Positive Linear Systems
Theory and Applications

LORENZO FARINA
SERGIO RINALDI

A Wiley-Interscience Series in Pure and Applied Mathematics
A Wiley-Interscience Publication
John Wiley & Sons, Inc.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto
Appendix A: Elements of Linear Algebra and Matrix Theory

A.1 REAL VECTORS AND MATRICES

An $n$-dimensional real vector $a$ is an ordered set of $n$ real numbers $a_1, a_2, \ldots, a_n$ that, conventionally, are written in column as follows

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

A real matrix of dimension $m \times n$ is a set of $mn$ real numbers $a_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ ordered by rows and columns

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Two matrices [vectors] of the same dimension can be summed by summing the corresponding elements. Analogously, a matrix [vector] can be multiplied by a real
number by multiplying all the elements of the matrix [vector] by such a number. If
m = n, the matrix A is said to be square. If n = 1, the matrix A is a column vector,
while if m = 1 the matrix A is a row vector. A matrix A can be transposed, by
exchanging the rows with the columns: The matrix A^T thus obtained has dimension
n x m and is called the transposed matrix of A and is given by

\[
A^T = \begin{pmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \cdots & a_{mn} \\
\end{pmatrix}
\]

The vector a^T is, therefore, the row vector

\[
a^T = (a_1, a_2, \ldots, a_n)
\]

Two matrices A and B can be multiplied provided that the number of columns of
A coincides with the number of rows of B. In other words, if A is of dimension
m x n and B is of dimension p x q, the product A times B is possible if and only
if n = p. The result is a matrix C = AB of dimension m x q whose element c_{ij}
can be obtained by multiplying the i-th row of A by the j-th column of B, that is,

\[
c_{ij} = \sum_{h=1}^{n} a_{ih} b_{hj}
\]

Obviously, it is possible to multiply a matrix A of dimension m x n by a vector b of
dimension n, the result being a vector c = Ab of dimension m whose i-th element
c_i is given by

\[
c_i = \sum_{h=1}^{n} a_{ih} b_h
\]

If we use the above rules, it is straightforward to see that

\[
(AB)^T = B^T A^T
\]

The class of square n x n matrices is particularly interesting. The identity matrix
I has unitary elements on its diagonal and is zero elsewhere, that is,

\[
I = \begin{pmatrix}
    1 & 0 & \cdots & 0 \\
    0 & 1 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & 1 \\
\end{pmatrix}
\]

If the product AB of two square matrices is the identity matrix, that is, AB = I, B
is called the inverse of A and denoted by A^{-1}. Moreover, the inverse matrix A^{-1}
of a matrix A is unique if it exists and A^{-1}A = AA^{-1}(= I). In this regard, it

is worth recalling that a matrix A is invertible if and only if its determinant
is nonzero, where the determinant of a matrix n x n is defined as follows:

\[
n = 1 : \det A = a_{11} \\
n = 2, 3, \ldots : \det A = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A_i
\]

where A_i is the (n - 1) x (n - 1) matrix obtained by deleting the i-th row and the
first column of A. Two square matrices A and B such that AB = BA are said to
commute. From the previous statement, a matrix commutes with its own inverse
(if it exists).

Matrices and vectors are particularly useful to represent problems in a compact
form. For example, a system of n linear equations with n unknowns

\[
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
\vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n
\]

becomes, in vector notation,

\[
Ax = b
\]

and its solution, provided the inverse matrix A^{-1} exists, can be synthetically expressed as

\[
x = A^{-1}b
\]  

(A.1)

This does not mean that for computing the solution of a system of n equations
with n unknowns it is convenient to compute the matrix A^{-1} and use formula
(A.1).

A.2 VECTOR SPACES

We now give the definition of vector space, but we warn the reader that in Definition
I, when no ambiguities are possible, the term vector is omitted.
**Definition 1 (space)**

A set \( X \) is called a space if any element \( x \) of \( X \) can be multiplied by a scalar \( a \) and the product \( ax \) is an element of \( X \), if any pair \((x, y)\) of elements of \( X \) can be summed up and the term \( x + y \) is an element of \( X \), and if the following properties hold:

\[
\begin{align*}
    x + y &= y + x \\
    (x + y) + z &= x + (y + z) \\
    0 + x &= x \\
    x + (-x) &= 0 \\
    1 \cdot x &= x \\
    0 \cdot x &= 0 \\
    a(x + y) &= a \cdot x + a \cdot y \\
    (a + b) \cdot x &= a \cdot x + b \cdot x \\
    a \cdot (b \cdot x) &= (a \cdot b) \cdot x
\end{align*}
\]

The third and fourth properties state that in a space there exists the additive identity (zero element or vector) and that each element \( x \) has its opposite \(-x\).

The simplest example of a space is the set of real numbers endowed with the usual arithmetic operations. Such a space, denoted by \( \mathbb{R} \), can be geometrically represented as an oriented straight line. An obvious extension is the space \( \mathbb{R}^n \) composed of the \( n \)-tuples of real numbers. The set of real matrices of dimensions \( m \times n \) is also a space since all the axioms required in Definition 1 hold, provided that one uses as zero matrix the matrix composed of zero elements. Other examples of spaces are the set of rational functions with real or complex coefficients, the set of polynomials with real or complex coefficients, the set of polynomials with degree smaller than \( n \) and real or complex coefficients, the set of real and continuous [integrable] [differentiable] functions in an interval \([a, b]\).

**Figure A.1** Examples of sum of two sets in \( \mathbb{R}^2 \): (a) sum of two squares; (b) sum of two circles (to be completed by the reader).

**Definition 2 (sum of sets)**

Given two sets \( X_1 \) and \( X_2 \) contained in a space \( X \), the set \( X = X_1 + X_2 \) composed of all the elements \( x = x_1 + x_2 \) with \( x_1 \in X_1 \) and \( x_2 \in X_2 \) is said to be the sum of the sets \( X_1 \) and \( X_2 \).

An example is illustrated in Fig. A.1(a), where the space \( X \) is \( \mathbb{R}^2 \). In contrast, in Fig. A.1(b) only the sets \( X_1 \) and \( X_2 \) are shown and the reader is invited to determine the set \( X_1 + X_2 \).

Other important operations on sets are the well-known union (\( \cup \)) and intersection (\( \cap \)), which, however, are not typical of vector spaces since no algebraic structure is required to perform them.

**Definition 3 (subspace)**

A set \( Z \subseteq X \) is a subspace of the space \( X \) if \( Z \) itself is a space.

The zero element of a space, as well as the whole space, satisfy Definition 3 and are therefore subspaces. They are called proper, while any other subspace is called proper. Every proper subspace in \( \mathbb{R}^3 \) is geometrically represented by a straight line or by a plane passing through the origin (being a space, the subspace must contain the zero element). Other examples of proper subspaces are the set of differentiable functions on the interval \([a, b]\) within the space of continuous functions in \([a, b]\) and the set of the polynomials of a degree smaller than \( n \) in the space of the polynomials.

**Theorem 1 (sum and intersection of two subspaces)**

The sum [intersection] of two subspaces is a subspace.

Obviously, the union of two subspaces is not, in general, a subspace.
A.3 DIMENSION OF A VECTOR SPACE

**Definition 7 (Linear Combination)**

A vector $x$ of a vector space $X$ is a linear combination of an $n$-tuple of vectors $x^1, x^2, \ldots, x^n$ if there exists an $n$-tuple of real numbers $a_1, a_2, \ldots, a_n$ such that

$$x = \sum_{i=1}^{n} a_i x^i$$

(A.3)

From this definition, it follows that the zero vector is a linear combination of any $n$-tuple of vectors, since it suffices to set $a_1 = a_2 = \cdots = a_n = 0$ for (A.3) to hold. Such a linear combination is called not proper while any linear combination with at least one nonzero $a_i$ is called proper. The above definition, referring to an $n$-tuple $(x^1, \ldots, x^n)$, can be extended in the following way:

**Definition 8 (Linear Combination)**

A vector $x$ of a vector space is a linear combination of a set $X$ when it is a linear combination of an $n$-tuple of vectors belonging to $X$.

