7.2 DIAGONALIZATION BY SIMILARITY TRANSFORMATIONS

The correct choice of a coordinate system (or basis) often can simplify the form of an equation or the analysis of a particular problem. For example, consider the obliquely oriented ellipse in Figure 7.2.1 whose equation in the xy-coordinate system is

$$13x^2 + 10xy + 13y^2 = 72.$$

By rotating the xy-coordinate system counterclockwise through an angle of 45°

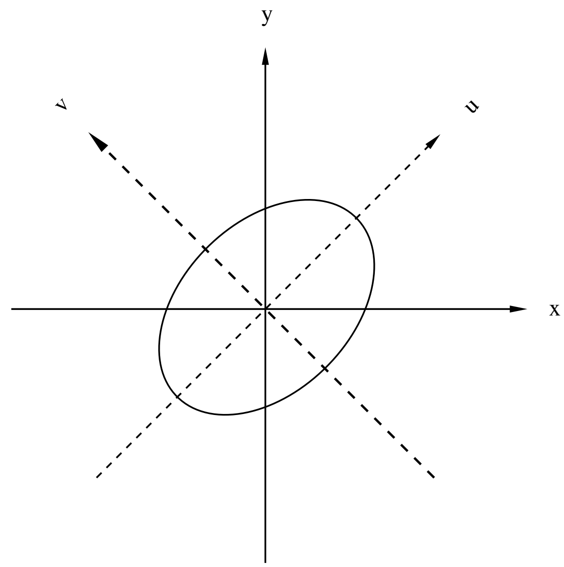

FIGURE 7.2.1

into a uv-coordinate system by means of (5.6.13) on p. 326, the cross-product term is eliminated, and the equation of the ellipse simplifies to become

$$\frac{u^2}{9} + \frac{v^2}{4} = 1.$$

It's shown in Example 7.6.3 on p. 567 that we can do a similar thing for quadratic equations in $\Re^n$.

Choosing or changing to the most appropriate coordinate system (or basis) is always desirable, but in linear algebra it is fundamental. For a linear operator $\mathbf{L}$ on a finite-dimensional space $\mathcal{V}$, the goal is to find a basis $\mathcal{B}$ for $\mathcal{V}$ such that the matrix representation of $\mathbf{L}$ with respect to $\mathcal{B}$ is as simple as possible. Since different matrix representations $\mathbf{A}$ and $\mathbf{B}$ of $\mathbf{L}$ are related by a similarity transformation $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{B}$ (recall §4.8),[70] the fundamental problem for linear operators is strictly a matrix issue—i.e., find a nonsingular matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{AP}$ is as simple as possible. The concept of similarity was first introduced on p. 255, but in the interest of continuity it is reviewed below.

---

70 While it is helpful to have covered the topics in §§4.7–4.9, much of the subsequent development is accessible without an understanding of this material.

Similarity

- Two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be ***similar*** whenever there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{B}$. The product $\mathbf{P}^{-1}\mathbf{AP}$ is called a ***similarity transformation*** on $\mathbf{A}$.
- **A Fundamental Problem.** Given a square matrix $\mathbf{A}$, reduce it to the simplest possible form by means of a similarity transformation.

Diagonal matrices have the simplest form, so we first ask, "Is every square matrix similar to a diagonal matrix?" Linear algebra and matrix theory would be simpler subjects if this were true, but it's not. For example, consider

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{7.2.1}$$

and observe that $\mathbf{A}^2 = \mathbf{0}$ ($\mathbf{A}$ is nilpotent). If there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$, where $\mathbf{D}$ is diagonal, then

$$\mathbf{D}^2 = \mathbf{P}^{-1}\mathbf{APP}^{-1}\mathbf{AP} = \mathbf{P}^{-1}\mathbf{A}^2\mathbf{P} = \mathbf{0} \implies \mathbf{D} = \mathbf{0} \implies \mathbf{A} = \mathbf{0},$$

which is false. Thus $\mathbf{A}$, as well as any other nonzero nilpotent matrix, is not similar to a diagonal matrix. Nonzero nilpotent matrices are not the only ones that can't be diagonalized, but, as we will see, nilpotent matrices play a particularly important role in nondiagonalizability.

So, if not all square matrices can be diagonalized by a similarity transformation, what are the characteristics of those that can? An answer is easily derived by examining the equation

$$\mathbf{P}^{-1}\mathbf{A}_{n\times n}\mathbf{P} = \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

which implies $\mathbf{A}\left[\mathbf{P}_{*1} \,|\, \cdots \,|\, \mathbf{P}_{*n}\right] = \left[\mathbf{P}_{*1} \,|\, \cdots \,|\, \mathbf{P}_{*n}\right] \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$ or, equivalently, $\left[\mathbf{AP}_{*1} \,|\, \cdots \,|\, \mathbf{AP}_{*n}\right] = \left[\lambda_1\mathbf{P}_{*1} \,|\, \cdots \,|\, \lambda_n\mathbf{P}_{*n}\right]$. Consequently, $\mathbf{AP}_{*j} = \lambda_j\mathbf{P}_{*j}$ for each j, so each $(\lambda_j, \mathbf{P}_{*j})$ is an eigenpair for $\mathbf{A}$. In other words, $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$ implies that $\mathbf{P}$ must be a matrix whose columns constitute n linearly independent eigenvectors, and $\mathbf{D}$ is a diagonal matrix whose diagonal entries are the corresponding eigenvalues. It's straightforward to reverse the above argument to prove the converse—i.e., if there exists a linearly independent set of n eigenvectors that are used as columns to build a nonsingular matrix $\mathbf{P}$, and if $\mathbf{D}$ is the diagonal matrix whose diagonal entries are the corresponding eigenvalues, then $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$. Below is a summary.

Diagonalizability

- A square matrix $\mathbf{A}$ is said to be ***diagonalizable*** whenever $\mathbf{A}$ is similar to a diagonal matrix.
- A ***complete set of eigenvectors*** for $\mathbf{A}_{n\times n}$ is any set of n linearly independent eigenvectors for $\mathbf{A}$. Not all matrices have complete sets of eigenvectors—e.g., consider (7.2.1) or Example 7.1.2. Matrices that fail to possess complete sets of eigenvectors are sometimes called ***deficient*** or ***defective*** matrices.
- $\mathbf{A}_{n\times n}$ is diagonalizable if and only if $\mathbf{A}$ possesses a complete set of eigenvectors. Moreover, $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ if and only if the columns of $\mathbf{P}$ constitute a complete set of eigenvectors and the λ_j 's are the associated eigenvalues—i.e., each $(\lambda_j, \mathbf{P}_{*j})$ is an eigenpair for $\mathbf{A}$.

Example 7.2.1

Problem: If possible, diagonalize the following matrix with a similarity transformation:

$$\mathbf{A} = \begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix}.$$

Solution: Determine whether or not $\mathbf{A}$ has a complete set of three linearly independent eigenvectors. The characteristic equation—perhaps computed by using (7.1.5)—is

$$\lambda^3 + 5\lambda^2 + 3\lambda - 9 = (\lambda - 1)(\lambda + 3)^2 = 0.$$

Therefore, $\lambda = 1$ is a simple eigenvalue, and $\lambda = -3$ is repeated twice (we say its algebraic multiplicity is 2). Bases for the eigenspaces $N(\mathbf{A} - 1\mathbf{I})$ and $N(\mathbf{A} + 3\mathbf{I})$ are determined in the usual way to be

$$N(\mathbf{A} - 1\mathbf{I}) = span\left\{\begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}\right\} \quad \text{and} \quad N(\mathbf{A} + 3\mathbf{I}) = span\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\right\},$$

and it's easy to check that when combined these three eigenvectors constitute a linearly independent set. Consequently, $\mathbf{A}$ must be diagonalizable. To explicitly exhibit the similarity transformation that diagonalizes $\mathbf{A}$, set

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}, \quad \text{and verify} \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix} = \mathbf{D}.$$

Since not all square matrices are diagonalizable, it's natural to inquire about the next best thing—i.e., can every square matrix be *triangularized* by similarity? This time the answer is *yes,* but before explaining why, we need to make the following observation.

Similarity Preserves Eigenvalues

Row reductions don't preserve eigenvalues (try a simple example). However, similar matrices have the same characteristic polynomial, so they have the same eigenvalues with the same multiplicities. **Caution!** Similar matrices need not have the same eigenvectors—see Exercise 7.2.3.

Proof. Use the product rule for determinants in conjunction with the fact that $\det\left(\mathbf{P}^{-1}\right) = 1/\det(\mathbf{P})$ (Exercise 6.1.6) to write

$$\begin{aligned}\det(\mathbf{A}-\lambda\mathbf{I}) &= \det\left(\mathbf{P}^{-1}\mathbf{B}\mathbf{P}-\lambda\mathbf{I}\right) = \det\left(\mathbf{P}^{-1}(\mathbf{B}-\lambda\mathbf{I})\mathbf{P}\right)\\ &= \det\left(\mathbf{P}^{-1}\right)\det(\mathbf{B}-\lambda\mathbf{I})\det(\mathbf{P}) = \det(\mathbf{B}-\lambda\mathbf{I}). \qquad \blacksquare\end{aligned}$$

In the context of linear operators, this means that the eigenvalues of a matrix representation of an operator $\mathbf{L}$ are invariant under a change of basis. In other words, the eigenvalues are intrinsic to $\mathbf{L}$ in the sense that they are independent of any coordinate representation.

Now we can establish the fact that every square matrix can be triangularized by a similarity transformation. In fact, as Issai Schur (p. 123) realized in 1909, the similarity transformation always can be made to be unitary.

Schur's Triangularization Theorem

Every square matrix is unitarily similar to an upper-triangular matrix. That is, for each $\mathbf{A}_{n\times n}$, there exists a unitary matrix $\mathbf{U}$ (not unique) and an upper-triangular matrix $\mathbf{T}$ (not unique) such that $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{T}$, and the diagonal entries of $\mathbf{T}$ are the eigenvalues of $\mathbf{A}$.

Proof. Use induction on n, the size of the matrix. For $n = 1$, there is nothing to prove. For $n > 1$, assume that all $n-1 \times n-1$ matrices are unitarily similar to an upper-triangular matrix, and consider an $n \times n$ matrix $\mathbf{A}$. Suppose that $(\lambda, \mathbf{x})$ is an eigenpair for $\mathbf{A}$, and suppose that $\mathbf{x}$ has been normalized so that $\|\mathbf{x}\|_2 = 1$. As discussed on p. 325, we can construct an elementary reflector $\mathbf{R} = \mathbf{R}^* = \mathbf{R}^{-1}$ with the property that $\mathbf{R}\mathbf{x} = \mathbf{e}_1$ or, equivalently, $\mathbf{x} = \mathbf{R}\mathbf{e}_1$ (set $\mathbf{R} = \mathbf{I}$ if $\mathbf{x} = \mathbf{e}_1$). Thus $\mathbf{x}$ is the first column in $\mathbf{R}$, so $\mathbf{R} = \left(\mathbf{x}\,|\,\mathbf{V}\right)$, and

$$\mathbf{R}\mathbf{A}\mathbf{R} = \mathbf{R}\mathbf{A}\left(\mathbf{x}\,|\,\mathbf{V}\right) = \mathbf{R}\left(\lambda\mathbf{x}\,|\,\mathbf{A}\mathbf{V}\right) = \left(\lambda\mathbf{e}_1\,|\,\mathbf{R}\mathbf{A}\mathbf{V}\right) = \begin{pmatrix}\lambda & \mathbf{x}^*\mathbf{A}\mathbf{V}\\ \mathbf{0} & \mathbf{V}^*\mathbf{A}\mathbf{V}\end{pmatrix}.$$

Since $\mathbf{V}^*\mathbf{AV}$ is $n-1 \times n-1$, the induction hypothesis insures that there exists a unitary matrix $\mathbf{Q}$ such that $\mathbf{Q}^*(\mathbf{V}^*\mathbf{AV})\mathbf{Q} = \tilde{\mathbf{T}}$ is upper triangular. If $\mathbf{U} = \mathbf{R}\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}$, then $\mathbf{U}$ is unitary (because $\mathbf{U}^* = \mathbf{U}^{-1}$), and

$$\mathbf{U}^*\mathbf{AU} = \begin{pmatrix} \lambda & \mathbf{x}^*\mathbf{AVQ} \\ \mathbf{0} & \mathbf{Q}^*\mathbf{V}^*\mathbf{AVQ} \end{pmatrix} = \begin{pmatrix} \lambda & \mathbf{x}^*\mathbf{AVQ} \\ \mathbf{0} & \tilde{\mathbf{T}} \end{pmatrix} = \mathbf{T}$$

is upper triangular. Since similar matrices have the same eigenvalues, and since the eigenvalues of a triangular matrix are its diagonal entries (Exercise 7.1.3), the diagonal entries of $\mathbf{T}$ must be the eigenvalues of $\mathbf{A}$. ■

Example 7.2.2

The Cayley–Hamilton[71] theorem asserts that every square matrix satisfies its own characteristic equation $p(\lambda) = 0$. That is, $p(\mathbf{A}) = \mathbf{0}$.

