AN ELEMENTARY DERIVATION OF THE JORDAN NORMAL FORM
WITH AN APPENDIX ON LINEAR SPACES

H. R. van der Vaart

Didactical Report

Work supported by Public Health Service Grant GM-678 from the National Institute of General Medical Sciences to the Biomathematics Training Program, Institute of Statistics Raleigh Section

Institute of Statistics
Mimeograph Series No. 460
February, 1966
INTRODUCTION

The aim of this mimeograph is purely didactical. The Jordan normal form of a matrix is generally regarded as an advanced topic. It is hard to find in the literature a complete, somewhat leisurely exposition, which in all its phases is essentially based on nothing more than the concepts of linear space and subspace, basis and direct sum, dimension, and the fundamental idea of mapping. In this sense the present exposition is elementary.

Most of the proofs assembled here are essentially known, but the present account constitutes a consistent effort to use none but the most elementary tools throughout the entire development. It is hoped that this report may help dispel whatever misunderstandings "users" of matrix theory may have about this topic.
CONTENTS

1. Preliminaries  
   1.1. Representation of a linear mapping by a square matrix 1-1  
   1.2. Change of basis 1-3  
   1.3. Special cases 1-5  
   1.4. Eigenvalues, eigenvectors, characteristic polynomial 1-6  
   1.5. Cayley-Hamilton theorem 1-8  

2. Canonical representations  
   2.1. A singular mapping on \( \mathcal{V} \) is nilpotent on a certain subspace of \( \mathcal{V} \) 2-1  
   2.2. \( \mathcal{V} \) is a direct sum of generalized eigenspaces 2-4  
   2.3. A diagonal block representation of linear mappings 2-7  
   2.4. A canonical representation of nilpotent linear mappings 2-10  
   2.5. The Jordan normal form of an \( n \times n \) matrix 2-16  

A. Appendix  
   A.1. Fields  
   A.2. Complex numbers  
   A.3. Linear spaces  
   A.4. Basis, dimension, direct sum  
   A.5. Linear mappings  

L. Some literature on linear algebra  

Page

0-2
1. Preliminaries

1.1. Representation of a linear mapping by a square matrix

We will consider only such linear mappings \( f \) as map a certain linear space \( (V, \text{say}) \) into the same space \( (V) \), \( f : V \to V \). We could choose one basis in \( V \) as the domain of \( f \) and another basis in \( V \) as (containing) the range of \( f \), but we will discuss only the case in which these two bases are the same.

Let now \( \{e'_1, e'_2, \ldots, e'_n\} \) be one such basis of \( V \). For each \( j = 1, \ldots, n \), \( f e'_j \), being a vector in \( V \), is a linear combination of basis vectors:

\[
(1.1,1) \quad f e'_j = \sum_{i=1}^{n} \alpha_{ij} e'_i = \sum_{i=1}^{n} e'_i \alpha_{ij} \quad (j = 1, \ldots, n)
\]

The matrix \( A = [\alpha_{ij}]^{i=1 \ldots n}_{j=1 \ldots n} \) completely determines the linear mapping \( f \). In fact, the vector \( c \in V \), which can be uniquely represented by \( c = \sum_{j=1}^{n} \gamma_j e'_j \), is carried into, say, \( d = \sum_{i=1}^{n} \delta_i e'_i = f c = \sum_{j=1}^{n} \gamma_j f e'_j = \sum_{i=1}^{n} e'_i \alpha_{ij} \gamma_j \).

Since \( \{e'_1, e'_2, \ldots, e'_n\} \) constitutes a basis of \( V \), it follows that

\[
(1.1,2) \quad \delta_i = \sum_{j=1}^{n} \alpha_{ij} \gamma_j \quad (i = 1, \ldots, n)
\]

which is the familiar formula for the transformation of vectors (such as \( c \)) when given by their coordinates, i.e. in the form \( c = \sum_{j=1}^{n} \gamma_j e'_j \), or in matrix notation:

\[
(1.1,2a) \quad c = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}
\]

Instead of the notation used in (1.1,2) we have the well known alternatives:

\[
(1.1,2b) \quad \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}, \text{ or}
\]

\[
(1.1,2c) \quad d = A c
\]
The matrix $A$ is said to be the **matrix of $\mathcal{L}$** (or the **matrix representing $\mathcal{L}$**) relative to the basis $\{e'_1, e'_2, \ldots, e'_n\}$. Note that the number of rows of $A$ and the number of columns of $A$ equal $n$, the dimension of $\mathcal{V}$. 
1.2. Change of basis.

Let \( \{ e'_1, e'_2, \ldots, e'_n \} \) be another basis of \( \mathcal{V} \), then clearly each \( e'_k \) is a linear combination of the \( e''_i \):

\[
e'_k = \sum_{i=1}^{n} \pi''_{ik} e''_i = \sum_{i=1}^{n} e''_i \pi''_{ik} \quad (k = 1, \ldots, n)
\]

(1.2.1)

Now

\[
\Sigma e''_h = \sum_k e''_k \beta_{kh} = \sum_i e''_i \pi''_{ih}, \quad \text{but also}
\]

\[
\Sigma e''_h = \sum_k e''_k \beta_{kh} = \sum_i \pi''_{ih} \beta_{kh},
\]

where \( B \) is the matrix of \( \Sigma \) re the basis \( \{ e''_1, e''_2, \ldots, e''_n \} \).

So obviously

\[
\sum_{i=1}^{n} \pi''_{ih} \pi''_{ik} = \sum_{i=1}^{n} \pi''_{ik} \beta_{ih}, \quad \text{or} \quad A \overline{\Sigma} = \overline{\Sigma} B,
\]

(1.2.2)

where \( \overline{\Sigma} \), as defined in eq. (1.2.1) is the matrix expressing the new basis vectors in terms of the old ones.

**Note.** As an alternative to eq. (1.2.1) it is also obvious that each \( e''_i \) is a linear combination of the \( e''_g \):

\[
e''_i = \sum_{g} e''_g \rho_{gi} \quad (i = 1, \ldots, n)
\]

(1.2.3)

Hence

\[
e''_k = \sum_{g} e''_g \rho_{gi} \pi''_{ik} = \sum_{g} e''_g \delta_{gk} \quad (k = 1, \ldots, n)
\]

where \( \delta_{gk} = 0 \) if \( g \neq k \), \( \delta_{gk} = 1 \) if \( g = k \). Thus, if \( I = [\delta_{gk}] \), \( P = [\rho_{gi}] \):

(1.2.4)

\[P \overline{\Sigma} = I, \quad \overline{\Sigma} \text{ has an inverse, } P = \overline{\Sigma}^{-1}.
\]

This allows us to replace (1.2.2) by

\[
\overline{\Sigma}^{-1} A \overline{\Sigma} = B.
\]

(1.2.5)

**Note.** Two matrices \( A \) and \( B \) satisfying an equation of the form (1.2.5) are said to be similar. They represent the same linear transformation \( \Sigma : \mathcal{V} \to \mathcal{V} \) re different bases.
Note. The concept of similar matrices is easiest grasped by a discussion of a linear transformation $L$ represented by different matrices relative to different bases. Practical situations are usually less symmetrical in that in practice a linear transformation is often introduced by its matrix representation to begin with (re some unidentified basis). By the same token vectors are often introduced in matrix form (cf. eq. (1.1,2a)), i.e., by their coordinates (re the same unidentified basis).

Let $L$ be such a matrix, and assume we have $n$ linearly independent vectors

$$
c_1 = \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1n} \end{bmatrix}, \quad c_2 = \begin{bmatrix} \gamma_{12} \\ \gamma_{22} \\ \vdots \\ \gamma_{2n} \end{bmatrix}, \quad \ldots, \quad c_n = \begin{bmatrix} \gamma_{1n} \\ \gamma_{2n} \\ \vdots \\ \gamma_{nn} \end{bmatrix},
$$

and $n^2$ scalars $\beta_{ij}$ such that

$$(1.2,6) \quad L c_i = \sum_j c_j \beta_{ji} = (c_1 \ c_2 \ \ldots \ c_n) \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \vdots \\ \beta_{nn} \end{bmatrix}, \quad (i = 1, 2, \ldots, n)$$

where $(c_1 \ c_2 \ \ldots \ c_n)$ denotes the matrix with first column $c_1$, second column $c_2$, etc. Writing $C$ for this matrix we find from (1.2,6):

$$(1.2,6a) \quad L (c_1 \ c_2 \ \ldots \ c_n) = (c_1 \ c_2 \ \ldots \ c_n) \begin{bmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \ldots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \ldots & \beta_{nn} \end{bmatrix},$$

hence

$$(1.2,6b) \quad L C = C B, \text{ or } C^{-1} L C = B, \text{ or } L = C B C^{-1}.$$ 

In other words, under the assumption (1.2,6), $L$ and $B$ are similar matrices, and the transforming matrix $C$ is built from the column matrices

$$c_i = \begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \\ \vdots \\ \gamma_{in} \end{bmatrix}.$$
1.3. Special cases.

i) The identity mapping $\mathcal{I}$, defined by

$$\mathcal{I}v = v \quad \text{for each} \quad v \in \mathcal{V},$$

is represented by one and the same matrix relative to all bases:

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Prove!

ii) $L$ carries each $v \in \mathcal{V}$ into the null vector $\mathcal{0}$ iff the matrix representing $L$ re some basis is the null matrix

$$\square = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The null matrix then represents this mapping $L$ re all bases. Prove! (This particular $L$ will be denoted by $\mathcal{0}$).

iii) Define $L^2 \overset{\text{def}}{=} L \circ L$, i.e., $L^2v = L(Lv)$ for each $v \in \mathcal{V}$. Similarly $L^k \overset{\text{def}}{=} L \circ L^{k-1}$, so "by induction" $L^s$ is now defined for every natural $s$. Now let the matrix $L$ represent $L$ relative to some basis.

Prove that the matrix $L^2$ represents $L^2$ re the same basis.

Prove that the matrix $L^s$ represents $L^s$ re the same basis,

where $s$ is any natural number.

Let $\phi$ be a (formal) polynomial, $\phi(\diamond) = \sum_{i=0}^{r} \alpha_i \diamond^i$. Then prove that

$$\phi(L)v = \sum_{i=0}^{r} \alpha_i L^i v \quad \text{for each} \quad v \in \mathcal{V}.$$

Note that

$$\begin{cases} \text{if } \diamond \text{ is a scalar, } \phi(\diamond) \text{ is a scalar,} \\
\text{if } \diamond \text{ is a mapping, } \phi(\diamond) \text{ is a mapping,} \\
\text{if } \diamond \text{ is a matrix, } \phi(\diamond) \text{ is a matrix.} \end{cases}$$
1.4. Eigenvalues, eigenvectors, characteristic polynomial

When a non-null vector \( v \in V \) and a scalar \( \lambda \) exist such that

\[ (1.4,1) \quad L v = \lambda v \]

then \( \lambda \) is called an eigenvalue of \( L \), \( v \) an eigenvector of \( L \). If \( V \) is an \( n \)-dimensional vector space, the eigenvalues of \( L \) are roots of an \( n \)-th degree polynomial equation.

**Proof.** Choose a basis of \( V \), say \( \{ e'_{1}, e'_{2}, \ldots, e'_{n} \} \). Let \( \gamma_{i} \) be the coordinates (re this basis) of \( v \), \( v = \sum_{i} \gamma_{i} e_{i}' \). Let \( A = [\alpha_{ij}] \) represent \( L \) re the same basis. Then

\[ L v = L \sum_{j} \gamma_{j} e'_{j} = \sum_{i} \alpha_{ij} \gamma_{i} e'_{i} \]

and \( L v = \lambda v \) iff

\[ (1.4,1a) \quad \sum_{ij} \alpha_{ij} \gamma_{j} = \lambda \gamma_{i} \]

i.e., iff

\[ (1.4,1b) \quad \begin{cases} 
\sum_{j} \alpha_{ij} \gamma_{j} = \lambda \gamma_{i} \\
\sum_{j} (\lambda \delta_{ij} - \alpha_{ij}) \gamma_{j} = 0 
\end{cases} \quad \text{for all } i = 1, \ldots, n, \]

where \( \delta_{ij} = 1 \) if \( i = j \), \( \delta_{ij} = 0 \) if \( i \neq j \). Equation (1.4,1b) has non-null solutions for \( \gamma_{j} \), i.e., equation (1.4,1) holds for non-null vectors of \( v \), iff \( \lambda \) satisfies the equation

\[ (1.4,2) \quad \begin{cases} 
\det ([\lambda \delta_{ij} - \alpha_{ij}]) = 0 \\
\det (\lambda I - A) = 0 
\end{cases} \]

Clearly this determinant is an \( n \)-degree polynomial in \( \lambda \), say \( \chi_{A}(\lambda) \):

\[ (1.4,2a) \quad \det (\lambda I - A) \xrightarrow{\text{def}} \chi_{A}(\lambda) \]

in which \( \lambda^{n} \) is the highest degree term ("leading coefficient is 1"). The polynomial \( \chi_{A}(\lambda) \) is called the **characteristic polynomial of the matrix** \( A \).

Now choose another basis of \( V \), say \( \{ e''_{1}, e''_{2}, \ldots, e''_{n} \} \), connected with \( \{ e'_{1}, e'_{2}, \ldots, e'_{n} \} \) by equation (1.2,1). Repeating the above argument, we find that the eigenvalues \( \lambda \) of \( L \) are also to satisfy the equation
where $\mathbf{B}$ represents the basis \{e''_1, e''_2, \ldots, e''_n\}. One would not like the roots of equation (1.4,2) to be different from those of eq. (1.4,2b).

