The best result of this type appeared in the paper of Omer Reingold, Salil Vadhan, and Avi Wigderson (*Annals of Math.*, Vol. 155 (1), 2001), and is stated next.

Theorem 1 (best known analysis of the Zig-Zag product): *Let $\lambda(X)$ denote the convergence rate of a random walk on the connected and non-bipartite graph X. Then, $\lambda(G \textcircled{z} H) \leq f(\lambda(G), \lambda(H))$, where*

$$f(x, y) \stackrel{\text{def}}{=} \frac{(1 - y^2) \cdot x}{2} + \sqrt{\left(\frac{(1 - y^2) \cdot x}{2}\right)^2 + y^2} \tag{1}$$

Clearly, $f(x, y) \leq (1 - y^2) \cdot x + y \leq x + y$ for $x, y \geq 0$. Furthermore, for $x \leq 1$, we have $f(x, y) \leq \frac{(1-y^2) \cdot x}{2} + \frac{1+y^2}{2} = 1 - \frac{(1-y^2) \cdot (1-x)}{2}$, which implies $1 - f(x, y) \geq \frac{(1-y^2) \cdot (1-x)}{2}$. Hence, for $y \leq \sqrt{1/3}$, we get $1 - f(x, y) \geq \frac{(2/3) \cdot (1-x)}{2} = (1 - x)/3$, and this is the result that is used in our presentation of the log-space UCONN algorithm (of Omer Reingold).[3]

We note that the proceeding version of RVW (*41st FOCS*, 2000) only claims that $f(x, y) \leq x + y$, whereas we shall only prove (see Lemma 2) that $f(x, y) \leq x + y + y^2$. We stress that while useful

[1] Notes for a follow-up communication to a group of graduate students at Weizmann Institute of Science.

[2] We first show that, for every $v \in V$, if (v, i) and (v, j) are neighbors in $G \textcircled{r} H$, then they are connected by an even length path in $G \textcircled{z} H$. This follows by considering the 3-step walks $(v, i) \to (v, j) \to (w, k) \to (w, \ell)$ and $(w, \ell) \to (w, k) \to (v, j) \to (v, j)$ on $G \textcircled{r} H$, where ℓ is an arbitrary neighbor of k in H. Next, we observe that each path (resp., cycle) is G corresponds to a (not necessarily simple) path (resp., cycle) in $G \textcircled{z} H$, and that the parity of the length of the path (resp., cycle) is preserved.

[3] Recall that $1 - f(\lambda(G), \lambda(H))$ is a lower bound on the spectral gap of $G \textcircled{z} H$ (i.e., on $1 - \lambda(G \textcircled{z} H)$), whereas $1 - \lambda(G)$ is the spectral gap of G. Indeed, we used the hypothesis that $\lambda(H) \leq \sqrt{1/3}$.

for other purposes, the bound $f(x,y) \le x+y+y^2$ (and even $f(x,y) \le x+y$) does not suffice for the analysis of the log-space UCONN algorithm (of Omer Reingold).

Intuition. The key observation is that, when starting in an arbitrary distribution $\overline{p}$ on the vertices of the Zig-Zag product $G \textcircled{z} H$, the distribution reached after taking a single random step on this graph is given by $BAB\overline{p}$, where B is the normalized matrix that represents a random step on the intra-cloud edges, and A is the adjacency matrix describing the inter-cloud edges. Note that B is an $|V| \cdot D$-by-$|V| \cdot D$ matrix that consists of $|V|$ copies of the matrix that corresponds to a random step on H, where these copies are placed on the diagonal (D-by-D) blocks of B. Likewise the vector $\overline{p}$ is the concatenation of $|V|$ vectors each representing the distribution restricted to the corresponding cloud (i.e., block v represents the restriction of $\overline{p}$ to the vertices in C_v).

Intuitively, applying the matrix B shrinks, in each block, the component of the restricted vector that is orthogonal to the uniform distribution on this cloud. On the other hand, applying the matrix A to a vector in which each block is uniform (but the different blocks may have different values), shrinks the component of this vector that is orthogonal to the uniform distribution on the entire graph. (Indeed, the first shrinkange is by a factor of at least $\lambda(H)$, whereas the second shrinkange is by a factor of at least $\lambda(G)$.)

