CS225: Pseudorandomness Prof. Salil Vadhan

Lecture 8: Random Walks on Expanders

March 1, 2007

Based on scribe notes by Mihai Pătraşcu.

1 Rapid Mixing of Random Walks

From the previous lecture, we know that one way of characterizing an expander graph G is by having a bound λ on their second eigenvalue, and in fact there exist constant-degree expanders where λ is a constant less than 1. From Lecture 4, we know that this implies that the random walk on G converges quickly from the uniform distribution. Specifically, a walk of length t started at any vertex ends at ℓ_2 distance at most λ^t from the uniform distribution. Thus after $t = O(\log N)$ steps, the distribution is very close to uniform (e.g. the probability of every vertex is $(1 \pm .01)/N$. Note that, if G has constant degree, the number of random bits invested here is $O(t) = O(\log N)$, which is within a constant factor of optimal; clearly $\log N - O(1)$ random bits are also necessary sample an almost uniform vertex. Thus, expander walks give a very good tradeoff between the number of random bits invested and the 'randomness' of the final vertex in the walk. Remarkably, expander walks give good randomness properties not only for the final vertex in the walk, but also of the sequence of vertices in the walk. Indeed, in several ways to be formalized below, this sequence of vertices 'behaves' like uniform independent samples of the vertex set.

A canonical application of expander walks is for randomness-efficient error reduction of randomized algorithms: Suppose we have an algorithm with constant error rate, which uses m random bits. Our goal is to reduce the error to 2^{-k}, with a minimal penalty in random bits and time. Independent repetitions of the algorithm suffers just an $O(k)$ penalty in time, but needs $O(km)$ random bits. We have already seen that with pairwise independence we can use just $O(m + k)$ random bits, but the time blows up by $O(2^k)$. Expander graphs let us have the best of both worlds, using just $m + O(k \lg D)$ random bits, and increasing the time only by $O(k)$. Note that for constant D, the number of random bits is $m + O(k)$, even better than what pairwise independence gives.

The general approach is to consider an expander graph with vertex set $\{0, 1\}^m$, where each vertex is associated with a setting of the random bits. We will choose a uniform random vertex v_1 and then do a random walk on length $t - 1$, visiting vertices $v_1, \ldots, v_t$. (Note that unlike the rapid mixing case, here we start at a uniformly random vertex.) This requires m random bits for the initial choice, and $\log D$ for each of the $t - 1$ steps. For every vertex v_i on the random walk, we will run the algorithm with the setting of the random coins v_i.

First, we consider the special case of **RP** algorithms. Thus, we accept if at least one execution of the algorithm accepts, and reject otherwise. If the input is not in the language, the algorithm never accepts, so we also reject. If the input is in the language, we want our random walk to hit at least one vertex which makes the algorithm accept. Let B denote the set of "bad" vertices giving bad coin tosses (which make the algorithm reject). By definition, the density of B is at most a half. Thus, our aim is to show that the probability that *all* the vertices in the walk $v_1, \ldots, v_t$ are in B vanishes exponentially fast in t.

The case $t = 2$ follows from the Expander Mixing Lemma given last time. If we choose a random edge in a λ spectral expander, the probability that both endpoints are in a set B is at most $\mu(B)^2 + \lambda \cdot \mu(B)$. So if $\lambda \ll \mu(B)$, then the probability is roughly $\mu(B)^2$, just like two independent random samples. The case of larger t is given by the following theorem.

Theorem 1 (Hitting Property of Expander Walks) *If G is a λ spectral expander, then for any $B \subset V(G)$ of density μ, the probability that a random walk $(V_1, \ldots, V_t)$ of t steps in G starting in a uniformly random vertex V_1 always remains in B is*

$$Pr[V_1, \ldots, V_t \in B] \leq (\mu + \lambda \cdot (1 - \mu))^t$$

.

Equivalently, a random walk 'hits' the complement of B with high probability. Thus, if μ and λ are constants less than 1, then the probability is $2^{-\Omega(t)}$, completing the analysis of the efficient error-reduction algorithm.

