MATH2040 Homework 5
Reference Solution

5.4.2(e). For the following linear operator T on the vector space V, determine whether the given subspace W is a T-invariant subspace of V.

$$V = M_{2\times 2}(\mathbb{R}),\ T(A) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} A, \text{ and } W = \{\ A \in V : A^{\mathsf{T}} = A\ \}$$

Solution: Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \in V$. Then $A \in W$, and $T(A) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \notin W$. This implies that $T(W) \nsubseteq W$, so W is not T-invariant.

5.4.4. Let T be a linear operator on a vector space V, and let W be a T-invariant subspace of V. Prove that W is $g(T)$-invariant for any polynomial $g(t)$.

Solution: Let $g \in \mathsf{P}(\mathbb{F})$. We may assume that $g(t) = \sum_{i=0}^{n} a_i t^i$ for some $n \in \mathbb{N}$ and $a_0, \ldots, a_n \in \mathbb{F}$.

Let $w \in W$. Suppose $T^k(w) \in W$ for some $k \in \mathbb{N}$. Since W is T-invariant, $T^{k+1}(w) = T(T^k(w)) \in W$. This implies by induction that $T^n(w) \in W$ for all $n \in \mathbb{N}$. So $g(T)(w) = \left(\sum_{i=0}^{n} a_i T^i\right)(w) = \sum_{i=0}^{n} a_i T^i(w) \in W$ as it is a linear combination of vectors in W. As w is arbitrary, $g(T)(W) \subseteq W$, so W is $g(T)$-invariant.

As g is arbitrary, W is $g(T)$-invariant for all polynomial g.

5.4.6. For each linear operator T on the vector space V, find an ordered basis for the T-cyclic subspace generated by the vector z.

(a) $V = \mathbb{R}^4$, $T(a, b, c, d) = (a + b, b - c, a + c, a + d)$, and $z = e_1$

(b) $V = \mathsf{P}_3(\mathbb{R})$, $T(f) = f''(x)$, and $z = x^3$

(c) $V = M_{2\times 2}(\mathbb{R})$, $T(A) = A^{\mathsf{T}}$, and $z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

(d) $V = M_{2\times 2}(\mathbb{R})$, $T(A) = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} A$, and $z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Solution:

(a)

$$\begin{aligned} Tz = Te_1 &= (1, 0, 1, 1) \notin \operatorname{Span}(\ \{\ e_1\ \}\ ) \\ T^2 e_1 = T(1, 0, 1, 1) &= (1, -1, 2, 2) \notin \operatorname{Span}(\ \{\ e_1, Te_1\ \}\ ) \\ T^3 e_1 = T(1, -1, 2, 2) &= (0, -3, 3, 3) = -3 \cdot (1, 0, 1, 1) + 3 \cdot (1, -1, 2, 2) \in \operatorname{Span}\big(\ \{\ e_1, Te_1, T^2 e_1\ \}\ \big) \end{aligned}$$

So $\{\ e_1, Te_1, T^2 e_1\ \} = \{\ (1, 0, 1, 1), (1, 0, 1, 1), (1, -1, 2, 2)\ \}$ is an ordered basis for the subspace.

(b)

$$\begin{aligned} Tz = Tx^3 &= 6x \notin \operatorname{Span}\big(\ \{\ x^3\ \}\ \big) \\ T^2 z = T(6x) &= 0 \in \operatorname{Span}\big(\ \{\ x^3, 6x\ \}\ \big) \end{aligned}$$

So $\{\ z, Tz\ \} = \{\ x^3, 6x\ \}$ is an ordered basis for the subspace.

(c)

$$Tz = T\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = z$$

As $z \neq 0$, $\{ z \} = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}$ is an ordered basis for the subspace.

(d)

$$\begin{aligned} Tz &= T\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \notin \text{Span}(\{ z \}) \\ T^2z &= T\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 6 & 6 \end{pmatrix} = 3Tz \in \text{Span}(\{ z, Tz \}) \end{aligned}$$

So $\{ z, Tz \} = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \right\}$ is an ordered basis for the subspace.

5.4.17. Let A be an $n \times n$ matrix. Prove that $\dim\left(\text{Span}(I_n, A, A^2, \ldots)\right) \leq n$.

Solution: Let $p(t) = \det(A - tI)$ be the characteristic polynomial of A. By the property of characteristic polynomial, we may assume that $p(t) = (-1)^n t^n + \sum_{i=0}^{n-1} a_i t^i$ for some $a_0, \ldots, a_{n-1} \in \mathbb{F}$. By Cayley–Hamilton theorem, $0 = p(A) = (-1)^n A^n + \sum_{i=0}^{n-1} a_i A^i$. As $(-1)^n \neq 0$, this implies that $\{ I_n = A^0, A, \ldots, A^n \}$ is linearly dependent, and $A^n = (-1)^n \sum_{i=0}^{n-1} a_i A^i$.

Let $C = \text{Span}(\{ I_n, A, \ldots, A^{n-1} \})$. We will show by induction that $A^m \in C$ for all integer $m \in \mathbb{N}$. This would implies that $\text{Span}(\{ I_n, A, A^2, \ldots \}) \subseteq C$ since $\{ I_n, A, A^2, \ldots \} \subseteq C$. As $C = \text{Span}(\{ I_n, A, \ldots, A^{n-1} \})$ is spanned by n elements, we would have $\dim(\text{Span}(\{ I_n, A, A^2, \ldots \})) \leq \dim(C) \leq n$.

