CHAPTER 7

Eigenvalues and Eigenvectors

7.1 ELEMENTARY PROPERTIES OF EIGENSYSTEMS

Up to this point, almost everything was either motivated by or evolved from the consideration of systems of linear *algebraic* equations. But we have come to a turning point, and from now on the emphasis will be different. Rather than being concerned with systems of *algebraic* equations, many topics will be motivated or driven by applications involving systems of linear *differential* equations and their discrete counterparts, difference equations.

For example, consider the problem of solving the system of two first-order linear differential equations, $du_1/dt = 7u_1 - 4u_2$ and $du_2/dt = 5u_1 - 2u_2$. In matrix notation, this system is

$$\begin{pmatrix} u_1' \\ u_2' \end{pmatrix} = \begin{pmatrix} 7 & -4 \\ 5 & -2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \text{or, equivalently,} \quad \mathbf{u}' = \mathbf{A}\mathbf{u}, \tag{7.1.1}$$

where $\mathbf{u}' = \begin{pmatrix} u_1' \\ u_2' \end{pmatrix}$, $\mathbf{A} = \begin{pmatrix} 7 & -4 \\ 5 & -2 \end{pmatrix}$, and $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$. Because solutions of a single equation $u' = \lambda u$ have the form $u = \alpha e^{\lambda t}$, we are motivated to seek solutions of (7.1.1) that also have the form

$$u_1 = \alpha_1 e^{\lambda t} \quad \text{and} \quad u_2 = \alpha_2 e^{\lambda t}. \tag{7.1.2}$$

Differentiating these two expressions and substituting the results in (7.1.1) yields

$$\begin{aligned} \alpha_1 \lambda e^{\lambda t} &= 7\alpha_1 e^{\lambda t} - 4\alpha_2 e^{\lambda t} \\ \alpha_2 \lambda e^{\lambda t} &= 5\alpha_1 e^{\lambda t} - 2\alpha_2 e^{\lambda t} \end{aligned} \Rightarrow \begin{aligned} \alpha_1 \lambda &= 7\alpha_1 - 4\alpha_2 \\ \alpha_2 \lambda &= 5\alpha_1 - 2\alpha_2 \end{aligned} \Rightarrow \begin{pmatrix} 7 & -4 \\ 5 & -2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \lambda \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

In other words, solutions of (7.1.1) having the form (7.1.2) can be constructed provided solutions for λ and $\mathbf{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ in the matrix equation $\mathbf{Ax} = \lambda\mathbf{x}$ can be found. Clearly, $\mathbf{x} = \mathbf{0}$ trivially satisfies $\mathbf{Ax} = \lambda\mathbf{x}$, but $\mathbf{x} = \mathbf{0}$ provides no useful information concerning the solution of (7.1.1). What we really need are scalars λ and *nonzero* vectors $\mathbf{x}$ that satisfy $\mathbf{Ax} = \lambda\mathbf{x}$. Writing $\mathbf{Ax} = \lambda\mathbf{x}$ as $(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0}$ shows that the vectors of interest are the *nonzero* vectors in $N(\mathbf{A} - \lambda\mathbf{I})$. But $N(\mathbf{A} - \lambda\mathbf{I})$ contains nonzero vectors if and only if $\mathbf{A} - \lambda\mathbf{I}$ is singular. Therefore, the scalars of interest are precisely the values of λ that make $\mathbf{A} - \lambda\mathbf{I}$ singular or, equivalently, the λ's for which $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. These observations motivate the definition of eigenvalues and eigenvectors.[66]

Eigenvalues and Eigenvectors

For an $n \times n$ matrix $\mathbf{A}$, scalars λ and vectors $\mathbf{x}_{n\times 1} \neq \mathbf{0}$ satisfying $\mathbf{Ax} = \lambda\mathbf{x}$ are called ***eigenvalues*** and ***eigenvectors*** of $\mathbf{A}$, respectively, and any such pair, $(\lambda, \mathbf{x})$, is called an ***eigenpair*** for $\mathbf{A}$. The set of *distinct* eigenvalues, denoted by $\sigma(\mathbf{A})$, is called the ***spectrum*** of $\mathbf{A}$.

- $\lambda \in \sigma(\mathbf{A}) \iff \mathbf{A} - \lambda\mathbf{I}$ is singular $\iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0$. (7.1.3)
- $\{\mathbf{x} \neq \mathbf{0} \mid \mathbf{x} \in N(\mathbf{A} - \lambda\mathbf{I})\}$ is the set of all eigenvectors associated with λ. From now on, $N(\mathbf{A} - \lambda\mathbf{I})$ is called an ***eigenspace*** for $\mathbf{A}$.
- Nonzero row vectors $\mathbf{y}^*$ such that $\mathbf{y}^*(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$ are called ***left-hand eigenvectors*** for $\mathbf{A}$ (see Exercise 7.1.18 on p. 503).

Geometrically, $\mathbf{Ax} = \lambda\mathbf{x}$ says that under transformation by $\mathbf{A}$, eigenvectors experience only changes in magnitude or sign—the orientation of $\mathbf{Ax}$ in $\Re^n$ is the same as that of $\mathbf{x}$. The eigenvalue λ is simply the amount of "stretch" or "shrink" to which the eigenvector $\mathbf{x}$ is subjected when transformed by $\mathbf{A}$. Figure 7.1.1 depicts the situation in $\Re^2$.

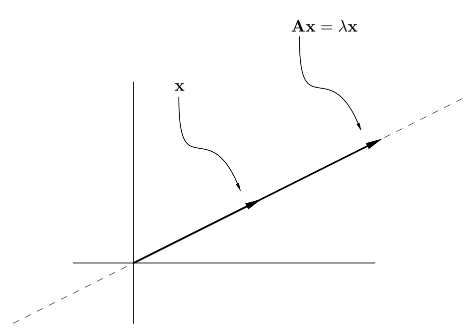


Figure 7.1.1

[66] The words *eigenvalue* and *eigenvector* are derived from the German word *eigen*, which means *owned by* or *peculiar to*. Eigenvalues and eigenvectors are sometimes called *characteristic values* and *characteristic vectors, proper values* and *proper vectors,* or *latent values* and *latent vectors.*

Let's now face the problem of finding the eigenvalues and eigenvectors of the matrix $\mathbf{A} = \begin{pmatrix} 7 & -4 \\ 5 & -2 \end{pmatrix}$ appearing in (7.1.1). As noted in (7.1.3), the eigenvalues are the scalars λ for which $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Expansion of $\det(\mathbf{A} - \lambda\mathbf{I})$ produces the second-degree polynomial

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 7-\lambda & -4 \\ 5 & -2-\lambda \end{vmatrix} = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3),$$

which is called the ***characteristic polynomial*** for $\mathbf{A}$. Consequently, the eigenvalues for $\mathbf{A}$ are the solutions of the ***characteristic equation*** $p(\lambda) = 0$ (i.e., the roots of the characteristic polynomial), and they are $\lambda = 2$ and $\lambda = 3$.

The eigenvectors associated with $\lambda = 2$ and $\lambda = 3$ are simply the nonzero vectors in the eigenspaces $N(\mathbf{A} - 2\mathbf{I})$ and $N(\mathbf{A} - 3\mathbf{I})$, respectively. But determining these eigenspaces amounts to nothing more than solving the two homogeneous systems, $(\mathbf{A} - 2\mathbf{I})\mathbf{x} = \mathbf{0}$ and $(\mathbf{A} - 3\mathbf{I})\mathbf{x} = \mathbf{0}$.

For $\lambda = 2$,

$$\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} 5 & -4 \\ 5 & -4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -4/5 \\ 0 & 0 \end{pmatrix} \implies \begin{array}{l} x_1 = (4/5)x_2 \\ x_2 \text{ is free} \end{array}$$

$$\implies N(\mathbf{A} - 2\mathbf{I}) = \left\{ \mathbf{x} \;\middle|\; \mathbf{x} = \alpha \begin{pmatrix} 4/5 \\ 1 \end{pmatrix} \right\}.$$

For $\lambda = 3$,

$$\mathbf{A} - 3\mathbf{I} = \begin{pmatrix} 4 & -4 \\ 5 & -5 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \implies \begin{array}{l} x_1 = x_2 \\ x_2 \text{ is free} \end{array}$$

$$\implies N(\mathbf{A} - 3\mathbf{I}) = \left\{ \mathbf{x} \;\middle|\; \mathbf{x} = \beta \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

In other words, the eigenvectors of $\mathbf{A}$ associated with $\lambda = 2$ are all nonzero multiples of $\mathbf{x} = (4/5 \quad 1)^T$, and the eigenvectors associated with $\lambda = 3$ are all nonzero multiples of $\mathbf{y} = (1 \quad 1)^T$. Although there are an infinite number of eigenvectors associated with each eigenvalue, each eigenspace is one dimensional, so, for this example, there is only one *independent* eigenvector associated with each eigenvalue.

