

**Exercise 5.4.17** If  $A$  is  $m \times n$  of rank  $r$ , show that  $A$  can be factored as  $A = PQ$  where  $P$  is  $m \times r$  with  $r$  independent columns, and  $Q$  is  $r \times n$  with  $r$  independent rows. [Hint: Let  $UAV = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$  by Theorem 2.5.3, and write  $U^{-1} = \begin{bmatrix} U_1 & U_2 \\ U_3 & U_4 \end{bmatrix}$  and  $V^{-1} = \begin{bmatrix} V_1 & V_2 \\ V_3 & V_4 \end{bmatrix}$  in block form, where  $U_1$  and  $V_1$  are  $r \times r$ .]

**Exercise 5.4.18**

- a. Show that if  $A$  and  $B$  have independent columns, so does  $AB$ .

- b. Show that if  $A$  and  $B$  have independent rows, so does  $AB$ .

**Exercise 5.4.19** A matrix obtained from  $A$  by deleting rows and columns is called a **submatrix** of  $A$ . If  $A$  has an invertible  $k \times k$  submatrix, show that  $\text{rank } A \geq k$ . [Hint: Show that row and column operations carry  $A \rightarrow \begin{bmatrix} I_k & P \\ 0 & Q \end{bmatrix}$  in block form.] Remark: It can be shown that  $\text{rank } A$  is the largest integer  $r$  such that  $A$  has an invertible  $r \times r$  submatrix.

## 5.5 Similarity and Diagonalization

In Section 3.3 we studied diagonalization of a square matrix  $A$ , and found important applications (for example to linear dynamical systems). We can now utilize the concepts of subspace, basis, and dimension to clarify the diagonalization process, reveal some new results, and prove some theorems which could not be demonstrated in Section 3.3.

Before proceeding, we introduce a notion that simplifies the discussion of diagonalization, and is used throughout the book.

### Similar Matrices

#### Definition 5.11 Similar Matrices

If  $A$  and  $B$  are  $n \times n$  matrices, we say that  $A$  and  $B$  are **similar**, and write  $A \sim B$ , if  $B = P^{-1}AP$  for some invertible matrix  $P$ .

Note that  $A \sim B$  if and only if  $B = QAQ^{-1}$  where  $Q$  is invertible (write  $P^{-1} = Q$ ). The language of similarity is used throughout linear algebra. For example, a matrix  $A$  is diagonalizable if and only if it is similar to a diagonal matrix.

If  $A \sim B$ , then necessarily  $B \sim A$ . To see why, suppose that  $B = P^{-1}AP$ . Then  $A = PBP^{-1} = Q^{-1}BQ$  where  $Q = P^{-1}$  is invertible. This proves the second of the following properties of similarity (the others are left as an exercise):

1.  $A \sim A$  for all square matrices  $A$ .
  2. If  $A \sim B$ , then  $B \sim A$ .
  3. If  $A \sim B$  and  $B \sim A$ , then  $A \sim C$ .
- (5.2)

These properties are often expressed by saying that the similarity relation  $\sim$  is an **equivalence relation** on the set of  $n \times n$  matrices. Here is an example showing how these properties are used.

**Example 5.5.1**

If  $A$  is similar to  $B$  and either  $A$  or  $B$  is diagonalizable, show that the other is also diagonalizable.

**Solution.** We have  $A \sim B$ . Suppose that  $A$  is diagonalizable, say  $A \sim D$  where  $D$  is diagonal. Since  $B \sim A$  by (2) of (5.2), we have  $B \sim A$  and  $A \sim D$ . Hence  $B \sim D$  by (3) of (5.2), so  $B$  is diagonalizable too. An analogous argument works if we assume instead that  $B$  is diagonalizable.

Similarity is compatible with inverses, transposes, and powers:

$$\text{If } A \sim B \text{ then } A^{-1} \sim B^{-1}, \quad A^T \sim B^T, \quad \text{and } A^k \sim B^k \text{ for all integers } k \geq 1.$$

The proofs are routine matrix computations using Theorem 3.3.1. Thus, for example, if  $A$  is diagonalizable, so also are  $A^T$ ,  $A^{-1}$  (if it exists), and  $A^k$  (for each  $k \geq 1$ ). Indeed, if  $A \sim D$  where  $D$  is a diagonal matrix, we obtain  $A^T \sim D^T$ ,  $A^{-1} \sim D^{-1}$ , and  $A^k \sim D^k$ , and each of the matrices  $D^T$ ,  $D^{-1}$ , and  $D^k$  is diagonal.

We pause to introduce a simple matrix function that will be referred to later.

**Definition 5.12 Trace of a Matrix**

The **trace**  $\text{tr } A$  of an  $n \times n$  matrix  $A$  is defined to be the sum of the main diagonal elements of  $A$ .

In other words:

$$\text{If } A = [a_{ij}], \text{ then } \text{tr } A = a_{11} + a_{22} + \cdots + a_{nn}.$$

It is evident that  $\text{tr}(A+B) = \text{tr } A + \text{tr } B$  and that  $\text{tr}(cA) = c \text{tr } A$  holds for all  $n \times n$  matrices  $A$  and  $B$  and all scalars  $c$ . The following fact is more surprising.

**Lemma 5.5.1**

Let  $A$  and  $B$  be  $n \times n$  matrices. Then  $\text{tr}(AB) = \text{tr}(BA)$ .

