MATHEMATICS I

Lecture Notes

Fernando F. Gonçalves

Revised Ed., 2016

ISEG
SINCE 1911
LISBOA
SCHOOL OF
ECONOMICS &
MANAGEMENT

Contents

Acknowledgements

I am most grateful to my colleague Gilson Silva for solving and gathering the solutions to the exercises proposed in these notes.

Lisbon, September 2016
Fernando F. Gonçalves

INTRODUCTION TO LINEAR ALGEBRA

1. Vectors

The linear space $\mathbb{R}^n$.

DEFINITION (The linear space $\mathbb{R}^n$). Consider the set $\mathbb{R}^n$, with n a positive integer, consisting of all sequences of n real numbers:

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

We use the short notation $\bar{x} = (x_1, x_2, \ldots, x_n)$ for elements of $\mathbb{R}^n$. In the present context, we call the elements of $\mathbb{R}^n$ *vectors*, and the individual numbers x_i ($i = 1, 2, \ldots, n$) the vector *components*.
We define two operations involving vectors. Let $(x_1, x_2, \ldots, x_n)$, $(y_1, y_2, \ldots, y_n)$ be arbitrary vectors in $\mathbb{R}^n$, and a an arbitrary real number (which, in this context, we call a *scalar*).

(1) *Addition of vectors*: $(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$;
(2) *Scalar multiplication*: $a(x_1, x_2, \ldots, x_n) = (ax_1, ax_2, \ldots, ax_n)$.

The set of vectors $\mathbb{R}^n$ together with the above operations is called the *real linear space* (or the *real vector space*) $\mathbb{R}^n$.

THEOREM 1.1. *Given vectors $\bar{x}$, $\bar{y}$, $\bar{z}$, and scalars a, b, the following properties hold:*

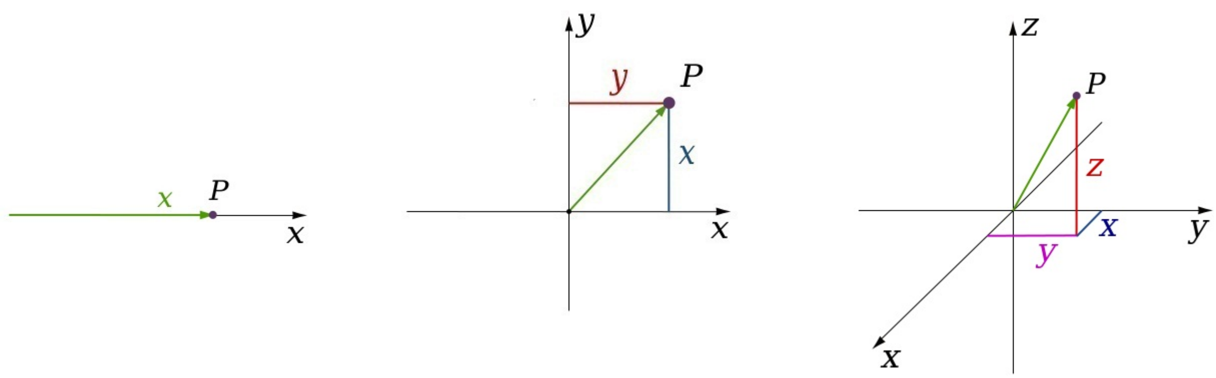
(1) Associativity of addition: $\bar{x} + (\bar{y} + \bar{z}) = (\bar{x} + \bar{y}) + \bar{z}$;
(2) Commutativity of addition: $\bar{x} + \bar{y} = \bar{y} + \bar{x}$;
(3) Identity element of addition: There exists a unique vector, called the zero vector *and denoted by $\bar{0}$, such that $\bar{x} + \bar{0} = \bar{x}$ for all $\bar{x} \in \mathbb{R}^n$;*
(4) Inverse elements of addition: For every vector $\bar{x} \in \mathbb{R}^n$, there exists a unique vector, called the additive inverse of $\bar{x}$ *or the* negative of $\bar{x}$ *and denoted by $-\bar{x}$, such that $\bar{x} + (-\bar{x}) = \bar{0}$;*
(5) Distributivity of scalar multiplication with respect to vector addition: $a(\bar{x} + \bar{y}) = a\bar{x} + a\bar{y}$;
(6) Distributivity of scalar multiplication with respect to addition of scalars: $(a + b)\bar{x} = a\bar{x} + b\bar{x}$;
(7) Compatibility of scalar multiplication with multiplication of scalars: $a(b\bar{x}) = (ab)\bar{x}$;
(8) Identity element of scalar multiplication: $1\bar{x} = \bar{x}$.

Clearly, the identity element of addition is the vector $\bar{0} = (0, 0, \ldots, 0)$, and the negative of a vector $\bar{x} = (x_1, x_2, \ldots, x_n)$ is the vector $-\bar{x} = (-x_1, -x_2, \ldots, -x_n)$.
The *subtraction of vectors* is defined by

$$\bar{x} - \bar{y} = \bar{x} + (-\bar{y}).$$

Let us consider the particular cases of the real linear spaces $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$. We want to obtain a geometric representation of vectors $x \in \mathbb{R}$, $(x, y) \in \mathbb{R}^2$, and $(x, y, z) \in \mathbb{R}^3$. For this, we follow the procedure:

- Set appropriate *Cartesian coordinate systems*: one-, two-, and three-dimensional Cartesian coordinate systems, respectively for the cases of $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$.
- Consider the *oriented line segments* (graphically, arrows) connecting the *terminal points* with coordinates x, (x, y), and (x, y, z) with the respective coordinate system *origin*.
- These line segments are the *position vectors* of points x, (x, y), and (x, y, z), and represent geometrically the vectors x, (x, y), and (x, y, z) (see Figure 1.1).



(a) Oriented line (b) Plane (c) Space

Figure 1.1. Geometric representation of vectors in Cartesian coordinate systems.

Note that, once appropriate Cartesian coordinate systems are set, we can establish a one-to-one correspondence between points of the oriented line, the plane, and the space, and vectors in $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$, respectively. Therefore, vectors can be uniquely represented by the position vectors of those points.
Note also that the above definitions of the operations addition of vectors and scalar multiplication are clearly consistent with the corresponding well-known geometric rules for adding vectors and multiplying a vector by a scalar.

Dot product, norm, distance, and orthogonal vectors.

DEFINITION (Dot product, norm, and distance). Let $\bar{x} = (x_1, x_2, \dots, x_n)$, $\bar{y} = (y_1, y_2, \dots, y_n)$ be vectors in $\mathbb{R}^n$.

(1) The *dot product* (or the *scalar product*) *of the vectors* $\bar{x}$ *and* $\bar{y}$, denoted by $\bar{x} \cdot \bar{y}$, is defined as

$$\bar{x} \cdot \bar{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

(2) The *norm of the vector* $\bar{x}$, denoted by $\|\bar{x}\|$, is

$$\|\bar{x}\| = \sqrt{\bar{x} \cdot \bar{x}} = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

(3) Taking $\bar{x}$ and $\bar{y}$ as position vectors in a n-dimensional Cartesian coordinate system, the *distance between the corresponding points* $\bar{x} = (x_1, x_2, \dots, x_n)$ *and* $\bar{y} = (y_1, y_2, \dots, y_n)$ is denoted by $\mathrm{d}(\bar{x}, \bar{y})$, and defined as

$$\mathrm{d}(\bar{x}, \bar{y}) = \mathrm{d}(\bar{y}, \bar{x}) = \|\bar{x} - \bar{y}\| = \sqrt{(\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y})} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}.$$

The above norm is called *Euclidean norm*, and the linear space $\mathbb{R}^n$ endowed with the dot product and norm as defined is called *Euclidean space*. Also, the above distance is called *Euclidean distance*, and we refer to the space of points in a Cartesian coordinate system associated with the linear space $\mathbb{R}^n$, together with the Euclidean distance, as an *affine Euclidean space*.
Note that the norm of a vector $\bar{x}$ can be geometrically interpreted as its *length*.

Example. Consider the vectors in $\mathbb{R}^3$: $\bar{x} = (1, 3, -5)$ and $\bar{y} = (4, -2, -1)$. The dot product $\bar{x} \cdot \bar{y}$ is

$$\bar{x} \cdot \bar{y} = \sum_{i=1}^{3} x_i y_i = 1 \times 4 + 3 \times (-2) + (-5) \times (-1) = 3.$$

The norm of $\bar{x}$ is

$$\|\bar{x}\| = \sqrt{\bar{x} \cdot \bar{x}} = \sqrt{\sum_{i=1}^{3} x_i^2} = \sqrt{1^2 + 3^2 + (-5)^2} = \sqrt{35}.$$

The distance between the corresponding points $\bar{x} = (1, 3, -5)$ and $\bar{y} = (4, -2, -1)$ is

$$\mathrm{d}(\bar{x}, \bar{y}) = \sqrt{\sum_{i=1}^{3} (x_i - y_i)^2} = \sqrt{(1-4)^2 + (3-(-2))^2 + (-5-(-1))^2} = \sqrt{50}.$$

Note that the vector

$$\frac{1}{\|\bar{x}\|}\bar{x} = \frac{1}{\sqrt{35}}(1,3,-5) = \left(\frac{1}{\sqrt{35}}, \frac{3}{\sqrt{35}}, -\frac{5}{\sqrt{35}}\right)$$

has norm 1:

$$\left\|\frac{1}{\|\bar{x}\|}\bar{x}\right\| == \left\|\left(\frac{1}{\sqrt{35}}, \frac{3}{\sqrt{35}}, -\frac{5}{\sqrt{35}}\right)\right\| = \sqrt{\left(\frac{1}{\sqrt{35}}\right)^2 + \left(\frac{3}{\sqrt{35}}\right)^2 + \left(-\frac{5}{\sqrt{35}}\right)^2} = 1.$$

Such a vector is called a *unit vector*.

DEFINITION (Angle between two vectors, orthogonal vectors). Let $\bar{x} = (x_1, x_2, \ldots, x_n)$, $\bar{y} = (y_1, y_2, \ldots, y_n)$ be vectors in $\mathbb{R}^n$, with $n \geq 2$.

(1) The *angle* θ between the vectors $\bar{x}$ and $\bar{y}$ is the least nonnegative value of θ satisfying

$$\bar{x} \cdot \bar{y} = \|\bar{x}\| \, \|\bar{y}\| \cos\theta.$$

(2) The vectors $\bar{x}$ and $\bar{y}$ are said to be *orthogonal* if

$$\bar{x} \cdot \bar{y} = 0.$$

Note that if $\bar{x}, \bar{y} \neq \bar{0}$ then

- $\theta = \arccos\left(\frac{\bar{x} \cdot \bar{y}}{\|\bar{x}\| \, \|\bar{y}\|}\right)$;
- $\bar{x} \cdot \bar{y} = 0 \Leftrightarrow \theta = \pi/2$;
- $\bar{x}, \bar{y}$ *collinear* $\Leftrightarrow |\bar{x} \cdot \bar{y}| = \|\bar{x}\| \, \|\bar{y}\|$. (Two vectors $\bar{x}$, $\bar{y}$ are said to be *collinear* if there exists a scalar a such that $\bar{x} = a\bar{y}$.)

Example. Consider the vectors in $\mathbb{R}^3$: $\bar{x} = (1, 2, -2)$, $\bar{y} = (-2, 1, 0)$, $\bar{z} = (-2, -4, 4)$, and $\bar{w} = (1, 0, 1)$. The vectors $\bar{x}$ and $\bar{y}$ are orthogonal as

$$\bar{x} \cdot \bar{y} = 1 \times (-2) + 2 \times 1 + (-2) \times 0 = 0.$$

The vectors $\bar{x}$ and $\bar{z}$ are collinear as $\bar{z} = -2\bar{x}$, and we have

$$|\bar{x} \cdot \bar{z}| = \|\bar{x}\| \, \|\bar{z}\| = 18.$$

The angle between the vectors $\bar{x}$ and $\bar{z}$ is

$$\theta = \arccos\left(\frac{\bar{x} \cdot \bar{z}}{\|\bar{x}\| \, \|\bar{z}\|}\right) = \arccos\left(\frac{-18}{3 \times 6}\right) = \arccos(-1) = \pi,$$

meaning that the vectors have opposite directions.

