Chapter 1

BASIC ELEMENTS OF LINEAR ALGEBRA

1. Matrices

Let **C** and **R** be the field of real numbers and the field of complex numbers, respectively. Moreover let **i** be the imaginary unit, defined by the property $\mathbf{i}^2 = -1$. Let $\mathbf{C}^{m\times n}$ be the set of matrices with complex entries, with m rows and n columns; in many cases it will be useful to denote as $\mathbf{R}^{m\times n}$ the subset of matrices with real entries. If $A \in \mathbf{C}^{n\times n}$, then A is called a *square matrix of order n*.

Usually matrices are denoted by a capital letter, while their entries are denoted by the same letter, in lower case, followed by the indices (*row* index and *column*) index: for example a_{ij} is an element of the matrix A. Usually one writes:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

The entries a_{ij} such that $i = j$ are called the *diagonal* or *principal* entries of A and they form the *principal* or *main* diagonal of A.

Given a matrix $A \in \mathbf{C}^{m\times n}$, the *conjugate transpose* of A is defined as the matrix $B \in \mathbf{C}^{n\times m}$ such that

$$b_{ij} = \overline{a}_{ji},$$

where $\overline{a}_{ji}$ is the conjugate of the complex number a_{ji}, and it is denoted by $B = A^H$. If $A \in \mathbf{R}^{m\times n}$, then the conjugate transpose of A is simply the *transpose* matrix, defined as

$$B = A^T, \qquad b_{ij} = a_{ji}.$$

A matrix $A \in \mathbf{C}^{n\times n}$ is:

diagonal if $a_{ij} = 0$ for $i \neq j$;

scalar if it is diagonal and $a_{ii} = \alpha \in \mathbf{C}$;

upper (lower) triangular if $a_{ij} = 0$ for $i > j$ (for $i < j$);

strictly upper (lower) triangular if $a_{ij} = 0$ if $i \geq j$ (for $i \leq j$);

tridiagonal if $a_{ij} = 0$ for $|i - j| > 1$.

The following operations on matrices:

matrix addition $(\mathbf{C}^{m\times n} \times \mathbf{C}^{m\times n} \rightarrow \mathbf{C}^{m\times n})$:

$$C = A + B, \qquad c_{ij} = a_{ij} + b_{ij},$$

multiplication of a matrix by a number or *scalar multiplication* $(\mathbf{C} \times \mathbf{C}^{m\times n} \rightarrow \mathbf{C}^{m\times n})$:

$$B = \alpha\, A, \qquad b_{ij} = \alpha\, a_{ij},$$

make $\mathbf{C}^{m\times n}$ a *vector space* over $\mathbf{C}$, where the matrix with all zero entries is the *identity* or *neutral* element. This matrix is denoted by $O_{m\times n}$ or shortly O if its dimensions are clear from the context. Matrix addition is associative and commutative, and scalar multiplication is distributive over addition.

The *row by column multiplication* or *matrix multiplication* of two matrices $A \in \mathbf{C}^{m\times n}$ and $B \in \mathbf{C}^{n\times p}$ is the matrix $C = A\,B \in \mathbf{C}^{m\times p}$, with elements

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

(remark that the number of columns of A is equal to the number of rows of B).

The matrix multiplication is associative and distributive over addition, but it is not commutative (see exercises 1.1 e 1.2). Moreover the following property holds

$$(A\,B)^H = B^H A^H.$$

The scalar matrix of order n with all diagonal entries equal to 1 is called *identity* and it is denoted by I_n or shortly I if the order n is clear from the context. This matrix has the following properties:

$$\left.\begin{aligned} I_m A &= A \\ A I_n &= A \end{aligned}\right\} \qquad \text{for any matrix } \; A \in \mathbf{C}^{m\times n}.$$

A matrix $A \in \mathbf{C}^{n\times n}$ is called:

normal if $A^H A = A\,A^H$;

hermitian if $A^H = A$;

unitary if $A^H A = A\,A^H = I$.

Let $A \in \mathbf{R}^{n\times n}$; if A is hermitian, then $A^T = A$ and A is called *symmetric*; if A is unitary, then $A^T A = A\,A^T = I$ and A is said *orthogonal.*

1.1 Example. The matrix

$$G = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, \qquad \phi \in \mathbf{R},$$

is unitary, because

$$G\,G^H = G^H G = \begin{bmatrix} \sin^2\phi + \cos^2\phi & 0 \\ 0 & \sin^2\phi + \cos^2\phi \end{bmatrix} = I.$$

Moreover, since it is real, G is also orthogonal. ∎

An important case of orthogonal matrices is given by *permutation matrices*, which are natrices obtained by permuting the rows of the identity matrix I. Permutation matrices have just only one element different from zero in each row and in each column, and this element is equal to 1.

A subset of $\mathbf{C}^{n\times n}$ i called *closed under multiplication*, if, given two matrices A e B belonging to the subset, also their prodfuct $A\,B$ belongs to the subset. The following subsets of $\mathbf{C}^{n\times n}$ are closed under multiplication:

– upper (lower) triangular matrices,

– strictly upper (lower) triangular matrices,

– unitary matrices.

Given a matrix $A \in \mathbf{C}^{m\times n}$, a matrix $B \in \mathbf{C}^{k\times h}$, $0 \le k < m$, $0 \le h < n$, is called *submatrix* of A if it is obtained from A by deleting $m-k$ rows and $n-h$ columns. Given a matrix $A \in \mathbf{C}^{n\times n}$, a square submatrix B of order $k \le n$ of A is called *principal* if the principal elements of B are also principal elements of A (that is, the rows and the columns of A which have not been deleted have the same indices). A principal submatrix B, of order k, of A is called *leading principal* if it is composed by the elements $a_{ij},\ i,\ j = 1, \dots, k$.