Let's complete the discussion concerning the system of differential equations $\mathbf{u}' = \mathbf{A}\mathbf{u}$ in (7.1.1). Coupling (7.1.2) with the eigenpairs $(\lambda_1, \mathbf{x})$ and $(\lambda_2, \mathbf{y})$ of $\mathbf{A}$ computed above produces two solutions of $\mathbf{u}' = \mathbf{A}\mathbf{u}$, namely,

$$\mathbf{u}_1 = e^{\lambda_1 t}\mathbf{x} = e^{2t}\begin{pmatrix} 4/5 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{u}_2 = e^{\lambda_2 t}\mathbf{y} = e^{3t}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

It turns out that all other solutions are linear combinations of these two particular solutions—more is said in §7.4 on p. 541.

Below is a summary of some general statements concerning features of the characteristic polynomial and the characteristic equation.

Characteristic Polynomial and Equation

- The ***characteristic polynomial*** of $\mathbf{A}_{n\times n}$ is $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$. The degree of $p(\lambda)$ is n, and the leading term in $p(\lambda)$ is $(-1)^n\lambda^n$.
- The ***characteristic equation*** for $\mathbf{A}$ is $p(\lambda) = 0$.
- The eigenvalues of $\mathbf{A}$ are the solutions of the characteristic equation or, equivalently, the roots of the characteristic polynomial.
- Altogether, $\mathbf{A}$ has n eigenvalues, but some may be complex numbers (even if the entries of $\mathbf{A}$ are real numbers), and some eigenvalues may be repeated.
- If $\mathbf{A}$ contains only real numbers, then its complex eigenvalues must occur in conjugate pairs—i.e., if $\lambda \in \sigma(\mathbf{A})$, then $\overline{\lambda} \in \sigma(\mathbf{A})$.

Proof. The fact that $\det(\mathbf{A} - \lambda\mathbf{I})$ is a polynomial of degree n whose leading term is $(-1)^n\lambda^n$ follows from the definition of determinant given in (6.1.1). If

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

then

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \sum_p \sigma(p)(a_{1p_1} - \delta_{1p_1}\lambda)(a_{2p_2} - \delta_{2p_2}\lambda)\cdots(a_{np_n} - \delta_{np_n}\lambda)$$

is a polynomial in λ. The highest power of λ is produced by the term

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

so the degree is n, and the leading term is $(-1)^n\lambda^n$. The discussion given earlier contained the proof that the eigenvalues are precisely the solutions of the characteristic equation, but, for the sake of completeness, it's repeated below:

$$\begin{aligned}\lambda \in \sigma(\mathbf{A}) &\iff \mathbf{A}\mathbf{x} = \lambda\mathbf{x} \text{ for some } \mathbf{x} \neq \mathbf{0} \iff (\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \neq \mathbf{0} \\ &\iff \mathbf{A} - \lambda\mathbf{I} \text{ is singular} \iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0.\end{aligned}$$

The fundamental theorem of algebra is a deep result that insures every polynomial of degree n with real or complex coefficients has n roots, but some roots may be complex numbers (even if all the coefficients are real), and some roots may be repeated. Consequently, $\mathbf{A}$ has n eigenvalues, but some may be complex, and some may be repeated. The fact that complex eigenvalues of real matrices must occur in conjugate pairs is a consequence of the fact that the roots of a polynomial with real coefficients occur in conjugate pairs. ∎

Example 7.1.1

Problem: Determine the eigenvalues and eigenvectors of $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

Solution: The characteristic polynomial is

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 + 1 = \lambda^2 - 2\lambda + 2,$$

so the characteristic equation is $\lambda^2 - 2\lambda + 2 = 0$. Application of the quadratic formula yields

$$\lambda = \frac{2 \pm \sqrt{-4}}{2} = \frac{2 \pm 2\sqrt{-1}}{2} = 1 \pm \mathrm{i},$$

so the spectrum of $\mathbf{A}$ is $\sigma(\mathbf{A}) = \{1+\mathrm{i},\, 1-\mathrm{i}\}$. Notice that the eigenvalues are complex conjugates of each other—as they must be because complex eigenvalues of real matrices must occur in conjugate pairs. Now find the eigenspaces.

For $\lambda = 1 + \mathrm{i}$,

$$\mathbf{A} - \lambda\mathbf{I} = \begin{pmatrix} -\mathrm{i} & -1 \\ 1 & -\mathrm{i} \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -\mathrm{i} \\ 0 & 0 \end{pmatrix} \implies N(\mathbf{A} - \lambda\mathbf{I}) = span\left\{\begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix}\right\}.$$

For $\lambda = 1 - \mathrm{i}$,

$$\mathbf{A} - \lambda\mathbf{I} = \begin{pmatrix} \mathrm{i} & -1 \\ 1 & \mathrm{i} \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \mathrm{i} \\ 0 & 0 \end{pmatrix} \implies N(\mathbf{A} - \lambda\mathbf{I}) = span\left\{\begin{pmatrix} -\mathrm{i} \\ 1 \end{pmatrix}\right\}.$$

In other words, the eigenvectors associated with $\lambda_1 = 1 + \mathrm{i}$ are all nonzero multiples of $\mathbf{x}_1 = (\mathrm{i} \quad 1)^T$, and the eigenvectors associated with $\lambda_2 = 1 - \mathrm{i}$ are all nonzero multiples of $\mathbf{x}_2 = (-\mathrm{i} \quad 1)^T$. In previous sections, you could be successful by thinking only in terms of real numbers and by dancing around those statements and issues involving complex numbers. But this example makes it clear that avoiding complex numbers, even when dealing with real matrices, is no longer possible—very innocent looking matrices, such as the one in this example, can possess complex eigenvalues and eigenvectors.

As we have seen, computing eigenvalues boils down to solving a polynomial equation. But determining solutions to polynomial equations can be a formidable task. It was proven in the nineteenth century that it's impossible to express the roots of a general polynomial of degree five or higher using radicals of the coefficients. This means that there does not exist a generalized version of the quadratic formula for polynomials of degree greater than four, and general polynomial equations cannot be solved by a finite number of arithmetic operations involving $+, -, \times, \div, \sqrt[n]{\ }$. Unlike solving $\mathbf{Ax} = \mathbf{b}$, the eigenvalue problem generally requires an infinite algorithm, so all practical eigenvalue computations are accomplished by iterative methods—some are discussed later.

For theoretical work, and for textbook-type problems, it's helpful to express the characteristic equation in terms of the principal minors. Recall that an $r \times r$ ***principal submatrix*** of $\mathbf{A}_{n\times n}$ is a submatrix that lies on the same set of r rows and columns, and an $r \times r$ ***principal minor*** is the determinant of an $r \times r$ principal submatrix. In other words, $r \times r$ principal minors are obtained by deleting the same set of $n-r$ rows and columns, and there are $\binom{n}{r} = n!/r!(n-r)!$ such minors. For example, the 1×1 principal minors of

$$\mathbf{A} = \begin{pmatrix} -3 & 1 & -3 \\ 20 & 3 & 10 \\ 2 & -2 & 4 \end{pmatrix} \tag{7.1.4}$$

are the diagonal entries -3, 3, and 4. The 2×2 principal minors are

$$\begin{vmatrix} -3 & 1 \\ 20 & 3 \end{vmatrix} = -29, \quad \begin{vmatrix} -3 & -3 \\ 2 & 4 \end{vmatrix} = -6, \quad \text{and} \quad \begin{vmatrix} 3 & 10 \\ -2 & 4 \end{vmatrix} = 32,$$

and the only 3×3 principal minor is $\det(\mathbf{A}) = -18$.

Related to the principal minors are the symmetric functions of the eigenvalues. The k^{th} ***symmetric function*** of $\lambda_1, \lambda_2, \ldots, \lambda_n$ is defined to be the sum of the product of the eigenvalues taken k at a time. That is,

$$s_k = \sum_{1 \le i_1 < \cdots < i_k \le n} \lambda_{i_1} \cdots \lambda_{i_k}.$$

For example, when $n = 4$,

$$\begin{aligned} s_1 &= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4, \\ s_2 &= \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_1\lambda_4 + \lambda_2\lambda_3 + \lambda_2\lambda_4 + \lambda_3\lambda_4, \\ s_3 &= \lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_2\lambda_4 + \lambda_1\lambda_3\lambda_4 + \lambda_2\lambda_3\lambda_4, \\ s_4 &= \lambda_1\lambda_2\lambda_3\lambda_4. \end{aligned}$$

The connection between symmetric functions, principal minors, and the coefficients in the characteristic polynomial is given in the following theorem.

