Lecture 10: Random Walks in Expanders

Topics in Error-Correcting Codes (Fall 2022)
University of Toronto
Swastik Kopparty
Scribe: Gal Gross

1 Random walks

Let $G = (V, E)$ be a d-regular graph of order n (i.e., $|V| = n$) which is a λ-absolute spectral expander, where $\lambda = 0.9d$ for concreteness[1] We briefly recall the relevant definition. The adjacency matrix of G is real symmetric, and so is diagonalizable over $\mathbb{R}$, with real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The fact that G is d-regular implies that $\lambda_1 = d$, and by definition λ-absoluteness means that $\lambda_2, \ldots, \lambda_n \in [-\lambda, \lambda]$. In what follows, we fix an orthonormal eigenbasis $\vec{b}_1, \ldots, \vec{b}_n$, with $\vec{b}_1 = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$.

Consider now a random walk on G. A *random walk of length* ℓ starting at $w_0 \in V$, chooses w_1 uniformly at random among the neighbours of w_0 in G. Given w_1 we then choose w_2 uniformly at random among the neighbours of w_1 in G. We continue this process until we have chosen $w_1, \ldots, w_\ell$.

Recall that in order to write down the adjacency matrix of G, we must have fixed some labeling on vertices of G; in what follows it will be convenient to assume that the labels are provided by the elements of V (e.g., $V = \{1, \ldots, n\}$). This labeling allows us to identify the $v \in V$ with the standard basis vector $\vec{e}_v \in \mathbb{R}^{|V|} = \mathbb{R}^n$. Let $\vec{P}_\ell \in \mathbb{R}^n$ be the *probability distribution* of vertex w_ℓ in the random walk of length ℓ. The v-th component of P_ℓ is denoted $P_\ell(v)$, it is simply the probability that $w_\ell = v$. In particular,

$$\vec{P}_0(v) = \begin{cases} 1 & \text{if } v = w_0; \\ 0 & \text{otherwise.} \end{cases}$$

By the definition of the random walk we have the recursive formula for $\ell > 0$:

$$P_\ell(v) = \frac{1}{d} \sum_{w \sim v} P_{\ell-1}(w),$$

where the notation $w \sim v$ means that w and v are neighbours; so that the sum is over the neighbourhood of v. In terms of the vector $\vec{P}_\ell$ we therefore have

$$\vec{P}_\ell = \frac{1}{d} A \vec{P}_{t-1}$$

and by induction

$$\vec{P}_\ell = \left(\frac{1}{d} A \right)^\ell \vec{P}_0.$$

[1]Fact: λ may be as small as $O(\sqrt{d})$.

In terms of our eigenbasis we have $A = \sum \lambda_i \vec{b}_i \vec{b}_i^T$ and for any vector $\vec{u} = \sum \alpha_i \vec{b}_i$ we have $A^\ell \vec{u} = \sum \alpha_i \lambda_i^\ell \vec{b}_i$; note that $\alpha_i = \left\langle \vec{u}, \vec{b}_i \right\rangle$. In particular, $\left\langle \vec{P}_0, \vec{b}_1 \right\rangle = \frac{1}{\sqrt{n}}$. Moreover, for any $\vec{P}_\ell$ we have $\sum \left\langle \vec{P}_\ell, \vec{b}_i \right\rangle^2 = \left\| \vec{P}_\ell \right\|^2 = 1$, since $\vec{P}_\ell$ represents a probability distribution.

