6.842 Randomness and Computation March 7, 2022

Lecture 11

Lecturer: Ronitt Rubinfeld *Scribe: René D. Reyes*

In this lecture, we keep exploring random walks (pun intended) and how they may be applied to solving problems when certain space constraints are present, as well as reducing the number of random bits used in algorithms. In order to better understand random walks, we also review some relevant liner algebra. Based on this, the notes are structured as follows:

1. Undirected Connectivity and Randomized Logspace (RL)
2. Linear Algebra Review
3. Linear Algebra and Random Walks

1 Undirected Connectivity and Randomized Logspace

We start off with an example of a well-known problem that is fairly simple to solve in polynomial time, but requires clever ideas to solve when we are limited to logarithmic space. This is the problem of checking whether two nodes in an undirected graph are in the same connected component, defined as follows:

Definition 1 (*Undirected s-t Connectivity*) *We define the* USTConn *problem:*

- ***Input:*** *undirected graph G and nodes s, t*
- ***Output:*** *"Yes" if s,t are in the same connected component in G*
  "No" otherwise

It was stated in class that there are many well-known techniques to solve this problem in polynomial time, including graph traversal algorithms like Depth-First Search. We will now start exploring known results about solving this problem when the amount of read/write space is limited to $O(\log n)$.

1.1 Randomized Log-Space

First, we define the model of computation used when talking about the class of problems solvable in deterministic logarithmic space, also known as **L**. First off, when computing the space complexity of these TMs, we do not include the space used for the input, but only for any additional tape cells that get written on. Therefore, it is best to imagine a TM with two tapes, one for the input which is read-only and another work tape which is limited to $O(\log n)$ space.

At a high level, we can think of this logarithmic space limit as being restricted to only storing a constant number of pointers into the input, as well as a constant number of counters that take on values in $\{0, \cdots, n\}$. While in class we did not show that USTConn $\in$ **L**, we take a step towards proving this by showing that if we use randomness, then undirected connectivity can be solved in logspace. Before showing this, we define this complexity class of randomized logspace (**RL**):

Definition 2 (*RL*) *We define **RL** to be the class of problems that can be computed by a two-tape Probabilistic Turing Machine, with one tape being a read-only and the second tape being an $O(\log n)$ read/write work tape. Furthermore, the machine must have one-sided error, meaning that if the answer is "no", it always outputs "no" but if the answer is "yes' it outputs this with high probability.*

Note that this class is the analog of **RP** to **L**. Now, we show how random walks can be used to show that USTConn $\in$ **RL**.

1.2 Undirected Connectivity in RL

Recall the following definition from Lecture 10, regarding random walks on graphs:

Definition 3 (*Cover Time*) *The cover time of a random walk on a graph G is:*

$$C(G) = \max_v(\text{Exp}[\#\ \textit{of steps to visit all nodes in } G \textit{ when start at } v])$$

Then, we showed the following result about how the cover time relates to the size of the graph:

Theorem 4 *For any graph G, $C(G)$ is $O(n \cdot m)$*

It is worth pointing out that since the number of edges m is upper bounded by n^2, then $C(G) \leq n^3$. Based on this, we will use random walks to show the following result:

Theorem 5 $\mathsf{USTConn} \in \mathbf{RL}$.

Proof Note that if s and t are in the same connected component, then a random walk starting at s will eventually reach t with high probability after sufficient steps. If they are not in the same connected component, then there is a probability of 1 that this will not occur. Moreover, we can compute a random walk in a probabilistic log-space machine by simply keeping track of the current node we are visiting along, a counter to keep track of the number of nodes that have been visited so far, and some additional space to determine how to pick a random neighbor. We now give the algorithm:

On input (G, s, t):

1. Start at node s and initialize a counter
2. Take a random walk for $c \cdot n^3$ steps, using the counter to keep track
3. If at any point t is visited, then *accept*
   If the counter reaches $c \cdot n^3$, then *reject*

We now analyze the space complexity of the algorithm and analyze its probability of success.

Space Complexity: Keeping track of the space counter takes $O(\log(c \cdot n^3)) = O(\log n)$. Then, keeping track of the current node also takes $O(\log n)$ bits. Finally, to compute a random neighbor we need to scan through the input to count the number of neighbors d of the current node using a counter that will take $O(\log n)$ space. Based on this counter, we can pick a random number i between 1 and d and then scan the input again until we find the i-th neighbor. Finally, we set the next visited node to be equal to this neighbor. Therefore, we see how the random walk can be computed using log-space.

