ORIE 6334 Bridging Continuous and Discrete Optimization September 18, 2019

Lecture 5

Lecturer: David P. Williamson *Scribe: Haici Tan*

1 Definitions and Eigenvalue Basics

Let $x = a + ib \in \mathbb{C}$ be a complex number, then we define $\overline{x} = a - ib$ to be its *conjugate*. For a matrix of complex numbers $A = (x_{ij}) \in \mathbb{C}^{m\times n}$, we define $A^* = (z_{ij}) \in \mathbb{C}^{n\times m}$ where $z_{ij} = \overline{x}_{ji}$ for all $i \le n$ and $j \le m$. A^* is then called the *conjugate transpose* of A.

For $x, y \in \mathbb{C}^n$, their inner product is defined as

$$\langle x, y\rangle \equiv x^* y = \sum_{i=1}^{n} \overline{x}_i y_i$$

For $A \in \mathbb{C}^{n\times n}$, $\lambda \in \mathbb{C}$ and $x \neq 0 \in \mathbb{C}^n$, if $Ax = \lambda x$, then x is an *eigenvector* of A and λ is the associated *eigenvalue.*

Note that $Ax = \lambda x$ if and only if $Ax - \lambda I x = 0$, which is equivalent to $(\lambda I - A)x = 0$. For $x \neq 0$, we have

$$\det(\lambda I - A) = 0.$$

$\det(\lambda I - A)$ for fixed A is a polynomial of degree n in λ. We call it the *characteristic polynomial* of A. There are exactly n solutions to $\det(\lambda I - A) = 0$ (with multiplicity). Each solution is an eigenvalue.

A matrix A is *Hermitian* if $A = A^*$. If $A \in \mathbb{R}^{n\times n}$, then A is symmetric. ($A = A^T$)

Hermitian matrices have the following two nice properties.

Lemma 1 *If A is Hermitian, then all its eigenvalues are real.*

Proof: Suppose λ and $x \neq 0$ satisfy $Ax = \lambda x$. Then,

$$\begin{aligned}\langle Ax, x\rangle = (Ax)^* x &= x^* A^* x \\ &= x^* A x \\ &= \langle x, Ax\rangle.\end{aligned}$$

Also, we have

$$\langle Ax, x\rangle = \langle \lambda x, x\rangle = \overline{\lambda}\langle x, x\rangle = \overline{\lambda}\|x\|^2$$

and

$$\langle x, Ax\rangle = \langle x, \lambda x\rangle = \lambda\langle x, x\rangle = \lambda\|x\|^2.$$

Since $x \neq 0$, $\lambda = \overline{\lambda}$, which means that λ is real. □

[0]This lecture was drawn from Trevisan, *Lecture Notes on Expansion, Sparsest Cut, and Spectral Graph Theory*, Chapter 1, and Lau's 2015 lecture notes, Lecture 1:`https://cs.uwaterloo.ca/~lapchi/cs798/notes/L01.pdf`.

Lemma 2 *Let A be a Hermitian matrix. Suppose x and y are eigenvectors of A with different eigenvalues λ and λ' ($\lambda \neq \lambda'$). Then, x and y are orthogonal.*

Proof: Since A is Hermitian, we have

$$\langle Ax, y\rangle = (Ax)^* y = x^* A^* y = x^* A y = \langle x, Ay\rangle$$

By Lemma 1, λ and λ' are real, so

$$\langle Ax, y\rangle = \langle \lambda x, y\rangle = \lambda\langle x, y\rangle$$

and

$$\langle x, Ay\rangle = \langle x, \lambda' y\rangle = \lambda'\langle x, y\rangle.$$

Then,

$$(\lambda - \lambda')\langle x, y\rangle = 0.$$

Because $\lambda \neq \lambda'$, x and y must be orthogonal. $\square$

2 Rayleigh Quotients and the Spectral Theorem

For the rest of the class, we are going to focus on real symmetric matrices. We assume that all matrices A that appear in this section are symmetric and $n \times n$. Our goal is to prove the following theorem, which will be extremely useful for the rest of the semester.

Theorem 3 *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be its eigenvalues (all real by Lemma 1) and $x_1, x_2, \cdots, x_n$ be orthonormal vectors (e.g. $\|x_i\|^2 = 1$, $\langle x_i, x_j\rangle = 0 \;\; \forall i \neq j$) such that $Ax_i = \lambda_i x_i$ for $i = 1, 2, \cdots, n$. Then, for all $0 \leq k \leq n-1$,*

$$\lambda_{k+1} = \min_{x \in \mathbb{R}^n : x \perp span(x_1, \cdots, x_k)} \frac{x^T A x}{x^T x}$$

and any minimizer is the associated eigenvector.

