0368-4159: First Course in Derandomization 3/12/18 – Lecture 8

Expanders

Lecturer: Amnon Ta-Shma

Scribe: Dean Doron

1 Undirected Graphs as Operators

Definition 1. *Let G be a (possibly weighted) undirected graph over n vertices with an adjacency matrix A_G. The normalized adjacency matrix, or the transition matrix, is the matrix $A = A_G D^{-1}$ where D is the diagonal degree matrix, i.e., $D[i,i] = \sum_j A_G[i,j]$ for every $i \in [n]$, and so*

$$A[i,j] \;=\; \frac{1}{d(j)} A_G[i,j].$$

If G is d-regular then A is Hermitian, and is simply $\frac{1}{d}A_G$.

Theorem 2. *Let G be an undirected graph over n vertices and let A be its normalized adjacency matrix. Let $\lambda_n, \ldots, \lambda_1$ be the eigenvalues of A. Then:*

1. *$\lambda_1, \ldots, \lambda_n$ are real.*
2. *$\lambda_1 = 1$.*
3. *$\lambda_n \geq -1$.*
4. *$\lambda_1 = \ldots = \lambda_k = 1$ and $\lambda_{k+1} < 1$ if and only if G has exactly k connected components.*
5. *$\lambda_n = -1$ if and only if at least one of the connected components of G is bipartite.*

Claim 3 (bipartite graphs)**.** *Let G be a d-regular undirected bipartite graph over n vertices and let A_G be its adjacency matrix. Then, the eigenvalues of A_G are symmetric around 0. That is, every positive eigenvalue λ_k has a negative eigenvalue $-\lambda_k$ and vice versa.*

As another exercise prove:

Claim 4. *Let A be the normalized adjacency matrix of a regular undirected graph over n vertices and let $\lambda_n \leq \ldots \leq \lambda_1$ be the eigenvalues of A. Then, $\lambda_2 = \max_{x \perp \mathbf{1}} \frac{x^\dagger A x}{x^\dagger x}$.*

Throughout, we denote $\bar{\lambda}(G) = \max_{i \neq 1} |\lambda_i|$. We also let $\mathbf{1}$ be the all-ones vector, J be the all-ones matrix and $\mathbf{1}_X$ is the vector which is 1 over some index set X and 0 elsewhere.

Claim 5. *Let G be an undirected graph over n vertices and let A be its normalized adjacency matrix. Then, $\bar{\lambda}(G) = \left\|A - \frac{1}{n}J\right\|$.*

To prove the claim note that J and G commute (prove!) and share a common orthonormal eigenvalue basis of eigenvectors. Then the claim is also immediate (prove!). A similar (almost identical) claim appears (with a proof) in Claim 16, and if you do not want to prove the claim yourself you can look there.

2 Random walks over expanders mix fast

In a random walk over a graph G, we start with some initial vertex v_0 and at each step we move from vertex v to an adjacent vertex in $\Gamma(v)$ with probability proportional to the degree of v. Namely, if A is the normalized adjacency matrix of G, we move from vertex i to vertex j with probability $A[j,i]$.

Suppose we start a random walk at a vertex chosen by a probability distribution p. After taking one step, the probability of being at vertex i is $\sum_j p_j A[i,j]$ so the probability distribution after one step is described by Ap.

Iterating the above reasoning, we see that, after a t-step random walk whose initial vertex is chosen according to p, the last vertex reached is distributed according to $A^t p$. We say that π is a *stationary distribution* if $A\pi = \pi$, i.e., no further steps change the distribution.

