CHAPTER 4

Minimal Polynomial and Cayley-Hamilton
Theorem

Notations

- $\mathbb{R}$ is the set of real numbers.
- $\mathbb{C}$ is the set of complex numbers.
- $\mathbb{Q}$ is the set of rational numbers.
- $\mathbb{Z}$ is the set of integers.
- $\mathbb{N}$ is the set of non-negative integers.
- $\mathbb{Z}^+$ is the set of positive integers.
- $\text{Re}(z)$, $\text{Im}(z)$, $\overline{z}$ and $|z|$ are the real part, imaginary part, conjugate and modulus of a complex number $z$.
- $M_{m \times n}(F)$ is the set of $m \times n$ matrices with entries in $F$.
- $F(x)$ is the set of polynomials in $x$ with coefficients in $x$.
- $\text{Row}(A)$, $\text{Col}(A)$ and $\text{Nul}(A)$ are the row, column and null spaces of a matrix $A$, respectively.
- $\text{rank}(A)$ is the rank of a matrix $A$.
- $I_n = [e_1 \ e_2 \ ... \ e_n]$ is the $n \times n$ identity matrix, where $e_1, e_2, ..., e_n$ are the column vectors of $I_n$.
- $K(T)$, $R(T)$ and $\text{rank}(T)$ are the kernel, range and rank of a linear transformation $T$.
- $\gcd(f_1(x), f_2(x), \ldots, f_m(x))$ is the greatest common divisor of polynomials $f_1(x), f_2(x), \ldots, f_m(x)$.
- $\text{lcm}(f_1(x), f_2(x), \ldots, f_m(x))$ is the least common multiple of polynomials $f_1(x), f_2(x), \ldots, f_m(x)$.

1. Basics on Polynomials

Let $F$ be a field. The space $F[x]$ of polynomials of $x$ with coefficients in $F$ is a vector space over $F$ equipped with vector addition $f(x) + g(x)$ and scalar multiplication $cf(x)$ for $f(x), g(x) \in F[x]$ and $c \in F$. It is more than a vector space as the product $f(x)g(x)$ of two polynomials is also defined. It is actually a ring and more precisely an algebra over $F$. We are not going to define ring or algebra as they are beyond the scope of this course. For us, $F[x]$ is a very concrete space and can be understood without much abstract machinery.

One of the most important properties of $F[x]$ is that one can do do long division in $F[x]$. For a pair of polynomials $f(x)$ and $g(x) \neq 0$, there exist
polynomials $q(x)$ and $r(x)$ such that $\deg r(x) < \deg g(x)$ and

$$f(x) = q(x)g(x) + r(x)$$

where $q(x)$ and $r(x)$ are called the quotient and the remainder for $f(x)$ divided by $g(x)$, respectively. Here we let $\deg 0 = -1$. The division is usually carried out in the same fashion as the division of two numbers. For examples,

$$\begin{array}{c|cccc}
\multicolumn{1}{r}{} & x^2 - 1 & x^3 - 2x + 1 & -x^3 & +x \\
\hline
x+2 & x^4 & -x^4 & -2x^3 & 2x^3 + 4x^2 \\
& -2x^3 & 4x^2 & -4x^2 & -8x \\
& & -8x & 8x & +16 \\
& & & & 15 \\
\end{array}$$

We say $g(x)$ divides $f(x)$ or $f(x)$ is divisible by $g(x)$ if $r(x) = 0$.

Every $c \in F^* = F \setminus \{0\}$ has a multiplicative inverse in $F[x]$ and hence divides every polynomial $f(x) \in F[x]$ so $c \in F^*$ is called a unit in $F[x]$.

When we divide $f(x)$ by $x - b$, we obtain

$$f(x) = (f(x) - f(b)) + f(b)$$

$$= ((a_0 + a_1 x + \cdots + a_n x^n) - (a_0 + a_1 b + \cdots + a_n b^n)) + f(b)$$

$$= a_1(x - b) + a_2(x^2 - b^2) + \cdots + a_n(x^n - b^n) + f(b)$$

$$= (x - b) \sum_{m=1}^{n} \frac{a_m x^m - b^m}{x - b} + f(b)$$

$$= (x - b) \sum_{j=0}^{m-1} \frac{x_{m-1-j} b^j}{r(x)}$$

where $x^m - b^m$ is divisible by $x - b$ because

$$x^m - b^m = (x - b) \sum_{j=0}^{m-1} x_{m-1-j} b^j$$

$$= (x - b)(x^{m-1} + x^{m-2} b + \cdots + x b^{m-2} + b^{m-1})$$

So $f(x)$ is divisible by $x - b$ if and only if $f(b) = 0$, i.e., $b$ is a root of $f(x)$. 

A polynomial \( f(x) \) of degree \( f(x) \geq 1 \) is **irreducible** if it is not divisible by any \( g(x) \in F[x] \) of \( 1 \leq \deg g(x) < \deg f(x) \), or equivalent, it is only divisible by a unit \( c \in F^* \) or \( cf(x) \) for \( c \in F^* \). Every polynomial of degree 1 is obviously irreducible. The notion of irreducibility is “sensitive” to the ground field \( F \). For example, \( x^2 + x + 1 \) is irreducible in \( \mathbb{R}[x] \) but reducible in \( \mathbb{C}[x] \) since

\[
(4.6) \quad x^2 + x + 1 = \left( x - \frac{-1 + \sqrt{3}i}{2} \right) \left( x - \frac{-1 - \sqrt{3}i}{2} \right).
\]

Using long division, we can prove that the **unique factorization** of a polynomial \( f(x) \) into a product of irreducible ones.

**Theorem 4.1 (Unique Factorization of Polynomials).** Every \( f(x) \in \mathbb{F}[x] \) is a product

\[
(4.7) \quad f(x) = c(f_1(x))^{d_1}(f_2(x))^{d_2} \ldots (f_m(x))^{d_m}
\]

where \( c \in F \), \( d_1, d_2, \ldots, d_m \) are positive integers and \( f_1(x), f_2(x), \ldots, f_m(x) \) are distinct irreducible polynomials with leading coefficient 1 (monic polynomials). Such factorization is unique up to a permutation of \( (f_i(x))^{d_i} \).

In (4.7), each \( f_i(x) \) is called an **irreducible (or prime) factor** of \( f(x) \) and \( d_i \) is its multiplicity in \( f(x) \).

Over \( \mathbb{C} \), every irreducible polynomial has degree 1 and every polynomial \( f(x) \in \mathbb{C}[x] \) is a product of polynomials of degree 1. This result is called **Fundamental Theorem of Algebra**.

**Theorem 4.2 (Fundamental Theorem of Algebra).** Every \( f(x) \in \mathbb{C}[x] \) is a product

\[
(4.8) \quad f(x) = c(x - r_1)(x - r_2) \ldots (x - r_n)
\]

for some \( c, r_1, r_2, \ldots, r_n \in \mathbb{C} \).