The vectors $\bar{x}$ and $\bar{w}$ are neither orthogonal nor collinear. The angle defined by these vectors is

$$\gamma = \arccos\left(\frac{\bar{x}\cdot\bar{w}}{\|\bar{x}\|\,\|\bar{w}\|}\right) = \arccos\left(\frac{-1}{3\times\sqrt{2}}\right) = \arccos\left(-\frac{\sqrt{2}}{6}\right) \simeq 1.8087375.$$

Linear combination and independence.

DEFINITION (Linear combination). Let

$$S = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\}$$

be a subset of the linear space $\mathbb{R}^n$. A vector $\bar{x}$ in $\mathbb{R}^n$ of the form

$$\bar{x} = \sum_{i=1}^{p} a_i\bar{x}_i = a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_p\bar{x}_p,$$

where $a_1, a_2, \ldots, a_p$ are scalars, is called a *linear combination of the elements of* S.

Example. Consider the set of vectors in $\mathbb{R}^3$

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

The following vectors are linear combination of the elements of S:

$$\begin{array}{rcrcrcrcr}
(2,2,1-\pi) &=& (1,2,1) &-& (-1,0,\pi) &+& 2(0,0,0) &+& 0(2,4,2);\\
(3,6,3) &=& 3(1,2,1) &+& 0(-1,0,\pi) &-& (0,0,0) &+& 0(2,4,2);\\
(-1,0,\pi) &=& 0(1,2,1) &+& (-1,0,\pi) &+& 2(0,0,0) &+& 0(2,4,2);\\
(0,0,0) &=& 0(1,2,1) &+& 0(-1,0,\pi) &+& 3(0,0,0) &+& 0(2,4,2).
\end{array}$$

On the contrary, the vector $(0,0,1)$ is not a linear combination of vectors of S as the equation

$$(0,0,1) = a_1(1,2,1) + a_2(-1,0,\pi) + a_3(0,0,0) + a_4(2,4,2)$$

is inconsistent. In fact, solving the above equation can be reduce to solve the system

$$\left\{\begin{array}{rcrcrcl}
a_1 &-& a_2 &+& 2a_4 &=& 0\\
2a_1 & & &+& 4a4 &=& 0\\
a_1 &+& \pi a_2 &+& 2a_4 &=& 1
\end{array}\right. \Leftrightarrow \left\{\begin{array}{rcrr}
a_2 &=& & 0\\
a_1 &=& -& 2a_4\\
0 &=& & 1
\end{array}\right.,$$

that is clearly impossible.

DEFINITION (Linear independence). A set $S = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\}$ of vectors in $\mathbb{R}^n$ is said to be *linearly independent* if the equation

$$\sum_{i=1}^{p} a_i\bar{x}_i = a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_p\bar{x}_p = \bar{0}$$

has the unique solution

$$a_1 = a_2 = \cdots = a_p = 0.$$

If S is not linearly independent then it is *linearly dependent*.

THEOREM 1.2. *A set $S = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\}$ of vectors in $\mathbb{R}^n$, with $p \geq 2$, is linearly dependent if and only if one of its vectors $\bar{x}_i$ is a linear combination of the remaining vectors in S.*

THEOREM 1.3. *Let $S = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\}$ be a set of vectors in $\mathbb{R}^n$.*

(1) If one of the vectors $\bar{x}_i \in S$ is the zero vector then S is linearly dependent.

(2) If one of the vectors $\bar{x}_i \in S$ is the product of a scalar by some other vector $\bar{x}_j \in S$ then S is linearly dependent.

THEOREM 1.4. *Let $S = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\}$ be a set of vectors in $\mathbb{R}^n$, and $\bar{x} \in \mathbb{R}^n$ a linear combination of the vectors in S*

$$\bar{x} = \sum_{i=1}^{p} a_i\bar{x}_i = a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_p\bar{x}_p.$$

The scalars $a_1, a_2, \ldots, a_p$ satisfying the equation are unique if and only if S is linearly independent.

Examples.

(a) Consider the set of vectors $S = \{(-2,0,1),(2,1,-1)\} \subseteq \mathbb{R}^3$. The set is linearly independent as the equation

$$a_1(-2,0,1) + a_2(2,1,-1) = (0,0,0)$$

has unique solution

$$\left\{\begin{array}{rcrcl} -2a_1 & + & 2a_2 & = & 0 \\ & + & a_2 & = & 0 \\ a_1 & - & a_2 & = & 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} a_1 & = & 0 \\ a_2 & = & 0 \end{array}\right. .$$

(b) Consider the set of vectors $T = \{(-2,2),(0,1),(1,-1)\} \subseteq \mathbb{R}^2$. The set is linearly dependent as the equation

$$a_1(-2,2) + a_2(0,1) + a_3(1,-1) = (0,0)$$

is equivalent to

$$\begin{cases} -2a_1 & & + & a_3 & = & 0 \\ 2a_1 & + & a_2 & - & a_3 & = & 0 \end{cases} \Leftrightarrow \begin{cases} a_2 & = & 0 \\ a_3 & = & 2a_1 \end{cases},$$

existing solutions other than the null solution $a_1 = a_2 = a_3 = 0$ (for example, $a_1 = 1$, $a_2 = 0$, $a_3 = 2$). Note that there exist vectors in T that are linear combination of the remaining vectors in T, such as

$$(-2, 2) = 0(0, 1) - 2(1, -1).$$

(c) Consider the set of vectors $U = \{(-2, 2, 5), (0, 0, 0), (1, -1, 2)\} \subseteq \mathbb{R}^3$. The set is linearly dependent as in contains the zero vector of $\mathbb{R}^3$.

(d) Consider the set of vectors $V = \{(-2, 2), (0, 1)\} \subseteq \mathbb{R}^2$. The vector $\bar{x} = (2, 2)$ is a linear combination of the vectors in V

$$a_1(-2, 2) + a_2(0, 1) = (2, 2).$$

In fact, the corresponding system

$$\begin{cases} -2a_1 & & & = & 2 \\ 2a_1 & + & a_2 & = & 2 \end{cases} \Leftrightarrow \begin{cases} a_1 & = & -1 \\ a_2 & = & 4 \end{cases}$$

is consistent. Note the values we obtained for a_1, a_2 are unique. Hence, V is linearly independent. This can be checked by observing that the equation

$$a_1(-2, 2) + a_2(0, 1) = (0, 0)$$

has the unique solution $a_1 = a_2 = 0$:

$$\begin{cases} -2a_1 & & & = & 0 \\ 2a_1 & + & a_2 & = & 0 \end{cases} \Leftrightarrow \begin{cases} a_1 & = & 0 \\ a_2 & = & 0 \end{cases}.$$

Exercises.

(1) Prove Theorem 1.1.

(2) Consider the vectors in $\mathbb{R}^3$

$$\bar{u} = (1, -1, -1), \quad \bar{v} = (1, 1, 2), \quad \bar{w} = (1, 0, 1), \quad \bar{x} = (0, 0, 0), \quad \bar{y} = (2, 2, 4), \quad \bar{z} = (-1, 0, 0).$$

a) Determine the dot product $\bar{u} \cdot \bar{v}$, and also the norms of the vectors $\bar{u}$ and $\bar{v}$.

b) Are the vectors $\bar{u}$ and $\bar{v}$ orthogonal? And the vectors $\bar{u}$ and $\bar{w}$? And $\bar{w}$ and $\bar{x}$?

c) Determine the angle between the vectors $\bar{v}$ and $\bar{y}$, and also between the vectors $\bar{w}$ and $\bar{z}$.

d) Determine the distances $\mathrm{d}(\bar{u}, \bar{x})$ and $\mathrm{d}(\bar{y}, \bar{z})$.

(3) Determine if the following sets of vectors are linearly independent:

a) $\{(3,1),(4,2)\} \subseteq \mathbb{R}^2$;

b) $\{(3,1),(4,-2),(7,2)\} \subseteq \mathbb{R}^2$;

c) $\{(0,-3,1),(2,4,1),(-2,8,5)\} \subseteq \mathbb{R}^3$;

d) $\{(0,3,1),(2,1,1),(4,2,2)\} \subseteq \mathbb{R}^3$;

e) $\{(0,3,1),(2,1,1),(0,0,0)\} \subseteq \mathbb{R}^3$;

f) $\{(0,3,1),(1,1,1),(-2,1,-1)\} \subseteq \mathbb{R}^3$;

g) $\{(-1,2,0,2),(5,0,1,1),(8,-6,1,-5)\} \subseteq \mathbb{R}^4$.

(4) Consider the vectors in $\mathbb{R}^3$

$$\bar{v} = (2,0,0), \quad \bar{w} = (1,1,1), \quad \bar{x} = (-1,0,0), \quad \bar{y} = (1,0,1), \quad \bar{z} = (0,1,1).$$

a) Write, if possible, $\bar{y}$ and $\bar{z}$ as linear combinations of the vectors $\bar{v}$, $\bar{w}$, and $\bar{x}$. What can you conclude about the linear independence of the set $\{\bar{v}, \bar{w}, \bar{x}\}$?

b) Write, if possible, $\bar{z}$ as a linear combination of the vectors $\bar{v}$ and $\bar{w}$, and draw the appropriate conclusions about the linear independence of the set $\{\bar{v}, \bar{w}\}$.

(5) Discuss the linear independence of the set $\{(1,-2),(\alpha,-1)\} \subseteq \mathbb{R}^2$ depending on the value of the real parameter α.

(6) Prove Theorems 1.2 to 1.4.

2. Matrices

Matrix algebra.

DEFINITION (Matrix). Given positive integers p and n, a *matrix A of size $p \times n$* or a *$p \times n$ matrix A* is a rectangular array of pn (real) numbers in a boxed display, consisting of p horizontal lines with n numbers each (the *rows* of the matrix) and n vertical lines with p numbers each (the *columns* of the matrix), written

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix}.$$

In the above display the first index gives the row and the second index the column, so that a_{ij} appears at the intersection of the i-th row and the j-th column. Sometimes, we express that matrix A is of size $p \times n$ by writing

$$A_{p \times n}.$$

When convenient, we write

$$A = [a_{ij}]_{p \times n},$$

and refer to a_{ij} as the *(i, j)-th element* or the *(i, j)-th entry* of the matrix.

Example. The matrix

$$A = \begin{bmatrix} -2 & 0 & 0 & \frac{1}{3} \\ 1 & -5 & 1 & 0 \\ \pi & -3 & 0 & -1 \end{bmatrix}$$

has size 3×4. The second row is the sequence of elements $(1, -5, 1, 0)$ and the first column is the sequence $(-2, 1, \pi)$. The element with row index 3 and column index 2 is $a_{32} = -3$.

DEFINITION (Diagonal elements). Given a matrix A of size $n \times n$, the elements $a_{ii},\ i = 1, 2, \ldots, n$, are called *diagonal elements*. The collection of the diagonal elements is called the *main diagonal* or simply the *diagonal*.

Example. Consider the 4×4 matrix

$$A = \begin{bmatrix} -2 & 0 & 0 & \frac{1}{3} \\ 1 & -5 & 1 & 0 \\ 11 & -3 & 0 & -1 \\ 0 & 3 & 1 & -1 \end{bmatrix}.$$

The diagonal elements are: $a_{11} = -2$, $a_{22} = -5$, $a_{33} = 0$ e $a_{44} = -1$. Jointly, these elements are the main diagonal of matrix A.

