Some notes on Linear Algebra

Simone Ugolini

Last update: 16:02, Wednesday 8th December, 2021

Contents

1 Some geometry of the plane and the space **7**
1.1 Introduction . . . 7
1.1.1 Cartesian plane . . . 7
1.1.2 Lines in the Cartesian plane passing through O=(0,0) . . . 8
1.2 Points and vectors of $\mathbb{R}^2$ . . . 10
1.2.1 Scalar multiplication . . . 10
1.2.2 Sum of vectors . . . 11
1.3 Lines in the Cartesian plane (general case) . . . 12
1.4 Linear independent and dependent vectors (2 vectors in $\mathbb{R}^2$) . . . 13
1.5 Points in the Cartesian space . . . 14
1.6 Vectors of $\mathbb{R}^3$ . . . 14
1.7 Parametric equations of a line in the space . . . 15
1.8 Parametric equations of a plane . . . 16
1.9 Cartesian equation of a plane . . . 17
1.10 Cartesian equations of a line . . . 17

2 The set $\mathbb{R}^n$ and real matrices **19**
2.1 The set $\mathbb{R}^n$ . . . 19
2.2 Vector space . . . 20
2.3 Euclidean norm in $\mathbb{R}^n$ . . . 20
2.4 Dot product and orthogonality . . . 21
2.4.1 Equation of a plane given an orthogonal vector and a point . . . 23
2.5 Matrices . . . 23
2.5.1 Sum of matrices and multiplication by scalars . . . 25
2.5.2 Product of matrices . . . 25
2.5.3 Diagonal and identity matrices . . . 27
2.5.4 Transpose of a matrix . . . 28
2.6 Equivalent linear systems and elementary row operations . . . 28
2.6.1 Row switching . . . 28
2.6.2 Row multiplication . . . 29
2.6.3 Row addition . . . 30

2.7 Invertible matrices 31
2.7.1 Equivalent linear systems 32
2.8 Gaussian elimination 33
2.9 Solving linear systems 36

3 Vector spaces and subspaces 39
3.1 Vector subspaces 47
3.1.1 Column space of a matrix 48
3.1.2 Isomorphic vector spaces 49
3.1.3 The vector space of polynomials 50

4 Determinants 53
4.1 Determinant 54
4.1.1 Properties with respect to elementary row operations 54
4.1.2 Determinant of square matrices of order 1 or 2 55
4.1.3 Laplace expansions 56
4.1.4 Sarrus rule for square matrices of order 3 57
4.2 Invertible matrices 57
4.2.1 How to compute the inverse of a matrix 58

5 Linear functions, eigenvalues and eigenvectors 59
5.1 Introduction 59
5.2 Matrix associated with a linear function 60
5.3 Change of basis 61
5.4 Eigenvalues and eigenvectors 61
5.4.1 Geometrical interpretation in the plane and in the space 63
5.5 Computing eigenvalues and eigenvectors 63
5.6 Diagonalizable matrices 66
5.6.1 Algebraic and geometric multiplicity of an eigenvalue 67
5.6.2 Power of a matrix 69
5.7 Orthogonal vectors and orthogonal bases 69
5.7.1 Orthogonal projection of a vector along another vector 70
5.7.2 Orthogonal and orthonormal bases 70
5.7.3 Gram-Schmidt process 72
5.7.4 Orthogonal matrices 72

6 Appendix 75
6.1 Orthogonal projection of a point onto a plane 75
6.2 Proof of Theorem 3.0.3 76
6.3 Proof of Theorem 3.0.7 77
6.4 Proof or Theorem 3.0.11 78

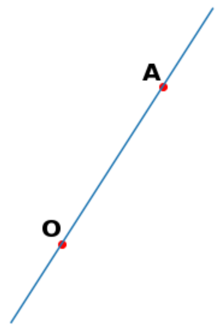
6.5 Proof of Theorem 3.0.14 78
6.6 Proof of Theorem 3.0.15 78
6.7 Something more on linear functions 79
6.7.1 Kernel and image of a linear function 79
6.7.2 Injective and surjective linear functions. Isomorphisms. 79
6.7.3 The rank-nullity theorem and its consequences 79
6.8 Principal component analysis (PCA) 80
6.8.1 Affine frames and coordinates 80
6.8.2 PCA 81

Chapter 1

Some geometry of the plane and the space

1.1 Introduction

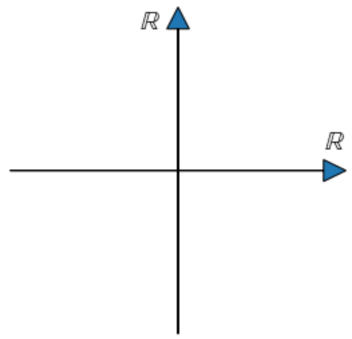
We consider two points O and A in the plane. We know that there exists a unique line passing through O and A. We could embed the points O and A in a Cartesian plane centered at O and find some equation of the line passing through O and A. Before doing so, we review the Cartesian plane.



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.

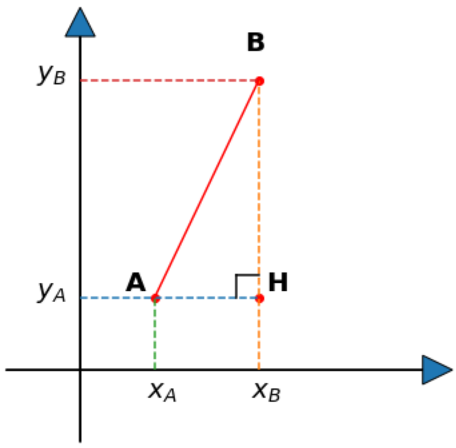


If we take two points $A = (x_A, y_A)$ and $B = (x_B, y_B)$ in the plane, then we can compute the distance between A and B using Pythagoras theorem. As illustrated in the figure aside, we construct a right triangle. Then, for the length $\|\overline{AB}\|$ of the line segment $\overline{AB}$ we have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

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

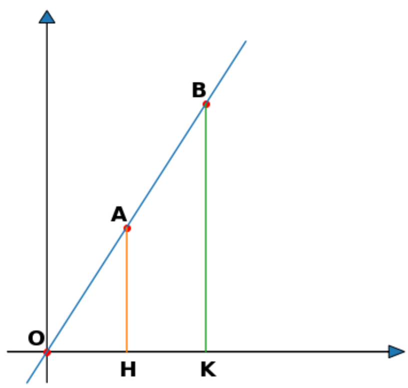
Therefore

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



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

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

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

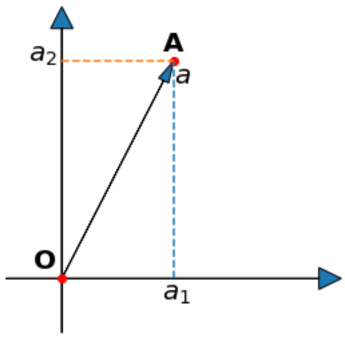
$$\begin{cases} x = t \cdot x_A \\ y = t \cdot y_A \end{cases} \qquad \text{with } t \in \mathbb{R}$$

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.

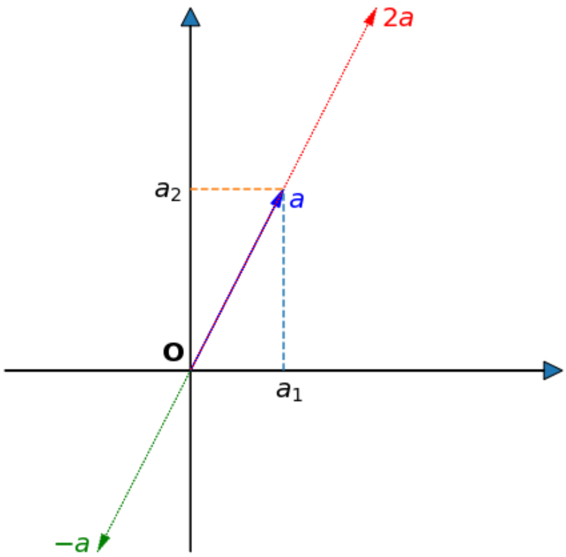


1.2.1 Scalar multiplication

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

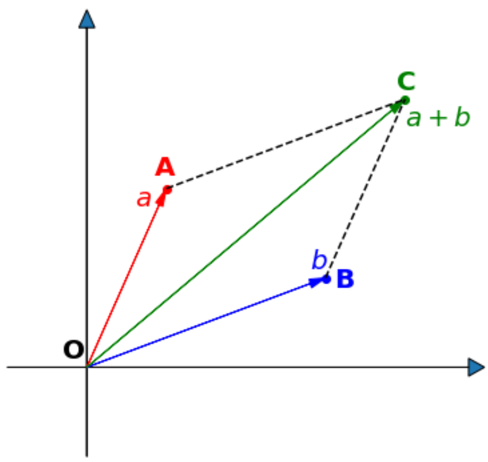
$$a + b = (a_1 + b_1, a_2 + b_2)$$

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:

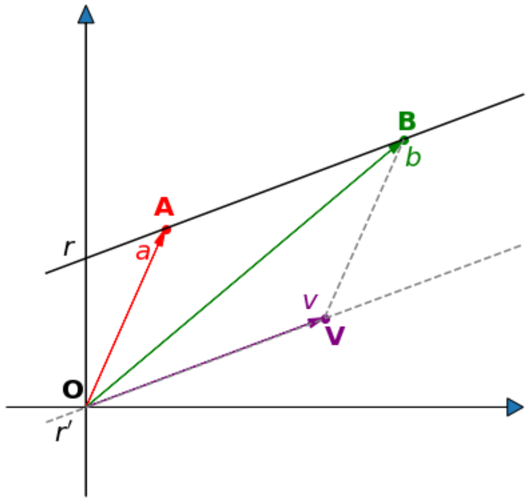
- $a + b = b + a$;
- $(a + b) + c = a + (b + c)$;
- $t(a + b) = ta + tb$;
- $a + O = a$;
- if $-a = (-a_1, -a_2)$, then $a + (-a) = O$.

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

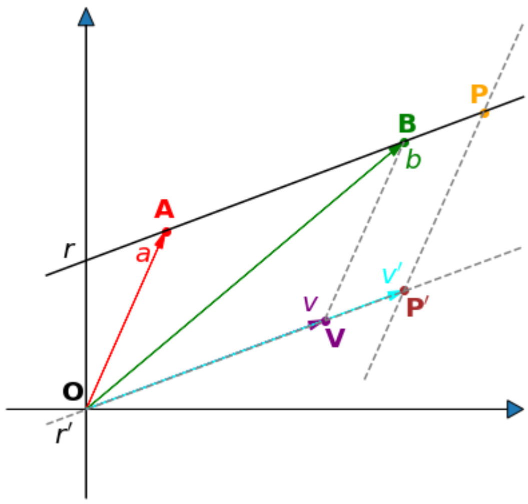
$$v = (x_B - x_A, y_B - y_A)$$

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

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

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 $t \in \mathbb{R}$.