Does every random walk over a graph approach some stationary distribution? If so, how fast? In the case of undirected regular expanders, the uniform distribution is the stationary distribution and we converge to it in a rate that depends on $\bar{\lambda}(G)$. Indeed, $\bar{\lambda}(G^t) = \bar{\lambda}(G)^t$ so if $\bar{\lambda}(G)$ is bounded away from 1, $\bar{\lambda}(G^t)$ approaches 0 and $A^t \to \lambda_1 v_1 v_1^\dagger = \frac{1}{n}J$. Thus, $A^t p \to \frac{1}{n}\mathbf{1}$ for every distribution p. Formally:

Lemma 6. *Let G be a regular graph over n vertices with normalized adjacency matrix A. Then, for every distribution p over the vertices and integer t, we have*

$$\left\| A^t p - \frac{1}{n}\mathbf{1} \right\| \;\leq\; \bar{\lambda}(G)^t.$$

Proof. Note that for every distribution, $\frac{1}{n}Jp = \frac{1}{n}\mathbf{1}$ and recall that $\bar{\lambda}(G)^t = \bar{\lambda}(G^t) = \left\|A^t - \frac{1}{n}J\right\|$. Denote by $\lambda_n \leq \ldots \leq \lambda_1$ the eigenvalues of A and $v_n, \ldots, v_1$ the corresponding eigenvectors. Recall that $v_1 = \frac{1}{\sqrt{n}}\mathbf{1}$, $\lambda_1 = 1$ and we can write $A = \sum_i \lambda_i v_i v_i^\dagger$ and $A^t = \sum_i \lambda_i^t v_i v_i^\dagger$. Thus:

$$\left\| A^t p - \frac{1}{n}\mathbf{1} \right\| \;\leq\; \left\| A^t p - \frac{1}{n}Jp \right\| \;\leq\; \left\| A^t - \frac{1}{n}J \right\| \|p\| \;\leq\; \bar{\lambda}(G)^t.$$

□

Proof. (An alternative proof) Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors of A with eigenvalues $1 = \lambda_1 \geq \ldots \geq \lambda_n \geq -1$. $v_1 = \frac{1}{\sqrt{n}}\mathbf{1}$.

Express p as $p = \sum \alpha_i v_i$. The fact that p is a probability distribution implies $\alpha_1 = \langle p, v_1 \rangle = \frac{1}{\sqrt{n}}\sum_i p_i = \frac{1}{\sqrt{n}}$. Thus, $\alpha_1 v_1 = \frac{1}{n}\mathbf{1}$ and $p - \frac{1}{n}\mathbf{1} = \sum_{i \neq 1} \alpha_i v_i$. It follows that $\left\|A^t p - \frac{1}{n}\mathbf{1}\right\| \leq \left\|A^t(p - \frac{1}{n}\mathbf{1})\right\| = \left\|A^t(\sum_{i\neq 1}\alpha_i v_i)\right\| = \left\|\sum_{i \neq 1} \alpha_i \lambda_i^t v_i\right\| = \sqrt{\sum_{i \neq 1} |\alpha_i|^2 |\lambda_i^{2t}|} \leq \sqrt{\sum_i \bar{\lambda}(G)^{2t} \alpha_i|^2} = \bar{\lambda}(G)^t \|p\| \leq \bar{\lambda}(G)^t$.. □

We often measure the distance between distribution in the ℓ_1-norm, as it is, up to a factor of 2, equivalent to the *total variation distance* between probability distributions – the maximum over all events of the difference between the probability of the event happening with respect to one distribution and the probability of it happening with respect to the other distribution. As $\|x\|_1 \leq \sqrt{|\mathsf{Supp}(x)|}\,\|x\|_2$, we get:

Corollary 7. *Let G be a regular graph over n vertices with normalized adjacency matrix A. Then, for every distribution p over the vertices and integer t, we have*

$$\left\| A^t p - \frac{1}{n}\mathbf{1} \right\|_1 \;\leq\; \sqrt{n} \cdot \bar{\lambda}(G)^t.$$

Specifically, for $t = \Omega\left(\frac{1}{1-\bar{\lambda}(G)} \ln \frac{n}{\varepsilon}\right)$ we have $\left\|A^t p - \frac{1}{n}\mathbf{1}\right\|_1 \leq \varepsilon$.

