2 Polynomials over a field

A polynomial over a field F is a sequence

$$(a_0, a_1, a_2, \ldots, a_n, \ldots) \quad \text{where} \quad a_i \in F \ \forall i$$

with $a_i = 0$ from some point on. a_i is called the i–th coefficient of f.

We define three special polynomials...

$$\begin{aligned} 0 &= (0, 0, 0, \ldots) \\ 1 &= (1, 0, 0, \ldots) \\ x &= (0, 1, 0, \ldots). \end{aligned}$$

The polynomial $(a_0, \ldots)$ is called a constant and is written simply as a_0.

Let $F[x]$ denote the set of all polynomials in x.

If $f \neq 0$, then the **degree** of f, written $\deg f$, is the greatest n such that $a_n \neq 0$. Note that the polynomial 0 has no degree.

a_n is called the 'leading coefficient' of f.

$F[x]$ forms a vector space over F if we define

$$\lambda(a_0,\, a_1, \ldots) = (\lambda a_0,\, \lambda a_1, \ldots),\ \lambda \in F.$$

DEFINITION 2.1

*(***Multiplication of polynomials***)*

Let $f = (a_0, a_1, \ldots)$ and $g = (b_0, b_1, \ldots)$. Then $fg = (c_0, c_1, \ldots)$ where

$$\begin{aligned} c_n &= a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0 \\ &= \sum_{i=0}^{n} a_i b_{n-i} \\ &= \sum_{\substack{0 \leq i, 0 \leq j \\ i+j=n}} a_i b_j. \end{aligned}$$

EXAMPLE 2.1

$$x^2 = (0,\, 0,\, 1,\, 0, \ldots),\ x^3 = (0,\, 0,\, 0,\, 1,\, 0, \ldots).$$

More generally, an induction shows that $x^n = (a_0, \ldots)$, where $a_n = 1$ and all other a_i are zero.

If $\deg f = n$, we have $f = a_0 1 + a_1 x + \cdots + a_n x^n$.

THEOREM 2.1 (Associative Law)
$f(gh) = (fg)h$

PROOF Take f, g as above and $h = (c_0, c_1, \ldots)$. Then $f(gh) = (d_0, d_1, \ldots)$, where

$$\begin{aligned} d_n &= \sum_{i+j=n} (fg)_i h_j \\ &= \sum_{i+j=n} \left(\sum_{u+v=i} f_u g_v \right) h_j \\ &= \sum_{u+v+j=n} f_u g_v h_j. \end{aligned}$$

Likewise $(fg)h = (e_0, e_1, \ldots)$, where

$$e_n = \sum_{u+v+j=n} f_u g_v h_j$$

Some properties of polynomial arithmetic:

$$\begin{aligned} fg &= gf \\ 0f &= 0 \\ 1f &= f \\ f(g+h) &= fg + fh \\ f \neq 0 \text{ and } g \neq 0 &\Rightarrow fg \neq 0 \\ &\quad \text{and } \deg(fg) = \deg f + \deg g. \end{aligned}$$

The last statement is equivalent to

$$fg = 0 \Rightarrow f = 0 \text{ or } g = 0.$$

The we deduce that

$$fh = fg \text{ and } f \neq 0 \Rightarrow h = g.$$

2.1 Lagrange Interpolation Polynomials

Let $P_n[F]$ denote the set of polynomials $a_0 + a_1 x + \cdots + a_n x^n$, where $a_0, \ldots, a_n \in F$. Then $a_0 + a_1 x + \cdots + a_n x^n = 0$ implies that $a_0 = 0, \ldots, a_n = 0$.

$P_n[F]$ is a subspace of $F[x]$ and $1, x, x^2, \ldots, x^n$ form the 'standard' basis for $P_n[F]$.

If $f \in P_n[F]$ and $c \in F$, we write

$$f(c) = a_0 + a_1 c + \cdots + a_n c^n.$$

This is the"value of f at c". This symbol has the following properties:

$$\begin{aligned} (f+g)(c) &= f(c) + g(c) \\ (\lambda f)(c) &= \lambda(f(c)) \\ (fg)(c) &= f(c)g(c) \end{aligned}$$

DEFINITION 2.2

Let $c_1, \ldots, c_{n+1}$ be distinct members of F. Then the **Lagrange interpolation polynomials** *$p_1, \ldots, p_{n+1}$ are polynomials of degree n defined by*

$$p_i = \prod_{\substack{j=1 \\ j \neq i}}^{n+1} \left(\frac{x - c_j}{c_i - c_j} \right), \ 1 \leq i \leq n+1.$$

EXAMPLE 2.2

$$\begin{aligned} p_1 &= & \left(\frac{x-c_2}{c_1-c_2}\right) & \left(\frac{x-c_3}{c_1-c_3}\right) & \cdots & \left(\frac{x-c_{n+1}}{c_1-c_{n+1}}\right) \\ p_2 &= \left(\frac{x-c_1}{c_2-c_1}\right) & \times & \left(\frac{x-c_3}{c_2-c_3}\right) & \cdots & \left(\frac{x-c_{n+1}}{c_2-c_{n+1}}\right) \\ & \textit{etc...} \end{aligned}$$

We now show that the Lagrange polynomials also form a basis for $P_n[F]$.
PROOF Noting that there are $n+1$ elements in the 'standard' basis, above, we see that $\dim P_n[F] = n+1$ and so it suffices to show that $p_1, \ldots, p_{n+1}$ are LI.

We use the following property of the polynomials p_i:

$$p_i(c_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Assume that

$$a_1 p_1 + \cdots + a_{n+1} p_{n+1} = 0$$

where $a_i \in F, \ 1 \leq i \leq n+1$. Evaluating both sides at $c_1, \ldots, c_{n+1}$ gives

$$\begin{aligned} a_1 p_1(c_1) + \cdots + a_{n+1} p_{n+1}(c_1) &= 0 \\ &\vdots \\ a_1 p_1(c_{n+1}) + \cdots + a_{n+1} p_{n+1}(c_{n+1}) &= 0 \end{aligned}$$

$\Rightarrow$

$$\begin{aligned} a_1 \times 1 + a_2 \times 0 + \cdots + a_{n+1} \times 0 &= 0 \\ a_1 \times 0 + a_2 \times 1 + \cdots + a_{n+1} \times 0 &= 0 \\ &\vdots \\ a_1 \times 0 + a_2 \times 0 + \cdots + a_{n+1} \times 1 &= 0 \end{aligned}$$

Hence $a_i = 0 \; \forall i$ as required.

COROLLARY 2.1

If $f \in P_n[F]$ then

$$f = f(c_1)p_1 + \cdots + f(c_{n+1})p_{n+1}.$$

PROOF*: We know that*

$$f = \lambda_1 p_1 + \cdots + \lambda_{n+1} p_{n+1} \quad \text{for some } \lambda_i \in F.$$

Evaluating both sides at $c_1, \ldots, c_{n+1}$ then, gives

$$\begin{aligned} f(c_1) &= \lambda_1, \\ &\vdots \\ f(c_{n+1}) &= \lambda_{n+1} \end{aligned}$$

as required.

COROLLARY 2.2

If $f \in P_n[F]$ and $f(c_1) = 0, \ldots, f(c_{n+1}) = 0$ where $c_1, \ldots, c_{n+1}$ are distinct, then $f = 0$. (I.e. a non-zero polynomial of degree n can have at most n roots.)

COROLLARY 2.3

If $b_1, \ldots, b_{n+1}$ are any *scalars in F, and $c_1, \ldots, c_{n+1}$ are again distinct, then there exists a* unique *polynomial $f \in P_n[F]$ such that*

$$f(c_1) = b_1, \ldots, f(c_{n+1}) = b_{n+1} \; ;$$

namely

$$f = b_1 p_1 + \cdots + b_{n+1} p_{n+1}.$$

EXAMPLE 2.3

Find the quadratic polynomial

$$f = a_0 + a_1 x + a_2 x^2 \ \in P_2[\mathbb{R}]$$

such that

$$f(1) = 8, f(2) = 5, f(3) = 4.$$

Solution: $f = 8p_1 + 5p_2 + 4p_3$ *where*

$$\begin{aligned} p_1 &= \frac{(x-2)(x-3)}{(1-2)(1-3)} \\ p_2 &= \frac{(x-1)(x-3)}{(2-1)(2-3)} \\ p_3 &= \frac{(x-1)(x-2)}{(3-1)(3-2)} \end{aligned}$$

2.2 Division of polynomials

DEFINITION 2.3

If $f, g \in F[x]$, we say f **divides** *g if $\exists h \in F[x]$ such that*

$$g = fh.$$

For this we write "$f \mid g$", and "$f \not| g$" denotes the negation "f does not divide g".

