

5.4 INVARIANT SUBSPACES AND THE CAYLEY–HAMILTON THEOREM

In Section 5.1, we observed that if v is an eigenvector of a linear operator T , then T maps the span of $\{v\}$ into itself. Subspaces that are mapped into themselves are of great importance in the study of linear operators (see, e.g., Exercises 28–32 of Section 2.1).

Definition. Let T be a linear operator on a vector space V . A subspace W of V is called a **T -invariant subspace** of V if $T(W) \subseteq W$, that is, if $T(v) \in W$ for all $v \in W$.

Example 1

Suppose that T is a linear operator on a vector space V . Then the following subspaces of V are T -invariant:

1. $\{0\}$
2. V
3. $R(T)$
4. $N(T)$
5. E_λ , for any eigenvalue λ of T .

The proofs that these subspaces are T -invariant are left as exercises. (See Exercise 3.) ♦

Example 2

Let T be the linear operator on R^3 defined by

$$T(a, b, c) = (a + b, b + c, 0).$$

Then the xy -plane $= \{(x, y, 0) : x, y \in R\}$ and the x -axis $= \{(x, 0, 0) : x \in R\}$ are T -invariant subspaces of R^3 . ♦

Let T be a linear operator on a vector space V , and let x be a nonzero vector in V . The subspace

$$W = \text{span}(\{x, T(x), T^2(x), \dots\})$$

is called the **T -cyclic subspace of V generated by x** . It is a simple matter to show that W is T -invariant. In fact, W is the “smallest” T -invariant subspace of V containing x . That is, any T -invariant subspace of V containing x must also contain W (see Exercise 11). Cyclic subspaces have various uses. We apply them in this section to establish the Cayley–Hamilton theorem. In Exercise 31, we outline a method for using cyclic subspaces to compute the characteristic polynomial of a linear operator without resorting to determinants. Cyclic subspaces also play an important role in Chapter 7, where we study matrix representations of nondiagonalizable linear operators.

Example 3

Let T be the linear operator on $\mathbb{R}^3$ defined by

$$T(a, b, c) = (-b + c, a + c, 3c).$$

We determine the T -cyclic subspace generated by $e_1 = (1, 0, 0)$. Since

$$T(e_1) = T(1, 0, 0) = (0, 1, 0) = e_2$$

and

$$T^2(e_1) = T(T(e_1)) = T(e_2) = (-1, 0, 0) = -e_1,$$

it follows that

$$\text{span}(\{e_1, T(e_1), T^2(e_1), \dots\}) = \text{span}(\{e_1, e_2\}) = \{(s, t, 0) : s, t \in \mathbb{R}\}. \quad \blacklozenge$$

Example 4

Let T be the linear operator on $P(R)$ defined by $T(f(x)) = f'(x)$. Then the T -cyclic subspace generated by x^2 is $\text{span}(\{x^2, 2x, 2\}) = P_2(R)$. $\blacklozenge$

The existence of a T -invariant subspace provides the opportunity to define a new linear operator whose domain is this subspace. If T is a linear operator on V and W is a T -invariant subspace of V , then the restriction T_W of T to W (see Appendix B) is a mapping from W to W , and it follows that T_W is a linear operator on W (see Exercise 7). As a linear operator, T_W inherits certain properties from its parent operator T . The following result illustrates one way in which the two operators are linked.

