6.842 Randomness and Computation	March 3, 2014
Lecture 8	
Lecturer: Ronitt Rubinfeld	*Scribe: Daniel Grier*

1 Useful Linear Algebra

Let $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ be a non-zero n-dimensional row vector and P an $n \times n$ matrix.

- We say $\mathbf{v}$ is an *eigenvector* of P with corresponding *eigenvalue* λ iff $\mathbf{v}P = \lambda\mathbf{v}$.
- The $\mathcal{L}_1$-norm of $\mathbf{v}$ (denoted $\|\mathbf{v}\|_1$) is $\sum_{i=1}^n v_i$.
- The $\mathcal{L}_2$-norm of $\mathbf{v}$ (denoted $\|\mathbf{v}\|_2$) is $\sqrt{\sum_{i=1}^n v_i^2}$.
- The *inner product* of two vectors $\mathbf{v}$ and $\mathbf{w}$ (denoted $\mathbf{v} \cdot \mathbf{w}$) is $\sum_{i=1}^n v_i w_i$.
- We say vectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots \mathbf{v}^{(m)}$ are *orthonormal* iff $\mathbf{v}^{(i)} \cdot \mathbf{v}^{(j)} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

Suppose P is an $n \times n$ matrix with positive entries, eigenvectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}$, and eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $\alpha \in \mathbb{R}$. Using the above definitions we derive the following facts:

Fact 1 *αP has eigenvectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}$ and eigenvalues $\alpha\lambda_1, \alpha\lambda_2, \ldots, \alpha\lambda_n$*

Proof $\mathbf{v}^{(i)}(\alpha P) = \alpha(\mathbf{v}^{(i)}P) = \alpha\lambda_i\mathbf{v}^{(i)}$. ■

Fact 2 *$P + I$ has eigenvectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}$ and eigenvalues $\lambda_1 + 1, \lambda_2 + 1, \ldots, \lambda_n + 1$*

Proof $\mathbf{v}^{(i)}(P + I) = \mathbf{v}^{(i)}P + \mathbf{v}^{(i)}I = \lambda_i\mathbf{v}^{(i)} + \mathbf{v}^{(i)} = (\lambda_i + 1)\mathbf{v}^{(i)}$. ■

Fact 3 *P^k has eigenvectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}$ and eigenvalues $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$*

Proof $\mathbf{v}^{(i)}P^k = (\mathbf{v}^{(i)}P)P^{k-1} = \lambda_i\mathbf{v}^{(i)}P^{k-1} = \lambda_i^2\mathbf{v}^{(i)}P^{k-2} = \ldots = \lambda_i^k\mathbf{v}^{(i)}$. ■

Fact 4 *If P is stochastic, then $|\lambda_i| \leq 1$ for all i.*

Proof For all i, let $I = \{j \mid v_j^{(i)} > 0\}$. Notice that we can force I to be non-empty. If $\mathbf{v}^{(i)}$ had all nonpositive entries, we could let $\mathbf{v}^{(i)} \leftarrow -\mathbf{v}^{(i)}$. Instead of trying to find a bound directly on λ_i, we attempt to find a bound on $\lambda_i \sum_{j \in I} \mathbf{v}_j^{(i)}$.

$$
\begin{aligned}
\lambda_i \sum_{j \in I} v_j^{(i)} &= \sum_{j \in I} \sum_{k=1}^{n} v_k^{(i)} P_{kj} \qquad \text{(select only the columns that produce positive value)} \\
&\leq \sum_{j,k \in I} v_k^{(i)} P_{kj} \qquad \text{(since } P \text{ has only positive entries)} \\
&= \sum_{k \in I} v_k^{(i)} \sum_{j \in I} P_{kj} \\
&\leq \sum_{k \in I} v_k^{(i)} \qquad \text{(since } P \text{ is stochastic)}
\end{aligned}
$$

This implies that $\lambda_i \leq 1$. Notice, however, that in forcing I to be non-empty we could have negated the value of the corresponding eigenvalue. Thus, what we should really conclude is that $|\lambda_i| \leq 1$. ■

Theorem 5 *Suppose P is a symmetric $n \times n$ transition matrix. P has eigenvectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}$ and corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that the eigenvectors are an orthonormal basis of $\mathbb{R}^n$, $1 = \lambda_1 \geq |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$, and $\mathbf{v}^{(1)} = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)$.*

The power of this theorem will be evident later when we use λ_2 to bound the size of all other eigenvalues (besides λ_1).