DEFINITION (Types of matrices).

(1) A matrix of size $p \times n$ is said to be a *square matrix* if $p = n$, and *rectangular* if $p \neq n$. A square matrix A of size $n \times n$ if often referred to as *square matrix of order* n denoted by $A(n)$ or A_n.

(2) A matrix of size $1 \times n$ is said to be a *row matrix*, and matrix of size $p \times 1$ a *column matrix*.

(3) A $p \times n$ matrix where all elements are zero is called the *null matrix* or the *zero matrix* of size $p \times n$, and is denoted by $0_{p \times n}$.

(4) A square matrix A is said to be *triangular* if all entries above or below the main diagonal are zero. In particular, A is called *lower triangular* if the entries above the diagonal are zero (that is if $a_{ij} = 0$ for $i < j$), and *upper triangular* if the entries below the diagonal are zero (that is if $a_{ij} = 0$ for $i > j$).

(5) A square matrix is called *diagonal* if all entries outside the main diagonal are zero, that is, if $a_{ij} = 0$ for $i \neq j$.

(6) A diagonal matrix whose diagonal elements all contain the same number is called *scalar*.

(7) A scalar matrix of order n whose diagonal elements are all equal to 1 is called the *identity matrix* of order n, and is denoted by $I(n)$ or I_n.

Example. Consider the matrices

$$A = \begin{bmatrix} -2 & 0 & 0 & \frac{1}{3} \\ 1 & -5 & 1 & 0 \\ -4 & -3 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -2 \\ 1 & 3 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 12 & 0 & 0 & -5 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix},$$

$$E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -5 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad J = \begin{bmatrix} 1 \end{bmatrix}.$$

A, B, C, and D are, respectively, a rectangular matrix of size 3×4, a square matrix of order 3, a row matrix of size 1×4, and a column matrix of size 3×1. E is the null matrix of size 2×3. F, G, and H are square matrices of order 3: F is upper triangular, G and H are diagonal, being H scalar. I e J, scalar matrices, are, respectively, the identity matrices of order 2 and 1.

DEFINITION (Equality of matrices). Given two matrices $A_{p\times n}$ and $B_{q\times m}$, we say that A and B are *equal* and write $A = B$ if

(1) $p = q$ and $n = m$, that is, matrices A and B have the same size;

(2) $a_{ij} = b_{ij}$, for all $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, n$, that is, the corresponding elements are equal.

Example. Consider the matrices

$$A = \begin{bmatrix} -2 & 0 \\ 1 & -5 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 3 \end{bmatrix}. \quad C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 3 \end{bmatrix}.$$

It is only interesting to check a possible equality for matrices of the same size: A, C, on one hand, and B, D, on the other hand. By simple inspection we see that $A \neq C$ and $B = D$.

DEFINITION (Addition of matrices). Given matrices A and B of size $p \times n$, the *sum* $A + B$ is a matrix of the same size, with elements $a_{ij} + b_{ij}$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, n$. This may be abbreviated as

$$A + B = [a_{ij} + b_{ij}]_{p\times n}.$$

Example. Given the matrices

$$A = \begin{bmatrix} -2 & 0 & 0 & 1 \\ 1 & -5 & 1 & 0 \\ 0 & -3 & 0 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 & 0 & -5 \\ 1 & 0 & -2 & 1 \\ -1 & 3 & 2 & 0 \end{bmatrix}.$$

we have

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

DEFINITION (Negative of a matrix). The *negative* (or the *additive inverse*) of a matrix A, of size $p \times n$, is a matrix of the same size, denoted by $-A$, whose elements are the negatives of the corresponding elements of A. In abbreviate form,

$$-A=[-a_{ij}]_{p\times n}.$$

Note that the *difference* of matrices A and B of the same size is defined as

$$A-B=A+(-B).$$

Example. Consider the matrices

$$A=\begin{bmatrix} -2 & 0 & 0 & 1 \\ 1 & -5 & 1 & 0 \\ 0 & -3 & 0 & -1 \end{bmatrix}, \qquad B=\begin{bmatrix} -2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

The negative of the matrix A is

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

We also have

$$B-A=B+(-A)=\begin{bmatrix} 0 & 0 & 0 & 0 \\ -1 & 5 & -1 & 0 \\ 1 & 4 & 1 & 2 \end{bmatrix}.$$

THEOREM 2.1. *Given matrices A, B, and C, of size $p \times n$, and denoting by 0 the null matrix of the same size,*

(1) $(A+B)+C=A+(B+C)$;

(2) $A+B=B+A$;

(3) $A+0=0+A=A$;

(4) $A+(-A)=(-A)+A=0$.

DEFINITION (Multiplication of a matrix by a scalar). Given a real number (or *scalar*) λ and a matrix A of size $p \times n$ the *product* λA (or $\lambda \cdot A$) is a matrix of size $p \times n$, whose elements are obtained by multiplying each element of A by λ. In abbreviate form,

$$\lambda A = [\lambda a_{ij}]_{p\times n}.$$

Example. Given the matrix

$$A = \begin{bmatrix} -2 & 0 \\ 1 & -5 \\ 0 & -3 \end{bmatrix},$$

we have

$$3A = \begin{bmatrix} -6 & 0 \\ 3 & -15 \\ 0 & -9 \end{bmatrix}, \quad 1A = \begin{bmatrix} -2 & 0 \\ 1 & -5 \\ 0 & -3 \end{bmatrix} = A, \quad -1A = \begin{bmatrix} 2 & 0 \\ -1 & 5 \\ 0 & 3 \end{bmatrix} = -A, \quad 0A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = 0.$$

THEOREM 2.2. *Given matrices A and B of size $p \times n$, and real numbers λ and μ, we have*

(1) $\lambda(A + B) = \lambda A + \lambda B$;

(2) $(\lambda + \mu)A = \lambda A + \mu A$;

(3) $\lambda(\mu A) = (\lambda\mu)A$;

(4) $1A = A$.

DEFINITION (Multiplication of matrices). Given matrices A of size $p \times n$, and B of size $n \times q$, the *product* AB (or $A \cdot B$) is a matrix of size $p \times q$ with elements

$$[AB]_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}, \qquad i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, q.$$

This may be abbreviated as

$$AB = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right]_{p\times q}.$$

Example. Given the matrices

$$A = \begin{bmatrix} -1 & 0 \\ 3 & 0 \\ 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} -2 & 0 & 1 & 0 \\ 1 & -5 & 0 & -1 \end{bmatrix},$$

of sizes 3×2 and 2×4, respectively, the product AB is the matrix of size 3×4

$$AB = \begin{bmatrix} 2 & 0 & -1 & 0 \\ -6 & 0 & 3 & 0 \\ 1 & -5 & 0 & -1 \end{bmatrix}.$$

THEOREM 2.3. *Let A and A^* be matrices of size $p \times n$, B and B^* matrices of size $n \times q$, C a matrix of size $q \times r$, and λ a real number. Then*

(1) $(AB)C = A(BC)$;

(2) $(A + A^*)B = AB + A^*B$;

(3) $A(B + B^*) = AB + AB^*$;

(4) $\lambda(AB) = (\lambda A)B = A(\lambda B)$;

(5) $I_pA = A = AI_n$.

The multiplication of matrices is not commutative, that is, given two matrices A and B, of the sizes $p \times n$ and $n \times q$, respectively, it cannot be assured that $AB = BA$ (even if both products exist and are of the same size).

DEFINITION (Commuting matrices). Given two matrices A and B, if $AB = BA$ the matrices A and B are said to *commute*.

Example. Consider the matrices

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} -2 & 0 & 1 \\ 1 & -5 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} -2 & 0 \\ 1 & -5 \\ 0 & 1 \end{bmatrix}.$$

The matrices A and E do not commute as, although the product EA exists, the product AE does not.
The products DE and ED both exist but the first is a 2×2 matrix whereas the second a 3×3 matrix, so that D and E do not commute.
The products AB and BA are both matrices of size 2×2 matrices, but distinct:

$$AB = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 9 & 6 \end{bmatrix} \neq BA = \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -3 & 6 \end{bmatrix}.$$

Therefore, the matrices A and B do not commute.
The matrices A and C commute:

$$AC = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} = CA.$$

DEFINITION (Exponentiation of matrices). Given a square matrix A of order n, and a positive integer k, the *power* A^k is the product of k factors of A

$$A^k = AA \cdots A.$$

Example. Consider the matrix

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

Then

$$A^2 = \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 3 & 7 \end{bmatrix}.$$

DEFINITION (Transposition). Given a matrix A of size $p \times n$, the *transpose of* A, denoted by A', is a matrix of size $n \times p$, with elements $a_{ji},\ j = 1, 2, \ldots, n,\ i = 1, 2, \ldots, p$. In abbreviate form,

$$A' = [a_{ji}]_{n \times p}.$$

Example. Consider the matrix of size 3×4

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

The transpose of A is the matrix obtained from A by "turning rows into columns and vice versa":

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

THEOREM 2.4. *Given matrices A and A^* of size $p \times n$, a matrix B of size $n \times q$, a square matrix C of order n, and a positive integer k,*

(1) $(A + A^*)' = A' + (A^*)'$;

(2) $(AB)' = B'A'$;

(3) $\left(C^k\right)' = (C')^k$;

(4) $I_n' = I_n$;

(5) $(A')' = A$;

(6) $(\lambda A)' = \lambda A'$.

DEFINITION (Symmetric and skew-symmetric matrices). A matrix A satisfying $A' = A$ is said to be *symmetric*. If $A' = -A$, A is called *skew-symmetric*.

It is clear that both symmetric and skew-symmetric matrices are necessarily square.

Example. Consider the matrices

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & -1 \\ 3 & -1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & -1 & -3 \\ 1 & 0 & 1 \\ 3 & -1 & 0 \end{bmatrix}.$$

Matrix A is clearly symmetric, and B skew-symmetric.

Invertible matrices.

DEFINITION (Invertible matrix). A matrix A is said to be *invertible* if there exists a matrix A^{-1}, called *inverse of* A, such that

$$AA^{-1} = A^{-1}A = I.$$

Obviously, if A is invertible then A and A^{-1} are square matrices of the same order.

THEOREM 2.5. *If a matrix A is invertible, its inverse A^{-1} is unique.*

Examples.

(a) The inverse B of matrix

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

if it exists, satisfies both the equalities $AB = I_2$ and $BA = I_2$. For the first equality we have

$$AB = I_2 \Leftrightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where a, b, c, and d are the elements to be determined of the matrix B. Multiplying the matrices on the left-hand side of the equality, we obtain

$$\begin{bmatrix} a & b \\ a+c & b+d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Solving the above equation amounts to solve the system of equations

$$\begin{cases} a &= 1 \\ b &= 0 \\ a+c &= 0 \\ b+d &= 1 \end{cases} \Leftrightarrow \begin{cases} a &= 1 \\ b &= 0 \\ c &= -1 \\ d &= 1 \end{cases}.$$

Using the obtained matrix

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

in the second equality, we can check that this equality is also satisfied

$$BA = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

We conclude that A is invertible and that its inverse is the matrix B obtained above: $A^{-1} = B$.

(b) Consider now the matrix

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

Following the same steps as in the previous example, we obtain

$$CD = I_2 \Leftrightarrow \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Leftrightarrow \begin{bmatrix} a+2c & b+2d \\ 2a+4c & 2b+4d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Writing the corresponding system of equations

$$\begin{cases} a+2c &= 1 \\ b+2d &= 0 \\ 2a+4c &= 0 \\ 2b+4d &= 1 \end{cases} \Leftrightarrow \begin{cases} a+2c &= 1 \\ b+2d &= 0 \\ a+2c &= 0 \\ b+2d &= \frac{1}{2} \end{cases},$$

we observe immediately that the system is inconsistent. Therefore, the matrix C is not invertible.