Problem: Show how the Cayley–Hamilton theorem follows from Schur's triangularization theorem.

Solution: Schur's theorem insures the existence of a unitary $\mathbf{U}$ such that $\mathbf{U}^*\mathbf{AU} = \mathbf{T}$ is triangular, and the development allows for the eigenvalues $\mathbf{A}$ to appear in any given order on the diagonal of $\mathbf{T}$. So, if $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ with λ_i repeated a_i times, then there is a unitary $\mathbf{U}$ such that

$$\mathbf{U}^*\mathbf{AU} = \mathbf{T} = \begin{pmatrix} \mathbf{T}_1 & \star & \cdots & \star \\ & \mathbf{T}_2 & \cdots & \star \\ & & \ddots & \vdots \\ & & & \mathbf{T}_k \end{pmatrix}, \quad \text{where} \quad \mathbf{T}_i = \begin{pmatrix} \lambda_i & \star & \cdots & \star \\ & \lambda_i & \cdots & \star \\ & & \ddots & \vdots \\ & & & \lambda_i \end{pmatrix}_{a_i \times a_i}.$$

Consequently, $(\mathbf{T}_i - \lambda_i\mathbf{I})^{a_i} = \mathbf{0}$, so $(\mathbf{T} - \lambda_i\mathbf{I})^{a_i}$ has the form

$$(\mathbf{T} - \lambda_i\mathbf{I})^{a_i} = \begin{pmatrix} \star & \cdots & \star & \cdots & \star \\ & \ddots & \vdots & & \vdots \\ & & \mathbf{0} & \cdots & \star \\ & & & \ddots & \vdots \\ & & & & \star \end{pmatrix} \longleftarrow i^{th} \text{ row of blocks.}$$

[71] William Rowan Hamilton (1805–1865), an Irish mathematical astronomer, established this result in 1853 for his quaternions, matrices of the form $\begin{pmatrix} a+b\mathrm{i} & c+d\mathrm{i} \\ -c+d\mathrm{i} & a-b\mathrm{i} \end{pmatrix}$ that resulted from his attempt to generalize complex numbers. In 1858 Arthur Cayley (p. 80) enunciated the general result, but his argument was simply to make direct computations for 2×2 and 3×3 matrices. Cayley apparently didn't appreciate the subtleties of the result because he stated that a formal proof "was not necessary." Hamilton's quaternions took shape in his mind while walking with his wife along the Royal Canal in Dublin, and he was so inspired that he stopped to carve his idea in the stone of the Brougham Bridge. He believed quaternions would revolutionize mathematical physics, and he spent the rest of his life working on them. But the world did not agree. Hamilton became an unhappy man addicted to alcohol who is reported to have died from a severe attack of gout.

This form insures that $(\mathbf{T}-\lambda_1\mathbf{I})^{a_1}(\mathbf{T}-\lambda_2\mathbf{I})^{a_2}\cdots(\mathbf{T}-\lambda_k\mathbf{I})^{a_k}=\mathbf{0}$. The characteristic equation for $\mathbf{A}$ is $p(\lambda)=(\lambda-\lambda_1)^{a_1}(\lambda-\lambda_2)^{a_2}\cdots(\lambda-\lambda_k)^{a_k}=0$, so

$$\begin{aligned}\mathbf{U}^*p(\mathbf{A})\mathbf{U}&=\mathbf{U}^*(\mathbf{A}-\lambda_1\mathbf{I})^{a_1}(\mathbf{A}-\lambda_2\mathbf{I})^{a_2}\cdots(\mathbf{A}-\lambda_k\mathbf{I})^{a_k}\mathbf{U}\\&=(\mathbf{T}-\lambda_1\mathbf{I})^{a_1}(\mathbf{T}-\lambda_2\mathbf{I})^{a_2}\cdots(\mathbf{T}-\lambda_k\mathbf{I})^{a_k}=\mathbf{0},\end{aligned}$$

and thus $p(\mathbf{A})=\mathbf{0}$. **Note:** A completely different approach to the Cayley–Hamilton theorem is discussed on p. 532.

Schur's theorem is not the complete story on triangularizing by similarity. By allowing nonunitary similarity transformations, the structure of the upper-triangular matrix $\mathbf{T}$ can be simplified to contain zeros everywhere except on the diagonal and the superdiagonal (the diagonal immediately above the main diagonal). This is the Jordan form developed on p. 590, but some of the seeds are sown here.

Multiplicities

For $\lambda\in\sigma(\mathbf{A})=\{\lambda_1,\lambda_2,\ldots,\lambda_s\}$, we adopt the following definitions.

- The ***algebraic multiplicity*** of λ is the number of times it is repeated as a root of the characteristic polynomial. In other words, $\textit{alg mult}_{\mathbf{A}}(\lambda_i)=a_i$ if and only if $(x-\lambda_1)^{a_1}\cdots(x-\lambda_s)^{a_s}=0$ is the characteristic equation for $\mathbf{A}$.
- When $\textit{alg mult}_{\mathbf{A}}(\lambda)=1$, λ is called a ***simple eigenvalue.***
- The ***geometric multiplicity*** of λ is $\dim N(\mathbf{A}-\lambda\mathbf{I})$. In other words, $\textit{geo mult}_{\mathbf{A}}(\lambda)$ is the maximal number of linearly independent eigenvectors associated with λ.
- Eigenvalues such that $\textit{alg mult}_{\mathbf{A}}(\lambda)=\textit{geo mult}_{\mathbf{A}}(\lambda)$ are called ***semisimple eigenvalues*** of $\mathbf{A}$. It follows from (7.2.2) on p. 511 that a simple eigenvalue is always semisimple, but not conversely.

Example 7.2.3

The algebraic and geometric multiplicity need not agree. For example, the nilpotent matrix $\mathbf{A}=\begin{pmatrix}0&1\\0&0\end{pmatrix}$ in (7.2.1) has only one distinct eigenvalue, $\lambda=0$, that is repeated twice, so $\textit{alg mult}_{\mathbf{A}}(0)=2$. But

$$\dim N(\mathbf{A}-0\mathbf{I})=\dim N(\mathbf{A})=1\quad\Longrightarrow\quad \textit{geo mult}_{\mathbf{A}}(0)=1.$$

In other words, there is only one linearly independent eigenvector associated with $\lambda=0$ even though $\lambda=0$ is repeated twice as an eigenvalue.

Example 7.2.3 shows that $\textit{geo mult}_{\mathbf{A}}(\lambda)<\textit{alg mult}_{\mathbf{A}}(\lambda)$ is possible. However, the inequality can never go in the reverse direction.

Multiplicity Inequality

For every $\mathbf{A} \in \mathcal{C}^{n \times n}$, and for each $\lambda \in \sigma(\mathbf{A})$,

$$geo\ mult_{\mathbf{A}}(\lambda) \leq alg\ mult_{\mathbf{A}}(\lambda). \tag{7.2.2}$$

Proof. Suppose $alg\ mult_{\mathbf{A}}(\lambda) = k$. Schur's triangularization theorem (p. 508) insures the existence of a unitary $\mathbf{U}$ such that $\mathbf{U}^*\mathbf{A}_{n\times n}\mathbf{U} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{0} & \mathbf{T}_{22} \end{pmatrix}$, where $\mathbf{T}_{11}$ is a $k \times k$ upper-triangular matrix whose diagonal entries are equal to λ, and $\mathbf{T}_{22}$ is an $n-k \times n-k$ upper-triangular matrix with $\lambda \notin \sigma(\mathbf{T}_{22})$. Consequently, $\mathbf{T}_{22} - \lambda\mathbf{I}$ is nonsingular, so

$$\begin{aligned} rank(\mathbf{A} - \lambda\mathbf{I}) = rank(\mathbf{U}^*(\mathbf{A} - \lambda\mathbf{I})\mathbf{U}) &= rank \begin{pmatrix} \mathbf{T}_{11} - \lambda\mathbf{I} & \mathbf{T}_{12} \\ \mathbf{0} & \mathbf{T}_{22} - \lambda\mathbf{I} \end{pmatrix} \\ &\geq rank(\mathbf{T}_{22} - \lambda\mathbf{I}) = n - k. \end{aligned}$$

The inequality follows from the fact that the rank of a matrix is at least as great as the rank of any submatrix—recall the result on p. 215. Therefore,

$$alg\ mult_{\mathbf{A}}(\lambda) = k \geq n - rank(\mathbf{A} - \lambda\mathbf{I}) = \dim N(\mathbf{A} - \lambda\mathbf{I}) = geo\ mult_{\mathbf{A}}(\lambda). \quad \blacksquare$$

Determining whether or not $\mathbf{A}_{n\times n}$ is diagonalizable is equivalent to determining whether or not $\mathbf{A}$ has a complete linearly independent set of eigenvectors, and this can be done if you are willing and able to compute all of the eigenvalues and eigenvectors for $\mathbf{A}$. But this brute force approach can be a monumental task. Fortunately, there are some theoretical tools to help determine how many linearly independent eigenvectors a given matrix possesses.

Independent Eigenvectors

Let $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ be a set of distinct eigenvalues for $\mathbf{A}$.

- If $\{(\lambda_1, \mathbf{x}_1), (\lambda_2, \mathbf{x}_2), \ldots, (\lambda_k, \mathbf{x}_k)\}$ is a set of eigenpairs for $\mathbf{A}$, then $\mathcal{S} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is a linearly independent set. (7.2.3)
- If $\mathcal{B}_i$ is a basis for $N(\mathbf{A} - \lambda_i\mathbf{I})$, then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ is a linearly independent set. (7.2.4)

Proof of (7.2.3). Suppose $\mathcal{S}$ is a dependent set. If the vectors in $\mathcal{S}$ are arranged so that $\mathcal{M} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ is a maximal linearly independent subset, then

$$\mathbf{x}_{r+1} = \sum_{i=1}^{r} \alpha_i \mathbf{x}_i,$$

and multiplication on the left by $\mathbf{A} - \lambda_{r+1}\mathbf{I}$ produces

$$\mathbf{0} = \sum_{i=1}^{r} \alpha_i \left(\mathbf{A}\mathbf{x}_i - \lambda_{r+1}\mathbf{x}_i\right) = \sum_{i=1}^{r} \alpha_i \left(\lambda_i - \lambda_{r+1}\right) \mathbf{x}_i.$$

Because $\mathcal{M}$ is linearly independent, $\alpha_i \left(\lambda_i - \lambda_{r+1}\right) = 0$ for each i. Consequently, $\alpha_i = 0$ for each i (because the eigenvalues are distinct), and hence $\mathbf{x}_{r+1} = \mathbf{0}$. But this is impossible because eigenvectors are nonzero. Therefore, the supposition that $\mathcal{S}$ is a dependent set must be false. ■

Proof of (7.2.4). The result of Exercise 5.9.14 guarantees that $\mathcal{B}$ is linearly independent if and only if

$$\mathcal{M}_j = N\left(\mathbf{A} - \lambda_j\mathbf{I}\right) \cap \Big[N\left(\mathbf{A} - \lambda_1\mathbf{I}\right) + N\left(\mathbf{A} - \lambda_2\mathbf{I}\right) + \cdots + N\left(\mathbf{A} - \lambda_{j-1}\mathbf{I}\right)\Big] = \mathbf{0}$$

for each $j = 1, 2, \ldots, k$. Suppose we have $\mathbf{0} \neq \mathbf{x} \in \mathcal{M}_j$ for some j. Then $\mathbf{A}\mathbf{x} = \lambda_j\mathbf{x}$ and $\mathbf{x} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_{j-1}$ for $\mathbf{v}_i \in N\left(\mathbf{A} - \lambda_i\mathbf{I}\right)$, which implies

$$\sum_{i=1}^{j-1}(\lambda_i - \lambda_j)\mathbf{v}_i = \sum_{i=1}^{j-1} \lambda_i\mathbf{v}_i - \lambda_j \sum_{i=1}^{j-1} \mathbf{v}_i = \mathbf{A}\mathbf{x} - \lambda_j\mathbf{x} = \mathbf{0}.$$

By (7.2.3), the $\mathbf{v}_i$'s are linearly independent, and hence $\lambda_i - \lambda_j = 0$ for each $i = 1, 2, \ldots, j-1$. But this is impossible because the eigenvalues are distinct. Therefore, $\mathcal{M}_j = \mathbf{0}$ for each j, and thus $\mathcal{B}$ is linearly independent. ■

These results lead to the following characterization of diagonalizability.

Diagonalizability and Multiplicities

A matrix $\mathbf{A}_{n\times n}$ is diagonalizable if and only if

$$\textit{geo mult}_{\mathbf{A}}\left(\lambda\right) = \textit{alg mult}_{\mathbf{A}}\left(\lambda\right) \tag{7.2.5}$$

for each $\lambda \in \sigma(\mathbf{A})$—i.e., if and only if every eigenvalue is semisimple.