Indeed $\det(\lambda \mathbf{I} - \mathbf{B}) = \det(\lambda \mathbf{I} - \mathbf{A})$ for all scalars $\lambda$; for, by equation (1.2,5):

$$\det(\lambda \mathbf{I} - \mathbf{B}) = \det(\lambda \mathbf{I} - \mathbf{T}^{-1} \mathbf{A} \mathbf{T}) = \det(\mathbf{T}^{-1}(\lambda \mathbf{I})\mathbf{T}^{-1} \mathbf{A} \mathbf{T}) = \det(\mathbf{T}^{-1}(\lambda \mathbf{I} - \mathbf{A})\mathbf{T}) =$$

$$\det(\mathbf{T}^{-1}) \det(\lambda \mathbf{I} - \mathbf{A}) \det \mathbf{T} = \det(\lambda \mathbf{I} - \mathbf{A})$$, for all $\lambda$.

Denote $\chi_B(\lambda) = \chi_A(\lambda)$ by $\chi(\lambda)$. The polynomial expression $\chi(\phi)$ is called the characteristic polynomial of the linear transformation $\mathbf{L} : \mathbf{V} \rightarrow \mathbf{V}$. It is seen that one can find it by computing the characteristic polynomial of the matrix representing $\mathbf{L}$ relative to any basis.

**Note.** Since our scalars are complex numbers, $\chi(\lambda)$ can always be written as a product of $n$ factors of form $(\lambda - \lambda_1)$, where the $\lambda_i$ are the (possibly complex valued) $n$ roots of the equation $\chi(\lambda) = 0$. Some of these $n$ roots may be equal. Assume that $p \leq n$ of them are different. Then $\chi(\lambda) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \ldots (\lambda - \lambda_p)^n$. 


1.5. Cayley - Hamilton theorem.

If \( \chi \) is the [[characteristic polynomial]] of the linear mapping \( \mathbf{L} : \mathbb{V} \rightarrow \mathbb{V} \), then \( \chi(\mathbf{L}) = 0 \).

**Proof.** According to sec. 1.3, point (iii), \( \chi(\mathbf{L}) \) is a linear mapping \( \mathbb{V} \rightarrow \mathbb{V} \), and if \( A \) represents \( \mathbf{L} \) on some basis then \( \chi(A) \) represents \( \chi(\mathbf{L}) \) relative to that basis. According to sec. 1.3, point (ii), we have to show that \( \chi(A) = \mathbf{0} \), the null matrix. Since \( \chi(\lambda) = \det(\lambda \mathbf{I} - A) \) one might think that a proof could be conducted as follows

\[
\chi(A) = \det(\lambda \mathbf{I} - A) = \det(\mathbf{0}) = 0.
\]

However, this is not good, since \( \det(\ldots) \) is scalar-valued, but \( \chi(A) \) is a matrix.

Let the matrix \( Q(\lambda) \) be the adjoint of \( (\lambda \mathbf{I} - A) \), that is,

\[
\text{cofactor of } (\lambda \delta_{ji} - \alpha_{ji}) \text{ in matrix } (\lambda \mathbf{I} - A) =
q_{ij}(\lambda) = (n-1)^{th} \text{ degree polynomial}; \quad \text{so } Q(\lambda) = \sum_{i=0}^{n-1} Q_i \lambda^i
\]

Then for all \( \lambda \) we have \((\lambda \mathbf{I} - A)^{-1} = \frac{Q(\lambda)}{\det(\lambda \mathbf{I} - A)}\), so

\[(1.5,1) \quad Q(\lambda). (\lambda \mathbf{I} - A) = \det(\lambda \mathbf{I} - A) = \chi_A(\lambda). \mathbf{I}.
\]

Write \( \chi_A(\lambda) = \sum_{j=0}^{n} \varepsilon_j \lambda^j \), where \( \varepsilon_n = 1 \), the \( \varepsilon_j \) are functions of \( A \).

Then eq. (1.5,1) can be written as

\[(1.5,1a) \quad (\sum_{i=0}^{n-1} Q_i \lambda^i). (\lambda \mathbf{I} - A) = \sum_{j=0}^{n} \varepsilon_j \lambda^j \mathbf{I} \quad \text{for all scalar } \lambda.
\]

Hence

\[
\begin{array}{ccc}
-Q_0 A + \varepsilon_0 \mathbf{I} & = & \mathbf{0}^0 \\
-Q_1 A + Q_0 & = & \varepsilon_1 \mathbf{I} \\
-Q_2 A + Q_1 & = & \varepsilon_2 \mathbf{I} \\
\vdots & = & \mathbf{0}^{n-1} \\
-Q_{n-1} A + Q_{n-2} & = & \varepsilon_{n-1} \mathbf{I} \\
+ Q_{n-1} & = & \varepsilon_n \mathbf{I} \\
\mathbf{0} & = & \sum_{i=0}^{n} \varepsilon_i A^i = \chi(A).
\end{array}
\]
Thus we have found an \(n\)th degree polynomial, the characteristic polynomial \(X\) of the linear mapping \(S : \mathcal{V} \to \mathcal{V}\) (\(\mathcal{V} n\)-dimensional), which at \(S\) has the value \(\mathcal{O}\). There are other polynomials, of degree greater than \(n\), which also have the value \(\mathcal{O}\) at \(S\), but it is far more interesting that some linear mappings \(S : \mathcal{V} \to \mathcal{V}\) admit polynomials \(\psi\) of degree less than \(n\) for which
\[
\psi(S) = \mathcal{O}.
\]

Calling a polynomial with at least one coefficient unequal to zero a non-trivial polynomial, we have the following definition: the non-trivial polynomial with leading coefficient \(1\) which is of least degree among all polynomials that assume the value \(\mathcal{O}\) at \(S\) (or \(\square\) at \(A\)) is called the **minimum polynomial** of the linear mapping \(S\) (or of the matrix \(A\)).

**Note.** Without proof we will cite the following statements:

- Every linear mapping \(S\) (or matrix \(A\)) has a unique minimum polynomial.
- Every polynomial which at \(S\) (or \(A\)) has the value \(\mathcal{O}\) (or \(\square\)) is divisible by the minimum polynomial (which points up a way to find the minimum polynomial from the characteristic polynomial).

Similar matrices have identical minimum polynomials.

**Example.**
\[
A = \begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0
\end{bmatrix}; \quad \lambda I - A = \begin{bmatrix}
\lambda & -1 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
0 & 0 & \lambda & -1 \\
0 & 0 & 0 & \lambda
\end{bmatrix}
\]
\[
\det(\lambda I - A) = \chi(\lambda) = \lambda^4; \quad \text{and indeed} \quad A^4 = \mathcal{O}.
\]

But already \(A^2 = \square\). Hence the minimum polynomial is
\[
\psi(\mathcal{O}) = \mathcal{O}^2.
\]

**Example.**
\[
B = \begin{bmatrix}
2 & 0 & 3 \\
2 & 1 & 2 \\
0 & 1 & -1
\end{bmatrix}.
\]
Find the characteristic polynomial \(\chi\).

Compute, number by number, the matrix \(\chi(B)\) and thus check the Cayley-Hamilton theorem.

Using the second statement in the above **Note** prove that the minimum polynomial of \(B\) is also \(\chi\).
2. Canonical representations *)

Again let \( \mathcal{Y} \) be an \( n \)-dimensional vector space. Let \( \mathcal{J} : \mathcal{Y} \to \mathcal{Y} \) be a linear mapping. Our aim will now be to find a basis of \( \mathcal{Y} \) such that the matrix of \( \mathcal{J} \) in this basis has as "simple" a form as possible. An important role in this process will be played by \textbf{singular} linear mappings, even though \( \mathcal{J} \) need not be singular at all. Singular mappings enter the picture through the following backdoor: let \( \lambda_i \) \( (i = 1, \ldots, p; \ p \leq n) ** \) be the eigenvalues of \( \mathcal{J} \); define \( \mathcal{L}_i = \mathcal{J} - \lambda_i \) \( (i = 1, \ldots, p) \); then, as a consequence of the definition of eigenvalue there exist non-null vectors \( \mathbf{v}_i \in \mathcal{Y} \) with

\[
(\mathcal{J} - \lambda_i \mathcal{I}) \mathbf{v}_i = \mathcal{L}_i \mathbf{v}_i = 0, \quad (i = 1, \ldots, p)
\]

i.e., \( \mathcal{L}_i \) is singular on \( \mathcal{Y} \). In the process of constructing a simplifying basis the mappings \( \mathcal{L}_i \) play an important role.

We will first show that \( \mathcal{Y} \) is the direct sum of \( p \) subspaces \( \mathcal{R}_i \) such that

1) \( \mathcal{L}_i \) is nilpotent on \( \mathcal{R}_i \) and non-singular on

\[
\bigoplus_{\substack{j=1\ j\neq i}}^{n} \mathcal{R}_j,
\]

2) \( \mathcal{R}_i \) is invariant under \( \mathcal{J} \), i.e., \( \mathcal{J} \mathcal{R}_i \subseteq \mathcal{R}_i \).

The second property permits us to choose a basis in \( \mathcal{Y} \) such that the matrix of \( \mathcal{J} \) in this basis consists of diagonal blocks and zeros outside these blocks.

Since according to the first property \( \mathcal{L}_i \) is nilpotent on the subspace \( \mathcal{R}_i \) \( (i = 1, \ldots, p) \), the remaining problem is then to choose bases in \( \mathcal{R}_i \) such that the corresponding matrices of the nilpotent mappings \( \mathcal{L}_i \) are "simple". The solution of this problem will enable us to obtain "simple" forms for the above-mentioned diagonal blocks.

*) Before starting on this chapter, the reader may want to at least glance through the Appendix, especially sections A.3, A.4, A.5, in order to get acquainted with or brush up on matters of definition and notation.

**) Some of the roots of the equation \( X(\lambda) = 0 \) may be equal. Then the total number, say \( p \), of \textit{different} eigenvalues is less than \( n \).
2.1. A singular mapping on \( \mathcal{Y} \) is nilpotent on a certain subspace of \( \mathcal{Y} \)

Let \( \mathcal{Y} \) be an \( n \)-dimensional linear space. Let \( f : \mathcal{Y} \rightarrow \mathcal{Y} \) be singular; consider the following sequences of null-spaces and image spaces:

\( \mathcal{R}(f), \mathcal{R}(f^2), \mathcal{R}(f^3), ... \)

\( \mathcal{F}(f), \mathcal{F}(f^2), \mathcal{F}(f^3), ... \)

1) \( v \in \mathcal{R}(f) \Rightarrow f^2 v = 0 = f^2 v = f f v = f f v 
\in \mathcal{R}(f^2) = v \in \mathcal{R}(f^2) \) (why?), etc.

So \( \mathcal{R}(f) \subseteq \mathcal{R}(f^2) \subseteq \mathcal{R}(f^3) \subseteq ... \)

Whenever \( \subseteq \) holds as in \( \mathcal{R}(f^s) \subseteq \mathcal{R}(f^{s+1}) \), it follows (see section A.4, under 'Dimension of Subspace') that \( \dim \mathcal{R}(f^s) < \dim \mathcal{R}(f^{s+1}) \); hence \( f^k \subseteq \mathcal{Y} \) for all integer \( k > 0 \), these dimensions must be less than or equal to \( n \); hence the strict inequality signs cannot go on indefinitely, and many integers \( k > 0 \) must exist such that \( \mathcal{R}(f^k) = \mathcal{R}(f^{k+1}) \). Let \( m \) be the smallest such integer. Then it follows that \( \mathcal{R}(f^m) = \mathcal{R}(f^{m+s}) \) for all integer \( s > 0 \).

Proof: Since \( \mathcal{R}(f^{m+1}) = \mathcal{R}(f^m) \), it follows that \( f^m v = 0 \) whenever \( f^{m+1} v = 0 \). Hence \( f^{m+s} v = f^{m+1} (f^{s-1} v) = 0 \)

\[ f^m (f^{s-1} v) = 0 = f^{m+s-1} v. \]

So \( \mathcal{R}(f^{m+s}) \subseteq \mathcal{R}(f^{m+s-1}) \), and so on down the line. That is:

\[ \mathcal{R}(f^{m+s}) \subseteq \mathcal{R}(f^{m+s-1}) \subseteq \mathcal{R}(f^{m+s-2}) \subseteq ... \subseteq \mathcal{R}(f^{m+1}) = \mathcal{R}(f^m). \]

From the beginning of this point (i) we already know that all these inequality signs hold when reversed. Hence they are all strict equalities.

ii) By definition \( \mathcal{R}(f^m) = \{0\} \), i.e., \( f \) is nil onto \( \mathcal{R}(f^m) \) (cf. end of sec. A.5).

iii) \( \mathcal{R}(f^2) = f^2 \mathcal{Y} \subseteq f \mathcal{Y} = \mathcal{R}(f) \), where the inequality follows from \( f \mathcal{Y} \subseteq \mathcal{Y} (= f : \mathcal{Y} \rightarrow \mathcal{Y}) \).

Similarly \( \mathcal{R}(f^3) \subseteq \mathcal{R}(f^2) \), etc. So:

\[ \mathcal{R}(f) \supseteq \mathcal{R}(f^2) \supseteq \mathcal{R}(f^3) \supseteq ... \]

Since \( \dim \mathcal{R}(f^k) + \dim \mathcal{R}(f^k) = n \) for all integer \( k > 0 \) (cf. sec. A.5, 'Theorem'), it follows from point (i) that \( \dim \mathcal{R}(f^m) = \dim \mathcal{R}(f^{m+s}) \), (cf. sec.
A.4 'Dimension of Subspace') hence \( \mathcal{R}(f^m) = \mathcal{R}(f^{m+s}) \) for all integer \( s > 0 \).
So \( \mathcal{L}(\mathcal{R}(f^m)) = f^{m+1} = \mathcal{R}(f^{m+1}) = \mathcal{R}(f^m) \), i.e., (cf. sec. A.5, under 'Non-singular vs. singular'):
\[ \mathcal{L} \text{ is non-singular on } \mathcal{R}(f^m). \]
Note the definition of \( m \) in the above point (i).

iv) The intersection \( \mathcal{R}(f^k) \cap \mathcal{R}(f^k) \) is not necessarily restricted to \( \{0\} \) even though \( \dim \mathcal{R}(f^k) + \dim \mathcal{R}(f^k) = n \) for all natural \( k \). Let, for instance \( \mathcal{R}(f) \subset \mathcal{R}(f^2) \), then there exists \( v \in \mathbb{V} \) such that \( v \notin \mathcal{R}(f) \), \( v \in \mathcal{R}(f^2) \), i.e., \( f^2v \neq 0 \), \( f^2v = 0 \), \( f^2v \in \mathcal{R}(f) \), yet obviously \( f^2v \in \mathcal{R}(f) \). However, if \( m \) (as in point (i)) is the smallest integer with \( \mathcal{R}(f^m) = \mathcal{R}(f^{m+1}) \), then \( \mathcal{R}(f^m) \cap \mathcal{R}(f^m) = \{0\} \).