THEOREM 2.6. *Given square matrices A and B of order n, invertible, and a positive integer k, we have*

(1) $(AB)^{-1} = B^{-1}A^{-1}$;

(2) $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$;

(3) $(A')^{-1} = \left(A^{-1}\right)'$;

(4) $\left(A^{-1}\right)^{-1} = A$.

Example. We can use the properties of matrix algebra and of invertible matrices to solve in order to X the *matrix equation*

$$A + B'X' - X' = 2A - B'.$$

We assume that all the matrices involved have the proper sizes so that the relevant operations are allowed. We assume also that matrix $B' - I$ is invertible. We then have

$$\begin{aligned} A + B'X' - X' = 2A - B' \Leftrightarrow (B' - I)X' = A - B' &\Leftrightarrow X' = (B' - I)^{-1}(A - B') \\ &\Leftrightarrow X = ((B' - I)^{-1}(A - B'))' \\ &\Leftrightarrow X = (A - B')'((B' - I)^{-1})' \\ &\Leftrightarrow X = (A - B')'((B' - I)')^{-1} \\ &\Leftrightarrow X = (A' - B)(B - I)^{-1}. \end{aligned}$$

Rank.

Consider matrix A of size $p \times n$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix}.$$

We now interpret the rows of a matrix A as row matrices

$$A_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}, \quad i = 1, 2, \ldots, p,$$

and the columns as column matrices

$$a_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{pj} \end{bmatrix}, \quad j = 1, 2, \ldots, n.$$

Moreover, these row and column matrices can be viewed as vectors in $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively. Assuming this point of view, the notions and results presented in Chapter 1 for vectors, still hold for rows and columns of a matrix (in particular, the definitions and results concerning linear combination and dependency hold).

Examples.

(a) Consider the 2×3 matrix

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

The row matrix

$$\begin{bmatrix} 1 & 15 & -5 \end{bmatrix}$$

is not a linear combination of the rows of A, as there are not real numbers λ_1 and λ_2 such that

$$\begin{bmatrix} 1 & 15 & -5 \end{bmatrix} = \lambda_1 \begin{bmatrix} 2 & 0 & 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 & 3 & -1 \end{bmatrix}.$$

In fact, solving the above equation can be reduce to solve the system

$$\begin{cases} 2\lambda_1 & + & \lambda_2 & = & 1 \\ & & 3\lambda_2 & = & 15 \\ \lambda_1 & - & \lambda_2 & = & -5 \end{cases},$$

which is clearly inconsistent. But the column matrix

$$\begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

is a linear combination of the columns of matrix A as

$$\begin{bmatrix} 4 \\ 5 \end{bmatrix} = 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

(b) Consider the 2×3 matrix

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

The rows are linearly independent as the equation

$$\lambda_1 \begin{bmatrix} -2 & 0 & 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

has solution

$$\left\{ \begin{array}{rcrcl} -2\lambda_1 & + & 2\lambda_2 & = & 0 \\ & & \lambda_2 & = & 0 \\ \lambda_1 & - & 2\lambda_2 & = & 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} \lambda_1 & = & 0 \\ & & \\ \lambda_2 & = & 0 \end{array} \right. .$$

But the columns are dependent as

$$\lambda_1 \begin{bmatrix} -2 \\ 2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

is equivalent to

$$\left\{ \begin{array}{rcrcrcl} -2\lambda_1 & & & + & \lambda_3 & = & 0 \\ 2\lambda_1 & + & \lambda_2 & - & \lambda_3 & = & 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcr} \lambda_2 & = & 0 \\ \lambda_3 & = & 2\lambda_1 \end{array} \right. ,$$

existing solutions other than $\lambda_1 = \lambda_2 = \lambda_3 = 0$ (for example, $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = 2$). Notice that there exist columns that are linear combination of the remaining columns. For example,

$$\begin{bmatrix} -2 \\ 2 \end{bmatrix} = 0 \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} .$$

(c) Consider the matrix

$$C = \begin{bmatrix} -2 & 0 & 1 \\ 2 & 0 & -1 \\ 5 & 0 & 2 \end{bmatrix} .$$

As the second column is null, the columns are linearly dependent. The rows are also dependent as the first row equals the product of -1 and the second row: $C_1 = -1C_2$.

(d) Consider the 2×2 matrix

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

The column

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

is a linear combination of the columns of D since

$$\lambda_1 \begin{bmatrix} -2 \\ 2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

if and only if

$$\left\{\begin{array}{rcrcr} -2\lambda_1 & & & = & 2 \\ 2\lambda_1 & + & \lambda_2 & = & 2 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcr} \lambda_1 & = & -1 \\ \lambda_2 & = & 4 \end{array}\right. .$$

Notice the values we obtained for λ_1, λ_2 are unique. Hence, the columns of D are independent. In fact, the zero linear combination of the columns of D has the unique solution $\lambda_1 = \lambda_2 = 0$:

$$\lambda_1 \begin{bmatrix} -2 \\ 2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

if and only if

$$\left\{\begin{array}{rcrcr} -2\lambda_1 & & & = & 0 \\ 2\lambda_1 & + & \lambda_2 & = & 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcr} \lambda_1 & = & 0 \\ \lambda_2 & = & 0 \end{array}\right. .$$

DEFINITION (Row and column rank). Given a matrix A, the maximum number of linearly independent rows (columns) of A is called the *row (column) rank of* A.

THEOREM 2.7. *Given a matrix A, the row and column ranks of A are the same.*

DEFINITION (Rank). The common value of the row and column ranks of a matrix A is called the *rank of* A, denoted by $\mathrm{r}(A)$.

DEFINITION (Equivalent matrices). Two $p \times n$ matrices are said to be *equivalent* if they have the same rank.

We want to develop a procedure in order to determine the rank of any given matrix.

DEFINITION (Elementary row and column operations). Given a matrix, we designate *elementary row (column) operations*:

(1) Swapping positions of two rows (columns);

(2) Multiplying a row (column) by a number different from zero;

(3) Adding one row (column) to another.

DEFINITION (Elementary matrices). An *elementary matrix* of size $m \times m$ is a matrix obtained from the identity matrix I_m by applying to it a single elementary operation.

Example. The matrices below are the elementary matrices obtained from I_3 by applying, respectively, the elementary operations "interchanging the second and third columns", "multiplying the first row by 2", and "adding the third row to the first":

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

THEOREM 2.8. *An elementary row (column) operation on a $p \times n$ matrix A can be achieved by pre-multiplying (post-multiplying) A by the elementary matrix obtained from I_p (I_n) by applying to I_p (I_n) precisely the same row (column) operations.*

THEOREM 2.9. *If a matrix B is obtained from a matrix A by means of a finite sequence of elementary row or column operations then $\mathrm{r}(A) = \mathrm{r}(B)$ (and matrices A and B are equivalent).*

Example. Let

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

We want to consecutively apply to A the elementary operations "interchanging the first and third columns", "multiplying the first row by -1", and "adding the third column to the first". This can be achieved by pre- and post-multiplying A by suitable elementary matrices:

$$P_1AQ_1Q_2 = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & -1 \\ 3 & 0 & 0 \\ 0 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -3 & -1 \\ -1 & 1 & 0 \\ 3 & 0 & 3 \\ 1 & 2 & 0 \end{bmatrix} = B.$$

Alternatively, we can apply the elementary operations directly to matrix A:

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

By applying elementary operations to matrix A, no matter what is the algorithm used, we obtained an equivalent matrix, the matrix B.

DEFINITION (Upper triangular form). A $p \times n$ matrix A is said to be in *upper triangular form* if it writes

$$A = \begin{bmatrix} T_m & Q \\ 0 & 0 \end{bmatrix},$$

where T_m designates an upper triangular submatrix of order m ($m \leq \min(p, n)$) with nonzero diagonal elements, Q any submatrix, and 0 null submatrices (with Q and the submatrices 0 of the appropriate sizes).

Example. The matrices below are in upper triangular form:

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

is in upper triangular form.

THEOREM 2.10. *Any nonzero matrix can be reduced to upper triangular form by using elementary row and column operations.*

Note that to obtain an upper triangular form it suffices to use row operations, with possible interchanging of columns.

THEOREM 2.11. *The rank of a matrix in upper triangular form is the order m of the upper triangular submatrix T_m.*

Example. To determine the rank of the matrix

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

we reduce the matrix,

$$\begin{bmatrix} \textcircled{1} & 3 & 1 \\ 0 & 0 & -1 \\ 3 & 0 & 0 \\ 0 & 2 & 1 \end{bmatrix} \xrightarrow[r_3-3r_1]{} \begin{bmatrix} 1 & 3 & 1 \\ 0 & 0 & -1 \\ 0 & -9 & -3 \\ 0 & 2 & 1 \end{bmatrix} \xrightarrow[c_2 \leftrightarrow c_3]{} \begin{bmatrix} 1 & 1 & 3 \\ 0 & \textcircled{-1} & 0 \\ 0 & -3 & -9 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow[\substack{r_3-3r_2 \\ r_4+r_2}]{} \begin{bmatrix} 1 & 1 & 3 \\ 0 & -1 & 0 \\ 0 & 0 & -9 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\xrightarrow[-\frac{1}{9}r_3]{} \begin{bmatrix} 1 & 1 & 3 \\ 0 & -1 & 0 \\ 0 & 0 & \textcircled{1} \\ 0 & 0 & 2 \end{bmatrix} \xrightarrow[r_4-2r_3]{} \begin{bmatrix} 1 & 1 & 3 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

As the upper triangular submatrix we obtained is of order 3 we have that $\mathrm{r}(A) = 3$.

THEOREM 2.12. *A square matrix A of order n is invertible if and only if* $\mathrm{r}(A) = n$.

THEOREM 2.13. *Given a matrix A,* $\mathrm{r}(A') = \mathrm{r}(A)$.

Exercises.

(1) Consider the matrices

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -5 & 1 \\ -4 & -3 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & -2 & -5 \\ 1 & 3 & 2 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$E = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$H = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

a) Determine the type of each one of the above matrices.

b) Determine the negative of matrix B.

c) Determine the diagonal elements of matrix F.

d) With respect to the matrices G e H, determine the pairs of corresponding elements.

(2) Determine, if possible:

a) $\begin{bmatrix} 1 & 2 & -3 & 4 \\ 0 & -5 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 3 & -5 & 6 & 1 \\ 2 & 0 & -2 & -3 \end{bmatrix}$;

b) $\begin{bmatrix} 0 & 2 & 2 \\ 0 & -5 & 1 \end{bmatrix} - \begin{bmatrix} -3 & -5 \\ 1 & 0 \end{bmatrix}$;

c) $-2 \cdot \begin{bmatrix} 1 & -2 & 2 \\ 0 & -5 & \frac{3}{5} \end{bmatrix}$;

d) $\begin{bmatrix} 0 & 2 & -3 \end{bmatrix} + \begin{bmatrix} -3 & -5 & 0 \end{bmatrix}$.

(3) Consider the matrices

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -5 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -2 \\ 1 & 3 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$E = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad H = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Determine, if possible:

a) $0A - F + 2G$;

b) $C + 0D$;

c) $2(B + H)$;

d) $3(2C + E) - C$.

(4) Consider the matrices

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -5 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & -1 \\ 3 & 0 & 0 \end{bmatrix}.$$

Determine the matrix X that satisfies the equation

$$3(A - 2X) = (-B - X) + C.$$

(5) Determine, is possible, the products:

a) $\begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & -5 \\ 1 & 0 \end{bmatrix}$; b) $\begin{bmatrix} 0 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$;

c) $\begin{bmatrix} 1 & -2 & 2 \\ 0 & -5 & \frac{3}{5} \end{bmatrix} \cdot \begin{bmatrix} 1 & -2 & 2 \end{bmatrix}$; d) $\begin{bmatrix} 0 & 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$;

e) $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 2 & -2 \end{bmatrix}$.