Every \( f(x) \in \mathbb{R}[x] \) is a product

\[
(4.9) \quad f(x) = c(x - r_1)(x - r_2) \ldots (x - r_l) \left( x^2 + a_1x + b_1 \right) \left( x^2 + a_2x + b_2 \right) \ldots \left( x^2 + a_mx + b_m \right)
\]

for some \( c, r_1, r_2, \ldots, r_l \in \mathbb{R} \) and \( a_1, b_1, a_2, b_2, \ldots, a_m, b_m \in \mathbb{R} \) satisfying that

\[
a_j^2 - 4b_j < 0 \quad \text{for} \quad j = 1, 2, \ldots, m.
\]

**Proof.** There are many proofs of (4.8). All require some amount of complex analysis. We will not do it here. The interested readers may consult a standard textbook on the subject.

We will only show how to derive (4.9) from (4.8). Let us consider a real polynomial \( f(x) \in \mathbb{R}[x] \) as a polynomial \( f(x) \in \mathbb{C}[x] \). Then we have (4.8) for some \( r_1, r_2, \ldots, r_n \in \mathbb{C} \). Taking complex conjugate on both sides, we obtain

\[
(4.10) \quad \overline{f(x)} = \overline{c}(x - \overline{r_1})(x - \overline{r_2}) \ldots (x - \overline{r_n}).
\]
Since \( f(x) \) is a real polynomial, \( \overline{f(x)} = f(x) \). Also \( c = c \) since \( c \) is the lead coefficient of \( f(x) \). So we have
\[
(x - r_1)(x - r_2) \ldots (x - r_n) = (x - \overline{r}_1)(x - \overline{r}_2) \ldots (x - \overline{r}_n).
\]
In other words, the complex roots of \( f(x) \) come in conjugate pairs. That is, if \( x - r \) divides \( f(x) \), \( x - \overline{r} \) also divides \( f(x) \). Let us write
\[
f(x) = c(x - r_1)(x - r_2) \ldots (x - r_l)
\]
\[
(x - s_1)(x - \overline{s}_1)(x - s_2)(x - \overline{s}_2) \ldots (x - s_m)(x - \overline{s}_m)
\]
\[
= c(x - r_1)(x - r_2) \ldots (x - r_l)
\]
\[
(x^2 + a_1x + b_1)(x^2 + a_2x + b_2) \ldots (x^2 + a_mx + b_m)
\]
where \( r_1, r_2, \ldots, r_l \in \mathbb{R} \), \( s_1, s_2, \ldots, s_m \in \mathbb{C} \setminus \mathbb{R} \)
and
\[
x^2 + a_jx + b_j = (x - s_1)(x - \overline{s}_1) = x^2 - (s_j + \overline{s}_j)x + s_j\overline{s}_j
\]
for \( j = 1, 2, \ldots, m \).

Although every complex polynomial \( f(x) \) is simply factorized into polynomials of degree 1 in (4.8), there is no algorithm to find the exact values of its roots \( r_j \). There is no close formula for roots of polynomials of degree at least 5. In general, we can only approximate \( r_j \) using numerical methods. On the other hand, given two polynomials \( f(x) \) and \( g(x) \), there are efficient algorithms to find their greatest common divisor.

The greatest common divisor of \( m \) polynomials \( f_1(x), f_2(x), \ldots, f_m(x) \), written as \( \gcd(f_1(x), f_2(x), \ldots, f_m(x)) \), is a polynomial of the highest degree that divides all \( f_1(x), f_2(x), \ldots, f_m(x) \). Similarly, the least common multiple of \( f_1(x), f_2(x), \ldots, f_m(x) \), written as \( \text{lcm}(f_1(x), f_2(x), \ldots, f_m(x)) \), is a polynomial of the lowest degree that is divisible by all \( f_1(x), f_2(x), \ldots, f_m(x) \). We say two polynomials \( f_1(x) \) and \( f_2(x) \) are coprime if \( \gcd(f_1(x), f_2(x)) = 1 \).

If each \( f_i(x) \) is factorized into irreducible polynomials as
\[
f_i(x) = (p_1(x))^{d_{i1}}(p_2(x))^{d_{i2}} \ldots (p_n(x))^{d_{im}},
\]
then
\[
\gcd(f_1(x), f_2(x), \ldots, f_m(x))
\]
\[
= (p_1(x))^{\min(d_{11})}(p_2(x))^{\min(d_{12})} \ldots (p_n(x))^{\min(d_{1m})}
\]
\[
\text{lcm}(f_1(x), f_2(x), \ldots, f_m(x))
\]
\[
= (p_1(x))^{\max(d_{11})}(p_2(x))^{\max(d_{12})} \ldots (p_n(x))^{\max(d_{1m})}.
\]
From this, we can derive formulas relating \( \gcd \) and \( \text{lcm} \) as
\[
\text{lcm}(f(x), g(x)) = \frac{f(x)g(x)}{\gcd(f(x), g(x))}.
\]
More general, they are related via inclusion-exclusion principle: We write \( \gcd(S) \) and \( \text{lcm}(S) \) for the greatest common divisor and the least common
The greatest common divisor of two polynomials can be effectively computed by *Euclidean Algorithm*, which is basically a sequence of long divisions. Given two polynomials $f_1(x)$ and $f_2(x)$, we do long divisions
\begin{equation}
\begin{aligned}
f_1(x) &= q_1(x)f_2(x) + f_3(x), \quad \text{deg } f_3(x) < \text{deg } f_2(x) \\
f_2(x) &= q_2(x)f_3(x) + f_4(x), \quad \text{deg } f_4(x) < \text{deg } f_3(x) \\
&\quad \vdots \\
f_{n-1}(x) &= q_{n-1}(x)f_n(x) + f_{n+1}(x), \quad \text{deg } f_{n+1}(x) < \text{deg } f_n(x)
\end{aligned}
\end{equation}
and stop when $f_{n+1}(x) = 0$. Then $\gcd(f_1(x), f_2(x)) = f_n(x)$. The greatest common divisor of $m$ polynomials $g_1(x), g_2(x), \ldots, g_m(x)$ can be computed inductively by
\begin{equation}
\gcd(g_1(x), g_2(x), \ldots, g_m(x)) = \gcd\left(\gcd(g_1(x), g_2(x)), \ldots, g_m(x)\right).
\end{equation}

For examples,
\begin{equation}
\begin{aligned}
x^3 - 1 &= (x^2 - 3x + 2) \cdot (x + 3) + (7x - 7) \\
x^2 - 3x + 2 &= (7x - 7) \cdot \left(\frac{1}{2}x - \frac{3}{2}\right) + 0
\end{aligned}
\end{equation}
and hence $\gcd(x^3 - 1, x^2 - 3x + 2) = x - 1$;
\begin{equation}
\begin{aligned}
x^4 - 2x^2 + 1 &= (x^3 - x^2 + x - 1) \cdot (x + 1) + (-2x^2 + 2) \\
x^3 - x^2 + x - 1 &= (-2x^2 + 2) \cdot (-\frac{1}{2}x + \frac{1}{2}) + (2x - 2) \\
-2x^2 + 2 &= (2x - 2) \cdot (-x - 1) + 0
\end{aligned}
\end{equation}
and hence $\gcd(x^4 - 2x^2 + 1, x^3 - x^2 + x - 1) = x - 1$.

