

# Section 5: Eigenvalues, Diagonalization and Symmetric Matrices

MATH241 Lecture Notes, notes pages 53–65

**Notation.** Scalars are plain italic ( $a, \alpha, z$ ), vectors are bold ( $\mathbf{u}, \mathbf{x}$ ), and matrices are italic capitals ( $A, B$ ; zero matrix  $O$ , identity  $I$ ). The handwritten notes underline vectors once and matrices twice.

These notes are built on top of Orhan Arikan's linear algebra course notes.

— notes page 53 —

## The Eigenvalue Problem

(Chapter 11 in the textbook, Lecture notes.)

A square matrix  $A$  acts on vectors: it transforms a vector  $\mathbf{x}$  into the vector  $A\mathbf{x}$ , a map from  $\mathbb{C}^n$  to  $\mathbb{C}^n$ . Most vectors are turned and stretched by this map. Some special vectors keep their direction and are only scaled. These are the *eigenvectors*, and the eigenvalue problem is to find them: for a square matrix  $A$ , find the vectors  $\mathbf{x} \neq \mathbf{0}$  and scalars  $\lambda$  with

### The eigenvalue equation

$$A\mathbf{x} = \lambda\mathbf{x} \quad \text{for a scalar } \lambda.$$

To solve it, move everything to one side:

$$A\mathbf{x} = \lambda\mathbf{x} \Rightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Rightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}.$$

For a given  $\lambda$  this is a homogeneous system, so it is always consistent, and  $\mathbf{x} = \mathbf{0}$  is always a solution. But we are interested in non-zero solutions. A homogeneous system has non-zero solutions exactly when its matrix is not invertible, that is, when

$$\underbrace{\det(A - \lambda I) = 0}_{\text{a polynomial equation in } \lambda}.$$

So non-trivial solutions of  $(A - \lambda I)\mathbf{x} = \mathbf{0}$  exist only for those  $\lambda$  that solve this equation.

Def'n A pair  $(\lambda, \mathbf{x})$  with  $\mathbf{x} \neq \mathbf{0}$  that solves  $A\mathbf{x} = \lambda\mathbf{x}$  is an *eigenpair*:  $\lambda$  is an *eigenvalue* of  $A$  and  $\mathbf{x}$  is a corresponding *eigenvector*.

Def'n The polynomial  $\det(A - \lambda I)$  is called the *characteristic polynomial* of  $A$ . Its roots are exactly the eigenvalues of  $A$ .

— notes page 54 —

ex/ Find the eigenvalues and the corresponding eigenvectors (the eigenspaces) of

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}.$$

The characteristic polynomial factors as

$$\det(A - \lambda I) = \begin{vmatrix} 2 - \lambda & 2 & 1 \\ 1 & 3 - \lambda & 1 \\ 1 & 2 & 2 - \lambda \end{vmatrix} = -\lambda^3 + 7\lambda^2 - 11\lambda + 5 = -(\lambda - 5)(\lambda - 1)^2.$$

So  $\lambda_1 = 5$  is a simple eigenvalue, and  $\lambda_2 = 1$  is a repeated eigenvalue with multiplicity 2. For each eigenvalue we solve  $(A - \lambda I)\mathbf{x} = \mathbf{0}$ .

For  $\lambda_1 = 5$ :

$$(A - 5I)\mathbf{x} = \begin{bmatrix} -3 & 2 & 1 \\ 1 & -2 & 1 \\ 1 & 2 & -3 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

Gauss elimination gives

$$\begin{bmatrix} -3 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{array}{l} x_1 \text{ \& } x_2 \text{ are pivot variables,} \\ x_3 \text{ is a free variable.} \end{array}$$

Setting  $x_3 = \alpha$  gives  $x_2 = \alpha$  and then  $x_1 = \alpha$ , so  $\mathbf{x} = \begin{bmatrix} \alpha \\ \alpha \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ . The eigenspace of  $\lambda_1 = 5$  is

$$\text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

For  $\lambda_2 = 1$ :

$$(A - I)\mathbf{x} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

Gauss elimination leaves a single non-zero row:

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow x_1 \text{ is a pivot variable, and } x_2 \text{ \& } x_3 \text{ are free variables.}$$

Setting  $x_3 = \alpha$  and  $x_2 = \beta$  gives  $x_1 = -2\beta - \alpha$ , so

$$\mathbf{x} = \begin{bmatrix} -2\beta - \alpha \\ \beta \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

The eigenspace associated with  $\lambda_2 = 1$  is two-dimensional:

$$\text{span} \left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

### Notes

- (i) If  $\mathbf{x}$  is an eigenvector, then so is  $c\mathbf{x}$  for any  $c \neq 0$ . Eigenvectors are only determined up to a scale factor.
- (ii) If  $\lambda$  is an eigenvalue with multiplicity  $k$ , its eigenspace may have a dimension lower than  $k$ . The next example shows this.