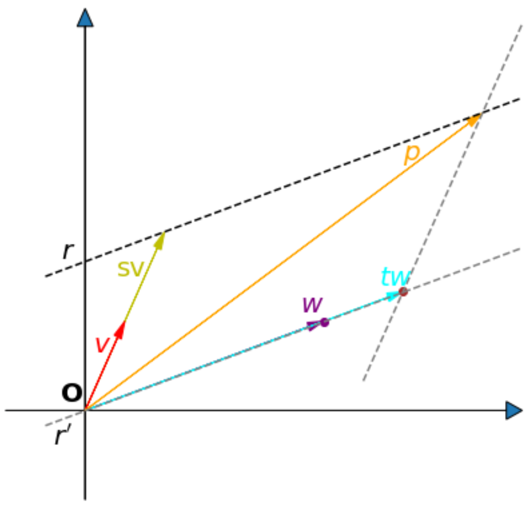
As a particular case we notice that if one of vectors v and w is the zero vector, then v and w are linearly dependent. In fact, the zero vector lies on any line passing thorugh the origin.

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

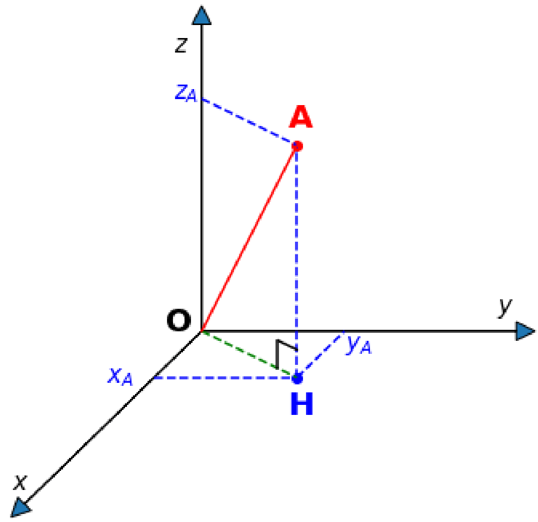
$$p = sv + tw.$$

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 $\overset{\triangle}{OHA}$. 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

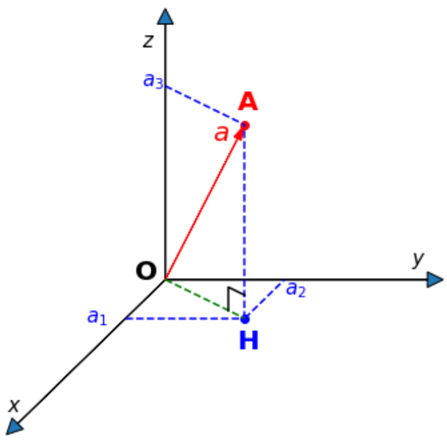
$$\|\overline{OA}\| = \sqrt{\|\overline{OH}\|^2 + \|\overline{AH}\|^2} = \sqrt{x_A^2 + y_A^2 + z_A^2}$$

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

$$\begin{aligned} ta &= (ta_1, ta_2, ta_3); \\ a + b &= (a_1 + b_1, a_2 + b_2, a_3 + b_3). \end{aligned}$$

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:

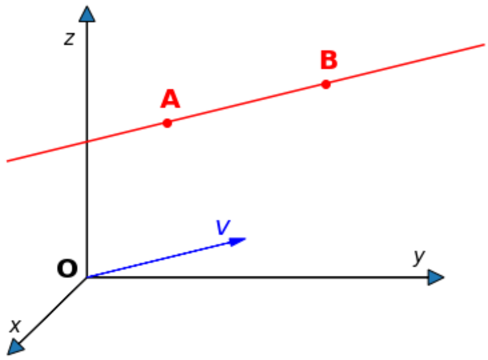
$$(x, y, z) = (x_P, y_P, z_P) + t(v_1, v_2, v_3) \quad \text{with } t \in \mathbb{R}.$$

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

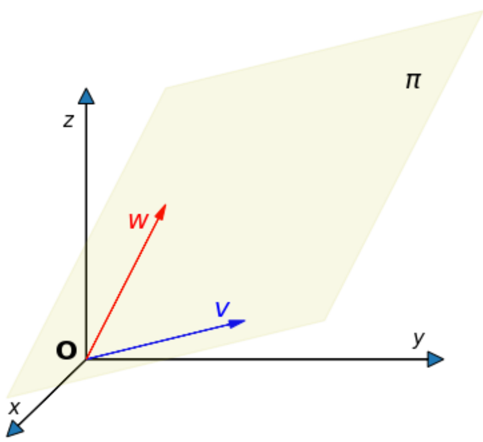
$$sv + tw \quad \text{with } s, t \in \mathbb{R}.$$

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

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

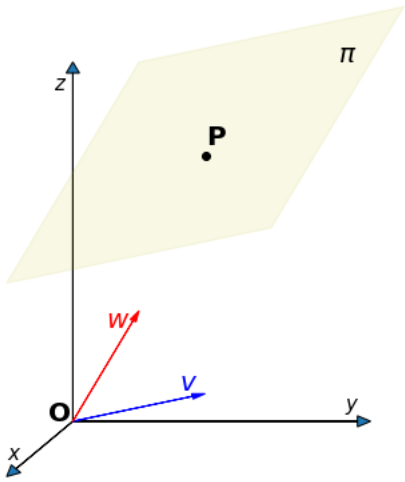
or

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



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

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

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

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

of all ordered sequences of n real numbers. An element $(x_1, x_2, \dots, x_n)$ 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 $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $t \in \mathbb{R}$, then

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

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

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

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

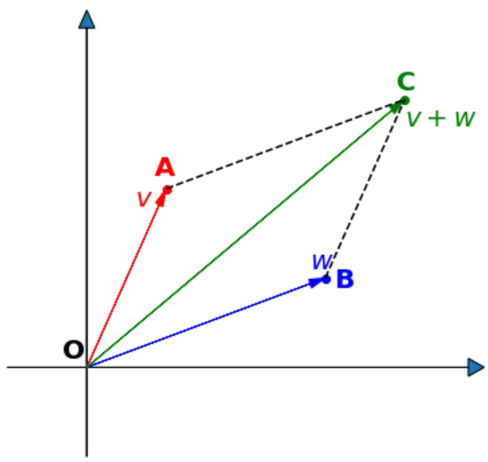
$$\|\overline{OC}\| \leq \|\overline{OA}\| + \|\overline{AC}\|.$$

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.

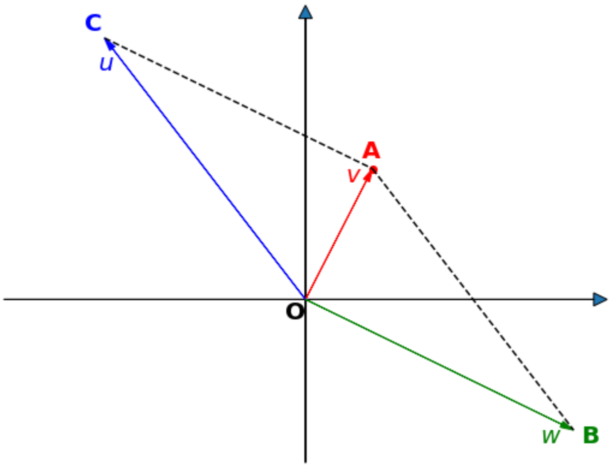
Theorem 2.4.1

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 $\overset{\triangle}{AOB}$ has a right angle in O.



According to Pythagoras theorem the triangle $\overset{\triangle}{AOB}$ 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

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

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 $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$, while $x_1, \dots, x_n$ 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}:

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

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

We 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

It 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

In 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

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

2.5.4 Transpose of a matrix

If A is a $m \times n$ matrix, then the transpose A^T of A is the matrix obtained from A switching the rows with the columns. Hence A^T is a $n \times m$ matrix.

Example 2.5.13

Let

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

Then

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

Here are some useful properties.

1. If A is a matrix, then $(A^T)^T = A$.

2. If A is a $m \times n$ matrix and B is a $n \times l$ matrix, then

$$(AB)^T = B^T A^T.$$

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,

$$S_{ij}S_{ij} = I;$$
$$D_i(k)D_i(k^{-1}) = D_i(k^{-1})D_i(k) = I;$$
$$E_{ij}(k)E_{ij}(-k) = E_{ij}(-k)E_{ij}(k) = I.$$

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

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

Remark 2.7.3

We 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

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

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

Since 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) if

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 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. $\operatorname{rank}([A|b]) \neq \operatorname{rank}(A)$. 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. $\operatorname{rank}([A|b]) = \operatorname{rank}(A) = n$. 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. $\operatorname{rank}([A|b]) = \operatorname{rank}(A) < n$. 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 $\operatorname{rank}(A) = \operatorname{rank}([A|b])$. Moreover, in case there are solutions, the following hold.

- If $\operatorname{rank}(A) = n$ then there exists a unique solution.
- If $\operatorname{rank}(A) < n$, then there are infinitely many solutions which depend on $n - \operatorname{rank}(A)$ 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 Section 6.2. □

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 $v_1, \dots, v_k$ are vectors of V and $c_1, \dots, c_k$ are real numbers, then we say that

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k$$

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

Definition 3.0.6: set generated by some vectors

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

$$S = \{c_1v_1 + c_2v_2 + \cdots + c_kv_k : c_1, \dots, c_k \in \mathbb{R}\}$$

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

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

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 $v_1, \dots, v_k$ vectors of V. Then $\text{span}\{v_1, \dots, v_k\}$ is a vector subspace of V.

Proof. See Section 6.3. □

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 $\mathbb{R}^3$.

Definition 3.0.9: linearly independent vectors

Let V be a vector space. Let $v_1, \dots, v_k$ be vectors of V.
We say that the vectors $v_1, \dots, v_k$ are linearly independent if the only solution of the equation

$$c_1 v_1 + \dots + c_k v_k = O \tag{3.1}$$

in the unknowns $c_1, \dots, c_k$ is

$$c_1 = \dots = c_k = 0.$$

Remark 3.0.10

If the equation (3.1) has some solutions other than $c_1 = \dots c_k = 0$, then we say that $v_1, \dots, v_k$ are linearly dependent. The following theorem expresses two equivalent conditions for linearly dependent vectors.

Theorem 3.0.11

Let $v_1, \dots, v_k$ be k vectors of a vector space V. Then the following conditions are equivalent.

1. The equation $c_1 v_1 + \dots + c_k v_k = O$ in the unknowns $c_1, \dots, c_k$ has at least one solution with at least one of the coefficients different from 0.
2. At least one of the vectors $v_1, \dots, v_k$ can be expressed as a linear combination of the other $k - 1$ vectors.