The diameter of a graph is the maximum minimal distance between two vertices in the graph. For an undirected regular graph G, if $\bar{\lambda}(G)$ is constant bounded away from 1, the graph's diameter is logarithmic. More generally:

Lemma 8. *Let G be a d-regular undirected connected graph over n vertices. Then, the diameter of G is at most $1 + \log_{\frac{1}{\bar{\lambda}(G)}} n$.*

The proof follows from $\left\|A^t p - \frac{1}{n}\mathbf{1}\right\|_\infty \leq \left\|A^t p - \frac{1}{n}\mathbf{1}\right\| \leq \bar{\lambda}(G)^t$.

3 The Expander Mixing Lemma

We first show that an expander behaves like a random graph in the following sense: The number of edges between every two *large* subsets $S, T \subseteq [n]$ is close to what we would have expected in a random graph of average degree d, i.e., $\frac{d}{n}|S||T|$.

Lemma 9 (Expander Mixing Lemma)**.** *Let $G = (V = [n], E)$ be a d-regular graph and let $S, T \subseteq [n]$. Then,*

$$\left| |E(S,T)| - \frac{d|S||T|}{n} \right| \;\leq\; \bar{\lambda}(G) \cdot d\sqrt{|S|(1-|S|/n)|T|(1-|T|/n)}$$

where $|E(S,T)|$ is the number of edges between the two sets.

Proof. Let A be the normalized adjacency matrix of G, so we have

$$|E(S,T)| \;=\; d \cdot \mathbf{1}_T^\dagger A \mathbf{1}_S.$$

We decompose $\mathbf{1}_S$ and $\mathbf{1}_T$ to a component parallel to $\mathbf{1}$ (the 1-eigenvector of A) and a perpendicular component. Write $\mathbf{1}_S = \frac{|S|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_S^\perp$ where

$$\mathbf{1}_S^\perp[i] \;=\; \begin{cases} n - |S| & i \in S \\ -|S| & i \notin S. \end{cases}$$

and notice that $\mathbf{1}_S^\perp \perp \mathbf{1}$. Similarly we write $\mathbf{1}_T = \frac{|T|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_T^\perp$. Then, using the fact that $A\mathbf{1} = \mathbf{1}$:

$$\begin{aligned} E(S,T) &= d \cdot \left(\frac{|T|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_T^\perp\right)^\dagger A \left(\frac{|S|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_S^\perp\right) \\ &= d \cdot \frac{|S||T|}{n^2}\mathbf{1}^\dagger\mathbf{1} + \frac{1}{n^2}\left(\mathbf{1}_T^\perp\right)^\dagger A \mathbf{1}_S^\perp. \end{aligned}$$

As both $\mathbf{1}_S$ and $\mathbf{1}_S$ are perpendicular to the 1-eigenvector,

$$\left|\left(\mathbf{1}_T^{\perp}\right)^{\dagger} A\mathbf{1}_S^{\perp}\right| \;\leq\; \bar{\lambda}(G)\cdot\left\|\mathbf{1}_T^{\perp}\right\|\cdot\left\|\mathbf{1}_S^{\perp}\right\|.$$

A simple calculation shows that $\left\|\mathbf{1}_S^{\perp}\right\| = \sqrt{n|S|(n-|S|)}$ and likewise for $\left\|\mathbf{1}_T^{\perp}\right\|$, so overall

$$\begin{aligned}\left||E(S,T)| - \frac{d|S||T|}{n}\right| &\leq \bar{\lambda}(G)\cdot\frac{d}{n^2}\cdot\sqrt{n|S|(n-|S|)}\sqrt{n|T|(n-|T|)}\\ &= \bar{\lambda}(G)\cdot d\cdot\sqrt{|S|(1-|S|/n)}\sqrt{|T|(1-|T|/n)},\end{aligned}$$

as desired. □

Corollary 10. *With respect to densities (dividing by dn), we can express the above result as*

$$\left|\Pr_{e=(i,j)\in E}[i\in S\wedge j\in T] - \rho(S)\rho(T)\right| \;\leq\; \bar{\lambda}(G)\cdot\sqrt{\rho(S)(1-\rho(S))\rho(T)(1-\rho(T))},$$

where for $A\subseteq B$ we denote $\rho(A) = |A|/|B|$.