For the case $\vec{u} = \vec{P}_0$ we have by our recursive formula

$$\vec{P}_\ell = \left(\frac{1}{d} A \right)^\ell \vec{P}_0 = \alpha_1 \vec{b}_1 + \sum_{i=2}^{n} \left(\frac{\lambda_i}{d} \right)^\ell \vec{b}_i.$$

Letting $\vec{U}_n$ denote the *uniform distribution vector*, we see that $\vec{U}_n = \frac{1}{\sqrt{n}} \vec{b}_1 = \alpha_1 \vec{b}_1$. By the Pythagorean theorem,

$$\left\| \vec{P}_\ell - \vec{U}_n \right\|^2 = \left\| \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{d} \right)^\ell \vec{v}_i \right\|^2 = \sum_{i=2}^{n} \alpha_i^2 \left(\frac{\lambda_i}{d} \right)^{2\ell} \leq \left(\frac{\lambda}{d} \right)^{2\ell} \sum_{i=2}^{n} \alpha_i^2 \leq \left(\frac{\lambda}{d} \right)^{2\ell}.$$

(The first inequality follows from the λ-absolute expander assumption that $\lambda_2, \ldots, \lambda_n \in [-\lambda, \lambda]$; the second inequality follows from the fact that $\sum \alpha_i^2 = 1$.)

Since λ was assumed to be $0.9d$, we see that $\vec{P}_\ell$ very quickly becomes very close to uniform distribution. That is, random walks on expanders have high mixing. Quantitatively, for $\lambda \leq 0.9d$ and $\ell = O(\lg k)$ we have $\left\| \vec{P}_\ell - \vec{U}_n \right\|_2 \leq k^{-100}$ and by standard inequalities $\left\| \vec{P}_\ell - \vec{U}_n \right\| \leq k^{-99}$.

Remark: The same proof shows that for any connected d-regular non-bipartite graph random walks quickly approximate the uniform distribution.

2 Subset-avoiding random walks

Let $G = (V, E)$ be as before. Fix some small subset $S \subseteq V$ and a starting node $v_0 \in V$, say $|S| = 0.1n$. We'd like to bound the probability that a random walk of length ℓ starting at w_0 completely avoids the set S. I.e.,

$$\Pr[w_0, \ldots, w_\ell \notin S].$$

In full generality, this problem depends too much on the relationship between S and w_0. (For a trivial example, if $w_0 \in S$, the probability is always 0.) Thus, for a chance at analyzing the situation we need to introduce some randomness. We can take S to be random, but that would defeat the purpose of having a bound which only depends on the size of S. We therefore take w_0 to be random.

In the same notation for vectors and matrices as in the previous section, taking w_0 to be random is equivalent to $\vec{P}_0$ being the uniform distribution vector $\begin{bmatrix} 1/n & 1/n & \cdots & 1/n \end{bmatrix}^T$. Can we express the probability $\Pr[w_0 \notin S]$ in terms of this vector?

Let M be the $n \times n$ matrix whose v-th column is v if $v \notin S$ and 0 otherwise. Thus M is just the identity matrix with every vector (representing an element) in S replaced with the 0 vector. The vector $M\vec{P}_0$ is thus the vector $\vec{P}_0$ after changing every v-th entry for $v \in S$ to 0. Taking the sum of the elements $\vec{1}^T M\vec{p}_0$ we obtain the probability $\Pr[w_0 \notin S]$. (Here $\vec{1}$ denote the all 1s vector $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$.)

The previous paragraph is seemingly an overly complicated way to compute

$$\Pr[w_0 \notin S] = \frac{n - |S|}{n} = 0.9.$$

However, the advantage of doing everything in terms of matrices and vectors is that the procedure generalizes to $\Pr[w_0, w_1 \notin S]$. Indeed, $w_0, w_1 \notin S$ means that in our random-walk we should only consider starting-points not in S, this gives us $\frac{1}{d}A(M\vec{P}_0)$. Out of this result, we should only keep vectors not in S, so we multiply by M again: $M\frac{1}{d}A(M\vec{P}_0)$. Finally, to calculate the probability we sum the entries of the vector:

$$\Pr[w_0, w_1 \notin S] = \vec{1}^T M \frac{1}{d} A M \vec{P}_0.$$

By induction we therefore have

$$\Pr[w_0, w_1, \ldots, w_\ell \notin S] = \vec{1}^T \left(M \frac{1}{d} A \right)^\ell M\vec{P}_0 = \vec{1}^T M \left(\frac{1}{d} A M \right)^\ell \vec{P}_0.$$

The point of the rewriting the equality in the last step is that $\vec{P}_0$ is a unit vector $\left\| \vec{P}_0 \right\| = 1$. Our goal is therefore to estimate the expression above, which can be contextualized as the 1-norm of the matrix $\frac{1}{d}AM$. It turns out there are better tools for estimating the $(2,2)$-matrix-norm (i.e., operator norm) of $\frac{1}{d}AM$; and there are general theorems from analysis which relate the two norms.

