ORIE 6334 Spectral Graph Theory	September 22, 2016

Lecture 10

Lecturer: David P. Williamson *Scribe: Julian Sun*

Recall that if we can find a vector $\mathbf{x}$ such that $< \mathbf{x}, e >= 0$ and such that $\mathbf{x}^T \mathscr{L} \mathbf{x} \leq (\lambda_2 + \epsilon)\mathbf{x}^T\mathbf{x}$, then we showed in Trevisan's algorithm that we can find $S \subset V$ such that such that the conductance $\phi(S) \leq \sqrt{2(\lambda_2 + \epsilon)}$. In this lecture we will give an algorithm that finds such a vector.

1 Eigenvector of the Largest Eigenvalue

We begin by finding the eigenvector corresponding to the largest eigenvalue of a symmetric positive semidefinite matrix A.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be orthonormal eigenvectors, with $\lambda_1 \geq \cdots \geq \lambda_n$ being the corresponding eigenvalues of A. Consider the following algorithm.

Algorithm 1: Power Method

```
Pick v_0 by drawing v_0(i) ~ N(0,1) for all i
for j ← 1 to k do
    v_j ← A v_{j-1}
return v_k
```

If $A \in \mathbb{R}^{n \times n}$ has m nonzero entries, then this algorithm runs in $O(k(m+n))$ time.

Let us introduce a preliminary lemma regarding the vector $\mathbf{v}_0$.

Lemma 1 *Let $\mathbf{x}$ be a vector such that $\|\mathbf{x}\| = 1$. Then for $\mathbf{v}$ such that $\mathbf{v}(i) \sim N(0,1)$ for all i,*

$$\Pr\left[|\mathbf{x}^T\mathbf{v}| \geq \frac{1}{2}\right] \geq 2\Phi\left(-\frac{1}{2}\right) \geq 0.6,$$

where $N(0,1)$ is the standard normal distribution, and Φ is the cdf of the standard normal. Also

$$\Pr\left[\|\mathbf{v}\|^2 \leq 2n\right] \geq 1 - e^{-n/6}.$$

With that in mind, we can show that the vector we have constructed in Power Method has Rayleigh ratio reasonably close to that of the eigenvector of the largest eigenvalue.

---

[0]This lecture is derived from Trevisan Chapter 4, `https://people.eecs.berkeley.edu/~luca/books/expanders.pdf` and Vishnoi, Chapter 8, `http://research.microsoft.com/en-us/um/people/nvishno/site/Lxb-Web.pdf`.

Lemma 2 *Let* $\mathbf{v}$ *be such that* $\mathbf{v}(i) \sim N(0,1)$ *for all* i, $|\mathbf{v}^T\mathbf{x}_1| \geq \frac{1}{2}$, $A \succeq 0$. *Then for* $k > 0$, $\epsilon > 0$, $\mathbf{y} = A^k\mathbf{v}$,

$$\frac{\mathbf{y}^T A\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \geq \lambda_1(1-\epsilon)\frac{1}{1+4\|\mathbf{v}\|^2(1-\epsilon)^{2k}}.$$

If we can show that the lemma is true, then for $k = O(\frac{\log n}{\epsilon})$, $\|\mathbf{v}\|^2 \leq 2n$, we have

$$\frac{\mathbf{y}^T A\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \geq \lambda_1(1-\epsilon)\frac{1}{1+8n(1-\epsilon)^{2k}} \geq \lambda_1(1-\epsilon)\frac{1}{1+\frac{8}{n}} \geq \lambda_1(1-2\epsilon)$$

for n sufficiently large.

