Matrix Analysis (Lecture 2)

Yikun Zhang[*]

April 12, 2018

Abstract

Eigenvalues, eigenvectors, and similarity serve as the main topic of Chapter 1 of Horn's book[1]. In the last lecture, we introduced the origin of similar matrices, which are different basis representations of a single linear transformation, and scrutinized the existence and cardinality of eigenvalues given a matrix $A \in M_n$. At that time, computing the zeroes of its characteristic polynomial is a fundamental avenue to uncover eigenvalues and calculate their corresponding algebraic multiplicities of A. In this lecture, we are supposed to discuss the properties of similarity and diagonalizable matrices and their connections with eigenvalues and eigenvectors in detail. Meanwhile, the concept of left eigenvectors and geometric multiplicity will be unfolded and utilized to determine the diagonalizability of a matrix.

1 Similarity (Page 57-69, Page 17, Page 21)

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

---

[*]School of Mathematics, Sun Yat-sen University

Definition 1.1. *Let $A, B \in M_n$ be given. B is said to be similar to A if there exists a nonsingular $S \in M_n$ such that*

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

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

Remark. A square matrix P is a *permutation matrix* if exactly one entry in each row and column is equal to 1 and all other entries are 0.

Observation 1.2. *Similarity is an* equivalence relation *on M_n; that is, similarity is* reflexive*,* symmetric*, and* transitive*.*

Proof. (Reflexiveness): $A = I^{-1}AI$, where I is the identity matrix in M_n.
(Symmetry): If $B \sim A$, there exists a nonsingular matrix S such that $B = S^{-1}AS$. Thus, $A = SBS^{-1} = (S^{-1})^{-1}BS^{-1}$, showing that $A \sim B$.
(Transitiveness): If $B \sim A$ and $C \sim B$, then there exist nonsingular matrices S_1, S_2 such that $B = S_1^{-1}AS_1, C = S_2^{-1}BS_2$. Therefore, we find that $C = S_2^{-1}BS_2 = S_2^{-1}S_1^{-1}AS_1S_2 = (S_1S_2)^{-1}A(S_1S_2)$, where S_1S_2 is still a nonsingular matrix by computing its determinant. $\square$

Like any equivalence relation, similarity partitions the set M_n into disjoint equivalence classes. Within each dissimilarity equivalence class, matrices share many important properties. Here we discuss some of them in detail.

Theorem 1.3. *Let $A, B \in M_n$. If B is similar to A, then A and B have the same characteristic polynomial.*

Proof. Compute

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

$\square$

Remark. Having the same eigenvalues is a necessary but not sufficient condition for similarity. Consider $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, which have the same eigenvalues but are not similar. This is because $\operatorname{rank} A = 1$ while $\operatorname{rank} B = 0$. A necessary condition for similar matrices is that they have the same rank.

Corollary 1.4. *Let $A, B \in M_n$ and suppose that A is similar to B. Then*

(a) A and B have the same eigenvalues.

(b) If B is a diagonal matrix, its main diagonal entries are the eigenvalues of A.

(c) $B = 0$ (a diagonal matrix) if and only if $A = 0$.

(d) $B = I$ (a diagonal matrix) if and only if $A = I$.

Proof. (a) and (b) follow from Theorem 1.3. (c,d) If $A \sim B$, then $A = S^{-1}BS, B = SAS^{-1}$, where S is a nonsingular matrix. Then $A = 0 \Leftrightarrow B = 0$ and $A = I \Leftrightarrow B = I$. □

Example 1.5. Suppose that $A, B \in M_n$ are similar and let $q(t)$ be a given polynomial. It turns out that $q(A)$ and $q(B)$ are similar. This is because $B = S^{-1}AS$, where S is nonsingular. Assume that $q(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$. Then $q(A) = a_n A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I$ and

$$\begin{aligned} S^{-1}q(A)S &= S^{-1}(a_n A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I)S \\ &= a_n(S^{-1}AS)^n + a_{n-1}(S^{-1}AS)^{n-1} + \cdots + a_1(S^{-1}AS) + a_0 I \\ &= a_n B^n + a_{n-1}B^{n-1} + \cdots + a_1 B + a_0 I \\ &= q(B). \end{aligned}$$

In particular, we show that $A + \alpha I$ and $B + \alpha I$ are similar for any $c \in \mathbf{C}$ by choosing $q(t) = t + \alpha$.

Example 1.6. Let $A, S \in M_n$ and suppose that S is nonsingular. Then we conclude that $S_k(S^{-1}AS) = S_k(A)$ for all $k = 1, ..., n$. (Recall that Definition (3.9) in Lecture 1) states that $S_k(A) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \prod_{j=1}^{k} \lambda_{i_j}$, where $\lambda_1, ..., \lambda_n$ are eigenvalues of A.) This can be deduced from Theorem 1.4, which illustrates that A and $S^{-1}AS$ have the same eigenvalues. Moreover, since $E_k(A) = S_k(A)$ (Theorem (3.10) in Lecture 1), we also have $E_k(S^{-1}AS) = E_k(A)$ for all $k = 1, ..., n$. Thus, all the principal minor sums (Definition (3.7)) are similarity invariants, not just the determinant and trace.

Since diagonal matrices are especially simple and have very nice properties, we would like to know which matrices are similar to diagonal matrices.

Definition 1.7. *If $A \in M_n$ is similar to diagonal matrix, then A is said to be diagonalizable.*

After defining diagonalizable matrices, we would like to figure out an algorithm to achieve the diagonal form for a given diagonalizable matrix, which allows for the following theorem.

Theorem 1.8. *Let $A \in M_n$ be given.*

(a) A is similar to a block matrix of the form

$$\begin{bmatrix} \Lambda & C \\ 0 & D \end{bmatrix}, \quad \Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_k),\ D \in M_{n-k},\ 1 \leq k < n, \tag{1}$$

where $\lambda_1, ..., \lambda_k$ are eigenvalues of A if and only if there are k linearly independent vectors in $\mathbf{C}^n$, each of which is an eigenvector of A.

(b) The matrix A is diagonalizable if and only if there are n linearly independent vectors, each of which is an eigenvector of A.

(c) If $x^{(1)}, ..., x^{(n)}$ are linearly independent eigenvectors of A and if $S = [x^{(1)} \cdots x^{(n)}]$, then $S^{-1}AS$ is a diagonal matrix, whose diagonal entries are all of the corresponding eigenvalues of A.

Proof. (a) Suppose that $k < n$, the n-vectors $x^{(1)}, ..., x^{(k)}$ are linearly independent, and $Ax^{(i)} = \lambda_i x^{(i)}$ for each $i = 1, ..., k$. Let $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_k)$, let $S_1 = [x^{(1)} \cdots x^{(n)}]$, and choose any $S_2 \in M_{n,n-k}$ such that $S = [S_1\ S_2]$ is nonsingular. Compute

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

The assertion about the eigenvalues follow from an extension of the characteristic polynomial: $p_A(t) = p_\Lambda(t)p_D(t)$ if $k < n$.

Conversely, if S is nonsingular, $S^{-1}AS = \begin{bmatrix} \Lambda & C \\ 0 & D \end{bmatrix}$, and we partition $S = [S_1\ S_2]$ with $S_1 \in M_{n,k}$, then S_1 has linearly independent columns and

$$[AS_1\ AS_2] = AS = S \begin{bmatrix} \Lambda & C \\ 0 & D \end{bmatrix} = [S_1\Lambda\ S_1C + S_2D].$$

Thus, $AS_1 = S_1\Lambda$, so each column of S_1 is an eigenvector of A.
(b,c) If $k = n$ and we have a basis $\{x^{(1)}, ..., x^{(n)}\}$ of $\mathbf{C}^n$ such that $Ax^{(i)} = \lambda_i x^{(i)}$ for each $i = 1, ..., n$, let $\Lambda = \text{diag}(\lambda_1, ..., \lambda_n)$ and let $S = [x^{(1)} \cdots x^{(n)}]$, which is nonsingular. Our previous argument shows that $S^{-1}AS = \Lambda$. Conversely, if S is nonsingular and $S^{-1}AS = \Lambda$, then $AS = S\Lambda$, so each column of S is an eigenvector of A. In addition, $p_A(t) = p_\Lambda(t)$ when $k = n$ shows that all the diagonal entries of Λ are corresponding eigenvalues. □

Remark. Theorem 1.8 furnishes us a plausible but somehow inefficient approach to diagonalize a diagonalizable matrix: align n linearly independent eigenvectors of $A \in M_n$ by columns to construct the similarity matrix S. However, except for some small matrices, this is *not* a practical computational procedure. In practice, there are miscellaneous algorithms aiming at different types of matrices. Here we introduce a numerical method to computing the greatest eigenvalue and corresponding eigenvector, *power iteration*, which is commonly used when A is sparse.