The expression $\frac{x^T A x}{x^T x}$ is called the *Rayleigh quotient*. One reason this theorem is very useful is that it allows us to get an upper bound on λ_{k+1}: the Rayleigh quotient of x for any $x \perp \text{span}(x_1, \cdots, x_k)$ yields an upper bound. We will be using this technique for bounding eigenvalues *ad nauseum*.

In order to prove the theorem, we first prove the following lemma.

Lemma 4 *Let $A \in \mathbb{R}^{n\times n}$ be symmetric and $k \leq n-1$. Let $x_1, \cdots, x_k$ be orthogonal eigenvectors of A. Then there exists an eigenvector x_{k+1} orthogonal to $x_1, \cdots, x_k$.*

Proof: Let V be a $(n-k)$-dimensional subspace of $\mathbb{R}^n$ that contains all $x \in \mathbb{R}^n$ such that $x \perp \text{span}(x_1, \cdots, x_k)$. For any $x \in V$, $Ax \in V$ since for all $i = 1, \cdots, k$,

$$\langle x_i, Ax\rangle = x_i^T A x = (A^T x_i)^T x = (Ax_i)^T x = (\lambda x_i)^T x = \lambda\langle x_i, x\rangle = 0.$$

Let $b_1, \cdots, b_{n-k}$ be an orthonormal basis of V. Define

$$B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_{n-k} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times (n-k)}.$$

For any $z \in \mathbb{R}^{n-k}$, $Bz \in V$ since Bz is a linear combination of vectors in V.

Also, for all $z \in V$,

$$BB^T z = B \begin{bmatrix} b_1^T z \\ b_2^T z \\ \vdots \\ b_{n-k}^T z \end{bmatrix} = \langle b_1, z\rangle b_1 + \cdots + \langle b_{n-k}, z\rangle b_{n-k} = z \tag{1}$$

since B is a orthonormal basis of V.

Let λ be an eigenvalue of $A' = B^T AB \in \mathbb{R}^{(n-k)\times(n-k)}$ with associated eigenvector y. Then,

$$B^T ABy = \lambda y.$$

We know $By \in V$, so $A(By) \in V$. By (1),

$$BB^T(ABy) = ABy.$$

On the other hand,

$$BB^T ABy = B(B^T ABy) = \lambda By,$$

so

$$ABy = \lambda By.$$

Since B is non-singular and $y \neq 0$, $By \neq 0$, so By is an eigenvector of A. Note that By is orthogonal to $x_1, \cdots, x_k$ because $By \in V$. □

An easy corollary of Lemma 4 is the Spectral Theorem.

Corollary 5 *(Spectral Theorem) For a symmetric matrix $A \in \mathbb{R}^{n\times n}$ with (real) eigenvalues $\lambda_1, \cdots, \lambda_n$, there exist orthonormal vectors $x_1, \cdots, x_n$ such that x_i is the eigenvector associated with λ_i.*

Now we can prove our main theorem.

Proof of Theorem 3: Given eigenvalues $\lambda_1, \cdots, \lambda_k$ and associated vectors $x_1, \cdots, x_k$, we can use Lemma 4 repeatedly to find orthonormal eigenvectors $x_{k+1}, \cdots, x_n$. We sort the remaining eigenvalues so that $\lambda_{k+1} \leq \cdots \leq \lambda_n$. Note that

$$\frac{x_{k+1}^T A x_{k+1}}{x_{k+1}^T x_{k+1}} = \frac{\lambda_{k+1}(x_{k+1}^T x_{k+1})}{x_{k+1}^T x_{k+1}} = \lambda_{k+1}.$$

Consider any other feasible solution x. Let V be the subspace containing all $y \in \mathbb{R}^n$ such that $y \perp \text{span}(x_1, x_2, \cdots, x_k)$. Then $x_{k+1}, \cdots, x_n$ is a basis of V. Assume that

