CSE 521: Design and Analysis of Algorithms I **Spring 2016**

Lecture 8: Eigenvalues, Eigenvectors and Spectral Theorem

Lecturer: Shayan Oveis Gharan *April 20th* *Scribe: Sudipto Mukherjee*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

8.1 Introduction

Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix. We say λ is an eigenvalue of M with eigenvector v, if

$$Mv = \lambda v.$$

Theorem 8.1. *If M is symmetric, then all its eigenvalues are real.*

Proof. Suppose

$$Mv = \lambda v.$$

We want to show that λ has imaginary value 0. For a complex number $x = a + ib$, the conjugate of x, is defined as follows: $x^* = a - ib$. So, all we need to show is that $\lambda = \lambda^*$. The conjugate of a vector is the conjugate of all of its coordinate.

Taking the conjugate transpose of both sides of the above equality, we have

$$v^* M = \lambda^* v^*, \tag{8.1}$$

where we used that $M^T = M$.

So, on one hand,

$$v^* Mv = v^*(Mv) = v^*(\lambda v) = \lambda(v^* v).$$

and on the other hand, by (8.1)

$$v^* Mv = \lambda^* v^* v.$$

So, we must have $\lambda = \lambda^*$. □

8.2 Characteristic Polynomial

If M does not have 0 as one of its eigenvalues, then $\det(M) \neq 0$. An equivalent statement is that, if M all columns of M are linearly independent, then $\det(M) \neq 0$.

If λ is an eigenvalue of M, then $Mv = \lambda v$, so $(M - \lambda I)v = 0$. In other words, $M - \lambda I$ has a zero eigenvalue and $\det(M - \lambda I) = 0$.

Definition 8.2. *Characteristic polynomial of a matrix M is given by $\det(xI - M)$, which is a polynomial of degree n in the variable x.*

By the above argument any eigenvalue of M is a root of $\det(xI - M)$. Since any degree n polynomial has n roots, M must have exactly n eigenvalues. Furthermore, since the coefficient of x^n in $\det(xI - M)$ is 1, we can write,

$$\det(xI - M) = (x - \lambda_1)(x - \lambda_2)\dots(x - \lambda_n),$$

where $\lambda_1, \lambda_2, \dots \lambda_n$, are the eigenvalues of M.

Let us give an example to better understand the characteristic polynomial. Let

$$M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Then,

$$\det(xI - M) = \det \begin{bmatrix} x-1 & -2 \\ x-1 & -2 \end{bmatrix} = (x-1)^2 - (-2)(-2) = (x-3)(x+1).$$

So, by the above theory, $3, -1$ must be the eigenvalues of M. Indeed these are the eigenvalues with the following eigenvectors,

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The following corollary immediately follows from the above discussion.

Corollary 8.3. *For any matrix M,* $\det(M) = \prod_{i=1}^{n} \lambda_i$

To prove the above corollary it is enough to let $x = 0$ in the characteristic polynomial. Then,

$$\det(0I - M) = \det(-M) = (-1)^n \det(M) = (-1)^n \prod_{i=1}^{n} \lambda_i.$$

8.3 Spectral Theorem

For two vectors u, v we use

$$\langle u, v \rangle := \sum_i u(i)v(i).$$

We say u is *orthogonal* to v, $u \perp v$, if $\langle u, v \rangle = 0$. We say a family of vectors $\{v_1, \dots, v_n\}$ are *orthonormal* if for all $i \neq j$, v_i, v_j are orthogonal and each v_i has norm 1.

Theorem 8.4. *For any symmetric matrix $M \in \mathbb{R}^{n \times n}$, $\exists \lambda_1, \lambda_2, \dots \lambda_n \in \mathbb{R}$ with corresponding orthonormal eigenvectors $v_1, v_2, \dots v_n$. Furthermore, M can be written as follows:*

$$M := \sum_i \lambda_i v_i v_i^T.$$

Note that for any vector v, vv^T is an $n \times n$ matrix where for all i, j, $(vv^T)_{i,j} = v(i)v(j)$.

If λ is an eigenvalue of M with eigenvector v, then λ^k is an eigenvalue of M^k. This is because

$$M^k v = M^{k-1}\lambda v = M^{k-2}\lambda^2 v = \lambda^k v.$$

In addition observe that the same vector v is an eigenvector of λ^k. Consequently, by the spectral theorem we can write,

$$M^k = \sum_i \lambda_i^k v_i v_i^T.$$

We can also use spectral theorem to write any functions of the matrix M. For example, we can write $M^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i} v_i v_i^T$.

Claim 8.5. *Let M^{-1} be as defined above. Then,*

$$M^{-1}M = I.$$

Proof. By the above definition,

$$\begin{aligned} M^{-1}M &= \left(\sum_{i=1}^n \frac{1}{\lambda_i} v_i v_i^T\right)\left(\sum_{i=1}^n \lambda_i v_i v_i^T\right) \\ &= \sum_{i,j} \frac{\lambda_i}{\lambda_j}(v_i v_i^T)(v_j v_j^T) \\ &= \sum_{i,j} \frac{\lambda_i}{\lambda_j} v_i (v_i^T v_j) v_j^T) \\ &= \sum_i 1.v_i(\underbrace{v_i^T v_i}_{1}) v_i^T) = \sum_i v_i v_i^T = I. \end{aligned}$$

In the fourth equality we used that for all $i \neq j$, v_i, v_j are orthogonal, and in the fifth equality we used that each v_i has norm 1. It is a nice exercise to prove the last equality. □

8.4 Trace

The *Trace* of a symmetric matrix M is defined as follows:

$$\text{Tr}(M) = \sum \sum_{i=1}^n M_{i,i}.$$

Theorem 8.6. *For any symmetric matrix M, with eigenvalues $\lambda_1, \ldots, \lambda_n$, we have* $\text{Tr}(M) = \sum_i \lambda_i$.