**Proof.** Write  $A = [a_{ij}]$  and  $B = [b_{ij}]$ . For each  $i$ , the  $(i, i)$ -entry  $d_i$  of the matrix  $AB$  is given as follows:  $d_i = a_{i1}b_{1i} + a_{i2}b_{2i} + \cdots + a_{in}b_{ni} = \sum_j a_{ij}b_{ji}$ . Hence

$$\text{tr}(AB) = d_1 + d_2 + \cdots + d_n = \sum_i d_i = \sum_i \left( \sum_j a_{ij}b_{ji} \right)$$

Similarly we have  $\text{tr}(BA) = \sum_i (\sum_j b_{ij}a_{ji})$ . Since these two double sums are the same, Lemma 5.5.1 is proved.  $\square$

As the name indicates, similar matrices share many properties, some of which are collected in the next theorem for reference.

**Theorem 5.5.1**

If  $A$  and  $B$  are similar  $n \times n$  matrices, then  $A$  and  $B$  have the same determinant, rank, trace, characteristic polynomial, and eigenvalues.

**Proof.** Let  $B = P^{-1}AP$  for some invertible matrix  $P$ . Then we have

$$\det B = \det(P^{-1}) \det A \det P = \det A \text{ because } \det(P^{-1}) = 1/\det P$$

Similarly,  $\text{rank } B = \text{rank}(P^{-1}AP) = \text{rank } A$  by Corollary 5.4.3. Next Lemma 5.5.1 gives

$$\text{tr}(P^{-1}AP) = \text{tr}[P^{-1}(AP)] = \text{tr}[(AP)P^{-1}] = \text{tr } A$$

As to the characteristic polynomial,

$$\begin{aligned} c_B(x) &= \det(xI - B) = \det\{x(P^{-1}IP) - P^{-1}AP\} \\ &= \det\{P^{-1}(xI - A)P\} \\ &= \det(xI - A) \\ &= c_A(x) \end{aligned}$$

Finally, this shows that  $A$  and  $B$  have the same eigenvalues because the eigenvalues of a matrix are the roots of its characteristic polynomial.  $\square$

**Example 5.5.2**

Sharing the five properties in Theorem 5.5.1 does not guarantee that two matrices are similar. The matrices  $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  and  $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  have the same determinant, rank, trace, characteristic polynomial, and eigenvalues, but they are not similar because  $P^{-1}IP = I$  for any invertible matrix  $P$ .

**Diagonalization Revisited**

Recall that a square matrix  $A$  is **diagonalizable** if there exists an invertible matrix  $P$  such that  $P^{-1}AP = D$  is a diagonal matrix, that is if  $A$  is similar to a diagonal matrix  $D$ . Unfortunately, not all matrices are diagonalizable, for example  $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  (see Example 3.3.10). Determining whether  $A$  is diagonalizable is closely related to the eigenvalues and eigenvectors of  $A$ . Recall that a number  $\lambda$  is called an **eigenvalue** of  $A$  if  $A\mathbf{x} = \lambda\mathbf{x}$  for some nonzero column  $\mathbf{x}$  in  $\mathbb{R}^n$ , and any such nonzero vector  $\mathbf{x}$  is called an **eigenvector** of  $A$  corresponding to  $\lambda$  (or simply a  $\lambda$ -eigenvector of  $A$ ). The eigenvalues and eigenvectors of  $A$  are closely related to the **characteristic polynomial**  $c_A(x)$  of  $A$ , defined by

$$c_A(x) = \det(xI - A)$$

If  $A$  is  $n \times n$  this is a polynomial of degree  $n$ , and its relationship to the eigenvalues is given in the following theorem (a repeat of Theorem 3.3.2).

**Theorem 5.5.2**

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

1. The eigenvalues  $\lambda$  of  $A$  are the roots of the characteristic polynomial  $c_A(x)$  of  $A$ .
2. The  $\lambda$ -eigenvectors  $\mathbf{x}$  are the nonzero solutions to the homogeneous system

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

of linear equations with  $\lambda I - A$  as coefficient matrix.

**Example 5.5.3**

Show that the eigenvalues of a triangular matrix are the main diagonal entries.

**Solution.** Assume that  $A$  is triangular. Then the matrix  $xI - A$  is also triangular and has diagonal entries  $(x - a_{11}), (x - a_{22}), \dots, (x - a_{nn})$  where  $A = [a_{ij}]$ . Hence Theorem 3.1.4 gives

$$c_A(x) = (x - a_{11})(x - a_{22}) \cdots (x - a_{nn})$$

and the result follows because the eigenvalues are the roots of  $c_A(x)$ .

Theorem 3.3.4 asserts (in part) that an  $n \times n$  matrix  $A$  is diagonalizable if and only if it has  $n$  eigenvectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  such that the matrix  $P = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$  with the  $\mathbf{x}_i$  as columns is invertible. This is equivalent to requiring that  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  is a basis of  $\mathbb{R}^n$  consisting of eigenvectors of  $A$ . Hence we can restate Theorem 3.3.4 as follows:

**Theorem 5.5.3**

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

1.  $A$  is diagonalizable if and only if  $\mathbb{R}^n$  has a basis  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  consisting of eigenvectors of  $A$ .
2. When this is the case, the matrix  $P = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$  is invertible and  $P^{-1}AP = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  where, for each  $i$ ,  $\lambda_i$  is the eigenvalue of  $A$  corresponding to  $\mathbf{x}_i$ .