Coefficients in the Characteristic Equation

If $\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \cdots + c_{n-1}\lambda + c_n = 0$ is the characteristic equation for $\mathbf{A}_{n\times n}$, and if s_k is the k^{th} symmetric function of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$, then

- $c_k = (-1)^k \sum(\text{all } k \times k \text{ principal minors}),$ (7.1.5)
- $s_k = \sum(\text{all } k \times k \text{ principal minors}),$ (7.1.6)
- $trace\,(\mathbf{A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n = -c_1,$ (7.1.7)
- $\det(\mathbf{A}) = \lambda_1\lambda_2\cdots\lambda_n = (-1)^n c_n.$ (7.1.8)

Proof. At least two proofs of (7.1.5) are possible, and although they are conceptually straightforward, each is somewhat tedious. One approach is to successively use the result of Exercise 6.1.14 to expand $\det(\mathbf{A} - \lambda\mathbf{I})$. Another proof rests on the observation that if

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = (-1)^n\lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_{n-1}\lambda + a_n$$

is the characteristic *polynomial* for $\mathbf{A}$, then the characteristic *equation* is

$$\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \cdots + c_{n-1}\lambda + c_n = 0, \quad \text{where} \quad c_i = (-1)^n a_i.$$

Taking the r^{th} derivative of $p(\lambda)$ yields $p^{(r)}(0) = r!a_{n-r}$, and hence

$$c_{n-r} = \frac{(-1)^n}{r!}p^{(r)}(0). \tag{7.1.9}$$

It's now a matter of repeatedly applying the formula (6.1.19) for differentiating a determinant to $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$. After r applications of (6.1.19),

$$p^{(r)}(\lambda) = \sum_{i_j \neq i_k} D_{i_1\cdots i_r}(\lambda),$$

where $D_{i_1\cdots i_r}(\lambda)$ is the determinant of the matrix identical to $\mathbf{A} - \lambda\mathbf{I}$ except that rows $i_1, i_2, \ldots, i_r$ have been replaced by $-\mathbf{e}_{i_1}^T, -\mathbf{e}_{i_2}^T, \ldots, -\mathbf{e}_{i_r}^T$, respectively. It follows that $D_{i_1\cdots i_r}(0) = (-1)^r \det(\mathbf{A}_{i_1\cdots i_r})$, where $\mathbf{A}_{i_1 i_2\cdots i_r}$ is identical to $\mathbf{A}$ except that rows $i_1, i_2, \ldots, i_r$ have been replaced by $\mathbf{e}_{i_1}^T, \mathbf{e}_{i_2}^T, \ldots, \mathbf{e}_{i_r}^T$, respectively, and $\det(\mathbf{A}_{i_1\cdots i_r})$ is the $n-r \times n-r$ principal minor obtained by deleting rows and columns $i_1, i_2, \ldots, i_r$ from $\mathbf{A}$. Consequently,

$$\begin{aligned} p^{(r)}(0) &= \sum_{i_j \neq i_k} D_{i_1\cdots i_r}(0) = (-1)^r \sum_{i_j \neq i_k} \det(\mathbf{A}_{i_1\cdots i_r}) \\ &= r! \times (-1)^r \sum (\text{all } n-r \times n-r \text{ principal minors}). \end{aligned}$$

The factor $r!$ appears because each of the $r!$ permutations of the subscripts on $\mathbf{A}_{i_1\cdots i_r}$ describes the same matrix. Therefore, (7.1.9) says

$$c_{n-r} = \frac{(-1)^n}{r!}p^{(r)}(0) = (-1)^{n-r} \sum (\text{all } n-r \times n-r \text{ principal minors}).$$

To prove (7.1.6), write the characteristic equation for $\mathbf{A}$ as

$$(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) = 0, \tag{7.1.10}$$

and expand the left-hand side to produce

$$\lambda^n - s_1\lambda^{n-1} + \cdots + (-1)^k s_k\lambda^{n-k} + \cdots + (-1)^n s_n = 0. \tag{7.1.11}$$

(Using $n = 3$ or $n = 4$ in (7.1.10) makes this clear.) Comparing (7.1.11) with (7.1.5) produces the desired conclusion. Statements (7.1.7) and (7.1.8) are obtained from (7.1.5) and (7.1.6) by setting $k = 1$ and $k = n$. ∎

Example 7.1.2

Problem: Determine the eigenvalues and eigenvectors of

$$\mathbf{A} = \begin{pmatrix} -3 & 1 & -3 \\ 20 & 3 & 10 \\ 2 & -2 & 4 \end{pmatrix}.$$

Solution: Use the principal minors computed in (7.1.4) along with (7.1.5) to obtain the characteristic equation

$$\lambda^3 - 4\lambda^2 - 3\lambda + 18 = 0.$$

A result from elementary algebra states that if the coefficients α_i in

$$\lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0 = 0$$

are integers, then every integer solution is a factor of α_0. For our problem, this means that if there exist integer eigenvalues, then they must be contained in the set $\mathcal{S} = \{\pm 1, \pm 2, \pm 3, \pm 6, \pm 9, \pm 18\}$. Evaluating $p(\lambda)$ for each $\lambda \in \mathcal{S}$ reveals that $p(3) = 0$ and $p(-2) = 0$, so $\lambda = 3$ and $\lambda = -2$ are eigenvalues for $\mathbf{A}$. To determine the other eigenvalue, deflate the problem by dividing

$$\frac{\lambda^3 - 4\lambda^2 - 3\lambda + 18}{\lambda - 3} = \lambda^2 - \lambda - 6 = (\lambda - 3)(\lambda + 2).$$

Thus the characteristic equation can be written in factored form as

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

so the spectrum of $\mathbf{A}$ is $\sigma(\mathbf{A}) = \{3, -2\}$ in which $\lambda = 3$ is repeated—we say that the ***algebraic multiplicity*** of $\lambda = 3$ is two. The eigenspaces are obtained as follows.

For $\lambda = 3$,

$$\mathbf{A} - 3\mathbf{I} \longrightarrow \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies N(\mathbf{A} - 3\mathbf{I}) = span\left\{ \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \right\}.$$

For $\lambda = -2$,

$$\mathbf{A} + 2\mathbf{I} \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \implies N(\mathbf{A} + 2\mathbf{I}) = span\left\{ \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \right\}.$$

Notice that although the algebraic multiplicity of $\lambda = 3$ is two, the dimension of the associated eigenspace is only one—we say that $\mathbf{A}$ is ***deficient*** in eigenvectors. As we will see later, deficient matrices pose significant difficulties.

Example 7.1.3

Continuity of Eigenvalues. A classical result (requiring complex analysis) states that the roots of a polynomial vary continuously with the coefficients. Since the coefficients of the characteristic polynomial $p(\lambda)$ of $\mathbf{A}$ can be expressed in terms of sums of principal minors, it follows that the coefficients of $p(\lambda)$ vary continuously with the entries of $\mathbf{A}$. Consequently, the eigenvalues of $\mathbf{A}$ must vary continuously with the entries of $\mathbf{A}$. **Caution!** Components of an eigenvector need not vary continuously with the entries of $\mathbf{A}$—e.g., consider $\mathbf{x} = (\epsilon^{-1}, 1)$ as an eigenvector for $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & \epsilon \end{pmatrix}$, and let $\epsilon \to 0$.

Example 7.1.4

Spectral Radius. For square matrices $\mathbf{A}$, the number

$$\rho(\mathbf{A}) = \max_{\lambda \in \sigma(\mathbf{A})} |\lambda|$$

is called the ***spectral radius*** of $\mathbf{A}$. It's not uncommon for applications to require only a bound on the eigenvalues of $\mathbf{A}$. That is, precise knowledge of each eigenvalue may not called for, but rather just an upper bound on $\rho(\mathbf{A})$ is all that's often needed. A rather crude (but cheap) upper bound on $\rho(\mathbf{A})$ is obtained by observing that $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ for every matrix norm. This is true because if $(\lambda, \mathbf{x})$ is any eigenpair, then $\mathbf{X} = \left[\mathbf{x} \,|\, \mathbf{0} \,|\, \cdots \,|\, \mathbf{0}\right]_{n \times n} \neq \mathbf{0}$, and $\lambda\mathbf{X} = \mathbf{A}\mathbf{X}$ implies $|\lambda| \; \|\mathbf{X}\| = \|\lambda\mathbf{X}\| = \|\mathbf{A}\mathbf{X}\| \leq \|\mathbf{A}\| \; \|\mathbf{X}\|$, so

$$|\lambda| \leq \|\mathbf{A}\| \quad \text{for all } \lambda \in \sigma(\mathbf{A}). \tag{7.1.12}$$

This result is a precursor to a stronger relationship between spectral radius and norm that is hinted at in Exercise 7.3.12 and developed in Example 7.10.1 (p. 619).

The eigenvalue bound (7.1.12) given in Example 7.1.4 is cheap to compute, especially if the 1-norm or ∞-norm is used, but you often get what you pay for. You get one big circle whose radius is usually much larger than the spectral radius $\rho(\mathbf{A})$. It's possible to do better by using a set of Gerschgorin[67] circles as described below.