$x = \alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n$. We have

$$\begin{aligned}
\frac{x^TAx}{x^Tx} &= \frac{x^T(\alpha_{k+1}\lambda_{k+1}x_{k+1} + \cdots + \alpha_n\lambda_n x_n)}{x^Tx} \\
&= \frac{(\alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n)^T(\alpha_{k+1}\lambda_{k+1}x_{k+1} + \cdots + \alpha_n\lambda_n x_n)}{(\alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n)^T(\alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n)} \\
&= \frac{\alpha_{k+1}^2\lambda_{k+1} + \cdots + \alpha_n^2\lambda_n}{\alpha_{k+1}^2 + \cdots + \alpha_n^2} \\
&\geq \lambda_{k+1}\frac{\alpha_{k+1}^2 + \cdots + \alpha_n^2}{\alpha_{k+1}^2 + \cdots + \alpha_n^2} \\
&= \lambda_{k+1}.
\end{aligned}$$

Hence,

$$\lambda_{k+1} = \min_{x\in\mathbb{R}^n : x\perp \mathrm{span}(x_1,\cdots,x_k)} \frac{x^TAx}{x^Tx}.$$

□

In fact, we can further extend Theorem 3 and reach the following conclusion.

Theorem 6 *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be its eigenvalues and $x_1, x_2, \cdots, x_n$ be associated orthonormal eigenvectors. Then,*

$$\begin{aligned}
\lambda_k &= \min_{x\in\mathbb{R}^n : x\perp span(x_1,\cdots,x_{k-1})} \frac{x^TAx}{x^Tx} \\
&= \min_{x\in\mathbb{R}^n : x\in span(x_k,\cdots,x_n)} \frac{x^TAx}{x^Tx} \\
&= \max_{x\in\mathbb{R}^n : x\perp span(x_{k+1},\cdots,x_n)} \frac{x^TAx}{x^Tx} \\
&= \max_{x\in\mathbb{R}^n : x\in span(x_1,\cdots,x_k)} \frac{x^TAx}{x^Tx}.
\end{aligned}$$

We omit the proof of Theorem 6 since it is similar to the proof of Theorem 3.

Some special cases of the theorem are

$$\lambda_n = \max_{x\in\mathbb{R}^n} \frac{x^TAx}{x^Tx}$$

and

$$\lambda_1 = \min_{x\in\mathbb{R}^n} \frac{x^TAx}{x^Tx}.$$

3 Inverse and Pseudo-inverse

Since $x_1, \cdots, x_n$ are orthonormal, for any $x \in \mathbb{R}^n$, we can write x as

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n.$$

Then,
$$\langle x, x_i \rangle = \langle \alpha_1 x_1 + \cdots + \alpha_n x_n, x_i \rangle = \alpha_i \langle x_i, x_i \rangle = \alpha_i.$$
Therefore,
$$\begin{aligned} x &= \langle x, x_1 \rangle x_1 + \cdots \langle x, x_n \rangle x_n \\ &= x_1(x_1^T x) + \cdots + x_n(x_n^T x) \\ &= (x_1 x_1^T + \cdots + x_n x_n^T) x \end{aligned}$$
for all $x \in \mathbb{R}^n$. Hence,
$$x_1 x_1^T + \cdots + x_n x_n^T = I. \tag{2}$$
By (2),
$$Ax = AIx = A(x_1 x_1^T + \cdots + x_n x_n^T)x = (\lambda_1 x_1 x_1^T + \cdots + \lambda_n x_n x_n^T)x,$$
so
$$A = \lambda_1 x_1 x_1^T + \cdots + \lambda_n x_n x_n^T. \tag{3}$$
We know that A^{-1} exists iff all eigenvalues of A are non-zero. Also,
$$(\lambda_1 x_1 x_1^T + \cdots + \lambda_n x_n x_n^T)(\frac{1}{\lambda_1} x_1 x_1^T + \cdots + \frac{1}{\lambda_n} x_n x_n^T) = x_1 x_1^T + \cdots x_n x_n^T = I.$$
Thus,
$$A^{-1} = \frac{1}{\lambda_1} x_1 x_1^T + \cdots + \frac{1}{\lambda_n} x_n x_n^T.$$
When A is singular, we define the *pseudo-inverse* of A analogously:
$$A^\dagger \equiv \sum_{i:\lambda_i \neq 0} \frac{1}{\lambda_i} x_i x_i^T.$$

One of the reasons that spectral graph theory has become an intense area of study in theoretical computer science in the last few years is that researchers (starting with Spielman and Teng) have shown how to compute $A^\dagger b$ quickly for some cases of A and b. This has led to further research on how to solve this product quickly plus additional research on what can be done with a quick solver of this type. We will hear more about this later in the term.