3.1 Expanders have no small cuts

An often desirable feature of a graph is that no deletion of few edges can cause the graph to be disconnected. It is indeed the case with expanders. Given an undirected d-regular graph $G=(V,E)$ we define the *edge expansion* of a cut $(S, V\setminus S)$ as

$$h(S) \;=\; \frac{|E(S,V\setminus S)|}{d\cdot\min\{|S|,|V\setminus S|\}},$$

and we let $h(G) = \min_{S\subseteq V} h(S)$.

exercise: Let $G=(V,E)$ be a d-regular undirected graph over n vertices. Use the expander mixing lemma to prove $h(G)\geq\frac{1-\bar{\lambda}(G)}{2}$.

We want to prove the stronger theorem:

Theorem 11. *Let $G=(V,E)$ be a d-regular undirected graph over n vertices and let λ_2 be the second eigenvalue of its normalized adjacency matrix A. Then, $h(G)\geq\frac{1-\lambda_2}{2}$.*

That is, for every $S\subseteq[V]$ of cardinality at most $\frac{n}{2}$, $|E(S,V\setminus S)|\geq\frac{d(1-\lambda_2)}{2}|S|$.

This theorem is one side of "Cheeger's Inequality". The other, harder, side is $h(G)\leq\sqrt{2(1-\lambda_2)}$ and we will not prove it. Morally, Cheeger's Inequality tells us that algebraic expansion and edge expansion are equivalent up to some loss in parameters.

Before we prove the theorem, we prove the following useful claim:

Claim 12. *Let M be a symmetric $n\times n$ operator, v a real length n vector. Let D be the $n\times n$ diagonal matrix with $D[i,i]=\sum_j M[i,j]$. Then*

$$\sum_{i,j} M[i,j](v_i-v_j)^2 \;=\; 2v^{\dagger}(D-M)v.$$

Proof. A straightforward computation shows that:

$$\begin{aligned}\sum_{i,j} M[i,j](v_i - v_j)^2 &= \sum_{i,j} M[i,j](v_i^2 + v_j^2) - 2\sum_{i,j} M[i,j]v_i v_j \;=\; 2\sum_{i,j} M[i,j]v_i^2 - 2\sum_i v_i \sum_j M[i,j]v_j \\ &= 2\sum_i v_i^2 \sum_j M[i,j] - 2\sum_i v_i(Mv)_i \;=\; 2v^\dagger D v - 2\sum_i v_i (Mv)_i \\ &= 2v^\dagger D v - 2v^\dagger M v \;=\; 2v^\dagger(D-M)v.\end{aligned}$$

□

Proof. We need to prove that $\lambda_2 \geq 1 - 2h(S)$ for every S with $|S| \leq \frac{n}{2}$. Equivalently, we can find a $v \perp \mathbf{1}$ for which $\frac{v^\dagger A v}{v^\dagger v} \geq 1 - 2h(S)$. Define a vector v such that:

$$v_i \;=\; \begin{cases} -n + |S| & i \in S \\ |S| & i \notin S. \end{cases}$$

First, notice that $v \perp \mathbf{1}$, as $\sum_i v_i = |S|(-n+|S|) + |S|(n-|S|) = 0$. Also, we have

$$v^\dagger v \;=\; |S|(-n+|S|)^2 + (n-|S|)|S|^2 \;=\; n|S|(n-|S|).$$

In our case,

$$\sum_{i,j} A[i,j](v_i - v_j)^2 \;=\; \frac{1}{d}\sum_{(i,j)\in E(S,\overline{S})} (|S| - (|S|-n))^2 \;=\; \frac{n^2}{d} 2|E(S, V\setminus S)|,$$

so $v^\dagger A v \;=\; v^\dagger v - \frac{n^2}{d}|E(S,V\setminus S)|$, and

$$\frac{v^\dagger A v}{v^\dagger v} \;=\; 1 - \frac{n^2|E(S,V\setminus S)|}{d\cdot v^\dagger v} \;=\; 1 - \frac{n|E(S,V\setminus S)|}{d\cdot |S|(n-|S|)} \;\geq\; 1 - \frac{2|E(S,V\setminus S)|}{d\cdot|S|} \;=\; 1 - 2h(S).$$

□

4 Deterministic amplification

Most of the material in this section (and a lot that is not in this section) is covered in a survey of Goldreich [1] and the monograph of Luby and Wigderson [2].