Proof. Let \( w \in \mathcal{R}(f^m) \cap \mathcal{R}(f^m) \). Then \( f^mw = 0 \), and there exists \( v \in \mathbb{V} \) with \( f^mv = w \); so \( 0 = f^mw = f^2mv = f^mv = w \); QED.
\[ \text{(since } \mathcal{R}(f^{m+s}) = \mathcal{R}(f^m) \text{ for all } s > 0 \). \]

v) \( \mathbb{V} = \mathcal{R}(f^m) \oplus \mathcal{R}(f^m) \), \( m \) defined as in points (i) and (iv).

Proof. \( \mathcal{R}(f^m) \) is a linear subspace of \( \mathbb{V} \),
\( \mathcal{R}(f^m) \) is a linear subspace of \( \mathbb{V} \),
\( \dim \mathcal{R}(f^m) + \dim \mathcal{R}(f^m) = \dim \mathbb{V} \),
\( \mathcal{R}(f^m) \cap \mathcal{R}(f^m) = \{0\} \).

Claim follows by sec. A.4: 'Sufficient and necessary condition for \( \mathbb{V} = U_1 \oplus U_2 \)'.

An alternative way of describing the same result is, of course
\[ \mathbb{V} = \mathcal{R}(f^m) \oplus f^m \mathbb{V} \].
2.2. \( V \) is a direct sum of generalized eigenspaces

Let \( \lambda_1, \lambda_2, \ldots, \lambda_p \) be the \( p \) different eigenvalues of the linear mapping \( \mathcal{F} : V \rightarrow V \) (\( V \) still \( n \)-dimensional). For all \( i = 1, 2, \ldots, p \):

- define \( L_i = \mathcal{F} - \lambda_i I \); \( L_i \) is singular on \( V \);
- go through the steps taken in section 2.1;
- \( m_i \) is the smallest natural number with \( \mathbb{R}(L_i^{m_i}) = \mathbb{R}(L_i^{m_i+1}) \);
- for \( \mathbb{R}(L_i^{m_i}) \) write \( \mathbb{R}_i \) (these subspaces generalize the single eigenvectors occurring if all \( n \) eigenvalues are different; they will be called generalized eigenspaces);

for \( \mathbb{R}(L_i^{m_i}) \) write \( \mathbb{R}_i \);

- \( L_i \) is nilpotent on \( \mathbb{R}_i \), non-singular on \( \mathbb{R}_i \); \( L_i \mathbb{R}_i = \{0\} \); \( L_i \mathbb{R}_i = \mathbb{R}_i \);

\( V = \mathbb{R}_1 \oplus \mathbb{R}_i \); \( n = \dim \mathbb{R}_1 + \dim \mathbb{R}_i \).

In addition we will show

i) \( \mathbb{R}_i \subseteq \mathbb{R}_i \); \( \mathbb{R}_i \) is invariant under \( \mathcal{F} \) (\( i = 1, \ldots, p \))

Proof. \( v \in \mathbb{R}_i \Rightarrow L_i^{m_i} v = 0 \Rightarrow L_i \mathbb{R}_i = \{0\} \Rightarrow L_i \mathbb{R}_i = \mathbb{R}_i \)

So all \( \mathcal{F} \)-images of vectors in \( \mathbb{R}_i \) belong to \( \mathbb{R}_i \); i.e., \( \mathcal{F} \mathbb{R}_i \subseteq \mathbb{R}_i \).

ii) \( \mathbb{R}_i \subseteq \mathbb{R}_i \); \( \mathbb{R}_i \) is invariant under \( \mathcal{F} \).

Proof. By point (iii) of section 2.1:

\[ L_i (\mathbb{R}(L_i^{m_i})) = \mathbb{R}(L_i^{m_i}) \Rightarrow L_i \mathbb{R}_i = \mathbb{R}_i \Rightarrow (\mathcal{F} - \lambda_i I) \mathbb{R}_i = \mathbb{R}_i \], i.e.,

\[ v \in \mathbb{R}_i \Rightarrow (\mathcal{F} - \lambda_i I) v \in \mathbb{R}_i \Rightarrow \mathcal{F} v \text{ is a linear combination of } v (v \in \mathbb{R}_i) \]

and some vector of \( \mathbb{R}_i \) is \( v \in \mathbb{R}_i \).

iii) \( L_i \mathbb{R}_i = \mathbb{R}_i \) for \( i \neq j \); \( L_k L_j \mathbb{R}_i = \mathbb{R}_i \) for \( i \neq j, i \neq k, k \neq j \).
Proof. We know that $\mathcal{M}_i \subseteq \mathcal{M}_j$, hence $\mathcal{M}_j \subseteq \mathcal{M}_i$, even if $i \neq j$.

Hence if we can prove that $\mathcal{L}_j$ is non-singular on $\mathcal{M}_i$ ($i \neq j$), it will follow that $\mathcal{M}_j \mathcal{M}_i = \mathcal{M}_i = \mathcal{M}_j \mathcal{M}_i = \mathcal{M}_i$.

In fact, assume $v \in \mathcal{M}_i = \mathcal{R}(\mathcal{L}_i)$ and also $v \in \mathcal{R}(\mathcal{L}_j)$. Then:

$$(\lambda_j - \lambda_i)^m_i v = (\mathcal{L}_i - \mathcal{L}_j)^m_i v = \mathcal{L}_i^m_i v + \sum_{k=1}^{m_j} (-1)^k \mathcal{L}_i^k \mathcal{L}_j^k v,$$

where the first term is zero since $v \in \mathcal{R}(\mathcal{L}_i^m_i)$, and the second term is zero since $v \in \mathcal{R}(\mathcal{L}_j)$. Since $\lambda_j \neq \lambda_i$, it follows that whenever $v \in \mathcal{R}(\mathcal{L}_i^m_i) \cap \mathcal{R}(\mathcal{L}_j^m_j)$ then $v = 0$. Hence no non-null vector in $\mathcal{M}_i$ is annihilated by $\mathcal{L}_j = \mathcal{J} - \lambda_j \mathcal{J}$; i.e. $\mathcal{L}_j$ is non-singular on $\mathcal{M}_i$. (This rather ingenious proof from Nehring, p. 103, Th. 8.2).

Now, obviously, since $\mathcal{L}_j^{m_j} \mathcal{M}_i = \mathcal{M}_i$ ($i \neq j$) it follows immediately that $\mathcal{L}_k^{m_k} \mathcal{L}_j^{m_j} \mathcal{M}_i = \mathcal{L}_k^{m_k} \mathcal{M}_i = \mathcal{M}_i$ ($i \neq j$, $j \neq k$, $i \neq k$). Generalization to more "factors" of the form $\mathcal{L}_j^{m_j}$ is obvious.

iv) The set of all vectors in $\mathcal{V}$ annihilated by $\mathcal{L}_i^{m_i}$ is the same as the set of all vectors in $\mathcal{L}_j^{m_j} \mathcal{L}_j^{m_j} \ldots \mathcal{L}_k^{m_k} \mathcal{V}$ annihilated by $\mathcal{L}_i^{m_i}$; hence the latter set is still $\mathcal{M}_i$ (provided all $i, j_1, j_2, \ldots, j_k$ are different).

Proof. Because of point (iii) in section 2.2.

Hence $\mathcal{M}_i \subseteq \mathcal{L}_j^{m_j} \mathcal{L}_j^{m_j} \ldots \mathcal{L}_k^{m_k} \mathcal{V}$.

v) $\mathcal{V} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \ldots \oplus \mathcal{M}_p \oplus (\mathcal{L}_p \mathcal{L}_p^{m_p-1} \ldots \mathcal{L}_1 \mathcal{V})$.

Proof. According to point (v) of section 2.1 we have: if $\mathcal{L}$ is singular on $\mathcal{V}$ then $\mathcal{V}$ is the direct sum of the set of all vectors in $\mathcal{V}$ annihilated by $\mathcal{L}^m$ and the $\mathcal{L}^m$-image of $\mathcal{V}$ (where $m$ is defined as in point (i) of section 2.1). Thus ($\mathcal{V}$ for $\mathcal{V}$, $\mathcal{M}_i$ for $\mathcal{F}$):
\[ \mathbf{Y} = \mathbb{R}_1 \oplus \mathbf{l}_1^m \mathbf{Y} \]

Now take \( \mathbf{l}_1^m \mathbf{Y} \) for \( \mathbf{Y} \) and \( \mathbf{l}_2 \) for \( \mathbf{L} \). Using the result of point (iv) in this section we find

\[ \mathbf{l}_1^m \mathbf{Y} = \mathbb{R}_2 \oplus \mathbf{l}_2^m \mathbf{l}_1^m \mathbf{Y} \]

Now take \( \mathbf{l}_2^m \mathbf{l}_1^m \mathbf{Y} \) for \( \mathbf{Y} \) and \( \mathbf{l}_3 \) for \( \mathbf{L} \); etc. The claim made then follows from 'Corollary Associativity' at the end of sec. A.4.

vi) \[ \mathbf{Y} = \mathbb{R}_1 \oplus \mathbb{R}_2 \oplus \cdots \oplus \mathbb{R}_p \]

Proof. According to point (iii) of section 2.1 \( \mathbb{R}_1 \mathbb{R}_1 = \mathbb{R}_1 \), i.e.

\[ \mathbf{l}_1^m \mathbf{Y} = \mathbf{l}_1^m \mathbf{Y} \]  

By repeated application of this result and of the commutativity of the "factors" in \( \mathbf{l}_1^m \mathbf{Y} \) it follows that

\[ \mathbf{l}_p^m \mathbf{l}_{p-1}^m \cdots \mathbf{l}_1^m \mathbf{Y} = \mathbf{l}_p^m \mathbf{l}_{p-1}^m \cdots \mathbf{l}_1^m \mathbf{Y} \]

So putting \( \mathbf{l}_p^m \cdots \mathbf{l}_1^m \mathbf{Y} = \mathbf{W} \) we find

\[ \mathbf{l}_p^{n_p} \cdots \mathbf{l}_2^{n_2} \mathbf{l}_1^{n_1} = \mathbf{W} \]  

Similarly \( \mathbf{l}_p^{n_p} \cdots \mathbf{l}_2^{n_2} \mathbf{l}_1^{n_1} = \mathbf{W} \),

where \( \mathbf{l}_p^{n_p} \cdots \mathbf{l}_2^{n_2} \mathbf{l}_1^{n_1} = (\mathbf{J} - \lambda_1 \mathbf{J})^{n_p} \cdots (\mathbf{J} - \lambda_2 \mathbf{J})^{n_2} (\mathbf{J} - \lambda_3 \mathbf{J})^{n_1} = \mathbf{W} \)

so that \( \mathbf{Y} = \{d\} \) is the value of the characteristic polynomial at \( \mathbf{J} \). Thus

\[ \mathbf{W} = \{d\} \].
2.3. A diagonal block matrix representation of linear mappings

\[ R \] was defined to be \( R((J - \lambda I)^{m_1}) = \{ v \in V : (J - \lambda I)^{m_1} v = 0 \} = \{ v \in V : (J - \lambda I)^k v = 0 \text{ for some } k \} \quad (i = 1, \ldots, p) \]

with \( m_1 \) defined as in the beginning of section 2.2. But what is the dimension of \( R \)? We shall answer this question at the same time as we find a diagonal block matrix representation of \( J \), i.e., a particular type of basis.

Denote \( \dim R \) by \( \gamma_i \) \((i = 1, \ldots, p)\). Then because of sec. 2.2, pt. (vi), and sec. A.4, under 'Dimension of a direct sum', \( \sum_{i=1}^{p} \gamma_i = n = \dim V \). Introducing the notation

\[ \mu_i = \begin{cases} 0 & \text{if } i = 1 \\ \sum_{h=1}^{i-1} \gamma_h & \text{if } i = 2, \ldots, p \end{cases} = \mu_i + \gamma_i = \gamma_1 + \gamma_2 + \cdots + \gamma_i \quad (i = 1, \ldots, p), \]

let \( \{ x_{\mu_i+1}, x_{\mu_i+2}, \ldots, x_{\mu_i+\gamma_i} \} \) be a basis of \( R \). Then because of sec. 2.2, pt. (vi):

\[ B = \{ x_1, x_{\gamma_1}, x_{\mu_2+1}, \ldots, x_{\mu_2+\gamma_2}, \ldots, x_{\mu_p+1}, \ldots, x_{\mu_p+\gamma_p} \} \]

is a basis of \( V \).

Since \( J R \subseteq R \) (sec. 2.2, pt. (1)) we have

\[ J x_{\mu_i+s} = \sum_{s=1}^{\gamma_i} \alpha_{i+s, i+s} x_{\mu_i+s} \quad (r=1, \ldots, \gamma_i; i=1, \ldots, p), \]

that is

\[ J x_h = \sum_{g=1}^{p} \alpha_{gh} x_g \quad (h = 1, \ldots, n) \]

with \( \alpha_{gh} = 0 \) unless \((g, h)\) belongs to any one of the diagonal blocks described by

\[ \left\{ \begin{array}{l} \mu_i + 1 \leq g \leq \mu_i + \gamma_i \\ \mu_i + 1 \leq h \leq \mu_i + \gamma_i \end{array} \right\} \quad (i = 1, \ldots, p) \]

Hence the matrix representing \( J \) re the basis \( B \) is of the form:
where $A_i$ is $(v_i \times v_i)$ matrix $(i = 1, \ldots, p)$.

