CHAPTER 1

Eigenvalues, eigenvectors, and similarity

1.0 Introduction and notation

†In this and the following chapters, we motivate some key issues discussed in the chapter with examples of how they arise, either conceptually or in application.

Throughout the book, we use (typically without further comment) the notation and terminology introduced in Chapter 0. Readers should consult the index to find a definition of an unfamiliar term; unfamiliar notation can usually be identified by using the Notation table that follows the References.

1.0.1 Change of basis and similarity. Every invertible matrix is a change-of-basis matrix, and every change-of-basis matrix is invertible (0.10). Thus, if $\mathcal{B}$ is a given basis of a vector space V, if T is a given linear transformation on V, and if $A = {}_{\mathcal{B}}[T]_{\mathcal{B}}$ is the $\mathcal{B}$ basis representation of T, the set of all possible basis representations of T is

$$\begin{aligned}&\{{}_{\mathcal{B}_1}[I]_{\mathcal{B}}\ {}_{\mathcal{B}}[T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{B}_1}\ :\ \mathcal{B}_1 \text{ is a basis of } V\}\\ &\quad = \{S^{-1}AS\colon S \in M_n(\mathbf{F}) \text{ is invertible}\}\end{aligned}$$

This is just the set of all matrices that are *similar* to the given matrix A. Similar but not identical matrices are therefore just different basis representations of a single linear transformation.

One would expect similar matrices to share many important properties—at least, those properties that are intrinsic to the underlying linear transformation—and this is an important theme in linear algebra. It is often useful to step back from a question about a given matrix to a question about some intrinsic property of the linear transformation of which the matrix is only one of many possible representations.

The notion of similarity is a key concept in this chapter.

1.0.2 Constrained extrema and eigenvalues. A second key concept in this chapter is the notion of *eigenvector* and *eigenvalue*. Nonzero vectors x such that Ax is a scalar multiple of x play a major role in analyzing the structure of a matrix or linear transformation, but such vectors arise in the more elementary context of maximizing (or minimizing) a real symmetric quadratic form subject to a geometric constraint: For a given real symmetric $A \in M_n(\mathbf{R})$,

$$\text{maximize } x^T Ax, \quad \text{subject to} \quad x \in \mathbf{R}^n, \quad x^T x = 1 \tag{1.0.3}$$

A conventional approach to such a constrained optimization problem is to introduce the Lagrangian $L = x^T Ax - \lambda x^T x$. Necessary conditions for an extremum then are

$$0 = \nabla L = 2(Ax - \lambda x) = 0$$

Thus, if a vector $x \in \mathbf{R}^n$ with $x^T x = 1$ (and hence $x \neq 0$) is an extremum of $x^T Ax$, it must satisfy the equation $Ax = \lambda x$, and hence Ax is a real scalar multiple of x. Such a pair λ, x is called an *eigenvalue–eigenvector pair* for A.

Problems

1. Use Weierstrass's Theorem (see Appendix E) to explain why the constrained extremum problem (1.0.3) has a solution, and conclude that every real symmetric matrix has at least one real eigenvalue. *Hint:* $f(x) = x^T Ax$ is a continuous function on the compact set $\{x \in \mathbf{R}^n : x^T x = 1\}$.

2. Suppose that $A \in M_n(\mathbf{R})$ is symmetric. Show that $\max\{x^T Ax : x \in \mathbf{R}^n, x^T x = 1\}$ is the *largest* real eigenvalue of A.

1.1 The eigenvalue–eigenvector equation

A matrix $A \in M_n$ can be thought of as a linear transformation from $\mathbf{C}^n$ into $\mathbf{C}^n$, namely

$$A : x \to Ax \tag{1.1.1}$$

but it is also useful to think of it as an array of numbers. The interplay between these two concepts of A, and what the array of numbers tells us about the linear transformation, is a central theme of matrix analysis and a key to applications.

A fundamental concept in matrix analysis is the set of *eigenvalues* of a square complex matrix.

1.1.2 Definition. Let $A \in M_n$. If a scalar λ and a nonzero vector x satisfy the equation

$$Ax = \lambda x, \qquad x \in \mathbf{C}^n, x \neq 0, \lambda \in \mathbf{C} \tag{1.1.3}$$

then λ is called an *eigenvalue* of A and x is called an *eigenvector of A associated with λ*. The pair λ, x is an *eigenvalue–eigenvector pair for A*.

The scalar λ and the vector x in the preceding definition occur inextricably as a pair. It is a key element of the definition that *an eigenvector can never be the zero vector*.

Exercise. Consider the diagonal matrix $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$. Explain why the standard basis vectors e_i, $i = 1, \ldots, n$, are eigenvectors of D. With what eigenvalue is each eigenvector e_i associated?

The equation (1.1.3) can be rewritten as $\lambda x - Ax = (\lambda I - A)x = 0$, a square system of homogeneous linear equations. If this system has a nontrivial solution, then λ is an eigenvalue of A and the matrix $\lambda I - A$ is singular. Conversely, if $\lambda \in \mathbf{C}$ and if $\lambda I - A$ is singular, then there is a nonzero vector x such that $(\lambda I - A)\, x = 0$, so $Ax = \lambda x$, that is, λ, x is an eigenvalue–eigenvector pair for A.

1.1.4 Definition. The *spectrum* of $A \in M_n$ is the set of all $\lambda \in \mathbf{C}$ that are eigenvalues of A; we denote this set by $\sigma(A)$.

For a given $A \in M_n$, we do not know at this point whether $\sigma(A)$ is empty, or, if it is not, whether it contains finitely or infinitely many complex numbers.

Exercise. If x is an eigenvector associated with an eigenvalue λ of A, show that any nonzero scalar multiple of x is an eigenvector of A associated with λ.

If x is an eigenvector of $A \in M_n$ associated with λ, it is often convenient to normalize it, that is, to form the unit vector $\xi = x / \|x\|_2$, which is still an eigenvector of A associated with λ. Normalization does not select a unique eigenvector associated with λ, however: $\lambda, e^{i\theta}\xi$ is an eigenvalue-eigenvector pair for A for all $\theta \in \mathbf{R}$.

Exercise. If $Ax = \lambda x$, observe that $\bar{A}\bar{x} = \bar{\lambda}\bar{x}$. Explain why $\sigma(\bar{A}) = \overline{\sigma(A)}$. If $A \in M_n(\mathbf{R})$ and $\lambda \in \sigma(A)$, explain why $\bar{\lambda} \in \sigma(A)$ as well.

Even if they had no other importance, eigenvalues and eigenvectors would be interesting algebraically: according to (1.1.3), the eigenvectors are just those vectors such that multiplication by the matrix A is the same as multiplication by the scalar λ.

Example. Consider the matrix

$$A = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix} \in M_2 \tag{1.1.4a}$$

Then $3 \in \sigma(A)$ and $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ is an associated eigenvector since

$$A \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Also, $5 \in \sigma(A)$. Find an eigenvector associated with the eigenvalue 5.

Sometimes the structure of a matrix makes an eigenvector easy to perceive, so the associated eigenvalue can be computed easily.

Exercise. Let J_n be the n-by-n matrix whose entries are all equal to 1. Consider the n-vector e whose entries are all equal to 1, and let $x_k = e - ne_k$, in which $\{e_1, \ldots, e_n\}$ is the standard basis for $\mathbf{C}^n$. For $n = 2$, show that e and x_1 are linearly independent eigenvectors of J_2 and that 2 and 0, respectively, are the corresponding eigenvalues. For $n = 3$, show that e, x_1, and x_2 are linearly independent eigenvectors of J_3 and that 2, 0, and 0, respectively, are the corresponding eigenvalues. In general, show that $e, x_1, \ldots, x_{n-1}$ are linearly independent eigenvectors of J_n and that $n, 0, \ldots, 0$, respectively, are the corresponding eigenvalues.

Exercise. Show that 1 and 4 are eigenvalues of the matrix

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

Hint: Use eigenvectors. Write $A = 4I - J_3$ and use the preceding exercise.

Evaluation of a polynomial of degree k

$$p(t) = a_k t^k + a_{k-1} t^{k-1} + \cdots + a_1 t + a_0, \quad a_k \neq 0 \tag{1.1.5a}$$

with real or complex coefficients at a matrix $A \in M_n$ is well defined since we may form linear combinations of integral powers of a given square matrix. We define

$$p(A) \equiv a_k A^k + a_{k-1} A^{k-1} + \cdots + a_1 A + a_0 I \tag{1.1.5b}$$

in which we observe the universal convention that $A^0 \equiv I$. A polynomial (1.1.5a) of degree k is said to be *monic* if $a_k = 1$; since $a_k \neq 0$, $a_k^{-1} p(t)$ is always monic. Of course, a monic polynomial cannot be the zero polynomial.

There is an alternative way to represent $p(A)$ that has very important consequences. The *Fundamental Theorem of Algebra* (Appendix C) ensures that any monic polynomial (1.1.5a) of degree $k \geq 1$ can be represented as a product of exactly k complex or real linear factors:

$$p(t) = (t - \alpha_1) \cdots (t - \alpha_k) \tag{1.1.5c}$$

This representation of $p(t)$ is unique up to permutation of its factors. It tells us that $p(\alpha_j) = 0$ for each $j = 1, \ldots, k$, so that each α_j is a *root* of the equation $p(t) = 0$; one also says that each α_j is a *zero* of $p(t)$. Conversely, if β is a complex number such that $p(\beta) = 0$, then $\beta \in \{\alpha_1, \ldots, \alpha_k\}$, so a polynomial of degree $k \geq 1$ has at most k distinct zeroes. In the product (1.1.5c), some factors might be repeated, e.g., $p(t) = t^2 + 2t + 1 = (t + 1)(t + 1)$. The number of times a factor $(t - \alpha_j)$ is repeated is the *multiplicity* of α_j as a zero of $p(t)$. The factorization (1.1.5c) gives a factorization of $p(A)$:

$$p(A) = (A - \alpha_1 I) \cdots (A - \alpha_k I) \tag{1.1.5d}$$

The eigenvalues of $p(A)$ are linked to the eigenvalues of A in a simple way.

1.1.6 Theorem. Let $p(t)$ be a given polynomial of degree k. If λ, x is an eigenvalue–eigenvector pair of $A \in M_n$, then $p(\lambda), x$ is an eigenvalue–eigenvector pair of $p(A)$. Conversely, if $k \geq 1$ and if μ is an eigenvalue of $p(A)$, then there is some eigenvalue λ of A such that $\mu = p(\lambda)$.

Proof: We have

$$p(A)x \equiv a_k A^k x + a_{k-1} A^{k-1} x + \cdots + a_1 A x + a_0 x, \quad a_k \neq 0$$

and $A^j x = A^{j-1} A x = A^{j-1} \lambda x = \lambda A^{j-1} x = \cdots = \lambda^j x$ by repeated application of the eigenvalue–eigenvector equation. Thus,

$$p(A)x = a_k \lambda^k x + \cdots + a_0 x = (a_k \lambda^k + \cdots + a_0) x = p(\lambda) x$$

Conversely, if μ is an eigenvalue of $p(A)$ then $p(A) - \mu I$ is singular. Since $p(t)$ has degree $k \geq 1$, the polynomial $q(t) \equiv p(t) - \mu$ has degree $k \geq 1$ and we can factor it as $q(t) = (t - \beta_1) \cdots (t - \beta_k)$ for some complex or real $\beta_1, \ldots, \beta_k$. Since $p(A) - \mu I = q(A) = (A - \beta_1 I) \cdots (A - \beta_k I)$ is singular, some factor $A - \beta_j I$ is singular, which means that β_j is an eigenvalue of A. But $0 = q(\beta_j) = p(\beta_j) - \mu$, so $\mu = p(\beta_j)$, as claimed. $\square$

Exercise. Suppose $A \in M_n$. If $\sigma(A) = \{-1, 1\}$, what is $\sigma(A^2)$? *Caution*: The first assertion in Theorem 1.1.6 permits you to identify a point in $\sigma(A)$, but you must invoke the second assertion in order find out if it is the *only* point in $\sigma(A)$.

Exercise. Consider $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. What is A^2? Show that e_1 is an eigenvector of A and of A^2, both associated with the eigenvalue $\lambda = 0$. Show that e_2 is an eigenvector of A^2 but *not* of A. Explain why the "converse" part of Theorem 1.1.6 speaks only about eigenvalues of $p(A)$, not eigenvectors. Show that A has no eigenvectors other than scalar multiples of e_1 and explain why $\sigma(A) = \{0\}$.

1.1.7 Observation. A matrix $A \in M_n$ is singular if and only if $0 \in \sigma(A)$.

Proof: The matrix A is singular if and only if $Ax = 0$ for some $x \neq 0$. This happens if and only if $Ax = 0x$ for some $x \neq 0$, that is, if and only if $\lambda = 0$ is an eigenvalue of A. □

1.1.8 Observation. Let $A \in M_n$ and $\lambda, \mu \in \mathbf{C}$ be given. Then $\lambda \in \sigma(A)$ if and only if $\lambda + \mu \in \sigma(A + \mu I)$.

Proof: If $\lambda \in \sigma(A)$ there is a nonzero vector x such that $Ax = \lambda x$ and hence $(A + \mu I)x = Ax + \mu x = \lambda x + \mu x = (\lambda + \mu)x$. Thus, $\lambda + \mu \in \sigma(A + \mu I)$. Conversely, if $\lambda + \mu \in \sigma(A + \mu I)$ there is a nonzero vector y such that $Ay + \mu y = (A + \mu I)y = (\lambda + \mu)y = \lambda y + \mu y$. Thus, $Ay = \lambda y$ and $\lambda \in \sigma(A)$. □

We are now prepared to make a very important observation: *every complex matrix has a nonempty spectrum*, that is, for each $A \in M_n$ there is some scalar $\lambda \in \mathbf{C}$ and some nonzero $x \in \mathbf{C}^n$ such that $Ax = \lambda x$.

1.1.9 Theorem. Let $A \in M_n$ be given. Then A has an eigenvalue. In fact, for each given nonzero $y \in \mathbf{C}^n$ there is a polynomial $g(t)$ of degree at most $n - 1$ such that $g(A)y$ is an eigenvector of A.

Proof: Let m be the *least* integer k such that $\{y, Ay, A^2y, \ldots, A^ky\}$ is linearly dependent. Then $m \geq 1$ since $y \neq 0$, and $m \leq n$ since *any* $n + 1$ vectors in $\mathbf{C}^n$ are linearly dependent. Let $a_0, a_1, \ldots, a_m$ be scalars, not all zero, such that

$$a_m A^m y + a_{m-1} A^{m-1} y + \cdots + a_1 Ay + a_0 y = 0 \tag{1.1.10}$$

If $a_m = 0$, then (1.1.10) implies that $\{y, Ay, A^2y, \ldots, A^{m-1}y\}$ is linearly dependent, contradicting the minimality of m. Thus, $a_m \neq 0$ and we may consider the polynomial $p(t) = t^m + (a_{m-1}/a_m)t^{m-1} + \cdots + (a_1/a_m)t + (a_0/a_m)$. The identity (1.1.10) ensures that $p(A)y = 0$, so $0, y$ is an eigenvalue–eigenvector pair for $p(A)$. Theorem 1.1.6 ensures that one of the m zeroes of $p(t)$ is an eigenvalue of A.

Suppose λ is a zero of $p(t)$ that is an eigenvalue of A and factor $p(t) = (t - \lambda)g(t)$, in which $g(t)$ is a polynomial of degree $m - 1$. If $g(A)y = 0$, the minimality of m would be contradicted again, so $g(A)y \neq 0$. But $0 = p(A)y = (A - \lambda I)(g(A)y)$, so the nonzero vector $g(A)y$ is an eigenvector of A associated with the eigenvalue λ. □

The preceding argument shows that for a given $A \in M_n$ we can find a polynomial of degree at most n such that *at least one* of its zeroes is an eigenvalue of A. In the next section, we introduce a polynomial $p_A(t)$ of degree exactly n such that each of its zeroes is an eigenvalue of A and each eigenvalue of A is a zero of $p_A(t)$, that is, $p_A(\lambda) = 0$ if and only if $\lambda \in \sigma(A)$.

Problems

1. Suppose that $A \in M_n$ is nonsingular. According to (1.1.7), this is equivalent to assuming that $0 \notin \sigma(A)$. For each $\lambda \in \sigma(A)$, show that $\lambda^{-1} \in \sigma(A^{-1})$. If $Ax = \lambda x$ and $x \neq 0$, show that $A^{-1}x = \lambda^{-1}x$.

2. Let $A \in M_n$ be given. (a) Show that the sum of the entries in each row of A is 1 if and only if $1 \in \sigma(A)$ and the vector $e = [1, 1, \ldots, 1]^T$ is an associated eigenvector, that is, $Ae = e$. (b) Suppose that the sum of the entries in each row of A is 1. If A is nonsingular, show that the sum of the entries in each row of A^{-1} is also 1. Moreover, for any given polynomial $p(t)$, show that the sums of the entries in each row $p(A)$ are equal. Equal to what?

3. Let $A \in M_n(\mathbf{R})$. Suppose that λ is a real eigenvalue of A and that $Ax = \lambda x, x \in \mathbf{C}^n, x \neq 0$. Let $x = u + iv$, in which $u, v \in \mathbf{R}^n$ are the respective real and imaginary parts of x. (0.2.5) Show that $Au = \lambda u$ and $Av = \lambda v$. Explain why at least one of u, v must be nonzero and conclude that A has a real eigenvector associated with λ. Must both u and v be eigenvectors of A? Can A have a real eigenvector associated with a non-real eigenvalue?

4. Consider the block diagonal matrix

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \qquad A_{ii} \in M_{n_i}$$

Show that $\sigma(A) = \sigma(A_{11}) \cup \sigma(A_{22})$. There are three things you must show: (a) if λ is an eigenvalue of A, then it is an eigenvalue of either A_{11} or of A_{22}; (b) if λ is an eigenvalue of A_{11}, then it is an eigenvalue of A; and (c) if λ is an eigenvalue of A_{22}, then it is an eigenvalue of A.

5. Let $A \in M_n$ be a given idempotent matrix, so $A^2 = A$. Show that each eigenvalue of A is either 0 or 1. Explain why I is the only nonsingular idempotent matrix.

6. Show that all eigenvalues of a nilpotent matrix are 0. Give an example of a nonzero nilpotent matrix. Explain why 0 is the only nilpotent idempotent matrix.

7. If $A \in M_n$ is Hermitian, show that all eigenvalues of A are *real. Hint:* Let λ, x be an eigenvalue–eigenvector pair of A. Then $x^*Ax = \lambda x^*x$. But $x^*x > 0$ and $\overline{x^*Ax} = x^*A^*x = x^*Ax$ is real.

8. Explain how the argument in Theorem (1.1.9) fails if we try to use it to show that every square real matrix has a real eigenvalue.

9. Use the definition (1.1.3) to show that the real matrix $A = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ has no real eigenvalue. However, Theorem 1.1.9 says that A has a complex eigenvalue. Actually, there are two; what are they?

