Mathematical Tools in Computer Science, Winter 2006

Lecture 3

Lecturer: Nati Linial *Scribe: Noa Eidelstein*

The structure of this lecture is as follows. We start with Rayleigh-Ritz theorem, and the spectrum of a symmetric (Hermitian) matrix. Then we turn to non-negative matrices and prove the Perron-Frobenius theorem. We deduce an important corollary about stochastic matrices and disscuss Markov chains.

1 Variational Approach to Eigenvalues

Definitions:

1. The spectrum of a matrix is the collection of its eigenvalues.
2. A spectral decomposition of a vector is a representation of this vector as a linear combination of eigenvectors.

Facts:

- All the eigenvalues of a real symmetric (Hermitian) matrix are real.
- Two eigenvectors of a real symmetric (Hermitian) matrix that correspond to different eigenvalues are orthogonal.

  Proof: assume $Ax = \lambda x, Ay = \mu y$, and look at $y^t Ax$: $y^t Ax =< y, Ax >= \lambda < y, x >$. On the other hand, $y^t Ax =< y^t A, x >= \mu < y, x >$. Since $\lambda \neq \mu$, we get $< y, x >= 0$.
- A real symmetric matrix A can be written in the form $A = V \Lambda V^t$ where V is an orthogonal matrix (whose coloumns are the eigenvectors of A), and Λ is a diagonal matrix (whose diagonal elements are the eigenvalues of A).

Theorem 1 (Rayleigh-Ritz) *Let $A \in M_n(\mathbb{R})$ be a symmetric real matrix, and let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be its eigenvalues (in descending order). Then*

$$\lambda_1 = \max_{x \in \mathbb{R}^n} \frac{x^t Ax}{||x||} = \max_{||x||=1} x^t Ax,$$

and the eigenvector corresponding to λ_1 is the vector v_1 on which the maximum is attained. Furthermore, let $v_1, \ldots, v_k$ denote the first k eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_k$, and let $V_k^\perp$ denote the linear subspace orthogonal to their span (i.e., $V_k^\perp = \{v \mid v \perp v_1, \ldots, v_k\}$). Then

$$\lambda_{k+1} = \max_{x \in V_k^\perp} \frac{x^t Ax}{||x||} = \max_{||x||=1, x \in V_k^\perp} x^t Ax$$

Proof 1
We will find the points where the derivative of $\frac{x^tAx}{||x||^2}$ is zero. First,

$$\begin{aligned}\frac{\partial}{\partial x_i}\left(x^tAx\right) &= \frac{\partial}{\partial x_i}\left(\sum_{r,s} a_{r,s}x_rx_s\right) = \sum_i (Ax)_i \\ \frac{\partial}{\partial x_i}\left(\|x\|^2\right) &= \frac{\partial}{\partial x_i}\left(\sum_r {x_r}^2\right) = 2x_i.\end{aligned}$$

Remember that $\left(\frac{f}{g}\right)' = 0 \iff f'g = fg'$. Thus, $\left(\frac{xAx^t}{\|x\|^2}\right)'_i = 0$ for every i if and only if:

$$\forall i, (Ax)_i \|x\|^2 = \left(x^tAx\right)x_i \qquad \iff \qquad \|x\|^2Ax = \left(x^tAx\right)x \qquad \iff \qquad Ax = \lambda x.$$

All this shows that the derivative equals zero in the eigenvectors of A. The maximal value is recieved when we take the eigenvector that correspond to the maximal eigenvalue. With the same reasoning we can find λ_{k+1}. ■

Proof 2
Recall that a symmetric real matrix $A \in M_n(\mathbb{R})$ can be written as $A = V\Lambda V^t$ where V is an orthogonal matrix (whose columns are eigenvectors) and Λ is a diagonal matrix (whose diagonal contains the eigenvalues).

Thus,

$$x^tAx = x^tV\Lambda V^tx = (V^tx)^t\Lambda V^tx = \sum_{i=1}^n \lambda_i(V^tx)_i^2 \le \max_i\{\lambda_i\}\cdot\sum_{i=1}^n(V^tx)_i^2 = \lambda_1\sum_{i=1}^n(V^tx)_i^2$$

Since V is orthogonal, we have $\sum_{i=1}^n (xV)_i^2 = ||V^tx||^2 = ||x||^2$ and therefore, $x^tAx \le \lambda_1||x||^2$.