Consequently, according to section 1.4 the characteristic polynomial of $J$ is

$$\det(\lambda I - A) = \prod_{i=1}^{p} \det(\lambda I_i - A_i),$$

where $I_i$ is $(v_i \times v_i)$ identity matrix. Note that $A_i$ is the matrix representing the restriction of $J$ to $\mathbb{R}_i$ (cf. equation 2.3,1), hence $\det(\lambda I_i - A_i)$ is the characteristic polynomial of this restriction of $J$ to $\mathbb{R}_i$. On the other hand, $(\lambda J - J)$ is singular on $\mathbb{R}_i$ if and only if $\lambda = \lambda_i$ (as is easily seen by means of the argument used in sec. 2.2, pt. (iii))

so the characteristic polynomial of the restriction of $J$ to $\mathbb{R}_i$ must be

$$(\lambda - \lambda_i)^{n_i}$$

(degree of characteristic polynomial of a linear mapping : $U \to U$ equals dim $U) = \det(\lambda I_i - A_i)$. Hence, the characteristic polynomial of $J$ is

$$\det(\lambda I - A) = \prod_{i=1}^{p} \det(\lambda I_i - A_i) = \prod_{i=1}^{p} (\lambda - \lambda_i)^{n_i}.$$ Therefore $n_i = \dim \mathbb{R}_i$ is the exponent of the power of $(\lambda - \lambda_i)$ in the characteristic polynomial of $J$; for $i = 1, 2, \ldots, p$ (In sec. 1.4 we used $n_i$ for these exponents).

We shall now show that the next step in finding a basis in which $J$ has a 'nice' matrix representation depends on the possibility of finding such a basis for a nilpotent linear mapping. In fact, since $A_i$ represents the restriction of $J$ to $\mathbb{R}_i$, relative to a basis of $\mathbb{R}_i$, we know that $(\lambda_i I_i - A_i)$ represents the restriction of $(\lambda_i J - J)$ relative to the same basis. Now $(\lambda_i J - J)$ is nilpotent on $\mathbb{R}_i$, because of $(\lambda_i J - J)^{m_i} \mathbb{R}_i = \{0\}$, so the matrix $\lambda_i I_i - A_i \overset{\text{def}}{=} F_i$ is nilpotent, more precisely: $F_i^{m_i} = 0$. Suppose we find a basis for $\mathbb{R}_i$ re which the restriction of $(J - \lambda_i J)$ to $\mathbb{R}_i$ has a nice matrix representation, say $M_i$.

We know that then a matrix $C_i$ exists such that $C_i^{-1} F_i C_i = M_i$, which entails that
\[ c_1^{-1} (\lambda_1 I_1 + F_1) c_1 = c_1^{-1} A_1 c_1 \]

\[ \lambda_1 I_1 + M_1 = c_1^{-1} A_1 c_1 , \]

hence

\[
\begin{pmatrix}
\lambda_1 I_1 + M_1 \\
\lambda_2 I_2 + M_2 \\
\vdots \\
\lambda_p I_p + M_p
\end{pmatrix}
= \begin{pmatrix}
c_1^{-1} \\
\cdots \\
c_p^{-1}
\end{pmatrix}
\begin{pmatrix}
A_1 \\
\cdots \\
A_p
\end{pmatrix}
\begin{pmatrix}
c_1 \\
\cdots \\
c_p
\end{pmatrix}
\]

The matrix in the first member is obviously similar to \( A \), that is, it represents \( A \) in a suitable basis. If the \( M_i \) are in a 'nice' form, then certainly the \( \lambda_i I_i + M_i \) are too. So the next section is about a canonical representation of nilpotent linear mappings. The section after that will combine all pieces, and demonstrate how a canonical form can be constructed for an arbitrary square matrix.
2.4. A canonical representation of nilpotent linear mappings

Let $U$ be an $n$-dimensional vector space and $L$ be a nilpotent linear mapping on $U$, i.e., an integer $k > 0$ exists such that

$$L^k U = \{0\} \Rightarrow \mathcal{N}(L^k) = U$$

for $k \geq k$. Note that a nilpotent $L$ must be singular: if $L^k U = U$ then $L^k U = U$ for all $k \Rightarrow$ contradiction. Let, as before, $m$ be the smallest integer with $\mathcal{N}(L^m) = \mathcal{N}(L^{m+1})$, so that

$$\mathcal{N}(L) \subset \mathcal{N}(L^2) \subset \ldots \subset \mathcal{N}(L^{m-1}) \subset \mathcal{N}(L^m) = \mathcal{N}(L^{m+s}) \subset U; \quad s > 0.$$ 

Clearly $m$ is also the smallest integer for which $\mathcal{N}(L^m) = U$, so the above-defined integer $k$ equals $m$.

Now we will construct a basis of $U$ so that the matrix of $L$ is that basis is nice and simple.

First observe that $\dim \mathcal{N}(L^m) = \dim U = n$. Next observe that all $s_i$ ($i = 1, \ldots, m$) defined below are strictly positive:

$$\begin{align*}
\left\{ \begin{array}{l}
 n - \dim \mathcal{N}(L^{m-1}) \overset{\text{def}}{=} s_m > 0 \\
 \dim \mathcal{N}(L^i) - \dim \mathcal{N}(L^{i-1}) \overset{\text{def}}{=} s_i > 0 \quad (i = 2, \ldots, m-1) \\
 \dim \mathcal{N}(L) \overset{\text{def}}{=} s_1 > 0
\end{array} \right.
\end{align*}$$

(2.4.1)

Clearly

(2.4.1a)

$$\sum_{i=1}^{m} s_i = n.$$ 

Define $J_1$ as the $1 \times 1$ matrix $[0]$.

Define $J_k$ as the $k \times k$ matrix which has all entries equal to $0$, except those in the first superdiagonal which consists of a string of $(k - 1)$ 1's. For example

$$J_4 = \begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0
\end{bmatrix}$$
We will prove:

1) \( s_m \leq s_{m-1} \leq s_{m-2} \leq \cdots \leq s_2 \leq s_1 \)

2) A basis of \( U \) exists such that the matrix of \( J \) re this basis is of the following form:

\[
\begin{array}{cccc}
    s_m & s_{m-1} & \cdots & s_2 \\
    & s_{m-1} & \cdots & s_2 \\
    & & \ddots & \vdots \\
    & & & s_2 \\
\end{array}
\]

Whenever \( s_i - s_{i+1} = 0 \) it is understood that there are no blocks \( J_i \).

Proof. To simplify notation we will discuss a special case. Assume

\[ m = 4, \; s_4 = 2, \; s_3 - s_4 = 1, \; s_2 - s_3 = 1, \; s_1 - s_2 = 2, \; \text{so } n = \sum_1^4 s_i = 15. \]

Then since \( \mathcal{R}(s^3) \subset \mathcal{R}(s^4) = U \) and \( \dim \mathcal{R}(s^3) \) is \( s_4 = 2 \) less than \( \dim U \), we know (sec. A.4, 'Extension of a linearly independent set to a basis') that there is (even an infinite number of such sets) a set of \( s_4 = 2 \) vectors \( e_{\mathcal{R}(s^4)} \) and \( f_{\mathcal{R}(s^3)} \), which are mutually independent and no linear combination of which \( e_{\mathcal{R}(s^4)} \). . .
Let $b_{41}$ and $b_{42}$ be two such vectors. Then

1°) $s^0 b_{41} = \emptyset$ and $s^k b_{41} \neq \emptyset$, $s^k b_{41} \in \mathbb{R}(s^{4-k})$,

$s^k b_{41} \notin \mathbb{R}(s^{3-k})$ for $0 \leq k \leq 3$ and $i = 1, 2$. ($\mathbb{R}, s^0 \overset{\text{def}}{=} \{\emptyset\}$)

2°) The vectors in the first two columns of the following array are linearly dependent:

\[
\begin{array}{ccc}
    b_{41} & b_{42} & b_{41} & b_{42} & b_{31} & b_{31} & b_{21} & b_{21} & b_{11} & b_{11} & b_{12} & b_{12} \\
    s b_{41} & s b_{42} & s b_{32} & s b_{31} & s b_{21} & b_{31} & s b_{21} & b_{11} & b_{12} & b_{12} & b_{12} & b_{12}
\end{array}
\]

Indeed, assume $\sum_{i=0}^3 \alpha_{41i} s^i b_{41} + \sum_{i=0}^3 \alpha_{42i} s^i b_{42} = \emptyset \quad (*)$

Apply $s^3$ to both sides and find

\[s^3 (\alpha_{410} b_{41} + \alpha_{420} b_{42}) = \emptyset\]

Hence $\alpha_{410} b_{41} + \alpha_{420} b_{42} \in \mathbb{R}(s^3)$; but $b_{41}$ and $b_{42}$ were explicitly assumed to have the property that no linear combination would belong to $\mathbb{R}(s^3)$; so $\alpha_{410} = 0 = \alpha_{420}$. Now apply $s^2$ to both sides and similarly find that $\alpha_{411} = 0 = \alpha_{421}$, etc. Claim under (2°) is proved.

If $b_{41}$ and $b_{42}$ are linearly independent, they belong to $\mathbb{R}(s^3)$, not to $\mathbb{R}(s^2)$, and no linear combination of them belongs to $\mathbb{R}(s^2)$ (assume the opposite, then $s^3 b_{41}$ and $s^3 b_{42}$ would be linearly dependent, which would contradict point (2°) which was just proved). This proves that

\[\dim \mathbb{R}(s^3) - \dim \mathbb{R}(s^2) = s_3 \geq s_4\]
In our case \( s_4 = 2 \), \( s_3 = 3 \). So there exists one other vector \( e^{(3)} \), \( b_{31} \) say (again there is a huge choice), such that \( \xi^0 b_{41}, \xi^0 b_{42}, b_{31} \) are linearly independent and no linear combination of them belongs to \( \mathbb{R}(3^2) \). The reader will prove

\[ 3^0) \xi^k b_{31} \in \mathbb{R}(3^3-k) ; \xi^k b_{31} \notin \mathbb{R}(3^2-k) \text{ for } 0 \leq k \leq 2. \ (\mathbb{R}(3^0) \text{ def } \emptyset) \]

\[ 4^0) \text{ The vectors in the first three columns of the above array are independent.} \]

\[ \xi^2 b_{41}, \xi^2 b_{42}, \xi b_{31}, \text{ are linearly independent, } \xi \mathbb{R}(3^2), \xi \mathbb{R}(3), \text{ no linear combination belongs to } \mathbb{R}(3). \ \text{Therefore } s_2 \geq s_3. \text{ In our case } s_2 - s_3 = 1. \]

So there exists one other vector which shares these features with \( \xi^2 b_{41}, \xi^2 b_{42}, \xi b_{31} \); call it \( b_{31} \). The reader will prove that \( \xi b_{31} \in \mathbb{R}(3), \xi b_{21} \notin \mathbb{R}(3) \), and also that the vectors in the first four columns of the above array are linearly independent.

Finally the fact that \( s_1 - s_2 = 2 \) entails that two more vectors, say \( b_{11} \) and \( b_{12} \), have to be taken in, in order that the bottom line of this array may be a basis for \( \mathbb{R}(3) \).

As a result we have 15 linearly independent vectors, which therefore form a basis of \( \mathbb{U} \); the 6 in the bottom line are a basis of \( \mathbb{A}(3) \), the 4 in the next line complement this basis into a basis of \( \mathbb{A}(3^2) \); the 3 in the next line complement this basis into a basis of \( \mathbb{A}(3^3) \); the 2 at the top complement this basis into a basis of \( \mathbb{A}(3^4) = \mathbb{U} \).

Note that the number of vectors in each row and in each column depends on \( m, s_1, s_2, ..., s_m \); i.e., on the dimensions of the sets \( \mathbb{A}(3), \mathbb{A}(3^2), \mathbb{A}(3^3), \ldots \); not in the least on the matrix representing \( \xi \). This means that the lengths of the rows and columns of this array are the same for all matrices which are similar to each other.

Now enumerate the vectors of the above array in the following order: first column, bottom to top; second column, bottom to top, etc. For instance \( \xi b_{21} \) would be \( v_{12} \), \( b_{21} \) would be \( v_{13} \) (Note that the subscripts of the \( b \) are double subscripts, of the \( v \) are single: the subscript of \( b_{21} \) is two-one; of \( v_{12} \) is twelve). These vectors \( v \) now satisfy the following equations:
\[ v_4 = v_3, \quad v_8 = v_7 \\
\quad v_2 = v_1, \quad v_6 = v_5, \quad v_{11} = v_{10} \\
\quad v_7 = v_6, \quad v_{10} = v_9, \quad v_{13} = v_{12} \\
\quad v_1 = 0, \quad v_5 = 0, \quad v_9 = 0, \quad v_{12} = 0, \quad v_{14} = 0, \quad v_{15} = 0 \]

Hence, in accordance with equation (1.1.1) the matrix \([\alpha_{ij}]\) representing \( f \) with respect to the basis \([v_1, v_2, \ldots, v_{15}]\) is as follows:

\[
\begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{bmatrix}
\]

and 0 at all other entries.

