Notes for the course
Mathematics for Data Science
Università di Trento

Simone Ugolini

Last update: 14:19, Saturday 17^{th} October, 2020

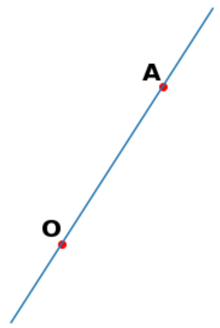
Contents

Chapter 1

Some geometry of the plane and the space

1.1 Introduction

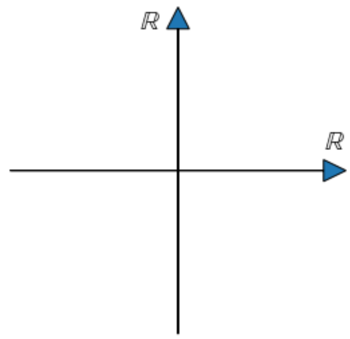
We consider two points O and A in the plane. We know that there exists a unique line passing through O and A. We could embed the points O and A in a Cartesian plane centered at O and find some equation of the line passing through O and A. Before doing so, we review the Cartesian plane.



1.1.1 Cartesian plane

We denote by $\mathbb{R}$ the set of real numbers. We represent real numbers on the real axis, namely an oriented line, where we fix a unit of measure.

If we take two orthogonal real lines intersecting at a point O, then we construct the Cartesian plane.

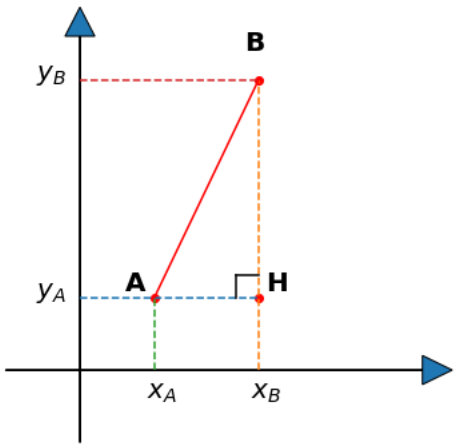


If we take two points $A = (x_A, y_A)$ and $B = (x_B, y_B)$ in the plane, then we can compute the distance between A and B using Pythagoras theorem. As illustrated in the figure aside, we construct a right triangle. Then, for the length $\|\overline{AB}\|$ of the line segment $\overline{AB}$ we have that

$$\|\overline{AB}\| = \sqrt{\|\overline{AH}\|^2 + \|\overline{BH}\|^2}$$

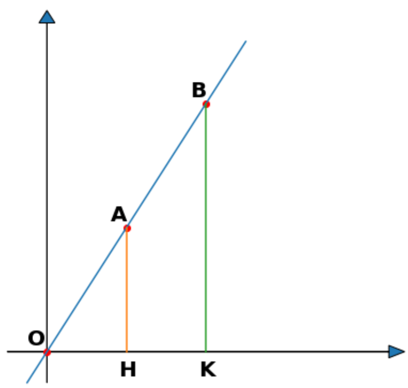
Therefore

$$\|\overline{AB}\| = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$



1.1.2 Lines in the Cartesian plane passing through O=(0,0)

We embed two points O and A in a Cartesian plane in such a way that O is the center of the plane. Then we take another point B on the line passing through O and A. We construct two triangles $\overset{\triangle}{AHO}$ and $\overset{\triangle}{BKO}$, where the points H and K are respectively the orthogonal projection of A and B onto the x-axis.



We can notice that the triangles $\overset{\triangle}{AHO}$ and $\overset{\triangle}{BKO}$ are similar since their corresponding angles have the same measure. Therefore the sides of the triangles are proportional, namely there exists a constant $t > 0$ such that

$$\|\overline{OK}\| = t \cdot \|\overline{OH}\|;$$
$$\|\overline{BK}\| = t \cdot \|\overline{AH}\|.$$

Since $\|\overline{OH}\| = x_A$ and $\|\overline{AH}\| = y_A$, we have that

$$\begin{cases} x_B = tx_A \\ y_B = tx_A \end{cases}$$

Indeed, the equations above do not depend on the choice of the point B. In fact, in the picture above the coordinates of the point B have the same sign of the coordinates of the point A. One could also consider the case that the coordinates of B and the coordinates of A have opposite signs. In this case the triangles $\overset{\triangle}{AHO}$ and $\overset{\triangle}{BKO}$ are still similar and there exists a constant $k > 0$ such that

$$\|\overline{OK}\| = k \cdot \|\overline{OH}\|;$$
$$\|\overline{BK}\| = k \cdot \|\overline{AH}\|.$$

By the way, since $x_B < 0$ and $y_B < 0$, we have that $\|\overline{OK}\| = |x_B| = -x_B$ and $\|\overline{BK}\| = |y_B| = -y_B$. Hence

$$\begin{cases} -x_B = k \cdot x_A \\ -y_B = k \cdot y_A \end{cases} \Leftrightarrow \begin{cases} x_B = (-k) \cdot x_A \\ y_B = (-k) \cdot y_A \end{cases}$$

Therefore if we define $t := -k$ we can say that

$$\begin{cases} x_B = t \cdot x_A \\ y_B = t \cdot y_A \end{cases}$$

In conclusion, the points (x, y) of the line passing through the points O and A are described by the parametric equations

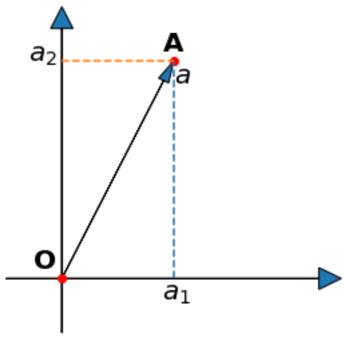
$$\begin{cases} x = t \cdot x_A \\ y = t \cdot y_A \end{cases} \qquad \text{with } t \in \mathbb{R}$$

1.2 Points and vectors of $\mathbb{R}^2$

We define $\mathbb{R}^2$ as the set of all the ordered pairs of real numbers:

$$\mathbb{R}^2 = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}\}.$$

The elements of $\mathbb{R}^2$ can be represented as points in the Cartesian plane. Moreover, we can associate an arrow with every pair of $\mathbb{R}^2$. If we consider the pair (a_1, a_2), then we can draw an arrow which joins the point O to the point $A = (a_1, a_2)$. If we want to distinguish the two objects, we denote by $a = (a_1, a_2)$ the arrow.