Proof. See Section 6.4. $\square$

Now we give another important definition.

Definition 3.0.12: set of generators

Let V be a vector space. Let $v_1, \dots, v_k$ be vectors of V.
We say that the set $S = \{v_1, \dots, v_k\}$ is a set of generators for V if any vector $v \in V$ can be expressed as a linear combination of $v_1, \dots, v_k$ or, equivalently, if $V = \langle v_1, \dots, v_k \rangle$ (we can also write $V = \text{span}(S)$).

Example 3.0.13

Let $V = \mathbb{R}^2$. Consider the following sets:

$$S_1 = \{(1,3)\};$$
$$S_2 = \{(1,0),(0,1)\};$$
$$S_3 = \{(1,0),(1,1),(0,1)\}.$$

- The set S_1 is not a set of generators of V. For example, $(1,0) \notin \text{span}(S_1)$.
- The set S_2 is a set of generators of V. In fact, let $(a,b) \in \mathbb{R}^2$. Then

$$(a,b) = a(1,0) + b(0,1).$$

- The set S_3 is also a set of generators, since any $(a,b) \in \mathbb{R}^2$ can be written as

$$(a,b) = a(1,0) + 0(1,1) + 1(0,1).$$

  We notice in passing that there are other possible linear combinations of the vectors of S_3 which yields the vector (a,b). For example

$$(a,b) = 0(1,0) + a(1,1) + (b-a)(0,1).$$

Now we state two useful theorems on linearly independent vectors and set of generators.

Theorem 3.0.14

Let $S = \{v_1, \dots, v_k\}$ be a set containing k vectors of a vector space V. If one of the vectors of S is equal to the zero vector, then the vectors in S are linearly dependent.

Proof. See Section 6.5. □

Theorem 3.0.15

Let $S = \{v_1, \ldots, v_k\}$ be a set of generators of a vector space V, namely $V = \text{span}(S)$. Suppose that, for some $j \in \{1, \ldots, k\}$, the vector v_j can be expressed as a linear combination of the remaining $k-1$ vectors of S. Then

$$V = \text{span}(S \backslash \{v_j\}) = \text{span}\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k\}.$$

Proof. See Section 6.6. □

Example 3.0.16

Let $S = \{(1,0), (1,1), (0,1)\} \subseteq \mathbb{R}^2$. We know that S is a set of generators of $\mathbb{R}^2$. Moreover

$$(1,1) = (1,0) + (0,1).$$

Hence $\mathbb{R}^2 = \text{span}(S \backslash \{(1,1)\}) = \text{span}\{(1,0), (0,1)\}$.

Definition 3.0.17: basis

Let V be a vector space. Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a set of n vectors of V. We say that $\mathcal{B}$ is a basis for V if

1. the vectors $b_1, \ldots, b_n$ are linearly independent;
2. $\mathcal{B}$ is a set of generators for V, namely $V = \text{span}(\mathcal{B})$.

Example 3.0.18

Let $V = \mathbb{R}^2$. Consider the following sets:

$$\begin{aligned} S_1 &= \{(1,3)\}; \\ S_2 &= \{(1,0), (0,1)\}; \\ S_3 &= \{(1,0), (1,1), (0,1)\}. \end{aligned}$$

- The set S_1 is not a basis for $\mathbb{R}^2$ because S_1 is not a set of generator s for $\mathbb{R}^2$.
- The set S_2 is a basis for $\mathbb{R}^2$ because the vectors in S_2 are linearly independent (check it!) and are a set of generators for $\mathbb{R}^2$.
- The set S_3 is not a basis for $\mathbb{R}^2$ because the vectors in S_3 are not linearly independent. For example, $(1,1) = (1,0) + (0,1)$.

A basis for a vector space can be characterized in an alternative way, as expressed in the following theorem.

Theorem 3.0.19

Let V be a vector space. Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a set of n vectors in V. Then the following conditions are equivalent.

- $\mathcal{B}$ is a basis of V.
- Every vector in V can be expressed in a unique way as a linear combination of vectors of $\mathcal{B}$.

Proof. See the lecture slides. □

Now we state two very important theorems.

Theorem 3.0.20

Let V be a vector space. Then there exists a basis $\mathcal{B}$ for V.

Theorem 3.0.21

Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a basis for a vector space V. If $\mathcal{B}'$ is another basis of V, then $\mathcal{B}'$ contains n vectors.

Definition 3.0.22: dimension

The number of vectors in any basis of a vector space V is called the dimension of V.

Example 3.0.23

1. $\mathcal{B} = \{(1,0), (0,1)\}$ is a basis of $\mathbb{R}^2$. Therefore $\dim(\mathbb{R}^2) = 2$.
2. Consider in $\mathbb{R}^n$ the vectors
$$\begin{aligned} e_1 &= (1, 0, 0 \ldots, 0); \\ e_2 &= (0, 1, 0, \ldots, 0); \\ &\vdots \\ e_n &= (0, 0, 0, \ldots, 1). \end{aligned}$$
Then $\mathcal{E} = \{e_1, \ldots, e_n\}$ is a basis of $\mathbb{R}^n$.
3. Consider in $\mathbb{R}^{m,n}$ the matrices B_{ij}, for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots n\}$, whose coefficients are all equal to 0 except for the coefficient in row i and

column j which is equal to 1. For example

$$B_{11} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \quad B_{12} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Then the set

$$\mathcal{B} = \{B_{ij} : 1 \leq i \leq m,\ 1 \leq j \leq n\}$$

is a basis of $\mathbb{R}^{m,n}$. Therefore $\dim(\mathbb{R}^{m,n}) = mn$.

Sometimes it is useful to extend a set of linearly independent vectors to a basis. First we state a preliminary result.

Theorem 3.0.24

Let $S = \{v_1, \dots, v_k\}$ be a set of k linearly independent vectors in a vector space V. Let $w \in V$. If $w \notin \operatorname{span}(S)$, then the set

$$T = S \cup \{w\}$$

is a set of $k + 1$ linearly independent vectors.

Theorem 3.0.25: extending linearly independent vectors to a basis

Let V be a vector space with $\dim(V) = n$. Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of V. If $S = \{v_1, \dots, v_k\}$ is a set of k linearly independent vectors in V, then either S is a basis of V or we can extend S to a basis of V adding to S some vectors of $\mathcal{B}$.

When we extend a set of linearly independent vectors to a basis, or when we simply want to check if some vectors are linearly independent, the following result can be useful.

Theorem 3.0.26

Let A be a $m \times n$ matrix. Then the columns of A are linearly independent if and only if $\operatorname{rank}(A) = n$.

Example 3.0.27

Let $S = \{(1, 0, 2), (0, 1, 0), (2, 3, 0)\} \subseteq \mathbb{R}^3$.
We can check whether the vectors ins S are linearly independent constructing the

auxiliary matrix

$$M = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 2 & 0 & 0 \end{bmatrix} \xrightarrow{E_{31}(-2)} \begin{bmatrix} \mathbf{1} & 0 & 2 \\ 0 & \mathbf{1} & 3 \\ 0 & 0 & -\mathbf{4} \end{bmatrix} = M'.$$

We notice that M' is in REF and has three pivots. Therefore $\text{rank}(M) = 3$ and the vectors in S are linearly independent.

Now we state a theorem, which can be used to check if a set of vectors is a basis.

Theorem 3.0.28

Let V be a vector space with $\dim(V) = n$. If $S = \{v_1, \ldots, v_n\}$ is a set of linearly independent vectors of V, then S is a basis of V.

Example 3.0.29

Let $S = \{(1,0,1),(1,1,0),(0,0,1)\} \subseteq \mathbb{R}^3$. Since S contains 3 vectors and $\dim(\mathbb{R}^3) = 3$, according to the last theorem S is a basis of $\mathbb{R}^3$ if and only if the vectors in S are linearly independent.
We construct the auxiliary matrix

$$M = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \xrightarrow{E_{31}(-1)} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{E_{32}(1)} \begin{bmatrix} \mathbf{1} & 1 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} \end{bmatrix} = M'.$$

Since $\text{rank}(M) = 3$, the vectors in S are linearly independent.

We know that in general a set of generators might not be a basis. By the way, such a set contains a basis.

Theorem 3.0.30: extraction of a basis

Let V be a vector space with $\dim(V) = n$. Let $S = \{v_1, \ldots, v_k\}$ be a set of generators for V. Then either S is a basis of V or we can extract from S a subset $T \subseteq S$ such that T is a basis of V.

Proof. See lecture slides. □

Example 3.0.31

Consider in $\mathbb{R}^3$ the vector subspace $V = \text{span}(S)$, where

$$S = \{(1,0,2),(2,0,4),(0,1,0),(1,3,2)\}.$$

Since S is a set of generators for V, we can extract from S a basis of V. We notice that $(2,0,4) = 2(1,0,2)$. Therefore we can safely remove $(2,0,4)$ from S.
The vector $(0,1,0)$ is not a multiple of $(1,0,2)$ (easy check).
As regards the last vector of S, we check that it cannot be expressed as a linear combination of $(1,0,2)$ and $(0,1,0)$, constructing the augmented matrix

$$[A|b] = \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 3 \\ 2 & 0 & 2 \end{array}\right] \xrightarrow{E_{31}(-2)} \left[\begin{array}{cc|c} \mathbf{1} & 0 & 1 \\ 0 & \mathbf{1} & 3 \\ 0 & 0 & 0 \end{array}\right]$$

Since $\text{rank}(A) = \text{rank}([A|b]) = 2$, we have that $(1,3,2) \in \text{span}\{(1,0,2),(0,1,0)\}$ and we can remove $(1,3,2)$ from S.
Hence a basis of V is

$$T = \{(1,0,2),(0,1,0)\}.$$

3.1 Vector subspaces

Let V be vector space with $\dim(V) = n$. What can be the dimension of a vector subspace $W \subseteq V$?

Theorem 3.1.1

Let V be a vector space with $\dim(V) = n$. Let $W \subseteq V$ be a vector subspace of V. Then $\dim(W) \leq n$.

Proof. See lecture slides. □

Remark 3.1.2

If W is a subspace of a vector space V with $\dim(V) = n$, then

$$0 \leq \dim(W) \leq n.$$

In particular, we notice that

- if $\dim(W) = 0$, then $W = \{O\}$, namely W contains only the zero vector;

- if $\dim(W) = n$, then $W = V$.

Subspaces of $\mathbb{R}^2$

Since $\dim(\mathbb{R}^2) = 2$, if $W \subseteq \mathbb{R}^2$ then $0 \leq \dim(W) \leq 2$.