— notes page 55 —

ex/ Let  $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ . Then

$$\det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 0 & 1 \\ 1 & 1 - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^3 = 0.$$

So  $\lambda_1 = 1$  is the only eigenvalue, with multiplicity 3. Its eigenspace is the set of solutions of

$$(A - I)\mathbf{x} = \mathbf{0} \Rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Here  $x_1$  and  $x_3$  are pivot variables and  $x_2$  is free. The equations force  $x_1 = 0$  and  $x_3 = 0$ , and  $x_2 = \alpha$  is arbitrary, so  $\mathbf{x} = \begin{bmatrix} 0 \\ \alpha \\ 0 \end{bmatrix} = \alpha \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ . The eigenspace is

$$\text{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\},$$

which is only one-dimensional even though the eigenvalue has multiplicity three.

## A Special Case: Symmetric Matrices

Symmetric matrices have especially nice eigenvalues and eigenvectors. We collect four facts about them, each building on the one before.

### Theorem 11.3.1

If  $A$  is real and symmetric, all its eigenvalues are real.

Proof Let  $(\lambda, \mathbf{x})$  be an eigenpair of  $A$ , possibly with complex entries, and write  $\mathbf{x}^H$  for the conjugate transpose of  $\mathbf{x}$ . Multiply  $A\mathbf{x} = \lambda\mathbf{x}$  on the left by  $\mathbf{x}^H$ . Because  $A$  is real and symmetric,  $A^H = A$ , so  $\mathbf{x}^H(A\mathbf{x}) = (A\mathbf{x})^H\mathbf{x}$ . Therefore

$$\begin{aligned} A\mathbf{x} = \lambda\mathbf{x} &\Rightarrow \mathbf{x}^H(A\mathbf{x}) = \lambda\mathbf{x}^H\mathbf{x} \\ &= (A\mathbf{x})^H\mathbf{x} = (\lambda\mathbf{x})^H\mathbf{x} = \lambda^*\mathbf{x}^H\mathbf{x}. \end{aligned}$$

Comparing the two ends gives

$$\Rightarrow \lambda\mathbf{x}^H\mathbf{x} = \lambda^*\mathbf{x}^H\mathbf{x} \Rightarrow (\lambda - \lambda^*)\mathbf{x}^H\mathbf{x} = 0.$$

Since  $\mathbf{x} \neq \mathbf{0}$ , we have  $\mathbf{x}^H\mathbf{x} > 0$ , so  $\lambda = \lambda^*$ , and  $\lambda$  is real. ■

### Theorem 11.3.2

If  $\lambda$  is an eigenvalue of a real symmetric matrix  $A$  with multiplicity  $k$ , then its eigenspace is  $k$ -dimensional.

Proof This is proved after Theorem 11.4.4.

— notes page 56 —

### Theorem 11.3.3

If  $A$  is real and symmetric, eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof Let  $(\lambda_1, \mathbf{x}_1)$  and  $(\lambda_2, \mathbf{x}_2)$  be eigenpairs of  $A$  with  $\lambda_1 \neq \lambda_2$ . As before, move  $A$  across the product using  $A^H = A$ , and recall that  $\lambda_2$  is real by Theorem 11.3.1, so  $\lambda_2^* = \lambda_2$ :

$$\begin{aligned} A\mathbf{x}_1 = \lambda_1\mathbf{x}_1 &\Rightarrow \mathbf{x}_2^H(A\mathbf{x}_1) = \lambda_1\mathbf{x}_2^H\mathbf{x}_1 \\ &= (A\mathbf{x}_2)^H\mathbf{x}_1 = \lambda_2^*\mathbf{x}_2^H\mathbf{x}_1 = \lambda_2\mathbf{x}_2^H\mathbf{x}_1. \end{aligned}$$

Comparing the ends,

$$\Rightarrow \lambda_1 \mathbf{x}_2^H \mathbf{x}_1 = \lambda_2 \mathbf{x}_2^H \mathbf{x}_1 \Rightarrow (\lambda_1 - \lambda_2) \mathbf{x}_2^H \mathbf{x}_1 = 0.$$

Since  $\lambda_1 - \lambda_2 \neq 0$ , we must have  $\mathbf{x}_2^H \mathbf{x}_1 = 0$ . ■

### Theorem 11.3.4

If  $A_{n \times n}$  is real and symmetric, then its eigenvectors provide an orthogonal basis for  $\mathbb{R}^n$ .