Proof: Let $\mathbf{v} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n$, where $\alpha_i = \langle \mathbf{v}, \mathbf{x}_i\rangle$. Note that this implies

$$\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v} = (\alpha_1\mathbf{x}_1 + \cdots \alpha_n\mathbf{x}_n)^T(\alpha_1\mathbf{x}_1 + \cdots \alpha_n\mathbf{x}_n) = \sum_{i=1}^n \alpha_i^2.$$

We assumed that $|\alpha_1| \geq \frac{1}{2}$. Recall that the eigenvectors of A^k are still $\mathbf{x}_1, \ldots, \mathbf{x}_n$, with eigenvalues being $\lambda_1^k, \ldots, \lambda_n^k$. Since $A \succeq 0$, it follows that $\lambda_1, \ldots, \lambda_n \geq 0$. Hence $\lambda_1^k \geq \cdots \geq \lambda_n^k \geq 0$. So

$$\mathbf{y} = A^k\mathbf{v} = A^k(\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n) = \alpha_1\lambda_1^k\mathbf{x}_1 + \cdots + \alpha_n\lambda_n^k\mathbf{x}_n,$$

and hence

$$\mathbf{y}^T A\mathbf{y} = \alpha_1^2\lambda_1^{2k+1} + \cdots + \alpha_n^2\lambda_n^{2k+1}.$$

Similarly,

$$\mathbf{y}^T\mathbf{y} = \alpha_1^2\lambda_1^{2k} + \cdots + \alpha_n^2\lambda_n^{2k}.$$

Let l be the largest index such that $\lambda_l \geq (1-\epsilon)\lambda_1$. Then

$$\mathbf{y}^T A\mathbf{y} \geq \sum_{i=1}^l \alpha_i^2\lambda_i^{2k+1} \geq (1-\epsilon)\lambda_1\sum_{i=1}^l \alpha_i^2\lambda_i^{2k}.$$

We can also write

$$\mathbf{y}^T\mathbf{y} = \sum_{i=1}^l \alpha_i^2\lambda_i^{2k} + \sum_{i=l+1}^n \alpha_i^2\lambda_i^{2k},$$

where by definition of l and by the fact that $|\alpha_1| \geq \frac{1}{2}$,

$$\begin{aligned}
\sum_{i=l+1}^n \alpha_i^2\lambda_i^{2k} &\leq \lambda_1^{2k}(1-\epsilon)^{2k}\sum_{i=l+1}^n \alpha_i^2 \\
&\leq \lambda_1^{2k}(1-\epsilon)^{2k}\|\mathbf{v}\|^2 \\
&\leq 4\alpha_1^2\lambda_1^{2k}(1-\epsilon)^{2k}\|\mathbf{v}\|^2 \\
&\leq 4(1-\epsilon)^{2k}\|\mathbf{v}\|^2\sum_{i=1}^l \alpha_i^2\lambda_i^{2k}.
\end{aligned}$$

Hence it follows that

$$\mathbf{y}^T\mathbf{y} \leq (1 + 4(1-\epsilon)^{2k}\|\mathbf{v}\|^2) \sum_{i=1}^{l} \alpha_i^2 \lambda_i^{2k},$$

and therefore

$$\frac{\mathbf{y}^T A\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \geq \frac{(1-\epsilon)\lambda_1 \sum_{i=1}^{l} \alpha_i^2 \lambda_i^{2k}}{(1 + 4(1-\epsilon)^{2k}\|\mathbf{v}\|^2) \sum_{i=1}^{l} \alpha_i^2 \lambda_i^{2k}} = \lambda_1(1-\epsilon)\frac{1}{1 + 4\|\mathbf{v}\|^2(1-\epsilon)^{2k}},$$

as desired. □

2 Eigenvector of the Second Largest Eigenvalue

Observe that if $\mathbf{x}_1$ the eigenvector associated with the largest eigenvalue of the matrix A, then recall that

$$\lambda_2 = \max_{\mathbf{x}:\langle \mathbf{x},\mathbf{x}_1\rangle = 0} \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

So if we want to find a vector with Rayleigh ratio close to that of $\mathbf{x}_2$, we can just make sure that the vector $\mathbf{x}$ satisfies the condition $\langle \mathbf{x},\mathbf{x}_1\rangle = 0$ and apply the Power Method.