1.2 Example. Let us consider the matrix $A \in \mathbf{R}^{3\times 3}$:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

The matrix

$$\begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$$

is a square submatrix of order 2 of A, the matrix

$$\begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix}$$

is a principal submatrix of order 2 of A, the matrix

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$$

is a leading principal submatrix of order 2 of A. ■

2. Vectors

If $A \in \mathbf{C}^{m\times 1}(A \in \mathbf{C}^{1\times m})$, then the matrix is made up of a single column (row) and is called *column (row) vector with m elements.*

Usually a *vector* is assumed to be a column vector and the vector space $\mathbf{C}^{m\times 1}$ of vectors with m elements is denoted by $\mathbf{C}^m$. In most cases a vector is denoted by a bold letter in lower case, and each element is denoted by a letter in lower case, followed by a subscript: for example x_i is the i-th element of the vector $\mathbf{x}$. One writes

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad \text{or also} \quad \mathbf{x} = [x_1,\, x_2,\, \dots,\, x_m]^T.$$

The vector with all elements zero is denoted by $\mathbf{0}$. If $\mathbf{x} \in \mathbf{C}^m$, then $\mathbf{x}^H \in \mathbf{C}^{1\times m}$ is the row vector whose elements are the conjugates of the corresponding elements of $\mathbf{x}$.

Particular cases of matrix multiplication::

Multiplication of a matrix by a vector $(\mathbf{C}^{m\times n} \times \mathbf{C}^n \to \mathbf{C}^m)$:

$$\mathbf{y} = A\mathbf{x}, \qquad y_i = \sum_{j=1}^{n} a_{ij}\, x_j, \qquad i = 1, \dots, m;$$

inner product of vectors $(\mathbf{C}^m \times \mathbf{C}^m \to \mathbf{C})$:

$$\alpha = \mathbf{x}^H \mathbf{y}, \qquad \alpha = \sum_{i=1}^{n} \overline{x}_i\, y_i;$$

outer product of vectors $(\mathbf{C}^m \times \mathbf{C}^{1\times n} \to \mathbf{C}^{m\times n})$:

$$A = \mathbf{x}\,\mathbf{y}^H, \qquad a_{ij} = x_i \overline{y}_j, \qquad i = 1, \dots, m, \quad j = 1, \dots, n.$$

The vector $\dfrac{1}{\alpha}\,\mathbf{x}$, $\alpha \neq 0, \alpha \in \mathbf{C}$, sometimes is denoted by $\dfrac{\mathbf{x}}{\alpha}$.

1.3 Esempio. Given the vectors

$$\mathbf{x} = [1,\ \mathbf{i},\ -\mathbf{i}]^T \qquad \text{e} \qquad \mathbf{y} = [\mathbf{i},\ 1,\ \mathbf{i}]^T,$$

we have

$$\mathbf{x}^H\mathbf{y} = -1,$$

$$\mathbf{x}\,\mathbf{y}^H = \begin{bmatrix} \text{-}\mathbf{i} & 1 & \text{-}\mathbf{i} \\ 1 & \mathbf{i} & 1 \\ \text{-}1 & \text{-}\mathbf{i} & \text{-}1 \end{bmatrix}. \qquad \blacksquare$$

The inner product of vectors is a *scalar product* over $\mathbf{C}^n$ and enjoys the following properties (see exercise 1.26):

1. $\mathbf{x}^H\mathbf{x}$ is real nonnegative, it is zero if and only if $\mathbf{x} = \mathbf{0}$;
2. $\overline{\mathbf{x}^H\mathbf{y}} = \mathbf{y}^H\mathbf{x}$;
3. $\mathbf{x}^H(\alpha\mathbf{y}) = \alpha\,\mathbf{x}^H\mathbf{y}$ for $\alpha \in \mathbf{C}$;
4. $\mathbf{x}^H(\mathbf{y}+\mathbf{z}) = \mathbf{x}^H\mathbf{y} + \mathbf{x}^H\mathbf{z}$ for $\mathbf{z} \in \mathbf{C}^n$.

The real number $\sqrt{\mathbf{x}^H\mathbf{x}}$ is the *euclidean length* of the vector $\mathbf{x}$, therefore the vector $\dfrac{\mathbf{x}}{\sqrt{\mathbf{x}^H\mathbf{x}}}$ has length 1. In $\mathbf{R}^n$, if $\mathbf{x}$ has length 1, the product $\mathbf{x}^H\mathbf{y}$ gives the *projection* of $\mathbf{y}$ onto the straight line passing through the origin and the point representing the vector $\mathbf{x}$. Since the inequality of *Cauchy-Schwarz* holds

$$|\mathbf{x}^H\mathbf{y}|^2 \le (\mathbf{x}^H\mathbf{x})\ (\mathbf{y}^H\mathbf{y}), \tag{1}$$

(see exercise 1.30) it is possible to define the *angle* θ between the two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$:

$$\theta = \arccos \frac{\mathbf{x}^H\mathbf{y}}{\sqrt{(\mathbf{x}^H\mathbf{x})\ (\mathbf{y}^H\mathbf{y})}}.$$

It is easy to verify that in $\mathbf{R}^2$ e in $\mathbf{R}^3$ this definition corresponds to the geometric notion of angle, as one can see in the case of $\mathbf{R}^2$ in figure 1.1.

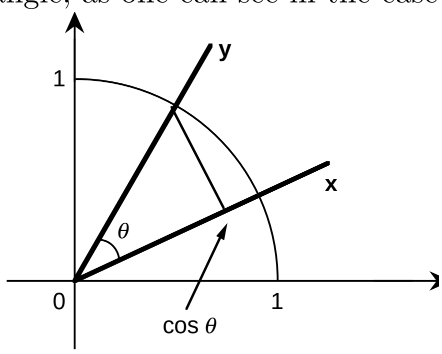

Fig.1.1 - Angle between two vectors.