One can prove that the set of all linear combinations of a set $X$ is a subspace and that such a subspace coincides with the subspace $[X]$ previously defined (see Definition 6). Therefore, Theorem 3 holds.

**Theorem 3 (Subspace Generated by a Set)**

Given a set $X$, the subspace $[X]$ generated by $X$ is the set of all linear combinations of $X$.

**Definition 9 (Linear Independence)**

The vectors $x^1, \ldots, x^n$ are said to be linearly independent if the zero vector is not a proper linear combination of them, that is, if the equation

$$\sum_{i=1}^{n} a_i x^i = 0$$

holds only for $a_1 = a_2 = \cdots = a_n = 0$.

An analogous definition (Definition 10) holds for the linear independence of vectors of a set $X$.

**Definition 10 (Linear Dependence)**

The vectors $x^1, \ldots, x^n$ are said to be linearly dependent if they are not linearly independent.
**Definition 11 (basis of a space)**

A set \( \mathcal{X} \) of vectors is a basis of a vector space \( X \) if every element of \( X \) can be uniquely written as a linear combination of the elements of \( \mathcal{X} \).

The set
\[
\mathcal{X} = \{1, 1 + \lambda, 1 + \lambda + \lambda^2, \ldots, 1 + \lambda + \ldots + \lambda^{n-1}\}
\]
is, for example, a basis of the space of polynomials with a degree \(< n\), while the set
\[
\mathcal{X} = \{1, 1 + \lambda, \ldots, 1 + \lambda + \ldots + \lambda^k, \ldots\}
\]
is a basis of the space of polynomials.

Due to the uniqueness property required by Definition 11, the following important property holds:

**Theorem 4 (linear independence of the elements of a basis)**

The vectors of a basis are linearly independent vectors.

In general, the basis of a vector space is not unique; for example, the set \( \mathcal{X} = (1, \lambda, \ldots, \lambda^{n-1}) \), which differs from the previously cited one, is another basis of the space of polynomials with a degree \(< n\). However, Theorem 5 holds.

**Theorem 5 (invariance of the number of elements of a basis)**

If a vector space has a basis composed of a finite number of elements, any other basis is composed of the same number of elements.

This last result justifies Definition 12.

**Definition 12 (finite dimensional and functional spaces)**

A vector space with a basis composed of \( n \) elements is called finite dimensional (the dimension of the space is \( n \)). All other spaces are called functional (or infinite dimensional) spaces.

Examples of finite dimensional spaces are the space \( \mathbb{R}^n \), the space of polynomials with a degree \(< n \), and the space of \( p \times q \) matrices (the latter having dimension \( n = pq \)). Examples of functional spaces are the space of polynomials and the space of continuous functions.

Obviously, the concept of dimension can be extended to subspaces. The set of the polynomials is, for example, an infinite dimensional subspace of the space of the continuous functions, while the set of real numbers is a subspace of finite dimension of the space of polynomials, which is infinite dimensional.

---

**A.4 CHANGE OF BASIS**

Given a basis \( \mathcal{X} = \{e^1, \ldots, e^n\} \) of a vector space \( X \) of dimension \( n \), each vector \( x \) of \( X \) can be uniquely written as a linear combination of the basis, that is,
\[
x = \sum_{i=1}^{n} a_i e^i
\]
The abstract vector \( x \) is, therefore, represented by the vector \( a \in \mathbb{R}^n \) given by
\[
a = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}^T
\]

If the basis is changed, one obtains a different representation of the vector \( x \). In fact, if one considers a new basis \( \mathcal{X}' = \{e'^1, \ldots, e'^n\} \), then
\[
x = \sum_{i=1}^{n} \bar{a}_i e'^i
\]
and the abstract vector \( x \) is represented by the new vector \( \bar{a} \in \mathbb{R}^n \) given by
\[
\bar{a} = \begin{pmatrix} \bar{a}_1 & \bar{a}_2 & \cdots & \bar{a}_n \end{pmatrix}^T
\]

It is, therefore, important to know how to obtain \( \bar{a} \) from \( a \), once the change of basis is specified. To this end, the following important result holds:

**Theorem 6 (change of basis)**

Let \( a \) and \( \bar{a} \) be the representations of the same vector \( x \), in the bases \( \mathcal{X} \) and \( \mathcal{X}' \), respectively. Then,
\[
\bar{a} = T a
\]
where \( T \) is the \( n \times n \) square matrix whose \( i \)th column is the representation of the \( i \)th element of the basis \( \mathcal{X} \) in the basis \( \mathcal{X}' \).

To prove this theorem, it suffices to note that a relationship of the kind \( A.4 \) must hold for each vector \( x \in X \). Thus, consider as vector \( x \) the \( i \)th vector \( e^i \) of the basis \( X \), which is represented in the basis \( X \) itself by a vector with zero entries except the \( i \)th one, which is equal to 1. Hence, by performing the matrix product one finds out that the representation of \( e^i \) in the basis \( \mathcal{X}' \) is the \( i \)th column of \( T \).

An obvious but important property of matrix \( T \) is that it is composed of linearly independent vectors (since the elements of a basis are such). Moreover, from Theorem 6 it follows that matrix \( T \) is invertible and that the \( i \)th column of matrix \( T^{-1} \) is the representation of the \( i \)th element of the basis \( \mathcal{X} \) in the basis \( \mathcal{X}' \).
A.5 LINEAR TRANSFORMATIONS AND MATRICES

**Definition 13 (linear transformation)**

Given two vector spaces $X$ and $Y$, a transformation $L : X \rightarrow Y$ is said to be linear if

$$L(ax^1 + bx^2) = aLx^1 + bLx^2$$

(A.5)

Note that in (A.5) the two $+$ symbols work in different spaces: the first one works on elements of $X$, while the second works on elements of $Y$.

Examples of linear transformations are the integration over an interval $[a, b]$, which transforms elements of the space of the integrable functions into the space of real numbers, the approximation of a polynomial by truncation, which transforms elements of the space of the polynomials into elements of the space of the polynomials with a degree $< n$, and the derivation, which transforms the space of the polynomials with a degree $< n$ into the space of the polynomials with a degree $< n - 1$.

Every transformation $y = Lx$ admits a particular representation as a basis $\mathcal{X}$ of $X$ and a basis $\mathcal{Y}$ of $Y$ are chosen. A remarkable case is the one in which $X$ and $Y$ have finite dimensions, since the transformation can be represented by a matrix. In fact, *Theorem 7* holds.

**Theorem 7 (matrices for the representation of transformations)**

Let $L$ be a linear transformation $X \rightarrow Y$ between finite dimensional spaces and let $\mathcal{X} = \{x^1, \ldots, x^n\}$ and $\mathcal{Y} = \{y^1, \ldots, y^m\}$ be bases of $X$ and $Y$ respectively. Let $a$ be the representation of $x$ in the basis $\mathcal{X}$ and $b$ be the representation of $Lx$ in the basis $\mathcal{Y}$. Then,

$$b = Aa$$

where $A$ is the $m \times n$ matrix whose $i$th column is the representation of $Lx^i$ in the basis $\mathcal{Y}$.

The proof of this theorem is analogous to that of *Theorem 6*.