Proof There are two cases.

- (i) If  $A$  has  $n$  distinct eigenvalues  $\lambda_1, \dots, \lambda_n$ , their eigenvectors are orthogonal by Theorem 11.3.3, so they form an orthogonal basis.
- (ii) If  $A$  has repeated eigenvalues, say distinct eigenvalues  $\lambda_1, \dots, \lambda_m$  where  $\lambda_i$  has multiplicity  $k_i$ , then each eigenspace is  $k_i$ -dimensional by Theorem 11.3.2. So each eigenspace has an orthogonal basis  $\{\mathbf{u}_{i1}, \dots, \mathbf{u}_{ik_i}\}$  (for example by Gram-Schmidt). Eigenspaces belonging to different eigenvalues are orthogonal to each other by Theorem 11.3.3, so putting all these bases together gives an orthogonal basis of  $\mathbb{R}^n$ :

$$\bigcup_{i=1}^m \{\mathbf{u}_{i1}, \dots, \mathbf{u}_{ik_i}\}.$$

There are exactly  $n$  vectors in total, because  $\sum_{i=1}^m k_i = n$ . ■

— notes page 57 —

## Diagonalization

A diagonal matrix is the easiest kind of matrix to work with. Diagonalization tries to turn a given matrix into one by changing coordinates. If we can find a matrix  $Q$  such that

### Diagonalization

$$Q^{-1}AQ = D \quad \leftarrow \text{diagonal}$$

then we say that  $A$  is *diagonalizable* and that  $Q$  *diagonalizes*  $A$ . The next theorem says which matrices are diagonalizable and what  $Q$  and  $D$  are.

### Theorem 11.4.1

Let  $A$  be an  $n \times n$  matrix.

### Theorem 11.4.1

- (i)  $A$  is diagonalizable if and only if  $A$  has  $n$  linearly independent eigenvectors.
- (ii) If  $A$  has  $n$  linearly independent eigenvectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $Q = [\mathbf{e}_1 \ \dots \ \mathbf{e}_n]$ , then  $Q^{-1}AQ = D$  is diagonal with  $d_i = \lambda_i$ , the eigenvalue belonging to  $\mathbf{e}_i$ .

Proof (i) We show the two directions separately.

( $\Rightarrow$ ) Suppose  $A$  is diagonalizable, so  $Q^{-1}AQ = D$  for some  $Q = [\mathbf{e}_1 \ \dots \ \mathbf{e}_n]$ . Multiply by  $Q$  on the left and compare columns:

$$Q^{-1}AQ = D \Rightarrow AQ = QD = [\mathbf{e}_1 \ \dots \ \mathbf{e}_n] \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}$$

$$\Rightarrow A[\mathbf{e}_1 \cdots \mathbf{e}_n] = [d_1\mathbf{e}_1 \ d_2\mathbf{e}_2 \ \cdots \ d_n\mathbf{e}_n].$$

So  $A\mathbf{e}_i = d_i\mathbf{e}_i$  for every  $i$ . No column of an invertible matrix is zero, so  $\mathbf{e}_i \neq \mathbf{0}$ , and therefore  $\mathbf{e}_i$  is an eigenvector with eigenvalue  $d_i$ . Since  $Q$  is invertible, its columns  $\mathbf{e}_i$  are linearly independent. This gives  $n$  linearly independent eigenvectors. ■

( $\Leftarrow$ ) Suppose  $A$  has  $n$  linearly independent eigenvectors  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ . Put them in the columns of  $Q$  and the eigenvalues on the diagonal of  $\Lambda$ . The statements  $A\mathbf{e}_i = \lambda_i\mathbf{e}_i$  for all  $i$  together say

$$A[\mathbf{e}_1 \cdots \mathbf{e}_n] = \underbrace{[\mathbf{e}_1 \cdots \mathbf{e}_n]}_Q \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_\Lambda.$$

The columns of  $Q$  are linearly independent, so  $Q$  is invertible, and

$$\Rightarrow Q^{-1}AQ = \Lambda \leftarrow \text{diagonal} \Rightarrow A \text{ is diagonalizable.} \quad \blacksquare$$

— notes page 58 —

### Theorem 11.4.2

When are there enough eigenvectors? One easy case is when all the eigenvalues are different.