**Theorem 4.3 (Bezout Identity).** For two polynomials $f_1(x)$ and $f_2(x)$, there exist polynomials $g_1(x)$ and $g_2(x)$ such that
\begin{equation}
g_1(x)f_1(x) + g_2(x)f_2(x) = \gcd(f_1(x), f_2(x)).
\end{equation}

**Proof.** Applying Euclidean algorithm to $f_1(x)$ and $f_2(x)$, we obtain
\begin{equation}
\begin{aligned}
f_1(x) &= q_1(x)f_2(x) + f_3(x), \quad \text{deg } f_3(x) < \text{deg } f_2(x) \\
f_2(x) &= q_2(x)f_3(x) + f_4(x), \quad \text{deg } f_4(x) < \text{deg } f_3(x) \\
&\quad \vdots \\
f_{n-1}(x) &= q_{n-1}(x)f_n(x) + f_{n+1}(x), \quad f_{n+1}(x) = 0
\end{aligned}
\end{equation}
Then \( f_n(x) = \gcd(f_1(x), f_2(x)) \) and a Bezout identity (4.22) can be inductively derived by

\[
a_n = 1, b_n = 0 \Rightarrow a_nf_n + b_nf_{n+1} = \gcd(f_1, f_2)
\]

\[
a_{n-1} = b_n, b_{n-1} = a_n - q_{n-1}b_n \Rightarrow a_{n-1}f_{n-1} + b_{n-1}f_n = \gcd(f_1, f_2)
\]

\[
\vdots \Rightarrow \vdots
\]

\[
a_1 = b_2, b_1 = a_2 - q_1b_2 \Rightarrow a_1f_1 + b_1f_2 = \gcd(f_1, f_2)
\]

\[
\square
\]

For example, in (4.20), we have

\[
1 \cdot (7x - 7) + 0 \cdot 0 = 7x - 7
\]

\[
0 \cdot (x^2 - 3x + 2) + 1 \cdot (7x - 7) = 7x - 7
\]

\[
1 \cdot (x^3 - 1) + (-x + 3) \cdot (x^2 - 3x + 2) = 7x - 7
\]

with the last line being a Bezout identity of \( x^3 - 1 \) and \( x^2 - 3x + 2 \). In (4.21), we have

\[
1 \cdot (2x - 2) + 0 \cdot 0 = 2x - 2
\]

\[
0 \cdot (-2x^2 + 2) + 1 \cdot (2x - 2) = 2x - 2
\]

\[
1 \cdot (x^3 - x^2 + x - 1) + \left(\frac{1}{2}x - \frac{1}{2}\right) \cdot (-2x^2 + 2) = 2x - 2
\]

\[
\left(\frac{1}{2}x - \frac{1}{2}\right) \cdot (x^4 - 2x^2 + 1) + \left(\frac{3}{2} - \frac{x^2}{2}\right) \cdot (x^3 - x^2 + x - 1) = 2x - 2
\]

with the last line being a Bezout identity of \( x^4 - 2x^2 + 1 \) and \( x^3 - x^2 + x - 1 \).

In the special case that \( \gcd(f_1(x), f_2(x)) = 1 \), there exist polynomials \( g_1(x) \) and \( g_2(x) \) such that

\[
g_1(x)f_1(x) + g_2(x)f_2(x) = 1.
\]

Indeed, \( \gcd(f_1(x), f_2(x)) = 1 \) if and only if (4.27) holds for some \( g_j(x) \).

Bezout identity holds for multiple polynomials as well.

**Corollary 4.4.** For \( m \) polynomials \( f_1(x), f_2(x), \ldots, f_m(x) \), there exist polynomials \( g_1(x), g_2(x), \ldots, g_m(x) \) such that

\[
\sum_{j=1}^{m} g_j(x)f_j(x) = \gcd(f_1(x), f_2(x), \ldots, f_m(x)).
\]

**Proof.** We prove (4.28) by induction. It is true for \( m = 2 \) by (4.22).

Suppose that (4.28) holds for \( m - 1 \). Then there exist polynomials \( h_1(x), h_2(x), \ldots, h_{m-1}(x) \) such that

\[
\sum_{j=1}^{m-1} h_j(x)f_j(x) = \gcd(f_1(x), f_2(x), \ldots, f_{m-1}(x)).
\]
Applying (4.22) to $\gcd(f_1(x), f_2(x), \ldots, f_{m-1}(x))$ and $f_m(x)$, we obtain two polynomials $p(x)$ and $q(x)$ such that

\[
p(x) \gcd(f_1(x), f_2(x), \ldots, f_{m-1}(x)) + q(x)f_m(x) = \gcd(f_1(x), f_2(x), \ldots, f_{m-1}(x), f_m(x))
\]

(4.30)

Combining (4.29) and (4.30), we obtain

\[
p(x) \sum_{j=1}^{m-1} h_j(x)f_j(x) + q(x)f_m(x) = \gcd(f_1(x), f_2(x), \ldots, f_m(x)).
\]

(4.31)

Setting $g_j(x) = p(x)h_j(x)$ for $j = 1, 2, \ldots, m - 1$ and $g_m(x) = q(x)$, we obtain (4.28).

**Example 4.5 (Partial Fractions).** We can always write a rational function $g(x)/f(x)$ as a sum of partial fractions, which is actually an application of Bezout identity. Suppose that

\[
f(x) = (x - \lambda_1)^{d_1}(x - \lambda_2)^{d_2} \cdots (x - \lambda_m)^{d_m}
\]

for $\lambda_1, \lambda_2, \ldots, \lambda_m$ distinct. Applying Bezout identity (4.28) to

\[
f_1(x) = \frac{f(x)}{x - \lambda_1} = (x - \lambda_2)^{d_2} \cdots (x - \lambda_m)^{d_m}
\]

(4.33)

\[
f_2(x) = \frac{f(x)}{x - \lambda_2} = (x - \lambda_1)^{d_1}(x - \lambda_3)^{d_3} \cdots (x - \lambda_m)^{d_m}
\]

we obtain

\[
\sum_{j=1}^{m} g_j(x)f_j(x) = \gcd(f_1(x), f_2(x), \ldots, f_m(x)) = 1
\]

(4.34)

Therefore,

\[
\frac{g(x)}{f(x)} = \sum_{j=1}^{m} \frac{g_j(x)f_j(x)g(x)}{f(x)} = \sum_{j=1}^{m} \frac{g_j(x)g(x)}{(x - \lambda_j)^{d_j}}.
\]

(4.35)

For each $j$, we divide $g_j(x)g(x)$ by $(x - \lambda_j)^{d_j}$ by long division:

\[
\frac{g_j(x)g(x)}{(x - \lambda_j)^{d_j}} = q_j(x) + \frac{r_j(x)}{(x - \lambda_j)^{d_j}}.
\]

