Lectures - Week 7
Eigenvalues

The Algebraic Eigenvalue Problem

There are two major problems in linear algebra – solving a linear system $A\vec{x} = \vec{b}$ and the eigenvalue problem which is stated below.

Given an $n \times n$ matrix A, find a scalar λ and a nonzero vector $\vec{x}$ such that

$$A\vec{x} = \lambda\vec{x}$$

The scalar λ is called an *eigenvalue* of A and $\vec{x}$ the corresponding *eigenvector*. The pair $(\lambda, \vec{x})$ is called an eigenpair.

We first make some remarks.

1. If $(\lambda, \vec{x})$ is an eigenpair of A then so is $(\lambda, c\vec{x})$ for any nonzero c.
2. If $A\vec{x} = \lambda\vec{x}$ then $A\vec{x} - \lambda I\vec{x} = 0$ which implies $(A - \lambda I)\vec{x} = 0$; i.e., $\vec{x} \in \mathcal{N}(A - \lambda I)$.
3. Even if A is real, its eigenvalues may be complex.
4. If A is not invertible and if $\vec{x} \in \mathcal{N}(A)$ then $\vec{x}$ is an eigenvector of A corresponding to a zero eigenvalue.
5. If A is invertible, then there are no zero eigenvalues because there is nothing in the null space of A.
6. If A is invertible with eigenpair $(\lambda, \vec{x})$, then $(\frac{1}{\lambda}, \vec{x})$ is an eigenpair of A^{-1}.
7. If A has eigenpair $(\lambda, \vec{x})$, then $(\lambda^k, \vec{x})$ is an eigenpair of A^k.

Our goals are to determine methods to calculate eigenvalues and eigenvectors and to determine what properties the eigenvalues of a class of matrices have; for example is there a class of matrices which are guaranteed to have real eigenvalues.

Because the eigenvector is in the null space of $A - \lambda I$ we know that this matrix does not have full rank and thus is not invertible, i.e., it is singular. If a matrix B is singular then its determinant is zero. Consequently for our eigenvalue problem

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

This equation is called the *characteristic equation*. Because the determinant of a square matrix B equals the determinant of its transpose, the determinants of $(A - \lambda I)$ and $(A - \lambda I)^T$ are the same. Consequently the determinant of $(A^T - \lambda I)$ is the same and thus

8. A and A^T have the same eigenvalues.

Example Write the characteristic equation for

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix}$$

and use it to find the eigenvalues.

First we form $(A - \lambda I)$ and then take its determinant. We have

$$(A - \lambda I) = \begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 2-\lambda & 1 \\ 3 & -\lambda \end{pmatrix}$$

and then its determinant is

$$\det \begin{pmatrix} 2-\lambda & 1 \\ 3 & -\lambda \end{pmatrix} = (2-\lambda)(-\lambda) - 3 = \lambda^2 - 2\lambda - 3$$

So the eigenvalues are just the roots of the equation $\lambda^2 - 2\lambda - 3 = 0$. We can either factor this or use the quadratic formula

$$\lambda^2 - 2\lambda - 3 = 0 \Rightarrow (\lambda - 3)(\lambda + 1) = 0 \Rightarrow \lambda = 3, \lambda = -1$$

We see that this 2×2 matrix has two eigenvalues and they are distinct.

In this example the characteristic equation is a quadratic polynomial which we have to find the roots of using the quadratic formula. For a 2×2 matrix the characteristic polynomial is always a quadratic polynomial. We know that the two roots of a quadratic polynomial can be (i) real and distinct, (ii) real and repeated, or (iii) complex and complex conjugates of each other. In general, the characteristic equation of an $n \times n$ matrix is a polynomial of degree n in λ so it is a nonlinear equation which has n roots which may be real, repeated or complex. Recall that we have formulas for finding the roots of a quadratic polynomial and in fact there are also formulas for third and fourth degree polynomials but none for higher degrees. What are the repercussions of this? It means that we can NOT find a method for directly determining the eigenvalues because we can't solve an nth degree polynomial (for $n \geq 5$) exactly. This means that we will have to devise *iterative methods* for determining eigenvalues/eigenvectors; i.e., we will start with an initial guess and generate a sequence which we hope converges to the desired eigenpair. It might be tempting to find the eigenvalues by finding the roots of the characteristic equation but this turns out to be a terrible approach computationally.

It turns out that for some matrices (diagonal and triangular) the eigenvalues are just the diagonal entries. Why is this the case? For a diagonal matrix the matrix $(A - \lambda I)$ is still diagonal and the characteristic equation is just

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda) = 0$$

so the roots are just the diagonal entries. The same is true for a triangular matrix; $(A - \lambda I)$ is still triangular and the determinant of a triangular matrix is the product of its diagonal entries.