10. Provide details for the following example, which shows that a linear operator on an infinite dimensional complex vector space might have *no* eigenvalues. Let $V = \{(a_1, a_2, \ldots, a_k, \ldots) : a_i \in \mathbf{C}, i = 1, 2, \ldots\}$ be the vector space of all formal infinite sequences of complex numbers and define the *right shift operator* S on V by $S(a_1, a_2, \ldots) = (0, a_1, a_2, \ldots)$. Verify that S is a linear. If $Sx = \lambda x$, show that $x = 0$.

11. Let $A \in M_n$ and $\lambda \in \sigma(A)$ be given. Then $A - \lambda I$ is singular, so $(A - \lambda I) \operatorname{adj}(A - \lambda I) = (\det(A - \lambda I))I = 0$. [See (0.8.2)] Explain why there is some $y \in \mathbf{C}^n$ ($y = 0$ is possible) such that $A - \lambda I = xy^*$, so every nonzero column of $\operatorname{adj}(A - \lambda I)$ is an eigenvector of A associated with the eigenvalue λ. Why is this observation useful only if $\operatorname{rank}(A - \lambda I) = n - 1$?

12. Suppose that λ is an eigenvalue of $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in M_2$. Use Problem 11 to show that if either column of $\left[\begin{smallmatrix} d-\lambda & -b \\ -c & a-\lambda \end{smallmatrix}\right]$ is nonzero, then it is an eigenvector of A associated with λ. Why must one of the columns be a scalar multiple of the other? Use this method to find eigenvectors of the matrix (1.1.4a) associated with the eigenvalues 3 and 5.

13. Let $A \in M_n$ and $\lambda \in \sigma(A)$ be given. Suppose $x \neq 0$ and $Ax = \lambda x$, so that $(\operatorname{adj} A)Ax = (\operatorname{adj} A)\lambda x$. If $\lambda \neq 0$, explain why $(\operatorname{adj} A)x = (\lambda^{-1} \det A)x$. If $\lambda = 0$, refer to (0.8.2) and explain why there is some $y \in \mathbf{C}^n$ (possibly $y = 0$) such that $\operatorname{adj} A = xy^*$, so $(\operatorname{adj} A)x = (y^*x)x$. In either case, notice that x is an eigenvector of $\operatorname{adj} A$.

1.2 The characteristic polynomial and algebraic multiplicity

Natural questions to ask about the eigenvalues of $A \in M_n$ are: How many are there? How may they be characterized in a systematic way?

Rewrite the eigenvalue-eigenvector equation (1.1.3) as

$$(\lambda I - A)x = 0, \qquad x \neq 0 \tag{1.2.1}$$

Thus, $\lambda \in \sigma(A)$ if and only if $\lambda I - A$ is singular, that is, if and only if

$$\det(\lambda I - A) = 0 \tag{1.2.2}$$

1.2.3 Definition. Thought of as a formal polynomial in t, the *characteristic polynomial* of $A \in M_n$ is

$$p_A(t) \equiv \det(tI - A)$$

We refer to the equation $p_A(t) = 0$ as the *characteristic equation* of A.

1.2.4 Observation. The characteristic polynomial of each $A = [a_{ij}] \in M_n$ has degree n and $p_A(t) = t^n - (\operatorname{tr} A)t^{n-1} + \cdots + (-1)^n \det A$. Moreover, $p_A(\lambda) = 0$ if and only if $\lambda \in \sigma(A)$, so $\sigma(A)$ contains at most n complex numbers.

Proof: Each summand in the presentation (0.3.2.1) of the determinant of $tI - A$ is a product of exactly n entries of $tI - A$, each from a different row and column, so each summand is a polynomial in t of degree at most n. The degree of a summand can be n only if every factor in the product involves t, which happens only for the summand

$$(t - a_{11}) \cdots (t - a_{nn}) = t^n - (a_{11} + \cdots + a_{nn})t^{n-1} + \cdots \tag{1.2.4a}$$

that is the product of the diagonal entries. Any other summand must contain a factor $-a_{ij}$ with $i \neq j$, so the diagonal entries $(t - a_{ii})$ (in the same row as a_{ij}) and $(t - a_{ii})$ (in the same column as a_{ij}) cannot also be factors; this summand therefore cannot have degree larger than $n-2$. Thus, the coefficients of t^n and t^{n-1} in the polynomial $p_A(t)$ arise only from the summand (1.2.4a). The constant term in $p_A(t)$ is just $p_A(0) = \det(0I - A) = \det(-A) = (-1)^n \det A$. The remaining assertion is the equivalence of (1.2.1) and (1.2.2), together with the fact that a polynomial of degree $n \geq 1$ has at most n distinct zeroes. □

Exercise. Show that the roots of $\det(A - tI) = 0$ are the same as those of $\det(tI - A) = 0$, and that $\det(A - tI) = (-1)^n \det(tI - A) = (-1)^n(t^n + \cdots)$

The characteristic polynomial could alternatively be defined as $\det(A - tI) = (-1)^n t^n + \cdots$. The convention we have chosen ensures that the coefficient of t^n in the characteristic polynomial is always $+1$.

Exercise. Let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in M_2$. Show that the characteristic polynomial of A is

$$p_A(t) = t^2 - (a+d)t + (ad - bc) = t^2 - (\operatorname{tr} A)\, t + \det A$$

With $r \equiv (a-d)^2 + 4bc$ and letting $\sqrt{r}$ be a fixed square root of r, deduce that each of

$$\lambda_1 = \tfrac{1}{2}\left(a + d + \sqrt{r}\right) \quad \text{and} \quad \lambda_2 = \tfrac{1}{2}\left(a + d - \sqrt{r}\right) \tag{1.2.4b}$$

is an eigenvalue of A. Verify that $\operatorname{tr} A = \lambda_1 + \lambda_2$ and $\det A = \lambda_1 \lambda_2$. Explain why $\lambda_1 \neq \lambda_2$ if and only if $r \neq 0$. If $A \in M_2(\mathbf{R})$, show that (a) the eigenvalues of A are real if and only if $r \geq 0$; (b) the eigenvalues are real if $bc \geq 0$; and (c) if $r < 0$ then $\lambda_1 = \overline{\lambda_2}$, that is, λ_1 is the complex conjugate of λ_2.

The preceding exercise illustrates that an eigenvalue λ of a matrix $A \in M_n$ with $n > 1$ can be a multiple zero of $p_A(t)$ (equivalently, a multiple root of its characteristic equation). Indeed, the characteristic polynomial of $I \in M_n$ is

$$p_I(t) = \det(tI - I) = \det((t-1)I) = (t-1)^n \det I = (t-1)^n$$

so the eigenvalue $\lambda = 1$ has multiplicity n as a zero of $p_I(t)$. How should we account for such repetitions in an enumeration of the eigenvalues?

For a given $A \in M_n$ with $n > 1$, factor its characteristic polynomial as $p_A(t) = (t - \alpha_1) \cdots (t - \alpha_n)$. We know that each zero α_i of $p_A(t)$ (regardless of its multiplicity) is an eigenvalue of A. A computation reveals that

$$p_A(t) = t^n - (\alpha_1 + \cdots + \alpha_n)t^{n-1} + \cdots + (-1)^n \alpha_1 \cdots \alpha_n \tag{1.2.4c}$$

so a comparison of (1.2.4) and (1.2.4c) tells us that the sum of the zeroes of $p_A(t)$ is the trace of A, and the product of the zeroes of $p_A(t)$ is the determinant of A. If each zero of $p_A(t)$ has multiplicity one, that is, if $\alpha_i \neq \alpha_j$ whenever $i \neq j$, then $\sigma(A) = \{\alpha_1, \ldots, \alpha_n\}$, so $\operatorname{tr} A$ *is the sum of the eigenvalues* of A and $\det A$ *is the product of the eigenvalues* of A. If these two statements are to remain true even if some zeroes of $p_A(t)$ have multiplicity greater than one, we must enumerate the eigenvalues of A according to their multiplicities as roots of the characteristic equation.

1.2.5 Definition. Let $A \in M_n$. The *multiplicity* of an eigenvalue λ of A is its multiplicity as a zero of the characteristic polynomial $p_A(t)$. For clarity, we sometimes refer to the multiplicity of an eigenvalue as its *algebraic multiplicity*.

Henceforth, *the eigenvalues of* $A \in M_n$ will always mean *the eigenvalues together with their respective (algebraic) multiplicities*. Thus, the zeroes of the characteristic polynomial of A (including their multiplicities) are the same as the eigenvalues of A (including their multiplicities):

$$p_A(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n) \tag{1.2.6}$$

in which $\lambda_1, \ldots, \lambda_n$ are the n eigenvalues of A, listed in any order. When we refer to *the distinct eigenvalues* of A, we mean the elements of $\sigma(A)$.

We can now say without qualification that *each matrix $A \in M_n$ has exactly n eigenvalues among the complex numbers*; the trace and determinant of A are the sum and product, respectively, of its eigenvalues. If A is real, some or all of its eigenvalues might not be real.

Exercise. Consider a real matrix $A \in M_n(\mathbf{R})$. Explain why all the coefficients of $p_A(t)$ are real. Suppose A has a non-real eigenvalue λ. Why is $\bar{\lambda}$ also an eigenvalue of A, and why are the algebraic multiplicities of λ and $\bar{\lambda}$ the same? If x, λ is an eigenvector-eigenvalue pair for A, explain why $\bar{x}, \bar{\lambda}$ is also an eigenvector-eigenvalue pair. Notice that x and $\bar{x}$ are eigenvectors of A that are associated with *distinct* eigenvalues λ and $\bar{\lambda}$.

1.2.7 Example. Let $x, y \in \mathbf{C}^n$. What are the eigenvalues and determinant of $I + xy^*$? Using (0.8.5.11) and the fact that $\operatorname{adj}(\alpha I) = \alpha^{n-1} I$, we compute

$$\begin{aligned} p_{I+xy^*}(t) &= \det(tI - (I + xy^*)) = \det((t-1)I - xy^*) \\ &= \det((t-1)I) - y^* \operatorname{adj}((t-1)I)x \\ &= (t-1)^n - (t-1)^{n-1} y^* x = (t-1)^{n-1}(t - (1 + y^* x)) \end{aligned}$$

Thus, the eigenvalues of $I + xy^*$ are $1 + y^*x$ and 1 (with multiplicity $n - 1$), so $\det(I + xy^*) = (1 + y^*x)(1)^{n-1} = 1 + y^*x$.

1.2.8 Example. Let $x, y \in \mathbf{C}^n$, $x \neq 0$, and $A \in M_n$. Suppose that $Ax = \lambda x$ and let the eigenvalues of A be $\lambda, \lambda_2, \ldots, \lambda_n$. What are the eigenvalues of $A + xy^*$? Using (0.8.5.11) again, we compute

$$\begin{aligned} p_{A+xy^*}(t) &= \det(tI - (A + xy^*)) = \det((tI - A) - xy^*) \\ &= \det(tI - A) - y^* \operatorname{adj}(tI - A)x \end{aligned}$$

Multiply both sides by $(t - \lambda)$, use the identity $(t - \lambda) \operatorname{adj}(tI - A)x = \det(tI - A)x$ (Problem 13 in (1.2)), and compute

$$\begin{aligned} (t - \lambda)\, p_{A+xy^*}(t) &= (t - \lambda) \det(tI - A) - y^*(t - \lambda) \operatorname{adj}(tI - A)x \\ &= (t - \lambda) \det(tI - A) - \det(tI - A) y^* x \\ &= (t - \lambda) p_A(t) - p_A(t) y^* x \end{aligned}$$

Thus, we have the polynomial identity

$$(t-\lambda)\, p_{A+xy^*}(t) = (t-(\lambda + y^*x))\, p_A(t)$$

The zeroes of the left-hand side are λ together with the n eigenvalues of $A + xy^*$. The zeroes of the right-hand side are $\lambda + y^*x, \lambda, \lambda_2, \ldots, \lambda_n$. It follows that the eigenvalues of $A + xy^*$ are $\lambda + y^*x, \lambda_2, \ldots, \lambda_n$.

Since we now know that each n-by-n complex matrix has finitely many eigenvalues, we may make the following definition.

1.2.9 Definition. Let $A \in M_n$. The *spectral radius* of A is $\rho(A) \equiv \max\{|\lambda| : \lambda \in \sigma(A)\}$.

Exercise. Explain why every eigenvalue of $A \in M_n$ lies in the closed bounded disk $\{z : z \in \mathbf{C} \text{ and } |z| \le \rho(A)\}$ in the complex plane.

Exercise. Suppose $A \in M_n$ has at least one nonzero eigenvalue. Explain why $\min\{|\lambda| : \lambda \in \sigma(A) \text{ and } \lambda \neq 0\} > 0$.

Exercise. Underlying both of the two preceding exercises is the fact that $\sigma(A)$ is a nonempty finite set. Explain why.

Sometimes the structure of a matrix makes the characteristic polynomial easy to calculate. This is the case for diagonal or triangular matrices.

Exercise. Consider an upper triangular matrix

$$T = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ & \ddots & \vdots \\ 0 & & t_{nn} \end{bmatrix} \in M_n$$

Show that $p_T(t) = (t - t_{11}) \cdots (t - t_{nn})$, so the eigenvalues of T are its diagonal entries $t_{11}, t_{22}, \ldots, t_{nn}$. What if T is lower triangular? What if T is diagonal?

Exercise. Suppose that $A \in M_n$ is block upper triangular

$$A = \begin{bmatrix} A_{11} & & \bigstar \\ & \ddots & \\ \mathbf{0} & & A_{kk} \end{bmatrix}, \quad A_{ii} \in M_{n_i} \text{ for } i = 1, \ldots, n_k$$

Explain why $p_A(t) = p_{A_{11}}(t) \cdots p_{A_{kk}}(t)$ and the eigenvalues of A are the eigenvalues of A_{11}, together with those of $A_{22}, \ldots$, together with those of A_{kk} including all their respective algebraic multiplicities. This observation is the basis of many algorithms to compute eigenvalues. Explain why the preceding exercise is a special case of this one.

1.2.10 Definition. The sum of all the k-by-k principal minors of $A \in M_n$ (there are $\binom{n}{k}$ of them) is denoted by $E_k(A)$.

We have already encountered principal minor sums as two coefficients of the characteristic polynomial

$$p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_2t^2 + a_1t + a_0$$

If $k = 1$, then $\binom{n}{k} = n$ and $E_1(A) = a_{11} + \cdots + a_{nn} = \operatorname{tr} A = -a_{n-1}$; if $k = n$, then $\binom{n}{k} = 1$ and $E_n(A) = \det A = (-1)^n a_0$. The broader connection between coefficients and principal minor sums is a consequence of the fact that the coefficients are explicit functions of certain derivatives of $p_A(t)$ at $t = 0$:

$$a_k = \frac{1}{k!}p_A^{(k)}(0), \quad k = 0, 1, \ldots, n-1 \tag{1.2.11}$$

Use (0.8.10.2) to evaluate the derivative

$$p'_A(t) = \operatorname{tr}\operatorname{adj}(tI - A)$$

Observe that $\operatorname{tr}\operatorname{adj} A$ is the sum of all the principal minors of A of size $n-1$, so $\operatorname{tr}\operatorname{adj} A = E_{n-1}(A)$. Then

$$\begin{aligned} a_1 &= p'_A(t)|_{t=0} = \operatorname{tr}\operatorname{adj}(tI - A)|_{t=0} = \operatorname{tr}\operatorname{adj}(-A) \\ &= (-1)^{n-1}\operatorname{tr}\operatorname{adj}(A) = (-1)^{n-1}E_{n-1}(A) \end{aligned}$$

Now observe that $\operatorname{tr}\operatorname{adj}(tI - A) = \sum_{i=1}^n p_{A_{(i)}}(t)$ is the sum of the characteristic polynomials of the n principal submatrices of A of size $n-1$, which we denote by $A_{(1)}, \ldots, A_{(n)}$. Use (0.8.10.2) again to evaluate

$$p''_A(t) = \frac{d}{dt}\operatorname{tr}\operatorname{adj}(tI - A) = \sum_{i=1}^n \frac{d}{dt}p_{A_{(i)}}(t) = \sum_{i=1}^n \operatorname{tr}\operatorname{adj}(tI - A_{(i)}) \tag{1.2.12}$$

Each summand $\operatorname{tr}\operatorname{adj}(tI - A_{(i)})$ is the sum of the $n-1$ principal minors of size $n-2$ of a principal minor of $tI - A$, so each summand is a sum of certain principal minors of $tI - A$ of size $n-2$. Each of the $\binom{n}{n-2}$ principal minors of $tI - A$ of size $n-2$ appears twice in (1.2.12): the principal minor with rows and columns k and ℓ omitted appears when $i = k$ as well as when $i = \ell$. Thus,

$$\begin{aligned} a_2 &= \frac{1}{2}\, p''_A(t)|_{t=0} = \frac{1}{2}\sum_{i=1}^n \operatorname{tr}\operatorname{adj}(tI - A_{(i)})\big|_{t=0} = \frac{1}{2}\sum_{i=1}^n \operatorname{tr}\operatorname{adj}(-A_{(i)}) \\ &= \frac{1}{2}(-1)^{n-2}\sum_{i=1}^n \operatorname{tr}\operatorname{adj}(A_{(i)}) = \frac{1}{2}(-1)^{n-2}\left(2E_{n-2}(A)\right) \\ &= (-1)^{n-2}E_{n-2}(A) \end{aligned}$$

Repeating this argument reveals that $p_A^{(k)}(0) = k!(-1)^{n-k}E_{n-k}(A)$, $k = 0, 1, \ldots, n-1$, so the coefficients of the characteristic polynomial (1.2.11) are

$$a_k = \frac{1}{k!}p_A^{(k)}(0) = (-1)^{n-k}E_{n-k}(A), \quad k = 0, 1, \ldots, n-1$$

and hence

$$p_A(t) = t^n - E_1(A)t^{n-1} + \cdots + (-1)^{n-1}E_{n-1}t + (-1)^n E_n \quad (1.2.13)$$

With the identity (1.2.6) in mind, we make the following definition:

1.2.14 Definition. The kth *elementary symmetric function* of n complex numbers $\lambda_1, \ldots, \lambda_n$, $k \le n$, is

$$S_k(\lambda_1, \ldots, \lambda_n) \equiv \sum_{1 \le i_1 < \cdots < i_k \le n} \prod_{j=1}^{k} \lambda_{i_j}$$

Notice that the sum has $\binom{n}{k}$ summands. If $A \in M_n$ and $\lambda_1, \ldots, \lambda_n$ are its eigenvalues, we define $S_k(A) \equiv S_k(\lambda_1, \ldots, \lambda_n)$.

Exercise. What are $S_1(\lambda_1, \ldots, \lambda_n)$ and $S_n(\lambda_1, \ldots, \lambda_n)$? Explain why none of the functions $S_k(\lambda_1, \ldots, \lambda_n)$ changes if the list $\lambda_1, \ldots, \lambda_n$ is re-indexed and re-arranged.

A calculation with (1.2.6) reveals that

$$p_A(t) = t^n - S_1(A)t^{n-1} + \cdots + (-1)^{n-1}S_{n-2}(A)t + (-1)^{n-1}S_n \quad (1.2.15)$$

Comparison of (1.2.6) with (1.2.15) gives the following identities between elementary symmetric functions of eigenvalues of a matrix and sums of its principal minors.