Algorithm 2: Power2

```
Pick v by drawing v(i) ~ N(0,1) for all i
v_0 ← v − ⟨v, x_1⟩x_1
for j ← 1 to k do
    v_j ← A v_{j−1}
return v_k
```

In this case we have $\mathbf{v}_0 = \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n$, and thus $\|\mathbf{v}_0\|^2 \leq \|\mathbf{v}\|^2 \leq 2n$ with high probability. Also,

$$\mathbf{v}_k = \alpha_2\lambda_2^k\mathbf{x}_2 + \cdots + \alpha_n\lambda_n^k\mathbf{x}_n.$$

So if $\mathbf{y} = A^k\mathbf{v}_0$, and $|\alpha_2| \geq \frac{1}{2}$ (still with probability ≥ 0.6), then

$$\langle \mathbf{y},\mathbf{x}_1\rangle = 0,$$

and

$$\frac{\mathbf{y}^T A\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \geq \lambda_2(1-\epsilon)\frac{1}{1 + 4\|\mathbf{v}_0\|^2(1-\epsilon)^{2k}},$$

with proof identical to that in the previous section.

3 Eigenvector of the Second Smallest Eigenvalue of $\mathscr{L}$

Now, how we compute the second smallest eigenvalue of $\mathscr{L}$, as we originally set out to do? We can alternatively look at the second largest eigenvalue of $\mathscr{M} = 2I - \mathscr{L}$.

In this case, if $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ are the eigenvalues of $\mathscr{L}$, then $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ are the eigenvalues of $\mathscr{M}$, with $\mu_i = 2 - \lambda_i$. In this case since $\mu_n \geq 0$, it follows that $\mathscr{M} \succeq 0$.

Now, the eigenvector for the largest eigenvalue of $\mathscr{M}$ (and for the smallest eigenvalue of $\mathscr{L}$) is $\mathbf{x}_1 = D^{1/2}\mathbf{e}$. Then we can run Power2 on $\mathscr{M}$ to find $\mathbf{y}$ such that $\langle \mathbf{y}, \mathbf{x}_1 \rangle = 0$, and

$$\frac{\mathbf{y}^T \mathscr{M} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq (1-\epsilon)\mu_2 = (1-\epsilon)(2-\lambda_2).$$

We also have that

$$\mathbf{y}^T \mathscr{M} \mathbf{y} = \mathbf{y}^T (2I - \mathscr{L}) \mathbf{y} = 2\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathscr{L} \mathbf{y}.$$

Hence

$$\mathbf{y}^T \mathscr{L} \mathbf{y} \leq (\lambda_2 + 2\epsilon - \epsilon\lambda_2)\mathbf{y}^T \mathbf{y} \leq (2\epsilon)\mathbf{y}^T \mathbf{y},$$

and so

$$\frac{\mathbf{y}^T \mathscr{L} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \leq \lambda_2 + 2\epsilon.$$

Note that if we want $\frac{\mathbf{y}^T \mathscr{L} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \leq 2\lambda_2$, we'll need $\epsilon = \frac{\lambda_2}{2}$. Recall that the running time of Power2 is $O((m \log n)/\epsilon)$; thus if λ_2 is very small (for instance, $\lambda_2 = O(\frac{1}{n})$), the running time could be much higher than nearly linear in m. So let's explore other methods!

Recall the pseudo inverse $\mathscr{L}^\dagger$ of $\mathscr{L}$:

$$\mathscr{L}^\dagger = \sum_{i:\lambda_i \neq 0} \frac{1}{\lambda_i} \mathbf{x}_i \mathbf{x}_i^T.$$

Note that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are still the eigenvectors of $\mathscr{L}^\dagger$ since

$$\left(\sum \frac{1}{\lambda_i} \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{x}_j = \frac{1}{\lambda_j} \mathbf{x}_j$$

if $j \neq 1$, and

$$\left(\sum \frac{1}{\lambda_i} \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{x}_j = 0\mathbf{x}_j$$

if $j = 1$.

So here's an idea: we can run Power Method on $\mathscr{L}^\dagger$! Since $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$, it follows that $\frac{1}{\lambda_2}$ is the largest eigenvalue of $\mathscr{L}^\dagger$.