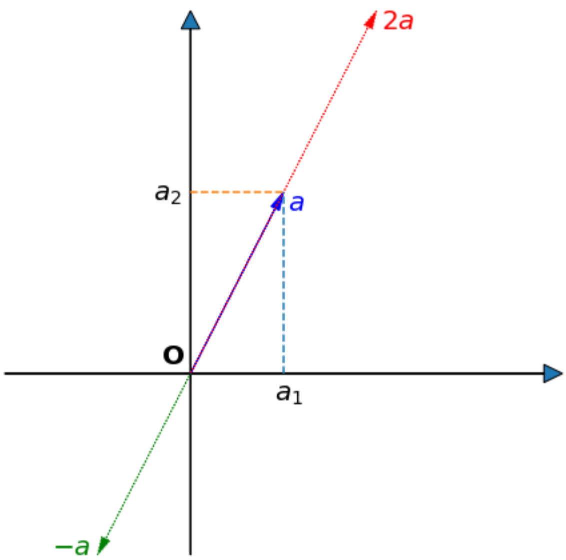


1.2.1 Scalar multiplication

Let $a = (a_1, a_2)$ be a vector of $\mathbb{R}^2$ and $t \in \mathbb{R}$. Then we define

$$ta = (ta_1, ta_2).$$

Geometrically, when we multipy a vector by a scalar t, we stretch or shrink the vector, eventually changing its direction.



1.2.2 Sum of vectors

Let $a = (a_1, a_2)$ and $b = (b_1, b_2)$ be two vectors of $\mathbb{R}^2$. The sum of a and b is defined as

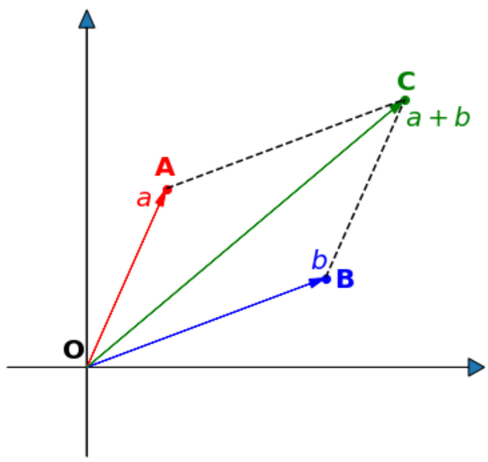
$$a + b = (a_1 + b_1, a_2 + b_2)$$

The vector $a + b$ is the diagonal of the parallelogram generated by a and b.
We can check that the quadrilateral OBCA is really a parallelogram. To do that, it suffices to verify that

$$\|\overline{OA}\| = \|\overline{BC}\| \quad \text{and} \quad \|\overline{OB}\| = \|\overline{AC}\|$$

We just check that $\|\overline{OA}\| = \|\overline{BC}\|$:

$$\begin{aligned} \|\overline{OA}\| &= \sqrt{a_1^2 + a_2^2}; \\ \|\overline{BC}\| &= \sqrt{(a_1 + b_1 - b_1)^2 + (a_2 + b_2 - b_2)^2} \\ &= \sqrt{a_1^2 + a_2^2}. \end{aligned}$$



Properties of the operations on vectors of $\mathbb{R}^2$.

Let $a = (a_1, a_2)$, $b = (b_1, b_2)$ and $c = (c_1, c_2)$ be three vectors of $\mathbb{R}^2$. Let $O = (0, 0)$ be the zero vecctor. Let $t \in \mathbb{R}$. The following properties hold:

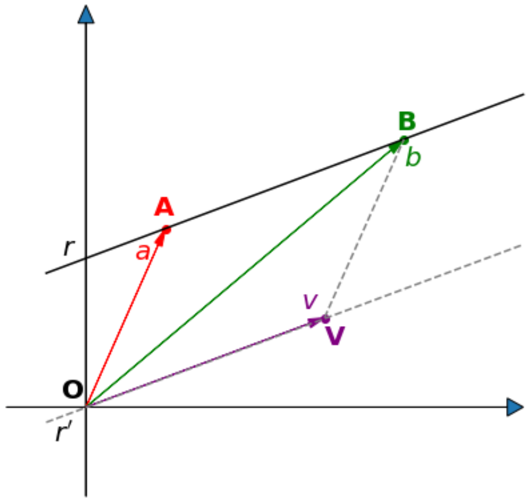
- $a + b = b + a$;
- $(a + b) + c = a + (b + c)$;
- $t(a + b) = ta + tb$;
- $a + O = a$;
- if $-a = (-a_1, -a_2)$, then $a + (-a) = O$.

1.3 Lines in the Cartesian plane (general case)

Now we consider two points $A = (x_A, y_A)$ and $B = (x_B, y_B)$ in the Cartesian plane and the corresponding vectors $a = (x_A, y_A)$ and $b = (x_B, y_B)$. We draw the line r passing through A and B. Moreover we draw the line r' passing through O and parallel to r and define the vector

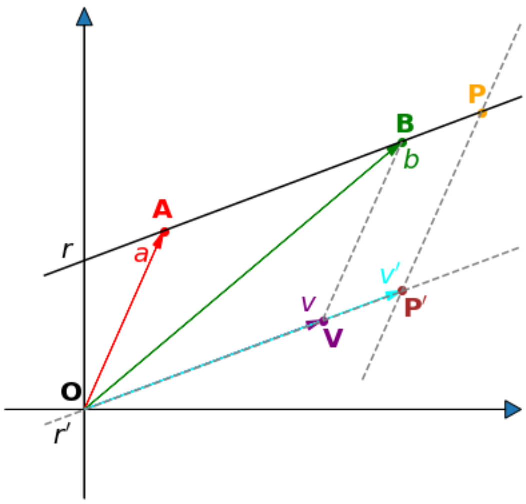
$$v = (x_B - x_A, y_B - y_A)$$

joining the point O to the point $V = (x_B - x_A, y_B - y_A)$. The quadrilateral $OABV$ is a parallelogram.