Proof. Suppose $geo\ mult_{\mathbf{A}}(\lambda_i) = alg\ mult_{\mathbf{A}}(\lambda_i) = a_i$ for each eigenvalue λ_i. If there are k distinct eigenvalues, and if $\mathcal{B}_i$ is a basis for $N(\mathbf{A} - \lambda_i\mathbf{I})$, then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ contains $\sum_{i=1}^{k} a_i = n$ vectors. We just proved in (7.2.4) that $\mathcal{B}$ is a linearly independent set, so $\mathcal{B}$ represents a complete set of linearly independent eigenvectors of $\mathbf{A}$, and we know this insures that $\mathbf{A}$ must be diagonalizable. Conversely, if $\mathbf{A}$ is diagonalizable, and if λ is an eigenvalue for $\mathbf{A}$ with $alg\ mult_{\mathbf{A}}(\lambda) = a$, then there is a nonsingular matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{pmatrix} \lambda\mathbf{I}_{a\times a} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix},$$

where $\lambda \notin \sigma(\mathbf{B})$. Consequently,

$$rank\,(\mathbf{A} - \lambda\mathbf{I}) = rank\ \mathbf{P}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} - \lambda\mathbf{I} \end{pmatrix}\mathbf{P}^{-1} = rank\,(\mathbf{B} - \lambda\mathbf{I}) = n - a,$$

and thus

$$geo\ mult_{\mathbf{A}}(\lambda) = \dim N(\mathbf{A} - \lambda\mathbf{I}) = n - rank\,(\mathbf{A} - \lambda\mathbf{I}) = a = alg\ mult_{\mathbf{A}}(\lambda). \quad \blacksquare$$

Example 7.2.4

Problem: Determine if either of the following matrices is diagonalizable:

$$\mathbf{A} = \begin{pmatrix} -1 & -1 & -2 \\ 8 & -11 & -8 \\ -10 & 11 & 7 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix}.$$

Solution: Each matrix has exactly the same characteristic equation

$$\lambda^3 + 5\lambda^2 + 3\lambda - 9 = (\lambda - 1)(\lambda + 3)^2 = 0,$$

so $\sigma(\mathbf{A}) = \{1, -3\} = \sigma(\mathbf{B})$, where $\lambda = 1$ has algebraic multiplicity 1 and $\lambda = -3$ has algebraic multiplicity 2. Since

$$geo\ mult_{\mathbf{A}}(-3) = \dim N(\mathbf{A} + 3\mathbf{I}) = 1 < alg\ mult_{\mathbf{A}}(-3),$$

$\mathbf{A}$ is *not* diagonalizable. On the other hand,

$$geo\ mult_{\mathbf{B}}(-3) = \dim N(\mathbf{B} + 3\mathbf{I}) = 2 = alg\ mult_{\mathbf{B}}(-3),$$

and $geo\ mult_{\mathbf{B}}(1) = 1 = alg\ mult_{\mathbf{B}}(1)$, so $\mathbf{B}$ *is* diagonalizable.

If $\mathbf{A}_{n\times n}$ happens to have n distinct eigenvalues, then each eigenvalue is simple. This means that $geo\ mult_{\mathbf{A}}(\lambda) = alg\ mult_{\mathbf{A}}(\lambda) = 1$ for each λ, so (7.2.5) produces the following corollary guaranteeing diagonalizability.

Distinct Eigenvalues

If no eigenvalue of $\mathbf{A}$ is repeated, then $\mathbf{A}$ is diagonalizable. (7.2.6)
Caution! The converse is not true—see Example 7.2.4.

Example 7.2.5

Toeplitz[72] **matrices** have constant entries on each diagonal parallel to the main diagonal. For example, a 4×4 Toeplitz matrix $\mathbf{T}$ along with a *tridiagonal* Toeplitz matrix $\mathbf{A}$ are shown below:

$$\mathbf{T} = \begin{pmatrix} t_0 & t_1 & t_2 & t_3 \\ t_{-1} & t_0 & t_1 & t_2 \\ t_{-2} & t_{-1} & t_0 & t_1 \\ t_{-3} & t_{-2} & t_{-1} & t_0 \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} t_0 & t_1 & 0 & 0 \\ t_{-1} & t_0 & t_1 & 0 \\ 0 & t_{-1} & t_0 & t_1 \\ 0 & 0 & t_{-1} & t_0 \end{pmatrix}.$$

Toeplitz structures occur naturally in a variety of applications, and tridiagonal Toeplitz matrices are commonly the result of discretizing differential equation problems—e.g., see §1.4 (p. 18) and Example 7.6.1 (p. 559). The Toeplitz structure is rich in special properties, but tridiagonal Toeplitz matrices are particularly nice because they are among the few nontrivial structures that admit formulas for their eigenvalues and eigenvectors.

Problem: Show that the eigenvalues and eigenvectors of

$$\mathbf{A} = \begin{pmatrix} b & a & & & \\ c & b & a & & \\ & \ddots & \ddots & \ddots & \\ & & c & b & a \\ & & & c & b \end{pmatrix}_{n \times n} \qquad \text{with} \qquad a \neq 0 \neq c$$

are given by

$$\lambda_j = b + 2a\sqrt{c/a}\,\cos\left(\frac{j\pi}{n+1}\right) \qquad \text{and} \qquad \mathbf{x}_j = \begin{pmatrix} (c/a)^{1/2}\sin\left(1j\pi/(n+1)\right) \\ (c/a)^{2/2}\sin\left(2j\pi/(n+1)\right) \\ (c/a)^{3/2}\sin\left(3j\pi/(n+1)\right) \\ \vdots \\ (c/a)^{n/2}\sin\left(nj\pi/(n+1)\right) \end{pmatrix}$$

[72] Otto Toeplitz (1881–1940) was a professor in Bonn, Germany, but because of his Jewish background he was dismissed from his chair by the Nazis in 1933. In addition to the matrix that bears his name, Toeplitz is known for his general theory of infinite-dimensional spaces developed in the 1930s.

for $j = 1, 2, \ldots, n,$ and conclude that $\mathbf{A}$ is diagonalizable.

Solution: For an eigenpair $(\lambda, \mathbf{x})$, the components in $(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0}$ are $cx_{k-1} + (b-\lambda)x_k + ax_{k+1} = 0,\ k = 1, \ldots, n$ with $x_0 = x_{n+1} = 0$ or, equivalently,

$$x_{k+2} + \left(\frac{b-\lambda}{a}\right) x_{k+1} + \left(\frac{c}{a}\right) x_k = 0 \quad \text{for } k = 0, \ldots, n-1 \text{ with } x_0 = x_{n+1} = 0.$$

These are second-order homogeneous difference equations, and solving them is similar to solving analogous differential equations. The technique is to seek solutions of the form $x_k = \xi r^k$ for constants ξ and r. This produces the quadratic equation $r^2 + (b-\lambda)r/a + c/a = 0$ with roots r_1 and r_2, and it can be argued that the general solution of $x_{k+2} + ((b-\lambda)/a)x_{k+1} + (c/a)x_k = 0$ is

$$x_k = \begin{cases} \alpha r_1^k + \beta r_2^k & \text{if } r_1 \neq r_2, \\ \alpha \rho^k + \beta k \rho^k & \text{if } r_1 = r_2 = \rho, \end{cases} \qquad \text{where } \alpha \text{ and } \beta \text{ are arbitrary constants.}$$

For the eigenvalue problem at hand, r_1 and r_2 must be distinct—otherwise $x_k = \alpha\rho^k + \beta k\rho^k$, and $x_0 = x_{n+1} = 0$ implies each $x_k = 0$, which is impossible because $\mathbf{x}$ is an eigenvector. Hence $x_k = \alpha r_1^k + \beta r_2^k$, and $x_0 = x_{n+1} = 0$ yields

$$\left\{ \begin{array}{l} 0 = \alpha + \beta \\ 0 = \alpha r_1^{n+1} + \beta r_2^{n+1} \end{array} \right\} \implies \left(\frac{r_1}{r_2}\right)^{n+1} = \frac{-\beta}{\alpha} = 1 \implies \frac{r_1}{r_2} = \mathrm{e}^{\mathrm{i}2\pi j/(n+1)},$$

so $r_1 = r_2 \mathrm{e}^{\mathrm{i}2\pi j/(n+1)}$ for some $1 \le j \le n$. Couple this with

$$r^2 + \frac{(b-\lambda)r}{a} + \frac{c}{a} = (r - r_1)(r - r_2) \implies \begin{cases} r_1 r_2 = c/a \\ r_1 + r_2 = -(b-\lambda)/a \end{cases}$$

to conclude that $r_1 = \sqrt{c/a}\,\mathrm{e}^{\mathrm{i}\pi j/(n+1)},\ \ r_2 = \sqrt{c/a}\,\mathrm{e}^{-\mathrm{i}\pi j/(n+1)}$, and

$$\lambda = b + a\sqrt{c/a}\left(\mathrm{e}^{\mathrm{i}\pi j/(n+1)} + \mathrm{e}^{-\mathrm{i}\pi j/(n+1)}\right) = b + 2a\sqrt{c/a}\,\cos\left(\frac{j\pi}{n+1}\right).$$

Therefore, the eigenvalues of $\mathbf{A}$ must be given by

$$\lambda_j = b + 2a\sqrt{c/a}\,\cos\left(\frac{j\pi}{n+1}\right), \quad j = 1, 2, \ldots, n.$$

Since these λ_j 's are all distinct ($\cos\theta$ is a strictly decreasing function of θ on $(0, \pi)$, and $a \neq 0 \neq c$), $\mathbf{A}$ must be diagonalizable—recall (7.2.6). Finally, the k^{th} component of any eigenvector associated with λ_j satisfies $x_k = \alpha r_1^k + \beta r_2^k$ with $\alpha + \beta = 0$, so

$$x_k = \alpha \left(\frac{c}{a}\right)^{k/2}\left(\mathrm{e}^{\mathrm{i}\pi jk/(n+1)} - \mathrm{e}^{-\mathrm{i}\pi jk/(n+1)}\right) = 2\mathrm{i}\alpha\left(\frac{c}{a}\right)^{k/2}\sin\left(\frac{jk\pi}{n+1}\right).$$

Setting $\alpha = 1/2\mathrm{i}$ yields a particular eigenvector associated with λ_j as

$$\mathbf{x}_j = \begin{pmatrix} (c/a)^{1/2}\sin\left(1j\pi/(n+1)\right) \\ (c/a)^{2/2}\sin\left(2j\pi/(n+1)\right) \\ (c/a)^{3/2}\sin\left(3j\pi/(n+1)\right) \\ \vdots \\ (c/a)^{n/2}\sin\left(nj\pi/(n+1)\right) \end{pmatrix}.$$

Because the λ_j 's are distinct, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a complete linearly independent set—recall (7.2.3)—so $\mathbf{P} = \left(\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \cdots \,|\, \mathbf{x}_n\right)$ diagonalizes $\mathbf{A}$.

It's often the case that a right-hand and left-hand eigenvector for some eigenvalue is known. Rather than starting from scratch to find additional eigenpairs, the known information can be used to reduce or "deflate" the problem to a smaller one as described in the following example.

Example 7.2.6

Deflation. Suppose that right-hand and left-hand eigenvectors $\mathbf{x}$ and $\mathbf{y}^*$ for an eigenvalue λ of $\mathbf{A} \in \Re^{n\times n}$ are already known, so $\mathbf{Ax} = \lambda\mathbf{x}$ and $\mathbf{y}^*\mathbf{A} = \lambda\mathbf{y}^*$. Furthermore, suppose $\mathbf{y}^*\mathbf{x} \neq 0$—such eigenvectors are guaranteed to exist if λ is simple or if $\mathbf{A}$ is diagonalizable (Exercises 7.2.23 and 7.2.22).

Problem: Use $\mathbf{x}$ and $\mathbf{y}^*$ to deflate the size of the remaining eigenvalue problem.

Solution: Scale $\mathbf{x}$ and $\mathbf{y}^*$ so that $\mathbf{y}^*\mathbf{x} = 1$, and construct $\mathbf{X}_{n\times n-1}$ so that its columns are an orthonormal basis for $\mathbf{y}^\perp$. An easy way of doing this is to build a reflector $\mathbf{R} = \left[\tilde{\mathbf{y}} \,|\, \mathbf{X}\right]$ having $\tilde{\mathbf{y}} = \mathbf{y}/\left\|\mathbf{y}\right\|_2$ as its first column as described on p. 325. If $\mathbf{P} = \left[\mathbf{x} \,|\, \mathbf{X}\right]$, then straightforward multiplication shows that

$$\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{y}^* \\ \mathbf{X}^*(\mathbf{I} - \mathbf{x}\mathbf{y}^*) \end{pmatrix} \quad \text{and} \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix},$$

where $\mathbf{B} = \mathbf{X}^*\mathbf{A}\mathbf{X}$ is $n-1 \times n-1$. The eigenvalues of $\mathbf{B}$ constitute the remaining eigenvalues of $\mathbf{A}$ (Exercise 7.1.4), and thus an $n \times n$ eigenvalue problem is deflated to become one of size $n-1 \times n-1$.

Note: When $\mathbf{A}$ is symmetric, we can take $\mathbf{x} = \mathbf{y}$ to be an eigenvector with $\left\|\mathbf{x}\right\|_2 = 1$, so $\mathbf{P} = \mathbf{R} = \mathbf{R}^{-1}$, and $\mathbf{RAR} = \begin{pmatrix} \lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}$ in which $\mathbf{B} = \mathbf{B}^T$.