BPP is the class of decision problems solvable by a probabilistic Turing machine in polynomial time with a two-sided bounded error. RP and coRP are its one-sided variants. Formally:

Definition 13. *For $a < b$, a language $L \in \mathsf{BPP}[a,b]$ if there exists a polynomial-time probabilistic TM $M(x,y)$, where:*

- *If $x \in L$ then $\Pr_y[M(x,y) = 1] \geq b$.*
- *If $x \notin L$ then $\Pr_y[M(x,y) = 1] \leq a$.*

We denote $\mathsf{BPP} = \mathsf{BPP}[\frac{1}{3}, \frac{2}{3}]$, $\mathsf{RP} = \mathsf{BPP}[0, \frac{1}{2}]$ *and* $\mathsf{coRP} = \mathsf{BPP}[\frac{1}{2}, 1]$.

Suppose we have $L \in \mathsf{BPP}[a-\varepsilon, a+\varepsilon]$, for some constant a and $\varepsilon = \varepsilon(n)$, accepted by a TM M that on input of length n uses $t(n)$ random bits. If we run M k times, each time with fresh, independent, random bits and eventually output according to whether the average of k answers exceeded a, the error probability should decrease exponentially.

If we denote X_i as the answer in the i-th run, when $x \in L$ we err if $\frac{1}{k}\sum_{i=1}^{k} X_i < a$. By Chernoff, the probability for this to happen is bounded by $e^{-\Omega(\varepsilon^2 k)}$. Likewise for $x \notin L$. Thus, to bring the error to δ, we can take $k = O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$. Thus, we can amplify any polynomially large gap $\varepsilon = n^{-\alpha}$ to an exponentially small error $\delta = 2^{-n^c}$ in polynomial time, and therefore also using polynomially many random bits. The question we ask is whether we can re-use random bits and reduce the error without using too many additional random bits.

Throughout, we are given x and a black-box access to $M(x,y)$. We are allowed to pick $y_1,\dots,y_T$ in some way, and answer according to $M(x,y_1),\dots,M(x,y_T)$. Denote $m = |y|$. So far we have seen that with independent trials, with T queries and mT random coins we can amplify $(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)$ to $(\delta, 1-\delta)$ error with $T = O(\frac{1}{\varepsilon^2}\log\frac{1}{\delta})$.

4.1 Via pair-wise independence

Let us start with $k=2$. Pick $y_1,\dots,y_T$ from a pairwise independent distribution where each y_i is uniform over $\Sigma = \{0,1\}^m$. For every $i \in [T]$, let Y_i be the boolean random variable that is 1 iff $M(x,y_i)$ answered correctly. Denote $\mu_i = \mathbb{E}[Y_i] \geq \frac{1}{2}+\varepsilon$. We answer according to the median of the T trials. By Chebyshev and pairwise independence,

$$\begin{aligned}\Pr[\text{we are wrong}] &\leq \Pr\left[\left|\sum_{i=1}^{T} Y_i - \mu_i\right| \geq \varepsilon T\right] \\ &\leq \frac{\mathrm{Var}[\sum_i Y_i]}{\varepsilon^2 T^2} \leq \frac{(\frac{1}{2}-\varepsilon)(\frac{1}{2}+\varepsilon)}{\varepsilon^2 T} \leq \frac{1}{\varepsilon^2 T} = \delta.\end{aligned}$$

We thus choose $T = \frac{1}{\varepsilon^2\delta}$. The sample space is of size at most 2^{2m} so overall $2m$ random coins are used. If we want to amplify a non-negligible gap to a constant gap, it is sufficient to use pairwise independence.