Trivially, $A^0 = I_n, \ldots, A^{n-1} \in C$. By the argument above, we also have $A^n \in C$. Suppose for some integer $k \in \mathbb{N}$ we have $A^k \in C$. Then $A^k = \sum_{i=0}^{n-1} c_i A^i$ for some $c_0, \ldots, c_{n-1} \in \mathbb{F}$, so $A^{k+1} = AA^k = \sum_{i=0}^{n-1} c_i A^{i+1} \in C$ as $A, A^2, \ldots, A^n \in C$. By induction, we have $A^m \in C$ for all $m \in \mathbb{Z}^+$.

Hence $\dim(\text{Span}(\{ I_n, A, A^2, \ldots \})) \leq n$.

Note

$\text{Span}(\{ I_n, A, A^2, \ldots \}) = C$. The dimension of C is also the degree of the minimal polynomial of A.

5.4.18. Let A be an $n \times n$ matrix with characteristic polynomial $f(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \ldots + a_1 t + a_0$

(a) Prove that A is invertible if and only if $a_0 \neq 0$

(b) Prove that if A is invertible, then $A^{-1} = (-1/a_0)[(-1)^n A^{n-1} + a_{n-1}A^{n-2} + \ldots + a_1 I_n]$

(c) Use (b) to compute A^{-1} for $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & -1 \end{pmatrix}$

Solution:

(a) By the result of Question 5.1.20 in Homework 4, the proposition holds.

(b) By Cayley–Hamilton theorem, we have $0 = p(A) = (-1)^n A^n + a_{n-1}A^{n-1} + \ldots + a_1 A + a_0 I_n$. Since A is invertible, $a_0 \neq 0$, so $I_n = (-1/a_0)\left((-1)^n A^n + \ldots + a_1 A\right) = (-1/a_0)A\left((-1)^n A^{n-1} + a_{n-1}A^{n-2} + \ldots + a_1 I_n\right)$, and thus

$$\begin{aligned} A^{-1} = A^{-1}I_n &= A^{-1}(-1/a_0)A\left((-1)^n A^{n-1} + a_{n-1}A^{n-2} + \ldots + a_1 I_n\right) \\ &= (-1/a_0)\left((-1)^n A^{n-1} + a_{n-1}A^{n-2} + \ldots + a_1 I_n\right) \end{aligned}$$

(c) The characteristic polynomial of A is $p(t) = \det(A - tI_3) = -(t-1)(t-2)(t+1) = -t^3 + 2t^2 + t - 2$. By part (b), this implies that $A^{-1} = \frac{-1}{-2}(-A^2 + 2A + I_3) = \frac{1}{2}\left(-\begin{pmatrix} 1 & 6 & 6 \\ 0 & 4 & 3 \\ 0 & 0 & 1 \end{pmatrix} + 2\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & -1 & -2 \\ 0 & 1/2 & 3/2 \\ 0 & 0 & -1 \end{pmatrix}$.

5.4.19. Let A denote the $k \times k$ matrix

$$\begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & -a_{k-2} \\ 0 & 0 & \dots & 1 & -a_{k-1} \end{pmatrix}$$

where $a_0, a_1, \dots, a_{k-1}$ are arbitrary scalars. Prove that the characteristic polynomial of A is $(-1)^k(a_0 + a_1 t + \dots + a_{k-1}t^{k-1} + t^k)$

Solution: We will show the proposition by induction on the size k. Trivially the proposition holds on the case $k = 1$ and $k = 2$.

Suppose for some integer $n \geq 2$ we have that the characteristic polynomial of $\begin{pmatrix} 0 & 0 & \dots & 0 & -b_0 \\ 1 & 0 & \dots & 0 & -b_1 \\ 0 & 1 & \dots & 0 & -b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & -b_{n-2} \\ 0 & 0 & \dots & 1 & -b_{n-1} \end{pmatrix} \in M_{n\times n}(\mathbb{F})$ is

$$\det\left(\begin{pmatrix} 0 & 0 & \dots & 0 & -b_0 \\ 1 & 0 & \dots & 0 & -b_1 \\ 0 & 1 & \dots & 0 & -b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & -b_{n-2} \\ 0 & 0 & \dots & 1 & -b_{n-1} \end{pmatrix} - tI_n\right) = \det\begin{pmatrix} -t & 0 & \dots & 0 & -b_0 \\ 1 & -t & \dots & 0 & -b_1 \\ 0 & 1 & \dots & 0 & -b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -t & -b_{n-2} \\ 0 & 0 & \dots & 1 & -b_{n-1} - t \end{pmatrix} = (-1)^n(b_0 + b_1 t + \dots + b_{n-1}t^{n-1} + t^n)$$

for arbitrary scalars $b_0, \dots, b_{n-1}$. Let $A = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & -a_{n-1} \\ 0 & 0 & \dots & 1 & -a_n \end{pmatrix} \in M_{(n+1)\times(n+1)}(\mathbb{F})$ with scalars $a_0, \dots, a_n \in \mathbb{F}$. Then by expanding along the last column, we can see that the characteristic polynomial of A is

$$p(t) = \det(A - tI_{n+1}) = \det\begin{pmatrix} -t & 0 & \dots & 0 & -a_0 \\ 1 & -t & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -t & -a_{n-1} \\ 0 & 0 & \dots & 1 & -a_n - t \end{pmatrix}$$