Example 1.9 (Power Iteration Algorithm[2]). Assume that A has an eigenvalue that is strictly greater in magnitude than its other eigenvalues and the initial vector b_0 has a nonzero component in the direction of an eigenvector associated with the dominant eigenvalue. Then the recursive sequence

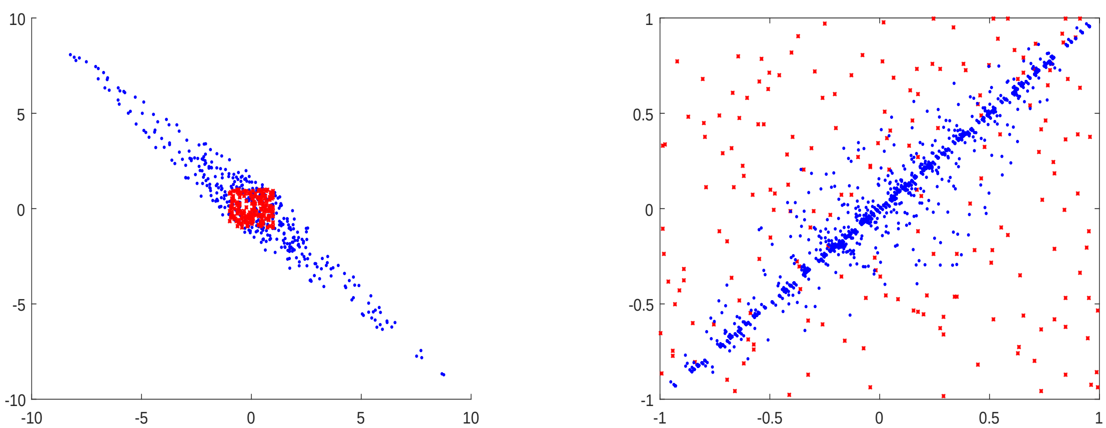
$$b_{k+1} = \frac{Ab_k}{||Ab_k||}$$

converges to an eigenvector associated with the dominant eigenvalue. To obtain the dominant eigenvalue (spectral radius), we resort to the Rayleigh quotient

$$\rho(A) = \frac{b_k^T Ab_k}{b_k^T b_k} = \frac{b_{k+1}^T b_k}{b_k^T b_k}.$$

The Matlab code for the power iterative algorithm can be found on the course website. In two-dimensional case[3], we randomly simulate $n = 200$ points within the square $\{(x, y) : |x| \leq 1, |y| \leq 1\}$ and apply a linear transformation $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ to each of them iteratively. The eigenvalues of A are $\lambda_1 = 1, \lambda_2 = 3$ and the associated eigenvectors are $(1, 1)^T, (1, -1)^T$, respectively. It turns out that the resulting points converge to the direction of the principal eigenvector $(1, -1)^T$ as the number of iterative times increases. In a similar fashion, we can determine the principal eigenvector of A^{-1}. See Figure 1 for details.

Here we briefly sketch the proof of the correctness of the algorithm.



(a) Iterations with A (b) Iterations with A^{-1}

Figure 1: Visualization of Original Simulated Points (red asterisk points) and the Resulting Points Operated by A (blue dotted points).

Proof. Recall that a matrix $A \in M_n$ is similar to its *Jordan normal form*, namely, a block diagonal matrix

$$J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{bmatrix}$$

where each block J_i is a square matrix of the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

and $\lambda_i \in \sigma(A), i = 1, ..., p$. Thus, A can be decomposed into $A = VJV^{-1}$, where V is nonsingular and the first column of V is an eigenvalue of A corresponding to the dominant eigenvalue, says λ_1. Since the dominant eigenvalue of A is unique, the first Jordan block of J is the 1-by-1 matrix $[\lambda_1]$. Since V

is nonsingular, we write the starting vector b_0 as a linear combination of the columns of V, i.e.,

$$b_0 = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$

By assumption, b_0 has a nonzero component in the direction of an eigenvector associated with the dominant eigenvalue, so $c_1 \neq 0$. The recurrence relation for b_{k+1} can be rewritten as

$$\begin{aligned}
b_{k+1} &= \frac{Ab_k}{||Ab_k||} = \frac{A^{k+1}b_0}{||A^{k+1}b_0||} \\
&= \frac{(VJV^{-1})^{k+1}b_0}{||(VJV^{-1})^{k+1}b_0||} = \frac{VJ^{k+1}V^{-1}b_0}{||VJ^{k+1}V^{-1}b_0||} \\
&= \frac{VJ^{k+1}V^{-1}(c_1v_1 + c_2v_2 + \cdots + c_nv_n)}{||VJ^{k+1}V^{-1}(c_1v_1 + c_2v_2 + \cdots + c_nv_n)||} \\
&= \frac{VJ^{k+1}(c_1e_1 + c_2e_2 + \cdots + c_ne_n)}{||VJ^{k+1}(c_1e_1 + c_2e_2 + \cdots + c_ne_n)||} \\
&= \left(\frac{\lambda_1}{|\lambda_1|}\right)^{k+1} \cdot \frac{c_1}{|c_1|} \cdot \frac{v_1 + \frac{1}{c_1}V(\frac{1}{\lambda_1}J)^{k+1}(c_2e_2 + \cdots + c_ne_n)}{||v_1 + \frac{1}{c_1}V(\frac{1}{\lambda_1}J)^{k+1}(c_2e_2 + \cdots + c_ne_n)||}
\end{aligned}$$

The expression above simplifies as $k \to \infty$, since

$$\left(\frac{1}{\lambda_1}J\right)^{k+1} = \begin{bmatrix} [1] & & & \\ & \left(\frac{1}{\lambda_1}J_2\right)^{k+1} & & \\ & & \ddots & \\ & & & \left(\frac{1}{\lambda_1}J_m\right)^{k+1} \end{bmatrix} \to \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix},$$

where we use the fact that the eigenvalues of $\frac{1}{\lambda_1}J_i$ are less than 1 in magnitude, so $\left(\frac{1}{\lambda_1}J_i\right)^{k+1} \to 0, i = 1, ..., p$, as $k \to \infty$.
It follows that as $k \to \infty$

$$\frac{1}{c_1}V\left(\frac{1}{\lambda_1}J\right)^{k+1}(c_2e_2 + \cdots + c_ne_n) \to 0.$$

Therefore, b_{k+1} can be written in a form that emphasizes its relationship with v_1 when k is large, i.e.,

$$b_{k+1} = \left(\frac{\lambda_1}{|\lambda_1|}\right)^{k+1} \frac{c_1}{|c_1|} \cdot \frac{v_1 + \frac{1}{c_1}V(\frac{1}{\lambda_1}J)^{k+1}(c_2e_2 + \cdots + c_ne_n)}{||v_1 + \frac{1}{c_1}V(\frac{1}{\lambda_1}J)^{k+1}(c_2e_2 + \cdots + c_ne_n)||} = e^{i\phi_{k+1}} \frac{c_1}{|c_1|} \frac{v_1}{||v_1||} + r_{k+1},$$

where $e^{i\phi_{k+1}} = \left(\frac{\lambda_1}{|\lambda_1|}\right)^{k+1}$ and $||r_{k+1}|| \to 0$ as $k \to \infty$.
The sequence $\{b_k\}$ is bounded, so it contains a convergent subsequence. Note that the eigenvector corresponding to the dominant eigenvalue is only unique up to a scalar, so although the sequence $\{b_k\}$ may not converge, b_k is nearly an eigenvector of A for large k.

In particular, if A is diagonalizable, it can be decomposed into $A = V\Lambda V^{-1}$, where Λ is diagonal and the columns of V, $v_1, ..., v_n$, are linearly independent eigenvectors of A. Suppose that λ_1 is the dominant eigenvalue. The initial vector b_0 can be written into

$$b_0 = c_1 v_1 + c_2 v_2 + \cdots + c_m v_m.$$

(If b_0 is chosen randomly with uniform probability, then $c \neq 0$ with probability 1.) Now,

$$\begin{aligned} A^k b_0 &= c_1 A^k v_1 + c_2 A^k v_2 + \cdots + c_n A^k v_n \\ &= c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \cdots + c_m \lambda_n^k v_n \\ &= c_1 \lambda_1^k \left(v_1 + \frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + \frac{c_n}{c_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n \right) \to c_1 \lambda_1^k v_1, \end{aligned}$$

as $k \to \infty$. Hence b_k converges to (a multiple of) the eigenvector v_1. □

Remark. (a) The convergent rate of the algorithm is determined by the ratio $\left|\frac{\lambda_2}{\lambda_1}\right|$, where λ_2 denotes the second dominant eigenvalue;

(b) If the set of all linearly independent eigenvectors associated with the dominant eigenvalue has the cardinality more than 1, then the resulting vector of the power iteration algorithm would be the linear combination of these eigenvectors in the list. Equivalently, if we randomly generate many points (or vectors) within the vector space, the points after iterations will spread out the space spanned by the eigenvectors corresponding to the dominant eigenvalues.