Probabilistic Analysis: As mentioned earlier, if s and t are not in the same connected component, then the random walk will never reach t. This means that if the real answer is "no", then:

$$\Pr[\text{Algorithm outputs "no" when real anser is "yes"}] = 1$$

In the case that they are in the same connected component, then we know that the time it takes the random walk to visit t is upper-bounded by the cover time of this connected component, which we call $C(G_s)$. By theorem 4, we know that $C(G_s) \leq n^3$. Then, since we run the random walk for $c \cdot n^3$ steps, the probability that the algorithm outputs "no" is equivalent to the probability that the cover time is c times greater than its expectation. This means that we can bound the probability of failure using Markov's inequality as follows:

$$\Pr[\text{output "no" when real answer is "yes"}] \leq \Pr[\text{time to cover graph} \geq c \cdot C(G_s)] \leq \frac{1}{c}$$

Therefore, it is sufficient to set $c \geq 3$ to get a success probability greater than $\frac{1}{2}$.

We have thus shown that the algorithm using random uses logarithmic space and succeeds with high probability, so $\mathsf{USTConn} \in RL$ ■

1.3 Some More Facts about Connectivity and Log-Space

It turns out that the *USTConn* problem can actually be shown to be in **L**, deterministic log-space via a derandomization technique due to Reingold. The proof was not covered in lecture.

We also discussed that whether or not the STConn problem for *directed* graphs is in **L** is an open question. If this were proven, then it would mean that **RL**= **L**, due to completeness of this problem. Some facts we do know are that $\mathbf{RL} \subseteq \mathbf{L}^{3/2}$, and more recently $\mathbf{RL} \subseteq \mathsf{SPACE}\left(\frac{log^{3/2}n}{\sqrt{\log\log n}}\right)$. This second fact was published in 2021 by William Hoza.

2 Linear Algebra Review

We now move on to the second half of the material covered in lecture, which was some review of linear algebra that is relevant to studying the behavior of Markov Chains (MCs). Now, we will revisit some of the theory from this field, and immediately see how they yield results about MCs.

First, some definitions.

Definition 6 *Let A be a square matrix. We say v is an eigenvector of A with corresponding eigenvalue λ iff:*

$$vA = \lambda v$$

Here, note that we are left-multiplying the vector. While we may be used to right-multiplying matrices in most Linear Algebra contexts, this is convention when talking about transition matrices of MCs.

We define more useful terminology:

Definition 7 *The L_2-norm of a vector $v = (v_1, \cdots, v_n)$ is equal to $||v||_2 = \sqrt{v \cdot v}$, meaning the square root of the vector's inner product with itself.*

Definition 8 *We say $v^{(1)}, \cdots, v^{(m)}$ are orthonormal if for all i, j pairs:*

$$v^{(i)} \cdot v^{(j)} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

We now give an example of how this all relates to MCs, particularly through transition matrices.

Example 9 *Let P be the transition matrix of a d-regular undirected graph. Then, since in-degree=outdegree, we know that P is doubly stochastic, which means that all of its rows and columns sum up to 1.*

Then, we mentioned the two following eigenvectors of P:

- $(\frac{1}{n}, \cdots, \frac{1}{n})P = 1 \cdot (\frac{1}{n}, \cdots, \frac{1}{n})$. *Since the entries of this vector add up to 1, then it can be interpreted as a probability distribution*
- $(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}})P = 1 \cdot (\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}})$. *While this is not a probability distribution, we also care about this eigenvector because it has L_2-norm equal to 1.*

2.1 Important Theorem and Facts

Now we discuss another theorem, whose importance was really emphasized in lecture:

Theorem 10 *If a transition matrix P is real and symmetric, then there exist eigenvectors $v^{(1)}, \cdots, v^{(n)}$ which form an orthonormal basis, and the corresponding eigenvalues are:*

$$1 = \lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$$

Once again, we did not prove this here, but the idea behind it follows from the fact that P is stochastic as well as the Real Spectral Theorem.