#### Theorem 11.4.2

If  $A_{n \times n}$  has distinct eigenvalues  $\lambda_1, \dots, \lambda_n$ , then the corresponding eigenvectors are linearly independent.

Proof Suppose  $\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n = \mathbf{0}$ . We will show that this forces  $\alpha_1 = \cdots = \alpha_n = 0$ , which means that  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  is LI. Written with a column of ones, the assumption says

$$[\alpha_1\mathbf{e}_1 \ \cdots \ \alpha_n\mathbf{e}_n] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{0}.$$

Applying  $A$  to the assumption multiplies each term by its own eigenvalue,  $A(\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n) = \lambda_1\alpha_1\mathbf{e}_1 + \cdots + \lambda_n\alpha_n\mathbf{e}_n = \mathbf{0}$ , and applying  $A$  again and again gives, for every power up to  $n-1$ ,

$$A^{n-1}(\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n) = \lambda_1^{n-1}\alpha_1\mathbf{e}_1 + \cdots + \lambda_n^{n-1}\alpha_n\mathbf{e}_n = \mathbf{0}.$$

Collecting all of these equations into one matrix equation,

$$[\alpha_1\mathbf{e}_1 \ \cdots \ \alpha_n\mathbf{e}_n] \underbrace{\begin{bmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \cdots & \lambda_n^{n-1} \end{bmatrix}}_{\text{Vandermonde matrix}} = [\mathbf{0} \ \mathbf{0} \ \cdots \ \mathbf{0}] = O_{n \times n}.$$

A Vandermonde matrix has a non-zero determinant when the  $\lambda_i$  are all different (a standard fact about these matrices), so it is invertible. Multiplying on the right by its inverse gives

$$\Rightarrow [\alpha_1\mathbf{e}_1 \ \cdots \ \alpha_n\mathbf{e}_n] = O.$$

Each eigenvector is non-zero,  $\mathbf{e}_i \neq \mathbf{0}$ , so each  $\alpha_i = 0$ . Hence  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  is LI. ■

### Theorem 11.4.3

#### Theorem 11.4.3

If  $A_{n \times n}$  has  $n$  distinct eigenvalues, it is diagonalizable.

Proof  $n$  distinct eigenvalues give  $n$  linearly independent eigenvectors (Theorem 11.4.2), and  $n$  linearly independent eigenvectors make  $A$  diagonalizable (Theorem 11.4.1). ■

### Theorem 11.4.4

Symmetric matrices are always diagonalizable, even when eigenvalues repeat, and moreover the matrix  $Q$  can be chosen orthogonal.

#### Theorem 11.4.4

Every real symmetric matrix  $A$  is diagonalizable, by an orthogonal matrix  $Q$  (that is,  $Q^T Q = I$ ).

Proof We give a proof that does not use Theorem 11.3.2; this is what makes the proof of Theorem 11.3.2 below legitimate.

— notes page 59 —

The proof is by induction on the size  $n$  of the matrix.

For a  $1 \times 1$  symmetric matrix,  $A = [a_{11}]$  is already diagonal, so it is certainly diagonalizable (with  $Q = 1$ ).

Now assume that every symmetric matrix of size  $(n-1) \times (n-1)$  is diagonalizable in the following way:

$$Q^T A Q = D \quad \leftarrow \text{diagonal} \quad \text{where } Q^T Q = I \text{ or } Q^{-1} = Q^T.$$

We show that every  $n \times n$  symmetric matrix  $A_{n \times n}$  then also has an orthogonal  $Q_{n \times n}$  that diagonalizes it.