**Example 1 (differentiation of polynomials)**

Consider the differentiation $D$ of polynomials $x$ with a degree $< n$. Let $\mathcal{X} = \{1, \lambda, \ldots, \lambda^{n-1}\}$ and $\mathcal{Y} = \{1, \lambda, \ldots, \lambda^{n-2}\}$ be the bases of the spaces of the polynomials with a degree $< n$ and $n - 1$, respectively. In these bases, the differentiation $y = Dx$ is represented by the $(n - 1) \times n$ matrix

$$A = \begin{pmatrix}
0 & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & 2 & 0 & \ldots & 0 \\
0 & 0 & 0 & 3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & n - 1
\end{pmatrix}$$

**Example 2**

Continuing with *Example 2*, suppose one wants to find the representation $\tilde{A}$ of the transformation $y = Dx$ when the bases in the two spaces are

$$\mathcal{X} = \{1, 1 + \lambda, \ldots, 1 + \lambda + \ldots + \lambda^{n-1}\}, \quad \mathcal{Y} = \{1, \lambda, \ldots, \lambda^{n-2}\}$$

The matrix $\tilde{A}$ can be obtained directly through *Theorem 7*, or, alternatively, by using (A.6) and the representation $A$ of *Example 1*. Following this last possibility, since $\mathcal{Y} = \tilde{Y}$ we have $P = I$, while from *Theorem 6* it follows that

$$Q = \begin{pmatrix}
1 & 1 & 1 & \ldots & 1 \\
0 & 1 & 1 & \ldots & 1 \\
0 & 0 & 1 & \ldots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{pmatrix}$$

Therefore,
If $n$ and $m$ are the dimensions of $X$ and $Y$, then

$$\rho(L) \leq \min(m, n) \quad \nu(I) \leq n$$

An important relationship between rank and nullity of a linear transformation is given by Theorem 9.

**Theorem 9 (relationship between rank and nullity)**

Let $L$ be a linear transformation $X \to Y$ between an $n$-dimensional space $X$ and a space $Y$. Then

$$\rho(L) + \nu(L) = n$$

According to the previous remarks on vectors and transformations, it is important to know how the image and null space of a transformation can be represented and how the rank and nullity can be obtained from the matrix representing the transformation. To this end, Theorem 10 can be used.

**Theorem 10 (representation of the image)**

The image of a transformation represented in a given basis by a matrix $A$ is the subspace generated by the column vectors of the matrix $A$.

**Example 3**

Let,

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

In $\mathbb{R}^3$, the subspace $\mathcal{I}(A)$ is represented by the plane passing through the origin and determined by the three vectors

$$a_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

as shown in Fig. A.5. This subspace is two dimensional, thus $\rho(A) = 2$ and $\nu(A) = 1$.

As a consequence of Theorem 10 and of the definition of rank, one obtains the following criterion (Theorem 11) for the determination of the rank of a transformation.

**Theorem 11 (property of the rank)**

The rank of a transformation $L$ is given by the maximum number of linearly independent columns of any of its representations $A$. 

\[\begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & 1 \end{pmatrix} \]

All the matrices representing the same transformation $L$ are called equivalent. A particular but interesting case is that of the transformation of a space $X$ into itself, that is, $L : X \to X$. In this case, it is often appropriate to choose the same basis for the space $X$, considered as the set of the elements to be transformed and as the set of the transformed elements. In such a case, $Q = P^{-1}$ so that (A.6) becomes

$$\tilde{A} = PAP^{-1}$$

Expression (A.7) is frequently used in linear algebra, where the two matrices $A$ and $\tilde{A}$ are called similar.

### A.6 IMAGE AND NULL SPACE

In this section, we present the definitions and the main properties of the image and null space of a linear transformation.

**Definition 14 (image and null space of a transformation)**

The image (or range) of a linear transformation $L : X \to Y$ is the set $\mathcal{I}(L)$ of all the elements $y \in Y$ such that $y = Lx$ for some $x \in X$. The null space (or kernel) of the transformation is the set $\mathcal{N}(L)$ of all the elements $x \in X$ such that $Lx = 0$.

**Theorem 8 (property of the image and null space)**

The image and the null space of a linear transformation $L : X \to Y$ are two subspaces of $Y$ and $X$, respectively.

**Definition 15 (rank and nullity)**

The dimensions of the image and null space of a linear transformation, called rank and nullity, are denoted by $\rho(L)$ and $\nu(L)$, that is,

$$\rho(L) = \dim \mathcal{I}(L) \quad \nu(L) = \dim \mathcal{N}(L)$$

If $n$ and $m$ are the dimensions of $X$ and $Y$, then

$$\rho(L) \leq \min(m, n) \quad \nu(I) \leq n$$

An important relationship between rank and nullity of a linear transformation is given by Theorem 9.

**Theorem 9 (relationship between rank and nullity)**

Let $L$ be a linear transformation $X \to Y$ between an $n$-dimensional space $X$ and a space $Y$. Then

$$\rho(L) + \nu(L) = n$$

According to the previous remarks on vectors and transformations, it is important to know how the image and null space of a transformation can be represented and how the rank and nullity can be obtained from the matrix representing the transformation. To this end, Theorem 10 can be used.

**Theorem 10 (representation of the image)**

The image of a transformation represented in a given basis by a matrix $A$ is the subspace generated by the column vectors of the matrix $A$.

**Example 3**

Let,

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

In $\mathbb{R}^3$, the subspace $\mathcal{I}(A)$ is represented by the plane passing through the origin and determined by the three vectors

$$a_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

as shown in Fig. A.5. This subspace is two dimensional, thus $\rho(A) = 2$ and $\nu(A) = 1$.

As a consequence of Theorem 10 and of the definition of rank, one obtains the following criterion (Theorem 11) for the determination of the rank of a transformation.

**Theorem 11 (property of the rank)**

The rank of a transformation $L$ is given by the maximum number of linearly independent columns of any of its representations $A$.
One can also prove that the number of linearly independent columns of a matrix $A$ is equal to the number of linearly independent rows. Therefore, the rank of a transformation $L$ can be simply determined by checking the linear independence of the columns or rows of any representation of $L$. If matrix $A$ is square and the columns (and, hence, the rows) of $A$ are linearly independent one can say that the transformation $L$ is nonsingular (or invertible) in accordance with Definition 16.

**DEFINITION 16 (nonsingular transformation)**

A linear transformation $L : X \rightarrow X$ is called nonsingular (or invertible) if

$$T(L) = X$$

For each nonsingular transformation, it is possible to define an inverse transformation $L^{-1}$ as the one for which

$$L^{-1}(Lx) = x$$

Therefore, the transformation $L^{-1}L$ is the identity transformation that leaves all the elements of $X$ unchanged. It is easy to prove also that $LL^{-1}$ is the identity transformation and that if $L$ is represented by $A$, the representation of $L^{-1}$ is the matrix $A^{-1}$.

As for the rank of the transformation $L = L_1L_2$, where $L_1 = Y \rightarrow Z$ and $L_2 = X \rightarrow Y$, the following result (Theorem 12) holds:

**THEOREM 12 (rank of the product of transformations)**

If $L = L_1L_2$, then

$$\rho(L) \leq \min[\rho(L_1), \rho(L_2)]$$

Moreover, if $L_1$ is nonsingular $\rho(L) = \rho(L_2)$ and if $L_2$ is nonsingular, $\rho(L) = \rho(L_1)$.

A consequence of Theorem 12 is that the product $L = L_1L_2$ of two nonsingular transformations is nonsingular. It can also be proved that

$$(L_1L_2)^{-1} = L_2^{-1}L_1^{-1}$$

**A.7 INVARIANT SUBSPACES, EIGENVECTORS, AND EIGENVALUES**

**DEFINITION 17 (transformed set)**

Given a set $\mathcal{X}$ of space $X$ and a linear transformation $L : X \rightarrow Y$, the set

$$L\mathcal{X} = \{y : y = Lx, x \in \mathcal{X}\}$$

is called the transformed set of $\mathcal{X}$.

Obviously, the transformed set $LZ$ of a subspace $Z$ is a subspace and, in particular, the image $T(L)$ of a transformation is the transformed set of the whole space.

**DEFINITION 18 (invariant set)**

A set $\mathcal{X} \subset X$ is said to be invariant with respect to the linear transformation $L : X \rightarrow X$ if $L\mathcal{X} \subset \mathcal{X}$.

Definition 18 also clearly applies to the subspaces $Z$, which are therefore called invariant when the transformation $L$ transforms them into subspaces $LZ$ contained in or coinciding with $Z$. Particularly important in applications are the one-dimensional invariant subspaces that are related with the notions of eigenvector and eigenvalue, as pointed out by Definition 19.

