# LINEAR ALGEBRA NOTES

# MP274 1991

K. R. MATTHEWS
LaTeXed by Chris Fama

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF QUEENSLAND

1991

Comments to the author at krm@maths.uq.edu.au

# Contents

# 1 Linear Transformations

We will study mainly finite-dimensional vector spaces over an arbitrary field $F$—i.e. vector spaces with a basis. (Recall that the dimension of a vector space $V$ ($\dim V$) is the number of elements in a basis of $V$.)

**DEFINITION 1.1**

(**Linear transformation**)

Given vector spaces $U$ and $V$, $T : U \mapsto V$ is a linear transformation (LT) if

$$T(\lambda u + \mu v) = \lambda T(u) + \mu T(v)$$

for all $\lambda, \mu \in F$, and $u, v \in U$. Then $T(u+v) = T(u)+T(v)$, $T(\lambda u) = \lambda T(u)$ and

$$T\left(\sum_{k=1}^{n} \lambda_k u_k\right) = \sum_{k=1}^{n} \lambda_k T(u_k).$$

**EXAMPLES 1.1**

Consider the linear transformation

$$T = T_A : V_n(F) \mapsto V_m(F)$$

where $A = [a_{ij}]$ is $m \times n$, defined by $T_A(X) = AX$.

Note that $V_n(F)$ = the set of all $n$-dimensional column vectors $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ of $F$—sometimes written $F^n$.

Note that if $T : V_n(F) \mapsto V_m(F)$ is a linear transformation, then $T = T_A$, where $A = [T(E_1)|\cdots|T(E_n)]$ and

$$E_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, E_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

**Note:**

$$v \in V_n(F), \qquad v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 E_1 + \cdots + x_n E_n$$

If $V$ is a vector space of all infinitely differentiable functions on $\mathbb{R}$, then

$$T(f) = a_0 \mathrm{D}^n f + a_1 \mathrm{D}^{n-1} f + \cdots + a_{n-1} \mathrm{D} f + a_n f$$

defines a linear transformation $T : V \mapsto V$.

The set of $f$ such that $T(f) = 0$ (i.e. the kernel of $T$) is important.

Let $T : U \mapsto V$ be a linear transformation. Then we have the following definition:

**DEFINITIONS 1.1**

(**Kernel of a linear transformation**)

$$\operatorname{Ker} T = \{u \in U \mid T(u) = 0\}$$

(**Image of** $T$)

$$\operatorname{Im} T = \{v \in V \mid \exists u \in U \text{ such that } T(u) = v\}$$

**Note**: $\operatorname{Ker} T$ is a subspace of $U$. Recall that $W$ is a subspace of $U$ if

1. $0 \in W$,
2. $W$ is closed under addition, and
3. $W$ is closed under scalar multiplication.

PROOF. that $\operatorname{Ker} T$ is a subspace of $U$:

1. $T(0) + 0 = T(0) = T(0 + 0) = T(0) + T(0)$. Thus $T(0) = 0$, so $0 \in \operatorname{Ker} T$.
2. Let $u, v \in \operatorname{Ker} T$; then $T(u) = 0$ and $T(v) = 0$. So $T(u + v) = T(u) + T(v) = 0 + 0 = 0$ and $u + v \in \operatorname{Ker} T$.
3. Let $u \in \operatorname{Ker} T$ and $\lambda \in F$. Then $T(\lambda u) = \lambda T(u) = \lambda 0 = 0$. So $\lambda u \in \operatorname{Ker} T$.

**EXAMPLE 1.1**

$$\begin{aligned} \operatorname{Ker} T_A &= N(A), \text{ the null space of } A \\ &= \{X \in V_n(F) \mid AX = 0\} \\ \text{and } \operatorname{Im} T_A &= C(A), \text{ the column space of } A \\ &= \langle A_{*1}, \ldots, A_{*n} \rangle \end{aligned}$$

Generally, if $U = \langle u_1, \ldots, u_n \rangle$, then $\operatorname{Im} T = \langle T(u_1), \ldots, T(u_n) \rangle$.

**Note:** Even if $u_1, \ldots, u_n$ form a basis for $U$, $T(u_1), \ldots, T(u_n)$ may not form a basis for $\operatorname{Im} T$. I.e. it may happen that $T(u_1), \ldots, T(u_n)$ are linearly dependent.

## 1.1 Rank + Nullity Theorems (for Linear Maps)

**THEOREM 1.1 (General rank + nullity theorem)**
*If $T : U \mapsto V$ is a linear transformation then*

$$\operatorname{rank} T + \operatorname{nullity} T = \dim U.$$

PROOF.

1. $\operatorname{Ker} T = \{0\}$.
   Then $\operatorname{nullity} T = 0$.
   We first show that the vectors $T(u_1), \ldots, T(u_n)$, where $u_1, \ldots, u_n$ are a basis for $U$, are LI (linearly independent):
   Suppose $x_1 T(u_1) + \cdots + x_n T(u_n) = 0$ where $x_1, \ldots, x_n \in F$.
   Then

$$\begin{aligned} T(x_1 u_1 + \cdots + x_n u_n) &= 0 \qquad \text{(by linearity)} \\ x_1 u_1 + \cdots + x_n u_n &= 0 \qquad (\text{since } \operatorname{Ker} T = \{0\}) \\ x_1 = 0, \ldots, x_n = 0 & \qquad (\text{since } u_i \text{ are LI}) \end{aligned}$$

   Hence $\operatorname{Im} T = \langle T(u_1), \ldots, T(u_n) \rangle$ so

$$\operatorname{rank} T + \operatorname{nullity} T = \dim \operatorname{Im} T + 0 = n = \dim V.$$

2. $\operatorname{Ker} T = U$.
   So $\operatorname{nullity} T = \dim U$.
   Hence $\operatorname{Im} T = \{0\} \Rightarrow \operatorname{rank} T = 0$

$$\begin{aligned} \Rightarrow \operatorname{rank} T + \operatorname{nullity} T &= 0 + \dim U \\ &= \dim U. \end{aligned}$$

3. $0 < \operatorname{nullity} T < \dim U$.
   Let $u_1, \ldots, u_r$ be a basis for $\operatorname{Ker} T$ and $n = \dim U$, so $r = \operatorname{nullity} T$ and $r < n$.
   Extend the basis $u_1, \ldots, u_r$ to form a basis $u_1, \ldots, u_r, u_{r+1}, \ldots, u_n$ of

$U$ (refer to last year's notes to show that this can be done).
Then $T(u_{r+1}), \ldots, T(u_n)$ span $\operatorname{Im} T$. For

$$\begin{aligned}\operatorname{Im} T &= \langle T(u_1), \ldots, T(u_r), T(u_{r+1}), \ldots, T(u_n)\rangle \\ &= \langle 0, \ldots, 0, T(u_{r+1}), \ldots, T(u_n)\rangle \\ &= \langle T(u_{r+1}), \ldots, T(u_n)\rangle\end{aligned}$$

So assume

$$\begin{aligned}& x_1T(u_{r+1}) + \cdots + x_{n-r}T(u_n) = 0 \\ \Rightarrow\ & T(x_1u_{r+1} + \cdots + x_{n-r}u_n) = 0 \\ \Rightarrow\ & x_1u_{r+1} + \cdots + x_{n-r}u_n \in \operatorname{Ker} T \\ \Rightarrow\ & x_1u_{r+1} + \cdots + x_{n-r}u_n = y_1u_1 + \cdots + y_ru_r \\ & \qquad \text{for some } y_1, \ldots, y_r \\ \Rightarrow\ & (-y_1)u_1 + \cdots + (-y_r)u_r + x_1u_{r+1} + \cdots + x_{n-r}u_n = 0\end{aligned}$$

and since $u_1, \ldots, u_n$ is a basis for $U$, all coefficients vanish.

Thus

$$\begin{aligned}\operatorname{rank} T + \operatorname{nullity} T &= (n-r) + r \\ &= n \\ &= \dim U.\end{aligned}$$

We now apply this theorem to prove the following result:

**THEOREM 1.2 (Dimension theorem for subspaces)**

$$\dim(U \cap V) + \dim(U + V) = \dim U + \dim V$$

*where $U$ and $V$ are subspaces of a vector space $W$.*

(Recall that $U + V = \{u + v \mid u \in U, v \in V\}$.)
For the proof we need the following definition:

**DEFINITION 1.2**
*If $U$ and $V$ are any two vector spaces, then the direct sum is*

$$U \oplus V = \{(u, v) \mid u \in U, v \in V\}$$

*(i.e. the cartesian product of $U$ and $V$) made into a vector space by the component-wise definitions:*

1. $(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2)$,
2. $\lambda(u, v) = (\lambda u, \lambda v)$, *and*
3. $(0, 0)$ *is an identity for* $U \oplus V$ *and* $(-u, -v)$ *is an additive inverse for* $(u, v)$.

We need the following result:

**THEOREM 1.3**

$$\dim(U \oplus V) = \dim U + \dim V$$

PROOF.

**Case 1:** $U = \{0\}$

**Case 2:** $V = \{0\}$

Proof of cases 1 and 2 are left as an exercise.

**Case 3:** $U \neq \{0\}$ and $V \neq \{0\}$

Let $u_1, \ldots, u_m$ be a basis for $U$, and
$v_1, \ldots, v_n$ be a basis for $V$.

We assert that $(u_1, 0), \ldots, (u_m, 0), (0, v_1), \ldots, (0, v_n)$ form a basis for $U \oplus V$.

Firstly, spanning:

Let $(u, v) \in U \oplus V$, say $u = x_1u_1 + \cdots + x_mu_m$ and $v = y_1v_1 + \cdots + y_nv_n$. Then

$$\begin{aligned} (u, v) &= (u, 0) + (0, v) \\ &= (x_1u_1 + \cdots + x_mu_m, 0) + (0, y_1v_1 + \cdots + y_nv_n) \\ &= x_1(u_1, 0) + \cdots + x_m(u_m, 0) + y_1(0, v_1) + \cdots + y_n(0, v_n) \\ \text{So} \quad U \oplus V &= \langle (u_1, 0), \ldots, (u_m, 0), (0, v_1), \ldots, (0, v_n) \rangle \end{aligned}$$

Secondly, independence: assume $x_1(u_1, 0) + \cdots + x_m(u_m, 0) + y_1(0, v_1) + \cdots + y_n(0, v_n) = (0, 0)$. Then

$$\begin{aligned} (x_1u_1 + \cdots + x_mu_m, y_1v_1 + \cdots + y_nv_n) &= 0 \\ \Rightarrow \quad x_1u_1 + \cdots + x_mu_m &= 0 \\ \text{and } y_1v_1 + \cdots + y_nv_n &= 0 \\ \Rightarrow \quad x_i &= 0, \ \forall i \\ \text{and } y_i &= 0, \ \forall i \end{aligned}$$

Hence the assertion is true and the result follows.

PROOF.

Let $T : U \oplus V \mapsto U + V$ where $U$ and $V$ are subspaces of some $W$, such that $T(u, v) = u + v$.

Thus $\operatorname{Im} T = U + V$, and

$$\begin{aligned} \operatorname{Ker} T &= \{(u, v) \mid u \in U, v \in V, \text{ and } u + v = 0\} \\ &= \{(t, -t) \mid t \in U \cap V\} \end{aligned}$$

Clearly then, dim $\operatorname{Ker} T = \dim(U \cap V)$[1] and so

$$\begin{aligned} \operatorname{rank} T + \operatorname{nullity} T &= \dim(U \oplus V) \\ \Rightarrow \quad \dim(U + V) + \dim(U \cap V) &= \dim U + \dim V. \end{aligned}$$

## 1.2 Matrix of a Linear Transformation

**DEFINITION 1.3**

*Let $T : U \mapsto V$ be a LT with bases $\beta : u_1, \ldots, u_n$ and $\gamma : v_1, \ldots, v_m$ for $U$ and $V$ respectively.*

*Then*

$$T(u_j) = \begin{matrix} a_{1j}v_1 \\ + \\ a_{2j}v_2 \\ + \\ \vdots \\ + \\ a_{mj}v_m \end{matrix} \qquad \text{for some} \qquad \begin{matrix} a_{1j} \\ \vdots \\ a_{mj} \end{matrix} \in F.$$

*The $m \times n$ matrix*

$$A = [a_{ij}]$$

*is called the matrix of $T$ relative to the bases $\beta$ and $\gamma$ and is also written*

$$A = [T]_{\beta}^{\gamma}$$

**Note:** *The $j$-th column of $A$ is the co-ordinate vector of $T(u_j)$, where $u_j$ is the $j$-th vector of the basis $\beta$.*

*Also if $u = x_1u_1 + \cdots + x_nu_n$, the co-ordinate vector* $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ *is denoted by* $[u]_{\beta}$.

[1] True if $U \cap V = \{0\}$; if not, let $S = \operatorname{Ker} T$ and $u_1, \ldots, u_r$ be a basis for $U \cap V$. Then $(u_1, -u_1), \ldots, (u_r, -u_r)$ form a basis for S and hence dim $\operatorname{Ker} T = \dim S$.

**EXAMPLE 1.2**

*Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2}(F)$ *and let* $T : M_{2\times 2}(F) \mapsto M_{2\times 2}(F)$ *be defined by*

$$T(X) = AX - XA.$$

*Then* $T$ *is linear*[2]*, and* $\operatorname{Ker} T$ *consists of all* $2 \times 2$ *matrices* $A$ *where* $AX = XA$*.*

*Take* $\beta$ *to be the basis* $E_{11}$*,* $E_{12}$*,* $E_{21}$*, and* $E_{22}$*, defined by*

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

*(so we can define a matrix for the transformation, consider these henceforth to be column vectors of four elements).*

*Calculate* $[T]^{\beta}_{\beta} = B$ *:*

$$\begin{aligned} T(E_{11}) &= AE_{11} - E_{11}A \\ &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} 0 & -b \\ c & 0 \end{bmatrix} \\ &= 0E_{11} - bE_{12} + cE_{21} + 0E_{22} \end{aligned}$$

*and similar calculations for the image of other basis vectors show that*

$$B = \begin{bmatrix} 0 & -c & b & 0 \\ -b & a-d & 0 & b \\ c & 0 & d-a & -c \\ 0 & c & b & 0 \end{bmatrix}$$

EXERCISE: *Prove that* $\operatorname{rank} B = 2$ *if* $A$ *is not a scalar matrix (i.e. if* $A \neq tI_n$*).*

*Later, we will show that* $\operatorname{rank} B = \operatorname{rank} T$*. Hence*

$$\operatorname{nullity} T = 4 - 2 = 2$$

---

[2]

$$\begin{aligned} T(\lambda X + \mu Y) &= A(\lambda X + \mu Y) - (\lambda X + \mu Y)A \\ &= \lambda(AX - XA) + \mu(AY - YA) \\ &= \lambda T(X) + \mu T(Y) \end{aligned}$$

**Note:** *$I_2, A \in \operatorname{Ker} T$ which has dimension 2. Hence if $A$ is not a scalar matrix, since $I_2$ and $A$ are LI they form a basis for $\operatorname{Ker} T$. Hence*

$$AX = XA \quad \Rightarrow \quad X = \alpha I_2 + \beta A.$$

**DEFINITIONS 1.2**

*Let $T_1$ and $T_2$ be LT's mapping $U$ to $V$.*
*Then $T_1 + T_2 : U \mapsto V$ is defined by*

$$(T_1 + T_2)(x) = T_1(x) + T_2(x) \quad ; \forall x \in U$$

*For $T$ a LT and $\lambda \in F$, define $\lambda T : U \mapsto V$ by*

$$(\lambda T)(x) = \lambda T(x) \; \forall x \in U$$

Now ...

$$\begin{aligned} [T_1 + T_2]_\beta^\gamma &= [T_1]_\beta^\gamma + [T_2]_\beta^\gamma \\ [\lambda T]_\beta^\gamma &= \lambda [T]_\beta^\gamma \end{aligned}$$

**DEFINITION 1.4**

$$\operatorname{Hom}(U, V) = \{T | T : U \mapsto V \text{ is a LT}\}.$$

$\operatorname{Hom}(U, V)$ *is sometimes written $L(U, V)$.*

The zero transformation $0 : U \mapsto V$ is such that $0(x) = 0$, $\forall x$.
If $T \in \operatorname{Hom}(U, V)$, then $(-T) \in \operatorname{Hom}(U, V)$ is defined by

$$(-T)(x) = -(T(x)) \quad \forall x \in U.$$

Clearly, $\operatorname{Hom}(U, V)$ is a vector space.
Also

$$\begin{aligned} [0]_\beta^\gamma &= 0 \\ \text{and} \quad [-T]_\beta^\gamma &= -[T]_\beta^\gamma \end{aligned}$$

The following result reduces the computation of $T(u)$ to matrix multiplication:

**THEOREM 1.4**

$$[T(u)]_\gamma = [T]_\beta^\gamma [u]_\beta$$

PROOF.

Let $A = [T]_{\beta}^{\gamma}$, where $\beta$ is the basis $u_1, \ldots, u_n$, $\gamma$ is the basis $v_1, \ldots, v_m$, and

$$T(u_j) = \sum_{i=1}^{m} a_{ij} v_i.$$

Also let $[u]_{\beta} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

Then $u = \sum_{j=1}^{n} x_j u_j$, so

$$\begin{aligned} T(u) &= \sum_{j=1}^{n} x_j T(u_j) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} v_i \\ &= \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij} x_j \right) v_i \\ \Rightarrow [T(u)]_{\gamma} &= \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} \\ &= A[u]_{\beta} \end{aligned}$$

**DEFINITION 1.5**

*(***Composition** *of LTs)*

*If $T_1 : U \mapsto V$ and $T_2 : V \mapsto W$ are LTs, then $T_2 T_1 : U \mapsto W$ defined by*

$$(T_2 T_1)(x) = T_2(T_1(x)) \qquad \forall x \in U$$

*is a LT.*

**THEOREM 1.5**

*If $\beta$, $\gamma$ and $\delta$ are bases for $U$, $V$ and $W$, then*

$$[T_2 T_1]_{\beta}^{\delta} = [T_2]_{\gamma}^{\delta} [T_1]_{\beta}^{\gamma}$$

PROOF. Let $u \in U$. Then

$$\begin{aligned} [T_2T_1(u)]_\delta &= [T_2T_1]^\delta_\beta [u]_\beta \\ \text{and} &= [T_2(T_1(u))]_\delta \\ &= [T_2]^\delta_\gamma [T_1(u)]_\gamma \end{aligned}$$

Hence

$$[T_2T_1]^\delta_\beta [u]_\beta = [T_2]^\delta_\gamma [T_1]^\gamma_\beta [u]_\beta \tag{1}$$

(note that we can't just "cancel off" the $[u]_\beta$ to obtain the desired result!)

Finally, if $\beta$ is $u_1, \ldots, u_n$, note that $[u_j]_\beta = E_j$ (since $u_j = 0u_1 + \cdots + 0u_{j-1} + 1u_j + 0u_{j+1} + \cdots + 0u_n$) then for an appropriately sized matrix $B$,

$$BE_j = B_{*j}, \quad \text{the } j\text{th column of } B.$$

Then (1) shows that the matrices

$$[T_2T_1]^\delta_\beta \quad \text{and} \quad [T_2]^\delta_\gamma [T_1]^\gamma_\beta$$

have their first, second, ..., $n$th columns respectively equal.

**EXAMPLE 1.3**

*If $A$ is $m \times n$ and $B$ is $n \times p$, then*

$$T_A T_B = T_{AB}.$$

**DEFINITION 1.6**

*(the* **identity transformation***)*

*Let $U$ be a vector space. Then the identity transformation $I_U : U \mapsto U$ defined by*

$$I_U(x) = x \quad \forall x \in U$$

*is a linear transformation, and*

$$[I_U]^\beta_\beta = I_n \quad \text{if } n = \dim U.$$

*Also note that $I_{V_n(F)} = T_{I_n}$.*

**THEOREM 1.6**

*Let $T : U \mapsto V$ be a LT. Then*

$$I_V T = T I_U = T.$$

Then

$$T_{I_m} T_A = T_{I_m A} = T_A = T_A T_{AI_n} = T_{AI_n}$$

and consequently we have the familiar result

$$I_m A = A = AI_n.$$

**DEFINITION 1.7**

(**Invertible LTs**)

*Let $T : U \mapsto V$ be a LT.*

*If $\exists S : V \mapsto U$ such that $S$ is linear and satisfies*

$$ST = I_U \quad \text{and} \quad TS = I_V$$

*then we say that $T$ is* **invertible** *and that $S$ is an* **inverse** *of $T$.*

*Such inverses are unique and we thus denote $S$ by $T^{-1}$.*

*Explicitly,*

$$S(T(x)) = x \; \forall x \in U \quad \text{and} \quad T(S(y)) = y \; \forall y \in V$$

*There is a corresponding definition of an* **invertible matrix**: *$A \in M_{m\times n}(F)$ is called invertible if $\exists B \in M_{n\times m}(F)$ such that*

$$AB = I_m \quad \text{and} \quad BA = I_n$$

Evidently

**THEOREM 1.7**

*$T_A$ is invertible iff $A$ is invertible (i.e. if $A^{-1}$ exists). Then,*

$$(T_A)^{-1} = T_{A^{-1}}$$

**THEOREM 1.8**

*If $u_1, \ldots, u_n$ is a basis for $U$ and $v_1, \ldots, v_n$ are vectors in $V$, then there is one and only one linear transformation $T : U \to V$ satisfying*

$$T(u_1) = v_1, \ldots, T(u_n) = v_n,$$

*namely $T(x_1u_1 + \cdots + x_nu_n) = x_1v_1 + \cdots + x_nv_n$.*

(In words, a linear transformation is determined by its action on a basis.)

## 1.3 Isomorphisms

**DEFINITION 1.8**

*A linear map $T : U \mapsto V$ is called an* **isomorphism** *if $T$ is 1-1 and onto, i.e.*

1. $T(x) = T(y) \Rightarrow x = y \; \forall x, y \in U$, *and*
2. $\operatorname{Im} T = V$, *that is, if* $v \in V$, $\exists u \in U$ *such that* $T(u) = v$.

**Lemma:** A linear map $T$ is 1-1 iff $\operatorname{Ker} T = \{0\}$.

PROOF:

1. ($\Rightarrow$) Suppose $T$ is 1-1 and let $x \in \operatorname{Ker} T$.
   We have $T(x) = 0 = T(0)$, and so $x = 0$.
2. ($\Leftarrow$) Assume $\operatorname{Ker} T = \{0\}$ and $T(x) = T(y)$ for some $x, y \in U$.
   Then

$$\begin{aligned} T(x-y) &= T(x) - T(y) = 0 \\ \Rightarrow x - y &\in \operatorname{Ker} T \\ \Rightarrow x - y &= 0 \Rightarrow x = y \end{aligned}$$

**THEOREM 1.9**

*Let $A \in M_{m\times n}(F)$. Then $T_A : V_n(F) \to V_m(F)$ is*

*(a) onto:* $\Leftrightarrow \dim C(A) = m \Leftrightarrow$ *the rows of $A$ are LI;*

*(b) 1–1:* $\Leftrightarrow \dim N(A) = 0 \Leftrightarrow \operatorname{rank} A = n \Leftrightarrow$ *the columns of $A$ are LI.*

**EXAMPLE 1.4**

*Let $T_A : V_n(F) \mapsto V_n(F)$ with $A$ invertible; so $T_A(X) = AX$.*
*We will show this to be an isomorphism.*

1. *Let $X \in \operatorname{Ker} T_A$, i.e. $AX = 0$. Then*

$$\begin{aligned} A^{-1}(AX) &= A^{-1}0 \\ \Rightarrow I_n X &= 0 \\ \Rightarrow X &= 0 \\ \Rightarrow \operatorname{Ker} T &= \{0\} \\ &\Leftrightarrow T \text{ is 1-1.} \end{aligned}$$

*2. Let $Y \in V_n(F)$ : then,*

$$\begin{aligned} T(A^{-1}Y) &= A(A^{-1}Y) \\ &= I_n Y = Y \\ \text{so} \quad \operatorname{Im} T_A &= V_n(F) \end{aligned}$$

**THEOREM 1.10**

*If $T$ is an isomorphism between $U$ and $V$, then*

$$\dim U = \dim V$$

PROOF.

Let $u_1, \ldots, u_n$ be a basis for $U$. Then

$$T(u_1), \ldots, T(u_n)$$

is a basis for $V$ (i.e. $\langle u_i \rangle = U$ and $\langle T(u_i) \rangle = V$, with $u_i$, $v_i$ independent families), so

$$\dim U = n = \dim V$$

**THEOREM 1.11**

$$\Phi : \operatorname{Hom}(U, V) \mapsto M_{m \times n}(F) \quad \textit{defined by} \quad \Phi(T) = [T]_\beta^\gamma$$

*is an isomorphism.*

*Here $\dim U = n$, $\dim V = m$, and $\beta$ and $\gamma$ are bases for $U$ and $V$, respectively.*

**THEOREM 1.12**

$$\begin{aligned} & T : U \mapsto V \textit{ is invertible} \\ \Leftrightarrow \quad & T \textit{ is an isomorphism between } U \textit{ and } V. \end{aligned}$$

PROOF.

$\Rightarrow$ Assume $T$ is invertible. Then

$$\begin{aligned} T^{-1}T &= I_U \\ \text{and} \quad TT^{-1} &= I_V \\ \Rightarrow T^{-1}(T(x)) &= x \; \forall x \in U \\ \text{and } T(T^{-1}(y)) &= y \; \forall y \in V \end{aligned}$$

1. We prove $\operatorname{Ker} T = \{0\}$.
   Let $T(x) = 0$. Then
   $$T^{-1}(T(x)) = T^{-1}(0) = 0 = x$$
   So $T$ is 1-1.
2. We show $\operatorname{Im} T = V$.
   Let $y \in V$. Now $T(T^{-1}(y)) = y$, so taking $x = T^{-1}(y)$ gives
   $$T(x) = y.$$
   Hence $\operatorname{Im} T = V$.

$\Leftarrow$ Assume T is an isomorphism, and let S be the inverse map of T
$$S : V \mapsto U$$
Then $ST = I_U$ and $TS = I_V$. It remains to show that $S$ is linear.

We note that
$$x = S(y) \Leftrightarrow y = T(x)$$
And thus, using linearity of $T$ only, for any $y_1, y_2 \in V$, $x_1 = S(y_1)$, and $x_2 = S(y_2)$ we obtain
$$\begin{aligned} S(\lambda y_1 + \mu y_2) &= S(\lambda T(x_1) + \mu T(x_2)) \\ &= S(T(\lambda x_1 + \mu x_2)) \\ &= \lambda x_1 + \mu x_2 \\ &= \lambda S(y_1) + \mu S(y_2) \end{aligned}$$

**COROLLARY 1.1**
*If $A \in M_{m\times n}(F)$ is invertible, then $m = n$.*

PROOF.
Suppose $A$ is invertible. Then $T_A$ is invertible and thus an isomorphism between $V_n(F)$ and $V_m(F)$.
Hence $\dim V_n(F) = \dim V_m(F)$ and hence $m = n$.