Recall that for an $n \times n$ matrix Q, the *operator norm* is the minimum number γ such that

$$\|Q\vec{u}\| \leq \gamma \|\vec{u}\| \tag{1}$$

for all $\vec{u} \in \mathbb{R}^n$. Thus, it is a measure of the maximum amount by which Q can "stretch" a vector (where the direction may also change). Thus, we always have

$$\|Q\vec{u}\| \leq \gamma \|\vec{u}\| .$$

Dividing both sides of (1) by $\|\vec{u}\|$, we see that

$$\gamma = \sup_{\|\vec{u}\|=1} \|Q\vec{u}\| . \tag{2}$$

In analysis, one proves that the supremum is in fact a maximum. Since $\|\vec{x}\|^2 = \langle \vec{x}, \vec{x} \rangle$, we also have

$$\gamma^2 = \max_{\|\vec{u}\|=1} \langle Q\vec{u}, Q\vec{u} \rangle .$$

We shall need the follow fact[2]

[2]For a proof, see the end of this document.

Fact 0. *For Q real symmetric matrix with operator norm γ,*

$$\gamma = \max_{\|\vec{u}\|=1} \langle \vec{u}, Q\vec{u} \rangle .$$

We are now ready to carry on with the computation. We are interested in the operator norm of $\frac{1}{d}AM$. It is easy to show that the norm is ≤ 1, but we'd like to prove it is < 1, since we want to show that iterating the process above (random walk) significantly reduces the probability. This turns out to be difficult to do for the matrix $\frac{1}{d}AM$, so we change the problem by symmetrizing the expression. Since M was obtained from the identity matrix by changing some diagonal entries to 0, it is symmetric and $M^T = M$ and $M^2 = M$. We can therefore rewrite $\Pr[w_0, w_1, \cdots, w_\ell \notin S]$ one more time:

$$\Pr[w_0, w_1, \ldots, w_\ell \notin S] = \vec{1}^T \left(M\frac{1}{d}A \right)^\ell M\vec{P}_0 = \vec{1}^T (M\frac{1}{d}AM)^\ell \vec{P}_0$$

and so we shall therefore estimate the operator norm of $M\frac{1}{d}AM$.

By 0, we need to bound

$$\max_{\|\vec{u}\|=1} \left\langle \vec{u}, M\frac{1}{d}AM\vec{u} \right\rangle$$

and since $M^T = M$, we have

$$\left\langle \vec{u}, M\frac{1}{d}AM\vec{u} \right\rangle = \left\langle M\vec{u}, \frac{1}{d}AM\vec{u} \right\rangle .$$

Let us therefore write $M\vec{u}$ in our orthonormal basis:

$$M\vec{u} = \sum_{i=1}^{n} \alpha_i \vec{b}_i$$

with $\alpha_i = \left\langle M\vec{u}, \vec{b}_i \right\rangle$. In particular, again using the symmetry of M, and Cauchy-Schwartz,

$$\alpha_1 = \frac{1}{\sqrt{n}} \left\langle M\vec{u}, \vec{b}_i \right\rangle = \frac{1}{\sqrt{n}} \langle \vec{u}, \vec{u} \rangle M\vec{b}_1 = \sqrt{\frac{n - |S|}{n}}.$$

Since $\vec{b}_1, \ldots, \vec{b}_n$ is an eigenbasis for A, we have

$$\frac{1}{d}AM\vec{u} = \sum_{i=1}^{n} \alpha_i \frac{\lambda_i}{d} \vec{b}_i.$$