Some properties:

$$f \mid g \text{ and } g \neq 0 \Rightarrow \deg f \leq \deg g$$

and thus of course

$$f \mid 1 \Rightarrow \deg f = 0.$$

2.2.1 Euclid's Division Theorem

Let $f, g \in F[x]$ and $g \neq 0$.

Then $\exists q, r \in F[x]$ such that

$$f = qg + r, \tag{3}$$

where $r = 0$ or $\deg r < \deg g$. Moreover q and r are unique.

Outline of Proof:

If $f = 0$ or $\deg f < \deg g$, (3) is trivially true (taking $q = 0$ and $r = f$).
So assume $\deg f \geq \deg g$, where

$$f = a_m x^m + a_{m-1} x^{m-1} + \cdots a_0,$$
$$g = b_n x^n + \cdots + b_0$$

and we have a long division process, viz:

$$\begin{array}{r|l} & a_m b_n^{-1} x^{m-n} \quad + \quad \cdots \\ b_n x^n + \cdots + b_0 & \underline{a_m x^m \quad + \quad a_{m-1} x^{m-1} \quad + \cdots + \quad a_0} \\ & \underline{a_m x^m} \\ & \qquad\qquad\qquad\qquad\qquad\qquad \text{etc.}\ldots \end{array}$$

(See S. Perlis, Theory of Matrices, p.111.)

2.2.2 Euclid's Division Algorithm

$$\begin{array}{rcll} f & = & q_1 g + r_1 & \text{with} \quad \deg r_1 < \deg g \\ g & = & q_2 r_1 + r_2 & \text{with} \quad \deg r_2 < \deg r_1 \\ r_1 & = & q_3 r_2 + r_3 & \text{with} \quad \deg r_3 < \deg r_2 \\ \cdots & \vdots & \cdots & \cdots \quad \cdots \\ r_{n-2} & = & q_n r_{n-1} + r_n & \text{with} \quad \deg r_n < \deg r_{n-1} \\ r_{n-1} & = & q_{n+1} r_n & \end{array}$$

Then $r_n = \gcd(f, g)$, the **greatest common divisor** of f and g—i.e. r_n is a polynomial d with the property that

1. $d \mid f$ and $d \mid g$, and
2. $\forall e \in F[x]$, $e \mid f$ and $e \mid g \Rightarrow e \mid d$.

(This defines $\gcd(f, g)$ uniquely up to a constant multiple.)

We select the *monic* (i.e. leading coefficient = 1) gcd as "the" gcd.

Also, $\exists u, v \in F[x]$ such that

$$\begin{aligned} r_n &= \gcd(f, g) \\ &= uf + vg \end{aligned}$$

—find u and v by 'forward substitution' in Euclid's algorithm; viz.

$$\begin{aligned} r_1 &= f + (-q_1) g \\ r_2 &= g + (-q_2) r_1 \end{aligned}$$

$$
\begin{aligned}
&= g + (-q_2)(f + (-q_1)g) \\
&= g + (-q_2)f + (q_1 q_2)g \\
&= (-q_2)f + (1 + q_1 q_2)g \\
&\vdots \\
r_n &= \underbrace{(\ldots)}_{u} f + \underbrace{(\ldots)}_{v} g.
\end{aligned}
$$

In general, $r_k = s_k f + t_k g$ for $-1 \leq k \leq n$, where

$$r_{-1} = f,\ r_0 = g,\ s_{-1} = 1,\ s_0 = 0,\ t_{-1} = 0,\ t_0 = 1$$

and

$$s_k = -q_k s_{k-1} + s_{k-2},\ t_k = -q_k t_{k-1} + t_{k-2}$$

for $1 \leq k \leq n$. (Proof by induction.)

The special case $\gcd(f, g) = 1$ (i.e. f and g are **relatively prime**) is of great importance: here $\exists u, v \in F[x]$ such that

$$uf + vg = 1.$$

EXERCISE 2.1

Find $\gcd(3x^2 + 2x + 4, 2x^4 + 5x + 1)$ *in* $\mathbb{Q}[x]$ *and express it as* $uf + vg$ *for two polynomials* u *and* v.

2.3 Irreducible Polynomials

DEFINITION 2.4

Let f *be a non-constant polynomial. Then, if*

$$g \mid f \Rightarrow \begin{array}{ll} & g \text{ is a constant} \\ \text{or} & g = \text{constant} \times f \end{array}$$

we call f *an* **irreducible polynomial**.

Note: (**Remainder theorem**)

$f = (x - a)q + f(a)$ where $a \in F$. So $f(a) = 0$ iff $(x - a) \mid f$.

EXAMPLE 2.4

$f(x) = x^2 + x + 1 \in \mathbb{Z}_2[x]$ *is irreducible, for* $f(0) = f(1) = 1 \neq 0$, *and hence there are no polynomials of degree 1 which divide* f.

THEOREM 2.2

Let f be irreducible. Then if $f \not| g$, $\gcd(f,g) = 1$ and $\exists u, v \in F[x]$ such that

$$uf + vg = 1.$$

PROOF Suppose f is irreducible and $f \not| g$. Let $d = \gcd(f,g)$ so

$$d \mid f \quad \text{and} \quad d \mid g.$$

Then either $d = cf$ for some constant c, or $d = 1$. But if $d = cf$ then

$$\begin{aligned} & f \mid d \quad \text{and} \quad d \mid g \\ \Rightarrow \quad & f \mid g \quad \text{—a contradiction.} \end{aligned}$$

So $d = 1$ as required.

COROLLARY 2.4

If f is irreducible and $f \mid gh$, then $f \mid g$ or $f \mid h$.

PROOF*: Suppose f is irreducible and $f \mid gh$, $f \not| g$. We show that $f \mid h$.*

By the above theorem, $\exists u, v$ such that

$$\begin{aligned} uf + vg &= 1 \\ \Rightarrow ufh + vgh &= h \\ \Rightarrow f &\mid h \end{aligned}$$

THEOREM 2.3

Any non-constant polynomial is expressible as a product of irreducible polynomials where representation is unique up to the order of the irreducible factors.

Some examples:

$$\begin{aligned} (x+1)^2 &= x^2 + 2x + 1 \\ &= x^2 + 1 \quad \text{in} \mathbb{Z}_2[x] \\ (x^2 + x + 1)^2 &= x^4 + x^2 + 1 \quad \text{in } \mathbb{Z}_2[x] \\ (2x^2 + x + 1)(2x + 1) &= x^3 + x^2 + 1 \quad \text{in} \mathbb{Z}_3[x] \\ &= (x^2 + 2x + 2)(x + 2) \quad \text{in} \mathbb{Z}_3[x]. \end{aligned}$$

PROOF

Existence of factorization: If $f \in F[x]$ is not a constant polynomial, then f being irreducible implies the result.

Otherwise, $f = f_1 F_1$, with $0 < \deg f_1, \deg F_1 < \deg f$. If f_1 and F_1 are irreducible, stop. Otherwise, keep going.

Eventually we end with a decomposition of f into irreducible polynomials.

Uniqueness: Let

$$cf_1 f_2 \cdots f_m = dg_1 g_2 \cdots g_n$$

be two decompositions into products of constants (c and d) and monic irreducibles (f_i, g_j). Now

$$f_1 \mid f_1 f_2 \cdots f_m \Longrightarrow f_1 \mid g_1 g_2 \cdots g_n$$

and since f_i, g_i are irreducible we can cancel f_1 and some g_j.