A previous Example (1.5) illuminates that the polynomial evaluation on matrices preserves similarity. Likewise, we can show that it also preserves diagonalizability.

Example 1.10. Let $q(t)$ be a given polynomial. If A is diagonalizable, then there exists a nonsingular matrix S such that $D = S^{-1}AS$ is diagonal. Assume $q(t) = a_n t^n + \cdots + a_1 t + a_0$. Then

$$\begin{aligned} S^{-1}q(A)S &= S^{-1}(a_n A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I)S \\ &= a_n(S^{-1}AS)^n + a_{n-1}(S^{-1}AS)^{n-1} + \cdots + a_1(S^{-1}AS) + a_0 I \\ &= a_n D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0 I, \end{aligned}$$

which is diagonal. However, the converse does not hold. Consider $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $q(t) = t^2 + 1$. Then $q(A) = I_2$ is diagonal, while A is not diagonalizable, since there is only one linearly independent eigenvector $(1,0)^T$ associated with its unique eigenvalue.

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

Lemma 1.11. *Let $\lambda_1, ..., \lambda_k$ be $k \geq 2$ distinct eigenvalues of $A \in M_n$ (that is, $\lambda_i \neq \lambda_j$ if $i \neq j$ and $1 \leq i, j \leq k$), and suppose that $x^{(i)}$ is an eigenvector associated with λ_i for each $i = 1, ..., k$. Then the vectors $x^{(1)}, ..., x^{(k)}$ are linearly independent.*

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

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

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

Theorem 1.12. *If $A \in M_n$ has n distinct eigenvalues, then A is diagonalizable.*

Proof. Let $x^{(i)}$ be an eigenvector associated with the eigenvalue λ_i for each $i = 1, ..., n$. Since all the eigenvalues are distinct, Lemma 1.11 ensures that the vectors $x^{(1)}, ..., x^{(n)}$ are linearly independent. Theorem 1.8 then ensures that A is diagonalizable. $\square$

Remark. Having distinct eigenvalues is *sufficient* for diagonalizability, but certainly, it is not *necessary*. The simplest counterexample is the identity matrix I_n. Also, in Lecture 1, Example 2.3 shows that J_n whose entries are all equal to 1 has n linearly independent eigenvectors but $\sigma(J_n) = \{n, 0\}$.

In general, matrices $A, B \in M_n$ *do not* commute, but if A and B are both diagonal, they always commute. The latter assumption can be generalized to a concept called *simultaneous diagonalizability*.

Definition 1.13. *Two matrices $A, B \in M_n$ are said to be* simultaneously diagonalizable *if there is a single nonsingular $S \in M_n$ such that $S^{-1}AS$ and $S^{-1}BS$ are both diagonal.*

Proposition 1.14. *If $A, B \in M_n$ are simultaneously diagonalizable, then they commute.*

Proof. By definition, there exists a nonsingular matrix S such that $D_1 = S^{-1}AS, D_2 = S^{-1}BS$, where D_1, D_2 are diagonal. Then

$$\begin{aligned} AB &= (SD_1S^{-1})(SD_2S^{-1}) = SD_1D_2S^{-1} \\ &= SD_2D_1S^{-1} = (SD_2S^{-1})(SD_1S^{-1}) \\ &= BA, \end{aligned}$$

where we use the fact that diagonal matrices always commute. $\square$

Surprisingly, the converse of Proposition 1.14 is also true and the proof requires the following lemma.

Lemma 1.15. *Let $B_1 \in M_{n_1}, ..., B_d \in M_{n_d}$ be given and let B be the direct sum*

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

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

Proof. ($\Leftarrow$) If for each $i = 1, ..., d$ there is a nonsingular $S_i \in M_{n_i}$ such that $S_i^{-1} B_i S_i$ is diagonal, and if we define $S = S_1 \oplus \cdots \oplus S_d$, then one checks that $S^{-1}BS$ is diagonal.
($\Rightarrow$) We proceed by induction. There is nothing to prove for $d = 1$. Suppose that $d \geq 2$ and that the assertion has been established for direct sums with $d-1$ or fewer direct summands. Let $C = B_1 \oplus \cdots \oplus B_{d-1}$, $n = n_1 + \cdots + n_d$, and $m = n_d$. Let $S \in M_{n+m}$ be nonsingular and such that

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

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

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

Then $Bs_i = \lambda_i s_i$ implies that $C\xi_i = \lambda_i \xi_i$ and $B_d \eta_i = \lambda_i \eta_i$ for $i = 1, 2, ..., n+m$. The row rank of $[\xi_i\, \cdots\, \xi_{n+m}] \in M_{n,n+m}$ is n because this matrix comprises the first n rows of the nonsingular matrix S. Thus, its column rank is also n, so the list $\xi_1, ..., \xi_{n+m}$ contains a linearly independent list of n vectors, each of which is an eigenvector of C. Theorem 1.8 ensures that C is diagonalizable and the induction hypothesis ensures that its direct summands $B_1, ..., B_{d-1}$ are all diagonalizable. The row rank of $[\eta_1\, \cdots\, \eta_{n+m}] \in M_{m,n+m}$ is m, so the list $\eta_1, ..., \eta_{n+m}$ contains a linearly independent list of m vectors; Theorem 1.8 enables that B_d is diagonalizable as well. □

Now we are prepared to prove the converse of Proposition 1.14 and summarize both directions into the following theorem.

Theorem 1.16. *Let $A, B \in M_n$ be diagonalizable. Then A and B commute if and only if they are simultaneously diagonalizable.*

Proof. ($\Leftarrow$) It has been proved in Proposition 1.14.
($\Rightarrow$) Assume that A and B commute. We perform a similarity transformation S on both A and B that diagonalizes A (but not necessarily B) and groups together any repeated eigenvalues of A. Denote $S^{-1}AS$ and $S^{-1}BS$ by $\tilde{A}$ and $\tilde{B}$, respectively. If $\mu_1, ..., \mu_d$ are the distinct eigenvalues of A and $n_1, ..., n_d$ are their respective multiplicities, then we may assume that

$$\tilde{A} = \begin{bmatrix} \mu_1 I_{n_1} & & & 0 \\ & \mu_2 I_{n_2} & & \\ & & \ddots & \\ 0 & & & \mu_d I_{n_d} \end{bmatrix}, \mu_i \neq \mu_j \text{ if } i \neq j \tag{2}$$

Since $AB = BA$,

$$\tilde{A}\tilde{B} = (S^{-1}AS)(S^{-1}BS) = S^{-1}ABS = S^{-1}BAS = (S^{-1}BS)(S^{-1}AS) = \tilde{B}\tilde{A}.$$

Thus, partitioning $\tilde{B}$ conformally to $\tilde{A}$, we obtain that $\mu_i B_{ij} = B_{ij}\mu_j$ for each $i, j = 1, ..., d$. This equality holds if and only if $B_{ij} = 0$ whenever $i \neq j$, since $\mu_i \neq \mu_j$. It shows that

$$\tilde{B} = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_d \end{bmatrix}, \text{ each } B_i \in M_{n_i} \tag{3}$$

is a block diagonal conformal to A. Since B is diagonalizable, $\tilde{B}$ is also diagonalizable and Lemma 1.15 ensures that each B_i is diagonalizable. Let $T_i \in M_{n_i}$ be nonsingular and such that $T_i^{-1}B_iT_i$ is diagonal for each $i = 1, ..., d$; let

$$T = \begin{bmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \ddots & \\ 0 & & & T_d \end{bmatrix} \tag{4}$$

Then $T_i^{-1}\mu_i I_{n_i} T_i = \mu_i I_{n_i}$, so $T^{-1}\tilde{A}T = \tilde{A}$ and $T^{-1}\tilde{B}T$ are both diagonal. Therefore, ST is a similarity matrix that can diagonalize both A and B. $\square$

Remark. The two matrix partitions are said to be *conformal* if $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{n,p}(\mathbf{F})$ are partitioned so that the two partitions of $\{1, ..., n\}$ coincide.

We want to have a version of Theorem 1.16 involving arbitrarily many commuting diagonalizable matrices. Central to our investigation is the notion of an invariant subspace and the companion notion of a block triangular matrix.

Definition 1.17. *(a) A family $\mathcal{F} \subseteq M_n$ of matrices is a nonempty finite or infinite set of matrices; a* commuting family *is a family of matrices in which every pair of matrices commutes.*
(b) For a given $A \in M_n$, a subspace $W \subseteq \mathbf{C}^n$ is A-invariant *if $Aw \in W$ for every $w \in W$. (A subspace $W \subseteq \mathbf{C}^n$ is* trivial *if either $W = \{0\}$ or $W = \mathbf{C}^n$; otherwise, it is* nontrivial*.)*

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

Example 1.18. For $A \in M_n$, each nonzero element of a one-dimensional A-invariant subspace of $\mathbf{C}^n$ is an eigenvector of A. This is because we may assume that the subspace W is spanned by a nonzero vector x_1, i.e., $W = \text{span}(\{x_1\})$. Then for each $x \in W, x \neq 0$, we have $x = c_1 x_1$, where $c_1 \in \mathbf{C}^*$. Since W is A-invariant, there exists an $c_2 \in \mathbf{C}$ such that $Ax = c_2 x = \frac{c_2}{c_1}(c_1 x_1) = \frac{x_2}{x_1} x$, yielding that x is indeed an eigenvector of A.