**THEOREM 1.13**
*If $\dim U = \dim V$ and $T : U \mapsto V$ is a LT, then*
$$\begin{aligned} &T \text{ is 1-1 (injective)} \quad \Leftrightarrow \quad T \text{ is onto (surjective)} \\ &\qquad\qquad (\quad \Leftrightarrow \quad T \text{ is an isomorphism} \quad ) \end{aligned}$$

PROOF.

$\Rightarrow$ Suppose $T$ is 1-1.
Then $\operatorname{Ker} T = \{0\}$ and we have to show that $\operatorname{Im} T = V$.

$$\begin{aligned} \operatorname{rank} T + \operatorname{nullity} T &= \dim U \\ \Rightarrow \operatorname{rank} T + 0 &= \dim V \\ \text{i.e. } \dim(\operatorname{Im} T) &= \dim V \\ \Rightarrow \operatorname{Im} T &= V \quad \text{as } T \subseteq V. \end{aligned}$$

$\Leftarrow$ Suppose $T$ is onto.
Then $\operatorname{Im} T = V$ and we must show that $\operatorname{Ker} T = \{0\}$. The above argument is reversible:

$$\begin{aligned} \operatorname{Im} T &= V \\ \operatorname{rank} T &= \dim V \\ &= \dim U \\ &= \operatorname{rank} T + \operatorname{nullity} T \\ \Rightarrow \operatorname{nullity} T &= 0 \\ \text{or} \quad \operatorname{Ker} T &= \{0\} \end{aligned}$$

**COROLLARY 1.2**
*Let $A, B \in M_{n\times n}(F)$. Then*

$$AB = I_n \Rightarrow BA = I_n.$$

PROOF Suppose $AB = I_n$. Then $\operatorname{Ker} T_B = \{0\}$. For

$$\begin{aligned} BX = 0 &\Rightarrow A(BX) = A0 = 0 \\ &\Rightarrow I_n X = 0 \Rightarrow X = 0. \end{aligned}$$

But $\dim U = \dim V = n$, so $T_B$ is an isomorphism and hence invertible.
Thus $\exists C \in M_{n\times n}(F)$ such that

$$\begin{aligned} T_B T_C &= I_{V_n(F)} = T_C T_B \\ \Rightarrow BC &= I_n = CB, \end{aligned}$$

noting that $I_{V_n(F)} = T_{I_n}$.

Now, knowing $AB = I_n$,

$$\begin{aligned} \Rightarrow A(BC) &= A \\ (AB)C &= A \\ I_n C &= A \\ \Rightarrow C &= A \\ \Rightarrow BA &= I_n \end{aligned}$$

**DEFINITION 1.9**

*Another standard isomorphism: Let* $\dim V = m$, *with basis* $\gamma = v_1, \ldots, v_m$. *Then* $\phi_\gamma : V \mapsto V_m(F)$ *is the isomorphism defined by*

$$\phi_\gamma(v) = [v]_\gamma$$

**THEOREM 1.14**

$$\operatorname{rank} T = \operatorname{rank} [T]_\beta^\gamma$$

PROOF

$$\begin{array}{ccc} U & \xrightarrow{T} & V \\ \phi_\beta \downarrow & & \downarrow \phi_\gamma \\ V_n(F) & \xrightarrow{T_A} & V_m(F) \end{array}$$

With

$$\begin{array}{l} \beta : u_1, \ldots, u_n \\ \gamma : v_1, \ldots, v_m \end{array} \text{ a basis for } \begin{array}{l} U \\ V, \end{array}$$

let $A = [T]_\beta^\gamma$. Then the commutative diagram is an abbreviation for the equation

$$\phi_\gamma T = T_A \phi_\beta. \tag{2}$$

Equivalently

$$\phi_\gamma T(u) = T_A \phi_\beta(u) \; \forall u \in U$$

or

$$[T(u)]_\gamma = A[u]_\beta$$

which we saw in Theorem 1.4.

But $\operatorname{rank}(ST) = \operatorname{rank} T$ if $S$ is invertible and $\operatorname{rank}(TR) = \operatorname{rank} T$ if $R$ is invertible. Hence, since $\phi_\beta$ and $\phi_\gamma$ are both invertible,

$$(2) \;\Rightarrow\; \operatorname{rank} T = \operatorname{rank} T_A = \operatorname{rank} A$$

and the result is proven.

**Note:**

Observe that $\phi_\gamma(T(u_j)) = A_{*j}$, the $j$th column of $A$. So $\operatorname{Im} T$ is mapped under $\phi_\gamma$ into $C(A)$. Also $\operatorname{Ker} T$ is mapped by $\phi_\beta$ into $N(A)$. Consequently we get bases for $\operatorname{Im} T$ and $\operatorname{Ker} T$ from bases for $C(A)$ and $N(A)$, respectively.

$$\begin{aligned}(u \in \operatorname{Ker} T \Leftrightarrow T(u) = 0 \quad &\Leftrightarrow \quad \phi_\gamma(T(u)) = 0 \\ &\Leftrightarrow \quad T_A \phi_\beta(u) = 0 \\ &\Leftrightarrow \quad \phi_\beta(u) \in N(A).)\end{aligned}$$

**THEOREM 1.15**

*Let $\beta$ and $\gamma$ be bases for some vector space $V$. Then, with $n = \dim V$,*

$$[I_V]_\beta^\gamma$$

*is non-singular and its inverse*

$$\left\{[I_V]_\beta^\gamma\right\}^{-1} = [I_V]_\gamma^\beta.$$

PROOF

$$\begin{aligned} I_V I_V &= I_V \\ \Rightarrow [I_V I_V]_\beta^\beta &= [I_V]_\beta^\beta = I_n \\ &= [I_V]_\gamma^\beta [I_V]_\beta^\gamma. \end{aligned}$$

The matrix $P = [I_V]_\beta^\gamma = [p_{ij}]$ is called the *change of basis matrix*. For if $\beta : u_1, \ldots, u_n$ and $\gamma : v_1, \ldots, v_n$ then

$$\begin{aligned} u_j &= I_V(u_j) \\ &= p_{1j}v_1 + \cdots + p_{nj}v_n \quad \text{for } j = 1, \ldots, n. \end{aligned}$$

It is also called the *change of co-ordinate matrix*, since

$$[v]_\gamma = [I_V(v)]_\beta^\gamma [v]_\beta$$

i.e. if

$$\begin{aligned} v &= x_1u_1 + \cdots + x_nu_n \\ &= y_1v_1 + \cdots + y_nv_n \end{aligned}$$

then

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = P \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

or, more explicitly,

$$\begin{aligned} y_1 &= p_{11}x_1 + \cdots + p_{1n}x_n \\ &\vdots \\ y_n &= p_{n1}x_1 + \cdots + p_{1n}x_n. \end{aligned}$$

**THEOREM 1.16 (Effect of changing basis on matrices of LTs)**
*Let $T : V \mapsto V$ be a LT with bases $\beta$ and $\gamma$. Then*

$$[T]_\beta^\beta = P^{-1}[T]_\gamma^\gamma P$$

*where*

$$P = [I_V]_\beta^\gamma$$

*as above.*

PROOF

$$\begin{aligned} I_V T &= T = T I_V \\ \Rightarrow [I_V T]_\beta^\gamma &= [T I_V]_\beta^\gamma \\ \Rightarrow [I_V]_\beta^\gamma [T]_\beta^\beta &= [T]_\gamma^\gamma [I_V]_\beta^\gamma \end{aligned}$$

**DEFINITION 1.10**
*(***Similar matrices***)*
*If $A$ and $B$ are two matrices in $M_{m\times n}(F)$, then if there exists a non-singular matrix $P$ such that*

$$B = P^{-1}AP$$

*we say that $A$ and $B$ are* **similar** *over $F$.*

## 1.4 Change of Basis Theorem for $T_A$

In the MP274 course we are often proving results about linear transformations $T : V \mapsto V$ which state that a basis $\beta$ can be found for $V$ so that $[T]_\beta^\beta = B$, where $B$ has some special property. If we apply the result to the linear transformation $T_A : V_n(F) \mapsto V_n(F)$, the change of basis theorem applied to $T_A$ tells us that $A$ is similar to $B$. More explicitly, we have the following:

**THEOREM 1.17**

*Let $A \in M_{n\times n}(F)$ and suppose that $v_1, \ldots, v_n \in V_n(F)$ form a basis $\beta$ for $V_n(F)$. Then if $P = [v_1|\cdots|v_n]$ we have*

$$P^{-1}AP = [T_A]_\beta^\beta.$$

PROOF. Let $\gamma$ be the standard basis for $V_n(F)$ consisting of the unit vectors $E_1, \ldots, E_n$ and let $\beta : v_1, \ldots, v_n$ be a basis for $V_n(F)$. Then the change of basis theorem applied to $T = T_A$ gives

$$[T_A]_\beta^\beta = P^{-1}[T_A]_\gamma^\gamma P,$$

where $P = [I_V]_\beta^\gamma$ is the change of coordinate matrix.

Now the definition of $P$ gives

$$\begin{aligned} v_1 = I_V(v_1) &= p_{11}E_1 + \cdots + p_{n1}E_n \\ &\vdots \\ v_n = I_V(v_n) &= p_{1n}E_1 + \cdots + p_{nn}E_n, \end{aligned}$$

or, more explicitly,

$$v_1 = \begin{bmatrix} p_{11} \\ \vdots \\ p_{n1} \end{bmatrix}, \quad \ldots, \quad \begin{bmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{bmatrix}.$$

In other words, $P = [v_1|\cdots|v_n]$, the matrix whose columns are $v_1, \ldots, v_n$ respectively.

Finally, we observe that $[T_A]_\gamma^\gamma = A$.

## 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 $\mathrm{ch}_A$ and $m_A$ have degree $n$ and moreover $m_A$ divides $\mathrm{ch}_A$.

**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 $1 \times 1$ 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 $\beta$ 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 $\beta$:*

PROOF ($\mathrm{ch}_T$ is independent of the basis.)

If $\gamma$ 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* $\mathrm{ch}_A$ *, 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 $\Phi$ 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 $\mathrm{ch}_A$, 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* $\beta$ *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 $\beta$ 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* $\lambda$ *is an eigenvalue of* $A$ *with multiplicity* $m$, *(that is* $(x-\lambda)^m$ *is the exact power of* $x-\lambda$ *which divides* $\text{ch}_A$ *), 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 $\lambda$, while $m$ is referred to as the *algebraic multiplicity* of $\lambda$.

PROOF. Let $v_1, \ldots, v_r$ be a basis for $N(A - \lambda I_n)$, where $\lambda$ 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 $\text{ch}_A$, 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 $\beta_i$ is a basis for $V_i$ for $i \leq i \leq t$ and $\beta = \beta_1 \cup \cdots \cup \beta_t$, it follows that $\beta$ 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 $\beta$ 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.*

# 3 Invariant subspaces

**DEFINITIONS 3.1**

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

$$v_1 + \cdots + v_t = 0 \Rightarrow v_1 = 0, \ldots, v_t = 0 \qquad \forall v_1 \in V_1, \ldots, v_t \in V_t.$$

*We say that $V$ is the (***internal***) **direct sum** of the subspaces $V_1, \ldots, V_t$ if*

(a) *$V_1, \ldots, V_t$ are independent and*

(b) *$V = V_1 + \cdots + V_t$.*

*I.e. every element $v \in V$ is* uniquely *expressible as*

$$v = v_1 + \cdots + v_t$$

*with $v_i \in V_i$.*

*Then $V$ is isomorphic to the (external) direct sum $V_1 \oplus \cdots \oplus V_t$ under the isomorphism $v \mapsto (v_1, \ldots, v_t)$ and we write $V = V_1 \oplus \cdots \oplus V_t$.*

**THEOREM 3.1**

*If $V = V_1 \oplus \cdots \oplus V_t$ (an internal direct sum) and $\beta_1, \ldots, \beta_t$ are bases for $V_1, \ldots, V_t$ respectively, then*

$$\beta = \beta_1 \cup \cdots \cup \beta_t,$$

*the sequence formed by juxtaposing the separate bases, is a basis for $V$. Also*

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

Proof: Left as an exercise.

DEFINITION. Let $T : V \mapsto V$ be a LT and $W$ a subspace of $V$. Then if

$$w \in W \Rightarrow T(w) \in W,$$

we say $W$ is a ***T*-invariant subspace** of $V$. We can then consider the linear transformation $T_W : W \to W$ defined by

$$T_W(w) = T(w) \qquad \forall w \in W.$$

If $\beta'$ is a basis for $W$, $\{0\} \subset W \subset V$, and $\beta$ is an extension to a basis of $V$, then

$$[T]_\beta^\beta = \left[\begin{array}{c|c} [T_W]_{\beta'}^{\beta'} & B_1 \\ \hline 0 & B_2 \end{array}\right].$$

A situation of great interest is when we have $T$-invariant subspaces $W_1, \ldots, W_t$ and $V = W_1 \oplus \cdots \oplus W_t$. For if $\beta = \beta_1 \cup \cdots \cup \beta_t$, where $\beta_i$ is a basis for $W_i$, we see that

$$[T]_\beta^\beta = [T_{W_1}]_{\beta_1}^{\beta_1} \oplus \cdots \oplus [T_{W_t}]_{\beta_t}^{\beta_t}.$$

There are two important examples of $T$–invariant subspaces that arise in our study of Jordan and rational canonical forms - $\operatorname{Ker} p^t(T)$ and $T$–cyclic subspaces.

## 3.1 $T$–cyclic subspaces

**DEFINITION 3.1**

*The unique monic polynomial $f$ in $F[x]$ of least degree satisfying*

$$f(T)(v) = 0$$

*is called the* **minimum polynomial of the vector** $v \in V$ *relative to the transformation $T : V \mapsto V$ and is denoted $m_{T,v}$.*

Then $f(T)(v) = 0 \Rightarrow m_{T,v} \mid f$, so $m_{T,v} \mid m_T$. Also $m_{T,v} = 1 \Leftrightarrow v = 0$ and so if $v \neq 0$, $\deg m_{T,v} \geq 1$.

EXAMPLE. Let $T = T_A$, where $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. Also let

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Then $Av_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \neq c_0 v_1$, so $m_{T,\,v_1} \neq x - c_0$. Next $A^2 v_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so $m_{T,\,v_1} = x^2$. Also $Av_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so $m_{T,\,v_2} = x$.

**DEFINITION 3.2**

*(**$T$-cyclic subspace** generated by $v$.)*

*If $v \in V$, the set of all vectors of the form $f(T)(v)$, $f \in F[x]$, forms a subspace of $V$ called the $T$-cyclic subspace generated by $v$. It is denoted by $C_{T,v}$.*

PROOF. Exercise.

Also, $C_{T,v}$ is a $T$-invariant subspace of $V$. For

$$\begin{aligned} w \in C_{T,v} &\Rightarrow w = f(T)(v) \\ &\Rightarrow T(w) = T(f(T)(v)) = (Tf(T))(v) = ((xf)(T))(v) \in C_{T,v}. \end{aligned}$$

We see that $v = 0$ if and only if $C_{T,v} = \{0\}$.

**THEOREM 3.2**

*Let $v \neq 0,\ v \in V$. Then $C_{T,v}$ has the basis $\beta$*

$$v, T(v), T^2(v), \ldots, T^{k-1}(v)$$

*where $k = \deg m_{T,v}$. ($\beta$ is called the $T$-**cyclic basis** generated by $v$. ) Note that $\dim C_{T,v} = \deg m_{T,v}$.*

*Finally,*

$$\left[T_{C_{T,v}}\right]_\beta^\beta = C(m_{T,v}),$$

*the companion matrix of the minimum polynomial of $v$.*

PROOF.

1. The $T$-cyclic basis is a basis for $C_{T,v}$:

   Spanning:

   Let $w \in \langle v, T(v), \ldots, T^{k-1}(v)\rangle$, so

   $$\begin{aligned} w &= w_0 v + w_1 T(v) + \cdots + w_{k-1} T^{k-1}(v) \\ &= (w_0 I_V + \cdots + w_{k-1} T^{k-1})(v) \\ &= g(T)(v), \end{aligned}$$

   where $g = w_0 + \cdots + w_{k-1} x^{k-1}$, so $w \in C_{T,v}$. Hence

   $$\langle v, T(v), \ldots, T^{k-1}(v)\rangle \subseteq C_{T,v}.$$

   Conversely, suppose that $w \in C_{T,v}$ so

   $$w = f(T)(v)$$

   and

   $$f = q m_{T,v} + r$$

   where $r = a_0 + a_1 x + \cdots + a_{k-1} x^{k-1}$ and $a_0, \ldots, a_{k-1} \in F$. So

   $$\begin{aligned} f(T)(v) &= q(T) m_{T,v}(T)(v) + r(T)(v) \\ &= q(T) m_{T,v}(T)(v) + a_0 v + a_1 T(v) + \cdots + a_{k-1} T^{k-1}(v) \\ &= a_0 v + a_1 T(v) + \cdots + a_{k-1} T^{k-1}(v) \\ &\in \langle v, T(v), \ldots, T^{k-1}(v)\rangle. \end{aligned}$$

Independence:

Assume

$$a_0 v + a_1 T(v) + \cdots + a_{k-1} T^{k-1}(v) = 0,$$

where $a_0, \ldots, a_{k-1} \in F$; that is, $f(T)(v) = 0$ where

$$f = a_0 + a_1 x + \cdots + a_{k-1} x^{k-1}.$$

Hence $m_{T,v} \mid f$ and since

$$\deg f = k - 1 < k = \deg m_{T,v},$$

we have $f = 0$ and thus $a_i = 0 \ \forall i$.

2. $[T_{C_{T,v}}]_\beta^\beta = C(m_{T,v})$:

Let $L = T_{C_{T,v}}$, the restriction of $T$ to $C_{T,v}$.

We want to find $[L]_\beta^\beta$. So

$$\begin{aligned}
L(v) &= T(v) = 0v + 1T(v) + 0T^2(v) + \cdots + 0T^{k-1}(v) \\
L(T(v)) &= T^2(v) = 0v + 0T(v) + 1T^2(v) + \cdots + 0T^{k-1}(v) \\
&\vdots \\
L(T^{k-2}(v)) &= T^{k-1}(v) = 0v + 0T(v) + 0T^2(v) + \cdots + 1T^{k-1}(v)
\end{aligned}$$

Finally, to calculate $L(T^{k-1}(v)) = T^k(v)$, we let

$$m_{T,v} = a_0 + a_1 x + \cdots + a_{k-1} x^{k-1} + x^k.$$

Then $m_{T,v}(T)(v) = 0$ and hence

$$L(T^{k-1}(v)) = T^k(v) = -a_0 v - a_1 T(v) - \cdots - a_{k-1} T^{k-1}(v).$$

Hence

$$[L]_\beta^\beta = \begin{bmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & & & -a_1 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix} = C(m_{T,v}),$$

as required.

**THEOREM 3.3**

*Suppose that* $m_{T,\,v} = (x-c)^k$. *Then the vectors*

$$v,\ (T - cI_V)(v), \ldots, (T - cI_V)^{k-1}(v)$$

*form a basis* $\beta$ *for* $W = C_{T,\,v}$ *which we call the* elementary Jordan basis.

*Also*

$$[T_W]_\beta^\beta = J_k(c).$$

*More generally suppose* $m_{T,v} = p^k$, *where* $p$ *is a monic irreducible polynomial in* $F[x]$, *with* $n = \deg p$. *Then the vectors*

$$\begin{array}{cccc} v, & T(v), & \ldots, & T^{n-1}(v) \\ p(T)(v), & Tp(T)(v), & \ldots, & T^{n-1}p(T)(v) \\ \vdots & \vdots & \vdots & \vdots \\ p^{k-1}(T)(v), & Tp^{k-1}(T)(v), & \ldots, & T^{n-1}p^{k-1}(T)(v), \end{array}$$

*form a basis for* $W = C_{T,v}$, *which reduces to the elementary Jordan basis when* $p = x - c$. *Also*

$$[T_W]_\beta^\beta = H(p^k),$$

*where* $H(p^k)$ *is a* **hypercompanion** *matrix, which reduces to the elementary Jordan matrix* $J_k(c)$ *when* $p = x - c$:

$$H(p^k) = \begin{bmatrix} C(p) & 0 & \cdots & 0 \\ N & C(p) & \cdots & 0 \\ 0 & N & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & N & C(p) \end{bmatrix},$$

*where there are* $k$ *blocks on the diagonal and* $N$ *is a square matrix of same size as* $C(p)$ *which is everywhere zero, except in the top right–hand corner, where there is a* 1. *The overall effect is an unbroken subdiagonal of* $1's$.

### 3.1.1 A nice proof of the Cayley-Hamilton theorem

(From Insel, Friedberg and Spence.)

Let $f = \mathrm{ch}_T$, for some $T : V \mapsto V$. We must show that $f(T) = 0_V$—i.e. that $f(T)(v) = 0\ \forall v \in V$.

This is immediate if $v = 0$, so assume $v \neq 0$ and let $W = C_{T,v}$. Let $\beta'$ be a basis of $W$ and $\beta$ be an extension of $\beta'$ to a basis of $V$. Then

$$[T]_\beta^\beta = \left[\begin{array}{c|c} [T_W]_{\beta'}^{\beta'} & B_1 \\ \hline 0 & B_2 \end{array}\right]$$

and $\mathrm{ch}_T = \mathrm{ch}_{T_W} \cdot \mathrm{ch}_{B_2}$. So $\mathrm{ch}_{T_W} \mid \mathrm{ch}_T$ and since we know that

$$\mathrm{ch}_{T_W} = m_{T,v},$$

we have $m_{T,v} \mid \mathrm{ch}_T$.

Hence $\mathrm{ch}_T = g m_{T,v}$ and

$$\mathrm{ch}_T(T)(v) = (g(T)m_{T,v}(T))(v) = g(T)(m_{T,v}(T)(v)) = g(T)(0) = 0.$$

## 3.2 An Algorithm for Finding $m_T$

We use the factorization of $\mathrm{ch}_T$ into monic irreducibles.

**THEOREM 3.4**

*Suppose* $T : V \mapsto V$,

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

*where* $b_1, \dots, b_t \geq 1$, *and* $p_1, \dots, p_t$ *are distinct monic irreducibles.*

*Then for* $i = 1, \dots, t$ *we have*

(a)
$$V \supset \mathrm{Im}\, p_i(T) \supset \cdots \supset \mathrm{Im}\, p_i^{b_i-1}(T) \supset \mathrm{Im}\, p_i^{b_i}(T) = \cdots$$

(b)
$$\{0\} \subset \mathrm{Ker}\, p_i(T) \subset \cdots \subset \mathrm{Ker}\, p_i^{b_i-1}(T) \subset \mathrm{Ker}\, p_i^{b_i}(T) = \cdots.$$

**Note**: In terms of nullities, conclusion (b) says that

$$0 < \nu(p_i(T)) < \cdots < \nu(p_i^{b_i-1}(T)) < \nu(p_i^{b_i}(T)) = \cdots$$

so this gives us a method of calculating $b_i$.

Presently we'll show that if $\mathrm{ch}_T = p_1^{a_1} \dots p_t^{a_t}$, then

$$\mathrm{nullity}\,(p_i^{b_i}(T)) = a_i \deg p_i.$$

Hence $b_i$ is also characterised as the smallest integer $h$ such that

$$\mathrm{nullity}\,(p_i^{h}(T)) = a_i \deg p_i.$$

Also note that (a) and (b) are equivalent, and it is the latter that we prove.

A notational simplification—**the left $F[x]$-module notation**.

If $f \in F[x]$ and $v \in V$, we define

$$fv = f(T)(v).$$

It is easy to verify that

1.
$$(f+g)v = fv + gv \qquad \forall f, g \in F[x], v \in V;$$

2.
$$f(v+w) = fv + fw \qquad \forall f \in F[x], v, w \in V;$$

3.
$$(fg)v = f(gv) \qquad \forall f, g \in F[x], v \in V;$$

4.
$$1v = v \qquad \forall v \in V.$$

These axioms, together with the four axioms for addition on $V$, turn $V$ into what is called a "left $F[x]$-module". (So there are deeper considerations lurking in the background—ideas of greater generality which make the algorithm we unravel for the rational canonical form also apply to other things such as the theorem that any finite abelian group is a direct product of cyclic prime power subgroups.)

(i) We first prove that $\{0\} \subset \operatorname{Ker} p_i(T)$. We write $p = p_i$, $b = b_i$ for brevity; no confusion should arise since $i$ is fixed.

PROOF. $m_T = pf$, $f \in F[x]$ and $f(T) \neq 0_V$. Hence $\exists v \in V$ such that $fv \neq 0$. Then
$$p(fv) = (pf)v = m_T v = 0,$$
so $fv \in \operatorname{Ker} p(T)$.

(ii) We next prove that
$$\operatorname{Ker} p^b(T) = \operatorname{Ker} p^{b+1}(T).$$

The containment
$$\operatorname{Ker} p^b(T) \subseteq \operatorname{Ker} p^{b+1}(T)$$
is obvious, so we need only show that
$$\operatorname{Ker} p^b(T) \supseteq \operatorname{Ker} p^{b+1}(T).$$

Let $w \in \operatorname{Ker} p^{b+1}(T)$, i.e. $p^{b+1}w = 0$. Now if $m_T = p^b q$, then $\gcd(p^b, q) = 1$. So $\exists u, v \in F[x]$ such that $1 = up^b + vq$. Hence
$$p^b = up^{2b} + vm_T.$$

Hence

$$\begin{aligned} p^b(T) &= (up^{2b})(T) + v(T)m_T(T) \\ &= (up^{2b})(T) \end{aligned}$$

and thus

$$p^b w = (p^{b-1})(p^{b+1}w) = p^{b-1}0 = 0$$

and $w \in \operatorname{Ker} p^b$, as required.

