Lecture 8 : Eigenvalues and Eigenvectors

Lecturer: Kamesh Munagala *Scribe: Kamesh Munagala*

Hermitian Matrices

It is simpler to begin with matrices with complex numbers. Let $x = a + ib$, where a, b are real numbers, and $i = \sqrt{-1}$. Then, $x^* = a - ib$ is the complex conjugate of x. In the discussion below, all matrices and numbers are complex-valued unless stated otherwise.

Let M be an $n \times n$ square matrix with complex entries. Then, λ is an eigenvalue of M if there is a non-zero vector $\vec{v}$ such that

$$M\vec{v} = \lambda\vec{v}$$

This implies $(M - \lambda I)\vec{v} = 0$, which also means the determinant of $M - \lambda I$ is zero. Since the determinant is a degree n polynomial in λ, this shows that any M has n real or complex eigenvalues.

A complex-valued matrix M is said to be *Hermitian* if for all i, j, we have $M_{ij} = M^*_{ji}$. If the entries are all real numbers, this reduces to the definition of *symmetric matrix*.

In the discussion below, we will need the notion of inner product. Let $\vec{v}$ and $\vec{w}$ be two vectors with complex entries. Define their inner product as

$$\langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i^* w_i$$

Since $(x^*)^* = x$ for any complex number x, we have

$$\langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i^* w_i = \sum_{i=1}^{n} (w_i^* v_i)^* = (\langle \vec{w}, \vec{v} \rangle)^*$$

Furthermore, we define

$$\langle \vec{v}, \vec{v} \rangle = \sum_{i=1}^{n} v_i^* v_i = \|\vec{v}\|^2$$

Claim 1. *M is Hermitian iff all its eigenvalues are real. If further M is real and symmetric, then all its eigenvectors have real entries as well.*

Proof. Using the fact that $M_{ij}* = M_{ji}$, we have the following relations:

$$\begin{aligned}
\langle M\vec{v}, \vec{v} \rangle &= \sum_i \sum_j (M_{ij} v_j)^* v_i \\
&= \sum_i \sum_j M_{ji} v_j^* v_i \\
&= \sum_j v_j^* \sum_i (M_{ji} v_i) \\
&= \langle \vec{v}, M\vec{v} \rangle
\end{aligned}$$

Suppose $M\vec{v} = \lambda\vec{v}$. Then

$$\langle M\vec{v}, \vec{v}\rangle = \langle \lambda\vec{v}, \vec{v}\rangle = \lambda^* \|\vec{v}\|^2$$

and

$$\langle \vec{v}, M\vec{v}\rangle = \langle \vec{v}, \lambda\vec{v}\rangle = \lambda \|\vec{v}\|^2$$

Since these expressions are equal, this means $\lambda^* = \lambda$, which means λ is real.

Suppose M is real and symmetric. Let λ be some eigenvalue, which by the above definition is real. Then we have $M\vec{v} = \lambda\vec{v}$ for some complex $\vec{v}$. Since M is real, this means, the above relation also holds for both the real and complex parts of $\vec{v}$. Therefore, if $\vec{w}$ is the real part of $\vec{v}$, then $M\vec{w} = \lambda\vec{w}$. This implies all eigenvectors are real if M is real and symmetric. □

From now on, we will only focus on matrices with real entries.

Claim 2. *For a real, symmetric matrix M, let $\lambda \neq \lambda'$ be two eigenvalues. Then the corresponding eigenvectors are orthogonal.*

Proof. Let $M\vec{v} = \lambda\vec{v}$ and $M\vec{w} = \lambda'\vec{w}$. Since M is symmetric, it is easy to check that

$$\langle M\vec{v}, \vec{w}\rangle = \langle \vec{v}, M\vec{w}\rangle = \sum_{i,j} M_{ij} v_i w_j$$

But

$$\langle M\vec{v}, \vec{w}\rangle = \lambda\langle \vec{v}, \vec{w}\rangle$$

and

$$\langle \vec{v}, M\vec{w}\rangle = \lambda'\langle \vec{v}, \vec{w}\rangle$$

Since $\lambda \neq \lambda'$, this implies $\langle \vec{v}, \vec{w}\rangle = 0$, which means the eigenvectors are orthogonal. □

We state the next theorem without proof.