An elegant and more geometrical way of expressing diagonalizability is now presented to help simplify subsequent analyses and pave the way for extensions.

Spectral Theorem for Diagonalizable Matrices

A matrix $\mathbf{A}_{n\times n}$ with spectrum $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ is diagonalizable if and only if there exist matrices $\{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_k\}$ such that

$$\mathbf{A} = \lambda_1\mathbf{G}_1 + \lambda_2\mathbf{G}_2 + \cdots + \lambda_k\mathbf{G}_k, \tag{7.2.7}$$

where the $\mathbf{G}_i$'s have the following properties.

- $\mathbf{G}_i$ is the projector onto $N(\mathbf{A} - \lambda_i\mathbf{I})$ along $R(\mathbf{A} - \lambda_i\mathbf{I})$. (7.2.8)
- $\mathbf{G}_i\mathbf{G}_j = \mathbf{0}$ whenever $i \neq j$. (7.2.9)
- $\mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_k = \mathbf{I}$. (7.2.10)

The expansion (7.2.7) is known as the ***spectral decomposition*** of $\mathbf{A}$, and the $\mathbf{G}_i$'s are called the ***spectral projectors*** associated with $\mathbf{A}$.

Proof. If $\mathbf{A}$ is diagonalizable, and if $\mathbf{X}_i$ is a matrix whose columns form a basis for $N(\mathbf{A} - \lambda_i\mathbf{I})$, then $\mathbf{P} = \big(\mathbf{X}_1 \,|\, \mathbf{X}_2 \,|\, \cdots \,|\, \mathbf{X}_k\big)$ is nonsingular. If $\mathbf{P}^{-1}$ is partitioned in a conformable manner, then we must have

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \Big(\mathbf{X}_1 \,|\, \mathbf{X}_2 \,|\, \cdots \,|\, \mathbf{X}_k\Big)\begin{pmatrix} \lambda_1\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \lambda_2\mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \lambda_k\mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{Y}_1^T \\ \hline \mathbf{Y}_2^T \\ \hline \vdots \\ \hline \mathbf{Y}_k^T \end{pmatrix} \tag{7.2.11}$$

$$= \lambda_1\mathbf{X}_1\mathbf{Y}_1^T + \lambda_2\mathbf{X}_2\mathbf{Y}_2^T + \cdots + \lambda_k\mathbf{X}_k\mathbf{Y}_k^T$$
$$= \lambda_1\mathbf{G}_1 + \lambda_2\mathbf{G}_2 + \cdots + \lambda_k\mathbf{G}_k.$$

For $\mathbf{G}_i = \mathbf{X}_i\mathbf{Y}_i^T$, the statement $\mathbf{P}\mathbf{P}^{-1} = \mathbf{I}$ translates to $\sum_{i=1}^k \mathbf{G}_i = \mathbf{I}$, and

$$\mathbf{P}^{-1}\mathbf{P} = \mathbf{I} \implies \mathbf{Y}_i^T\mathbf{X}_j = \begin{cases} \mathbf{I} & \text{when } i = j, \\ \mathbf{0} & \text{when } i \neq j, \end{cases} \implies \begin{cases} \mathbf{G}_i^2 = \mathbf{G}_i, \\ \mathbf{G}_i\mathbf{G}_j = \mathbf{0} & \text{when } i \neq j. \end{cases}$$

To establish that $R(\mathbf{G}_i) = N(\mathbf{A} - \lambda_i\mathbf{I})$, use $R(\mathbf{AB}) \subseteq R(\mathbf{A})$ (Exercise 4.2.12) and $\mathbf{Y}_i^T\mathbf{X}_i = \mathbf{I}$ to write

$$R(\mathbf{G}_i) = R(\mathbf{X}_i\mathbf{Y}_i^T) \subseteq R(\mathbf{X}_i) = R(\mathbf{X}_i\mathbf{Y}_i^T\mathbf{X}_i) = R(\mathbf{G}_i\mathbf{X}_i) \subseteq R(\mathbf{G}_i).$$

Thus $R(\mathbf{G}_i) = R(\mathbf{X}_i) = N(\mathbf{A} - \lambda_i\mathbf{I})$. To show $N(\mathbf{G}_i) = R(\mathbf{A} - \lambda_i\mathbf{I})$, use $\mathbf{A} = \sum_{j=1}^k \lambda_j\mathbf{G}_j$ with the already established properties of the $\mathbf{G}_i$'s to conclude

$$\mathbf{G}_i(\mathbf{A} - \lambda_i\mathbf{I}) = \mathbf{G}_i\left(\sum_{j=1}^k \lambda_j\mathbf{G}_j - \lambda_i\sum_{j=1}^k \mathbf{G}_j\right) = \mathbf{0} \implies R(\mathbf{A} - \lambda_i\mathbf{I}) \subseteq N(\mathbf{G}_i).$$

But we already know that $N(\mathbf{A} - \lambda_i\mathbf{I}) = R(\mathbf{G}_i)$, so

$$\dim R(\mathbf{A} - \lambda_i\mathbf{I}) = n - \dim N(\mathbf{A} - \lambda_i\mathbf{I}) = n - \dim R(\mathbf{G}_i) = \dim N(\mathbf{G}_i),$$

and therefore, by (4.4.6), $R(\mathbf{A} - \lambda_i\mathbf{I}) = N(\mathbf{G}_i)$. Conversely, if there exist matrices $\mathbf{G}_i$ satisfying (7.2.8)–(7.2.10), then $\mathbf{A}$ must be diagonalizable. To see this, note that (7.2.8) insures $\dim R(\mathbf{G}_i) = \dim N(\mathbf{A} - \lambda_i\mathbf{I}) = \textit{geo mult}_{\mathbf{A}}(\lambda_i)$, while (7.2.9) implies $R(\mathbf{G}_i) \cap R(\mathbf{G}_j) = \mathbf{0}$ and $R\big(\sum_{i=1}^k \mathbf{G}_i\big) = \sum_{i=1}^k R(\mathbf{G}_i)$ (Exercise 5.9.17). Use these with (7.2.10) in the formula for the dimension of a sum (4.4.19) to write

$$\begin{aligned} n = \dim R(\mathbf{I}) &= \dim R(\mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_k) \\ &= \dim [R(\mathbf{G}_1) + R(\mathbf{G}_2) + \cdots + R(\mathbf{G}_k)] \\ &= \dim R(\mathbf{G}_1) + \dim R(\mathbf{G}_2) + \cdots + \dim R(\mathbf{G}_k) \\ &= \textit{geo mult}_{\mathbf{A}}(\lambda_1) + \textit{geo mult}_{\mathbf{A}}(\lambda_2) + \cdots + \textit{geo mult}_{\mathbf{A}}(\lambda_k). \end{aligned}$$

Since $\textit{geo mult}_{\mathbf{A}}(\lambda_i) \le \textit{alg mult}_{\mathbf{A}}(\lambda_i)$ and $\sum_{i=1}^k \textit{alg mult}_{\mathbf{A}}(\lambda_i) = n$, the above equation insures that $\textit{geo mult}_{\mathbf{A}}(\lambda_i) = \textit{alg mult}_{\mathbf{A}}(\lambda_i)$ for each i, and, by (7.2.5), this means $\mathbf{A}$ is diagonalizable. ■

Simple Eigenvalues and Projectors

If $\mathbf{x}$ and $\mathbf{y}^*$ are respective right-hand and left-hand eigenvectors associated with a *simple* eigenvalue $\lambda \in \sigma(\mathbf{A})$, then

$$\mathbf{G} = \mathbf{x}\mathbf{y}^*/\mathbf{y}^*\mathbf{x} \tag{7.2.12}$$

is the projector onto $N(\mathbf{A} - \lambda\mathbf{I})$ along $R(\mathbf{A} - \lambda\mathbf{I})$. In the context of the spectral theorem (p. 517), this means that $\mathbf{G}$ is the spectral projector associated with λ.

Proof. It's not difficult to prove $\mathbf{y}^*\mathbf{x} \neq 0$ (Exercise 7.2.23), and it's clear that $\mathbf{G}$ is a projector because $\mathbf{G}^2 = \mathbf{x}(\mathbf{y}^*\mathbf{x})\mathbf{y}^*/(\mathbf{y}^*\mathbf{x})^2 = \mathbf{G}$. Now determine $R(\mathbf{G})$. The image of any $\mathbf{z}$ is $\mathbf{G}\mathbf{z} = \alpha\mathbf{x}$ with $\alpha = \mathbf{y}^*\mathbf{z}/\mathbf{y}^*\mathbf{x}$, so

$$R(\mathbf{G}) \subseteq \textit{span}\{\mathbf{x}\} = N(\mathbf{A} - \lambda\mathbf{I}) \quad \text{and} \quad \dim R(\mathbf{G}) = 1 = \dim N(\mathbf{A} - \lambda\mathbf{I}).$$

Thus $R(\mathbf{G}) = N(\mathbf{A} - \lambda\mathbf{I})$. To find $N(\mathbf{G})$, recall $N(\mathbf{G}) = R(\mathbf{I} - \mathbf{G})$ (see (5.9.11), p. 386), and observe that $\mathbf{y}^*(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0} \implies \mathbf{y}^*(\mathbf{I} - \mathbf{G}) = \mathbf{0}$, so

$$R(\mathbf{A} - \lambda\mathbf{I})^\perp \subseteq R(\mathbf{I} - \mathbf{G})^\perp = N(\mathbf{G})^\perp \Longrightarrow N(\mathbf{G}) \subseteq R(\mathbf{A} - \lambda\mathbf{I}) \text{ (Exercise 5.11.5).}$$

But $\dim N(\mathbf{G}) = n - \dim R(\mathbf{G}) = n - 1 = n - \dim N(\mathbf{A} - \lambda\mathbf{I}) = \dim R(\mathbf{A} - \lambda\mathbf{I})$, so $N(\mathbf{G}) = R(\mathbf{A} - \lambda\mathbf{I})$. ■

Example 7.2.7

Problem: Determine the spectral projectors for $\mathbf{A} = \begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix}$.

Solution: This is the diagonalizable matrix from Example 7.2.1 (p. 507). Since there are two distinct eigenvalues, $\lambda_1 = 1$ and $\lambda_2 = -3$, there are two spectral projectors,

$$\mathbf{G}_1 = \text{ the projector onto } N(\mathbf{A} - 1\mathbf{I}) \text{ along } R(\mathbf{A} - 1\mathbf{I}),$$
$$\mathbf{G}_2 = \text{ the projector onto } N(\mathbf{A} + 3\mathbf{I}) \text{ along } R(\mathbf{A} + 3\mathbf{I}).$$

There are several different ways to find these projectors.

1. Compute bases for the necessary nullspaces and ranges, and use (5.9.12).
2. Compute $\mathbf{G}_i = \mathbf{X}_i\mathbf{Y}_i^T$ as described in (7.2.11). The required computations are essentially the same as those needed above. Since much of the work has already been done in Example 7.2.1, let's complete the arithmetic. We have

$$\mathbf{P} = \left(\begin{array}{c|cc} 1 & 1 & 1 \\ 2 & 1 & 0 \\ -2 & 0 & 1 \end{array}\right) = \left(\mathbf{X}_1 \,|\, \mathbf{X}_2\right), \quad \mathbf{P}^{-1} = \left(\begin{array}{ccc} 1 & -1 & -1 \\ \hline -2 & 3 & 2 \\ 2 & -2 & -1 \end{array}\right) = \left(\begin{array}{c} \mathbf{Y}_1^T \\ \hline \mathbf{Y}_2^T \end{array}\right),$$

so

$$\mathbf{G}_1 = \mathbf{X}_1\mathbf{Y}_1^T = \begin{pmatrix} 1 & -1 & -1 \\ 2 & -2 & -2 \\ -2 & 2 & 2 \end{pmatrix}, \quad \mathbf{G}_2 = \mathbf{X}_2\mathbf{Y}_2^T = \begin{pmatrix} 0 & 1 & 1 \\ -2 & 3 & 2 \\ 2 & -2 & -1 \end{pmatrix}.$$

Check that these are correct by confirming the validity of (7.2.7)–(7.2.10).

3. Since $\lambda_1 = 1$ is a simple eigenvalue, (7.2.12) may be used to compute $\mathbf{G}_1$ from any pair of associated right-hand and left-hand eigenvectors $\mathbf{x}$ and $\mathbf{y}^T$. Of course, $\mathbf{P}$ and $\mathbf{P}^{-1}$ are not needed to determine such a pair, but since $\mathbf{P}$ and $\mathbf{P}^{-1}$ have been computed above, we can use $\mathbf{X}_1$ and $\mathbf{Y}_1^T$ to make the point that *any* right-hand and left-hand eigenvectors associated with $\lambda_1 = 1$ will do the job because they are all of the form $\mathbf{x} = \alpha\mathbf{X}_1$ and $\mathbf{y}^T = \beta\mathbf{Y}_1^T$ for $\alpha \neq 0 \neq \beta$. Consequently,

$$\mathbf{G}_1 = \frac{\mathbf{x}\mathbf{y}^T}{\mathbf{y}^T\mathbf{x}} = \frac{\alpha\begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}\beta\,(1 \quad -1 \quad -1)}{\alpha\beta} = \begin{pmatrix} 1 & -1 & -1 \\ 2 & -2 & -2 \\ -2 & 2 & 2 \end{pmatrix}.$$

Invoking (7.2.10) yields the other spectral projector as $\mathbf{G}_2 = \mathbf{I} - \mathbf{G}_1$.