Let  $(\lambda, \mathbf{x})$  be an eigenpair of  $A_{n \times n}$ . The eigenvalue  $\lambda$  is real by Theorem 11.3.1, and we choose the eigenvector real, with length one:  $\mathbf{x}^T \mathbf{x} = 1$ . Extend  $\mathbf{x}$  to an orthonormal basis  $\{\mathbf{y}_1, \dots, \mathbf{y}_{n-1}, \mathbf{x}\}$  of  $\mathbb{R}^n$ , and collect the new vectors in the  $n \times (n-1)$  matrix  $Y = [\mathbf{y}_1 \cdots \mathbf{y}_{n-1}]$ . It has the following properties:

- (i)  $Y^T Y = I_{n-1}$ , since  $\mathbf{y}_i^T \mathbf{y}_j = \delta_{ij}$ .
- (ii)  $\mathbf{x}^T Y = \mathbf{0}^T$ , since  $\mathbf{x}^T \mathbf{y}_i = 0$ .
- (iii)  $Y^T A Y$  is symmetric, because  $A$  is symmetric.

So  $Y^T A Y$  is a symmetric matrix of size  $(n-1) \times (n-1)$ . By the induction hypothesis there is an orthogonal  $Q_{(n-1) \times (n-1)}$  that diagonalizes it:

$$Q_{(n-1) \times (n-1)}^T Y^T A Y Q_{(n-1) \times (n-1)} = D. \quad (\star 1)$$

Now define  $Q_{n \times n} = [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}]$ . We show that it is orthogonal and that it diagonalizes  $A$ . First, orthogonality:

— notes page 60 —

$$Q_{n \times n}^T Q_{n \times n} = [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}]^T [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}]$$

$$\begin{aligned}
 &= \begin{bmatrix} \mathbf{x}^T \\ Q_{(n-1) \times (n-1)}^T Y^T \end{bmatrix} [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}] \\
 &= \begin{bmatrix} \mathbf{x}^T \mathbf{x} & \mathbf{x}^T Y Q_{(n-1) \times (n-1)} \\ Q_{(n-1) \times (n-1)}^T Y^T \mathbf{x} & Q_{(n-1) \times (n-1)}^T Y^T Y Q_{(n-1) \times (n-1)} \end{bmatrix}.
 \end{aligned}$$

The four blocks are:

$$\begin{aligned}
 &> \mathbf{x}^T \mathbf{x} = 1. \\
 &> \mathbf{x}^T Y Q_{(n-1) \times (n-1)} = \mathbf{0}^T, \text{ since } \mathbf{x}^T Y = \mathbf{0}^T. \\
 &> Q_{(n-1) \times (n-1)}^T Y^T \mathbf{x} = \mathbf{0}, \text{ since } Y^T \mathbf{x} = \mathbf{0}. \\
 &> Q_{(n-1) \times (n-1)}^T Y^T Y Q_{(n-1) \times (n-1)} = Q_{(n-1) \times (n-1)}^T \underbrace{(Y^T Y)}_{I_{n-1}} Q_{(n-1) \times (n-1)} = Q_{(n-1) \times (n-1)}^T Q_{(n-1) \times (n-1)} = I_{n-1}.
 \end{aligned}$$

So

$$\Rightarrow Q_{n \times n}^T Q_{n \times n} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & I_{(n-1) \times (n-1)} \end{bmatrix} = I_{n \times n}.$$

Second, we compute  $Q^T A Q$ :

$$\begin{aligned}
 Q_{n \times n}^T A_{n \times n} Q_{n \times n} &= [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}]^T A_{n \times n} [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}] \\
 &= \begin{bmatrix} \mathbf{x}^T A \mathbf{x} & \mathbf{x}^T A Y Q_{(n-1) \times (n-1)} \\ (\mathbf{x}^T A Y Q_{(n-1) \times (n-1)})^T & (Y Q_{(n-1) \times (n-1)})^T A Y Q_{(n-1) \times (n-1)} \end{bmatrix}.
 \end{aligned}$$

Its blocks are:

$$\begin{aligned}
 &> \mathbf{x}^T A \mathbf{x} = \mathbf{x}^T \lambda \mathbf{x} = \lambda. \\
 &> \underbrace{\mathbf{x}^T A}_{\lambda \mathbf{x}^T} Y Q_{(n-1) \times (n-1)} = \lambda \underbrace{\mathbf{x}^T Y}_{\mathbf{0}^T} Q_{(n-1) \times (n-1)} = \mathbf{0}^T. \\
 &> Q_{(n-1) \times (n-1)}^T (Y^T A Y) Q_{(n-1) \times (n-1)} = D, \text{ by equation } (\star 1).
 \end{aligned}$$

Therefore

$$\Rightarrow Q_{n \times n}^T A_{n \times n} Q_{n \times n} = \begin{bmatrix} \lambda & \mathbf{0}^T \\ \mathbf{0} & D \end{bmatrix} \text{ is diagonal, } Q_{n \times n}^T Q_{n \times n} = I.$$

Thus  $A_{n \times n}$  is diagonalizable by an orthogonal matrix. ■

— notes page 61 —

### Revisit Theorem 11.3.2

Now that we know every real symmetric matrix is diagonalizable, we can prove the fact that was postponed earlier.