1.2.16 Theorem. Let $A \in M_n$. Then

$$S_k(A) = E_k(A), \quad k = 1, \ldots, n$$

The next theorem shows that a singular complex matrix can always be shifted slightly to become nonsingular. This important fact often permits us to use continuity arguments to deduce results about singular matrices from properties of nonsingular matrices.

1.2.17 Theorem. Let $A \in M_n$ be given. Then there is some $\delta > 0$ such that $A + \varepsilon I$ is nonsingular whenever $\varepsilon \in \mathbf{C}$ and $0 < |\varepsilon| < \delta$.

Proof: Observation 1.1.8 ensures that $\lambda \in \sigma(A)$ if and only if $\lambda + \varepsilon \in \sigma(A + \varepsilon I)$. Therefore, $0 \in \sigma(A + \varepsilon I)$ if and only if $\lambda + \varepsilon = 0$ for some

$\lambda \in \sigma(A)$, that is, if and only if $\varepsilon = -\lambda$ for some $\lambda \in \sigma(A)$. If all the eigenvalues of A are zero, take $\delta = 1$. If some eigenvalue of A is nonzero, let $\delta = \min\{|\lambda| : \lambda \in \sigma \text{ and } \lambda \neq 0\}$. If we choose any ε such that $0 < |\varepsilon| < \delta$, we are assured that $-\varepsilon \notin \sigma(A)$, so $0 \notin \sigma(A + \varepsilon I)$ and $A + \varepsilon I$ is nonsingular. $\square$

There is a useful connection between the derivatives of a polynomial $p(t)$ and the multiplicity of its zeroes: α is a zero of $p(t)$ with multiplicity $k \geq 1$ if and only if we can write $p(t)$ in the form

$$p(t) = (t - \alpha)^k q(t)$$

in which $q(t)$ is a polynomial such that $q(\alpha) \neq 0$. Differentiating this identity gives $p'(t) = k(t-\alpha)^{k-1}q(t) + (t-\alpha)^k q'(t)$, which shows that $p'(\alpha) = 0$ if and only if $k > 1$. If $k \geq 2$, then $p''(t) = k(k-1)(t-\alpha)^{k-2}q(t) +$ polynomial terms each involving a factor $(t-\alpha)^m$ with $m \geq k-1$, so $p''(\alpha) = 0$ if and only if $k > 2$. Repetition of this calculation shows that α is a zero of $p(t)$ of multiplicity k if and only if $p(\alpha) = p'(\alpha) = \cdots = p^{(k-1)}(\alpha) = 0$ and $p^{(k)}(\alpha) \neq 0$.

1.2.18 Theorem. Let $A \in M_n$ and suppose $\lambda \in \sigma(A)$ has algebraic multiplicity k. Then $\operatorname{rank}(A - \lambda I) \geq n - k$ with equality for $k = 1$.

Proof: Apply the preceding observation to the characteristic polynomial $p_A(t)$ of a matrix $A \in M_n$ that has an eigenvalue λ with multiplicity $k \geq 1$. If we let $B = A - \lambda I$, then zero is an eigenvalue of B with multiplicity k and hence $p_B^{(k)}(0) \neq 0$. But $p_B^{(k)}(0) = k!(-1)^{n-k}E_{n-k}(B)$, so $E_{n-k}(B) \neq 0$. In particular, *some* principal minor of $B = A - \lambda I$ of size $n - k$ is nonzero, so $\operatorname{rank}(A - \lambda I) \geq n - k$. If $k = 1$ we can say more: $A - \lambda I$ is singular, so $n > \operatorname{rank}(A - \lambda I) \geq n - 1$, which means that $\operatorname{rank}(A - \lambda I) = n - 1$ if the eigenvalue λ has algebraic multiplicity one. $\square$

Problems

1. Let $A \in M_n$. Use the identity $S_n(A) = E_n(A)$ to verify (1.1.7).

2. For matrices $A \in M_{m,n}$ and $B \in M_{n,m}$, show by direct calculation that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. For any $A \in M_n$ and nonsingular $S \in M_n$, deduce that $\operatorname{tr}(S^{-1}AS) = \operatorname{tr} A$. For any $A, B \in M_n$ use multiplicativity of the determinant function to show that $\det(S^{-1}AS) = \det A$, that is, the determinant is a similarity invariant.

3. Let $D \in M_n$ be a diagonal matrix. Compute the characteristic polynomial $p_D(t)$ and show that $p_D(D) = 0$.

4. Suppose $A \in M_n$ is idempotent. Use (1.2.15) and Problem 5 in (1.1) to show that every coefficient of $p_A(t)$ is an integer (positive, negative, or zero).

5. Use Problem 6 in (1.1) to show that the trace of a nilpotent matrix is 0. What is the characteristic polynomial of a nilpotent matrix?

6. If $A \in M_n$ and $\lambda \in \sigma(A)$ has multiplicity 1, we know that $\operatorname{rank}(A - \lambda I) = n-1$. Consider the converse: If $\operatorname{rank}(A-\lambda I) = n-1$, must λ be an eigenvalue of A? Must it have multiplicity 1?

7. Use (1.2.13) to determine the characteristic polynomial of the tridiagonal matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Consider how this procedure could be used to compute the characteristic polynomial of a general n-by-n tridiagonal matrix.

8. Let $A \in M_n$ and $\lambda \in \mathbf{C}$ be given. Suppose the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$. Explain why $p_{A+\lambda I}(t) = p_A(t-\lambda)$ and deduce from this identity that the eigenvalues of $A + \lambda I$ are $\lambda_1 + \lambda, \ldots, \lambda_n + \lambda$.

9. Explicitly compute $S_2(\lambda_1, \ldots, \lambda_6)$, $S_3(\lambda_1, \ldots, \lambda_6)$, $S_4(\lambda_1, \ldots, \lambda_6)$, and $S_5(\lambda_1, \ldots, \lambda_6)$.

10. If $A \in M_n(\mathbf{R})$ and if n is *odd*, show that A has at least one real eigenvalue. *Hint:* Any nonreal complex zeroes of a polynomial with real coefficients occur in conjugate pairs; $p_A(t)$ has real coefficients if $A \in M_n(\mathbf{R})$.

11. Let V be a vector space over a field **F**. An eigenvalue of a linear transformation $T : V \to V$ is a scalar $\lambda \in \mathbf{F}$ such that there is a nonzero vector $v \in V$ with $Tv = \lambda v$. Show that if **F** is the field of complex numbers and if V is finite-dimensional, then every linear transformation T has an eigenvalue. Give examples to show that if either hypothesis is weakened (finite dimensionality of V or $\mathbf{F} = \mathbf{C}$), then T may not have an eigenvalue. *Hint:* Let $\mathcal{B}$ be a basis for V and consider $[T]_{\mathcal{B}}$.

12. Let $x = [x_i]$, $y = [y_i] \in \mathbf{C}^n$, and $a \in \mathbf{C}$ be given and let $A = \left[\begin{smallmatrix} 0_n & x \\ y^* & a \end{smallmatrix}\right] \in M_{n+1}$. Show that $p_A(t) = t^{n-1}(t^2 - at - y^*x)$ in two ways: (a) Use Cauchy's expansion (0.8.5.10) to calculate $\det \left[\begin{smallmatrix} tI_n & -x \\ -y^* & t-a \end{smallmatrix}\right]$. (b) Explain why $\operatorname{rank} A \leq 2$ and use (1.2.13). Why do only $E_1(A)$ and $E_2(A)$ need to be calculated and only principal submatrices of the form $\left[\begin{smallmatrix} 0 & x_i \\ \bar{y}_i & a \end{smallmatrix}\right]$ need to be considered? What are the eigenvalues of A?

13. Let $x, y \in \mathbf{C}^n$, $a \in \mathbf{C}$, and $B \in M_n$. Consider the bordered matrix $A = \left[\begin{smallmatrix} B & x \\ y^* & a \end{smallmatrix}\right] \in M_{n+1}$. (a) Use (0.8.5.10) to show that

$$p_A(t) = (t-a)p_B(t) - y^*(\operatorname{adj}(tI - B))x \tag{1.2.19}$$

(b) If $B = \lambda I_n$, deduce that

$$p_A(t) = (t-\lambda)^{n-1}(t^2 - (a+\lambda)t + a\lambda - y^*x) \tag{1.2.20}$$

and conclude that the eigenvalues of $\left[\begin{smallmatrix} \lambda I_n & x \\ y^* & a \end{smallmatrix}\right]$ are λ with multiplicity $n-1$, together with $\frac{1}{2}(a + \lambda \pm ((a-\lambda)^2 + 4y^*x)^{1/2})$.

14. Let $n \geq 3$, $B \in M_{n-2}$, and $\lambda, \mu \in \mathbf{C}$. Consider the block matrix

$$A = \begin{bmatrix} \lambda & \bigstar & \bigstar \\ 0 & \mu & 0 \\ 0 & \bigstar & B \end{bmatrix}$$

in which the $\bigstar$ entries are not necessarily zero. Show that $p_A(t) = (t-\lambda)(t-\mu)p_B(t)$. *Hint*: Evaluate $\det(tI - A)$ by cofactors along the first column, and then use cofactors along the first row in the next step.

15. Suppose $A(t) \in M_n$ is a given continuous matrix-valued function and each of the vector valued functions $x_1(t), \ldots, x_n(t) \in \mathbf{C}^n$ satisfies the system of ordinary differential equations $x_j'(t) = A(t)x_j(t)$. Let $X(t) = [x_1(t) \ \ldots \ x_n(t)]$ and let $W(t) \equiv \det X(t)$. Use (0.8.10) and (0.8.2.11) and provide details for the following argument:

$$\begin{aligned} W'(t) &= \sum_{j=1}^{n} \det\left(X(t) \xleftarrow[j]{} x_j'(t)\right) = \operatorname{tr}\left[\det\left(X(t) \xleftarrow[i]{} x_j'(t)\right)\right]_{i,j=1}^{n} \\ &= \operatorname{tr}\left((\operatorname{adj} X(t))\, X'(t)\right) = \operatorname{tr}\left((\operatorname{adj} X(t))\, A(t)X(t)\right) = W(t) \operatorname{tr} A(t) \end{aligned}$$

Thus, $W(t)$ satisfies the scalar differential equation $W'(t) = \operatorname{tr} A(t)W(t)$, whose solution is *Abel's formula* for the *Wronskian*

$$W(t) = W(t_0)e^{\int_{t_0}^{t} \operatorname{tr} A(s)\, ds}$$

Conclude that if the vectors $x_1(t), \ldots, x_n(t)$ are linearly independent for $t = t_0$, then they are linearly independent for all t. How did you use the identity $\operatorname{tr} BC = \operatorname{tr} CB$ (Problem 2)?

16. Let $A \in M_n$ and $x, y \in \mathbf{C}^n$ be given. Let $f(t) = \det(A + txy^T)$. Use (0.8.5.11) to show that $f(t) = \det A + \beta t$, a linear function of t. What is

β? For any $t_1 \neq t_2$, show that $\det A = (t_2 f(t_1) - t_1 f(t_2))/(t_2 - t_1)$. Now consider

$$A = \begin{bmatrix} d_1 & b & \cdots & b \\ c & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ c & \cdots & c & d_n \end{bmatrix} \in M_n$$

$x = y = e$ (all ones), $t_1 = b$, and $t_2 = c$. Let $q(t) = (d_1 - t)\cdots(d_n - t)$. Show that $\det A = (bq(c) - cq(b))/(b - c)$ if $b \neq c$, and $\det A = q(b) - bq'(b)$ if $b = c$. If $d_1 = \cdots = d_n = 0$, show that $p_A(t) = (b(t+c)^n - c(t+b)^n)/(b-c)$ if $b \neq c$, and $p_A(t) = (t+b)^{n-1}(t-(n-1)b)$ if $b = c$.

17. Let $A, B \in M_n$ and let $C = \left[\begin{smallmatrix} 0_n & A \\ B & 0_n \end{smallmatrix}\right]$. Use (0.8.5.13-14) to show that $p_C(t) = p_{AB}(t^2) = p_{BA}(t^2)$, and explain carefully why this implies that AB and BA have the same eigenvalues. Explain why this confirms that $\operatorname{tr} AB = \operatorname{tr} BA$ and $\det AB = \det BA$. Also explain why $\det(I+AB) = \det(I+BA)$.

18. Let $A \in M_3$. Explain why $p_A(t) = t^3 - (\operatorname{tr} A)t^2 + (\operatorname{tr}\operatorname{adj} A)t - \det A$.

19. Suppose all the entries of $A = [a_{ij}] \in M_n$ are either zero or one, and suppose all the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are positive real numbers. Explain why $\det A$ is a positive integer, and provide details for the following:

$$\begin{aligned} n &\geq \operatorname{tr} A = \frac{1}{n}(\lambda_1 + \cdots + \lambda_n)n \geq n(\lambda_1 \cdots \lambda_n)^{1/n} \\ &= n(\det A)^{1/n} \geq n \end{aligned}$$

Conclude that all $\lambda_i = 1$, all $a_{ii} = 1$, and $\det A = 1$.

20. For any $A \in M_n$, show that $\det(I + A) = 1 + E_1(A) + \cdots + E_n(A)$.

21. Let $A \in M_n$ and nonzero vectors $x, y \in \mathbb{C}^n$ be given. Suppose that $c \in \mathbb{C}$, $y^*x = 1$, $Ax = \lambda x$, and the eigenvalues of A are $\lambda, \lambda_2, \ldots, \lambda_n$. Show that the eigenvalues of the *Google matrix* $A(c) = cA + (1-c)\lambda xy^*$ are $\lambda, c\lambda_2, \ldots, c\lambda_n$. *Hint:* Example 1.2.8.

1.3 Similarity

We know that a similarity transformation of a matrix in M_n corresponds to representing its underlying linear transformation on $\mathbf{C}^n$ in another basis. Thus, studying similarity can be thought of as studying properties that are intrinsic to one linear transformation, or the properties that are common to all its basis representations.

1.3.1 Definition. Let $A, B \in M_n$ be given. We say that B is *similar* to A if there exists a nonsingular $S \in M_n$ such that

$$B = S^{-1}AS$$

The transformation $A \to S^{-1}AS$ is called a *similarity transformation* by the *similarity matrix* S. We say that B is *permutation similar* to A if there is a permutation matrix P such that $B = P^TAP$. The relation "B is similar to A" is sometimes abbreviated $B \sim A$.

1.3.2 Observation. Similarity is an *equivalence relation* on M_n; that is, similarity is *reflexive*, *symmetric*, and *transitive*; see (0.11).

Like any equivalence relation, similarity partitions the set M_n into disjoint equivalence classes. Each equivalence class is the set of all matrices in M_n similar to a given matrix, a representative of the class. All matrices in an equivalence class are similar, and matrices in different classes are not similar. The crucial observation is that matrices in a similarity class share many important properties. Some of these are mentioned here; a complete description of the *similarity invariants* (e.g., Jordan canonical form) is in Chapter 3.

1.3.3 Theorem. Let $A, B \in M_n$. If B is similar to A, then A and B have the same characteristic polynomial.

Proof: Compute

$$\begin{aligned} p_B(t) &= \det(tI - B) \\ &= \det(tS^{-1}S - S^{-1}AS) = \det(S^{-1}(tI - A)S) \\ &= \det S^{-1} \det(tI - A) \det S = (\det S)^{-1}(\det S)\det(tI - A) \\ &= \det(tI - A) = p_A(t) \end{aligned}$$

$\square$

1.3.4 Corollary. If $A, B \in M_n$ and if A and B are similar, then they have the same eigenvalues.

Exercise. Show that the only matrix similar to the zero matrix is the zero matrix, and the only matrix similar to the identity matrix is the identity matrix.

1.3.5 Example. Having the same eigenvalues is a necessary but not sufficient condition for similarity. Consider

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

which have the same eigenvalues but are not similar.

Exercise. Suppose that $A, B \in M_n$ are similar and let $q(t)$ be a given polynomial. Show that $q(A)$ and $q(B)$ are similar. In particular, show that $A + \alpha I$ and $B + \alpha I$ are similar for any $\alpha \in \mathbf{C}$.

Exercise. Let $A, B, C, D \in M_n$. Suppose that $A \sim B$ and $C \sim D$, both via the same similarity matrix S. Show that $A + C \sim B + D$ and $AC \sim BD$.

Exercise. Let $A, S \in M_n$ and suppose that S is nonsingular. Show that $S_k(S^{-1}AS) = S_k(A)$ for all $k = 1, \dots, n$ and explain why $E_k(S^{-1}AS) = E_k(A)$ for all $k = 1, \dots, n$. Thus, all the principal minor sums (1.2.10) are similarity invariants, not just the determinant and trace.

Exercise. Explain why rank is a similarity invariant: If $B \in M_n$ is similar to $A \in M_n$, then rank B = rank A. *Hint:* See (0.4.6).

Since diagonal matrices are especially simple and have very nice properties, it is of interest to know for which $A \in M_n$ there is a diagonal matrix in the similarity equivalence class of A, that is, which matrices are similar to diagonal matrices.

1.3.6 Definition. If $A \in M_n$ is similar to a diagonal matrix, then A is said to be *diagonalizable*.

1.3.7 Theorem. Let $A \in M_n$ be given. Then A is similar to a block matrix of the form

$$\begin{bmatrix} \Lambda & C \\ 0 & D \end{bmatrix}, \quad \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_k),\ D \in M_{n-k},\ 1 \le k < n \quad (1.3.7.1)$$

if and only if there is a linearly independent set of k vectors in $\mathbf{C}^n$, each of which is an eigenvector of A. The matrix A is diagonalizable if and only if there is a linearly independent set of n vectors, each of which is an eigenvector of A. If $\{x^{(1)}, \dots, x^{(n)}\}$ is a basis of $\mathbf{C}^n$ consisting of eigenvectors of A and if $S = [x^{(1)} \ \dots \ x^{(n)}]$, then $S^{-1}AS$ is a diagonal matrix. If A is similar to a matrix of the form (1.3.7.1), then the diagonal entries of Λ are eigenvalues of A; if A is similar to a diagonal matrix Λ, then the diagonal entries of Λ are all of the eigenvalues of A.

Proof: Suppose that $k < n$, $\{x^{(1)}, \dots, x^{(k)}\} \subset \mathbf{C}^n$ is linearly independent, and $Ax^{(i)} = \lambda_i x^{(i)}$ for each $i = 1, \dots, k$. Let $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_k)$, let $S_1 = [x^{(1)} \ \dots \ x^{(k)}]$ and choose any $S_2 \in M_n$ such that $S = [S_1 \ S_2]$ is nonsingular. Calculate

$$\begin{aligned} S^{-1}AS &= S^{-1}[Ax^{(1)} \ \ldots \ Ax^{(k)} \ AS_2] = S^{-1}[\lambda_1 x^{(1)} \ \ldots \ \lambda_k x^{(k)} \ AS_2] \\ &= [\lambda_1 S^{-1}x^{(1)} \ \ldots \ \lambda_k S^{-1}x^{(k)} \ S^{-1}AS_2] = [\lambda_1 e_1 \ \ldots \ \lambda_k e_k \ S^{-1}AS_2] \\ &= \begin{bmatrix} \Lambda & C \\ 0 & D \end{bmatrix}, \quad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_k), \ \begin{bmatrix} C \\ D \end{bmatrix} = S^{-1}AS_2 \end{aligned}$$

Conversely, if S is nonsingular, $S^{-1}AS = \left[\begin{smallmatrix} \Lambda & C \\ 0 & D \end{smallmatrix}\right]$, and we partition $S = [S_1 \ S_2]$ with $S_1 \in M_{n,k}$, then S_1 has independent columns and $[AS_1 \ AS_2] = AS = S\left[\begin{smallmatrix} \Lambda & C \\ 0 & D \end{smallmatrix}\right] = [S_1\Lambda \ S_1C + S_2D]$. Thus, $AS_1 = S_1\Lambda$, so each column of S_1 is an eigenvector of A.