We got $\lambda_1 \ge \frac{\max_{x\in\mathbb{R}^n} x^tAx}{||x||^2}$ and since the quotient actually reaches λ_1 (on the corresponding eigenvector) we get the desired result.

As for the rest of the eigenvalues and eigenvectors, for $i \le k$ it holds that $(V^tx)_i = 0$ so

$$x^tAx = \sum_{i=k+1}^n \lambda_i(V^tx)_i^2 \le \max_{i>k}\{\lambda_i\}\cdot\sum_{i=k+1}^n (V^tx)_i^2 \le \lambda_{k+1}||x||^2$$

and the argument continues as for λ_1.

Note that since the eigenvectors are orthogonal, the constraint $x \perp v_1, \ldots, v_k$ is equivalent to constraining x to be in the span of the remaining eigenvectors $v_{k+1}, \ldots, v_n$. ■

The quotient $\max_{x\in\mathbb{R}^n} \frac{x^tAx}{||x||^2}$ is called *Rayleigh quotient.* Note that the same argumentation could be carried out using the minimum of the Rayleigh quotient to get λ_n, then $\lambda_{n-1}, ..., \lambda_1$.

A closely related theorem is the Courant-Fischer theorem.

Theorem 2 (Courant-Fischer) *Let $A \in M_n(\mathbb{R})$ be a symmetric real matrix, and let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be its eigenvalues (in descending order), then for $i \in \{1, \ldots, n\}$:*

$$\lambda_{i+1} = \min_{F \,:\, dim(F)=i} \; \max_{x \perp F} \frac{x^t A x}{||x||^2},$$

and the eigenvector corresponding to λ_i is the vector on which the maximum is attained. Similarly,

$$\lambda_i = \max_{G \,:\, dim(G)=n-i} \; \min_{x \perp G} \frac{x^t A x}{||x||^2}$$

and the eigenvector corresponding to λ_i is the vector on which the minimum is attained.

2 Perron-Frobenius Theorem

We defined: A matrix $A_{n \times n}$ is a real positive matrix if each entry of A is a real positive number. We denote by $\rho(A)$ the spectral radius of A, that is,

$$\rho(A) \;=\; \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\} \;\; .$$

Many dynamical models can be described as iterative processes in which we begin with some initial vector $\mathbf{x}^0$ and then evolve recursively by the rule, $\mathbf{x}^{t+1} = A\mathbf{x}^t$. When the matrix A is positive, Perron's theorem tells us that the iterative process will converge exactly if $\rho(A) = 1$.

Theorem 3 (Perron) *Let $A \in M_n(\mathbb{R}^+)$ be a real positive matrix. Then, $\rho(A)$ is a simple eigenvalue of A and its corresponding eigenvector is a positive vector. In addition, if $\mathbf{x}_r$ and $\mathbf{x}_\ell$ are the right and left eigenvalues that correspond to $\rho(A)$, namely, $A\mathbf{x}_r = \rho(A)\mathbf{x}_r$ and $\mathbf{x}_\ell^T A = \rho(A)\mathbf{x}_\ell^T$, then*

$$\lim_{m \to \infty} \left(\frac{A}{\rho(A)} \right)^m \;=\; \frac{\mathbf{x}_r \, \mathbf{x}_\ell^t}{< \mathbf{x}_r, \mathbf{x}_l >} \;\; .$$

There is a connection between a non-negative matrix with $n \times n$ entries and a graph with n vertices V. Every entry A_{ij} in the matrix represents the weight of an edge between vertex v_i and v_j. Frobenius (1918) generalized Perron's theorem from positive matrices to non-negative matrices. In this case, the result in the theorem is still valid as long as the (directed) graph $G = (V, E)$ corresponds to A satisfies the following two conditions:

1. G is strongly connected. That is, for all two vertices $i, j \in V$, there is a directed path from i to j and from j to i.

2. V can not be partitioned into subsets $S_1, \ldots, S_d$ such that $E = (S_1 \times S_2) \cup \ldots \cup (S_{d-1} \times S_d) \cup (S_d \times S_1)$.