4. An even easier solution is obtained from the spectral theorem by writing

$$\mathbf{A} - \mathbf{I} = (1\mathbf{G}_1 - 3\mathbf{G}_2) - (\mathbf{G}_1 + \mathbf{G}_2) = -4\mathbf{G}_2,$$
$$\mathbf{A} + 3\mathbf{I} = (1\mathbf{G}_1 - 3\mathbf{G}_2) + 3\,(\mathbf{G}_1 + \mathbf{G}_2) = 4\mathbf{G}_1,$$

so that

$$\mathbf{G}_1 = \frac{(\mathbf{A} + 3\mathbf{I})}{4} \quad \text{and} \quad \mathbf{G}_2 = \frac{-(\mathbf{A} - \mathbf{I})}{4}.$$

Can you see how to make this rather ad hoc technique work in more general situations?

5. In fact, the technique above is really a special case of a completely general formula giving each $\mathbf{G}_i$ as a function $\mathbf{A}$ and λ_i as

$$\mathbf{G}_i = \frac{\prod_{\substack{j=1 \\ j \neq i}}^{k} (\mathbf{A} - \lambda_j \mathbf{I})}{\prod_{\substack{j=1 \\ j \neq i}}^{k} (\lambda_i - \lambda_j)}.$$

This "interpolation formula" is developed on p. 529.

Below is a summary of the facts concerning diagonalizability.

Summary of Diagonalizability

For an $n \times n$ matrix $\mathbf{A}$ with spectrum $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$, the following statements are equivalent.

- $\mathbf{A}$ is similar to a diagonal matrix—i.e., $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$.
- $\mathbf{A}$ has a complete linearly independent set of eigenvectors.
- Every λ_i is semisimple—i.e., $geo\ mult_{\mathbf{A}}(\lambda_i) = alg\ mult_{\mathbf{A}}(\lambda_i)$.
- $\mathbf{A} = \lambda_1\mathbf{G}_1 + \lambda_2\mathbf{G}_2 + \cdots + \lambda_k\mathbf{G}_k$, where
  - ▷ $\mathbf{G}_i$ is the projector onto $N(\mathbf{A} - \lambda_i\mathbf{I})$ along $R(\mathbf{A} - \lambda_i\mathbf{I})$,
  - ▷ $\mathbf{G}_i\mathbf{G}_j = \mathbf{0}$ whenever $i \neq j$,
  - ▷ $\mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_k = \mathbf{I}$,
  - ▷ $\mathbf{G}_i = \prod_{\substack{j=1 \\ j \neq i}}^{k} (\mathbf{A} - \lambda_j\mathbf{I}) \Big/ \prod_{\substack{j=1 \\ j \neq i}}^{k} (\lambda_i - \lambda_j)$ (see (7.3.11) on p. 529).
  - ▷ If λ_i is a simple eigenvalue associated with right-hand and left-hand eigenvectors $\mathbf{x}$ and $\mathbf{y}^*$, respectively, then $\mathbf{G}_i = \mathbf{x}\mathbf{y}^*/\mathbf{y}^*\mathbf{x}$.

Exercises for section 7.2

7.2.1. Diagonalize $\mathbf{A} = \begin{pmatrix} -8 & -6 \\ 12 & 10 \end{pmatrix}$ with a similarity transformation, or else explain why $\mathbf{A}$ can't be diagonalized.

7.2.2. (a) Verify that $alg\ mult_{\mathbf{A}}(\lambda) = geo\ mult_{\mathbf{A}}(\lambda)$ for each eigenvalue of

$$\mathbf{A} = \begin{pmatrix} -4 & -3 & -3 \\ 0 & -1 & 0 \\ 6 & 6 & 5 \end{pmatrix}.$$

(b) Find a nonsingular $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{AP}$ is a diagonal matrix.

7.2.3. Show that similar matrices need not have the same eigenvectors by giving an example of two matrices that are similar but have different eigenspaces.

7.2.4. $\lambda = 2$ is an eigenvalue for $\mathbf{A} = \begin{pmatrix} 3 & 2 & 1 \\ 0 & 2 & 0 \\ -2 & -3 & 0 \end{pmatrix}$. Find $alg\ mult_{\mathbf{A}}(\lambda)$ as well as $geo\ mult_{\mathbf{A}}(\lambda)$. Can you conclude anything about the diagonalizability of $\mathbf{A}$ from these results?

7.2.5. If $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$, explain why $\mathbf{B}^k = \mathbf{P}^{-1}\mathbf{A}^k\mathbf{P}$.

7.2.6. Compute $\lim_{n\to\infty} \mathbf{A}^n$ for $\mathbf{A} = \begin{pmatrix} 7/5 & 1/5 \\ -1 & 1/2 \end{pmatrix}$.

7.2.7. Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t\}$ be a set of linearly independent eigenvectors for $\mathbf{A}_{n\times n}$ associated with respective eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_t\}$, and let $\mathbf{X}$ be any $n \times (n-t)$ matrix such that $\mathbf{P}_{n\times n} = \big(\mathbf{x}_1 \,|\, \cdots \,|\, \mathbf{x}_t \,|\, \mathbf{X}\big)$ is nonsingular. Prove that if $\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{y}_1^* \\ \vdots \\ \mathbf{y}_t^* \\ \mathbf{Y}^* \end{pmatrix}$, where the $\mathbf{y}_i^*$'s are rows and $\mathbf{Y}^*$ is $(n-t)\times n$, then $\{\mathbf{y}_1^*, \mathbf{y}_2^*, \ldots, \mathbf{y}_t^*\}$ is a set of linearly independent left-hand eigenvectors associated with $\{\lambda_1, \lambda_2, \ldots, \lambda_t\}$, respectively (i.e., $\mathbf{y}_i^*\mathbf{A} = \lambda_i\mathbf{y}_i^*$).

7.2.8. Let $\mathbf{A}$ be a diagonalizable matrix, and let $\rho(\star)$ denote the spectral radius (recall Example 7.1.4 on p. 497). Prove that $\lim_{k\to\infty} \mathbf{A}^k = \mathbf{0}$ if and only if $\rho(\mathbf{A}) < 1$. **Note:** It is demonstrated on p. 617 that this result holds for nondiagonalizable matrices as well.

7.2.9. Apply the technique used to prove Schur's triangularization theorem (p. 508) to construct an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}^T\mathbf{AP}$ is upper triangular for $\mathbf{A} = \begin{pmatrix} 13 & -9 \\ 16 & -11 \end{pmatrix}$.

7.2.10. Verify the Cayley–Hamilton theorem for $\mathbf{A} = \begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix}$.
Hint: This is the matrix from Example 7.2.1 on p. 507.

7.2.11. Since each row sum in the following symmetric matrix $\mathbf{A}$ is 4, it's clear that $\mathbf{x} = (1,1,1,1)^T$ is both a right-hand and left-hand eigenvector associated with $\lambda = 4 \in \sigma(\mathbf{A})$. Use the deflation technique of Example 7.2.6 (p. 516) to determine the remaining eigenvalues of

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

7.2.12. Explain why $\mathbf{AG}_i = \mathbf{G}_i\mathbf{A} = \lambda_i\mathbf{G}_i$ for the spectral projector $\mathbf{G}_i$ associated with the eigenvalue λ_i of a diagonalizable matrix $\mathbf{A}$.

7.2.13. Prove that $\mathbf{A} = \mathbf{c}_{n\times 1}\mathbf{d}^T_{1\times n}$ is diagonalizable if and only if $\mathbf{d}^T\mathbf{c} \neq 0$.

7.2.14. Prove that $\mathbf{A} = \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix}$ is diagonalizable if and only if $\mathbf{W}_{s\times s}$ and $\mathbf{Z}_{t\times t}$ are each diagonalizable.

7.2.15. Prove that if $\mathbf{AB} = \mathbf{BA}$, then $\mathbf{A}$ and $\mathbf{B}$ can be *simultaneously* triangularized by a unitary similarity transformation—i.e., $\mathbf{U}^*\mathbf{AU} = \mathbf{T}_1$ and $\mathbf{U}^*\mathbf{BU} = \mathbf{T}_2$ for some unitary matrix $\mathbf{U}$. **Hint:** Recall Exercise 7.1.20 (p. 503) along with the development of Schur's triangularization theorem (p. 508).

7.2.16. For diagonalizable matrices, prove that $\mathbf{AB} = \mathbf{BA}$ if and only if $\mathbf{A}$ and $\mathbf{B}$ can be *simultaneously* diagonalized—i.e., $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}_1$ and $\mathbf{P}^{-1}\mathbf{BP} = \mathbf{D}_2$ for some $\mathbf{P}$. **Hint:** If $\mathbf{A}$ and $\mathbf{B}$ commute, then so do $\mathbf{P}^{-1}\mathbf{AP} = \begin{pmatrix} \lambda_1\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}$ and $\mathbf{P}^{-1}\mathbf{BP} = \begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix}$.

7.2.17. Explain why the following "proof" of the Cayley–Hamilton theorem is not valid. $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) \implies p(\mathbf{A}) = \det(\mathbf{A} - \mathbf{AI}) = \det(\mathbf{0}) = 0$.

7.2.18. Show that the eigenvalues of the finite difference matrix (p. 19)

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}_{n\times n} \quad \text{are } \lambda_j = 4\sin^2\frac{j\pi}{2(n+1)},\ 1 \le j \le n.$$

7.2.19. Let $\mathbf{N} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}_{n \times n}$.

(a) Show that $\lambda \in \sigma\left(\mathbf{N}+\mathbf{N}^T\right)$ if and only if $\mathrm{i}\lambda \in \sigma\left(\mathbf{N}-\mathbf{N}^T\right)$.
(b) Explain why $\mathbf{N}+\mathbf{N}^T$ is nonsingular if and only if n is even.
(c) Evaluate $\det\left(\mathbf{N}-\mathbf{N}^T\right)/\det\left(\mathbf{N}+\mathbf{N}^T\right)$ when n is even.

7.2.20. A Toeplitz matrix having the form

$$\mathbf{C} = \begin{pmatrix} c_0 & c_{n-1} & c_{n-2} & \cdots & c_1 \\ c_1 & c_0 & c_{n-1} & \cdots & c_2 \\ c_2 & c_1 & c_0 & \cdots & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \cdots & c_0 \end{pmatrix}_{n \times n}$$

is called a ***circulant matrix***. If $p(x) = c_0 + c_1 x + \cdots + c_{n-1}x^{n-1}$, and if $\{1, \xi, \xi^2, \ldots, \xi^{n-1}\}$ are the n^{th} roots of unity, then the results of Exercise 5.8.12 (p. 379) insure that

$$\mathbf{F}_n \mathbf{C} \mathbf{F}_n^{-1} = \begin{pmatrix} p(1) & 0 & \cdots & 0 \\ 0 & p(\xi) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(\xi^{n-1}) \end{pmatrix}$$

in which $\mathbf{F}_n$ is the Fourier matrix of order n. Verify these facts for the circulant below by computing its eigenvalues and eigenvectors directly:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

7.2.21. Suppose that $(\lambda, \mathbf{x})$ and $(\mu, \mathbf{y}^*)$ are right-hand and left-hand eigenpairs for $\mathbf{A} \in \Re^{n \times n}$—i.e., $\mathbf{Ax} = \lambda\mathbf{x}$ and $\mathbf{y}^*\mathbf{A} = \mu\mathbf{y}^*$. Explain why $\mathbf{y}^*\mathbf{x} = 0$ whenever $\lambda \neq \mu$.

7.2.22. Consider $\mathbf{A} \in \Re^{n \times n}$.

(a) Show that if $\mathbf{A}$ is diagonalizable, then there are right-hand and left-hand eigenvectors $\mathbf{x}$ and $\mathbf{y}^*$ associated with $\lambda \in \sigma(\mathbf{A})$ such that $\mathbf{y}^*\mathbf{x} \neq 0$ so that we can make $\mathbf{y}^*\mathbf{x} = 1$.
(b) Show that not every right-hand and left-hand eigenvector $\mathbf{x}$ and $\mathbf{y}^*$ associated with $\lambda \in \sigma(\mathbf{A})$ must satisfy $\mathbf{y}^*\mathbf{x} \neq 0$.
(c) Show that (a) need not be true when $\mathbf{A}$ is not diagonalizable.

7.2.23. Consider $\mathbf{A} \in \Re^{n\times n}$ with $\lambda \in \sigma(\mathbf{A})$.

(a) Prove that if λ is simple, then $\mathbf{y}^*\mathbf{x} \neq 0$ for every pair of respective right-hand and left-hand eigenvectors $\mathbf{x}$ and $\mathbf{y}^*$ associated with λ regardless of whether or not $\mathbf{A}$ is diagonalizable. **Hint:** Use the core-nilpotent decomposition on p. 397.