(4.36)
where \( \deg r_j < d_j \). Then we expand each \( r_j(x) \) as a Taylor polynomial in \( (x - \lambda_j)^k \):

\[
\frac{g_j(x)g(x)}{(x - \lambda_j)^{d_j}} = q_j(x) + \frac{r_j(x)}{(x - \lambda_j)^{d_j}}
\]

\[(4.37)\]

\[
= q_j(x) + \frac{1}{(x - \lambda_j)^{d_j}} \sum_{k=0}^{d_j-1} \frac{r_j^{(k)}(\lambda_j)}{k!} (x - \lambda_j)^k
\]

\[
= q_j(x) + \sum_{k=0}^{d_j-1} \frac{r_j^{(k)}(\lambda_j)}{k!} \frac{1}{(x - \lambda_j)^{d_j-k}}.
\]

In this way, we have written

\[(4.38)\]

\[
g(x) = \sum_{j=1}^{m} q_j(x) + \sum_{j=1}^{m} \sum_{k=0}^{d_j-1} \frac{r_j^{(k)}(\lambda_j)}{k!} \frac{1}{(x - \lambda_j)^{d_j-k}}
\]

as a sum of partial fractions.

From time to time, we need to check whether a polynomial \( f(x) \) has a multiple irreducible factor, i.e., one with multiplicity \( \geq 2 \). This can be verified efficiently by computing \( \gcd(f(x), f'(x)) \) by Euclidean algorithm.

**Proposition 4.6.** A real or complex polynomial \( f(x) \) has an irreducible factor with multiplicity \( \geq 2 \) if and only if

\[(4.39)\]

\[\gcd(f(x), f'(x)) = 1\]

or equivalently, \( g_1(x)f(x) + g_2(x)f'(x) = 1 \) for some polynomials \( g_1(x) \) and \( g_2(x) \).

**Proof.** For simplicity, we assume that \( f(x) \) has leading coefficient 1. Suppose that \( f(x) \) is factorized into irreducible polynomials as

\[(4.40)\]

\[f = f_1^{d_1} f_2^{d_2} \cdots f_m^{d_m}\]

for distinct irreducible polynomials \( f_1, f_2, \ldots, f_m \).

We differentiate \( (4.40) \) on both sides:

\[
f = (f_1^{d_1} f_2^{d_2} \cdots f_m^{d_m})'
\]

\[(4.41)\]

\[= d_1 f_1^{d_1-1} f_2^{d_2} \cdots f_m^{d_m} + d_2 f_1^{d_1} f_2^{d_2-1} f_3^{d_3} \cdots f_m^{d_m}
\]

\[+ \cdots + d_m f_1^{d_1} f_2^{d_2} \cdots f_m^{d_m-1} f_1^{d_1-1} \cdots f_m^{d_m-1}
\]

\[= f_1^{d_1-1} f_2^{d_2-1} \cdots f_m^{d_m-1} g
\]

where

\[(4.42)\]

\[g(x) = d_1 f_1'(x) f_2(x) \cdots f_m(x) + d_2 f_1(x) f_2'(x) f_3(x) \cdots f_m(x)
\]

\[+ \cdots + d_m f_1(x) f_2(x) \cdots f_m(x) f_m'(x).
\]
Since \( f_k(x) \) does not divide \( f'_k(x) \), \( f_k(x) \) does not divide \( g(x) \) for all \( k = 1, 2, \ldots, m \). And since \( f_k(x) \) is irreducible,

\[
(4.43) \quad \gcd(f_1f_2\ldots f_m, g) = 1.
\]

Therefore,

\[
(4.44) \quad \gcd(f, f') = f^{d_1-1}f_2^{d_2-1}\ldots f_m^{d_m-1}\gcd(f_1f_2\ldots f_m, g) = f^{d_1-1}f_2^{d_2-1}\ldots f_m^{d_m-1}.
\]

Thus, \( d_1 = d_2 = \ldots = d_m = 1 \) if and only if \( \gcd(f(x), f'(x)) = 1 \). \( \square \)

For example, \( f(x) = x^n + x + 1 \) has no multiple roots in \( \mathbb{C} \) since

\[
\begin{align*}
\gcd(f(x), f'(x)) &= \gcd(x^n + x + 1, nx^{n-1} + 1) \\
&= \gcd\left(\frac{n-1}{n} x + 1, nx^{n-1} + 1\right) \\
&= \gcd\left(\frac{n-1}{n} x + 1, (-1)^{n-1} \frac{n^n}{(n-1)^{n-1}} + 1\right) \\
&= 1.
\end{align*}
\]

In Theorem 3.15, the hypothesis on polynomial \( f(x) \) in (3.83) or (3.86) simply means that \( f(x) \) has no multiple roots if we work over \( \mathbb{C} \). Combining with (4.39), we obtain another criterion for matrices \( A \) or linear endomorphisms \( T \) to be diagonalizable.

**Proposition 4.7.** A matrix \( A \in M_{n \times n}(\mathbb{C}) \) or a linear endomorphism \( T : V \to V \) on a finite-dimensional vector space over \( \mathbb{C} \) is diagonalizable if and only if there exists a nonzero polynomial \( f(x) \in \mathbb{C}[x] \) such that \( \gcd(f(x), f'(x)) = 1 \) and \( f(A) = 0 \) (\( f(T) = 0 \)).

## 2. Minimal Polynomial

The **minimal polynomial** of a square matrix \( A \) or a linear endomorphism \( T : V \to V \) on a finite-dimensional vector space \( V \) is a nonzero polynomial \( f(x) \) of the lowest degree such that \( f(A) = 0 \) or \( f(T) = 0 \). Let us first show that there exists at least one polynomial such that \( f(A) = 0 \) or \( f(T) = 0 \).

**Theorem 4.8 (Existence of Minimal Polynomials).** For every \( n \times n \) matrix \( A \) or every linear endomorphism \( T : V \to V \) on a vector space \( V \) of dimension \( n \), there exists a polynomial \( f(x) \) of degree \( 1 \leq \deg f(x) \leq n^2 \) such that \( f(A) = 0 \) or \( f(T) = 0 \).

**Proof.** Since the vector space \( M_{n \times n}(F) \) (or \( L(V, V) \)) has dimension \( n^2 \), \( I, A, A^2, \ldots, A^{n^2} \) (\( I, T, T^2, \ldots, T^{n^2} \)) are linearly dependent over \( F \). So

\[
(4.46) \quad \sum_{m=0}^{n^2} c_m A^m = 0 \quad (\sum_{m=0}^{n^2} c_m T^m = 0)
\]
for some \(c_0, c_1, c_2, \ldots, c_{n^2} \in F\), not all zero, i.e., \(f(A) = 0 \ (f(T) = 0)\) for \(n^2\) \(f(x) = \sum_{m=0}^{n^2} c_m x^m = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n^2} x^{n^2}\).