If $k = n$ and we have a basis $\{x^{(1)}, \ldots, x^{(n)}\}$ of $\mathbf{C}^n$ such that $Ax^{(i)} = \lambda_i x^{(i)}$ for each $i = 1, \ldots, n$, let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and let $S = [x^{(1)} \ \ldots \ x^{(n)}]$, which is nonsingular. Our previous calculation shows that $S^{-1}AS = \Lambda$. Conversely, if S is nonsingular and $S^{-1}AS = \Lambda$ then $AS = S\Lambda$, so each column of S is an eigenvector of A.

The final assertions about the eigenvalues follow from an examination of the characteristic polynomials: $p_A(t) = p_\Lambda(t)p_D(t)$ if $k < n$ and $p_A(t) = p_\Lambda(t)$ if $k = n$. □

The proof of Theorem 1.3.7 is, in principle, an algorithm for diagonalizing a diagonalizable matrix $A \in M_n$: find all n of the eigenvalues of A; find n associated (and linearly independent!) eigenvectors; and construct the matrix S. However, except for small examples, this is *not* a practical computational procedure.

Exercise. Show that $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ is *not* diagonalizable. *Hint*: If it were diagonalizable, it would be similar to the zero matrix. Alternatively, how many linearly independent eigenvectors are associated with the eigenvalue 0?

Exercise. Let $q(t)$ be a given polynomial. If A is diagonalizable, show that $q(A)$ is diagonalizable. If $q(A)$ is diagonalizable, must A be diagonalizable? Why?

Exercise. If λ is an eigenvalue of $A \in M_n$ that has multiplicity $m \geq 1$, show that A is not diagonalizable if rank $(A - \lambda I) > n - m$.

Exercise. If there is a linearly independent set of k vectors in $\mathbf{C}^n$, each of which is an eigenvector of $A \in M_n$ associated with a given eigenvalue λ, explain carefully why the (algebraic) multiplicity of λ is at least k.

Diagonalizability is assured if *all* the eigenvalues are distinct. The basis for this fact is the following important lemma about *some* of the eigenvalues.

1.3.8 Lemma. Let $\lambda_1, \ldots, \lambda_k$ be $k \geq 2$ distinct eigenvalues of $A \in M_n$ (that is, $\lambda_i \neq \lambda_j$ if $i \neq j$ and $1 \leq i, j \leq k$), and suppose that $x^{(i)}$ is an eigenvector associated with λ_i for each $i = 1, \ldots, k$. Then $\{x^{(1)}, \ldots, x^{(k)}\}$ is linearly independent.

Proof: Suppose there are complex scalars $\alpha_1, \ldots \alpha_k$ such that $\alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \cdots + \alpha_r x^{(r)} = 0$. Let $B_1 = (A - \lambda_2 I)(A - \lambda_3 I) \cdots (A - \lambda_k I)$ (the product omits $A - \lambda_1 I$). Since $x^{(i)}$ is an eigenvector associated with the eigenvalue λ_i for each $i = 1, \ldots, n$, we have $B_1 x^{(i)} = (\lambda_i - \lambda_2)(\lambda_i - \lambda_3) \cdots (\lambda_i - \lambda_k) x^{(i)}$, which is zero if $2 \leq i \leq k$ (one of the factors is zero) and nonzero if $i = 1$ (no factor is zero and $x^{(1)} \neq 0$). Thus,

$$\begin{aligned} 0 &= B_1 \left(\alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \cdots + \alpha_k x^{(k)} \right) \\ &= \alpha_1 B_1 x^{(1)} + \alpha_2 B_1 x^{(2)} + \cdots + \alpha_k B_1 x^{(k)} \\ &= \alpha_1 B_1 x^{(1)} + 0 + \cdots + 0 = \alpha_1 B_1 x^{(1)} \end{aligned}$$

which ensures that $\alpha_1 = 0$ since $B_1 x^{(1)} \neq 0$. Repeat this argument for each $j = 2, \ldots, k$, defining B_j by a product like that defining B_1, but in which the factor $A - \lambda_j I$ is omitted. For each j we find that $\alpha_j = 0$, so $\alpha_1 = \cdots = \alpha_k = 0$ and hence $\{x^{(1)}, \ldots, x^{(k)}\}$ is an independent set. $\square$

1.3.9 Theorem. If $A \in M_n$ has n distinct eigenvalues, then A is diagonalizable.

Proof: Let $x^{(i)}$ be an eigenvector associated with the eigenvalue λ_i for each $i = 1, \ldots, n$. Since all the eigenvalues are distinct, Lemma 1.3.8 ensures that $\{x^{(1)}, \ldots, x^{(n)}\}$ is linearly independent. Theorem 1.3.7 then ensures that A is diagonalizable. $\square$

Having distinct eigenvalues is *sufficient* for diagonalizability, but of course it is not *necessary*.

Exercise. Give an example of a diagonalizable matrix that does not have distinct eigenvalues.

Exercise. Let $A, P \in M_n$ and suppose that P is a permutation matrix, so every entry of P is either 0 or 1 and $P^T = P^{-1}$; see (0.9.5). Show that the permutation similarity PAP^{-1} reorders the diagonal entries of A. For any given diagonal matrix $D \in M_n$ explain why there is a permutation similarity PDP^{-1} that puts the diagonal entries of D into any given order. In particular, explain why P can be chosen so that any repeated diagonal entries occur contiguously.

In general, matrices $A, B \in M_n$ do *not* commute, but if A and B are both diagonal, they always commute. The latter observation can be generalized somewhat; the following lemma is helpful in this regard.

1.3.10 Lemma. Let $B_1 \in M_{n_1}, \ldots, B_d \in M_{n_d}$ be given and let B be the direct sum

$$B = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_d \end{bmatrix} = B_1 \oplus \cdots \oplus B_d$$

Then B is diagonalizable if and only if each of $B_1, \ldots, B_d$ is diagonalizable.

Proof: If for each $i = 1, \ldots, d$ there is a nonsingular $S_i \in M_{n_i}$ such that $S_i^{-1} B_i S_i$ is diagonal, and if we define $S = S_1 \oplus \cdots \oplus S_d$, then one checks that $S^{-1}BS$ is diagonal.

For the converse, we proceed by induction. There is nothing to prove for $d = 1$. Suppose that $d \geq 2$ and that the assertion has been established for direct sums with $d - 1$ or fewer direct summands. Let $C = B_1 \oplus \cdots \oplus B_{d-1}$, let $n = n_1 + \cdots + n_{d-1}$, and let $m = n_d$. Let $S \in M_{n+m}$ be nonsingular and such that

$$S^{-1}BS = S^{-1}(C \oplus B_d)\, S = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{n+m})$$

Rewrite this identity as $BS = S\Lambda$. Partition $S = [s_1\, s_2 \ldots s_{n+m}]$ with

$$s_i = \begin{bmatrix} \xi_i \\ \eta_i \end{bmatrix} \in \mathbf{C}^{n+m}, \qquad \xi_i \in \mathbf{C}^n, \eta_i \in \mathbf{C}^m, i = 1, 2, \ldots, n+m$$

Then $Bs_i = \lambda_i s_i$ implies that $C\xi_i = \lambda_i \xi_i$ and $B_d \eta_i = \lambda_i \eta_i$ for $i = 1, 2, \ldots, n+m$. The row rank of $[\xi_1 \ \ldots \ \xi_{n+m}] \in M_{n,n+m}$ is n because this matrix comprises the first n rows of the nonsingular matrix S. Thus, its column rank is also n, so the set $\{\xi_1, \ldots, \xi_{n+m}\}$ contains a subset of n independent vectors, each of which is an eigenvector of C. Theorem 1.3.7 ensures that C is diagonalizable and the induction hypothesis ensures that its direct summands $B_1, \ldots, B_d$ are all diagonalizable. The row rank of $[\eta_1 \ \ldots \ \eta_{n+m}] \in M_{n,n+m}$ is m, so the set $\{\eta_1, \ldots, \eta_{n+m}\}$ contains a subset of m independent vectors; it follows that B_d is diagonalizable as well. $\square$

1.3.11 Definition. Two matrices $A, B \in M_n$ are said to be *simultaneously diagonalizable* if there is a single nonsingular $S \in M_n$ such that $S^{-1}AS$ and $S^{-1}BS$ are both diagonal.

Exercise. Let $A, B, S \in M_n$ and suppose that S is nonsingular. Show that A commutes with B if and only if $S^{-1}AS$ commutes with $S^{-1}BS$.

Exercise. If $A, B \in M_n$ are simultaneously diagonalizable, show that they commute. *Hint:* Diagonal matrices commute.

Exercise. Show that if $A \in M_n$ is diagonalizable and $\lambda \in \mathbf{C}$, then A and λI are simultaneously diagonalizable.

1.3.12 Theorem. Let $A, B \in M_n$ be diagonalizable. Then A and B commute if and only if they are simultaneously diagonalizable.

Proof: Assume that A and B commute, perform a similarity transformation on both A and B that diagonalizes A (but not necessarily B) and groups together any repeated eigenvalues of A. If $\mu_1, \ldots, \mu_d$ are the distinct eigenvalues of A and $n_1, \ldots, n_d$ are their respective multiplicities, then we may assume that

$$A = \begin{bmatrix} \mu_1 I_{n_1} & & 0 \\ & \mu_2 I_{n_2} & \\ & & \ddots & \\ 0 & & & \mu_d I_{n_d} \end{bmatrix}, \quad \mu_i \neq \mu_j \text{ if } i \neq j \tag{1.3.13}$$

Since $AB = BA$, (0.7.7) ensures that

$$B = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_d \end{bmatrix}, \quad \text{each } B_i \in M_{n_i} \tag{1.3.14}$$

is block diagonal conformal to A. Since B is diagonalizable, (1.3.10) ensures that each B_i is diagonalizable. Let $T_i \in M_{n_i}$ be nonsingular and such that $T_i^{-1} B_i T_i$ is diagonal for each $i = 1, \ldots, d$; let

$$T = \begin{bmatrix} T_1 & & 0 \\ & T_2 & \\ & & \ddots & \\ 0 & & & T_d \end{bmatrix} \tag{1.3.15}$$

Then $T_i^{-1} \mu_i I_{n_i} T_i = \mu_i I_{n_i}$, so $T^{-1}AT = A$ and $T^{-1}BT$ are both diagonal. The converse is included in an earlier exercise. $\square$

We want to have a version of Theorem 1.3.12 involving arbitrarily many commuting diagonalizable matrices. Central to our investigation is the notion of an invariant subspace and the companion notion of a block triangular matrix.

1.3.16 Definitions. A *family* $\mathcal{F} \subseteq M_n$ of matrices is a nonempty finite or infinite set of matrices; a *commuting family* is a family of matrices in which every pair of matrices commutes. For a given $A \in M_n$, a subspace $W \subseteq \mathbf{C}^n$ is A-*invariant* if $Aw \in W$ for every $w \in W$. A subspace $W \subseteq \mathbf{C}^n$ is *trivial* if either $W = \{0\}$ or $W = \mathbf{C}^n$; otherwise, it is *nontrivial*. For a given family

$\mathcal{F} \subseteq M_n$, a subspace $W \subseteq \mathbf{C}^n$ is *$\mathcal{F}$-invariant* if W is A-invariant for each $A \in \mathcal{F}$. A given family $\mathcal{F} \subseteq M_n$ is *reducible* if some nontrivial subspace of C^n is F-invariant; otherwise, $\mathcal{F}$ is *irreducible*.

Exercise. For $A \in M_n$, show that each nonzero element of a one-dimensional A-invariant subspace of $\mathbf{C}^n$ is an eigenvector of A.

Exercise. Suppose that $n \geq 2$ and $S \in M_n$ is nonsingular. Partition $S = [S_1 \ S_2]$, in which $S_1 \in M_{n,k}$ and $S_2 \in M_{n,n-k}$ with $1 < k < n$. Explain why

$$S^{-1}S_1 = [e_1 \ \ldots \ e_k] = \begin{bmatrix} I_k \\ 0 \end{bmatrix} \text{ and } S^{-1}S_2 = [e_{k+1} \ \ldots \ e_n] = \begin{bmatrix} 0 \\ I_{n-k} \end{bmatrix}$$

Invariant subspaces and block triangular matrices are two sides of the same valuable coin: the former is the linear algebra side, while the latter is the matrix analysis side. Let $A \in M_n$ with $n \geq 2$ and suppose that $W \subseteq \mathbf{C}^n$ is a k-dimensional subspace with $1 < k < n$. Choose a basis $s_1, \ldots, s_k$ of W and let $S_1 = [s_1 \ \ldots \ s_k] \in M_{n,k}$. Choose any $s_{k+1}, \ldots, s_n$ such that $s_1, \ldots, s_n$ is a basis for $\mathbf{C}^n$, let $S_2 = [s_{k+1} \ \ldots \ s_n] \in M_{n,n-k}$, and let $S = [S_1 \ S_2]$; S has linearly independent columns, so it is nonsingular. If W is A-invariant, then $As_j \in W$ for each $j = 1, \ldots, k$, so each As_j is a linear combination of $s_1, \ldots, s_k$, that is, $AS_1 = S_1B$ for some $B \in M_k$. If $AS_1 = S_1B$, then $AS = [AS_1 \ AS_2] = [S_1B \ AS_2]$ and hence

$$\begin{aligned} S^{-1}AS &= \begin{bmatrix} S^{-1}S_1B & S^{-1}AS_2 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} I_k \\ 0 \end{bmatrix} B & S^{-1}AS_2 \end{bmatrix} \\ &= \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}, \quad B \in M_k, \ 1 \leq k \leq n-1 \end{aligned} \tag{1.3.17}$$

The conclusion is that A is similar to a block triangular matrix (1.3.17) if it has a k-dimensional invariant subspace. But we can say a little more: we know that $B \in M_k$ has an eigenvalue, so suppose $B\xi = \lambda\xi$ for some scalar λ and a nonzero $\xi \in \mathbf{C}^k$. Then $0 \neq S_1\xi \in W$ and $A(S_1\xi) = (AS_1)\xi = S_1B\xi = \lambda(S_1\xi)$, which means that A has an eigenvector in W.

Conversely, if $S = [S_1 \ S_2] \in M_n$ is nonsingular, $S_1 \in M_{n,k}$, and $S^{-1}AS$ has the block triangular form (1.3.17), then

$$AS_1 = AS \begin{bmatrix} I_k \\ 0 \end{bmatrix} = S \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \begin{bmatrix} I_k \\ 0 \end{bmatrix} = \begin{bmatrix} S_1 & S_2 \end{bmatrix} \begin{bmatrix} B_1 \\ 0 \end{bmatrix} = S_1B$$

so the (k-dimensional) span of the columns of S_1 is A-invariant. We summarize the foregoing discussion in the following observation.

1.3.18 Observation. Suppose that $n \geq 2$. A given $A \in M_n$ is similar to a block triangular matrix of the form (1.3.17) if and only if some nontrivial subspace of $\mathbf{C}^n$ is A-invariant. Moreover, if $W \subseteq \mathbf{C}^n$ is a nonzero A-invariant subspace, then some vector in W is an eigenvector of A. A given family $\mathcal{F} \subseteq M_n$ is reducible if and only if there is some $k \in \{2, \ldots, n-1\}$ and a nonsingular $S \in M_n$ such that $S^{-1}AS$ has the form (1.3.17) for every $A \in \mathcal{F}$.

The following lemma is at the heart of many subsequent results.

1.3.19 Lemma. Let $\mathcal{F} \subset M_n$ be a commuting family. Then some nonzero vector in $\mathbf{C}^n$ is an eigenvector of every $A \in \mathcal{F}$.

Proof: There is always a nonzero $\mathcal{F}$-invariant subspace, namely, $\mathbf{C}^n$. Let $m = \min\{\dim V : V \text{ is a nonzero } \mathcal{F}\text{-invariant subspace of } \mathbf{C}^n\}$ and let W be any given $\mathcal{F}$-invariant subspace such that $\dim W = m$. Let any $A \in \mathcal{F}$ be given. Since W is $\mathcal{F}$-invariant, it is A-invariant, so (1.3.18) ensures that there is some nonzero $x_0 \in W$ and some $\lambda \in \mathbf{C}$ such that $Ax_0 = \lambda x_0$. Consider the subspace $W_{A,\lambda} \equiv \{x \in W : Ax = \lambda x\}$. Then $x_0 \in W_{A,\lambda}$ so $W_{A,\lambda}$ is a nonzero subspace of W. For any $B \in \mathcal{F}$ and any $x \in W_{A,\lambda}$, $\mathcal{F}$-invariance of W ensures that $Bx \in W$. Using commutativity of $\mathcal{F}$, we compute

$$A(Bx) = (AB)x = (BA)x = B(Ax) = B(\lambda x) = \lambda(Bx)$$

which shows that $Bx \in W_{A,\lambda}$. Thus, $W_{A,\lambda}$ is $\mathcal{F}$-invariant and nonzero, so $\dim W_{A,\lambda} \geq m$. But $W_{A,\lambda} \subseteq W$, so $\dim W_{A,\lambda} \leq m$ and hence $W = W_{A,\lambda}$. We have now shown that for each $A \in \mathcal{F}$ there is some scalar λ_A such that $Ax = \lambda_A x$ for all $x \in W$, so every nonzero vector in W is an eigenvector of every matrix in $\mathcal{F}$. $\square$

Exercise. Consider the nonzero $\mathcal{F}$-invariant subspace W in the preceding proof. Explain why $m = \dim W = 1$.

Exercise. Suppose $\mathcal{F} \subset M_n$ is a commuting family. Show that there is a nonsingular $S \in M_n$ such that for every $A \in \mathcal{F}$, $S^{-1}AS$ has the block triangular form (1.3.17) with $k = 1$.

Lemma 1.3.19 concerns commuting families of arbitrary nonzero cardinality. Our next result shows that Theorem 1.3.12 can be extended to arbitrary commuting families of diagonalizable matrices.

1.3.20 Definition. A family $\mathcal{F} \subset M_n$ is said to be *simultaneously diagonalizable if* there is a single nonsingular $S \in M_n$ such that $S^{-1}AS$ is diagonal for every $A \in \mathcal{F}$.

1.3.21 Theorem. Let $\mathcal{F} \subset M_n$ be a family of diagonalizable matrices. Then $\mathcal{F}$ is a commuting family if and only if it is a simultaneously diagonalizable family. Moreover, for any given $A_0 \in \mathcal{F}$ and for any given ordering $\lambda_1, \ldots, \lambda_n$ of the eigenvalues of A_0, there is a nonsingular $S \in M_n$ such that $S^{-1}A_0S = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $S^{-1}BS$ is diagonal for every $B \in \mathcal{F}$.