When $\mathbf{x}^H\mathbf{y} = 0$, the two vectors $\mathbf{x}$ and $\mathbf{y}$ are called *orthogonal.*

1.4 Definition. The vecors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbf{C}^m, n \le m$, are called *linearly independent* if from the condition

$$\sum_{i=1}^{n} \alpha_i \mathbf{x}_i = \mathbf{0}, \quad \alpha_i \in \mathbf{C},$$

it follows that

$$\alpha_i = 0, \quad i = 1, \ldots, n. \qquad \blacksquare$$

n vectors, which are not linearly independent, are called *linearly dependent*; in this case, if $\alpha_k \neq 0$, one has

$$\mathbf{x}_k = \sum_{\substack{i=1 \\ i \neq k}}^{n} \beta_i \mathbf{x}_i, \quad \text{where} \quad \beta_i = -\frac{\alpha_i}{\alpha_k}, \quad i = 1, \ldots, n, \ i \neq k.$$

1.5 Definition. Let S be a subspace of $\mathbf{C}^n$. k vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in S$ form a *basis* of S if any vector $\mathbf{v} \in S$ can be expressed, in a unique way, as linear combination of the vectors of the basis

$$\mathbf{v} = \sum_{i=1}^{k} \alpha_i \mathbf{x}_i.$$

We say also that S is *generated* by the basis $\mathbf{x}_1, \ldots, \mathbf{x}_k$. ■

A particularly relevant basis of $\mathbf{C}^n$ is the so-called *canonical basis*, made up by the vectors

$$\mathbf{e}_i = [0, \ldots, 0, \underset{\substack{\uparrow \\ i}}{1}, 0, \ldots, 0]^T, \quad i = 1, \ldots, n,$$

which are the columns of the identity matrix of order n.

The k vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of a basis are linearly independent; moreover all the bases of a subspace have the same number of elements, and this number, denoted by $\dim S$, is called *dimension* of the subspace. The space $\mathbf{C}^n$, as vector space over the field $\mathbf{C}$, has dimension n, and any set of n linearly independent vectors of $\mathbf{C}^n$ is a basis of $\mathbf{C}^n$.

Let S e T be two subspaces di $\mathbf{C}^n$. Then the sum

$$S + T = \{\mathbf{s} + \mathbf{t}, \ \mathbf{s} \in S, \ \mathbf{t} \in T\}$$

and the intesection $S \cap T$ are subspaces too. Their dimensions obey to the following relation

$$\dim(S + T) = \dim S + \dim T - \dim(S \cap T), \qquad (2)$$

which implies

$$\max\{\dim S, \dim T\} \le \dim(S + T) \le \min\{\dim S + \dim T, n\}, \qquad (3)$$

$$\max\{0, \dim S + \dim T - n\} \le \dim(S \cap T) \le \min\{\dim S, \dim T\}. \qquad (4)$$

If $S \cap T = \{\,\mathbf{0}\,\}$, the subspace $X = S + T$ is called *direct sum* of S and T, and is usually denoted by $S \oplus T$. In this case

$$\dim X = \dim S + \dim T,$$

and the vectors $\mathbf{x}$ of X can be expressed in a unique way as the sum

$$\mathbf{x} = \mathbf{s} + \mathbf{t}, \quad \mathbf{s} \in S,\ \mathbf{t} \in T.$$

1.6 Definition. Let S be a subspace of $\mathbf{C}^n$. The subspace

$$S^{\perp} = \{\mathbf{u} \in \mathbf{C}^n \ : \ \mathbf{u}^H \mathbf{v} = 0 \text{ per ogni } \mathbf{v} \in S\}$$

is called *subspace orthogonal* to S. The following relations hold

$$S \cap S^{\perp} = \{\mathbf{0}\},$$

$$S \oplus S^{\perp} = \mathbf{C}^n,$$

$$\dim S^{\perp} = n - \dim S.$$

thus any vector $\mathbf{x} \in \mathbf{C}^n$ can be expressed in a unique way as

$$\mathbf{x} = \mathbf{s} + \mathbf{t},\ \mathbf{s} \in S,\ \mathbf{t} \in S^{\perp}. \qquad (5)$$

The vector $\mathbf{s}$ is called *orthogonal projection* of $\mathbf{x}$ onto S. ∎

1.7 Example. In $\mathbf{R}^3$ let S be the subspace generated by the vector

$$\mathbf{x}_1 = [0,\ 0,\ 1]^T.$$

The vectors of S are all those vectors having two zeros as first and second entries, thus its dimension is 1. The space $S^{\perp}$, formed by the vectors having zero as third entry, is generated by the vectors

$$\mathbf{x}_2 = [1,\ 0,\ 0]^T \qquad \text{e} \qquad \mathbf{x}_3 = [0,\ 1,\ 0]^T$$

and its dimension is 2. The figure 1.2 gives a geometric explanation of the relation (5) for this example. (5). ∎

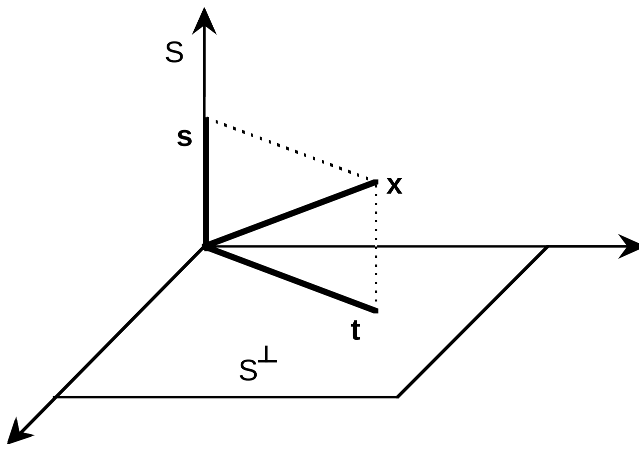

Fig. 1.2 - Orthogonal projection of $\mathbf{x}$ onto S.