A technical aspect. In light of the foregoing, we focus on the aforementioned orthogonal vectors; that is, the vectors that represent the components that are shrinked by either B or A. We shall use the fact that the convergence rate of a random walk on a regular graph X (equiv., the second eigenvalue of the corresponding normalized adjaceny matrix M) is captured by the maximum taken over all vectors $\overline{\alpha}$ orthogonal to the uniform vector of the ratio $\langle M\overline{\alpha}, \overline{\alpha}\rangle$ over $\langle \overline{\alpha}, \overline{\alpha}\rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product of vectors.[4] (Actually, we are interested in $\overline{\alpha} = \overline{p} - \overline{e}_1$, where $\overline{p}$ is an arbitrary probability vector (i.e., has non-negative entries that sum-up to 1) and $\overline{e}_1$ is a probability vector that represents the uniform distribution.)

Lemma 2 (the second eigenvalue of $G \textcircled{z} H$): *Let A and B be as above. Then, for every vector $\overline{\alpha}$ that is orthogonal to the uniform vector it holds that*

$$\langle BAB\overline{\alpha}, \overline{\alpha}\rangle \le (\lambda(G) + \lambda(H) + \lambda(H)^2) \cdot \langle \overline{\alpha}, \overline{\alpha}\rangle.$$

Proof: Following the foregoing intuition, we write $\overline{\alpha}$ as the sum of a vector in which each D-long block is uniform and a vector in which each block is orthogonal to the uniform D-long vector. Specifically, let $\overline{q}$ be a vector in which each entry in the v^{th} block holds the average of the entries of the v^{th} block of $\overline{\alpha}$, and $\overline{r} = \overline{\alpha} - \overline{q}$. Note that the $\overline{r}$ is orthogonal to $\overline{q}$, because, for each $v \in V$, the v^{th} block of $\overline{r}$ is a vector of sum 0 whereas the v^{th} block of $\overline{q}$ is uniform. Observing that $\langle BAB\overline{\alpha}, \overline{\alpha}\rangle = \langle AB\overline{\alpha}, B\overline{\alpha}\rangle$, we get

$$\begin{aligned}
\langle BAB\overline{\alpha}, \overline{\alpha}\rangle &= \langle AB(\overline{q}+\overline{r}), B(\overline{q}+\overline{r})\rangle \\
&= \langle A(\overline{q}+B\overline{r}), \overline{q}+B\overline{r}\rangle \\
&= \langle A\overline{q}, \overline{q}\rangle + 2 \cdot \langle A\overline{q}, B\overline{r}\rangle + \langle AB\overline{r}, B\overline{r}\rangle \\
&\le \langle A\overline{q}, \overline{q}\rangle + 2 \cdot \|A\overline{q}\| \cdot \|B\overline{r}\| + \|AB\overline{r}\| \cdot \|B\overline{r}\|,
\end{aligned}$$

[4] Letting $\overline{\alpha} = \sum_{i>1} c_i \overline{e}_i$, where the $\overline{e}_i$'s are the eigenvectors that are orthogonal to the uniform vector and the λ_i's be the corresponding eigenvalues, we have $M\overline{\alpha} = \sum_{i>1} c_i \lambda_i \overline{e}_i$, and $\langle M\overline{\alpha}, \overline{\alpha}\rangle = \sum_{i>1} c_i^2 \lambda_i$ follows. On the other hand, $\langle \overline{\alpha}, \overline{\alpha}\rangle = \sum_i c_i^2$.

where the second equality is due to the fact that B applies a random step of H to each block of $\overline{q}$, whereas each such block is a uniform vector, the third equality is due to $\langle AB\overline{r}, \overline{q}\rangle = \langle B\overline{r}, A\overline{q}\rangle$, and the inequality is due to the Cauchy-Schwarz inequality. Using the fact that A, being a permutation matrix, preserves the length of vectors, we get

$$\langle BAB\overline{\alpha}, \overline{\alpha}\rangle \leq \langle A\overline{q}, \overline{q}\rangle + 2 \cdot \|\overline{q}\| \cdot \|B\overline{r}\| + \|B\overline{r}\|^2. \tag{2}$$

On the other hand, note that $\overline{r}$ is orthogonal to $\overline{q}$, and so $\langle \overline{\alpha}, \overline{\alpha}\rangle = \langle \overline{q}, \overline{q}\rangle + \langle \overline{r}, \overline{r}\rangle = \|\overline{q}\|^2 + \|\overline{r}\|^2$.

Claim 2.1 (the effect of A on $\overline{q}$): $\langle A\overline{q}, \overline{q}\rangle \leq \lambda(G) \cdot \langle \overline{q}, \overline{q}\rangle$.