Since \(c_m\) are not all zero, \(f(x) \neq 0\). Moreover, \(\deg f(x) > 0\); otherwise, we have \(c_0 \neq 0\), \(c_1 = c_2 = \cdots = c_{n^2} = 0\) and \(f(A) = c_0 I = 0\), which is impossible. Therefore, \(f(A) = 0 \ (f(T) = 0)\) for some \(f(x) \in F[x]\) satisfying \(1 \leq \deg f(x) \leq n^2\). \(\square\)

So the minimal polynomial of \(A \in M_{n \times n}(F)\) has degree at most \(n^2\). As we will see, this upper bound is way too high and the optimal upper bound for all \(A\) is \(n\), which is essentially Cayley-Hamilton theorem.

**Theorem 4.9 (Uniqueness of Minimal Polynomials).** For a square matrix \(A\) or a linear endomorphism \(T\) on a finite-dimensional vector space, the following holds:

1. If \(f_1(A) = f_2(A) = 0\) or \(f_1(T) = f_2(T) = 0\) for two polynomials \(f_1(x)\) and \(f_2(x)\), then \(g(A) = 0\) or \(g(T) = 0\) for the greatest common divisor \(g(x) = \gcd(f_1(x), f_2(x))\) of \(f_1(x)\) and \(f_2(x)\);
2. If \(f_1(x)\) is a minimal polynomial of \(A\) or \(T\) and \(f_2(x)\) is a polynomial such that \(f_2(A) = 0\) or \(f_2(T) = 0\), then \(f_1(x)\) divides \(f_2(x)\);
3. The minimal polynomial of \(A\) or \(T\) is unique up to a scalar.

**Proof.** We will prove these statements for square matrices \(A\). The very same argument works for linear endomorphisms \(T\).

There exist polynomials \(g_1(x)\) and \(g_2(x)\) such that

\[ g_1(x)f_1(x) + g_2(x)f_2(x) = g(x) = \gcd(f_1(x), f_2(x)) \]  

So \(g(A) = g_1(A)f_1(A) + g_2(A)f_2(A) = 0\). This proves part (1).

For part (2), let \(d = \deg f_1(x)\) and \(g(x) = \gcd(f_1(x), f_2(x))\). By part (1), \(g(A) = 0\). Since \(g(x)\) divides \(f_1(x)\), \(\deg g(x) \leq d\). And since \(f_1(x)\) is minimal, \(\deg g(x) = d\). So \(f_1(x) = cg(x)\) for some \(c \neq 0\) and \(f_1(x)\) divides \(f_2(x)\).

For part (3), suppose that \(f_1(x)\) and \(f_2(x)\) are two minimal polynomials of \(A\). Then by part (2), \(f_1(x)\) divides \(f_2(x)\) and \(f_2(x)\) divides \(f_1(x)\). Therefore, \(f_1(x) = cf_2(x)\) for some \(c \neq 0\). \(\square\)

### 3. Cayley-Hamilton Theorem

In this section, we are going to prove Cayley-Hamilton Theorem:

**Theorem 4.10 (Cayley-Hamilton Theorem).** For a square matrix \(A\) or a linear endomorphism \(T\) on a finite-dimensional vector space \(V\), \(f(A) = 0\) or \(f(T) = 0\) for the characteristic polynomial \(f(x)\) of \(A\) or \(T\). Or equivalently, the minimal polynomial of \(A\) or \(T\) divides its characteristic polynomial.
We will prove this theorem over \( \mathbb{C} \). Clearly, if \( f(A) = 0 \) for the characteristic polynomial \( f(x) = \det(xI - A) \) and all complex square matrices \( A \), the same is true for all real square matrices as well. Actually, Cayley-Hamilton holds over all fields \( F \).

We start with a useful observation:

**Proposition 4.11.** Let \( A \) be a square matrix in the form of

\[
A = \begin{bmatrix}
A_1 & & \\
 & A_2 & \\
& & \ddots \\
& & & A_m
\end{bmatrix}
\]

where \( A_1, A_2, \ldots, A_m \) are square matrices. Then the characteristic polynomial of \( A \) is

\[
\det(xI - A) = \det(xI - A_1) \det(xI - A_2) \ldots \det(xI - A_m)
\]

and a minimal polynomial \( f(x) \) of \( A \) is

\[
f(x) = \text{lcm}(f_1(x), f_2(x), \ldots, f_m(x))
\]

where \( f_1(x), f_2(x), \ldots, f_m(x) \) are minimal polynomials of \( A_1, A_2, \ldots, A_m \), respectively.

**Proof.** We can prove (4.50) easily by

\[
\det(xI - A) = \det \begin{bmatrix}
xI - A_1 \\
& xI - A_2 \\
& & \ddots \\
& & & xI - A_m
\end{bmatrix}
= \det(xI - A_1) \det(xI - A_2) \ldots \det(xI - A_m).
\]

To see (4.51), we observe that

\[
f(A) = \begin{bmatrix}
f(A_1) \\
& f(A_2) \\
& & \ddots \\
& & & f(A_m)
\end{bmatrix}
\]

So \( f(A) = 0 \) if and only if \( f(A_1) = f(A_2) = \ldots = f(A_m) = 0 \). This implies that a minimal polynomial \( f_j(x) \) of \( A_j \) must divide a minimal polynomial \( f(x) \) of \( A \) for all \( j = 1, 2, \ldots, m \). That is, \( f(x) \) is a common multiple of \( f_1(x), f_2(x), \ldots, f_m(x) \). And since \( f(x) \) has the lowest degree such that \( f(A) = 0 \), \( f(x) \) is the least common multiple of \( f_1(x), f_2(x), \ldots, f_m(x) \). \( \square \)

**Example 4.12 (Jordan Matrices).** We may use the above proposition to construct matrices with the required minimal polynomials. These examples will help us understand better the concept of minimal polynomials and Cayley-Hamilton theorem.
A Jordan block is a square matrix in the form

\[
J_{\lambda,n} = \begin{bmatrix}
\lambda & 1 & & \\
& \lambda & 1 & \\
& & \ddots & \ddots \\
& & & \lambda \\
\end{bmatrix}_{n \times n}
\]

For example,

\[
J_{\lambda,1} = [\lambda], \quad J_{2,2} = \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix}, \quad J_{-1,3} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \ldots
\]

It is easy to verify that

\[
J_{\lambda,n} = \lambda I_n + J_{0,n} \quad \text{and} \quad J_{\lambda,1,n} J_{\lambda,2,n} = J_{\lambda,2,n} J_{\lambda,1,n}.
\]

It is not hard to figure out the minimal polynomial of \(J_{0,n}\). Note that

\[
J_{0,n} e_1 = 0, \quad J_{0,n} e_2 = e_1, \ldots, \quad J_{0,n} e_n = e_{n-1}.
\]

Or we just write

\[
J_{0,n} e_j = e_{j-1}
\]

for \(j = 1, 2, \ldots, n\) if we set \(e_j = 0\) for \(j \leq 0\). Therefore,

\[
J_{0,n}^m e_j = e_{j-m} \Rightarrow J_{0,n}^n e_j = e_{j-n} = 0 \quad \text{for all} \quad j \Rightarrow J_{0,n}^n = 0.
\]

So \(f(J_{0,n}) = 0\) for \(f(x) = x^n\). On the other hand, there does not exist a nonzero polynomial \(f(x)\) of degree \(< n\) such that \(f(J_{0,n}) = 0\) since

\[
f(J_{0,n}) e_n = (a_0 I + a_1 J_{0,n} + \cdots + a_{n-1} J_{0,n}^{n-1}) e_n
\]

\[
= \sum_{m=0}^{n-1} a_m J_{0,n}^m e_n = \sum_{m=0}^{n-1} a_m e_{n-m} \neq 0
\]

for \(f(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1} \neq 0\). So \(x^n\) is a minimal polynomial of \(J_{0,n}\). And by (4.56), \((x - \lambda)^n\) is a minimal polynomial of \(J_{\lambda,n}\).