Now we consider another point P on the line r and sketch the line passing through P and parallel to the vector a. Such a line intersect the line r' in a point P'. The vector v' which joins O to P' is a multiple of the vector v. Therefore

$$v' = tv \quad \text{for some } t \in \mathbb{R}$$



The quadrilateral $OAPP'$ is a parallelogram. If p is the vector which joins the points O and P we have that

$$p = a + v'.$$

If $P = (x_P, y_P)$, then

$$(x_P, y_P) = (x_A, y_A) + tv \quad \text{with } t \in \mathbb{R}.$$

Hence we obtain the parametric equations of a line passing through two points A

and B in the plane. We define the direction vector $v = (v_1, v_2)$, where

$$v_1 = x_B - x_A \quad \text{and} \quad v_2 = y_B - y_A$$

Then the line is formed by all points (x, y) such that

$$\begin{cases} x = x_A + tv_1 \\ y = y_A + tv_2 \end{cases} \quad \text{with } t \in \mathbb{R}$$

1.4 Linear independent and dependent vectors (2 vectors in $\mathbb{R}^2$)

If $v = (v_1, v_2)$ is a non-zero vector of $\mathbb{R}^2$, then any multiple tv, with $t \in \mathbb{R}$, lies on the same line passing through the origin and the point (v_1, v_2). Indeed, we say that such vectors are *linearly dependent*. In general, two vectors v and w in $\mathbb{R}^2$ are linearly dependent if at least one of the following conditions holds:

- $v = tw$ for some $t \in \mathbb{R}$;
- $w = tv$ for some $t \in \mathbb{R}$.

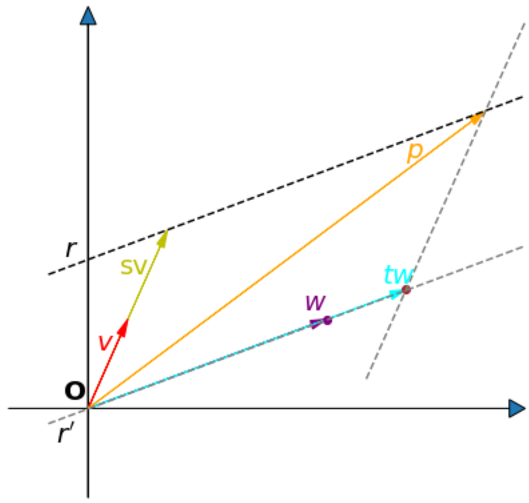
As a particular case we notice that if one of vectors v and w is the zero vector, then v and w are linearly dependent. In fact, the zero vector lies on any line passing thorugh the origin.

If two vectors v and w are not linearly dependent, then they are *linearly independent*.

If we take two linearly independent vectors v and w in $\mathbb{R}^2$, then any other vector p of $\mathbb{R}^2$ can be expressed as a *linear combination* of v and w. This latter amounts to saying that there exist two real numbers s and t such that

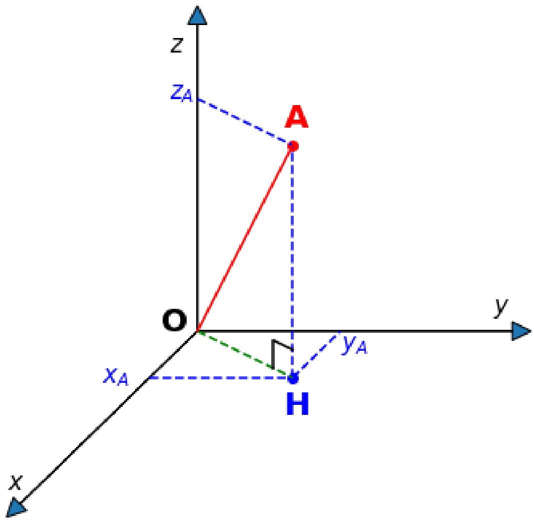
$$p = sv + tw.$$

In order to find the numbers s and t we can consider the geometrical construction aside.



1.5 Points in the Cartesian space

Now we consider three orthogonal real lines in the space, which we denote as the x, y and z axes. Any point A can be identified by three coordinates, namely $A = (x_A, y_A, z_A)$ for some x_A, y_A and z_A in $\mathbb{R}$.



We can compute the Euclidean distance of the point A from the point O. The distance is equal to the length $\|\overline{OA}\|$ of the line segment $\overline{OA}$. We can consider the right triangle $O\overset{\triangle}{H}A$. We have that

$$\|\overline{OH}\| = \sqrt{x_A^2 + y_A^2} \quad \text{and} \quad \|\overline{AH}\| = |z_A|.$$

We notice that it is safer to take the absolute value of z_A because z_A could also be negative. By Pythagoras theorem we get

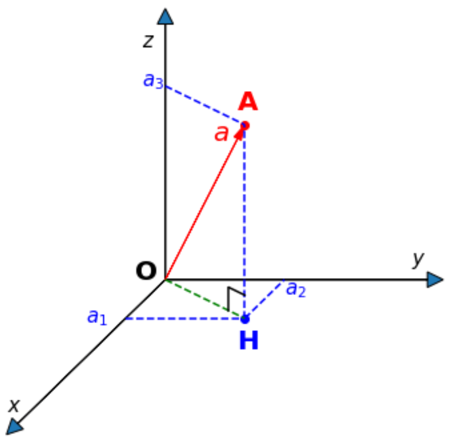
$$\|\overline{OA}\| = \sqrt{\|\overline{OH}\|^2 + \|\overline{AH}\|^2} = \sqrt{x_A^2 + y_A^2 + z_A^2}$$

More in general, if we take two points $A = (x_A, y_A, z_A)$ and $B = (x_B, y_B, z_B)$, then their Euclidean distance is

$$d(A, B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}$$

1.6 Vectors of $\mathbb{R}^3$

We can associate a point in the Cartesian space with any triple of real numbers $(a_1, a_2, a_3) \in \mathbb{R}^3$. Indeed we can also draw an arrow joining the origin O to the point $A = (a_1, a_2, a_3)$. When we refer to such an arrow we write a small letter $a = (a_1, a_2, a_3)$.



We can define the scalar multiplication and the sum of vectors in analogy with the case of vectors of $\mathbb{R}^2$. If $t \in \mathbb{R}$ and $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3) \in \mathbb{R}^3$, then

$$ta = (ta_1, ta_2, ta_3);$$
$$a + b = (a_1 + b_1, a_2 + b_2, a_3 + b_3).$$

1.7 Parametric equations of a line in the space

Let $v = (v_1, v_2, v_3)$ be a non-zero vector of $\mathbb{R}^3$. The parametric equations of a line passing through the origin O and having direction vector v are

$$(x, y, z) = tv \quad \Leftrightarrow \quad (x, y, z) = t(v_1, v_2, v_3) \quad \text{with } t \in \mathbb{R}.$$

More in general I could consider the parametric equations of a line passing through a point $P = (x_P, y_P, z_P)$ and with direction vector v:

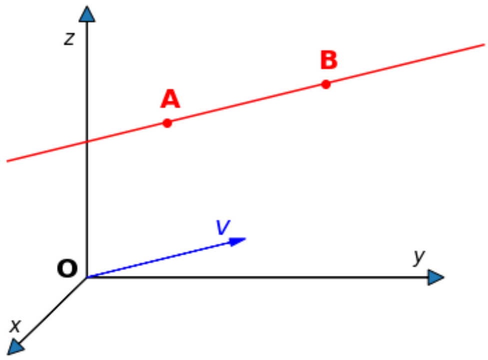
$$(x, y, z) = (x_P, y_P, z_P) + t(v_1, v_2, v_3) \quad \text{with } t \in \mathbb{R}.$$

If I am given two distinct points $A = (x_A, y_A, z_A)$ and $B = (x_B, y_B, z_B)$, then I can define a direction vector

$$v = (v_1, v_2, v_3) = (x_B - x_A, y_B - y_A, z_B - z_A).$$

The parametric equations are

$$(x, y, z) = (x_A, y_A, z_A) + t(v_1, v_2, v_3) \quad \text{with } t \in \mathbb{R}.$$



1.8 Parametric equations of a plane

Let $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$ be two non-zero linearly independent vectors, namely two vectors v and w in $\mathbb{R}^3$ such that

$$v \neq kw \text{ for all } k \in \mathbb{R} \quad \text{and} \quad w \neq kv \text{ for all } k \in \mathbb{R}.$$

We can consider all the linear combinations of the vectors v and w, namely all the vectors in the form

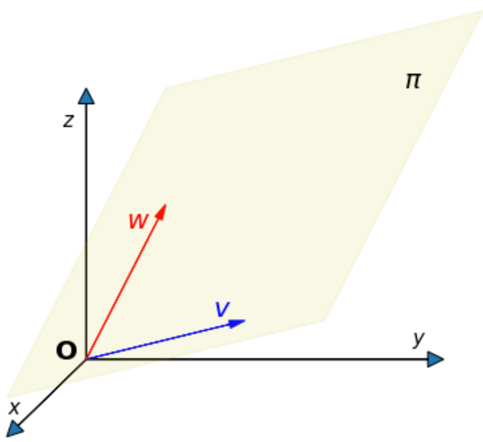
$$sv + tw \quad \text{with } s, t \in \mathbb{R}.$$

The points (x, y, z) in the space which can be expressed as linear combinations of the vectors v and w form the plane π generated by v and w and passing through the origin. Therefore, the parametric equations are

$$(x, y, z) = s(v_1, v_2, v_3) + t(w_1, w_2, w_3)$$

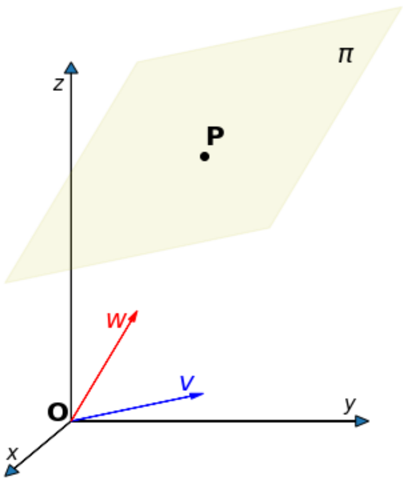
or

$$\begin{cases} x = sv_1 + tw_1 \\ y = sv_2 + tw_2 \\ z = sv_3 + tw_3 \end{cases} \qquad \text{with } s, t \in \mathbb{R}.$$



More in general, the parametric equations of the plane passing trhough a point $P = (x_P, y_P, z_P)$ and parallel to v and w are

$$\begin{cases} x = x_P + sv_1 + tw_1 \\ y = y_P + sv_2 + tw_2 \\ z = z_P + sv_3 + tw_3 \end{cases} \qquad \text{with } s, t \in \mathbb{R}.$$



1.9 Cartesian equation of a plane

As we have seen above, there are three parametric equations defining a plane with two parameters. If we eliminate the parameters s and t from two of the parametric equations, we can find a relation between x, y and z from the third parametric equation. Such a relation has the form

$$ax + by + cz = d$$

where a, b, c, d are four real numbers such that at least one of a, b, c is not zero. The last equation is a *Cartesian equation* of the plane.

1.10 Cartesian equations of a line

If we consider two planes in the space, their intersection can be empty or non-empty. In the case the intersection is not empty, then either the two planes have all the points in common or they intersect along a line. This fact follows from the general theory of linear systems, which will be discussed later.

If we take two planes π and π' of equations

$$\pi : ax + by + cz = d \quad \text{and} \quad \pi' : a'x + b'y + c'z = d'$$

such that π and π' have non-empty intersection but do not have all points in common, then

$$\begin{cases} ax + by + cz = d \\ a'x + b'y + c'z = d' \end{cases}$$

are the linear equations of a line.

Chapter 2

The set $\mathbb{R}^n$ and real matrices

2.1 The set $\mathbb{R}^n$

Till now we have considered the set $\mathbb{R}$ of real numbers, the set $\mathbb{R}^2$ of pairs of real numbers and the set $\mathbb{R}^3$ of triples of real numbers. Indeed, we can consider a positive integer n, namely an element

$$n \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}.$$

We recall that $\mathbb{N}$ denotes the set of natural numbers, namely the non-negative integers $0, 1, 2$, etc.

We can define the set

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\}$$

of all ordered sequences of n real numbers. An element $(x_1, x_2, \ldots, x_n)$ is called a n-tuple of real numbers.

We can define two operations on the n-tuples of $\mathbb{R}^n$.

- *Multiplication by scalars.* If $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $t \in \mathbb{R}$, then

$$tx = (tx_1, tx_2, \ldots, tx_n).$$

- *Sum of n-tuples.* If $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are n-tuples of $\mathbb{R}^n$, then

$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

The set $\mathbb{R}^n$ with the two operations defined above is a *vector space*. A vector space is an important algebraic structure. We will find more examples of vector spaces later. We give now the formal definition.

2.2 Vector space

A real vector space is a set V together with an operation $+$ of sum of vectors and an operation $\cdot$ of multiplication by scalars such that 8 axioms are satisfied.

- Axioms for $+$
  1. If $v, w \in V$, then $v + w = w + v$.
  2. If $u, v, w \in V$, then $(u + v) + w = u + (v + w)$.
  3. There exists a vector $O \in V$ such that $v + O = O + v$ for all $v \in V$.
  4. For any $v \in V$ there exists a vector $-v$ such that $v + (-v) = O$.
- Axioms for $\cdot$
  1. For any $v \in V$, we have that $1 \cdot v = v$.
  2. If $s, t \in \mathbb{R}$ and $v \in V$, then $(s + t) \cdot v = s \cdot v + t \cdot v$.
  3. If $s, t \in \mathbb{R}$ and $v \in V$, then $(st) \cdot v = s \cdot (t \cdot v)$.
  4. If $t \in \mathbb{R}$ and $v, w \in V$, then $t \cdot (v + w) = t \cdot v + t \cdot w$.