$$= \sum_{i=0}^{n-1}(-1)^{(i+1)+(n+1)}(-a_i)\det\begin{pmatrix} -tI_i + J_i & 0_{i\times(n-i)} \\ 0_{(n-i)\times i} & I_{n-i} - tJ_{n-i}^{\intercal} \end{pmatrix} + (-1)^{(n+1)+(n+1)}(-a_n - t)\det\begin{pmatrix} -t & 0 & \dots & 0 \\ 1 & -t & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & -t \end{pmatrix}$$

$$= \sum_{i=0}^{n-1}(-1)^{n+i+1}a_i\det(I_{n-i} - tJ_{n-i}^{\intercal})\det\Big((-tI_i + J_i) - 0_{i\times(n-i)}\big(I_{n-i} - tJ_{n-i}^{\intercal}\big)^{-1}0_{(n-i)\times i}\Big) + (-1)^{n+1}(a_n + t)t^n$$

$$= \sum_{i=0}^{n-1}(-1)^{n+i+1}a_i\det(-tI_i + J_i) + (-1)^{n+1}(a_n + t)t^n$$

$$= \sum_{i=0}^{n-1}(-1)^{n+1}a_i t^i + (-1)^{n+1}(a_n + t)t^n = (-1)^{n+1}(a_0 + \dots + a_n t^n + t^{n+1})$$

where J_k is the $k \times k$ lower shift matrix. As $a_0, \dots, a_n$ are arbitrary, by induction the proposition holds for all $k \in \mathbb{Z}^+$.

Note

The general form of A is $\begin{pmatrix} 0_{1\times(k-1)} & -a_0 \\ I_{k-1} & \begin{pmatrix} -a_1 & \dots & -a_{k-1} \end{pmatrix}^\mathsf{T} \end{pmatrix}$.

As $a_0, \dots, a_{k-1}$ are arbitrary, this implies that every nonzero polynomial with the correct leading coefficient is a characteristic polynomial of its companion matrix.

5.4.22. Let T be a linear operator on a two-dimensional vector space V and suppose that $T \neq c\mathrm{Id}$ for any scalar c. Show that if U is any linear operator on V such that $UT = TU$, then $U = g(T)$ for some polynomial $g(t)$.

Idea: Since V is two-dimensional, by Cayley–Hamilton every power of T with exponent larger than 1 can be expressed as a linear combination of Id and T. So if the proposition holds, we should be able to find a g that is of degree at most 1.

Again, due to the dimension it suffices to show the proposition on a linearly independent set of two vectors (which would be a basis). How do we find the right vectors? Suppose somehow we have already got one of the vectors $x \in V$. This means that $Ux = ax + bTx$ for some a, b. To show the proposition, we need to find another vector $y \in V$ that is not a multiple of x (so that $\{ x, y \}$ is linearly independent), and $Uy = ay + bTy$, with exactly the same coefficients. It would be hard to ensure this property unless x and y are related in some way. However, as U, T commute, we will always have $U(Tx) = T(Ux) = T(ax + bTx) = aTx + bT(Tx)$. So we may select $y = Tx$, as long as x makes $\{ x, Tx \}$ linearly independent, or equivalently V is T-cyclic generated by x.

How do we make sure that V is T-cyclic? It would be hard to pick one such x without knowing what T is, so we may as well prove it by contradiction. Assuming there is no such x, for each nonzero v we will always have Tv being a scalar multiple of v, for otherwise they are linearly independent. This means that every nonzero vector is an eigenvector of T. It remains to show that this implies that T is a multiple of Id (so that we can have a contradiction).

Solution: We first show that V is T-cyclic.

Suppose V is not T-cyclic. Let $v \in V$ be nonzero. Then $V \neq \mathrm{Span}(\{ v, Tv, T^2v, \dots \})$. Since V is two-dimensional, we must have that $\{ v, Tv \}$ is linearly dependent, for otherwise $\mathrm{Span}(\{ v, Tv \}) \subseteq V$ is also two-dimensional and so is V itself, implying that V is T-cyclic. Then there exists scalars a, b, depending on v and not all zero, such that $av + bTv = 0$. Since $v \neq 0$, we cannot have $b = 0$, for this would implies that $av = 0$ with $a \neq 0$ and $v \neq 0$. Hence $Tv = c_v \cdot v$ with $c_v = -\frac{a}{b}$ where c_v also depends on v. Since v is nonzero, there is only such scalar that satisfies the relation $Tv = cv$, so $c(v) = c_v$ is a well-defined function on the set $V \setminus \{0\}$.

We will show that $c(v)$, as a function on $V \setminus \{0\}$, is constant. Let $x, y \in V$ be nonzero. Then

- Suppose $\{ x, y \}$ is linearly independent. Then $x, y, x+y$ are all nonzero, so $c(x+y)\cdot(x+y) = T(x+y) = Tx + Ty = c(x)\cdot x + c(y)\cdot y$ and thus $(c(x+y) - c(x))x + (c(x+y) - c(y))y = 0$. As $\{ x, y \}$ is linearly independent, this implies that $c(x+y) - c(x) = c(x+y) - c(y) = 0$ and so $c(x) = c(x+y) = c(y)$.
- Suppose $\{ x, y \}$ is linearly dependent. As x, y are both nonzero, there exists nonzero $\lambda \in \mathbb{F}$ such that $y = \lambda x$. So $(\lambda c(\lambda x)) \cdot x = c(\lambda x) \cdot (\lambda x) = T(\lambda x) = \lambda T x = \lambda c(x) x$. As $\lambda \neq 0$ and $x \neq 0$, we must have $c(y) = c(\lambda x) = c(x)$

So for all nonzero x, y we always have $c(x) = c(y)$. This means that for some scalar $c \in \mathbb{F}$, we have for all nonzero $v \in V$ that $c(v) = c$ and thus $Tv = cv$. As $T0 = 0 = c \cdot 0$, we have $Tv = cv$ for all $v \in V$ and thus $T = c\mathrm{Id}$. This contradicts the assumption on T.