1.8 Definition. n nonzero vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbf{C}^m$ are called *orthogonal* if $\mathbf{x}_i^H \mathbf{x}_j = 0$ for $i \neq j$; they are called *orthonormal* if they are orthogonal and moreover $\mathbf{x}_i^H \mathbf{x}_i = 1$, i.e. if they have length 1 or, shortly, if they are *normalized*. In this case the following notation is often used,

$$\mathbf{x}_i^H \mathbf{x}_j = \delta_{ij},$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

is the *Kronecker delta*. ∎

We remark that n orthogonal vectors are also linearly independent.

1.9 Example. The vectors

$$\mathbf{x} = [1,\, \mathbf{i},\, -\mathbf{i}]^T \qquad \text{e} \qquad \mathbf{y} = [\mathbf{i},\, 1,\, \mathbf{i}]^T,$$

introduced in the example 1.3 are linearly independent, but are not orthogonal, since $\mathbf{x}^H \mathbf{y} = -1 \neq 0$. The vectors

$$\mathbf{u} = \frac{1}{\sqrt{3}}\, \mathbf{x} \qquad \text{e} \qquad \mathbf{v} = \frac{1}{\sqrt{8}}\, [-2\mathbf{i},\, -1 - \mathbf{i},\, 1 - \mathbf{i}]^T$$

are orthonormal, since $\mathbf{u}^H \mathbf{u} = 1$, $\mathbf{v}^H \mathbf{v} = 1$ e $\mathbf{u}^H \mathbf{v} = 0$. The vector

$$\mathbf{z} = \mathbf{i}\mathbf{x} + \mathbf{y} = [2\mathbf{i},\, 0,\, \mathbf{i} + 1]^T$$

is a linear combination of $\mathbf{x}$ and $\mathbf{y}$, thus the vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ are linearly dependent. ∎

Among all possible bases of $\mathbf{C}^n$, a particular rôle is played by the *orthonormal bases*, i.e. those ones whose vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are orthonormal.

If we choose a basis $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of a subspace S of $\mathbf{C}^n$, we can construct an orthonormal basis $\mathbf{y}_1, \ldots, \mathbf{y}_k$ through *Gram-Schmidt orthogonalization* based on the following theorem.

1.10 Theorem. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbf{C}^n$, $k \le n$, be k linearly independent vectors. The vectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$, built in this way*

$$\begin{aligned}
&\mathbf{t}_1 = \mathbf{x}_1 \qquad \mathbf{y}_1 = \mathbf{t}_1 / \sqrt{\mathbf{t}_1^H \mathbf{t}_1},\\
&\mathbf{t}_i = \mathbf{x}_i - \sum_{j=1}^{i-1} (\mathbf{y}_j^H \mathbf{x}_i)\, \mathbf{y}_j, \qquad \mathbf{y}_i = \mathbf{t}_i / \sqrt{\mathbf{t}_i^H \mathbf{t}_i}, \qquad i = 2, \ldots, k,
\end{aligned}$$

are orthonormal.

Proof. The vectors $\mathbf{y}_i$ are normalized. In order to prove their orthogonality we use induction on k. For $k = 2$, since

$$\mathbf{t}_2^H \mathbf{y}_1 = \mathbf{x}_2^H \mathbf{y}_1 - (\mathbf{x}_2^H \mathbf{y}_1)\, \mathbf{y}_1^H \mathbf{y}_1 = 0,$$

it follows that $\mathbf{y}_2^H \mathbf{y}_1 = 0$. For $k > 2$, we assume that the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}$ are orthonormal, and then we show that $\mathbf{t}_k$ is orthogonal to $\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}$. In detail, from the equation

$$\mathbf{y}_j^H \mathbf{y}_i = 0 \quad \text{per} \quad j,\ i \le k-1,\ i \ne j,$$

it follows:

$$\begin{aligned}
\mathbf{t}_k^H \mathbf{y}_i &= \mathbf{x}_k^H \mathbf{y}_i - \sum_{j=1}^{k-1} (\mathbf{x}_k^H \mathbf{y}_j)\, \mathbf{y}_j^H \mathbf{y}_i\\
&= \mathbf{x}_k^H \mathbf{y}_i - (\mathbf{x}_k^H \mathbf{y}_i)\, \mathbf{y}_i^H \mathbf{y}_i = 0.
\end{aligned}$$
∎

1.11 Example. The vectors of $\mathbf{C}^n$

$$\mathbf{x}_i = [\underbrace{1, \ldots, 1}_{i \text{ entries}}, 0, \ldots, 0]^T, \quad i = 1, \ldots, n,$$

form a basis of $\mathbf{C}^n$, but this basis is not orthonormal. By applying the Gram-Schmidt process to the vectors $\mathbf{x}_i$, we obtain the vectors $\mathbf{e}_i$, $i = 1, \ldots, n$, of the canonical basis of $\mathbf{C}^n$. The vectors of $\mathbf{C}^n$

$$\mathbf{x}_1 = \mathbf{e}_1 + \mathbf{e}_2, \quad \mathbf{x}_2 = \mathbf{e}_2 + \mathbf{e}_3, \quad \ldots, \quad \mathbf{x}_{n-1} = \mathbf{e}_{n-1} + \mathbf{e}_n, \quad \mathbf{x}_n = \mathbf{e}_n + \mathbf{e}_1$$

are linearly independent. By applying the Gram-Schmidt process we obtain the vectors