**DEFINITION 19 (eigenvector and eigenvalue)**

Let $L : X \rightarrow X$ be a linear transformation. A nonzero vector $x \in X$ is called an eigenvector of $L$ if there exists a scalar $\lambda$ such that

$$Lx = \lambda x$$

The scalar $\lambda$ associated with the eigenvector $x$ is called eigenvalue of $L$.

Obviously, the expression $Lx = \lambda x$ holds for every representation $A$ of $L$, so that

$$Aa = \lambda a$$
where $\alpha$ is the representation of the eigenvector.

An eigenvector $x$ determines a one-dimensional invariant subspace (the subspace $[x]$ generated by $x$) and each nonzero vector of such a subspace is an eigenvector. Thus, an infinite number of eigenvectors is associated with each eigenvalue $\lambda$; clearly, what is important is the one-dimensional invariant subspace associated with the eigenvalue. However, it is worth noting that different one-dimensional invariant subspaces can be associated with the same eigenvalue $\lambda$. A trivial example is represented by the identity transformation in which each vector is an eigenvector associated with the unitary eigenvalue. On the other hand, there exist transformations that do not admit eigenvectors (and, consequently, eigenvalues). For example, the clockwise rotation in $\mathbb{R}^2$ of an angle $\alpha$ is a linear transformation and if $\alpha \neq \pi \pi$ no one-dimensional invariant subspace exists. Finally, an example in which there is only one-dimensional invariant subspace and, hence, only one eigenvalue is given by the differentiation in the space of the polynomials. Indeed, the one-dimensional subspace of constant polynomials is transformed into the origin and, is therefore invariant, but since there are no other one-dimensional invariant subspaces there exists only one eigenvalue ($\lambda = 0$).

From the definition of an eigenvector, it follows that a scalar $\lambda$ is an eigenvalue if and only if

$$\exists x \neq 0 : (\lambda I - L)x = 0$$

that is, if and only if the null space of the transformation $(\lambda I - L)$ has dimension $n \geq 1$. Therefore, if $A$ is a representation of $L$ and $I$ is the identity matrix, $\lambda$ is an eigenvalue of $L$ if and only if it satisfies

$$\det(\lambda I - A) = 0$$

**Definition 20 (characteristic polynomial)**

Let $L : X \to X$ be a linear transformation represented by a matrix $A$ in a basis $X$ of $X$. The polynomial $\Delta_L(\cdot)$ given by

$$\Delta_L(\lambda) = \det(\lambda I - A)$$

is called a characteristic polynomial of the transformation $L$.

**Definition 20** makes sense only if the polynomial $\Delta_L(\cdot)$ is independent of the basis. But this is actually the case since a change of basis from $X$ to $\tilde{X}$ implies that the representation changes from $A$ to $\tilde{A}$ with

$$\tilde{A} = TAT^{-1}$$

and $T$ nonsingular, so that

$$\det(\lambda I - \tilde{A}) = \det(\lambda I - TAT^{-1}) =$$

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

**Definition 21 (characteristic equation and characteristic roots)**

Given a linear transformation $L : X \to X$, the equation

$$\Delta_L(\lambda) = 0$$

is called a characteristic equation. The roots of this equation in the scalar field of the vector space are called characteristic roots.

In view of the previous statements, **Theorem 13** holds.

**Theorem 13 (eigenvalues and characteristic roots)**

The eigenvalues of a linear transformation $L : X \to X$ coincide with its characteristic roots.

**Theorem 13** justifies the "algebraic" definition of an eigenvalue that is often given instead of the "geometric" definition adopted here. It is worth noting, however, that the "algebraic" definition (an eigenvalue is a solution of the characteristic equation) can easily induce the mistake of considering as eigenvalues scalar values that do not belong to the field over which the vector space $X$ is defined. To better clarify this point, consider **Example 4**.

**Example 4 (differentiation of polynomials)**

Consider the transformation $L : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates in counterclockwise direction each vector of the plane by an angle $\pi/2$. Obviously, this transformation does not admit eigenvectors and, hence, eigenvalues. However, the transformation $L$ is represented by the matrix

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

so that

$$\Delta_L(\lambda) = \det \begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix} = \lambda^2 + 1$$

and the characteristic equation is $\lambda^2 + 1 = 0$, whose solutions are $\lambda = \pm i$. However, such (imaginary) solutions should not be regarded as characteristic roots since they do not belong to the field of real numbers.

In the case where scalars are complex numbers, the solutions over the complex field of the characteristic equation are the eigenvalues of the transformation $L$. In the sequel, we shall always refer to such a case; in particular, if the matrix representing a transformation is real, we shall consider it as belonging to the set of matrices with complex entries. In this way, one can state that the transformation $L : X \to X$, represented by a real matrix, has a number of eigenvalues equal to the dimension of the space $X$, since a polynomial equation of degree $n$ with real coefficients has $n$ roots over the complex field. Clearly, the $n$ eigenvalues
are not always distinct. Thus, one must distinguish between simple and multiple eigenvalues and introduce the notion of algebraic multiplicity of an eigenvalue. In general, there are \(k(\leq n)\) distinct eigenvalues \(\lambda_1, \lambda_2, \ldots, \lambda_k\), and each one of them has an algebraic multiplicity \(a_i, i = 1, \ldots, k\) (the eigenvalues with unitary algebraic multiplicity are called simple). Thus, the characteristic polynomial can be factorized as follows:

\[
\Delta_L(\lambda) = (\lambda - \lambda_1)^{a_1}(\lambda - \lambda_2)^{a_2}\cdots(\lambda - \lambda_k)^{a_k}
\]

Finally, it is worth noting that complex eigenvalues are always present in an even number because if \(\lambda = a + ib\) is a complex eigenvalue, its conjugate \(\lambda = a - ib\) is also an eigenvalue.

**Definition 22 (generalized eigenvector)**

Given a linear transformation \(L : X \to X\) represented by a matrix \(A\), a vector \(x \neq 0\) is a generalized eigenvector of order \(k\) associated with the eigenvalue \(\lambda\) if

\[
(A - \lambda I)^{k-1} x \neq 0, \quad (A - \lambda I)^k x = 0
\]

Note that the generalized eigenvectors of order 1 are the eigenvectors considered in **Definition 19**, since \((A - \lambda I)^0 = I\).

Once a generalized eigenvector of order \(k\) has been found, it is possible to construct a set of \(k\) linearly independent generalized eigenvectors, as shown by **Theorem 14**.

**Theorem 14 (chain of generalized eigenvectors)**

Let \(x\) be a generalized eigenvector of order \(k\) associated with an eigenvalue \(\lambda\). Then, the \(k\) vectors

\[
x^k = x \\
x^{k-1} = (A - \lambda I)x^k \\
x^{k-2} = (A - \lambda I)x^{k-1} \\
\vdots \\
x^1 = (A - \lambda I)x^2
\]

(A.8)

are linearly independent generalized eigenvectors (\(x^i\) is of order \(i\)).

The fact that the vectors \(x^i\) given by (A.8) are generalized eigenvectors of order \(i\), can be easily checked, since

\[
(A - \lambda I)^i x^i = (A - \lambda I)^{i+1} x^{i+1} = \cdots = (A - \lambda I)^k x^k = 0
\]

while

\[
(A - \lambda I)^{i-1} x^i = (A - \lambda I)^{k-1} x^k \neq 0
\]

since \(x^k = x\) is an eigenvector of order \(k\). Moreover,

\[
A x^1 = \lambda x^1
\]

since \(x^1\) is an eigenvector of order 1, while for \(i \geq 2\) the following relationship holds (obtained from (A.8)):

\[
A x^i = x^{i-1} + \lambda x^i
\]

The linear independence of the generalized eigenvectors can, then, be proved by contradiction.

But more can be said on generalized eigenvectors as shown by **Theorem 15**.