(b) Show that $\mathbf{y}^*\mathbf{x} = 0$ is possible when λ is not simple.

7.2.24. For $\mathbf{A} \in \Re^{n\times n}$ with $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$, show $\mathbf{A}$ is diagonalizable if and only if $\Re^n = N(\mathbf{A} - \lambda_1\mathbf{I}) \oplus N(\mathbf{A} - \lambda_2\mathbf{I}) \oplus \cdots \oplus N(\mathbf{A} - \lambda_k\mathbf{I})$. **Hint:** Recall Exercise 5.9.14.

7.2.25. The Real Schur Form. Schur's triangularization theorem (p. 508) insures that every square matrix $\mathbf{A}$ is unitarily similar to an upper-triangular matrix—say, $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{T}$. But even when $\mathbf{A}$ is real, $\mathbf{U}$ and $\mathbf{T}$ may have to be complex if $\mathbf{A}$ has some complex eigenvalues. However, the matrices (and the arithmetic) can be constrained to be real by settling for a block-triangular result with 2×2 or scalar entries on the diagonal. Prove that for each $\mathbf{A} \in \Re^{n\times n}$ there exists an orthogonal matrix $\mathbf{P} \in \Re^{n\times n}$ and real matrices $\mathbf{B}_{ij}$ such that

$$\mathbf{P}^T\mathbf{A}\mathbf{P} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1k} \\ \mathbf{0} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_{kk} \end{pmatrix}, \quad \text{where } \mathbf{B}_{jj} \text{ is } 1\times 1 \text{ or } 2\times 2.$$

If $\mathbf{B}_{jj} = [\lambda_j]$ is 1×1, then $\lambda_j \in \sigma(\mathbf{A})$, and if $\mathbf{B}_{jj}$ is 2×2, then $\sigma(\mathbf{B}_{jj}) = \{\lambda_j, \overline{\lambda}_j\} \subseteq \sigma(\mathbf{A})$.

7.2.26. When $\mathbf{A} \in \Re^{n\times n}$ is diagonalizable by a similarity transformation $\mathbf{S}$, then $\mathbf{S}$ may have to be complex if $\mathbf{A}$ has some complex eigenvalues. Analogous to Exercise 7.2.25, we can stay in the realm of real numbers by settling for a block-diagonal result with 1×1 or 2×2 entries on the diagonal. Prove that if $\mathbf{A} \in \Re^{n\times n}$ is diagonalizable with real eigenvalues $\{\rho_1, \ldots, \rho_r\}$ and complex eigenvalues $\{\lambda_1, \overline{\lambda}_1, \lambda_2, \overline{\lambda}_2, \ldots, \lambda_t, \overline{\lambda}_t\}$ with $2t + r = n$, then there exists a nonsingular $\mathbf{P} \in \Re^{n\times n}$ and $\mathbf{B}_j$'s $\in \Re^{2\times 2}$ such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_t \end{pmatrix}, \quad \text{where} \quad \mathbf{D} = \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_r \end{pmatrix},$$

and where $\mathbf{B}_j$ has eigenvalues λ_j and $\overline{\lambda}_j$.

7.2.27. For $\mathbf{A} \in \mathcal{C}^{n\times n}$, prove that $\mathbf{x}^*\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{x} \in \mathcal{C}^{n\times 1} \Rightarrow \mathbf{A} = \mathbf{0}$. Show that $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{x} \in \Re^{n\times 1} \not\Rightarrow \mathbf{A} = \mathbf{0}$, even if $\mathbf{A}$ is real.

Solutions for exercises in section 7. 2

7.2.1. The characteristic equation is $\lambda^2-2\lambda-8=(\lambda+2)(\lambda-4)=0$, so the eigenvalues are $\lambda_1=-2$ and $\lambda_2=4$. Since no eigenvalue is repeated, (7.2.6) insures $\mathbf{A}$ must be diagonalizable. A similarity transformation $\mathbf{P}$ that diagonalizes $\mathbf{A}$ is constructed from a complete set of independent eigenvectors. Compute a pair of eigenvectors associated with λ_1 and λ_2 to be

$$\mathbf{x}_1=\begin{pmatrix}-1\\1\end{pmatrix},\ \mathbf{x}_2=\begin{pmatrix}-1\\2\end{pmatrix},\quad \text{and set}\quad \mathbf{P}=\begin{pmatrix}-1&-1\\1&2\end{pmatrix}.$$

Now verify that

$$\mathbf{P}^{-1}\mathbf{AP}=\begin{pmatrix}-2&-1\\1&1\end{pmatrix}\begin{pmatrix}-8&-6\\12&10\end{pmatrix}\begin{pmatrix}-1&-1\\1&2\end{pmatrix}=\begin{pmatrix}-2&0\\0&4\end{pmatrix}=\mathbf{D}.$$

7.2.2. (a) The characteristic equation is $\lambda^3-3\lambda-2=(\lambda-2)(\lambda+1)^2=0$, so the eigenvalues are $\lambda=2$ and $\lambda=-1$. By reducing $\mathbf{A}-2\mathbf{I}$ and $\mathbf{A}+\mathbf{I}$ to echelon form, compute bases for $N(\mathbf{A}-2\mathbf{I})$ and $N(\mathbf{A}+\mathbf{I})$. One set of bases is

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

Therefore,

$$\begin{aligned} geo\ mult_{\mathbf{A}}(2)&=\dim N(\mathbf{A}-2\mathbf{I})=1=alg\ mult_{\mathbf{A}}(2),\\ geo\ mult_{\mathbf{A}}(-1)&=\dim N(\mathbf{A}+\mathbf{I})\ =2=alg\ mult_{\mathbf{A}}(-1).\end{aligned}$$

In other words, $\lambda=2$ is a simple eigenvalue, and $\lambda=-1$ is a semisimple eigenvalue.

(b) A similarity transformation $\mathbf{P}$ that diagonalizes $\mathbf{A}$ is constructed from a complete set of independent eigenvectors, and these are obtained from the above bases. Set $\mathbf{P}=\begin{pmatrix}-1&-1&-1\\0&1&0\\2&0&1\end{pmatrix}$, and compute $\mathbf{P}^{-1}=\begin{pmatrix}1&1&1\\0&1&0\\-2&-2&-1\end{pmatrix}$ and verify that $\mathbf{P}^{-1}\mathbf{AP}=\begin{pmatrix}2&0&0\\0&-1&0\\0&0&-1\end{pmatrix}$.

7.2.3. Consider the matrix $\mathbf{A}$ of Exercise 7.2.1. We know from its solution that $\mathbf{A}$ is similar to $\mathbf{D}=\begin{pmatrix}-2&0\\0&4\end{pmatrix}$, but the two eigenspaces for $\mathbf{A}$ are spanned by $\begin{pmatrix}-1\\1\end{pmatrix}$ and $\begin{pmatrix}-1\\2\end{pmatrix}$, whereas the eigenspaces for $\mathbf{D}$ are spanned by the unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$.

7.2.4. The characteristic equation of $\mathbf{A}$ is $p(\lambda) = (\lambda-1)(\lambda-2)^2$, so $alg\ mult_{\mathbf{A}}(2) = 2$. To find $geo\ mult_{\mathbf{A}}(2)$, reduce $\mathbf{A} - 2\mathbf{I}$ to echelon form to find that

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

so $geo\ mult_{\mathbf{A}}(2) = \dim N(\mathbf{A} - 2\mathbf{I}) = 1$. Since there exists at least one eigenvalue such that $geo\ mult_{\mathbf{A}}(\lambda) < alg\ mult_{\mathbf{A}}(\lambda)$, it follows (7.2.5) on p. 512 that $\mathbf{A}$ cannot be diagonalized by a similarity transformation.

7.2.5. A formal induction argument can be given, but it suffices to "do it with dots" by writing

$$\begin{aligned}\mathbf{B}^k &= (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})(\mathbf{P}^{-1}\mathbf{A}\mathbf{P})\cdots(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) \\ &= \mathbf{P}^{-1}\mathbf{A}(\mathbf{P}\mathbf{P}^{-1})\mathbf{A}(\mathbf{P}\mathbf{P}^{-1})\cdots(\mathbf{P}\mathbf{P}^{-1})\mathbf{A}\mathbf{P} = \mathbf{P}^{-1}\mathbf{A}\mathbf{A}\cdots\mathbf{A}\mathbf{P} = \mathbf{P}^{-1}\mathbf{A}^k\mathbf{P}.\end{aligned}$$

7.2.6. $\lim_{n\to\infty} \mathbf{A}^n = \begin{pmatrix} 5 & 2 \\ -10 & -4 \end{pmatrix}$. Of course, you could compute $\mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \ldots$ in hopes of seeing a pattern, but this clumsy approach is not definitive. A better technique is to diagonalize $\mathbf{A}$ with a similarity transformation, and then use the result of Exercise 7.2.5. The characteristic equation is $0 = \lambda^2 - (19/10)\lambda + (1/2) = (\lambda-1)(\lambda-(9/10))$, so the eigenvalues are $\lambda = 1$ and $\lambda = .9$. By reducing $\mathbf{A}-\mathbf{I}$ and $\mathbf{A} - .9\mathbf{I}$ to echelon form, we see that

$$N(\mathbf{A} - \mathbf{I}) = span\left\{\begin{pmatrix} -1 \\ 2 \end{pmatrix}\right\} \quad \text{and} \quad N(\mathbf{A} - .9\mathbf{I}) = span\left\{\begin{pmatrix} -2 \\ 5 \end{pmatrix}\right\},$$

so $\mathbf{A}$ is indeed diagonalizable, and $\mathbf{P} = \begin{pmatrix} -1 & -2 \\ 2 & 5 \end{pmatrix}$ is a matrix such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & .9 \end{pmatrix} = \mathbf{D}$ or, equivalently, $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. The result of Exercise 7.2.5 says that $\mathbf{A}^n = \mathbf{P}\mathbf{D}^n\mathbf{P}^{-1} = \mathbf{P}\begin{pmatrix} 1 & 0 \\ 0 & .9^n \end{pmatrix}\mathbf{P}^{-1}$, so

$$\lim_{n\to\infty} \mathbf{A}^n = \mathbf{P}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\mathbf{P}^{-1} = \begin{pmatrix} -1 & -2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -5 & -2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ -10 & -4 \end{pmatrix}.$$

7.2.7. It follows from $\mathbf{P}^{-1}\mathbf{P} = \mathbf{I}$ that $\mathbf{y}_i^*\mathbf{x}_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$ as well as $\mathbf{y}_i^*\mathbf{X} = \mathbf{0}$ and $\mathbf{Y}^*\mathbf{x}_i = \mathbf{0}$ for each $i = 1, \ldots, t$, so

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \mathbf{y}_1^* \\ \vdots \\ \mathbf{y}_t^* \\ \mathbf{Y}^* \end{pmatrix}\mathbf{A}\Big(\mathbf{x}_1 \,|\, \cdots \,|\, \mathbf{x}_t \,|\, \mathbf{X}\Big) = \begin{pmatrix} \lambda_1 & \cdots & 0 & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \lambda_t & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Y}^*\mathbf{A}\mathbf{X} \end{pmatrix} = \mathbf{B}.$$

Therefore, examining the first t rows on both sides of $\mathbf{P}^{-1}\mathbf{A} = \mathbf{B}\mathbf{P}^{-1}$ yields $\mathbf{y}_i^*\mathbf{A} = \lambda_i \mathbf{y}_i^*$ for $i = 1, \ldots, t$.

7.2.8. If $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, then $\mathbf{P}^{-1}\mathbf{A}^k\mathbf{P} = \operatorname{diag}\left(\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\right)$ for $k = 0, 1, 2, \ldots$ or, equivalently, $\mathbf{A}^k = \mathbf{P}\operatorname{diag}\left(\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\right)\mathbf{P}^{-1}$. Therefore, $\mathbf{A}^k \to \mathbf{0}$ if and only if each $\lambda_i^k \to 0$, which is equivalent to saying that $|\lambda_i| < 1$ for each i. Since $\rho(\mathbf{A}) = \max_{\lambda_i \in \sigma(\mathbf{A})} |\lambda_i|$ (recall Example 7.1.4 on p. 497), it follows that $\mathbf{A}^k \to \mathbf{0}$ if and only if $\rho(\mathbf{A}) < 1$.

7.2.9. The characteristic equation for $\mathbf{A}$ is $\lambda^2 - 2\lambda + 1$, so $\lambda = 1$ is the only distinct eigenvalue. By reducing $\mathbf{A} - \mathbf{I}$ to echelon form, we see that $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ is a basis for $N(\mathbf{A} - \mathbf{I})$, so $\mathbf{x} = (1/5)\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ is an eigenvector of unit length. Following the procedure on p. 325, we find that $\mathbf{R} = \begin{pmatrix} 3/5 & 4/5 \\ 4/5 & -3/5 \end{pmatrix}$ is an elementary reflector having $\mathbf{x}$ as its first column, and $\mathbf{R}^T\mathbf{A}\mathbf{R} = \mathbf{R}\mathbf{A}\mathbf{R} = \begin{pmatrix} 1 & 25 \\ 0 & 1 \end{pmatrix}$.