$$\begin{aligned}
\mathbf{y}_1 &= \frac{1}{\sqrt{2}}\,[1,\ 1,\ 0,\ \ldots,\ 0]^T,\\
\mathbf{y}_2 &= \frac{1}{\sqrt{2}\sqrt{3}}\,[1,\ -1,\ -2,\ 0,\ \ldots,\ 0]^T,\\
\mathbf{y}_3 &= \frac{1}{\sqrt{3}\sqrt{4}}\,[1,\ -1,\ 1,\ 3,\ 0,\ \ldots,\ 0]^T,\\
&\vdots\\
\mathbf{y}_{n-1} &= \frac{1}{\sqrt{n-1}\sqrt{n}}\,[1,\ -1,\ \ldots,\ (-1)^n,\ (-1)^n(n-1)]^T,\\
\mathbf{y}_n &= \frac{1}{\sqrt{n}}\,[1,\ -1,\ 1,\ \ldots,\ (-1)^n,\ (-1)^{n+1}]^T,
\end{aligned}$$

which form an orthonormal basis of $\mathbf{C}^n$. ■

3. Positive definite matrices

If $A \in \mathbf{C}^{n\times n}$ is a hermitian matrix, that is $A = A^H$, and $\mathbf{x} \in \mathbf{C}^n$, the number

$$\alpha = \mathbf{x}^H A\mathbf{x}$$

is real. This is due to the fact that, being A hermitian, we have:

$$\overline{\alpha} = \overline{\mathbf{x}^H A\mathbf{x}} = (\mathbf{x}^H A\mathbf{x})^H = \mathbf{x}^H A^H\mathbf{x} = \mathbf{x}^H A\mathbf{x} = \alpha.$$

1.12 Definition. Let $A \in \mathbf{C}^{n\times n}$ be a hermitian matrix. If for any $\mathbf{x} \in \mathbf{C}^n$, $\mathbf{x} \neq \mathbf{0}$, the real number $\alpha = \mathbf{x}^H A\mathbf{x}$ has the same sign, we say that the matrix A is *definite*, and, in particular:

if $\mathbf{x}^H A\mathbf{x} > 0$ $\quad A$ is *positive definite*,
if $\mathbf{x}^H A\mathbf{x} \geq 0$ $\quad A$ is *positive semidefinite*,
if $\mathbf{x}^H A\mathbf{x} \leq 0$ $\quad A$ is *negative semidefinite*,
if $\mathbf{x}^H A\mathbf{x} < 0$ $\quad A$ is *negative definite*. ■

1.13 Example. The hermitian matrix

$$A = \begin{bmatrix} 3 & \mathbf{i} \\ -\mathbf{i} & 3 \end{bmatrix}$$

is positive definite, because for any $\mathbf{x} = [x_1\ ,\ x_2]^T \neq \mathbf{0}$ we have:

$$\mathbf{x}^H A\mathbf{x} = |x_1 - \mathbf{i}x_2|^2 + 2|x_2 - \mathbf{i}x_1|^2 > 0.$$ ■

1.14 Theorem. *If a matrix $A \in \mathbf{C}^{n\times n}$ is positive definite, also its principal submatrices are positive definite as well.*

Proof. Let B a principal submatrix of A obtained by deleting $(n-i)$ rows and the cooresponding $(n-i)$ columns. For any vector $\mathbf{x} \in \mathbf{C}^i$, $\mathbf{x} \neq 0$, let us consider the vector $\mathbf{y} \in \mathbf{C}^n$ having zero entries in the positions corresponding to the rows deleted and the same entries of $\mathbf{x}$ in the positions corresponding to the rows left. Thus, since A is positive definite, we have

$$\mathbf{x}^H B\mathbf{x} = \mathbf{y}^H A\mathbf{y} > 0. \qquad \blacksquare$$

The principal submatrices of order 1 are made by a single element, therefore all the principal entries of a positive definite matrix, besides being real because the matrix is hermitian, are also positive.

4. The determinant

1.15 Definition. Let $A \in \mathbf{C}^{n\times n}$. The *determinant* of A is defined as the number

$$\det A = \sum_{\pi\in P} \operatorname{sgn}(\pi)\, a_{1,\pi_1}\, a_{2,\pi_2} \ldots a_{n,\pi_n}$$

where P is the set of the $n!$ vectors $\pi = [\pi_1, \pi_2, \ldots, \pi_n]^T$, returned by all the permutations of the vector $[1,\, 2, \ldots, n]^T$; any permutation can be expressed as the composition of a finite number of index transpositions, in infinitely many ways, but the parity of this number of transpositions depends only on the permutation, and is referred as the parity of the permutation; the value $\operatorname{sgn}(\pi)$ is $+1$ or -1 depending if the parity is even or odd. ■

The determinant of a matrix can be expressed in a simpler way by using the *Laplace expansion*. Let A_{ij} be the square submatrix of order $n-1$ obtained fron the matrix A by deleting the i-th row and the j-th column, for any index i we have:

$$\det A = \begin{cases} a_{11} & \text{if } n = 1, \\ \displaystyle\sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} & \text{if } n > 1. \end{cases} \tag{6}$$

Let $A,\, B \in \mathbf{C}^{n\times n}$, $\alpha \in \mathbf{C}$; the following properties hold:

$$\det A = \prod_{i=1}^{n} a_{ii} \qquad \text{if } A \text{ is diagonal or triangular;}$$

$\det I = 1$;

$\det A^T = \det A$

$\det A^H = \overline{\det A}$;

$\det(A\,B) = \det A \det B$ (*Binet*'s theorem);

$\det B = \alpha \det A$, if B is obtained from A by multiplying by α a row (or a column);

$\det(\alpha A) = \alpha^n \det A$;

$\det B = -\det A$, if B is obtained from A by exchanging two rows (or columns);

$\det B = \det A$, if B is obtained from A by adding to a row (or column) another row (or column) multiplied by a number;

$\det A = 0$, if two or more rows (or columns) of A are linearly dependent.

Since $\det A = \det A^T$, the Laplace expansion for the computation of the determinant of A can be applied summing over the row index i in the formula (6).