Therefore, using the fact that $\vec{b}_1, \ldots, \vec{b}_n$ is orthonormal,

$$\left\langle M\vec{u}, \frac{1}{d}AM\vec{u} \right\rangle = \left\langle \sum_{i=1}^{n} \alpha_i \vec{b}_i, \sum_{i=1}^{n} \alpha_i \frac{\lambda_i}{d} \vec{b}_i \right\rangle = \sum_{i=1}^{n} \alpha_i^2 \frac{\lambda_i}{d}.$$

We know $\lambda_1 = d$ and $\lambda_2, \ldots, \lambda_n \leq \lambda$ (by the λ-absolute expansion) so that

$$\sum_{i=1}^{n} \alpha_i^2 \frac{\lambda_i}{d} \leq \alpha_1^2 + \frac{\lambda}{d} \sum_{i=2}^{n} \alpha_i^2.$$

Finally, since $\vec{u}$ was assumed to be a unit vector $\sum_{i=1}^{n} \alpha_i^2 = 1$, and we know what α_1 is. Plugging this information we obtain

$$\alpha_1^2 + \frac{\lambda}{d} \sum_{i=2}^{n} \alpha_i^2 \leq \frac{n - |S|}{n} + \frac{\lambda}{d}$$

and this is the bound on the operator norm we were looking for.

Quantitatively, for some $\varepsilon, \eta < 1$ such that $|S| < \varepsilon n$ and $\lambda < \eta d$, as long as $\varepsilon - \eta > 0$ we see that the bound above is strictly less than 1, and so we have exponential decay of the probability $\Pr[w_0, \ldots, w_\ell \notin S]$.

3 Error-reduction for randomized algorithm

Fix some function $f : \{0,1\}^n \to \{0,1\}$. A computable function $A : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ is a *randomized algorithm* for f if for every $\vec{x} \in \{0,1\}^n$

$$\Pr[A(\vec{x}, \vec{r}) = f(\vec{x})] \geq 0.9,$$

where the probability is taken over all $\vec{r} \in \{0,1\}^m$.

Thus, given m random bits, the algorithm A computes f with error 0.1. The standard way to reduce the error is to run $A(\vec{x}, \vec{r})$ for many independently chosen $\vec{r}$ and then return the majority vote.

There is a related notion of random computation with 1-sided error. A computable function $A : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ is a *randomized algorithm* (for f) *with one-sided error* if: for every $\vec{x} \in \{0,1\}^n$,

- if $f(x) = 0$, $A(\vec{x}, \vec{r}) = 0$ for every $\vec{r} \in \{0,1\}^m$;
- if $f(x) = 1$, $\Pr[A(\vec{x}, \vec{r}) = 1] \geq 0.9$.

Once again, the standard way to reduce the error is to run $A(\vec{x}, \vec{r})$ for ℓ independently chosen $\vec{r}$, and then take the "or" of the result:

$$A(\vec{x}, \vec{r}_1) \vee A(\vec{x}, \vec{r}_2) \vee \cdots \vee A(\vec{x}, \vec{r}_\ell).$$

This procedure will reduce the error from 0.1 to $(0.1)^\ell$, with the price that we now need $m\ell$ random bits.

Another idea, which costs less random bits, is to use a random walk on expander graphs. In detail, take G a d-regular graph on 2^m vertices labeled by $\{0,1\}^m$, which is also λ-absolute expander with