Remark 2.2.1: I

these notes we consider only real vector spaces. More in general, a vector space can be defined over any field $\mathbb{K}$ (for example the field $\mathbb{C}$ of complex numbers).

Remark 2.2.2: S

far the only example of vector space we know is $\mathbb{R}^n$. Indeed, one can check that the axioms of vector space are verified for $\mathbb{R}^n$.

2.3 Euclidean norm in $\mathbb{R}^n$

Let $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. The Euclidean norm of v is

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

We notice that the Euclidean norm of a vector of $\mathbb{R}^2$ or $\mathbb{R}^3$ is equal to the length of the vector in the Cartesian plane or space.

Properties of the Euclidean norm

Let $v, w \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Then

1. $\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = O$.
2. $\|tv\| = |t| \cdot \|v\|$.
3. $\|v + w\| \leq \|v\| + \|w\|$. (triangle's inequality)

The first two properties are immediate to check. As regards the last property, in general the proof is not immediate. By the way, it is easy to check the validity at least in the case of vectors of $\mathbb{R}^2$ or $\mathbb{R}^3$. We can resort to Euclidean geometry and recall that in any triangle the length of a side cannot be bigger than the sum of the lengths of the other two sides.

In the figure aside we consider two vectors v and w and their sum. Now we focus on the triangle $O\overset{\triangle}{A}C$. We have that

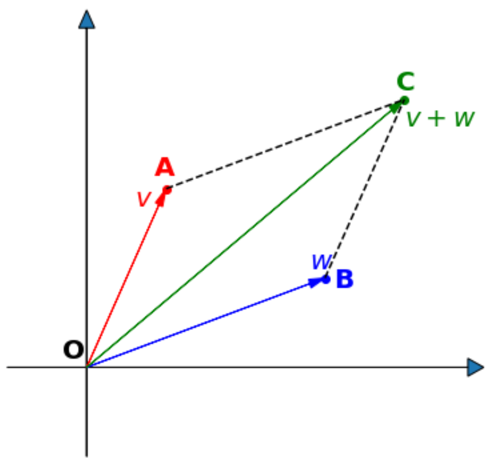
$$\|\overline{OC}\| \leq \|\overline{OA}\| + \|\overline{AC}\|.$$

Since

$$\begin{aligned} \|\overline{OC}\| &= \|v + w\| \\ \|\overline{OA}\| &= \|v\| \\ \|\overline{AC}\| &= \|w\| \end{aligned}$$

we get

$$\|v + w\| \leq \|v\| + \|w\|.$$



2.4 Dot product and orthogonality

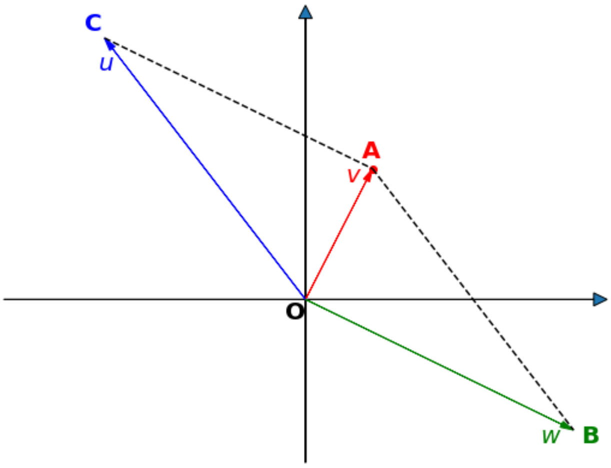
First we recall Pythagoras theorem and its reverse.

Theorem 2.4.1

Let $A\overset{\triangle}{B}C$ be a triangle. Then there is a right angle in B if and only if

$$\|\overline{AC}\|^2 = \|\overline{AB}\|^2 + \|\overline{BC}\|^2.$$

The dot product can be employed to check when two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are orthogonal.



Consider two vectors v and w in the plane ending at the points A and B respectively. Then we construct the parallelogram, which has O, A and B as three of its vertices. We denote by C the fourth vertex. Moreover we denote by u the vector joining the point O to C.
We have that v and w are orthogonal if and only if the triangle $A\overset{\triangle}{O}B$ has a right angle in O.

According to Pythagoras theorem the triangle $A\overset{\triangle}{O}B$ has a right angle in O if and only if

$$\|\overline{AB}\|^2 = \|\overline{OA}\|^2 + \|\overline{OB}\|^2$$

We notice that

$$\|\overline{OA}\| = \|v\| \quad \text{and} \quad \|\overline{OB}\| = \|w\|.$$

For the law of the parallelogram $u + w = v$. Therefore $u = v - w$. Since

$$\|\overline{OC}\| = \|\overline{AB}\| \quad \text{and} \quad \|\overline{OC}\| = \|v - w\|$$

we have that

$$\begin{aligned}
\|\overline{AB}\|^2 = \|\overline{OA}\|^2 + \|\overline{OB}\|^2 &\Leftrightarrow \\
\|v - w\|^2 = \|v\|^2 + \|w\|^2 &\Leftrightarrow \\
(v - w) \cdot (v - w) = \|v\|^2 + \|w\|^2 &\Leftrightarrow \\
v \cdot v - 2v \cdot w + w \cdot w = \|v\|^2 + \|w\|^2 &\Leftrightarrow \\
\|v\|^2 - 2v \cdot w + \|w\|^2 = \|v\|^2 + \|w\|^2 &\Leftrightarrow \\
v \cdot w = 0.
\end{aligned}$$

We have proved (for vectors in the Cartesian plane, but the proof is the same also in the space) the

Theorem 2.4.2

Two vectors v and w in $\mathbb{R}^2$ or $\mathbb{R}^3$ are orthogonal if and only if $v \cdot w = 0$.

Remark 2.4.3: W

ile in $\mathbb{R}^2$ and $\mathbb{R}^3$ the geometric notion of orthogonality can be related to the dot product, in $\mathbb{R}^n$ we say *by definition* that two vectors v and w are orthogonal if and only if $v \cdot w = 0$.