Theorem 1. *Let M be a $n \times n$ real symmetric matrix, and let $\lambda_1, \ldots, \lambda_n$ denote its eigenvalues. Then, there exist n real-valued vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}$ such that:*

- $\|\vec{v_i}\| = 1$ *for all* $i = 1, 2, \ldots, n$;
- $\langle \vec{v_i}, \vec{v_j}\rangle = 0$ *for all* $i \neq j \in \{1, 2, \ldots, n\}$; *and*
- $M\vec{v_i} = \lambda_i \vec{v_i}$ *for all* $i = 1, 2, \ldots, n$.

Computing Eigenvalues

In this section, we assume M is real and symmetric.

Lemma 1. *Let M be a real symmetric matrix. Let λ_1 denote its largest eigenvalue and $\vec{v_1}$ denote the corresponding eigenvector with unit norm. Then*

$$\lambda_1 = \sup_{\vec{x} \in \mathbf{R}^n, \|\vec{x}\| = 1} \vec{x}^T M \vec{x} = \vec{v_1}^T M \vec{v_1}$$

Proof. First note that since $\|\vec{v_1}\| = 1$, we have:

$$\vec{v_1}^T M \vec{v_1} = \lambda_1 \vec{v_1}^T \vec{v_1} = \lambda_1$$

This means

$$\lambda_1 \leq \sup_{\vec{x} \in \mathbf{R}^n, \|\vec{x}\|=1} \vec{x}^T M \vec{x}$$

Suppose the supremum is achieved at vector $\vec{y}$. Let $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}$ denote the orthogonal eigenvectors of unit length corresponding to the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ respectively. Then we can write $\vec{y}$ in this basis as:

$$\vec{y} = \sum_{i=1}^{n} \alpha_i \vec{v_i}$$

Since $\vec{y}$ has unit length, this means

$$\langle \vec{y}, \vec{y} \rangle = \sum_{i=1}^{n} \alpha_i^2 = 1$$

Note next that

$$M\vec{y} = \sum_{i=1}^{n} \alpha_i \lambda_i \vec{v_i}$$

This means

$$\vec{y}^T M \vec{y} = \langle \sum_{i=1}^{n} \alpha_i \vec{v_i}, \sum_{i=1}^{n} \alpha_i \lambda_i \vec{v_i} \rangle = \sum_{i=1}^{n} \alpha_i^2 \lambda_i$$

This means

$$\sup_{\vec{x} \in \mathbf{R}^n, \|\vec{x}\|=1} \vec{x}^T M \vec{x} = \vec{y}^T M \vec{y} = \sum_{i=1}^{n} \alpha_i^2 \lambda_i \leq \lambda_1 \sum_{i=1}^{n} \alpha_i^2 = \lambda_1$$

This means the supremum has value exactly λ_1 and is achieved for $\vec{x} = \vec{v_1}$. □

We can continue the same argument to show the following corollaries:

Corollary 2. *Let λ_2 denote the second largest eigenvalue of a real, symmetric matrix M, and let $\vec{v_1}$ denote the first eigenvector. Then*

$$\lambda_2 = \sup_{\vec{x} \in \mathbf{R}^n, \|\vec{x}\|=1, \langle \vec{x}, \vec{v_1} \rangle = 0} \vec{x}^T M \vec{x}$$

Corollary 3. *Let M be a real symmetric matrix, and λ_n denote its smallest eigenvalue. Then*

$$\lambda_n = \inf_{\vec{x} \in \mathbf{R}^n, \|\vec{x}\|=1} \vec{x}^T M \vec{x}$$

Orthonormal Square Matrices

In order to interpret what the above results mean, we first review orthonormal matrices. A square matrix P is orthonormal if its rows (columns) are orthogonal vectors of unit length. More formally, we have

$$P^T P = P P^T = I$$

Note that since the matrix is square and the rows are orthogonal, they cannot be expressed as linear combinations of each other. This means the matrix is invertible. Multiplying the above expressions by P^{-1}, it is easy to check that

$$P^T = P^{-1}$$

In order to interpret what P does, note the following. For any vectors $\vec{v}$ and $\vec{w}$,

$$\langle P\vec{v}, P\vec{w}\rangle = \vec{v}^T P^T P\vec{w} = \vec{v}^T\vec{w} = \langle \vec{v}, \vec{w}\rangle$$

This means applying P preserves the angles between vectors, as well as their lengths. This means applying P performs a *rotation* of the space.