[67] S. A. Gerschgorin illustrated the use of Gerschgorin circles for estimating eigenvalues in 1931, but the concept appears earlier in work by L. Lévy in 1881, by H. Minkowski (p. 278) in 1900, and by J. Hadamard (p. 469) in 1903. However, each time the idea surfaced, it gained little attention and was quickly forgotten until Olga Taussky (1906–1995), the premier woman of linear algebra, and her fellow German emigrè Alfred Brauer (1894–1985) became captivated by the result. Taussky (who became Olga Taussky-Todd after marrying the numerical analyst John Todd) and Brauer devoted significant effort to strengthening, promoting, and popularizing Gerschgorin-type eigenvalue bounds. Their work during the 1940s and 1950s ended the periodic rediscoveries, and they made Gerschgorin (who might otherwise have been forgotten) famous.

Gerschgorin Circles

- The eigenvalues of $\mathbf{A} \in \mathcal{C}^{n\times n}$ are contained the union $\mathcal{G}_r$ of the n ***Gerschgorin circles*** defined by
$$|z - a_{ii}| \leq r_i, \quad \text{where} \quad r_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \text{ for } i = 1, 2, \ldots, n. \tag{7.1.13}$$
In other words, the eigenvalues are trapped in the collection of circles centered at a_{ii} with radii given by the sum of absolute values in $\mathbf{A}_{i*}$ with a_{ii} deleted.
- Furthermore, if a union $\mathcal{U}$ of k Gerschgorin circles does not touch any of the other $n-k$ circles, then there are exactly k eigenvalues (counting multiplicities) in the circles in $\mathcal{U}$. (7.1.14)
- Since $\sigma(\mathbf{A}^T) = \sigma(\mathbf{A})$, the deleted absolute row sums in (7.1.13) can be replaced by deleted absolute column sums, so the eigenvalues of $\mathbf{A}$ are also contained in the union $\mathcal{G}_c$ of the circles defined by
$$|z - a_{jj}| \leq c_j, \quad \text{where} \quad c_j = \sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}| \text{ for } j = 1, 2, \ldots, n. \tag{7.1.15}$$
- Combining (7.1.13) and (7.1.15) means that the eigenvalues of $\mathbf{A}$ are contained in the intersection $\mathcal{G}_r \cap \mathcal{G}_c$. (7.1.16)

Proof. Let $(\lambda, \mathbf{x})$ be an eigenpair for $\mathbf{A}$, and assume $\mathbf{x}$ has been normalized so that $\|\mathbf{x}\|_\infty = 1$. If x_i is a component of $\mathbf{x}$ such that $|x_i| = 1$, then

$$\lambda x_i = [\lambda\mathbf{x}]_i = [\mathbf{A}\mathbf{x}]_i = \sum_{j=1}^{n} a_{ij}x_j \implies (\lambda - a_{ii})x_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij}x_j,$$

and hence

$$|\lambda - a_{ii}| = |\lambda - a_{ii}|\,|x_i| = \left|\sum_{j\neq i} a_{ij}x_j\right| \leq \sum_{j\neq i} |a_{ij}|\,|x_j| \leq \sum_{j\neq i} |a_{ij}| = r_i.$$

Thus λ is in one of the Gerschgorin circles, so the union of all such circles contains $\sigma(\mathbf{A})$. To establish (7.1.14), let $\mathbf{D} = \text{diag}\,(a_{11}, a_{22}, \ldots, a_{nn})$ and $\mathbf{B} = \mathbf{A} - \mathbf{D}$, and set $\mathbf{C}(t) = \mathbf{D} + t\mathbf{B}$ for $t \in [0, 1]$. The first part shows that the eigenvalues of $\lambda_i(t)$ of $\mathbf{C}(t)$ are contained in the union of the Gerschgorin circles $\mathcal{C}_i(t)$ defined by $|z - a_{ii}| \leq t\,r_i$. The circles $\mathcal{C}_i(t)$ grow continuously with t from individual points a_{ii} when $t = 0$ to the Gerschgorin circles of $\mathbf{A}$ when $t = 1$, so, if the circles in the isolated union $\mathcal{U}$ are centered at $a_{i_1 i_1}, a_{i_2 i_2}, \ldots, a_{i_k i_k}$, then for every $t \in [0,1]$ the union $\mathcal{U}(t) = \mathcal{C}_{i_1}(t) \cup \mathcal{C}_{i_2}(t) \cup \cdots \cup \mathcal{C}_{i_k}(t)$ is disjoint from the union $\overline{\mathcal{U}(t)}$ of the other $n-k$ Gerschgorin circles of $\mathbf{C}(t)$. Since (as mentioned in Example 7.1.3) each eigenvalue $\lambda_i(t)$ of $\mathbf{C}(t)$ also varies continuously with t, each $\lambda_i(t)$ is on a continuous curve Γ_i having one end at $\lambda_i(0) = a_{ii}$ and the other end at $\lambda_i(1) \in \sigma(\mathbf{A})$. But since $\mathcal{U}(t) \cap \overline{\mathcal{U}(t)} = \phi$ for all $t \in [0,1]$, the curves $\Gamma_{i_1}, \Gamma_{i_2}, \ldots, \Gamma_{i_k}$ are entirely contained in $\mathcal{U}$, and hence the end points $\lambda_{i_1}(1), \lambda_{i_2}(1), \ldots, \lambda_{i_k}(1)$ are in $\mathcal{U}$. Similarly, the other $n-k$ eigenvalues of $\mathbf{A}$ are in the union of the complementary set of circles. ■

Example 7.1.5

Problem: Estimate the eigenvalues of $\mathbf{A} = \begin{pmatrix} 5 & 1 & 1 \\ 0 & 6 & 1 \\ 1 & 0 & -5 \end{pmatrix}$.

- A crude estimate is derived from the bound given in Example 7.1.4 on p. 497. Using the ∞-norm, (7.1.12) says that $|\lambda| \leq \|\mathbf{A}\|_\infty = 7$ for all $\lambda \in \sigma(\mathbf{A})$.
- Better estimates are produced by the Gerschgorin circles in Figure 7.1.2 that are derived from row sums. Statements (7.1.13) and (7.1.14) guarantee that one eigenvalue is in (or on) the circle centered at -5, while the remaining two eigenvalues are in (or on) the larger circle centered at $+5$.

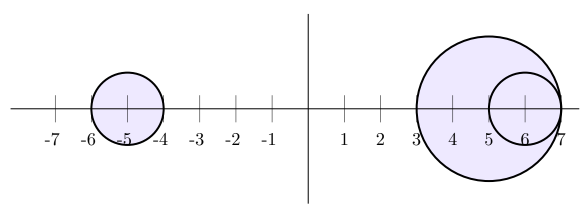


FIGURE 7.1.2. GERSCHGORIN CIRCLES DERIVED FROM ROW SUMS.

- The best estimate is obtained from (7.1.16) by considering $\mathcal{G}_r \cap \mathcal{G}_c$.

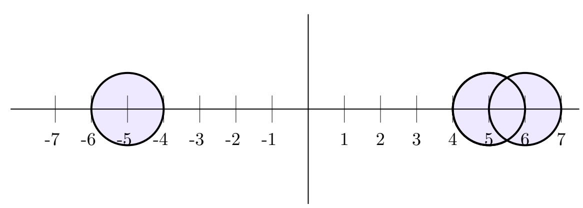


FIGURE 7.1.3. GERSCHGORIN CIRCLES DERIVED FROM $\mathcal{G}_r \cap \mathcal{G}_c$.

In other words, one eigenvalue is in the circle centered at -5, while the other two eigenvalues are in the union of the other two circles in Figure 7.1.3. This is corroborated by computing $\sigma(\mathbf{A}) = \{5,\ (1 \pm 5\sqrt{5})/2\} \approx \{5,\ 6.0902,\ -5.0902\}$.

Example 7.1.6

Diagonally Dominant Matrices Revisited. Recall from Example 4.3.3 on p. 184 that $\mathbf{A}_{n\times n}$ is said to be ***diagonally dominant*** (some authors say *strictly* diagonally dominant) whenever

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \quad \text{for each } i = 1, 2, \ldots, n.$$

Gerschgorin's theorem (7.1.13) guarantees that diagonally dominant matrices cannot possess a zero eigenvalue. But $0 \notin \sigma(\mathbf{A})$ if and only if $\mathbf{A}$ is nonsingular (Exercise 7.1.6), so Gerschgorin's theorem provides an alternative to the argument used in Example 4.3.3 to prove that *all diagonally dominant matrices are nonsingular*.[68] For example, the 3×3 matrix $\mathbf{A}$ in Example 7.1.5 is diagonally dominant, and thus $\mathbf{A}$ is nonsingular. Even when a matrix is not diagonally dominant, Gerschgorin estimates still may be useful in determining whether or not the matrix is nonsingular simply by observing if zero is excluded from $\sigma(\mathbf{A})$ based on the configuration of the Gerschgorin circles given in (7.1.16).