(6) Let

$$A = \begin{bmatrix} 1 & 2 & -3 \\ 5 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -1 & 2 \\ 4 & 2 & 5 \\ 2 & 0 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 4 & 1 & 2 \\ 0 & 3 & 2 \\ 1 & -2 & 3 \end{bmatrix}.$$

Determine the matrices:

a) $A + B$;

b) $A - B$;

c) AB;

d) BA;

e) $(AB)C$.

(7) Consider the matrices

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -2 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix},$$

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

Determine, if possible, $C(A + 3IB)D$.

(8) Given the matrices A e B, determine under what conditions we have

$$(A+B)(A-B) = A^2 - B^2.$$

(9) Consider the matrices

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -5 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 0 & -1 \\ -1 & -1 & -3 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix},$$

$$E = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad F = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}.$$

a) Determine the transposes of the given matrices.

b) Which of the above matrices are symmetric? And skew-symmetric?

(10) Determine the values of the real parameter a for which the following matrix is symmetric

$$\begin{bmatrix} a & a^2-1 & -3 \\ a+1 & 2 & a^2+4 \\ -3 & 4a & -1 \end{bmatrix}$$

(11) Let A, B be square matrices of order n.

a) Show that AA' is symmetric.

b) Show that if the matrices A, B, and AB are symmetric then A and B commute.

(12) Consider the matrices

$$A = \begin{bmatrix} 3 & 0 \\ 2 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{3} & 0 \\ \frac{2}{3} & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & -3 \\ 2 & 1 & -3 \\ 2 & 2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} -1 & 1 & 0 \\ \frac{8}{7} & -1 & \frac{3}{7} \\ -\frac{2}{7} & 0 & \frac{1}{7} \end{bmatrix}.$$

Show that $A^{-1} = B$ and that $C^{-1} = D$.

(13) Consider the matrices

$$A = \begin{bmatrix} 2 & -1 & -1 \\ a & \frac{1}{4} & b \\ \frac{1}{8} & \frac{1}{8} & -\frac{1}{8} \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 6 \\ 1 & 3 & 2 \end{bmatrix}.$$

Determine the values of a and b such that $A = B^{-1}$.

(14) Consider the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}.$$

a) Check that $A^{-1} = B$. Without making any further computations, determine B^{-1}.

b) Check that $C^{-1} \neq D$.

c) Does D^{-1} exist?

(15) Consider the diagonal matrices

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

Check that $A^{-1} = B$. More generally, determine the inverse of a diagonal matrix whose diagonal contains no element equal to zero.

(16) Let A, B, C, D be square matrices of order n, with A, C invertible. Solve in order to X the equation

$$A(B + X)C = D.$$

(17) Let A, B, C, D be square matrices of the same order, with A, C invertible. Solve the matrix equation

$$A^{-1}(B + X')(2CA')' = D'C'.$$

(18) Prove Theorem 2.5.

(19) Consider the matrix

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

a) Determine the row linear combination $2A_1 - A_2 + 3A_3$.

b) Determine the column linear combination $a_1 + a_2 - 3a_4$.

c) Are the rows linearly independent? And the columns?

d) What is the rank of the matrix A?

(20) Consider the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 8 & 16 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} -2 & 1 & 3 \\ 1 & -1 & 1 \end{bmatrix},$$

$$D = \begin{bmatrix} 0 & -2 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad E = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & -4 & 7 \\ -1 & -2 & -1 & -2 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 2 & 4 & 0 & -1 \\ 1 & -1 & 2 & 2 \end{bmatrix}, \quad H = \begin{bmatrix} 0 & 3 & 0 & 0 \\ 2 & 4 & 0 & -1 \\ 0 & -1 & 2 & 2 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & 1 & 2 \\ -1 & 0 & 1 & -2 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$

$$J = \begin{bmatrix} 1 & -2 & -1 & 1 \\ 2 & 1 & 1 & 2 \\ -1 & 1 & -1 & -3 \\ -2 & -5 & -2 & 0 \end{bmatrix}.$$

Reduce the given matrices to upper triangular form and determine their ranks.

(21) Discuss the rank of the following matrices depending on the values of the relevant real parameters:

$$\text{a)} \quad \begin{bmatrix} x & 0 & x^2-2 \\ 0 & 1 & 1 \\ -1 & x & x-1 \end{bmatrix}; \quad \text{b)} \quad \begin{bmatrix} t+3 & 5 & 6 \\ -1 & t-3 & -6 \\ 1 & 1 & t+4 \end{bmatrix}; \quad \text{c)} \quad \begin{bmatrix} 1 & x & y & 0 \\ 0 & z & w & 1 \\ 1 & x & y & 1 \\ 0 & z & w & 1 \end{bmatrix}.$$

(22) Find an example, for the particular case of matrices of size 2×2, illustrating the fact that, in general, $\mathrm{r}(AB) \neq \mathrm{r}(BA)$.

3. Determinants

Definitions.

Consider the set P_n of all *permutations* of $\{1, 2, \ldots, n\}$, that is, the set of all one-to-one mappings of $\{1, 2, \ldots, n\}$ onto itself. It is useful to write a permutation $\sigma \in P_n$ as

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}.$$

If we assume that the elements $1, 2, \ldots, n$ in the domain of the permutation are always written in this order, the permutation σ may be represented by simply listing the images

$$(\sigma_1, \sigma_2, \ldots, \sigma_n),$$

where σ_i denotes $\sigma(i)$ for $i = 1, 2, \ldots, n$.
In a permutation $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ we say that the pair (σ_i, σ_j) is an *inversion* if $i < j$ and $\sigma_i > \sigma_j$. The number of inversions in σ is denoted by $I(\sigma)$, and σ is said to be *even* or *odd* depending on the parity of $I(\sigma)$.

DEFINITION (Signum of a permutation). For every permutation σ the *signum of* σ, ε_σ, is defined by

$$\varepsilon_\sigma = (-1)^{I(\sigma)} = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}.$$

Example. Consider the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 5 & 3 \end{pmatrix}$$

or, with a simpler notation,

$$(2, 4, 1, 5, 3).$$

The inversions in σ are $(2, 1)$, $(4, 1)$, $(4, 3)$, and $(5, 3)$. Their number is $I(\sigma) = 4$, σ is even, and its signum is

$$\varepsilon_\sigma = (-1)^{I(\sigma)} = (-1)^4 = 1.$$

DEFINITION (Elementary product). Given a square matrix A of order n, we call *elementary product* to any product of n of its elements choosing exactly one for each row with no repetition of columns

$$a_{1\sigma_1}a_{2\sigma_2}\cdots a_{n\sigma_n}.$$

Example. If A is a matrix of order 3 the following are elementary products

$$a_{11}a_{22}a_{33} \quad \text{and} \quad a_{13}a_{22}a_{31}.$$

Let us assign to each elementary product in A_n the signum of the permutation of the column indices

$$\varepsilon_\sigma a_{1\sigma_1}a_{2\sigma_2}\cdots a_{n\sigma_n}.$$

The sum of all such products is called the *determinant* of matrix A.

DEFINITION (Determinant). Let A be a square matrix of order n. The *determinant of* A, denoted by $|A|$ or $\det(A)$ is the real number defined by

$$|A| = \sum_{\sigma\in P_n} \varepsilon_\sigma a_{1\sigma_1}a_{2\sigma_2}\cdots a_{n\sigma_n}.$$

In the particular, for the cases where $n = 1, 2, 3$ we have:

(1) If $n = 1$,

$$|A| = |a_{11}| = a_{11};$$

(2) If $n = 2$,

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21};$$

(3) If $n = 3$,

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \begin{aligned} = \; & a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} \\ & -a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}. \end{aligned}$$

Example. Let us compute the determinants of the matrices

$$A = \begin{bmatrix} -1 \end{bmatrix}, \qquad B = \begin{bmatrix} -1 & -3 \\ 4 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 1 & 1 & -1 \end{bmatrix}.$$

We have

$$|A| = |-1| = -1, \qquad |B| = \begin{vmatrix} -1 & -3 \\ 4 & 2 \end{vmatrix} = -1 \cdot 2 - (-3 \cdot 4) = 10,$$

e

$$|C| = \begin{vmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 1 & 1 & -1 \end{vmatrix} \quad \begin{aligned} = \quad & 1 \cdot 1 \cdot (-1) + 3 \cdot 0 \cdot 1 + 1 \cdot 0 \cdot 2 \\ & -3 \cdot 1 \cdot 1 - 1 \cdot 2 \cdot 1 - (-1) \cdot 0 \cdot 0 = -6. \end{aligned}$$

Properties of the determinants.

THEOREM 3.1. *Let A and B be square matrices of order n, α a real number, and p a natural number. The following holds:*

(1) If A has a null row (column) or two equal or proportional rows (columns) or if, more generally, a row (column) is a linear combination of the remaining rows (columns) then $|A| = 0$;

(2) If A is a (upper or lower) triangular matrix then the determinant $|A|$ is the product of the diagonal elements of A;

(3) If row i in matrix A is written

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} = \begin{bmatrix} a'_{i1} + a''_{i1} & a'_{i2} + a''_{i2} & \cdots & a'_{in} + a''_{in} \end{bmatrix}$$

then

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a'_{i1} & a'_{i2} & \cdots & a'_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a''_{i1} & a''_{i2} & \cdots & a''_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix};$$

Mutatis mutantis the same holds for columns;

(4) $|A'| = |A|$;

(5) $|\alpha A| = \alpha^n |A|$;

(6) $|AB| = |A| \cdot |B|$;

(7) $|A^p| = |A|^p$.

Examples. Consider the matrices

$$A = \begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & -1 \\ 0 & 0 & 5 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}.$$

(a) As the columns of matrix A are proportional, we have

$$|A| = \begin{vmatrix} -1 & 2 \\ 3 & -6 \end{vmatrix} = 0.$$

(b) As matrix B is (upper) triangular, we have

$$|B| = \begin{vmatrix} 1 & 0 & 2 \\ 0 & -1 & -1 \\ 0 & 0 & 5 \end{vmatrix} = 1 \cdot (-1) \cdot 5 = -5.$$

(c) Writing the second column of matrix A as the sum of two rows we obtain

$$|A| = \begin{vmatrix} -1 & 1+1 \\ 3 & -5-1 \end{vmatrix} = \begin{vmatrix} -1 & 1 \\ 3 & -5 \end{vmatrix} + \begin{vmatrix} -1 & 1 \\ 3 & -1 \end{vmatrix} = 2 - 2 = 0.$$

(d) Noting that

$$|C| = \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = 1 \cdot 1 - (-1 \cdot 1) = 2$$

we have

$$|C'| = \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} = 1 \cdot 1 - (1 \cdot (-1)) = 2 = |C|.$$

(e) $|3C| = \begin{vmatrix} 3 & -3 \\ 3 & 3 \end{vmatrix} = 3 \cdot 3 - (-3 \cdot 3) = 18 = 3^2|C|.$

(f) Noting that $|C| = 2$ and that

$$|D| = \begin{vmatrix} 1 & 0 \\ 3 & -1 \end{vmatrix} = 1 \cdot (-1) - (0 \cdot 3) = -1$$

we have $|CD| = \begin{vmatrix} -2 & 1 \\ 4 & -1 \end{vmatrix} = -2 \cdot (-1) - (1 \cdot 4) = -2 = |C| \cdot |D|.$

(g) $|C^2| = \begin{vmatrix} 0 & -2 \\ 2 & 0 \end{vmatrix} = 0 \cdot 0 - (-2 \cdot 2) = 4 = |C|^2.$

THEOREM 3.2. *Let A be a square matrix of order n. The following statements are equivalent:*

(1) A *is invertible;*

(2) $|A| \neq 0$;

(3) $\mathrm{r}(A) = n$.

