Lecture 5. Diagonalization of Matrices
October 2022

Invariant Subspaces

DEFINITION 5.1 A subspace U of a vector space V is called an **invariant subspace** of a linear operator $f : V \to V$ if $f(U) \subset U$, that is, if for any $\mathbf{x} \in U$ we have $f(\mathbf{x}) \in U$. This subspace is also called f-**invariant**.

DEFINITION 5.2 Let $f : V \to V$ be a linear operator. If U is any invariant subspace of V, then $f : U \to U$ is a linear operator on the subspace U, called the **restriction** of f to U and denoted by $f|_U$.

LEMMA 5.3 Let $f : V \to V$ be a linear operator. Then:

1. $\{\mathbf{0}\}$ and V are invariant subspaces.
2. Both $\operatorname{Ker} f$ and $\operatorname{Im} f = f(V)$ are invariant subspaces.
3. If U and W are invariant subspaces, so are $f(U)$, $U \cap W$, and $U + W$.

EXAMPLE 5.4 Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be a rotation about z-axis through an angle of θ, $0 < \theta < \pi$. Then, except for $\{\mathbf{0}\}$ and $\mathbb{R}^3$, there are two invariant subspaces: they are z-axis and xy-plane.

EXAMPLE 5.5 If $\mathbf{v}$ is an eigenvector of a linear operator $f : V \to V$, then $\operatorname{span}\{\mathbf{v}\}$ is an invariant subspace.

THEOREM 5.6 Let $f : V \to V$ be a linear operator with $\dim V = n$ and suppose that U is any invariant subspace of V. Let $B_1 = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ be any basis of U extended to a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_k, \mathbf{b}_{k+1}, \dots, \mathbf{b}_n\}$ of V in any way. Then the matrix A of f with respect to B has the block triangular form

$$A = \begin{pmatrix} A_1 & Y \\ 0 & Z \end{pmatrix},$$

where Z is $(n-k) \times (n-k)$ and A_1 is the matrix of the restriction of f to U.

Proof follows from the fact that

$$f(\mathbf{b}_i) = t_1\mathbf{b}_1 + t_2\mathbf{b}_2 + \cdots + t_k\mathbf{b}_k + 0\mathbf{b}_{k+1} + \cdots + 0\mathbf{b}_n, \qquad 1 \leqslant i \leqslant k. \quad \square$$

COROLLARY 5.7 Let $f : V \to V$ be a linear operator with $\dim V = n$. Suppose $V = U_1 \oplus U_2$, where both U_1 and U_2 are invariant. If $B_1 = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ and $B_2 = \{\mathbf{b}_{k+1}, \dots, \mathbf{b}_n\}$ are bases of U_1 and U_2 respectively, then with respect to the basis $B = B_1 \cup B_2$, the matrix A of f has the block diagonal form

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

where A_1 and A_2 are the matrices of the restrictions of f to U_1 and U_2 respectively.

THEOREM 5.8 A linear operator $f : V \to V$ has a block diagonal matrix

$$A = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_s \end{pmatrix}$$

if and only if V is a direct sum of invariant subspases U_i, $1 \leqslant i \leqslant s$, i.e., $V = U_1 \oplus \cdots \oplus U_s$, and the basis of V is a union of bases of U_i's.

THEOREM 5.9 If operators f and g on a vector space V commute, i.e., $f \circ g = g \circ f$, then

1. $\operatorname{Ker} f$ and $\operatorname{Im} f$ are g-invariant;
2. $\operatorname{Ker} g$ and $\operatorname{Im} g$ are f-invariant.

Proof. We will prove the first statement. Let $\mathbf{x} \in \operatorname{Ker} f$. Then $f(\mathbf{x}) = \mathbf{0}$ and, therefore, $f(g(\mathbf{x})) = g(f(\mathbf{x})) = \mathbf{0}$, which implies $g(\mathbf{x}) \in \operatorname{Ker} f$. This proves that $\operatorname{Ker} f$ is invariant under g.

If $\mathbf{y} \in \operatorname{Im} f$, then there exists a vector $\mathbf{x} \in V$ such that $f(\mathbf{x}) = \mathbf{y}$. So we have: $g(\mathbf{y}) = g(f(\mathbf{x})) = f(g(\mathbf{x})) \in \operatorname{Im} f$. This proves that $\operatorname{Im} f$ is invariant under g. □

REMARK 5.10 Since any linear operator f always commutes with $P(f)$, where $P(f)$ is a polynomial of f, Theorem 5.9 applies.

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

THEOREM 5.11 Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ of a linear operator f. Then $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ is a linearly independent set.