(iii)

$$\begin{aligned} &\operatorname{Ker} p^h(T) = \operatorname{Ker} p^{h+1}(T) \Rightarrow \operatorname{Ker} p^{h+1}(T) = \operatorname{Ker} p^{h+2}(T), \\ \text{i.e.} \quad &\operatorname{Ker} p^h(T) \supseteq \operatorname{Ker} p^{h+1}(T) \Rightarrow \operatorname{Ker} p^{h+1}(T) \supseteq \operatorname{Ker} p^{h+2}(T). \end{aligned}$$

PROOF. Suppose that $\operatorname{Ker} p^h(T) \supseteq \operatorname{Ker} p^{h+1}(T)$. Then

$$\begin{aligned} v \in \operatorname{Ker} p^{h+2}(T) &\Rightarrow p^{h+2}v = 0 \\ &\Rightarrow p^{h+1}(pv) = 0 \Rightarrow pv \in \operatorname{Ker} p^{h+1}(T) \\ &\Rightarrow pv \in \operatorname{Ker} p^h(T) \Rightarrow p^h(pv) = 0 \\ &\Rightarrow p^{h+1}v = 0 \Rightarrow v \in \operatorname{Ker} p^{h+1}(T). \end{aligned}$$

So it follows by induction from (ii) that

$$\operatorname{Ker} p_i^{b_i}(T) = \operatorname{Ker} p_i^{b_{i+1}}(T) = \cdots.$$

(iv)

$$\operatorname{Ker} p^{b-1}(T) \subset \operatorname{Ker} p^b(T)$$

and this forces a chain of proper inclusions:

$$\{0\} \subset \operatorname{Ker} p(T) \subset \cdots \subset \operatorname{Ker} p^{b-1}(T) \subset \operatorname{Ker} p^b(T) = \cdots$$

which is the desired result. For
$p^{b-1}q(T) \neq 0_V$, so $\exists v$ such that $p^{b-1}qv \neq 0$. Then

$$qv \notin \operatorname{Ker} p^{b-1}(T),$$

but $qv \in \operatorname{Ker} p^b(T)$ as

$$p^b qv = m_T v = 0.$$

## 3.3 Primary Decomposition Theorem

**THEOREM 3.5 (Primary Decomposition)**
*If $T : V \mapsto V$ is a LT with $m_T = p_1^{b_1} \dots p_t^{b_t}$, where $p_1, \dots, p_t$ are monic irreducibles, then*

$$V = \operatorname{Ker} p_1^{b_1}(T) \oplus \cdots \oplus \operatorname{Ker} p_t^{b_t}(T),$$

*a direct sum of $T$-invariant subspaces. Moreover for $1 \le i \le t$,*

$$\nu(p_i(T)^{b_i}) = a_i \deg p_i,$$

*where $ch_T = p_1^{a_1} \cdots p_t^{a_t}$.*

REMARK. The same proof gives a slightly more general result:
If $p = p_1^{b_1} \cdots p_t^{b_t}$, then

$$\operatorname{Ker} p(T) = \operatorname{Ker} p_1^{b_1}(T) \oplus \cdots \operatorname{Ker} p_t^{b_t}(T).$$

We subsequently give an application of the decomposition theorem in this form to the solution of the $n$–th order linear differential equations with constant coefficients. (See Hoffman and Kunze, pages 184–185.)
PROOF. Let $m_T = p_i^{b_i} q_i \ \forall i = 1, \dots, t$. Then

$$(q_i q_j)(T) = 0_V \quad \text{if } i \neq j \text{ as} \quad m_T \mid q_i q_j \quad \text{if } i \neq j.$$

Now $\gcd(q_1, \dots, q_t) = 1$, so $\exists f_1, \dots, f_t \in F[x]$ such that

$$1 = f_1 q_1 + \cdots + f_t q_t$$

and with $T_i = (f_i q_i)(T)$ we have

$$I_V = T_1 + \cdots + T_t. \tag{6}$$

Also

$$\begin{aligned} T_i T_j &= (f_i q_i)(T)(f_j q_j)(T) \\ &= (f_i f_j)(T)(q_i q_j)(T) \\ &= 0_V \qquad \text{if } i \neq j, \end{aligned}$$

Then

$$V = \bigoplus_{i=1}^{t} \operatorname{Im} T_i.$$

For $T_i^2 = T_i(T_1 + \cdots + T_t) = T_i I_V = T_i$. Next, $V = \operatorname{Im} T_1 + \cdots + \operatorname{Im} T_t$. For

$$v \in V \Rightarrow v = I_V(v) = T_1(v) + \cdots + T_t(v) \in \operatorname{Im} T_1 + \cdots + \operatorname{Im} T_t.$$

Next assume $v_1 + \cdots + v_t = 0,\ v_i \in \operatorname{Im} T_i,\ 1 \leq i \leq t$. Then $v_i = T_i(u_i)$ and

$$\begin{aligned} T_1(u_1) + \cdots + T_t(u_t) &= 0 \\ T_i(T_1(u_1) + \cdots + T_t(u_t)) &= T(0) = 0 \\ T_i T_i(u_i) &= 0 \\ v_i = T_i(u_i) &= 0. \end{aligned}$$

We now show that

$$\operatorname{Im} T_i = \operatorname{Ker} p_i^{b_i}(T).$$

"$\subseteq$" Let $v \in \operatorname{Im} T_i$. Then

$$\begin{aligned} v &= f_i q_i w \\ \Rightarrow p_i^{b_i} v &= p_i^{b_i} f_i q_i w \\ &= f_i (p_i^{b_i} q_i) w \\ &= 0. \end{aligned}$$

"$\supseteq$" Suppose $p_i^{b_i} v = 0$.

Now if $j \neq i$, we have $p_i^{b_i} \mid f_j q_j$, so

$$T_j(v) = f_j q_j v = 0.$$

So

$$\begin{aligned} v &= I_V(v) = T_1(v) + \cdots + T_t(v) = T_i(v) \\ &\in \operatorname{Im} T_i, \end{aligned}$$

as required.

Finally, let $V_i = \operatorname{Ker} p_i^{b_i}(T)$ and $L_i = T_{V_i}$. Then because $V_1, \ldots, V_t$ are $T$–invariant subspaces of $V$, we have

$$ch_T = ch_{L_1} \cdots ch_{L_t}.$$

Now $p_i^{b_i}(T)(v) = 0$ if $v \in V_i$, so $p_i^{b_i}(L_i) = 0_{V_i}$. Hence $m_{L_i}$ has the form $m_{L_i} = p_i^{e_i}$. Hence $ch_{L_i}$ has the form $ch_{L_i} = p_i^{d_i}$. Hence

$$ch_T = p_1^{a_1} \cdots p_t^{a_t} = p_1^{d_1} \cdots p_t^{d_t}$$

and consequently $d_i = a_i$.

Finally,

$$\dim V_i = \deg ch_{L_i} = \deg p_i^{a_i} = a_i \deg p_i.$$

(Incidentally, we mention that $m_T = \text{lcm}\,(m_{L_1}, \ldots, m_{L_t})$. Hence

$$m_T = p_1^{b_1} \cdots p_t^{b_t} = p_1^{e_1} \cdots p_t^{e_t}$$

and consequently $e_i = b_i$. Hence $m_{L_i} = p_i^{b_i}$.)

**THEOREM 3.6 (Commuting diagonable linear transformations)**

*If $T_1, \ldots, T_n : V \to V$ are commuting diagonable linear transformations, then there exists a basis $\beta$ for $V$ such that each of $[T_1]_\beta^\beta, \ldots [T_m]_\beta^\beta$ are each diagonal.*

*(Matrix version) If $A_1, \ldots, A_m$ are commuting diagonable matrices of the same size, then there exists a non-singular matrix $P$ such that the matrices*

$$P^{-1}A_1P, \ldots P^{-1}A_mP$$

*are simultaneously diagonal.*

PROOF. (From Samelson page 158). We prove the result when $m = 2$, the general case follows by an easy iteration. Suppose $T_1$ and $T_2$ are commuting diagonable linear transformations on $V$. Because $m_{T_1}$ splits as a product of distinct linear factors, the primary decomposition theorem gives a direct sum decomposition as a sum of the $T_1$–eigenspaces:

$$V = U_1 \oplus \cdots \oplus U_t.$$

It turns out that not only are the subspaces $U_i$ $T_1$–invariant, they are $T_2$–invariant. For if $U_i = \text{Ker}\,(T_1 - cI_V)$, then

$$\begin{aligned} v \in U_i &\Rightarrow T_1(v) = cv \\ &\Rightarrow T_2(T_1(v)) = cT_2(v) \\ &\Rightarrow T_1(T_2(v)) = cT_2(v) \\ &\Rightarrow T_2(v) \in U_i. \end{aligned}$$

Now because $T_2$ is diagonable, $V$ has a basis consisting of $T_2$–eigenvectors and it is an easy exercise to show that in a direct sum of $T_2$–invariant subspaces, each non-zero "component" of a $T_2$–eigenvector is itself a $T_2$–eigenvector; moreover each non–zero component is a $T_1$– eigenvector. Hence $V$ is spanned by a family of vectors which are simultaneously $T_1$–eigenvectors and $T_2$–eigenvectors. If $\beta$ is a subfamily which forms a basis for $V$, then $[T_1]_\beta^\beta$ and $[T_2]_\beta^\beta$ are diagonal.

**THEOREM 3.7 (Fitting's lemma)**

*Suppose $T : V \to V$ is a linear transformation over $T$ and*

$$\text{Ker} \subset \text{Ker}\, T^2 \subset \cdots \text{Ker}\, T^n = \text{Ker}\, T^{n+1} = \cdots$$

*Then $V = \text{Im}\, T^n \oplus \text{Ker}\, T^n$.*

**COROLLARY 3.1**

*If $T : V \to V$ is an* indecomposable *linear transformation (that is the only $T$–invariant subspaces of $V$ are $\{0\}$ and $V$), then $T$ is either* nilpotent *(that is $T^n = 0_V$ for some $n \geq 1$) or $T$ is an isomorphism.*

# 4 The Jordan Canonical Form

The following subspaces are central for our treatment of the Jordan and rational canonical forms of a linear transformation $T : V \to V$.

**DEFINITION 4.1**

*With $m_T = p_1^{b_1} \dots p_t^{b_t}$ as before and $p = p_i$, $b = b_i$ for brevity, we define*

$$N_{h,p} = \operatorname{Im} p^{h-1}(T) \cap \operatorname{Ker} p(T).$$

REMARK. In numerical examples, we will need to find a spanning family for $N_{h,\,p}$. This is provided by Problem Sheet 1, Question 11(a): we saw that if $T : U \to V$ and $S : V \to W$ are linear transformations, then If $\operatorname{Ker} p^h(T) = \langle u_1, \dots, u_n \rangle$, then

$$N_{h,\,p} = \langle p^{h-1}u_1, \dots, p^{h-1}u_n \rangle,$$

where we have taken $U = V = W$ and replaced $S$ and $T$ by $p(T)$ and $p^{h-1}(T)$ respectively, so that $ST = p^h(T)$. Also

$$\dim(\operatorname{Im} T \cap \operatorname{Ker} S) = \nu(ST) - \nu(T).$$

Hence

$$\begin{aligned} \nu_{h,p} &= \dim N_{h,p} \\ &= \dim(\operatorname{Im} p^{h-1}(T) \cap \operatorname{Ker} p(T)) \\ &= \nu(p^h(T)) - \nu(p^{h-1}(T)). \end{aligned}$$

**THEOREM 4.1**

$$N_{1,p} \supseteq N_{2,p} \supseteq \cdots \supseteq N_{b,p} \neq \{0\} = N_{b+1,p} = \cdots.$$

PROOF. Successive containment follows from

$$\operatorname{Im} L^{h-1} \supseteq \operatorname{Im} L^h$$

with $L = p(T)$.

The fact that $N_{b,p} \neq \{0\}$ and that $N_{b+1,p} = \{0\}$ follows directly from the formula

$$\dim N_{h,p} = \nu(p^h(T)) - \nu(p^{h-1}(T)).$$

For simplicity, assume that $p$ is linear, that is that $p = x - c$. The general story (when $\deg p > 1$) is similar, but more complicated; it is delayed until the next section.

Telescopic cancellation then gives

**THEOREM 4.2**

$$\nu_{1,p} + \nu_{2,p} + \cdots + \nu_{b,p} = \nu(p^b(T)) = a,$$

*where $p^a$ is the exact power of $p$ dividing $ch_T$.*

Consequently we have the decreasing sequence

$$\nu_{1,p} \geq \nu_{2,p} \geq \cdots \geq \nu_{b,p} \geq 1.$$

**EXAMPLE 4.1**

*Suppose $T : V \mapsto V$ is a LT such that $p^4 || m_T$, $p = x - c$ and*

$$\begin{array}{ll} \nu(p(T)) = 3, & \nu(p^2(T)) = 6, \\ \nu(p^3(T)) = 8, & \nu(p^4(T)) = 10. \end{array}$$

*So*

$$\mathrm{Ker}\, p(T) \subset \mathrm{Ker}\, p^2(T) \subset \mathrm{Ker}\, p^3(T) \subset \mathrm{Ker}\, p^4(T) = \mathrm{Ker}\, p^5(T) = \cdots.$$

*Then*

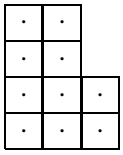
$$\begin{array}{ll} \nu_{1,p} = 3, & \nu_{2,p} = 6 - 3 = 3, \\ \nu_{3,p} = 8 - 6 = 2, & \nu_{4,p} = 10 - 8 = 2 \end{array}$$

*so*

$$N_{1,p} = N_{2,p} \supset N_{3,p} = N_{4,p} \neq \{0\}.$$

## 4.1 The Matthews' dot diagram

We would represent the previous example as follows:



$\nu_{4,p}$ Dots represent dimension:

$\nu_{3,p}$

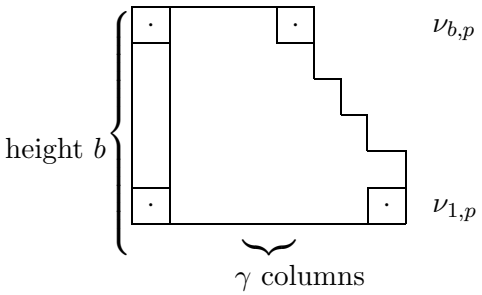
$\nu_{2,p}$ $3 + 3 + 2 + 2 = 10$

$\nu_{1,p}$ $\Rightarrow 10 = 4 + 4 + 2$

The conjugate partition of 10 is 4+4+2 (sum of column heights of diagram), and this will soon tell us that there is a corresponding contribution to the **Jordan canonical form** of this transformation, namely

$$J_4(c) \oplus J_4(c) \oplus J_2(c).$$

In general,




and we label the conjugate partition by

$$e_1 \geq e_2 \geq \cdots \geq e_\gamma.$$

Finally, note that the total number of dots in the dot diagram is $\nu(p^b(T))$, by Theorem 4.2.

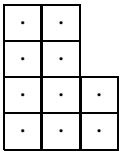
**THEOREM 4.3**

*$\exists v_1, \ldots, v_\gamma \in V$ such that*

$$p^{e_1-1}v_1, p^{e_2-1}v_2, \ldots, p^{e_\gamma-1}v_\gamma$$

*form a basis for* $\operatorname{Ker} p(T)$.

PROOF. Special case, but the construction is quite general.



choose a basis $p^3v_1, p^3v_2$ for $N_{4,p}$

extend to a basis $p^3v_1, p^3v_2, pv_3$ for $N_{2,p}$

Then $p^3v_1$, $p^3v_2$, $pv_3$ is a basis for $N_{1,p} = \operatorname{Ker} p(T)$.

**THEOREM 4.4 (Secondary decomposition)**

(i)
$$m_{T,v_i} = p^{e_i}$$

(ii)
$$\operatorname{Ker} p^b(T) = C_{T,v_1} \oplus \cdots \oplus C_{T,v_\gamma}$$

PROOF.

(i) We have $p^{e_i-1}v_i \in \operatorname{Ker} p(T)$, so $p^{e_i}v_i = 0$ and hence $m_{T,v_i} \mid p^{e_i}$. Hence $m_{T,v_i} = p^f$, where $0 \leq f \leq e_i$.

But $p^{e_i-1}v_i \neq 0$, as it is part of a basis. Hence $f \geq e_i$ and $f = e_i$ as required.

(ii) (a)

$$C_{T,v_i} \subseteq \operatorname{Ker} p^b(T).$$

For $p^{e_i}v_i = 0$ and so $p^{e_i}(fv_i) = 0 \quad \forall f \in F[x]$. Hence as $e_i \leq b$, we have

$$p^b(fv_i) = p^{b-e_i}(p^{e_i}fv_i) = p^{b-e_i}0 = 0$$

and $fv_i \in \operatorname{Ker} p^b(T)$. Consequently $C_{T,v_i} \subseteq \operatorname{Ker} p^b(T)$ and hence

$$C_{T,v_1} + \cdots + C_{T,v_\gamma} \subseteq \operatorname{Ker} p^b(T).$$

(b) We presently show that the subspaces $C_{T,v_j}$, $j = 1, \ldots, \gamma$ are independent, so

$$\begin{aligned} \dim(C_{T,v_1} + \cdots + C_{T,v_\gamma}) &= \sum_{j=1}^{\gamma} \dim C_{T,v_j} \\ &= \sum_{j=1}^{\gamma} \deg m_{T,v_j} = \sum_{j=1}^{\gamma} e_j \\ &= \nu(p^b(T)) \\ &= \dim \operatorname{Ker} p^b(T). \end{aligned}$$

Hence

$$\begin{aligned} \operatorname{Ker} p^b(T) &= C_{T,v_1} + \cdots + C_{T,v_\gamma} \\ &= C_{T,v_1} \oplus \cdots \oplus C_{T,v_\gamma}. \end{aligned}$$

The independence of the $C_{T,v_i}$ is stated as a lemma:
**Lemma**: Let $v_1, \ldots, v_\gamma \in V$, $e_1 \geq \cdots \geq e_\gamma \geq 1$;
$m_{T,v_j} = p^{e_j} \quad 1 \leq j \leq \gamma$; $p = x - c$;
Also $p^{e_1-1}v_1, \ldots, p^{e_\gamma-1}v_\gamma$ are LI. Then

$$\begin{aligned} & f_1v_1 + \cdots + f_\gamma v_\gamma = 0\,; \quad f_1, \ldots, f_\gamma \in F[x] \\ \Rightarrow \quad & p^{e_j} \mid f_j \quad 1 \leq j \leq \gamma. \end{aligned}$$

PROOF: (induction on $e_1$)

Firstly, consider $e_1 = 1$. Then

$$e_1 = e_2 = \cdots = e_\gamma = 1.$$

Now $m_{T,v_j} = p^{e_j}$ and

$$p^{e_1-1}v_1, \ldots, p^{e_\gamma-1}v_\gamma$$

are LI, so $v_1, \ldots, v_\gamma$ are LI. So assume

$$f_1v_1 + \cdots + f_\gamma v_\gamma = 0 \qquad f_1, \ldots, f_\gamma \in F[x]. \tag{7}$$

and by the remainder theorem

$$f_j = (x - c)q_j + f_j(c). \tag{8}$$

Thus

$$\begin{aligned} f_jv_j &= q_j(x-c)v_j + f_j(c)v_j \\ &= f_j(c)v_j. \end{aligned}$$

So (7) implies

$$\begin{aligned} f_1(c)v_1 + \cdots + f_\gamma(c)v_\gamma(c) &= 0 \\ \Rightarrow f_j(c) &= 0 \qquad \forall j = 1, \ldots, \gamma \end{aligned}$$

and (8) implies

$$(x - c) \mid f_j \qquad \forall j$$

which is the result.
Now let $e_1 > 1$ and assume the lemma is true for $e_1 - 1$. If

$$\begin{gathered} m_{T,v_j} = p^{e_j}; \\ p^{e_1-1}v_1, \ldots, p^{e_\gamma-1}v_\gamma \text{ are LI,} \\ \text{and} \qquad f_1v_1 + \cdots + f_\gamma v_\gamma = 0 \end{gathered} \tag{9}$$

as before, we have

$$f_1(pv_1) + \cdots + f_\gamma(pv_\gamma) = 0 \tag{10}$$

where $m_{T,pv_j} = p^{e_j-1}$.
Now let $\delta$ be the greatest positive integer such that $e_\delta > 1$; i.e. $e_{\delta+1} = 1$, but $e_\delta > 1$. Applying the induction hypothesis to (10), in the form

$$f_1(pv_1) + \cdots + f_\delta(pv_\delta) = 0$$

we obtain

$$p^{e_j-1} \mid f_j \qquad \forall j = 1, \ldots, \delta,$$

so we may write

$$f_j = p^{e_j-1} g_j,$$

(where if $g_j = f_j$ if $j > \delta$). Now substituting in (9),

$$g_1 p^{e_1-1} v_1 + \cdots + g_\gamma p^{e_\gamma - 1} v_\gamma = 0. \tag{11}$$

But

$$m_{T, p^{e_j-1} v_j} = p$$

so (11) and the case $e_1 = 1$ give

$$p \mid g_j \qquad \forall j,$$

as required.

**A summary**:

If $m_T = (x - c_1)^{b_1} \ldots (x - c_t)^{b_t} = p_1^{b_1} \ldots p_t^{b_t}$, then there exist vectors $v_{ij}$ and positive integers $e_{ij}$ $(1 \leq i \leq t,\ 1 \leq j \leq \gamma_i)$, where $\gamma_i = \nu(T - c_i I_V)$, satisfying

$$b_i = e_{i1} \geq \cdots \geq e_{i\gamma_i}, \quad m_{T, v_{ij}} = p_i^{e_{ij}}$$

and

$$V = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C_{T, v_{ij}}.$$

We choose the elementary Jordan bases

$$\beta_{ij} : v_{ij}, (T - c_i I_V)(v_{ij}), \ldots, (T - c_i I_V)^{e_{ij}-1}(v_{ij})$$

for $C_{T, v_{ij}}$. Then if

$$\beta = \bigcup_{i=1}^{t} \bigcup_{j=1}^{\gamma_i} \beta_{ij},$$

$\beta$ is a basis for $V$ and we have

$$[T]_\beta^\beta = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}(c_i) = J.$$

A direct sum of elementary Jordan matrices such as $J$ is called a Jordan canonical form of $T$.

If $T = T_A$ and $P = [v_{11} | \ldots\ldots | v_{t\gamma_t}]$, then

$$P^{-1} A P = J$$

and $J$ is called a Jordan canonical form of $A$.

## 4.2 Two Jordan Canonical Form Examples

### 4.2.1 Example (a):

Let $A = \begin{bmatrix} 4 & 0 & 1 & 0 \\ 2 & 2 & 3 & 0 \\ -1 & 0 & 2 & 0 \\ 4 & 0 & 1 & 2 \end{bmatrix} \in M_{4\times 4}(\mathbb{Q})$.

We find $ch_T = (x-2)^2(x-3)^2 = p_1^2 p_2^2$, where $p_1 = x-2,\ p_2 = x-3$.

**CASE 1,** $p_1 = x-2$:

$$p_1(A) = A - 2I_4 = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 2 & 0 & 3 & 0 \\ -1 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

so $\nu(p_1(A)) = \gamma_1 = 2$. Hence $b_1 = 1$ and the corresponding dot diagram has height 1, width 2, with associated Jordan blocks $J_1(2) \oplus J_1(2)$:

$N_{1,\,x-2}$

We find $v_{11} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $v_{12} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ form a basis for $\operatorname{Ker} p_1(T_A) = N(A - 2I_4)$ and $m_{T_A,\,v_{11}} = m_{T_A,\,v_{12}} = x-2$. Also

$$\operatorname{Ker}\,(p_1^{b_1}(T_A)) = N(p_1(A)) = N(A-2I_4) = C_{T_A,\,v_{11}} \oplus C_{T_A,\,v_{12}}.$$

Note that $C_{T_A,\,v_{11}}$ and $C_{T_A,\,v_{12}}$ have Jordan bases $\beta_{11} : v_{11}$ and $\beta_{12} : v_{12}$ respectively.

**CASE 2,** $p_2 = x-3$:

$$p_2(A) = A - 3I_4 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 2 & -1 & 3 & 0 \\ -1 & 0 & -1 & 0 \\ 4 & 0 & 1 & -1 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

so $\nu(p_2(A)) = 1 = \gamma_2$; also $\nu(p_2^2(A)) = 2$. Hence $b_2 = 2$ and we get a corresponding dot diagram consisting of two vertical dots, with associated Jordan block $J_2(3)$:

$N_{2,\,x-3}$
$N_{1,\,x-3}$

We have to find a basis of the form $p_2(T_A)(v_{21}) = (A-3I_4)v_{21}$ for $\operatorname{Ker} p_2(T_A) = N(A-3I_4)$.

To find $v_{21}$ we first get a basis for $N(A-3I_4)^2$. We have

$$p_2^2(A) = (A-3I_4)^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -3 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 1 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & -10 & -3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and we find $X_1 = \begin{bmatrix} 2 \\ 10 \\ 1 \\ 0 \end{bmatrix}$ and $X_2 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix}$ is such a basis. Then we have

$$\begin{aligned} N_{2,\,p_2} &= \langle p_2X_1,\, p_2X_2\rangle \\ &= \langle p_2(A)X_1,\, p_2(A)X_2\rangle = \langle (A-3I_4)X_1,\, (A-3I_4)X_2\rangle \\ &= \left\langle \begin{bmatrix} 3 \\ -3 \\ -3 \\ 9 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 3 \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} 3 \\ -3 \\ -3 \\ 9 \end{bmatrix} \right\rangle. \end{aligned}$$

Hence we can take $v_{21} = X_1$. Then $m_{T_A,\,v_{21}} = (x-3)^2$. Also

$$\operatorname{Ker} p_2^{b_2}(T_A)) = N(p_2^2(A)) = N(A-3I_4)^2 = C_{T_A,\,v_{21}}.$$

Moreover $C_{T_A,\,v_{21}}$ has Jordan basis $\beta_{21} : v_{21},\ (A-3I_4)v_{21}$.

Finally we have $V_4(\mathbb{Q}) = C_{T_A,\,v_{11}} \oplus C_{T_A,\,v_{12}} \oplus C_{T_A,\,v_{21}}$ and $\beta = \beta_{11} \cup \beta_{12} \cup \beta_{21}$ is a basis for $V_4(\mathbb{Q})$. Then with

$$P = [v_{11}|v_{12}|v_{21}|(A-3I_4)v_{21}] = \begin{bmatrix} 0 & 0 & 2 & 3 \\ 1 & 0 & 10 & -3 \\ 0 & 0 & 1 & -3 \\ 0 & 1 & 0 & 9 \end{bmatrix}$$

we have

$$P^{-1}AP = [T_A]_\beta^\beta = J_1(2) \oplus J_1(2) \oplus J_2(3) = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}.$$

### 4.2.2 Example (b):

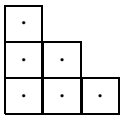
Let $A \in M_{6\times 6}(F)$ have the property that $ch_A = x^6,\ m_A = x^3$ and

$$\nu(A) = 3,\ \nu(A^2) = 5,\ (\nu(A^3) = 6).$$

Next, with $\nu_{h,\,x} = \dim_F N_{h,\,x}$ we have

$$\begin{aligned} \nu_{1,\,x} &= \nu(A) = 3 = \gamma_1; \\ \nu_{2,\,x} &= \nu(A^2) - \nu(A) = 5 - 3 = 2; \\ \nu_{3,\,x} &= \nu(A^3) - \nu(A^2) = 6 - 5 = 1. \end{aligned}$$

Hence the dot diagram corresponding to the (only) monic irreducible factor $x$ of $m_A$ is




Hence we read off that $\exists$ a non-singular $P \in M_{6\times 6}(F)$ such that $P^{-1}AP = J_3(0) \oplus J_2(0) \oplus J_1(0)$. To find such a matrix $P$ we proceed as follows:

(i) First find a basis for $N_{3,\,x}$. We do this by first finding a basis for $N(A^3)$: $X_1,\ X_2,\ X_3,\ X_4,\ X_5,\ X_6$. Then

$$N_{3,\,x} = \langle A^2X_1,\ A^2X_2,\ A^2X_3,\ A^2X_4,\ A^2X_5,\ A^2X_6\rangle.$$

We now apply the LRA (left–to–right algorithm) to the above spanning family to get a basis $A^2v_{11}$ for $N_{3,\,x}$, where $A^2v_{11}$ is the first non–zero vector in the spanning family.