Example. Consider the matrices

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

The determinants of A and B are, respectively,

$$|A| = \begin{vmatrix} 0 & -1 \\ 1 & 1 \end{vmatrix} = 0 \cdot 1 - (-1 \cdot 1) = 1 \qquad \text{and} \qquad |B| = \begin{vmatrix} -1 & 1 \\ 1 & -1 \end{vmatrix} = -1 \cdot (-1) - (1 \cdot 1) = 0.$$

As $|A| = 1 \neq 0$ we conclude that $\mathrm{r}(A) = 2$ (and A is invertible). Also, as $|B| = 0 \neq 0$ we have that $\mathrm{r}(B) < 2$ (and B is not invertible).

THEOREM 3.3. *Let A be a square matrix of order n. If A is invertible then*

$$\left|A^{-1}\right| = |A|^{-1} = \frac{1}{|A|}.$$

Example. Consider the matrix

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

Note that $|A| = 2 \neq 0$ so that A is invertible. It can be seen that the inverse of matrix A is

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

and then

$$\left|A^{-1}\right| = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix} = \frac{1}{2} \cdot \frac{1}{2} - \left(\frac{1}{2} \cdot \left(-\frac{1}{2}\right)\right) = \frac{1}{2} = \frac{1}{|A|}.$$

Next result establishes what is the effect on the determinant $|A|$ when elementary row or column operations are applied to matrix A.

THEOREM 3.4. *Let A be a square matrix of order n. The following holds:*

(1) If B is obtained from A by swapping positions of two rows (columns) then $|B| = -|A|$;

(2) If B is obtained from A by multiplying a row (column) by a real number λ then $|B| = \lambda|A|$;

(3) If B is obtained from A by adding to one row (column) the product of another row (column) by a real number then $|B| = |A|$.

Since a (non-null) square matrix can be always reduced to an upper triangular matrix by applying elementary row and column operations, we have just obtained a simple method to compute a general determinant (recall that the determinant of an upper triangular matrix is just the product of its diagonal elements).

Example. Let us compute the determinant of matrix

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

We have

$$\begin{vmatrix} \textcircled{1} & 3 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 3 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{vmatrix} \underset{r_3-3r_1}{=} \begin{vmatrix} 1 & 3 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & -9 & -3 & 1 \\ 0 & 2 & 1 & 0 \end{vmatrix} \underset{c_2 \leftrightarrow c_4}{=} - \begin{vmatrix} 1 & 0 & 1 & 3 \\ 0 & \textcircled{1} & -1 & 0 \\ 0 & 1 & -3 & -9 \\ 0 & 0 & 1 & 2 \end{vmatrix}$$

$$\underset{r_3-r_2}{=} - \begin{vmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -2 & -9 \\ 0 & 0 & 1 & 2 \end{vmatrix} \underset{\frac{1}{2}r_3}{=} -2 \begin{vmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & \textcircled{-1} & -\frac{9}{2} \\ 0 & 0 & 1 & 2 \end{vmatrix}$$

$$\underset{r_4+r_3}{=} -2 \begin{vmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & -\frac{9}{2} \\ 0 & 0 & 0 & -\frac{5}{2} \end{vmatrix} = -5.$$

Laplace's expansion.

DEFINITION (Minor and cofactor). Let A be a square matrix of order n.

(1) The *minor of A associated to the element a_{ij}* or the (i,j) *minor of A*, denoted by M_{ij}, is the determinant of the $(n-1)\times(n-1)$ matrix that results from deleting the i-th row and the j-th column of matrix A.

(2) The *cofactor of A associated to the element a_{ij}* or the (i,j) *cofactor of A* is the number

$$A_{ij} = (-1)^{i+j} M_{ij}.$$

Example. Let

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

We have, for example,

$$M_{12} = \begin{vmatrix} 0 & -2 \\ -2 & 1 \end{vmatrix} = 0\cdot 1 - (-2\cdot(-2)) = -4$$

and

$$A_{12} = (-1)^{1+2}\cdot M_{12} = (-1)^{1+2}\cdot \begin{vmatrix} 0 & -2 \\ -2 & 1 \end{vmatrix} = (-1)\cdot(-4) = 4.$$

THEOREM 3.5 (Laplace's expansion). *Let A be a square matrix of order n. The determinant of A is the weighted sum of the elements of any row or column, with the weights being the corresponding cofactors:*

(1) $|A| = \sum_{k=1}^{n} a_{ik}A_{ik} = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}$ *(expansion along row i),*

(2) $|A| = \sum_{k=1}^{n} a_{kj}A_{kj} = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj}$ *(expansion along column j).*

Example. Consider the matrix

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

Let us proceed to the expansion of $|A|$ along, for example, the second column:

$$|A| = 1(-1)^{1+2}\cdot \begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} + 2(-1)^{2+2}\cdot \begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} + 1(-1)^{3+2}\cdot \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = -1.$$

Adjugate and inverse of a square matrix.

DEFINITION (Cofactor and adjugate matrices). Let A be a square matrix of order n.

(1) The *cofactor matrix of* A is the matrix $\hat{A} = [A_{ij}]$ whose elements are the cofactors of the corresponding elements of A.

(2) The transpose of $\hat{A}$ is called the *adjugate matrix of* A or the *classical adjoint of* A, and is denoted by $\operatorname{adj} A$. That is

$$\operatorname{adj} A = \hat{A}'.$$

THEOREM 3.6 (Inverse of a square matrix). *Let A be a square matrix of order n, invertible. Then*

$$A^{-1} = \frac{1}{|A|} \cdot \operatorname{adj} A.$$

Example. Consider the matrix

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

To determine the cofactor matrix $\hat{A}$ and then the adjugate matrix $\operatorname{adj} A$ we have to determine the cofactors of each element of A:

$$A_{11} = (-1)^{1+1} \cdot 2 = 2, \quad A_{12} = (-1)^{1+2} \cdot (-1) = 1, \quad A_{21} = (-1)^{2+1} \cdot 1 = -1, \quad A_{22} = (-1)^{2+2} \cdot 1 = 1.$$

Hence,

$$\hat{A} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \qquad \text{and} \qquad \operatorname{adj} A = \hat{A}' = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}.$$

As $|A| = 3$, we have

$$A^{-1} = \frac{1}{|A|} \cdot \operatorname{adj} A = \frac{1}{3} \cdot \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Exercises.

(1) Compute the following determinants (a, b, c, and d are real numbers):

a) $\begin{vmatrix} -2 & 4 \\ 1 & -2 \end{vmatrix}$ b) $\begin{vmatrix} 2 & -2 \\ 1 & -1 \end{vmatrix}$ c) $\begin{vmatrix} 2 & 0 \\ 1 & -5 \end{vmatrix}$

d) $\begin{vmatrix} 2 & 0 \\ 1 & 0 \end{vmatrix}$ e) $\begin{vmatrix} 2 & -1 \\ 1 & -2 \end{vmatrix}$ f) $\begin{vmatrix} 2 & 1 & -1 \\ 1 & 2 & -2 \\ 2 & 1 & -1 \end{vmatrix}$

g) $\begin{vmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 2 & 1 & 3 \end{vmatrix}$ h) $\begin{vmatrix} 3 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 2 & 1 \end{vmatrix}$ i) $\begin{vmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{vmatrix}$

j) $\begin{vmatrix} 1 & 3 & -2 \\ 2 & 0 & 2 \\ 1 & 1 & 1 \end{vmatrix}$ k) $\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ \frac{3}{2} & \frac{1}{2} & 0 & 3 \end{vmatrix}$ l) $\begin{vmatrix} 0 & 1 & 1 & -1 & 0 \\ 0 & -1 & 2 & -1 & 1 \\ 0 & 0 & 2 & 1 & 10 \\ 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 3 \end{vmatrix}$

m) $\begin{vmatrix} 0 & 0 & a \\ 0 & b & 0 \\ c & 0 & 0 \end{vmatrix}$ n) $\begin{vmatrix} 0 & 0 & 0 & a \\ 0 & 0 & b & 0 \\ 0 & c & 0 & 0 \\ d & 0 & 0 & 0 \end{vmatrix}$ 0) $\begin{vmatrix} 2 & 0 & 3 & -1 \\ 0 & 4 & 0 & 0 \\ 0 & 1 & -1 & 2 \\ 3 & 2 & 5 & -3 \end{vmatrix}$

p) $\begin{vmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 4 & 0 & 3 & 4 \\ 6 & 2 & 3 & 1 & 2 \end{vmatrix}$

(2) Compute the following determinants by using Laplace's expansion (a, b, c, and d are real numbers):

a) $\begin{vmatrix} 1 & 2 & 1 & 3 \\ 2 & 1 & 3 & 0 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 4 & 2 \end{vmatrix}$ b) $\begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{vmatrix}$ c) $\begin{vmatrix} a & 3 & 0 & 5 \\ 0 & b & 0 & 2 \\ 1 & 2 & c & 3 \\ 0 & 0 & 0 & d \end{vmatrix}$

(3) Show that the following determinants are zero (a, b, and c are real numbers):

a) $\begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 6 & 8 \end{vmatrix}$ b) $\begin{vmatrix} 1 & a & b+c \\ 1 & b & c+a \\ 1 & c & a+b \end{vmatrix}$ c) $\begin{vmatrix} a-b & a-b & a^2-b^2 \\ 1 & 1 & a+b \\ b & 1 & a \end{vmatrix}$

(4) Show that the following equalities are satisfied (x_i, y_i, for $i = 1, 2, 3$, are real numbers):

a) $\begin{vmatrix} 1 & x_1 & x_2 \\ 1 & y_1 & x_2 \\ 1 & y_1 & y_2 \end{vmatrix} = (y_1 - x_1)(y_2 - x_2)$

b) $\begin{vmatrix} 1 & x_1 & x_2 & x_3 \\ 1 & y_1 & x_2 & x_3 \\ 1 & y_1 & y_2 & x_3 \\ 1 & y_1 & y_2 & y_3 \end{vmatrix} = (y_1 - x_1)(y_2 - x_2)(y_3 - x_3)$

(5) Consider the matrices

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 5 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 1 & 1 \end{bmatrix}$$

Compute $|A'B|$.

(6) Let

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

Compute the determinant of the matrix $C = (BA)'$.

(7) Determine the adjugate and, if possible, the inverse of each one of the following matrices:

$$A = \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -2 \\ 1 & -2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 3 \end{bmatrix}, \quad D = \begin{bmatrix} -4 & -3 & -3 \\ 1 & 0 & 1 \\ 4 & 4 & 3 \end{bmatrix},$$

$$E = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 2 \\ 2 & 0 & 3 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 0 \\ 2 & 0 & 2 \end{bmatrix}, \quad G = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(8) Consider the matrix

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

Find the values of the real parameter a such that matrix A is invertible. For $a = 1$ determine A^{-1}.

4. Systems of linear equations

Matrix form and solvability.

Let us consider the *system of linear equations* (or the *linear system*) with p equations and n unknowns

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots \\ a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pn}x_n = b_p \end{cases},$$

where the a_{ij}, $i = 1, \ldots, p$, $j = 1, \ldots, n$, are the system *coefficients*, the b_i, $i = 1, \ldots, p$, the *free terms*, and the x_j, $j = 1, \ldots, n$, the *variables* (or unknowns).
The system can be represented by the *matrix equation*

$$Ax = b,$$

where A is the *coefficient matrix* of the system (of size $p \times n$)

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \ldots & a_{pn} \end{bmatrix},$$

x is the *column matrix of the unknowns* (of size $n \times 1$), and b the *column matrix of the free terms* (of size $p \times 1$)

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}.$$

We call a *solution* to the system to any vector in $\mathbb{R}^n$

$$\left(x_1^0, x_2^0, \ldots, x_n^0\right)$$

satisfying the system. The set of all solutions is called the *general solution* to the system.