7.2.10. From Example 7.2.1 on p. 507 we see that the characteristic equation for $\mathbf{A}$ is $p(\lambda) = \lambda^3 + 5\lambda^2 + 3\lambda - 9 = (\lambda - 1)(\lambda + 3)^2 = 0$. Straightforward computation shows that

$$p(\mathbf{A}) = (\mathbf{A} - \mathbf{I})(\mathbf{A} + 3\mathbf{I})^2 = \begin{pmatrix} 0 & -4 & -4 \\ 8 & -12 & -8 \\ -8 & 8 & 4 \end{pmatrix}\begin{pmatrix} 16 & -16 & -16 \\ 32 & -32 & -32 \\ -32 & 32 & 32 \end{pmatrix} = \mathbf{0}.$$

7.2.11. Rescale the observed eigenvector as $\mathbf{x} = (1/2)(1, 1, 1, 1)^T = \mathbf{y}$ so that $\mathbf{x}^T\mathbf{x} = 1$. Follow the procedure described in Example 5.6.3 (p. 325), and set $\mathbf{u} = \mathbf{x} - \mathbf{e}_1$ to construct

$$\mathbf{R} = \mathbf{I} - \frac{2\mathbf{u}\mathbf{u}^T}{\mathbf{u}^T\mathbf{u}} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} = \mathbf{P} = \left[\mathbf{x} \,|\, \mathbf{X}\right] \quad (\text{since } \mathbf{x} = \mathbf{y}).$$

Consequently, $\mathbf{B} = \mathbf{X}^T\mathbf{A}\mathbf{X} = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}$, and $\sigma(\mathbf{B}) = \{2, \sqrt{2}, -\sqrt{2}\}$.

7.2.12. Use the spectral theorem with properties $\mathbf{G}_i\mathbf{G}_j = \mathbf{0}$ for $i \neq j$ and $\mathbf{G}_i^2 = \mathbf{G}_i$ to write $\mathbf{A}\mathbf{G}_i = (\lambda_1\mathbf{G}_1 + \lambda_2\mathbf{G}_2 + \cdots + \lambda_k\mathbf{G}_k)\mathbf{G}_i = \lambda_i\mathbf{G}_i^2 = \lambda_i\mathbf{G}_i$. A similar argument shows $\mathbf{G}_i\mathbf{A} = \lambda_i\mathbf{G}_i$.

7.2.13. Use (6.2.3) to show that $\lambda^{n-1}(\lambda - \mathbf{d}^T\mathbf{c}) = 0$ is the characteristic equation for $\mathbf{A}$. Thus $\lambda = 0$ and $\lambda = \mathbf{d}^T\mathbf{c}$ are the eigenvalues of $\mathbf{A}$. We know from (7.2.5) that $\mathbf{A}$ is diagonalizable if and only if the algebraic and geometric multiplicities agree for each eigenvalue. Since $\mathit{geo\ mult}_{\mathbf{A}}(0) = \dim N(\mathbf{A}) = n - \mathit{rank}(\mathbf{A}) = n - 1$, and since

$$\mathit{alg\ mult}_{\mathbf{A}}(0) = \begin{cases} n - 1 & \text{if } \ \mathbf{d}^T\mathbf{c} \neq 0, \\ n & \text{if } \ \mathbf{d}^T\mathbf{c} = 0, \end{cases}$$

it follows that $\mathbf{A}$ is diagonalizable if and only if $\mathbf{d}^T\mathbf{c} \neq 0$.

7.2.14. If $\mathbf{W}$ and $\mathbf{Z}$ are diagonalizable—say $\mathbf{P}^{-1}\mathbf{W}\mathbf{P}$ and $\mathbf{Q}^{-1}\mathbf{Z}\mathbf{Q}$ are diagonal—then $\begin{pmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}$ diagonalizes $\mathbf{A}$. Use an indirect argument for the converse. Suppose $\mathbf{A}$ is diagonalizable but $\mathbf{W}$ (or $\mathbf{Z}$) is not. Then there is an eigenvalue $\lambda \in \sigma(\mathbf{W})$ with $geo\ mult_{\mathbf{W}}(\lambda) < alg\ mult_{\mathbf{W}}(\lambda)$. Since $\sigma(\mathbf{A}) = \sigma(\mathbf{W}) \cup \sigma(\mathbf{Z})$ (Exercise 7.1.4), this would mean that

$$\begin{aligned} geo\ mult_{\mathbf{A}}(\lambda) &= \dim N(\mathbf{A}-\lambda\mathbf{I}) = (s+t) - rank(\mathbf{A}-\lambda\mathbf{I}) \\ &= (s - rank(\mathbf{W}-\lambda\mathbf{I})) + (t - rank(\mathbf{Z}-\lambda\mathbf{I})) \\ &= \dim N(\mathbf{W}-\lambda\mathbf{I}) + \dim N(\mathbf{Z}-\lambda\mathbf{I}) \\ &= geo\ mult_{\mathbf{W}}(\lambda) + geo\ mult_{\mathbf{Z}}(\lambda) \\ &< alg\ mult_{\mathbf{W}}(\lambda) + alg\ mult_{\mathbf{Z}}(\lambda) \\ &< alg\ mult_{\mathbf{A}}(\lambda), \end{aligned}$$

which contradicts the fact that $\mathbf{A}$ is diagonalizable.

7.2.15. If $\mathbf{AB} = \mathbf{BA}$, then, by Exercise 7.1.20 (p. 503), $\mathbf{A}$ and $\mathbf{B}$ have a common eigenvector—say $\mathbf{Ax} = \lambda\mathbf{x}$ and $\mathbf{Bx} = \mu\mathbf{x}$, where $\mathbf{x}$ has been scaled so that $\|\mathbf{x}\|_2 = 1$. If $\mathbf{R} = [\mathbf{x} \mid \mathbf{X}]$ is a unitary matrix having $\mathbf{x}$ as its first column (Example 5.6.3, p. 325), then

$$\mathbf{R}^*\mathbf{A}\mathbf{R} = \begin{pmatrix} \lambda & \mathbf{x}^*\mathbf{A}\mathbf{X} \\ \mathbf{0} & \mathbf{X}^*\mathbf{A}\mathbf{X} \end{pmatrix} \quad \text{and} \quad \mathbf{R}^*\mathbf{B}\mathbf{R} = \begin{pmatrix} \mu & \mathbf{x}^*\mathbf{B}\mathbf{X} \\ \mathbf{0} & \mathbf{X}^*\mathbf{B}\mathbf{X} \end{pmatrix}.$$

Since $\mathbf{A}$ and $\mathbf{B}$ commute, so do $\mathbf{R}^*\mathbf{A}\mathbf{R}$ and $\mathbf{R}^*\mathbf{B}\mathbf{R}$, which in turn implies $\mathbf{A}_2 = \mathbf{X}^*\mathbf{A}\mathbf{X}$ and $\mathbf{B}_2 = \mathbf{X}^*\mathbf{B}\mathbf{X}$ commute. Thus the problem is deflated, so the same argument can be applied inductively in a manner similar to the development of Schur's triangularization theorem (p. 508).

7.2.16. If $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}_1$ and $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \mathbf{D}_2$ are both diagonal, then $\mathbf{D}_1\mathbf{D}_2 = \mathbf{D}_2\mathbf{D}_1$ implies that $\mathbf{AB} = \mathbf{BA}$. Conversely, suppose $\mathbf{AB} = \mathbf{BA}$. Let $\lambda \in \sigma(\mathbf{A})$ with $alg\ mult_{\mathbf{A}}(\lambda) = a$, and let $\mathbf{P}$ be such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \lambda\mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}$, where $\mathbf{D}$ is a diagonal matrix with $\lambda \notin \sigma(\mathbf{D})$. Since $\mathbf{A}$ and $\mathbf{B}$ commute, so do $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ and $\mathbf{P}^{-1}\mathbf{B}\mathbf{P}$. Consequently, if $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix}$, then

$$\begin{pmatrix} \lambda\mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}\begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix} = \begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix}\begin{pmatrix} \lambda\mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \implies \begin{cases} \lambda\mathbf{X} = \mathbf{X}\mathbf{D}, \\ \mathbf{D}\mathbf{Y} = \lambda\mathbf{Y}, \end{cases}$$

so $(\mathbf{D}-\lambda\mathbf{I})\mathbf{X} = \mathbf{0}$ and $(\mathbf{D}-\lambda\mathbf{I})\mathbf{Y} = \mathbf{0}$. But $(\mathbf{D}-\lambda\mathbf{I})$ is nonsingular, so $\mathbf{X} = \mathbf{0}$ and $\mathbf{Y} = \mathbf{0}$, and thus $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix}$. Since $\mathbf{B}$ is diagonalizable, so is

$\mathbf{P}^{-1}\mathbf{BP}$, and hence so are $\mathbf{W}$ and $\mathbf{Z}$ (Exercise 7.2.14). If $\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_w & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_z \end{pmatrix}$, where $\mathbf{Q}_w$ and $\mathbf{Q}_z$ are such that $\mathbf{Q}_w^{-1}\mathbf{W}\mathbf{Q}_w = \mathbf{D}_w$ and $\mathbf{Q}_z^{-1}\mathbf{Z}\mathbf{Q}_z = \mathbf{D}_z$ are each diagonal, then

$$(\mathbf{PQ})^{-1}\mathbf{A}(\mathbf{PQ}) = \begin{pmatrix} \lambda\mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_z^{-1}\mathbf{D}\mathbf{Q}_z \end{pmatrix} \quad \text{and} \quad (\mathbf{PQ})^{-1}\mathbf{B}(\mathbf{PQ}) = \begin{pmatrix} \mathbf{D}_w & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_z \end{pmatrix}.$$

Thus the problem is deflated because $\mathbf{A}_2 = \mathbf{Q}_z^{-1}\mathbf{D}\mathbf{Q}_z$ and $\mathbf{B}_2 = \mathbf{D}_z$ commute and are diagonalizable, so the same argument can be applied to them. If $\mathbf{A}$ has k distinct eigenvalues, then the desired conclusion is attained after k repetitions.

7.2.17. It's not legitimate to equate $p(\mathbf{A})$ with $\det(\mathbf{A} - \mathbf{AI})$ because the former is a matrix while the latter is a scalar.

7.2.18. This follows from the eigenvalue formula developed in Example 7.2.5 (p. 514) by using the identity $1 - \cos\theta = 2\sin^2(\theta/2)$.

7.2.19. (a) The result in Example 7.2.5 (p. 514) shows that the eigenvalues of $\mathbf{N}+\mathbf{N}^T$ and $\mathbf{N}-\mathbf{N}^T$ are $\lambda_j = 2\cos(j\pi/n+1)$ and $\lambda_j = 2\mathrm{i}\cos(j\pi/n+1)$, respectively.

(b) Since $\mathbf{N} - \mathbf{N}^T$ is skew symmetric, it follows from Exercise 6.1.12 (p. 473) that $\mathbf{N}-\mathbf{N}^T$ is nonsingular if and only if n is even, which is equivalent to saying $\mathbf{N} - \mathbf{N}^T$ has no zero eigenvalues (recall Exercise 7.1.6, p. 501), and hence, by part (a), the same is true for $\mathbf{N} + \mathbf{N}^T$.

(b: Alternate) Since the eigenvalues of $\mathbf{N}+\mathbf{N}^T$ are $\lambda_j = 2\cos(j\pi/n+1)$ you can argue that $\mathbf{N}+\mathbf{N}^T$ has a zero eigenvalue (and hence is singular) if and only if n is odd by showing that there exists an integer α such that $j\pi/n+1 = \alpha\pi/2$ for some $1 \le j \le n$ if and only if n is odd.

(c) Since a determinant is the product of eigenvalues (recall (7.1.8), p. 494), $\det(\mathbf{N} - \mathbf{N}^T)/\det(\mathbf{N} + \mathbf{N}^T) = (\mathrm{i}\lambda_1 \cdots \mathrm{i}\lambda_n)/(\lambda_1 \cdots \lambda_n) = \mathrm{i}^n = (-1)^{n/2}$.

7.2.20. The eigenvalues are $\{2, 0, 2, 0\}$. The columns of $\mathbf{F}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -\mathrm{i} & -1 & \mathrm{i} \\ 1 & -1 & 1 & -1 \\ 1 & \mathrm{i} & -1 & -\mathrm{i} \end{pmatrix}$ are corresponding eigenvectors.

7.2.21. $\mathbf{Ax} = \lambda\mathbf{x} \implies \mathbf{y}^*\mathbf{Ax} = \lambda\mathbf{y}^*\mathbf{x}$ and $\mathbf{y}^*\mathbf{A} = \mu\mathbf{y}^* \implies \mathbf{y}^*\mathbf{Ax} = \mu\mathbf{y}^*\mathbf{x}$. Therefore, $\lambda\mathbf{y}^*\mathbf{x} = \mu\mathbf{y}^*\mathbf{x} \implies (\lambda - \mu)\mathbf{y}^*\mathbf{x} = 0 \implies \mathbf{y}^*\mathbf{x} = 0$ when $\lambda \ne \mu$.