A Jordan matrix (or a matrix in Jordan canonical form) is a square matrix in the form

\[
\begin{bmatrix}
J_{\lambda_1,n_1} \\
& J_{\lambda_2,n_2} \\
& & \ddots \\
& & & J_{\lambda_k,n_k}
\end{bmatrix}
\]
For example, a diagonal matrix is a Jordan matrix:

\[
\begin{bmatrix}
\lambda_1 & & & \\
& \lambda_2 & & \\
& & \ddots & \\
& & & \lambda_n
\end{bmatrix}
= \begin{bmatrix}
J_{\lambda_1,1} & & & \\
& J_{\lambda_2,1} & & \\
& & \ddots & \\
& & & J_{\lambda_n,1}
\end{bmatrix}.
\]

For example, the following matrices are Jordan matrices:

\[
\begin{bmatrix}
1 & 1 & & \\
& 1 & 1 & \\
& & 1 & \\
& & & 1
\end{bmatrix}
= \begin{bmatrix}
1 & 1 & & \\
& 1 & 1 & \\
& & 1 & \\
& & & 1
\end{bmatrix}
= \begin{bmatrix}
J_{1,2} & \\
\end{bmatrix}.
\]

By Proposition 4.11 and our previous analysis, we see that the characteristic polynomial and minimal polynomial of the Jordan matrix (4.61) are

\[
\det(xI - J_{\lambda_1,n_1}) \det(xI - J_{\lambda_2,n_2}) \cdots \det(xI - J_{\lambda_k,n_k})
= (x - \lambda_1)^{n_1} (x - \lambda_2)^{n_2} \cdots (x - \lambda_k)^{n_k}
\]

and

\[
\text{lcm} \left( (x - \lambda_1)^{n_1}, (x - \lambda_2)^{n_2}, \ldots, (x - \lambda_k)^{n_k} \right),
\]

respectively. So both Jordan matrices in (4.63) have the same characteristic polynomial \((x - 1)^4\); but their minimal polynomials are

\[
\text{lcm} \left( (x - 1)^2, (x - 1)^2 \right) = (x - 1)^2 \text{ and}
\]

\[
\text{lcm} \left( (x - 1)^3, x - 1 \right) = (x - 1)^3,
\]

respectively.

We may use the above to construct square matrices with given minimal polynomials. For example, to construct a 10 \times 10 matrix with minimal polynomial \(x^2(x - 1)^3(x - 2)^4\), we choose a Jordan matrix (4.61) with \(\lambda_j\) being one of 0, 1, 2 satisfying

\[
n_1 + n_2 + \cdots + n_k = 10
\]

\[
\max_{\lambda_j=0}(n_j) = 2, \ max_{\lambda_j=1}(n_j) = 3, \ and \ max_{\lambda_j=2}(n_j) = 4.
\]
In other words, among the Jordan blocks $J_{\lambda_i,n_j}$ in (4.61), we need one $J_{0,2}$, one $J_{1,3}$ and one $J_{2,4}$. So

(4.68)

$$A = \begin{bmatrix}
J_{0,2} & & \\
& J_{1,3} & \\
& & J_{2,4}
\end{bmatrix} = \begin{bmatrix}
0 & 1 & & \\
& 1 & 1 & 1 \\
& 2 & 1 & & \\
& & & 0
\end{bmatrix}$$

is a $10 \times 10$ matrix with minimal polynomial $x^2(x - 1)^3(x - 2)^4$. Such Jordan matrix is not unique: $\lambda$ can be any of 0, 1, 2 and the Jordan blocks $J_{\lambda_i,n_j}$ can be permuted on the diagonal.

We can also formulate Proposition 4.11 for linear endomorphisms. For that purpose, let us introduce the concept of $T$-invariant subspace. For a linear endomorphism $T : V \to V$, we call a subspace $W$ $T$-invariant if $T(W) \subset W$. Clearly, the kernel and range of $T$ are $T$-invariant. More generally, we have

**Proposition 4.13.** For two linear endorphisms $T_1$ and $T_2$ on a vector space $V$, if $T_1 \circ T_2 = T_2 \circ T_1$, then $K(T_1)$ and $R(T_1)$ are $T_2$-invariant and $K(T_2)$ and $R(T_2)$ are $T_1$-invariant.

In particular, for a linear endomorphism $T$ on a vector space $V$ over $F$ and a polynomial $f(x) \in F[x]$, the subspaces $K(f(T))$ and $R(f(T))$ are $T$-invariant.

**Proof.** For every $v \in K(T_1)$, $T_1(v) = 0$. Therefore,

(4.69)

$$T_2(T_1(v)) = 0 \Rightarrow T_2 \circ T_1(v) = 0 \Rightarrow T_1 \circ T_2(v) = 0 \Rightarrow T_1(T_2(v)) = 0 \Rightarrow T_2(v) \in K(T_1).$$

Hence $T_2(K(T_1)) \subset K(T_1)$ and $K(T_1)$ is $T_2$-invariant.

For $R(T_1)$, we have

(4.70)

$$T_2(R(T_1)) = T_2(T_1(V)) = T_2 \circ T_1(V) = T_2(T_2(V)) \subset T_1(V) = R(T_1)$$

because $T_2(V) \subset V$. Hence $R(T_1)$ is $T_2$-invariant.
The same argument proves that $K(T_2)$ and $R(T_2)$ are $T_1$-invariant. Applying this to $T_1 = T$ and $T_2 = f(T)$, we conclude that $K(f(T))$ and $R(f(T))$ are $T$-invariant.

It is also easy to see that the sum and the intersections of $T$-invariant subspaces are also $T$-invariant.

Likewise, for an $n \times n$ matrix $A \in M_{n \times n}(F)$, we may also call a subspace $W$ $A$-invariant if $T_A(W) \subset W$ for $T_A(v) = Av$, although this term is not widely acceptable for matrices.