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

$$t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_{k+1}\mathbf{x}_{k+1} = \mathbf{0}. \tag{1}$$

We must show that each $t_i = 0$. Apply f to both sides of (1) and use the fact that $f(\mathbf{x}_i) = \lambda_i \mathbf{x}_i$ to get

$$t_1\lambda_1\mathbf{x}_1 + t_2\lambda_2\mathbf{x}_2 + \cdots + t_{k+1}\lambda_{k+1}\mathbf{x}_{k+1} = \mathbf{0}. \tag{2}$$

If we multiply (1) by λ_1 and subtract the result from (2), the first terms cancel and we obtain

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

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

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

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

Theorem 5.11 gives a useful condition for when a matrix is diagonalizable.

COROLLARY 5.12 If V is n-dimensional and $f : V \to V$ is a linear operator with n distinct eigenvalues, then f is diagonalizable.

Proof. Choose one eigenvector for each of the n distinct eigenvalues. Then these eigenvectors are independent by Theorem 5.11, and so are a basis of V. $\square$

EXAMPLE 5.13 Show that

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ -1 & 1 & 0 \end{pmatrix}$$

is diagonalizable.

Solution. A routine computation shows that $p_A(\lambda) = (1-\lambda)(3-\lambda)(-1-\lambda)$ and so has distinct eigenvalues 1, 3, and -1. Hence Corollary 5.12 applies. $\square$

Algebraic and Geometric Multiplicities of Eigenvalues

DEFINITION 5.14 The **algebraic multiplicity** $\mu(\lambda_i)$ of the eigenvalue λ_i of f is its multiplicity as a root of the characteristic polynomial, that is, the largest integer $m \geqslant 1$ such that $(\lambda - \lambda_i)^m$ divides evenly that polynomial.

DEFINITION 5.15 Let $f : V \to V$ be a linear operator and 1_V be an identity operator. If λ_i is an eigenvalue of f, then the subspace $E_{\lambda_i} = \operatorname{Ker}(f - \lambda_i 1_V) \subset V$ is called the **eigenspace** of f associated with eigenvalue λ_i.

The eigenspace corresponding to λ_i can also be defined by

$$E_{\lambda_i} = \{\mathbf{x} \in V \mid f(\mathbf{x}) = \lambda_i \mathbf{x}\}.$$

DEFINITION 5.16 The dimension of the eigenspace E_{λ_i} associated with λ_i, or equivalently the maximum number of linearly independent eigenvectors associated with λ_i, is referred to as a **geometric multiplicity** $\gamma(\lambda_i)$ of the eigenvalue λ_i of f.

Because of the definition of eigenvalues and eigenvectors, geometric multiplicity of an eigenvalue must be at least one, that is, each eigenvalue has at least one associated eigenvector. Furthermore, a geometric multiplicity of an eigenvalue cannot exceed its algebraic multiplicity. Additionally, recall that an algebraic multiplicity cannot exceed n. Now we are going to prove the corresponding result.

LEMMA 5.17 Let λ_i be an eigenvalue of a linear operator f. Then

$$1 \leqslant \gamma(\lambda_i) \leqslant \mu(\lambda_i) \leqslant n.$$

Proof. Write $\gamma(\lambda_i) = d$. Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d\}$ be a basis of E_{λ_i}. Then this basis can be extended to a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_d, \ldots, \mathbf{x}_n\}$ of V and by Theorem 5.6 the matrix of f has block form

$$\begin{pmatrix} \lambda_i I_d & B \\ 0 & A_1 \end{pmatrix}$$

in block form, where I_d denotes the $d \times d$-identity matrix. Find the characteristic polynomial of f:

$$p(\lambda) = \det \begin{pmatrix} (\lambda_i - \lambda) I_d & -B \\ 0 & A_1 - \lambda I_{n-d} \end{pmatrix} = \det[(\lambda_i - \lambda) I_d] \cdot \det(A_1 - \lambda I_{n-d}).$$

Therefore, $d \leqslant \mu(\lambda_i)$, because $\mu(\lambda_i)$ is the highest power of $(\lambda - \lambda_i)$ that divides $p(\lambda)$. $\square$

It is impossible to ignore the question when equality holds in Lemma 5.17 for each eigenvalue λ.

DEFINITION 5.18 We say that a polynomial

$$P_n(x) = a_n x^n + \cdots + a_1 x + a_0, \qquad n > 0, \ a_n \neq 0$$

with $a_i \in \mathbb{R}$, $i = 0, \ldots, n$, **factors completely** over $\mathbb{R}$ if it can be represented as

$$P_n(x) = a_n (x - x_1)(x - x_2) \ldots (x - x_n),$$

where the x_i are real numbers (not necessarily distinct).