Before proving the theorem, we discuss general approaches to analyzing spectral expanders and random walks on them. Typically, the first step is to express the quantities of interest linear-algebraically, involving applications of the random-walk (or adjacency) matrix M to some vectors v. For example, last time when proving the Expander Mixing Lemma, we expressed the fraction of edges between sets S and T as $\chi_S^t M \chi_T$ (up to some normalization factor). Then we can proceed in one of the two following ways:

Vector Decomposition Decompose the input vector v as $v = v^{\|} + v^{\perp}$, where $v^{\|} = (\langle v, u\rangle / \langle u, u\rangle)u$ is the component of v in the direction of the uniform distribution u and $v^{\perp}$ is the component of v orthogonal to u. Then this induces a similar orthogonal decomposition of the output vector Mv into $Mv = Mv^{\|} + Mv^{\perp} = (Mv)^{\|} + (Mv)^{\perp}$, where $Mv^{\|} = v^{\|}$ and $\|Mv^{\perp}\| \leq \lambda \cdot \|v^{\perp}\|$. Thus, from information about how v's lengths are divided into the uniform and non-uniform components, we deduce information about how Mv is divided into the uniform and non-uniform components. This is the approach we took in the proof of the Expander Mixing Lemma.

Matrix Decomposition This corresponds to a different decomposition of the output vector Mv that can be expressed in a way that is independent of the decomposition of the input vector v. Specifically, we can write

$$Mv = (1 - \lambda)v^{\|} + (\lambda v^{\|} + Mv^{\perp}) = (1 - \lambda)Jv + \lambda Ev = ((1 - \lambda)J + \lambda E)v,$$

where J the matrix that projects onto direction u and the error matrix E satisfies $\|Ev\| \leq \|v\|$. The advantage of this decomposition is that we can apply it even when we have no information about how v decomposes (only its length), and the fact that M is a convex combination of J and E means that we can often treat each of these components separately and then just apply triangle inequality. However, it is less refined than the vector decomposition approach, and sometimes gives weaker bounds. Indeed, if we use it to prove the Expander Mixing Lemma (without decomposing χ_S and χ_T), we would get a slightly worse error term of $\lambda\sqrt{\mu(S)\mu(T)} + \lambda\mu(S)\mu(T)$.

The Matrix Decomposition Approach can be formalized as follows.

Definition 2 *The* norm *of an $N \times N$ matrix M is defined to be*

$$\|M\| = \max_{x \in \mathbb{R}^N} \frac{\|xM\|}{\|x\|}$$

If M is symmetric, then $\|M\|$ equals the largest *eigenvalue of M.*

Some basic properties of the matrix norm are that $\|cA\| = |c| \cdot \|A\|$, $\|A + B\| \leq \|A\| + \|B\|$, and $\|A \cdot B\| \leq \|A\| \cdot \|B\|$ for every two matrices A, B, and $c \in \mathbb{R}$. From the discussion above, we have the following lemma:

Lemma 3 *Let G be a regular digraph on N vertices with random-walk matrix M. Then G is a λ spectral expander iff $M = (1-\lambda)J + \lambda E$, where J is the $N \times N$ matrix where every entry is $1/N$ (i.e. the random-walk matrix for the complete graph with self-loops) and $\|E\| \leq 1$.*

This lemma has a nice intuition: we can think of a random step on a λ spectral expander as being a random step on the complete graph with probability $1-\lambda$ and "not doing damage" with probability λ. This intuition would be accurate if E were a stochastic matrix, but it is typically not (e.g. it may have negative entries). Still, note that the bound given in Theorem 1 exactly matches this intuition: in every step, the probability of remaining in B is at most $(1-\lambda) \cdot \mu + \lambda = \mu + \lambda \cdot (1-\mu)$.

Now we can return to the proof of the theorem.

Proof: We need a way to express getting stuck in B linear-algebraically. For that, we define P to be a diagonal matrix, with $P_{i,i} = 1$ if $i \in B$ and $P_{i,i} = 0$ otherwise. An example of using P would be to say that the probability a distribution π picks a node in B is $|P\pi|_1$. We use $|\cdot|_1$ for the ℓ_1 norm, $|x|_1 = \sum |x_i|$, which in our case is equal to the sum of the components of the vector (since all values are positive).