**Note.** The actual construction of such a basis is harder than it might seem because of the condition that, e.g., no linear combination of \( b_{41} \) and \( b_{42} \) may belong to \( A(f^3) \). Although the existence of such vectors can be asserted without having an actual basis of \( A(f^3) \) available, this checking requires more doing (either one has first to choose bases for \( A(f^k) \), \( k = 1, 2, \ldots \), (cf. Nehring, p. 104) against which this type of conditions can then be checked; or one selects e.g. \( b_{41} \) and \( b_{42} \) on a trial-and-error basis checking if \( f^3(\beta_1 b_{41} + \beta_2 b_{42}) = 0 \) indeed entails \( \beta_1 = 0 = \beta_2 \), and if it does not, then chooses another \( b_{41} \) and/or \( b_{42} \).

**Note.** The **Segre Characteristic** of a nilpotent mapping gives the subsequent orders of the blocks \( J_k \); in the above example \((4, 4, 3, 2, 1, 1)\).
The Weyr characteristic of a nilpotent mapping is \( \{ s_1, s_2, s_3, \ldots, s_m \} \), in the above example \( \{6, 4, 3, 2\} \). Obviously the Weyr characteristic follows from the row lengths of our above array of vectors, the Segre characteristic from its column lengths.
2.5. The Jordan normal form of an $n \times n$ matrix

We shall now combine the results of sections 2.3 and 2.4. Let $\mathcal{J}$ be a linear transformation mapping an 8-dimensional linear space into itself, with characteristic polynomial

$$
\chi(\lambda) = (\lambda + 1)^5 (\lambda - 2)^3
$$

The so-called Jordan normal matrix of $\mathcal{J}$ has then for its main diagonal

$$
\begin{array}{cccccc}
-1 & -1 & -1 & -1 & 2 & 2 \\
2 & 2 & & & & \\
\end{array}
$$

in keeping with the formula at the end of section 2.3. $(n_1 = 5 ; \ n_2 = 3)$

The first diagonal block $(5 \times 5)$ will then be completed by adding $M_1$ to it, where $M_1$ is a 'nice' representation of the restriction of $\mathcal{J} - \lambda_1 \mathcal{J}$ ($= \mathcal{J} + \mathcal{J}$ in our case) to $\mathcal{N}_1$. Since $\mathcal{J} - \lambda_1 \mathcal{J}$ ($= \mathcal{J} + \mathcal{J}$) is nilpotent on the 5-dimensional linear space $\mathcal{N}_1$ (which has the role of $\mathcal{U}$ in section 2.4), $M_1$ can be found by applying sec. 2.4. We saw that $M_1$ will consist of zeros except possibly for some 1's on the super diagonal, depending on the value of $m_1$ and on the dimensions of the null-spaces of the powers of $F_1 = A_1 - \lambda_1 I_1$, i.e., of the restriction of $\mathcal{J} - \lambda_1 \mathcal{J}$ to $\mathcal{N}_1$.

Suppose in our case $m_1 = 5$ (i.e., $(\mathcal{J} + \mathcal{J})^5 \mathcal{N}_1 = \{0\}$ for $\lambda \geq 5$). Then the reader will prove that $\dim \mathcal{N}( (\mathcal{J} + \mathcal{J})^4 ) = i$, $\varepsilon_i = i$ for $1 \leq i \leq 5$. So

$$
M_1 = \begin{bmatrix}
0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
\end{bmatrix}
$$

\[\]
Suppose \( m_1 = 2 \), and \( \dim \mathfrak{A} (\mathfrak{S} + \mathfrak{J}) = 3 \). Show that now

\[
M_1 = \begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0
\end{bmatrix}
\]

Note that in the first case the scheme connecting Segre and Weyr characteristics is as follows: and in the second case:

\[
\begin{array}{cccc}
| s_5 = 1 & \cdot & \cdot & \cdot \\
| s_4 = 1 & \cdot & \cdot & \cdot \\
| s_3 = 1 & \cdot & \cdot & \cdot \\
| s_2 = 1 & \cdot & \cdot & \cdot \\
| s_1 = 1 & \cdot & \cdot & \cdot \\
\hline
\text{s} & \text{s} & \text{s} & \text{s}
\end{array}
\]

As to the second diagonal block \((3 \times 3)\), it will have to be completed by adding \( M_2 \) to it, where \( M_2 \) is a nice representation of the restriction of \( \mathfrak{T} - \lambda_2 \mathfrak{J} = (\mathfrak{T} - 2 \mathfrak{J}) \) to \( \mathfrak{G}_2 \). Suppose \( m_2 = 1 \), i.e., \((\mathfrak{T} - 2 \mathfrak{J}) \mathfrak{G}_2 = \{0\} \). Then \( \dim \mathfrak{G}_2 = 3 = s_1 \), and the above pattern is:

\[
\begin{array}{cccc}
W & s_1 = 3 & \cdot & \cdot \\
\hline
& 1 & 1 & 1 \\
& s & &
\end{array}
\]

and

\[
M_2 = \begin{bmatrix}
0 & 0 \\
0 & 0 \\
0 & 0
\end{bmatrix}
\]

A transformation \( \mathfrak{T} \) which combines \( \cdot \), \( \cdot \), for \((\mathfrak{T} - \lambda_1 \mathfrak{J})\); \( \lambda_1 = -1 \); with \( \cdot \), \( \cdot \), for \((\mathfrak{T} - \lambda_2 \mathfrak{J})\); \( \lambda_2 = 2 \); has Jordan normal representation:

\[
\begin{bmatrix}
-1 & 1 & & \\
-1 & 0 & & \\
-1 & 1 & & \\
-1 & 0 & & \\
-1 & 0 & & \\
2 & 0 & & \\
2 & 0 & & 
\end{bmatrix}
\]
A transformation $T$ which combines for $(\lambda_1 \neq 0; \lambda_1 = -1)$ with

... for $(\lambda_2 \neq 0; \lambda_2 = 2)$ has Jordan normal representation:

$$
\begin{bmatrix}
-1 & 1 \\
-1 & 1 \\
-1 & 1 \\
-1 & 0 \\
2 & 0 \\
2 & 0
\end{bmatrix}
$$

**Exercise.** What is the Jordan normal form of the matrix

$$
\begin{bmatrix}
1 & 1 & 0 & 1 \\
0 & 2 & 0 & 0 \\
-1 & 1 & 2 & 1 \\
-1 & 1 & 0 & 3
\end{bmatrix}
$$

and of the matrix

$$
\begin{bmatrix}
3 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 1 & -1 \\
0 & 0 & 3 & -1 & 0 \\
0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 2
\end{bmatrix}
$$
The principal purpose of this appendix is to make this report to a certain degree self-contained. To this end we have listed axioms, definitions, and standard results in so far as they have been used in the main part of this report. Sometimes a proof is provided, sometimes a plausibility argument or a heuristic motivation. For complete proofs see the textbooks listed in the section on literature.

Since in the definition of linear spaces the algebraic concept of a field plays an essential role, section A.1 discusses the axioms of fields. The field needed for the derivation of the Jordan normal form consists of the complex numbers. These have many additional properties, of which a few (viz., those needed in section A.3 and A.4) are given in section A.2. Then section A.3 gives the axioms for linear spaces, section A.4 discusses relevant properties of linear spaces, section A.5 of linear mappings.
A.1. Fields

Denote the set of complex numbers by \( \phi \). The system of complex numbers has the following properties.

I.1. A mapping: \( \phi \times \phi \rightarrow \phi \), called addition is defined.

Notation \( \alpha_1 + \alpha_2 = \beta \). For any \( \alpha_1 \in \phi \), \( \alpha_2 \in \phi \) their sum also \( \in \phi \).

I.2. Associative law of addition. For any \( \alpha_1 \in \phi \), \( \alpha_2 \in \phi \), \( \alpha_3 \in \phi \):

\[
(\alpha_1 + \alpha_2) + \alpha_3 = \alpha_1 + (\alpha_2 + \alpha_3)
\]

I.3. Existence of solution of certain equations: for any \( \alpha \in \phi \), \( \beta \in \phi \):

there exists at least one \( \xi \in \phi \) with \( \alpha + \xi = \beta \)

and at least one \( \eta \in \phi \) with \( \eta + \alpha = \beta \).

Corollaries:

a: existence of at least one right additive identity element, \( 0_r \in \phi \);

i.e., for any \( \alpha \in \phi \) there is at least one \( 0_r \in \phi \) with \( \alpha + 0_r = \alpha \).

Note. Here the possibility is left open that there might be more than one right identity element for a given element of \( \phi \), or different ones for different elements.

b: existence of at least one left additive identity element, \( 0_L \in \phi \);

i.e., for any \( \alpha \in \phi \) there is at least one \( 0_L \in \phi \) with \( 0_L + \alpha = \alpha \).

Note. Similar to the preceding note. However, because in I.3 the existence of solutions to both equations is postulated:

c: a \{ right \} additive identity element for one element \( \in \phi \) is a \{ right \} additive identity element for all elements \( \in \phi \).

Indeed, let \( 0_r \) be a right identity for \( \alpha_1 \): \( \alpha_1 + 0_r = \alpha_1 \), let \( \eta \) be a solution to \( \eta + \alpha_1 = \alpha_2 \), \( \alpha_2 \neq \alpha_1 \);

then \( \alpha_2 + 0_r = \eta + \alpha_1 + 0_r = \eta + \alpha_1 = \alpha_2 = \alpha_2 + 0_r = \alpha_2 \).

Similar proof for \( 0_L \). Now we shall prove:
d: uniqueness of additive identity element: there is only one right identity, only one left identity, and they are the same, denoted by 0.

Indeed, let \(0'_r\) and \(0''_r\) be two right identity elements, and \(0'_l\) some left identity element, then because of corollary 2 (elaborate)

\[0'_r = 0'_l + 0'_r = 0'_l, \text{ and } 0''_r = 0'_l + 0''_r = 0'_l, \text{ hence } 0'_r = 0''_r = 0_r,\]
say. Similar proof holds for \(0'_l = 0''_l = 0_l,\) say. Finally,

\[0'_l + 0_l = 0_r = 0,\]
say. 0 is the additive identity element, called zero.

e: for any \(a \in \Phi\) there is at least one right additive inverse \((-a)_r\) such that \(a + (-a)_r = 0\).

f: for any \(a \in \Phi\) there is at least one left additive inverse \((-a)_l\) such that \((-a)_l + a = 0\).

g: uniqueness of additive inverse: for any \(a \in \Phi\) there is only one right additive inverse, only one left additive inverse, and the two are the same, denoted by \((-a)_\).

In fact, let \(a + (-a)'_r = 0, \ a + (-a)''_r = 0,\) \((-a)'_l + a = 0,\) then

\[(-a)'_l + a + (-a)'_r = (-a)'_l + a + (-a)'_r = (-a)''_r = (-a)'_r = (-a)_r,\]

which proves that \((-a)'_r = (-a)''_r = (-a)_r,\) say.

Similar proof for uniqueness of left additive inverse.

Finally: \((-a)'_l = (-a)'_l + 0 = (-a)'_l + a + (-a)'_r = 0 + (-a)_r = (-a)_r = (-a)\)

Note. The right and left inverse are the same, even without commutativity.

Exercise. Prove that \((a+\beta) = (-\beta) + (-a).\) Prove that \((-(-\alpha)) = \alpha.\)

Note. Notation: \(a - \beta \overset{\text{def}}{=} a + (-\beta).\) Show that \(a - \alpha = 0.\)
h: uniqueness of solution of $\alpha + \xi = \beta$

uniqueness of solution of $\eta + \alpha = \beta$

Indeed, given $\alpha \in \phi$, $\beta \in \phi$, let $\alpha + \xi' = \beta = \alpha + \xi''$

Then $(-\alpha) + \alpha + \xi' = (-\alpha) + \beta = (-\alpha) + \alpha + \xi''$

So $\xi' = (-\alpha) + \beta = \xi'' = \xi$

Similarly $\eta = \beta + (-\alpha)$

Note. Without commutativity $\xi$ need not be equal to $\eta$.

Note. The three postulates I.1, I.2, I.3 are those of a(n additive) group.

I.4. Commutative law of addition. For all $\alpha_1 \in \phi$, $\alpha_2 \in \phi$:

$$\alpha_1 + \alpha_2 = \alpha_2 + \alpha_1$$

Note. The four postulates I.1, I.2, I.3, I.4 are those of a commutative group, also called an Abelian group (under addition).

Complex numbers can also be multiplied:

II.1. A mapping: $\phi \times \phi \rightarrow \phi$, called multiplication is defined.

Notation $\alpha_1 \alpha_2 = \gamma$ (alternative $\alpha_1 \cdot \alpha_2 = \gamma$)

For any $\alpha_1 \in \phi$, $\alpha_2 \in \phi$ their product also $\in \phi$.

II.2. Associative law of multiplication. For any $\alpha_1 \in \phi$, $\alpha_2 \in \phi$, $\alpha_3 \in \phi$:

$$(\alpha_1 \alpha_2) \alpha_3 = \alpha_1 (\alpha_2 \alpha_3)$$

The addition and multiplication interact according to:

III. Distributive law. For any $\alpha_1 \in \phi$, $\alpha_2 \in \phi$, $\alpha_3 \in \phi$:

III.1. $\alpha_1 (\alpha_2 + \alpha_3) = \alpha_1 \alpha_2 + \alpha_1 \alpha_3$

III.2. $(\alpha_2 + \alpha_3) \alpha_1 = \alpha_2 \alpha_1 + \alpha_3 \alpha_1$

Corollaries.

8. $\alpha(\beta - \gamma) = \alpha \beta - \alpha \gamma$ for any $\alpha \in \phi$, $\beta \in \phi$, $\gamma \in \phi$.

Indeed, both members can be shown to satisfy $\eta + \alpha \gamma = \alpha \beta$.

(Use definition given of the notation $(\alpha - \beta)$ at the end of Corollary 8 to point I.3).

$$(\beta - \gamma) \alpha = \beta \alpha - \gamma \alpha$$ for any $\alpha \in \phi$, $\beta \in \phi$, $\gamma \in \phi$.

Indeed, both members can be shown to satisfy $\eta + \gamma \alpha = \beta \alpha$. 

b. \( \alpha \cdot 0 = 0 \) for any \( \alpha \in \Phi \).