Therefore, V is T-cyclic. So there exists nonzero $v \in V$ such that $V = \mathrm{Span}(\{ v, Tv, T^2v, \dots \})$. As $v \neq 0$ and V is two-dimensional, we must have that $\beta = \{ v, Tv \}$ is a basis of V.

Since $Uv \in V$, we have $Uv = cv + dTv$ for some scalars c, d. Then $U(Tv) = UTv = TUv = T(cv + dTv) = c(Tv) + dT(Tv)$. So $Ux = cx + dTx = (c\mathrm{Id} + dT)(x)$ for all $x \in \beta$. As β is a basis of V and U is linear, with the polynomial $g(t) = c + dt$ we must have $U = c\mathrm{Id} + dT = g(T)$.

Therefore $U = g(T)$ for some polynomial g.

Note

The part where we show that V is T-cyclic is the answer for Question 5.4.21, which is also a hint for this question if you read the textbook for a bit.

Note that in the first part a, b are not unique (as $\lambda a, \lambda b$ also verify the relation for all $\lambda \neq 0$), but their ratio $-c = a/b$ is.

This answer also proves the following: if T is linear on a vector space (not necessarily finite-dimensional) where every nonzero vector is an eigenvector, T must be a scalar multiple of Id.

5.4.23. Let T be a linear operator on a finite-dimensional vector space V, and let W be a T-invariant subspace of V. Suppose that $v_1, v_2, \ldots, v_k$ are eigenvectors of T corresponding to distinct eigenvalues. Prove that if $v_1 + v_2 + \ldots + v_k$ is in W, then $v_i \in W$ for all i.

Solution: To show the proposition, we will use induction on k. Trivially, the base case where $k = 1$ holds.

Suppose for some integer $n \geq 1$ we have $w_1, \ldots, w_n \in W$ whenever $w_1, \ldots, w_n \in V$ are eigenvectors of T corresponding to distinct eigenvalues such that $w_1 + \ldots + w_n \in W$. Let $v_1, \ldots, v_{n+1} \in V$ be eigenvectors of T corresponding to distinct eigenvalues such that $v_1 + \ldots + v_{n+1} \in W$. For $i \in \{ 1, \ldots, n+1 \}$ let $\lambda_i \in \mathbb{F}$ be the eigenvalue corresponding to v_i. Then $Tv_i = \lambda_i v_i$ for all $i \in \{ 1, \ldots, n+1 \}$.

Since $v_1 + \ldots + v_{n+1} \in W$ and W is T-invariant, we have $\lambda_1 v_1 + \ldots + \lambda_{n+1} v_{n+1} = T(v_1 + \ldots + v_{n+1}) \in W$. As W is a subspace of V, we also have $\lambda_{n+1}(v_1 + \ldots + v_{n+1}) \in W$, hence $(\lambda_1 - \lambda_{n+1})v_1 + \ldots + (\lambda_n - \lambda_{n+1})v_n = (\lambda_1 v_1 + \ldots + \lambda_{n+1} v_{n+1}) - \lambda_{n+1}(v_1 + \ldots + v_{n+1}) \in W$. Since $\lambda_1, \ldots, \lambda_{n+1}$ are distinct, we have $\lambda_1 - \lambda_{n+1}, \ldots, \lambda_n - \lambda_{n+1}$ are all nonzero, hence $(\lambda_1 - \lambda_{n+1})v_1, \ldots, (\lambda_n - \lambda_{n+1})v_n$ are all eigenvectors of T that correspond to distinct eigenvalues. By induction assumption, $(\lambda_1 - \lambda_{n+1})v_1, \ldots, (\lambda_n - \lambda_{n+1})v_n \in W$ and thus $v_1, \ldots, v_n \in W$. This also implies that $v_{n+1} = (v_1 + \ldots + v_{n+1}) - v_1 - \ldots - v_n \in W$.

By induction, the proposition holds for all $n \in \mathbb{Z}^+$.

5.4.24. Prove that the restriction of a diagonalizable linear operator T to any nontrivial T-invariant subspace is also diagonalizable.

Idea: How do we approach this problem? Although we know that the characteristic polynomial p_W of T_W on a nontrivial T-invariant subspace W is a factor of the original characteristic polynomial and so it splits, this tells us nothing about the eigenspaces of T_W. However, from an exercise in a previous homework, we know that the eigenspaces of T_W must be the intersections of W with the corresponding eigenspaces of T. It then remains to characterize diagonalizablility with a relation between eigenspaces that works well with intersections. The hint in the textbook about using Question 5.4.23 also gives away about which relation to use.

Solution: Let T be a diagonalizable linear operator on a finite-dimensional space V. Then by Theorem 5.11 in textbook, V is a direct sum of the eigenspaces of T. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T, and the corresponding eigenspaces be $E_1, \ldots, E_k$. Then $V = \bigoplus_{i=1}^k E_i$, so $V = \sum_{i=1}^k E_i$ and $E_i \cap \sum_{j \neq i} E_j = \{0\}$ for all i.