Example For the matrix given in the previous example, find the corresponding eigenvectors. Are they linearly independent?

To find the eigenvector we must find a vector in $\mathcal{N}(A - \lambda I)$ for each λ. Forming each matrix and row reducing gives us the solution

$$(A - 3I)\vec{x} = \vec{0} \Rightarrow \begin{pmatrix} -1 & 1 \\ 3 & -3 \end{pmatrix} \vec{x} = \vec{0} \Rightarrow \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \vec{x} = \vec{0} \Rightarrow \vec{x} = C_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$(A + I)\vec{x} = \vec{0} \Rightarrow \begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix} \vec{x} = \vec{0} \Rightarrow \begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix} \vec{x} = \vec{0} \Rightarrow \vec{x} = C_2 \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

Clearly they are linearly independent because one is not a constant times the other.

In the previous example we had two distinct eigenvalues and an eigenvector corresponding to each; because the two eigenvectors are linearly independent they form a basis for $\mathbf{R}^2$. Do we always have n linearly independent eigenvectors? To consider the situations that can arise we consider some examples.

Example Find the eigenpairs for each of the following matrices

$$A_1 = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \qquad A_3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

A_1 is an upper triangular matrix so its eigenvalues are its diagonal entries or equivalently we have the characteristic equation

$$(2 - \lambda)^2 = 0 \Rightarrow \lambda = 2, 2$$

In this case the eigenvalues are repeated. To find the eigenvectors we look at the null space of $A_1 - 2I$

$$A_1 - 2I = \begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}$$

which has a one dimensional null space and an eigenvector is $(1, 0)^T$. Thus this matrix has a repeated eigenvalue but only one eigenvector.

For A_2 the matrix is diagonal and has the repeated eigenvalue 2 but it has two linearly independent eigenvectors $(1, 0)$ and $(0, 1)$. So if we compare A_1 and A_2 we see that they have the same eigenvalue which is repeated twice but A_2 has two linearly independent eigenvectors whereas A_1 only has one.

The characteristic equation for A_4 is

$$\lambda^2 + 1 = 0 \Rightarrow \lambda^2 = -1 \Rightarrow \lambda = \pm i$$

So A_4 is a real matrix but it has imaginary eigenvalues. Clearly its eigenvectors must be complex too. We have

$$A_4 - iI = \begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \rightarrow \begin{pmatrix} -i & 1 \\ 0 & 0 \end{pmatrix}$$

and its eigenvector is $(-i, 1)$. Also

$$A_4 + iI = \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix} \to \begin{pmatrix} i & 1 \\ 0 & 0 \end{pmatrix}$$

and its eigenvector is $(i, 1)^T$.

It turns out that an $n \times n$ matrix has n eigenvalues counted according to multiplicity (i.e., how many times they are repeated). We use the term *algebraic multiplicity* to indicate the number of times a root of the characteristic equation is repeated. We use the term *geometric multiplicity* to indicate the number of linearly independent eigenvectors corresponding to a particular eigenvalue. So in our example, the eigenvalue 2 has algebraic multiplicity two for A_1 and A_2. For A_1 the geometric multiplicity is one whereas for A_2 it is two. The following theorem tells us that distinct eigenvalues have one linearly independent eigenvector.

Theorem For each distinct eigenvalue of an $n \times n$ matrix there is one eigenvector corresponding to it; i.e., both the algebraic and geometric multiplicities are one. If the algebraic multiplicity of an eigenvalue is r there can be at most r linearly independent eigenvectors corresponding to it.

Sometimes we have an eigenvector and we would like to use it to determine the corresponding eigenvalue. There is an easy way to do this called the *Rayleigh Quotient*. Multiplying our eigenvalue problem by $\vec{x}^T$ and solving for λ gives us

$$A\vec{x} = \lambda\vec{x} \Rightarrow \vec{x}^T A\vec{x} = \lambda\vec{x}^T\vec{x} \Rightarrow \lambda = \frac{\vec{x}^T A\vec{x}}{\vec{x}^T\vec{x}}$$

because we know $\vec{x}^T\vec{x} \neq 0$.

We would like our matrices to have a complete set of n linearly independent eigenvectors but as we have seen, this is not always the case. We will see that some classes of matrices are guaranteed to have a complete set. We now want to look at classes of matrices we have studied and see what can be said about their eigenvalues.

Lemma If A is an $n \times n$ symmetric (real) matrix then the eigenvalues of A are real.