Proof: If $\mathcal{F}$ is simultaneously diagonalizable, then it is a commuting family by a previous exercise. We prove the converse by induction on n. If $n = 1$, there is nothing to prove since every family is both commuting and diagonal. Let us suppose that $n \geq 2$ and that, for each $k = 1, 2, \ldots, n-1$, any commuting family of k-by-k diagonalizable matrices is simultaneously diagonalizable. If every matrix in $\mathcal{F}$ is a scalar matrix, there is nothing to prove, so we may assume that $A \in \mathcal{F}$ is a given n-by-n diagonalizable matrix with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ and $k \geq 2$, that $AB = BA$ for every $B \in \mathcal{F}$, and that each $B \in \mathcal{F}$ is diagonalizable. Using the argument in (1.3.12), we reduce to the case in which A has the form (1.3.13). Since every $B \in \mathcal{F}$ commutes with A, (0.7.7) ensures that each $B \in \mathcal{F}$ has the form (1.3.14). Let $B, \hat{B} \in \mathcal{F}$, so $B = B_1 \oplus \cdots \oplus B_k$ and $\hat{B} = \hat{B}_1 \oplus \cdots \oplus \hat{B}_k$ in which each of $B_i, \hat{B}_i$ has the same size and that size is at most $n - 1$. Commutativity and diagonalizability of B and $\hat{B}$ imply commutativity and diagonalizability of B_i and $\hat{B}_i$ for each $i = 1, \ldots, d$. By the induction hypothesis, there are k similarity matrices $T_1, T_2, \ldots, T_k$ of appropriate size, each of which diagonalizes the corresponding block of every matrix in $\mathcal{F}$. Then the direct sum (1.3.15) diagonalizes every matrix in $\mathcal{F}$.

We have shown that there is a nonsingular $T \in M_n$ such that $T^{-1}BT$ is diagonal for every $B \in \mathcal{F}$. Then $T^{-1}A_0T = P \operatorname{diag}(\lambda_1, \ldots, \lambda_n)P^T$ for some permutation matrix P, $P^T(T^{-1}A_0T)P = (TP)^{-1}A_0(TP) = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $(TP)^{-1}B(TP) = P^T(T^{-1}BT)P$ is diagonal for every $B \in \mathcal{F}$ (0.9.5). $\square$

Remarks: We defer two important issues until Chapter 3: (1) Given $A, B \in M_n$, how can we determine if A is similar to B? (2) How can we tell if a given matrix is diagonalizable without knowing its eigenvectors?

Although AB and BA need not be the same (and need not be the same size even when both products are defined), their eigenvalues are as much the same as possible. Indeed, if A and B are both square, then AB and BA have exactly the same eigenvalues. These important facts follow from a simple but very useful observation.

Exercise. Let $X \in M_{m,n}$ be given. Explain why $\left[\begin{smallmatrix} I_m & X \\ 0 & I_n \end{smallmatrix}\right] \in M_{m+n}$ is nonsingular and verify that its inverse is $\left[\begin{smallmatrix} I_m & -X \\ 0 & I_n \end{smallmatrix}\right]$.

1.3.22 Theorem. Suppose that $A \in M_{m,n}$ and $B \in M_{n,m}$ with $m \leq n$. Then the n eigenvalues of BA are the m eigenvalues of AB together with $n - m$ zeroes; that is, $p_{BA}(t) = t^{n-m} p_{AB}(t)$. If $m = n$ and at least one of A or B is nonsingular, then AB and BA are similar.

Proof: A computation reveals that

$$\begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix} \begin{bmatrix} AB & 0 \\ B & 0_n \end{bmatrix} \begin{bmatrix} I_m & A \\ 0 & I_n \end{bmatrix} = \begin{bmatrix} 0_m & 0 \\ B & BA \end{bmatrix}$$

and the preceding exercise ensures that $C_1 = \left[\begin{smallmatrix} AB & 0 \\ B & 0_n \end{smallmatrix}\right]$ and $C_2 = \left[\begin{smallmatrix} 0_m & 0 \\ B & BA \end{smallmatrix}\right]$ are similar. The eigenvalues of C_1 are the eigenvalues of AB together with n zeroes. The eigenvalues of C_2 are the eigenvalues of BA together with m zeroes. Since the eigenvalues of C_1 and C_2 are the same, the first assertion of the theorem follows. The final assertion follows from the observation that $AB = A(BA)A^{-1}$ if A is nonsingular and $m = n$. $\square$

Theorem 1.3.22 has many applications, several of which emerge in the following chapters. Here are just four.

1.3.23 Example. Eigenvalues of a low-rank matrix. Suppose $A \in M_n$ is factored as $A = XY^T$, in which $X, Y \in M_{n,r}$ and $r < n$. Then the eigenvalues of A are the same as those of the r-by-r matrix $Y^T X$, together with $n - r$ zeroes. For example, consider the n-by-n matrix $J_n = ee^T$ whose entries are all ones. Its eigenvalues are the eigenvalue of the 1-by-1 matrix $e^T e = [n]$, namely, n, together with $n - 1$ zeroes. The eigenvalues of any matrix of the form $A = xy^T$ with $x, y \in \mathbf{C}^n$ ($\operatorname{rank} A$ is at most 1) are $y^T x$, together with $n - 1$ zeroes. The eigenvalues of any matrix of the form $A = xy^T + zw^T = [x\ z][y\ w]^T$ with $x, y, z, w \in \mathbf{C}^n$ ($\operatorname{rank} A$ is at most 2) are the two eigenvalues of $[y\ w]^T [x\ z] = \left[\begin{smallmatrix} y^T x & y^T z \\ w^T x & w^T z \end{smallmatrix}\right]$ (1.2.4b) together with $n - 2$ zeroes.

1.3.24 Example. Cauchy's determinant identity. Let a nonsingular $A \in M_n$

and $x, y \in \mathbf{C}^n$ be given. Then

$$\begin{aligned}
\det(A + xy^T) &= (\det A)\left(\det(I + A^{-1}xy^T)\right) \\
&= (\det A)\prod_{i=1}^{n} \lambda_i(I + A^{-1}xy^T) \\
&= (\det A)\prod_{i=1}^{n} \left(1 + \lambda_i(A^{-1}xy^T)\right) \\
&= (\det A)\left(1 + y^T A^{-1}x\right) \quad \text{(use (1.3.23))} \\
&= \det A + y^T\left((\det A)\, A^{-1}\right)x = \det A + y^T\,(\operatorname{adj} A)\, x
\end{aligned}$$

Cauchy's identity $\det(A + xy^T) = \det A + y^T\,(\operatorname{adj} A)\, x$, valid for any $A \in M_n$, now follows by continuity. For a different approach, see (0.8.5).

1.3.25 Example. For any $n \geq 2$, consider the n-by-n real symmetric Hankel matrix

$$A = [i + j]_{i,j=1}^{n} = \begin{bmatrix} 2 & 3 & 4 & \cdots \\ 3 & 4 & 5 & \cdots \\ 4 & 5 & 6 & \cdots \\ \vdots & & & \ddots \end{bmatrix} = ve^T + ev^T = [v\ e][e\ v]^T$$

in which every entry of $e \in \mathbf{R}^n$ is 1 and $v = [1\ 2\ \ldots\ n]^T$. The eigenvalues of A are the same as those of

$$B = [e\ v]^T[v\ e] = \begin{bmatrix} e^Tv & e^Te \\ v^Tv & v^Te \end{bmatrix} = \begin{bmatrix} \frac{n(n+1)}{2} & n \\ \frac{n(n+1)(2n+1)}{6} & \frac{n(n+1)}{2} \end{bmatrix}$$

together with $n - 2$ zeroes. According to (1.2.4b), the eigenvalues B (one positive and one negative) are

$$n(n+1)\left[\frac{1}{2} \pm \sqrt{\frac{2n+1}{6(n+1)}}\right]$$

1.3.26 Example. For any $n \geq 2$, consider the n-by-n real skew-symmetric Toeplitz matrix

$$A = [i - j]_{i,j=1}^{n} = \begin{bmatrix} 0 & -1 & -2 & \cdots \\ 1 & 0 & -1 & \cdots \\ 2 & 1 & 0 & \cdots \\ \vdots & & & \ddots \end{bmatrix} = ve^T - ev^T = [v\ \ -e][e\ v]^T$$

Except for $n-2$ zeroes, the eigenvalues of A are the same as those of

$$B = [e\ v]^T[v\ -e] = \begin{bmatrix} e^Tv & -e^Te \\ v^Tv & -v^Te \end{bmatrix}$$

which, using (1.2.4b) again, are $\pm\frac{ni}{2}\sqrt{\frac{n^2-1}{3}}$.

Theorem 1.3.22 on the eigenvalues of AB vs. BA is only part of the story; we return to that story in (3.2.11).

If $A \in M_n$ is diagonalizable and $A = S\Lambda S^{-1}$, then aS also diagonalizes A for any $a \neq 0$. Thus, a diagonalizing similarity is never unique. Nevertheless, *every* similarity of A to a particular diagonal matrix can be obtained from just one given similarity.

1.3.27 Theorem. Suppose $A \in M_n$ is diagonalizable, let $\mu_1, \ldots, \mu_d$ be its distinct eigenvalues with respective multiplicities $n_1, \ldots, n_d$, let $S, T \in M_n$ be nonsingular, and suppose $A = S\Lambda S^{-1}$, in which Λ is a diagonal matrix of the form (1.3.13). (a) Then $A = T\Lambda T^{-1}$ if and only if $T = S(R_1 \oplus \cdots \oplus R_d)$ in which each $R_i \in M_{n_i}$ is nonsingular. (b) If $S = [S_1\ \ldots\ S_d]$ and $T = [T_1\ \ldots\ T_d]$ are partitioned conformally to Λ, then $A = S\Lambda S^{-1} = T\Lambda T^{-1}$ if and only if for each $i = 1, \ldots, d$ the column space of S_i is the same as the column space of T_i. (c) If A has n distinct eigenvalues and $S = [s_1\ \ldots\ s_n]$ and $T = [t_1\ \ldots\ t_n]$ are partitioned according to their columns, then $A = S\Lambda S^{-1} = T\Lambda T^{-1}$ if and only if there is a nonsingular diagonal matrix $R = \mathrm{diag}(r_1, \ldots, r_n)$ such that $T = SR$ if and only if for each $i = 1, \ldots, n$ the column s_i is a nonzero scalar multiple of the corresponding column t_i.

Proof: We have $S\Lambda S^{-1} = T\Lambda T^{-1}$ if and only if $(S^{-1}T)\Lambda = \Lambda(S^{-1}T)$ if and only if $S^{-1}T$ is block diagonal conformal to Λ (0.7.7), that is, if and only if $S^{-1}T = R_1 \oplus \cdots \oplus R_d$ and each $R_i \in M_{n_i}$ is nonsingular. For (b), observe that if $1 \leq k \leq n$ then the column space of $X \in M_{n,k}$ is contained in the column space of $Y \in M_{n,k}$ if and only if there is some $C \in M_k$ such that $X = YC$; if, in addition, $\mathrm{rank}\, X = \mathrm{rank}\, Y = k$, then C must be nonsingular. The assertion (c) is a special case of (a) and (b). $\square$

Could two real matrices be similar *only* via a complex matrix? The following theorem answers that question.

1.3.28 Theorem. Let real matrices $A, B \in M_n(\mathbf{R})$ be given and suppose that there is a nonsingular $S \in M_n$ such that $A = SBS^{-1}$, that is, A and B are similar over $\mathbf{C}$. Then there is a nonsingular $T \in M_n(\mathbf{R})$ such that $A = TBT^{-1}$, that is, A and B are similar over $\mathbf{R}$.

Proof: Let $S = C + iD$ be nonsingular and $C, D \in M_n(\mathbf{R})$. Then $A = SBS^{-1}$ if and only if $A(C + iD) = AS = SB = (C + iD)B$. Equating the real and imaginary parts of this identity shows that $AC = CB$ and $AD = DB$. If C is nonsingular, take $T = C$. Otherwise, consider the polynomial $p(t) = \det(C+tD)$, which is not identically constant since $p(0) = \det C = 0$ and $p(i) = \det S \neq 0$. Since $p(t)$ has only finitely many zeros in the complex plane, there is a real t_0 such that $p(t_0) \neq 0$. Take $T = C + t_0 D$. □

Our final theorem about similarity shows that the *only* relationship between the eigenvalues and main diagonal entries of a complex matrix is that their respective sums be equal.

1.3.29 Theorem. Let an integer $n \geq 2$ and complex scalars $\lambda_1, \ldots, \lambda_n$ and $d_1, \ldots, d_n$ be given. There is an $A \in M_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ and main diagonal entries $d_1, \ldots, d_n$ if and only if $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n d_i$. If $\lambda_1, \ldots, \lambda_n$ and $d_1, \ldots, d_n$ are all real and have the same sums, there is an $A \in M_n(\mathbf{R})$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ and main diagonal entries $d_1, \ldots, d_n$.

Proof: We know that $\operatorname{tr} A = E_1(A) = S_1(A)$ for any $A \in M_n$ (1.2.16), which establishes the necessity of the stated condition. We must prove its sufficiency.

If $k \geq 2$ and if $\lambda_1, \ldots, \lambda_k$ and $d_1, \ldots, d_k$ are any given complex scalars such that $\sum_{i=1}^k \lambda_i = \sum_{i=1}^k d_i$, we claim that the upper bidiagonal matrix

$$T(\lambda_1, \ldots, \lambda_k) = \begin{bmatrix} \lambda_1 & 1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in M_k$$

is similar to a matrix with diagonal entries $d_1, \ldots, d_k$; that matrix has the property asserted. Let $L(s,t) = \left[\begin{smallmatrix} 1 & 0 \\ s-t & 1 \end{smallmatrix}\right]$, so $L(s,t)^{-1} = \left[\begin{smallmatrix} 1 & 0 \\ t-s & 1 \end{smallmatrix}\right]$.

Consider first the case $k = 2$, so $\lambda_1 + \lambda_2 = d_1 + d_2$. Compute the similarity $L(\lambda_1, d_1) T(\lambda_1, \lambda_2) L(\lambda_1, d_1)^{-1} =$

$$\begin{aligned} & \begin{bmatrix} 1 & 0 \\ \lambda_1 - d_1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ d_1 - \lambda_1 & 1 \end{bmatrix} \\ = & \begin{bmatrix} d_1 & \bigstar \\ \bigstar & \lambda_1 + \lambda_2 - d_1 \end{bmatrix} = \begin{bmatrix} d_1 & \bigstar \\ \bigstar & d_2 \end{bmatrix} \end{aligned}$$

in which we use the hypothesis $\lambda_1 + \lambda_2 - d_1 = d_1 + d_2 - d_1 = d_2$. This verifies our claim for $k = 2$.

We proceed by induction. Assume that our claim has been proved for some

$k \geq 2$ and that $\sum_{i=1}^{k+1} \lambda_i = \sum_{i=1}^{k+1} d_i$. Partition $T(\lambda_1, \ldots, \lambda_{k+1}) = [T_{ij}]_{i,j=1}^2$, in which $T_{11} = T(\lambda_1, \lambda_2)$, $T_{12} = E_2$, $T_{21} = 0$, and $T_{22} = T(\lambda_3, \ldots, \lambda_{k+1})$, with $E_2 = [e_2 \; 0 \; \ldots \; 0] \in M_{2,k-1}$ and $e_2 = [0 \; 1]^T \in \mathbf{C}^2$. *Let* $\mathcal{L} = L(\lambda_1, d_1) \oplus I_{k-1}$ and compute $\mathcal{L}T(\lambda_1, \ldots, \lambda_{k+1})\mathcal{L}^{-1} =$

$$\begin{bmatrix} L(\lambda_1, d_1) & 0 \\ 0 & I_{k-1} \end{bmatrix} \begin{bmatrix} T(\lambda_1, \lambda_2) & E_2 \\ 0 & T(\lambda_3, \ldots, \lambda_{k+1}) \end{bmatrix} \begin{bmatrix} L(d_1, \lambda_1) & 0 \\ 0 & I_{k-1} \end{bmatrix}$$

$$= \begin{bmatrix} \begin{bmatrix} d_1 & \bigstar \\ \bigstar & \lambda_1 + \lambda_2 - d_1 \end{bmatrix} & E_2 \\ 0 & T(\lambda_3, \ldots, \lambda_{k+1}) \end{bmatrix}$$

$$= \begin{bmatrix} d_1 & \bigstar \\ \bigstar & T(\lambda_1 + \lambda_2 - d_1, \lambda_3, \ldots, \lambda_{k+1}) \end{bmatrix} = \begin{bmatrix} d_1 & \bigstar \\ \bigstar & D \end{bmatrix}$$

The sum of the eigenvalues of $D = T(\lambda_1 + \lambda_2 - d_1, \lambda_3, \ldots, \lambda_{k+1}) \in M_k$ is $\sum_{i=1}^{k+1} \lambda_i - d_1 = \sum_{i=1}^{k+1} d_i - d_1 = \sum_{i=2}^{k+1} d_i$, so the induction hypothesis ensures that there is a nonsingular $S \in M_k$ such that the diagonal entries of SDS^{-1} are $d_2, \ldots, d_{k+1}$. Then $\left[\begin{smallmatrix} 1 & 0 \\ 0 & S \end{smallmatrix}\right] \left[\begin{smallmatrix} d_1 & \bigstar \\ \bigstar & D \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 0 \\ 0 & S \end{smallmatrix}\right]^{-1} = \left[\begin{smallmatrix} d_1 & \bigstar \\ \bigstar & SDS^{-1} \end{smallmatrix}\right]$ has diagonal entries $d_1, d_2, \ldots, d_{k+1}$.

If $\lambda_1, \ldots, \lambda_n$ and $d_1, \ldots, d_n$ are all real, all of the matrices and similarities in the preceding constructions are real. $\square$

Exercise. Write out the details of the inductive step $k = 2 \Rightarrow k = 3$ in the preceding proof.

Problems

1. Let $A, B \in M_n$. Suppose that A and B are diagonalizable and commute. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A and let $\mu_1, \ldots, \mu_n$ be the eigenvalues of B. (a) Show that the eigenvalues of $A+B$ are $\lambda_1 + \mu_{i_1}, \lambda_2 + \mu_{i_2}, \ldots, \lambda_n + \mu_{i_n}$, for some permutation $i_1, \ldots, i_n$ of $1, \ldots, n$. (b) If B is nilpotent, explain why A and $A + B$ have the same eigenvalues. (c) What are the eigenvalues of AB?

2. If $A, B \in M_n$ and if A and B commute, show that any polynomial in A commutes with any polynomial in B.

3. If $A \in M_n$, $SAS^{-1} = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, and $p(t)$ is a polynomial, show that $p(A) = S^{-1}p(\Lambda)S$ and that $p(\Lambda) = \text{diag}(p(\lambda_1), \ldots, p(\lambda_n))$. This provides a simple way to evaluate $p(A)$ if one can diagonalize A.

4. If $A \in M_n$ has distinct eigenvalues $\alpha_1, \ldots, \alpha_n$ and commutes with a given matrix $B \in M_n$, show that B is diagonalizable and that there is a polynomial $p(t)$ of degree at most $n - 1$ such that $B = p(A)$. *Hint:* Review the proof of Theorem (1.3.12) and show that B and A are simultaneously diagonalizable. See (0.9.11) and explain why there is a (Lagrange interpolating) polynomial $p(t)$ of degree at most $n-1$ such that the eigenvalues of B are $p(\alpha_1), \ldots, p(\alpha_n)$.