Invariant subspaces and block triangular matrices are two sides of the same valuable coin: The former is the linear algebra side, while the latter is the matrix analysis side. We now exploit the connection between invariant subspaces and triangular block matrices.

Proposition 1.19. *Suppose that $n \geq 2$. A given matrix $A \in M_n$ is similar to a block matrix of the form*

$$\begin{bmatrix} B & C \\ 0 & D \end{bmatrix}, \quad B \in M_k, 1 \leq k \leq n-1 \tag{5}$$

if and only if some nontrivial subspace of $\mathbf{C}^n$ is A-invariant. Moreover, if $W \subseteq \mathbf{C}^n$ is a nonzero A-invariant subspace, then some vector in W is an eigenvector of A. A given family $\mathcal{F} \subseteq M_n$ is reducible if and only if there is some $k \in \{1, ..., n-1\}$ and a nonsingular $S \in M_n$ such that $S^{-1}AS$ has the form (5) for every $A \in \mathcal{F}$.

Proof. Let $A \in M_n$ with $n \geq 2$ and suppose that $W \subseteq \mathbf{C}^n$ is a k-dimensional subspace with $1 \geq k < n$. Choose a basis $s_1, ..., s_k$ of W and let $S_1 = [s_1 \cdots s_n] \in M_{n,k}$. Choose any $s_{k+1}, ..., s_n$ such that $s_1, ..., s_n$ is a basis for $\mathbf{C}^n$, let $S_2 = [s_{k+1} \cdots s_n] \in M_{n,n-k}$, and let $S = [S_1\ S_2]$; S has linearly independent columns, so it is nonsingular.
($\Leftarrow$) If W is A-invariant, then $As_j \in W$ for each $j = 1, ..., k$, so each As_j is a linear combination of $s_1, ..., s_k$, that is, $AS_1 = S_1 B$ for some $B \in M_k$. Thus, $AS = [AS_1\ AS_2] = [S_1 B\ AS_2]$ and hence

$$\begin{aligned} S^{-1}AS = [S^{-1}S_1 B\ S^{-1}AS_2] &= \begin{bmatrix} \begin{bmatrix} I_k \\ 0 \end{bmatrix} B & S^{-1}AS_2 \end{bmatrix} \\ &= \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}, \quad B \in M_k, 1 \leq k \leq n-1 \end{aligned} \tag{6}$$

But we can say a little more: We know that $B \in M_k$ has an eigenvalue, so suppose that $B\xi = \lambda\xi$ for some scalar λ, and a nonzero $\xi \in \mathbf{C}^k$. Then $0 \neq S_1\xi \in W$ ($\operatorname{rank} S_1 = k$ and $S_1\xi$ is a linear combination of $s_1, ..., s_k$) and $A(S_1\xi) = (AS_1)\xi = S_1B\xi = \lambda(S_1\xi)$, which means that A has an eigenvector in W.
($\Rightarrow$) If $S = [S_1\ S_2] \in M_n$ is nonsingular, $S_1 \in M_{n,k}$, and $S^{-1}AS$ has the block triangular form (5), then

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

so the (k-dimensional) span of the columns of S_1 is A-invariant. □

Remark. Proposition 1.19 is somehow similar to Theorem 1.8, except that now B in the block matrix form (5) is no longer diagonal in that we are only given some nontrivial A-invariant subspace of $\mathbf{C}^n$.

Now we digress to talk about a lemma that is at the heart of many subsequent results.

Lemma 1.20. *Let $\mathcal{F} \subset M_n$ be a commuting family. Then some nonzero vector in $\mathbf{C}^n$ is an eigenvector of every $A \in \mathcal{F}$.*

Proof. There always exists a nonzero $\mathcal{F}$-invariant subspace, namely, $\mathbf{C}^n$. Let $m = \min\{\dim V : V \text{ is a nonzero } \mathcal{F}\text{-invariant subspace of } \mathbf{C}^n\}$ and let W be any given $\mathcal{F}$-invariant subspace such that $\dim W = m$. Let any $A \in \mathcal{F}$ be given.
Since W is $\mathcal{F}$-invariant, it is also A-invariant, so Proposition 1.19 ensures that there is some nonzero $x_0 \in W$ and some $\lambda \in \mathbf{C}^n$ such that $Ax_0 = \lambda x_0$. Consider the subspace $W_{A,\lambda} = \{x \in W : Ax = \lambda x\}$. Then $x_0 \in W_{A,\lambda}$, so $W_{A,\lambda}$ is a nonzero subspace of W. For any $B \in \mathcal{F}$ and any $x \in W_{A,\lambda}$, $\mathcal{F}$-invariance of W ensures that $Bx \in W$. Using commutativity of $\mathcal{F}$, we compute

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

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

Remark. Indeed, the nonzero $\mathcal{F}$-invariant subspace W in the preceding proof has dimension 1, i.e., $m = \dim W = 1$. This is because for any $x \in W, x \neq 0$, for any $A \in \mathcal{F}$, there exists a scalar $\lambda_A \in \mathbf{C}$ such that $Ax = \lambda_A x \in \text{span}(\{x\})$, yielding that $\text{span}(\{x\})$ is also an $\mathcal{F}$-invariant subspace. By the minimality of W, we have $\dim W \leq \dim \text{span}\{x\}$. On the other hand, since $x \in W$, we also find that $\text{span}(\{x\}) \subset W$. Hence $\dim W = \dim \text{span}(\{x\}) = 1$.

Example 1.21. Suppose that $\mathcal{F} \subset M_n$ is a commuting family. Then by Lemma 1.20, there exists a nonzero vector $x \in W$ such that it is an eigenvector of every $A \in \mathcal{F}$. If we choose any $x_2, ..., x_n$ such that $S = [x\ x_2\ \cdots\ x_n] \in M_n$ is nonsingular, then for every $A \in \mathcal{F}$

$$S^{-1}AS = S^{-1}\begin{bmatrix}\lambda x & Ax_2 & \cdots & Ax_n\end{bmatrix} = \begin{bmatrix}\lambda & B \\ 0 & C\end{bmatrix},$$

which is the block triangular form (5) with $k = 1$.

Lemma 1.20 concerns commuting families of arbitrary nonzero cardinality. Our next result shows that Theorem 1.16 can be extended to arbitrary commuting families of diagonalizable matrices.

Definition 1.22. *A family $\mathcal{F} \subset M_n$ is said to be* simultaneously diagonalizable *if there is a single nonsingular $S \in M_n$ such that $S^{-1}AS$ is diagonal for every $A \in \mathcal{F}$.*

Theorem 1.23. *Let $\mathcal{F} \subset M_n$ be a family of diagonalizable matrices. Then $\mathcal{F}$ is a commuting family if and only if it is a simultaneously diagonalizable family.*
In particular, for any given $A_0 \in \mathcal{F}$ and for any given ordering $\lambda_1, ..., \lambda_n$ of the eigenvalues of A_0, there is a nonsingular $S \in M_n$ such that $S^{-1}A_0S = \text{diag}(\lambda_1, ..., \lambda_n)$ and $S^{-1}BS$ is diagonal for every $B \in \mathcal{F}$.

Proof. ($\Leftarrow$) If $\mathcal{F}$ is simultaneously diagonalizable, then it is a commuting family by Proposition 1.14.
($\Rightarrow$) We prove it by induction on n. If $n = 1$, there is nothing to prove, since matrices are just scalars and every family is both commuting and diagonal. Suppose that $n \geq 2$ and that, for each $k = 1, 2, ..., n-1$, any commuting family of diagonalizable matrices in M_k is simultaneously diagonalizable. If every matrix in $\mathcal{F}$ is a scalar matrix, namely, αI, $\alpha \in \mathbf{C}$, there is nothing to

prove, since scalar matrices are already diagonal and we simply take $S = I$. So, we may assume that $A \in \mathcal{F}$ is a given diagonalizable matrix in M_n with distinct eigenvalues $\lambda_1, ..., \lambda_k$ and $k \geq 2$, that $BA = AB$ for every $B \in \mathcal{F}$, and that each $B \in \mathcal{F}$ is diagonalizable. Using the argument in Theorem 1.16, we reduce to the case in which A has the form (2)

$$A = \begin{bmatrix} \lambda_1 I_{n_1} & & & 0 \\ & \lambda_2 I_{n_2} & & \\ & & \ddots & \\ 0 & & & \lambda_k I_{n_k} \end{bmatrix}, \lambda_i \neq \lambda_j \text{ if } i \neq j$$

Since every $B \in \mathcal{F}$ commutes with A, each $B \in \mathcal{F}$ has the form (3)

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

as in Theorem 1.16. Let $B, \hat{B} \in \mathcal{F}$, so $B = B_1 \oplus \cdots \oplus B_k$ and $\hat{B} = \hat{B}_1 \oplus \cdots \oplus \hat{B}_k$, in which each of $B_i, \hat{B}_i$ has the same size and that size is at most $n-1$. By Lemma 1.15, commutativity and diagonalizability of B and $\hat{B}$ imply commutativity and diagonalizability of B_i and $\hat{B}_i$ for each $i = 1, ..., d$. By the induction hypothesis, there are k similarity matrices $T_1, T_2, ..., T_k$ of appropriate size, each of which diagonalizes the corresponding block of *every* matrix in $\mathcal{F}$. Then the direct sum $T_1 \oplus T_2 \oplus \cdots \oplus T_k$ diagonalizes every matrix in $\mathcal{F}$.