The next result is a basic tool for determining when a matrix is diagonalizable. It reveals an important connection between eigenvalues and linear independence: Eigenvectors corresponding to distinct eigenvalues are necessarily linearly independent.

**Theorem 5.5.4**

Let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  be eigenvectors corresponding to distinct eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_k$  of an  $n \times n$  matrix  $A$ . Then  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$  is a linearly independent set.

**Proof.** We use induction on  $k$ . If  $k = 1$ , then  $\{\mathbf{x}_1\}$  is independent because  $\mathbf{x}_1 \neq \mathbf{0}$ . In general, suppose the theorem is true for some  $k \geq 1$ . Given eigenvectors  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k+1}\}$ , suppose a linear combination vanishes:

$$t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_{k+1}\mathbf{x}_{k+1} = \mathbf{0} \quad (5.3)$$

We must show that each  $t_i = 0$ . Left multiply (5.3) by  $A$  and use the fact that  $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$  to get

$$t_1\lambda_1\mathbf{x}_1 + t_2\lambda_2\mathbf{x}_2 + \cdots + t_{k+1}\lambda_{k+1}\mathbf{x}_{k+1} = \mathbf{0} \quad (5.4)$$

If we multiply (5.3) by  $\lambda_1$  and subtract the result from (5.4), the first terms cancel and we obtain

$$t_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + t_3(\lambda_3 - \lambda_1)\mathbf{x}_3 + \cdots + t_{k+1}(\lambda_{k+1} - \lambda_1)\mathbf{x}_{k+1} = \mathbf{0}$$

Since  $\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{k+1}$  correspond to distinct eigenvalues  $\lambda_2, \lambda_3, \dots, \lambda_{k+1}$ , the set  $\{\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{k+1}\}$  is independent by the induction hypothesis. Hence,

$$t_2(\lambda_2 - \lambda_1) = 0, \quad t_3(\lambda_3 - \lambda_1) = 0, \quad \dots, \quad t_{k+1}(\lambda_{k+1} - \lambda_1) = 0$$

and so  $t_2 = t_3 = \cdots = t_{k+1} = 0$  because the  $\lambda_i$  are distinct. Hence (5.3) becomes  $t_1\mathbf{x}_1 = \mathbf{0}$ , which implies that  $t_1 = 0$  because  $\mathbf{x}_1 \neq \mathbf{0}$ . This is what we wanted.  $\square$

Theorem 5.5.4 will be applied several times; we begin by using it to give a useful condition for when a matrix is diagonalizable.

### Theorem 5.5.5

*If  $A$  is an  $n \times n$  matrix with  $n$  distinct eigenvalues, then  $A$  is diagonalizable.*

**Proof.** Choose one eigenvector for each of the  $n$  distinct eigenvalues. Then these eigenvectors are independent by Theorem 5.5.4, and so are a basis of  $\mathbb{R}^n$  by Theorem 5.2.7. Now use Theorem 5.5.3.  $\square$

### Example 5.5.4

Show that  $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ -1 & 1 & 0 \end{bmatrix}$  is diagonalizable.

**Solution.** A routine computation shows that  $c_A(x) = (x-1)(x-3)(x+1)$  and so has distinct eigenvalues 1, 3, and  $-1$ . Hence Theorem 5.5.5 applies.

However, a matrix can have multiple eigenvalues as we saw in Section 3.3. To deal with this situation, we prove an important lemma which formalizes a technique that is basic to diagonalization, and which will be used three times below.

**Lemma 5.5.2**

Let  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$  be a linearly independent set of eigenvectors of an  $n \times n$  matrix  $A$ , extend it to a basis  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_n\}$  of  $\mathbb{R}^n$ , and let

$$P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

be the (invertible)  $n \times n$  matrix with the  $\mathbf{x}_i$  as its columns. If  $\lambda_1, \lambda_2, \dots, \lambda_k$  are the (not necessarily distinct) eigenvalues of  $A$  corresponding to  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  respectively, then  $P^{-1}AP$  has block form

$$P^{-1}AP = \begin{bmatrix} \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k) & B \\ 0 & A_1 \end{bmatrix}$$

where  $B$  has size  $k \times (n - k)$  and  $A_1$  has size  $(n - k) \times (n - k)$ .

**Proof.** If  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  is the standard basis of  $\mathbb{R}^n$ , then

$$\begin{aligned} \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} &= I_n = P^{-1}P = P^{-1} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \\ &= \begin{bmatrix} P^{-1}\mathbf{x}_1 & P^{-1}\mathbf{x}_2 & \cdots & P^{-1}\mathbf{x}_n \end{bmatrix} \end{aligned}$$

Comparing columns, we have  $P^{-1}\mathbf{x}_i = \mathbf{e}_i$  for each  $1 \leq i \leq n$ . On the other hand, observe that

$$P^{-1}AP = P^{-1}A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} (P^{-1}A)\mathbf{x}_1 & (P^{-1}A)\mathbf{x}_2 & \cdots & (P^{-1}A)\mathbf{x}_n \end{bmatrix}$$

Hence, if  $1 \leq i \leq k$ , column  $i$  of  $P^{-1}AP$  is

$$(P^{-1}A)\mathbf{x}_i = P^{-1}(\lambda_i\mathbf{x}_i) = \lambda_i(P^{-1}\mathbf{x}_i) = \lambda_i\mathbf{e}_i$$

This describes the first  $k$  columns of  $P^{-1}AP$ , and Lemma 5.5.2 follows.  $\square$

Note that Lemma 5.5.2 (with  $k = n$ ) shows that an  $n \times n$  matrix  $A$  is diagonalizable if  $\mathbb{R}^n$  has a basis of eigenvectors of  $A$ , as in (1) of Theorem 5.5.3.