Proof sketch: Recall that $\overline{q}$ is uniform on each block. Hence, $A\overline{q}$ is identical to applying the normalized adjacency matrix of G to an $|V|$-long vector in which the v^{th} entry equals the sum of the (identical) entries in the v^{th} block of $\overline{q}$. Denoting the corresponding matrix and vector by A' and $\overline{q}'$, respectively, we observe that $\langle A\overline{q}, \overline{q}\rangle = \langle A'\overline{q}', \overline{q}'\rangle / D$, whereas $\langle \overline{q}, \overline{q}\rangle = \langle \overline{q}', \overline{q}'\rangle / D$. (Note that for a uniform D-long vector $\overline{\rho} = (\rho, ..., \rho)$, it holds that $\langle \overline{\rho}, \overline{\rho}\rangle = D \cdot \rho^2 = \langle D \cdot \rho, D \cdot \rho\rangle / D$.) ■

Claim 2.2 (the effect of B on $\overline{r}$): $\|B\overline{r}\| \leq \lambda(H) \cdot \|\overline{r}\|$.

Proof sketch: Recall that each block of $\overline{r}$ is orthogonal to the uniform D-long vector, and that B applies the normalized adjacency matrix of H to each block. Hence, each block of $B\overline{r}$ shrinks by a factor of $\lambda(H)$ when compared to its length in $\overline{r}$.[5] ■

Conclusion. Combining Eq. (2) with the two claims, letting $q = \|\overline{q}\| / \|\overline{\alpha}\|$ and $r = \|\overline{r}\| / \|\overline{\alpha}\|$, and recalling that $\langle \overline{\alpha}, \overline{\alpha}\rangle = \|\overline{q}\|^2 + \|\overline{r}\|^2$ (equiv., $q^2 + r^2 = 1$), we get

$$\begin{aligned} \frac{\langle BAB\overline{\alpha}, \overline{\alpha}\rangle}{\langle \overline{\alpha}, \overline{\alpha}\rangle} &\leq \frac{\lambda(G) \cdot \langle \overline{q}, \overline{q}\rangle + 2\lambda(H) \cdot \|\overline{q}\| \cdot \|\overline{r}\| + (\lambda(H) \cdot \|\overline{r}\|)^2}{\|\overline{q}\|^2 + \|\overline{r}\|^2} && (3) \\ &= \lambda(G) \cdot q^2 + 2\lambda(H) \cdot qr + \lambda(H)^2 \cdot r^2 \\ &\leq \lambda(G) + \lambda(H) + \lambda(H)^2, && (4) \end{aligned}$$

where the last inequality follows from the fact that $q^2, r^2 \leq 1$ and $qr \leq 1/2$ for $q^2 + r^2 = 1$. ■

Towards proving Theorem 1. The core of the proof (of Lemma 2) is captured by Claims 2.1 and 2.2, whereas the rest of the proof (of Lemma 2) consists of setting-up the stage for these claims and deriving conclusions from them. The latter aspect is performed in a rather straightforward and wasteful manner, and this is the reason that the result captured by Lemma 2 falls short of proving Theorem 1. Specifically, Eq. (2) was wastefully derived by using the Cauchy-Schwarz inequality, and Eq. (4) was derived by using $q^2, r^2 \leq 1$ and $qr \leq 1/2$ for $q^2 + r^2 = 1$. In contract, tighter bounds can be derived by a more careful treatment of the various vectors involved (e.g., $A\overline{q}, B\overline{r}$ and $AB\overline{r}$); that is, paying attention to their relative lengths and to the angle between them.[6]

---

[5] Analogously to Footnote 4, let $\overline{\alpha} = \sum_{i>1} c_i \overline{e}_i$, and note that $M\overline{\alpha} = \sum_{i>1} c_i \lambda_i \overline{e}_i$. Then, $\langle M\overline{\alpha}, M\overline{\alpha}\rangle = \sum_{i>1} c_i^2 \lambda_i^2$, and $\|M\overline{\alpha}\| \leq \max_{i>1}\{|\lambda_i|\} \cdot \|\overline{\alpha}\|$ follows.

[6] Specifically, the analysis in RVW works with $\langle BAB\overline{\alpha}, \overline{\alpha}\rangle = \langle A\overline{q} + AB\overline{r}, \overline{q} + B\overline{r}\rangle$ rather than using Eq. (2), and refers to the angle between $\overline{q} + B\overline{r}$ and the space of eigenvectors of A associated with the eigenvalue 1, the angle between $\overline{q}$ and $\overline{\alpha}$, and the angle between $\overline{q}$ and $\overline{q} + B\overline{r}$.