2 Mixing Times of Markov Chains

For $\epsilon > 0$, the *mixing time* $T(\epsilon)$ of a Markov chain with transition matrix P and stationary distribution Π is the minimum t such that $\|\Pi - \Pi^0 P^t\|_2 < \epsilon$ for all initial distributions Π^0. We say that a Markov chain is *rapidly mixing* if $T(\epsilon) = \text{poly}(\log n, \log \frac{1}{\epsilon})$ where n is the number of states.

Theorem 6 *Suppose P is the transition matrix of an undirected, nonbipartite, d-regular, connected Markov chain with starting distribution Π^0. The stationary distribution of the Markov chain is unique and equal to $\frac{1}{n}(1, 1, \ldots, 1)$. Furthermore, $\|\Pi^0 P^t - \Pi\|_2 \leq |\lambda_2|^t$ where λ_2 is the eigenvalue corresponding to the eigenvectors obtained from Theorem 5.*

Before we prove this theorem, it might help to take a moment to decipher what it tells us. First, we know that any ergodic Markov chain has a unique stationary distribution. However, the above Markov chain does not necessarily need to be ergodic, but it still has a unique (known) stationary distribution. For instance, the cycle of length k for any k falls into this category. As we will see later, this theorem provides an important method of determining how quickly a Markov chain converges to its stationary distribution. For example, when λ_2 is a constant less than 1, we have that the Markov chain is rapidly mixing (actually, t only depends on ϵ).

Proof Since P is undirected and d-regular, P is symmetric. Thus, P is real and symmetric, justifying our use of Theorem 5 to produce eigenvectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}$ with corresponding eigenvectors $1 = \lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$. Since these eigenvector form an orthonormal basis of $\mathbb{R}^n$, we can express Π^0 as a linear combination of the $v^{(i)}$'s. So,

$$
\begin{aligned}
\Pi^0 &= \sum_{i=1}^{n} \alpha_i \mathbf{v}^{(i)} \\
\implies \Pi^0 P^t &= \sum_{i=1}^{n} \alpha_i \mathbf{v}^{(i)} P^t \\
&= \sum_{i=1}^{n} \alpha_i \lambda_i^t \mathbf{v}^{(i)} \qquad \text{(using Fact 3)} \\
&= \alpha_1 \lambda_1^t \mathbf{v}^{(1)} + \sum_{i=2}^{n} \alpha_i \lambda_i^t \mathbf{v}^{(i)} \\
&= \alpha_1 \mathbf{v}^{(1)} + \sum_{i=2}^{n} \alpha_i \lambda_i^t \mathbf{v}^{(i)}
\end{aligned}
$$

Using the orthonormality of the basis, we can find the value of α_1. Recall from Theorem 5 that $\mathbf{v}^{(1)} = \frac{1}{\sqrt{n}}(1,1,\ldots,1)$.

$$\begin{aligned}
\Pi^0 \cdot \mathbf{v}^{(1)} &= \alpha_1 \mathbf{v}^{(1)} \cdot \mathbf{v}^{(1)} + \sum_{i=2}^{n} \alpha_i \lambda_i^t \mathbf{v}^{(i)} \cdot \mathbf{v}^{(1)} \\
\frac{1}{\sqrt{n}} \Pi^0 \cdot (1,1,\ldots,1) &= \alpha_1 \qquad \text{(since the } \mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)} \text{ are orthonormal)} \\
\frac{1}{\sqrt{n}} &= \alpha_1 \qquad \text{(since } \Pi^0 \text{ is a probability distribution)}
\end{aligned}$$

So, $\alpha_1 \mathbf{v}^{(1)} = \frac{1}{n}(1,1,\ldots,1)$. We claim now that this is fact the stationary distribution of the Markov chain. That is,

$$\begin{aligned}
\|\Pi^0 P^t - \frac{1}{n}(1,1,\ldots,1)\| &= \|\sum_{i=2}^{n} \alpha_i \lambda_i^t \mathbf{v}^{(i)}\| \qquad \text{(using above calculations)} \\
&= \sqrt{\sum_{i=2}^{n} \alpha_i \lambda_i^t \mathbf{v}^{(i)} \cdot \sum_{i=2}^{n} \alpha_i \lambda_i^t \mathbf{v}^{(i)}} \\
&= \sqrt{\sum_{i=2}^{n} \alpha_i^2 \lambda_i^{2t}} \qquad \text{(by orthonormality of basis vectors)} \\
&\le |\lambda_2|^t \sqrt{\sum_{i=2}^{n} \alpha_i^2} \qquad \text{(since } |\lambda_2| > |\lambda_i|\text{)} \\
&\le |\lambda_2|^t \|\Pi^0\|_2 \qquad \left(\text{since } \sqrt{\sum_{i=1}^{n} \alpha_i^2} = \|\Pi^0\|_2\right) \\
&\le |\lambda_2|^t \|\Pi^0\|_1 \qquad \text{(since } \mathcal{L}_1\text{-norm is at least } \mathcal{L}_2\text{-norm when entries at most 1)} \\
&= |\lambda_2|^t
\end{aligned}$$