**Definition 5.13 Eigenspace of a Matrix**

If  $\lambda$  is an eigenvalue of an  $n \times n$  matrix  $A$ , define the **eigenspace** of  $A$  corresponding to  $\lambda$  by

$$E_\lambda(A) = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = \lambda\mathbf{x}\}$$

This is a subspace of  $\mathbb{R}^n$  and the eigenvectors corresponding to  $\lambda$  are just the nonzero vectors in  $E_\lambda(A)$ . In fact  $E_\lambda(A)$  is the null space of the matrix  $(\lambda I - A)$ :

$$E_\lambda(A) = \{\mathbf{x} \mid (\lambda I - A)\mathbf{x} = \mathbf{0}\} = \text{null}(\lambda I - A)$$

Hence, by Theorem 5.4.2, the basic solutions of the homogeneous system  $(\lambda I - A)\mathbf{x} = \mathbf{0}$  given by the gaussian algorithm form a basis for  $E_\lambda(A)$ . In particular

$$\dim E_\lambda(A) \text{ is the number of basic solutions } \mathbf{x} \text{ of } (\lambda I - A)\mathbf{x} = \mathbf{0} \quad (5.5)$$

Now recall (Definition 3.7) that the **multiplicity**<sup>11</sup> of an eigenvalue  $\lambda$  of  $A$  is the number of times  $\lambda$  occurs as a root of the characteristic polynomial  $c_A(x)$  of  $A$ . In other words, the multiplicity of  $\lambda$  is the largest integer  $m \geq 1$  such that

$$c_A(x) = (x - \lambda)^m g(x)$$

for some polynomial  $g(x)$ . Because of (5.5), the assertion (without proof) in Theorem 3.3.5 can be stated as follows: A square matrix is diagonalizable if and only if the multiplicity of each eigenvalue  $\lambda$  equals  $\dim [E_\lambda(A)]$ . We are going to prove this, and the proof requires the following result which is valid for *any* square matrix, diagonalizable or not.

### Lemma 5.5.3

Let  $\lambda$  be an eigenvalue of multiplicity  $m$  of a square matrix  $A$ . Then  $\dim [E_\lambda(A)] \leq m$ .

**Proof.** Write  $\dim [E_\lambda(A)] = d$ . It suffices to show that  $c_A(x) = (x - \lambda)^d g(x)$  for some polynomial  $g(x)$ , because  $m$  is the highest power of  $(x - \lambda)$  that divides  $c_A(x)$ . To this end, let  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d\}$  be a basis of  $E_\lambda(A)$ . Then Lemma 5.5.2 shows that an invertible  $n \times n$  matrix  $P$  exists such that

$$P^{-1}AP = \begin{bmatrix} \lambda I_d & B \\ 0 & A_1 \end{bmatrix}$$

in block form, where  $I_d$  denotes the  $d \times d$  identity matrix. Now write  $A' = P^{-1}AP$  and observe that  $c_{A'}(x) = c_A(x)$  by Theorem 5.5.1. But Theorem 3.1.5 gives

$$\begin{aligned} c_A(x) = c_{A'}(x) &= \det(xI_n - A') = \det \begin{bmatrix} (x - \lambda)I_d & -B \\ 0 & xI_{n-d} - A_1 \end{bmatrix} \\ &= \det [(x - \lambda)I_d] \det [(xI_{n-d} - A_1)] \\ &= (x - \lambda)^d g(x) \end{aligned}$$

where  $g(x) = c_{A_1}(x)$ . This is what we wanted. □

It is impossible to ignore the question when equality holds in Lemma 5.5.3 for each eigenvalue  $\lambda$ . It turns out that this characterizes the diagonalizable  $n \times n$  matrices  $A$  for which  $c_A(x)$  **factors completely** over  $\mathbb{R}$ . By this we mean that  $c_A(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$ , where the  $\lambda_i$  are *real* numbers (not necessarily distinct); in other words, every eigenvalue of  $A$  is real. This need not happen (consider  $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ ), and we investigate the general case below.

### Theorem 5.5.6

The following are equivalent for a square matrix  $A$  for which  $c_A(x)$  factors completely.

1.  $A$  is diagonalizable.
2.  $\dim [E_\lambda(A)]$  equals the multiplicity of  $\lambda$  for every eigenvalue  $\lambda$  of the matrix  $A$ .

<sup>11</sup>This is often called the *algebraic* multiplicity of  $\lambda$ .

**Proof.** Let  $A$  be  $n \times n$  and let  $\lambda_1, \lambda_2, \dots, \lambda_k$  be the distinct eigenvalues of  $A$ . For each  $i$ , let  $m_i$  denote the multiplicity of  $\lambda_i$  and write  $d_i = \dim [E_{\lambda_i}(A)]$ . Then

$$c_A(x) = (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \dots (x - \lambda_k)^{m_k}$$

so  $m_1 + \dots + m_k = n$  because  $c_A(x)$  has degree  $n$ . Moreover,  $d_i \leq m_i$  for each  $i$  by Lemma 5.5.3.