(ii) Now extend the linearly independent family $A^2v_{11}$ to a basis for $N_{2,\,x}$. We do this by first finding a basis $Y_1,\ Y_2,\ Y_3,\ Y_4,\ Y_5$ for $N(A^2)$. Then

$$N_{2,\,x} = \langle AY_1,\ AY_2,\ AY_3,\ AY_4,\ AY_5\rangle.$$

We now attach $A^2v_{11}$ to the head of this spanning family:

$$N_{2,\,x} = \langle A^2v_{11},\ AY_1,\ AY_2,\ AY_3,\ AY_4,\ AY_5\rangle$$

and apply the LRA to find a basis for $N_{2,\,x}$ which includes $A^2X_1$. This will have the form $A^2v_{11},\ Av_{12}$, where $Av_{12}$ is the first vector in the list $AY_1, \ldots, AY_5$ which is not a linear combination of $A^2v_{11}$.

(iii) Now extend the linearly independent family $A^2v_{11},\ Av_{12}$ to a basis for $N_{1,\,x} = N(A)$. We do this by first finding a basis $Z_1,\ Z_2,\ Z_3$ for $N(A)$.

Then place the linearly independent family $A^2v_{11}$, $Av_{12}$ at the head of this spanning family:

$$N_{1,x} = \langle A^2v_{11},\, Av_{12},\, Z_1,\, Z_2,\, Z_3\rangle.$$

The LRA is then applies to the above spanning family selects a basis of the form $A^2v_{11}$, $Av_{12}$, $v_{13}$, where $v_{13}$ is the first vector among $Z_1$, $Z_2$, $Z_3$ which is not a linear combination of $A^2v_{11}$ and $Av_{12}$.

Then $m_{T_A,\,v_{11}} = x^3,\ m_{T_A,\,v_{12}} = x^2,\ m_{T_A,\,v_{13}} = x$. Also

$$\operatorname{Ker} p_1^{b_1}(T_A) = N(A^3) = C_{T_A,\,v_{11}} \oplus C_{T_A,\,v_{12}} \oplus C_{T_A,\,v_{13}}.$$

Finally, if we take Jordan bases

$$\begin{array}{rcl} \beta_{11} & : & v_{11},\, Av_{11},\, A^2v_{11}; \\ \beta_{12} & : & v_{12},\, Av_{12}; \\ \beta_{13} & : & v_{13} \end{array}$$

for the three T–cyclic subspaces $C_{T_A,\,v_{11}}$, $C_{T_A,\,v_{12}}$, $C_{T_A,\,v_{13}}$, respectively, we then get the basis

$$\begin{array}{rcl} \beta & = & \beta_{11} \cup \beta_{12} \cup \beta_{13} \\ & = & v_{11},\, Av_{11},\, A^2v_{11};\, v_{12},\, Av_{12};\, v_{13} \end{array}$$

for $V_6(F)$. Then if

$$P = [v_{11}|Av_{11}|A^2v_{11}|v_{12}|Av_{12}|v_{13}]$$

we have

$$\begin{array}{rcl} P^{-1}AP = [T_A]_\beta^\beta & = & J_3(0) \oplus J_2(0) \oplus J_1(0) \\ & = & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \end{array}$$

## 4.3 Uniqueness of the Jordan form

Let $\beta$ be a basis for $V$ for which $[T]_\beta^\beta$ is in Jordan canonical form

$$J = J_{e_1}(\lambda_1) \oplus \cdots J_{e_s}(\lambda_s).$$

If we change the order of the basis vectors in $\beta$, we produce a corresponding change in the order of the elementary Jordan matrices. It is customary to assume our Jordan forms arranged so as to group together into a block those elementary Jordan matrices having the same eigenvalue $c_i$:

$$J = J_1 \oplus \cdots \oplus J_t,$$

where

$$J_i = \bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}(c_i).$$

Moreover within this $i$–th block $J_i$, we assume the sizes $e_{i1}, \ldots, e_{i\gamma_i}$ of the elementary Jordan matrices decrease monotonically:

$$e_{i1} \geq \ldots \geq e_{i\gamma_i}.$$

We prove that with this convention, the above sequence is uniquely determined by $T$ and the eigenvalue $c_i$.

We next observe that

$$ch_T = ch_J = \prod_{i=1}^{t} ch_{J_i} = \prod_{i=1}^{t}\prod_{j=1}^{\gamma_i}(x-c_i)^{e_{ij}} = \prod_{i=1}^{t}(x-c_i)^{e_{i1}+\cdots+e_{i\gamma_i}}.$$

Hence $c_1, \ldots, c_t$ are determined as the distinct eigenvalues of $T$.

**DEFINITION 4.2**

*The numbers* $e_{i1}, \ldots, e_{i\gamma_i}$, $1 \leq i \leq t$, *are called the* Segre *characteristic of* $T$, *while the numbers* $\nu_{1,x-c_i}, \ldots, \nu_{b_i,x-c_i}$, $1 \leq i \leq t$ *are called the* Weyr *characteristic of* $T$.

*The polynomials* $(x-c_i)^{e_{ij}}$ *are called the* elementary divisors *of* $T$.

**LEMMA 4.1**

*Let*

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

*Then*

$$\nu(A^h) = \begin{cases} h & \text{if } 1 \leq h \leq e-1, \\ e & \text{if } e \leq h. \end{cases}$$

Proof. $A^h$ has 1 on the $h$–th sub–diagonal, 0 elsewhere, if $1 \leq h \leq e-1$, whereas $A^h = 0$ if $h \geq e$.

Consequently

$$\nu(A^h) - \nu(A^{h-1}) = \begin{cases} 1 & \text{if } 1 \leq h \leq e, \\ 0 & \text{if } e < h. \end{cases}$$

We now can prove that the sequence $e_{i1} \geq \dots e_{i\gamma_i}$ is determined uniquely by $T$ and the eigenvalue $c_i$.

Let $p_k = x - c_k$ and

$$A = [T]^{\beta}_{\beta} = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}(c_i).$$

Then

$$\begin{aligned} \nu(p_k^h(T)) &= \nu(p_k^h(A)) \\ &= \nu(\bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} p_k^h(J_{e_{ij}}(c_i))) \\ &= \nu(\bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}^h(c_i - c_k)) \\ &= \sum_{i=1}^{t} \sum_{j=1}^{\gamma_i} \nu(J_{e_{ij}}^h(c_i - c_k)), \end{aligned}$$

where we have used the fact that

$$p_k(J_{e_{ij}}(c_i)) = J_{e_{ij}}(c_i) - c_k I_n = J_{e_{ij}}(c_i - c_k).$$

However $J_{e_{ij}}(c_i - c_k)$ is a non–singular matrix if $i \neq k$, so

$$\nu(J_{e_{ij}}^h(c_i - c_k)) = 0$$

if $i \neq k$. Hence

$$\nu(p_k^h(T)) = \sum_{j=1}^{\gamma_k} \nu(J_{e_{kj}}^h(0)).$$

Hence

$$\begin{aligned}\nu_{h,\,x-c_k} = \nu(p_k^h(T)) - \nu(p_k^{h-1}(T)) &= \sum_{j=1}^{\gamma_k} \Big(\nu(J_{e_{kj}}^h(0)) - \nu(J_{e_{kj}}^{h-1}(0))\Big) \\ &= \sum_{\substack{j=1 \\ h \le e_{kj}}}^{\gamma_k} 1.\end{aligned}$$

Consequently $\nu_{h,\,x-c_k} - \nu_{h+1,\,x-c_k}$ is the number of $e_{kj}$ which are equal to $h$. Hence by taking $h = 1, \ldots,$ we see that the sequence $e_{k1}, \ldots, e_{k\gamma_k}$ is determined by $T$ and $c_k$ and is in fact the contribution of the eigenvalue $c_k$ to the Segre characteristic of $T$ .

REMARK. If $A$ and $B$ are similar matrices over $F$, then $B = P^{-1}AP$ say. Also $A$ and $B$ have the same characteristic polynomials. Then if $c_k$ is an eigenvalue of $A$ and $B$ and $p_k = x - c_k$, we have

$$p_k^h(T_B) = p_k^h(B) = P^{-1}p_k^h(A)P = P^{-1}p_k^h(T_A)P$$

and hence

$$\nu(p_k^h(T_B)) = \nu(p_k^h(T_A))$$

for all $h \ge 1$.

Consequently the Weyr characteristics of $T_A$ and $T_B$ will be identical. Hence the corresponding dot diagrams and so the Segre characteristics will also be identical. Hence $T_A$ and $T_B$ have the same Jordan form.

**EXAMPLE 4.2**

*Let $A = J_2(0) \oplus J_2(0)$ and $B = J_2(0) \oplus J_1(0) \oplus J_1(0)$. Then*

$$ch_A = ch_B = x^4 \quad \text{and} \quad m_A = m_B = x^2.$$

*However $A$ is not similar to $B$. For both matrices are in Jordan form and the Segre characteristics for $T_A$ and $T_B$ are* 2, 2 *and* 2, 1, 1, *respectively.*

EXERCISE List all possible Jordan canonical forms of $2 \times 2$ and $3 \times 3$ matrices and deduce that if $A$ and $B$ have the same characteristic and same minimum polynomials, then $A$ and $B$ are similar if $A$ and $B$ are $2 \times 2$ or $3 \times 3$.

REMARK. Of course if $A$ and $B$ have the same Jordan canonical form, then $A$ and $B$ are similar.

We now present some interesting applications of the Jordan canonical form.

## 4.4 Non–derogatory matrices and transformations

If $\mathrm{ch}_A = m_A$, we say that the matrix $A$ is **non-derogatory**.

**THEOREM 4.5**

*Suppose that $ch_T$ splits completely in $F[x]$. Then $\mathrm{ch}_T = m_T \Leftrightarrow \exists$ a basis $\beta$ for $V$ such that*

$$[T]_\beta^\beta = J_{b_1}(c_1) \oplus \ldots \oplus J_{b_t}(c_t),$$

*where $c_1, \ldots, c_t$ are distinct elements of $F$.*

PROOF.

$\Leftarrow$

$$\begin{aligned} \mathrm{ch}_T &= \prod_{i=1}^{t} \mathrm{ch}_{J_{b_i}(c_i)} = \prod_{i=1}^{t} (x - c_i)^{b_i}, \\ m_T &= \mathrm{lcm}\,((x-c_1)^{b_1}, \ldots, (x-c_t)^{b_t}) = (x-c_1)^{b_1} \ldots (x-c_t)^{b_t} = \mathrm{ch}_T\,. \end{aligned}$$

$\Rightarrow$ Suppose that $\mathrm{ch}_T = m_T = (x-c_1)^{a_1} \cdots (x-c_t)^{a_t}$.

We deduce that the dot diagram for each $p_i = (x - c_i)$ consists of a single column of $b_i$ dots, where $p_i^{b_i} || m_T$; that is,

$$\dim_F N_{h,p_i} = 1 \qquad \text{for } h = 1, 2, \ldots, b_i.$$

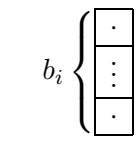
Then, for each $i = 1, 2, \ldots, t$ we have the following sequence of positive integers:

$$1 \leq \nu(p_i(T)) < \nu(p_i^2(T)) < \cdots < \nu(p_i^{b_i}(T)) = a_i.$$

But $a_i = b_i$ here, as we are assuming that $\mathrm{ch}_T = m_T$. In particular, it follows that $\nu(p_i^h(T)) = h$ for $h = 1, 2, \ldots, b_i$ and $h = 1$ gives

$$\nu(p_i(T)) = 1 = \gamma_i.$$

So the bottom row of the $i$-th dot diagram has only one element; it looks like this:

$b_i$ 

and we get the secondary decomposition

$$\operatorname{Ker} p_i^{b_i}(T) = C_{T,v_{i1}}.$$

Further, if $\beta = \beta_{11} \cup \cdots \cup \beta_{t1}$, where $\beta_{i1}$ is the elementary Jordan basis for $C_{T,v_{i1}}$, then

$$\begin{aligned} [T]_\beta^\beta &= \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}(c_i)) \\ &= \bigoplus_{i=1}^{t} J_{b_i}(c_i), \end{aligned}$$

as required.

## 4.5 Calculating $A^m$, where $A \in M_{n\times n}(\mathbb{C})$.

**THEOREM 4.6**

*Let $c \in F$.*

*(a)*

$$J_n^m(c) = \begin{bmatrix} c^m & 0 & \cdots & \cdots & & 0 \\ \binom{m}{1}c^{m-1} & c^m & \cdots & \cdots & & 0 \\ \binom{m}{2}c^{m-2} & \binom{m}{1}c^{m-1} & \cdots & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \binom{m}{m} & & \cdots & \cdots & & 0 \\ 0 & \binom{m}{m} & \cdots & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \binom{m}{m} & \cdots & \binom{m}{1}c^{m-1} & c^m \end{bmatrix}$$

*if $1 \le m \le n-1$;*

*(b)*

$$J_n^m(c) = \begin{bmatrix} c^m & 0 & \cdots & 0 & 0 \\ \binom{m}{1}c^{m-1} & c^m & \cdots & 0 & 0 \\ \binom{m}{2}c^{m-2} & \binom{m}{1}c^{m-1} & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ \binom{m}{n-1}c^{m-n+1} & \binom{m}{n-2}c^{m-n+2} & \cdots & \binom{m}{1}c^{m-1} & c^m \end{bmatrix}$$

*if* $n-1 \leq m$, *where* $\binom{m}{k}$ *is the binomial coefficient*

$$\binom{m}{k} = \frac{m!}{k!(m-k)!} = \frac{m(m-1)\cdots(m-n+1)}{k!}.$$

PROOF. $J_n(c) = cI_n + N$, where $N$ has the special property that $N^k$ has 1 on the $k$–th sub–diagonal and 0 elsewhere, for $0 \leq k \leq n-1$.

Then because $cI_n$ and $N$ commute, we can use the binomial theorem:

$$\begin{aligned} J_n^m(c) &= (cI_n + N)^m \\ &= \sum_{k=0}^{m} \binom{m}{k} (cI_n)^{m-k} N^k \\ &= \sum_{k=0}^{m} \binom{m}{k} c^{m-k} N^k. \end{aligned}$$

(a). Let $1 \leq m \leq n-1$. Then in the above summation, the variable $k$ must satisfy $0 \leq k \leq n-1$. Hence $J_n^m(c)$ is an $n \times n$ matrix having $\binom{m}{k} c^{m-k}$ on the $k$–th sub–diagonal, $0 \leq k \leq m$ and 0 elsewhere.
(b). Let $n-1 \leq m$. Then

$$J_n^m(c) = \sum_{k=0}^{m} \binom{m}{k} c^{m-k} N^k = \sum_{k=0}^{n-1} \binom{m}{k} c^{m-k} N^k,$$

as $N^k = 0$ if $n \leq k$. Hence $J_n^m(c)$ is an $n \times n$ matrix having $\binom{m}{k} c^{m-k}$ on the $k$–th sub–diagonal, $0 \leq k \leq n-1$ and 0 elsewhere.

**COROLLARY 4.1**

*Let* $F = \mathbb{C}$. *Then*

$$\lim_{m \to \infty} J_n^m(c) = 0 \quad \textit{if} \quad |c| < 1.$$

PROOF. Suppose that $|c| < 1$. Let $n-1 \leq m$. Then

$$J_n^m(c) = \sum_{k=0}^{n-1} \binom{m}{k} c^{m-k} N^k.$$

But for fixed $k$, $0 \leq k \leq n-1$, $c^{m-k} \to 0$ as $m \to \infty$. For

$$\binom{m}{k} = \frac{m(m-1)\cdots(m-k+1)}{k!}$$

is a polynomial in $m$ of degree $k$ and

$$|m^j c^{m-k}| = |m^j e^{(m-k)\log c}| = m^j e^{(m-k)\log|c|} \to 0 \quad \text{as} \quad m \to \infty,$$

as $\log c = \log|c| + i \arg c$ and $\log|c| < 0$.

The last corollary gives a more general result:

**COROLLARY 4.2**

*Let $A \in M_{n\times n}(\mathbb{C})$ and suppose that all the eigenvalues of $A$ are less than 1 in absolute value. Then*

$$\lim_{m\to\infty} A^m = 0.$$

PROOF. Suppose $ch_A = (x-c_1)^{a_1}\cdots(x-c_t)^{a_t}$, where $c_1,\ldots,c_t$ are the distinct eigenvalues of $A$ and $|c_1|<1,\ldots,|c_t|<1$.

Then if $J$ is the Jordan canonical form of $A$, there exists a non–singular matrix $P \in M_{n\times n}(\mathbb{C})$, such that

$$P^{-1}AP = J = \bigoplus_{i=1}^{t}\bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}(c_i).$$

Hence

$$P^{-1}A^mP = (P^{-1}AP)^m = J^m = \bigoplus_{i=1}^{t}\bigoplus_{j=1}^{\gamma_i} J^m_{e_{ij}}(c_i).$$

Hence $P^{-1}A^mP \to 0$ as $m\to\infty$, because $J^m_{e_{ij}}(c_i)\to 0$.

## 4.6 Calculating $e^A$, where $A \in M_{n\times n}(\mathbb{C})$.

We first show that the matrix limit

$$\lim_{M\to\infty}\left(I_n + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{M!}A^M\right)$$

exists. We denote this limit by $e^A$ and write

$$e^A = I_n + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{m!}A^m + \cdots = \sum_{m=0}^{\infty}\frac{1}{m!}A^m.$$

To justify this definition, we let $A^m = [a_{ij}^{(m)}]$. We have to show that

$$\left(I_n + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{M!}A^M\right)_{ij} = a_{ij}^{(0)} + \frac{1}{1!}a_{ij}^{(1)} + \cdots + \frac{1}{M!}a_{ij}^{(M)}$$

tends to a limit as $M \to \infty$; in other words, we have to show that the series

$$\sum_{m=0}^{\infty} \frac{1}{m!} a_{ij}^{(m)}$$

converges. To do this, suppose that

$$|a_{ij}| \leq \rho, \quad \forall\, i,\, j.$$

Then it is an easy induction to prove that

$$|a_{ij}^{(m)}| \leq n^{m-1}\rho^m \quad \text{if} \quad m \geq 1.$$

Then the above series converges by comparison with the series

$$\sum_{m=0}^{\infty} \frac{1}{m!} n^{m-1}\rho^m.$$

## 4.7 Properties of the exponential of a complex matrix

**THEOREM 4.7**

*(i)* $e^0 = I_n$;

*(ii)* $e^{diag\,(\lambda_1,\ldots,\lambda_n)} = diag\,(e^{\lambda_1},\ldots,e^{\lambda_n})$;

*(iii)* $e^{P^{-1}AP} = P^{-1}e^A P$;

*(iv)* $e^{\bigoplus_{i=1}^t A_i} = \bigoplus_{i=1}^t e^{A_i}$;

*(v)* *if $A$ is diagonable and has principal idempotent (spectral) decomposition:*

$$A = c_1E_1 + \cdots + c_tE_t,$$

*then*

$$e^A = e^{c_1}E_1 + \cdots + e^{c_t}E_t;$$

*(vi)*

$$\frac{d}{dt}\, e^{tA} = Ae^{tA},$$

*if $A$ is a constant matrix;*

*(vii)* $e^A = p(A)$, *where* $p \in \mathbb{C}[x]$;

*(viii)* $e^A$ *is non–singular and*

$$(e^A)^{-1} = e^{-A};$$

*(ix)* $e^A e^B = e^{A+B}$ *if* $AB = BA$*;*

*(x)*

$$e^{J_n(c)} = \begin{bmatrix} e^c & 0 & 0 & \cdots & & 0 \\ e^c/1! & e^c & 0 & \cdots & & 0 \\ e^c/2! & e^c/1! & e^c & \cdots & & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ e^c/(n-2)! & & \ddots & e^c/1! & e^c & 0 \\ e^c/(n-1)! & e^c/(n-2)! & \cdots & e^c/2! & e^c/1! & e^c \end{bmatrix}.$$

*(xi)*

$$e^{tJ_n(c)} = \begin{bmatrix} e^{tc} & 0 & 0 & \cdots & & 0 \\ te^{tc}/1! & e^{tc} & 0 & \cdots & & 0 \\ t^2e^{tc}/2! & e^{tc}/1! & e^{tc} & \cdots & & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ t^{n-2}e^{tc}/(n-2)! & & \ddots & te^{tc}/1! & e^{tc} & 0 \\ t^{n-1}e^{tc}/(n-1)! & t^{n-2}e^{tc}/(n-2)! & \cdots & t^2e^{tc}/2! & te^{tc}/1! & e^c \end{bmatrix}.$$

*(xii) If*

$$P^{-1}AP = J = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}(c_i).$$

*then*

$$P^{-1}e^AP = J = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} e^{J_{e_{ij}}(c_i)}.$$

PROOF.

(i)

$$e^0 = \sum_{m=0}^{\infty} \frac{1}{m!} 0^k = I_n;$$

(ii) Let $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then

$$\begin{aligned} A^m &= \operatorname{diag}(\lambda_1^m, \ldots, \lambda_n^m) \\ \sum_{m=0}^{\infty} \frac{1}{m!} A^m &= \operatorname{diag}\left( \sum_{m=0}^{\infty} \frac{\lambda_1^m}{m!}, \ldots, \sum_{m=0}^{\infty} \frac{\lambda_n^m}{m!} \right) \\ &= \operatorname{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n}). \end{aligned}$$

(iii)

$$\begin{aligned} e^{P^{-1}AP} &= \sum_{m=0}^{\infty} \frac{1}{m!} (P^{-1}AP)^m \\ &= \sum_{m=0}^{\infty} \frac{1}{m!} (P^{-1}A^m P) \\ &= P^{-1} \left( \sum_{m=0}^{\infty} \frac{1}{m!} A^m \right) P \\ &= P^{-1} e^A P. \end{aligned}$$

(iv) and (v) are left as exercises.

(vi) Using the earlier notation, $A^m = [a_{ij}^{(m)}]$, we have

$$\begin{aligned} e^{tA} &= \sum_{m=0}^{\infty} \frac{1}{m!} (tA)^m \\ &= \sum_{m=0}^{\infty} \frac{t^m}{m!} A^m \\ &= \left[ \sum_{m=0}^{\infty} \frac{t^m a_{ij}^{(m)}}{m!} \right] \\ \frac{d}{dt} e^{tA} &= \left[ \frac{d}{dt} \sum_{m=0}^{\infty} \frac{t^m a_{ij}^{(m)}}{m!} \right] \\ &= \left[ \sum_{m=1}^{\infty} \frac{t^{m-1} a_{ij}^{(m)}}{(m-1)!} \right] \\ &= \left[ \sum_{m=0}^{\infty} \frac{t^m a_{ij}^{(m+1)}}{(m)!} \right] \end{aligned}$$

$$= \sum_{m=0}^{\infty} \frac{t^m}{m!} A^{m+1}$$
$$= Ae^{tA}.$$

(vii) Let $\deg m_A = r$. Then the matrices $I_n$, $A, \ldots, A^{r-1}$ are linearly independent over $\mathbb{C}$, as if

$$m_A = x^r - a_{r-1}x^{r-1} - \cdots - a_0,$$

then

$$m_A(A) = 0 \Rightarrow A^r = a_0 I_n + a_1 A + \cdots + a_{r-1}A^{r-1}.$$

Consequently for each $m \geq 1$, we can express $A^m$ as a linear combination over $\mathbb{C}$ of $I_n$, $A$, $\ldots$, $A^{r-1}$:

$$A^m = a_0^{(m)} I_n + a_1^{(m)} A + \cdots + a_{r-1}^{(m)} A^{r-1}$$

and hence

$$\sum_{m=0}^{M} \frac{1}{m!} A^m = \sum_{m=0}^{M} \frac{a_0^{(m)}}{m!} I_n + \sum_{m=0}^{M} \frac{a_1^{(m)}}{m!} A + \cdots + \sum_{m=0}^{M} \frac{a_{r-1}^{(m)}}{m!} A^{r-1},$$

or

$$[t_{ij}^{(M)}] = s_{0M} I_n + s_{1M} A + \cdots + s_{r-1\,M} A^{r-1},$$

say.

Now $[t_{ij}^{(M)}] \to e^A$ as $M \to \infty$.

Also the above matrix equation can be regarded as $n^2$ equations in

$$s_{0M},\ s_{1M}, \ldots, s_{r-1,\,M}.$$

Also the linear independence of $I_n$, $A, \ldots, A^{r-1}$ implies that this sytem has a unique solution. Consequently we can express $s_{0M}$, $s_{1M}, \ldots, s_{r-1,\,M}$ as linear combinations *with coefficients independent of* $M$ of the sequences $t_{ij}^{(M)}$. Hence, because each of the latter sequences converges, it follows that each of the sequences $s_{0M}$, $s_{1M}, \ldots, s_{r-1,\,M}$ converges to $s_0$, $s_1, \ldots, s_{r-1}$, respectively. Consequently

$$\sum_{k=0}^{r-1} s_{kM} A^k \to \sum_{k=0}^{r-1} s_k A^k$$

and

$$e^A = s_0 I_n + s_1 A + \cdots s_{r-1} A^{r-1},$$

a polynomial in $A$.

(viii) – (ix) Suppose that $AB = BA$. Then $e^{tB}$ is a polynomial in $B$ and hence $A$ commutes with $e^{tB}$. Similarly, $A$ and $B$ commute with $e^{A+B}$. Now let

$$C(t) = e^{t(A+B)}e^{-tB}e^{-tA},\ t \in \mathbb{R}.$$

Then $C(0) = I_n$. Also

$$\begin{aligned} C'(t) &= (A+B)e^{t(A+B)}e^{-tB}e^{-tA} \\ &\quad + e^{t(A+B)}(-B)e^{-tB}e^{-tA} \\ &\quad + e^{t(A+B)}e^{-tB}(-A)e^{-tA} \\ &= 0. \end{aligned}$$

Hence $C(t)$ is a constant matrix and $C(0) = C(1)$. That is

$$I_n = e^{A+B}e^{-B}e^{-A}, \tag{12}$$

for any matrices $A$ and $B$ which commute.

The special case $B = -A$ then gives

$$I_n = e^0 e^A e^{-A} = e^A e^{-A},$$

thereby proving that $e^A$ is non–singular and $(e^A)^{-1} = e^{-A}$.

Then multiplying both sides of equation (12) on the left by $e^A e^B$ gives the equation $e^A e^B = e^{A+B}$.

In §4.8 we give an application to the solution of a system of differential equations.

(x) Let $J_n(c) = cI_n + N$, where $N = J_n(0)$. Then

$$\begin{aligned} e^{J_n(c)} &= e^{cI_n+N} = e^{cI_n}e^N \\ &= (e^c I_n)\sum_{m=0}^{\infty}\frac{1}{m!}N^m \\ &= \sum_{m=0}^{n-1}\frac{e^c}{m!}N^m. \end{aligned}$$

(xi) Similar to above.