Repeating this for $f_2, \ldots, f_m$, we eventually obtain $m = n$ and $c = d$—in other words, each expression is simply a rearrangement of the factors of the other, as required.

THEOREM 2.4

Let F_q be a field with q elements. Then if $n \in \mathbb{N}$, there exists an irreducible polynomial of degree n in $F[x]$.

PROOF First we introduce the idea of the **Riemann zeta function**:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - \frac{1}{p^s}}.$$

To see the equality of the latter expressions note that

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i = 1 + x + x^2 + \cdots$$

and so

$$\begin{aligned} \text{R.H.S.} &= \prod_{p \text{ prime}} \left(\sum_{i=0}^{\infty} \frac{1}{p^{is}} \right) \\ &= \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \cdots\right)\left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \cdots\right) \\ &= 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots \end{aligned}$$

—note for the last step that terms will be of form

$$\left(\frac{1}{p_1^{a_1}\cdots p_R^{a_R}}\right)^s$$

up to some prime p_R, with $a_i \geq 0\ \forall i = 1, \ldots, R.$ and as $R \to \infty$, the prime factorizations

$$p_1^{a_1}\cdots p_R^{a_R}$$

map onto the natural numbers, $\mathbb{N}$.

We let N_m denote the number of monic irreducibles of degree m in $F_q[x]$. For example, $N_1 = q$ since $x + a, a \in F_q$ are the irreducible polynomials of degree 1.

Now let $|f| = q^{\deg f}$, and $|0| = 0$. Then we have

$$|fg| = |f|\,|g| \quad \text{since } \deg fg = \deg f + \deg g$$

and, because of the uniqueness of factorization theorem,

$$\sum_{f \text{ monic}} \frac{1}{|f|^s} = \prod_{\substack{f \text{ monic and} \\ \text{irreducible}}} \frac{1}{1 - \frac{1}{|f|^s}}.$$

Now the left hand side is

$$\begin{aligned}
&\sum_{n=0}^{\infty} \sum_{\substack{f \text{ monic and} \\ \deg f = n}} \frac{1}{|f|^s} \\
&= \sum_{n=0}^{\infty} \frac{q^n}{q^{ns}} \\
&\qquad \text{(there are } q^n \text{ monic polynomials of degree } n\text{)} \\
&= \sum_{n=0}^{\infty} \frac{1}{q^{n(s-1)}} \\
&= \frac{1}{1 - \frac{1}{q^{s-1}}} \\
\text{and R.H.S.} &= \prod_{n=1}^{\infty} \frac{1}{\left(1 - \frac{1}{q^{ns}}\right)^{N_n}}.
\end{aligned}$$

Equating the two, we have

$$\frac{1}{1-\frac{1}{q^{s-1}}} = \prod_{n=1}^{\infty} \frac{1}{\left(1-\frac{1}{q^{ns}}\right)^{N_n}}. \tag{4}$$

We now take logs of both sides, and then use the fact that

$$\log\left(\frac{1}{1-x}\right) = \sum_{n=1}^{\infty} \frac{x^n}{n} \quad \text{if } |x| < 1;$$

so (4) becomes

$$\begin{aligned}
\log \frac{1}{1-q^{-(s-1)}} &= \prod_{n=1}^{\infty} \frac{1}{\left(1-\frac{1}{q^{ns}}\right)^{N_n}} \\
\Rightarrow \sum_{k=1}^{\infty} \frac{1}{kq^{(s-1)k}} &= -\sum_{n=1}^{\infty} N_n \log\left(1-\frac{1}{q^{ns}}\right) \\
&= \sum_{n=1}^{\infty} N_n \sum_{m=1}^{\infty} \frac{1}{mq^{mns}} \\
\text{so} \quad \sum_{k=1}^{\infty} \frac{q^k}{kq^{sk}} &= \sum_{n=1}^{\infty} N_n \sum_{m=1}^{\infty} \frac{n}{mnq^{mns}} \\
&= \sum_{k=1}^{\infty} \frac{\sum_{mn=k} nN_n}{kq^{ks}}.
\end{aligned}$$

Putting $x = q^s$, we have

$$\sum_{k=1}^{\infty} \frac{q^k x^k}{k} = \sum_{k=1}^{\infty} x^k \times \sum_{mn=k} nN_n,$$

and since both sides are power series, we may equate coefficients of x^k to obtain

$$q^k = \sum_{mn=k} nN_n = \sum_{n|k} nN_n. \tag{5}$$

We can deduce from this that $N_n > 0$ as $n \to \infty$ (see Berlekamp's "*Algebraic Coding Theory*").

Now note that $N_1 = q$, so if k is a prime—say $k = p$, (5) gives

$$\begin{aligned}
q^p &= N_1 + pN_p = q + pN_p \\
\Rightarrow N_p &= \frac{q^p - q}{p} > 0 \qquad \text{as } q > 1 \text{ and } p \geq 2.
\end{aligned}$$

This proves the theorem for $n = p$, a prime.

But what if k is not prime? Equation (5) also tells us that

$$q^k \geq kN_k.$$

Now let $k \geq 2$. Then

$$\begin{aligned}
q^k &= kN_k + \sum_{\substack{n|k \\ n \neq k}} nN_n \\
&\leq kN_k + \sum_{\substack{n|k \\ n \neq k}} q^n \qquad (\text{as } nN_n \leq q^n) \\
&\leq kN_k + \sum_{n=1}^{\lfloor k/2 \rfloor} q^n \\
&< kN_k + \sum_{n=0}^{\lfloor k/2 \rfloor} q^n \qquad (\text{adding } 1) \\
&= kN_k + \frac{q^{\lfloor k/2 \rfloor + 1} - 1}{q - 1} \qquad (\text{sum of geometric series}).
\end{aligned}$$

But

$$\frac{q^{t+1} - 1}{q - 1} < q^{t+1} \qquad \text{if } q \geq 2,$$

so

$$\begin{aligned}
q^k &< kN_k + q^{\lfloor k/2 \rfloor + 1} \\
\Rightarrow N_k &> \frac{q^k - q^{\lfloor k/2 \rfloor + 1}}{k} \\
&\geq 0 \quad \text{if } q^k \geq q^{\lfloor k/2 \rfloor + 1}.
\end{aligned}$$

Since $q > 1$ (we cannot have a field with a single element, since the additive and multiplicative identities cannot be equal by one of the axioms), the latter condition is equivalent to

$$k \geq \lfloor k/2 \rfloor + 1$$

which is true and the theorem is proven.

2.4 Minimum Polynomial of a (Square) Matrix

Let $A \in M_{n\times n}(F)$, and $g = \mathrm{ch}_A$. Then $g(A) = 0$ by the Cayley–Hamilton theorem.

DEFINITION 2.5

Any non–zero polynomial g of minimum degree and satisfying $g(A) = 0$ is called a **minimum polynomial** *of A.*

Note: If f is a minimum polynomial of A, then f cannot be a constant polynomial. For if $f = c$, a constant, then $0 = f(A) = cI_n$ implies $c = 0$.

THEOREM 2.5

If f is a minimum polynomial of A and $g(A) = 0$, then $f \mid g$. (In particular, $f \mid ch_A$.)

PROOF Let $g(A) = 0$ and f be a minimum polynomial. Then

$$g = qf + r,$$

where $r = 0$ or $\deg r < \deg f$. Hence

$$\begin{aligned} g(A) &= q(A) \times 0 + r(A) \\ 0 &= r(A). \end{aligned}$$

So if $r \neq 0$, the inequality $\deg r < \deg f$ would give a contradict the definition of f. Consequently $r = 0$ and $f \mid g$.

Note: It follows that if f and g are minimum polynomials of A, then $f|g$ and $g|f$ and consequently $f = cg$, where c is a scalar. Hence there is a unique monic minimum polynomial and we denote it by m_A.

EXAMPLES (of minimum polynomials):

1. $A = 0 \Leftrightarrow m_A = x$
2. $A = I_n \Leftrightarrow m_A = x - 1$
3. $A = cI_n \Leftrightarrow m_A = x - c$
4. $A^2 = A$ and $A \neq 0$ and $A \neq I_n \Leftrightarrow m_A = x^2 - x$.