**Theorem 15 (independence of chains of generalized eigenvectors)**

The generalized eigenvectors corresponding to distinct eigenvalues are linearly independent. On the other hand, if \(x\) and \(y\) are two generalized eigenvectors of order \(h\) and \(k\) associated with the same eigenvalue, the chains \(x^1, x^2, \ldots, x^h\), and \(y^1, y^2, \ldots, y^k\) obtained from \(x\) and \(y\) using (A.8) are linearly independent if \(x^1\) and \(y^1\) are such.

On the basis of the previous results, one can prove that, given an \(n \times n\) matrix \(A\), the following procedure always yields \(n\) linearly independent generalized eigenvectors (the importance of such a procedure will be clear in the next paragraph).

**Procedure for the computation of \(n\) linearly independent (generalized) eigenvectors of an \(n \times n\) matrix**

1. Determine the distinct eigenvalues \(\lambda_1, \ldots, \lambda_k\) and their algebraic multiplicities \(a_1, \ldots, a_k\).

2. For each eigenvalue \(\lambda_i\) determine \(a_i\) linearly independent (generalized) eigenvectors as follows:

   2a. Evaluate the matrices \((A - \lambda_i I)^h, h = 1, 2, \ldots\), and so on, until the rank of \((A - \lambda_i I)^m\) is equal to the rank of \((A - \lambda_i I)^{m+1}\).

   2b. Determine a generalized eigenvector \(x\) of order \(m\) by finding a nonzero solution of

   \[
   (A - \lambda_i I)^m x = 0
   \]

   (A.9)

   2c. Evaluate the chain of eigenvectors \(x^1, x^2, \ldots, x^m\) by means of Eq. (A.8) starting with \(k = m\).

   2d. If \(m = a_i\), the vectors \(x^1, x^2, \ldots, x^m\), are the required ones. If, on the contrary, \(m < a_i\), then determine a new generalized eigenvector \(y\) associated with \(\lambda_i\) with order \(m\), such that the corresponding \(y^1\), determined using Eq. (A.8), is linearly independent of \(x^1\) [for this find
a solution \( y \) of (A.9) such that \( y \) is linearly independent of \( x \). If this is not possible, find an eigenvector of rank \((m - 1)\) or \((m - 2)\), and so forth, always by choosing \( y \) linearly independent of \( x \) and continue like so, until \( \alpha_i \) eigenvectors have been found.

**Example 5**

Consider the matrix

\[
A = \begin{pmatrix}
1 & 2 & 0 & 1 \\
0 & 2 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

and suppose we want to determine four linearly independent eigenvectors using the procedure described above.

The characteristic polynomial \( \Delta_A(\lambda) \) is given by

\[
\Delta_A(\lambda) = \det(\lambda I - A) = (\lambda - 1)^3(\lambda - 2)
\]

Hence, there are two distinct eigenvalues \((\lambda_1 = 1, \lambda_2 = 2)\) with algebraic multiplicity \(\alpha_1 = 3\) and \(\alpha_2 = 1\).

As for the eigenvalue \(\lambda_1\) we have

\[
(A - \lambda_1 I) = \begin{pmatrix}
0 & 2 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\]

\[
(A - \lambda_1 I)^2 = \begin{pmatrix}
0 & 2 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\]

\[
(A - \lambda_1 I)^3 = \begin{pmatrix}
0 & 2 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\]

so that, \( \rho(A - \lambda_1 I) = 2 \) and \( \rho((A - \lambda_1 I)^2) = \rho((A - \lambda_1 I)^3) = 1 \). Thus, we must determine an eigenvector \( x \), of order 2, and, for this, \( x \) must satisfy the following relationships:

\[
(A - \lambda_1 I)^2 x = 0 \quad (A - \lambda I)x \neq 0
\]

A vector satisfying these relationships is

\[
x = (1 \ 0 \ 0 \ 1)^T
\]

Hence, Eq. (A.8) yields two eigenvectors

\[
x^2 = x = (1 \ 0 \ 0 \ 1)^T
\]

\[
x^1 = (A - \lambda_1 I)x^2 = (1 \ 0 \ 0 \ 0)^T
\]

Since \( \alpha_1 = 3 \), we must determine another linearly independent eigenvector associated with \( \lambda_1 \). Such an eigenvector cannot be of order 2, since, if this would be the case, there would be two other linearly independent eigenvectors associated with \( \lambda_1 \), which, together with the eigenvector associated with \( \lambda_2 \) would give five linearly independent eigenvectors, a contradiction since we are dealing with a \( 4 \times 4 \) matrix. Therefore, the new eigenvector \( x^3 \) must be of rank 1 and linearly independent of \( x^1 \). Such an eigenvector is, for example,

\[
x^3 = (0 \ 0 \ 1 \ 0)^T
\]

Finally, as for the simple eigenvalue \( \lambda_2 \), we have

\[
(A - \lambda_2 I) = \begin{pmatrix}
-1 & 2 & 0 & 1 \\
0 & 0 & 0 & 0 \\
0 & -1 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\]

and an eigenvector \( x^4 \) is given by (check that \( Ax^4 = \lambda_2 x^4 \))

\[
x^4 = (2 \ 1 \ -1 \ 0)^T
\]

In conclusion, the four eigenvectors are

\[
x^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad x^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad x^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad x^4 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ -1 \end{pmatrix}
\]

and all of them are of order 1 except \( x^2 \), which is of order 2.

---

**A.8 Jordan Canonical Form**

In this section, we show how the “simplest” matrix similar to a given matrix \( A \) can be determined. Before describing the most general case, we consider the simple but frequent case of a matrix \( A \) with \( n \) distinct eigenvalues \( \lambda_1, \ldots, \lambda_n \).
THEOREM 16 (diagonalization of a matrix with distinct eigenvalues)

Let \( A \) be an \( n \times n \) matrix with \( n \) distinct eigenvalues \( \lambda_i, i = 1, \ldots, n \), and \( n \) linearly independent eigenvectors \( x^i, i = 1, \ldots, n \). The diagonal matrix

\[
A_D = \begin{pmatrix}
\lambda_1 & 0 & \cdots & \cdots & 0 \\
0 & \lambda_2 & \cdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & \lambda_n
\end{pmatrix}
\]

is similar to the matrix \( A \) and \( A_D = TAT^{-1} \) where the \( i \)-th column of the matrix \( T^{-1} \) is the \( i \)-th eigenvector \( x^i \).

The proof of this theorem is very simple. In fact,

\[
Ax^i = \lambda_i x^i \quad (A.10)
\]

since the eigenvectors \( x^i \) are of order 1. Consequently, the transformation \( A \) is represented in the basis \( \mathcal{X} = \{x^1, x^2, \ldots, x^n\} \) by a matrix \( A_D \) whose \( i \)-th column is the representation of the vector \( Ax^i \) in the \( \mathcal{X} \) basis. From Eq. (A.10), it follows that the vector \( Ax^i \) is represented by a vector with zero entries except the \( i \)-th one, which is equal to \( \lambda_i \). From this, Theorem 16 follows.

A slightly more general case is that of matrix \( A \) with multiple eigenvalues but with \( n \) linearly independent eigenvectors of order 1. If we use the set of these eigenvectors as a basis \( \mathcal{X} \), the transformation is still represented by a matrix \( A_D \) in diagonal form. More precisely, if \( a_i \) is the multiplicity of the \( i \)-th eigenvalue (\( i = 1, \ldots, k, k < n \)), matrix \( A_D \) will be a block diagonal matrix

\[
A_D = \begin{pmatrix}
D_1 & 0 & \cdots & \cdots & 0 \\
0 & D_2 & \cdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & D_k
\end{pmatrix}
\]

where each block \( D_i \) is a square diagonal matrix of dimension \( a_i \times a_i \), namely,

\[
D_i = \begin{pmatrix}
\lambda_i & 0 & \cdots & \cdots & 0 \\
0 & \lambda_i & \cdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & \lambda_i
\end{pmatrix}
\]

We now consider the most general case in which some of the \( n \) eigenvectors \( x^1, x^2, \ldots, x^n \) (obtained by means of the previously outlined procedure) are of order \( > 1 \).