Example. Consider the system

$$\left\{ \begin{array}{rcrcr} x & - & 2y & = & 1 \\ -2x & + & 4y & = & -2 \end{array} \right. .$$

We see that the vector $(x, y) = (1, 0)$ solves the system. Therefore $(1, 0)$ is a (particular) solution to the system.

To find the general solution we use the usual equivalence rules for systems of equations

$$\left\{ \begin{array}{rcrcr} x & - & 2y & = & 1 \\ -2x & + & 4y & = & -2 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcrcr} x & & & = & 2y+1 \\ -2x & + & 4y & = & -2 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcr} x & = & 2y+1 \\ -2 & = & -2 \end{array} \right.$$

and obtain

$$(x, y) = (2y + 1, y) \quad \text{for all} \quad y \in \mathbb{R}^n.$$

Note that the particular solution $(1, 0)$ above is obtained from the general solution by making $y = 0$.

A linear system is said to be *consistent* if it has, at least one solution. Otherwise, that is, if there is no solution to the system, the system is called *inconsistent*.

Moreover, a consistent system is said to be *independent* if its solution is unique and *dependent* if the system admits more than one solution.

Notice that the system can also be written

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{pn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}.$$

This last representation suggests the following interpretation:

- The system is solvable if and only if the column b is a linear combination of the columns of A;
- The system has a unique solution if and only if the column b writes as a unique linear combination of the columns of A.

The above conditions can be easily checked by computing the ranks of appropriate matrices.

THEOREM 4.1. *Let $Ax = b$ be a system of linear equations, and $[A|b]$ its so-called augmented matrix*

$$[A|b] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} & b_p \end{array}\right].$$

(1) If $\mathrm{r}(A) = \mathrm{r}([A|b])$, the system is consistent. Otherwise, that is, if $\mathrm{r}(A) < \mathrm{r}([A|b])$, the system is inconsistent.

(2) If the system is consistent and also $\mathrm{r}(A) = n$, then it is independent. Otherwise, that is, if $\mathrm{r}(A) < n$, the system is dependent (and $d = n - \mathrm{r}(A)$ is the number of degrees of freedom of the system).

The following scheme summarizes the above:

$$\begin{cases} \mathrm{r}(A) = \mathrm{r}([A|B]) \quad \text{— consistent system} \begin{cases} \mathrm{r}(A) = n \quad \text{— independent system} \\ \mathrm{r}(A) < n \quad \text{— dependent system (with } d = n - \mathrm{r}(A) \\ \qquad\qquad\quad \text{the number of degrees of freedom)} \end{cases} \\ \mathrm{r}(A) < \mathrm{r}([A|B]) \quad \text{— inconsistent system} \end{cases}$$

DEFINITION (Homogeneous system). If in a system $Ax = b$ we have $b = 0$, the system is called *homogeneous*. Otherwise, that is, if $b \neq 0$, the system is called *nonhomogeneous* or *inhomogeneous*.

THEOREM 4.2. *A homogeneous system is always consistent.*

Gaussian elimination.

Classifying and solving linear systems.

Let $[A|b]$ be the augmented matrix representing a linear system.

(1) *Gaussian elimination*: By using elementary row operations, with possible interchanging of columns of A, we reduce matrix $[A|b]$ to a rank-equivalent matrix $[A^*|b^*]$, where A^* is in upper triangular form (this procedure is nothing else than the well-known *method of elimination of variables* for solving linear systems).

(2) We evaluate the consistency of the system by checking if $\mathrm{r}(A) = \mathrm{r}([A|b])$.

(3) If the system is consistent, the independence is evaluated by checking if $\mathrm{r}(A) = n$.

(4) Finally, in the case the system is consistent, we carry on further Gaussian elimination on matrix $[A^*|b^*]$ in order to obtain an identity matrix in the place of the upper triangular submatrix of A^*. The system's solution can then be immediately found.

Examples.

(a) Consider the system

$$\begin{cases} x + 2y = 1 \\ 3x + y + z = 0 \\ y + z = -3 \end{cases}.$$

We construct the system's augmented matrix $[A|b]$ and, by Gaussian elimination,

$$\left[\begin{array}{ccc|c} \textcircled{1} & 2 & 0 & 1 \\ 3 & 1 & 1 & 0 \\ 0 & 1 & 1 & -3 \end{array}\right] \xrightarrow[r_2-3r_1]{} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & -5 & 1 & -3 \\ 0 & 1 & 1 & -3 \end{array}\right] \xrightarrow[r_2\leftrightarrow r_3]{} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & \textcircled{1} & 1 & -3 \\ 0 & -5 & 1 & -3 \end{array}\right]$$

$$\xrightarrow[r_3+5r_2]{} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & 6 & -18 \end{array}\right].$$

As $\mathrm{r}(A) = \mathrm{r}([A|b]) = n = 3$, we have that the system is consistent and independent. To obtain the solution, we carry on Gaussian elimination until the upper triangular submatrix is transformed into an identity matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & 6 & -18 \end{array}\right] \xrightarrow[\frac{1}{6}r_3]{} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & \textcircled{1} & -3 \end{array}\right] \xrightarrow[r_2-r_3]{} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & \textcircled{1} & 0 & 0 \\ 0 & 0 & 1 & -3 \end{array}\right] \xrightarrow[r_1-2r_2]{} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -3 \end{array}\right].$$

The solution is $(x, y, z) = (1, 0, -3)$.

(b) Consider the linear system

$$\begin{cases} x + 2y + 3z = 3 \\ \quad\quad y + 2z = 2 \end{cases},$$

with augmented matrix

$$[A|b] = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 3 \\ 0 & 1 & 2 & 2 \end{array}\right].$$

The matrix A is in an upper triangular form, and $\mathrm{r}(A) = \mathrm{r}([A|b]) = 2$, so that the system is consistent. As $n = 3 > 2 = \mathrm{r}(A)$, the system is dependent with $d = n - \mathrm{r}(A) = 3 - 2 = 1$ degrees of freedom. In order to obtain the solution we carry on further Gaussian elimination

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 3 \\ 0 & \textcircled{1} & 2 & 2 \end{array}\right] \xrightarrow[r_1 - 2r_2]{} \left[\begin{array}{ccc|c} 1 & 0 & -1 & -1 \\ 0 & 1 & 2 & 2 \end{array}\right].$$

We finally obtain

$$\begin{cases} x \quad\quad - z = -1 \\ \quad y + 2z = 2 \end{cases} \Leftrightarrow \begin{cases} x = z - 1 \\ y = -2z + 2 \end{cases},$$

and the solution to the system is $(x, y, z) = (z - 1, -2z + 2, z)$, with z taking any real value.

Application: invertibility of a matrix and determining the inverse.

We know that a square matrix A_n is invertible if the matrix equation

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

is solvable, that is, if there are real numbers x_{ij}, with $i, j = 1, 2, \dots, p$, satisfying the equation.
Also, evaluating the solvability of and solving, if possible, the equation amounts to evaluating the solvability of and solving the uncoupled matrix equations

$$\left\{\begin{array}{l}
\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \\
\qquad\qquad \vdots \qquad\qquad\quad \vdots \quad \vdots \quad \vdots \\
\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{nn} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}
\end{array}\right. .$$

As each one of the above equations has the same coefficient matrix, and noticing that the algorithm for Gaussian elimination does not depend on the free term column, the following procedure is available for evaluating the invertibility of a matrix and, if possible, for finding its inverse.

Invertibility of a matrix and determining the inverse

(1) The matrix A is augmented with the identity matrix of the same order: $[A|I]$.

(2) The matrix A is reduced to upper triangular form by using row elementary operations on matrix $[A|I]$.

(3) If $\mathrm{r}(A) = n$, matrix A is invertible. Otherwise, A has no inverse.

(4) In the case that A is invertible, we carry on further Gaussian elimination (still using only row operations) until matrix $[I|A^{-1}]$ is obtained.

Example. Consider the matrix

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

To determine if the matrix is invertible and, if it is, to compute its inverse A^{-1}, we use Gaussian elimination on matrix $[A|I]$ (performing exclusively elementary row operations)

$$\left[\begin{array}{cc|cc} \textcircled{1} & -1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right] \xrightarrow[r_2-r_1]{} \left[\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & 2 & -1 & 1 \end{array}\right] \xrightarrow[\frac{1}{2}r_2]{} \left[\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & \textcircled{1} & -\frac{1}{2} & \frac{1}{2} \end{array}\right] \xrightarrow[r_1+r_2]{} \left[\begin{array}{cc|cc} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} \end{array}\right].$$

We then have

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Cramer's rule.

DEFINITION (Cramer's system). A linear system $Ax = b$ is said to be a *Cramer's system* if

(1) A is a square matrix (that is, the number of equations equals the number of unknowns);

(2) $|A| \neq 0$.

Note that a Cramer's system is always consistent and independent.

The following theorem gives the Cramer's rule for solving linear systems.

THEOREM 4.3. *Given a Cramer's system, with matrix form*

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix},$$

the solution is given by

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & \dots & a_{1n} \\ b_2 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ b_n & a_{n2} & \dots & a_{nn} \end{vmatrix}}{|A|}, x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & \dots & a_{1n} \\ a_{21} & b_2 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & b_n & \dots & a_{nn} \end{vmatrix}}{|A|}, \dots, x_n = \frac{\begin{vmatrix} a_{11} & a_{12} & \dots & b_1 \\ a_{21} & a_{22} & \dots & b_2 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & b_n \end{vmatrix}}{|A|}.$$

Example. Let us consider the system

$$\begin{cases} x + y + z = 1 \\ x + 2y + 3z = 2 \\ 2x + y - z = 1 \end{cases}.$$

It is a Cramer's system, as the coefficient matrix A is a square matrix and

$$|A| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 2 & 1 & -1 \end{vmatrix} = -1 \neq 0.$$

The solution to the system is

$$x = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & 3 \\ 1 & 1 & -1 \end{vmatrix}}{-1} = 0, \qquad y = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 2 & 1 & -1 \end{vmatrix}}{-1} = 1, \qquad z = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \end{vmatrix}}{-1} = 0.$$

If the system $Ax = b$ is consistent but dependent, the Cramer's rule can still be adapted for finding the solution.

For this, we consider a square submatrix of A of the largest size such that its determinant is different from zero. The system is then rewritten in order to the variables corresponding to columns of the chosen submatrix.

Example. Let us consider the system

$$\begin{cases} x + 2y + z = 1 \\ -x + y + z = 2 \end{cases}.$$

The coefficient matrix

$$A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

has rank $\mathrm{r}(A) = 2$. The determinant of the submatrix containing the first two columns of A is different from zero:

$$\begin{vmatrix} 1 & 2 \\ -1 & 1 \end{vmatrix} = 3 \neq 0.$$

Rewriting the system, we obtain the Cramer's system

$$\begin{cases} x + 2y = 1 - z \\ -x + y = 2 - z \end{cases},$$

with solution

$$x = \frac{\begin{vmatrix} 1-z & 2 \\ 2-z & 1 \end{vmatrix}}{3} = -1 + \frac{1}{3}z, \qquad y = \frac{\begin{vmatrix} 1 & 1-z \\ -1 & 2-z \end{vmatrix}}{3} = 1 - \frac{2}{3}z,$$

with z taking any real value.

Exercises.