4.2 Via k-wise independence

We proceed with $k=4$. For every $i \in [T]$, let X_i be the output of the i-th run and let $X = \sum_i X_i$, $\mu_i = \mathbb{E}[X_i]$ and $\mu = \sum_i \mu_i$. By Markov,

$$\Pr[|X-\mu| \geq A] \leq \Pr[(X-\mu)^4 \geq A^4] \leq \frac{\mathbb{E}[(X-\mu)^4]}{A^4}.$$

Denote $Z_i = X_i - \mu_i$, $\mathbb{E}[Z_i] = 0$. By linearity,

$$\mathbb{E}[(X-\mu)^4] = \mathbb{E}[(\sum_i Z_i)^4] = \sum_{i_1,i_2,i_3,i_4} \mathbb{E}[Z_{i_1}Z_{i_2}Z_{i_3}Z_{i_4}].$$

By four-wise independence, whenever all i_1,i_2,i_3,i_4 are different, $\mathbb{E}[Z_{i_1}Z_{i_2}Z_{i_3}Z_{i_4}] = E[Z_{i_1}] \cdot E[Z_{i_2}] \cdot E[Z_{i_3}] \cdot E[Z_{i_3}]$. However, for every i, $E[Z_i] = 0$, and so the term vanishes. In fact, this is true for every term i_1, i_2, i_3, i_4 in which some term appears with an odd power. Thus, the only terms that survive are those where every term appears an even number of times. Thus,

$$\begin{aligned}\mathbb{E}[(X-\mu)^4] &= \sum_a \mathbb{E}[Z_a^4] + \binom{4}{2} \sum_{1\le a<b\le T} \mathbb{E}[Z_a^2]\,\mathbb{E}[Z_b^2] \\ &= \sum_a \mathbb{E}[Z_a^4] + \binom{4}{2} \sum_{1\le a<b\le T} \operatorname{Var}[Z_a]\operatorname{Var}[Z_b].\end{aligned}$$

As for every i, $\operatorname{Var}[Z_i] = \mu_i(1-\mu_i) \le 1$,

$$\mathbb{E}[(X-\mu)^4] \;\le\; T + \binom{4}{2}\binom{T}{2} \;\le\; 4T^2.$$

We then obtain:

$$\begin{aligned}\Pr[\text{we are wrong}] &\le \Pr\left[\left|\sum_{i=1}^{T} Y_i - \mu_i\right| \ge \varepsilon T\right] \\ &\le \frac{\mathbb{E}[(X-\mu)^4]}{\varepsilon^4 T^4} \;\le\; \frac{4T^2}{\varepsilon^4 T^4} \;=\; \frac{4}{\varepsilon^4 T^2} \;=\; \delta.\end{aligned}$$

So, with four-wise independence, we get an error of $O(T^{-2})$. Specifically, we take $T = \frac{2}{\varepsilon^2}\sqrt{\frac{1}{\delta}}$. For arbitrary $2k$-independence, similar analysis shows that the error decreases like $O(T^{-k})$.

Lemma 14. *Let X be the average of T k-wise independent random variables for an even integer k, and let $\mu = \mathbb{E}[X]$. Then,*

$$\Pr[|X-\mu| \ge \varepsilon] \;\le\; \left(\frac{k^2}{4T\varepsilon^2}\right)^{\frac{k}{2}}.$$

The situation we have so far:

Table 1: Amplifying $(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)$ to $(\delta, 1-\delta)$ if r random bits are initially required

	Number of samples	Number of random bits
Truly random	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r \cdot O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
k-wise independence	$O(\frac{1}{\varepsilon^2}\frac{k^2}{\delta^{\frac{2}{k}}})$	$O(kr + k\log\frac{1}{\varepsilon} + \log\frac{1}{\delta})$
Pairwise independence	$O(\frac{1}{\varepsilon^2}\frac{1}{\delta})$	$O(r + \log\frac{1}{\delta\varepsilon})$

4.3 Via expanders

We start with a one-sided error $(0, \alpha)$ algorithm. With full independence, $O(\frac{1}{\alpha}\log\frac{1}{\delta})$ trials are sufficient (Check, and compare to the two sided error). Now, consider an expander $G = (V = \{0,1\}^m, E)$ with a constant degree D and a constant $\lambda = \min\left\{\lambda_2(G), -\lambda_{|V|}(G)\right\} < 1$.