(1)  $\Rightarrow$  (2). By (1),  $\mathbb{R}^n$  has a basis of  $n$  eigenvectors of  $A$ , so let  $t_i$  of them lie in  $E_{\lambda_i}(A)$  for each  $i$ . Since the subspace spanned by these  $t_i$  eigenvectors has dimension  $t_i$ , we have  $t_i \leq d_i$  for each  $i$  by Theorem 5.2.4. Hence

$$n = t_1 + \dots + t_k \leq d_1 + \dots + d_k \leq m_1 + \dots + m_k = n$$

It follows that  $d_1 + \dots + d_k = m_1 + \dots + m_k$  so, since  $d_i \leq m_i$  for each  $i$ , we must have  $d_i = m_i$ . This is (2).

(2)  $\Rightarrow$  (1). Let  $B_i$  denote a basis of  $E_{\lambda_i}(A)$  for each  $i$ , and let  $B = B_1 \cup \dots \cup B_k$ . Since each  $B_i$  contains  $m_i$  vectors by (2), and since the  $B_i$  are pairwise disjoint (the  $\lambda_i$  are distinct), it follows that  $B$  contains  $n$  vectors. So it suffices to show that  $B$  is linearly independent (then  $B$  is a basis of  $\mathbb{R}^n$ ). Suppose a linear combination of the vectors in  $B$  vanishes, and let  $\mathbf{y}_i$  denote the sum of all terms that come from  $B_i$ . Then  $\mathbf{y}_i$  lies in  $E_{\lambda_i}(A)$ , so the nonzero  $\mathbf{y}_i$  are independent by Theorem 5.5.4 (as the  $\lambda_i$  are distinct). Since the sum of the  $\mathbf{y}_i$  is zero, it follows that  $\mathbf{y}_i = \mathbf{0}$  for each  $i$ . Hence all coefficients of terms in  $\mathbf{y}_i$  are zero (because  $B_i$  is independent). Since this holds for each  $i$ , it shows that  $B$  is independent.  $\square$

### Example 5.5.5

If  $A = \begin{bmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{bmatrix}$  and  $B = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & -2 \\ -1 & 0 & -2 \end{bmatrix}$  show that  $A$  is diagonalizable but  $B$  is not.

**Solution.** We have  $c_A(x) = (x+3)^2(x-1)$  so the eigenvalues are  $\lambda_1 = -3$  and  $\lambda_2 = 1$ . The corresponding eigenspaces are  $E_{\lambda_1}(A) = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$  and  $E_{\lambda_2}(A) = \text{span}\{\mathbf{x}_3\}$  where

$$\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

as the reader can verify. Since  $\{\mathbf{x}_1, \mathbf{x}_2\}$  is independent, we have  $\dim(E_{\lambda_1}(A)) = 2$  which is the multiplicity of  $\lambda_1$ . Similarly,  $\dim(E_{\lambda_2}(A)) = 1$  equals the multiplicity of  $\lambda_2$ . Hence  $A$  is diagonalizable by Theorem 5.5.6, and a diagonalizing matrix is  $P = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3]$ .

Turning to  $B$ ,  $c_B(x) = (x+1)^2(x-3)$  so the eigenvalues are  $\lambda_1 = -1$  and  $\lambda_2 = 3$ . The corresponding eigenspaces are  $E_{\lambda_1}(B) = \text{span}\{\mathbf{y}_1\}$  and  $E_{\lambda_2}(B) = \text{span}\{\mathbf{y}_2\}$  where

$$\mathbf{y}_1 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{y}_2 = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix}$$

Here  $\dim(E_{\lambda_1}(B)) = 1$  is *smaller* than the multiplicity of  $\lambda_1$ , so the matrix  $B$  is *not* diagonalizable, again by Theorem 5.5.6. The fact that  $\dim(E_{\lambda_1}(B)) = 1$  means that there is no possibility of finding *three* linearly independent eigenvectors.



## Complex Eigenvalues

All the matrices we have considered have had real eigenvalues. But this need not be the case: The matrix  $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$  has characteristic polynomial  $c_A(x) = x^2 + 1$  which has no real roots. Nonetheless, this matrix is diagonalizable; the only difference is that we must use a larger set of scalars, the complex numbers. The basic properties of these numbers are outlined in Appendix A.

Indeed, nearly everything we have done for real matrices can be done for complex matrices. The methods are the same; the only difference is that the arithmetic is carried out with complex numbers rather than real ones. For example, the gaussian algorithm works in exactly the same way to solve systems of linear equations with complex coefficients, matrix multiplication is defined the same way, and the matrix inversion algorithm works in the same way.

But the complex numbers are better than the real numbers in one respect: While there are polynomials like  $x^2 + 1$  with real coefficients that have no real root, this problem does not arise with the complex numbers: *Every* nonconstant polynomial with complex coefficients has a complex root, and hence factors completely as a product of linear factors. This fact is known as the fundamental theorem of algebra.<sup>12</sup>

### Example 5.5.6

Diagonalize the matrix  $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ .

**Solution.** The characteristic polynomial of  $A$  is

$$c_A(x) = \det(xI - A) = x^2 + 1 = (x - i)(x + i)$$

where  $i^2 = -1$ . Hence the eigenvalues are  $\lambda_1 = i$  and  $\lambda_2 = -i$ , with corresponding eigenvectors  $\mathbf{x}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$  and  $\mathbf{x}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}$ . Hence  $A$  is diagonalizable by the complex version of Theorem 5.5.5,

and the complex version of Theorem 5.5.3 shows that  $P = [\mathbf{x}_1 \quad \mathbf{x}_2] = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$  is invertible

and  $P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$ . Of course, this can be checked directly.

We shall return to complex linear algebra in Section 8.7.

<sup>12</sup>This was a famous open problem in 1799 when Gauss solved it at the age of 22 in his Ph.D. dissertation.

## Symmetric Matrices<sup>13</sup>

On the other hand, many of the applications of linear algebra involve a real matrix  $A$  and, while  $A$  will have complex eigenvalues by the fundamental theorem of algebra, it is always of interest to know when the eigenvalues are, in fact, real. While this can happen in a variety of ways, it turns out to hold whenever  $A$  is symmetric. This important theorem will be used extensively later. Surprisingly, the theory of *complex* eigenvalues can be used to prove this useful result about *real* eigenvalues.

Let  $\bar{z}$  denote the conjugate of a complex number  $z$ . If  $A$  is a complex matrix, the **conjugate matrix**  $\bar{A}$  is defined to be the matrix obtained from  $A$  by conjugating every entry. Thus, if  $A = [z_{ij}]$ , then  $\bar{A} = [\bar{z}_{ij}]$ . For example,

$$\text{If } A = \begin{bmatrix} -i+2 & 5 \\ i & 3+4i \end{bmatrix} \text{ then } \bar{A} = \begin{bmatrix} i+2 & 5 \\ -i & 3-4i \end{bmatrix}$$

Recall that  $\overline{z+w} = \bar{z} + \bar{w}$  and  $\overline{zw} = \bar{z}\bar{w}$  hold for all complex numbers  $z$  and  $w$ . It follows that if  $A$  and  $B$  are two complex matrices, then

$$\overline{A+B} = \bar{A} + \bar{B}, \quad \overline{AB} = \bar{A}\bar{B} \quad \text{and} \quad \overline{\lambda A} = \bar{\lambda}\bar{A}$$

hold for all complex scalars  $\lambda$ . These facts are used in the proof of the following theorem.

### Theorem 5.5.7

Let  $A$  be a symmetric real matrix. If  $\lambda$  is any complex eigenvalue of  $A$ , then  $\lambda$  is real.<sup>14</sup>

**Proof.** Observe that  $\bar{A} = A$  because  $A$  is real. If  $\lambda$  is an eigenvalue of  $A$ , we show that  $\lambda$  is real by showing that  $\bar{\lambda} = \lambda$ . Let  $\mathbf{x}$  be a (possibly complex) eigenvector corresponding to  $\lambda$ , so that  $\mathbf{x} \neq \mathbf{0}$  and  $A\mathbf{x} = \lambda\mathbf{x}$ . Define  $c = \mathbf{x}^T \bar{\mathbf{x}}$ .

$$\text{If we write } \mathbf{x} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \text{ where the } z_i \text{ are complex numbers, we have}$$

$$c = \mathbf{x}^T \bar{\mathbf{x}} = z_1 \bar{z}_1 + z_2 \bar{z}_2 + \cdots + z_n \bar{z}_n = |\bar{z}_1|^2 + |\bar{z}_2|^2 + \cdots + |\bar{z}_n|^2$$

Thus  $c$  is a real number, and  $c > 0$  because at least one of the  $z_i \neq 0$  (as  $\mathbf{x} \neq \mathbf{0}$ ). We show that  $\bar{\lambda} = \lambda$  by verifying that  $\lambda c = \bar{\lambda} c$ . We have

$$\lambda c = \lambda (\mathbf{x}^T \bar{\mathbf{x}}) = (\lambda \mathbf{x})^T \bar{\mathbf{x}} = (A\mathbf{x})^T \bar{\mathbf{x}} = \mathbf{x}^T A^T \bar{\mathbf{x}}$$

At this point we use the hypothesis that  $A$  is symmetric and real. This means  $A^T = A = \bar{A}$  so we continue the calculation:

<sup>13</sup>This discussion uses complex conjugation and absolute value. These topics are discussed in Appendix A.

<sup>14</sup>This theorem was first proved in 1829 by the great French mathematician Augustin Louis Cauchy (1789–1857).

$$\begin{aligned}
 \lambda c &= \mathbf{x}^T A^T \bar{\mathbf{x}} = \mathbf{x}^T (\overline{A \mathbf{x}}) = \mathbf{x}^T (\overline{A \mathbf{x}}) = \mathbf{x}^T (\overline{\lambda \mathbf{x}}) \\
 &= \mathbf{x}^T (\overline{\lambda} \bar{\mathbf{x}}) \\
 &= \overline{\lambda} \mathbf{x}^T \bar{\mathbf{x}} \\
 &= \overline{\lambda} c
 \end{aligned}$$

as required. □

The technique in the proof of Theorem 5.5.7 will be used again when we return to complex linear algebra in Section 8.7.

### Example 5.5.7

Verify Theorem 5.5.7 for every real, symmetric  $2 \times 2$  matrix  $A$ .