EXAMPLE 2.5

$F = \mathbb{Q}$ and

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}.$$

Now

$$\begin{aligned} A &\neq c_0 I_3,\ c_0 \in \mathbb{Q},\ \textit{so } m_A \neq x - c_0, \\ A^2 &= 3A - 2I_3 \\ \Rightarrow m_A &= x^2 - 3x + 2 \end{aligned}$$

This is an special case of a general algorithm:

(**Minimum polynomial algorithm**) Let $A \in M_{n\times n}(F)$. Then we find the least positive integer r such that A^r is expressible as a linear combination of the matrices

$$I_n,\ A, \ldots, A^{r-1},$$

say

$$A^r = c_0 + c_1 A + \cdots + c_{r-1}A^{r-1}.$$

(Such an integer must exist as $I_n,\ A, \ldots, A^{n^2}$ form a linearly dependent family in the vector space $M_{n\times n}(F)$ and this latter space has dimension equal to n^2.)

Then $m_A = x^r - c_{r-1}x^{r-1} - \cdots - c_1 x - c_0$.

THEOREM 2.6

If $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in F[x]$, then $m_{C(f)} = f$, where

$$C(f) = \begin{bmatrix} 0 & 0 & & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & & 0 & -a_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

.

PROOF For brevity denote $C(f)$ by A. Then post-multiplying A by the respective unit column vectors $E_1, \ldots, E_n$ gives

$$\begin{aligned} AE_1 &= E_2 \\ AE_2 &= E_3 \Rightarrow A^2E_1 = E_3 \\ &\vdots \\ AE_{n-1} &= E_n \Rightarrow A^{n-1}E_1 = E_n \\ AE_n &= -a_0E_1 - a_2E_2 - \cdots - a_{n-1}E_n \\ &= -a_0E_1 - a_2AE_1 - \cdots - a_{n-1}A^{n-1}E_1 = A^nE_1, \end{aligned}$$

so

$$\Rightarrow f(A)E_1 = 0 \Rightarrow \text{ first column of } f(A) \text{ zero}$$

Now although matrix multiplication is not commutative, multiplication of two matrices, each of which is a polynomial in a given square matrix A, is commutative. Hence $f(A)g(A) = g(A)f(A)$ if $f, g \in F[x]$. Taking $g = x$ gives

$$f(A)A = Af(A).$$

Thus

$$f(A)E_2 = f(A)AE_1 = Af(A)E_1 = 0$$

and so the second column of A is zero. Repeating this for $E_3, \ldots, E_n$, we see that

$$f(A) = 0$$

and thus $m_A | f$.

To show $m_A = f$, we assume $\deg m_A = t < n$; say

$$m_A = x^t + b_{t-1}x^{t-1} + \cdots + b_0.$$

Now

$$\begin{aligned} m_A(A) &= 0 \\ \Rightarrow A^t + b_{t-1}A^{t-1} + \cdots + b_0 I_n &= 0 \\ \Rightarrow (A^t + b_{t-1}A^{t-1} + \cdots + b_0 I_n)E_1 &= 0, \end{aligned}$$

and recalling that $AE_1 = E_2$ etc., and $t < n$, we have

$$E_{t+1} + b_{t-1}E_t + \cdots + b_1 E_2 + b_0 E_1 = 0$$

which is a contradiction—since the E_i are independent, the coefficient of E_{t+1} cannot be 1.

Hence $m_A = f$.

Note: It follows that $\mathrm{ch}_A = f$. Because both ch_A and m_A have degree n and moreover m_A divides ch_A.

EXERCISE 2.2

If $A = J_n(a)$ for $a \in F$, an **elementary Jordan matrix** *of size n, show*

that $m_A = (x-a)^n$ *where*

$$A = J_n(a) = \begin{bmatrix} a & 0 & & & 0 \\ 1 & a & \cdots & & \\ 0 & 1 & & & \\ & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & a & 0 \\ 0 & 0 & & 1 & a \end{bmatrix}$$

(i.e. A *is an* $n \times n$ *matrix with* a*'s on the diagonal and 1's on the subdiagonal).*

Note*: Again, the minimum polynomial happens to equal the characteristic polynomial here.*

DEFINITION 2.6

*(***Direct Sum of Matrices***)*

Let $A_1, \ldots, A_t$ *be matrices over* F*. Then the direct sum of these matrices is defined as follows:*

$$A_1 \oplus A_2 \oplus \cdots \oplus A_t = \begin{bmatrix} A_1 & 0 & \ldots & \\ 0 & A_2 & & \\ \vdots & & \ddots & \vdots \\ & \cdots & 0 & A_t \end{bmatrix}.$$

Properties:

1. $$(A_1 \oplus \cdots \oplus A_t) + (B_1 \oplus \cdots \oplus B_t) = (A_1 + B_1) \oplus \cdots \oplus (A_t + B_t)$$

2. *If* $\lambda \in F$,
$$\lambda(A_1 \oplus \cdots \oplus A_t) = (\lambda A_1) \oplus \cdots \oplus (\lambda A_t)$$

3. $$(A_1 \oplus \cdots \oplus A_t)(B_1 \oplus \cdots \oplus B_t) = (A_1 B_1) \oplus \cdots \oplus (A_t B_t)$$

4. *If* $f \in F[x]$ *and* $A_1, \ldots, A_t$ *are square,*
$$f(A_1 \oplus \cdots \oplus A_t) = f(A_1) \oplus \cdots \oplus f(A_t)$$

DEFINITION 2.7

If $f_1, \ldots, f_t \in F[x]$*, we call* $f \in F[x]$ *a least common multiple (*lcm*) of* $f_1, \ldots, f_t$ *if*

1. $f_1 \mid f, \ldots f_t \mid f$, *and*

2. $f_1 \mid e, \ldots f_t \mid e \Rightarrow f \mid e$.

This uniquely defines the lcm *up to a constant multiple and so we set "the"* lcm *to be the monic* lcm .

EXAMPLES 2.1

If $fg \neq 0$, $\operatorname{lcm}(f, g) \mid fg$.

(Recursive property)

$$\operatorname{lcm}(f_1, \ldots, f_{t+1}) = \operatorname{lcm}(\operatorname{lcm}(f_1, \ldots, f_t), f_{t+1}).$$

THEOREM 2.7

$$m_{A_1 \oplus \cdots \oplus A_t} = \operatorname{lcm}(m_{A_1}, \ldots, m_{A_t}),$$

Also

$$\operatorname{ch}_{A_1 \oplus \cdots \oplus A_t} = \prod_{i=1}^{t} \operatorname{ch}_{A_i} .$$

PROOF Let $f =$ L.H.S. and $g =$ R.H.S. Then

$$\begin{aligned} & f(A_1 \oplus \cdots \oplus A_t) = 0 \\ \Rightarrow \quad & f(A_1) \oplus \cdots \oplus f(A_t) = 0 \oplus \cdots \oplus 0 \\ \Rightarrow \quad & f(A_1) = 0, \ldots, f(A_t) = 0 \\ \Rightarrow \quad & m_{A_1} \mid f, \ldots, m_{A_t} \mid f \\ \Rightarrow \quad & g \mid f. \end{aligned}$$

Conversely,

$$\begin{aligned} & m_{A_1} \mid g, \ldots, m_{A_t} \mid g \\ \Rightarrow \quad & g(A_1) = 0, \ldots, g(A_t) = 0 \\ \Rightarrow \quad & g(A_1) \oplus \cdots \oplus g(A_t) = 0 \oplus \cdots \oplus 0 \\ \Rightarrow \quad & g(A_1 \oplus \cdots \oplus A_t) = 0 \\ \Rightarrow \quad & f = m_{A_1 \oplus \cdots \oplus A_t} \mid g. \end{aligned}$$

Thus $f = g$.

EXAMPLE 2.6

Let $A = C(f)$ *and* $B = C(g)$.
Then $m_{A \oplus B} = \operatorname{lcm}(f, g)$.