5. The inverse matrix

1.16 Definitions. Let $A \in \mathbf{C}^{n\times n}$, we define:

inverse matrix of A a matrix $B \in \mathbf{C}^{n\times n}$ such that

$$AB = BA = I,$$

adjoint matrix of A the matrix $\text{adj}A \in \mathbf{C}^{n\times n}$, whose entry in position (i, j) is given by

$$(-1)^{i+j} \det A_{ji},$$

where A_{ji} is the submatrix obtained from A by deleting the j-th row and the i-th column. ■

A matrix A which does not admit an inverse matrix is called *singular*. The following relation holds (see the exercise 1.48)

$$A\ \text{adj}A = (\det A) I.$$

It follows that A is not singular (or *nonsingular*) if $\det A \neq 0$, thus if A is nonsingular, the inverse matrix, which is denoted as A^{-1}, is unique and it can be expressed as

$$A^{-1} = \frac{1}{\det A}\,\mathrm{adj}A.$$

The following properties hold:

$(A^H)^{-1} = (A^{-1})^H$ (the notation A^{-H} can be used);

$A^{-1} = A^H$ if A is unitary, that means $A^H A = AA^H = I$;

$\det A^{-1} = 1/\det A$;

$(A\,B)^{-1} = B^{-1}A^{-1}$.

The following subsets of $\mathbf{C}^{n\times n}$ are *closed with respect to the inversion*, that is if A is a nonsingular matrix in the subset, also A^{-1} is in the subset as well:

- hermitian matrices,
- unitary matrices,
- normal matrices,
- positive (negative) definite matrices,
- superior (inferior) triangular matrices,
- diagonal matrices.

6. Linear systems

Let $A \in \mathbf{C}^{m\times n}$, and consider the following subspaces:

$$S(A) = \{\mathbf{y} \in \mathbf{C}^m \ : \ \mathbf{y} = A\mathbf{x},\ \mathbf{x} \in \mathbf{C}^n\},$$

called *range* of A and

$$N(A) = \{\mathbf{x} \in \mathbf{C}^n \ : \ A\mathbf{x} = \mathbf{0}\},$$

called *kernel* or *null space* of A. It is well known that

$$S(A)^\perp = N(A^H),$$

and therefore

$$\dim S(A) + \dim N(A^H) = m.$$

The number $\dim S(A)$ is called *rank* of A, it is equal to the number of rows (and of columns, see the exercise 1.35) linearly independent of A. Since the rank of A and the rank of A^H are equal, we have

$$\dim S(A) + \dim N(A) = n. \tag{7}$$

More generally, if T is a subspace of $\mathbf{C}^n$, after setting

$$S_T(A) = \{\mathbf{y} \in \mathbf{C}^m \ : \ \mathbf{y} = A\mathbf{x},\ \mathbf{x} \in T\},$$

$$N_T(A) = \{\mathbf{x} \in T \ : \ A\mathbf{x} = \mathbf{0}\} = N(A) \cap T,$$

we have

$$\dim S_T(A) + \dim N_T(A) = \dim T.$$

1.17 Example. Let

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

The subspace $S(A)$ is spanned by the vectors

$$\mathbf{y}_1 = [1,\ 1,\ 1]^T \quad \text{and} \quad \mathbf{y}_2 = [1,\ -1,\ 0]^T,$$

therefore

$$\text{rank of } A = \dim S(A) = 2.$$

The kernel of A is the subspace spanned by the vectors

$$\mathbf{x}_1 = [1,\ 0,\ -1,\ 0]^T \quad \text{and} \quad \mathbf{x}_2 = [0,\ 1,\ 0,\ -1]^T,$$

therefore

$$\dim N(A) = 2.$$

The subspace $S(A^T)$ is spanned by the vectors

$$\mathbf{x}_3 = [1,\ 1,\ 1,\ 1]^T \quad \text{and} \quad \mathbf{x}_4 = [1,\ -1,\ 1,\ -1]^T,$$

and we have

$$\text{rank of } A^T = \dim S(A^T) = \dim S(A) = 2.$$

The kernel of A^T is the subspace spanned by the vector

$$\mathbf{y}_3 = [1,\ 1,\ -2]^T$$

and we have

$$\dim N(A^T) = 1. \qquad \blacksquare$$

1.18 Example. If $\mathbf{x},\ \mathbf{y} \in \mathbf{C}^n, \mathbf{x},\ \mathbf{y} \neq \mathbf{0}$, the matrix (called *dyad*)

$$A = \mathbf{x}\,\mathbf{y}^H$$

has rank 1. This is due to the fact that the columns of A are the vectors

$$\overline{y}_1\,\mathbf{x},\ \overline{y}_2\,\mathbf{x},\ \ldots,\ \overline{y}_n\,\mathbf{x},$$

which are pairwise linearly dependent. ■

If $m = n$, A is nonsingular if and only if rank of $A = n$, and from (7) it follows that A is nonsingular if and only if

$$\dim N(A) = 0,$$

and this means that the kernel of A contains only the null vector.

If rank of $A = r = \min\{m, n\}$, then the matrix A is said to have *maximum rank*. In this case the matrix $A^H A \in \mathbf{C}^{n\times n}$ has rank r and if $r = n$, the matrix $A^H A$ is nonsingular. Counterwise if $A^H A$ is nonsingular, then $m \geq n$ and rank of A is maximum.

1.19 Definition. Let $A \in \mathbf{C}^{m\times n}$, $\mathbf{b} \in \mathbf{C}^m$; we define *linear system of m equations in n unknowns* the system

$$A\mathbf{x} = \mathbf{b}, \tag{8}$$

where $\mathbf{x} \in \mathbf{C}^n$ is the vector of the unknowns, A is the *system matrix* and $\mathbf{b}$ is the *constant terms vector*. The system is said *consistent* if some solution exists. ■

1.20 Theorem. *The following conditions are equivalent*:

a) *the system (8) is consistent*,

b) $\mathbf{b} \in S(A)$,

c) *the matrix A and the matrix $[A|\mathbf{b}]$, obtained by appending to A the vector $\mathbf{b}$ as $n+1$-th column, have the same rank.* ■

If the system (8) is consistent and $\mathbf{x}$ is a solution, then any solution of (8) can be expressed as $\mathbf{x} + \mathbf{y}$, where $\mathbf{y}$ is such that $A\mathbf{y} = \mathbf{0}$, that is $\mathbf{y} \in N(A)$. Therefore the solution is unique if and only if $\dim N(A) = 0$.