Let $W \subseteq V$ be a nontrivial T-invariant subspace. Then $W \supseteq W \cap E_i$ for all i, so $W \supseteq \sum (W \cap E_i)$. Let $w \in W \subseteq V$. Then for each i there exists $v_i \in E_i$ such that $w = \sum v_i$. By permuting the indices we may assume that $v_1, \ldots, v_l$ are nonzero and $v_{l+1} = \ldots = v_k = 0$ with $l \in \{ 0, \ldots, k \}$, with the obvious convention that $l = 0$ implies $v_1 = \ldots = v_k = 0$ and $l = k$ implies that $v_1, \ldots, v_k$ are all nonzero. Then $v_1 + \ldots + v_l = w \in W$. By definition, $v_1, \ldots, v_l$ are eigenvectors that correspond to distinct eigenvalues, so by the result of Question 5.4.23, $v_1, \ldots, v_l \in W$. Trivially, $v_{l+1} = \ldots = v_k = 0 \in W$. This implies that $v_i \in W \cap E_i$ for each i, and so $w \in \sum (W \cap E_i)$. As w is arbitrary, $W \subseteq \sum (W \cap E_i)$ and so $W = \sum (W \cap E_i)$.

For each i, since $\{0\} = W \cap \{0\} = W \cap \left(E_i \cap \sum_{j \neq i} E_j \right) = (W \cap E_i) \cap \sum_{j \neq i} E_j \supseteq (W \cap E_i) \cap \sum_{j \neq i} (W \cap E_j) \supseteq \{0\}$, we have $(W \cap E_i) \cap \sum_{j \neq i} (W \cap E_j) = \{0\}$. Hence $W = \bigoplus_{i=1}^k (W \cap E_i)$. By the result of Question 2.1.32 in Homework 2, $E_{\lambda_i}(T_W) = \mathsf{N}\left(T_W - \lambda_i \mathrm{Id}_W \right) = \mathsf{N}\left((T - \lambda_i \mathrm{Id})_W \right) = \mathsf{N}\left(T - \lambda_i \mathrm{Id} \right) \cap W = E_i \cap W$ for each i. In particular, $S = \{ W \cap E_i : W \cap E_i \neq \{0\} \}$ is the complete set of eigenspaces of T_W, and $W = \bigoplus_{E \in S} E$. By Theorem 5.11 in textbook, this implies that T_W is diagonalizable.

As W is arbitrary, the restriction of T to any T-invariant subspace is also diagonalizable.

7.1.7(a). Let U be a linear operator on a finite-dimensional vector space V. Prove that $\mathsf{N}(U) \subseteq \mathsf{N}(U^2) \subseteq \ldots \subseteq \mathsf{N}(U^k) \subseteq \mathsf{N}(U^{k+1}) \subseteq \ldots$.

Solution: Let $k \in \mathbb{Z}^+$ and $v \in \mathsf{N}(U^k)$. Then $U^k(v) = 0$, so $U^{k+1}(v) = U(U^k(v)) = U(0) = 0$. Hence $v \in \mathsf{N}(U^{k+1})$. As v is arbitrary, $\mathsf{N}(U^k) \subseteq \mathsf{N}(U^{k+1})$. As k is arbitrary, we have $\mathsf{N}(U) \subseteq \mathsf{N}(U^2) \subseteq \ldots$.

Note

See also Question 2.3.16 in Homework 3.

7.1.7(b). Let U be a linear operator on a finite-dimensional vector space V. Prove that if $\mathrm{rank}(U^m) = \mathrm{rank}(U^{m+1})$ for some positive integer m, then $\mathrm{rank}(U^m) = \mathrm{rank}(U^k)$ for any positive integer $k \geq m$.

Solution: Let $k \in \mathbb{Z}^+$ and $y \in \mathsf{R}\left(U^{k+1} \right)$. Then $y = U^{k+1}(x) = U^k(U(x)) \in \mathsf{R}\left(U^k \right)$ for some $x \in V$. As y is arbitrary, $\mathsf{R}\left(U^k \right) \supseteq \mathsf{R}\left(U^{k+1} \right)$. As k is arbitrary, $\mathsf{R}\left(U^m \right) \supseteq \mathsf{R}\left(U^{m+1} \right) \supseteq \ldots$.

Since $\mathsf{R}\left(U^m \right), \mathsf{R}\left(U^{m+1} \right)$ are subspaces of a finite-dimensional space V and $\dim(\mathsf{R}\left(U^m \right)) = \text{rank}(U^m) = \text{rank}(U^{m+1}) = \dim(\mathsf{R}\left(U^{m+1} \right))$, $\mathsf{R}\left(U^m \right) = \mathsf{R}\left(U^{m+1} \right)$.

Suppose $\mathsf{R}\left(U^m \right) = \mathsf{R}\left(U^{m+k} \right)$ for some $k \in \mathbb{Z}^+$. Let $y \in \mathsf{R}\left(U^{m+k} \right)$. Then there exists $x \in V$ such that $y = U^{m+k}(x) = U^k(U^m(x))$ with $U^m(x) \in \mathsf{R}\left(U^m \right) = \mathsf{R}\left(U^{m+1} \right)$, so there exists $z \in V$ such that $U^m(x) = U^{m+1}(z)$ and thus $y = U^k(U^m(x)) = U^k(U^{m+1}(z)) = U^{m+k+1}(z) \in \mathsf{R}\left(U^{m+k+1} \right)$. As y is arbitrary, $\mathsf{R}\left(U^{m+k} \right) \subseteq \mathsf{R}\left(U^{m+k+1} \right)$, and thus $\mathsf{R}\left(U^m \right) = \mathsf{R}\left(U^{m+k} \right) = \mathsf{R}\left(U^{m+k+1} \right)$.