- If $\dim(W) = 0$, then $W = \{(0, 0)\}$.
- If $\dim(W) = 1$, then $W = \text{span}\{w\}$ for some non zero vector $w \in \mathbb{R}^2$. Geometrically W is a line passing through the origin in the Cartesian plane.
- If $\dim(W) = 2$, then $W = \mathbb{R}^2$

Subspaces of $\mathbb{R}^3$

Since $\dim(\mathbb{R}^3) = 3$, if $W \subseteq \mathbb{R}^3$ then $0 \leq \dim(W) \leq 3$.

- If $\dim(W) = 0$, then $W = \{(0, 0, 0)\}$.
- If $\dim(W) = 1$, then $W = \text{span}\{w\}$ for some non zero vector $w \in \mathbb{R}^3$. Geometrically W is a line passing through the origin in the Cartesian space.
- If $\dim(W) = 2$, then $W = \text{span}\{w_1, w_2\}$ where w_1 and w_2 are two linearly independent vectors. Geometrically W is a plane containing the origin in the Cartesian space.
- If $\dim(W) = 3$, then $W = \mathbb{R}^3$.

3.1.1 Column space of a matrix

Let A be a $m \times n$ matrix. We denote by $v_1, \ldots, v_n$ the columns of A:

$$A = [v_1|v_2|\ldots|v_n].$$

The column space $\mathcal{C}(A)$ of A is the $\text{span}\{v_1, \ldots, v_n\}$. We notice that the columns of A are vectors of $\mathbb{R}^m$. Hence $\mathcal{C}(A)$ is a vector subspace of $\mathbb{R}^m$ and $\dim(\mathcal{C}(A)) \leq m$. We can extract a basis of $\mathcal{C}(A)$ from the columns of A.

Theorem 3.1.3

Let M be a $m \times n$ matrix. Let M' be a REF of M. Then the columns of M corresponding to the pivot columns of M' are a basis of $\mathcal{C}(M)$. Hence $\dim(\mathcal{C}(M)) = \text{rank}(M)$.

Example 3.1.4

Let $S = \{(1,0,1,0),(0,1,1,0)\}$.
We can easily verify that the vectors in S are linearly independent. We can extend S to a basis of $\mathbb{R}^4$ constructing the auxiliary matrix M whose columns are the two vectors of S and the four vectors of the standard basis of $\mathbb{R}^4$. Then we reduce M to a REF M'.

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

$$\xrightarrow{E_{32}(-1)} \begin{bmatrix} \mathbf{1} & 0 & 1 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 1 & 0 & 0 \\ 0 & 0 & -\mathbf{1} & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix} = M'$$

We notice that the columns 1, 2, 3 and 6 of M' contains the pivots. Therefore a basis $\mathcal{B}$ of $\mathbb{R}^4$ extending S is given by the corresponding columns of M, namely

$$\mathcal{B} = \{(1,0,1,0),(0,1,1,0),(1,0,0,0),(0,0,0,1)\}.$$

3.1.2 Isomorphic vector spaces

We recall that if V is a vector space with $\dim(V) = n$ and $\mathcal{B} = \{b_1, \ldots, b_n\}$ is a basis of V, then any vector $v \in V$ can be written in a unique way as a linear combination of the vectors $b_1, \ldots, b_n$, namely

$$v = x_1 b_1 + x_2 b_x + \ldots x_n b_n$$

for some $x_1, \ldots, x_n \in \mathbb{R}$ which are uniquely determined. The coefficients $x_1, \ldots, x_n$ are called the coordinates of v with respect to the basis $\mathcal{B}$. Therefore, we can associate the unique n-tuple of coordinates with any $v \in V$:

$$\begin{aligned} \mathbb{R}^n &\longleftrightarrow V \\ (x_1, \ldots, x_n) &\longleftrightarrow x_1 v_1 + \cdots + x_n v_n. \end{aligned}$$

We say that $\mathbb{R}^n$ is isomorphic to V because doing operations on the vectors of V is equivalent to performing the same operations on the corresponding n-tuples of $\mathbb{R}^n$.

Example 3.1.5

The standard basis of $\mathbb{R}^{2,2}$ is $\mathcal{B} = \{B_1, B_2, B_3, B_4\}$, where

$$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad B_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad B_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

We can write any matrix $A \in \mathbb{R}^{2,2}$ in a unique way as a linear combination of the matrices in $\mathcal{B}$, namely

$$A = x_1B_1 + x_2B_2 + x_3B_3 + x_4B_4$$

for some $x_1, \dots, x_4 \in \mathbb{R}$. For example,

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} = 1B_1 + 2B_2 + 0B_3 + 3B_4.$$

Hence we can associate with A the vector

$$(1, 2, 0, 3) \in \mathbb{R}^4.$$

3.1.3 The vector space of polynomials

So far the only examples of vector spaces we have seen are the spaces $\mathbb{R}^n$, $\mathbb{R}^{m,n}$ and their subspaces. Now we introduce the set $\mathbb{R}[x]$ of all polynomials in one variable having real coefficients. We recall that a one variable real polynomial is an expression of the form

$$p(x) = a_0 + a_1x + \cdots + a_nx^n,$$

where $n \in \mathbb{N}$ and $a_0, \dots, a_n \in \mathbb{R}$. If all the coefficients are equal to 0, then we say that the polynomial is the zero polynomial. Otherwise, we suppose that $a_n \neq 0$ and we say that the polynomial $p(x)$ has degree n.

We can add two polynomials and we can also multiply polynomials by scalars. In such a way the set $\mathbb{R}[x]$ becomes a vector space.

Example 3.1.6

Consider the following polynomials:

$$\begin{aligned} p(x) &= 1 + x^2 \\ q(x) &= x + 3x^4 \end{aligned}$$

The polynomial $p(x)$ has degree 2 while the degree of $q(x)$ is 4. The sum of the two polynomials is

$$p(x) + q(x) = 1 + x + x^2 + 3x^4.$$

The space $\mathbb{R}[x]$ has infinite dimension. In fact, any polynomial can be written in a unique way as a linear combination of the powers

$$x^0, x^1, x^2, \dots .$$

Therefore a basis for $\mathbb{R}[x]$ is given by

$$\mathcal{B} = \{1, x, x^2, \dots\} = \{x^k : k \in \mathbb{N}\}.$$

For any $n \in \mathbb{N}$ we can consider the vector subspace of the real polynomials having degree at most n:

$$\mathbb{R}_n[x] = \{p(x) : \ p(x) \text{ has degree at most } n\}.$$

We have that $\dim(\mathbb{R}_n[x]) = n + 1$. Indeed $\mathbb{R}_n[x]$ is isomorphic to $\mathbb{R}^{n+1}$ through the correspondence

$$\begin{aligned} \mathbb{R}^{n+1} &\longleftrightarrow \mathbb{R}_n[x] \\ (a_0, \dots, a_n) &\longleftrightarrow a_0 + a_1 x + \dots + a_n x^n \end{aligned}$$

The standard basis for $\mathbb{R}_n[x]$ is $\{1, x, \dots, x^n\}$.

Example 3.1.7

Consider in $\mathbb{R}_2[x]$ the polynomials

$$p_1(x) = 1 + x, \quad p_2(x) = x + x^2, \quad p_3(x) = 1 + 2x + x^2.$$

We can wonder if such polynomials are a basis of $\mathbb{R}_2[x]$. We can solve this problem in $\mathbb{R}^3$, using the correspondence

$$\begin{aligned} \mathbb{R}^3 &\longleftrightarrow \mathbb{R}_2[x] \\ v_1 = (1, 1, 0) &\longleftrightarrow p_1(x) \\ v_2 = (0, 1, 1) &\longleftrightarrow p_2(x) \\ v_3 = (1, 2, 1) &\longleftrightarrow p_3(x) \end{aligned}$$

We construct and reduce to REF the auxiliary matrix M whose columns are v_1, v_2, v_3:

$$M = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{E_{32}(-1)} \begin{bmatrix} \mathbf{1} & 0 & 1 \\ 0 & \mathbf{1} & 1 \\ 0 & 0 & 0 \end{bmatrix} = M'.$$

Since M' has two pivots, $\operatorname{rank}(M) = 2$. Hence the vectors v_1, v_2, v_3 are not linearly independent. The same holds for the polynomials $p_1(x), p_2(x), p_3(x)$, which do not form a basis for $\mathbb{R}_2[x]$.

Chapter 4

Determinants

We recall that a square matrix A of order n is invertible if and only if there exists a square matrix B of order n such that

$$AB = BA = I_n$$

where I_n is the identity matrix of order n.

If A is a square matrix of order 1, then $A = \begin{bmatrix} a \end{bmatrix}$ for some $a \in \mathbb{R}$. It is easily seen that A is invertible if and only if $a \neq 0$ and $A^{-1} = \begin{bmatrix} a^{-1} \end{bmatrix}$.

If A is a square matrix of order $n > 1$, there is not a so easy condition for invertibility which can be related to the coefficients of the matrix. For example, none of the coefficients of the matrix

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

is equal to 0. By the way A is not invertible. In fact, suppose that $AB = I_2$ and

$$B = \begin{bmatrix} b_1 & b_3 \\ b_2 & b_4 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 & b_3 \\ b_2 & b_4 \end{bmatrix} = \begin{bmatrix} b_1 + b_2 & b_3 + b_4 \\ b_1 + b_2 & b_3 + b_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence

$$b_1 + b_2 = 1 \quad \text{and} \quad b_1 + b_2 = 0,$$

which is absurd.

A condition for the invertibility of a square matrix A is that the determinant of A is different from 0.

4.1 Determinant

We can associate with any square matrix A of order n its determinant, which we denote by $\det(A)$. The formal definition of determinant is not straightforward. Our approach in this notes is just computational, namely we list the fundamental properties of the determinants and we see some techniques for computing determinants.

4.1.1 Properties with respect to elementary row operations

- If A is a REF matrix and A has n pivots, then $\det(A)$ is equal to the product of the pivots.
- If A is a REF matrix and A has less than n pivots, then $\det(A) = 0$.
- If B is obtained from A switching two rows of A, then $\det(B) = -\det(A)$.
- If B is obtained from A multiplying one row of A by a scalar $k \neq 0$, then $\det(B) = k\det(A)$.
- If B is obtained from A adding to a row of A a multiple of another row, then $\det(B) = \det(A)$.