The following cases can occur:

1. If $n = m$, and the matrix A is nonsingular, then $S(A) = \mathbf{C}^n$ and $N(A) = \{\mathbf{0}\}$. In this case the system is consistent, the solution is unique and can be expressed as

$$\mathbf{x} = A^{-1}\mathbf{b},$$

and, by using the *Cramer's rule*, also as

$$x_i = \frac{\det A_i}{\det A}, \quad i = 1, \ldots, n,$$

where A_i is the matrix obtained from A by replacing the i-th column with the vector $\mathbf{b}$. If $\mathbf{b} = \mathbf{0}$ (*homogeneous system*), the system admits only the zero vector as solution. If, on the contrary, the matrix A is singular, the system may be not consistent. Anyway, the system is consistent if it is homogeneous, because after appending the zero vector $\mathbf{b}$ to the matrix A we obtain a matrix with the same rank as A.

2. If $m < n$, that is there are more unknowns than equations, the system, if consistent, admits infinitely many solutions since $\dim S(A) \leq m$ and then $\dim N(A) \geq n - m > 0$.

3. If $n < m$, that is there are more equations than unknowns, the system can be consistent only if there are at least $m - n$ equations which are linear combinations of the left ones.

7. Block matrices

Often it can be simpler to describe a matrix in terms of its submatrices instead of defining its entries. For instance the matrix

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

can be described in this much more compact fashion:

$$A = \begin{bmatrix} I_2 & E \\ E^T & I_3 \end{bmatrix},$$

where $E \in \mathbf{R}^{2\times 3}$ is the matrix

$$E = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

We say that A is *block partitioned* or also that A is a 2×2 block matrix. In general a $p \times q$ block matrix is a matrix of this form:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1q} \\ A_{21} & A_{22} & \dots & A_{2q} \\ \vdots & \vdots & & \vdots \\ A_{p1} & A_{p2} & \dots & A_{pq} \end{bmatrix},$$

where $A_{ij} \in \mathbf{C}^{m_i \times n_j}$, and m_i, n_j are positive integers, for $i = 1, \ldots, p$, $j = 1, \ldots, q$, therefore $A \in \mathbf{C}^{m \times n}$, with

$$m = \sum_{i=1}^{p} m_i, \quad n = \sum_{j=1}^{q} n_j.$$

A frequent case is when some blocks are row or column vectors, as for the matrix $A \in \mathbf{C}^{n \times n}$

$$A = \begin{bmatrix} \alpha & \mathbf{v}^H \\ & \\ \mathbf{u} & B \end{bmatrix}$$

where $\alpha \in \mathbf{C}$, $\mathbf{u}$, $\mathbf{v} \in \mathbf{C}^{n-1}$, $B \in \mathbf{C}^{(n-1)\times(n-1)}$.

Many of the definitions given in previous sections can be easily extended to block matrices. For instance the block matrix

$$A = \begin{bmatrix} A_{11} & O & O \\ A_{21} & A_{22} & O \\ A_{31} & A_{32} & A_{33} \end{bmatrix},$$

is called *block lower triangular*. The multiplication of two blocks matrices A e B can be defined in terms of block row by block column products. For instance, if

$$A = \begin{bmatrix} A_{11} & A_{12} \\ & \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ & \\ B_{21} & B_{22} \end{bmatrix}, \quad C = \begin{bmatrix} C_{11} & C_{12} \\ & \\ C_{21} & C_{22} \end{bmatrix},$$

are 2×2 block matrices such that $C = AB$, then we have

$$C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j}, \quad i,\ j = 1, 2.$$

This property holds in the general case of block matrices, provided that the number of blocks and their sizes are compatible.

8. Reducible matrices

1.21 Definition. A matrix A of order $n \geq 2$ is called *reducible* if there exists a permutation matrix Π and an integer number k, $0 < k < n$, such that

$$B = \Pi A \Pi^T = \begin{bmatrix} A_{11} & A_{12} \\ & \\ O & A_{22} \end{bmatrix} \begin{matrix} \} \;\; k \text{ righe} \\ \\ \} \; n-k \text{ righe} \end{matrix} \tag{9}$$

where $A_{11} \in \mathbf{C}^{k\times k}$ e $A_{22} \in \mathbf{C}^{(n-k)\times(n-k)}$. If the matrix A is not reducible, A is called *irreducible.* ■

If a matrix A is reducible, there many permutation matrices Π can exist which allow to transform the matrix A into a matrix B having the form (9). If the matrix A of the linear system (8) is reducible, since the matrix Π in (9) is orthogonal, we have

$$\Pi A \Pi^T \Pi \mathbf{x} = \Pi \mathbf{b},$$

and after setting $\mathbf{y} = \Pi \mathbf{x}$ e $\mathbf{c} = \Pi \mathbf{b}$, we have

$$B\mathbf{y} = \mathbf{c}.$$

By partitioning the vectors

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \\ \mathbf{y}_2 \end{bmatrix} \begin{matrix} \} \quad k \text{ entries} \\ \\ \} \; n-k \text{ entries} \end{matrix} \qquad \mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \\ \mathbf{c}_2 \end{bmatrix} \begin{matrix} \} \quad k \text{ entries} \\ \\ \} \; n-k \text{ entries} \end{matrix}$$

where $\mathbf{y}_1, \mathbf{c}_1 \in \mathbf{C}^k, \mathbf{y}_2, \mathbf{c}_2 \in \mathbf{C}^{n-k}$, the system (8) can be written in the form

$$\begin{cases} A_{11}\,\mathbf{y}_1 + A_{12}\,\mathbf{y}_2 = \mathbf{c}_1 \\ \\ \qquad\quad A_{22}\,\mathbf{y}_2 = \mathbf{c}_2. \end{cases}$$

The resolution of the system (8) with a coefficient matrix of order n is addressed to the resolution of two linear systems, the first with a coefficient matrix of order $n-k$, the second with a coefficient matrix of order k.