## 4.8 Systems of differential equations

**THEOREM 4.8**

*If $X = X(t)$ satisfies the system of differential equations*

$$\dot{X} = AX,$$

*for $t \geq t_0$, where $A$ is a constant matrix, then*

$$X = e^{(t-t_0)A}X(t_0).$$

PROOF. Suppose $\dot{X} = AX$ for $t \geq t_0$. Then

$$\begin{aligned}
\frac{d}{dt}\left(e^{-tA}X\right) &= (-Ae^{-tA})X + e^{-tA}\dot{X} \\
&= (-Ae^{-tA})X + e^{-tA}(AX) \\
&= (-Ae^{-tA})X + (Ae^{-tA})X \\
&= (-Ae^{-tA} + Ae^{-tA})X \\
&= 0X = 0.
\end{aligned}$$

Hence the vector $e^{-tA}X$ is constant for $t \geq t_0$. Thus

$$e^{-tA}X = e^{-t_0A}X(t_0)$$

and

$$X = e^{tA}e^{-t_0A}X(t_0) = e^{(t-t_0)A}X(t_0).$$

**EXAMPLE 4.3**

*Solve $\dot{X} = AX$, where*

$$A = \begin{bmatrix} 0 & 4 & -2 \\ -1 & -5 & 3 \\ -1 & -4 & 2 \end{bmatrix}.$$

**Solution:** *$\exists P$ with*

$$\begin{aligned}
P^{-1}AP &= J_2(-1) \oplus J_1(-1) \\
&= \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}
\end{aligned}$$

*and*

$$P^{-1}(tA)P = \begin{bmatrix} -t & 0 & 0 \\ t & -t & 0 \\ 0 & 0 & -t \end{bmatrix}.$$

*Thus*

$$\begin{aligned} P^{-1}e^{tA}P &= e^{tJ_2(-1)} \oplus J_1(-1) \\ &= e^{tJ_2(-1)} \oplus e^{tJ_1(-1)} \\ &= \left[\begin{array}{cc|c} e^{-t} & 0 & 0 \\ te^{-t} & e^{-t} & 0 \\ \hline 0 & 0 & e^{-t} \end{array}\right] = K(t), \text{ say.} \end{aligned}$$

*So* $e^{tA} = PK(t)P^{-1}$. *Now*

$$\begin{aligned} X = e^{tA}X_0 &= e^{-t}P \begin{bmatrix} 1 & 0 & 0 \\ t & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\ &= e^{-t}P \begin{bmatrix} a \\ at+b \\ c \end{bmatrix}, \end{aligned}$$

*where for brevity we have set* $\begin{bmatrix} a \\ b \\ c \end{bmatrix} = P^{-1}X_0$.

## 4.9 Markov matrices

**DEFINITION 4.3**

*A real $n \times n$ matrix $A = [a_{ij}]$ is called a* **Markov** *matrix, or* **row–stochastic matrix** *if*

*(i)* $a_{ij} \geq 0$ *for* $1 \leq i, j \leq n$;

*(ii)* $\sum_{j=1}^{n} a_{ij} = 1$ *for* $1 \leq i \leq n$.

**Remark:** (ii) is equivalent to $AJ_n = J_n$, where $J_n = [1, \ldots, 1]^t$. So 1 is always an eigenvalue of a Markov matrix.

**EXERCISE 4.1**

*If $A$ and $B$ are $n \times n$ Markov matrices, prove that $AB$ is also a Markov matrix.*

**THEOREM 4.9**

*Every eigenvalue $\lambda$ of a Markov matrix satisfies $|\lambda| \leq 1$.*

PROOF Suppose $\lambda \in \mathbb{C}$ is an eigenvalue of $A$ and $X \in V_n(\mathbb{C})$ is a corresponding eigenvector. Then

$$AX = \lambda X. \tag{13}$$

Let $k$ be such that $|x_j| \leq |x_k|$, $\forall j$, $1 \leq j \leq n$. Then equating the $k$–th component of each side of equation (13) gives

$$\sum_{j=1}^{n} a_{kj} x_j = \lambda x_k. \tag{14}$$

Hence

$$\begin{aligned} |\lambda x_k| &= |\lambda| \cdot |x_k| = |\sum_{j=1}^{n} a_{kj} x_j| \leq \sum_{j=1}^{n} a_{kj} |x_j| && (15) \\ &\leq \sum_{j=1}^{n} a_{kj} |x_k| = |x_k|. && (16) \end{aligned}$$

Hence $|\lambda| \leq 1$.

**DEFINITION 4.4**

*A* **positive Markov matrix** *is one with all positive elements (i.e. strictly greater than zero). For such a matrix $A$ we may write "$A > 0$".*

**THEOREM 4.10**

*If $A$ is a positive Markov matrix, then* 1 *is the only eigenvalue of modulus* 1. *Moreover* nullity $(A - I_n) = 1$.

PROOF Suppose $|\lambda| = 1$, $AX = \lambda X$, $X \in V_n(\mathbb{C})$, $X \neq 0$.

Then inequalities (15) and (16) reduce to

$$|x_k| = \Big| \sum_{j=1}^{n} a_{kj} x_j \Big| \leq \sum_{j=1}^{n} a_{kj} |x_j| \leq \sum_{j=1}^{n} a_{kj} |x_k| = |x_k|. \tag{17}$$

Then inequalities (17) and a sandwich principle, give

$$|x_j| = |x_k| \quad \text{for} \quad 1 \leq j \leq n. \tag{18}$$

Also, as equality holds in the triangle inequality section of inequalities (17), this forces all the complex numbers $a_{kj} x_j$ to lie in the same direction:

$$\begin{aligned} a_{kj} x_j &= t_j a_{kk} x_k, \ , t_j > 0, \ 1 \leq j \leq n, \\ x_j &= \tau_j x_k, \end{aligned}$$

where $\tau_j = (t_j a_{kk})/a_{kj} > 0$.

Then equation (18) implies $\tau_j = 1$ and hence $x_j = x_k$ for $1 \leq j \leq n$.

Consequently $X = x_k J_n$, thereby proving that $N(A - I_n) = \langle J_n \rangle$.

Finally, equation (14) implies

$$\sum_{j=1}^{n} a_{kj} x_j = \lambda x_k = \sum_{j=1}^{n} a_{kj} x_k = x_k,$$

so $\lambda = 1$.

**COROLLARY 4.3**

*If $A$ is a positive Markov matrix, then $A^t$ has* 1 *as the only eigenvalue of modulus* 1. *Also* nullity $(A^t - I_n) = 1$.

PROOF The eigenvalues of $A^t$ are precisely the same as those of $A$, even up to multiplicities. For

$$\mathrm{ch}_{A^t} = \det(xI_n - A^t) = \det(xI_n - A)^t = \det(xI_n - A) = \mathrm{ch}_A .$$

Also $\nu(A^t - I_n) = \nu(A - I_n)^t = \nu(A - I_n) = 1$.

**THEOREM 4.11**

*If $A$ is a positive Markov matrix, then*

*(i)* $(x-1)||m_A$;

*(ii)* $A^m \to B$, *where* $B = \begin{bmatrix} \frac{X^t}{\vdots} \\ \hline X^t \end{bmatrix}$ *is a positive Markov matrix and where* $X$ *is uniquely defined as the (positive) vector satisfying* $A^tX = X$ *whose components sum to* 1.

**Remark:** In view of part (i) and the equation $\nu(A - I_n) = 1$, it follows that $(x-1)||\operatorname{ch}_A$.

PROOF As $\nu(A - I_n) = 1$, the Jordan form of $A$ has the form $J_b(1) \oplus K$, where $(x-1)^b||m_A$. Here $K$ is the direct sum of all Jordan blocks corresponding to all the eigenvalues of $A$ other than 1 and hence $K^m \to 0$.

Now suppose that $b > 1$; then $J_b(1)$ has size $b > 1$. Then $\exists P$ such that

$$\begin{aligned} P^{-1}AP &= J_b(1) \oplus K, \\ P^{-1}A^mP &= J_b^m(1) \oplus K^m. \end{aligned}$$

Hence the $2 \times 1$ element of $J_b^m(1)$ equals $\binom{m}{1} \to \infty$ as $m \to \infty$.

However the elements of $A^m$ are $\leq 1$, as $A^m$ is a Markov matrix. Consequently the elements of $P^{-1}A^mP$ are bounded as $m \to \infty$. This contradiction proves that $b = 1$.

Hence $P^{-1}A^mP \to I_1 \oplus 0$ and $A^m \to P(I_1 \oplus 0)P^{-1} = B$.

We see that $\operatorname{rank} B = \operatorname{rank}(I_1 \oplus 0) = 1$.

Finally it is easy to prove that $B$ is a Markov matrix. So

$$B = \begin{bmatrix} \frac{t_1X^t}{\vdots} \\ \hline t_nX^t \end{bmatrix}$$

for some non–negative column vector $X$ and where $t_1, \ldots, t_n$ are positive. We can assume that the entries of $X$ sum to 1. It then follows that $t_1 = \cdots = t_n = 1$ and hence

$$B = \begin{bmatrix} \frac{X^t}{\vdots} \\ \hline X^t \end{bmatrix}. \tag{19}$$

Now $A^m \to B$, so $A^{m+1} = A^m \cdot A \to BA$. Hence $B = BA$ and

$$A^t B^t = B^t. \tag{20}$$

Then equations (19) and (20) imply

$$A^t[X|\cdots|X] = [X|\cdots|X]$$

and hence $A^t X = X$.

However $X \geq 0$ and $A^t > 0$, so $X = A^t X > 0$.

**DEFINITION 4.5**

*We have thus proved that there is a positive eigenvector $X$ of $A^t$ corresponding to the eigenvalue 1, where the components of $X$ sum to 1. Then because we know that the eigenspace $N(A^t - I_n)$ is one–dimensional, it follows that this vector is unique.*

*This vector is called the* **stationary vector** *of the Markov matrix $A$.*

**EXAMPLE 4.4**

*Let*

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

*Then*

$$A^t - I_3 \text{ row–reduces to } \begin{bmatrix} 1 & 0 & -4/9 \\ 0 & 1 & -2/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

*Hence* $N(A^t - I_3) = \left\langle \begin{bmatrix} 4/9 \\ 2/3 \\ 1 \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} 4/19 \\ 6/19 \\ 9/19 \end{bmatrix} \right\rangle$ *and*

$$\lim_{m\to\infty} A^m = \frac{1}{19} \begin{bmatrix} 4 & 6 & 9 \\ 4 & 6 & 9 \\ 4 & 6 & 9 \end{bmatrix}.$$

*We remark that* $ch_A = (x-1)(x^2 - 1/24)$.

**DEFINITION 4.6**

*A Markov Matrix is called* **regular** *or* **primitive** *if* $\exists k \geq 1$ *such that* $A^k > 0$.

**THEOREM 4.12**

*If $A$ is a primitive Markov matrix, then $A$ satisfies the same properties enunciated in the last two theorems for positive Markov matrices.*

PROOF Suppose $A^k > 0$. Then $(x-1)||\operatorname{ch}_{A^k}$ and hence $(x-1)||\operatorname{ch}_A$, as

$$\operatorname{ch}_A = (x-c_1)^{a_1}\cdots(x-c_t)^{a_t} \Rightarrow \operatorname{ch}_{A^k} = (x-c_1^k)^{a_1}\cdots(x-c_t^k)^{a_t}. \qquad (21)$$

and consequently $(x-1)||m_A$.

Also as 1 is the only eigenvalue of $A^k$ with modulus 1, it follows from equation (21) that 1 is the only eigenvalue of $A$ with modulus 1.

The proof of the second theorem goes through, with the difference that to prove the positivity of $X$ we observe that $A^tX = X$ implies $(A^k)^tX = X$.

**EXAMPLE 4.5**

*The following Markov matrix is primitive (its fourth power is positive) and is related to the $5x+1$ problem:*

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}.$$

*Its stationary vector is $[\frac{1}{15}, \frac{2}{15}, \frac{8}{15}, \frac{4}{15}]^t$.*

*We remark that $ch_A = (x-1)(x+1/2)(x^2+1/4)$.*

## 4.10 The Real Jordan Form

### 4.10.1 Motivation

If $A$ is a real $n \times n$ matrix, the characteristic polynomial of $A$ will in general have real roots and complex roots, the latter occurring in complex pairs. In this section we show how to derive a canonical form $B$ for $A$ which has real entries. It turns out that there is a simple formula for $e^B$ and this is useful in solving $\dot{X} = AX$, as it allows one to directly express the complete solution of the system of differential equations in terms of real exponentials and sines and cosines.

We first introduce a real analogue of $J_n(a+ib)$. It's the matrix $K_n(a,\, b) \in M_{2n\times 2n}(\mathbb{R})$ defined as follows:

Let $D = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} = aI_2 + bJ$ where $J^2 = -I_2$ ($J$ is a matrix version of $i = \sqrt{-1}$, while $D$ corresponds to the complex number $a + ib$) then

$$\begin{aligned} e^D &= e^{aI_2+bJ} \\ &= e^{aI_2}e^{bJ} \\ &= e^a I_2 \left[ I_2 + \frac{bJ}{1!} + \frac{(bJ)^2}{2!} + \cdots \right] \\ &= e^a \left[ \left\{ I_2 - \frac{b^2}{2!}I_2 + \frac{b^4}{4!}I_2 + \cdots \right\} + \left\{ \frac{b}{1!}J - \frac{b^3}{3!}J + \cdots \right\} \right] \\ &= e^a \left[ (\cos b)I_2 + (\sin b)J \right] \\ &= e^a \begin{bmatrix} \cos b & \sin b \\ -\sin b & \cos b \end{bmatrix}. \end{aligned}$$

Replacing $a$ and $b$ by $ta$ and $tb$, where $t \in \mathbb{R}$, gives

$$e^{tD} = e^{at} \begin{bmatrix} \cos bt & \sin bt \\ -\sin bt & \cos bt \end{bmatrix}.$$

**DEFINITION 4.7**

*Let $a$ and $b$ be real numbers and $K_n(a,\, b) \in M_{2n\times 2n}(\mathbb{R})$ be defined by*

$$K_n(a,\, b) = \left[\begin{array}{c|c|c} D & 0 & \cdots \\ \hline I_2 & D & \\ \hline 0 & I_2 & \\ & & \ddots \\ & & D \end{array}\right]$$

*where* $D = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$. *Then it is easy to prove that*

$$e^{K_n(a,b)} = \left[\begin{array}{c|c|ccc} e^D & 0 & \cdots & & \\ \hline e^D/1! & e^D & & & \\ \hline e^D/2! & e^D/1! & \ddots & & \\ \vdots & & \ddots & \ddots & \\ e^D/(n-1)\ ! & \cdots \quad \cdots & & e^D/1! & e^D \end{array}\right].$$

**EXAMPLE 4.6**

$$K_2(0,\,1) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

*and*

$$e^{tK_2(0,1)} = \begin{bmatrix} \cos t & \sin t & 0 & 0 \\ -\sin t & \cos t & 0 & 0 \\ t\cos t & t\sin t & \cos t & \sin t \\ -t\sin t & t\cos t & -\sin t & \cos t \end{bmatrix}.$$

### 4.10.2 Determining the real Jordan form

If $A = [a_{ij}]$ is a complex matrix, let $\overline{A} = [\overline{a}_{ij}]$. Then

1. $$\overline{A \pm B} = \overline{A} \pm \overline{B},\ \overline{cA} = \overline{c}\overline{A}\ c \in \mathbb{C},\ \overline{AB} = \overline{A} \cdot \overline{B}.$$

2. If $A \in M_{n\times n}(\mathbb{R})$ and $a_0, \ldots, a_r \in \mathbb{C}$, then
$$\overline{a_0 I_n + \cdots a_r A^r} = \overline{a}_0 I_n + \cdots + \overline{a}_r A^r.$$

3. If $W$ is a subspace of $V_n(\mathbb{C})$, then so is $\overline{W} = \{\overline{w} | w \in W\}$.
Moreover if $W = \langle w_1, \ldots, w_r\rangle$, then
$$\overline{W} = \langle \overline{w}_1, \ldots, \overline{w}_r\rangle.$$

4. If $w_1, \ldots, w_r$ are linearly independent vectors in $V_n(\mathbb{C})$, then so are $\overline{w}_1, \ldots, \overline{w}_r$. Hence if $w_1, \ldots, w_r$ form a basis for a subspace $W$, then $\overline{w}_1, \ldots, \overline{w}_r$ form a basis for $\overline{W}$.

5. Let $A$ be a real $n \times n$ matrix and $c \in \mathbb{C}$. Then

(a)

$$W = N((A - cI_n)^h) \Rightarrow \overline{W} = N((A - \overline{c}I_n)^h).$$

(b)

$$W = W_1 \oplus \cdots \oplus W_r \Rightarrow \overline{W} = \overline{W}_1 \oplus \cdots \oplus \overline{W}_r.$$

(c)

$$W = C_{T_A, v} \Rightarrow \overline{W} = C_{T_A, \overline{v}}.$$

(d)

$$W = \bigoplus_{i=1}^{r} C_{T_A, v_i} \Rightarrow \overline{W} = \bigoplus_{i=1}^{r} C_{T_A, \overline{v}_i}.$$

(e)

$$m_{T_A, v} = (x - c)^e \Rightarrow m_{T_A, \overline{v}} = (x - \overline{c})^e.$$

Let $A \in M_{n \times n}(\mathbb{R})$. Then $m_A \in \mathbb{R}[x]$ and so any complex roots will occur in conjugate pairs.

Suppose that $c_1, \ldots, c_r$ are the distinct real eigenvalues and $c_{r+1}, \ldots, c_{r+s}$, $\bar{c}_{r+1}, \ldots, \bar{c}_{r+s}$ are the distinct non-real roots and

$$\begin{aligned} m_A &= (x - c_1)^{b_1} \ldots (x - c_r)^{b_r} (x - c_{r+1})^{b_{r+1}} \ldots (x - c_{r+s})^{b_{r+s}} \\ &\quad \times (x - \bar{c}_{r+1})^{b_{r+1}} \ldots (x - \bar{c}_{r+s})^{b_{r+s}}. \end{aligned}$$

For each complex eigenvalue $c_i$, $r+1 \leq i \leq r+s$, there exists a secondary decomposition

$$N(A - c_i I_n)^{b_i} = \bigoplus_{j=1}^{\gamma_i} C_{T_A, v_{ij}}, \quad m_{T_A, v_{ij}} = (x - c_i)^{e_{ij}}$$

Hence we have a corresponding secondary decomposition for the eigenvalue $\bar{c}_i$:

$$N(A - \bar{c}_i I_n)^{b_i} = \bigoplus_{j=1}^{\gamma_i} C_{T_A, \bar{v}_{ij}}, \quad m_{T_A, \bar{v}_{ij}} = (x - \bar{c}_i)^{e_{ij}}.$$

For brevity, let $c = c_i$, $v = v_{ij}$, $e = e_{ij}$. Let

$$P_1 = v, P_2 = (A - cI_n)P_1, \ \ldots, P_e = (A - cI_n)P_{e-1}$$

and

$$P_1 = X_1 + iY_1, \ P_2 = X_2 + iY_2, \ \ldots, \ P_e = X_e + iY_e; \ c = a + ib.$$

Then we have the following equations, posed in two different ways:

$$\begin{array}{rcl|rcl} AP_1 & = & cP_1 + P_2 & AX_1 & = & aX_1 - bY_1 + X_2 \\ & & & AY_1 & = & bX_1 + aY_1 + Y_2 \\ & \vdots & & & \vdots & \\ AP_e & = & cP_e & AX_e & = & aX_e - bY_e \\ & & & AY_e & = & bX_e + aY_e. \end{array}$$

In matrix terms we have

$$A[X_1|Y_1|X_2|Y_2|\cdots|X_e|Y_e] =$$

$$[X_1|Y_1|X_2|Y_2|\cdots|X_e|Y_e] \left[\begin{array}{cc|cc|c|c} a & b & & & \cdots & \\ -b & a & & & & \\ \hline 1 & 0 & a & b & & \\ 0 & 1 & -b & a & & \\ \vdots & & & & \ddots & \\ & & & & & \begin{array}{|cc} a & b \\ -b & a \end{array} \\ 0 & & & & & \end{array}\right].$$

The large "real jordan form" matrix is the $2e \times 2e$ matrix $K_e(a, b)$.
**Note**: If $e = 1$, no $I_2$ block is present in this matrix.

The spaces $C_{T_A, v}$ and $C_{T_A, \bar{v}}$ are independent and have bases $P_1, \ldots, P_e$ and $\bar{P}_1, \ldots, \bar{P}_e$, respectively.

Consequently the vectors

$$P_1, \ldots, P_e, \ \bar{P}_1, \ldots, \bar{P}_e$$

form a basis for $C_{T_A, v} + C_{T_A, \bar{v}}$. It is then an easy exercise to deduce that the real vectors $X_1$, $Y_1, \ldots, X_e$, $Y_e$ form a basis $\beta$ for the $T$–invariant subspace

$$W = C_{T_A, v} + C_{T_A, \bar{v}}.$$

Writing $T = T_A$ for brevity, the above right hand batch of equations tells us that $[T_W]_\beta^\beta = K_e(a, b)$. There will be $s$ such real bases corresponding to each of the complex eigenvalues $c_{r+1} \ldots, c_{r+s}$.

Joining together these bases with the real elementary Jordan bases arising from any real eigenvalues $c_1, \ldots, c_r$ gives a basis $\beta$ for $V_n(\mathbb{C})$ such that if $P$ is the non–singular real matrix formed by these basis vectors, then

$$P^{-1}AP = [T_A]_\beta^\beta = J \oplus K,$$

where

$$J = \bigoplus_{i=1}^{r} \bigoplus_{j=1}^{\gamma_i} J_{e_{ij}}(c_i), \quad K = \bigoplus_{i=r+1}^{r+s} \bigoplus_{j=1}^{\gamma_i} K_{e_{ij}}(a_i,\, b_i),$$

where $c_i = a_i + ib_i$ for $r+1 \leq i \leq r+s$.

The matrix $J \oplus K$ is said to be in real Jordan canonical form.

**EXAMPLE 4.7**

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ -2 & -1 & -1 & -1 \end{bmatrix} \quad \text{so} \quad \begin{aligned} m_A &= (x^2+1)^2 \\ &= (x-i)^2(x+i)^2. \end{aligned}$$

*Thus with $p_1 = x - i$, we have the dot diagram*

$$\begin{array}{|c|} \hline \cdot \\ \hline \cdot \\ \hline \end{array} \quad \begin{aligned} &N_{2,p_1} \\ &N_{1,p_1} = N(A - iI_4). \end{aligned}$$

*Thus we find an elementary Jordan basis for $N_{1,p_1}$:*

$$X_{11} + iY_{11},\ (A - iI_4)(X_{11} + iY_{11}) = X_{12} + iY_{12}$$

*yielding*

$$\begin{aligned} AX_{11} &= -Y_{11} + X_{12} \\ AY_{11} &= X_{11} + Y_{12}. \end{aligned} \tag{22}$$

*Now we know*

$$\begin{aligned} & m_{T_A, X_{11}+iY_{11}} = (x-i)^2 \\ \Rightarrow\ & (A - iI_4)^2(X_{11} + iY_{11}) = 0 \\ \Rightarrow\ & (A - iI_4)(X_{12} + iY_{12}) = 0 \\ \Rightarrow\ & \begin{aligned} AX_{12} &= -Y_{12} \\ AY_{12} &= X_{12}. \end{aligned} \end{aligned} \tag{23}$$

*Writing the four real equations (22) and (23) in matrix form, with*

$$P = [X_{11}|Y_{11}|X_{12}|Y_{12}],$$

*then $P$ is non-singular and*

$$P^{-1}AP = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{array}\right].$$

*The numerical determination of $P$ is left as a tutorial problem.*

### 4.10.3 A real algorithm for finding the real Jordan form

Referring to the last example, if we write $Z = \begin{bmatrix} A & I_4 \\ -I_4 & A \end{bmatrix}$, then

$$\begin{aligned} Z\begin{bmatrix} X_{11} \\ \hline Y_{11} \end{bmatrix} &= \begin{bmatrix} X_{12} \\ \hline Y_{12} \end{bmatrix}, \\ Z\begin{bmatrix} -Y_{11} \\ \hline X_{11} \end{bmatrix} &= \begin{bmatrix} -Y_{12} \\ \hline X_{12} \end{bmatrix}, \\ Z\begin{bmatrix} X_{12} \\ \hline Y_{12} \end{bmatrix} &= \begin{bmatrix} 0 \\ \hline 0 \end{bmatrix}, \\ Z\begin{bmatrix} -Y_{12} \\ \hline X_{12} \end{bmatrix} &= \begin{bmatrix} 0 \\ \hline 0 \end{bmatrix}. \end{aligned}$$

Then the vectors

$$\begin{bmatrix} X_{11} \\ \hline Y_{11} \end{bmatrix}, \begin{bmatrix} -Y_{11} \\ \hline X_{11} \end{bmatrix}, Z\begin{bmatrix} X_{11} \\ \hline Y_{11} \end{bmatrix}, Z\begin{bmatrix} -Y_{11} \\ \hline X_{11} \end{bmatrix}$$

actually form an $\mathbb{R}$–basis for $N(Z)$. This leads to a method for finding the real Jordan canonical form using real matrices. (I am indebted to Dr. B.D. Jones for introducing me to the $Z$ matrix approach.)

More generally, we observe that a collection of equations of the form

$$\begin{aligned} AX_{ij1} &= a_iX_{ij1} - b_iY_{ij1} + X_{ij2} \\ AY_{ij1} &= b_iX_{ij1} + a_iY_{ij1} + Y_{ij2} \\ &\vdots \\ AX_{ije_{ij}} &= a_iX_{ije_{ij}} - b_iY_{ije_{ij}} \\ AY_{ije_{ij}} &= b_iX_{ije_{ij}} + a_iY_{ije_{ij}} \end{aligned}$$

can be written concisely in real matrix form, giving rise to an elementary Jordan basis corresponding to an elementary divisor $x^{e_{ij}}$ for the following real matrix: Let

$$Z_i = \begin{bmatrix} A - a_iI_n & b_iI_n \\ -b_iI_n & A - a_iI_n \end{bmatrix}.$$

Then

$$\begin{aligned} Z_i\begin{bmatrix} X_{ij1} \\ \hline Y_{ij1} \end{bmatrix} &= \begin{bmatrix} X_{ij2} \\ \hline Y_{ij2} \end{bmatrix} \\ &\vdots \\ Z_i\begin{bmatrix} X_{ije_{ij}} \\ \hline Y_{ije_{ij}} \end{bmatrix} &= \begin{bmatrix} 0 \\ \hline 0 \end{bmatrix}. \end{aligned}$$

**LEMMA 4.2**

*If $V$ is a $\mathbb{C}$–vector space with basis $v_1, \ldots, v_n$, then $V$ is also an $\mathbb{R}$–vector space with basis*

$$v_1,\ iv_1, \ldots, v_n,\ iv_n.$$

*Hence*

$$\dim_{\mathbb{R}} V = 2 \dim_{\mathbb{C}} V.$$

**DEFINITION 4.8**

*Let $A \in M_{n\times n}(\mathbb{R})$ and $c = a + ib$ be a complex eigenvalue of $A$ with $b \neq 0$. Let $Z \in M_{2n\times 2n}(\mathbb{R})$ be defined by*

$$Z = \begin{bmatrix} A - aI_n & bI_n \\ -bI_n & A - aI_n \end{bmatrix} = (A - aI_n) \otimes I_n - I_n \otimes (bJ).$$

*Also let $p = x - c$.*

**LEMMA 4.3**

*Let $\Phi : V_{2n}(\mathbb{R}) \to V_n(\mathbb{C})$ be the mapping defined by*

$$\Phi\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = X + iY, \quad X,\ Y \in V_n(\mathbb{R}).$$

*Then*

*(i)* $\Phi$ *is an* $\mathbb{R}$*– isomorphism;*

*(ii)* $\Phi\left(\left[\frac{-Y}{X}\right]\right) = i(X + iY)$*;*

*(iii)* $\Phi\left(Z^h \left[\frac{X}{Y}\right]\right) = p^h(A)(X + iY)$*;*

*(iv)* $\Phi\left(Z^h \left[\frac{-Y}{X}\right]\right) = ip^h(A)(X + iY)$*;*

*(v)* $\Phi$ *maps* $N(Z^h)$ *onto* $N(p^h(A)$*;*

**COROLLARY 4.4**

*If*

$$p^{e_1-1}(A)(X_1 + iY_1), \ldots, p^{e_\gamma-1}(A)(X_\gamma + iY_\gamma)$$

*form a $\mathbb{C}$–basis for $N(p(A))$, then*

$$Z^{e_1-1}\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix},\ Z^{e_1-1}\begin{bmatrix} -Y_1 \\ X_1 \end{bmatrix}, \ldots, Z^{e_\gamma-1}\begin{bmatrix} X_\gamma \\ Y_\gamma \end{bmatrix},\ Z^{e_\gamma-1}\begin{bmatrix} -Y_\gamma \\ X_\gamma \end{bmatrix}$$

*form an $\mathbb{R}$–basis for $N(Z)$ and conversely.*

**Remark:** Consequently the dot diagram for the eigenvalue 0 for the matrix $Z$ has the same height as that for the eigenvalue $c$ of $A$, with each row expanded to twice the length.