By induction, $\mathsf{R}\left(U^m \right) = \mathsf{R}\left(U^k \right)$ for all $k \geq m$, and hence $\text{rank}(U^m) = \text{rank}(U^k)$ for all $k \geq m$.

7.1.7(c). Let U be a linear operator on a finite-dimensional vector space V. Prove that If $\text{rank}(U^m) = \text{rank}(U^{m+1})$ for some positive integer m, then $\mathsf{N}\left(U^m \right) = \mathsf{N}\left(U^k \right)$ for any positive integer $k \geq m$.

Solution: Let $k \geq m$ be integer. By the previous question (Question 7.1.7(b)), $\text{rank}(U^m) = \text{rank}(U^k)$. Since V is finite-dimensional, by dimension theorem $\dim(\mathsf{N}\left(U^m \right)) = \text{nullity}(U^m) = \dim(V) - \text{rank}(U^m) = \dim(V) - \text{rank}(U^k) = \text{nullity}(U^k) = \dim(\mathsf{N}\left(U^k \right))$. By part (a), $\mathsf{N}\left(U^m \right) \subseteq \mathsf{N}\left(U^k \right)$, so $\mathsf{N}\left(U^m \right) = \mathsf{N}\left(U^k \right)$.

As k is arbitrary, $\mathsf{N}\left(U^m \right) = \mathsf{N}\left(U^k \right)$ for all $k \geq m$.

7.1.7(d). Let V be a finite-dimensional vector space. Let T be a linear operator on V, and let λ be an eigenvalue of T. Prove that if $\text{rank}((T - \lambda \text{Id})^m) = \text{rank}((T - \lambda \text{Id})^{m+1})$ for some integer m, then $K_\lambda = \mathsf{N}\left((T - \lambda \text{Id})^m \right)$.

Solution: By the previous part (Question 7.1.7(c)), $\mathsf{N}\left((T - \lambda \text{Id})^m \right) = \mathsf{N}\left((T - \lambda \text{Id})^k \right)$ for all $k \geq m$. By the result of Question 7.1.7(a), $\mathsf{N}\left(T - \lambda \text{Id} \right) \subseteq \ldots \subseteq \mathsf{N}\left((T - \lambda \text{Id})^m \right)$. Hence $K_\lambda = \bigcup_{n \in \mathbb{Z}^+} \mathsf{N}\left((T - \lambda \text{Id})^n \right) = \bigcup_{n=1}^{m} \mathsf{N}\left((T - \lambda \text{Id})^n \right) \cup \bigcup_{n \geq m+1} \mathsf{N}\left((T - \lambda \text{Id})^n \right) = \mathsf{N}\left((T - \lambda \text{Id})^m \right) \cup \mathsf{N}\left((T - \lambda \text{Id})^m \right) = \mathsf{N}\left((T - \lambda \text{Id})^m \right)$.

Practice Problems

5.4.1. Label the following statements as true or false.

(a) There exists a linear operator T with no T-invariant subspace.

(b) If T is a linear operator on a finite-dimensional vector space V and W is a T-invariant subspace of V. then the characteristic polynomial of T_W divides the characteristic polynomial of T.

(c) Let T be a linear operator on a finite-dimensional vector space V, and let v and w be in V. If W is the T-cyclic subspace generated by V, W' is the T-cyclic subspace generated by w, and $W = W$, then $v = w$.

(d) If T is a linear operator on a finite-dimensional vector space V, then for any $v \in V$ the T-cyclic subspace generated by v is the same as the T-cyclic subspace generated by $T(v)$.

(e) Let T be a linear operator on an n-dimensional vector space. Then there exists a polynomial $g(t)$ of degree n such that $g(T) = T_0$.

(f) Any polynomial of degree n with leading coefficient $(-1)^n$ is the characteristic polynomial of some linear operator.

(g) If T is a linear operator on a finite-dimensional vector space V, and if V is the direct sum of k T-invariant subspaces, then there is an ordered basis β for V such that $[T]_\beta$ is a direct sum of k matrices.

Solution:

(a) False

(b) True

(c) False

(d) False

(e) True

(f) True. See Question 5.4.19

(g) True

5.4.15. Use the Cayley-Hamilton theorem to prove its corollary for matrices.

Solution: Let $A \in M_{n\times n}(\mathbb{F})$ with $n \in \mathbb{Z}^+$. Then L_A is a linear operator on $\mathbb{F}^n$. Let α be the standard basis of $\mathbb{F}^n$. Then the characteristic polynomial of L_A is $p(t) = \det([\mathsf{L}_A]_\alpha - tI_n) = \det(A - tI_n)$, which is also the characteristic polynomial of A.

Assume that $p(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \ldots + a_0$ with $a_0, \ldots, a_{n-1} \in \mathbb{F}$. Then by Cayley–Hamilton theorem, $p(\mathsf{L}_A) = (-1)^n \mathsf{L}_A^n + a_{n-1}\mathsf{L}_A^{n-1} + \ldots + a_0 \mathrm{Id} = 0$. Hence $0 = [p(\mathsf{L}_A)]_\alpha = (-1)^n[\mathsf{L}_A^n]_\alpha + a_{n-1}[\mathsf{L}_A^{n-1}]_\alpha + \ldots + a_0[\mathrm{Id}]_\alpha = (-1)^n[\mathsf{L}_A]_\alpha^n + a_{n-1}[\mathsf{L}_A]_\alpha^{n-1} + \ldots + a_0 I_n = (-1)^n A^n + a_{n-1}A^{n-1} + \ldots + a_0 A^0 = p(A)$.