5. Give an example of two commuting matrices that are not simultaneously diagonalizable. Does this contradict Theorem (1.3.12)? Why?

6. (a) If $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, show that $p_\Lambda(\Lambda)$ is the zero matrix. (b) Suppose $A \in M_n$ is diagonalizable. Explain why $p_A(t) = p_\Lambda(t)$ and $p_\Lambda(A) = Sp_\Lambda(\Lambda)S^{-1}$. Conclude that $p_A(A)$ is the zero matrix.

7. A matrix $A \in M_n$ is a *square root* of $B \in M_n$ if $A^2 = B$. Show that every diagonalizable $B \in M_n$ has a square root. Does $B = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ have a square root? Why? *Hint*: If $A^2 = B$ and $Ax = \lambda x$, show that $\lambda^4 x = A^4 x = B^2 x = 0$ and explain why both eigenvalues of A are zero. Then $\operatorname{tr} A = 0$ so $A = \left[\begin{smallmatrix} a & b \\ c & -a \end{smallmatrix}\right]$. Also, $\det A = 0$, so $a^2 + bc = 0$. Then $A^2 = ?$

8. If $A, B \in M_n$ and if at least one has distinct eigenvalues (no assumption, even of diagonalizability, about the other), provide details for the following geometric argument that A and B commute if and only if they are simultaneously diagonalizable: One direction is easy; for the other, suppose B has distinct eigenvalues and $Bx = \lambda x$ with $x \neq 0$. Then $B(Ax) = A(Bx) = A\lambda x = \lambda Ax$, so $Ax = \mu x$ for some $\mu \in \mathbf{C}$ (Why? See (1.2.18)). Thus, we can diagonalize A with the same matrix of eigenvectors that diagonalizes B. Of course, the eigenvalues of A need not be distinct.

9. Consider the singular matrices $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right]$. Show that AB and BA are not similar, but that they do have the same eigenvalues.

10. Let $A \in M_n$ be given, and let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of A. For each $i = 1, 2, \ldots, k$ suppose $\{x_1^{(i)}, x_2^{(i)}, \ldots, x_{n_i}^{(i)}\}$ is an independent set of $n_i \geq 1$ eigenvectors of A corresponding to the eigenvalue λ_i. Show that $\{x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)}\} \cup \cdots \cup \{x_1^{(k)}, x_2^{(k)}, \ldots, x_{n_k}^{(k)}\}$ is an independent set. *Hint:* If some linear combination is zero, say $0 = \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij} x_j^{(i)} = \sum_{i=1}^k y^{(i)}$, use (1.3.8) to show that each $y^{(i)} = 0$.

11. Provide details for the following alternative proof of Lemma (1.3.19): (a) Suppose that $A, B \in M_n$ commute, $x \neq 0$, and $Ax = \lambda x$. Consider $x, Bx, B^2x, B^3x, \ldots$. Suppose $B^k x$ is the first element of this sequence that is dependent upon its predecessors; $\mathcal{S} = \operatorname{Span}\{x, Bx, B^2x, \ldots, B^{k-1}x\}$ is B-invariant and hence contains an eigenvector of B. But $AB^j x = B^j Ax = B^j \lambda x = \lambda B^j x$, so every nonzero vector in $\mathcal{S}$ is an eigenvector for A. Conclude that A and B have a common eigenvector. (b) If $\mathcal{F} = \{A_1, A_2 \ldots, A_m\} \subset M_n$ is a finite commuting family, use induction to show that it has a common eigenvector: If $y \neq 0$ is a common eigenvector for $A_1, A_2 \ldots, A_{m-1}$, consider

$y, A_m y, A_m^2 y, A_m^3 y, \ldots$. (c) If $\mathcal{F} \subset M_n$ is a non-finite commuting family, then no subset of $\mathcal{F}$ containing more than n^2 matrices can be linearly independent. Select a maximal independent set and explain why a common eigenvector for this finite set is a common eigenvector for $\mathcal{F}$.

12. Let $A, B \in M_n$, and suppose that either A or B is nonsingular. If AB is diagonalizable, show that BA is also diagonalizable. Consider $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ to show that this need not be true if both A and B are singular.

13. Show that two diagonalizable matrices are similar if and only if their characteristic polynomials are the same. Is this true for two matrices that are not both diagonalizable? *Hint*: Consider $\left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$.

14. Suppose $A \in M_n$ is diagonalizable. (a) Prove that the rank of A is equal to the number of its nonzero eigenvalues. (b) Prove that $\operatorname{rank} A = \operatorname{rank} A^k$ for all $k = 1, 2, \ldots$. (c) Prove that A is nilpotent if and only if $A = 0$. (d) If $\operatorname{tr} A = 0$, prove that $\operatorname{rank} A \neq 1$. (e) Use each of the four preceding results to show that $B = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ is not diagonalizable.

15. Let $A \in M_n$ and a polynomial $p(t)$ be given. If A is diagonalizable, show that $p(A)$ is diagonalizable. What about the converse?

16. Let $A \in M_n$ and suppose that $n > \operatorname{rank} A = r \geq 1$. If A is similar to $B \oplus 0_{n-r}$ (so $B \in M_r$ is nonsingular), show that A has a nonsingular r-by-r principal submatrix (that is, A is rank principal (0.7.6)). If A is rank principal, must it be similar to $B \oplus 0_{n-r}$? *Hint*: $p_A(t) = t^{n-r}(t^r - t^{r-1} \operatorname{tr} B + \cdots \pm \det B)$, so $E_r(A) \neq 0$. (1.2.13) Consider $\left[\begin{smallmatrix} 1 & i \\ i & -1 \end{smallmatrix}\right]$.

17. Let $A, B \in M_n$ be given. Prove that there is a nonsingular $T \in M_n(\mathbf{R})$ such that $A = TBT^{-1}$ if and only if there is a nonsingular $S \in M_n$ such that both $A = SBS^{-1}$ and $\bar{A} = S\bar{B}S^{-1}$. *Hint:* Let $S = C + iD$ with $C = (S + \bar{S})/2$ and $D = (S - \bar{S})/(2i)$. Then $AS = SB$ and $A\bar{S} = \bar{S}B$ imply that $AC = CB$ and $AD = DB$. Proceed as in the preceding problem.

18. Suppose $A, B \in M_n$ are coninvolutory, that is, $A\bar{A} = B\bar{B} = I$. Show that A and B are similar over $\mathbf{C}$ if and only if they are similar over $\mathbf{R}$.

19. Let $B, C \in M_n$ and define $\mathcal{A} = \left[\begin{smallmatrix} B & C \\ C & B \end{smallmatrix}\right] \in M_{2n}$. Let $Q = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} I_n & I_n \\ I_n & -I_n \end{smallmatrix}\right]$ and verify that $Q^{-1} = Q = Q^T$. Show that $Q^{-1}\mathcal{A}Q = (B + C) \oplus (B - C)$. Explain why (a) $\det \mathcal{A} = \det(B^2 + CB - BC - C^2)$; (b) $\operatorname{rank} \mathcal{A} = \operatorname{rank}(B + C) + \operatorname{rank}(B - C)$; (c) If C is nilpotent, then the eigenvalues of $\mathcal{A}$ are the eigenvalues of B, each with doubled multiplicity; (d) If B is symmetric and C is skew symmetric, then the eigenvalues of $\mathcal{A}$ are the eigenvalues of $B + C$, each with doubled multiplicity.

20. Represent any $A, B \in M_n$ as $A = A_1 + iA_2$ and $B = B_1 + iB_2$, in which $A_1, A_2, B_1, B_2 \in M_n(\mathbf{R})$. Define $R_1(A) = \left[\begin{smallmatrix} A_1 & A_2 \\ -A_2 & A_1 \end{smallmatrix}\right] \in M_{2n}(\mathbf{R})$. Show that:
(a) $A = B$ if and only if $R_1(A) = R_1(B)$, $R_1(A+B) = R_1(A) + R_1(B)$, $R_1(AB) = R_1(A)R_1(B)$, and $R(I_n) = I_{2n}$;
(b) if A is nonsingular then $R_1(A)$ is nonsingular, and $R_1(A^{-1}) = R_1(A)^{-1}$;
(c) if S is nonsingular, then $R_1(SAS^{-1}) = R_1(S)R_1(A)R_1(S)^{-1}$;
(d) if A and B are similar, then $R_1(A)$ and $R_1(B)$ are similar.
Let the eigenvalues of A be $\lambda_1, \ldots, \lambda_n$, let $S = \left[\begin{smallmatrix} I_n & iI_n \\ 0 & I_n \end{smallmatrix}\right]$, and let $U = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} I_n & iI_n \\ iI_n & I_n \end{smallmatrix}\right]$. Show that:
(e) $S^{-1} = \bar{S}$ and $U^{-1} = \bar{U} = U^*$;
(f) $S^{-1}R_1(A)S = \left[\begin{smallmatrix} A & 0 \\ -A_2 & \bar{A} \end{smallmatrix}\right]$ and $U^{-1}R_1(A)U = \left[\begin{smallmatrix} A & 0 \\ 0 & \bar{A} \end{smallmatrix}\right]$;
(g) the eigenvalues of $R_1(A)$ are the same as the eigenvalues of $A \oplus \bar{A}$, which are $\lambda_1, \ldots, \lambda_n, \overline{\lambda_1}, \ldots, \overline{\lambda_n}$;
(h) $\det R_1(A) = |\det A|^2$ and $\operatorname{rank} R_1(A) = 2 \operatorname{rank} A$;
(i) if $R_1(A)$ is nonsingular then A is nonsingular.
(j) iI_n is not similar to $-iI_n$, but $R_1(iI_n)$ is similar to $R_1(-iI_n)$, so the implication in (d) can not be reversed.
(k) $p_{R_1(A)}(t) = p_A(t)p_{\bar{A}}(t)$.
(l) $R_1(A^*) = R_1(A)^T$, so A is Hermitian if and only if $R_1(A)$ is (real) symmetric.
(m) A commutes with A^* if and only if $R_1(A)$ commutes with $R_1(A)^T$, that is, the complex matrix A is normal if and only if the real matrix $R_1(A)$ is normal. (2.5)
The block matrix $R_1(A)$ is an example of a *real representation* of A.

21. Using the same notation as in the preceding problem, define $R_2(A) = \left[\begin{smallmatrix} A_1 & A_2 \\ A_2 & -A_1 \end{smallmatrix}\right] \in M_{2n}(\mathbf{R})$. Let $V = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} -iI_n & -iI_n \\ I_n & -I_n \end{smallmatrix}\right]$, and consider $R_2(iI_n) = \left[\begin{smallmatrix} 0 & I_n \\ I_n & 0 \end{smallmatrix}\right]$ and $R_2(I_n) = \left[\begin{smallmatrix} I_n & 0 \\ 0 & -I_n \end{smallmatrix}\right]$. Show that:
(a) $V^{-1} = V^*$, $R_2(I_n)^{-1} = R_2(I_n) = R_2(I_n)^*$, $R_2(iI_n)^{-1} = R_2(iI_n) = R_2(iI_n)^*$, and $R_2(iI_n) = V^{-1}R_2(I_n)V$;
(b) $A = B$ if and only if $R_2(A) = R_2(B)$, and $R_2(A+B) = R_2(A) + R_2(B)$;
(c) $R_2(A) = V\left[\begin{smallmatrix} 0 & \bar{A} \\ A & 0 \end{smallmatrix}\right]V^{-1}$;
(d) $\det R_2(A) = (-1)^n|\det A|^2$;
(e) $R_2(A)$ is nonsingular if and only if A is nonsingular;
(f) $R_2(AB) = R_2(A \cdot I_n \cdot B) = R_2(A)R_2(I_n)R_2(B)$;
(g) $R_2(\bar{A}) = R_2(I_n)R_2(A)R_2(I_n)$, so $R_2(\bar{A})$ is similar to $R_2(A)$;
(h) $-R_2(A) = R_2(-A) = R_2(iI_n \cdot A \cdot iI_n) = (R_2(iI_n)R_2(I_n)) \cdot R_2(A) \cdot (R_2(iI_n)R_2(I_n))^{-1}$, so $R_2(-A)$ is similar to $R_2(A)$;
(i) $R_2(A)R_2(B) = V(\bar{A}B \oplus A\bar{B})V^{-1}$;

(j) if A is nonsingular, then $R_2(A)^{-1} = R_2(\bar{A}^{-1})$;
(k) $R_2(A)^2 = R_1(\bar{A}A)$;
(l) if S is nonsingular, then $R_2(S A \bar{S}^{-1}) = (R_2(S)R_2(I_n)) \cdot R_2(A) \cdot (R_2(S)R_2(I_n))^{-1}$, so $R_2(SA\bar{S}^{-1})$ is similar to $R_2(A)$. In fact, the converse is true: if $R_2(A)$ is similar to $R_2(B)$, then there is a nonsingular S such that $B = SA\bar{S}^{-1}$; see Chapter XXX.
(m) $R_2(A^T) = R_2(A)^T$, so A is (complex) symmetric if and only if $R_2(A)$ is (real) symmetric.
(n) A is unitary if and only if $R_2(A)$ is real orthogonal. *Hint*: (m) and (j).
The block matrix $R_2(A)$ is a second example of a *real representation* of A.

22. Let $A, B \in M_n$. Show that A and B are similar if and only if there are $X, Y \in M_n$, at least one of which is nonsingular, such that $A = XY$ and $B = YX$.

23. Let $B \in M_n$ and $C \in M_{n,m}$ and define $\mathcal{A} = \left[\begin{smallmatrix} B & C \\ 0 & 0_m \end{smallmatrix}\right] \in M_{n+m}$. Show that $\mathcal{A}$ is similar to $B \oplus 0_m$ if and only if $\operatorname{rank}[B\ C] = \operatorname{rank} B$, that is, if and only if there is some $X \in M_{n,m}$ such that $C = BX$. *Hint:* Consider a similarity of $\mathcal{A}$ via $\left[\begin{smallmatrix} I_n & X \\ 0 & I_m \end{smallmatrix}\right]$.

24. For a given integer $n \geq 3$, let $\theta = 2\pi/n$ and let $A = [\cos(j\theta + k\theta)]_{j,k=1}^n \in M_n(\mathbf{R})$. Show that $A = [x\ y][x\ y]^T$, in which $x = [\alpha\ \alpha^2\ \ldots\ \alpha^n]^T$, $y = [\alpha^{-1}\ \alpha^{-2}\ \ldots\ \alpha^{-n}]^T$, and $\alpha = e^{2\pi i/n}$. Show that the eigenvalues of A are $n/2$ and $-n/2$, together with $n-2$ zeroes.

25. Let $x, y \in \mathbf{C}^n$ be given and suppose that $y^*x \neq -1$. (a) Verify that $(I + xy^*)^{-1} = I - cxy^*$, in which $c = (1 + y^*x)^{-1}$. (b) Let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and suppose that $y^*x = 0$. Explain why the eigenvalues of

$$A = (I + xy^*)\Lambda(I - xy^*) = \Lambda + xy^*\Lambda - \Lambda xy^* - (y^*\Lambda x)xy^*$$

are $\lambda_1, \ldots, \lambda_n$. Notice that A has integer entries if the entries of x, y, and Λ are integers. Use this observation to construct an interesting 3-by-3 matrix with integer entries and eigenvalues 1, 2, and 7; verify that your construction has the asserted eigenvalues.

26. Let $e_1, ..., e_n$ and $\varepsilon_1, ..., \varepsilon_m$, denote the standard orthonormal bases of $\mathbf{C}^n$ and $\mathbf{C}^m$, respectively. Consider the n-by-m block matrix $P = [P_{ij}] \in M_{mn}$ in which each block $P_{ij} \in M_{m,n}$ is given by $P_{ij} = \varepsilon_j e_i^T$. (a) Show that P is a permutation matrix. (b) Similarity of any matrix $A \in M_{mn}$ by P gives a matrix $\tilde{A} \equiv PAP^T$ whose entries are a re-arrangement of the entries of A. Appropriate partitioning of both A and $\tilde{A}$ permits us to describe this re-arrangement in a simple way. Write $A = [A_{ij}] \in M_{mn}$ as an m-by-m block matrix in which each block $A_{kl} \equiv [a_{ij}^{(k,l)}] \in M_n$, and write $\tilde{A} = [\tilde{A}_{ij}]$ as an n-by-n block matrix in which each block $\tilde{A}_{ij} \in M_m$. Explain why the i, j entry of $\tilde{A}_{pq}$ is the p, q entry of A_{ij} for all $i, j = 1, ..., m$ and all $p, q = 1, ..., n$, that is, $\tilde{A}_{pq} = [a_{pq}^{(i,j)}]$. Since A and $\tilde{A}$ are permutation similar, they have the same eigenvalues, determinant, etc. *Hint:* $\tilde{A}_{pq} = \sum_{i,j=1}^{n} P_{pi} A_{ij} P_{qj}^T = \sum_{i,j=1}^{n} \left(e_p^T A_{ij} e_q\right) \varepsilon_i \varepsilon_j^T$ (c) Various special patterns in the entries of A result in special patterns in the entries of $\tilde{A}$ (and *vice versa*). For example, explain why: (i) All of the blocks A_{ij} are upper triangular if and only if $\tilde{A}$ is block upper triangular. (ii) All of the blocks A_{ij} are upper Hessenberg if and only if $\tilde{A}$ is block upper Hessenberg. (iii) All of the blocks A_{ij} are diagonal if and only if $\tilde{A}$ is block diagonal. (iv) A is block upper triangular and all of the blocks A_{ij} are upper triangular if and only if $\tilde{A}$ is block diagonal and all of its main diagonal blocks are upper triangular.

27. (Continuation of Problem 26) Let $A = [A_{kl}] \in M_{mn}$ be a given m-by-m block matrix with each $A_{kl} \equiv [a_{ij}^{(k,l)}] \in M_n$, and suppose each block A_{kl} is upper triangular. Explain why the eigenvalues of A are the same as those of $\tilde{A}_{11} \oplus \cdots \oplus \tilde{A}_{nn}$, in which $\tilde{A}_{pp} \equiv [a_{pp}^{(i,j)}]$ for $p = 1, ..., n$. Thus, the eigenvalues of A depend only on the main diagonal entries of the blocks A_{ij}. In particular, $\det A = (\det \tilde{A}_{11}) \cdots (\det \tilde{A}_{nn})$. What can you say about the eigenvalues and determinant of A if the diagonal entries of each block A_{ij} are constant (so there are scalars α_{kl} such that $a_{ii}^{(k,l)} = \alpha_{kl}$ for all $i = 1, ..., n$ and all $k, l = 1, ..., m$)?

28. Let $A \in M_{m,n}$ and $B \in M_{n,m}$ be given. Prove that $\det(I_m + AB) = \det(I_n + BA)$.

29. Let $A = [a_{ij}] \in M_n$. Suppose each $a_{ii} = 0$ for $i = 1, \ldots, n$ and $a_{ij} \in \{-1, 1\}$ for all $i \neq j$. Explain why $\det A$ is an integer. Use Cauchy's identity (1.3.24) to show that if any -1 entry of A is changed to $+1$, then the parity of $\det A$ is unchanged, that is it remains even if it was even and it remains odd if it was odd. Show that the parity of $\det A$ is the same as the parity of $\det(J_n - I)$, which is opposite to the parity of n. Conclude that A is nonsingular if n is even.