#### Theorem 11.3.2

If  $\lambda$  is an eigenvalue of a real symmetric matrix  $A$  with multiplicity  $k$ , then its eigenspace is  $k$ -dimensional.

proof By Theorem 11.4.4, a real symmetric  $A_{n \times n}$  is diagonalizable by an orthogonal matrix:

$$Q^T A Q = Q^{-1} A Q = D.$$

Since  $Q^T(A - \lambda I)Q = D - \lambda I$  and  $\det Q \det Q^T = 1$ , the characteristic polynomial is  $\det(A - \lambda I) = \prod_{i=1}^n (d_i - \lambda)$ . So an eigenvalue of multiplicity  $k$  appears exactly  $k$  times on the diagonal of  $D$ . The  $k$  matching columns of  $Q$  are eigenvectors (because  $AQ = QD$ ) and they are linearly independent, so the eigenspace has dimension at least  $k$ . It cannot be larger: eigenspaces of different eigenvalues are linearly independent of each other, and the multiplicities add up to  $n$ , so the dimensions cannot add up to more than  $n$ . Hence the eigenspace has dimension exactly  $k$ . ■

## Closing Note

Even if  $A$  is not diagonalizable, it can still be brought to its Jordan form (over  $\mathbb{C}$ ), which is upper triangular, with the eigenvalues on the diagonal and only 0's and 1's directly above the diagonal:

$$P^{-1}AP = \underbrace{J}_{\text{Jordan form for } A}.$$

The textbook discusses this on page 582 (exercise 11.4.10).

## Cayley–Hamilton Theorem

A matrix satisfies its own characteristic equation. Let  $A_{n \times n}$  have the characteristic equation

$$\lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1} \lambda + \alpha_n = 0.$$

### Cayley–Hamilton theorem

Replacing  $\lambda$  by  $A$  (and the constant term by  $\alpha_n I$ ) gives the zero matrix:

$$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I = O.$$

— notes page 62 —

proof We prove the theorem for the case where  $A$  has a full set of eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  (the theorem is true in general, but this case shows the idea). Collect the eigenvectors in  $V$  and the eigenvalues on the diagonal of  $\Lambda$ :

$$A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}}_V \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}}_{\Lambda}.$$

Some of the  $\lambda_i$  may be equal, but the columns of  $V$  are linearly independent, so  $V$  is invertible, and

$$V^{-1}AV = \Lambda.$$

Powers of  $A$  are easy to follow through this change of coordinates, because the  $V$  and  $V^{-1}$  in the middle cancel:

$$(V^{-1}AV)^2 = V^{-1}A \underbrace{VV^{-1}}_I AV = V^{-1}A^2V = \Lambda^2$$

and in general

$$(V^{-1}AV)^k = V^{-1}A^kV = \Lambda^k = \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix}.$$

Each eigenvalue  $\lambda_i$  satisfies the characteristic equation, so the diagonal matrix of those polynomial values is zero:

$$\begin{bmatrix} \lambda_1^n + \alpha_1 \lambda_1^{n-1} + \cdots + \alpha_{n-1} \lambda_1 + \alpha_n & & \\ & \ddots & \\ & & \lambda_n^n + \alpha_1 \lambda_n^{n-1} + \cdots + \alpha_{n-1} \lambda_n + \alpha_n \end{bmatrix} = O.$$

This diagonal matrix is the same as the polynomial evaluated at  $\Lambda$ . Rewriting each power of  $\Lambda$  as  $V^{-1}A^kV$  gives

$$\begin{aligned} &= \Lambda^n + \alpha_1 \Lambda^{n-1} + \cdots + \alpha_{n-1} \Lambda + \alpha_n I = O \\ &\Rightarrow V^{-1}A^nV + \alpha_1 V^{-1}A^{n-1}V + \cdots + \alpha_{n-1} V^{-1}AV + \alpha_n V^{-1}IV = O \\ &\Rightarrow V^{-1}(A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I)V = O. \end{aligned}$$