Exercises for section 7.1

7.1.1. Determine the eigenvalues and eigenvectors for the following matrices.

$$\mathbf{A} = \begin{pmatrix} -10 & -7 \\ 14 & 11 \end{pmatrix}. \quad \mathbf{B} = \begin{pmatrix} 2 & 16 & 8 \\ 4 & 14 & 8 \\ -8 & -32 & -18 \end{pmatrix}. \quad \mathbf{C} = \begin{pmatrix} 3 & -2 & 5 \\ 0 & 1 & 4 \\ 0 & -1 & 5 \end{pmatrix}.$$

$$\mathbf{D} = \begin{pmatrix} 0 & 6 & 3 \\ -1 & 5 & 1 \\ -1 & 2 & 4 \end{pmatrix}. \quad \mathbf{E} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Which, if any, are deficient in eigenvectors in the sense that there fails to exist a complete linearly independent set?

7.1.2. Without doing an eigenvalue–eigenvector computation, determine which of the following are eigenvectors for

$$\mathbf{A} = \begin{pmatrix} -9 & -6 & -2 & -4 \\ -8 & -6 & -3 & -1 \\ 20 & 15 & 8 & 5 \\ 32 & 21 & 7 & 12 \end{pmatrix},$$

and for those which are eigenvectors, identify the associated eigenvalue.

(a) $\begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$. (b) $\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$. (c) $\begin{pmatrix} -1 \\ 0 \\ 2 \\ 2 \end{pmatrix}$. (d) $\begin{pmatrix} 0 \\ 1 \\ -3 \\ 0 \end{pmatrix}$.

[68] In fact, this result was the motivation behind the original development of Gerschgorin's circles.

7.1.3. Explain why the eigenvalues of triangular and diagonal matrices

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

are simply the diagonal entries—the t_{ii} 's and λ_i 's.

7.1.4. For $\mathbf{T} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}$, prove $\det(\mathbf{T} - \lambda\mathbf{I}) = \det(\mathbf{A} - \lambda\mathbf{I})\det(\mathbf{C} - \lambda\mathbf{I})$ to conclude that $\sigma\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} = \sigma(\mathbf{A}) \cup \sigma(\mathbf{C})$ for square $\mathbf{A}$ and $\mathbf{C}$.

7.1.5. Determine the eigenvectors of $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. In particular, what is the eigenspace associated with λ_i?

7.1.6. Prove that $0 \in \sigma(\mathbf{A})$ if and only if $\mathbf{A}$ is a singular matrix.

7.1.7. Explain why it's apparent that $\mathbf{A}_{n\times n} = \begin{pmatrix} n & 1 & 1 & \cdots & 1 \\ 1 & n & 1 & \cdots & 1 \\ 1 & 1 & n & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & n \end{pmatrix}$ doesn't have a zero eigenvalue, and hence why $\mathbf{A}$ is nonsingular.

7.1.8. Explain why the eigenvalues of $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ are real and nonnegative for every $\mathbf{A} \in \mathcal{C}^{m\times n}$. **Hint:** Consider $\|\mathbf{A}\mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2$. When are the eigenvalues of $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ strictly positive?

7.1.9. (a) If $\mathbf{A}$ is nonsingular, and if $(\lambda, \mathbf{x})$ is an eigenpair for $\mathbf{A}$, show that $(\lambda^{-1}, \mathbf{x})$ is an eigenpair for $\mathbf{A}^{-1}$.

(b) For all $\alpha \notin \sigma(\mathbf{A})$, prove that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ if and only if $\mathbf{x}$ is an eigenvector of $(\mathbf{A} - \alpha\mathbf{I})^{-1}$.

7.1.10. (a) Show that if $(\lambda, \mathbf{x})$ is an eigenpair for $\mathbf{A}$, then $(\lambda^k, \mathbf{x})$ is an eigenpair for $\mathbf{A}^k$ for each positive integer k.

(b) If $p(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_k x^k$ is any polynomial, then we define $p(\mathbf{A})$ to be the matrix

$$p(\mathbf{A}) = \alpha_0\mathbf{I} + \alpha_1\mathbf{A} + \alpha_2\mathbf{A}^2 + \cdots + \alpha_k\mathbf{A}^k.$$

Show that if $(\lambda, \mathbf{x})$ is an eigenpair for $\mathbf{A}$, then $(p(\lambda), \mathbf{x})$ is an eigenpair for $p(\mathbf{A})$.

7.1.11. Explain why (7.1.14) in Gerschgorin's theorem on p. 498 implies that $\mathbf{A} = \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 12 & 0 & -4 \\ 1 & 0 & -1 & 0 \\ 0 & 5 & 0 & 0 \end{pmatrix}$ must have at least two real eigenvalues. Corroborate this fact by computing the eigenvalues of $\mathbf{A}$.

7.1.12. If $\mathbf{A}$ is ***nilpotent*** ($\mathbf{A}^k = \mathbf{0}$ for some k), explain why $trace\,(\mathbf{A}) = 0$. **Hint:** What is $\sigma(\mathbf{A})$?

7.1.13. If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are eigenvectors of $\mathbf{A}$ associated with the same eigenvalue λ, explain why every nonzero linear combination

$$\mathbf{v} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n$$

is also an eigenvector for $\mathbf{A}$ associated with the eigenvalue λ.

7.1.14. Explain why an eigenvector for a square matrix $\mathbf{A}$ cannot be associated with two distinct eigenvalues for $\mathbf{A}$.

7.1.15. Suppose $\sigma(\mathbf{A}_{n\times n}) = \sigma(\mathbf{B}_{n\times n})$. Does this guarantee that $\mathbf{A}$ and $\mathbf{B}$ have the same characteristic polynomial?

7.1.16. Construct 2×2 examples to prove the following statements.

(a) $\lambda \in \sigma(\mathbf{A})$ and $\mu \in \sigma(\mathbf{B}) \;\not\Longrightarrow\; \lambda + \mu \in \sigma(\mathbf{A}+\mathbf{B})$.

(b) $\lambda \in \sigma(\mathbf{A})$ and $\mu \in \sigma(\mathbf{B}) \;\not\Longrightarrow\; \lambda\mu \in \sigma(\mathbf{AB})$.

7.1.17. Suppose that $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ are the eigenvalues for $\mathbf{A}_{n\times n}$, and let $(\lambda_k, \mathbf{c})$ be a particular eigenpair.

(a) For $\lambda \notin \sigma(\mathbf{A})$, explain why $(\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{c} = \mathbf{c}/(\lambda_k - \lambda)$.

(b) For an arbitrary vector $\mathbf{d}_{n\times 1}$, prove that the eigenvalues of $\mathbf{A} + \mathbf{c}\mathbf{d}^T$ agree with those of $\mathbf{A}$ except that λ_k is replaced by $\lambda_k + \mathbf{d}^T\mathbf{c}$.

(c) How can $\mathbf{d}$ be selected to guarantee that the eigenvalues of $\mathbf{A}+\mathbf{c}\mathbf{d}^T$ and $\mathbf{A}$ agree except that λ_k is replaced by a specified number μ?

7.1.18. Suppose that $\mathbf{A}$ is a square matrix.

(a) Explain why $\mathbf{A}$ and $\mathbf{A}^T$ have the same eigenvalues.

(b) Explain why $\lambda \in \sigma(\mathbf{A}) \iff \overline{\lambda} \in \sigma(\mathbf{A}^*)$.
Hint: Recall Exercise 6.1.8.

(c) Do these results imply that $\lambda \in \sigma(\mathbf{A}) \iff \overline{\lambda} \in \sigma(\mathbf{A})$ when $\mathbf{A}$ is a square matrix of *real* numbers?

(d) A nonzero row vector $\mathbf{y}^*$ is called a ***left-hand*** eigenvector for $\mathbf{A}$ whenever there is a scalar $\mu \in \mathcal{C}$ such that $\mathbf{y}^*(\mathbf{A} - \mu\mathbf{I}) = \mathbf{0}$. Explain why μ must be an eigenvalue for $\mathbf{A}$ in the "right-hand" sense of the term when $\mathbf{A}$ is a square matrix of *real* numbers.

7.1.19. Consider matrices $\mathbf{A}_{m\times n}$ and $\mathbf{B}_{n\times m}$.

(a) Explain why $\mathbf{AB}$ and $\mathbf{BA}$ have the same characteristic polynomial if $m = n$. **Hint:** Recall Exercise 6.2.16.

(b) Explain why the characteristic polynomials for $\mathbf{AB}$ and $\mathbf{BA}$ can't be the same when $m \neq n$, and then explain why $\sigma(\mathbf{AB})$ and $\sigma(\mathbf{BA})$ agree, with the possible exception of a zero eigenvalue.

7.1.20. If $\mathbf{AB} = \mathbf{BA}$, prove that $\mathbf{A}$ and $\mathbf{B}$ have a common eigenvector.

Hint: For $\lambda \in \sigma(\mathbf{A})$, let the columns of $\mathbf{X}$ be a basis for $N(\mathbf{A} - \lambda\mathbf{I})$ so that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{BX} = \mathbf{0}$. Explain why there exists a matrix $\mathbf{P}$ such that $\mathbf{BX} = \mathbf{XP}$, and then consider any eigenpair for $\mathbf{P}$.