The construction: Choose y_1 uniformly at random and take a random walk on G of length $T-1$ to obtain $y_2, \ldots, y_T$. Accept iff one of $M(x, y_i)$ accepted. Fix $x \in \{0,1\}^n$. If $x \notin L$ then we always reject, so we assume from now on that $x \in L$. Let $Bad \subseteq \{0,1\}^m$ be the set of strings that are bad for x. That is, $Bad = \{y \in \{0,1\}^m \mid M(x,y) = 0\}$. Thus,

$$\Pr[\text{we are wrong}] \;=\; \Pr\left[\bigwedge_{i=1}^{T}(y_i \in Bad)\right].$$

Then:

Theorem 15. *Using our above notations,*

$$\Pr\left[\bigwedge_{i=1}^{T}(y_i \in Bad)\right] \;\le\; (\beta + (1-\beta)\lambda)^T,$$

where $\beta = \frac{|Bad|}{|V|}$.

In our case, $\beta \le \alpha$ and $(\beta+(1-\beta)\lambda) = 1-(1-\lambda)(1-\beta) < 1$. Thus, with $m+\log D\cdot(T-1) = m+O(T)$ random coins we can amplify, say, $(0,\frac{1}{2})$ to $(0, 1-2^{-\Omega(T)})$.

Proof. The proof has two main components. First, we need to translate the condition $\bigwedge_{i=1}^{T}(y_i \in Bad)$ to an algebraic terminology, and then we analyze it.

The translation to algebraic terminology. Let M be the transition matrix of G and denote $|V| = 2^m = N$. Pick $y_1 \in V$ uniformly at random. That is, the initial distribution over the vertices is $u = \frac{1}{N}\mathbf{1}_N$. Define an $N \times N$ diagonal matrix B with $B[y,y] = 1$ if $y \in Bad$ and 0 otherwise. In this terminology, $|\langle \mathbf{1}, Bu\rangle|$ is the probability a random element belongs to BAD (and so is β). $|\langle \mathbf{1}, BMBu\rangle|$ is the probability in a random walk of length two, both samples belong to BAD. Similarly, $|\langle \mathbf{1}, (BM)^k Bu\rangle|$ is the probability in a random walk of length $k+1$ the walk is confined to the set BAD, i.e., all samples belong to BAD.

Reducing the analysis to understanding a single step : As B is a projection, $B^2 = B$, and so $(BM)^k Bu = (BMB)^k Bu$. Also, the vector is supported only on coordinates from Bad, Cauchy-Schwartz implies

$$|\langle \mathbf{1}, (BMB)^T Bu\rangle| \;\le\; \sqrt{\beta N}\left\|(BMB)^T Bu\right\|_2$$

and since $\|AB\|_2 \le \|A\|_2 \|B\|_2$,

$$\begin{aligned}
|\langle \mathbf{1}, (BMB)^T Bu\rangle| \;&\le\; \sqrt{\beta N}\,\|BMB\|_2^T\,\|Bu\|_2 \\
&=\; \sqrt{\beta N}\sqrt{\frac{\beta}{N}}\,\|BMB\|_2^T \\
&=\; \beta\,\|BMB\|_2^T \le \|BMB\|_2^T .
\end{aligned}$$

Summing up, it is enough to bound $\|BMB\|_2$, i.e., it is enough to analyze a single step.

Thus, we are left with analyzing a single step. We will show, $\|BMB\|_2 \le \beta + (1-\beta)\lambda$.