Note: If

$$\begin{aligned} f &= cp_1^{a_1} \dots p_t^{a_t} \\ g &= dp_1^{b_1} \dots p_t^{b_t} \end{aligned}$$

where $c, d \neq 0$ are in F and $p_1, \dots, p_t$ are distinct monic irreducibles, then

$$\begin{aligned} \gcd(f,g) &= p_1^{\min(a_1,b_1)} \dots p_t^{\min(a_t,b_t)}, \\ \operatorname{lcm}(f,g) &= p_1^{\max(a_1,b_1)} \dots p_t^{\max(a_t,b_t)} \end{aligned}$$

Note

$$\min(a_i, b_i) + \max(a_i, b_i) = a_i + b_i.$$

so

$$\gcd(f,\, g) \operatorname{lcm}(f,\, g) = fg.$$

EXAMPLE 2.7

If $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, *then* $m_A = (x - c_1) \cdots (x - c_t)$, *where* $c_1, \dots, c_t$ *are the distinct members of the sequence* $\lambda_1, \dots, \lambda_n$.

PROOF. For A is the direct sum of the 1×1 matrices $\lambda_1, \dots, \lambda_n$ having minimum polynomials $x - \lambda_1, \dots, \lambda_n$. Hence

$$m_A = \operatorname{lcm}(x - \lambda_1, \dots, x - \lambda_n) = (x - c_1) \cdots (x - c_t).$$

We know that $m_A \mid \operatorname{ch}_A$. Hence if

$$\operatorname{ch}_A = p_1^{a_1} \dots p_t^{a_t}$$

where $a_1 > 0, \dots, a_t > 0$, and $p_1, \dots, p_t$ are distinct monic irreducibles, then

$$m_A = p_1^{b_1} \dots p_t^{b_t}$$

where $0 \leq b_i \leq a_i$, $\forall i = 1, \dots, t$.

We soon show that each $b_i > 0$, i.e. if $p \mid \operatorname{ch}_A$ and p is irreducible then $p \mid m_A$.

2.5 Construction of a field of p^n elements

(where p is prime and $n \in \mathbb{N}$)

Let f be a monic irreducible polynomial of degree n in $\mathbb{Z}_p[x]$—that is, $F_q = \mathbb{Z}_p$ here.
For instance,

$$\begin{aligned} n = 2, p = 2 &\Rightarrow x^2 + x + 1 = f \\ n = 3, p = 2 &\Rightarrow x^3 + x + 1 = f \text{ or } x^3 + x^2 + 1 = f. \end{aligned}$$

Let $A = C(f)$, the companion matrix of f. Then we know $f(A) = 0$.

We assert that the set of all matrices of the form $g(A)$, where $g \in \mathbb{Z}_p[x]$, forms a field consisting of precisely p^n elements. The typical element is

$$b_0 I_n + b_1 A + \cdots + b_t A^t$$

where $b_0, \ldots, b_t \in \mathbb{Z}_p$.

We need only show existence of a multiplicative inverse for each element except 0 (the additive identity), as the remaining axioms clearly hold.

So let $g \in \mathbb{Z}_p[x]$ such that $g(A) \neq 0$. We have to find $h \in \mathbb{Z}_p[x]$ satisfying

$$g(A)h(A) = I_n.$$

Note that $g(A) \neq 0 \Rightarrow f \not| g$, since

$$f \mid g \Rightarrow g = ff_1$$

and hence

$$g(A) = f(A)f_1(A) = 0f_1(A) = 0.$$

Then since f is irreducible and $f \not| g$, there exist u, $v \in \mathbb{Z}_p[x]$ such that

$$uf + vg = 1.$$

Hence $u(A)f(A) + v(A)g(A) = I_n$ and $v(A)g(A) = I_n$, as required.

We now show that our new field is a $\mathbb{Z}_p$–vector space with basis consisting of the matrices

$$I_n,\ A, \ldots, A^{n-1}.$$

Firstly the spanning property: By Euclid's division theorem,

$$g = fq + r$$

where $q, r \in \mathbb{Z}_p[x]$ and $\deg r < \deg g$. So let

$$r = r_0 + r_1 x + \cdots + r_{n-1} x^{n-1}$$

where $r_0, \ldots, r_{n-1} \in \mathbb{Z}_p$. Then

$$\begin{aligned} g(A) &= f(A)q(A) + r(A) \\ &= 0q(A) + r(A) \\ &= r(A) \\ &= r_0 I_n + r_1 A + \cdots + r_{n-1} A^{n-1} \end{aligned}$$

Secondly, linear independence over Z_p: Suppose that

$$r_0 I_n + r_1 A + \cdots + r_{n-1} A^{n-1} = 0,$$

where r_0, $r_1, \ldots, r_{n-1} \in \mathbb{Z}_p$. Then $r(A) = 0$, where

$$r = r_0 + r_1 x + \cdots + r_{n-1} x^{n-1}.$$

Hence $m_A = f$ divides r. Consequently $r = 0$, as $\deg f = n$ whereas $\deg r < n$ if $r \neq 0$.

Consequently, there are p^n such matrices $g(A)$ in the field we have constructed.

Numerical Examples

EXAMPLE 2.8

Let $p = 2$, $n = 2$, $f = x^2 + x + 1 \in \mathbb{Z}_2[x]$, and $A = C(f)$. Then

$$A = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

and

$$\begin{aligned} F_4 &= \{\, a_0 I_2 + a_1 A \mid a_0, a_1 \in \mathbb{Z}_2 \,\} \\ &= \{\, 0,\, I_2,\, A,\, I_2 + A \,\}. \end{aligned}$$

We construct addition and multiplication tables for this field, with $B = I_2 + A$ (as an exercise, check these):

$\oplus$	0	I_2	A	B
0	0	I_2	A	B
I_2	I_2	0	B	A
A	A	B	0	I_2
B	B	A	I_2	0

$\otimes$	0	I_2	A	B
0	0	0	0	0
I_2	0	I_2	A	B
A	0	A	B	I_2
B	0	B	I_2	A

EXAMPLE 2.9

Let $p = 2$, $n = 3$, $f = x^3 + x + 1 \in \mathbb{Z}_2[x]$. Then

$$A = C(f) = \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

and our eight-member field F_8 (usually denoted by $GF(8)$ ["GF" corresponds to "Galois Field", in honour of Galois]) is

$$\begin{aligned} F_8 &= \{\, a_0 I_3 + a_1 A + a_2 A^2 \mid a_0, a_1, a_2 \in \mathbb{Z}_2 \,\} \\ &= \{\, 0, I_3, A, A^2, I_3 + A, I_3 + A^2, A + A^2, I_3 + A + A^2 \,\}. \end{aligned}$$

Now find $(A^2 + A)^{-1}$.

Solution*: use Euclid's algorithm.*

$$x^3 + x + 1 = (x + 1)(x^2 + x) + 1.$$

Hence

$$\begin{aligned} x^3 + x + 1 + (x + 1)(x^2 + x) &= 1 \\ A^3 + A + I_3 + (A + I_3)(A^2 + A) &= I_3 \\ (A + I_3)(A^2 + A) &= I_3. \end{aligned}$$

Hence $(A^2 + A)^{-1} = A + I_3$.

THEOREM 2.8

Every finite field has precisely p^n elements for some prime p—the least positive integer with the property that

$$\underbrace{1 + 1 + 1 + \cdots + 1}_{p} = 0.$$

p is then called the **characteristic** *of the field.*

Also, if $x \in F$, a field of q elements, then it can be shown that if $x \neq 0$, then

$$x^{q-1} = 1.$$

In the special case $F = \mathbb{Z}_p$, this reduces to **Fermat's Little Theorem**:

$$x^{p-1} \equiv 1 \pmod{p},$$

if p is prime not dividing x.

2.6 Characteristic and Minimum Polynomial of a Transformation

DEFINITION 2.8

(Characteristic polynomial of $T : V \mapsto V$)

Let β be a basis for V and $A = [T]_\beta^\beta$.

Then we define $\mathrm{ch}_T = \mathrm{ch}_A$. This polynomial is independent of the basis β:

PROOF (ch_T is independent of the basis.)

If γ is another basis for V and $B = [T]_\gamma^\gamma$, then we know $A = P^{-1}BP$ where P is the change of basis matrix $[I_V]_\beta^\gamma$.