7.1.21. For fixed matrices $\mathbf{P}_{m\times m}$ and $\mathbf{Q}_{n\times n}$, let $\mathbf{T}$ be the linear operator on $\mathcal{C}^{m\times n}$ defined by $\mathbf{T}(\mathbf{A}) = \mathbf{PAQ}$.

(a) Show that if $\mathbf{x}$ is a right-hand eigenvector for $\mathbf{P}$ and $\mathbf{y}^*$ is a left-hand eigenvector for $\mathbf{Q}$, then $\mathbf{xy}^*$ is an eigenvector for $\mathbf{T}$.

(b) Explain why $\mathit{trace}(\mathbf{T}) = \mathit{trace}(\mathbf{P})\,\mathit{trace}(\mathbf{Q})$.

7.1.22. Let $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be a diagonal real matrix such that $\lambda_1 < \lambda_2 < \cdots < \lambda_n$, and let $\mathbf{v}_{n\times 1}$ be a column of real nonzero numbers.

(a) Prove that if α is real and nonzero, then λ_i is not an eigenvalue for $\mathbf{D} + \alpha\mathbf{vv}^T$. Show that the eigenvalues of $\mathbf{D} + \alpha\mathbf{vv}^T$ are in fact given by the solutions of the *secular equation* $f(\xi) = 0$ defined by

$$f(\xi) = 1 + \alpha \sum_{i=1}^{n} \frac{v_i^2}{\lambda_i - \xi}.$$

For $n = 4$ and $\alpha > 0$, verify that the graph of $f(\xi)$ is as depicted in Figure 7.1.4, and thereby conclude that the eigenvalues of $\mathbf{D} + \alpha\mathbf{vv}^T$ interlace with those of $\mathbf{D}$.

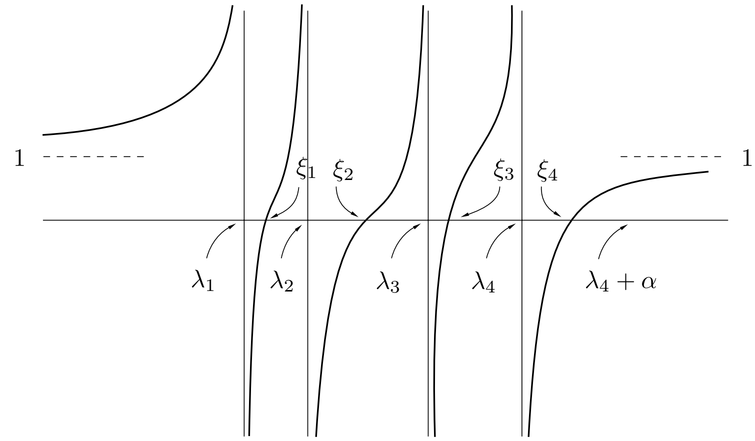


FIGURE 7.1.4

(b) Verify that $(\mathbf{D} - \xi_i \mathbf{I})^{-1}\mathbf{v}$ is an eigenvector for $\mathbf{D} + \alpha\mathbf{v}\mathbf{v}^T$ that is associated with the eigenvalue ξ_i.

7.1.23. Newton's Identities. Let $\lambda_1, \ldots, \lambda_n$ be the roots of the polynomial $p(\lambda) = \lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \cdots + c_n$, and let $\tau_k = \lambda_1^k + \lambda_2^k + \cdots + \lambda_n^k$. Newton's identities say $c_k = -(\tau_1 c_{k-1} + \tau_2 c_{k-2} + \cdots + \tau_{k-1}c_1 + \tau_k)/k$. Derive these identities by executing the following steps:

(a) Show $p'(\lambda) = p(\lambda)\sum_{i=1}^n (\lambda - \lambda_i)^{-1}$ (logarithmic differentiation).

(b) Use the geometric series expansion for $(\lambda - \lambda_i)^{-1}$ to show that for $|\lambda| > \max_i |\lambda_i|$,

$$\sum_{i=1}^{n} \frac{1}{(\lambda - \lambda_i)} = \frac{n}{\lambda} + \frac{\tau_1}{\lambda^2} + \frac{\tau_2}{\lambda^3} + \cdots.$$

(c) Combine these two results, and equate like powers of λ.

7.1.24. Leverrier–Souriau–Frame Algorithm.[69] Let the characteristic equation for $\mathbf{A}$ be given by $\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \cdots + c_n = 0$, and define a sequence by taking $\mathbf{B}_0 = \mathbf{I}$ and

$$\mathbf{B}_k = -\frac{trace\,(\mathbf{A}\mathbf{B}_{k-1})}{k}\mathbf{I} + \mathbf{A}\mathbf{B}_{k-1} \quad \text{for} \quad k = 1, 2, \ldots, n.$$

Prove that for each k,

$$c_k = -\frac{trace\,(\mathbf{A}\mathbf{B}_{k-1})}{k}.$$

Hint: Use Newton's identities, and recall Exercise 7.1.10(a).

[69] This algorithm has been rediscovered and modified several times. In 1840, the Frenchman U. J. J. Leverrier provided the basic connection with Newton's identities. J. M. Souriau, also from France, and J. S. Frame, from Michigan State University, independently modified the algorithm to its present form—Souriau's formulation was published in France in 1948, and Frame's method appeared in the United States in 1949. Paul Horst (USA, 1935) along with Faddeev and Sominskii (USSR, 1949) are also credited with rediscovering the technique. Although the algorithm is intriguingly beautiful, it is not practical for floating-point computations.

Solutions for Chapter 7

Solutions for exercises in section 7. 1

7.1.1. $\sigma(\mathbf{A}) = \{-3, 4\}$

$$N(\mathbf{A}+3\mathbf{I}) = span\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\} \quad \text{and} \quad N(\mathbf{A}-4\mathbf{I}) = span\left\{\begin{pmatrix} -1/2 \\ 1 \end{pmatrix}\right\}$$

$\sigma(\mathbf{B}) = \{-2, 2\}$ in which the algebraic multiplicity of $\lambda = -2$ is two.

$$N(\mathbf{B}+2\mathbf{I}) = span\left\{\begin{pmatrix} -4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}\right\} \text{ and } N(\mathbf{B}-2\mathbf{I}) = span\left\{\begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix}\right\}$$

$\sigma(\mathbf{C}) = \{3\}$ in which the algebraic multiplicity of $\lambda = 3$ is three.

$$N(\mathbf{C}-3\mathbf{I}) = span\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right\}$$

$\sigma(\mathbf{D}) = \{3\}$ in which the algebraic multiplicity of $\lambda = 3$ is three.

$$N(\mathbf{D}-3\mathbf{I}) = span\left\{\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\right\}$$

$\sigma(\mathbf{E}) = \{3\}$ in which the algebraic multiplicity of $\lambda = 3$ is three.

$$N(\mathbf{E}-3\mathbf{I}) = span\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$$

Matrices $\mathbf{C}$ and $\mathbf{D}$ are deficient in eigenvectors.

7.1.2. Form the product $\mathbf{Ax}$, and answer the question, "Is $\mathbf{Ax}$ some multiple of $\mathbf{x}$?" When the answer is *yes,* then $\mathbf{x}$ is an eigenvector for $\mathbf{A}$, and the multiplier is the associated eigenvalue. For this matrix, (a), (c), and (d) are eigenvectors associated with eigenvalues 1, 3, and 3, respectively.

7.1.3. The characteristic polynomial for $\mathbf{T}$ is

$$\det(\mathbf{T}-\lambda\mathbf{I}) = (t_{11}-\lambda)(t_{22}-\lambda)\cdots(t_{nn}-\lambda),$$

so the roots are the t_{ii} 's.

7.1.4. This follows directly from (6.1.16) because

$$\det(\mathbf{T}-\lambda\mathbf{I}) = \begin{vmatrix} \mathbf{A}-\lambda\mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{C}-\lambda\mathbf{I} \end{vmatrix} = \det(\mathbf{A}-\lambda\mathbf{I})\det(\mathbf{C}-\lambda\mathbf{I}).$$

7.1.5. If λ_i is not repeated, then $N(\mathbf{A}-\lambda_i\mathbf{I}) = span\{\mathbf{e}_i\}$. If the algebraic multiplicity of λ_i is k, and if λ_i occupies positions $i_1, i_2, \ldots, i_k$ in $\mathbf{D}$, then

$$N(\mathbf{A}-\lambda_i\mathbf{I}) = span\{\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_k}\}.$$

7.1.6. $\mathbf{A}$ singular $\iff \det(\mathbf{A}) = 0 \iff 0$ solves $\det(\mathbf{A}-\lambda\mathbf{I}) = 0 \iff 0 \in \sigma(\mathbf{A})$.

7.1.7. Zero is not in or on any Gerschgorin circle. You could also say that $\mathbf{A}$ is nonsingular because it is diagonally dominant—see Example 7.1.6 on p. 499.