As A, n are arbitrary, we have $p(A) = 0$ for all matrix $A \in M_{n\times n}(\mathbb{F})$ where p is the characteristic polynomial of A.

Note

As mentioned in the remark in the textbook, it is invalid to claim that the proposition holds as $p(A) = \det(A - AI) = \det(0_{n\times n}) = 0$. See this question on MSE for some discussions.

5.4.16. Let T be a linear operator on a finite-dimensional vector space V.

(a) Prove that if the characteristic polynomial of T splits, then so does the characteristic polynomial of the restriction of T to any T-invariant subspace of V.

(b) Deduce that if the characteristic polynomial of T splits, then any nontrivial T-invariant subspace of V contains an eigenvector of T.

Solution:

(a) Let $W \subseteq V$ be a T-invariant subspace of V, and $p(t), p_W(t)$ be the characteristic polynomials of T, T_W respectively. By theorem 5.21 in textbook, we have p_W divides p, so there exists a polynomial q such that $p = p_W q$. Trivially, we must have $p_W, q \neq 0$. As $\mathbb{F}$ is a field, $\mathbb{F}[t]$ is a PID and so a UFD. Hence we may assume that $p_W = up_1 \ldots p_n$ and $q = u'q_1 \ldots q_m$ where u, u' are units and thus scalars and $p_1, \ldots, p_n, q_1, \ldots, q_m$ are primes for some $n, m \in \mathbb{N}$ possibly zero. This implies that $p = (uu')p_1 \ldots p_n q_1 \ldots q_m$. As p splits, $p_1, \ldots, p_n, q_1, \ldots, q_m$ must all be linear factors. This implies that $p_W = up_1 \ldots p_n$ is a product of linear factors (and scalars) and thus splits.

(b) Let $W \subseteq V$ be a nontrivial T-invariant subspace of V. Let $p_W(t)$ be the characteristic polynomial of T_W. By the previous part, p_W splits. Let $u, n, p_1, \ldots, p_n$ be defined as in part (a). Then $n = \deg p_W = \dim(W) \geq 1$. This implies that p_W is a multiple of a linear factor. This implies that p_W has a root λ in $\mathbb{F}$. By the property of characteristic polynomial, $E_\lambda(T_W)$ is nontrivial. Hence T_W (and so T) has a eigenvector in W.

5.4.25. (a) Prove the converse to Exercise 18(a) of Section 5.2: If T and U are diagonalizable linear operators on a finite-dimensional vector space V such that $UT = TU$, then T and U are simultaneously diagonalizable.

(b) State and prove a matrix version of (a).

Solution:

(a) Let λ be an eigenvalue of T. Let $v \in E_\lambda(T)$. Then $Tv = \lambda v$, so $T(Uv) = UTv = U(\lambda v) = \lambda Uv$, $Uv \in E_\lambda(T)$. As v is arbitrary, $UE_\lambda(T) \subseteq E_\lambda(T)$, and so $E_\lambda(T)$ is U-invariant.

Since λ is an eigenvalue of T, $E_\lambda(T)$ is nontrivial. As T is diagonalizable, the characteristic polynomial of T splits. So by Question 5.4.24, $U_{E_\lambda(T)}$ is diagonalizable, and thus there exists a basis β_λ of $E_\lambda(T)$ consisting of eigenvectors of $U_{E_\lambda(T)}$ (and so of U).

Let $\{ \lambda_1, \ldots, \lambda_k \}$ be the complete set of distinct eigenvalues of T. As T is diagonalizable, $V = \bigoplus_{i=1}^k E_{\lambda_i}(T)$. For each $i \in \{ 1, \ldots, k \}$ let β_{λ_i} be the basis of $E_{\lambda_i}(T)$ defined above, and $\beta = \bigcup_{i=1}^k \beta_{\lambda_i}$. Then β is a union of bases of subspaces whose direct sum is the whole space V, so β is a basis of V. By definition, β is consisting of vectors which are eigenvectors of both T and U. So T and U are simultaneously diagonalizable (as witnessed by β).

(b) Let $A, B \in M_{n\times n}(\mathbb{F})$ be diagonalizable matrices that commute. Then they are simultaneously diagonalizable.

The proof is as follows:

Let $A, B \in M_{n\times n}(\mathbb{F})$ be diagonalizable matrices that commute. Then by Question 5.2.17 (in Homework 4), $\mathsf{L}_A, \mathsf{L}_B$ are diagonalizable linear operators and commute. By part (a), L_A and L_B are simultaneously diagonalizable. So again by Question 5.2.17, $A = [\mathsf{L}_A]_\alpha$ and $B = [\mathsf{L}_B]_\alpha$ are simultaneously diagonalizable where α is the standard basis of $\mathbb{F}^n$.

Note

See also this note and section 13 of this reference.

7.1.1. Label the following statements as true or false.

(a) Eigenvectors of a linear operator T are also generalized eigenvectors of T.

(b) It is possible for a generalized eigenvector of a linear operator T to correspond to a scalar that is not an eigenvalue of T.

(c) Any linear operator on a finite-dimensional vector space has a Jordan canonical form.

(d) A cycle of generalized eigenvectors is linearly independent.

(e) There is exactly one cycle of generalized eigenvectors corresponding to each eigenvalue of a linear operator on a finite-dimensional vector space.