THEOREM 17 (Jordan form)

Let \( A \) be an \( n \times n \) matrix with \( k \) distinct eigenvalues \( \lambda_1, \ldots, \lambda_k \) with multiplicities \( a_1, \ldots, a_k \) and assume that \( a_i \) linearly independent eigenvectors associated with each \( \lambda_i \) are known. Moreover, suppose that these eigenvectors are divided into \( g_1 \) groups (or chains) of the form \((A.8)\) with dimensions \( n_i^1 \geq n_i^2 \geq \cdots \geq n_i^{g_i} \) (in the following, \( n_i^1 \) will be denoted by \( n_i \)). The following matrix:

\[
J_i^h = \begin{pmatrix}
\lambda_i & 1 & 0 & \cdots & 0 \\
0 & \lambda_i & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \cdots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & \cdots & \lambda_i
\end{pmatrix}
\]

of dimension \( n_i^h \times n_i^h \) is associated with the \( h \)-th group of eigenvectors associated with \( \lambda_i \). Thus a Jordan block

\[
J_i = \begin{pmatrix}
J_i^1 & 0 & \cdots & \cdots & 0 \\
0 & J_i^2 & & & \cdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & J_i^{g_i}
\end{pmatrix}
\]

of dimension \( a_i \times a_i \) is associated with each eigenvalue. The block-diagonal matrix

\[
A_J = \begin{pmatrix}
J_1 & 0 & \cdots & \cdots & 0 \\
0 & J_2 & & & \cdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & J_k
\end{pmatrix}
\]

is similar to the matrix \( A \), that is, \( A_J = TAT^{-1} \) and the columns of the matrix \( T^{-1} \) are the \( n \) eigenvectors \( x^1, x^2, \ldots, x^n \). Matrix \( A_J \) is known as the Jordan canonical form.

The proof of this theorem is very similar to that of the previous one.

EXAMPLE 6

Suppose we want to determine the Jordan canonical form of the matrix

\[
A = \begin{pmatrix}
1 & 2 & 0 & 1 \\
0 & 2 & 0 & 0 \\
0 & -1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

which has two distinct eigenvalues \( \lambda_1 = 1 \) (\( a_1 = 3 \)) and \( \lambda_2 = 2 \) (\( a_2 = 1 \)). We have already seen in the examples that two chains of eigenvectors are associated
with $\lambda_1$ so that $n_1 = n_1^1 = 2$ and $n_1^2 = 1$. Matrix $A_J$ is, therefore, given by

$$A_J = \begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 2
\end{pmatrix}$$

while matrix $T^{-1}$ is (see the eigenvectors $x^i, i = 1, \ldots, 4$, in Example 5)

$$T^{-1} = \begin{pmatrix}
1 & 1 & 0 & 2 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & -1 \\
0 & 1 & 0 & 0
\end{pmatrix}$$

so that

$$T = \begin{pmatrix}
1 & -2 & 0 & -1 \\
0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0
\end{pmatrix}$$

The reader can easily check that $TAT^{-1} = A_J$.

The dimension $n_1$ of the first block of each Jordan block is equal to the dimension of the longest chain of eigenvectors associated with the $i$th eigenvalue $\lambda_i$ and is called the index of the $i$th eigenvalue. The number $g_i$ of blocks composing each Jordan block $J_i$ is called a geometric multiplicity of the eigenvalue $\lambda_i$. The reason for this terminology is that an invariant subspace $X^i$ of dimension $g_i$ is associated with $\lambda_i$, and each vector $x$ belonging to $X^i$ is an eigenvector associated with $\lambda_i$, that is,

$$A_J x = \lambda_i x \quad \forall x \in X^i$$

Such a subspace $X^i$ is the subspace generated by all the eigenvectors of order 1 associated with $\lambda_i$, which are as many as the chains of eigenvectors associated with $\lambda_i$ and, consequently, as many as the blocks composing $J_i$.

### A.9 ANNIHILATING POLYNOMIAL AND MINIMAL POLYNOMIAL

We have already seen that it is interesting to associate with a matrix $A$ a polynomial, namely, the characteristic polynomial

$$\Delta_A(\lambda) = \det(\lambda I - A)$$

Moreover, we have verified that such a polynomial is invariant with respect to the similarity transformation $A = TAT^{-1}$, that is, $\Delta_A(\cdot) = \Delta_{\tilde{A}}(\cdot)$. This is an important property, since it allows us to associate the characteristic polynomial to the transformation $L : X \to X$. Other polynomials enjoying the same property are the annihilating polynomials. Before defining them, let us remark that if $p(\cdot)$ is a polynomial

$$p(\lambda) = k_0 + k_1 \lambda + k_2 \lambda^2 + \cdots + k_m \lambda^m$$

then, $p(A)$ is the matrix

$$p(A) = k_0 I + k_1 A + k_2 A^2 + \cdots + k_m A^m$$

Moreover, given a matrix $A$ and one of its similar matrices $\tilde{A} = TAT^{-1}$, we have

$$(\tilde{A})^i = (TAT^{-1})(TAT^{-1}) \cdots (TAT^{-1}) = TA^i T^{-1}$$

so that

$$p(\tilde{A}) = T p(A) T^{-1}$$

**DEFINITION 23** (annihilating polynomial and minimal polynomial)

A polynomial $\vartheta_A(\cdot)$ is said to be an annihilating polynomial of the matrix $A$ if $\vartheta_A(A) = 0$. Among all monic annihilating polynomials $\vartheta_A(\cdot)$

$$\vartheta_A(\lambda) = \lambda^m + a_1 \lambda^{m-1} + \cdots + a_m$$

that with the lowest degree $m$ is the minimal polynomial and is denoted by $\Psi_A(\cdot)$.

Since the minimal polynomial $\Psi_A(\cdot)$ is particularly important in applications, it is interesting to know how it can be found. **Theorem 18** links the minimal polynomial of a matrix with its eigenvalues. Since the index of an eigenvalue is the dimension of its largest Jordan block, the minimal polynomial is easily determined, once the Jordan form $A_J$ of matrix $A$ is known (the opposite statement is false, namely the Jordan form contains more "information" than the minimal polynomial).

**Theorem 18** (eigenvalues and minimal polynomial)

Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of a matrix $A$ of dimension $n \times n$ and let $n_1, \ldots, n_k$ be the indices of such eigenvalues. Then, the minimal polynomial $\Psi_A(\cdot)$ is given by

$$\Psi_A(\cdot) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k} = \prod_{i=1}^{k} (\lambda - \lambda_i)^{n_i}$$

The proof of this theorem is not very complicated and is based on the Jordan form discussed in Section A.8.
An important consequence of Theorem 18 is that the minimal polynomial of an \( n \times n \) matrix has a degree that is \( \leq n \), since \( \sum_{i=1}^{k} n_i \leq n \). Moreover, \( \psi_A(\cdot) \) divides \( \Delta_A(\cdot) \) because

\[
\Delta_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{n_i}
\]

and \( n_i \geq n_i \), so that \( \psi_A(A) = 0 \) implies \( \Delta_A(A) = 0 \) (Cayley–Hamilton's theorem, known since 1858).

Theorem 18 often allows one to determine the Jordan form of a matrix \( A \) without computing the eigenvectors. For example, matrix \( A \) considered in Example 5 has two distinct eigenvalues \( \lambda_1 = 1 \) and \( \lambda_2 = 2 \), the first with algebraic multiplicity \( n_1 = 3 \). Consequently, the minimal polynomial \( \psi_A(\cdot) \) is one of the three following polynomials:

\[
\begin{align*}
p_1(\lambda) &= (\lambda - 1)(\lambda - 2) \\
p_2(\lambda) &= (\lambda - 1)^2(\lambda - 2) \\
p_3(\lambda) &= (\lambda - 1)^3(\lambda - 2)
\end{align*}
\]

Since \( p_1(A) \neq 0 \) and \( p_2(A) = 0 \) we have \( \psi_A(A) = p_3(\lambda) \) so that the Jordan form of \( A \) is

\[
A_J = \begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 2
\end{pmatrix}
\]

as already found in Example 6.