2.4.1 Equation of a plane given an orthogonal vector and a point

If $n = (n_1, n_2, n_3)$ is a non-zero vector in $\mathbb{R}^3$, then we can define the plane passing through the origin and orthogonal to n as the set of all points (x, y, z) such that the vector (x, y, z) is orthogonal to n. Therefore the equation of the plane is

$$(x, y, z) \cdot n = 0 \qquad \iff \qquad n_1 x + n_2 y + n_3 z = 0$$

More in general, the equation of a plane containing a point $A = (x_A, y_A, z_A)$ and orthogonal to n is given by all points (x, y, z) such that $(x - x_A, y - y_A, z - z_A)$ is orthogonal to n. Therefore the equation is

$$n_1(x - x_A) + n_2(y - y_A) + n_3(z - z_A) = 0$$

2.5 Matrices

A linear system of m linear equations in n unknowns can be written in the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ \ddots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Here we notice that the coefficients a_{ij} are real numbers for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$, while $x_1, \ldots, x_n$ are the unknowns.

We can associate with the linear system above a table with m rows and n columns, containing all the coefficients a_{ij}:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ & \ddots & & & \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{bmatrix}$$

The table A is called the *matrix of coefficients* of the linear system. Since A has m rows and n columns, we say that A is a $m \times n$ matrix.

We can also associate with the linear system the *column vector of constant terms*

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

and the *column vector of the unknowns*

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Definition 2.5.1: row matrix

A matrix $1 \times n$ is called row matrix or row vector, while a matrix $m \times 1$ is called column matrix or column vector.

Definition 2.5.2: square matrix

We say that a matrix $n \times n$ is a square matrix.

Definition 2.5.3

Let m and n be two positive integers. We denote by $\mathbb{R}^{m,n}$ the set of all the matrices $m \times n$ with real coefficients.

2.5.1 Sum of matrices and multiplication by scalars

Let A and B be two matrices in $\mathbb{R}^{m,n}$. Then we define the sum

$$A+B=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \ddots & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}+\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ & \ddots & & \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

$$=\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ & \ddots & & \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{bmatrix}$$

Remark 2.5.4: W

notice that the sum of two matrices is defined if and only if A and B have the same number of rows m and the same number of columns n.

Now we define the multiplication by scalars. Let A be a matrix and $k \in \mathbb{R}$. Then

$$kA=\begin{bmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ ka_{21} & ka_{22} & \dots & ka_{2n} \\ & \ddots & & \\ ka_{m1} & ka_{m2} & \dots & ka_{mn} \end{bmatrix}$$

Remark 2.5.5: I

is possible to check that the set $\mathbb{R}^{m,n}$ together with the sum of matrices and multiplication by scalars is a real vector space.

2.5.2 Product of matrices

Let A and B be two matrices. The multiplication AB is defined if and only if the number of columns of A is equal to the number of rows of B.

First we define the product of a row vector A by a column vector B. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix}$$

Then AB is a 1×1 matrix $C = \begin{bmatrix} c_{11} \end{bmatrix}$ where

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + \dots + a_{1n}b_{n1}.$$

Now we consider a $m \times n$ matrix A and a $n \times l$ matrix B. Then the product $C = AB$ is the matrix $m \times l$ such that the coefficient c_{ij}, for $1 \leq i \leq m$ and $1 \leq j \leq n$, is given by the product of the i-th row of A with the j-th column of B.

Example 2.5.6

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ -1 & 3 \end{bmatrix}.$$

Since A is 2×3 and B is 3×2 we can compute $C = AB$. The matrix C is 2×2:

$$\begin{bmatrix} 1 \cdot 2 + 2 \cdot 1 + 3 \cdot (-1) & 1 \cdot 0 + 2 \cdot 1 + 3 \cdot 3 \\ 0 \cdot 2 + 1 \cdot 1 + (-1) \cdot (-1) & 0 \cdot 0 + 1 \cdot 1 + (-1) \cdot 3 \end{bmatrix} = \begin{bmatrix} 1 & 11 \\ 2 & -2 \end{bmatrix}.$$

We can also compute the product BA and the result is the 3×3 matrix

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

Remark 2.5.7: I

the last example we noticed that both AB and BA may exist, but they can have different sizes. We also notice that when both AB and BA exist and have the same sizes, still they can be different. Consider for example the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad AB = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Some properties of sums and product of matrices

Let A, B, C be three matrices and k a real number such that all the operations below are defined. Then

1. $(A+B)C = AC + BC$.
2. $A(B+C) = AB + AC$.
3. $(AB)C = A(BC)$.
4. $k(AB) = (kA)B = A(kB)$.

2.5.3 Diagonal and identity matrices

We introduce two important families of matrices.

Definition 2.5.8: diagonal matrix

A square matrix D is diagonal if $d_{ij} = 0$ for $i \neq j$.

Example 2.5.9

The following are diagonal matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Definition 2.5.10: identity matrix

Let $n \in \mathbb{N}^*$. Then the identity matrix of order n is the diagonal matrix I_n whose elements on the main diagonal are all equal to 1.

Example 2.5.11

The following are the identity matrices of order up to 4:

$$I_1 = \begin{bmatrix} 1 \end{bmatrix}, \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Remark 2.5.12: I

A is a $m \times n$ matrix, then

$$I_m A = A \quad \text{and} \quad A I_n = A.$$

2.6 Equivalent linear systems and elementary row operations

Two linear systems are equivalent if the have the same solutions. First we deal with homogeneous linear systems, namely linear systems in the form

$$Ax = O$$

where the constant vector is equal to the zero vector. Later we will consider general linear systems.

There are some elementary operations which can be performed on the rows of the matrix of coefficients of a homogeneous linear system and which do not change the set of solutions of the system itself.

2.6.1 Row switching

Switching two rows of the matrix A amounts to switching the equations in the linear system. We can switch the rows i and j of A multiplying on the left the matrix A by a matrix which we denote by S_{ij} and which is obtained from the identity I_m switching the i-th with the j-th row.

For example, if we want to switch the rows 2 and 4 of the matrix

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

we multiply A on the left by the matrix

$$S_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

We have

$$S_{24}A = \begin{bmatrix} 1 & 2 \\ 7 & 8 \\ 5 & 6 \\ 3 & 4 \end{bmatrix}.$$

In general, if x is a solution of the linear system

$$Ax = O,$$

then we can multiply both sides of the equation by S_{ij} and we get

$$S_{ij}Ax = S_{ij}O \Leftrightarrow A'x = O$$

where $A' = S_{ij}A$. Therefore x is a solution of $A'x = O$. Viceversa, if x is a solution of $A'x = O$, then

$$S_{ij}A'x = S_{ij}O \Leftrightarrow S_{ij}S_{ij}Ax = O \Leftrightarrow Ax = O,$$

namely x is a solution of $Ax = O$.