(f) Let T be a linear operator on a finite-dimensional vector space whose characteristic polynomial splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of T. If, for each i, β_i is a basis for K_{λ_i}, then $\beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$ is a Jordan canonical basis for T.

(g) For any Jordan block J, the operator L_J has Jordan canonical form J.

(h) Let T be a linear operator on an n-dimensional vector space whose characteristic polynomial splits. Then, for any eigenvalue λ of T, $K_\lambda = \mathsf{N}\left((T - \lambda \mathrm{Id})^n \right)$.

Solution:

(a) True

(b) False

(c) True

(d) True. See Corollary of Theorem 7.6 in textbook.

(e) False

(f) False

(g) True

(h) True

7.1.7(e). Let V be a finite-dimensional vector space. Let T be a linear operator on V whose characteristic polynomial splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of T. Prove that T is diagonalizable if and only if $\mathrm{rank}(T - \lambda_i \mathrm{Id}) = \mathrm{rank}((T - \lambda_i \mathrm{Id})^2)$ for $1 \leq i \leq k$.

Solution: Suppose T is diagonalizable. Let $i \in \{ 1, \ldots, k \}$. Then by the corollary of Theorem 7.4 in textbook, we have $E_{\lambda_i} = K_{\lambda_i}$ and so $\dim(\mathsf{N}(T - \lambda \mathrm{Id})) = \dim(E_{\lambda_i}) = \dim(K_{\lambda_i}) \geq \dim(\mathsf{N}\left((T - \lambda_i \mathrm{Id})^2 \right))$ as $K_{\lambda_i} \supseteq \mathsf{N}\left((T - \lambda_i \mathrm{Id})^2 \right)$. By the result of Question 7.1.7(a), $\mathsf{N}(T - \lambda_i \mathrm{Id}) = \mathsf{N}\left((T - \lambda_i \mathrm{Id})^2 \right)$, so $\mathrm{nullity}(T - \lambda_i \mathrm{Id}) = \dim(\mathsf{N}(T - \lambda_i \mathrm{Id})) = \dim(\mathsf{N}\left((T - \lambda_i \mathrm{Id})^2 \right)) = \mathrm{nullity}((T - \lambda_i \mathrm{Id})^2)$. By dimension theorem, $\mathrm{rank}(T - \lambda_i \mathrm{Id}) = \dim(V) - \mathrm{nullity}(T - \lambda_i \mathrm{Id}) = \dim(V) - \mathrm{nullity}((T - \lambda_i \mathrm{Id})^2) = \mathrm{rank}((T - \lambda_i \mathrm{Id})^2)$. As i is arbitrary, $\mathrm{rank}(T - \lambda_i \mathrm{Id}) = \mathrm{rank}((T - \lambda_i \mathrm{Id})^2)$ for all $i \in \{ 1, \ldots, k \}$.

Suppose $\mathrm{rank}(T - \lambda_i \mathrm{Id}) = \mathrm{rank}((T - \lambda_i \mathrm{Id})^2)$ for all $i \in \{ 1, \ldots, k \}$. By the result of Question 7.1.7(d), $K_{\lambda_i} = \mathsf{N}(T - \lambda_i I) = E_{\lambda_i}$ for all $i \in \{ 1, \ldots, k \}$. By the corollary of Theorem 7.4 in textbook, T is diagonalizable.

7.1.7(f). Prove that if T is a diagonalizable linear operator on a finite dimensional vector space V and W is a T-invariant subspace of V, then T_W is diagonalizable.

Solution: Since T is diagonalizable, the characteristic polynomial of T splits. Let $\lambda_1, \ldots, \lambda_k$ be all the distinct eigenvalues of T for some $k \in \mathbb{Z}^+$. Since W is T-invariant, the characteristic polynomial of T_W divides the characteristic polynomial of T, and so the eigenvalues of T_W are all contained in $\{ \lambda_1, \ldots, \lambda_k \}$.

Let $i \in \{ 1, \ldots, k \}$. By the result of Question 7.1.7(e), we have $\mathrm{rank}(T - \lambda_i \mathrm{Id}) = \mathrm{rank}((T - \lambda_i \mathrm{Id})^2)$, so by dimension theorem (as in part (e)), $\mathsf{N}(T - \lambda_i \mathrm{Id}) = \mathsf{N}\left((T - \lambda_i \mathrm{Id})^2 \right)$. By the result of Question 2.1.32 in Homework 2, $\mathsf{N}(T_W - \lambda_i \mathrm{Id}_W) = \mathsf{N}((T - \lambda_i \mathrm{Id})_W) = \mathsf{N}(T - \lambda_i \mathrm{Id}) \cap W = \mathsf{N}\left((T - \lambda_i \mathrm{Id})^2 \right) \cap W = \mathsf{N}\left((T - \lambda_i \mathrm{Id})^2_W \right) = \mathsf{N}\left((T_W - \lambda_i \mathrm{Id}_W)^2 \right)$. By dimension theorem, $\mathrm{rank}(T_W - \lambda_i \mathrm{Id}_W) = \mathrm{rank}((T_W - \lambda_i \mathrm{Id}_W)^2)$. As i is arbitrary, for each eigenvalue μ of T_W we have $\mathrm{rank}(T_W - \mu \mathrm{Id}_W) = \mathrm{rank}((T_W - \mu \mathrm{Id}_W)^2)$, so by the result of Question 7.1.7(e), T_W is diagonalizable.