**Proposition 4.14.** Let $T$ be a linear endomorphism on a finite-dimensional vector space $V$ over $F$. If

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_m$$

for some $T$-invariant subspaces $V_1, V_2, \ldots, V_m$ of $V$, then the characteristic polynomial of $T$ is

$$\det(xI - T) = \det(xI - T_1) \det(xI - T_2) \cdots \det(xI - T_m)$$

and a minimal polynomial $f(x)$ of $T$ is

$$f(x) = \text{lcm}(f_1(x), f_2(x), \ldots, f_m(x))$$

where $T_k : V_k \to V_k$ are the restrictions of $T$ to $V_k$ for $k = 1, 2, \ldots, m$ and $f_1(x), f_2(x), \ldots, f_m(x)$ are minimal polynomials of $T_1, T_2, \ldots, T_m$, respectively.

**Proof.** We choose a basis $B_k$ for each $V_k$. Since $V$ is a direct sum of $V_k$, $B = B_1 \cup B_2 \cup \ldots \cup B_m$ is a basis of $V$. We order the vectors in $B$ in the way of

$$B = \left\{ \begin{array}{c}
\{v_1, v_2, \ldots, v_{b_1}\} \\
\{v_{b_1+1}, v_{b_1+2}, \ldots, v_{b_2}\} \\
\vdots \\
\{v_{b_{m-1}+1}, v_{b_{m-1}+2}, \ldots, v_{b_m}\} \\
\end{array} \right\}$$

We claim that

$$[T]_{B \leftarrow B} = \begin{bmatrix}
[T_1]_{B_1 \leftarrow B_1} & & \\
& [T_2]_{B_2 \leftarrow B_2} & \\
& & \ddots \\
& & & [T_m]_{B_m \leftarrow B_m}
\end{bmatrix}.$$

To simplify our notations, we write $A_k = [T_k]_{B_k \leftarrow B_k}$ for $k = 1, 2, \ldots, m$.

For a vector $w \in V$, let us write

$$w = w_1 + w_2 + \cdots + w_m$$
for \( w_k \in V_k \) for \( k = 1, 2, \ldots, m \). Then

\[
[w]_B = \begin{bmatrix}
[w_1]_{B_1} \\
[w_2]_{B_2} \\
\vdots \\
[w_m]_{B_m}
\end{bmatrix}
\]

Since \( V_k \) is \( T \)-invariant, \( T(V_k) \subset V_k \) and hence \( T(w_k) = T_k(w_k) \). Therefore,

\[
T(w) = T(w_1) + T(w_2) + \cdots + T(w_m)
\]

(4.78)

\[
= T_1(w_1) + T_2(w_2) + \cdots + T_m(w_m)
\]

and hence

\[
[T(w)]_B = \begin{bmatrix}
[T_1(w_1)]_{B_1} \\
[T_2(w_2)]_{B_2} \\
\vdots \\
[T_m(w_m)]_{B_m}
\end{bmatrix} = \begin{bmatrix}
A_1[w_1]_{B_1} \\
A_2[w_2]_{B_2} \\
\vdots \\
A_m[w_m]_{B_m}
\end{bmatrix} = A[w]_B
\]

(4.79)

for all \( w \in V \). So (4.75) follows. Then (4.72) and (4.73) follow from (4.52) and (4.53), respectively, in Proposition 4.11.

Now we are ready to prove Cayley-Hamilton. Let us prove it for linear endomorphisms. Of course, if it holds for linear endomorphisms, it holds for square matrices.

Let \( T \) be a linear endomorphism on a vector space \( V \) over \( F \) of \( \dim V = n \) and let \( f(x) \in F[x] \) be a minimal polynomial of \( T \). For simplicity, we assume that \( f(x) \) is monic, i.e., has leading coefficient 1. Suppose that

\[
f(x) = (x - \lambda_1)^{d_1}(x - \lambda_2)^{d_2} \cdots (x - \lambda_m)^{d_m}
\]

(4.80)

for some positive integers \( d_1, d_2, \ldots, d_m \) and \( m \) distinct numbers \( \lambda_1, \lambda_2, \ldots, \lambda_m \) in \( F \). Note that we can always do this over \( F = \mathbb{C} \).

**Proposition 4.15.** Let \( T : V \to V \) be a linear endomorphism on a finite-dimensional vector space \( V \) over \( F \) and let \( f(x) \in F[x] \) be a minimal polynomial of \( T \) given by (4.80). Then \( \lambda_1, \lambda_2, \ldots, \lambda_m \) are all the eigenvalues of \( T \).

**Proof.** Suppose that one of \( \lambda_1, \lambda_2, \ldots, \lambda_m \) is not an eigenvalue of \( T \). Without the loss of generality, let us assume that \( \lambda_1 \) is not an eigenvalue of \( T \). Then \( K(T - \lambda_1 I) \neq 0 \) and \( T - \lambda_1 I \) is invertible. Hence

\[
f(T) = 0 \Rightarrow (T - \lambda I)g(T) = 0
\]

(4.81)

\[
\Rightarrow (T - \lambda I)^{-1}(T - \lambda I)g(T) = 0 \Rightarrow g(T) = 0
\]
where \( g(x) = (x - \lambda_1)^{-1}f(x) = (x - \lambda_1)^{d_1-1}(x - \lambda_2)^{d_2} \ldots (x - \lambda_m)^{d_m} \). Since \( \deg g(x) < \deg f(x) \), this is a contradiction as \( f(x) \) is a minimal polynomial of \( T \). So all of \( \lambda_1, \lambda_2, \ldots, \lambda_m \) are eigenvalues of \( T \).

Suppose that \( T \) has an eigenvalue \( \lambda \) other than \( \lambda_1, \lambda_2, \ldots, \lambda_m \). Then \( T(v) = \lambda v \) for some \( v \neq 0 \). It follows that
\[
(4.82) \quad f(T)(v) = f(\lambda)v
\]
by (3.80). Since \( \lambda \) is different from any of \( \lambda_1, \lambda_2, \ldots, \lambda_m \), \( f(\lambda) \neq 0 \) and hence \( f(T)(v) \neq 0 \). This contradicts the fact that \( f(T) = 0 \).

In conclusion, \( \lambda_1, \lambda_2, \ldots, \lambda_m \) are all the eigenvalues of \( T \). \( \square \)

Next, let us show the generalized eigenspace theorem. For an eigenvalue \( \lambda \) of \( T \), the space \( K((T - \lambda I)^d) \) is called a generalized eigenspace of \( T \) for positive integer \( d \). We will show that
\[
(4.83) \quad V = K((T - \lambda_1 I)^{d_1}) \oplus K((T - \lambda_2 I)^{d_2}) \oplus \cdots \oplus K((T - \lambda_m I)^{d_m})
\]
if \( f(T) = 0 \) for a polynomial \( f(x) \) given by (4.80).