We now state (without proof) that the nonbipartite property of P ensures that $|\lambda_2| < 1$. Thus, $|\lambda_2|^t$ goes to 0 as t goes to infinity. Thus, $\frac{1}{n}(1,1,\ldots,1)$ must be the stationary distribution for Π^0! Since there is no dependence on Π^0, we conclude that this is the unique stationary distribution for any starting distribution. ■

3 Using Markov Chains to Reduce Randomness

Recall our previous methods for reducing error for problems in RP. By repeating the algorithm k times, we used $O(k \cdot r)$ bits of randomness. Using ideas from pairwise independence, we were able to reduce to the randomness further to $O(k + r)$. We now give an approach using random walks on Markov chains that uses $r + O(k)$ bits of randomness.

We concern ourselves with problems that have one-sided error. That is, for algorithm $\mathcal{A}$ deciding language L we have

1. $\forall x \in L$, $\Pr[\mathcal{A}(x) = 1] \ge \frac{99}{100}$

2. $\forall x \notin L$, $\Pr[\mathcal{A}(x) = 0] = 1$

The idea is to associate all (random) strings in $\{0,1\}^n$ with nodes of a graph G. If $x \notin L$, we do not care which path we take on the graph because $\mathcal{A}$ will never accept. However, if $x \in L$, we wish to design G in such a way that a random walk starting from a random node is likely to arrive at *any* random string for which the algorithm accepts.

Lemma 7 *There exists a graph G on 2^r nodes with the following properties*

- *constant degree d-regular, connected, nonbipartite*
- *transition matrix for random walk on G has $\lambda_2 \le \frac{1}{10}$.*
- *uniform stationary distribution (since d-regular)*

Algorithm 1 RP error reduction algorithm

$w \leftarrow \{0,1\}^r$
repeat
  $w \leftarrow$ neighbor of w in G
  Run $\mathcal{A}(x)$ using randomness w
  if $\mathcal{A}(x)$ outputs "$x \in L$" **then**
    return "$x \in L$" and halt
  end if
until $\mathcal{A}$ does not accept k times
return "$x \notin L$"

Use Algorithm 1 to reduce error for RP problems, and note that all assignments are done uniformly at random. Examining the algorithm, r bits of randomness are used to choose the initial w and $\log d$ bits of randomness are used to choose a random neighbor on each iteration. Thus, the total amount of randomness used in the algorithm is $r + k \log d = r + O(k)$ since d is constant.

Claim 8 *Probability of error of Algorithm 1 is at most $\frac{1}{5^k}$ for $x \in L$. If $x \notin L$, probability of error is 0.*

Proof If $x \notin L$, then $\mathcal{A}$ never accepts, so probability of error is 0. If $x \in L$, then at least $\frac{99}{100}2^r$ choices of random bits have accepting paths in $\mathcal{A}$. Let B be the set capturing those random strings which are "bad". That is, $B = \{w \mid \mathcal{A}(x) \text{ with randomness } w \text{ rejects}\}$. By the above observation, $|B| \le \frac{2^r}{100}$.

To use the linear algebraic properties of G, we need a linear algebraic way to describe the random walks that stay within B. We define N as a $2^r \times 2^r$ diagonal matrix (i.e. the only non-zero elements are on the diagonal). The ith diagonal of N is 1 if $i \in B$ and is 0 otherwise.

Let Π be any probability distribution. We arrive at the following ideas.

$$\|\Pi N\|_1 = \Pr_{w \sim \Pi}[\mathcal{A}(x) \text{ rejects}]$$

$$\|\Pi P N\|_1 = \Pr_{w \sim \Pi}[\mathcal{A}(x) \text{ rejects after taking a random step}]$$

$$\vdots$$

$$\|\Pi (PN)^i\|_1 = \Pr_{w \sim \Pi}[\mathcal{A}(x) \text{ rejects on each of } i \text{ random steps}]$$

Notice that the expression $\|\Pi(PN)^i\|_1$ ignores the possibility that the initial w drawn from Π is a "bad" random string. However, since this only hurts our estimates, we are okay to ignore it.