To find suitable vectors $X_1, Y_1, \ldots, X_\gamma, Y_\gamma$, we employ the usual algorithm for finding the Jordan blocks corresponding to the eigenvalue 0 of the matrix $Z$, with the extra proviso that we always ensure that the basis for $N_{h,x}$ is chosen to have the form

$$Z^{h-1}\begin{bmatrix} X_1 \\ \hline Y_1 \end{bmatrix}, Z^{h-1}\begin{bmatrix} -Y_1 \\ \hline X_1 \end{bmatrix}, \ldots, Z^{h-1}\begin{bmatrix} X_r \\ \hline Y_r \end{bmatrix}, Z^{h-1}\begin{bmatrix} -Y_r \\ \hline X_r \end{bmatrix},$$

where $r = (\text{ nullity } Z^h - \text{ nullity } Z^{h-1})/2$.

This can be ensured by extending a spanning family for $N(Z^h)$:

$$\begin{bmatrix} X_1 \\ \hline Y_1 \end{bmatrix}, \ldots, \begin{bmatrix} X_{\nu(Z^h)} \\ \hline Y_{\nu(Z^h)} \end{bmatrix}$$

to the form

$$\begin{bmatrix} X_1 \\ \hline Y_1 \end{bmatrix}, \begin{bmatrix} -Y_1 \\ \hline X_1 \end{bmatrix}, \ldots, \begin{bmatrix} X_{\nu(Z^h)} \\ \hline Y_{\nu(Z^h)} \end{bmatrix}, \begin{bmatrix} -Y_{\nu(Z^h)} \\ \hline X_{\nu(Z^h)} \end{bmatrix}.$$

**EXAMPLE 4.8**

$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ -2 & -1 & -1 & -1 \end{bmatrix} \in M_{4\times 4}(\mathbb{R})$ *has* $m_A = (x^2+1)^2$. *Find a real non–singular matrix* $P$ *such that* $P^{-1}AP$ *is in real Jordan form.*

**Solution:**

$$Z = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -2 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ -2 & -1 & -1 & -1 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & -2 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & -2 & -1 & -1 & -1 \end{bmatrix}$$

$$\textit{basis for } N(Z^2): \begin{bmatrix} 1 & 1 & 1/2 & 1/2 \\ -2 & -1/2 & 0 & -1/2 \\ 2 & 1/2 & 1 & 3/2 \\ -2 & -3/2 & -2 & -3/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

*blown–up basis for* $N(Z^2)$ :

$$\begin{bmatrix} 1 & -1 & 1 & 0 & 1/2 & 0 & 1/2 & 0 \\ -2 & 0 & -1/2 & -1 & 0 & 0 & -1/2 & 0 \\ 2 & 0 & 1/2 & 0 & 1 & -1 & 3/2 & 0 \\ -2 & 0 & -3/2 & 0 & -2 & 0 & -3/2 & -1 \\ 1 & 1 & 0 & 1 & 0 & 1/2 & 0 & 1/2 \\ 0 & -2 & 1 & -1/2 & 0 & 0 & 0 & -1/2 \\ 0 & 2 & 0 & 1/2 & 1 & 1 & 0 & 3/2 \\ 0 & -2 & 0 & -3/2 & 0 & -2 & 1 & -3/2 \end{bmatrix}$$

$$\rightarrow \textit{ left–to–right basis for } N(Z^2): \begin{bmatrix} 1 & -1 & 1 & 0 \\ -2 & 0 & -1/2 & -1 \\ 2 & 0 & 1/2 & 0 \\ -2 & 0 & -3/2 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & -2 & 1 & -1/2 \\ 0 & 2 & 0 & 1/2 \\ 0 & -2 & 0 & -3/2 \end{bmatrix}$$

*We then derive a spanning family for* $N_{2,\,x}$*:*

$$Z \times \textit{ basis matrix} = \begin{bmatrix} 0 & 0 & 1/2 & 0 \\ 0 & 0 & -1/2 & -1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & -1/2 & -1/2 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & -1/2 & 1/2 \\ 0 & 0 & 1/2 & -1/2 \end{bmatrix} \quad \rightarrow \quad \textit{basis for } N_{2,\,x}:$$

$$\begin{bmatrix} 1/2 & 0 \\ -1/2 & -1/2 \\ 1/2 & 1/2 \\ -1/2 & -1/2 \\ 0 & 1/2 \\ 1/2 & -1/2 \\ -1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

*Consequently we read off that* $Z \begin{bmatrix} X_{11} \\ Y_{11} \end{bmatrix} = \begin{bmatrix} X_{12} \\ Y_{12} \end{bmatrix}$ *is a basis for* $N_{2,\,x} = N_{1,\,x} = N(Z)$. *where*

$$P = [X_{11}|Y_{11}|X_{12}|Y_{12}] = \begin{bmatrix} 1 & 0 & 1/2 & 0 \\ -1/2 & 1 & -1/2 & 1/2 \\ 1/2 & 0 & 1/2 & -1/2 \\ -3/2 & 0 & -1/2 & 1/2 \end{bmatrix}.$$

*Then*

$$P^{-1}AP = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix},$$

*which is in real Jordan form.*

# 5 The Rational Canonical Form

Here $p$ is a monic irreducible factor of the minimum polynomial $m_T$ and is not necessarily of degree one.

Let $F_p$ denote the field constructed earlier in the course, consisting of all matrices of the form $f(B)$, $f \in F[x]$, where $B = C(p)$, the companion matrix of $p$. (We saw that if $\deg p = n$, then

$$F_p = \{a_0 I_n + \cdots + a_{n-1} B^{n-1} | \, a_0, \ldots, a_{n-1} \in F\}.$$

Let $\overline{f} = f(B)$, where $f \in F[x]$. Then this new symbol has the following properties:

(i) $\overline{f} + \overline{g} = \overline{f+g}$; $\overline{f}\,\overline{g} = \overline{fg}$;

(ii) $\overline{f} = \overline{0} \Leftrightarrow p \mid f$;

(iii) $\overline{f} = \overline{g} \Leftrightarrow p \mid (f-g)$;

(iv) $\overline{f}^{-1}$ exists $\Leftrightarrow p$ does not divide $f$.

**Note**: If $p = x - c$, then $F_p = F$.

**THEOREM 5.1**

*$N_{h,p}$ becomes a vector space over $F_p$ if we define*

$$\overline{f}v = fv = f(T)(v).$$

First we must verify that the above definition is well–defined, that is, independent of the particular polynomial $f$ used to define the field element $\overline{f}$. So suppose $\overline{f} = \overline{g}$. Then $f = g + kp$, $k \in F[x]$. Hence

$$fv = (g + kp)v = gv + k(pv) = gv + k0 = gv,$$

as $v \in \operatorname{Im} p^{h-1}(T) \cap \operatorname{Ker} p(T)$ and consequently $pv = 0$.

The four addition axioms hold as $V$ is already a vector space over $F$; The remaining vector space axioms then follow from the left $F[x]$–module axioms:

(i) $(\overline{f} + \overline{g}) = \overline{(f+g)}v = (f+g)v = fv + gv = \overline{f}v + \overline{g}v$;

(ii) $\overline{f}(v+w) = f(v+w) = fv + fw = \overline{f}v + \overline{f}w$;

(iii) $\overline{f}(\overline{g}v) = \overline{f}(gv) = f(gv) = (fg)v = (\overline{fg})v$;

(iv) $\overline{1}v = 1v = v$.

**Remark:** An $F$–basis for $N_{h,\,p}$ will be an $F_p$–spanning family for $N_{h,\,p}$, but will not, in general, be an $F_p$–basis for $N_{h,\,p}$. The precise connection between $F$–independence and $F_p$–independence is given by the following theorem:

**THEOREM 5.2**

*Vectors $v_1, \ldots, v_r$ form an $F_p$–basis for $N_{h,\,p}$ if and only if the vectors*

$$\begin{array}{cccc} v_1, & T(v_1), & \ldots, & T^{n-1}(v_1) \\ v_2, & T(v_2), & \ldots, & T^{n-1}(v_2) \\ \vdots & \vdots & \vdots & \vdots \\ v_r, & T(v_r), & \ldots, & T^{n-1}(v_r) \end{array}$$

*form an $F$–basis for $N_{h,\,p}$.*

**COROLLARY 5.1**

$$\nu_{h,\,p} = \dim_{F_p} N_{h,\,p} = \frac{1}{\deg p} \dim_F N_{h,\,p} = \frac{\nu(p^h(T)) - \nu(p^{h-1}(T))}{\deg p}.$$

The exposition for $p = x - c$ now goes over to general $p$, with small changes. We again have the decreasing sequence of dimensions:

$$\nu_{1,\,p} \geq \cdots \geq \nu_{b,\,p} \geq 1,$$

where $\nu_{1,\,p} = \dim_{F_p} \operatorname{Ker} p(T) = \frac{\nu(p(T))}{\deg p}$.

Also

$$\nu_{1,\,p} + \cdots + \nu_{b,\,p} = \frac{\nu(p^b(T))}{\deg p}, \tag{24}$$

where $p^b \parallel \mathrm{m}_T$.

There is a corresponding dot diagram where the number of dots in the $h$–th row from the bottom represents the integer $\nu_{h,\,p}$. We also have a similar theorem to an earlier one, in terms of the conjugate partition

$$e_1 \geq \cdots \geq e_\gamma \geq 1$$

of the partition (24) above, where $\gamma = \nu_{1,\,p} = \dim_{F_p} \operatorname{Ker} p(T) = \frac{\nu(p(T))}{\deg p}$.

**THEOREM 5.3**

*Vectors $v_1, \ldots, v_\gamma \in V$ can be found with the property that*

$$p^{e_1-1}v_1, \ldots, p^{e_\gamma-1}v_\gamma$$

*form an $F_p$–basis for* $\operatorname{Ker} p(T)$. *Moreover*

(i) $m_{T,\,v_j} = p^{e_j}$;

(ii) $\operatorname{Ker} p^b(T) = C_{T,\,v_1} \oplus \cdots \oplus C_{T,\,v_\gamma}$.

In conclusion, if $m_T = p_1^{b_1} \cdots p_t^{b_t}$, we now have the direct sum decomposition

$$V = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C_{T,\,v_{ij}},$$

where $m_{T,\,v_{ij}} = p_i^{e_{ij}}$ and

$$e_{i1} = b_i \geq \ldots \geq e_{i\gamma_i}$$

form the conjugate partition for the dot diagram corresponding to $p_i$. Here

$$\gamma_i = \frac{\nu(p_i(T))}{\deg p_i}.$$

Taking $T$–cyclic bases $\beta_{ij}$ for $C_{T,\,v_{ij}}$, then gives a basis

$$\beta = \bigcup_{i=1}^{t} \bigcup_{j=1}^{\gamma_i} \beta_{ij}$$

for $V$. Moreover

$$[T]_\beta^\beta = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C(p_i^{e_{ij}}).$$

The matrix on the right is said to be in **rational canonical form**.

If instead, we take the following basis $\beta'_{ij}$ for $C_{T,\,v_{ij}}$

$$\beta'_{ij} : \begin{cases} v_{ij}, & T(v_{ij}), & \ldots, & T^{n-1}(v_{ij}) \\ p_i(T)(v_{ij}), & Tp_i(T)(v_{ij}), & \ldots, & T^{n-1}p_i(T)(v_{ij}) \\ \vdots & \vdots & \vdots & \vdots \\ p_i^{e_{ij}-1}(T)(v_{ij}), & Tp_i^{e_{ij}-1}(T)(v_{ij}), & \ldots, & T^{n-1}p_i^{e_{ij}-1}(T)(v_{ij}), \end{cases}$$

(with $n = \deg p_i$) which reduces to the Jordan basis when $p_i = x - c_i$, it is not difficult to verify that we get a corresponding matrix $H(p_i^{e_{ij}})$ called a **hypercompanion** matrix, which reduces to the elementary Jordan matrix $J_{e_{ij}}(c_i)$ when $p_i = x - c_i$:

$$H(p_i^{e_{ij}}) = \begin{bmatrix} C(p_i) & 0 & \cdots & 0 \\ N & C(p_i) & \cdots & 0 \\ 0 & N & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & N & C(p_i) \end{bmatrix},$$

where there are $e_{ij}$ blocks on the diagonal and $N$ is a square matrix of same size as $C(p_i)$ which is everywhere zero, except in the top right–hand corner, where there is a 1. The overall effect is an unbroken subdiagonal of $1'$s.

We then get the corresponding rational canonical form:

$$[T]_{\beta'}^{\beta'} = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} H(p_i^{e_{ij}}).$$

Computational Remark:

We can do our computations completely over $F$, without going into $F_p$, as follows. Suppose $v_1, \ldots, v_r$ form an $F$–spanning family for $N_{h,\,p}$. Then we could, in principle, perform the LRA over $F_p$ on this spanning family and find an $F_p$–basis $v_{c_1}, \ldots, v_{c_R}$. A little thought reveals that if we had instead applied the LRA algorithm over $F$ to the expanded sequence:

$$v_1,\, T(v_1),\, \ldots, T^{n-1}(v_1); \ldots;\, v_r,\, T(v_r), \ldots, T^{n-1}(v_r),$$

we would have obtained the $F$–basis for $N_{h,\,p}$:

$$v_{c_1},\, T(v_{c_1}),\, \ldots, T^{n-1}(v_{c_1}); \ldots;\, v_{c_R},\, T(v_{c_R}), \ldots, T^{n-1}(v_{c_R})$$

from which we select the desired $F_p$–basis $v_{c_1}, \ldots, v_{c_R}$.

Let $A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 2 & 2 \\ 2 & 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \in M_{6\times 6}(\mathbb{Z}_3)$.

Here $m_A = p^2,\ p = x^2 + x + 2 \in F[x],\ F = \mathbb{Z}_3$.

$$p(A) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \nu(p(A)) = 4,\ \nu_{1,p} = \tfrac{\nu(p(A))}{\deg p} = 2.$$

$$p^2(A) = 0,\ \nu(p^2(A)) = 6,\ \nu_{2,p} = \frac{\nu(p^2(A)) - \nu(p(A))}{\deg p} = \frac{6-4}{2} = 1.$$

Hence we have a corresponding $F_p$ dot diagram:


We have to find an $F_p$–basis $p(A)v_{11}$ for $N_{2,p}$ and extend this to an $F_p$–basis $p(A)v_{11},\ v_{12}$ for $N(p(A))$.

An $F$–basis for $N(p^2(A))$ is $E_1, \ldots, E_6$. Then

$$N_{2,p} = \langle p(A)E_1, \ldots, p(A)E_6 \rangle$$

and the LRA give $p(A)E_2$ as an $F_p$–basis for $N_{2,p}$ so we can take $v_{11} = E_2$.

We find the columns of the following matrix form an $F$–basis for $N(p(A))$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We place $p(A)E_2$ in front and then pad the resulting matrix to get

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 2 \\ 2 & 0 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 \\ 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 2 & 1 & 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

The first four columns $p(A)E_2,\ Ap(A)E_2,\ E_1,\ AE_1$ of this matrix form a LR $F$–basis for $N(p(A))$ and hence $p(A)E_2,\ E_1$ form an $F_p$–basis for $N(p(A))$. So we can take $v_{12} = E_1$.

Then $V_6(\mathbb{Z}_3) = N(p^2(A)) = C_{T_A,\, v_{11}} \oplus C_{T_A,\, v_{12}}$.
Then joining hypercompanion bases for $C_{T_A,\, v_{11}}$ and $C_{T_A,\, v_{12}}$:

$$v_{11},\ Av_{11},\ p(A)v_{11},\ Ap(A)v_{11} \quad \text{and} \quad v_{12},\ Av_{11}$$

gives a basis $v_{11}$, $Av_{11}$, $p(A)v_{11}$, $Ap(A)v_{11}; v_{12}$, $Av_{11}$ for $V_6(\mathbb{Z}_3)$. Finally if $P$ is the non–singular matrix whose columns are these vectors, we transform $A$ into direct sum of hypercompanion matrices:

$$P^{-1}AP = H(p^2) \oplus H(p) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

Explicitly, we have

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

## 5.1 Uniqueness of the Rational Canonical Form

Suppose that $T : V \to V$ is a linear transformation over $F$ and that $\beta$ is a basis for $V$ such that

$$[T]_\beta^\beta = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C(p_i^{e_{ij}}). \tag{25}$$

where

$$e_{i1} \geq \ldots \geq e_{i\gamma_i} \geq 1 \tag{26}$$

and $p_1, \ldots, p_t$ are distinct monic irreducible polynomials.

We show that the polynomials $p_i$ and the sequences (26) are determined by the transformation $T$.

First, it is not difficult to show that

$$\beta = \bigcup_{i=1}^{t} \bigcup_{j=1}^{\gamma_i} \beta_{ij},$$

where

$$\beta_{ij} : v_{ij},\, T(v_{ij}), \ldots, T^{n_{ij}-1}(v_{ij})$$

and $n_{ij} = \deg p_i^{e_{ij}}$ and $m_{T,\,v_{ij}} = p_i^{e_{ij}}$. Then we have the direct sum decomposition

$$V = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C_{T,\,v_{ij}}.$$

Also if we write $b_i = e_{i1}$, we have

$$\operatorname{Ker} p_i^{b_i}(T) = \bigoplus_{j=1}^{\gamma_i} C_{T,\,v_{ij}}$$

and hence

$$V = \bigoplus_{i=1}^{t} \operatorname{Ker} p_i^{b_i}(T).$$

Then from equation (25) above, it follows that

$$\mathrm{m}_T = \operatorname{lcm} p_i^{e_{ij}} = p_1^{b_1} \cdots p_t^{b_t},$$

thereby determining $p_1, \ldots, p_t$ up to order.

Then it can be shown that if $1 \leq h \leq b_i$, then $N_{h,\,p_i}$ has $F_{p_i}$ basis

$$p_i^{e_{i1}-1} v_{i1}, \ldots, p_i^{e_{ij_h}-1} v_{ij_h},$$

where $e_{i1}, \ldots, e_{ij_h}$ are the integers not less than $h$.

There are consequently $\dim_{F_{p_i}} N_{h,\,p_i} = \nu_{h,\,p_i}$ such integers and hence the number of integers $e_{i1}, \ldots, e_{i\gamma_i}$ equal to $h$ is equal to $\nu_{h,\,p_i} - \nu_{h+1,\,p_i}$, which depends only on $T$. In other words, for each $i$, the sequence $e_{i1}, \ldots, e_{i\gamma_i}$ depends only on $T$.

## 5.2 Deductions from the Rational Canonical Form

**THEOREM 5.4**

$$\frac{\nu(p_i^{b_i}(T))}{\deg p_i} = a_i$$

*where* $p_i^{a_i} || \operatorname{ch}_T$ , *and* $p_i^{b_i} || m_T$.

Note that this determines $b_i$—we may evaluate

$$\frac{\nu(p_i^h(T))}{\deg p_i}$$

for $h = 1, 2, \ldots$ until we get a value of $a_i$. Then that $h = b_i$.
PROOF $\exists$ a basis for $V$ such that

$$A = [T]_\beta^\beta = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C(p_i^{e_{ij}}).$$

So

$$\mathrm{ch}_T = \prod_{i=1}^{t} \prod_{j=1}^{\gamma_i} \mathrm{ch}_{B_{i,j}}$$

where, for brevity, we write $B_{i,j} = C(p_i^{e_{ij}})$. Hence

$$\begin{aligned} \mathrm{ch}_T &= \prod_{i=1}^{t} \prod_{j=1}^{\gamma_i} p_i^{e_{ij}} \\ &= \prod_{i=1}^{t} p_i^{\sum_{j=1}^{\gamma_i} e_{ij}} \\ &= \prod_{i=1}^{t} p_i^{\frac{\nu(p_i^{b_i}(T))}{\deg p_i}} \end{aligned}$$

as required.

**THEOREM 5.5**
$\mathrm{ch}_T = m_T \Leftrightarrow \exists$ *a basis* $\beta$ *for* $V$ *such that*

$$[T]_\beta^\beta = C(p_1^{b_1}) \oplus \ldots \oplus C(p_t^{b_t})$$

*where* $p_1, \ldots, p_t$ *are distinct monic irreducibles and* $b_1 \geq \ldots b_t \geq 1$.

Note that if $\mathrm{ch}_A = m_A$ (i.e. $T = T_A$ in the above), we say that the matrix $A$ is **non-derogatory**.
PROOF

$\Leftarrow$

$$\begin{aligned} \mathrm{ch}_T &= \prod_{i=1}^{t} \mathrm{ch}_{C(p_i^{b_i})} = \prod_{i=1}^{t} p_i^{b_i}, \\ m_T &= \mathrm{lcm}\,(p_1^{b_1}, \ldots, p_t^{b_t}) = p_1^{b_1} \ldots p_t^{b_t} = \mathrm{ch}_T\,. \end{aligned}$$

$\Rightarrow$ Suppose that $\mathrm{ch}_T = m_T$.

We deduce that the dot diagram for each $p_i$ consists of a single column of $b_i$ dots, where $p_i^{b_i} || m_T$; that is,

$$\dim_{F_p} N_{h,p_i} = 1 \qquad \text{for } h = 1, 2, \ldots, b_i.$$

Observe that

$$\frac{\nu(p_i^h(T))}{\deg p_i} \in \mathbb{N},$$

for it may be written

$$\begin{aligned} & \sum_{j=1}^{h} \frac{\nu(p_i^j(T)) - \nu(p_i^{j-1}(T))}{\deg p_i} \\ = & \sum_{j=1}^{h} \dim_{F_p} N_{j,p_i} \in \mathbb{N}. \end{aligned}$$

Then, for each $i = 1, 2, \ldots, t$ we have the following sequence of positive integers:

$$1 \leq \frac{\nu(p_i(T))}{\deg p_i} < \frac{\nu(p_i^2(T))}{\deg p_i} < \ldots < \frac{\nu(p_i^{b_i}(T))}{\deg p_i} = a_i.$$

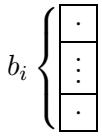
But $a_i = b_i$ here, as we are assuming that $\mathrm{ch}_T = m_T$. In particular, it follows that

$$\frac{\nu(p_i^h(T))}{\deg p_i} = h \qquad \text{for } h = 1, 2, \ldots, b_i$$

and $h = 1$ gives

$$\frac{\nu(p_i(T))}{\deg p_i} = 1 = \gamma_i.$$

So the bottom row of the $i$-th dot diagram has only one element; it looks like this:

$b_i$ 

and we get the secondary decomposition

$$\operatorname{Ker} p_i^{b_i}(T) = C_{T,v_{i1}}.$$

Further, if $\beta = \beta_{11} \cup \cdots \cup \beta_{t1}$, where $\beta_{i1}$ is the $T$–cyclic basis for $C_{T,v_{i1}}$, then

$$\begin{aligned} [T]_\beta^\beta &= \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C(p_i^{e_{ij}}) \\ &= \bigoplus_{i=1}^{t} C(p_i^{b_i}) \\ &= C(p_1^{b_1}) \oplus \ldots \oplus C(p_t^{b_t}) \end{aligned}$$

as required.

**THEOREM 5.6**

*$m_T = p_1 p_2 \ldots p_t$, a product of distinct monic irreducibles, if and only if $\exists$ a basis $\beta$ for $V$ such that*

$$\begin{aligned} [T]_\beta^\beta = \quad & \underbrace{C(p_1) \oplus \ldots \oplus C(p_1)}_{\gamma_1 \text{ times}} \oplus \ldots\ldots \\ \oplus \quad & \underbrace{C(p_t) \oplus \ldots \oplus C(p_t)}_{\gamma_t \text{ times}}. \end{aligned} \tag{27}$$

**Note**: This is a generalization of an earlier result, namely that a transformation is diagonable if and only if its minimum polynomial splits into a product of distinct linear factors.

PROOF

$\Leftarrow$ Assume $\exists \beta$ such that (27) holds. Then

$$\begin{aligned} m_T &= \operatorname{lcm}(\underbrace{p_1, \ldots, p_1}_{\gamma_1}, \ldots, \underbrace{p_t, \ldots, p_t}_{\gamma_t}) \\ &= \operatorname{lcm}(p_1, \ldots, p_t) \\ &= p_1 p_2 \ldots p_t. \end{aligned}$$

$\Rightarrow$ Assume $m_T = p_1 \ldots p_t$. Then $b_i = 1$ for $i = 1, \ldots, t$ (i.e. the $i$-th dot diagram has height 1) and $\exists \beta$ such that

$$[T]_\beta^\beta = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C(p_i),$$

as $e_{ij} = 1 \ \forall i, j$.

## 5.3 Elementary divisors and invariant factors

### 5.3.1 Elementary Divisors

**DEFINITION 5.1**

*The polynomials $p_i^{e_{ij}}$ occurring in the rational canonical form of $T$ are called the* **elementary divisors** *of $T$. Similarly the elementary divisors of a matrix $A \in M_{n\times n}(F)$ are the polynomials $p_i^{e_{ij}}$ occurring in the rational canonical form of $A$.*

**THEOREM 5.7**

*Linear transformations $T_1, T_2 : V \to V$ have the same elementary divisors if and only if there exists an isomorphism $L : V \to V$ such that*

$$T_2 = L^{-1}T_1L.$$

PROOF

"only if". Suppose that $T_1$ and $T_2$ have the same elementary divisors. Then $\exists$ bases $\beta$, $\gamma$ for $V$ such that

$$[T_1]_\beta^\beta = [T_2]_\gamma^\gamma = A.$$

Then we have the equations

$$\begin{aligned} \phi_\beta T_1 &= T_A\phi_\beta \\ \phi_\gamma T_2 &= T_A\phi_\gamma. \end{aligned}$$

Hence

$$\phi_\beta T_1\phi_\beta^{-1} = T_A = \phi_\gamma T_2\phi_\gamma^{-1},$$

so

$$\phi_\gamma^{-1}\phi_\beta T_1\phi_\beta^{-1}\phi_\gamma = T_2,$$

or

$$L^{-1}T_1L = T_2,$$

where $L = \phi_\beta^{-1}\phi_\gamma$ is an isomorphism.