**Solution.** If  $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$  we have  $c_A(x) = x^2 - (a+c)x + (ac - b^2)$ , so the eigenvalues are given by  $\lambda = \frac{1}{2}[(a+c) \pm \sqrt{(a+c)^2 - 4(ac - b^2)}]$ . But here

$$(a+c)^2 - 4(ac - b^2) = (a-c)^2 + 4b^2 \geq 0$$

for any choice of  $a$ ,  $b$ , and  $c$ . Hence, the eigenvalues are real numbers.

## Exercises for 5.5

**Exercise 5.5.1** By computing the trace, determinant, and rank, show that  $A$  and  $B$  are *not* similar in each case.

a.  $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

b.  $A = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$

c.  $A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}, B = \begin{bmatrix} 3 & 0 \\ 1 & -1 \end{bmatrix}$

d.  $A = \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}, B = \begin{bmatrix} 2 & -1 \\ 3 & 2 \end{bmatrix}$

e.  $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ -3 & 6 & -3 \end{bmatrix}$

f.  $A = \begin{bmatrix} 1 & 2 & -3 \\ 1 & -1 & 2 \\ 0 & 3 & -5 \end{bmatrix}, B = \begin{bmatrix} -2 & 1 & 3 \\ 6 & -3 & -9 \\ 0 & 0 & 0 \end{bmatrix}$

**Exercise 5.5.2** Show that  $\begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 0 & 1 & 1 \\ 1 & 1 & 0 & -1 \\ 4 & 3 & 0 & 0 \end{bmatrix}$  and

$\begin{bmatrix} 1 & -1 & 3 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & -1 & 4 & 1 \\ 5 & -1 & -1 & -4 \end{bmatrix}$  are *not* similar.

**Exercise 5.5.3** If  $A \sim B$ , show that:

- a.  $A^T \sim B^T$                       b.  $A^{-1} \sim B^{-1}$   
 c.  $rA \sim rB$  for  $r$  in  $\mathbb{R}$       d.  $A^n \sim B^n$  for  $n \geq 1$

**Exercise 5.5.4** In each case, decide whether the matrix  $A$  is diagonalizable. If so, find  $P$  such that  $P^{-1}AP$  is diagonal.

a. 
$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

b. 
$$\begin{bmatrix} 3 & 0 & 6 \\ 0 & -3 & 0 \\ 5 & 0 & 2 \end{bmatrix}$$

c. 
$$\begin{bmatrix} 3 & 1 & 6 \\ 2 & 1 & 0 \\ -1 & 0 & -3 \end{bmatrix}$$

d. 
$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 2 & 3 & 1 \end{bmatrix}$$

**Exercise 5.5.5** If  $A$  is invertible, show that  $AB$  is similar to  $BA$  for all  $B$ .

**Exercise 5.5.6** Show that the only matrix similar to a scalar matrix  $A = rI$ ,  $r$  in  $\mathbb{R}$ , is  $A$  itself.

**Exercise 5.5.7** Let  $\lambda$  be an eigenvalue of  $A$  with corresponding eigenvector  $\mathbf{x}$ . If  $B = P^{-1}AP$  is similar to  $A$ , show that  $P^{-1}\mathbf{x}$  is an eigenvector of  $B$  corresponding to  $\lambda$ .

**Exercise 5.5.8** If  $A \sim B$  and  $A$  has any of the following properties, show that  $B$  has the same property.

- Idempotent, that is  $A^2 = A$ .
- Nilpotent, that is  $A^k = 0$  for some  $k \geq 1$ .
- Invertible.

**Exercise 5.5.9** Let  $A$  denote an  $n \times n$  upper triangular matrix.

- If all the main diagonal entries of  $A$  are distinct, show that  $A$  is diagonalizable.
- If all the main diagonal entries of  $A$  are equal, show that  $A$  is diagonalizable only if it is *already* diagonal.

- c. Show that  $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$  is diagonalizable but that  $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$  is not diagonalizable.

**Exercise 5.5.10** Let  $A$  be a diagonalizable  $n \times n$  matrix with eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  (including multiplicities). Show that:

- $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$
- $\operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$

**Exercise 5.5.11** Given a polynomial  $p(x) = r_0 + r_1x + \cdots + r_nx^n$  and a square matrix  $A$ , the matrix  $p(A) = r_0I + r_1A + \cdots + r_nA^n$  is called the **evaluation** of  $p(x)$  at  $A$ . Let  $B = P^{-1}AP$ . Show that  $p(B) = P^{-1}p(A)P$  for all polynomials  $p(x)$ .

**Exercise 5.5.12** Let  $P$  be an invertible  $n \times n$  matrix. If  $A$  is any  $n \times n$  matrix, write  $T_P(A) = P^{-1}AP$ . Verify that:

- $T_P(I) = I$
- $T_P(AB) = T_P(A)T_P(B)$
- $T_P(A + B) = T_P(A) + T_P(B)$
- $T_P(rA) = rT_P(A)$
- $T_P(A^k) = [T_P(A)]^k$  for  $k \geq 1$
- If  $A$  is invertible,  $T_P(A^{-1}) = [T_P(A)]^{-1}$ .
- If  $Q$  is invertible,  $T_Q[T_P(A)] = T_{PQ}(A)$ .