To prove that the eigenvalue is real we must demonstrate that $\lambda = \lambda^*$ where $*$ denotes the complex conjugate (i.e., everywhere there is an i replace it with $-i$; thus, e.g., the complex conjugate of $2 - 3i$ is $2 + 3i$) We first pre-multiply our eigenvalue problem by $(x^*)^T$ and solve for λ to get

$$(\vec{x}^*)^T A\vec{x} = \lambda(\vec{x}^*)^T\vec{x} \Rightarrow \lambda = \frac{(\vec{x}^*)^T A\vec{x}}{(\vec{x}^*)^T\vec{x}}$$

We want to demonstrate that λ^* is equal to the same expression. Now we take the complex conjugate of $A\vec{x} = \lambda\vec{x}$ and then the transpose to get

$$A^*\vec{x}^* = \lambda^*\vec{x}^* \Rightarrow (A^*x^*)^T = (\lambda^*\vec{x}^*)^T \Rightarrow (\vec{x}^*)^T A^T = \lambda^*(\vec{x}^*)^T$$

where we have used the fact that A is real so $A^* = A$. Because A is symmetric $A^T = A$ and post-multiplying by $\vec{x}$ gives

$$(\vec{x}^*)^T A\vec{x} = \lambda^*(\vec{x}^*)^T\vec{x} \Rightarrow \lambda^* = \frac{(\vec{x}^*)^T A\vec{x}}{(\vec{x}^*)^T\vec{x}}$$

Thus λ and λ^* are equal to the same thing so $\lambda = \lambda^*$.

Lemma If A is an $n \times n$ symmetric positive definite matrix then the eigenvalues of A are real and greater than zero.

From the previous lemma we know that the eigenvalues are real thus we only need to demonstrate that they are positive. Because A is positive definite

$$\vec{x}^T A\vec{x} > 0 \quad \text{for all } \vec{x} \neq 0$$

But from the Rayleigh Quotient

$$\lambda = \frac{\vec{x}^T A\vec{x}}{\vec{x}^T\vec{x}} > 0$$

because $\vec{x}^T\vec{x} > 0$ for $\vec{x} \neq 0$.

Lemma If A is an $n \times n$ orthogonal matrix then the eigenvalues of A have magnitude one. Note that this says the eigenvalues can be for example ± 1, $\pm i$, $\frac{1}{\sqrt{2}}(1 \pm i)$, because the magnitude of a complex number $a + bi$ is given by $\sqrt{a^2 + b^2}$.

Taking the Euclidean norm of s $A\vec{x} = \lambda\vec{x}$ gives

$$\|A\vec{x}\|_2 = \|\lambda\vec{x}\|_2 \Rightarrow \|A\vec{x}\|_2^2 = |\lambda|^2\|\vec{x}\|_2^2$$

Now we know that an orthogonal matrix preserves the Euclidean length but we will demonstrate it again here

$$\vec{x}^T A^T A\vec{x} = |\lambda|^2\vec{x}^T\vec{x} \Rightarrow |\lambda|^2 = \frac{\vec{x}^T\vec{x}}{\vec{x}^T\vec{x}} = 1$$

where we have used the orthogonality of A.

Similarity Transformations

In some algorithms for finding eigenvalues we want to transform our matrix into one with a simpler form (such as a diagonal or triangular matrix) which has the same eigenvalues as the original matrix. This way we can "read off" the eigenvalues of the simpler matrix. So our question now is to determine when two matrices have the same eigenvalues.

Definition The matrices A, B are *similar* if there exists an invertible matrix M such that

$$B = M^{-1}AM$$

Note that this also implies

$$MBM^{-1} = A \Rightarrow A = P^{-1}BP \qquad \text{where } P = M^{-1}$$

If the matrix M is orthogonal, then we say the matrices are orthogonally similar, i.e.,

$$B = Q^T AQ \qquad \text{where } Q \text{ is an orthogonal matrix}$$

The next result demonstrates that similar matrices have the same eigenvalues but not the same eigenvectors; however, their eigenvectors are related.

Theorem If A and B are similar matrices, i.e., $B = M^{-1}AM$, then they have the same eigenvalues. Also if $A\vec{x} = \lambda\vec{x}$ and $B\vec{y} = \sigma y$ then $\vec{x} = M\vec{y}$.

Proof We begin by taking the eigenvalue problem for B and then using the fact that A, B are similar to obtain

$$B\vec{y} = \sigma y \Rightarrow M^{-1}AM\vec{y} = \sigma\vec{y} \Rightarrow AM\vec{y} = \sigma M\vec{y}$$

and thus $A\vec{x} = \sigma\vec{x}$ where $\vec{x} = M\vec{y}$ which means that if σ is an eigenvalue of B then it is an eigenvalue of A. The eigenvector of A corresponding to σ is found by multiplying the eigenvector of B corresponding to σ by M.

This tells us that if we use similarity transformations to convert our original matrix into one which is diagonal or triangular then the eigenvalues are the same and can be easily obtained. We first want to know when a matrix can be made similar to a diagonal or triangular matrix.