EXAMPLE 5.19 Consider two matrices

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Show that $p_B(\lambda)$ factors completely over $\mathbb{R}$, but $p_A(\lambda)$ does not.

THEOREM 5.20 (Diagonalization Theorem) The following are equivalent for a linear operator $f : V \to V$ for which $p(\lambda)$ factors completely.

(a) f is diagonalizable.
(b) $\gamma(\lambda_i) = \mu(\lambda_i)$ for every eigenvalue λ_i of f.

Proof. Let V be n-dimensional and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of f. For each i, let m_i and d_i denote the algebraic and geometric multiplicities of λ_i, respectively. Then

$$p(\lambda) = (-1)^n (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \ldots (\lambda - \lambda_k)^{m_k}$$

so $m_1 + \cdots + m_k = n$ because $p(\lambda)$ has degree n. Moreover, $d_i \leqslant m_i$ for each i by Lemma 5.17.

$\boxed{\text{(a)} \Longrightarrow \text{(b)}}$ By (a), V has a basis of n eigenvectors of f, so d_i of them lie in E_{λ_i} for each i. Since $\dim E_{\lambda_i} = d_i$, then

$$n = d_1 + \cdots + d_k \leqslant m_1 + \cdots + m_k = n.$$

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

$\boxed{\text{(b)} \implies \text{(a)}}$ Let B_i denote a basis of E_{λ_i} for each i. Each B_i contains m_i vectors by (b), and the B_i are pairwise disjoint (the λ_i are distinct):

$$B_1 = \{\mathbf{b}_1, \dots, \mathbf{b}_{m_1}\}, \quad \dots, \quad B_k = \{\mathbf{b}_{n-m_k+1}, \dots, \mathbf{b}_n\}.$$

Let $B = B_1 \cup \cdots \cup B_k$. Since B contains n vectors, it suffices to show that B is linearly independent (then B is a basis of V).

Suppose a linear combination of the vectors in B vanishes:

$$t_1\mathbf{b}_1 + \cdots + t_{m_1}\mathbf{b}_{m_1} + \quad \dots \quad + t_{n-m_k+1}\mathbf{b}_{n-m_k+1} + \cdots + t_n\mathbf{b}_n = \mathbf{0} \tag{3}$$

Let $\mathbf{y}_i$ denote the sum of all terms that come from B_i. Then we can rewrite (3) as

$$(t_1\mathbf{b}_1 + \cdots + t_{m_1}\mathbf{b}_{m_1}) + \cdots + (t_{n-m_k+1}\mathbf{b}_{n-m_k+1} + \cdots + t_n\mathbf{b}_n) = \mathbf{y}_1 + \cdots + \mathbf{y}_k = \mathbf{0}.$$

Since $\mathbf{y}_i$ lies in E_{λ_i} for each i, it follows that the nonzero $\mathbf{y}_i$ are independent by Theorem 5.11 (as the λ_i are distinct). The sum of the $\mathbf{y}_i$ is zero, thus $\mathbf{y}_i = \mathbf{0}$ for each i. Hence all coefficients of terms in $\mathbf{y}_i$ are zero (because B_i is independent). Since this holds for each i, it shows that B is independent. This yields $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$. By Theorem 5.8 relative to B, operator f has a block diagonal form. But each block is itself diagonal, since f acts on E_{λ_i} as a scalar multiplication by λ_i. This completes the proof. $\square$

Symmetric Matrices

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

DEFINITION 5.21 Let $\bar{z}$ denote the conjugate of a complex number z. If A is a complex matrix, the **conjugate matrix** $\bar{A}$ is defined to be the matrix obtained from A by conjugating every entry. Thus, if $A = (z_{ij})$, then $\bar{A} = (\bar{z}_{ij})$.

LEMMA 5.22 If A and B are two complex matrices, then

$$\overline{A+B} = \bar{A} + \bar{B}, \qquad \overline{AB} = \bar{A}\bar{B}, \qquad \overline{\lambda A} = \bar{\lambda}\bar{A}$$

hold for all complex scalars λ.

Proof follows from the fact that $\overline{z+w} = \bar{z} + \bar{w}$ and $\overline{zw} = \bar{z}\bar{w}$ hold for all complex numbers z and w. $\square$

THEOREM 5.23 Let A be a real symmetric matrix. If λ is any complex eigenvalue of A, then λ is real.

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

If we write $\mathbf{x} = (z_1\ z_2\ \ldots\ z_n)^T$, where the z_i are complex numbers, we have

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

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

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

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

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

as required. $\square$