CS 744: Pseudorandomness Generators

Lecture 3: Some Linear Algebra

January 11, 14, 2015

1 Linear Algebra: Mixing Times of Random Walks on undirected graphs

First, we will define a notion called the hitting time of a random walk. Following Vadhan 12, we will adopt a slightly different notion from the common one.

Definition 1. *The* hitting time *of a random walk on a directed multigraph G is defined as*

$$hit(G) = \max_{i,j \in G} \min\{t \mid Pr\ [a\ random\ walk\ from\ i\ hits\ j\ in\ at\ most\ t\ steps] > \frac{1}{2}\}.$$

It is customary to define the quantity below as the hitting time.

$$\text{ehit}(G) = \max_{i,j \in G} \mathbf{E}\{t \mid i \text{ reaches } j \text{ in } t \text{ steps. }\}.$$

Let us denote by $E[i,j]$, the quantity

$$\mathbf{E}\{t \mid i \text{ reaches } j \text{ in } t \text{ steps. }\}$$

By Markov inequality, it follows that

$$\Pr\{i \text{ does not reach } j \text{ in } 2E[i,j] \text{ steps. }\} \leq \frac{1}{2}.$$

That is, the minimum number of steps before the probability that you hit j starting from i is at least $1/2$, is at most $2E[i,j]$. Thus, $\text{hit}(G)$ is at most 2 $\text{ehit}(G)$.

We will prove the following theorem.

Theorem 2. *For every connected undirected multigraph G with n vertices and maximum degree G, $hit(G) = O(d^2 n^3 \log n)$.*

Let us define an initial probability vector $\pi = (\pi_1, \ldots, \pi_n)$ on the graph G. Let the transition probability matrix for each step be $M_{n\times n}$. Since we are assuming that G is a d-regular undirected graph, then M can be described as

$$M(i,j) = \begin{cases} \frac{1}{d} & \text{if } (i,j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

After taking one step according to the transition matrix M, the distribution on the vertices becomes πM, and inductively, after t steps, the probability on the matrix becomes πM^t. We will show that πM^t converges to the uniform distribution on n vertices. Whenever we talk of convergence, we have to talk about what notion of distance we are using. We will use the ℓ_2 norm — the norm of a vector x is $\|x\|_2 = \sqrt{\langle x, x\rangle}$, where $\langle x, y\rangle$ is the inner product of the vectors x and y.

We will now show that $\|\pi M^t - u\|_2$ decreases with t.

Definition 3.

$$\lambda(G) = \max_{\pi \in P} \frac{\|\pi M - u\|_2}{\|\pi - u\|_2}.$$

Lemma 4. *For an undirected d-regular graph G, $0 \leq \lambda(G) \leq 1$.*

We will prove this later in the course.

We want to calculate hit(G), the hitting time. Now we will slightly change course, and talk about the *mixing time of the random walk*, that is, the time by which the random walk reaches the probability distribution in the limit, in the following lemma. Of course, in the normal course of events, the existence of a limit distribution is often in question. In this case, however, we will establish not only that such a distribution exists, but also that we know what it is - the uniform distribution on n vertices.

If the limit distribution is the uniform distribution on n vertices, then there is a significant probability $(1/\Theta(n))$ that you can start from u and reach any vertex after a sufficiently large number of steps t. In particular, there is a significant probability that you reach v after t steps.

Thus the connection between the mixing time and the hitting time goes through the convenient fact that the limit distribution is u.

The connection between $\lambda(G)$ and the mixing time is as follows.

Lemma 5. *For any undirected d-regular graph G and its random walk matrix M, for any probability distribution π on the set of vertices of G, we have, for any t,*

$$\|\pi M^t - u\|_2 \leq \lambda(G)^t \|\pi - u\|_2 \leq \lambda(G)^t.$$

If $\lambda(G)$ is strictly less than 1, then the random walk mixes to the uniform distribution. The smaller $\lambda(G)$ is, the faster the convergence.

Proof. First, we show $\lambda(G)^t\|\pi - u\|_2 \leq \lambda(G)^t$. It suffices to show $\|\pi - u\|_2 \leq 1$. We have

$$\begin{aligned}\|\pi - u\|_2^2 &= \sum_{i=1}^{n}\left(\pi_i - \frac{1}{n}\right)^2 \\ &= \sum_{i=1}^{n}{\pi_i}^2 - 2\frac{1}{n}\sum_{i=1}^{n}\pi_i + \frac{n}{n^2} \\ &= \sum_{i=1}^{n}{\pi_i}^2 - \frac{1}{n} \leq 1 - \frac{1}{n}.\end{aligned}$$

Now, we show that for every probability vector π, for every $t \in \mathbb{N}$, the claim $\|\pi M^t - u\|_2 \leq \lambda(G)^t\|\pi - u\|_2$ holds, by induction on t. Clearly, for $t = 1$, the claim is true by the definition of $\lambda(G)$. Now, assume that the claim holds for $t = n$.

Consider $\|\pi M^{n+1} - u\|_2$. This can be written as $\|(\pi M^n)M^n - u\|_2$. Since M is a random walk matrix, πM^n is always a probability distribution on the vertices. Let us call this distribution ν.

By the definition of $\lambda(G)$, it follows that $\|\nu M - u\|_2 \leq \lambda(G)\|\nu - u\|_2$. Since the induction hypothesis holds for all probability distributions, we have

$$\|\nu - u\|_2 \;=\; \|\pi M^n - u\|_2 \;\leq\; \lambda(G)^n\|\pi - u\|_2,$$

hence $\|\pi M^{n+1} - u\|_2 \leq \lambda(G)^{n+1}\|\pi - u\|_2$. □

Addendum

The following facts may be useful.

Definition 6. *A matrix $M_{n\times n}$ is called* stochastic *if for every column j in the matrix, $\sum_{i=1}^{n} M_{i,j} = 1$. A stochastic matrix is called* doubly stochastic*, if, in addition, the sum of entries in any row is 1.*

Proposition 7. *If $\pi_{1\times n}$ is a probability distribution, and M is a stochastic matrix, then $\pi_{n\times 1}$ is also a probability distribution.*

This can be proved by direct computation.

Lemma 8. *If $\pi_{1\times n}$ is a probability distribution, then $\pi - u$ is a vector in $\mathbb{R}^n$ which is orthogonal to u.*

Proof. The i^{th} co-ordinate in $\pi - u$ is $\pi_i - \frac{1}{n}$. Then,

$$\begin{aligned}\langle \pi - u, u \rangle &= \sum_{i=1}^{n} (\pi_i - u_i) u_i \\ &= \frac{1}{n} \left[\sum_{i=1}^{n} \pi_i - \sum_{i=1}^{n} \frac{1}{n} \right] \\ &= \frac{1}{n}[1 - 1] = 0.\end{aligned}$$

Hence $(\pi - u) \perp u$. □

Note the following points, though.

1. $\pi - u$ is never a probability vector - $\sum_i (\pi_i - u_i)$ is $\sum_i \pi_i - \sum_i u_i$, which is zero. Hence the sum of the coordinates of $\pi - u$ is not 1.

2. If we take two arbitrary probability vectors π and ν, then $\pi - \nu$ need not be orthogonal to ν. For example, take the 2-dimensional vectors $(1, 0)$ and $(0, 1)$. Then $(1, -1)$ is not orthogonal to $(0, 1)$. In general, for two arbitrary vectors x and y in $\mathbb{R}^n$, $x - y$ need not be orthogonal to y.