2.6.2 Row multiplication

We can multiply all the coefficients of the row i of the matrix A by a non-zero scalar k without affecting the solutions of the system $Ax = O$. We can perform this operation multiplying on the left the matrix A by the matrix $D_i(k)$, which is a diagonal matrix obtained from the identity I_m replacing the 1 on the row i with the scalar k.

For example, if I want to multiply the second row of the matrix

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

by $k = 3$, then I multiply A on the left by the matrix

$$D_2(3) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and I get the matrix

$$D_2(3)A = \begin{bmatrix} 1 & 2 \\ 9 & 12 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}.$$

In general, if x is a solution of the linear system

$$Ax = O,$$

then we can multiply both sides of the equation by $D_i(k)$ (with $k \neq 0$) and we get

$$D_i(k)Ax = D_i(k)O \Leftrightarrow A'x = O$$

where $A' = D_i(k)A$. Therefore x is a solution of $A'x = O$. Viceversa, if x is a solution of $A'x = O$, then

$$D_i(k^{-1})A'x = D_i(k^{-1})O \Leftrightarrow D_i(k^{-1})D_i(k)Ax = O \Leftrightarrow Ax = O,$$

namely x is a solution of $Ax = O$. Indeed, one can check that

$$D_i(k^{-1})D_i(k) = I_m.$$

2.6.3 Row addition

We can add to one row of A another row or a multiple of another row without affecting the solutions of the system $Ax = O$. If we want to add to the i-th row of A the j-th row multiplied by a scalar k, then we multiply on the left the matrix A by the matrix $E_{ij}(k)$, which is a matrix obtained from the identity I_m replacing the coefficient in position ij by k.

For example, if I want to add to the third row of the matrix

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

the first row multiplied by -5, then I multiply on the left the matrix A by

$$E_{31}(-5) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -5 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and I get the matrix

$$E_{31}(-5)A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & -4 \\ 7 & 8 \end{bmatrix}.$$

In general, if x is a solution of the linear system

$$Ax = O,$$

then we can multiply both sides of the equation by $E_{ij}(k)$ and we get

$$E_{ij}(k)Ax = E_{ij}(k)O \Leftrightarrow A'x = O$$

where $A' = E_{ij}(k)A$. Therefore x is a solution of $A'x = O$. Viceversa, if x is a solution of $A'x = O$, then

$$E_{ij}(-k)A'x = E_{ij}(-k)O \Leftrightarrow E_{ij}(-k)E_{ij}(k)Ax = O \Leftrightarrow Ax = O,$$

namely x is a solution of $Ax = O$. Indeed, one can check that

$$E_{ij}(-k)E_{ij}(k) = I_m.$$

2.7 Invertible matrices

We have introduced above the elementary row operations and the corresponding elementary matrices. As we have noticed, any elementary row operations can be reverted. Indeed, for any elementary matrix P there exists another elementary matrix Q such that $PQ = QP = I$, where is an identity matrix. More precisely,

$$\begin{aligned} S_{ij}S_{ij} &= I; \\ D_i(k)D_i(k^{-1}) = D_i(k^{-1})D_i(k) &= I; \\ E_{ij}(k)E_{ij}(-k) = E_{ij}(-k)E_{ij}(k) &= I. \end{aligned}$$

Elementary matrices are examples of *invertible matrices*.

Definition 2.7.1

A square matrix A of order n is invertible if there exists a square matrix B of order n such that $AB = BA = I_n$.

Remark 2.7.2: T

e matrix B of the definition will be denoted from now on by A^{-1}.

Remark 2.7.3: W

know that elementary matrices are invertible. Later we will discuss in details some conditions for the invertibility of a square matrix. Moreover we will study some methods for the computations of the inverse.

2.7.1 Equivalent linear systems

As we noticed above, a homogeneous system $Ax = O$ is equivalent to any other homogeneous system $A'x = O$, where the matrix A' is obtained multiplying on the left the matrix A by elementary matrices.

More in general, if I consider any linear system $Ax = b$, then I get an equivalent linear system multiplying both sides of the equation by an elementary matrix P. In fact, P is invertible and

$$Ax = b \;\Rightarrow\; PAx = Pb \;\Rightarrow P^{-1}PAx = P^{-1}Pb \;\Rightarrow IAx = IPb \;\Rightarrow\; Ax = Pb.$$

Since all the information of a linear system are stored in the matrix A and in the column vector b, it is more convenient to apply the elementary row operations on the *augmented matrix* $[A|b]$, which is just equal to the matrix obtained extending A with one more column, namely the column vector b. Therefore, we can compute $P[A|b]$ and get the augmented matrix $[A'|b']$ of the linear system $A'x = b'$ equivalent to $Ax = b$.

2.8 Gaussian elimination

Applying elementary row operations on the augmented matrix $[A|b]$ we get the augmented matrix $[A'|b']$ of an equivalent linear system. Our goal is to transform $[A|b]$ into a "more convenient" form $[A'|b']$ which reveals immediately if the initial system is solvable and how many solutions it has. The sought more convenient form is the so-called "row echelon form" of $[A|b]$.

Definition 2.8.1: Row echelon form

Let M be a $m \times n$ matrix. We say that M is in row echelon form (REF) if

1. the first non-zero element (*pivot*) of any row is strictly to the right of the first non-zero element of the previous row;
2. the zero-rows are at the bottom of the matrix M.

Example 2.8.2

The matrix

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 3 & 0 & 1 \\ 4 & 0 & 0 & 1 \end{bmatrix}$$

is not in REF. In fact, the first non-zero element of the third row (4) is not to the the right of the first non-zero element of the second row (3).
For the same reason, the matrix

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 3 & 0 & 1 \\ 0 & 4 & 0 & 1 \end{bmatrix}$$

is not in REF.
The following is an example of matrix in REF (the coefficients in red are the pivots):

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

Also the matrix below is in REF (notice that there are only two pivots because the last row is a zero-row):

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

Gauss proved that any matrix can be reduced into REF just by means of elementary operations. While I omit here all the steps of Gaussian algorithm, the strategy consists in creating for any row a pivot eliminating the non-zero elements in the rows below in the same column containing a pivot.

For example, if I consider the matrix

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

then I can eliminate the initial 2 on the second row multiplying on the left the matrix M by $E_{21}(-2)$:

$$E_{21}(-2)M = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & -3 & -6 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

If I consider another matrix

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

then the only way to get rid of the initial 2 on the second row and get a REF consists in switching the first and second row:

$$S_{12}M = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We notice that in general it is possible to get different REF for a same matrix using a different sequence of elementary row operations. The following result holds (we omit the proof).

Theorem 2.8.3

If M' and M'' are two REF matrices obtained from a same matrix M through elementary row operations, then M' and M'' have the same number of pivots and the pivots are in the same positions.

