Recall the following Theorems, Propositions and definitions:

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 $v \in W$.

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

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

is called the $T$-cyclic subspace of $V$ generated by $x$. Recall that another way of describing a $T$-cyclic subspace is as the smallest $T$-invariant subspace of $V$ containing $x$.

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 $T$.

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) $\{v, T(v), T^2(v), \ldots, T^{k-1}(v)\}$ is a basis for $W$.

(b) If $a_0 v + a_1 T(v) + \ldots + a_{k-1} T^{k-1}(v) + T^{k}(v) = 0$ then the characteristic polynomial of $T_w$ is

$$f(t) = (-1)^k (a_0 + a_1 t + \ldots + a_{k-1} t^{k-1} + t^k).$$
For the following Definition and Theorem we refer to Appendix E in the back of our text book.

**Definition:** Let

\[ f(x) = a_0 + a_1(x) + \ldots + a_n x^n \]

be a polynomial with coefficients from a field \( F \). If \( T \) is a linear operator on a vector space \( V \) over \( F \), we define

\[ f(T) = a_0 I + a_1 T + \ldots + a_n T^n. \]

**Theorem E.3:** Let \( f(x) \) be a polynomial with coefficients from a field \( F \), and let \( T \) be a linear operator on a vector space \( V \) over \( F \). then \( f(T) \) is a linear operator on \( V \).

**Claim:** Let \( T \) be a linear operator on \( V \) then \( T^k \) is linear for all \( k \geq 1 \).

**Proof of Claim:** Let \( T \) be linear then \( T(cx + y) = cT(x) + T(y) \). For our base case we show that \( T^2 \) is linear.

\[
T^2(cx + y)
= T(T(cx + y))
= T(cT(x) + T(y))
= cT^2(x) + T^2(y)
\]

Which we can see is linear.

Assume that this holds for \( T^k \) where \( k < n \) that is it holds for \( k = 1, 2, \ldots, n - 1 \)

We want to show it holds for \( k = n \).

\[
T^k(cx + y)
= T^{k-1}(T(cx + y))
= T^{k-1}(cT(x) + T(y))
= cT^k(x) + T^{k-1}T(y)
= cT^k(x) + T^{k}(y)
\]

which is linear. Thus we have proved our claim.
Proof of Theorem E.3:
Define a function \( f : \mathbb{R} \to \mathbb{R} \) by \( f(x) = a_0 + a_1(x) + \ldots + a_n x^n \) and let \( T \) be a linear operator on \( V \). Since \( T \) is linear then
\[
T(cx + y) = cT(x) + T(y).
\]
We need to show that \( f(T) \) is linear.

\[
f(T)(cx + y) = a_0(cx + y) + a_1T(cx + y) + \ldots + a_n T^n(cx + y)
\]

By our claim which we proved above each \( T^k(cx + y) \) is linear so we have that

\[
f(T)(cx + y) = a_0(cI(x) + I(y)) + a_1(cT(x) + T(y)) + \ldots + a_n (cT^n(x) + T^n(y)
= ca_0 I(x) + a_0 I(y) + ca_1T(x) + a_1T(y) + \ldots + ca_n T^n(x) + a_n T(y)
= c(a_0 I(x) + a_1T(x) + \ldots + a_n T^n(x)) + (a_0 I(y) + a_1T(y) + \ldots + a_n T^n(y))
= cf(T)(x) + f(T)(y)
\]

So \( f(T) \) is linear. This completes the proof.
Theorem 5.23 (The Cayley-Hamilton theorem)
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 want to show that \( f(T)(v) = 0 \) for all \( v \in V \). Suppose \( v = 0 \). Of course, \( T(0) = 0 \) since \( T \) is linear. Also by theorem E.3 (appendix E) \( f(T) \) is a linear operator whenever \( f(t) \) is a polynomial with coefficients from a field \( F \). Therefore, since \( f(T) \) is linear we have that \( f(T)(0) = 0 \).

Let \( W \) be a \( T \)-cyclic subspace generated by \( v \) where \( v \) is any nonzero vector and suppose that \( \dim(W) = k \). By theorem 5.22(a), there exists scalars \( a_0, a_1, \ldots, a_{k-1} \) such that

\[
a_0 v + a_1 T(v) + \ldots + a_{k-1} T^{k-1}(v) + T^k(v) = 0
\]

So by theorem 5.22(b) we have that

\[
g(t) = (-1)^k (a_0 + a_1 t + \ldots + a_{k-1} t^{k-1} + t^k)
\]

is the characteristic polynomial of \( T_w \). Combining these two equations will give us

\[
g(T)(v) = (-1)^k (a_0 I + a_1 T + \ldots + a_{k-1} T^{k-1} + T^k)(v) = 0
\]

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

\[
f(T)(v) = (q(T)g(T))(v) = q(T)(v)g(T)(v) = q(T)(v) = 0
\]