**Exercise 5.5.13**

- Show that two diagonalizable matrices are similar if and only if they have the same eigenvalues with the same multiplicities.
- If  $A$  is diagonalizable, show that  $A \sim A^T$ .
- Show that  $A \sim A^T$  if  $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

**Exercise 5.5.14** If  $A$  is  $2 \times 2$  and diagonalizable, show that  $C(A) = \{X \mid XA = AX\}$  has dimension 2 or 4. [Hint: If  $P^{-1}AP = D$ , show that  $X$  is in  $C(A)$  if and only if  $P^{-1}XP$  is in  $C(D)$ .]

**Exercise 5.5.15** If  $A$  is diagonalizable and  $p(x)$  is a polynomial such that  $p(\lambda) = 0$  for all eigenvalues  $\lambda$  of  $A$ , show that  $p(A) = 0$  (see Example 3.3.9). In particular, show  $c_A(A) = 0$ . [Remark:  $c_A(A) = 0$  for all square matrices  $A$ —this is the Cayley-Hamilton theorem, see Theorem 11.1.2.]

**Exercise 5.5.16** Let  $A$  be  $n \times n$  with  $n$  distinct real eigenvalues. If  $AC = CA$ , show that  $C$  is diagonalizable.

**Exercise 5.5.17** Let  $A = \begin{bmatrix} 0 & a & b \\ a & 0 & c \\ b & c & 0 \end{bmatrix}$  and

$$B = \begin{bmatrix} c & a & b \\ a & b & c \\ b & c & a \end{bmatrix}.$$

- Show that  $x^3 - (a^2 + b^2 + c^2)x - 2abc$  has real roots by considering  $A$ .
- Show that  $a^2 + b^2 + c^2 \geq ab + ac + bc$  by considering  $B$ .

**Exercise 5.5.18** Assume the  $2 \times 2$  matrix  $A$  is similar to an upper triangular matrix. If  $\operatorname{tr} A = 0 = \operatorname{tr} A^2$ , show that  $A^2 = 0$ .

**Exercise 5.5.19** Show that  $A$  is similar to  $A^T$  for all  $2 \times 2$  matrices  $A$ . [Hint: Let  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . If  $c = 0$  treat the cases  $b = 0$  and  $b \neq 0$  separately. If  $c \neq 0$ , reduce to the case  $c = 1$  using Exercise 5.5.12(d).]

**Exercise 5.5.20** Refer to Section 3.4 on linear recurrences. Assume that the sequence  $x_0, x_1, x_2, \dots$  satisfies

$$x_{n+k} = r_0x_n + r_1x_{n+1} + \cdots + r_{k-1}x_{n+k-1}$$

for all  $n \geq 0$ . Define

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ r_0 & r_1 & r_2 & \cdots & r_{k-1} \end{bmatrix}, V_n = \begin{bmatrix} x_n \\ x_{n+1} \\ \vdots \\ x_{n+k-1} \end{bmatrix}.$$

Then show that:

- $V_n = A^n V_0$  for all  $n$ .
- $c_A(x) = x^k - r_{k-1}x^{k-1} - \cdots - r_1x - r_0$
- If  $\lambda$  is an eigenvalue of  $A$ , the eigenspace  $E_\lambda$  has dimension 1, and  $\mathbf{x} = (1, \lambda, \lambda^2, \dots, \lambda^{k-1})^T$  is an eigenvector. [Hint: Use  $c_A(\lambda) = 0$  to show that  $E_\lambda = \mathbb{R}\mathbf{x}$ .]
- $A$  is diagonalizable if and only if the eigenvalues of  $A$  are distinct. [Hint: See part (c) and Theorem 5.5.4.]
- If  $\lambda_1, \lambda_2, \dots, \lambda_k$  are distinct real eigenvalues, there exist constants  $t_1, t_2, \dots, t_k$  such that  $x_n = t_1\lambda_1^n + \cdots + t_k\lambda_k^n$  holds for all  $n$ . [Hint: If  $D$  is diagonal with  $\lambda_1, \lambda_2, \dots, \lambda_k$  as the main diagonal entries, show that  $A^n = PD^nP^{-1}$  has entries that are linear combinations of  $\lambda_1^n, \lambda_2^n, \dots, \lambda_k^n$ .]

**Exercise 5.5.21** Suppose  $A$  is  $2 \times 2$  and  $A^2 = 0$ . If  $\operatorname{tr} A \neq 0$  show that  $A = 0$ .

## 5.6 Best Approximation and Least Squares

Often an exact solution to a problem in applied mathematics is difficult to obtain. However, it is usually just as useful to find arbitrarily close approximations to a solution. In particular, finding “linear approximations” is a potent technique in applied mathematics. One basic case is the situation where a system of linear equations has no solution, and it is desirable to find a “best approximation” to a solution to the system. In this section best approximations are defined and a method for finding them is described. The result is then applied to “least squares” approximation of data.

Suppose  $A$  is an  $m \times n$  matrix and  $\mathbf{b}$  is a column in  $\mathbb{R}^m$ , and consider the system

$$A\mathbf{x} = \mathbf{b}$$

of  $m$  linear equations in  $n$  variables. This need not have a solution. However, given any column  $\mathbf{z} \in \mathbb{R}^n$ , the distance  $\|\mathbf{b} - A\mathbf{z}\|$  is a measure of how far  $A\mathbf{z}$  is from  $\mathbf{b}$ . Hence it is natural to ask whether there is a column  $\mathbf{z}$  in  $\mathbb{R}^n$  that is as close as possible to a solution in the sense that

$$\|\mathbf{b} - A\mathbf{z}\|$$