Lemma 9 *For all Π (not necessarily probability distributions), $\|\Pi P N\|_2 \le \frac{1}{5}\|\Pi\|_2$.*

Before we prove the lemma, let us see how it implies the theorem. Let Π be the uniform distribution on $\{0,1\}^r$. The $\mathcal{L}_2$-norm of the uniform distribution is $\sqrt{\sum_{i=1}^{2^r}(\frac{1}{2^r})^2} = \sqrt{\frac{1}{2^r}}$.

$$\begin{aligned}
\Pr[\text{Algorithm 1 incorrect}] &\leq \|\Pi(PN)^k\|_1 \\
&\leq \sqrt{2^r}\|\Pi(PK)^k\|_2 \qquad \text{(by Cauchy-Schwarz)} \\
&\leq \sqrt{2^r}\|\Pi\|_2 \frac{1}{5^k} \qquad \text{(applying lemma } k \text{ times)} \\
&= \frac{1}{5^k} \qquad \text{(using above calculation of norm of uniform distribution)}
\end{aligned}$$

So Algorithm 1 only uses $r + O(k)$ bits of randomness and still guarantees that the error probability decreases exponentially in k. ■

Proof (of Lemma 9) This is where we use the nice linear algebraic properties of G. Since G is real and symmetric we can apply Theorem 5 to the transition matrix P of G. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{2^r}$ be the eigenvectors of P. Since $\mathbf{v}_1 = \frac{1}{2^r}(1,1,\ldots,1), \|\mathbf{v}_1\|_2 = 1$. Use the fact that the eigenvectors form a basis to write $\Pi = \sum_{i=1}^{2^r} \alpha_i \mathbf{v}_i$. So,

$$\begin{aligned}
\|\Pi PN\|_2 &= \|\sum_{i=1}^{2^r} \alpha_i \mathbf{v}_i PN\|_2 \\
&= \|\sum_{i=1}^{2^r} \alpha_i \lambda_i \mathbf{v}_i N\|_2 \\
&\leq \|\alpha_1\lambda_1\mathbf{v}_1 N\|_2 + \|\sum_{i=2}^{2^r} \alpha_i \lambda_i \mathbf{v}_i N\|_2 \qquad \text{(by triangle inequality)}
\end{aligned}$$

We will proceed by bounding each term separately. Intuitively, the first term should be small because we are unlikely to draw a "bad" string drawing uniformly from $\{0,1\}^r$. The second term should be small because the eigenvalues are small.

$$\begin{aligned}
\|\alpha_1\lambda_1\mathbf{v}_1 N\|_2 &= \|\alpha_1\mathbf{v}_1 N\|_2 \qquad \text{(since } \lambda_i = 1\text{)} \\
&= |\alpha_1| \sqrt{\sum_{i\in B}\left(\frac{1}{\sqrt{2^r}}\right)^2} \qquad \left(\text{since } \mathbf{v}_1 = \frac{1}{\sqrt{2^r}}(1,1,\ldots,1)\right) \\
&= |\alpha_1| \sqrt{\frac{|B|}{2^r}} \\
&\leq \frac{|\alpha_1|}{10} \qquad \left(\text{since } \frac{|B|}{2^r} \leq \frac{1}{100}\right) \\
&\leq \frac{\|\Pi\|_2}{10} \qquad \left(\text{since } \|\Pi\|_2 = \sqrt{\sum_{i=1}^{2^r} \alpha_i^2}\right)
\end{aligned}$$

$$
\begin{aligned}
\|\sum_{i=2}^{2^r} \alpha_i \lambda_i \mathbf{v}_i N\|_2 &\le \|\sum_{i=2}^{2^r} \alpha_i \lambda_i \mathbf{v}_i\|_2 \qquad \left(\text{since } \|\mathbf{v}N\|_2 = \sqrt{\sum_{i\in B} v_i^2} \le \sqrt{\sum_{i=1}^{2^r} v_i^2} = \|\mathbf{v}\|_2\right) \\
&= \sqrt{\sum_{i=2}^{2^r} (\alpha_i \lambda_i)^2} \\
&\le \sqrt{\sum_{i=2}^{2^r} \alpha_i^2 \left(\frac{1}{10}\right)^2} \qquad (\text{since } \lambda_i \le \frac{1}{10}) \\
&\le \frac{\|\Pi\|_2}{10}
\end{aligned}
$$

So, $\|\Pi P N\|_2 \le \frac{\|\Pi\|_2}{5}$. ■