Indeed, \( \alpha \cdot 0 = \alpha (\beta - \beta) = \alpha \beta - \alpha \beta = 0 \)

\( 0 \cdot \alpha = 0 \) for any \( \alpha \in \Phi \).

Indeed, \( 0 \cdot \alpha = (\beta - \beta) \alpha = \beta \alpha - \beta \alpha = 0 \).

c. For any \( \alpha \in \Phi \), \( \beta \in \Phi \)

\( (-\alpha) \cdot \beta = -(\alpha \beta) \), for both are the additive inverse of \( (\alpha \beta) \)

\( \alpha \cdot (-\beta) = -(\alpha \beta) \), same argument

\( (-\alpha) \cdot (-\beta) = \alpha \beta \), for both are the additive inverse of \( (-\alpha) \cdot \beta = -(\alpha \beta) \).

Note. The postulates I.1, I.2, I.3, I.4, II.1, II.2, III are those of a ring.

If for \( \alpha \in \Phi \), \( \beta \in \Phi \), \( \alpha \beta \) cannot be zero unless either \( \alpha = 0 \) or \( \beta = 0 \),

the ring is said to be without zero divisors.

If multiplication is also commutative (postulate II.4 below), the ring is said to be commutative.

A commutative ring without zero divisors is often called an integral domain.

II.3. The set \( \Phi \setminus \{0\} = \Phi' \), say, is not empty and is a group under multiplication,

that is:

Existence of solution of certain equations: for any \( \alpha \in \Phi' \), \( \beta \in \Phi' \),

there exists at least one \( \xi \in \Phi' \) with \( \alpha \xi = \beta \)

and at least one \( \eta \in \Phi' \) with \( \eta \alpha = \beta \)

Note. If \( \beta \in \Phi \), \( \beta \notin \Phi' \), i.e., \( \beta = 0 \), then III, corollary b shows that \( \xi = 0 \), \( \eta = 0 \) are solutions. If \( \alpha = 0 \), \( \beta \neq 0 \), no \( \xi \in \Phi \) or \( \eta \in \Phi \), respectively, can satisfy these equations (same coroll. b).

If \( \alpha = 0 \), \( \beta = 0 \), every \( \xi \in \Phi \), \( \eta \in \Phi \) satisfies these equations.

If \( \beta \neq 0 \), neither \( \xi \) nor \( \eta \) can be zero.

Corollaries.

a: existence of at least one right multiplicative identity element, \( 1_r \in \Phi' \),

i.e., for any \( \alpha \in \Phi' \) there is at least one \( 1_r \in \Phi' \) with \( 1_r \alpha = \alpha \).

Note. See note in I.3, a.

b: existence of at least one left multiplicative identity element, \( 1_l \in \Phi' \),

i.e., for any \( \alpha \in \Phi' \) there is at least one \( 1_l \in \Phi' \) with \( 1_l \alpha = \alpha \).
Note. See note in I.3, b. Now because in II.3 the existence of solutions to both equations is postulated:

c: a \{\text{right}\} multiplicative identity element for one element \( e \in \mathfrak{g} \) is a \{\text{right}\} multiplicative identity element for all elements \( e \in \mathfrak{g} \).

Proof. Replica of I.3, c, with \( \cdot \) instead of \( + \), and \( 1 \) instead of \( 0 \).

d: uniqueness of multiplicative identity element: there is only one right identity, only one left identity, and they are the same, denoted by \( 1 \), called unity.

Proof. Replica of I.3, d, with \( \cdot \) instead of \( + \), and \( 1 \) instead of \( 0 \).

e: for any \( \alpha \in \mathfrak{g} \) there is at least one right multiplicative inverse, \( \alpha_\cdot^{-1} \) such that \( \alpha \cdot \alpha_\cdot^{-1} = 1 \).

f: for any \( \alpha \in \mathfrak{g} \) there is at least one left multiplicative inverse, \( \alpha_\cdot^{-1} \) such that \( \alpha_\cdot^{-1} \cdot \alpha = 1 \).

g: uniqueness of multiplicative inverse: for any \( \alpha \in \mathfrak{g} \) there is only one right multiplicative inverse, only one left multiplicative inverse, and the two are the same, denoted by \( \alpha^{-1} \).

Proof. The reader will model the proof after I.3, g.

Note. The right and left inverse are the same, even without commutativity.

Exercise. Prove that \((\alpha \beta)^{-1} = \beta^{-1} \cdot \alpha^{-1}\). Prove \((\alpha^{-1})^{-1} = \alpha\).

Notation. \( \frac{\alpha}{\beta} \overset{\text{def}}{=} \alpha \cdot \beta^{-1}; \frac{\alpha}{\beta} \overset{\text{def}}{=} \alpha \cdot \beta^{-1} \). Show that \( \frac{\alpha}{\alpha} = 1 \).

h: No zero divisors. Let \( \alpha \beta = 0 \). Assume \( \alpha \neq 0 \), \( \beta \neq 0 \) (i.e., assume our ring has zero divisors) Then \( \alpha^{-1} \cdot \alpha \beta = \beta = 0 \); contradiction.

Note. A ring satisfying postulate II.3 is called a division ring or a skew-field. We saw that a skew-field cannot have zero divisors.
II.4. Commutative law of multiplication. For all $\alpha_1 \in \Phi, \alpha_2 \in \Phi$:

$$\alpha_1 \alpha_2 = \alpha_2 \alpha_1$$

Note. A skew-field satisfying postulate II.4 is called a field or a rational domain. (Some authors use field instead of skew-field, and commutative field instead of field).

A ring satisfying postulate II.4 is a commutative ring.

Note. The postulates I, II, III are shared by complex numbers, real numbers, rational numbers. Of the various properties which $\Phi$ has in addition to I, II, III, we will discuss only those cited in the discussion of Linear spaces and those used in deriving the Jordan normal form.
A.2. Complex numbers

The field axioms listed in section A.1 are satisfied by the systems of rational numbers, of real numbers, and of complex numbers, to name a few examples. We have assumed that the reader is familiar with these various types of numbers. In the remaining sections of the appendix and in the main part of this report we need only the following two properties of complex numbers:

a/ The field of complex numbers is algebraically closed, i.e., for any positive integer \( n \), for any polynomial, \( P_n \) say, of the \( n \)th degree, there exist

\[
\begin{align*}
\{ & \text{a complex number } \gamma \\
& \text{p (\leq n) complex numbers } \xi_1, \xi_2, \ldots, \xi_p \\
& \text{p positive integers } n_1, n_2, \ldots, n_p \text{ with } \sum_{k=1}^{p} n_k = n
\end{align*}
\]

such that for all complex \( z \):

\[
P_n(z) = \gamma \cdot \prod_{k=1}^{p} (z - \xi_k)^{n_k}
\]

(**)

In fact, according to the 'Fundamental Theorem of Algebra' all that needs to be postulated in order that this equality may hold for all polynomials \( P_n \), is that the equation \( z^2 + 1 = 0 \) has at least one root: this implies the validity of equality (**)

b/ A one-to-one correspondence exists between the set of complex numbers and the set of ordered pairs of real numbers. The traditional notation for complex numbers \( z \) is:

\[ z = x + iy \quad \text{where} \quad x \in \mathbb{R}, \quad y \in \mathbb{R}, \quad i^2 = -1 \]

Let \( z_1 = x_1 + iy_1 \), \( z_2 = x_2 + iy_2 \), then

\[
\begin{align*}
z_1 + z_2 &= (x_1 + x_2) + i(y_1 + y_2) \\
z_1 z_2 &= (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1)
\end{align*}
\]
A.3. Linear spaces.

A linear space (or alternatively: vector space) is made up of a set \( \mathcal{V} \), whose elements (denoted by lower case Latin letters) are called vectors, and a field, in our case \( \mathbb{F} \), whose elements (denoted by lower case Greek letters) are called scalars, and certain rules according to which the vectors may be related to each other and to the scalars. Following a common abuse of notation we will use the single letter \( \mathcal{Y} \) not only for the set of vectors, but also for the linear space.

More precisely, we say that \( \mathcal{Y} \) is a linear space over the field \( \mathbb{F} \) iff

I.1. A mapping: \( \mathcal{Y} \times \mathcal{Y} \to \mathcal{Y} \), called (vector) addition is defined. Notation

\[ v_1 + v_2 = w \]

For any \( v_1 \in \mathcal{Y}, v_2 \in \mathcal{Y} \), their sum also \( \in \mathcal{Y} \). Again following a common abuse of notation, we will use the same sign, \( + \), for the addition of vectors as for the addition of scalars. No confusion will arise, and certain formulae will even look simpler for it.

I.2. Associative law of (vector) addition. For any \( v_1 \in \mathcal{Y}, v_2 \in \mathcal{Y}, v_3 \in \mathcal{Y} \):

\[ (v_1 + v_2) + v_3 = v_1 + (v_2 + v_3) \]

I.3. Existence of solution of certain equations: for any \( a \in \mathcal{Y}, b \in \mathcal{Y} \):

there exists at least one \( x \in \mathcal{Y} \) with \( a + x = b \)

and at least one \( y \in \mathcal{Y} \) with \( y + a = b \)

Corollaries (proofs same as in sec. A.1, under I.3; here we cite only the most important ones):

\( d \): unique identity element: there is one and only one element \( \in \mathcal{Y} \), denoted by \( \theta \), called the null vector, such that for any \( a \in \mathcal{Y} \):

\[ a + \theta = a \]

\[ \theta + a = a \]

\( g \): unique inverse: for any \( a \in \mathcal{Y} \), there is one and only one element, written \( (-a) \), such that

\[ a + (-a) = \theta \]

\[ (-a) + a = \theta \]

Notation: \( a - b \text{ def } a + (-b) \).
h: uniqueness of solution of \( a + x = b \) \}

uniqueness of solution of \( y + a = b \) \}

I.4. Commutativity of (vector) addition. For any \( v_1 \in \mathbb{V}, v_2 \in \mathbb{V} \):
\[ v_1 + v_2 = v_2 + v_1 \]

Note. The postulates I.1, I.2, I.3, and I.4 together obviously require \( \mathbb{V} \) to be a commutative group under addition.

II.1. A mapping: \( \phi \times \mathbb{V} \rightarrow \mathbb{V} \), called scalar multiplication, is defined.

Notation: given \( \alpha \in \phi, v \in \mathbb{V} \), the value of the mapping is \( \alpha v \in \mathbb{V} \)
(sometimes for convenience written as \( \alpha v \)). This mapping satisfies the conditions:

II.2. \( \alpha(x + y) = \alpha x + \alpha y \) for any \( \alpha \in \phi, x \in \mathbb{V}, y \in \mathbb{V} \).

Corollaries:

\( a \): \( \alpha \theta = \theta \), for any \( \alpha \in \phi \).

In fact, \( x = x + \theta = \alpha x + \alpha \theta = \alpha (x + \theta) = \alpha x + \alpha \theta = \alpha \theta = \theta \).

\( b \): \( \alpha(-x) = -(\alpha x) \), for any \( \alpha \in \phi, x \in \mathbb{V} \).

In fact, \( \theta = \alpha \theta = \alpha (x + (-x)) = \alpha x + \alpha(-x) \).

So \( \alpha(-x) \) is the additive inverse of \( \alpha x = \alpha(-x) = -(\alpha x) \).

II.3. \( (\alpha + \beta)x = \alpha x + \beta x \) for any \( \alpha \in \phi, \beta \in \phi, x \in \mathbb{V} \).

Corollaries:

\( a \): \( Ox = \theta \), for any \( x \in \mathbb{V} \).

In fact, \( \alpha = \alpha + 0 = \alpha x = (\alpha + 0)x = \alpha x + 0x = 0x = \theta \).

\( b \): \( (-\alpha)x = -(\alpha x) \), for any \( \alpha \in \phi, x \in \mathbb{V} \).

In fact, \( \theta = 0x = (\alpha + (-\alpha))x = \alpha x + (-\alpha)x \).

So \( (-\alpha)x \) is the additive inverse of \( \alpha x = (-\alpha)x = -(\alpha x) \).

II.4. \( \alpha(\beta x) = (\alpha \beta)x \), for any \( \alpha \in \phi, \beta \in \phi, x \in \mathbb{V} \).

II.5. \( 1x = x \), for any \( x \in \mathbb{V} \).

Corollary:

\( a \): \( -1x = -x \) for any \( x \in \mathbb{V} \).

Indeed, by corollary \( b \) of II.3 above: \( (-1)x = (-1) \alpha x = -x \).
Note. Point II.1 above postulates that a mapping $\phi \times \mathcal{V} \to \mathcal{V}$, called scalar multiplication, is defined, and II.2 through II.5 proceed to postulate certain properties of this mapping. This does not clarify yet how to define scalar multiplication in specific instances of vector spaces $\mathcal{V}$ over fields $\phi$. One example: let $\mathcal{V}$ be the space of ordered triples of complex numbers (and $\phi$ be the complex number system). Now let $\alpha \in \phi$, $(\xi_1, \xi_2, \xi_3) \in \mathcal{V}$, then the definition

$$\alpha (\xi_1, \xi_2, \xi_3) \overset{\text{def}}{=} (\alpha \xi_1, \alpha \xi_2, \alpha \xi_3),$$

or in matrix notation:

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} \overset{\text{def}}{=} \begin{bmatrix} \alpha \xi_1 \\ \alpha \xi_2 \\ \alpha \xi_3 \end{bmatrix}$$

is easily seen to satisfy all postulates II.1 through II.5.
A.4. Basis, dimension, direct sum

Linear combination of k vectors. Given k vectors \( x_i \in \mathcal{V} \) (\( i = 1, \ldots, k \)), each vector of form \( \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \) (where \( \alpha_1, \ldots, \alpha_k \) are scalars, i.e. belong to \( \mathcal{F} \)) is said to be a linear combination of \( x_1, x_2, \ldots, x_k \). (Note: If the field \( \mathcal{F} \) consists of the rational numbers only, the coefficients \( \alpha_i \) can only be rational numbers). Prove from definition of linear space: if \( x_i \in \mathcal{V} \) then \( \alpha_1 x_1 + \cdots + \alpha_k x_k \in \mathcal{V} \).