$\lambda \leq 0.01d$. Pick $w_0 \in \{0,1\}^m$ uniformly at random and take a random walk on G of length $\ell - 1$. Use the vertices of the random walk instead of the independently chosen $\vec{r}$. That is, we return

$$A(\vec{x}, w_0) \vee A(\vec{x}, w_1) \vee \cdots \vee A(\vec{x}, w_\ell).$$

We think of the set S from the previous section as the subset of $\{0,1\}^m$ such that $A(\vec{x}, \vec{r})$ returns the correct result (so we are guaranteed that $|S| \geq 0.9\,|\{0,1\}^m|$ since the error-rate of A is < 0.1). The probability that the disjunctive expression above returns the wrong answer (in the case $f(x) = 1$) is the same as the probability that $w_0, \ldots, w_\ell \notin S$, which according to our estimations from Section 2 scales as

$$\left(\frac{n - |S|}{n} + \frac{\lambda}{d}\right)^\ell \approx (0.1 + 0.01)^\ell$$

(for the parameters we've chosen). Thus, we get a comparable error-reduction to iterating the algorithm. However, in this procedure we've spend m random bits to choose w_0, and then $\log d$ random bits to choose w_{i+1} from among the d neighbours of w_i. Thus, the total cost of randomness is $m + \ell \log d$ bits, an improvement over the $m\ell$ random bits required for ℓ independent choices of $\vec{r}$.

For this procedure to be efficiently computable, we need to produce the expander graph in $\mathcal{P}oly(m)$ time, which is $\mathcal{P}olylog(|V|)$. This is a more stringent requirement than what we usually ask of an "explicit construction" which should produce an expander graph in time $\mathcal{P}oly(|V|)$.

Next class we'll see an "explicit construction" (i.e., in time $\mathcal{P}oly(|V|)$) of expander graphs using the zigzag product of graphs.

Appendix: Fact 0 follows from the spectral theorem for symmetric matrices

Fact 0. *For any symmetric matrix Q,*

$$\max_{\|\vec{u}\|=1} \langle Q\vec{u}, Q\vec{u} \rangle = \left(\max_{\|\vec{u}\|=1} \langle \vec{u}, Q\vec{u} \rangle \right)^2.$$

Proof. We first prove the claim for the special case of a diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$. Let $d = \max\{d_1, \ldots, d_n\}$ be the largest eigenvalue. Then, for any $\vec{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$ with $\|\vec{u}\| = 1$, i.e., $\sum_{i=1}^n u_i^2 = 1$, we have

$$\langle D\vec{u}, D\vec{u} \rangle = \sum_{i=1}^n (d_i u_i)^2 \leq d^2 \sum_{i=1}^n u_i^2 = d^2.$$

On the other hand, there is some $\vec{e}_i$ such that $\|D\vec{e}_i\|^2 = d^2$. We conclude that

$$\max_{\|\vec{u}\|=1} \langle Q\vec{u}, Q\vec{u} \rangle = d^2.$$

Exactly the same reasoning shows that

$$\max_{\|\vec{u}\|=1} \langle \vec{u}, Q\vec{u} \rangle = d,$$

which proves the claim for diagonal matrices.

For an arbitrary real symmetric matrix Q, the spectral theorem says that Q is diagonalizable by some orthogonal matrix S, which is necessarily an isometry. Thus, $D = S^{-1}QS$, and for any $\vec{u}$ whatsoever we have

$$\|D\vec{u}\| = \|Q\vec{u}\|.$$

In particular, the operator norm of Q is the same as that of D (the largest eigenvalue of Q). Now, $Q = SDS^{-1}$ and since S is orthogonal $S^T = S^{-1}$, so

$$\langle \vec{u}, Q\vec{u} \rangle = \left\langle \vec{u}, SDS^{-1}\vec{u} \right\rangle = \left\langle S^{-1}\vec{u}, DS^{-1}\vec{u} \right\rangle.$$

Finally, since S is an isometry, taking the maximum over all $\vec{u}$ with $\|\vec{u}\| = 1$ is the same as taking the maximum over all $S^{-1}\vec{u}$ with $\left\|S^{-1}\vec{u}\right\| = 1$ so we see that

$$\max_{\|\vec{u}\|=1} \langle \vec{u}, Q\vec{u} \rangle = \max_{\|\vec{u}\|=1} \langle \vec{u}, D\vec{u} \rangle.$$

This conclude the proof. $\square$