Example 4.1.1

Consider the following matrices:

$$A_1 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 4 \\ 0 & 0 & 7 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} \quad A_3 = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 4 \\ 1 & 0 & 0 \end{bmatrix}$$

- A_1 has order 3, is in REF and has three pivots. Then
$$\det(A_1) = 1 \cdot 2 \cdot 7 = 14.$$
- A_2 has order 3, is in REF and has less than three pivots. Then
$$\det(A_2) = 0.$$
- A_3 has order 3 but it is not in REF. We reduce A_3 to REF:
$$A_3 \xrightarrow{S_{12}} \begin{bmatrix} 1 & 1 & 4 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix} \xrightarrow{E_{31}(-1)} \begin{bmatrix} 1 & 1 & 4 \\ 0 & 1 & 2 \\ 0 & -1 & -4 \end{bmatrix} \xrightarrow{E_{32}(1)} \begin{bmatrix} 1 & 1 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & -2 \end{bmatrix} = A_3'$$

The matrix A_3' is in REF and has three pivots. We have obtained A_3' through two elementary operations of type 3 and one operation of type 1 (switching row 1 and 2). Therefore

$$\det(A_3) = -\det(A_3') = -(-2) = 2.$$

From the properties above we deduce immediately the following result.

Theorem 4.1.2

Let A be a square matrix of order n in REF. Then

$$\det(A) \neq 0 \Leftrightarrow \operatorname{rank}(A) = n$$

When we reduce a matrix A to a REF A' we perform a sequence of elementary row operations. When we apply a row operation either we do not modify the determinant or we multiply by -1 or some non-zero $k \in \mathbb{R}$ the determinant. Therefore the following holds.

Theorem 4.1.3

Let A be a square matrix. If A' is a REF of A then

$$\det(A') = c \cdot \det(A) \quad \text{for some } c \in \mathbb{R}^*.$$

As an immediate consequence we get the result below.

Theorem 4.1.4

Let A be a square matrix of order n. Then

$$\det(A) \neq 0 \Leftrightarrow \operatorname{rank}(A) = n.$$

Proof. Let A' be a REF of A. Then $\det(A') = c \cdot \det(A)$ for some $c \in \mathbb{R}^*$. Hence

$$\det(A) \Leftrightarrow \det(A') \neq 0 \Leftrightarrow A' \text{ has } n \text{ pivots} \Leftrightarrow \operatorname{rank}(A) = n.$$ □

4.1.2 Determinant of square matrices of order 1 or 2

- Let $A = [a]$ be a square matrix of order 1. Since A is a REF matrix we have that

$$\det(A) = a.$$

- We can compute the determinant of a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

through the formula

$$\det(A) = ad - bc.$$

Such a formula can be deduced reducing A to REF.

4.1.3 Laplace expansions

The following two theorems due to Laplace hold.

Theorem 4.1.5: Laplace expansion along a row

Let A be a square matrix of order n. Let $k \in \{1, \dots, n\}$. Then

$$\det(A) = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det(A_{kj})$$

where A_{kj} is the matrix obtained from A removing the row k and the column j.

Theorem 4.1.6: Laplace expansion along a column

Let A be a square matrix of order n. Let $k \in \{1, \dots, n\}$. Then

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+k} a_{ik} \det(A_{ik})$$

where A_{ik} is the matrix obtained from A removing the row i and the column k.

Remark 4.1.7

Since $\det(A)$ does not depend on the row or the column of the expansion, it is convenient to consider the row(s) or the column(s) with more zeros.

Example 4.1.8

Consider the matrix

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

Since in the third row there are two zeros, it is convenient to compute $\det(A)$ using the Laplace expansion along the third row. We have

$$\det(A) = (-1)^{3+3} 1 \cdot \det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = 4 - 6 = -2.$$

4.1.4 Sarrus rule for square matrices of order 3

For square matrices of order 3 (**ONLY** order 3) we can use Sarrus rule. If

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

then $\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}$.

4.2 Invertible matrices

We recall that a square matrix A of order n is invertible if there exists a square matrix B of order n such that

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

We state an important theorem due to Binet.

Theorem 4.2.1: Binet formula

Let A and B be two square matrices of order n. Then

$$\det(AB) = \det(A) \cdot \det(B).$$

Using Binet formula one can deduce the following result.

Theorem 4.2.2

Let A be a square matrix of order n. If A is invertible then $\det(A) \neq 0$.

Indeed, if $\det(A) \neq 0$, a square matrix A is invertible. In fact, the following result holds.

Theorem 4.2.3

Let A be a square matrix of order n. If $\det(A) \neq 0$, then there exists a square matrix B of order n such that $AB = I_n$.

Remark 4.2.4

According to the theorem above, we can construct a matrix B such that $AB = I_n$. Indeed, it can be proved that also $BA = I_n$, namely B is the inverse of A.

From the previous result we deduce the following theorem, which expresses a sufficient and necessary condition for the invertibility of a matrix.

Theorem 4.2.5

Let A be a square matrix of order n. Then

$$A \text{ is invertible} \Leftrightarrow \det(A) \neq 0.$$

4.2.1 How to compute the inverse of a matrix

Suppose that A is a square matrix of order n such that $\det(A) \neq 0$. Constructing the inverse A^{-1} of A amounts to finding the n columns of a square matrix B such that $AB = I_n$. This is equivalent to solving n linear systems, where the matrix of the coefficients is A and the column vectors of constant terms are the n columns of the identity matrix. This can be done efficiently reducing the block matrix $[A|I_n]$ to RREF. In fact, since $\operatorname{rank}(A) = n$, the RREF of $[A|I_n]$ is $[I_n|B]$ for some matrix B. The matrix B we obtain is the inverse of A.

Example 4.2.6

Consider the matrix

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

The matrix A is invertible since $\det(A) = -6 \neq 0$. We can compute A^{-1} as follows:

$$[A|I_2] \xrightarrow{E_{21}(-3)} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -6 & -3 & 1 \end{array}\right] \xrightarrow{D_2(-1/6)} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{1}{2} & -\frac{1}{6} \end{array}\right]$$
$$\xrightarrow{E_{12}(-2)} \left[\begin{array}{cc|cc} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{1}{6} \end{array}\right]$$

Chapter 5

Linear functions, eigenvalues and eigenvectors

5.1 Introduction

We begin this chapter with a definition.

Definition 5.1.1

Let V and W be two vector spaces. We say that a function $f : V \to W$ is linear if

1. $f(u + v) = f(u) + f(v)$ for all $u, v \in V$;
2. $f(tv) = tf(v)$ for all $t \in \mathbb{R}$ and $v \in v$.

Remark 5.1.2: some properties

1. From the definition it follows that $f(O_V) = O_W$, where O_V and O_W are the zero vectors of V and W respectively. In fact

$$f(O_V) = f(O_V + O_V) = f(O_V) + f(O_V)$$

namely $f(O_V) = f(O_V) + f(O_V)$. Therefore, adding $-f(O_V)$ to both sides of the equation we get

$$O_W = f(O_V).$$

2. Let $f : V \to W$ be a linear function. Let $v_1, \ldots, v_k \in V$ and $x_1, \ldots x_k \in \mathbb{R}$. Then

$$f(x_1v_1 + x_2v_2 + \cdots + x_kv_k) = x_1f(v_1) + x_2f(v_2) + \cdots + x_kf(v_k).$$

In particular, if $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ is a basis of V, then

$$f(x_1 b_1 + x_2 b_2 + \cdots + x_n b_n) = x_1 f(b_1) + x_2 f(b_2) + \cdots + x_n f(b_n).$$

Therefore, once we know the values of $f(b_1), \ldots, f(b_n)$, we can find the value $f(v)$ for any vector $v \in V$ since v can be expressed in a unique way as a linear combination of the vectors of the basis $\mathcal{B}$.

5.2 Matrix associated with a linear function

Let V be a vector space with $\dim(V) = n$. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V. Let W be a vector space with $\dim(W) = m$. Let $\mathcal{C} = \{w_1, \ldots, w_m\}$ be a basis of W. We can associate with any vector $v \in V$ its coordinates with respect to $\mathcal{B}$:

$$v \to \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then we can associate with $f(v_1), f(v_2), \ldots, f(v_n)$ their coordinates with respect to the basis $\mathcal{C}$:

$$f(v_1) \to \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad f(v_2) \to \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \ldots, \quad f(v_n) \to \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Finally we associate with $f(v)$ its coordinates with respect to the basis $\mathcal{C}$:

$$f(v) \to \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Rephrasing $f(v) = x_1 f(v_1) + x_2 f(v_2) + \cdots + x_n f(v_n)$ in terms of the respective coordinates we get

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Therefore

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

The matrix A is the matrix associated with f with respect to the basis $\mathcal{B}$ on V and $\mathcal{C}$ on W.

5.3 Change of basis

Let V be a vector space with $\dim(V) = n$. Consider two bases of V:

$$\mathcal{B} = \{b_1, \dots, b_n\} \text{ and } \mathcal{B}' = \{b'_1, \dots, b'_n\}.$$

Let $id : V \to V$ be the identity function which takes any vector $v \in V$ to itself. The identity function is linear and we can write the matrix M of id with respect to the basis $\mathcal{B}$ on the domain and $\mathcal{B}'$ on the codomain. If

$$v = x_1 b_1 + \dots + x_n b_n,$$
$$v = x'_1 b'_1 + \dots + x'_n b'_n,$$

then

$$\begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} = M \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Moreover, the columns of M are linearly independent because each column contains the coordinates of a vector of $\mathcal{B}$ with respect to basis $\mathcal{B}'$. Therefore M is invertible. Hence

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = M^{-1} \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}.$$

The matrix M^{-1} is the matrix of change of basis from $\mathcal{B}'$ to $\mathcal{B}$.

5.4 Eigenvalues and eigenvectors

Consider a vector space V and a linear function $f : V \to V$. If $\mathcal{B}$ is a basis of V, then we can write the matrix M of f with respect to the basis $\mathcal{B}$ on the domain and the codomain. The matrix M is square. Moreover it can happen that M is also diagonal.

Example 5.4.1

Consider the linear function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$f(\alpha, \beta) = (\beta, \alpha) \quad \text{for all } (\alpha, \beta) \in \mathbb{R}^2.$$

Now we take the basis $\mathcal{B} = \{(1, 1), (-1, 1)\}$ of $\mathbb{R}^2$. We notice that

$$\begin{aligned} f(1, 1) &= (1, 1) = 1(1, 1) + 0(-1, 1); \\ f(-1, 1) &= (1, -1) = 0(1, 1) - 1(1, -1). \end{aligned}$$

Therefore, the matrix M of f with respect to the basis $\mathcal{B}$ is

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

In the last example the vectors of the basis $\mathcal{B}$ were eigenvectors of the function f, according to the following definition.

Definition 5.4.2: eigenvector and eigenvalue

Let V be a vector space and $f : V \to V$ a linear function. We say that a non-zero vector $v \in V$ is an eigenvector for f is there exists a $\lambda \in \mathbb{R}$ such that

$$f(v) = \lambda v.$$

The number λ is called the eigenvalue of the eigenvector v.

Remark 5.4.3

Notice that an eigenvector cannot be a zero vector by definition. By the way, an eigenvalue can also be equal to 0.