7.1.8. If $(\lambda, \mathbf{x})$ is an eigenpair for $\mathbf{A}^*\mathbf{A}$, then $\|\mathbf{A}\mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2 = \mathbf{x}^*\mathbf{A}^*\mathbf{A}\mathbf{x}/\mathbf{x}^*\mathbf{x} = \lambda$ is real and nonnegative. Furthermore, $\lambda > 0$ if and only if $\mathbf{A}^*\mathbf{A}$ is nonsingular or, equivalently, $n = rank(\mathbf{A}^*\mathbf{A}) = rank(\mathbf{A})$. Similar arguments apply to $\mathbf{A}\mathbf{A}^*$.

7.1.9. (a) $\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \implies \mathbf{x} = \lambda\mathbf{A}^{-1}\mathbf{x} \implies (1/\lambda)\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}$.

(b) $\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \iff (\mathbf{A}-\alpha\mathbf{I})\mathbf{x} = (\lambda-\alpha)\mathbf{x} \iff (\lambda-\alpha)^{-1}\mathbf{x} = (\mathbf{A}-\alpha\mathbf{I})^{-1}\mathbf{x}$.

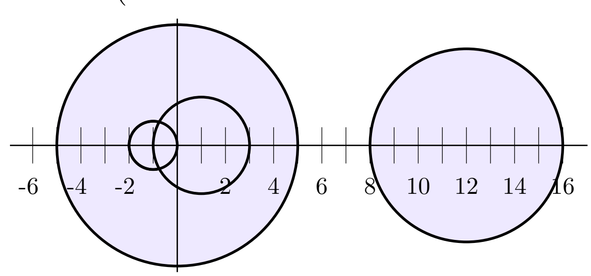
7.1.10. (a) Successively use $\mathbf{A}$ as a left-hand multiplier to produce

$$\begin{aligned} \mathbf{A}\mathbf{x} = \lambda\mathbf{x} &\implies \mathbf{A}^2\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} \\ &\implies \mathbf{A}^3\mathbf{x} = \lambda^2\mathbf{A}\mathbf{x} = \lambda^3\mathbf{x} \\ &\implies \mathbf{A}^4\mathbf{x} = \lambda^3\mathbf{A}\mathbf{x} = \lambda^4\mathbf{x} \\ &\quad \text{etc.} \end{aligned}$$

(b) Use part (a) to write

$$p(\mathbf{A})\mathbf{x} = \left(\sum_i \alpha_i \mathbf{A}^i\right)\mathbf{x} = \sum_i \alpha_i \mathbf{A}^i\mathbf{x} = \sum_i \alpha_i \lambda^i \mathbf{x} = \left(\sum_i \alpha_i \lambda^i\right)\mathbf{x} = p(\lambda)\mathbf{x}.$$

7.1.11. Since one Geschgorin circle (derived from row sums and shown below) is isolated



from the union of the other three circles, statement (7.1.14) on p. 498 insures that there is one eigenvalue in the isolated circle and three eigenvalues in the union of the other three. But, as discussed on p. 492, the eigenvalues of real matrices occur in conjugate pairs. So, the root in the isolated circle must be real and there must be at least one real root in the union of the other three circles. Computation reveals that $\sigma(\mathbf{A}) = \{\pm \mathrm{i},\, 2,\, 10\}$.

7.1.12. Use Exercise 7.1.10 to deduce that

$$\lambda \in \sigma(\mathbf{A}) \implies \lambda^k \in \sigma\left(\mathbf{A}^k\right) \implies \lambda^k = 0 \implies \lambda = 0.$$

Therefore, (7.1.7) insures that $trace(\mathbf{A}) = \sum_i \lambda_i = 0$.

7.1.13. This is true because $N(\mathbf{A} - \lambda\mathbf{I})$ is a subspace—recall that subspaces are closed under vector addition and scalar multiplication.

7.1.14. If there exists a nonzero vector $\mathbf{x}$ that satisfies $\mathbf{Ax} = \lambda_1\mathbf{x}$ and $\mathbf{Ax} = \lambda_2\mathbf{x}$, where $\lambda_1 \neq \lambda_2$, then

$$\mathbf{0} = \mathbf{Ax} - \mathbf{Ax} = \lambda_1\mathbf{x} - \lambda_2\mathbf{x} = (\lambda_1 - \lambda_2)\mathbf{x}.$$

But this implies $\mathbf{x} = \mathbf{0}$, which is impossible. Consequently, no such $\mathbf{x}$ can exist.

7.1.15. No—consider $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

7.1.16. Almost any example with rather random entries will do the job, but avoid diagonal or triangular matrices—they are too special.

7.1.17. (a) $\mathbf{c} = (\mathbf{A} - \lambda\mathbf{I})^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{c} = (\mathbf{A} - \lambda\mathbf{I})^{-1}(\mathbf{Ac} - \lambda\mathbf{c}) = (\mathbf{A} - \lambda\mathbf{I})^{-1}(\lambda_k - \lambda)\mathbf{c}$.

(b) Use (6.2.3) to compute the characteristic polynomial for $\mathbf{A} + \mathbf{cd}^T$ to be

$$\begin{aligned}
\det\left(\mathbf{A} + \mathbf{cd}^T - \lambda\mathbf{I}\right) &= \det\left(\mathbf{A} - \lambda\mathbf{I} + \mathbf{cd}^T\right) \\
&= \det(\mathbf{A} - \lambda\mathbf{I})\left(1 + \mathbf{d}^T(\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{c}\right) \\
&= \left(\pm\prod_{i=1}^{n}(\lambda_j - \lambda)\right)\left(1 + \frac{\mathbf{d}^T\mathbf{c}}{\lambda_k - \lambda}\right) \\
&= \left(\pm\prod_{j\neq k}(\lambda_j - \lambda)\right)\left(\lambda_k + \mathbf{d}^T\mathbf{c} - \lambda\right).
\end{aligned}$$

The roots of this polynomial are $\lambda_1, \ldots, \lambda_{k-1},\, \lambda_k + \mathbf{d}^T\mathbf{c},\, \lambda_{k+1}, \ldots, \lambda_n$.

(c) $\mathbf{d} = \dfrac{(\mu - \lambda_k)\mathbf{c}}{\mathbf{c}^T\mathbf{c}}$ will do the job.

7.1.18. (a) The transpose does not alter the determinant—recall (6.1.4)—so that

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\left(\mathbf{A}^T - \lambda\mathbf{I}\right).$$

(b) We know from Exercise 6.1.8 that $\overline{\det(\mathbf{A})} = \det(\mathbf{A}^*)$, so

$$\begin{aligned}\lambda \in \sigma(\mathbf{A}) &\Longleftrightarrow 0 = \det(\mathbf{A} - \lambda\mathbf{I}) \\ &\Longleftrightarrow 0 = \overline{\det(\mathbf{A} - \lambda\mathbf{I})} = \det\left((\mathbf{A} - \lambda\mathbf{I})^*\right) = \det\left(\mathbf{A}^* - \overline{\lambda}\mathbf{I}\right) \\ &\Longleftrightarrow \overline{\lambda} \in \sigma(\mathbf{A}^*).\end{aligned}$$

(c) Yes.

(d) Apply the reverse order law for conjugate transposes to obtain

$$\mathbf{y}^*\mathbf{A} = \mu\mathbf{y}^* \implies \mathbf{A}^*\mathbf{y} = \overline{\mu}\mathbf{y} \implies \mathbf{A}^T\mathbf{y} = \overline{\mu}\mathbf{y} \implies \overline{\mu} \in \sigma\left(\mathbf{A}^T\right) = \sigma(\mathbf{A}),$$

and use the conclusion of part (c) insuring that the eigenvalues of real matrices must occur in conjugate pairs.

7.1.19. (a) When $m = n$, Exercise 6.2.16 insures that

$$\lambda^n \det(\mathbf{AB} - \lambda\mathbf{I}) = \lambda^n \det(\mathbf{BA} - \lambda\mathbf{I}) \quad \text{for all } \lambda,$$

so $\det(\mathbf{AB} - \lambda\mathbf{I}) = \det(\mathbf{BA} - \lambda\mathbf{I})$.

(b) If $m \neq n$, then the characteristic polynomials of $\mathbf{AB}$ and $\mathbf{BA}$ are of degrees m and n, respectively, so they must be different. When m and n are different—say $m > n$—Exercise 6.2.16 implies that

$$\det(\mathbf{AB} - \lambda\mathbf{I}) = (-\lambda)^{m-n} \det(\mathbf{BA} - \lambda\mathbf{I}).$$

Consequently, $\mathbf{AB}$ has $m - n$ more zero eigenvalues than $\mathbf{BA}$.