In order to see if the matrix A is reducible, we can consider the *direct graph* of A, that is the graph which has as many nodes p_i, as the order n of A, and any ordered pair of nodes p_i (start node) and p_j (end node) are connected by a directed arc if a_{ij} is different from 0.

1.22 Example. The graph of the matrix

$$A = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 2 & \text{-}1 \\ \text{-}1 & 0 & 2 \end{bmatrix} \tag{10}$$

is represented in the figure 1.3. ■

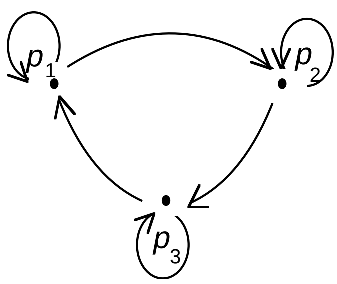

Fig. 1.3 - Directed graph of the matrix (10).

Two arcs of a directed graph are called *consecutive* if the end node of the first arc is the start node of the second one. A sequence of consecutive directed arcs is called *directed path.* A directed path is called *closed* if the start node of the first arc of the path is also the end node of the last arc.

1.23 Definition. A directed graph is called *strongly connected* if for any ordered pair of indices i, j, $1 \le i$, $j \le n$, with $i \ne j$, there exists a directed path wich starts fron the node p_i and ends at the node p_j. ■

1.24 Theorem. *A matrix A is reducible if and only if its directed graph is not strongly connected.*

Proof. First of all we note that the directed graphs of the matrix A and of the matrix $B = \Pi A \Pi^T$ differ only for a permutation of the indices of the nodes p_i. If the matrix A is reducible, and we consider the matrix B in (9) and some index i, with $k < i \le n$, then we see that there cannot be any directed path starting from p_i and ending at a node p_j, $j \le k$. Conversely, if the graph of A is not strongly connected, then there exists some node p_j from which it is not possible to reach another node of the graph, at least. Let $\mathcal{P}$ be the subset of nodes which can be reached starting from p_j and $\mathcal{Q}$ the subset of nodes which cannot be reached starting from p_j. The subsets $\mathcal{P}$ e $\mathcal{Q}$ form a partition of the set of all nodes and $\mathcal{Q}$ is not empty. Moreover there cannot exist directed paths starting from a node in $\mathcal{P}$ and ending at a node in $\mathcal{Q}$. The nodes can be rearranged with a permutation, in such a way that $\mathcal{Q} = \{p_1, p_2, \ldots, p_k\}$, with $k \ge 1$, and $\mathcal{P} = \{p_{k+1}, \ldots, p_n\}$. The matrix B obtained rearranging in the same way rows and columns of A is such that $b_{ij} = 0$ if $i > k$ e $j \le k$. ■

From theorem 1.24 it follows that if A is irreducible, then there exists a directed path which touches all the nodes of the graph.

1.25 Example. Given the linear system $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 2 & 3 & -2 & 1 \\ -1 & 0 & -2 & 0 \\ 1 & -1 & 1 & 4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ -2 \\ -1 \\ -2 \end{bmatrix}, \tag{11}$$

in order to see if A è reducible, we consider its directed graph, represented in figure 1.4:

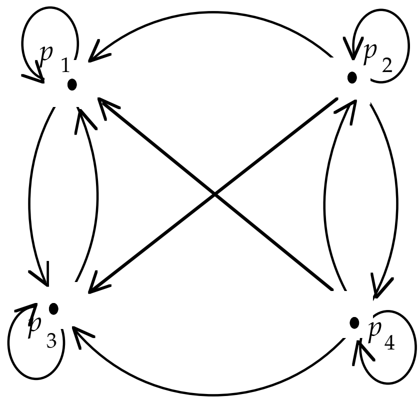

Fig. 1.4 - Directed graph of the matrix (11).

In this graph

p_1 is the end node of directed paths starting from the nodes p_1, p_2, p_3, p_4;

p_2 is the end node of directed paths starting from the nodes p_2, p_4;

p_3 is the end node of directed paths starting from the nodes p_1, p_2, p_3, p_4;

p_4 is the end node of directed paths starting from the nodes p_2, p_4.

At this point we permute the nodes p_1 and p_4, in this way the two nodes which cannot be reached from the other two are brought in the first two positions. The permutation matrix which represents this permutation of indices is

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

and is such that

$$B = \Pi A \Pi^T = \begin{bmatrix} 4 & -1 & 1 & 1 \\ 1 & 3 & -2 & 2 \\ 0 & 0 & -2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} -2 \\ -2 \\ -1 \\ 1 \end{bmatrix}.$$

By solving the two systems

$$\begin{bmatrix} -2 & -1 \\ -1 & 1 \end{bmatrix} \mathbf{y}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 4 & -1 \\ 1 & 3 \end{bmatrix} \mathbf{y}_1 = \begin{bmatrix} -2 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ -2 & 2 \end{bmatrix} \mathbf{y}_2,$$

we obtain first

$$\mathbf{y} = [-1,\ -1,\ 0,\ 1]^T,$$

and finally

$$\mathbf{x} = \Pi^T \mathbf{y} = [1,\ -1,\ 0,\ -1]^T. \qquad \blacksquare$$