Remark 5.4.4

Suppose that $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis of a vector space V and that any vector in $\mathcal{B}$ is an eigenvector of a linear function $f : V \to V$. Then the matrix M of f with respect to the basis $\mathcal{B}$ is diagonal. In fact,

$$f(v_1) = \lambda_1 v_1, \quad f(v_2) = \lambda_2 v_2, \ldots, f(v_n) = \lambda_n v_n$$

for some $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$. Then

$$M = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_{n-1} & 0 \\ 0 & 0 & \ldots & 0 & \lambda_n \end{bmatrix}.$$

5.4.1 Geometrical interpretation in the plane and in the space

Let $V = \mathbb{R}^2$ or $V = \mathbb{R}^3$. By definition, if $v \in V$ is an eigenvector of a linear function $f : V \to V$, then $f(v) = \lambda v$ for some $\lambda \in \mathbb{R}$. Therefore v and $f(v)$ belong to the same line passing through the origin. Geometrically, f stretches or shrinks the vector v. Moreover, in the case $\lambda < 0$, the vectors v and $f(v)$ have opposite directions.

5.5 Computing eigenvalues and eigenvectors

Let $f : V \to V$ be a linear function. Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a basis of V. We can construct the matrix A of f with respect to the basis $\mathcal{B}$. We know that $v \in V$ is an eigenvector of f if and only if $f(v) = \lambda v$ for some $\lambda \in \mathbb{R}$. We can associate with v and $f(v)$ their coordinate vectors with respect to the basis $\mathcal{B}$:

$$v \to x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad f(v) \to y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

We have that $y = Ax$. Moreover, since $f(v) = \lambda v$, we also have that $y = \lambda x$. Hence $Ax = \lambda x$, namely

$$(A - \lambda I_n)x = O.$$

Hence x is a non-zero solution of a homogeneous linear system whose matrix of coefficients is $A - \lambda I_n$. Since a homogeneous linear system has at least the zero solution, there exists a non-zero solution if and only if the linear system has infinitely many solutions. This happens if and only if $n > \operatorname{rank}(A - \lambda I_n)$ or, equivalently,

$$\det(A - \lambda I_n) = 0.$$

This latter condition can be used in order to find the eigenvalues and eigenvectors of f. We can proceed as follows.

1. We define the characteristic polynomial
$$p(\lambda) = \det(A - \lambda I_n),$$
which is a polynomial of degree n in the variable λ.

2. We solve the equation $p(\lambda) = 0$. The solutions of $p(\lambda) = 0$ are the roots of $p(\lambda)$ and are the eigenvalues of A.

3. For any eigenvalue λ we can compute all the solutions of the linear system
$$(A - \lambda I_n)x = O.$$
The set of all the solutions of this linear system is a vector subspace of $\mathbb{R}^n$ (see Theorem 3.0.3), which we call the eigenspace of the eigenvalue λ and we denote by V_λ. We notice that all the vectors in V_λ, except for the zero vector, are eigenvectors of f.

Remark 5.5.1

Here are some important remarks on eigenvalues and eigenvectors.

1. Since $p(\lambda)$ is a polynomial of degree n, the matrix A can have at most n distinct eigenvalues.

2. The polynomial $p(\lambda)$ might have no real root. In such a case, there is no real eigenvalue for A.

3. The eigenvalues of f do not depend on the basis of V we choose to construct the matrix A of f (a theorem will be stated below). Therefore we can say that the eigenvalues of f are the eigenvalues of A, where A is the matrix of f with respect to an arbitrary basis of V.

The procedure we use to compute the eigenvalues of f is based upon the polynomial
$$p(\lambda) = \det(A - \lambda I_n).$$
Since the matrix A depends on the basis we choose for the vector space V, one could wonder if also the eigenvalues depend on the basis. Actually, they are independent of the basis, as stated in the following theorem.

Theorem 5.5.2

Let $f : V \to V$ be a linear function, where V is a vector space with $\dim(V) = n$. Suppose that A is the matrix of f with respect to a basis $\mathcal{B}$ and A' is the matrix of f with respect to another basis $\mathcal{B}'$. If

$$p(\lambda) = \det(A - \lambda I_n)$$
$$q(\lambda) = \det(A' - \lambda I_n)$$

then $p(\lambda) = q(\lambda)$.

Proof. First we notice that we can associate with any vector $v \in V$ its coordinates

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ with respect to the basis } \mathcal{B},$$

$$x' = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} \text{ with respect to the basis } \mathcal{B}'.$$

Similarly, we can associate with $f(v)$ its coordinates

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ with respect to the basis } \mathcal{B},$$

$$y' = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{bmatrix} \text{ with respect to the basis } \mathcal{B}'.$$

We have that

$$y = Ax \quad \text{and} \quad y' = A'x'.$$

If M is the matrix of change of basis from $\mathcal{B}$ to $\mathcal{B}'$ then

$$x' = Mx \quad \text{and} \quad y' = My.$$

Therefore

$$y' = A'x' \Leftrightarrow My = A'Mx \Leftrightarrow y = M^{-1}A'Mx$$

In particular we have that $A = M^{-1}A'M$. Now we can prove that $p(\lambda) = q(\lambda)$.

$$\begin{aligned}
p(\lambda) &= \det(A - \lambda I_n) = \det(M^{-1}A'M - \lambda I_n) = \det(M^{-1}A'M - \lambda M^{-1}I_nM)\\
&= \det(M^{-1}(A' - \lambda I_n)M) = \det(M^{-1}) \cdot \det(A' - \lambda I_n) \cdot \det(M)\\
&= \det(M^{-1}) \cdot \det(M) \cdot \det(A' - \lambda I_n) = \det(M^{-1}M) \cdot \det(A' - \lambda I_n)\\
&= \det(I_n) \cdot \det(A' - \lambda I_n) = \det(A' - \lambda I_n)\\
&= q(\lambda).
\end{aligned}$$

□

5.6 Diagonalizable matrices

Let $f : V \to V$ be a linear function, where V is a vector space with $\dim(V) = n$. Suppose that A is the matrix of f with respect to a basis $\mathcal{B}$. If there exists a basis $\mathcal{B}' = \{v_1, \ldots, v_n\}$ of V formed by eigenvectors of f, then

$$f(v_1) = \lambda_1 v_1,\ f(v_2) = \lambda_2 v_2,\ \ldots,\ f(v_n) = \lambda_n v_n$$

for some $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$. Hence the matrix D of f with respect to $\mathcal{B}'$ is

$$D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix},$$

namely D is a diagonal matrix whose coefficients on the main diagonal are the eigenvalues of f. Now, if we suppose to write $v_1, \ldots, v_n$ as column vectors whose components are their coordinates with respect to the basis $\mathcal{B}$, we have that

$$Av_1 = \lambda_1 v_1,\ Av_2 = \lambda_2 v_2,\ \ldots,\ Av_n = \lambda v_n.$$

Therefore we can construct two matrices, having as columns respectively the vectors Av_i and $\lambda_i v_i$ for $i \in \{1, \ldots, n\}$. These two matrices are equal and we have that

$$\left[Av_1|Av_2|\ldots|Av_n\right] = \left[\lambda_1 v_1|\lambda_2 v_2|\ldots|\lambda_n v_n\right].$$

Equivalently we have that

$$A \cdot \left[v_1|v_2|\ldots|v_n\right] = \left[v_1|v_2|\ldots|v_n\right] \cdot D.$$

If we define the matrix $M = \left[v_1|v_2|\dots|v_n\right]$, we have that

$$AM = MD$$

or equivalently

$$M^{-1}AM = D.$$

We notice that each column of M contains the coordinates of some vector v_i with respect to the basis $\mathcal{B}$, namely M is the matrix of change of basis from $\mathcal{B}'$ to $\mathcal{B}$. We can also say that A is diagonalizable, according to the following definition.

Definition 5.6.1: diagonalizable matrix

A square matrix A of order n is diagonalizable if and only if there exists an invertible matrix M such that $M^{-1}AM = D$, where D is a diagonal matrix of order n.

The following theorem holds.

Theorem 5.6.2

A square matrix A of order n is diagonalizable if and only if there exists a basis of $\mathbb{R}^n$ formed by eigenvectors of A.

Remark 5.6.3

According to Theorem 5.6.2 we can diagonalize A once we know a basis of $\mathbb{R}^n$ formed by eigenvectors of A. Indeed, if we consider n column vectors $v_1, \dots, v_n \in \mathbb{R}^n$ which are eigenvectors of A and form a basis of $\mathbb{R}^n$, then we can construct the matrix

$$M = \left[v_1|v_2|\dots|v_n\right]$$

In this case we have that

$$M^{-1}AM = D,$$

where D is a diagonal matrix, whose coefficients on the main diagonal are the eigenvalues of $v_1, \dots, v_n$. In the following section we will give some useful conditions for the existence of such a basis of eigenvectors.

5.6.1 Algebraic and geometric multiplicity of an eigenvalue

Let A be a square matrix of order n. Let $\alpha \in \mathbb{R}$ be an eigenvalue of A.

- We say that α has algebraic multiplicity $k \in \mathbb{N}^*$ if

$$p(\lambda) = (\lambda - \alpha)^k \cdot g(\lambda)$$

where $g(\lambda)$ is a polynomial such that $g(\alpha) \neq 0$.

- We say that $d \in \mathbb{N}^*$ is the geometric multiplicity of α if $d = \dim(V_\alpha)$, where V_α is the eigenspace of α.

The algebraic and geometric multiplicities are related as follows.

Theorem 5.6.4

Let $\alpha \in \mathbb{R}$ be an eigenvalue of a square matrix A. If k is the algebraic multiplicity of α and d is the geometric multiplicity of α, then

$$1 \leq d \leq k.$$

Now we state an important theorem which furnishes necessary and sufficient conditions for the diagonalizability of a square matrix.

Theorem 5.6.5

Let A be a real square matrix of order n. Then A is diagonalizable if and only if all the eigenvalues of A are real and the geometric multiplicity of any eigenvalue is equal to its algebraic multiplicity.

Remark 5.6.6

In the special case that A has n distinct eigenvalues, then the characteristic polynomial $p(\lambda)$ can be factored as

$$p(\lambda) = \pm(\lambda - \lambda_1) \cdot (\lambda - \lambda_2) \cdots (\lambda - \lambda_n).$$

Hence, all the eigenvalues of A have algebraic multiplicity equal to 1 and the same holds for the geometric multiplicity. Therefore A is diagonalizable.

Example 5.6.7

Consider the square matrix

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

The eigenvalues of A are 1 and -1. Since the eigenvalues of A are distinct, the matrix A is diagonalizable. The bases of the two eigenspaces are respectively

$$\mathcal{B}_1 = \{(1,1)\};$$
$$\mathcal{B}_{-1} = \{(-1,1)\}.$$

We can also notice that the vectors $(1,1)$ and $(-1,1)$ are linearly independent.