### A.10 NORMED SPACES

Vector spaces have often more "structure" than that considered so far (sum of elements and scalar multiplication). The most important concept, which allows the study of vector spaces from a "topological" point of view, is that of norm, which generalizes the usual notion of distance in three-dimensional Euclidean spaces.

**Definition 24 (normed space)**

A normed vector space is a vector space \( X \) for which there exists a transformation, denoted by \( \| \cdot \| \) and called norm, that transforms each element \( x \in X \) in a real number \( \| x \| \) and has the following three properties:

\[
\begin{align*}
\| x \| &> 0 \quad \forall x \neq 0 \quad \text{and} \quad \| 0 \| = 0 \\
\| x + y \| &\leq \| x \| + \| y \| \\
\| ax \| &= |a| \cdot \| x \|
\end{align*}
\]

where \( |a| \) is the modulus of the scalar \( a \).

The three properties of the norm have a straightforward geometric interpretation; in particular, the second one is known as the triangular inequality. Obviously, the concept of norm applies both to finite dimensional spaces and to functional spaces. An example of a normed functional space is the space \( C[a, b] \) of continuous functions \( x(\cdot) \) over the interval \([a, b] \) with the norm given by \( \max_{a \leq t \leq b} |x(t)| \).

It is worth noting that a space \( X \) can have many norms. For example, the space of continuous functions over the interval \([a, b] \), which generates the normed space \( C[a, b] \), may generate also other normed spaces such as those with the norm

\[
\| x(\cdot) \| = \int_a^b |x(t)| dt
\]

The best known normed functional space is the space \( L_p \), which is composed of the Lebesgue \( p \)-integrable functions \( (p \geq 1) \), namely, the functions \( x(\cdot) \) with a finite Lebesgue integral of \( |x(t)|^p \). In this space, the norm of \( x(\cdot) \) is

\[
\| x(\cdot) \|_p = \left( \int_b^a |x(t)|^p dt \right)^{1/p}
\]

Analogously, the spaces \( \mathbb{R}^n \) and \( \mathbb{C}^n \) of \( n \)-tuples of real and complex numbers \((x_1, \ldots, x_n)\) can be normed using the norm \((p \geq 1)\)

\[
\| x \|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}
\]

or, alternatively, the norm

\[
\| x \|_\infty = \max_{i} |x_i|
\]

Among these, the most significant ones are \( \| \cdot \|_1 \), \( \| \cdot \|_2 \) and \( \| \cdot \|_\infty \), which are represented in Fig. A.6 with their unitary contour lines (for the case \( n = 2 \)).

The norm \( \| \cdot \|_2 \) is the so-called Euclidean norm and, for this reason, the space \( \mathbb{R}^n \), normed with \( \| \cdot \|_2 \), is denoted by \( E^n \) and called Euclidean space. In the following (except when explicitly stated), we will assume that the norm adopted for \( \mathbb{R}^n \) or \( \mathbb{C}^n \) is \( \| \cdot \|_2 \) and, for the sake of brevity, we will simply write \( \| \cdot \| \).

In normed spaces, it is possible to define closed and open sets, bounded sets (composed of elements with bounded norm), complete sets, and compact sets (which in finite dimensional vector spaces are closed and bounded). All such notions are straightforward generalizations of the corresponding, and well-known, concepts in the plane.

The relevance of the norm becomes clear when studying transformations (linear or not) between vector spaces, because in the case where the spaces are normed, it is possible then to define continuous transformations.
DEFINITION 25 (continuous transformation)

A transformation $T(\cdot)$ (possibly nonlinear), which transforms the elements $x$ of a normed space $X$ into elements $T(x)$ of a normed space $Y$, is continuous at $x_0$ if for any $\varepsilon > 0$ there exists a $\delta(\varepsilon, x_0) > 0$ such that $\|T(x) - T(x_0)\| < \varepsilon$ for all $x$ such that $\|x - x_0\| < \delta$. If $T(\cdot)$ is continuous at every point $x_0$, the transformation is said to be continuous and if $\delta(\varepsilon, x_0)$ is independent of $x_0$, the transformation is said to be uniformly continuous.

It is pretty clear from Definition 25 that the continuity property is not only a characteristic of the transformation $T(\cdot)$, but also of the norms defined in the two spaces $X$ and $Y$. From this point of view, the problem of an appropriate choice of the norms in the spaces $X$ and $Y$ appears to be extremely critical. It is important, however, to observe that in finite dimensional vector spaces, the continuity of a transformation is independent of the norms of the spaces $X$ and $Y$ (in other words, in finite dimensional spaces all norms are equivalent).

The linear transformations $L : X \to Y$ form a vector space: It is then natural to ask whether such a space may be normed or not. The answer is given by Definition 26.

DEFINITION 26 (bounded transformations and their norm)

A linear transformation $L : X \to Y$ between normed spaces is said to be bounded if there exists a constant $M$ such that $\|Lx\| \leq M\|x\|$ for all $x \in X$. The smallest of such constants $M$ is the norm $\|L\|$ of the transformation $L$.

It is easy to see that $\|L\|$ has the following properties:
A.11 SCALAR PRODUCT AND ORTHOGONALITY

In some vector spaces another operation (besides the sum) between pairs of vectors, called a scalar product, can be defined. Such spaces admit a norm, while normed spaces may not admit a scalar product. The scalar product allows the definition of orthogonality of two vectors, as a natural extension of the well-known geometrical concept of orthogonal straight lines and planes. By using this concept, it is possible to decompose each space into the direct sum of a subspace and its orthogonal subspace.

Though the scalar product can be defined in a pretty general way, we will assume in the sequel that the scalar field is the field of the reals.

**Definition 27 (scalar product)**

A vector space is said to be a pre-Hilbert vector space if each pair of vectors \((x, y)\) is associated with a real number, called a scalar product and denoted by \((x|y)\), with the following properties:

\[
(x|y) = (y|x) \\
(x + y|z) = (x|z) + (y|z) \\
(ax|y) = a(x|y) \quad a \in \mathbb{R} \\
(x|x) > 0 \quad \text{per} \quad x \neq 0 \quad \text{and} \quad (0|0) = 0
\]

The simplest example of a pre-Hilbert space is \(\mathbb{R}^2\), which can be viewed as a plane in which the scalar product of two vectors \(x\) and \(y\) is given by the real number \(||x|| \cdot ||y|| \cdot \cos \theta\), where \(\theta\) is the angle between the two vectors \(x\) and \(y\) and \(||x||\) and \(||y||\) are the "lengths" of the two vectors (see Fig. A.7).

A trivial generalization is the space \(\mathbb{R}^n\) of the \(n\)-tuples of real numbers with

\[
(x|y) = \sum_{i=1}^{n} x_i y_i = x^T y
\]

where \(x^T\) denotes the row vector obtained by transposition of the column vector \(x\). As already announced, the pre-Hilbert spaces can be normed; in fact, Theorem 19 holds.

**Theorem 19 (norm of pre-Hilbert spaces)**

In a pre-Hilbert space, the quantity \((x|x)^{1/2}\) is a norm.

In the space \(\mathbb{R}^n\) with \((x|y) = x^T y\), one has

\[
(x|x)^{1/2} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} = ||x||_2
\]

![Figure A.7](image)

**Figure A.7** The scalar product of two vectors \(x\) and \(y\) is \(||x|| \cdot ||y|| \cdot \cos \theta\).

that is, the norm induced by the scalar product \((x|y) = x^T y\) is \(||x||_2\). From now on, the symbol \(|| \cdot ||\) will always refer to the norm induced by the scalar product, that is \(||x|| = (x|x)^{1/2}\).

It is also possible to extend to pre-Hilbert spaces the well-known "parallelogram law", which states that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of the sides. Indeed, it is easy to verify that

\[
||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2.
\]

**Definition 28** of orthogonality is of paramount importance.

**Definition 28 (orthogonality)**

In a pre-Hilbert space, two vectors \(x\) and \(y\) are said to be orthogonal (and denoted as \(x \perp y\)) if \((x|y) = 0\). Moreover, a vector \(y\) is said to be orthogonal to a set \(X\) if \((x|y) = 0 \ \forall \ x \in X\).