30. Suppose $A \in M_n$ is diagonalizable and $A = S\Lambda S^{-1}$ in which Λ has the form (1.3.13). If $f(z)$ is a complex valued function whose domain includes $\sigma(A)$, we *define* $f(A) \equiv Sf(\Lambda)S^{-1}$ in which $f(\Lambda) \equiv f(\mu_1)I_{n_1} \oplus \cdots \oplus f(\mu_d)I_{n_d}$. Does $f(A)$ depend on the choice of the diagonalizing similarity (which is never unique)? Use Theorem 1.3.27 to show that it does not, that is, if $A = S\Lambda S^{-1} = T\Lambda T^{-1}$, show that $Sf(\Lambda)S^{-1} = Tf(\Lambda)T^{-1}$. If A has real eigenvalues, show that $\cos^2(A) + \sin^2(A) = I$.

31. Let $a, b \in \mathbf{C}$. Show that the eigenvalues of $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ are $a \pm ib$.

32. Let $x \in \mathbf{C}^n$ be a given nonzero vector, and write $x = u + iv$, in which $u, v \in \mathbf{R}^n$. (0.2.5) Show that $\{x, \bar{x}\}$ is linearly independent (over $\mathbf{C}$) if and only if $\{u, v\}$ is linearly independent (over $\mathbf{R}$).

33. Suppose $A \in M_n(\mathbf{R})$ has a non-real eigenvalue λ and write $\lambda = a + ib$ with $a, b \in \mathbf{R}$ and $b \neq 0$; let x be an eigenvector associated with λ and write $x = u + iv$ with $u, v \in \mathbf{R}^n$. Explain why: u and v are both nonzero and $\bar{\lambda}$ is an eigenvalue of A associated with the eigenvector $\bar{x}$. Explain why: $\{x, \bar{x}\}$ is linearly independent (1.3.8) and hence $\{u, v\}$ is linearly independent. Show that $Au = \frac{1}{2}(Ax + A\bar{x}) = au - bv$ and $Av = bu + av$, so $A[u\, v] = [u\, v] \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$. Let $S = [u\ v\ S_1] \in M_n(\mathbf{R})$ be nonsingular and have u and v as its first and second columns. Explain why

$$\begin{aligned} S^{-1}AS &= S^{-1} \left[\ A[u\, v] \quad AS_1\ \right] = S^{-1} \left[\ [u\, v] \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \quad AS_1\ \right] \\ &= \begin{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} & \bigstar \\ 0 & A_1 \end{bmatrix}, \quad A_1 \in M_{n-2} \end{aligned}$$

Thus, a real square matrix with a non-real eigenvalue $\lambda = a + ib$ can be deflated by real similarity to a real block matrix in which the upper left 2-by-2 block $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ reveals the real and imaginary parts of λ.

34. If $A, B \in M_n$ are similar, show that $\operatorname{adj} A$ and $\operatorname{adj} B$ are similar.

35. A set $\mathcal{A} \subseteq M_n$ is an *algebra* if (i) $\mathcal{A}$ is a subspace, and (ii) $AB \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$. Provide details for the following assertions and assemble a proof of *Burnside's theorem on matrix algebras*: Let $n \geq 2$ and let $\mathcal{A} \subseteq M_n$ be a given algebra. Then $\mathcal{A} = M_n$ if and only if $\mathcal{A}$ is irreducible.
(a) If $n \geq 2$ and an algebra $\mathcal{A} \subseteq M_n$ is reducible, then $\mathcal{A} \neq M_n$. *Hint*: If $\mathcal{A}$ is reducible, use (1.3.17) to give an example of an $A \in M_n$ such that $A \notin \mathcal{A}$. This is the easy implication in Burnside's theorem; some work is required to show that if $\mathcal{A}$ is irreducible, then $\mathcal{A} = M_n$. In the following, $\mathcal{A} \subseteq M_n$ is a given algebra and $\mathcal{A}^* = \{A^* : A \in \mathcal{A}\}$.
(b) If $n \geq 2$ and $\mathcal{A}$ is irreducible, then $\mathcal{A} \neq \{0\}$. *Hint*: Every subspace is $\mathcal{A}$-invariant if $\mathcal{A} = \{0\}$.
(c) If $x \in \mathbf{C}^n$ is nonzero, then $\mathcal{A}x = \{Ax : A \in \mathcal{A}\}$ is an $\mathcal{A}$-invariant subspace of $\mathbf{C}^n$.
(d) If $n \geq 2$, $x \in \mathbf{C}^n$ is nonzero, and $\mathcal{A}$ is irreducible, then $\mathcal{A}x = \mathbf{C}^n$.
(e) For any given $x \in \mathbf{C}^n$, $\mathcal{A}^*x = \{A^*x : A \in \mathcal{A}\}$ is a subspace of $\mathbf{C}^n$.
(f) If $n \geq 2$, $x \in \mathbf{C}^n$ is nonzero, and $\mathcal{A}$ is irreducible, then $\mathcal{A}^*x = \mathbf{C}^n$. *Hint*:

If not, let z be a nonzero vector that is orthogonal to the subspace $\mathcal{A}^*x$, that is, $(A^*x)^*z = x^*Az = 0$ for all $A \in \mathcal{A}$. Use (d) to choose an $A \in \mathcal{A}$ such that $Az = x$.
(g) If $n \geq 2$ and $\mathcal{A}$ is irreducible, there is some $A \in \mathcal{A}$ such that $\operatorname{rank} A = 1$. *Hint*: If $d = \min\{\operatorname{rank} A : A \in \mathcal{A}$ and $A \neq 0\} > 1$, choose any $A_d \in \mathcal{A}$ with $\operatorname{rank} A_d = d$. Choose distinct i, j such that $\{Ae_i, Ae_j\}$ is linearly independent (pair of columns of A_d), so $A_d e_i \neq 0$ and $A_d e_j \neq \lambda A_d e_i$ for all $\lambda \in \mathbf{C}$. Choose $B \in \mathcal{A}$ such that $B(A_d e_i) = e_j$. Then $A_d B A_d e_i \neq \lambda A_d e_i$ for all $\lambda \in \mathbf{C}$. The range of A_d is $A_d B$-invariant ($A_d B(A_d x) = A_d(BA_d x)$) so it contains an eigenvector of $A_d B$ (1.3.18). Thus, there is an x such that $A_d x \neq 0$ and for some $\lambda_0 \in \mathbf{C}$, $(A_d B - \lambda_0 I)(A_d x) = 0$. Hence, $A_d B A_d - \lambda_0 A_d \in \mathcal{A}$, $A_d B A_d - \lambda_0 A_d \neq 0$, and $\operatorname{rank}(A_d B A_d - \lambda A_d) < d$. This contradiction implies that $d = 1$.
(h) If $n \geq 2$, $\mathcal{A}$ is irreducible, and there are nonzero $y, z \in \mathbf{C}^n$ such that $yz^* \in \mathcal{A}$, then $\mathcal{A}$ contains every rank one matrix. *Hint*: For any given nonzero $\eta, \zeta \in \mathbf{C}^n$, choose $A, B \in \mathcal{A}$ such that $\eta = Ay$ and $\zeta = B^*z$. Then $\eta\zeta^* = A(yz^*)B \in \mathcal{A}$.
(i) If $\mathcal{A}$ contains every rank one matrix, then $\mathcal{A} = M_n$. (0.4.4(i))

36. Suppose $A, B \in M_n$ are given and $n \geq 2$. The *algebra generated by* A *and* B is the span of the set of all words in A and B (2.2.5). (a) If A and B have no common eigenvector, explain why the algebra generated by A and B is all of M_n. (b) Let $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right]$. Show that A and B have no common eigenvector, so the algebra that they generate is all of M_2. Give a direct proof by exhibiting a basis of M_2 consisting of words in A and B.

Further Readings and Notes: Theorem 1.3.29 is due to L. Mirsky (1958); our proof is adapted from E. Carlen and E. Lieb, Short proofs of theorems of Horn and Mirsky on diagonals and eigenvalues of matrices, *Electron. J. Linear Algebra* XXXX. The proof of Burnside's theorem in Problem 35 is adapted from I. Halperin and P. Rosenthal, Burnside's Theorem on Algebras of Matrices, *Amer. Math. Monthly* 87 (1980) 810. For alternative approaches, see [RadRos] and V. Lomonosov and P. Rosenthal, The simplest proof of Burnside's theorem on matrix algebras, *Linear Algebra Appl.* 383 (2004) 45-47.

1.4 Left and right eigenvectors, and geometric multiplicity

The eigenvectors of a matrix are important not only for their role in diagonalization, but also for their utility in a variety of applications. We begin with an important observation about eigenvalues.

1.4.1 Observation. Let $A \in M_n$. (a) The eigenvalues of A and A^T are the same. (b) The eigenvalues of A^* are the complex conjugates of the eigenvalues of A.

Proof: Since $\det(tI - A^T) = \det(tI - A)^T = \det(tI - A)$, we have $p_{A^T}(t) = p_A(t)$, so $p_{A^T}(\lambda) = 0$ if and only if $p_A(\lambda) = 0$. Similarly, $\det(\bar{t}I - A^*) = \det[(tI - A)^*] = \overline{\det(tI - A)}$, so $p_{A^*}(\bar{t}) = \overline{p_A(t)}$, and $p_{A^*}(\bar{\lambda}) = 0$ if and only if $p_A(\lambda) = 0$. $\square$

Exercise. If $x, y \in \mathbf{C}^n$ are both eigenvectors of $A \in M_n$ corresponding to the eigenvalue λ, show that any nonzero linear combination of x and y is also an eigenvector corresponding to λ. Conclude that the set of all eigenvectors associated with a particular $\lambda \in \sigma(A)$, together with the zero vector, is a *subspace* of $\mathbf{C}^n$.

Exercise. The subspace described in the preceding exercise is the *null space* of $A - \lambda I$, that is, the solution set of the homogeneous linear system $(A - \lambda I)x = 0$. Explain why the dimension of this subspace is $n - \operatorname{rank}(A - \lambda I)$.

1.4.2 Definition. Let $A \in M_n$. For a given $\lambda \in \sigma(A)$, the set of all vectors $x \in \mathbf{C}^n$ satisfying $Ax = \lambda x$ is called the *eigenspace* of A corresponding to the eigenvalue λ. Every nonzero element of this eigenspace is an eigenvector of A corresponding to λ.

Exercise. Show that the eigenspace of A corresponding to an eigenvalue λ is an A-invariant subspace, but an A-invariant subspace need not be an eigenspace of A. Explain why a *minimal* A-invariant subspace (an A-invariant subspace that contains no strictly lower-dimensional, nonzero A-invariant subspace) W is the span of a *single* eigenvector of A, that is, $\dim W = 1$.

1.4.3 Definition. Let $A \in M_n$ and let λ be an eigenvalue of A. The dimension of the eigenspace of A corresponding to λ is the *geometric multiplicity* of λ. The multiplicity of λ as a zero of the characteristic polynomial of A is the *algebraic multiplicity* of λ. If the term *multiplicity* is used without qualification in reference to λ, it means the algebraic multiplicity. We say that λ is *simple* if its algebraic multiplicity is one; it is *semisimple* if its algebraic and geometric multiplicities are equal.

It is very useful to be able to think of the geometric multiplicity of an eigenvalue λ of $A \in M_n$ in more than one way: Since the geometric multiplicity is the dimension of the nullspace of $A - \lambda I$, it is equal to $n - \operatorname{rank}(A - \lambda I)$. It is also the maximum number of linearly independent eigenvectors associated with λ. Theorems 1.2.18 and 1.3.7 both contain an inequality between the geometric and algebraic multiplicities of an eigenvalue, but from two different viewpoints.

Exercise. Use Theorem 1.2.18 to explain why the algebraic multiplicity of an eigenvalue is greater than or equal to its geometric multiplicity. If the algebraic multiplicity is 1, why must the geometric multiplicity also be 1?

Exercise. Use Theorem 1.3.7 to explain why the geometric multiplicity of an eigenvalue is less than or equal to its algebraic multiplicity. If the algebraic multiplicity is 1, why must the geometric multiplicity also be 1?

Exercise. What are the algebraic and geometric multiplicities of the eigenvalue 0 of the matrix $\left[\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right]$?

1.4.4 Definitions. Let $A \in M_n$. We say that A is *defective* if the geometric multiplicity of some eigenvalue of A is strictly less than its algebraic multiplicity. If the geometric multiplicity of each eigenvalue of A is the same as its algebraic multiplicity, we say that A is *nondefective*. If each eigenvalue of A has geometric multiplicity 1, we say that A is *nonderogatory*.

A matrix is diagonalizable if and only if it is nondefective; it has distinct eigenvalues if and only if it is nonderogatory and nondefective.

1.4.5 Example. Even though A and A^T have the same eigenvalues, their eigenvectors corresponding to a given eigenvalue can be very different. For example, let $A = \left[\begin{smallmatrix} 2 & 3 \\ 0 & 4 \end{smallmatrix}\right]$. Then the (one-dimensional) eigenspace of A corresponding to the eigenvalue 2 is spanned by $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, while the corresponding eigenspace of A^T is spanned by $\left[\begin{smallmatrix} 1 \\ -3/2 \end{smallmatrix}\right]$.

1.4.6 Definition. A nonzero vector $y \in \mathbf{C}^n$ is a *left eigenvector* of $A \in M_n$ corresponding to an eigenvalue λ of A if $y^*A = \lambda y^*$. If necessary for clarity, we refer to the vector x in (1.1.3) as a *right eigenvector*; when the context does not require distinction, we continue to call x an *eigenvector*.

Exercise. Show that a left eigenvector y corresponding to an eigenvalue λ of $A \in M_n$ is a right eigenvector of A^* corresponding to $\bar{\lambda}$; also show that $\bar{y}$ is a right eigenvector of A^T corresponding to λ.

Exercise. Example (1.4.5) shows that a right eigenvector x of $A \in M_n$ need not also be a left eigenvector. But if it is, the corresponding right and left eigenvalues must be the same. Why? That is, if $x \neq 0$, $Ax = \lambda x$, and $x^*A = \mu x^*$, then why is $\lambda = \mu$? *Hint:* Evaluate x^*Ax in two ways.

Exercise. Suppose $A \in M_n$ is diagonalizable, S is nonsingular, and $S^{-1}AS = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Partition $S = [x_1 \ \ldots \ x_n]$ and $S^{-*} = [y_1 \ \ldots \ y_n]$ (0.2.5) according to their columns. The identity $AS = S\Lambda$ tells us that each column x_j of S is a right eigenvector of A corresponding to the eigenvalue λ_j. Explain why $(S^{-*})^*A = \Lambda(S^{-*})^*$, why each column y_j of S^{-*} is a left eigenvector of A corresponding to the eigenvalue λ_j, why $y_j^*x_j = 1$ for each $j = 1, \ldots, n$, and why $y_i^*x_j = 0$ whenever $i \neq j$.

One should not dismiss left eigenvectors as merely a parallel theoretical alternative to right eigenvectors. Each type of eigenvector can convey different information about a matrix, and it can be very useful to know how the two types of eigenvectors interact. In the preceding exercise, we learned that a diagonalizable $A \in M_n$ has n pairs of non-orthogonal left and right eigenvectors, and that left and right eigenvectors associated with different eigenvalues are orthogonal. We next examine a version of these results for matrices that are not necessarily diagonalizable.

1.4.7 Theorem. Let $A \in M_n$ and nonzero vectors $x, y \in \mathbf{C}^n$ be given.
(a) Suppose that $Ax = \lambda x$, $y^*A = \mu y^*$, and $\lambda \neq \mu$. Then $y^*x = 0$.
(b) Suppose that $Ax = \lambda x$, $y^*A = \lambda y^*$, and $y^*x \neq 0$. Then A is similar to

$$\begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix} \qquad \text{for some } B \in M_{n-1} \tag{1.4.8}$$

Conversely, if A is similar to a block matrix of the form (1.4.8), then it has a non-orthogonal pair of left and right eigenvectors associated with the eigenvalue λ.

Proof: (a) Let y be a left eigenvector of A corresponding to μ and let x be a right eigenvector of A corresponding to λ. Manipulate y^*Ax in two ways:

$$\begin{aligned} y^*Ax &= y^*(\lambda x) = \lambda(y^*x) \\ &= (\mu y^*)x = \mu(y^*x) \end{aligned}$$

Since $\lambda \neq \mu$, $\lambda y^*x = \mu y^*x$ only if $y^*x = 0$.
(b) Suppose that $Ax = \lambda x$, $y^*A = \lambda y^*$, and $y^*x \neq 0$. If we replace x by $x/(y^*x)$, we may assume that $y^*x = 1$. Let the columns of $S_1 \in M_{n,n-1}$ be any basis for the orthogonal complement of y (so $y^*S_1 = 0$) and consider $S = [x \ S_1] \in M_n$. Let $z = [z_1 \ \zeta^T]^T$ with $\zeta \in \mathbf{C}^{n-1}$ and suppose $Sz = 0$. Then

$$0 = y^*Sz = y^*(z_1 x + S_1\zeta) = z_1(y^*x) + (y^*S_1)\zeta = z_1$$

so $z_1 = 0$ and $0 = Sz = S_1\zeta$, which implies that $\zeta = 0$ since S_1 has full column rank. We conclude that S is nonsingular. Partition $S^{-*} = [\eta\ Z_1]$ with $\eta \in \mathbb{C}^n$ and compute

$$I_n = S^{-1}S = \begin{bmatrix} \eta^* \\ Z_1^* \end{bmatrix} \begin{bmatrix} x & S_1 \end{bmatrix} = \begin{bmatrix} \eta^* x & \eta^* S_1 \\ Z_1^* x & Z_1^* S_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I_{n-1} \end{bmatrix}$$

which contains four identities. The identity $\eta^* S_1 = 0$ implies that η is orthogonal to the orthogonal complement of y, so $\eta = \alpha y$ for some scalar α. The identity $\eta^* x = 1$ tells us that $\eta^* x = (\alpha y)^* x = \bar{\alpha}(y^* x) = \bar{\alpha} = 1$, so $\eta = y$. Using the identities $\eta^* S_1 = y^* S_1 = 0$ and $Z_1^* x = 0$ as well as the eigenvector properties of x and y, compute the similarity

$$\begin{aligned} S^{-1}AS &= \begin{bmatrix} y^* \\ Z_1^* \end{bmatrix} A \begin{bmatrix} x & S_1 \end{bmatrix} = \begin{bmatrix} y^* Ax & y^* AS_1 \\ Z_1^* Ax & Z_1^* AS_1 \end{bmatrix} \\ &= \begin{bmatrix} (\lambda y^*)x & (\lambda y^*)S_1 \\ Z_1^*(\lambda x) & Z_1^* AS_1 \end{bmatrix} = \begin{bmatrix} \lambda(y^* x) & \lambda(y^* S_1) \\ \lambda(Z_1^* x) & Z_1^* AS_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda & 0 \\ 0 & Z_1^* AS_1 \end{bmatrix} \end{aligned}$$

which verifies that A is similar to a matrix of the form (1.4.8).