The statement of Theorem 2.8.3 justifies the following important definition.

Definition 2.8.4: rank

If M is a $m \times n$ matrix, then the *rank* of M, which we denote by $\operatorname{rank}(M)$, is the number of pivots of any its REF.

Remark 2.8.5: S

nce any row and any column of a REF matrix can contain at most one pivot, we have that $\operatorname{rank}(M) \leq m$ and $\operatorname{rank}(M) \leq n$. Therefore

$$0 \leq \operatorname{rank}(M) \leq \min\{m, n\}.$$

Definition 2.8.6: pivot columns

If M is a $m \times n$ matrix in REF, then the pivot columns of M are the columns of M containing one pivot.

We introduce now a special row echelon form.

Definition 2.8.7: reduced row echelon form (RREF)

Let M be a $m \times n$ matrix. We say that M is in reduced row echelon form (RREF) is

1. M is in REF;
2. all the pivots of M are equal to 1;
3. the only non-zero element of any pivot-column is the pivot itself.

Example 2.8.8

The following matrix is in REF but not in RREF:

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

In fact, the pivots are all equal to 1 but the pivot columns contain some non-zero coefficients apart from the pivot.
The following matrix is is in RREF:

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

Computing a RREF of a matrix requires to apply a standard Gaussian elimination in order to get first a REF. Then, applying a backward Gaussian elimination, we can get rid of the non-zero element in any pivot-column.

While the REF of a matrix is not unique, the RREF is unique, as claimed in the following theorem which we do not prove.

Theorem 2.8.9

The RREF of a matrix M is unique.

2.9 Solving linear systems

Gaussian elimination, REF and RREF are useful for determining the existence and the number of solutions of linear systems.

Let $Ax = b$ be a linear system, where A is a matrix $m \times n$. We can reduce the augmented matrix $[A|b]$ to a REF $[A'|b']$. Then we can consider different cases.

1. $\text{rank}([A|b]) \neq \text{rank}(A)$. In such a case the column b' contains one pivot. Since the pivot of a row is the first non-zero element of the same row, this amounts to saying that in the linear system $A'x = b'$ there is an equation in the form
$$0 = c \quad \text{for some } c \in \mathbb{R}^*.$$
Hence $Ax = b$ has no solution.

2. $\text{rank}([A|b]) = \text{rank}(A) = n$. Then $[A'|b']$ is a $m \times n$ matrix, whose first n rows are
$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & \dots & 0 & b'_1 \\ 0 & 1 & 0 & \dots & 0 & b'_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & b'_n \end{array}\right]$$
Moreover, in case $m > n$, the remaining rows are all zero-rows.
Therefore the solution is unique and is equal to
$$\begin{cases} x_1 = b'_1 \\ x_2 = b'_2 \\ \vdots \\ x_n = b'_n \end{cases}$$

3. $\text{rank}([A|b]) = \text{rank}(A) < n$. In this case b' does not contain a pivot and we can express any variable corresponding to a pivot column in terms of variables corresponding to non-pivot columns. For this reason, these latter variables are called *free variables*.

The 3 cases above can be summarized in the following result.

Theorem 2.9.1: Rouché-Capelli theorem

Let $Ax = b$ be a linear system, where A is a matrix $m \times n$. Then $Ax = b$ has solutions if and only if $\text{rank}(A) = \text{rank}([A|b])$. Moreover, in case there are solutions, the following hold.

- If $\text{rank}(A) = n$ then there exists a unique solution.
- If $\text{rank}(A) < n$, then there are infinitely many solutions which depend on $n - \text{rank}(A)$ free variables (also called parameters).

Chapter 3

Vector spaces and subspaces

In Section 2.2 we defined the notion of real vector space. In this chapter we introduce some new notions related to vector spaces and their vectors.

Definition 3.0.1: vector subspace

Let V be a vector space. Let W be a subset of V. We say that W is a vector subspace of V is the following conditions are satisfied.

1. $O \in W$.
2. If $u, v \in W$, then $u + v \in W$.
3. If $u \in W$ and $t \in \mathbb{R}$, then $tu \in W$.

Remark 3.0.2

A vector subspace W of a vector space V is a vector space in its own right. Indeed, all 8 axioms of vector space hold for W.

There are some important examples of vector subspaces.

Theorem 3.0.3

Let U be the set of all the solutions of a homogeneous linear system $Ax = O$, where A is a matrix $m \times n$. Then U is a vector subspace of $\mathbb{R}^n$

Proof. See the lecture slides. □

Remark 3.0.4

If we consider the set of solutions U of a non-homogeneous linear system $Ax = b$, then U is not a vector subspace of $\mathbb{R}^n$. In fact, $O \notin U$ because

$$AO = O \neq b.$$

Now we give two definitions which we use immediately afterwards to introduce some important examples of vector subspaces.

Definition 3.0.5: linear combination

Let V be a vector space. Let $k \in \mathbb{N}^*$.
If $v_1, \ldots, v_k$ are vectors of V and $c_1, \ldots, c_k$ are real numbers, then we say that

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

is a linear combination of the vectors $v_1, \ldots, v_k$ with coefficients $c_1, \ldots, c_k$.

Definition 3.0.6: set generated by some vectors

Let V be a vector space.
If $v_1, \ldots, v_k$ are vectors of V, then we say that the set

$$S = \{c_1 v_1 + c_2 v_2 + \cdots + c_k v_k : c_1, \ldots, c_k \in \mathbb{R}\}$$

is the set generated by the k vectors $v_1, \ldots v_k$. Such a set is usually denoted by one of the following notations

$$\operatorname{span}\{v_1, \ldots, v_k\} \quad \text{or} \quad \langle v_1, \ldots, v_k \rangle$$

The set generated by some vectors of a vector space V is indeed a vector subspace of V.

Theorem 3.0.7

Let V be a vector space and $v_1, \ldots, v_k$ vectors of V. Then $\operatorname{span}\{v_1, \ldots, v_k\}$ is a vector subspace of V.

Proof. See the lecture slides. $\square$

Example 3.0.8

We consider some vector subspaces in $\mathbb{R}^2$ and $\mathbb{R}^3$.

- If v is a non-zero vector in $\mathbb{R}^2$ or $\mathbb{R}^3$, then $\langle v\rangle$ is a line passing through the origin. Such a line is a vector subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$.

- If v_1 and v_2 are two linearly independent vectors in $\mathbb{R}^3$, then $\langle v_1, v_2\rangle$ is a plane containing the origin. Such a plane is a vector subspace of $\mathbb{R}^3$.