Now, we give some more general facts about matrices and their proofs:

Theorem 11 *Let P be a square matrix with eigenvectors $v^{(1)}, \cdots, v^{(n)}$ and correspondoing eigenvalues $\lambda_1, \cdots, \lambda_n$. Then the following hold:*

1. *For any scalar α, αP has eigenvectors $v^{(1)}, \cdots v^{(n)}$ with corresponding eigenvalues $\alpha\lambda_1, \cdots, \alpha\lambda_n$.*
2. *$P + I$ has eigenvectors $v^{(1)}, \cdots v^{(n)}$ with corresponding eigenvalues $\lambda_1 + 1, \cdots, \lambda_n + 1$.*
3. *P^k has eigenvectors $v^{(1)}, \cdots v^{(n)}$ with corresponding eigenvalues $\lambda_1^k, \cdots, \lambda_n^k$.*
4. *If P is stochastic $\implies |\lambda_i| \leq 1$ for all $i = 1, \cdots, n$*

Proof

1. *We will show that each eigenvector of P is also an eigenvector of αP:*

$$vP = \lambda v \iff v \cdot \alpha P = \lambda \cdot \alpha \cdot v$$

   Which means that this is still an eigenvector, but now the corresponding eigenvalue is $\alpha\lambda$

2. *Let v be an eigenvector of P. Then: $v(P + I) = vP + vI = \lambda v + v = (\lambda + 1)v$. This shows that we get the same eigenvectors, but now with eigenvalue $\lambda + 1$.*

3. *We show this inductively. Note that:*

$$vP^k = (vP)P^{k-1} = \lambda v P^{k-1}$$

   If we repeat this grouping $k - 1$ more times, we end up getting:

$$\lambda v P^{k-1} = \cdots = \lambda^k v$$

   Which shows that v is still an eigenvector, now with eigenvalue λ^k

4. *For each eigenvector $v^{(i)}$, we define a set I containing all of the indices of $v^{(i)}$ that contain a positive element, so $I = \{j | v_j^{(i)} > 0\}$. Then, we have that:*

$$\begin{aligned}
\lambda_i \sum_{j \in I} v_j^{(i)} &= \sum_{j \in I} \sum_{k=1}^{n} v_k^{(i)} P_{kj} && \text{[given that the inner sum computes the j-th entry of } v^{(i)}P\text{]} \\
&\leq \sum_{j,k | j,k \in I} v_k^{(i)} P_{kj} && \text{[since entries of } v^{(i)} \text{ not in } I \text{ are} \leq 0 \text{ and } P_{kj} \geq 0\text{]} \\
&\leq \sum_{k \in I} v_k^{(i)} \sum_{j \in I} P_{kj} && \text{[By reorganizing the sum]} \\
&\leq \sum_{k \in I} v_k^{(i)} && \text{[Since } P \text{ is stochastic, } \textstyle\sum_{j \in I} P_{kj} \leq 1\text{]}
\end{aligned}$$

   By dividing on both sides by $\sum_{j \in I} v_j^{(i)}$, this gives $\lambda_i \leq 1$, which is what we wanted to show.

■

Remark *Note that results (1) and (2) are relevant to random walks given that if we have a transition matrix P and we add self-loops at each state such that we stay at the same node with probability $\frac{1}{2}$ and transition according to P with probability $\frac{1}{2}$, then the new transition matrix is $\frac{P+I}{2}$ and we know its eigenvalues are $\frac{\lambda_i+1}{2}$ for each i.*

Finally, we recall another important linear algebra that we will use when proving some results about MCs.

Theorem 12 *If $v^{(1)}, \cdot, v^{(n)}$ is an orthonormal basis, then any vector w is expressible as a linear combination of $v^{(i)}$'s:*

$$w = \sum_{i=1}^{n} \alpha_i v^{(i)}$$

And the L_2-norm of w is $||w||_2 = \sqrt{\sum_i \alpha_i^2}$.

Proof The first fact is due to the fact that any basis spans its corresponding vector space. We now show that the L_2-norm of w is what was stated:

$$\begin{aligned}
||w||_2 &= \sqrt{w \cdot w} && (1)\\
&= \sqrt{\left(\sum_{i=1}^{n} \alpha_i v^{(i)}\right) \cdot \left(\sum_{i=1}^{n} \alpha_i v^{(i)}\right)} && \text{[By expressing } w \text{ as a linear combination]} \quad (2)\\
&= \sqrt{\sum_{i,j} \alpha_i \cdot \alpha_j \cdot v^{(i)} \cdot v^{(j)}} && \text{[By expanding]} \quad (3)\\
&= \sqrt{\sum_{i} \alpha_i^2} && \text{[Since } v_i \cdot v_j \text{ is 1 when } i = j \text{ and 0 when } i \neq j] \quad (4)
\end{aligned}$$

