# Algorithms for Big Data (XI)

Chihao Zhang

Shanghai Jiao Tong University

Nov. 29, 2019

# Review

Last week, we start the topic on faster algorithms for numerical linear algebra.

We learnt an almost-linear algorithm to approximate Matrix Multiplication.

Next, we introduced spectral graph theory.

We will see how to design almost-linear algorithms for graph problems using spectral tools.

# Graph as a Matrix

Let $G = (V, E)$ be an undirected graph on $n$ vertices without self-loops and parallel edges.

Its adjacency matrix $A(G) = (a_{ij})_{i,j \in [n]}$ is symmetric.

We are interested in the eigenvalues and eigenvectors of $A$...

For symmetric matrices, the spectrum is well-structured.

# Spectral Decomposition Theorem

**Theorem**

An $n \times n$ symmetric matrix $A$ has $n$ real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthonormal. Moreover, it holds that

$$A = V \Lambda V^{\mathsf{T}},$$

where $V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots \mathbf{v}_n \end{bmatrix}$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

# Graph Laplacian for Regular Graphs

In the following, we assume the graph $G$ is d-regular. We will see how to generalize to irregular graphs later today.

Sometimes it is convenient to shift and scale the eigenvalues of $A$.

The Laplacian of $G$ is $L \triangleq dI - A$.

The normalized Laplacian of $G$ is $N \triangleq \frac{L}{d} = I - \frac{1}{d}A$.

We already verified the following identity:

$$\forall \mathbf{x} \in \mathbb{R}^{[n]} : \mathbf{x}^T L \mathbf{x} = \sum_{\{u,v\} \in E} (x(u) - x(v))^2 .$$

# Rayleigh Quotient

Let $\langle \cdot, \cdot \rangle$ denote the ordinary inner product of two vectors, i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathsf{T}} \mathbf{y}$.

Let $M \in \mathbb{R}^{n \times n}$ be a matrix. The Rayleigh quotient is

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad R_M(\mathbf{x}) = \frac{\langle \mathbf{x}, M\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

It is clear that if $\lambda$ is an eigenvalue of $M$ with eigenvector $\mathbf{v}$, then

$$R_M(\mathbf{v}) = \frac{\langle \mathbf{v}, M\mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} = \frac{\langle \mathbf{v}, \lambda\mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} = \lambda.$$

# Courant-Fischer Theorem

Let $M$ be a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be corresponding eigenvectors.

**Theorem (Courant-Fischer Theorem)**

$$\lambda_k = \min_{\text{k-dim } S \subseteq \mathbb{R}^n} \max_{\mathbf{x} \in S \setminus \{\mathbf{0}\}} R_M(\mathbf{x})$$

**Corollary**

$$\lambda_1 = \min_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} R_M(\mathbf{x}), \quad \lambda_n = \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} R_M(\mathbf{x}).$$

# Proof

We first show that

$$\min_{\text{k-dim } S \subseteq \mathbb{R}^n} \max_{\mathbf{x} \in S \setminus \{\mathbf{0}\}} R_M(\mathbf{x}) \leq \lambda_k.$$

We construct a k-dim space $S$ such that any $\mathbf{x} \in S \setminus \{\mathbf{0}\}$ satisfies $R_M(\mathbf{x}) \leq \lambda_k$.

$S = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ satisfies our need.

We then prove that any k-dim $S \subseteq \mathbb{R}^n$, there exists some $\mathbf{x} \in S \setminus \{\mathbf{0}\}$ satisfying $R_M(\mathbf{x}) \geq \lambda k$.

Choose nonzero $\mathbf{x} \in X \cap \text{span}(\mathbf{v}_k, \dots, \mathbf{v}_n)$.

# Eigenvalues for Laplacians

Recall $L$ is the Laplacian and $N$ is the normalized Laplacian.

**Theorem**

Assume $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $N$, then

- $\lambda_1 = 0$;
- $\lambda_n \leq 2$ and $\lambda_n = 2$ if and only if one of components of $G$ is bipartite;
- $\lambda_k = 0$ if and only if $G$ has at least $k$ components.

**Theorem**

$$\lambda_k = \max_{\substack{\mathbf{x} \perp span(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}) \\ \mathbf{x} \neq 0}} R_M(\mathbf{x})$$

# Laplacians for General Graphs

For a not necessarily simple graph $G$ with adjacency matrix $A$, define its Laplacian as

$$L = D - A$$

where $D = \mathrm{diag}(\deg(v_1), \ldots, \deg(v_n))$.

The normalized Laplacian is

$$N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}.$$

Both $L$ and $N$ are positive semi-definite.

# Rayleigh Quotient (for General normalized Laplacian)

Let $N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ be a normalized Laplacian, then

$$R_N(\mathbf{x}) = \frac{\langle \mathbf{x}, \left(D^{-\frac{1}{2}} L D^{-\frac{1}{2}}\right) \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\langle D^{-\frac{1}{2}} \mathbf{x}, L D^{-\frac{1}{2}} \mathbf{x} \rangle}{\langle D^{-\frac{1}{2}} \mathbf{x}, D D^{-\frac{1}{2}} \mathbf{x} \rangle} = \frac{\langle \mathbf{y}, L\mathbf{y} \rangle}{\langle \mathbf{y}, D\mathbf{y} \rangle},$$

where $\mathbf{y} = D^{-\frac{1}{2}} \mathbf{x}$.

It is an exercise to prove the theorem in the previous slide for general graphs.

It is useful to view $L$ as an operator, namely

$$L\mathbf{x}(i) = \deg(i) x(i) - \sum_{\{i,j\} \in E} x(j).$$

# Examples

The Laplacian of complete graph $K_n$: $E = \binom{[n]}{2}$.

- $\lambda_1 = 0, \lambda_2 = \lambda_3 = \cdots = \lambda_n = n$;
- $\mathbf{v}_1 = \mathbf{1}, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ can be a basis of $\mathrm{span}(\mathbf{1})^\perp$.

The Laplacian of a star $S_n$: $E = \{\{1, j\} : 2 \le j \le n\}$.

- $\lambda_0 = 0, \lambda_1 = \lambda_2 = \cdots = \lambda_{n-1} = 1, \lambda_n = n$;
- $\mathbf{v}_1 = \mathbf{1}, \mathbf{v}_i = \mathbf{e}_i - \mathbf{e}_{i+1}$ for $2 \le i < n$, $\mathbf{v}_n = (n-1)\mathbf{e}_1 - \sum_{2 \le j \le n} \mathbf{e}_j$.

# Random Walk (on Regular graphs)

Let $G = (V, E)$ be a d-regular graph.

One can naturally define a random walk: standing at vertex $i$, move to one of its randomly chosen neighbour $j$.

If we use denote $P = (p_{ij})$ such that $p_{ij} = \begin{cases} 1/d & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$, then

$$\mathbf{x}_{t+1} = P^T \mathbf{x}_t,$$

where $\mathbf{x}_t \in [0, 1]^{[n]}$ is the distribution of the location of you at time $t$.

We can make $P$ lazy by defining $\tilde{P} = \frac{1}{2}(I + P)$.

What is the spectrum of $\tilde{P}$?

$\tilde{P}$ is simply $\frac{1}{2}\left(I + \frac{A}{d}\right)$, so it satisfies

$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n = 1.$$

A distribution $\mathbf{x}$ is called stable if $\mathbf{x} = \tilde{P}^T \mathbf{x}$.

**Theorem**

If $G$ is connected, then $\mathbf{x}_t$ converges to a stable distribution whatever the initial one is.

Proof by Spectral Decomposition Theorem.