Proof We will prove Perron's theorem regarding positive matrices. Notice, that given two matrices V, W, in order to prove that $V = W$, we can show that for basis vector u_i,

$Vu_i = Wu_i$. We assume that the eigenvector of A form a basis for the vector space $\mathbb{R}^n$.[1] Let $u_1 = x_r, u_2, ..., u_n$ be the right eigenvectors of A, and let $\lambda_1 = \rho(A), \lambda_2, ..., \lambda_n$ be their corresponding eigenvalues.

$$\left(\frac{A}{\rho(A)}\right)^m u_i = \left(\frac{A}{\rho(A)}\right)^m = \left(\frac{\lambda_i}{\rho(A)}\right)^m u_i \longrightarrow \left\{ \begin{array}{ll} x_r & \text{if i=1} \\ 0 & \text{if i=2,...,n} \end{array} \right\}$$

when $m \to \infty$, since $|\frac{\lambda_i}{\rho(A)}| < 1$. This means that the limit exists, and let us denote this limit by L. Moreover, the limit L is a matrix of rank 1 (since the image is spanned by one vector, x_r). Every matrix of rank one equals uv^t for some vectors u, v. We will find u, v such that $L = uv^t$. On one hand,

$$Lu_i = lim_{m \to \infty} \left(\frac{A}{\rho(A)}\right)^m u_i = cx_r \text{ for } c \in \{0, 1\}.$$

On the other hand,

$$Lu_i = (uv^t)u_i = u < v, u_i > \in span\{u\}.$$

From this we get $u = x_r$, up to a constant multiplicity. Similarly let $v_1 = x_l, v_2, ..., v_n$ linearly independent set of left eigenvectors,

$$v_i \left(\frac{A}{\rho(A)}\right)^m \longrightarrow \left\{ \begin{array}{ll} x_l & \text{if i=1} \\ 0 & \text{if i=2,...,n} \end{array} \right\}$$

and thus $v = x_\ell$, up to a constant multiplicity. So we got $L = \mathbf{x}_r \mathbf{x}_\ell^t$, up to a constant multiplicity. A short calculation shows that the constant must be $< \mathbf{x}_r, \mathbf{x}_l >$ to ensure that the eigenvalue is 1.

■

2.1 Finding the largest eigenvalue $\rho(A)$

Look at the following optimization problem:

$$\lambda_0 = max\{\delta \in \mathbb{R} : \exists x \geq 0, x \neq 0, Ax \geq \delta x\}$$

Lemma 4 *$\lambda_0 = \rho(A)$ and the optimal vector x is the eigenvector of $\rho(A)$. If λ is any eigenvalue of A, then $\rho(A) \geq \lambda$* [2].

Proof Assume that $Ax > \lambda_0 x$ is not an equality. Meaning, there exists k such that $(Ax)_k > \lambda_0 x_k$. We will show that by changing x with $x + \epsilon e_k$, for ϵ small enough, we can enlarge λ_0. This will be a contrast to the definition of λ_0. So we have to show that $A(x + \epsilon e_k) > \lambda_0 (x + \epsilon e_k)$, or equivalently, that $Ax + \epsilon a^{(k)} > \lambda_0 x + \epsilon \lambda_0 e_k$, (where $a^{(k)}$ is the k-th coloumn of A). By assumption, $Ax \geq \lambda_0 x$. So we have to show that $a^{(k)} > \lambda_0 e_k$.

[1]If there is no basis of eigenvectors of A, then by a little perturbation on A we can have such a basis. This perturbation does not change the behavior. Therefore we can assume that there exists a basis of eigenvectors of A.

[2]Notice, that on Perron theorem $\rho(A) > \lambda$

We will show that this inequality holds in every coordinate. For $i \neq k$, $\left(a^{(k)}\right)_i = a_{ik} \geq 0 = \lambda_0 \left(e_k\right)_i$. For $i = k$, there is a strong inequality when $\epsilon = 0$. Thus, for ϵ small enough there is still a strong inequality. So we proved that λ_0 is an eigenvalue of A.

Now, suppose that λ is an eigenvalue of A, and denote the corresponding eigenvector by z (i.e. $Az = \lambda z$). Denote by $|z|$ the vector whose i-th coordinate is $|z_i|$ ($\Rightarrow |z| \geq 0$) .It is easy to see that $A|z| \geq \lambda|z|$. Thus, $\lambda \in \{\delta \in \mathbb{R} : \exists x \geq 0, x \neq 0, Ax \geq \delta x\}$, and from this follows that $\lambda_0 \geq \lambda$. Then, by definition, $\rho(A) = \lambda_0$.

■