Which is wanted to show. ■

3 Linear Algebra and Random Walks

Recall that in Lecture 10 we defined a stationary distribution of a MC as follows:

Definition 13 (*Stationary Distribution*) *A probability distribution Π is a stationary distribution of a MC with transition matrix P if it holds that $\Pi P = P$*

Note that this means we can think of stationary distributions as eigenvectors of the transition matrix with eigenvalue 1. Then, we also had the following theorem:

Theorem 14 *If P is ergodic (meaning that the MC is aperiodic and irreducible) then there is a unique stationary distribution Π.*

Now, we will use our linear algebra tools to introduce and prove results about how many steps it takes an MC to reach its stationary distribution.

3.1 Mixing Times

At a high-level, the mixing time of a MC corresponds to the amount of steps it takes to approach the stationary distribution. We can define this more formally by giving a concrete notion of what it means to "approach" the stationary distribution.

Definition 15 (*Mixing Time*) *For any $\varepsilon > 0$, we define the Mixing Time $T(\varepsilon)$ of a Markov Chain A with stationary distribution Π as the minimum t such that for all initial distributions $\Pi^{(0)}$:*

$$||\Pi - \Pi^{(0)} A^t||_1 < \varepsilon$$

Which means that after t steps, regardless of the initial distribution, the L_1-distance of the current distribution and the stationary distribution is less than ε.

Now, we are particularly interested in MCs where the mixing time is bounded, so that we are guaranteed to reach the stationary distribution efficiently. We can also define this formally as follows.

Definition 16 (*Rapidly Mixing MC*) *A Markov Chain A with n states is rapidly mixing if:*

$$T(\varepsilon) = \text{poly}(\log n, \log(1/\varepsilon))$$

Example 17 *Random walks over the following graphs are rapidly mixing MCs:*

- *The complete graph K_n. Note that regardless of starting node, there us a uniform probability of visiting any other node on the next step, so $T(\varepsilon) = 1$.*
- *A random graph is also rapidly mixing.*

Remark *Note that if a Markov Chain P has mixing time $T(\varepsilon)$, then the mixing time of $\frac{P+I}{2}$ is at most $2 \cdot T(\varepsilon)$*

Now that we have defined mixing times and seen some examples, we can show how the mixing time of a MC can be analyzed via the linear algebraic properties of its transition matrix. In class, we presented the following theorem and gave a proof sketch.

Theorem 18 *Let P be the transition matrix of a random walk on an undirected, non-k-partite (this can always be achieved by adding self-loops), d-regular connected graph. Let Π_0 be the initial distribution and Π be the stationary distribution $(\frac{1}{n}, \cdots, \frac{1}{n})$.*

Then we have the following bound on the L_2-norm of the difference between the stationary distribution and the distribution after t steps:

$$||\Pi - \Pi_0 P^t||_2 \leq |\lambda_2|^t$$

Which means that if $1 - \lambda_2$ is a constant, then this difference decreases exponentially.

Sketch of Proof In class, we saw the following proof sketch:

1. Since the graph is undirected and the transition matrix is stochastic, we have that P is real and symmetric. From Theorem 10, this tells us that there exists an orthonormal basis of eigenvectors $v^{(1)}, \cdots, v^{(n)}$ with corresponding eigenvalues $1 = \lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$.
2. Then, we can express any vector, particularly Π_0 as a linear combination of the eigenvectors:

$$\Pi_0 = \sum_{i=1}^{n} \alpha_i v^{(i)}$$

.

3. Then, since all of these are eigenvectors:

$$\Pi_0 P^t = \sum_{i=1}^{n} \alpha_i v^{(i)} P^t = \sum_{i=1}^{n} \alpha_i \lambda_i^t v^{(i)} = \alpha_1 \lambda_1 v^{(1)} + \alpha_2 \lambda_2 v^{(2)} + \cdots$$

4. Now, we can show that $\alpha_1 \cdot v_1$ corresponds to the stationary distribution, so we can use the previous result to write:

$$||\Pi_o \cdot P^t - \alpha_1 v_1||_2 = ||\sum_{i=2}^{n} \alpha_i \lambda_i^t v^{(i)}||$$

and use the fact that λ_2 has the greatest absolute value to show that this is upper-bounded by $|\lambda_2|^t$.

■