7.1.20. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$, and suppose $\mathbf{X}$ is $n \times g$. The equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{BX} = \mathbf{0}$ says that the columns of $\mathbf{BX}$ are in $N(\mathbf{A} - \lambda\mathbf{I})$, and hence they are linear combinations of the basis vectors in $\mathbf{X}$. Thus

$$[\mathbf{BX}]_{*j} = \sum_i p_{ij}\mathbf{X}_{*j} \implies \mathbf{BX} = \mathbf{XP}, \quad \text{where } \mathbf{P}_{g\times g} = [p_{ij}].$$

If $(\mu, \mathbf{z})$ is any eigenpair for $\mathbf{P}$, then

$$\mathbf{B}(\mathbf{Xz}) = \mathbf{XPz} = \mu(\mathbf{Xz}) \quad \text{and} \quad \mathbf{AX} = \lambda\mathbf{X} \implies \mathbf{A}(\mathbf{Xz}) = \lambda(\mathbf{Xz}),$$

so $\mathbf{Xz}$ is a common eigenvector.

7.1.21. (a) If $\mathbf{Px} = \lambda\mathbf{x}$ and $\mathbf{y}^*\mathbf{Q} = \mu\mathbf{y}^*$, then $\mathbf{T}(xy^*) = \mathbf{Pxy}^*\mathbf{Q} = \lambda\mu\mathbf{xy}^*$.

(b) Since $\dim \mathcal{C}^{m\times n} = mn$, the operator $\mathbf{T}$ (as well as any coordinate matrix representation of $\mathbf{T}$) must have exactly mn eigenvalues (counting multiplicities), and since there are exactly mn products $\lambda\mu$, where $\lambda \in \sigma(\mathbf{P})$, $\mu \in \sigma(\mathbf{Q})$, it follows that $\sigma(\mathbf{T}) = \{\lambda\mu \mid \lambda \in \sigma(\mathbf{P}), \mu \in \sigma(\mathbf{Q})\}$. Use the fact that the trace is the sum of the eigenvalues (recall (7.1.7)) to conclude that $trace\,(\mathbf{T}) = \sum_{i,j} \lambda_i \mu_j = \sum_i \lambda_i \sum_j \mu_j = trace\,(\mathbf{P})\, trace\,(\mathbf{Q})$.

7.1.22. (a) Use (6.2.3) to compute the characteristic polynomial for $\mathbf{D} + \alpha \mathbf{v}\mathbf{v}^T$ to be

$$\begin{aligned} p(\lambda) &= \det\left(\mathbf{D} + \alpha \mathbf{v}\mathbf{v}^T - \lambda \mathbf{I}\right) \\ &= \det\left(\mathbf{D} - \lambda \mathbf{I} + \alpha \mathbf{v}\mathbf{v}^T\right) \\ &= \det\left(\mathbf{D} - \lambda \mathbf{I}\right)\left(1 + \alpha \mathbf{v}^T (\mathbf{D} - \lambda \mathbf{I})^{-1} \mathbf{v}\right) \qquad (\ddagger) \\ &= \left(\prod_{j=1}^{n} (\lambda - \lambda_j)\right)\left(1 + \alpha \sum_{i=1}^{n} \frac{v_i^2}{\lambda_i - \lambda}\right) \\ &= \prod_{j=1}^{n} (\lambda - \lambda_j) + \alpha \sum_{i=1}^{n} \left(v_i \prod_{j \neq i} (\lambda - \lambda_j)\right). \end{aligned}$$

For each λ_k, it is true that

$$p(\lambda_k) = \alpha v_k \prod_{j \neq k} (\lambda_k - \lambda_j) \neq 0,$$

and hence no λ_k can be an eigenvalue for $\mathbf{D} + \alpha \mathbf{v}\mathbf{v}^T$. Consequently, if ξ is an eigenvalue for $\mathbf{D} + \alpha \mathbf{v}\mathbf{v}^T$, then $\det\left(\mathbf{D} - \xi \mathbf{I}\right) \neq 0$, so $p(\xi) = 0$ and $(\ddagger)$ imply that

$$0 = 1 + \alpha \mathbf{v}^T (\mathbf{D} - \xi \mathbf{I})^{-1} \mathbf{v} = 1 + \alpha \sum_{i=1}^{n} \frac{v_i^2}{\lambda_i - \xi} = f(\xi).$$

(b) Use the fact that $f(\xi_i) = 1 + \alpha \mathbf{v}^T (\mathbf{D} - \xi_i \mathbf{I})^{-1} \mathbf{v} = 0$ to write

$$\begin{aligned} \left(\mathbf{D} + \alpha \mathbf{v}\mathbf{v}^T\right)(\mathbf{D} - \xi_i \mathbf{I})^{-1} \mathbf{v} &= \mathbf{D}(\mathbf{D} - \xi_i \mathbf{I})^{-1} \mathbf{v} + \mathbf{v}\left(\alpha \mathbf{v}^T (\mathbf{D} - \xi_i \mathbf{I})^{-1} \mathbf{v}\right) \\ &= \mathbf{D}(\mathbf{D} - \xi_i \mathbf{I})^{-1} \mathbf{v} - \mathbf{v} \\ &= \Big(\mathbf{D} - (\mathbf{D} - \xi_i \mathbf{I})\Big)(\mathbf{D} - \xi_i \mathbf{I})^{-1} \mathbf{v} \\ &= \xi_i (\mathbf{D} - \xi_i \mathbf{I})^{-1} \mathbf{v}. \end{aligned}$$

7.1.23. (a) If $p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$, then

$$\ln p(\lambda) = \sum_{i=1}^{n} \ln(\lambda - \lambda_i) \implies \frac{p'(\lambda)}{p(\lambda)} = \sum_{i=1}^{n} \frac{1}{(\lambda - \lambda_i)}.$$

(b) If $|\lambda_i / \lambda| < 1$, then we can write

$$(\lambda - \lambda_i)^{-1} = \left(\lambda\left(1 - \frac{\lambda_i}{\lambda}\right)\right)^{-1} = \frac{1}{\lambda}\left(1 - \frac{\lambda_i}{\lambda}\right)^{-1} = \frac{1}{\lambda}\left(1 + \frac{\lambda_i}{\lambda} + \frac{\lambda_i^2}{\lambda^2} + \cdots\right).$$

Consequently,

$$\sum_{i=1}^{n} \frac{1}{(\lambda - \lambda_i)} = \sum_{i=1}^{n} \left(\frac{1}{\lambda} + \frac{\lambda_i}{\lambda^2} + \frac{\lambda_i^2}{\lambda^3} + \cdots \right) = \frac{n}{\lambda} + \frac{\tau_1}{\lambda^2} + \frac{\tau_2}{\lambda^3} + \cdots .$$

(c) Combining these two results yields

$$\begin{aligned} &n\lambda^{n-1} + (n-1)c_1\lambda^{n-2} + (n-2)c_2\lambda^{n-3} + \cdots + c_{n-1} \\ &= \left(\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \cdots + c_n\right)\left(\frac{n}{\lambda} + \frac{\tau_1}{\lambda^2} + \frac{\tau_2}{\lambda^3} + \cdots\right) \\ &= n\lambda^{n-1} + (nc_1 + \tau_1)\,\lambda^{n-2} + (nc_2 + \tau_1 c_1 + \tau_2)\,\lambda^{n-3} \\ &\qquad + \cdots + (nc_{n-1} + \tau_1 c_{n-2} + \tau_2 c_{n-3} + \cdots + \tau_{n-1}) \\ &\qquad + (nc_n + \tau_1 c_{n-1} + \tau_2 c_{n-2} \cdots + \tau_n)\,\frac{1}{\lambda} + \cdots , \end{aligned}$$

and equating like powers of λ produces the desired conclusion.

7.1.24. We know from Exercise 7.1.10 that $\lambda \in \sigma(\mathbf{A}) \implies \lambda^k \in \sigma\left(\mathbf{A}^k\right)$, so (7.1.7) guarantees that $trace\left(\mathbf{A}^k\right) = \sum_i \lambda_i^k = \tau_k$. Proceed by induction. The result is true for $k = 1$ because (7.1.7) says that $c_1 = -trace(\mathbf{A})$. Assume that

$$c_i = -\frac{trace(\mathbf{A}\mathbf{B}_{i-1})}{i} \quad \text{for} \quad i = 1, 2, \ldots, k-1,$$

and prove the result holds for $i = k$. Recursive application of the induction hypothesis produces

$$\begin{aligned} \mathbf{B}_1 &= c_1\mathbf{I} + \mathbf{A} \\ \mathbf{B}_2 &= c_2\mathbf{I} + c_1\mathbf{A} + \mathbf{A}^2 \\ &\vdots \\ \mathbf{B}_{k-1} &= c_{k-1}\mathbf{I} + c_{k-2}\mathbf{A} + \cdots + c_1\mathbf{A}^{k-2} + \mathbf{A}^{k-1}, \end{aligned}$$

and therefore we can use Newton's identities given in Exercise 7.1.23 to obtain

$$\begin{aligned} trace(\mathbf{A}\mathbf{B}_{k-1}) &= trace\left(c_{k-1}\mathbf{A} + c_{k-2}\mathbf{A}^2 + \cdots + c_1\mathbf{A}^{k-1} + \mathbf{A}^k\right) \\ &= c_{k-1}\tau_1 + c_{k-2}\tau_2 + \cdots + c_1\tau_{k-1} + \tau_k \\ &= -kc_k. \end{aligned}$$