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

Although AB and BA need not be the same (and need not be the same size even when both products are defined, as $A \in M_{m,n}, B \in M_{n,m}$ and $m \neq n$), their eigenvalues are as much the same as possible. Indeed, if A and B are both square, then AB and BA have exactly the same eigenvalues.

Theorem 1.24. *Suppose that $A \in M_{m,n}$ and $B \in M_{n,m}$ with $m \leq n$. Then the n eigenvalues of BA are the m eigenvalues of AB together with $n - m$ zeroes; that is, $p_{BA}(t) = t^{n-m}p_{AB}(t)$. If $m = n$ and at least one of A or B is nonsingular, then AB and BA are similar.*

Proof. A computation reveals that

$$\begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix} \begin{bmatrix} I_m & A \\ 0 & I_n \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & I_n \end{bmatrix} = I_{m+n},$$

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

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

Theorem 1.24 has many applications, several of which emerge in the following lectures. Here we discuss two of them.

Example 1.25 (Eigenvalues of a low-rank matrix)**.** Suppose that $A \in M_n$ is factored as $A = XY^T$, where $X, Y \in M_{n,r}$ and $r < n$. Then the eigenvalues of A are the same as those of the matrix $Y^T X \in M_{r,r}$, together with $n - r$ zeroes. For example, consider the all-ones matrix $J_n = ee^T$ (Example 2.3 in Lecture 1). Its eigenvalues are the eigenvalue of the 1-by-1 matrix $e^T e = [n]$, namely, n, together with $n-1$ zeroes. In fact,the eigenvalues of any matrix of the form $A = xy^T$ with $x, y \in \mathbf{C}^n$ (rank A is at most 1) are $y^T x$, together with $n - 1$ zeroes. The eigenvalues of any matrix of the form $A = xy^T + zw^T = [x\ z][y\ w]^T$ with $x, y, z, w \in \mathbf{C}^n$ (rank A is at most 2) are the two eigenvalues of $[y\ w]^T[x\ z] = \begin{bmatrix} y^T x & y^T z \\ w^T x & w^T z \end{bmatrix} \in M_2$, together with $n - 2$ zeroes.

In the last lecture, we apply *Schur complements* of a 2-by-2 block matrix to deduce the *Cauchy's formula for the determinant of a rank-one perturbation*, which is also called *Cauchy's determinant identity.* Here we utilize Theorem 1.24 to yield a more concise derivation.

Example 1.26 (Cauchy's determinant identity)**.** Let a nonsingular $A \in M_n$

and $x, y \in \mathbf{C}^n$ be given. Denote the i^{th} eigenvalue of A by $\lambda_i(A)$. Then

$$\begin{aligned}
\det(A + xy^T) &= (\det A)(\det(I + A^{-1}xy^T)) \\
&= (\det A)\prod_{i=1}^{n} \lambda_i(I + A^{-1}xy^T) \\
&= (\det A)\prod_{i=1}^{n}(1 + \lambda_i(A^{-1}xy^T)) \quad \text{(use Observation 2.10 in Lecture 1)} \\
&= (\det A)(1 + y^T A^{-1}x) \quad \text{(use Example 1.25)} \\
&= \det A + y^T((\det A)A^{-1})x = \det A + y^T(\operatorname{adj} A)x.
\end{aligned}$$

Cauchy's identity, $\det(A+xy^T) = \det A + y^T(\operatorname{adj} A)x$, is valid for any $A \in M_n$ by Theorem 3.11 in Lecture 1 and continuity (each entry of $\operatorname{adj} A$ and $\det A$ are multinomials in the entries of A).

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

Theorem 1.27. *Suppose that $A \in M_n$ is diagonalizable, let $\mu_1, ..., \mu_d$ be its distinct eigenvalues with respective multiplicities $n_1, ..., n_d$, let $S, T \in M_n$ be nonsingular, and suppose that $A = S\Lambda S^{-1}$, where Λ is a diagonal matrix of the form (2). Then*

(a) $A = T\Lambda T^{-1}$ if and only if $T = S(R_1 \oplus \cdots \oplus R_d)$ in which each $R_i \in M_{n_i}$ is nonsingular.

(b) If $S = [S_1 \cdots S_d]$ and $T = [T_1 \cdots T_d]$ are partitioned conformally to Λ, then $A = S\Lambda S^{-1} = T\Lambda T^{-1}$ if and only if for each $i = 1, ..., d$, the column space of S_i is the same as the column space of T_i.

(c) If A has n distinct eigenvalues and $S = [s_1 \cdots s_n]$ and $T = [t_1 \cdots t_n]$ are partitioned according to their columns, then $A = S\Lambda S^{-1} = T\Lambda T^{-1}$ if and only if there is a nonsingular diagonal matrix $R = \operatorname{diag}(r_1, ..., r_n)$ such that $T = SR$ if and only if, for each $i = 1, ..., n$, the column s_i is a nonzero scalar multiple of the corresponding column t_i.

Proof. (a) We have $S\Lambda S^{-1} = T\Lambda T^{-1}$ if and only if $(S^{-1}T)\Lambda = \Lambda(S^{-1}T)$ if and only if $S^{-1}T$ is block diagonal conformal to Λ, that is, if and only if

$S^{-1}T = R_1 \oplus \cdots \oplus R_d$ and each $R_i \in M_{n_i}$ is nonsingular. (See the argument in Theorem 1.16)
(b) Observe that $T_i = S_i R_i, i = 1, ..., d$ and R_i's are nonsingular by (a). This indicates that T_i is in the column space of S_i and alternatively, S_i is in the column space of T_i, since $S_i = T_i R_i^{-1}$ for each $i = 1, ..., d$. The converse follows from (a).
The assertion (c) is a special case of (a) and (b). □

Up to now, our discussions of similarity focus on matrices over the complex numbers $\mathbf{C}$. If, for instance, real matrices are similar via a complex matrix, are they similar via a real matrix? Is there a real version of Theorem 1.23 for commuting real matrices? The following lemma is the key to answering such questions.

Lemma 1.28. *Let $S \in M_n$ be nonsingular and let $S = C + iD$, where $C, D \in M_n(\mathbf{R})$. There is a real number τ such that $T = C + \tau D$ is nonsingular.*

Proof. If C is nonsingular, take $\tau = 0$. If C is singular, consider the polynomial $p(t) = \det(C + tD)$, which is not a constant (degree zero) polynomial, since $p(0) = \det C = 0 \neq \det S = p(i)$. Since $p(t)$ has only finitely many zeroes in the complex plane, there is a real τ such that $p(\tau) \neq 0$, so $C + \tau D$ is nonsingular. □

Theorem 1.29. *Let $\mathcal{F} = \{A_\alpha : \alpha \in \mathcal{I}\} \subset M_n(\mathbf{R})$ and $\mathcal{G} = \{B_\alpha : \alpha \in \mathcal{I}\} \subset M_n(\mathbf{R})$ be given families of real matrices. If there is a nonsingular $S \in M_n$ such that $A_\alpha = S B_\alpha S^{-1}$ for every $\alpha \in \mathcal{I}$, then there is a nonsingular $T \in M_n(\mathbf{R})$ such that $A_\alpha = T B_\alpha T^{-1}$ for every $\alpha \in \mathcal{I}$. In particular, two real matrices that are similar over $\mathbf{C}$ are similar over $\mathbf{R}$.*

Proof. Let $S = C + iD$ be nonsingular, where $C, D \in M_n(\mathbf{R})$. Lemma 1.28 ensures that there is a real number τ such that $T = C + \tau D$ is nonsingular. The similarity $A_\alpha = S B_\alpha S^{-1}$ is equivalent to the identity $A_\alpha(C + iD) = A_\alpha S = S B_\alpha = (C + iD) B_\alpha$. Equating the real and imaginary parts of this identity shows that $A_\alpha C = C B_\alpha$ and $A_\alpha D = D B_\alpha$, i.e., $A_\alpha(\tau D) = (\tau D) B_\alpha$, so $A_\alpha T = T B_\alpha$ and $A_\alpha = T B_\alpha T^{-1}$. □

An immediate consequence of the preceding theorem is a real version of Theorem 1.23.