Let M be the random walk matrix of G. The probability distribution for V_1 is the vector u. Now we can state the following crucial fact: the probability that the random walk does not leave B is precisely $|uP(MP)^{t-1}|_1$. This is an intuitive formula, which can also be shown inductively without difficulty. To do that, one can show that $\left(uP(MP)^{\ell}\right)_i$ is the probability that a random walk never leaves B and ends in node i in the first $\ell + 1$ steps. The base case is $\ell = 0$. If $i \in B$, $(uP)_i = 1/N$; if $i \notin B$, $(uP)_i = 0$. Now assume the hypothesis holds up to some ℓ. Then $\left(uP(MP)^{\ell}\right)_i$ is the probability that the random walk is at i after $\ell + 2$ steps, and never leaves B until possibly the last step. Multiplying by P, we zero out all components for nodes not in B and leave the others unchanged. Thus, we obtain the probability that the random walk is at i after $\ell + 2$ steps, and never leaves B.

To get a bound in terms of λ, we will now switch to the standard, Euclidean ℓ_2 norm. The intuition is that multiplying by M shrinks any component that is perpendicular to u; then multiplying by P shrink the component parallel to u, because it zeroes out some entries. Thus, we should be able to show that the norm $\|MP\|$ is less than 1. Actually, to get the best bound, we note that $uP(MP)^{t-1} = uP(PMP)^{t-1}$, because $P^2 = P$, so we instead bound $\|PMP\|$.

Thus:

Claim 4 $\|PMP\| \leq \mu + \lambda \cdot (1-\mu)$.

Proof of claim:

$$\begin{aligned}
\|PMP\| &= \|P((1-\lambda)J + \lambda E)P\| \\
&\leq (1-\lambda)\|PJP\| + \lambda\|PEP\| \\
&\leq (1-\lambda) \cdot \|PJP\| + \lambda
\end{aligned}$$

Thus, we only need to analyze the case of J, the random walk on the complete graph. Given any vector x, let $y = xP$. Note that $\|y\| \leq \|x\|$ and y has at most μN coordinates. Then

$$xPJP = yJP = ((\sum_i y_i)u)P = (\sum_i y_i)uP,$$

so

$$\|xPJP\| \leq |\sum_i y_i\| \cdot \|uP\| \leq \sqrt{\mu N} \cdot \|y\| \cdot \sqrt{\frac{\mu}{N}} \leq \mu \cdot \|x\|.$$

Thus,

$$\|PMP\| \leq (1-\lambda)\mu + \lambda = \mu + \lambda \cdot (1-\mu).$$

□

So the probability of never leaving B in a t-step random walk is

$$\begin{aligned}
|uP(MP)^{t-1}|_1 &\leq \sqrt{\mu N} \cdot \|uP(MP)^{t-1}\|, \\
&\leq \sqrt{\mu N} \cdot \|uP\| \cdot \|PMP\|^{t-1} \\
&\leq \sqrt{\mu N} \cdot \sqrt{\frac{\mu}{N}} \cdot (\mu + \lambda \cdot (1-\mu))^{t-1} \\
&\leq (\mu + \lambda \cdot (1-\mu))^t
\end{aligned}$$

■

The hitting properties described above suffice for reducing the error of **RP** algorithms. What about **BPP**? This is handled by the following.

Theorem 5 (Chernoff Bound for Expander Walks) *Let G be a λ spectral expander on N vertices, and let $f : [N] \to [0,1]$ be any function. Consider a random walk $V_1, \ldots, V_t$ in G from a uniform start vertex V_1. Then for any $\varepsilon > 0$*

$$\Pr\left[\left|\frac{1}{t}\sum_i f(V_i) - \mu(f)\right| > \lambda + \varepsilon\right] \leq 2e^{-\Omega(\varepsilon^2 t)}.$$

Note that this is just like the standard Chernoff Bound, except that our additive approximation error increases by λ. Thus, unlike the Hitting Property we proved above, this bound is only useful when λ is sufficiently small (as opposed to bounded away from 1). This can be achieved by taking an appropriate power of the initial expander. However, there is a better Chernoff Bound for Expander Walks, where λ does not appear in the approximation error, but the exponent in the probability of error is $\Omega((1-\lambda)\varepsilon^2 t)$ instead of $\Omega(\varepsilon^2 t)$. The bound above will suffice for our purposes (where ε is typically a constant, as in error reduction for **BPP**.)