Multiplying on the left by  $V$  and on the right by  $V^{-1}$  removes the outer factors:

$$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I = O. \quad \blacksquare$$

— notes page 63 —

Cayley–Hamilton is useful because it lets us reduce high powers of a matrix to low ones.

ex/ Let  $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$ . Its characteristic equation is

$$\begin{vmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 + 3\lambda + 2 = 0, \quad \lambda_1 = -1, \quad \lambda_2 = -2.$$

We can check the theorem directly:

$$A^2 = \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix}, \quad A^2 + 3A + 2I = \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix} + \begin{bmatrix} 0 & 3 \\ -6 & -9 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = O.$$

This says  $A^2 = -3A - 2I$ . So  $A^2$  is a combination of  $A$  and  $I$ , and every higher power can be reduced in the same way:

$$A^k = \beta_1 A + \beta_0 I \quad \text{where } \beta_1 \text{ and } \beta_0 \text{ depend on } k.$$

For an  $n \times n$  matrix in general,  $A^k = \sum_{i=0}^{n-1} \beta_i A^i$ .

How do we find the coefficients  $\beta_i$ ? Let  $\lambda_j$  be an eigenvalue of  $A$  with eigenvector  $\mathbf{v}_j$ . Applying both sides of  $A^k = \sum_i \beta_i A^i$  to  $\mathbf{v}_j$ , and using  $A^i \mathbf{v}_j = \lambda_j^i \mathbf{v}_j$ ,

$$\underbrace{A^k \mathbf{v}_j}_{\lambda_j^k \mathbf{v}_j} = \left( \sum_{i=0}^{n-1} \beta_i A^i \right) \mathbf{v}_j \quad \Rightarrow \quad \lambda_j^k \mathbf{v}_j = \left( \sum_{i=0}^{n-1} \beta_i \lambda_j^i \right) \mathbf{v}_j.$$

Since  $\mathbf{v}_j \neq \mathbf{0}$ , the scalars must agree. So every eigenvalue gives one equation for the unknown  $\beta_i$ :

**Finding the coefficients  $\beta_i$**

$$\lambda_j^k = \sum_{i=0}^{n-1} \beta_i \lambda_j^i \quad \text{for each eigenvalue } \lambda_j.$$

ex/ For the same matrix, compute  $A^5$ . Write  $A^5 = \beta_1 A + \beta_0 I$  with  $\lambda_1 = -1$  and  $\lambda_2 = -2$ . The two equations for  $\beta_0, \beta_1$  are

$$\begin{aligned} \lambda_1^5 &= \beta_1 \lambda_1 + \beta_0 \\ \lambda_2^5 &= \beta_1 \lambda_2 + \beta_0 \end{aligned} \Leftrightarrow \begin{bmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \lambda_1^5 \\ \lambda_2^5 \end{bmatrix}.$$

Plugging in the numbers and solving,

$$\begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} -1 \\ -32 \end{bmatrix} \Rightarrow \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = - \begin{bmatrix} -2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -32 \end{bmatrix} = \begin{bmatrix} 30 \\ 31 \end{bmatrix}.$$

Hence

$$A^5 = 31A + 30I = \begin{bmatrix} 0 & 31 \\ -62 & -93 \end{bmatrix} + \begin{bmatrix} 30 & 0 \\ 0 & 30 \end{bmatrix} = \begin{bmatrix} 30 & 31 \\ -62 & -63 \end{bmatrix}.$$

— notes page 64 —

ex/ A repeated eigenvalue. Let  $A = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$ , which has  $\lambda_1 = \lambda_2 = -1$ . Find  $A^{10}$ . Write  $A^{10} = \beta_1 A + \beta_0 I$ . The eigenvalue equation gives one condition:

$$\lambda_1^{10} = \beta_1 \lambda_1 + \beta_0 \Rightarrow 1 = -\beta_1 + \beta_0.$$