**Lemma 4.16.** Let \( T : V \to V \) be a linear endomorphism on a finite-dimensional vector space \( V \) over \( F \). For two polynomials \( f_1(x) \) and \( f_2(x) \) with \( \gcd(f_1(x), f_2(x)) = 1 \),
\[
(4.84) \quad K(f_1(T)f_2(T)) = K(f_1(T)) \oplus K(f_2(T)).
\]

**Proof.** By Bezout identity, \( g_1(x)f_1(x) + g_2(x)f_2(x) = 1 \) for some polynomials \( g_1(x) \) and \( g_2(x) \). Therefore,
\[
(4.85) \quad g_1(T)f_1(T) + g_2(T)f_2(T) = I.
\]
Then for every vector \( v \in K(f_1(T)) \cap K(f_2(T)), \)
\[
(4.86) \quad g_1(T)f_1(T)(v) + g_2(T)f_2(T)(v) = v \Rightarrow v = 0.
\]
Therefore, \( K(f_1(T)) \cap K(f_2(T)) = 0 \). So we can apply Lemma 3.17 to \( f_1(T) \) and \( f_2(T) \) to obtain (4.84). \( \square \)

**Theorem 4.17 (Generalized Eigenspace Theorem).** Let \( T : V \to V \) be a linear endomorphism on a finite-dimensional vector space \( V \) over \( F \) and let \( f(x) \in F[x] \) be a polynomial given by (4.80) such that \( f(T) = 0 \).

- \( V \) is the direct sum (4.83).
- If \( f(x) \) is a minimal polynomial of \( T \), then
\[
(4.87) \quad d_k = \min \left\{ b \in \mathbb{N} : K((T - \lambda_k I)^b) = K((T - \lambda_k I)^{b+1}) \right\}
\]
and
\[
(4.88) \quad d_k \leq \dim K((T - \lambda_k I)^{d_k})
\]
for \( k = 1, 2, \ldots, m \).
4. MINIMAL POLYNOMIAL AND CAYLEY-HAMILTON THEOREM

PROOF. Repeatedly applying Lemma [4.16] we obtain
\[ V = K(f(T)) = K((T - \lambda_1 I)^{d_1} (T - \lambda_2 I)^{d_2} \ldots (T - \lambda_m I)^{d_m}) \]
\[ = K((T - \lambda_1 I)^{d_1}) \oplus K((T - \lambda_2 I)^{d_2}) \oplus \ldots \oplus K((T - \lambda_m I)^{d_m}) \]
(4.89) \[ = K((T - \lambda_1 I)^{d_1}) \oplus K((T - \lambda_2 I)^{d_2}) \oplus \ldots \oplus K((T - \lambda_m I)^{d_m}) = \ldots \]

Let \( V_k = K((T - \lambda_k I)^{d_k}) \) for \( k = 1, 2, \ldots, m \). Then each \( V_k \) is a \( T \)-invariant subspace of \( V \). Let \( T_k : V_k \to V_k \) be the restriction of \( T \) to \( V_k \).

For every vector \( v \in V_k = K((T - \lambda_k I)^{d_k}) \), we have
\[ (T - \lambda_k I)^{d_k}(v) = 0 \Rightarrow (T_k - \lambda_k I)^{d_k}(v) = 0 \]
and hence \((T_k - \lambda_k I)^{d_k} = 0\). Let \( f_k(x) \) be a minimal polynomial of \( T_k \). Then \( f_k(x) \) divides \((x - \lambda_k)^{d_k}\). Therefore, \( f_k(x) = (x - \lambda_k)^{a_k} \) for some \( a_k \leq d_k \) if we choose \( f_k(x) \) to be monic. Then by (4.73),
\[ f(x) = \text{lcm}(f_1(x), f_2(x), \ldots, f_m(x)) \]
(4.90) \[ = \text{lcm} \left( (x - \lambda_1)^{a_1}, (x - \lambda_2)^{a_2}, \ldots, (x - \lambda_m)^{a_m} \right) \]
and we must have \( a_k = d_k \), i.e., \( f_k(x) = (x - \lambda_k)^{d_k} \) is a minimal polynomial of \( T_k \) for \( k = 1, 2, \ldots, m \).

On each \( V_k \), the sequence
(4.92) \[ 0 \subset K((T_k - \lambda_k I)) \subset K((T_k - \lambda_k I)^2) \subset \ldots \subset K((T_k - \lambda_k I)^n) \subset \ldots \]
will stabilize at some term. By Theorem 2.17,
\[ 0 \not\subseteq K((T_k - \lambda_k I)) \not\subseteq \ldots \not\subseteq K((T_k - \lambda_k I)^b) \]
(4.93) \[ = K((T_k - \lambda_k I)^{b+1}) = \ldots \]
\[ = K((T_k - \lambda_k I)^n) = \ldots \subset V_k \]
for some \( b \in \mathbb{N} \). Since \( K((T_k - \lambda_k I)^{d_k}) = V_k \), we must have \( b \leq d_k \) and
(4.94) \[ K((T_k - \lambda_k I)^b) = V_k \iff (T_k - \lambda_k I)^b = 0. \]
And since \((x - \lambda_k)^{d_k}\) is a minimal polynomial of \( T_k \), we must have \( b = d_k \) and hence (4.87).

Therefore,
(4.95) \[ 0 \not\subseteq K((T_k - \lambda_k I)) \not\subseteq \ldots \not\subseteq K((T_k - \lambda_k I)^{d_k}) = V_k. \]
Let \( r_d = \dim K((T_k - \lambda_k I)^{d}) \). Then
\[ \dim V_k = r_1 + (r_2 - r_1) + \cdots + (r_{d_k} - r_{d_k-1}) \]
(4.96) \[ = \sum_{i=1}^{d_k} (r_i - r_{i-1}) \geq d_k \]
since \( 0 = r_0 < r_1 < \ldots < r_{d_k} \). This proves (4.88). \( \square \)
Now we can prove Cayley-Hamilton Theorem.

**Proof of Cayley-Hamilton Theorem**

Let $T$ be a linear endomorphism on a finite-dimensional vector space $V$ over $\mathbb{C}$ and $f(x) \in \mathbb{C}[x]$ be the minimal polynomial of $T$ given by (4.80).

As in the proof of Theorem 4.17, we let $V_k = K((T - \lambda_k I)^{d_k})$ and let $T_k : V_k \to V_k$ be the restriction of $T$ to $V_k$ for $k = 1, 2, \ldots, m$. We have proved that $(x - \lambda_k)^{d_k}$ is a minimal polynomial of $T_k$. So $\lambda_k$ is the only eigenvalue of $T_k$ by Proposition 4.15. Therefore, the characteristic polynomial of $T_k$ is

$$\det(xI - T_k) = (x - \lambda_k)^{n_k}$$

for $n_k = \dim V_k$ and $k = 1, 2, \ldots, m$. Then the characteristic polynomial of $T$ is

$$\det(xI - T) = \det(xI - T_1) \det(xI - T_2) \ldots \det(xI - T_m)$$

$$= (x - \lambda_1)^{n_1}(x - \lambda_2)^{n_2} \ldots (x - \lambda_m)^{n_m}$$

by (4.72). By (4.88), $d_1 \leq n_1, d_2 \leq n_2, \ldots, d_m \leq n_m$. So $f(x)$ divides $\det(xI - T)$. We are done. \qed