Hence, they form a basis for $\mathbb{R}^2$. Actually, this is due to Theorem 5.6.8, which we are going to state.

Theorem 5.6.8

Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be k distinct eigenvalues of a matrix A. Let $v_1, v_2, \ldots, v_k$ be k eigenvectors associated with $\lambda_1, \lambda_2, \ldots, \lambda_k$ respectively. Then $v_1, v_2, \ldots, v_k$ are linearly independent.

5.6.2 Power of a matrix

Computing the power of a matrix is a time-consuming task. By the way, in some cases such a computation can be relatively easy. For example, if we consider a diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & 0 & \ldots & 0 \\ 0 & d_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & d_n \end{bmatrix}$$

then

$$D^k = \begin{bmatrix} d_1^k & 0 & 0 & \ldots & 0 \\ 0 & d_2^k & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & d_n^k \end{bmatrix}$$

Even if A is diagonalizable, computing A^k is relatively easy. In fact, since

$$M^{-1}AM = D,$$

we have that

$$A = MDM^{-1}.$$

Therefore

$$A^k = MDM^{-1}MDM^{-1} \ldots MDM^{-1} = MD^kM^{-1}.$$

5.7 Orthogonal vectors and orthogonal bases

We recall that two vectors v and w in $\mathbb{R}^n$ are orthogonal if and only if

$$v \cdot w = 0$$

where $v \cdot w$ is the dot product defined as

$$v \cdot w = \sum_{i=1}^{n} v_i w_i.$$

5.7.1 Orthogonal projection of a vector along another vector

Consider two vectors v and w in $\mathbb{R}^n$ with $w \neq O$. We say that a vector $p \in \mathbb{R}^n$ is the orthogonal projection of v along w if

1. $p = kw$ for some $k \in \mathbb{R}$;

2. $(v - p) \cdot w = 0$.

We can determine uniquely the vector p as follows:

$$\begin{aligned}(v-p)\cdot w = 0 \Leftrightarrow v\cdot w - p\cdot w = 0 \Leftrightarrow p\cdot w = v\cdot w \Leftrightarrow (kw)\cdot w = v\cdot w \\ \Leftrightarrow k(w\cdot w) = v\cdot w \Leftrightarrow k = \frac{v\cdot w}{w\cdot w}.\end{aligned}$$

Therefore the orthogonal projection of v along w is

$$p = \frac{v \cdot w}{w \cdot w} w.$$

We notice that in the particular case that w is a unit vector we have that $w \cdot w = 1$ and $p = (v \cdot w)w$.

5.7.2 Orthogonal and orthonormal bases

Definition 5.7.1

Let $\mathcal{B} = \{b_1, \ldots, b_k\}$ be a basis of a vector space V, with $V \subseteq \mathbb{R}^n$. We say that $\mathcal{B}$ is an orthogonal basis if $b_i \cdot b_j = 0$ for all i, j with $i \neq j$.

Definition 5.7.2

A basis $\mathcal{B} = \{b_1, \ldots, b_k\}$ of a vector space $V \subseteq \mathbb{R}^n$ is orthonormal if $\mathcal{B}$ is an orthogonal basis and $\|b_i\| = 1$ for any $i \in \{1, \ldots, k\}$ (we also say that any b_i is a normal vector or a unit vector).

Remark 5.7.3

In the definitions above V is a vector subspace of $\mathbb{R}^n$. Actually, V can also be equal to $\mathbb{R}^n$. In that case $k = n$.

Example 5.7.4

1. The standard basis $\mathcal{E}$ of $\mathbb{R}^n$ is an orthonormal basis.
2. The set $\mathcal{B} = \{(2,0,0),(0,1,1),(0,1,-1)\}$ is an orthogonal basis of $\mathbb{R}^3$, but it is not an orthonormal basis.

Definition 5.7.5

Let $S = \{v_1, \ldots, v_k\}$ be a set of vectors in $\mathbb{R}^n$. We say that the vectors in S are mutually orthogonal if $v_i \cdot v_j = 0$ for all $i, j \in \{1, \ldots, k\}$ such that $i \neq j$.

Theorem 5.7.6

If $S = \{v_1, \ldots, v_k\}$ is a set of mutually orthogonal non-zero vectors in $\mathbb{R}^n$, then the vectors in S are linearly independent.

Remark 5.7.7

In the case that $k = n$ in the theorem above we have that S is a basis of $\mathbb{R}^n$.

Finding the coordinates of a vector with respect to an orthonormal basis is quite easy.

Theorem 5.7.8

Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be an orthonormal basis of $\mathbb{R}^n$. Let $v \in \mathbb{R}^n$.
If $v = x_1 b_1 + \cdots + x_n b_n$, then

$$x_1 = v \cdot b_1, \quad x_2 = v \cdot b_2, \quad \ldots \quad x_n = v \cdot b_n.$$

5.7.3 Gram-Schmidt process

Gram-Schmidt process is an useful algorithm for the construction of orthogonal bases of $\mathbb{R}^n$ or subspaces of $\mathbb{R}^n$.

- **Input**: k linearly independent vectors $v_1, v_2, \ldots v_k \in \mathbb{R}^n$.
- **Output:** k orthogonal vectors $b_1, b_2, \ldots, b_k \in \mathbb{R}^n$ such that

$$\langle v_1, v_2, \ldots, v_k \rangle = \langle b_1, b_2, \ldots, b_k \rangle.$$

- **The algorithm.** We construct the vectors $b_1, \ldots, b_k$ as follows:

$$\begin{aligned}
b_1 &= v_1; \\
b_2 &= v_2 - \frac{v_2 \cdot b_1}{b_1 \cdot b_1} b_1; \\
b_3 &= v_3 - \frac{v_3 \cdot b_2}{b_2 \cdot b_2} b_2 - \frac{v_3 \cdot b_1}{b_1 \cdot b_1} b_1; \\
&\vdots \\
b_k &= v_k - \frac{v_k \cdot b_{k-1}}{b_{k-1} \cdot b_{k-1}} b_{k-1} - \cdots - \frac{v_k \cdot b_2}{b_2 \cdot b_2} b_2 - \frac{v_3 \cdot b_1}{b_1 \cdot b_1} b_1;
\end{aligned}$$

5.7.4 Orthogonal matrices

Definition 5.7.9

A square matrix A of order n is orthogonal if $AA^T = A^T A = I_n$.

Remark 5.7.10

1. The columns and the rows of an orthogonal matrix are mutually orthogonal.
2. The columns and the rows of an orthogonal matrix are unit vectors.
3. If A is orthogonal, then A is invertible and $A^{-1} = A^T$.

An important theorem relates symmetric matrices and diagonalizability.

Theorem 5.7.11: spectral theorem

Let A be a real square matrix. Then all the eigenvalues of A are real and there exists an orthogonal matrix P such that $P^T AP = D$, where D is a diagonal matrix.

Chapter 6

Appendix

6.1 Orthogonal projection of a point onto a plane

Consider in the $3d$-space a plane

$$\pi : ax + by + cz = d,$$

where $n = (a, b, c) \in \mathbb{R}^3 \backslash \{O\}$. Consider a point $P = (x_P, y_P, z_P)$ not belonging to π and the line

$$r : (x, y, z) = P + tn, \quad \text{with } t \in \mathbb{R}.$$

The line r intersects π in just one point. In fact,

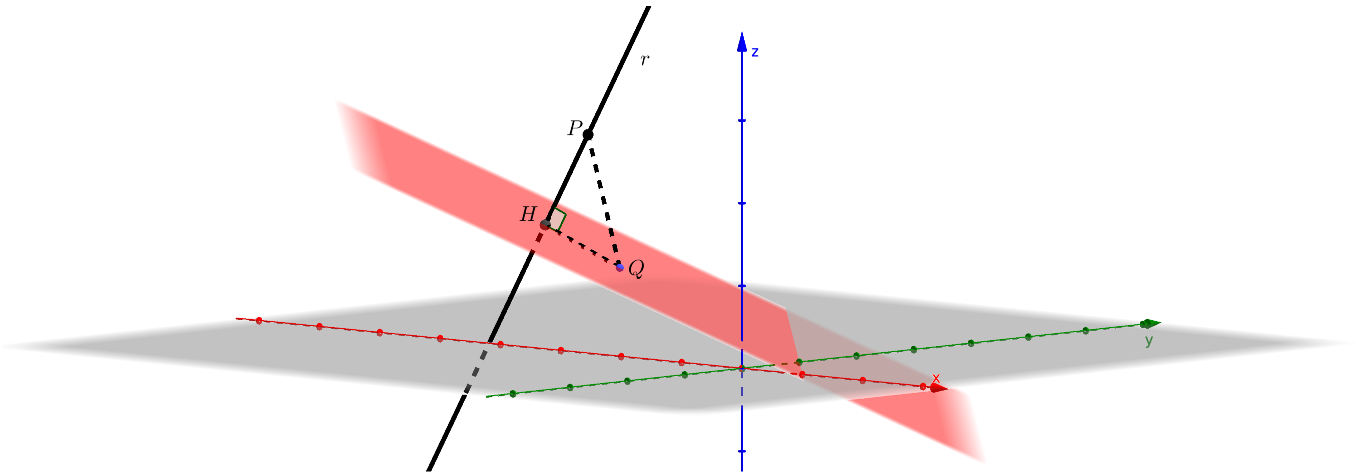
$$\begin{aligned}
&a(x_P + ta) + b(y_P + tb) + c(z_P + tc) = d \\
\Leftrightarrow &t(a^2 + b^2 + c^2) + (ax_P + by_P + cz_P) = d \\
\Leftrightarrow &t = \frac{d - ax_P - by_P - cz_P}{a^2 + b^2 + c^2}.
\end{aligned}$$

If we define $e = \frac{d - ax_P - by_P - cz_P}{a^2+b^2+c^2}$, then the only point of intersection H is

$$H = P + en.$$

Now we show that r is orthogonal to π, namely that the line segment $\overline{PH}$ is orthogonal to any line segment $\overline{QH}$ with $Q \in \pi$. This latter amounts to proving that

$$\|\overline{PH}\|^2 + \|\overline{HQ}\|^2 = \|\overline{PQ}\|^2. \tag{6.1}$$



For brevity we denote by

$$\begin{aligned} v &= H - P, \\ w &= Q - H, \\ v + w &= Q - P. \end{aligned}$$

Equation (6.1) is equivalent to

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2, \tag{6.2}$$

which can be expanded as

$$\|v\|^2 + \|w\|^2 = \|v\|^2 + 2(v \cdot w) + \|w\|^2.$$

This latter is true if and only if $v \cdot w = 0$. We notice that

$$\begin{aligned} v \cdot w &= (en) \cdot (Q - P) = e(n \cdot Q) - e(n \cdot P) \\ &= e(ax_Q + by_Q + cz_Q) - e(ax_P + by_P + cz_P) \\ &= ed - ed = 0. \end{aligned}$$

Since $v \cdot w = 0$, we conclude that Equation (6.1) holds and r is orthogonal to π.