Subspace. If \( \mathcal{V} \) is a linear space, not every subset \( \mathcal{U} \) of \( \mathcal{V} \) is a linear space. See the two figures; in both figures \( \mathcal{V} \) is the set of all points \((x, y)\) in the plane, a linear space over the field of real numbers.

\( \mathcal{U} \) is said to be a (linear) subspace of \( \mathcal{V} \) iff

1. \( \mathcal{U} \subseteq \mathcal{V} \) and
2. \( \mathcal{U} \) is itself a linear space (over the same field as \( \mathcal{V} \))

(hence \( \mathcal{U} \) contains all linear combinations of its own vectors.)
Subspace spanned. Let \( \mathcal{U} \) be a subspace of \( \mathcal{V} \), and let \( B \) be a finite set of
vectors in \( \mathcal{V} \). Then one says:

\[
\begin{cases}
\mathcal{U} \text{ is spanned by } B \\
\mathcal{U} \text{ is spanned by the vectors in } B \\
B \text{ spans } \mathcal{U} \\
\text{the vectors in } B \mbox{ span } \mathcal{U}
\end{cases}
\]

 iff

\[
\begin{cases}
(1) \text{ each linear combination of vectors } e \in B \text{ is a vector in } \mathcal{U} , \text{ and} \\
(2) \text{ each vector in } \mathcal{U} \text{ is equal to at least one linear combination of}
\quad \text{vectors } e \in B .
\end{cases}
\]

Example: \( \mathcal{U} \) same as above; \( B \) consists of
the 3 points indicated: \( B = \{ x_1, x_2, x_3 \} \). Show:
any point in the plane is equal to any one of
a number of linear combinations of these
three vectors: \( B \) spans \( \mathcal{U} \). Thus \( \alpha_1 \) and \( \beta_1 \)
can be found such that

\[
\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ ,
\]

whence

\[
(\alpha_1 - \beta_1)x_1 + (\alpha_2 - \beta_2)x_2 + (\alpha_3 - \beta_3)x = \Theta \text{ where not }
\]

\[
\alpha_1 = \beta_1, \alpha_2 = \beta_2, \alpha_3 = \beta_3 .
\]

Linear independence. In order that each vector in \( \mathcal{U} \) equal exactly one linear
combination of the \( k \) vectors \( x_1, x_2, \ldots, x_k \) in \( B \), it is necessary (why not
sufficient?) that \( \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_k x_k = \Theta \) implies \( \gamma_1 = 0 = \gamma_2 = \ldots = \gamma_k \). This
occurrence has received a special name. Consider a set \( B \) of \( k \) vectors
\( x_1, x_2, \ldots, x_k \); then;
The vectors \( x_1, x_2, \ldots, x_k \) are said to be linearly independent
The set \( B \) is said to be linearly independent

iff

\[
\gamma_1 x_1 + \ldots + \gamma_k x_k = \Theta \iff \gamma_1 = 0 = \gamma_2 = \ldots = \gamma_k ,
\]
i.e., iff the only linear combination of these \( k \) vectors which equals \( \Theta \) is the
trivial one (all coeff. zero). The null vector \( \Theta \) is never an element in a linearly
independent set (why?)
**Basis.** A finite set \( B \) of vectors is said to be a basis of the linear space \( \mathbb{U} \) iff 
\[
\exists \quad (1) \quad \text{B spans } \mathbb{U}, \text{ and } \\
(2) \quad \text{B is linearly independent.}
\]
So each vector of \( \mathbb{U} \) is a unique linear combination of the vectors in the basis. The coefficients in that linear combination are called the coordinates relative to that basis. Let \( e_1, e_2, \ldots, e_k \) be the vectors in some basis of \( \mathbb{U} \). If we have proved that some vector of \( \mathbb{U} \) equals:
\[
\sum_{i=1}^{k} \alpha_i e_i = \sum_{i=1}^{k} \beta_i e_i
\]
we can immediately conclude that \( \alpha_i = \beta_i \) (\( i = 1, \ldots, k \)). No basis ever contains the null vector (why?)

**Variety of bases. First example.** The linear space with ordered pairs of real numbers as vectors, and real numbers as scalars has a basis \( B' = \{(3, 0), (0, -1)\} \), but also a basis \( B'' = \{(1, 3), (-1, \frac{1}{2})\} \). **Second example.** The linear space with ordered triples of real numbers as vectors, and real numbers as scalars has a basis \( B^I = \{(1,0,0), (0,1,0), (0,0,1)\} \) and a basis \( B^II = \{(1,1,1), (0,1,1), (0,0,1)\} \).

However, once a basis has been chosen, the coordinates of any vector re that basis are uniquely determined. Find the coordinates of the vector \( (5, 4, 3) \) re \( B^{II} \).

**Dimension.** It can be proved that if a linear space \( \mathbb{V} \) has one finite basis \( B \), consisting of \( n \) vectors, then all its bases consist of \( n \) vectors. \( \mathbb{V} \) is then said to be \( n \)-dimensional, notation: \( \dim \mathbb{V} = n \).

**Dimension depends on field.** If \( \mathbb{V} \) is the set of vectors of a linear space, then the dimension of the linear space may very well depend on the field \( \phi \). **Example:**

Let \( \mathbb{Y} \) be the set of ordered triples of complex numbers 
\[
(\alpha_1 + i\beta_1, \alpha_2 + i\beta_2, \alpha_3 + i\beta_3).
\]
Show that \{(1,0,0), (0,1,0), (0,0,1)\} is a basis for the linear space with vectors in \(V\) and scalars in the set of complex numbers. However \{(1,0,0), (1,0,0), (0,1,0), (0,1,0), (0,0,1), (0,0,1)\} is a basis for the linear space with the same set of vectors, but with real numbers for scalars.

Extension of a linearly independent set to a basis. Let \(A\) be a set of \(k\) linearly independent vectors in an \(n\)-dimensional vector space \(V\). If \(k < n\), then \(n-k\) additional vectors exist which together with the \(k\) vectors in \(A\) form a basis of \(V\). These additional vectors are by no means unique (think of examples!) It is clear that no linear combination of the \(n-k\) additional vectors can be equal to any linear combination of the \(k\) vectors in \(A\).

Dimension of subspace. Let \(U\) be a subspace of a finite-dimensional linear space \(V\). Then

\[
\begin{align*}
\dim U &\leq \dim V \text{ and} \\
\dim U &= \dim V \text{ iff } U = V.
\end{align*}
\]

Note: If \(U \subseteq V\) then \(V\) must contain at least one vector not in \(U\); thus we find that a basis of \(V\) must have at least one more vector than any basis of \(U\). (For complete proof see Halmos, p. 18).

Corollary. Let \(U \subseteq V\) be a subspace of \(V\), \(\dim V = n\), \(\dim U = k < n\). Then let \(A\) be a basis of \(U\). Applying "extension of a linearly independent set to a basis" we find there are \((n-k)\) linearly independent vectors in \(V\), not in \(U\), none of whose linear combinations belong to \(U\).

Sum of two linear spaces. Let \(U_1\) and \(U_2\) be two linear spaces. \(U_1 + U_2\) is defined to be the set of all vectors \(x_1 + x_2\) with \(x_1 \in U_1\) and \(x_2 \in U_2\); it is itself a linear space.
If \( \{a_1, a_2, \ldots, a_k\} \) is a basis of \( U_1 \) and \( \{b_1, b_2, \ldots, b_l\} \) is a basis of \( U_2 \), then \( \{a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_l\} \) spans \( U_1 + U_2 \). Of course, \( \{a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_l\} \) need not be a basis (e.g., let \( U_1 \) and \( U_2 \) be two planes through the origin in \( \mathbb{R}^3 \)). If \( \{a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_l\} \) is not a basis, then \( \gamma_1, \ldots, \gamma_k \) and \( \delta_1, \ldots, \delta_l \) exist with

\[
\sum_{i=1}^{k} \gamma_i a_i = \sum_{j=1}^{l} \delta_j b_j
\]

such that at least one \( \gamma_i \) and one \( \delta_j \) are not zero (why 'and'?). So in this case \( U_1 \cap U_2 \) contains non-null vectors. (\( \sum \gamma_i a_i \in U_1 \); \( \sum \delta_j b_j \in U_2 \)). If, however, \( \{a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_l\} \) is a basis, then the only possibility for (*) to hold true is that \( \sum \gamma_i \) and \( \sum \delta_j \) are zero, and \( U_1 \cap U_2 = \{\theta\} \). This motivates the distinction of the joint occurrence of \( V = U_1 + U_2 \) and \( U_1 \cap U_2 = \{\theta\} \) by a special term:

**Direct sum of two linear spaces.** \( V \) is said to be the direct sum of \( U_1 \) and \( U_2 \) (notation: \( V = U_1 \oplus U_2 \)) iff \( V = U_1 + U_2 \) and \( U_1 \cap U_2 = \{\theta\} \).

**Uniqueness of vector decomposition relative to direct sums.** Let \( V = U_1 + U_2 \). Then in order that this sum be direct, it is necessary and sufficient that to each \( v \in V \) there corresponds only one vector \( x_1 \in U_1 \) and only one vector \( x_2 \in U_2 \) such that \( v = x_1 + x_2 \).

**Necessity.** Let \( x_1 \in U_1 \), \( x'_1 \in U_1 \), \( x_2 \in U_2 \), \( x'_2 \in U_2 \) such that \( v = x_1 + x_2 = x'_1 + x'_2 \). Then \( x_1 - x'_1 = x'_2 - x_2 \), a vector which clearly belongs to both \( U_1 \) and \( U_2 \). So, if \( U_1 \cap U_2 = \{\theta\} \), it follows that \( x_1 - x'_1 = \theta = x'_2 - x_2 \), which proves uniqueness.

**Sufficiency.** (reductio ad absurdum). If the sum is not direct, i.e., if \( U_1 \cap U_2 \) contains vectors other than \( \theta \), say \( u \), then \( v = x_1 + x_2 = (x'_1 + u) + (x'_2 - u) \), which proves non-uniqueness, since \( x_1 + u \in U_1 \) and \( x_2 - u \in U_2 \).
Variety of direct sums. Although we just showed that each vector belonging to a
direct sum \( U_1 \oplus U_2 \) is the sum of a uniquely determined vector \( x_1 \in U_1 \) and a
uniquely determined vector \( x_2 \in U_2 \), the reader should be cautioned that the choice
of subspaces \( U_1 \) and \( U_2 \) itself is by no means unique. Not only:

\[
U_1' \oplus U_2' = U_1'' \oplus U_2'' \quad \text{does not imply} \quad U_1' = U_1'' \quad \text{and} \quad U_2' = U_2''
\]

but even:

\[
U_1 \oplus U_2 = U_1' \oplus U_2' \quad \text{does not imply} \quad U_2 = U_2'
\]

The situation is entirely similar to the one described under a "variety of bases."
In fact, saying that \( \{x_1, x_2\} \) is a basis of a linear space \( U \) is tantamount to
saying that \( U = \langle x_1 \rangle \oplus \langle x_2 \rangle \), where \( \langle x_1 \rangle \) is the one-dimensional linear
space with basis \( \{x_1\} \). The reader may clarify the situation to himself by drawing
some pictures.

**Dimension of a direct sum.**

\[
\begin{cases}
V \text{ is a linear space} \\
U_1 \subseteq V \text{ is a linear space} \\
U_2 \subseteq V \text{ is a linear space} \\
\dim U_1 = n_1 \\
\dim U_2 = n_2 \\
V = U_1 \oplus U_2
\end{cases}
\]

then \( \dim V = n_1 + n_2 = \dim U_1 + \dim U_2 \).

Proof. Since \( \dim U_i = n_i \), a set \( \{x_{i1}, x_{i2}, \ldots, x_{in_i}\} \) of \( n_i \) linearly
independent vectors exists, which is a basis of \( U_i \) \( (i = 1, 2) \). Given
\( v \in V \), the preceding result shows that exactly one \( x_1 \in U_1 \), and exactly
one \( x_2 \in U_2 \) exist such that \( v = x_1 + x_2 \). Now \( x_i \) can be represented
(uniquely) as a linear combination of basis vectors in \( U_i \) \( (i = 1, 2) \). So
\( v \) can be represented (uniquely) as a linear combination of the vectors
\( x_{11}, x_{12}, \ldots, x_{1n_1}, x_{21}, x_{22}, \ldots, x_{2n_2} \). Hence these vectors do span \( V \).
They are also linearly independent, as follows along the lines of the
argument around equation (*) on page A-16. Consequently the $n_1 + n_2$
vectors $x_{11}', \ldots, x_{1n_1}', x_{21}', \ldots, x_{2n_2}'$ form a basis of $\mathcal{V}$, i.e.,
$$\dim \mathcal{V} = n_1 + n_2.$$  

Sufficient and necessary condition for $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$.

\[
\begin{align*}
\{ \mathcal{V} \} \\
\mathcal{U}_1 \subseteq \mathcal{V} \\
\mathcal{U}_2 \subseteq \mathcal{V} \\
\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset \\
\dim \mathcal{V} = \dim \mathcal{U}_1 + \dim \mathcal{U}_2
\end{align*}
\]

then $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$. The converse is obviously true.