"if". Suppose that $L^{-1}T_1L = T_2$. Then

$$m_{T_1} = m_{T_2} = p_1^{b_1}\cdots p_t^{b_t}, \quad \text{say;}$$

also for all $i$ and $h$, because

$$p_i^h(T_2) = p_i^h(L^{-1}T_1L) = L^{-1}p_i^h(T_1)L,$$

we have

$$\nu(p_i^h(T_2)) = \nu(p_i^h(T_1)).$$

Hence for each $p_i$, the corresponding dot diagrams for $T_1$ and $T_2$ are identical and consequently the elementary divisors for $T_1$ and $T_2$ are identical.

**COROLLARY 5.2**

*Let $A$, $B \in M_{n\times n}(F)$. Then $A$ is similar to $B$ if and only if $A$ and $B$ have the same elementary divisors.*

PROOF

$$\begin{aligned} A \text{ is similar to } B &\Leftrightarrow \exists\ P \text{ non–singular, with } P^{-1}AP = B \\ &\Leftrightarrow \exists\ P \text{ non–singular, with } T_P^{-1}T_AT_P = T_B \\ &\Leftrightarrow \exists\ L \text{ an isomorphism, with } L^{-1}T_AL = T_B. \end{aligned}$$

### 5.3.2 Invariant Factors

**THEOREM 5.8**

*Let $T : V \to V$ be a linear transformation over $F$. Then there exist non–constant monic polynomials $d_1, \ldots, d_s \in F[x]$, such that*

*(i) $d_k$ divides $d_{k+1}$ for $1 \le k \le s-1$;*

*(ii) vectors $v_1, \ldots, v_s \in V$ exist such that*

$$V = \bigoplus_{k=1}^{s} C_{T,\,v_k},$$

*where $m_{T,\,v_k} = d_k$.*

REMARK: If $\beta$ is the basis for $V$ obtained by stringing together the $T$–cyclic bases for each $C_{T,\,v_k}$, we obtain the matrix direct sum

$$[T]_\beta^\beta = \bigoplus_{k=1}^{s} C(d_k).$$

This matrix is also said to be in rational canonical form.
PROOF

Let $s = \max(\gamma_1, \ldots, \gamma_t)$ and if $1 \leq i \leq t$ and $\gamma_i < j \leq s$, define $e_{ij} = 0$ and $v_{ij} = 0$, the zero vector of $V$. Now arrange the polynomials $p_i^{e_{ij}}$, $1 \leq i \leq t$; $1 \leq j \leq s$ as a $t \times s$ rectangular array:

$$\begin{array}{|c|c|c|} \hline p_1^{e_{1s}} & \cdots & p_1^{e_{11}} \\ \hline \vdots & \vdots & \vdots \\ \hline p_t^{e_{ts}} & \cdots & p_t^{e_{t1}} \\ \hline \end{array}$$

Let

$$d_1 = p_1^{e_{1s}} \cdots p_t^{e_{ts}}, \ldots, d_s = p_1^{e_{11}} \cdots p_t^{e_{t1}}$$

be the products along columns of the array, from left to right. Then $d_1, \ldots, d_s$ are monic non–constant polynomials and

$$d_1 | d_2 | \cdots | d_s.$$

Also $C_{T, v_{ij}} = \{0\}$ if $v_{ij} = 0$, so $V$ is the direct sum of the following $ts$ $T$–cyclic subspaces:

$$\begin{array}{|c|c|c|} \hline C_{T, v_{1s}} & \cdots & C_{T, v_{11}} \\ \hline \vdots & \vdots & \vdots \\ \hline C_{T, v_{ts}} & \cdots & C_{T, v_{t1}} \\ \hline \end{array}$$

Then by Problem Sheet 5, Question 15(b), if we let

$$v_1 = v_{1s} + \cdots v_{ts}, \ldots, v_s = v_{11} + \cdots v_{t1},$$

we have $m_{T, v_1} = d_1, \ldots, m_{T, v_s} = d_s$ and

$$\begin{aligned} C_{T, v_1} &= C_{T, v_{1s}} \oplus \cdots \oplus C_{T, v_{ts}} \\ \vdots \quad & \quad \vdots \\ C_{T, v_s} &= C_{T, v_{11}} \oplus \cdots \oplus C_{T, v_{t1}}. \end{aligned}$$

Consequently

$$V = C_{T, v_1} \oplus \cdots \oplus C_{T, v_s}.$$

**DEFINITION 5.2**

*Polynomials $d_1, \ldots, d_s$ satisfying the conditions of the above theorem are called* **invariant factors** *of $T$.*

*There is a similar definition for matrices: if $A \in M_{n \times n}(F)$ is similar to a direct sum*

$$\bigoplus_{k=1}^{s} C(d_k),$$

*where $d_1, \ldots, d_s$ are non–constant monic polynomials in $F[x]$ such that $d_k$ divides $d_{k+1}$ for $1 \leq k \leq s-1$, then $d_1, \ldots, d_s$ are called* **invariant factors** *of $A$. So the invariant factors of $A$ are the invariant factors of $T_A$.*

**THEOREM 5.9**

*The invariant factors of a linear transformation $T : V \to V$ are uniquely defined by $T$.*

PROOF

Reverse the construction in the proof of the above theorem using Question 15(a) of Problem Sheet 5, thereby recapturing the rectangular array of elementary divisors, which in turn is uniquely determined by $T$.

**EXAMPLE 5.1**

*Suppose $T : V \to V$ has elementary divisors*

$$p_1^2,\ p_1^3,\ p_1^3;\ p_2,\ p_2^2,\ p_2^2,\ p_2^4;\ p_3,\ p_3,\ p_3^4,\ p_3^5,\ p_3^5.$$

*Form the rectangular array*

| | | | | |
|---|---|---|---|---|
| 1 | 1 | $p_1^2$ | $p_1^3$ | $p_1^3$ |
| 1 | $p_2$ | $p_2^2$ | $p_2^2$ | $p_2^4$ |
| $p_3$ | $p_3$ | $p_3^4$ | $p_3^5$ | $p_3^5$ |

*Then the invariant factors of $T$ are obtained by respectively multiplying along columns:*

$$\begin{aligned} d_1 &= p_3 \\ d_2 &= p_2 p_3 \\ d_3 &= p_1^2 p_2^2 p_3^4 \\ d_4 &= p_1^3 p_2^2 p_3^5 \\ d_5 &= p_1^3 p_2^4 p_3^5. \end{aligned}$$

**THEOREM 5.10**

*If $d_1, \ldots, d_s$ are the invariant factors of $T : V \to V$, then*

*(i)* $m_T = d_s$;

*(ii)* $\mathrm{ch}_T = d_1 \cdots d_s$.

PROOF

Suppose $B = [T]_\beta^\beta = \bigoplus_{k=1}^s C(d_k)$ is the canonical form corresponding to the invariant factors $d_1, \ldots, d_s$ of $T$. Then

$$\begin{aligned} m_T = m_B &= \operatorname{lcm}(m_{C(d_1)}, \ldots, m_{C(d_s)}) \\ &= \operatorname{lcm}(d_1, \ldots, d_s) = d_s. \end{aligned}$$

Also

$$\operatorname{ch}_T = \operatorname{ch}_B = \prod_{k=1}^s \operatorname{ch}_{C(d_k)} = \prod_{k=1}^s d_k.$$

We shall soon see that the invariant factors of a linear transformation or matrix are of independent interest. For example the invariant factors allow us to calculate the dimension of the vector space $Z_{L,M}$ consisting of all linear transformations $N : U \to V$ which satisfy the equation $MN = NL$, where $L : U \to U$ and $M : V \to V$ are given linear transformations over $F$.

It turns out that there is a more direct way of finding the invariant factors of $T$. To introduce this algorithm, we need to discuss an interesting equivalence relation on $M_{m\times n}(F[x])$, which in turn leads to the so-called **Smith canonical form** of a matrix over $F[x]$.

# 6 The Smith Canonical Form

## 6.1 Equivalence of Polynomial Matrices

**DEFINITION 6.1**

*A matrix $P \in M_{n\times n}(F[x])$ is called a* **unit** *in $M_{n\times n}(F[x])$ if $\exists\ Q \in M_{n\times n}(F[x])$ such that*

$$PQ = I_n.$$

*Clearly if $P$ and $Q$ are units, so is $PQ$.*

**THEOREM 6.1**

*A matrix $P \in M_{n\times n}(F[x])$ is a unit in $M_{n\times n}(F[x])$ if and only if $\det P = c$, where $c \in F$ and $c \neq 0$.*

proof

"only if". Suppose $P$ is a unit. Then $PQ = I_n$ and

$$\det PQ = \det P\ \det Q = \det I_n = 1.$$

However $\det P$ and $\det Q$ belong to $F[x]$, so both are in fact non–zero elements of $F$.

"if". Suppose $P \in M_{n\times n}(F[x])$ satisfies $\det P = c$, where $c \in F$ and $c \neq 0$. Then

$$P \operatorname{adj} P = (\det P) I_n = c I_n.$$

Hence $PQ = I_n$, where $Q = c^{-1} \operatorname{adj} P \in M_{n\times n}(F[x])$. Hence $P$ is a unit in $M_{n\times n}(F[x])$.

**EXAMPLE 6.1**

$P = \begin{bmatrix} 1+x & -x \\ x & 1-x \end{bmatrix} \in M_{2\times 2}(F[x])$ *is a unit, as* $\det P = 1$.

**THEOREM 6.2**

*Elementary row matrices in $M_{n\times n}(F[x])$ are units:*

*(i) $E_{ij}$: interchange rows $i$ and $j$ of $I_n$;*

*(ii) $E_i(t)$: multiply row $i$ of $I_n$ by $t \in F$, $t \neq 0$;*

*(iii) $E_{ij}(f)$: add $f$ times row $j$ of $I_n$ to row $i$, $f \in F[x]$.*

*In fact* $\det E_{ij} = -1;\ \det E_i(t) = t;\ \det E_{ij}(f) = 1$.

*Similarly for elementary column matrices in $M_{n\times n}(F[x])$:*

$$F_{ij},\ F_i(t),\ F_{ij}(f).$$

REMARK: It follows that a product of elementary matrices in $M_{n\times n}(F[x])$ is a unit. Later we will be able to prove that the converse is also true.

**DEFINITION 6.2**

*Let $A$, $B \in M_{m\times n}(F[x])$. Then $A$ is equivalent to $B$ over $F[x]$ if units $P \in M_{m\times m}(F[x])$ and $Q \in M_{n\times n}(F[x])$ exist such that*

$$PAQ = B.$$

**THEOREM 6.3**

*Equivalence of matrices over $F[x]$ defines an equivalence relation on $M_{m\times n}(F[x])$.*

### 6.1.1 Determinantal Divisors

**DEFINITIONS 6.1**

*Let $A \in M_{m\times n}(F[x])$. Then for $1 \le k \le \min(m, n)$, let $d_k(A)$ denote the* gcd *of all $k \times k$ minors of $A$.*

*$d_k(A)$ is sometimes called the $k^{th}$* **determinantal divisor of** *$A$.*

**Note**: $\gcd(f_1, \ldots, f_n) \neq 0 \Leftrightarrow$ *at least one of $f_1, \ldots, f_n$ is non–zero.*

*$\rho(A)$, the* **determinantal rank** *of $A$, is defined to be the largest integer $r$ for which there exists a non–zero $r \times r$ minor of $A$.*

**THEOREM 6.4**

*For $1 \le k \le \rho(A)$, we have $d_k(A) \neq 0$. Also $d_k(A)$ divides $d_{k+1}(A)$ for $1 \le k \le \rho(A) - 1$.*

proof

Let $r = \rho(A)$. Then there exists an $r \times r$ non–zero minor and hence $d_r(A) \neq 0$. Then because each $r \times r$ minor is a linear combination over $F[x]$ of $(r-1)\times(r-1)$ minors of $A$, it follows that some $(r-1)\times(r-1)$ minor of $A$ is also non–zero and hence $d_{r-1}(A) \neq 0$; also $d_{r-1}(A)$ divides each minor of size $r-1$ and consequently divides each minor of size $r$; hence $d_{r-1}(A)$ divides $d_r(A)$, the gcd of all minors of size $r$. This argument can be repeated with $r$ replaced by $r-1$ and so on.

**THEOREM 6.5**

*Let $A$, $B \in M_{m\times n}(F[x])$. Then if $A$ is equivalent to $B$ over $F[x]$, we have*

*(i)* $\rho(A) = \rho(B) = r$;

*(ii)* $d_k(A) = d_k(B)$ *for* $1 \leq k \leq r$.

proof

Suppose $PAQ = B$, where $P$ and $Q$ are units. First consider $PA$. The rows of $PA$ are linear combinations over $F[x]$ of the rows of $A$, so it follows that each $k \times k$ minor of $PA$ is a linear combination of the $k \times k$ minors of $A$. Similarly each column of $(PA)Q$ is a linear combinations over $F[x]$ of the columns of $PA$, so it follows that each $k \times k$ minor of $B = (PA)Q$ is a linear combination over $F[x]$ of the $k \times k$ minors of $PA$ and consequently of the $k \times k$ minors of $A$.

It follows that all minors of $B$ with size $k > \rho(A)$ must be zero and hence $\rho(B) \leq \rho(A)$. However $B$ is equivalent to $A$, so we deduce that $\rho(A) \leq \rho(B)$ and hence $\rho(A) = \rho(B)$.

Also $d_k(B)$ is a linear combination over $F[x]$ of all $k \times k$ minors of $B$ and hence of all $k \times k$ minors of $A$. Hence $d_k(A)|d_k(B)$ and by symmetry, $d_k(B)|d_k(A)$. Hence $d_k(A) = d_k(B)$ if $1 \leq k \leq r$.

## 6.2 Smith Canonical Form

**THEOREM 6.6 (Smith canonical form)**

*Every non–zero matrix* $A \in M_{m\times n}(F[x])$ *with* $r = \rho(A)$ *is equivalent to a matrix of the form*

$$D = \begin{bmatrix} f_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & f_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & f_r & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} = PAQ$$

*where* $f_1, \ldots, f_r \in F[x]$ *are monic,* $f_k|f_{k+1}$ *for* $1 \leq k \leq r-1$, $P$ *is a product of elementary row matrices, and* $Q$ *is a product of elementary column matrices.*

**DEFINITION 6.3**

*The matrix* $D$ *is said to be in* **Smith canonical form**.

proof

This is presented in the form of an algorithm which is in fact used by CMAT to find unit matrices $P$ and $Q$ such that $PAQ$ is in Smith canonical form.

Our account is based on that in the book "Rings, Modules and Linear Algebra," by B. Hartley and T.O. Hawkes.

We describe a sequence of elementary row and column operations over $F[x]$, which when applied to a matrix $A$ with $a_{11} \neq 0$ either yields a matrix $C$ of the form

$$C = \begin{bmatrix} f_1 & 0 \cdots 0 \\ 0 & \\ \vdots & C^* \\ 0 & \end{bmatrix}$$

where $f_1$ is monic and divides every element of $C^*$, or else yields a matrix $B$ in which $b_{11} \neq 0$ and

$$\deg b_{11} < \deg a_{11}. \tag{28}$$

Assuming this, we start with our non–zero matrix $A$. By performing suitable row and column interchanges, we can assume that $a_{11} \neq 0$. Now repeatedly perform the algorithm mentioned above. Eventually we must reach a matrix of type $C$, otherwise we would produce an infinite strictly decreasing sequence of non–negative integers by virtue of inequalities of type (28).

On reaching a matrix of type $C$, we stop if $C^* = 0$. Otherwise we perform the above argument on $C^*$ and so on, leaving a trail of diagonal elements as we go.

Two points must be made:

(i) Any elementary row or column operation on $C^*$ corresponds to an elementary operation on $C$, which does not affect the first row or column of $C$.

(ii) Any elementary operation on $C^*$ gives a new $C^*$ whose new entries are linear combinations over $F[x]$ of the old ones; consequently these new entries will still be divisible by $f_1$.

Hence in due course we will reach a matrix $D$ which is in Smith canonical form.

We now detail the sequence of elementary operations mentioned above. Case 1. $\exists\ a_{1j}$ in row 1 with $a_{11}$ not dividing $a_{1j}$. Then

$$a_{1j} = a_{11}q + b,$$

by Euclid's division theorem, where $b \neq 0$ and $\deg b < \deg a_{11}$. Subtract $q$ times column 1 from column $j$ and then interchange columns 1 and $j$. This yields a matrix of type $B$ mentioned above.

Case 2. $\exists\ a_{i1}$ in column 1 with $a_{11}$ not dividing $a_{i1}$. Proceed as in Case 1, operating on rows rather than columns, again reaching a matrix of type $B$.
Case 3. Here $a_{11}$ divides every element in the first row and first column. Then by subtracting suitable multiples of column 1 from the other columns, we can replace all the entries in the first row other than $a_{11}$ by 0. Similarly for the first column. We then have a matrix of the form

$$E = \begin{bmatrix} e_{11} & 0 \cdots 0 \\ 0 & \\ \vdots & E^* \\ 0 & \end{bmatrix}.$$

If $e_{11}$ divides every element of $E^*$, we have reached a matrix of type $C$. Otherwise $\exists\ e_{ij}$ not divisible by $e_{11}$. We then add row $i$ to row 1, thereby reaching Case 1.

**EXAMPLE 6.2**

*(of the Smith Canonical Form)*

$$A = \begin{bmatrix} 1+x^2 & x \\ x & 1+x \end{bmatrix}$$

*We want $D = PAQ$ in Smith canonical form. So we construct the augmented matrix*

| | | *work on rows* ↓ | | | | *work on columns* ↓ | |
|---|---|---|---|---|---|---|---|
| | | 1 | 0 | $1+x^2$ | $x$ | 1 | 0 |
| | | 0 | 1 | $x$ | $1+x$ | 0 | 1 |
| $R_1 \to R_1 - xR_2$ | $\Rightarrow$ | 1 | $-x$ | 1 | $-x^2$ | 1 | 0 |
| | | 0 | 1 | $x$ | $1+x$ | 0 | 1 |
| $C_2 \to C_2 + x^2C_1$ | $\Rightarrow$ | 1 | $-x$ | 1 | 0 | 1 | $x^2$ |
| | | 0 | 1 | $x$ | $1+x+x^3$ | 0 | 1 |
| $R_2 \to R_2 - xR_1$ | $\Rightarrow$ | 1 | $-x$ | 1 | 0 | 1 | $x^2$ |
| | | $-x$ | $1+x^2$ | 0 | $1+x+x^3$ | 0 | 1 |
| | | | ↑ $P$ | | ↑ $D$ | | ↑ $Q$ |

*Invariants are $f_1 = 1$, $f_2 = 1 + x + x^3$. Note also*

$$f_1 = d_1(A), \qquad f_2 = \frac{d_2(A)}{d_1(A)}.$$

### 6.2.1 Uniqueness of the Smith Canonical Form

**THEOREM 6.7**

*Every matrix $A \in M_{m\times n}(F[x])$ is equivalent to precisely one matrix is Smith canonical form.*

proof Suppose $A$ is equivalent to a matrix $B$ in Smith canonical form. That is,

$$B = \left[\begin{array}{ccc|c} f_1 & & & \\ & \ddots & & 0 \\ & & f_r & \\ \hline & 0 & & 0 \end{array}\right] \quad \text{and} \quad f_1 \mid f_2 \mid \cdots \mid f_r.$$

Then $r = \rho(A)$, the determinantal rank of $A$. But if $1 \le k \le r$,

$$d_k(A) = d_k(B) = f_1 f_2 \dots f_k$$

and so the $f_i$ are uniquely determined by

$$\begin{aligned} f_1 &= d_1(A) \\ f_2 &= \frac{d_2(A)}{d_1(A)} \\ &\vdots \\ f_r &= \frac{d_r(A)}{d_{r-1}(A)}. \end{aligned}$$

## 6.3 Invariant factors of a polynomial matrix

**DEFINITION 6.4**

*The polynomials $f_1, \dots, f_r$ in the Smith canonical form of $A$ are called the* **invariant factors** *of $A$.*[3]

**Note***: CMAT calls the invariant factors of $xI - B$, where $B \in M_{n\times n}(F)$, the "similarity invariants" of $B$.*

We next find these similarity invariants. They are

$$\underbrace{1, 1, \dots, 1}_{n-s}, d_1, \dots, d_s$$

where $d_1, \dots, d_s$ are what earlier called the invariant factors of $T_B$.

[3]**NB.** This is a slightly different, though similar, form of "invariant factor" to that we met a short while ago.

**LEMMA 6.1**

*The Smith canonical form of $xI_n - C(d)$ where $d$ is a monic polynomial of degree $n$ is*

$$\operatorname{diag}(\underbrace{1,\ldots,1}_{n-1},d).$$

proof Let $d = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in F[x]$, so

$$xI_n - C(d) = \begin{bmatrix} x & 0 & & & a_0 \\ -1 & x & \cdots & & a_1 \\ 0 & -1 & & & a_2 \\ & \vdots & \ddots & & \vdots \\ & & & x & a_{n-2} \\ 0 & & \cdots & -1 & x + a_{n-1} \end{bmatrix}.$$

Now use the row operation

$$R_1 \to R_1 + xR_2 + x^2R_3 + \cdots + x^{n-1}R_n$$

to obtain

$$\begin{bmatrix} 0 & 0 & & & d \\ -1 & x & \cdots & & a_1 \\ 0 & -1 & & & a_2 \\ & \vdots & \ddots & & \vdots \\ & & & x & a_{n-2} \\ 0 & & \cdots & -1 & x + a_{n-1} \end{bmatrix}$$

(think about it!) and then column operations

$$C_2 \to C_2 + xC_1, \ldots, C_{n-1} \to C_{n-1} + xC_{n-2}$$

and then

$$C_n \to C_n + a_1C_1 + a_2C_2 + \cdots + a_{n-2}C_{n-2} + (x + a_{n-1})C_{n-1}$$

yielding

$$\begin{bmatrix} 0 & 0 & \ldots & 0 & d \\ -1 & 0 & & & 0 \\ 0 & -1 & & & \\ & & \ddots & & \vdots \\ 0 & & \cdots & -1 & 0 \end{bmatrix}.$$

Trivially, elementary operations now form the matrix

$$\operatorname{diag}(\underbrace{1,\ldots,1}_{n-1},d).$$

**THEOREM 6.8**

*Let $B \in M_{n\times n}(F)$. Then if the invariant factors of $B$ are $d_1,\ldots,d_s$, then the invariant factors of $xI_n - B$ are*

$$\underbrace{1,\ldots,1}_{n-s},d_1,d_2,\ldots,d_s.$$

proof There exists non-singular $P \in M_{n\times n}(F)$ such that

$$P^{-1}BP = \bigoplus_{k=1}^{s} C(d_k).$$

Then

$$\begin{aligned} P^{-1}(xI_n - B)P &= xI_n - \bigoplus_{k=1}^{s} C(d_k) \\ &= \bigoplus_{k=1}^{s} (xI_{m_k} - C(d_k)) \qquad \text{where } m_k = \deg d_k. \end{aligned}$$

But by the lemma, each $xI_{m_k} - C(d_k)$ is equivalent over $F[x]$ to $\operatorname{diag}(1,\ldots,1,d_k)$ and hence $xI_n - B$ is equivalent to

$$\bigoplus_{k=1}^{s} \operatorname{diag}(1,\ldots,1,d_k) \sim \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & d_1 & & \\ & & & & \ddots & \\ & & & & & d_s \end{bmatrix}.$$

**EXAMPLE 6.3**

*Find the invariant factors of*

$$B = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \in M_{4\times 4}(\mathbb{Q})$$

*by finding the Smith canonical form of* $xI_4 - B$.

**Solution:**

$$xI_4 - B = \begin{bmatrix} x-2 & 0 & 0 & 0 \\ 1 & x-1 & 0 & 0 \\ 0 & 1 & x & 1 \\ -1 & -1 & -1 & x-2 \end{bmatrix}$$

*We start off with the row operations*

$$\begin{array}{lcl} R_1 & \to & R_1 - (x-2)R_2 \\ R_1 & \leftrightarrow & R_2 \\ R_4 & \to & R_4 + R_1 \end{array}$$

*and get*

$$\begin{bmatrix} 1 & x-1 & 0 & 0 \\ 0 & -(x-1)(x-2) & 0 & 0 \\ 0 & 1 & x & 1 \\ 0 & x-2 & -1 & x-2 \end{bmatrix}$$

$$\text{(column ops.)} \quad \Rightarrow \quad \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & -(x-1)(x-2) & 0 & 0 \\ 0 & & 1 & x \quad 1 \\ 0 & & x-2 & -1 \quad x-2 \end{array}\right]$$

$$\Rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & x & 1 \\ 0 & -(x-1)(x-2) & 0 & 0 \\ 0 & x-2 & -1 & x-2 \end{bmatrix}$$

$$\Rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & x & 1 \\ 0 & 0 & x(x-1)(x-2) & (x-1)(x-2) \\ 0 & 0 & -1-x(x-2) & 0 \\ & & \{= -(x-1)^2\} & \end{bmatrix}$$

$$\Rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x(x-1)(x-2) & (x-1)(x-2) \\ 0 & 0 & -(x-1)^2 & 0 \end{bmatrix}.$$

*Now, for brevity, we work just on the* $2 \times 2$ *block in the bottom right corner:*

$$\Rightarrow \quad \begin{bmatrix} (x-1)(x-2) & x(x-1)(x-2) \\ 0 & -(x-1)^2 \end{bmatrix}$$

$$
\begin{aligned}
C_2 \to C_2 - xC_1 \quad &\Rightarrow \quad \begin{bmatrix} (x-1)(x-2) & 0 \\ 0 & -(x-1)^2 \end{bmatrix} \\
R_1 \to R_1 + R_2 \quad &\Rightarrow \quad \begin{bmatrix} (x-1)(x-2) & (x-1)^2 \\ 0 & -(x-1)^2 \end{bmatrix} \\
C_2 \to C_2 - C_1 \quad &\Rightarrow \quad \begin{bmatrix} (x-1)(x-2) & x-1 \\ 0 & -(x-1)^2 \end{bmatrix} \\
C_1 \leftrightarrow C_2 \quad &\Rightarrow \quad \begin{bmatrix} x-1 & (x-1)(x-2) \\ -(x-1)^2 & 0 \end{bmatrix} \\
C_2 \to C_2 - (x-2)C_1 \quad &\Rightarrow \quad \begin{bmatrix} x-1 & 0 \\ -(x-1)^2 & (x-2)(x-1)^2 \end{bmatrix} \\
R_2 \to R_2 + (x-1)R_1 \quad &\Rightarrow \quad \begin{bmatrix} x-1 & 0 \\ 0 & (x-2)(x-1)^2 \end{bmatrix}
\end{aligned}
$$

*and here we stop, as we have a matrix in Smith canonical form. Thus*

$$
xI_4 - B \sim \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & x-1 & \\ & & & (x-1)^2(x-2) \end{bmatrix}
$$

*so the invariant factors of $B$ are the non-trivial ones of $xI_4 - B$, i.e.*

$$(x-1) \quad \text{and} \quad (x-1)^2(x-2).$$

*Also, the elementary divisors of $B$ are*

$$(x-1),\ (x-1)^2 \text{ and } (x-2)$$

*so the Jordan canonical form of $B$ is*

$$J_2(1) \oplus J_1(1) \oplus J_1(2).$$

**THEOREM 6.9**

*Let $A, B \in M_{n\times n}(F)$. Then $A$ is similar to $B$*

$$
\begin{aligned}
&\Leftrightarrow \quad xI_n - A \text{ is equivalent to } xI_n - B \\
&\Leftrightarrow \quad xI_n - A \text{ and } xI_n - B \text{ have the same} \\
&\qquad\qquad \text{Smith canonical form.}
\end{aligned}
$$

proof

$\Rightarrow$ Obvious. If $P^{-1}AP = B$, $P \in M_{n\times n}(F)$ then

$$\begin{aligned} P^{-1}(xI_n - A)P &= xI_n - P^{-1}AP \\ &= xI_n - B. \end{aligned}$$

$\Leftarrow$ If $xI_n - A$ and $xI_n - B$ are equivalent over $F[x]$, then they have the same invariant factors and so have the same non-trivial invariant factors. That is, $A$ and $B$ have the same invariant factors and hence are similar.