Proof: Let X_i be the random variable $f(V_i)$, and $X = \sum_i X_i$. Just like in the standard proof of the Chernoff Bound, we show that the expectation of the moment generating function $e^{rX} = \prod_i e^{rX_i}$ is not much larger than $e^{r\,\mathrm{E}[X]}$ and apply Markov's Inequality, for a suitable choice of r. However, here the factors e^{rX_i} are not independent, so the expectation does not commute with the product. Instead, we express $\mathrm{E}[e^{rX}]$ linear-algebraically as follows. Define a diagonal matrix P whose (i,i)'th entry is $e^{rf(i)}$. Then, similarly to the hitting proof above, we observe that

$$\mathrm{E}[e^{rX}] = \left|uP(MP)^{t-1}\right|_1 = \left|u(MP)^t\right|_1 \leq \sqrt{N} \cdot \|u\| \cdot \|MP\|^t \leq \|MP\|^t.$$

To see this, we simply note that each cross-term in the matrix product $uP(MP)^{t-1}$ corresponds to exactly one expander walk $v_1, \ldots, v_t$, with a coefficient equal to the probability of this walk times $\prod_i e^{f(v_i)}$. Again, we bound

$$\|MP\| \leq (1-\lambda) \cdot \|JP\| + \lambda \cdot \|EP\|.$$

Since J simply projects onto the uniform direction, we have

$$\begin{aligned} \|JP\|^2 &\leq \frac{\|uP\|^2}{\|u\|^2} \\ &= \frac{\sum_v (e^{r\cdot f(v)}/N)^2}{\sum_v (1/N)^2} \\ &= \frac{1}{N}\left(\sum_v e^{2rf(v)}\right) \\ &\leq \frac{1}{N}\left(\sum_v 1 + 2rf(v) + O(r^2)\right) \\ &= 1 + 2r\mu + O(r^2) \end{aligned}$$

for $r \leq 1$, and thus

$$\|JP\| \leq \sqrt{1 + 2r\mu + O(r^2)} \leq 1 + r\mu + O(r^2)$$

For the error term, we have

$$\|EP\| \leq \|P\| \leq e^r = 1 + r + O(r^2).$$

Thus,

$$\|MP\| \leq (1-\lambda)(1 + r\mu + O(r^2)) + \lambda \cdot (1 + r + O(r^2)) = 1 + (\mu+\lambda)r + O(r^2),$$

and we have

$$\mathrm{E}[e^{rX}] \leq (1 + (\mu+\lambda)r + O(r^2))^t \leq e^{(\mu+\lambda)rt + O(r^2 t)}.$$

By Markov's Inequality,

$$\Pr[X \geq (\mu + \lambda + \varepsilon)t] \leq e^{-\varepsilon rt + O(r^2 t)} = e^{-\Omega(\varepsilon^2 t)},$$

if we set $r = \varepsilon/c$ for a large enough constant c. ■

We now summarize the properties that expander walks give us for randomness-efficient error reduction and sampling.

For reducing the error of a **BPP** algorithm from $1/3$ to 2^{-k}, we have:

	Number of Repetitions	Number of Random Bits
Independent Repetitions	$O(k)$	$O(km)$
Pairwise Independent Repetitions	$O(2^k)$	$O(k+m)$
Expander Walks	$O(k)$	$m + O(k)$

For Sampling, where we are given an oracle to a function $f : \{0,1\}^m \to [0,1]$, we want to approximate $\mu(f)$ to within an additive error of ε, we have:

	Number of Samples	Number of Random Bits
Truly Random Sample	$O(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$	$O(\frac{m}{\epsilon^2} \log \frac{1}{\delta})$
Pairwise Independent Samples	$O(\frac{1}{\epsilon^2 \delta})$	$O(m + \log \frac{1}{\epsilon} + \log \frac{1}{\delta})$
Expander Walks	$O(\frac{1}{\varepsilon^2})$	$m + (\log \frac{1}{\delta}) \cdot (\log \frac{1}{\varepsilon})/\varepsilon^2$

The $\log(1/\varepsilon)$ factor in the number of random bits used by expander walks is actually not necessary and comes from the slightly weaker Chernoff Bound we proved. In any case, note that expander walks have a much better dependence on δ in the number of samples (as compared to pairwise independence), but have a worse dependence on ε in the number of random bits.

Before we end, we make an important remark: we have not actually given an algorithm for randomness-efficient error reduction! Our algorithm assumes an expander graph of size 2^m, i.e. exponential. Generating such a graph at random would use far too many coins. Even somehow generating it deterministically will not work, since we would have to write down an exponential-size object. In the next lectures, we will look at ways to implicitly construct an expander (without writing it down), and do random walks in such a graph.