Theorem 5.21. *Let T be a linear operator on a finite-dimensional vector space V , and let W be a T -invariant subspace of V . Then the characteristic polynomial of T_W divides the characteristic polynomial of T .*

Proof. Choose an ordered basis $\gamma = \{v_1, v_2, \dots, v_k\}$ for W , and extend it to an ordered basis $\beta = \{v_1, v_2, \dots, v_k, v_{k+1}, \dots, v_n\}$ for V . Let $A = [T]_\beta$ and $B_1 = [T_W]_\gamma$. Then, by Exercise 12, A can be written in the form

$$A = \begin{pmatrix} B_1 & B_2 \\ O & B_3 \end{pmatrix}.$$

Let $f(t)$ be the characteristic polynomial of T and $g(t)$ the characteristic polynomial of T_W . Then

$$f(t) = \det(A - tI_n) = \det \begin{pmatrix} B_1 - tI_k & B_2 \\ O & B_3 - tI_{n-k} \end{pmatrix} = g(t) \cdot \det(B_3 - tI_{n-k})$$

by Exercise 21 of Section 4.3. Thus $g(t)$ divides $f(t)$. $\blacksquare$

Example 5

Let T be the linear operator on $\mathbb{R}^4$ defined by

$$T(a, b, c, d) = (a + b + 2c - d, b + d, 2c - d, c + d),$$

and let $W = \{(t, s, 0, 0) : t, s \in \mathbb{R}\}$. Observe that W is a T -invariant subspace of $\mathbb{R}^4$ because, for any vector $(a, b, 0, 0) \in \mathbb{R}^4$,

$$T(a, b, 0, 0) = (a + b, b, 0, 0) \in W.$$

Let $\gamma = \{e_1, e_2\}$, which is an ordered basis for W . Extend γ to the standard ordered basis β for $\mathbb{R}^4$. Then

$$B_1 = [T_W]_\gamma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad A = [T]_\beta = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

in the notation of Theorem 5.21. Let $f(t)$ be the characteristic polynomial of T and $g(t)$ be the characteristic polynomial of T_W . Then

$$\begin{aligned} f(t) &= \det(A - tI_4) = \det \begin{pmatrix} 1-t & 1 & 2 & -1 \\ 0 & 1-t & 0 & 1 \\ 0 & 0 & 2-t & -1 \\ 0 & 0 & 1 & 1-t \end{pmatrix} \\ &= \det \begin{pmatrix} 1-t & 1 \\ 0 & 1-t \end{pmatrix} \cdot \det \begin{pmatrix} 2-t & -1 \\ 1 & 1-t \end{pmatrix} \\ &= g(t) \cdot \det \begin{pmatrix} 2-t & -1 \\ 1 & 1-t \end{pmatrix}. \quad \blacklozenge \end{aligned}$$

In view of Theorem 5.21, we may use the characteristic polynomial of T_W to gain information about the characteristic polynomial of T itself. In this regard, cyclic subspaces are useful because the characteristic polynomial of the restriction of a linear operator T to a cyclic subspace is readily computable.

Theorem 5.22. *Let T be a linear operator on a finite-dimensional vector space V , and let W denote the T -cyclic subspace of V generated by a nonzero vector $v \in V$. Let $k = \dim(W)$. Then*

- (a) $\{v, T(v), T^2(v), \dots, T^{k-1}(v)\}$ is a basis for W .
- (b) If $a_0v + a_1T(v) + \dots + a_{k-1}T^{k-1}(v) + T^k(v) = 0$, then the characteristic polynomial of T_W is $f(t) = (-1)^k(a_0 + a_1t + \dots + a_{k-1}t^{k-1} + t^k)$.

Proof. (a) Since $v \neq 0$, the set $\{v\}$ is linearly independent. Let j be the largest positive integer for which

$$\beta = \{v, T(v), \dots, T^{j-1}(v)\}$$

is linearly independent. Such a j must exist because V is finite-dimensional. Let $Z = \text{span}(\beta)$. Then β is a basis for Z . Furthermore, $T^j(v) \in Z$ by Theorem 1.7 (p. 39). We use this information to show that Z is a T -invariant subspace of V . Let $w \in Z$. Since w is a linear combination of the vectors of β , there exist scalars $b_0, b_1, \dots, b_{j-1}$ such that

$$w = b_0v + b_1T(v) + \dots + b_{j-1}T^{j-1}(v),$$

and hence

$$T(w) = b_0T(v) + b_1T^2(v) + \dots + b_{j-1}T^j(v).$$

Thus $T(w)$ is a linear combination of vectors in Z , and hence belongs to Z . So Z is T -invariant. Furthermore, $v \in Z$. By Exercise 11, W is the smallest T -invariant subspace of V that contains v , so that $W \subseteq Z$. Clearly, $Z \subseteq W$, and so we conclude that $Z = W$. It follows that β is a basis for W , and therefore $\dim(W) = j$. Thus $j = k$. This proves (a).