But here's a problem: it's not cheap to compute pseudoinverses. It takes $O(n^2)$ to write it down. ☹

However, we have a corresponding solution: notice that in the Power Method, we never need to actually write down the pseudoinverse. All we have to do is to compute $\mathscr{L}^\dagger \mathbf{v}$ for some $\mathbf{v}$ such that $\langle \mathbf{v}, \mathbf{x}_1 \rangle = 0$. ☺ We can do this approximately in $O(m \log^c n / \epsilon^{c'})$ time, for some constants c, c', as we will see in an upcoming lecture. And so here is our algorithm.

Algorithm 3: Power Method

Pick $\mathbf{v}$ by drawing $\mathbf{v}(i) \sim N(0,1)$ for all i
$\mathbf{v}_0 \leftarrow \mathbf{v} - \langle \mathbf{v}, \mathbf{x}_1 \rangle \mathbf{x}_1$, $\mathbf{x}_1 = D^{1/2}\mathbf{e}$
for $j \leftarrow 1$ **to** k **do**
  $\mathbf{v}_j \leftarrow \mathscr{L}^\dagger \mathbf{v}_{j-1}$
return $\mathbf{v}_k$

As before, $\mathbf{v}_0 = \alpha_2 \mathbf{x}_2 + \cdots + \alpha_n \mathbf{x}_n$, $\|\mathbf{v}_0\|^2 \leq 2n$ with high probability, and $|\alpha_2| \geq \frac{1}{2}$ with constant probability. Then

$$\mathbf{y} = (\mathscr{L}^\dagger)^k \mathbf{v}_0 = \sum_{i=2}^{n} \alpha_i \frac{1}{\lambda_i^k} \mathbf{x}_i.$$

Again we pick l to be the largest index such that $\lambda_l \leq (1+\epsilon)\lambda_2$, and in this case

$$\begin{aligned}
\mathbf{y}^T \mathscr{L} \mathbf{y} &= \left(\sum_{i=2}^{n} \alpha_i \frac{1}{\lambda_i^k} \mathbf{x}_i \right)^T \left(\sum_{i=2}^{n} \alpha_i \frac{1}{\lambda_i^{k-1}} \mathbf{x}_i \right) \\
&= \sum_{i=2}^{n} \alpha_i^2 \frac{1}{\lambda_i^{2k-1}} \\
&\leq \sum_{i=2}^{l} \alpha_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1}\lambda_2^{2k-1}} \sum_{i=l+1}^{n} \alpha_i^2 \\
&\leq \sum_{i=2}^{l} \alpha_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1}\lambda_2^{2k-1}} 2n \\
&\leq \sum_{i=2}^{l} \alpha_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{8n\alpha_2^2}{(1+\epsilon)^{2k-1}\lambda_2^{2k-1}} \\
&\leq \left(1 + \frac{8n}{(1+\epsilon)^{2k-1}}\right) \sum_{i=2}^{l} \alpha_i^2 \frac{1}{\lambda_i^{2k-1}}.
\end{aligned}$$

Also similarly to before, we can get that

$$\mathbf{y}^T \mathbf{y} = \sum_{i=2}^{n} \alpha_i^2 \frac{1}{\lambda_i^{2k}} \geq \sum_{i=2}^{l} \alpha_i^2 \frac{1}{\lambda_i^{2k}} \geq \frac{1}{(1+\epsilon)\lambda_2} \sum_{i=2}^{l} \alpha_i^2 \frac{1}{\lambda_i^{2k-1}}.$$

Therefore
$$\frac{\mathbf{y}^T \mathscr{L} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \leq \lambda_2(1+\epsilon)\left(1 + \frac{8n}{(1+\epsilon)^{2k-1}}\right) \leq (1+2\epsilon)\lambda_2$$
for $k = O(\log n/\epsilon)$ and n sufficiently large. In this case we can find λ_2 and $\mathbf{y}$ approximately, in time $O(m \log^{c+1} n/\epsilon^{c'+1})$.

We end with an open question by Vishnoi: Can we find this vector with Rayleigh ratio close to λ_2 in nearly linear time without having to multiply by $\mathscr{L}^{\dagger}$?