Given a real, symmetric matrix M with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, let Q denote the matrix whose rows are the corresponding eigenvectors of unit length. Since these eigenvectors are orthogonal, this implies Q is orthonormal. Let D be the matrix whose entries along the diagonal are the n eigenvalues, and other entries are zero. It is easy to check that:

$$MQ^T = Q^T D \qquad \Rightarrow \qquad M = Q^T D Q$$

Therefore, applying M to a vector v is the same as applying the rotation Q; then stretching dimension i by factor λ_i, and rotating back by Q^T. Note that λ_i could be negative, which flips the sign of the corresponding coordinate.

If we were to do this, which direction stretches the most? The dimension corresponding to λ_1 after the rotation by Q. But in the original space, this would be the direction corresponding to the first eigenvector $\vec{v_1}$. Now, for unit vector $\vec{x}$, we have

$$\vec{x^T} M\vec{x} = (Q\vec{x})^T D\,(Q\vec{x})$$

This quantity therefore corresponds to the length of the stretched rotated vector (since $Q\vec{x}$ is also a unit vector), which is λ_1 if $\vec{x} = \vec{v_1}$. This is the interpretation of the previous result.

Power Iteration

Assuming λ_1 is strictly larger than λ_2, there is a simple algorithm to compute the largest eigenvector. Note that

$$M^k = \left(Q^T D Q\right)^k = Q^T D^k Q$$

This means M^k has eigenvectors $\vec{v_1}, \vec{v_2}, \ldots$, and the corresponding eigenvalues are $\lambda_1^k, \lambda_2^k, \ldots$. Suppose $\vec{x} = \sum_{i=1}^n \alpha_i \vec{v_i}$, then

$$M^k\vec{x} = \sum_{i=1}^n \alpha_i \lambda_i^k \vec{v_i}$$

The claim is that as k becomes large, this vector points more and more in the direction of $\vec{v_1}$. Suppose we choose the initial vector $\vec{x}$ at random. Since it is a random vector, each $\alpha_i \sim \mathcal{N}\left(0, \frac{1}{n}\right)$. This means with large probability, $|\alpha_i| \in \left[\frac{1}{n}, 1\right]$. If we assume $\lambda_1 \geq c\lambda_2$, then, for $i \geq 2$, $\alpha_i \lambda_i^k \leq \lambda_1^k / c^k$, while $\alpha_1 \lambda_1^k \geq \lambda_1^k / n$. If we choose $c^k \geq n$, so that $k \geq \log_c n$, then the terms for $i \geq 2$ have vanishingly small coefficients, and the dominant term corresponds to $\vec{v_1}$. Assuming $c = 1 + \epsilon$, we need to choose $k \approx \frac{\log n}{\epsilon}$ for the second and larger eigenvectors to have vanishing contributions compared to the first.

This method is termed power iteration. It can be improved by repeated squaring, so that k grows in powers of 2. The number of iterations then drops to around $\log \frac{1}{\epsilon} + \log \log n$. So even an exponentially small gap, say $\epsilon = \frac{1}{2^n}$ between λ_1 and λ_2 is sufficient for the algorithm to converge to the direction $\vec{v_1}$ in polynomially many iterations.

Once we have computed $\vec{v_1}$, to compute $\vec{v_2}$, simply pick a random vector and project in the direction perpendicular to $\vec{v_1}$. Call this projected vector$\vec{x}$. Since $\vec{x} = \sum_{i=2}^{n} \alpha_i \vec{v_i}$, we again have

$$M^k \vec{x} = \sum_{i=2}^{n} \alpha_i \lambda_i^k \vec{v_i}$$

In the limit, this converges to the second eigenvector assuming the second eigenvalue is well-separated from the third. And so on.

Positive semidefinite Matrices

Positive semidefinite (PSD) matrices are a special case of real symmetric matrices. A matrix M is said to be PSD if $\vec{x}^T M \vec{x} \geq 0$ for all $\vec{x}$. As an example, if $M = A^T A$ for any matrix A, then it is easy to see that

$$\vec{x}^T M \vec{x} = (A\vec{x})^T (A\vec{x}) \geq 0$$

It is also easy to see that all eigenvalues of a PSD matrix are non-negative. To see this, note that

$$\vec{v_i}^T M \vec{v_i} = \lambda_i \langle \vec{v_i}^T, \vec{v_i} \rangle \geq 0 \qquad \Rightarrow \qquad \lambda_i \geq 0$$

The above characterization is if and only if: A real, symmetric matrix is PSD iff all its eigenvalues are non-negative.