Since  $\lambda_2 = \lambda_1$  this gives no second independent equation for  $\beta_1$  and  $\beta_0$ . But  $\lambda_1$  is a double root of the characteristic polynomial, so differentiating both sides of  $\lambda^{10} = \beta_1 \lambda + \beta_0$  with respect to  $\lambda$  also holds at  $\lambda_1$ , and that gives the missing equation:

$$\left. \frac{d}{d\lambda} \lambda^{10} \right|_{\lambda=\lambda_1} = \left. \frac{d}{d\lambda} (\beta_1 \lambda + \beta_0) \right|_{\lambda=\lambda_1} \Rightarrow 10\lambda_1^9 = \beta_1 \Rightarrow \beta_1 = -10.$$

Together,

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ -10 \end{bmatrix} \Rightarrow \beta_1 = -10, \quad \beta_0 = -9.$$

Hence

$$A^{10} = -10A - 9I = \begin{bmatrix} 0 & -10 \\ 10 & 20 \end{bmatrix} - \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} -9 & -10 \\ 10 & 11 \end{bmatrix}.$$

The same idea computes the matrix exponential  $e^{At}$ , which is important for differential equations later on.

ex/ Find  $e^{At}$  for  $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ . Since  $A$  is  $2 \times 2$  with  $\lambda_1 = \lambda_2 = 1$ , write  $e^{At} = \beta_1 A + \beta_0 I$ . The eigenvalue condition and its derivative with respect to  $\lambda$  at  $\lambda = 1$  give

$$e^{\lambda_1 t} = \beta_1 \lambda_1 + \beta_0 \Rightarrow e^t = \beta_1 + \beta_0, \quad \left. \frac{d}{d\lambda} e^{\lambda t} \right|_{\lambda=1} = \left. \frac{d}{d\lambda} (\beta_1 \lambda + \beta_0) \right|_{\lambda=1} \Rightarrow te^t = \beta_1.$$

So  $\beta_1 = te^t$  and  $\beta_0 = (1 - t)e^t$ , and

$$e^{At} = te^t \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} + (1 - t)e^t \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} e^t & -te^t \\ 0 & e^t \end{bmatrix}.$$

Alternatively, if  $A$  is diagonalizable, any polynomial in  $A$  can be computed from the eigenvalues. Let  $f(A) = \sum_{r=0}^p \beta_r A^r$  and  $Q^{-1}AQ = \Lambda$ , where  $\Lambda$  has the eigenvalues on its diagonal. Then

$$Q^{-1}f(A)Q = \sum_{r=0}^p \beta_r Q^{-1}A^r Q.$$

— notes page 65 —

Because  $Q^{-1}A^rQ = (Q^{-1}AQ)^r = \Lambda^r$ , the change of coordinates carries every power of  $A$  to the same power of  $\Lambda$ . So

$$\Rightarrow Q^{-1}f(A)Q = \sum_{r=0}^p \beta_r \Lambda^r, \quad \text{since } \underbrace{(Q^{-1}AQ)^r}_{\Lambda^r} = Q^{-1}A^rQ.$$

The right-hand side is diagonal, and its entries are the polynomial evaluated at each eigenvalue:

$$\Rightarrow Q^{-1}f(A)Q = \begin{bmatrix} \sum_{r=0}^p \beta_r \lambda_1^r & & \\ & \ddots & \\ & & \sum_{r=0}^p \beta_r \lambda_n^r \end{bmatrix} = \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix}.$$

Solving for  $f(A)$  gives a recipe for applying a function to a diagonalizable matrix: apply it to the eigenvalues, then change coordinates back.

#### A function of a matrix

$$f(A) = Q \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} Q^{-1}.$$

ex/ Calculate  $e^{At}$  for  $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$ , which has  $\lambda_1 = -1$  and  $\lambda_2 = -2$ . This time we also need the eigenvectors. Solving  $(A - \lambda_i I)\mathbf{e}_i = \mathbf{0}$  gives

$$\begin{aligned} (A - \lambda_1 I)\mathbf{e}_1 = \mathbf{0} &\Rightarrow \mathbf{e}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} & \begin{cases} Q = [\mathbf{e}_1 \ \mathbf{e}_2] = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \\ Q^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \end{cases} \\ (A - \lambda_2 I)\mathbf{e}_2 = \mathbf{0} &\Rightarrow \mathbf{e}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \end{aligned}$$

Applying  $f(\lambda) = e^{\lambda t}$  to the eigenvalues, we get

$$e^{At} = Q \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-2t} \end{bmatrix} Q^{-1} = \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{-2t} \end{bmatrix}.$$