Corollary 1.30. *Let $\mathcal{F} = \{A_\alpha : \alpha \in \mathcal{I}\} \subset M_n(\mathbf{R})$ be a family of real diagonalizable matrices with real eigenvalues. Then $\mathcal{F}$ is a commuting family if and only if there is a nonsingular real matrix T such that $T^{-1}A_\alpha T = \Lambda_\alpha$ is diagonal for every $A \in \mathcal{F}$. In particular, for any given $\alpha_0 \in \mathcal{I}$ and for any given ordering $\lambda_1, ..., \lambda_n$ of the eigenvalues of A_{α_0}, there is a nonsingular $T \in M_n(\mathbf{R})$ such that $T^{-1}A_{\alpha_0}T = \operatorname{diag}(\lambda_1, ..., \lambda_n)$ and $T^{-1}A_\alpha T$ is diagonal for every $\alpha \in \mathcal{I}$.*

Proof. ($\Rightarrow$) Apply Theorem 1.29 to the families $\mathcal{F} = \{A_\alpha : \alpha \in \mathcal{I}\}$ and $\mathcal{G} = \{\Lambda_\alpha : \alpha \in \mathcal{I}\}$ and use Theorem 1.23 for simultaneous diagonalizability.
($\Leftarrow$) Utilize Proposition 1.16. $\square$

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

Theorem 1.31 (Mirsky)**.** *Let an integer $n \geq 2$ and complex scalars $\lambda_1, ..., \lambda_n$ and $d_1, ..., d_n$ be given. There is an $A \in M_n$ with eigenvalues $\lambda_1, ..., \lambda_n$ and main diagonal entries $d_1, ..., d_n$ if and only if $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} d_i$. If $\lambda_1, ..., \lambda_n$ and $d_1, ..., d_n$ are all real and have the same sums, there is an $A \in M_n(\mathbf{R})$ with eigenvalues $\lambda_1, ..., \lambda_n$ and main diagonal entries $d_1, ..., d_n$.*

Proof. ($\Rightarrow$) We know that $\operatorname{tr} A = E_1(A) = S_1(A)$ for any $A \in M_n$ (Theorem 3.10 in Lecture 1), which establishes the necessity of the stated condition.
($\Leftarrow$) If $k \geq 2$ and if $\lambda_1, ..., \lambda_k$ and $d_1, ..., d_k$ are any given complex scalars such that $\sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} d_i$, we claim that the upper bidiagonal matrix

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

is similar to a matrix with diagonal entries $d_1, ..., d_k$; that matrix has the property asserted. Let $L(s,t) = \begin{bmatrix} 1 & 0 \\ s-t & 1 \end{bmatrix}$, so $L(s,t)^{-1} = \begin{bmatrix} 1 & 0 \\ t-s & 1 \end{bmatrix}$. Consider first the case $k=2$, so $\lambda_1+\lambda_2=d_1+d_2$. Compute the similarity

$$\begin{aligned} L(\lambda_1,d_1)T(\lambda_1,\lambda_2)L(\lambda_1,d_1)^{-1} &= \begin{bmatrix} 1 & 0 \\ \lambda_1-d_1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ d_1-\lambda_1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} d_1 & \bigstar \\ \bigstar & \lambda_1+\lambda_2-d_1 \end{bmatrix} = \begin{bmatrix} d_1 & \bigstar \\ \bigstar & d_2 \end{bmatrix} \end{aligned}$$

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

We proceed by induction. In order to clarify the inductive procedures, we first work on the inductive step $k=2\Rightarrow k=3$. If $\sum_{i=1}^{3}\lambda_i=\sum_{i=1}^{3}d_i$, compute the similarity

$$\begin{aligned} &\begin{bmatrix} L(\lambda_1,d_1) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} L(d_1,\lambda_1) & 0 \\ 0 & 1 \end{bmatrix} \\ &\qquad = \begin{bmatrix} d_1 & 1 & 0 \\ (d_1-\lambda_1)(\lambda_2-d_1) & \lambda_1+\lambda_2-d_1 & 1 \\ 0 & 0 & \lambda_3 \end{bmatrix} \\ &\qquad = \begin{bmatrix} d_1 & \bigstar \\ \bigstar & D \end{bmatrix}, \end{aligned} \tag{7}$$

where $D=T(\lambda_1+\lambda_2-d_1,\lambda_3)$. By the preceding argument (induction hypothesis) on $k=2$, we can apply the similarity transformation $S^{-1}=\begin{bmatrix} 1 & 0 \\ 0 & L(d_3,\lambda_3) \end{bmatrix}$ on the resulting matrix (7), since

$$\begin{aligned} L(d_3,\lambda_3)DL(d_3,\lambda_3)^{-1} &= \begin{bmatrix} 1 & 0 \\ d_3-\lambda_3 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1+\lambda_2-d_1 & 1 \\ 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \lambda_3-d_3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} d_2 & \bigstar \\ \bigstar & d_3 \end{bmatrix}. \end{aligned}$$

This completes our claim from $k=2$ to $k=3$.

Assume that our claim has been proved for some $k\geq 2$ and that $\sum_{i=1}^{k+1}\lambda_i=\sum_{i=1}^{k+1}d_i$. Partition $T(\lambda_1,...,\lambda_{k+1})=[T_{ij}]_{i,j=1}^{2}$, where $T_{11}=T(\lambda_1,\lambda_2), T_{12}=$

$E_2, T_{21} = 0$, and $T_{22} = T(\lambda_3, ..., \lambda_{k+1})$, with $E_2 = [e_2\ 0 \cdots 0] \in M_{2,k-1}$ and $e_2 = [0\ 1]^T \in \mathbf{C}^2$. Let $\mathcal{L} = L(\lambda_1, d_1) \oplus I_{k-1}$ and compute

$$\begin{aligned}\mathcal{L}T(\lambda_1, ..., \lambda_{k+1})\mathcal{L}^{-1}\\ &= \begin{bmatrix} L(\lambda_1, d_1) & 0 \\ 0 & I_{k-1} \end{bmatrix} \begin{bmatrix} T(\lambda_1, \lambda_2) & E_2 \\ 0 & T(\lambda_3, ..., \lambda_{k+1}) \end{bmatrix} \begin{bmatrix} L(d_1, \lambda_1) & 0 \\ 0 & I_{k-1} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} d_1 & \bigstar \\ \bigstar & \lambda_1 + \lambda_2 - d_1 \end{bmatrix} & E_2 \\ 0 & T(\lambda_3, ..., \lambda_{k+1}) \end{bmatrix} \\ &= \begin{bmatrix} d_1 & \bigstar \\ \bigstar & T(\lambda_1 + \lambda_2 - d_1, \lambda_3, ..., \lambda_{k+1}) \end{bmatrix} = \begin{bmatrix} d_1 & \bigstar \\ \bigstar & D \end{bmatrix}\end{aligned}$$

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

If $\lambda_1, ..., \lambda_n$ and $d_1, ..., d_n$ are all real, all of the matrices and similarities in the preceding constructions are real by Theorem 1.29. □

2 Left and right eigenvectors and geometric multiplicity (Page 75-80, Page 6)

The eigenvectors of a matrix are important not only for their role in diagonalization but also for their utility in a variety of applications. For example, Principal Component Analysis (PCA) is a statistical eigenvector-based method to fit an n-dimensional ellipsoid to the data, where each axis of the ellipsoid represents a principal component and all principal components are linearly correlated[4]. However, all the eigenvectors of a matrix may not span the whole vector space in some cases. In this section we characterize these defective situations by introducing the concept of geometric multiplicity of an eigenvalue. Moreover, the connections between left and right eigenvalues of a matrix will also be investigated. We begin with an important observation about eigenvalues.

Observation 2.1. *Let $A \in M_n$. (a) The eigenvalues of A and A^T are the same. (b) The eigenvalues of A^* are the complex conjugates of the eigenvalues of A.*

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

Proposition 2.2. *Let $A \in M_n$. The set of all eigenvectors associated with a particular eigenvalue $\lambda \in \sigma(A)$, together with the zero vector, is a* subspace *of $\mathbf{C}^n$.*

Proof. If $x, y \in \mathbf{C}^n$ are both eigenvectors of A associated with the eigenvalue λ and any $\alpha \in \mathbf{C}$, then $A(\alpha x + y) = \alpha Ax + Ay = \lambda(\alpha x + y)$, showing that $\alpha x + y$ is also an eigenvector of A. The set of all eigenvectors associated with λ, together with the zero vector, is closed under the operations of vector addition and scalar multiplication and is thus a subspace of $\mathbf{C}^n$. $\square$

The subspace described in the preceding proposition is the *null space* of $A - \lambda I$, that is, the solution set of the homogeneous linear system $(A - \lambda I)x = 0$. By the *rank-nullity theorem*, i.e., $\dim(\operatorname{range} A) + \dim(\operatorname{nullspace} A) = \operatorname{rank} A + \operatorname{nullity} A = n$, we conclude that the dimension of this subspace is $n - \operatorname{rank}(A - \lambda I)$. In reality, we have a formal terminology for this subspace.