Then

$$\begin{aligned}
\mathrm{ch}_A &= \mathrm{ch}_{P^{-1}BP} \\
&= \det(xI_n - P^{-1}BP) \quad \text{where } n = \dim V \\
&= \det(P^{-1}(xI_n)P - P^{-1}BP) \\
&= \det(P^{-1}(xI_n - B)P) \\
&= \det P^{-1}\, \mathrm{ch}_B\, \det P \\
&= \mathrm{ch}_B\,.
\end{aligned}$$

DEFINITION 2.9

If $f = a_0 + \cdots + a_t x^t$, where $a_0, \ldots, a_t \in F$, we define

$$f(T) = a_0 I_V + \cdots + a_t T^t.$$

Then the usual properties hold:

$f,\, g \in F[x] \Rightarrow (f+g)(T) = f(T)+g(T)$ and $(fg)(T) = f(T)g(T) = g(T)f(T)$.

LEMMA 2.1

$$f \in F[x] \Rightarrow [f(T)]_\beta^\beta = f\left([T]_\beta^\beta\right).$$

Note: The Cayley-Hamilton theorem for matrices says that $\mathrm{ch}_A(A) = 0$. Then if $A = [T]_\beta^\beta$, we have by the lemma

$$[\mathrm{ch}_T(T)]_\beta^\beta = \mathrm{ch}_T(A) = \mathrm{ch}_A(A) = 0,$$

so $\mathrm{ch}_T(T) = 0_V$.

DEFINITION 2.10

Let $T : V \to V$ be a linear transformation over F. Then any polynomial of least positive degree such that

$$f(T) = 0_V$$

is called a minimum polynomial of T.

We have corresponding results for polynomials in a transformation T to those for polynomials in a square matrix A:

$$g = qf + r \Rightarrow g(T) = q(T)f(T) + r(T).$$

Again, there is a unique monic minimum polynomial of T is denoted by m_T and called "the" minimum polynomial of T.

Also note that because of the lemma,

$$m_T = m_{[T]_\beta^\beta}.$$

For (with $A = [T]_\beta^\beta$)

(a) $m_A(A) = 0$, so $m_A(T) = 0_V$. Hence $m_T | m_A$.

(b) $m_T(T) = 0_V$, so $[m_T(T)]_\beta^\beta = 0$. Hence $m_T(A) = 0$ and so $m_A | m_T$.

EXAMPLES 2.2

$T = 0_V \Leftrightarrow m_T = x.$

$T = I_V \Leftrightarrow m_T = x - 1.$

$T = cI_V \Leftrightarrow m_T = x - c.$

$T^2 = T$ *and* $T \neq 0_V$ *and* $T \neq I_V \Leftrightarrow m_T = x^2 - x.$

2.6.1 $M_{n\times n}(F[x])$—Ring of Polynomial Matrices

EXAMPLE:

$$\begin{bmatrix} x^2+2 & x^5+5x+1 \\ x+3 & 1 \end{bmatrix} \in M_{2\times 2}(\mathbb{Q}[x])$$

$$= x^5 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + x^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + x \begin{bmatrix} 0 & 5 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix}$$

—we see that any element of $M_{n\times n}(F[x])$ is expressible as

$$x^m A_m + x^{m-1} A_{m-1} + \cdots + A_0$$

where $A_i \in M_{n\times n}(F)$. We write the coefficient of x^i after x^i, to distinguish these entities from corresponding objects of the following ring.

2.6.2 $M_{n\times n}(F)[y]$—Ring of Matrix Polynomials

This consists of all polynomials in y with coefficients in $M_{n\times n}(F)$.
EXAMPLE:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} y^5 + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} y^2 + \begin{bmatrix} 0 & 5 \\ 1 & 0 \end{bmatrix} y + \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix} \in M_{2\times 2}(F)[y].$$

THEOREM 2.9

The mapping

$$\Phi : M_{n\times n}(F)[y] \mapsto M_{n\times n}(F[x])$$

given by

$$\Phi(A_0 + A_1 y + \cdots + A_m y^m) = A_0 + xA_1 + \cdots + x^m A_m$$

where $A_i \in M_{n\times n}(F)$, is a 1–1 correspondence and has the following properties:

$$\begin{aligned} \Phi(X+Y) &= \Phi(X) + \Phi(Y) \\ \Phi(XY) &= \Phi(X)\Phi(Y) \\ \Phi(tX) &= t\Phi(X) \quad \forall t \in F. \end{aligned}$$

Also

$$\Phi(I_n y - A) = xI_n - A \quad \forall A \in M_{n\times n}(F).$$

THEOREM 2.10 ((Left) Remainder theorem for matrix polynomials)

Let $B_m y^m + \cdots + B_0 \in M_{n\times n}(F)[y]$ and $A \in M_{n\times n}(F)$.
Then

$$B_m y^m + \cdots + B_0 = (I_n y - A)Q + R$$

where

$$\begin{aligned} R &= A^m B_m + \cdots + AB_1 + B_0 \\ \text{and } Q &= C_{m-1} y^{m-1} + \cdots + C_0 \end{aligned}$$

where $C_{m-1}, \ldots, C_0$ are computed recursively:

$$\begin{aligned} B_m &= C_{m-1} \\ B_{m-1} &= -AC_{m-1} + C_{m-2} \\ &\vdots \\ B_1 &= -AC_1 + C_0. \end{aligned}$$

PROOF. First we verify that $B_0 = -AC_0 + R$:

$$\begin{array}{rcl} R = A^m B_m & = & A^m C_{m-1} \\ +A^{m-1}B_{m-1} & & -A^m C_{m-1} + A^{m-1}C_{m-2} \\ + & & + \\ \vdots & & \vdots \\ +AB_1 & & -A^2 C_1 + AC_0 \\ +B_0 & & B_0 \\ & = & B_0 + AC_0. \end{array}$$

Then

$$\begin{aligned} (I_n y - A)Q + R &= (I_n y)(C_{m-1}y^{m-1} + \cdots + C_0) \\ &\quad - A(C_{m-1}y^{m-1} + \cdots + C_0) + A^m B_m + \cdots + B_0 \\ &= C_{m-1}y^m + (C_{m-2} - AC_{m-1})y^{m-1} + \cdots + (C_0 - AC_1)y + \\ &\quad - AC_0 + R \\ &= B_m y^m + B_{m-1}y^{m-1} + \cdots + B_1 y + B_0. \end{aligned}$$

Remark. There is a similar "right" remainder theorem.

THEOREM 2.11

If p is an irreducible polynomial dividing ch_A *, then* $p \mid m_A$.

PROOF (From Burton Jones, "Linear Algebra").

Let $m_A = x^t + a_{t-1}x^{t-1} + \cdots + a_0$ and consider the matrix polynomial in y

$$\begin{aligned} \Phi^{-1}(m_A I_n) &= I_n y^t + (a_{t-1}I_n)y^{t-1} + \cdots + (a_0 I_n) \\ &= (I_n y - A)Q + A^t I_n + A^{t-1}(a_{t-1}I_n) + \cdots + a_0 I_n \\ &= (I_n y - A)Q + m_T(A) \\ &= (I_n y - A)Q. \end{aligned}$$

Now take Φ of both sides to give

$$m_A I_n = (xI_n - A)\Phi(Q)$$

and taking determinants of both sides yields

$$\{m_A\}^n = \mathrm{ch}_A \times \det \Phi(Q).$$

So letting p be an irreducible polynomial dividing ch_A, we have $p \mid \{m_A\}^n$ and hence $p \mid m_A$.

Alternative simpler proof (MacDuffee):

$m_A(x) - m_A(y) = (x-y)k(x,y)$, where $k(x,y) \in F[x,y]$. Hence

$$m_A(x)I_n = m_A(xI_n) - m_A(A) = (xI_n - A)k(xI_n, A).$$

Now take determinants to get

$$m_A(x)^n = ch_A(x) \det k(xI_n, A).$$

Exercise: If $\Delta(x)$ is the gcd of the elements of $adj(xI_n - A)$, use the equation $(xI_n - a)adj(xI_n - A) = ch_A(x)I_n$ and an above equation to deduce that $m_A(x) = ch_A(x)/\Delta(x)$.