Schur's Theorem Let A be an $n \times n$ matrix. Then there is an orthogonal matrix Q such that $Q^T AQ$ is triangular. Moreover, if A is symmetric, then $Q^T AQ$ is diagonal.

The proof of Schur's theorem is not constructive; i.e., it only tells us that there exists a Q but doesn't tell us how to get Q. We now look at what an invertible matrix P would look like which would make $P^{-1}AP$ diagonal, if that is possible.

Example We have already seen that the eigenpairs of

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix}$$

are

$$\left\{ -1, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right\} \qquad \left\{ 3, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

Form the product $P^{-1}AP$ where the columns of P are the eigenvectors of A. What do you conclude?

If we set P to have columns which are the eigenvectors of A then

$$P = \begin{pmatrix} -1 & 1 \\ 3 & 1 \end{pmatrix} \Rightarrow P^{-1} = \frac{-1}{4} \begin{pmatrix} 1 & -1 \\ -3 & -1 \end{pmatrix}$$

Now $P^{-1}AP$ is just

$$P^{-1}AP = \frac{-1}{4}\begin{pmatrix} 1 & -1 \\ -3 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ 3 & 1 \end{pmatrix} = \frac{-1}{4}\begin{pmatrix} 1 & -1 \\ -3 & -1 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ -3 & 3 \end{pmatrix}$$

and thus

$$P^{-1}AP = \frac{-1}{4}\begin{pmatrix} 4 & 0 \\ 0 & -12 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}$$

So the matrix that diagonalizes A is the matrix whose columns are the eigenvectors of A.

Lemma An $n \times n$ matrix A is similar to a diagonal matrix if and only if it has n linearly independent eigenvectors.

Proof First assume $P^{-1}AP = D$ where D is a diagonal matrix. This implies $AP = DP$. If we denote the columns of P as $\vec{p}_i$ then this implies $A\vec{p}_i = d_{ii}\vec{p}_i$. So the columns of P are eigenvectors of A and since P is invertible the columns are linearly independent.

For the other direction, assume that A has n linearly independent eigenvectors, say $\vec{v}_i$, $i = 1, \ldots, n$ corresponding to the eigenvalues λ_i. We want to show that it is similar to a diagonal matrix. Let P have columns $\vec{v}_i$, $i = 1, \ldots, n$. Because the columns are linearly independent, P is invertible. Now

$$P^{-1}AP = P^{-1}\left(A\vec{v}_1 \quad A\vec{v}_2 \quad \cdots \quad A\vec{v}_n \right) = P^{-1}\left(\lambda_1\vec{v}_1 \quad \lambda_2\vec{v}_2 \quad \cdots \quad \lambda_n\vec{v}_n \right)$$

and thus

$$P^{-1}AP = \left(\lambda_1 P^{-1}\vec{v}_1 \quad \lambda_2 P^{-1}\vec{v}_2 \quad \cdots \quad \lambda_n P^{-1}\vec{v}_n \right) = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

because $P^{-1}P = I$ so $P^{-1}\vec{v}_i$ is a column of the identity matrix.

The next result tells us that eigenvectors corresponding to different eigenvalues are linearly independent. Thus if an $n \times n$ matrix has n distinct eigenvalues, then it has n linearly independent eigenvectors and is thus diagonalizable.

Proposition Eigenvectors corresponding to distinct eigenvalues are linearly independent.

To understand why this is the case, lets take the simple situation of two eigenpairs $\{\lambda_1, \vec{v}_1\}$ and $\{\lambda_2, \vec{v}_2\}$ of A where we assume $\lambda_1 \neq \lambda_2$. We want to show that the only way we can combine $\vec{v}_1$ and $\vec{v}_2$ and get zero, is with zero coefficients. To demonstrate this by contradiction, assume there are nonzero c_1, c_2 such that

$$c_1\vec{v}_1 + c_2\vec{v}_2 = \vec{0}$$

Now multiplying this equation by A gives

$$c_1A\vec{v}_1 + c_2A\vec{v}_2 = \vec{0} \Rightarrow c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2 = \vec{0}$$

Multiplying our original equation by λ_1 and subtracting gives

$$c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2 - \lambda_1c_1\vec{v}_1 - \lambda_1c_2\vec{v}_2 = \vec{0} \Rightarrow c_2(\lambda_2 - \lambda_1)\vec{v}_2 = 0 \Rightarrow c_2 = 0$$

because $\lambda_2 - \lambda_1 \neq 0$ and $\vec{v}_2 \neq \vec{0}$ by assumption. If $c_2 = 0$ then $c_1\vec{v}_1 = 0$ and because $\vec{v}_1 \neq \vec{0}$, $c_1 = 0$ and we get our contradiction.

Lemma A symmetric matrix has real eigenvalues and n linearly independent eigenvectors.