(b) Now view β (from (a)) as an ordered basis for W . Let $a_0, a_1, \dots, a_{k-1}$ be the scalars such that

$$a_0v + a_1T(v) + \dots + a_{k-1}T^{k-1}(v) + T^k(v) = 0.$$

Observe that

$$[T_W]_\beta = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{k-1} \end{pmatrix},$$

which has the characteristic polynomial

$$f(t) = (-1)^k(a_0 + a_1t + \dots + a_{k-1}t^{k-1} + t^k)$$

by Exercise 19. Thus $f(t)$ is the characteristic polynomial of T_W , proving (b). ■

Example 6

Let T be the linear operator of Example 3, and let $W = \text{span}(\{e_1, e_2\})$, the T -cyclic subspace generated by e_1 . We compute the characteristic polynomial $f(t)$ of T_W in two ways: by means of Theorem 5.22 and by means of determinants.

(a) *By means of Theorem 5.22.* From Example 3, we have that $\{e_1, e_2\}$ is a cycle that generates W , and that $T^2(e_1) = -e_1$. Hence

$$1e_1 + 0T(e_1) + T^2(e_1) = 0.$$

Therefore, by Theorem 5.22(b),

$$f(t) = (-1)^2(1 + 0t + t^2) = t^2 + 1.$$

(b) *By means of determinants.* Let $\beta = \{e_1, e_2\}$, which is an ordered basis for W . Since $T(e_1) = e_2$ and $T(e_2) = -e_1$, we have

$$[T_W]_\beta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and therefore,

$$f(t) = \det \begin{pmatrix} -t & -1 \\ 1 & -t \end{pmatrix} = t^2 + 1. \quad \blacklozenge$$

The Cayley–Hamilton Theorem

As an illustration of the importance of Theorem 5.22, we prove a well-known result that is used in Chapter 7. The reader should refer to Appendix E for the definition of $f(T)$, where T is a linear operator and $f(x)$ is a polynomial.

Theorem 5.23 (Cayley–Hamilton). *Let T be a linear operator on a finite-dimensional vector space V , and let $f(t)$ be the characteristic polynomial of T . Then $f(T) = T_0$, the zero transformation. That is, T “satisfies” its characteristic equation.*

Proof. We show that $f(T)(v) = 0$ for all $v \in V$. This is obvious if $v = 0$ because $f(T)$ is linear; so suppose that $v \neq 0$. Let W be the T -cyclic subspace generated by v , and suppose that $\dim(W) = k$. By Theorem 5.22(a), there exist scalars $a_0, a_1, \dots, a_{k-1}$ such that

$$a_0v + a_1T(v) + \cdots + a_{k-1}T^{k-1}(v) + T^k(v) = 0.$$

Hence Theorem 5.22(b) implies that

$$g(t) = (-1)^k(a_0 + a_1t + \cdots + a_{k-1}t^{k-1} + t^k)$$

is the characteristic polynomial of T_W . Combining these two equations yields

$$g(T)(v) = (-1)^k(a_0I + a_1T + \cdots + a_{k-1}T^{k-1} + T^k)(v) = 0.$$

By Theorem 5.21, $g(t)$ divides $f(t)$; hence there exists a polynomial $q(t)$ such that $f(t) = q(t)g(t)$. So

$$f(T)(v) = q(T)g(T)(v) = q(T)(g(T)(v)) = q(T)(0) = 0. \quad \blacksquare$$