Definition 2.3 (eigenspace)**.** *Let $A \in M_n$. For a given $\lambda \in \sigma(A)$, the set of all vectors $x \in \mathbf{C}^n$ satisfying $Ax = \lambda x$ is called the* eigenspace *of A associated with the eigenvalue λ, denoted by $V_\lambda = \{x \in \mathbf{C}^n : Ax = \lambda x\}$. Every nonzero element of this eigenspace is an eigenvector of A associated with λ.*

Example 2.4. Even though A and A^T have the same eigenvalues, their eigenspaces associated with a given eigenvalue can be different. For example, let $A = \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix}$. Then the (one-dimensional) eigenspace of A associated with the eigenvalue 2 is spanned by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, while the eigenspace of A^T associated with the eigenvalue 2 is spanned by $\begin{bmatrix} 1 \\ -\frac{3}{2} \end{bmatrix}$.

Example 2.5. The eigenspace of A associated with an eigenvalue λ is an A-invariant subspace, since for any $x \in V_\lambda$ we still have $Ax = \lambda x \in V_\lambda$.

Conversely, an A-invariant subspace need not be an eigenspace of A, like $\{0\}$ and $\mathbf{C}^n$. As for a nontrivial counterexample, let $V_{\lambda_1}, V_{\lambda_2}$ be two eigenspace of A, where $\lambda_1 \neq \lambda_2$. The direct sum $V = V_{\lambda_1} \oplus V_{\lambda_2}$ is an A-invariant subspace, since for any $x \in V$, we can factorize x into $x = y + z$, where $y \in V_{\lambda_1}$ and $z \in V_{\lambda_2}$, and thus $Ax = A(y+z) = \lambda_1 y + \lambda_2 z \in V$. Nevertheless, $V = V_{\lambda_1} \oplus V_{\lambda_2}$ can never be an eigenspace of A.

It is easy to check that a *minimal* A-invariant subspace (an A-invariant subspace that contains no strictly lower-dimensional, nonzero A-invariant subspace) W is the span of a *single* eigenvector of A, i.e., $\dim W = 1$.

To formalize the dimension of an eigenspace of a matrix and characterize its diagonalizability, we make the following definitions.

Definition 2.6. *Let $A \in M_n$ and let λ be an eigenvalue of A.*

(a) The dimension of the eigenspace of A associated with λ is the geometric multiplicity *of λ.*

(b) The multiplicity of λ as a zero of the characteristic polynomial of A is the algebraic multiplicity *of λ.*

If the term multiplicity *is used without qualification in reference to λ, it means the algebraic multiplicity. We say that λ is* simple *if its algebraic multiplicity is 1; it is* semisimple *if its algebraic and geometric multiplicities are equal.*

Example 2.7. We have the following statements about the respective matrices and their eigenvalue $\lambda = 1$:

(a) $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$: geometric multiplicity = algebraic multiplicity = 1; λ is simple.

(b) $A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$: geometric multiplicity = algebraic multiplicity = 2; λ is semisimple.

(c) $A_3 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$: geometric multiplicity = 1; algebraic multiplicity = 2.

It can be very useful to think of the geometric multiplicity of an eigenvalue λ of $A \in M_n$ in more than one way: It is the dimension of the nullspace of $A - \lambda I$, which is equal to $n - \operatorname{rank}(A - \lambda I)$. Meanwhile, it is the maximum number of linearly independent eigenvectors associated with λ. The inequality between the geometric and algebraic multiplicity of an eigenvalue can be deduced from two different viewpoints.

On one hand, Theorem 3.12 in Lecture 1 states that for any $\lambda \in \sigma(A)$ its algebraic multiplicity k satisfies $k \geq n - \text{rank}(A - \lambda I)$, which is the corresponding geometric multiplicity, with the equality for $k = 1$.

On the other hand, Theorem 1.8 in this lecture points out that A is similar to a block matrix $\begin{bmatrix} \Lambda & C \\ 0 & D \end{bmatrix}$ with $\Lambda = \text{diag}(\lambda, ..., \lambda) \in M_m$, $D \in M_{n-m}$ if A has totally m linearly independent eigenvectors associated with λ. Then the characteristic polynomial of A becomes $p_A(t) = p_\Lambda(t)p_D(t)$ and $p_D(\lambda)$ may still be 0, yielding that the algebraic multiplicity of $\lambda \geq m =$ its geometric multiplicity. If the algebraic multiplicity is 1, then $p_A(t) = (t - \lambda)p_D(t)$ and $p_D(\lambda) \neq 0$, so there exists an $x \in \mathbf{C}^n, x \neq 0$ such that $Ax = \lambda x$.

Furthermore, information about eigenvalues of principal submatrices can refine the basic observation that the algebraic multiplicity of an eigenvalue cannot be less than its geometric multiplicity.

Theorem 2.8. *Let $A \in M_n$ and $\lambda \in \mathbf{C}$ be given, and let $k \geq 1$ be a given positive integer. Consider the following three statements:*

(a) λ is an eigenvalue of A with geometric multiplicity at least k.

(b) For each $m = n - k + 1, ..., n$, λ is an eigenvalue of every m-by-m principal submatrix of A.

(c) λ is an eigenvalue of A with algebraic multiplicity at least k.

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

Proof. (a)$\Rightarrow$(b): If λ is an eigenvalue of A with geometric multiplicity at least k, then $\text{rank}(A - \lambda I) \leq n - k$. Assume that $m > n - k$. Then every m-by-m minor of $A - \lambda I$ is zero. In particular, every principal m-by-m minor of $A - \lambda I$ is zero, so every m-by-m principal submatrix of $A - \lambda I$ is singular. Thus, λ is an eigenvalue of every m-by-m principal submatrix of A.
(b)$\Rightarrow$(c): If λ is an eigenvalue of every m-by-m principal submatrix of A for each $m \geq n - k + 1$, then every principal minor of $A - \lambda I$ of size at least $n - k + 1$ is zero, so each principal minor sum $E_j(A - \lambda I) = 0$ for all $j \geq n - k + 1$ (Definition 3.7 in Lecture 1). Since

$$p_{A-\lambda I}(t) = t^n - E_1(A-\lambda I)t^{n-1} + \cdots + (-1)^{n-1}E_{n-1}(A-\lambda I)t + (-1)^n E_n(A-\lambda I)$$

and $p^{(k)}_{A-\lambda I}(0) = k!(-1)^{n-k}E_{n-k}(A-\lambda I), k = 0, 1, ..., n-1$, we find that

$$p^{(i)}_{A-\lambda I}(0) = 0 \text{ for } i = 0, 1, ..., k-1.$$

But $p_{A-\lambda I}(t) = p_A(t+\lambda)$, so $p^{(i)}_A(\lambda) = 0$ for $i = 0, 1, ..., k-1$; that is, λ is a zero of $p_A(t)$ with multiplicity at least k. □

Definition 2.9. *Let $A \in M_n$.*

(a) A is defective *if the geometric multiplicity of some eigenvalue of A is strictly less than its algebraic multiplicity.*

(b) A is nondefective *if the geometric multiplicity of each eigenvalue of A is the same as its algebraic multiplicity.*

(c) A is nonderogatory *if each eigenvalue of A has geometric multiplicity 1; otherwise, it is* derogatory*.*

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

Up to now, we only come across the *right eigenvector*, that is, left multiplication by a matrix has the same effect as scalar multiplication. Symmetrically, we have the following definition.

Definition 2.10. *A nonzero vector $y \in \mathbf{C}^n$ is a* left eigenvector *of $A \in M_n$ associated with an eigenvalue λ of A if $y^*A = \lambda y^*$.*

Remark. (a) When the context does not require distinction we continue to call x an eigenvector.
(b) Given a right eigenvector of A associated with an eigenvalue λ, we can always obtain a left eigenvector of A associated with λ. By Observation 2.1, $\lambda \in \sigma(A)$ if and only if $\bar{\lambda} \in \sigma(A^*)$. Thus, we find that $A^*y = \bar{\lambda}y$, i.e., $y^*A = \lambda y^*$.

Example 2.11. Consider $A = \begin{bmatrix} -i & 0 \\ 1 & -i \end{bmatrix}, y = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. A direct computation shows that $y^*A = -iy^*$ and y is a left eigenvector of A associated with $-i$.