Proof. As usual, let $e_1, e_2, \ldots e_n$ be the standard basis vectors, i.e., for all i,

$$e_i(j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

Let $v_1, \ldots, v_n$ be the eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$. Then,

$$\begin{aligned} \text{Tr}(M) &= \sum_i e_i^T M e_i \\ &= \sum_i e_i^T \left(\sum_j \lambda_j v_j v_j^T\right) e_i \\ &= \sum_i \sum_j \lambda_i \langle e_i, v_j \rangle^2 = \sum_j \lambda_j \|v_j\|^2 = \sum_i \lambda_i. \end{aligned}$$

In the second to last equality we used that

$$\sum_{i=1}^{n} \langle e_i, v_j \rangle^2 = \sum_{i=1}^{n} v_j(i)^2 = \|v_j\|^2.$$

□

8.5 Rayleigh Quotient

So, far we defined eigenvectors as the solution of a system of linear equations, that is for eigenvalue λ, v is the solution to $(M - \lambda I)v = 0$. The Rayleigh Quotient, a.k.a., *variational characterization* of the eigenvalues allows us to write eigenvalues (and eigenvectors) as the solution to an optimization problem. Using this characterization one also define and study *approximate* eigenvectors of a matrix.

Theorem 8.7. *Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$. Then*

$$\lambda_1 = \min_{\|x\|=1} x^T M x, \tag{8.2}$$

furthermore, the minimizer is the eigenvector v_1 corresponding to λ_1. We also have

$$\lambda_2 = \min_{\substack{x : x \perp v_1, \\ \|x\|=1}} x^T M x.$$

In general, for any $k \geq 1$ we can write,

$$\lambda_k = \min_{dim(S)=k} \max_{\substack{x \in S \\ \|x\|=1}} x^T M x,$$

where the minimum is over all linear spaces S of dimension k.

Note that $x^T M x$ in the above theorem is also known as the *quadratic form*. It is a basic mathematical object in studying the matrices.

Proof. Here we prove (8.2). The rest of the equations can be proven by similar ideas. Let $v_1, \ldots, v_n$ be the eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$. First, observe that for $x = v_1$ we have

$$v_1^T M v_1 = v_1^T \lambda_1 v_1 = \lambda_1 \|v_1\|^2 = \lambda_1.$$

So, it remains to show that for any unit vector x, $x^T M x \geq \lambda_1$.

$$\begin{aligned} x^T M x &= \sum \lambda_i x^T v_i v_i^T x \\ &= \sum \lambda_i \langle x, v_i \rangle^2 \\ &\geq \sum \lambda_1 \langle x, v_i \rangle^2 \\ &= \lambda_1 \|x\|^2 = \lambda_1 \end{aligned}$$

The inequality uses the fact that λ_1 is the smallest eigenvalue of M and that for any i, $\langle x, v_i \rangle^2 \geq 0$. The second to last equality uses the fact that $v_1, \ldots, v_n$ form an orthonormal basis of $\mathbb{R}^n$. We can write any vector x in this bases as $x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i$. For such a representation $\|x\|^2$ is nothing but $\sum_i \langle v_i, x \rangle^2$. □

8.6 Positive Semidefinite Matrices

Definition 8.8. *We say a symmetric matrix $M \in \mathbb{R}^{n \times n}$ is positive semi-definite (denoted as $M \succeq 0$) iff $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$.*

Equivalently, a symmetric matrix $M \in \mathbb{R}^{n \times n}$ is positive semi-definite iff $x^T M x \ge 0 \, \forall \, x \ne 0$. The equivalence simply follows from the Rayleigh quotient. Note that the smallest eigenvalue of M is $\min x^T M x$ over all unit vectors x. So, if $x^T M x \ge 0$ for all x, the smallest eigenvalue of M is nonnegative. Conversely, if the smallest eigenvalue of M is nonnegative, $x^T M x \ge 0$ for all vectors x.

Similarly, we can define a negative semi-definite matrix denoted by $M \preceq 0$ as a matrix where all eigenvalues are nonpositive, $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \le 0$.

Using the above definition we define square root for PSD matrices. For a PSD matrix M,

$$\sqrt{M} = \sum_i \sqrt{\lambda_i} v_i v_i^T.$$

It is an exercise to show that $\sqrt{M} \cdot \sqrt{M} = M$.

8.7 Singular Value Decomposition

Non-symmetric square matrices still have n eigenvalues which are not necessarily real. This again follows from the fact that the characteristic polynomial has n roots. But unfortunately, the corresponding eigenvectors are not necessarily orthogonal. For a concrete example observe that the following matrix has eigenvalues $0, 3$ and the corresponding eigenvectors are not orthogonal.

$$\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = 0.$$

However, we can define a singular value decomposition for nonsymmetric matrices. This is known as the *singular value decomposition* (SVD) and we will talk about in the next lecture.