Proof. Let $\dim \mathcal{U}_1 = n_1$, $\dim \mathcal{U}_2 = n_2$. Then a basis $\{x_{11}', x_{12}', \ldots, x_{1n_1}'\}$ of $\mathcal{U}_1$ and a basis $\{x_{21}', x_{22}', \ldots, x_{2n_2}'\}$ of $\mathcal{U}_2$ exist. The set $\{x_{11}', x_{12}', \ldots, x_{1n_1}', x_{21}', x_{22}', \ldots, x_{2n_2}'\}$ of $n_1 + n_2$ vectors is linearly independent (see argument around equation (*) on page A-16), and clearly spans $\mathcal{U}_1 + \mathcal{U}_2 \subseteq \mathcal{V}$, hence is a basis of $\mathcal{U}_1 + \mathcal{U}_2$. So
$$\dim (\mathcal{U}_1 + \mathcal{U}_2) = n_1 + n_2 = \dim \mathcal{U}_1 + \dim \mathcal{U}_2 = \dim \mathcal{V}.$$  

According to the previous result on the dimension of subspaces it follows that $\mathcal{U}_1 + \mathcal{U}_2 = \mathcal{V}$. But we know that $\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$. Hence (definition of direct sum) $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$.

Commutativity. $\mathcal{U}_1 + \mathcal{U}_2 = \mathcal{U}_2 + \mathcal{U}_1$ (Property I.4 of section A.3)

Commutativity. $\mathcal{U}_1 \oplus \mathcal{U}_2 = \mathcal{U}_2 \oplus \mathcal{U}_1$

Proof. $\mathcal{U}_1 \oplus \mathcal{U}_2 = \mathcal{U}_1 + \mathcal{U}_2$ iff in addition $\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$.

$\mathcal{U}_2 \oplus \mathcal{U}_1 = \mathcal{U}_2 + \mathcal{U}_1$ iff in addition $\mathcal{U}_2 \cap \mathcal{U}_1 = \emptyset$.

Associativity. $\mathcal{U}_1 + (\mathcal{U}_2 + \mathcal{U}_3) = (\mathcal{U}_1 + \mathcal{U}_2) + \mathcal{U}_3$ (Property I.2 of section A.3)

Associativity. $\mathcal{U}_1 \oplus (\mathcal{U}_2 \oplus \mathcal{U}_3) = (\mathcal{U}_1 \oplus \mathcal{U}_2) \oplus \mathcal{U}_3$ in the sense that if the $\oplus$-signs are valid in one member of the equality, they are valid in the other member as well.
Proof. We shall assume the validity of $\oplus$ in the first member. Then take $y \in U_1 \oplus (U_2 \oplus U_3)$. Since vector decomposition relative to direct sums is unique, $x_1 \in U_1$ and $x_{23} \in (U_2 \oplus U_3)$ are uniquely determined by the condition that $y = x_1 + x_{23}$; again $x_2 \in U_2$ and $x_3 \in U_3$ are uniquely determined by the condition that $x_{23} = x_2 + x_3$; hence $x_1 \in U_1$, $x_2 \in U_2$, $x_3 \in U_3$ are uniquely determined by the condition that $y = x_1 + (x_2 + x_3) = x_1 + x_2 + x_3 = (x_1 + x_2) + x_3$; hence $(x_1 + x_2) \in U_1 + U_2$ and $x_3 \in U_3$ are uniquely determined by the condition that $y = (x_1 + x_2) + x_3$; so the sum $(U_1 + U_2) \oplus U_3$ is direct: we may write $(U_1 + U_2) \oplus U_3$; finally $U_1 \cap U_2 = \{0\}$ follows from $U_1 \cap (U_2 \oplus U_3) = \{0\}$, whence we are allowed to write $(U_1 \oplus U_2) \oplus U_3$ instead of $(U_1 + U_2) \oplus U_3$.

Notation associativity. As is usual in cases where associativity holds we will omit parentheses altogether and write $U_1 \oplus U_2 \oplus U_3$ for $U_1 \oplus (U_2 \oplus U_3)$ (since it was just proved the parentheses may be put up anyway we want to). The preceding proof suggests a handy definition for:

Direct sum of $k$ linear spaces. $\mathcal{V}$ is said to be the direct sum of $U_1, U_2, \ldots, U_k$ (notation $\mathcal{V} = U_1 \oplus U_2 \oplus \cdots \oplus U_k$, iff $\mathcal{V} = U_1 + U_2 + \cdots + U_k$ and in addition for any $y \in \mathcal{V}$ the condition that $y = x_1 + x_2 + \cdots + x_k$ uniquely determines the vectors $x_i \in U_i$ $(i = 1, 2, \ldots, k)$.

Corollary associativity. $U_1 \oplus (U_2 \oplus (U_3 \oplus U_4)) = U_1 \oplus U_2 \oplus U_3 \oplus U_4$.

$U_1 \oplus (U_2 \oplus (U_3 \oplus (U_4 \oplus (U_5 \oplus U_6)))) = U_1 \oplus U_2 \oplus U_3 \oplus U_4 \oplus U_5 \oplus U_6$. etc.

Proof. Similar to roughly the first half of the proof for three linear spaces summed directly.

Note. For $U_1 + U_2 + \cdots + U_k$ to be a direct sum of $U_1, \ldots, U_k$ it is not sufficient that $U_i \cap U_j = \{0\}$ for all pairs $(i, j)$ with $i \neq j$, see figure:
One (necessary and) sufficient condition in terms of intersections can be gathered from the above corollary: in case $k = 4$: $Y_3 \cap Y_4 = \{\emptyset\}$ and $Y_2 \cap (Y_3 + Y_4) = \{\emptyset\}$ and $Y_1 \cap (Y_2 + Y_3 + Y_4) = \{\emptyset\}$.
A.5. Linear mappings

The first condition for a mapping \( f : \mathcal{V} \rightarrow \mathcal{W} \) to be linear is that \( \mathcal{V} \) and \( \mathcal{W} \) be linear spaces. For our purposes we need only consider the case that \( \mathcal{W} = \mathcal{V} \).

**Definition.** Let \( \mathcal{V} \) be a linear space over a field \( \Phi \). The mapping \( f : \mathcal{V} \rightarrow \mathcal{V} \) is said to be **linear** iff

\[
 f(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 f x_1 + \alpha_2 f x_2
\]

for any \( x_1 \in \mathcal{V}, x_2 \in \mathcal{V}, \alpha_1 \in \Phi, \alpha_2 \in \Phi \).

**Note.** \( f \emptyset = \emptyset \)

**Image of a set.** Let \( \mathcal{U} \subseteq \mathcal{V} \), then the image of \( \mathcal{U} \) under \( f \) is defined by

\[
 f \mathcal{U} = \{ y \in \mathcal{V} ; \text{ there exists } x \in \mathcal{U} \text{ such that } y = f x \}
\]

**Inverse image of a point.** Let \( b \in \mathcal{V} \), then the (complete) inverse image of \( b \) under \( f \) is defined by

\[
 f^{-1} b = \{ x \in \mathcal{V} ; f x = b \}
\]

The **range** (or range space, or image set) of \( f \) is \( \mathcal{R}(f) \) \( \overset{\text{def}}{=} \ f \mathcal{V} \).

The **null-space** (nucleus, kernel) of \( f \) is \( \mathcal{N}(f) \) \( \overset{\text{def}}{=} \ f^{-1} \emptyset \).

By definition, \( \mathcal{R}(f) \) and \( \mathcal{N}(f) \) are subsets of \( \mathcal{V} \). They are even subspaces:

\( \mathcal{N}(f) \) is a subspace of \( \mathcal{V} \).

**Proof.** Let \( y_1 \in \mathcal{N}(f) \) and \( y_2 \in \mathcal{N}(f) \), then there exist \( x_1 \in \mathcal{V} \) and \( x_2 \in \mathcal{V} \) such that \( y_1 = f x_1 \), \( y_2 = f x_2 \), hence \( \alpha_1 y_1 + \alpha_2 y_2 = f(\alpha_1 x_1 + \alpha_2 x_2) \), so \( \alpha_1 y_1 + \alpha_2 y_2 \in \mathcal{N}(f) \). This takes care of the properties I.1 and II.1 of linear spaces. The reader will check the other properties.

\( \mathcal{R}(f) \) is a subspace of \( \mathcal{V} \).

**Proof.** Let \( x_1 \in \mathcal{R}(f) \), \( x_2 \in \mathcal{R}(f) \), then \( fx_1 = \emptyset \), \( fx_2 = \emptyset \), hence

\[
 f(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 f x_1 + \alpha_2 f x_2 = \emptyset, \text{ so } (\alpha_1 x_1 + \alpha_2 x_2) \in \mathcal{N}(f).
\]

Same remark as at the end of the preceding proof.
Suppose \( \dim \mathcal{R}(\mathcal{L}) = k \) (for some mappings \( k = 0 \), for others \( k > 0 \), or rather \( k \geq 1 \)). So we can choose a set of \( k \) independent vectors serving as a basis of \( \mathcal{R}(\mathcal{L}) \). Then we can extend this linearly independent set to a basis for \( \mathcal{Y} \), adding \((n-k)\) vectors such that no linear combination of them is equal to any linear combination of the original \( k \) vectors. What about the \( \mathcal{L}\)-image of the \((n-k)\) additional vectors? Evidently the \((n-k)\) images span \( \mathcal{L}\mathcal{Y} \) (why?) Is \( \dim \mathcal{L}\mathcal{Y} = \dim \mathcal{R}(\mathcal{L}) = n-k \)? The answer is yes:

**Theorem.** \( \dim \mathcal{R}(\mathcal{L}) + \dim \mathcal{R}(\mathcal{L}) = n = \dim \mathcal{Y} \).

**Proof.** Still calling \( \dim \mathcal{R}(\mathcal{L}) = k \), let \( \{z_1, \ldots, z_k\} \) be a basis of \( \mathcal{R}(\mathcal{L}) \).

Extend this basis to a basis \( \{z_1, \ldots, z_k, x_1, \ldots, x_{n-k}\} \) of \( \mathcal{Y} \). No linear combination of \( x_1, \ldots, x_{n-k} \) could then equal any linear combination of \( z_1, \ldots, z_k \) (unless all coefficients be zero). For any \( v \in \mathcal{Y} \), there exist (unique) \( \beta_i \) and \( \alpha_j \) such that

\[
\begin{align*}
v &= \sum_{i=1}^{k} \beta_i z_i + \sum_{j=1}^{n-k} \alpha_j x_j.
\end{align*}
\]

So \( \mathcal{L}v = \sum_{j=1}^{n-k} \alpha_j \mathcal{L}x_j \) which proves the above argument that \( \{\mathcal{L}x_1, \mathcal{L}x_2, \ldots, \mathcal{L}x_{n-k}\} \) spans \( \mathcal{L}\mathcal{Y} = \mathcal{R}(\mathcal{L}) \). Now let \( \gamma_j \) be such that

\[
\sum_{j=1}^{n-k} \gamma_j \mathcal{L}x_j = \mathcal{L}(\mathcal{E}_j \gamma_j x_j) = \Theta, \quad \text{i.e.,}
\]

such that \( \mathcal{E}_j \gamma_j x_j \in \mathcal{R}(\mathcal{L}) \). Then \( \mathcal{E}_j \gamma_j x_j \) would equal a linear combination of the \( z_i \), but we know that this implies \( \gamma_j = 0 \) for \( j = 1, \ldots, n-k \).

This means that the \( \mathcal{L}x_j \) are linearly independent, hence \( \dim \mathcal{R}(\mathcal{L}) = n-k = n - \dim \mathcal{R}(\mathcal{L}) \), which is what we wanted to prove.

**Note.** Section 2.1, point (iv), shows that not necessarily \( \mathcal{R}(\mathcal{L}) \cap \mathcal{R}(\mathcal{L}) = \{\Theta\} \), hence not necessarily \( \mathcal{Y} = \mathcal{R}(\mathcal{L}) \oplus \mathcal{R}(\mathcal{L}) \).
Non-singular vs. singular. A linear mapping \( L : V \to \mathcal{Y} \) is said to be non-singular iff \( \mathcal{N}(L) = L^{-1} \{ \emptyset \} \), singular iff \( \dim \mathcal{N}(L) \geq 1 \). Equivalently \( L : V \to \mathcal{Y} \) is non-singular iff \( \dim \mathcal{N}(L) = \dim L V = n = \dim \mathcal{Y} \), singular iff \( \dim \mathcal{N}(L) = \dim L V < n = \dim \mathcal{Y} \). Because \( L \) maps \( V \) into \( \mathcal{Y} \), \( L V \subseteq \mathcal{Y} \), so that, because of the result about dimensions of subspaces, \( L : V \to \mathcal{Y} \) is non-singular iff \( L V = \mathcal{Y} \), singular iff \( L V \subset \mathcal{Y} \) (in which case \( L \) maps many-to-one). Let \( \mathcal{U} \) be a subspace of \( \mathcal{Y} \). If \( L \mathcal{U} \subseteq \mathcal{U} \), \( \mathcal{U} \) is said to be invariant under \( L \). For such \( \mathcal{U} \), \( L \) is said to be singular on \( \mathcal{U} \) iff \( L \mathcal{U} \subset \mathcal{U} \), non-singular on \( \mathcal{U} \) iff \( L \mathcal{U} = \mathcal{U} \). In general, \( L \) is said to be non-singular on \( \mathcal{U} \) iff \( \{ v \in \mathcal{U} ; L v = \emptyset \} = \{ \emptyset \} \).

Nilpotent. \( L \) is said to be nilpotent on \( \mathcal{U} \) iff some natural number \( k \) exists such that \( L^k \mathcal{U} = \{ \emptyset \} \). (For definition of \( L^k \) see point (iii) of section 1.3.)
SOME LITERATURE ON LINEAR ALGEBRA *


Waeberden, B.L. van der (1959), Algebra II (4. Aufl. der Modernen Algebra); Berlin Springer; x + 275 pp. (Section III has a short derivation of the Jordan normal form on the basis of more advanced concepts).

Zamansky, Marc (1963), Introduction à l'algèbre et l'analyse modernes; 2ème édition; Paris, Dunod; xvi + 435 pp.

*) Only those titles are mentioned which have actually influenced our presentation.