Conversely, suppose there is a nonsingular S such that $A = S([\lambda] \oplus B)S^{-1}$. Let x be the first column of S, let y be the first column of S^{-*}, and partition $S = [x\ S_1]$ and $S^{-*} = [y\ Z_1]$. The $1,1$ entry of the identity $S^{-1}S = I$ tells us that $y^* x = 1$; the first column of the identity

$$[Ax\ AS_1] = AS = S([\lambda] \oplus B) = [\lambda x\ S_1 B]$$

tells us that $Ax = \lambda x$; and the first row of the identity

$$\begin{bmatrix} y^* A \\ Z_1^* A \end{bmatrix} = S^{-1}A = ([\lambda] \oplus B)S^{-1} = \begin{bmatrix} \lambda y^* \\ BZ_1^* \end{bmatrix}$$

tells us that $y^* A = \lambda y^*$. □

The assertion in Theorem 1.4.7(a) is the *principle of biorthogonality*. One might also ask what happens if left and right eigenvectors corresponding to the same eigenvalue are either orthogonal or equal; these cases are discussed in Theorem 2.4.11.1.

Eigenvectors transform under similarity in a simple way. The eigenvalues are, of course, unchanged by similarity.

1.4.9 Theorem. Let $A, B \in M_n$ and suppose $B = S^{-1}AS$ for some nonsingular S. If $x \in \mathbf{C}^n$ is a right eigenvector of B corresponding to an eigenvalue λ, then Sx is a right eigenvector of A corresponding to λ. If $y \in \mathbf{C}^n$ is a left eigenvector of B corresponding to λ, then $S^{-*}y$ is a left eigenvector of A corresponding to λ.

Proof: If $Bx = \lambda x$, then $S^{-1}ASx = \lambda x$, or $A(Sx) = \lambda(Sx)$. Since S is nonsingular and $x \neq 0$, $Sx \neq 0$, and hence Sx is an eigenvector of A. If $y^*B = \lambda y^*$, then $y^*S^{-1}AS = \lambda y^*$, or $(S^{-*}y)^*A = \lambda(S^{-*}y)^*$. □

Information about eigenvalues of principal submatrices can refine the basic observation that the algebraic multiplicity of an eigenvalue can not be less than its algebraic multiplicity.

1.4.10 Theorem. Let $A \in M_n$ and $\lambda \in \mathbf{C}$ be given, and let $k \geq 1$ be a given positive integer. Consider the following three statements:

(a) λ is an eigenvalue of A with geometric multiplicity at least k.
(b) For each $m = n - k + 1, \ldots, n$, λ is an eigenvalue of every m-by-m principal submatrix of A.
(c) λ is an eigenvalue of A with algebraic multiplicity at least k.

Then (a) implies (b), and (b) implies (c). In particular, the algebraic multiplicity of an eigenvalue is at least as great as its geometric multiplicity.

Proof: (a) $\Rightarrow$ (b): Let λ be an eigenvalue of A with geometric multiplicity at least k, which means that $\operatorname{rank}(A - \lambda I) \leq n - k$. Suppose $m > n - k$. Then every m-by-m minor of $A - \lambda I$ is zero. In particular, every principal m-by-m minor of $A - \lambda I$ is zero, so every m-by-m principal submatrix of $A - \lambda I$ is singular. Thus, λ is an eigenvalue of every m-by-m principal submatrix of A. (b) $\Rightarrow$ (c): Suppose λ is an eigenvalue of every m-by-m principal submatrix of A for each $m \geq n - k + 1$. Then every principal minor of $A - \lambda I$ of size at least $n - k + 1$ is zero, so each principal minor sum $E_j(A - \lambda I) = 0$ for all $j \geq n - k + 1$. Then (1.2.13) and (1.2.11) ensure that $p^{(i)}_{A-\lambda I}(0) = 0$ for $i = 0, 1, \ldots, k - 1$. But $p_{A-\lambda I}(t) = p_A(t + \lambda)$, so $p^{(i)}_A(\lambda) = 0$ for $i = 0, 1, \ldots, k - 1$, that is, λ is a zero of $p_A(t)$ with multiplicity at least k. □

An eigenvalue λ with geometric multiplicity one can have algebraic multiplicity two or more, but this can happen only if the left and right eigenvectors associated with λ are orthogonal. If λ has algebraic multiplicity one, however, then it has geometric multiplicity one and left and right eigenvectors associated with λ can never be orthogonal. Our approach to these results relies on the following lemma.

1.4.11 Lemma. Let $A \in M_n$, $\lambda \in \mathbf{C}$, and nonzero vectors $x, y \in \mathbf{C}^n$ be given. Suppose that $Ax = \lambda x$, $y^*A = \lambda y^*$, and λ has geometric multiplicity one. Then there is a nonzero $\gamma \in \mathbf{C}$ such that $\operatorname{adj}(\lambda I - A) = \gamma x y^*$.

Proof: We have $\operatorname{rank}(\lambda I - A) = n - 1$ and hence $\operatorname{rank} \operatorname{adj}(\lambda I - A) = 1$, that is, $\operatorname{adj}(\lambda I - A) = \xi \eta^*$ for some nonzero $\xi, \eta \in \mathbf{C}^n$; see (0.8.2). But $(\lambda I - A)(\operatorname{adj}(\lambda I - A)) = \det(\lambda I - A)I = 0$, so $(\lambda I - A)\xi\eta^* = 0$ and $(\lambda I - A)\xi = 0$, which implies that $\xi = \alpha x$ for some nonzero scalar α. Using the identity $(\operatorname{adj}(\lambda I - A))(\lambda I - A) = 0$ in a similar fashion, we conclude that $\eta = \beta y$ for some nonzero scalar β. Thus, $\operatorname{adj}(\lambda I - A) = \alpha\beta x y^*$. □

1.4.12 Theorem. Let $A \in M_n$, $\lambda \in \mathbf{C}$, and nonzero vectors $x, y \in \mathbf{C}^n$ be given. Suppose that $Ax = \lambda x$ and $y^*A = \lambda y^*$. (a) If λ has algebraic multiplicity one, then $y^*x \neq 0$. (b) If λ has geometric multiplicity one, then it has algebraic multiplicity one if and only if $y^*x \neq 0$.

Proof: In both cases (a) and (b), λ has geometric multiplicity one; the preceding lemma tells us that there is a nonzero $\gamma \in \mathbf{C}$ such that $\operatorname{adj}(\lambda I - A) = \gamma x y^*$. Then $p_A(\lambda) = 0$ and $p_A'(\lambda) = \operatorname{tr} \operatorname{adj}(\lambda I - A) = \gamma y^* x$; see (0.8.10.2). In (a) we assume that the algebraic multiplicity is one, so $p_A'(\lambda) \neq 0$ and hence $y^*x \neq 0$. In (b) we assume that $y^*x \neq 0$, so $p_A'(\lambda) \neq 0$ and hence the algebraic multiplicity is one. □

Problems

1. Let nonzero vectors $x, y \in M_n$ be given, let $A = xy^*$, and let $\lambda = y^*x$. Show that: (a) λ is an eigenvalue of A; (b) x is a right and y is a left eigenvector of A corresponding to λ; and (c) if $\lambda \neq 0$, then it is the *only* nonzero eigenvalue of A (algebraic multiplicity = 1). Explain why any vector that is orthogonal to y is in the null space of A. What is the geometric multiplicity of the eigenvalue 0? Explain why A is diagonalizable if and only if $y^*x \neq 0$.

2. Let $A \in M_n$ be skew-symmetric. Show that $p_A(t) = (-1)^n p_A(-t)$ and deduce that if λ is an eigenvalue of A with multiplicity k then so is $-\lambda$. If n is odd, explain why A must be singular. Explain why every principal minor of A with odd size is singular. Use the fact that a skew-symmetric matrix is rank principal (0.7.6) to show that $\operatorname{rank} A$ must be even.

3. Suppose $n \geq 2$ and let $T = [t_{ij}] \in M_n$ be upper triangular.(a) Let x be an eigenvector of T corresponding to the eigenvalue t_{nn}; explain why e_n is a left eigenvector corresponding to t_{nn}. If $t_{ii} \neq t_{nn}$ for each $i = 1, \ldots, n-1$, show that the last entry of x *must* be nonzero. (b) Let $k \in \{1, \ldots, n-1\}$. Show that there is an eigenvector x of T corresponding to the eigenvalue t_{kk} whose last $n-k$ entries of x are zero, that is, $x^T = [\xi^T\ 0]^T$ with $\xi \in \mathbf{C}^k$. If $t_{ii} \neq t_{kk}$ for all $i = 1, \ldots, k-1$, explain why the kth entry of x *must* be nonzero.

4. Suppose $A \in M_n$ is tridiagonal and has a zero main diagonal. Let $S = \operatorname{diag}(-1, 1, -1, \ldots, (-1)^n)$ and show that $S^{-1}AS = -A$. If λ is an eigenvalue of A with multiplicity k, explain why $-\lambda$ is also an eigenvalue of A with multiplicity k. If n is odd, show that A is singular.

5. Consider the block triangular matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \qquad A_{ii} \in M_{n_i}, \quad i = 1, 2$$

If $x \in \mathbf{C}^{n_1}$ is a right eigenvector of A_{11} corresponding to $\lambda \in \sigma(A_{11})$, and if $y \in \mathbf{C}^{n_2}$ is a left eigenvector of A_{22} corresponding to $\mu \in \sigma(A_{22})$, show that $\left[\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right] \in \mathbf{C}^{n_1+n_2}$ is a right eigenvector, and $\left[\begin{smallmatrix} 0 \\ y \end{smallmatrix}\right]$ is a left eigenvector, of A corresponding to λ and μ, respectively. Use this observation to show that the eigenvalues of A are the eigenvalues of A_{11} together with those of A_{22}.

6. Suppose $A \in M_n$ has an entry-wise positive left eigenvector and an entry-wise positive right eigenvector, both corresponding to a given eigenvalue λ with geometric multiplicity 1. (a) Show that A has no entry-wise nonnegative left or right eigenvectors corresponding to any eigenvalue different from λ. (b) If λ has geometric multiplicity one, show that it has algebraic multiplicity one.

7. In this problem we outline a simple version of the *power method* for finding the largest modulus eigenvalue and an associated eigenvector of $A \in M_n$. Suppose that $A \in M_n$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ and that there is exactly one eigenvalue λ_n of maximum modulus $\rho(A)$. If $x^{(0)} \in \mathbf{C}^n$ is *not* orthogonal to a left eigenvector associated with λ_n, show that the sequence

$$x^{(k+1)} = \frac{1}{(x^{(k)*}x^{(k)})^{1/2}} Ax^{(k)}, \quad k = 0, 1, 2, \ldots$$

approaches an eigenvector of A and the ratios of a given nonzero entry in the vectors $Ax^{(k)}$ and $x^{(k)}$ approach λ_n. *Hint:* Assume without loss of generality that $\lambda_n = 1$ and let $y^{(1)}, \ldots, y^{(n)}$ be linearly independent eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$. Write $x^{(0)}$ (uniquely: Why?) as $x^{(0)} = \alpha_1 y^{(1)} + \cdots + \alpha_n y^{(n)}$, with $\alpha_n \neq 0$. Then $x^{(k)} = c_k(\alpha_1 \lambda_1^k y^{(1)} + \cdots + \alpha_{n-1}\lambda_{n-1}^k y^{(n-1)} + \alpha_n y^{(n)})$ for some scalar $c_k \neq 0$. Since $|\lambda_i|^k \to 0, i = 1, \ldots, n-1$, $x^{(k)}$ converges to a scalar multiple of $y^{(n)}$.

8. Continue with the assumptions and notation of Problem 7. Further eigenvalues (and eigenvectors) of A can be calculated by combining the power method with *deflation*, which delivers a square matrix of size one smaller. Let $S \in M_n$ be nonsingular and have as its first column an eigenvector $y^{(n)}$ associated with the eigenvalue λ_n, which have been computed by the power method or otherwise. Show that

$$S^{-1}AS = \left[\begin{array}{c:c} \lambda_n & * \\ \hdashline 0 & A_1 \end{array}\right]$$

and that the eigenvalues of $A_1 \in M_{n-1}$ are $\lambda_1, \ldots, \lambda_{n-1}$. Another eigenvalue may be calculated from A_1 and the deflation repeated.

9. Let $A \in M_n$ have eigenvalues $\lambda_1, \ldots, \lambda_{n-1}, 0$, so that rank $A \leq n-1$, and suppose that the last row of A is a linear combination of the others.
(a) Partition

$$A = \begin{bmatrix} A_{11} & x \\ y^T & \alpha \end{bmatrix}$$

in which $A_{11} \in M_{n-1}$. Explain why there is a $b \in \mathbf{C}^{n-1}$ such that $y^T = b^T A_{11}$ and $\alpha = b^T x$. Interpret b in terms of a left eigenvector of A corresponding to the eigenvalue 0. (b) Show that $A_{11} + xb^T \in M_{n-1}$ has eigenvalues $\lambda_1, \ldots, \lambda_{n-1}$. This is another version of deflation: it produces a matrix of smaller size with the remaining eigenvalues. If one eigenvalue λ of A is known, then the process described in this problem can be applied to $P(A - \lambda I)P^{-1}$, for a suitable permutation P. *Hint:* Consider $S^{-1}AS$ with $S = \left[\begin{smallmatrix} I & 0 \\ b^T & 1 \end{smallmatrix}\right]$.

10. Let $T \in M_n$ be a nonsingular matrix whose columns are left eigenvectors of $A \in M_n$. Show that the columns of T^{-*} are right eigenvectors of A.

11. Suppose $A \in M_n$ is an unreduced upper Hessenberg matrix (0.9.9). Explain why $\operatorname{rank}(A - \lambda I) \geq n - 1$ for every $\lambda \in \mathbf{C}$ and deduce that every eigenvalue of A has geometric multiplicity one, that is, A is nonderogatory.

12. Let λ be an eigenvalue of $A \in M_n$. (a) Show that every set of $n-1$ columns of $A - \lambda I$ is linearly independent if and only if no eigenvector of A associated with λ has a zero entry. (b) If no eigenvector of A associated with λ has a zero entry, why must λ have geometric multiplicity one?

13. Under the hypotheses of Lemma 1.4.10, and assuming that the eigenvalues of A are $\lambda, \lambda_2, \ldots, \lambda_n$ and $\lambda \neq \lambda_i$ for all $i = 2, \ldots, n$, show that $\gamma = (\lambda - \lambda_2)(\lambda - \lambda_3) \cdots (\lambda - \lambda_n)/y^* x$.

14. Let $A \in M_n$ and let $t \in \mathbf{C}$. Explain why $(A - tI)\operatorname{adj}(A - tI) = \operatorname{adj}(A - tI)(A - tI) = p_A(t)I$. Now suppose λ is an eigenvalue of A. Show that: (a) every nonzero column of $\operatorname{adj}(A - \lambda I)$ is an eigenvector of A associated with λ; (b) every nonzero row of $\operatorname{adj}(A - \lambda I)$ is the conjugate transpose of a left eigenvector of A associated with λ; (c) $\operatorname{adj}(A-\lambda I) \neq 0$ if and only if λ has geometric multiplicity one; (d) If λ is an eigenvalue of $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, then each nonzero column of $\left[\begin{smallmatrix} d-\lambda & -b \\ -c & a-\lambda \end{smallmatrix}\right]$ is an eigenvector of A associated with λ; each nonzero row is the conjugate transpose of a left eigenvector of A associated with λ.

15. Suppose that λ is a simple eigenvalue of $A \in M_n$, and suppose that $x, y, z, w \in \mathbf{C}^n$, $Ax = \lambda x$, $y^*A = \lambda y^*$, $y^*z \neq 0$, and $w^*x \neq 0$. Show that $A - \lambda I + \kappa z w^*$ is nonsingular for all $\kappa \neq 0$. Explain why it is possible to take $z = x$. *Hint:* $(A - \lambda I + \kappa z w^*)u = 0 \Rightarrow \kappa(w^*u)y^*z = 0 \Rightarrow w^*u = 0 \Rightarrow u = \alpha x \Rightarrow w^*x = 0$.

16. Show that the complex tridiagonal Toeplitz matrix

$$A = \begin{bmatrix} a & b & & \\ c & a & \ddots & \\ & \ddots & \ddots & b \\ & & c & a \end{bmatrix} \in M_n, \quad bc \neq 0 \qquad (1.4.13)$$

is diagonalizable and has spectrum $\sigma(A) = \{a + 2\sqrt{bc}\cos(\frac{\pi\kappa}{n+1}) : \kappa = 1, \ldots, n\}$, in which $\operatorname{Re}\sqrt{bc} \geq 0$ and $\operatorname{Im}\sqrt{bc} > 0$ if bc is real and negative. *Hint*: Let $\lambda, x = [x_i]_{i=1}^n$ be an eigenvalue-eigenvector pair for A. Then $(A - \lambda I)x = 0 \Rightarrow cx_{k-1} + (a - \lambda)x_k + bx_{k+1} = 0 \Rightarrow x_{k+1} + \frac{a-\lambda}{b}x_k + \frac{c}{b}x_{k-1} = 0, k = 1, \ldots, n$, which is a second order difference equation with boundary conditions $x_0 = x_{n+1} = 0$, and indicial equation $t^2 + \frac{a-\lambda}{b}t + \frac{c}{b} = 0$ with roots r_1 and r_2. The general solution of the difference equation is (a) $x_k = \alpha r_1^k + \beta r_2^k$ if $r_1 \neq r_2$, or (b) $x_k = \alpha r_1^k + k\beta r_1^k$ if $r_1 = r_2$; α and β are determined by the boundary conditions. In either case, $r_1 r_2 = c/b$ (so $r_1 \neq 0 \neq r_2$) and $r_1 + r_2 = -(a-\lambda)/b$ (so $\lambda = a + b(r_1 + r_2)$). If $r_1 = r_2$ then $0 = x_0 = \alpha$ and $0 = x_{n+1} = (n+1)\beta r_1^{n+1} \Rightarrow x = 0$. Thus, $x_k = \alpha r_1^k + \beta r_2^k$ so $0 = x_0 = \alpha + \beta$ and $0 = x_{n+1} = \alpha(r_1^{n+1} - r_2^{n+1}) \Rightarrow (r_1/r_2)^{n+1} = 1 \Rightarrow r_1/r_2 = e^{\frac{2\pi i\kappa}{n+1}}$ for some $\kappa \in \{1, \ldots, n\}$. Since $r_1 r_2 = c/b$ we have $r_1 = \pm\sqrt{c/b}\, e^{\frac{\pi i\kappa}{n+1}}$ and $r_1 = \pm\sqrt{c/b}\, e^{\frac{-\pi i\kappa}{n+1}}$ (same choice of signs). Thus, $\lambda = a + b(r_1 + r_2) = a \pm 2\sqrt{bc}\cos(\frac{\pi\kappa}{n+1})$.

17. If $a = 2$ and $b = c = -1$ in (1.4.13), show that $\sigma(A) = \{4\sin^2(\frac{\pi\kappa}{2(n+1)}) : \kappa = 1, \ldots, n\}$.

Index