For orthogonal vectors \(x\) and \(y\), the celebrated Pythagoras' theorem holds, that is, \(||x + y||^2 = ||x||^2 + ||y||^2\).

**Definition 29 (orthogonal complement)**

Given a set \(X\) of a pre-Hilbert space, the orthogonal complement of \(X\), denoted by \(X^\bot\), is the set of all vectors orthogonal to \(X\).

Since the linear combination of orthogonal vectors to a set \(X\) is also orthogonal to the set \(X\), the orthogonal complement \(X^\bot\) of a set \(X\) is a subspace.

The results presented so far are valid both in finite dimensional spaces and in functional spaces. In contrast, the following results (Theorem 20) are valid only in finite-dimensional spaces:
Theorem 20 (orthogonal projection theorem)

Let $Z$ be a subspace of a finite dimensional space $X$. Given a vector $x \in X$, there exists a unique vector $z_0 \in Z$ such that

$$
\|x - z_0\| \leq \|x - z\| \quad \forall z \in Z
$$

Theorem 20 has straightforward geometric interpretation (see Fig. A.8) and is extremely important in applications. Moreover, it allows one to establish the decomposition Theorem 21.

Theorem 21 (decomposition theorem)

If $Z$ is a subspace of a finite dimensional space $X$, then

$$
X = Z \oplus Z^\perp
$$

Theorem 21 is illustrated in Fig. A.9 for the case $X = \mathbb{R}^3$.

The projection of a vector on a subspace $Z$ is, clearly, a linear transformation $P_z$ and, as such, it admits a representation with respect to any given basis of $X$. Suppose $\{e_1, \ldots, e^m\}$ is a basis of $Z$ and that $\{e^{m+1}, \ldots, e^n\}$ is a basis of $Z^\perp$. From Theorem 21, the vectors $\{e_1, \ldots, e^n\}$ represent a basis of $X$ and the projection on $Z$ leaves the vectors belonging to $Z$ unchanged and transforms into the zero vector all the vectors belonging to $Z^\perp$. Thus, the transformation $P_z$ is represented, in the basis $\{e_1, \ldots, e^n\}$, by the $n \times n$ matrix

$$
A = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}
$$

where $I_m$ is the identity matrix of dimension $m \times m$.

Figure A.8 The point $z_0 \in Z$ closest to $x$ is the orthogonal projection of $x$ on $Z$.

Figure A.9 Decomposition of a vector space into the direct sum of orthogonal subspaces.

Definition 30 (orthonormal vectors)

A set of vectors $\{x^1, \ldots, x^k\}$ is called orthonormal (as well as the vectors) if

$$
x^i \perp x^j \quad \forall i \neq j \quad \text{and} \quad \|x^i\| = 1 \quad \forall i.
$$

In an $n$-dimensional space, a set of $n$ orthonormal vectors is a basis (also called orthonormal) and such a basis is particularly useful in applications. Moreover, there exists a simple procedure (known as the Gram-Schmidt procedure) that allows one to obtain from a set $\{x^1, \ldots, x^k\}$ of linearly independent vectors an orthonormal set $\{e^1, \ldots, e^k\}$ with the property

$$
[x^1, \ldots, x^i] = [e^1, \ldots, e^i] \quad i = 1, \ldots, k \quad (A.11)
$$

In fact, let

$$
e^1 = \frac{1}{\|x^1\|} x^1
$$

Such a vector has unitary norm and $[x^1] = [e^1]$. Moreover, if (see Fig. A.10)

$$
x^2 = x^2 - (x^2|e^1)e^1
$$

the vector

$$
e^2 = \frac{1}{\|x^2\|} x^2
$$

has unitary norm, is orthogonal to $e^1$ and satisfies (A.11) for $i = 2$. By induction, one can show that the vectors

$$
e^i = \frac{1}{\|x^i\|} x^i \quad i = 1, \ldots, k
$$

where
Clearly, Eq. (A.13) must be appropriately modified whenever a change of basis $\mathcal{X} \rightarrow \tilde{\mathcal{X}}$ is performed. If $T$ is the matrix corresponding to such a change of basis, we have
\[
\tilde{a} = Ta \quad \tilde{b} = Tb
\]
and, therefore,
\[
(x|y) = \tilde{a}^T \tilde{b} = a^T T^T \tilde{S} T b
\]
so that, in view of (A.13),
\[
S = T^T \tilde{S} T
\]
Finally, it is worth noting that Eq. (A.13) simplifies greatly whenever one chooses as a basis $\mathcal{X}$ of $X$ an orthonormal basis. In fact, in this case, $(e^i|e^j) = 0$ for $i \neq j$ and $(e^i|e^i) = 1$, so that $S = I$. Thus, Eq. (A.13) becomes $(x|y) = a^T b$, which coincides with the expression already written for the scalar product in $\mathbb{R}^n$.

A.12 ADJOINT TRANSFORMATIONS

Though adjoint transformations could be presented with reference to functional spaces, we restrict ourselves to finite-dimensional spaces in which the scalar field is the field of the reals.

**Definition 31** (adjoint transformation)

Let $L : X \rightarrow Y$ be a linear transformation between two finite dimensional spaces in which a scalar product is defined. The adjoint transformation of $L$ (often called adjoint operator) is the transformation $L^* : Y \rightarrow X$ with the following property:

\[
(Lx|y) = (x|L^*y)
\]

It is easy to see that Definition 31 makes sense since the adjoint transformation exists and is unique and linear. Other properties of the adjoint transformation (not all easy to prove) are the following:

\[
(L_1 + L_2)^* = L_1^* + L_2^*
\]
\[
(aL)^* = aL^*
\]
\[
(L_1 L_2)^* = L_2^* L_1^*
\]
\[
\|L\| = \|L^*\|
\]
\[
L^{**} = L
\]
Moreover, if \( L^{-1} \), exists then
\[
(L^{-1})^* = (L^*)^{-1}
\]

The adjoint transformation allows one to identify the null space of a transformation with the orthogonal complement of the image of the adjoint transformation. In fact, Theorem 22 holds.

**Theorem 22 (null space and adjoint transformation)**

Let \( L : X \to Y \) be a linear transformation between two finite-dimensional spaces \( X \) and \( Y \). Then,
\[
N(L) = \mathcal{I}(L^*)^\perp
\]  
(A.14)

Recalling that \((X^\perp)^\perp = X\) and that \((L^*)^* = L\) Eq. (A.14) can be also written as
\[
\mathcal{I}(L) = N(L^*)^\perp \quad \mathcal{I}(L)^\perp = N(L^*) \quad N(L^\perp) = \mathcal{I}(L^*)
\]

In the particular case in which \( L : X \to Y \) is a matrix \( A \), the following result (Theorem 23) can be established:

**Theorem 23 (adjoint transformations and transposed matrices)**

If \( A : \mathbb{R}^n \to \mathbb{R}^m \), the adjoint transformation \( A^* \) is \( A^T \).

Such a result, together with Theorem 22, yields
\[
N(A) = \mathcal{I}(A^T)^\perp
\]
that is, the null space of a matrix \( A \) is the subspace complementary to the subspace generated by the columns of \( A^T \) (viz, the rows of \( A \)).

**Example 7**

Let us determine the null space of the transformation \( A : \mathbb{R}^3 \to \mathbb{R}^2 \) given by
\[
A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}
\]

From the above remark, it follows that the null space \( N(A) \) is the subspace complementary to the space generated by the vectors
\[
x^1 = (\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix})^T \quad x^2 = (\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix})^T
\]

Since these two vectors are linearly independent, the subspace they generate is two dimensional and, therefore, \( N(A) \) is one dimensional.

The subspace \( N(A) \) is generated, for example, by the vector
\[
x^3 = (\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix})^T
\]
which is orthogonal to \( x^1 \) and \( x^2 \), as illustrated in Fig. A.11.