Claim 16 ([3], Proposition 3.2)**.** *Let G be an undirected regular graph on n vertices, with $\lambda = \min\left\{\lambda_2(G), -\lambda_{|V|}(G)\right\}$ and its transition matrix is B. Then, $B = (1-\lambda)J + \lambda E$ for some E with $\|E\|_2 \leq 1$ and J that is the normalized all-ones matrix. I.e., B is a convex combination of J (that corresponds to a completely random walk) and E (that is some arbitrary error matrix).*

Proof. The first eigenvector of B is u the all one vector (possibly normalized) with eigenvalue 1. u is also an eigenvector of J with eigenvalue 1. We conclude that u is a common eigenvector of B, J and E and with eigenvalue 1 for all of them (Check!).

What about vectors in the orthogonal complement? Let $W^{\perp}$ denote all vectors perpendicular to x, i.e., all x such that $\langle x, u\rangle = 0$. Then $Jx = 0$ (Why?). Also, $W^{\perp}$ is invariant under B (Why?). Thus, $W^{\perp}$ is invariant also under E (Why?).

Thus, to bound the norm of E, it is enough to limit attention to $W^{\perp}$. For $v \in W^{\perp}$, $\|Ev\| = \frac{1}{\lambda}\|Av\| \leq \frac{\lambda}{\lambda}\|v\| = \|v\|$. Thus, $\|E\|_2 \leq 1$. □

Now, let us express BMB in this decomposition. We get

$$BMB = B((1-\lambda)J + \lambda E)B \ = \ (1-\lambda)BJB + \lambda BEB$$

The BJB part is the part corresponding to a true random walk step, the other part is "junk", and indeed we easily see that $\|BEB\|_2 \leq \|B\|_2 \|E\|_2 \|B\|_2 \leq 1$. Thus, we are now reduced to analyzing BJB, i.e., one true random walk step. For any $x \neq 0$, $x = \sum_i x_i e_i$. Then, $(BJBx)[i] = \frac{1}{N}\sum_{i \in Bad} x_i$ if $i \in Bad$ and 0 otherwise (check!). Thus, by Cauchy-Schwarz,

$$\|BJBx\|_2 = \sqrt{\beta N \left(\frac{1}{N}\sum_{i \in Bad} x_i\right)^2} = \sqrt{\frac{\beta}{N}}\sum_{i \in Bad} x_i \leq \sqrt{\frac{\beta}{N}}\sqrt{\beta N}\,\|x\|_2 = \beta,$$

which completes the proof. □

The two-sided error case is along the same ideas, but a bit more complicated. The analysis may use the useful *expander Chernoff bound.*

Theorem 17. *Let G be an undirected D-regular graph with $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$ and spectral gap $1 - \bar{\lambda}$ and let $f_i : V \to [0,1]$ for $i \in [T]$. Take a random walk $v_1, \ldots, v_T$ and let X_i be the random variable $f_i(v_i)$. Denote $\mu_i = \mathbb{E}[X_i]$ and $\bar{\mu} = \frac{1}{T}\sum_i \mu_i$. Then,*

$$\Pr\left[\left|\frac{1}{T}\sum_i X_i - \bar{\mu}\right| \geq \varepsilon\right] \ \leq \ 2e^{-\frac{1}{4}(1-\bar{\lambda})\varepsilon^2 T}.$$

We can then add the expander walk technique to our table, obtaining:

Table 2: Amplifying $(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)$ to $(\delta, 1-\delta)$ if r random bits are initially required

	Number of samples	Number of random bits
Truly random	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r \cdot O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
Expander walk	$O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$	$r + O(\frac{\log\frac{1}{\delta}}{\varepsilon^2})$
k-wise independence	$O(\frac{1}{\varepsilon^2}\frac{k^2}{\delta^{\frac{2}{k}}})$	$O(kr + k\log\frac{1}{\varepsilon} + \log\frac{1}{\delta})$
Pairwise independence	$O(\frac{1}{\varepsilon^2}\frac{1}{\delta})$	$O(r + \log\frac{1}{\delta\varepsilon})$

References

[1] Oded Goldreich. A sample of samplers: A computational perspective on sampling. *def*, 1:2n, 1997.

[2] Michael Luby and Avi Wigderson. *Pairwise independence and derandomization.* Citeseer, 1995.

[3] Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 436–447. Springer, 2005.