**Note**: It is possible to start from $xI_n - A$ and find $P \in M_{n\times n}(F)$ such that

$$P^{-1}AP = \bigoplus_{k=1}^{s} C(d_k)$$

where

$$P_1(xI_n - B)Q_1 = \text{ diag }(1, \ldots, 1, d_1, \ldots, d_s).$$

(See Perlis, Theory of matrices, p. 144, Corollary 8–1 and p. 137, Theorem 7–9.)

**THEOREM 6.10**

*Every unit in $M_{n\times n}(F[x])$ is a product of elementary row and column matrices.*

Proof: Problem sheet 7, Question 12.

# 7 Various Applications of Rational Canonical Forms

## 7.1 An Application to commuting transformations

**THEOREM 7.1 (Cecioni 1908, Frobenius 1910)**

*Let $L : U \mapsto U$ and $M : V \mapsto V$ be given LTs. Then the vector space $Z_{L,M}$ of all LTs $N : U \mapsto V$ satisfying*

$$MN = NL$$

*has dimension*

$$\sum_{k=1}^{s} \sum_{l=1}^{t} \deg \gcd(d_k, D_l),$$

*where $d_1, \ldots, d_s$ and $D_1, \ldots, D_t$ are the invariant factors of $L$ and $M$ respectively.*

**COROLLARY 7.1**

*Now take $U = V$ and $L = M$. Then $Z_{L,L}$ the vector space of LTs satisfying*

$$NL = LN,$$

*has dimension*

$$\sum_{k=1}^{s} (2s - 2k + 1) \deg d_k.$$

proof Omitted, but here's a hint:

$$\gcd(d_k, d_l) = \begin{cases} d_k \text{ if } k \leq l \text{ ; i.e. if } d_k \mid d_l \\ d_l \text{ if } k > l \text{ ; i.e. if } d_l \mid d_k. \end{cases}$$

**N.B.** Let $P_L$ be the vector space of all LTs of the form

$$f(L) : U \mapsto U \qquad f \in F[x].$$

Then $P_L \subseteq Z_{L,L}$ and we have the following...

**THEOREM 7.2**

$$P_L = Z_{L,L} \Leftrightarrow m_L = \mathrm{ch}_L\,.$$

proof First note that $\dim P_L = \deg m_L$ as

$$I_V, L, \ldots, L^{\deg m_L - 1}$$

form a basis for $P_L$. So, since $P_L \subseteq Z_{L,L}$ we have

$$\begin{aligned} P_L = Z_{L,L} &\Leftrightarrow \dim P_L = \dim Z_{L,L} \\ &\Leftrightarrow \deg m_L = \sum_{k=1}^{s} (2s - 2k + 1) \deg d_k \\ &\Leftrightarrow s = 1 \\ &\Leftrightarrow \mathrm{ch}_L = m_L. \end{aligned}$$

proof (a sketch) of Cecioni-Frobenius theorem.
We start with the invariant factor decompositions

$$U = \bigoplus_{k=1}^{s} C_{L,u_k} \qquad \text{and} \qquad V = \bigoplus_{l=1}^{t} C_{M,v_l}$$

where $m_{L,u_k} = d_k$ for $k = 1, \ldots, s$, and $m_{M,v_l} = D_l$ for $l = 1, \ldots, t$.
Let $MN = NL \ldots$

$$\begin{aligned} &\Rightarrow \quad M^n N = N L^n \qquad \forall n \geq 1 \\ &\Rightarrow \quad f(M)N = N f(L) \qquad \forall f \in F[x]. \end{aligned}$$

Define vectors $w_1, \ldots, w_s \in V$ by $w_k = N(u_k)$, and observe

$$\begin{aligned} d_k(M)(w_k) &= d_k(M)(N(u_k)) \\ &= N(d_k(L)(u_k)) \\ &= N(0) = 0. \end{aligned}$$

Then we have the
**Definition:** Let $W$ be the set of all $(w_1, \ldots, w_s)$ such that $w_1, \ldots, w_s \in V$ and

$$d_k(M)(w_k) = 0 \qquad \forall k = 1, \ldots, s.$$

We assert that $W$ is a vector space and the mapping

$$N \mapsto (w_1, \ldots, w_s)$$

is an isomorphism between $Z_{L,M}$ and $W$; proof is left as an exercise.

Now let

$$w_k = \sum_{l=1}^{t} c_{kl}(M)(v_l) \qquad k = 1, \ldots, s \text{ and } c_{kl} \in F[x].$$

**N.B.**

$$\begin{aligned} & f(M)(v_l) = g(M)(v_l) \text{ say} \\ \Leftrightarrow \quad & D_l \mid f - g. \end{aligned}$$

So, if we restrict $c_{kl}$ by the condition

$$\deg c_{kl} < \deg D_l \quad \text{if } c_{kl} \neq 0 \tag{29}$$

then the $c_{kl}$ are uniquely defined for each $k$.
Exercise: Now let

$$g_{kl} = \gcd(d_k, D_l).$$

Then from the condition $d_k(M)(w_k) = 0$, show that

$$\frac{D_l}{g_{kl}} \mid c_{kl} \tag{30}$$

i.e. that

$$c_{kl} = b_{kl}\frac{D_l}{g_{kl}} \qquad b_{kl} \in F[x]. \tag{31}$$

Then the matrices $[c_{kl}]$, where $c_{kl}$ satisfy (30), form a vector space (call it $X$) which is isomorphic to $W$.

Then in (31),

$$(29) \iff \deg b_{kl} < \deg g_{kl} \qquad \text{if } b_{kl} \neq 0.$$

Clearly then,

$$\dim X = \dim Z_{L,M} = \sum_{k=1}^{s}\sum_{l=1}^{t} \deg g_{kl}$$

as required.

**EXAMPLE 7.1**

*(of the vector space $X$, when $s = t = 2$)*

*Say*

$$[\deg g_{kl}] = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}.$$

*Then $X$ consists of all matrices of the form*

$$\begin{aligned}[c_{kl}] &= \begin{bmatrix} (a_0+a_1x)\cdot\frac{D_1}{g_{11}} & 0\cdot\frac{D_2}{g_{12}} \\ b_0\cdot\frac{D_1}{g_{21}} & (c_0+c_1x+c_2x^2)\cdot\frac{D_2}{g_{22}} \end{bmatrix} \\ &= a_0\begin{bmatrix} \frac{D_1}{g_{11}} & 0 \\ 0 & 0 \end{bmatrix} + a_1\begin{bmatrix} \frac{xD_1}{g_{11}} & 0 \\ 0 & 0 \end{bmatrix} + b_0\begin{bmatrix} 0 & 0 \\ \frac{D_1}{g_{21}} & 0 \end{bmatrix} + \cdots \end{aligned}$$

*... and so on.*

**EXAMPLE 7.2**

**The most general $3\times 3$ matrix which commutes with others.**

*Let $A \in M_{3\times 3}(\mathbb{Q})$ such that there exists non-singular $P \in M_{3\times 3}(\mathbb{Q})$ with*

$$\begin{aligned} P^{-1}AP &= C(x-1)\oplus C((x-1)^2) \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} = J, \text{ say,} \end{aligned}$$

*where $C(p)$ denotes the companion matrix of $p$, as usual.*

*Then $P = [u_1 \mid u_2 \mid T(u_2)]$ where $T = T_A$ and*

$$m_{T,u_1} = x-1, \qquad m_{T,u_2} = (x-1)^2.$$

*Also $V_3(\mathbb{Q}) = C_{T,u_1}\oplus C_{T,u_2}$.*

*Note that the invariant factors of $T$ are $(x-a)$ and $(x-1)^2$.*

*We find all $3\times 3$ matrices $B$ such that*

$$\begin{aligned} & BA = AB, \\ \text{i.e.} \quad & T_BT_A = T_AT_B. \end{aligned}$$

*Let $N = T_B$. Then $N$ must satisfy*

$$\begin{aligned} N(u_1) &= Bu_1 = c_{11}u_1 + c_{12}u_2 \qquad \text{and} \\ N(u_2) &= Bu_2 = c_{21}u_1 + c_{22}u_2 \qquad \text{where } c_{kl} \in \mathbb{Q}[x]. \end{aligned} \tag{32}$$

*Now*

$$[\deg \gcd(d_k, d_l)] = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

*so*

$$[c_{kl}] = \begin{bmatrix} a_0 & b_0(x-1) \\ c_0 & d_0+d_1x \end{bmatrix}$$

*where $a_0$ etc. $\in \mathbb{Q}$, so (32) gives*

$$\begin{aligned} Bu_1 &= a_0u_1 + b_0(x-1)u_2 \\ &= a_0u_1 - b_0u_2 + b_0T(u_2) \qquad (33) \\ Bu_2 &= c_0u_1 + (d_0 + d_1x)u_2 \\ &= c_0u_1 + d_0u_2 + d_1T(u_2). \qquad (34) \end{aligned}$$

*Noting that*

$$\begin{aligned} & m_{T,u_1} = x - 1 \Rightarrow T(u_1) = u_1 \\ \text{and} \quad & m_{T,u_2} = (x-1)^2 = x^2 - 2x + 1 \Rightarrow T^2(u_2) = 2T(u_2) - u_2, \end{aligned}$$

*we have from (34) that*

$$\begin{aligned} T(Bu_2) &= c_0T(u_1) + d_0T(u_2) + d_1T^2(u_2) \\ &= c_0u_1 - d_1u_2 + (d_0 + 2d_1)T(u_2). \end{aligned}$$

*In terms of matrices,*

$$\begin{aligned} B[u_1|u_2|T(u_2)] &= [u_1|u_2|T(u_2)] \begin{bmatrix} a_0 & c_0 & c_0 \\ -b_0 & d_0 & -d_1 \\ b_0 & d_1 & d_0 + 2d_1 \end{bmatrix} \\ \text{i.e.} \quad & BP = PK, \quad \text{say} \\ \text{or} \quad & B = PKP^{-1}. \end{aligned}$$

*This gives the most general matrix $B$ such that*

$$BA = AB.$$

**Note***: $BA = AB$ becomes*

$$\begin{aligned} PKP^{-1}\,PJP^{-1} &= PJP^{-1}\,PKP^{-1} \\ \Leftrightarrow \quad KJ &= JK. \end{aligned}$$

## 7.2 Tensor products and the Byrnes-Gauger theorem

We next apply the Cecioni-Frobenius theorem to derive a third criterion for deciding whether or not two matrices are similar.

**DEFINITION 7.1**

*(Tensor or Kronecker product)*

*If $A \in M_{m_1 \times n_1}(F)$ and $B \in M_{m_2 \times n_2}(F)$ we define*

$$A \otimes B = \left[\begin{array}{c|c|c} a_{11}B & a_{12}B & \cdots \\ \hline a_{21}B & a_{22}B & \cdots \\ \hline \vdots & \vdots & \ddots \end{array}\right] \in M_{m_1 m_2 \times n_1 n_2}(F).$$

*In terms of elements,*

$$(A \otimes B)_{(i,j),(k,l)} = a_{ij} b_{kl}$$

*—the element at the intersection of the $i$-th row block, $k$-th row sub-block, and the $j$-th column block, $l$-th column sub-block.*[4]

**EXAMPLE 7.3**

$$A \otimes I_p = \left[\begin{array}{ccc|c} a_{11} & & & \\ & \ddots & & \cdots \\ & & a_{11} & \\ \hline a_{21} & & & \\ & \ddots & & \cdots \\ & & a_{21} & \\ \hline & & & \\ & \vdots & & \ddots \end{array}\right],$$

$$I_p \otimes A = \left[\begin{array}{c|c|c} A & 0 & \cdots \\ \hline 0 & A & \cdots \\ \hline \vdots & \vdots & \ddots \end{array}\right].$$

*(Tensor-product-taking is obviously far from commutative!)*

### 7.2.1 Properties of the tensor product of matrices

(i) $(tA) \otimes B = A \otimes (tB) = t(A \otimes B),\ t \in F$;

(ii) $A \otimes B = 0 \Leftrightarrow A = 0$ or $B = 0$;

(iii) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$;

(iv) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$;

(v) $(B + C) \otimes D = (B \otimes D) + (C \otimes D)$;

[4] That is, the $((i-1)m_2 + k, (j-1)n_2 + l)$-th element in the tensor product is $a_{ij}b_{kl}$.

(vi) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$;

(vii) $(B \oplus C) \otimes D = (B \otimes D) \oplus (C \otimes D)$;

(viii) $P(A \otimes (B \oplus C))P^{-1} = (A \otimes B) \oplus (A \otimes C)$ for a suitable row permutation matrix $P$;

(ix) $\det(A \otimes B) = (\det A)^n (\det B)^m$ if $A$ is $m \times m$ and $B$ is $n \times n$;

(x) Let $f(x, y) = \sum_{i=0}^{m} \sum_{j=0}^{n} c_{ij} x^i y^j \in F[x, y]$ be a polynomial in $x$ and $y$ over $F$ and define

$$f(A; B) = \sum_{i=0}^{m} \sum_{j=0}^{n} c_{ij} (A^i \otimes B^j).$$

Then if $\mathrm{ch}_A = \prod_{k=1}^{s} (x - \lambda_k)$ and $\mathrm{ch}_B = \prod_{l=1}^{t} (x - \mu_l)$, we have

$$\mathrm{ch}_{f(A;B)} = \prod_{k=1}^{s} \prod_{l=1}^{t} (x - f(\lambda_k, \mu_l));$$

(xi) Taking $f(x, y) = xy$ gives

$$\mathrm{ch}_{A \otimes B} = \prod_{k=1}^{s} \prod_{l=1}^{t} (x - \lambda_k \mu_l);$$

(xii) Taking $f(x, y) = x - y$ gives

$$\mathrm{ch}_{(A \otimes I_n - I_m \otimes B)} = \prod_{k=1}^{s} \prod_{l=1}^{t} (x - (\lambda_k - \mu_l));$$

**Remark:** (ix) can be proved using the uniqueness theorem for alternating $m$–linear functions met in MP174; (x) follows from the the equations

$$P^{-1} A P = J_1 \quad \text{and} \quad Q^{-1} B Q = J_2,$$

where $J_1$ and $J_2$ are the Jordan forms of $A$ and $B$, respectively. Then $J_1$ and $J_2$ are lower triangular matrices with the eigenvalues $\lambda_k$, $1 \leq k \leq m$ and $\mu_l$, $1 \leq l \leq n$ of $A$ and $B$ as diagonal elements.

Then

$$P^{-1} A^i P = J_1^i \quad \text{and} \quad Q^{-1} B^j Q = J_2^j$$

and more generally

$$(P \otimes Q)^{-1} \sum_{i=0}^{s} \sum_{j=0}^{t} c_{ij}(A^i \otimes B^j)(P \otimes Q) = \sum_{i=0}^{s} \sum_{j=0}^{t} c_{ij}(J_1^i \otimes J_2^j).$$

The matrix on the right–hand side is lower triangular and has diagonal elements

$$f(\lambda_k, \mu_l),\ 1 \leq k \leq m,\ 1 \leq l \leq n.$$

**THEOREM 7.3**

*Let $\beta$ be the standard basis for $M_{m\times n}(F)$—i.e. the basis consisting of the matrices*

$$E_{11}, \ldots\ldots, E_{mn}$$

*and $\gamma$ be the standard basis for $M_{p\times n}(F)$.*

*Let $A$ be $p \times m$, and*

$$T_1 : M_{m\times n}(F) \mapsto M_{p\times n}(F)$$

*be defined by $T_1(X) = AX$. Then*

$$[T_1]_\beta^\gamma = A \otimes I_n.$$

*Similarly if $B$ is $n \times p$, and*

$$T_2 : M_{m\times n}(F) \mapsto M_{m\times p}(F)$$

*is defined by $T_2(Y) = YB$, then*

$$[T_2]_\beta^\delta = A \otimes I_n$$

*(where $\delta$ is the standard basis for $M_{m\times p}(F)$).*

proof Left for the intrepid reader. A hint:

$$E_{ij}E_{kl} = \begin{cases} 0 & \text{if } j \neq k, \\ E_{il} & \text{if } j = k \end{cases}$$

**COROLLARY 7.2**

*Let $A$ be $m \times m$,*
*$B$ be $n \times n$,*
*$X$ be $m \times n$, and*

$$T : M_{m\times n}(F) \mapsto M_{m\times n}(F)$$

*be defined by* $T(X) = AX - XB$.

*Then*

$$[T]_\beta^\beta = A \otimes I_n - I_m \otimes B^t,$$

*where* $\beta$ *is the standard basis for* $M_{m\times n}(F)$.

**DEFINITION 7.2**

*For brevity in the coming theorems, we define*

$$\nu_{A,B} = \nu(A \otimes I_n - I_m \otimes B^t)$$

*where* $A$ *is* $m \times m$ *and* $B$ *is* $n \times n$.

**THEOREM 7.4**

$$\begin{aligned} \nu_{A,B} &= \nu(A \otimes I_n - I_m \otimes B^t) \\ &= \sum_{k=1}^{s}\sum_{l=1}^{t} \deg \gcd(d_k, D_l) \end{aligned}$$

*where*

$$\begin{aligned} &d_1 \mid d_2 \mid \cdots \mid d_s \quad \text{and} \\ &D_1 \mid D_2 \mid \cdots \mid D_t \end{aligned}$$

*are the invariant factors of* $A$ *and* $B$ *respectively.*

proof With the transformation $T$ from corollary 7.2 above, we note that

$$\begin{aligned} \nu_{A,B} &= \text{nullity}\, T \\ &= \dim\{\, X \in M_{m\times n}(F) \mid AX = XB \,\} \\ &= \dim\{\, N \in \text{Hom}\,(V_n(F), V_m(F)) \mid T_A N = N T_B \,\} \end{aligned}$$

and the Cecioni-Frobenius theorem gives the result.

**LEMMA 7.1 (Byrnes-Gauger)**

*(This is needed in the proof of the Byrnes-Gauger theorem following.)*

*Suppose we have two monotonic increasing integer sequences:*

$$\begin{cases} m_1 \le m_2 \le \cdots \le m_s \\ n_1 \le n_2 \le \cdots \le n_s \end{cases} \quad \text{and}$$

*Then*

$$\sum_{k=1}^{s}\sum_{l=1}^{s}\{\min(m_k,m_l)+\min(n_k,n_l)-2\min(m_k,n_l)\}\geq 0.$$

*Further, equality occurs iff the sequences are identical.*

proof

**Case 1**: $k=l$.

The terms to consider here are of the form

$$m_k+n_k-2\min(m_k,n_k)$$

which is obviously $\geq 0$. Also, the term is equal to zero iff $m_k=n+k$.

**Case 2**: $k\neq l$; without loss of generality take $k<l$.

Here we pair the off-diagonal terms $(k,l)$ and $l,k$.

$$\begin{aligned}
&\{\min(m_k,m_l)+\min(n_k,n_l)-2\min(m_k,n_l)\}\\
&\qquad+\{\min(m_l,m_k)+\min(n_l,n_k)-2\min(m_l,n_k)\}\\
=\;&\{m_k+n_l-2\min(m_k,n_l)\}+\{m_l+n_k-2\min(m_l,n_k)\}\\
\geq\;&0,\qquad\text{obviously.}
\end{aligned}$$

Since the sum of the diagonal terms and the sum of the pairs of sums of off-diagonal terms are non-negative, the sum is non-negative. Also, if the sum is zero, so must be the sum along the diagonal terms, making

$$m_k=n_k\qquad\forall k.$$

**THEOREM 7.5 (Byrnes-Gauger)**

*If $A$ is $m\times m$ and $B$ is $n\times n$ then*

$$\nu_{A,A}+\nu_{B,B}\geq 2\nu_{A,B}$$

*with equality if and only if $m=n$ and $A$ and $B$ are similar.*

proof

$$\begin{aligned}
&\nu_{A,A}+\nu_{B,B}-2\nu_{A,B}\\
=\;&\sum_{k_1=1}^{s}\sum_{k_2=1}^{s}\deg\gcd(d_{k_1},d_{k_2})+\sum_{l_1=1}^{t}\sum_{l_2=1}^{t}\deg\gcd(D_{l_1},D_{l_2})\\
&\qquad-2\sum_{k=1}^{s}\sum_{l=1}^{t}\deg\gcd(d_k,D_l).
\end{aligned}$$

We now extend the definitions of $d_1, \ldots, d_s$ and $D_1, \ldots, D_t$ by renaming them as follows, with $N = \max(s,t)$:

$$\begin{aligned} \underbrace{1, \ldots, 1}_{N-s}, d_1, \ldots, d_s &\mapsto f_1, \ldots, f_N \\ \text{and} \quad \underbrace{1, \ldots, 1}_{N-t}, D_1, \ldots, D_t &\mapsto F_1, \ldots, F_N. \end{aligned}$$

This is so we may rewrite the above sum of three sums as a single sum, viz:

$$\begin{aligned} \nu_{A,A} + \nu_{B,B} - 2\nu_{A,B} &= \sum_{k=1}^{N} \sum_{l=1}^{N} \{\deg \gcd(f_k, f_l) + \deg \gcd(F_k, F_l) \\ &\qquad -2 \deg \gcd(f_k, F_l)\}. \end{aligned} \tag{35}$$

We now let $p_1, \ldots, p_r$ be the distinct monic irreducibles in $m_A m_B$ and write

$$\left.\begin{array}{rcccc} f_k = & p_1^{a_{k1}} & p_2^{a_{k2}} & \ldots & p_r^{a_{kr}} \\ F_k = & p_1^{b_{k1}} & p_2^{b_{k2}} & \ldots & p_r^{b_{kr}} \end{array}\right\} \qquad 1 \le k \le N$$

where the sequences $\{a_{ki}\}_{i=1}^r$, $\{b_{ki}\}_{i=1}^r$ are monotonic increasing non-negative integers. Then

$$\begin{aligned} & \gcd(f_k, F_l) = \prod_{i=1}^{r} p_i^{\min(a_{ki}, b_{li})} \\ \Rightarrow \qquad & \deg \gcd(f_k, F_l) = \sum_{i=1}^{r} \deg p_i \ \min(a_{ki}, b_{li}) \\ \text{and} \quad & \deg \gcd(f_k, f_l) = \sum_{i=1}^{r} \deg p_i \ \min(a_{ki}, a_{li}) \\ \text{and} \quad & \deg \gcd(F_k, F_l) = \sum_{i=1}^{r} \deg p_i \ \min(b_{ki}, b_{li}). \end{aligned}$$

Then equation (35) may be rewritten as

$$\begin{aligned} & \nu_{A,A} + \nu_{B,B} - 2\nu_{A,B} \\ &= \sum_{k=1}^{N} \sum_{l=1}^{N} \sum_{i=1}^{r} \deg p_i \{\min(a_{ki}, a_{li}) + \min(b_{ki}, b_{li}) \\ &\qquad -2 \min(a_{ki}, b_{li})\} \\ &= \sum_{i=1}^{r} \deg p_i \sum_{k=1}^{N} \sum_{l=1}^{N} \{\min(a_{ki}, a_{li}) + \min(b_{ki}, b_{li}) \\ &\qquad -2 \min(a_{ki}, b_{li})\}. \end{aligned}$$

The latter double sum is of the form in lemma 7.1 and so, since $\deg p_i > 0$, we have

$$\nu_{A,A} + \nu_{B,B} - 2\nu_{A,B} \geq 0,$$

proving the first part of the theorem.

Next we show that equality to zero in the above is equivalent to similarity of the matrices:

$$\begin{aligned}
& \nu_{A,A} + \nu_{B,B} - 2\nu_{A,B} = 0 \\
\Leftrightarrow \quad & \sum_{i=1}^{r} \deg p_i \sum_{k=1}^{N} \sum_{l=1}^{N} \{\min(a_{ki}, a_{li}) + \min(b_{ki}, b_{li}) \\
& \qquad -2\min(a_{ki}, b_{li})\} = 0 \\
\Leftrightarrow \quad & \text{sequences } \{a_{ki}\}, \{b_{ki}\} \text{ identical (by lemma 7.1)} \\
\Leftrightarrow \quad & A \text{ and } B \text{ have same invariant factors} \\
\Leftrightarrow \quad & A \text{ and } B \text{ are similar } (\Rightarrow m = n).
\end{aligned}$$

**EXERCISE 7.1**

*Show if if*

$$P^{-1}A_1P = A_2 \qquad \text{and} \qquad Q^{-1}B_1Q = B_2$$

*then*

$$\begin{aligned}
& (P^{-1} \otimes Q^{-1})(A_1 \otimes I_m - I_n \otimes B_1^t)(P \otimes Q) \\
= \quad & A_2 \otimes I_n - I_m \otimes B_2^t.
\end{aligned}$$

*(This is another way of showing that if $A$ and $B$ are similar then*

$$\nu_{A,A} + \nu_{B,B} - 2\nu_{A,B} = 0.)$$

# 8 Further directions in linear algebra

1. Dual space of a vector space; Tensor products of vector spaces; exterior algebra of a vector space. See C.W. Curtis, *Linear Algebra, an introductory approach* and T.S. Blyth, *Module theory*.

2. Quadratic forms, positive definite matrices (see L. Mirsky, *Introduction to linear Algebra*), singular value decomposition (see G. Strang, *Linear Algebra*).

3. Iterative methods for finding inverses and solving linear systems. See D.R. Hill and C.B. Moler, *Experiments in Computational Matrix Algebra*.

4. Positive matrices and Markov matrices are important in economics and statistics. For further reading on the structure of Markov matrices and more generally, non–negative matrices, the following books are recommended:

   [1] N.J. Pullman. *Matrix Theory and its Applications, 1976.* Marcel Dekker Inc. New York.

   [2] M. Pearl. *Matrix Theory and Finite Mathematics, 1973.* McGraw–Hill Book Company, New York.

   [3] H. Minc. *Nonnegative Matrices, 1988.* John Wiley and Sons, New York.

5. There are at least two research journals devoted to linear and multilinear algebra in our Physical Sciences Library: *Linear and Multilinear Algebra* and *Linear Algebra and its applications*.