EXAMPLES 2.3

With $A = 0 \in M_{n\times n}(F)$, *we have* $\mathrm{ch}_A = x^n$ *and* $m_A = x$.

$A = \mathrm{diag}\,(1,1,2,2,2) \in M_{5\times 5}(\mathbb{Q})$. *Here*

$$\mathrm{ch}_A = (x-1)^2(x-2)^3 \quad \textit{and} \quad m_A = (x-1)(x-2).$$

DEFINITION 2.11

A matrix $A \in M_{n\times n}(F)$ *is called* diagonable *over* F *if there exists a non–singular matrix* $P \in M_{n\times n}(F)$ *such that*

$$P^{-1}AP = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n),$$

where $\lambda_1, \ldots, \lambda_n$ *belong to* F.

THEOREM 2.12

If A *is diagonable, then* m_A *is a product of distinct linear factors.*

PROOF

If $P^{-1}AP = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$ (with $\lambda_1, \ldots, \lambda_n \in F$) then

$$\begin{aligned} m_A &= m_{P^{-1}AP} = m_{\mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)} \\ &= (x-c_1)(x-c_2)\ldots(x-c_t) \end{aligned}$$

where $c_1, \ldots, c_t$ are the distinct members of the sequence $\lambda_1, \ldots, \lambda_n$.

The converse is also true, and will (fairly) soon be proved.

EXAMPLE 2.10

$$A = J_n(a).$$

We saw earlier that $m_A = (x-a)^n$ *so if* $n \geq 2$ *we see that* A *is not diagonable.*

DEFINITION 2.12

*(***Diagonable LTs***)*

$T : V \mapsto V$ *is called* **diagonable** *over* F *if there exists a basis* β *for* V *such that* $[T]_\beta^\beta$ *is diagonal.*

THEOREM 2.13

A is diagonable $\Leftrightarrow$ T_A *is diagonable.*

PROOF (Sketch)

$\Rightarrow$ Suppose $P^{-1}AP = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$. Now pre-multiplying by P and letting $P = [P_1|\cdots|P_n]$ we see that

$$\begin{aligned} T_A(P_1) &= AP_1 = \lambda_1 P_1 \\ &\vdots \\ T_A(P_n) &= AP_n = \lambda_n P_n \end{aligned}$$

and we let β be the basis $P_1, \ldots, P_n$ over $V_n(F)$. Then

$$[T_A]_\beta^\beta = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

$\Leftarrow$ Reverse the argument and use Theorem 1.17.

THEOREM 2.14

Let $A \in M_{n\times n}(F)$. *Then if* λ *is an eigenvalue of* A *with multiplicity* m, *(that is* $(x-\lambda)^m$ *is the exact power of* $x-\lambda$ *which divides* ch_A *), we have*

$$\textit{nullity}\,(A - \lambda I_n) \leq m.$$

REMARKS. (1) If $m = 1$, we deduce that nullity $(A - \lambda I_n) = 1$. For the inequality

$$1 \leq \text{nullity}\,(A - \lambda I_n)$$

always holds.

(2) The integer nullity $(A - \lambda I_n)$ is called the *geometric multiplicity* of the eigenvalue λ, while m is referred to as the *algebraic multiplicity* of λ.

PROOF. Let $v_1, \ldots, v_r$ be a basis for $N(A - \lambda I_n)$, where λ is an eigenvalue of A having multiplicity m. Extend this linearly independent family to a basis $v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$ of $V_n(F)$. Then the following equations hold:

$$\begin{aligned} Av_1 &= \lambda v_1 \\ &\vdots \\ Av_r &= \lambda v_r \\ Av_{r+1} &= b_{11}v_1 + \cdots + b_{n1}v_n \\ &\vdots \\ Av_n &= b_{1n-r}v_1 + \cdots + b_{nn-r}v_n. \end{aligned}$$

These equations can be combined into a single matrix equation:

$$\begin{aligned} A[v_1|\cdots|v_r|v_{r+1}|\cdots|v_n] &= [Av_1|\cdots|Av_r|Av_{r+1}|\cdots|Av_n] \\ &= [\lambda v_1|\cdots|\lambda v_r|b_{11}v_1 + \cdots + b_{n1}v_n|\cdots|b_{1n-r}v_1 + \cdots + b_{nn-r}v_n] \\ &= [v_1|\cdots|v_n]\left[\begin{array}{c|c} \lambda I_r & B_1 \\ \hline 0 & B_2 \end{array}\right]. \end{aligned}$$

Hence if $P = [v_1|\cdots|v_n]$, we have

$$P^{-1}AP = \left[\begin{array}{c|c} \lambda I_r & B_1 \\ \hline 0 & B_2 \end{array}\right].$$

Then

$$\text{ch}_A \,=\, \text{ch}_{P^{-1}AP} \,=\, \text{ch}_{\lambda I_r} \cdot \text{ch}_{B_2} \,=\, (x - \lambda)^r \,\text{ch}_{B_2}$$

and because $(x - \lambda)^m$ is the exact power of $x - \lambda$ dividing ch_A, it follows that

$$\text{nullity}\,(A - \lambda I_n) = r \leq m.$$

THEOREM 2.15

Suppose that $\text{ch}_T = (x - c_1)^{a_1} \cdots (x - c_t)^{a_t}$. *Then* T *is diagonable if*

$$\text{nullity}\,(T - c_i I_v) = a_i \quad \textit{for } 1 \leq i \leq t.$$

PROOF. We first prove that the subspaces $\text{Ker}\,(T - c_i I_V)$ are independent.

(Subspaces $V_1, \ldots, V_t$ are called *independent* if

$$v_1 + \cdots + v_t = 0, v_i \in V_i, i = 1, \ldots t, \Rightarrow v_1 = 0, \ldots, v_t = 0.$$

Then $\dim(V_1 + \cdots + V_t) = \dim(V_1) + \cdots + \dim V_t)$.)

Assume that

$$v_1 + \cdots + v_t = 0,$$

where $v_i \in \text{Ker}\,(T - c_i I_v)$ for $1 \leq i \leq t$. Then

$$\begin{aligned} T(v_1 + \cdots + v_t) &= T(0) \\ c_1 v_1 + \cdots + c_t v_t &= 0. \end{aligned}$$

Similarly we deduce that

$$\begin{aligned} c_1^2 v_1 + \cdots + c_t^2 v_t &= 0 \\ &\vdots \\ c_1^{t-1} v_1 + \cdots + c_t^{t-1} v_t &= 0. \end{aligned}$$

We can combine these t equations into a single matrix equation

$$\begin{bmatrix} 1 & \cdots & 1 \\ c_1 & \cdots & c_t \\ & \vdots & \\ c_1^{t-1} & \cdots & c_t^{t-1} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_t \end{bmatrix} = \begin{bmatrix} o \\ \vdots \\ 0 \end{bmatrix}.$$

However the coefficient matrix is the Vandermonde matrix, which is non–singular as $c_i \neq c_j$ if $i \neq j$, so we deduce that $v_1 = 0, \cdots, v_t = 0$. Hence with $V_i = \text{Ker}\,(T - c_i I_V)$, we have

$$\dim(V_1 + \cdots + V_t) = \sum_{i=1}^{t} \dim V_i = \sum_{i=1}^{t} a_i = \dim V.$$

Hence

$$V = V_1 + \cdots + V_t.$$

Then if β_i is a basis for V_i for $i \leq i \leq t$ and $\beta = \beta_1 \cup \cdots \cup \beta_t$, it follows that β is a basis for V. Moreover

$$[T]_\beta^\beta = \bigoplus_{i=1}^{t} (c_i I_{a_i})$$

and T is diagonable.