7.2.22. (a) Suppose $\mathbf{P}$ is a nonsingular matrix such that $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$ is diagonal, and suppose that λ is the k^{th} diagonal entry in $\mathbf{D}$. If $\mathbf{x}$ and $\mathbf{y}^*$ are the k^{th} column and k^{th} row in $\mathbf{P}$ and $\mathbf{P}^{-1}$, respectively, then $\mathbf{x}$ and $\mathbf{y}^*$ must be right-hand and left-hand eigenvectors associated with λ such that $\mathbf{y}^*\mathbf{x} = 1$.

(b) Consider $\mathbf{A} = \mathbf{I}$ with $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$ for $i \ne j$.

(c) Consider $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

7.2.23. (a) Suppose not—i.e., suppose $\mathbf{y}^*\mathbf{x} = 0$. Then

$$\mathbf{x} \perp span(\mathbf{y}) = N(\mathbf{A} - \lambda\mathbf{I})^* \implies \mathbf{x} \in N(\mathbf{A} - \lambda\mathbf{I})^{*\perp} = R(\mathbf{A} - \lambda\mathbf{I}).$$

Also, $\mathbf{x} \in N(\mathbf{A} - \lambda\mathbf{I})$, so $\mathbf{x} \in R(\mathbf{A} - \lambda\mathbf{I}) \cap N(\mathbf{A} - \lambda\mathbf{I})$. However, because λ is a simple eigenvalue, the the core-nilpotent decomposition on p. 397 insures that $\mathbf{A} - \lambda\mathbf{I}$ is similar to a matrix of the form $\begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & 0_{1\times 1} \end{pmatrix}$, and this implies that $R(\mathbf{A} - \lambda\mathbf{I}) \cap N(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$ (Exercise 5.10.12, p. 402), which is a contradiction. Thus $\mathbf{y}^*\mathbf{x} \neq 0$.

(b) Consider $\mathbf{A} = \mathbf{I}$ with $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$ for $i \neq j$.

7.2.24. Let $\mathcal{B}_i$ be a basis for $N(\mathbf{A} - \lambda_i\mathbf{I})$, and suppose $\mathbf{A}$ is diagonalizable. Since *geo mult*$_{\mathbf{A}}(\lambda_i) =$ *alg mult*$_{\mathbf{A}}(\lambda_i)$ for each i, (7.2.4) implies $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ is a set of n independent vectors—i.e., $\mathcal{B}$ is a basis for $\Re^n$. Exercise 5.9.14 now guarantees that $\Re^n = N(\mathbf{A} - \lambda_1\mathbf{I}) \oplus N(\mathbf{A} - \lambda_2\mathbf{I}) \oplus \cdots \oplus N(\mathbf{A} - \lambda_k\mathbf{I})$. Conversely, if this equation holds, then Exercise 5.9.14 says $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ is a basis for $\Re^n$, and hence $\mathbf{A}$ is diagonalizable because $\mathcal{B}$ is a complete independent set of eigenvectors.

7.2.25. Proceed inductively just as in the development of Schur's triangularization theorem. If the first eigenvalue λ is real, the reduction is exactly the same as described on p. 508 (with everything being real). If λ is complex, then $(\lambda, \mathbf{x})$ and $(\overline{\lambda}, \overline{\mathbf{x}})$ are both eigenpairs for $\mathbf{A}$, and, by (7.2.3), $\{\mathbf{x}, \overline{\mathbf{x}}\}$ is linearly independent. Consequently, if $\mathbf{x} = \mathbf{u} + \mathrm{i}\mathbf{v}$, with $\mathbf{u}, \mathbf{v} \in \Re^{n\times 1}$, then $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent—otherwise, $\mathbf{u} = \xi\mathbf{v}$ implies $\mathbf{x} = (1 + \mathrm{i}\xi)\mathbf{u}$ and $\overline{\mathbf{x}} = (1 - \mathrm{i}\xi)\mathbf{u}$, which is impossible. Let $\lambda = \alpha + \mathrm{i}\beta$, $\alpha, \beta \in \Re$, and observe that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ implies $\mathbf{A}\mathbf{u} = \alpha\mathbf{u} - \beta\mathbf{v}$ and $\mathbf{A}\mathbf{v} = \beta\mathbf{u} + \alpha\mathbf{v}$, so $\mathbf{A}\mathbf{W} = \mathbf{W}\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$, where $\mathbf{W} = \big[\mathbf{u} \,|\, \mathbf{v}\big]$. Let $\mathbf{W} = \mathbf{Q}_{n\times 2}\mathbf{R}_{2\times 2}$ be a rectangular QR factorization (p. 311), and let $\mathbf{B} = \mathbf{R}\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}\mathbf{R}^{-1}$ so that $\sigma(\mathbf{B}) = \sigma\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} = \{\lambda, \overline{\lambda}\}$, and

$$\mathbf{A}\mathbf{W} = \mathbf{A}\mathbf{Q}\mathbf{R} = \mathbf{Q}\mathbf{R}\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \implies \mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{R}\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}\mathbf{R}^{-1} = \mathbf{B}.$$

If $\mathbf{X}_{n\times n-2}$ is chosen so that $\mathbf{P} = \big[\mathbf{Q} \,|\, \mathbf{X}\big]$ is an orthogonal matrix (i.e., the columns of $\mathbf{X}$ complete the two columns of $\mathbf{Q}$ to an orthonormal basis for $\Re^n$), then $\mathbf{X}^T\mathbf{A}\mathbf{Q} = \mathbf{X}^T\mathbf{Q}\mathbf{B} = \mathbf{0}$, and

$$\mathbf{P}^T\mathbf{A}\mathbf{P} = \begin{pmatrix} \mathbf{Q}^T\mathbf{A}\mathbf{Q} & \mathbf{Q}^T\mathbf{A}\mathbf{X} \\ \mathbf{X}^T\mathbf{A}\mathbf{Q} & \mathbf{X}^T\mathbf{A}\mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{B} & \mathbf{Q}^T\mathbf{A}\mathbf{X} \\ \mathbf{0} & \mathbf{X}^T\mathbf{A}\mathbf{X} \end{pmatrix}.$$

Now repeat the argument on the $n - 2 \times n - 2$ matrix $\mathbf{X}^T\mathbf{A}\mathbf{X}$. Continuing in this manner produces the desired conclusion.

7.2.26. Let the columns $\mathbf{R}_{n\times r}$ be linearly independent eigenvectors corresponding to the real eigenvalues ρ_j, and let $\{\mathbf{x}_1, \overline{\mathbf{x}}_1, \mathbf{x}_2, \overline{\mathbf{x}}_2, \ldots, \mathbf{x}_t, \overline{\mathbf{x}}_t\}$ be a set of linearly independent eigenvectors associated with $\{\lambda_1, \overline{\lambda}_1, \lambda_2, \overline{\lambda}_2, \ldots, \lambda_t, \overline{\lambda}_t\}$ so that the matrix $\mathbf{Q} = \big[\mathbf{R} \,|\, \mathbf{x}_1 \,|\, \overline{\mathbf{x}}_1 \,|\, \cdots \,|\, \mathbf{x}_t \,|\, \overline{\mathbf{x}}_t\big]$ is nonsingular. Write $\mathbf{x}_j = \mathbf{u}_j + \mathrm{i}\mathbf{v}_j$ for

$\mathbf{u}_j, \mathbf{v}_j \in \Re^{n\times 1}$ and $\lambda_j = \alpha_j + \mathrm{i}\beta_j$ for $\alpha, \beta \in \Re$, and let $\mathbf{P}$ be the real matrix $\mathbf{P} = \big[\mathbf{R}\,|\,\mathbf{u}_1\,|\,\mathbf{v}_1\,|\,\mathbf{u}_2\,|\,\mathbf{v}_2\,|\cdots|\,\mathbf{u}_t\,|\,\mathbf{v}_t\big]$. This matrix is nonsingular because Exercise 6.1.14 can be used to show that $\det(\mathbf{P}) = 2t(-\mathrm{i})^t \det(\mathbf{Q})$. For example, if $t = 1$, then $\mathbf{P} = \big[\mathbf{R}\,|\,\mathbf{u}_1\,|\,\mathbf{v}_1\big]$ and

$$\begin{aligned}\det(\mathbf{Q}) &= \det\big[\mathbf{R}\,|\,\mathbf{x}_1\,|\,\overline{\mathbf{x}}_1\big] = \det\big[\mathbf{R}\,|\,\mathbf{u}_1 + \mathrm{i}\mathbf{v}_1\,|\,\mathbf{u}_1 - \mathrm{i}\mathbf{v}_1\big]\\ &= \det\big[\mathbf{R}\,|\,\mathbf{u}_1\,|\,\mathbf{u}_1\big] + \det\big[\mathbf{R}\,|\,\mathbf{u}_1\,|-\mathrm{i}\mathbf{v}_1\big]\\ &\quad + \det\big[\mathbf{R}\,|\,\mathrm{i}\mathbf{v}_1\,|\,\mathbf{u}_1\big] + \det\big[\mathbf{R}\,|\,\mathrm{i}\mathbf{v}_1\,|\,\mathrm{i}\mathbf{v}_1\big]\\ &= -\mathrm{i}\det\big[\mathbf{R}\,|\,\mathbf{u}_1\,|\mathbf{v}_1\big] + \mathrm{i}\det\big[\mathbf{R}\,|\,\mathbf{v}_1\,|\,\mathbf{u}_1\big]\\ &= -\mathrm{i}\det\big[\mathbf{R}\,|\,\mathbf{u}_1\,|\mathbf{v}_1\big] - \mathrm{i}\det\big[\mathbf{R}\,|\,\mathbf{u}_1\,|\mathbf{v}_1\big] = 2(-\mathrm{i})\det(\mathbf{P}).\end{aligned}$$

Induction can now be used. The equations $\mathbf{A}(\mathbf{u}_j + \mathrm{i}\mathbf{v}_j) = (\alpha_j + \mathrm{i}\beta_j)(\mathbf{u}_j + \mathrm{i}\mathbf{v}_j)$ yield $\mathbf{A}\mathbf{u}_j = \alpha_j\mathbf{u}_j - \beta_j\mathbf{v}_j$ and $\mathbf{A}\mathbf{v}_j = \beta_j\mathbf{u}_j + \alpha_j\mathbf{v}_j$. Couple these with the fact that $\mathbf{AR} = \mathbf{RD}$ to conclude that

$$\mathbf{AP} = \big[\mathbf{RD}\,|\cdots|\,\alpha_j\mathbf{u}_j - \beta_j\mathbf{v}_j\,|\,\beta_j\mathbf{u}_j + \alpha_j\mathbf{v}_j\,|\cdots\big] = \mathbf{P}\begin{pmatrix}\mathbf{D} & 0 & \cdots & 0\\ 0 & \mathbf{B}_1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \mathbf{B}_t\end{pmatrix},$$

where

$$\mathbf{D} = \begin{pmatrix}\rho_1 & 0 & \cdots & 0\\ 0 & \rho_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \rho_r\end{pmatrix} \quad \text{and} \quad \mathbf{B}_j = \begin{pmatrix}\alpha_j & \beta_j\\ -\beta_j & \alpha_j\end{pmatrix}.$$

7.2.27. Schur's triangularization theorem says $\mathbf{U}^*\mathbf{AU} = \mathbf{T}$ where $\mathbf{U}$ is unitary and $\mathbf{T}$ is upper triangular. Setting $\mathbf{x} = \mathbf{U}\mathbf{e}_i$ in $\mathbf{x}^*\mathbf{A}\mathbf{x} = 0$ yields that $t_{ii} = 0$ for each i, so $t_{ij} = 0$ for all $i \geq j$. Now set $\mathbf{x} = \mathbf{U}(\mathbf{e}_i + \mathbf{e}_j)$ with $i < j$ in $\mathbf{x}^*\mathbf{A}\mathbf{x} = 0$ to conclude that $t_{ij} = 0$ whenever $i < j$. Consequently, $\mathbf{T} = \mathbf{0}$, and thus $\mathbf{A} = \mathbf{0}$. To see that $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0\ \forall\ \mathbf{x} \in \Re^{n\times 1} \not\Rightarrow\ \mathbf{A} = \mathbf{0}$, consider $\mathbf{A} = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$.

Solutions for exercises in section 7. 3

7.3.1. $\cos\mathbf{A} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$. The characteristic equation for $\mathbf{A}$ is $\lambda^2 + \pi\lambda = 0$, so the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 0$ and $\lambda_2 = -\pi$. Note that $\mathbf{A}$ is diagonalizable because no eigenvalue is repeated. Associated eigenvectors are computed in the usual way to be

$$\mathbf{x}_1 = \begin{pmatrix}1\\ 1\end{pmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{pmatrix}-1\\ 1\end{pmatrix},$$

so

$$\mathbf{P} = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix} \quad \text{and} \quad \mathbf{P}^{-1} = \frac{1}{2}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}.$$