Example 2.12. Let $x \in \mathbf{C}^n$ be nonzero, let $A \in M_n$ and suppose that $Ax = \lambda x$. If $x^*A = \mu x^*$, then $\lambda = \mu$. This is because $\mu x^*x = (x^*A)x = x^*Ax = x^*(Ax) = x^*(\lambda x) = \lambda x^*x$ and $x^*x \neq 0$.

Moreover, a left eigenvector y associated with an eigenvalue λ of $A \in M_n$ is a right eigenvector of A^* associated with $\bar{\lambda}$, since $y^*A = \lambda y^*$ implies that $(y^*A)^* = A^*y = \bar{\lambda}y$. Likewise, $\bar{y}$ is a right eigenvector of A^T associated with λ.

One should not dismiss left eigenvectors as merely a parallel theoretical alternative to right eigenvectors. Each type of eigenvector can convey different information about a matrix, and it can be very useful to know how the two types of eigenvectors interact.

Example 2.13. Suppose that $A \in M_n$ is diagonalizable, S is nonsingular, and $S^{-1}AS = \Lambda = \text{diag}(\lambda_1, ..., \lambda_n)$. Partition $S = [x_1 \cdots x_n]$ and $(S^{-1})^* = S^{-*} = [y_1 \cdots y_n]$ according to their columns. The identity $AS = S\Lambda$ tells us that each column x_j of S is a right eigenvector of A associated with the eigenvalue λ_j. Furthermore, $S^{-1}A = \Lambda S^{-1}$ indicates that $(S^{-*})^*A = \Lambda(S^{-*})^*$, so each column y_j of S^{-*} is a left eigenvector of A associated with the eigenvalue λ_j. Finally, $y_j^*x_j = 1$ for each $j = 1, ..., n$ and $y_i^*x_j = 0$ whenever $i \neq j$, where we use the fact that

$$S^{-1}S = \begin{bmatrix} y_1^* \\ \vdots \\ y_n^* \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = I_n.$$

We next examine a version of the results in the foregoing example 2.13 for matrices that are not necessarily diagonalizable.

Theorem 2.14. *Let $A \in M_n$, nonzero vectors $x, y \in \mathbf{C}^n$, and scalars $\lambda, \mu \in \mathbf{C}$ be given. Suppose that $Ax = \lambda x$ and $y^*A = \mu y^*$.*

(a) If $\lambda \neq \mu$, then $y^*x = 0$. This is called the principle of biorthogonality.

(b) If $\lambda = \mu$ and $y^*x \neq 0$, then there is a nonsingular $S \in M_n$ of the form $S = [x\ S_1]$ such that $S^{-*} = [y/(x^*y)\ Z_1]$ and

$$A = S \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix} S^{-1}, \quad B \in M_{n-1}. \tag{8}$$

Conversely, if A is similar to a block matrix of the form (8), then it has a nonorthogonal pair of left and right eigenvectors associated with the eigenvalue λ.

Proof. (a) By $Ax = \lambda x$ and $y^*A = \mu y^*$, we manipulate y^*Ax in two ways:

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

Since $\lambda \neq \mu$, $\lambda y^*x = \mu y^*x$ only if $y^*x = 0$.
(b) Suppose that $Ax = \lambda x$, $y^*A = \lambda y^*$, and $y^*x \neq 0$. If we replace y by $y/(x^*y)$, we may assume that $y^*x = 1$. Let the columns of $S_1 \in M_{n,n-1}$ be any basis for the orthogonal complement (see Definition 2.3 in Lecture 3) of y (so $y^*S_1 = 0$) and consider $S = [x\ S_1] \in M_n$. Suppose that $Sz = 0$ and $z = [z_1\ \zeta^T]^T$ with $\zeta \in \mathbf{C}^{n-1}$. Then

$$0 = y^*Sz = y^*(z_1x + S_1\zeta) = z_1(y^*x) + (y^*S_1)\zeta = z_1$$

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

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

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

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

which verifies (8).

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

$S^{-1}S = I_n$ shows that $y^*x = 1$, that is, x is nonorthogonal to y. The first column of the identity

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

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

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

tells us that $y^*A = \lambda y^*$. □

The eigenvalues of a matrix are unchanged by similarity, while its eigenvectors transform under similarity in a simple way.

Theorem 2.15. *Let $A, B \in M_n$ and suppose that $B = S^{-1}AS$ for some nonsingular S. If $x \in \mathbf{C}^n$ is a right eigenvector of B associated with an eigenvalue λ, then Sx is a right eigenvector of A associated with λ. If $y \in \mathbf{C}^n$ is a left eigenvector of B associated with λ, then $S^{-*}y$ is a left eigenvector of A associated with λ.*

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

An eigenvalue λ with geometric multiplicity 1 can have algebraic multiplicity 2 or more, but this can happen only if the left and right eigenvectors associated with λ are orthogonal. If λ has algebraic multiplicity 1, however, then it has geometric multiplicity 1; left and right eigenvectors associated with λ can never be orthogonal. Our approach to these results relies on the following lemma.

Lemma 2.16. *Let $A \in M_n, \lambda \in \mathbf{C}$, and nonzero vectors $x, y \in \mathbf{C}^n$ be given. Suppose that λ has geometric multiplicity 1 as an eigenvalue of A, $Ax = \lambda x$, and $y^*A = \lambda y^*$. Then there is a nonzero $\gamma \in \mathbf{C}$ such that* $\operatorname{adj}(\lambda I - A) = \gamma xy^*$.

Proof. We have $\operatorname{rank}(\lambda I - A) = n - 1$. On one hand, by the identity $\operatorname{adj}(\lambda I - A) \cdot (\lambda I - A) = \det(\lambda I - A) \cdot I = 0$, we know that $n - \operatorname{rank}\operatorname{adj}(\lambda I - A) \geq n - 1$, i.e., $\operatorname{rank}\operatorname{adj}(\lambda I - A) \leq 1$. On the other hand, since $\operatorname{rank}(\lambda I - A) = n - 1$, some $(n-1)$-by-$(n-1)$ principal submatrices are nonzero and $\operatorname{rank}\operatorname{adj}(\lambda I -$

$A) \geq 1$. Thus, $\operatorname{rank}\operatorname{adj}(\lambda I - A) = 1$, that is, $\operatorname{adj}(\lambda I - A) = \xi\eta^*$ for some nonzero $\xi, \eta \in \mathbf{C}^n$.
But $(\lambda I - A)(\operatorname{adj}(\lambda I - A)) = \det(\lambda I - A)I = 0$, so $(\lambda I - A)\xi\eta^* = 0$ and $(\lambda I - A)\xi = 0$, which implies that $\xi = \alpha x$ for some nonzero scalar α. Using the identity $(\operatorname{adj}(\lambda I - A))(\lambda I - A) = 0$, we also have that $\xi\eta^*(\lambda I - A) = 0$ and $\eta^*(\lambda I - A) = 0$, namely, $\eta = \beta y$ for some nonzero scalar β. Thus, $\operatorname{adj}(\lambda I - A) = \alpha\bar{\beta}xy^*$ and choose $\gamma = \alpha\bar{\beta}$. □

Theorem 2.17. *Let $A \in M_n, \lambda \in \mathbf{C}$, and nonzero vectors $x, y \in \mathbf{C}^n$ be given. Suppose that λ is an eigenvalue of A, $Ax = \lambda x$ and $y^*A = \lambda y^*$.*

(a) If λ has algebraic multiplicity 1, then $y^*x \neq 0$.

(b) If λ has geometric multiplicity 1, then it has algebraic multiplicity 1 if and only if $y^*x \neq 0$.

Proof. In both cases (a) and (b), λ has geometric multiplicity 1 (Theorem 3.12 in Lecture 1); the preceding lemma tells us that there is a nonzero $\gamma \in \mathbf{C}$ such that $\operatorname{adj}(\lambda I - A) = \gamma xy^*$. Then $p_A(\lambda) = 0$ and $p_A'(\lambda) = \operatorname{tr}\operatorname{adj}(\lambda I - A) = \gamma y^*x$ (Proposition 3.8 in Lecture 1). In (a) we assume that the algebraic multiplicity is 1, so $p_A'(\lambda) \neq 0$ and hence $y^*x \neq 0$. In (b) we assume that $y^*x \neq 0$, so $p_A'(\lambda) \neq 0$ and hence the algebraic multiplicity is 1. □

References

[1] Roger A. Horn, Charles R. Johnson (2012) *Matrix Analysis, Second Edition.* Cambridge University Press.

[2] Wikipedia: The Free Encyclopedia ***Power iteration.*** URL: https://en.wikipedia.org/wiki/Power_iteration Retrieved 12 April, 2018.

[3] Wikipedia: The Free Encyclopedia ***Eigenvalues and eigenvectors.*** URL: https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors Retrieved 12 April, 2018.

[4] Wikipedia: The Free Encyclopedia ***Principal component analysis.*** URL: https://en.wikipedia.org/wiki/Principal_component_analysis Retrieved 12 April, 2018.