6.2 Proof of Theorem 3.0.3

We check the three properties characterizing a vector subspace.

1. $O \in U$ (here O denotes the zero-vector in $\mathbb{R}^n$ written as column vector). The proof is immediate because $AO = O$.

2. If $v, w \in U$, then $v + w \in U$. By hypothesis we have that

$$\begin{aligned} Av &= O, \\ Aw &= O. \end{aligned}$$

We conclude that $v + w$ is a solution too because

$$A(v + w) = Av + Aw = O + O = O.$$

3. If $v \in U$ and $t \in \mathbb{R}$, then $tv \in U$. In fact, by hypothesis $Av = O$ and

$$A(tv) = t(Av) = O.$$

6.3 Proof of Theorem 3.0.7

Let $W = \text{span}\{v_1, \ldots, v_k\}$. We check the three properties characterizing a vector subspace.

1. $O \in W$. In fact,

$$O = 0v_1 + 0v_2 + \cdots + 0v_k$$

2. If $w_1, w_2 \in W$, then $w_1 + w_2 \in W$. In fact,

$$\begin{aligned} w_1 &= c_1 v_1 + \cdots + c_k v_k, \\ w_2 &= d_1 v_1 + \cdots + d_k w_k, \end{aligned}$$

for some $c_i, d_i \in \mathbb{R}$, with $1 \leq i \leq k$. Therefore

$$w_1 + w_2 = (c_1 + d_1)v_1 + \cdots + (c_k + d_k)v_k \in W.$$

3. If $t \in \mathbb{R}$ and $w \in W$, then $tw \in W$. In fact,

$$w = c_1 v_1 + \cdots + c_k v_k,$$

for some $c_i \in \mathbb{R}$, with $1 \leq i \leq k$. Therefore

$$tw = (tc_1)v_1 + \cdots + (tc_k)v_k \in W.$$

6.4 Proof or Theorem 3.0.11

Suppose first that condition 1 holds. Hence, $c_j \neq 0$ for some $j \in \{1, \dots, k\}$. Therefore we can write

$$v_j = -\frac{c_1}{c_j}v_1 + \cdots - \frac{c_{j-1}}{c_j}v_{j-1} - \frac{c_{j+1}}{c_j}v_{j+1} + \cdots - \frac{c_k}{c_j}v_k.$$

Now suppose that condition 2 holds. Hence, there exists $j \in \{1, \dots, k\}$ such that

$$v_j = d_1v_1 + \cdots + d_{j-1}v_{j-1} + d_{j+1}v_{j+1} + \cdots + d_kv_k$$

for some $d_i \in \mathbb{R}$. Therefore,

$$d_1v_1 + \cdots + d_{j-1}v_{j-1} - 1v_j + d_{j+1}v_{j+1} + \cdots + d_kv_k. = O.$$

If we define $c_1 = d_1, c_2 = d_2, \dots, c_j = -1, \dots, c_k = d_k$, we have that

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = O,$$

where at least the coefficient c_j is different from 0.

6.5 Proof of Theorem 3.0.14

Let $v_j = O$ for some $j \in \{1, \dots, k\}$. Then

$$0v_1 + \cdots + 0v_{j-1} + 1v_j + 0v_{j+1} + \cdots + 0v_k = O,$$

namely we have found a linear combination of the vectors of S, which gives the zero vector, where the coefficients are not all equal to 0. Then the vectors in S are not linearly independent.

6.6 Proof of Theorem 3.0.15

Since $S\backslash\{v_j\} \subseteq S$, we have that $\operatorname{span}(S\backslash\{v_j\}) \subseteq \operatorname{span}(S)$. Now we prove that $\operatorname{span}(S) \subseteq \operatorname{span}(S\backslash\{v_j\})$. Consider a vector $v \in V = \operatorname{span}(S)$. Such a vector can be written in the form

$$v = c_1v_1 + \cdots + c_kv_k$$

for some $c_1, c_2, \dots, c_k \in \mathbb{R}$. By hypothesis we can write

$$v_j = d_1v_1 + \cdots + d_{j-1}v_{j-1} + d_{j+1}v_{j+1} + \cdots + d_kv_k$$

for some coefficients $d_i \in \mathbb{R}$. Hence

$$\begin{aligned} v &= c_1v_1 + \cdots + c_jv_j + \cdots + c_kv_k \\ &= c_1v_1 + \cdots + c_j(d_1v_1 + \cdots + d_{j-1}v_{j-1} + d_{j+1}v_{j+1} + \cdots + d_kv_k) + \cdots + c_kv_k \\ &= (c_1 + c_jd_1)v_1 + \cdots + (c_{j-1} + c_jd_{j-1})v_{j-1} + (c_{j+1} + c_jd_{j+1})v_{j+1} + \cdots + (c_k + c_jd_k)v_k, \end{aligned}$$

namely we can write v as a linear combination of the vectors of $S\backslash\{v_j\}$.

6.7 Something more on linear functions

6.7.1 Kernel and image of a linear function

Let $f : V \to W$ be a linear function. The kernel of f (denoted by $\ker(f)$) is the subset of V defined as

$$\ker(f) = \{v \in V : f(v) = O_W\}.$$

The image of f (denoted by $\mathrm{Im}(f)$) is the subset of W defined as

$$\mathrm{Im}(f) = \{f(v) : v \in V\}.$$

The following holds.

Theorem 6.7.1

Let $f : V \to W$ be linear. Then

- $\ker(f)$ is a vector subspace of V;
- $\mathrm{Im}(f)$ is a vector subspace of W.

6.7.2 Injective and surjective linear functions. Isomorphisms.

Let $f : V \to W$ be a linear function. Then

- f is injective if, whenever $f(v_1) = f(v_2)$, then $v_1 = v_2$;
- f is surjective if $\mathrm{Im}(f) = W$.

The following characterization of injective linear functions holds.

Theorem 6.7.2

Let $f : V \to W$ be linear. Then f is injective if and only if $\ker(f) = \{O_V\}$.

A linear function, which is both injective and surjective, is said to be an *isomorphism*. We notice in passing that an injective and surjective function is said to be bijective. We recall that a bijective function is invertible. Hence, if $f : V \to W$ is an isomorphism, we can define $f^{-1} : W \to V$ as the function such that $f^{-1}(f(v)) = v$ for any $v \in V$ and $f(f^{-1}(w)) = w$ for any $w \in W$. It can be proved that f^{-1} is an isomorphism too, namely f^{-1} is linear, injective and surjective.

6.7.3 The rank-nullity theorem and its consequences

The following important result holds.

Theorem 6.7.3: rank-nullity theorem

If $f : V \to W$ is linear and V has finite dimension, then

$$\dim(V) = \dim(\ker(f)) + \dim(\mathrm{Im}(f)).$$

The following results follow from Theorem 6.7.3.

Corollary 6.7.1

If $f : V \to W$ is an isomorphism and $\dim(V) = n$, then $\dim(W) = n$.

Corollary 6.7.2

Let $f : V \to W$ be a linear function, where $\dim(V) = \dim(W) = n$. Then the following conditions are equivalent:

1. f is an isomorphism;
2. f is injective;
3. f is surjective.

6.8 Principal component analysis (PCA)

We recall that a symmetric real matrix is diagonalizable. For any real matrix A the following result holds.

Theorem 6.8.1

Let $A \in \mathbb{R}^{m,n}$. Then

1. the matrices AA^T and A^TA are symmetric;
2. the eigenvalues of A^TA and AA^T are non-negative.

6.8.1 Affine frames and coordinates

An affine frame in $\mathbb{R}^n$ consists of a point A and a basis $\mathcal{B} = \{b_1, \ldots, b_n\}$ of $\mathbb{R}^n$. For brevity we sometimes denote an affine frame with the pair $(A, \mathcal{B})$.

If $Af = (A, \mathcal{B})$ is an affine frame and P is a point of $\mathbb{R}^n$, then we can find the coordinates $(x_1, x_2, \ldots, x_n)$ of the vector $v = P - A$ with respect to the basis $\mathcal{B}$:

$$v = x_1 b_1 + \cdots + x_n b_n.$$

We say that $x_1, x_2, \ldots, x_n$ are the affine coordinates of P with respect to Af and we write

$$P = (x_1, \ldots, x_n)_{Af}$$

6.8.2 PCA

Let $s_1, s_2, \ldots, s_n$ be n samples of m-dimensional data:

$$s_1 = \begin{bmatrix} s_{11} \\ s_{21} \\ \vdots \\ s_{m1} \end{bmatrix}, \quad s_2 = \begin{bmatrix} s_{12} \\ s_{22} \\ \vdots \\ s_{m2} \end{bmatrix}, \quad \ldots, \quad s_n = \begin{bmatrix} s_{1n} \\ s_{2n} \\ \vdots \\ s_{mn} \end{bmatrix}.$$

We define the sample mean vector

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_m \end{bmatrix},$$

where

$$\begin{aligned} \mu_1 &= \frac{1}{n}(s_{11} + s_{12} + \cdots + s_{1n}), \\ \mu_2 &= \frac{1}{n}(s_{21} + s_{22} + \cdots + s_{2n}), \\ &\vdots \\ \mu_m &= \frac{1}{n}(s_{m1} + s_{m2} + \cdots + s_{mn}). \end{aligned}$$

If we define the matrices

$$\begin{aligned} S &= [s_1|s_2|\ldots|s_n] \in \mathbb{R}^{m,n}, \\ B &= [s_1 - \mu|s_2 - \mu|\ldots|s_n - \mu], \end{aligned}$$

then the covariance matrix is

$$C = \frac{1}{n-1} B B^T = \left(\frac{B}{\sqrt{n-1}}\right)\left(\frac{B}{\sqrt{n-1}}\right)^T \in \mathbb{R}^{m,m}.$$

According to Theorem 6.8.1, C is a symmetric matrix and all the eigenvalues of C are non-negative. There are m (possibly not distinct) eigenvalues, which we can arrange in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0.$$

We can diagonalize orthogonally the matrix C as

$$P^T C P = D.$$

The columns of $P = [p_1|p_2|\ldots|p_m]$ are said to be the *principal components* of C. If we take any sample $s \in \mathbb{R}^{m,1}$, we can find the coordinates of s with respect to the affine frame $(\mu, \mathcal{B})$, where $\mathcal{B}$ is the basis given by the columns of P:

$$s - \mu = x_1 p_1 + \cdots + x_m p_m.$$

We say that $x_1, \ldots, x_m$ are the *scores* of s with respect to the principal components.