EXAMPLE. Let

$$A = \begin{bmatrix} 5 & 2 & -2 \\ 2 & 5 & -2 \\ -2 & -2 & 5 \end{bmatrix}.$$

(a) We find that $\mathrm{ch}_A = (x-3)^2(x-9)$. Next we find bases for each of the eigenspaces $N(A-9I_3)$ and $N(A-3I_3)$:

First we solve $(A-3I_3)X = 0$. We have

$$A - 3I_3 = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Hence the eigenspace consists of vectors $X = [x,\, y,\, z]^t$ satisfying $x = -y+z$, with y and z arbitrary. Hence

$$X = \begin{bmatrix} -y+z \\ y \\ z \end{bmatrix} = y \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

so $X_{11} = [-1,\, 1,\, 0]^t$ and $X_{12} = [1,\, 0,\, 1]^t$ form a basis for the eigenspace corresponding to the eigenvalue 3.

Next we solve $(A-9I_3)X = 0$. We have

$$A - 9I_3 = \begin{bmatrix} -4 & 2 & -2 \\ 2 & -4 & -2 \\ -2 & -2 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Hence the eigenspace consists of vectors $X = [x,\, y,\, z]^t$ satisfying $x = -z$ and $y = -z$, with z arbitrary. Hence

$$X = \begin{bmatrix} -z \\ -z \\ z \end{bmatrix} = z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

and we can take $X_{21} = [-1,\, -1,\, 1]^t$ as a basis for the eigenspace corresponding to the eigenvalue 9.

Then $P = [X_{11}|X_{12}|X_{21}]$ is non–singular and

$$P^{-1}AP = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

THEOREM 2.16

If

$$m_T = (x - c_1) \dots (x - c_t)$$

for $c_1, \dots, c_t$ distinct in F, then T is diagonable and conversely. Moreover there exist unique linear transformations $T_1, \dots, T_t$ satisfying

$$\begin{aligned} I_V &= T_1 + \cdots + T_t, \\ T &= c_1 T_1 + \cdots + c_t T_t, \\ T_i T_j &= 0_V \text{ if } i \neq j, \\ T_i^2 &= T_i, \ 1 \leq i \leq t. \end{aligned}$$

Also $\operatorname{rank} T_i = a_i$, *where* $ch_T = (x - c_1)^{a_1} \cdots (x - c_t)^{a_t}$.

Remarks.

1. $T_1, \dots, T_t$ are called the *principal idempotents* of T.
2. If $g \in F[x]$, then $g(T) = g(c_1)T_1 + \cdots + g(c_t)T_t$. For example
$$T^m = c_1^m T_1 + \cdots + c_t^m T_t.$$
3. If $c_1, \dots, c_t$ are non–zero (that is the eigenvalues of T are non–zero), the T^{-1} is given by
$$T^{-1} = c_1^{-1} T_1 + \cdots + c_t^{-1} T_t.$$

Formulae 2 and 3 are useful in the corresponding matrix formulation. PROOF Suppose $m_T = (x - c_1) \cdots (x - c_t)$, where $c_1, \dots, c_t$ are distinct. Then $ch_T = (x - c_1)^{a_1} \cdots (x - c_t)^{a_t}$. To prove T is diagonable, we have to prove that $\text{nullity}\,(T - c_i I_V) = a_i, \ 1 \leq i \leq t$

Let $p_1, \dots, p_t$ be the Lagrange interpolation polynomials based on $c_1, \dots, c_t$, i.e.

$$p_i = \prod_{\substack{j=1 \\ j \neq i}}^{t} \left(\frac{x - c_j}{c_i - c_j} \right), \ 1 \leq i \leq t.$$

Then

$$g \in F[x] \Rightarrow g = g(c_1)p_1 + \cdots + g(c_t)p_t.$$

In particular,

$$g = 1 \Rightarrow 1 = p_1 + \cdots + p_t$$

and

$$g = x \Rightarrow x = c_1 p_1 + \cdots + c_t p_t.$$

Hence with $T_i = p_i(T)$,

$$\begin{aligned} I_V &= T_1 + \cdots + T_t \\ T &= c_1 T_1 + \cdots + c_t T_t. \end{aligned}$$

Next

$$\begin{aligned} & m_T = (x - c_1) \ldots (x - c_t) \mid p_i p_j \qquad \text{if } i \neq j \\ \Rightarrow \quad & (p_i p_j)(T) = 0_V \qquad \text{if } i \neq j \\ \Rightarrow \quad & p_i(T) p_j(T) = 0_V \quad \text{or} \quad T_i T_j = 0_V \qquad \text{if } i \neq j. \end{aligned}$$

Then $T_i^2 = T_i(T_1 + \cdots + T_t) = T_i I_V = T_i$.

Next

$$0_V = m_T(T) = (T - c_1 I_V) \cdots (T - c_t I_V).$$

Hence

$$\dim V = \text{ nullity } 0_V \leq \sum_{i=1}^{t} \text{nullity } (T - c_i I_V) \leq \sum_{i=1}^{t} a_i = \dim V.$$

Consequently $\text{nullity } (T - c_i I_V) = a_i$, $1 \leq i \leq t$ and T is therefore diagonable.

Next we prove that $\text{rank } T_i = a_i$. From the definition of p_i, we have

$$\text{nullity } p_i(T) \leq \sum_{\substack{j=1 \\ j \neq i}}^{t} \text{nullity } (T - c_j I_V) = \sum_{\substack{j=1 \\ j \neq i}}^{t} a_j = \dim V - a_i.$$

Also $p_i(T)(T - c_i I_V) = 0$, so $\text{Im } (T - c_i I_V) \subseteq \text{Ker } p_i(T)$. Hence

$$\dim V - a_i \leq \text{ nullity } p_i(T)$$

and consequently $\text{nullity } p_i(T) = \dim (V) - a_i$, so $\text{rank } p_i(T) = a_i$.

We next prove the uniqueness of $T_1, \ldots, T_t$. Suppose that $S_1, \ldots, S_t$ also satisfy the same conditions as $T_1, \ldots, T_t$. Then

$$\begin{aligned} T_i T &= T T_i = c_i T_i \\ S_j T &= T S_j = c_j S_j \\ T_i(T S_j) &= T_i(c_j S_j) = c_j T_i S_j = (T_i T) S_j = c_i T_i S_j \end{aligned}$$

so $(c_j - c_i)T_iS_j = 0_V$ and $T_iS_j = 0_V$ if $i \neq j$. Hence

$$\begin{aligned} T_i &= T_iI_V = T_i(\sum_{j=1}^{t} S_j) = T_iS_i \\ S_i &= I_VS_i = (\sum_{j=1}^{t} T_j)S_i = T_iS_i. \end{aligned}$$

Hence $T_i = S_i$.

Conversely, suppose that T is diagonable and let β be a basis of V such that

$$A = [T]_\beta^\beta = \text{diag}\,(\lambda_1, \ldots, \lambda_n).$$

Then $m_T = m_A = (x - c_1) \cdots (x - c_t)$, where $c_1, \ldots, c_t$ are the distinct members of the sequence $\lambda_1, \ldots, \lambda_n$.

COROLLARY 2.5

If

$$\text{ch}_T = (x - c_1) \ldots (x - c_t)$$

with c_i distinct members of F, then T is diagonable.
PROOF*: Here $m_T = \text{ch}_T$ and we use theorem 3.3.*

EXAMPLE 2.11

Let

$$A = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix} \qquad a, b \in F,\ ab \neq 0,\ 1 + 1 \neq 0.$$

Then A is diagonable if and only if $ab = y^2$ for some $y \in F$.

For $\text{ch}_A = x^2 - ab$, so if $ab = y^2$,

$$\text{ch}_A = x^2 - y^2 = (x + y)(x - y)$$

which is a product of distinct linear factors, as $y \neq -y$ here.

Conversely suppose that A is diagonable. Then as A is not a scalar matrix, it follows that m_A is not linear and hence

$$m_A = (x - c_1)(x - c_2),$$

where $c_1 \neq c_2$. Also $\text{ch}_A = m_A$, so $\text{ch}_A\,(c_1) = 0$. Hence

$$c_1^2 - ab = 0, \quad \text{or} \quad ab = c_1^2.$$

For example, take $F = \mathbb{Z}_7$ and let $a = 1$ and $b = 3$. Then $ab \neq y^2$ and consequently A is not diagonable.