(1) Classify the systems of equations, and solve them, when possible:

a) $\begin{cases} x + y = 0 \\ 2x + 2y = 4 \\ 3x + 3y = 1 \end{cases}$

b) $\begin{cases} x + y = 1 \\ 2x + 2y = 2 \end{cases}$

c) $\begin{cases} x + y + z = 2 \\ 2x + 2y + 2z = 4 \\ 3x + 3y + 3z = 6 \end{cases}$

d) $\begin{cases} x + y + z = 1 \\ 2x + 2y + 2z = 4 \\ 3x + 3z = -1 \end{cases}$

e) $\begin{cases} x + y + z = 1 \\ x + y + 2z = -1 \\ x - 2y = 2 \end{cases}$

f) $\begin{cases} x + y + z = 0 \\ x + y + 2z = 0 \\ x - 2y = 0 \end{cases}$

g) $\begin{cases} x + y + z + w = 2 \\ x + 2z - w = 1 \\ x + 2y + 3w = 3 \end{cases}$

h) $\begin{cases} x + 2y + z = 0 \\ x - y + z = 1 \\ -x - z = 2 \end{cases}$

(2) Classify the systems:

a) $\begin{cases} -2x - 3y + z = 3 \\ 4x + 6y - 2z = 1 \end{cases}$

b) $\begin{cases} x - y + 2z = 1 \\ 2x + y - z = 3 \\ x + 5y - 8z = 1 \\ 4x + 5y - 7z = 7 \end{cases}$

c) $\begin{cases} x + y - z = 1 \\ x - y + z = 0 \\ x + 2y - z = 0 \\ x - y - 2z = 1 \end{cases}$

d) $\begin{cases} x - y + 2z = 1 \\ 3x + 2z = 2 \\ 2x + y = 1 \\ x + 8y - 10z = -2 \end{cases}$

(3) Discuss the solvability of the systems, depending on the values of the real parameters a, b, and c:

a) $\begin{cases} x + y + z = 3 \\ x - y + z = 1 \\ x - y - z = a \end{cases}$ b) $\begin{cases} x + y + z = 3 \\ x - y + z = 1 \\ 2x - 2y + az = 2 \end{cases}$

c) $\begin{cases} y + az = 0 \\ x + by = 0 \\ by + az = 1 \end{cases}$ d) $\begin{cases} x + y + z = 1 \\ x - y + 2z = a \\ 2x + bz = 2 \end{cases}$

e) $\begin{cases} 2x + y = b \\ 3x + 2y + z = 0 \\ x + ay + z = 2 \end{cases}$ f) $\begin{cases} ax + y + (a+1)z = b \\ x + ay + z = 1 \\ ax + y - z = 0 \end{cases}$

(4) Solve the matrix equation $AX = B$, where

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 2 \\ 2 & 0 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 0 \\ 1 & 2 & 3 \end{bmatrix}.$$

(5) Consider the system of equations in the matrix form

$$\begin{bmatrix} 1 & 1 & 0 \\ -1 & -2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Compute the inverse of the coefficient matrix and use it to solve the system.

(6) Solve, if possible, the following systems by using the Cramer's rule:

a) $$\begin{cases} x + 2y + z = 1 \\ -x + 2y = 0 \\ -x + y + z = 2 \end{cases}$$

b) $$\begin{cases} x + 2y - z = 1 \\ x + z = 2 \\ x + 2y - 3z = 0 \end{cases}$$

c) $$\begin{cases} x + 2y + z - u = 1 \\ 3x + 6y + 2z - 2u = 0 \end{cases}$$

d) $$\begin{cases} x + 2y - z = -5 \\ 2x - y + z = 6 \\ x - y - 3z = -3 \end{cases}$$

e) $$\begin{cases} x + y = 3 \\ x + z = 2 \\ y + z + u = 6 \\ y + u = 1 \end{cases}$$

(7) Show that the following system has a unique solution for all real values of b_1, b_2, b_3, and find that solution:

$$\begin{cases} 3x_1 + x_2 = b_1 \\ x_1 - x_2 + 2x_3 = b_2 \\ 2x_1 + 3x_2 - x_3 = b_3 \end{cases}$$

(8) Show that the homogeneous system

$$\begin{cases} ax + by + cz = 0 \\ bx + cy + az = 0 \\ cx + ay + bz = 0 \end{cases}$$

has nontrivial solutions if and only if $a^3 + b^3 + c^3 - 3abc = 0$.

(9) Discuss the solvability of the linear system

$$\begin{cases} x + 2y + 3z = 1 \\ -x + ay - 21z = 2 \\ 3x + 7y + az = b \end{cases}$$

depending on the real parameters a and b.

(10) Consider the system

$$\begin{cases} x_1 + x_2 + x_3 = 2q \\ 2x_1 - 3x_2 + 2x_3 = 4q \\ 3x_1 - 2x_2 + px_3 = q \end{cases},$$

where p and q are arbitrary real constants. Determine the values of p and q such that the system is

a) Consistent and independent;

b) Consistent and dependent;

c) Inconsistent.

(11) Show that the system

$$\begin{cases} 2x + 3y = k \\ x + cy = 1 \end{cases}$$

has a unique solution except for a particular value c^* of the parameter c. Determine that solution. Show also that for $c = c^*$ the system is inconsistent except for a particular value k^* of k. Find the solution for $k = k^*$.

(12) Determine the values of the real parameters a and b such that the linear system

$$\begin{cases} 3x - y - z = 0 \\ -x + 2y - z = 0 \\ -x - y + az = b - 3 \end{cases}$$

is consistent and dependent.

(13) Discuss the solvability of the following system depending on the real parameter α:

$$\begin{cases} x + 2y + z + w = 0 \\ 2x + 4y + 2z + 3w = 1 \\ x + 2y + z + 2w = \alpha \end{cases}$$

CALCULUS

5. The real number system

Basic concepts of set theory.

We briefly review basic concepts and language of set theory.
Informally, a *set* is a collection of objects viewed as a single entity. The individual objects in a set are called *elements* or *members* of the set, and are said to *belong* or to *be contained* in the set. The set is said to *contain* or to *be composed of* its elements.
Sets are usually denoted by upper case letters of the English alphabet, $A, B, C, \dots, X, Y, Z$ and their elements by lower case letters, $a, b, c, \dots, x, y, z$. To state that the object x belongs to the set S we write $x \in S$ and to sate that it does not we write $x \notin S$.
The simplest way to describe a set is to list its elements inside a pair of curly braces (the *roster notation* for a set). For example, the set containing the elements a, b, and c can be represented by

$$\{a, b, c\}.$$

Notice that the above representation is only possible if the set has a finite number of elements or in the case the set in infinite when it is clear from the context which are the elements omitted from the list. For example, the set of all positive integers can be written

$$\{1, 2, 3, \dots\}$$

if it not ambiguous what the three dots stand for.

DEFINITION (Equality). Two sets A and B are said to be *equal* if they are composed of exactly the same elements, and we write

$$A = B.$$

If A and B are not equal we write

$$A \neq B.$$

DEFINITION (Subset). A set A is said to be a *subset* of a set B, and we write

$$A \subseteq B,$$

if all elements of A also belong to B. A is also said to *be contained in* B or that B *contains* A. The relation $\subseteq$ is referred to as *set inclusion*. A is said to be a *proper subset* of B if $A \subseteq B$ and $A \neq B$. In this case we write $A \subset B$.

Note that $A \subseteq B$ can also be written $B \supseteq A$ and $A \subset B$, $B \supset A$.

THEOREM 5.1. *Two sets A and B satisfy $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.*

The set inclusion is transitive (A, B, and C designate arbitrary sets):

- Transitive law: If $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.

In the usual applications, it is clear that there is a definite set, called the *universal set*, to which all elements of interest belong.

Examples.

(a) If the variable of interest is firm profits, the universal set is $S = \mathbb{R}$.

(b) If our interest goes to the number of employees in a firm, the universal set is $S = \mathbb{N} \cup \{0\}$.

(c) If our problem concerns stock prices, then $S = [0, +\infty)$.

This suggests another way to define a set. If we set the universal set S, a set A can be defined as containing all elements in S satisfying a given property p. Symbolically,

$$A = \{x \,|\, x \in S \textit{ and } x \textit{ satisfies } p\},$$

or, if is is clear which is the universal set S,

$$A = \{x \,|\, x \textit{ satisfies } p\},$$

or even,

$$A = \{x \,|\, p(x)\}.$$

Examples.

(a) $A = \{x \,|\, x \in \mathbb{N} \textit{ and } x < 4\} = \{1, 2, 3\}$.

(b) $B = \{x \,|\, x \in \mathbb{R} \textit{ and } x < 4\} = (-\infty, 4)$.

(c) In the set of the real numbers,

$$C = \{x \,|\, x^3 = -8\} = \{-2\}.$$

One special set is the *empty set* or the *void set*, denoted by $\emptyset$, that is the set containing no elements. Also, if a set contains only one element it called *elementary set*. For example, the set C in the above example is an elementary set. On the other hand, if its universal set was the set $\mathbb{N}$ of the positive integers, the set C would then be the empty set: $C = \emptyset$.
We can use operations to construct new sets from given sets.

DEFINITION (Set operations). Let A and B be arbitrary subsets of the universal set S.

(1) The *union* of A and B, denoted by $A \cup B$, is the set of all elements of S which are in A, in B, or in both. Symbolically,
$$A \cup B = \{x \mid x \in A \textit{ or } x \in B\}.$$
(2) The *intersection* of A and B, denoted by $A \cap B$, is the set of all elements of S common to both A and B. Symbolically,
$$A \cap B = \{x \mid x \in A \textit{ and } x \in B\}.$$
(3) The *difference* $A - B$, or the *complement* of B *relative to* A, is the set of all elements of A which do not belong to B. Symbolically,
$$A - B = \{x \mid x \in A \textit{ and } x \notin B\}.$$
In particular, the *complement* of B, denoted by B^c, is the complement of B relative to the universal set S. Symbolically,
$$S - B = \{x \mid x \notin B\}.$$

If the intersection of two sets A and B is the empty set (symbolically, $A \cap B = \emptyset$), the sets are said to be *disjoint*.
We list next some well-known properties of the set operations (A, B, and C designate arbitrary subsets of the universal set S).

- Commutative laws: $A \cup B = B \cup A, \quad A \cap B = B \cap A.$
- Associative laws: $A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C.$
- Distributive laws: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$
- $A \cup A = A, \quad A \cap A = A.$
- $A \cup \emptyset = A, \quad A \cap \emptyset = \emptyset.$
- $A \cup S = S, \quad A \cap S = A.$

Axioms of the real number system. Basic results.

The set $\mathbb{R}$ of the real numbers is the set, together with *two operations* ($+$ and $\cdot$) and *an order relation* ($<$), whose elements satisfy a set of 10 axioms. We begin by setting the operations of *addition* and *multiplication*, respectively,

$$(x, y) \to x + y \qquad \text{and} \qquad (x, y) \to x \cdot y,$$

where the *sum* $x + y$ and the *product* xy are uniquely determined.
Note that $x \cdot y$ can also be written xy.
Let $a, b, c \ldots, x, y, z$ be arbitrary numbers. We first give a set of axioms which define the set $\mathbb{R}$ as a *field*.

FIELD AXIOMS.

Axiom 1 (Commutative laws). $x + y = y + x, \quad xy = yx;$

Axiom 2 (Associative laws). $x + (y + z) = (x + y) + z, \quad x(yz) = (xy)z;$

Axiom 3 (Distributive law). $x(y + z) = xy + xz;$

Axiom 4 (Existence of identity elements). There exist two distinct and unique numbers 0 and 1 such that

$$x + 0 = 0 + x = x \quad \text{and} \quad 1 \cdot x = x \cdot 1 = x;$$

Axiom 5 (Existence of negatives). For every x there exists a number $-x$ satisfying

$$x + (-x) = -x + x = 0;$$

Axiom 6 (Existence of reciprocals). For every $x \neq 0$ there exists a number x^{-1} such that

$$xx^{-1} = x^{-1}x = 1.$$

The properties stated in the next theorem can be obtained form the above axioms.
For example, from Axioms 1, 3, and 4 we obtain

$$0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$$

so that, using additionally Axiom 5,

$$0 \cdot a = 0 \cdot a + 0 \cdot a \Leftrightarrow 0 \cdot a + (-(0 \cdot a)) = 0 \cdot a \Leftrightarrow 0 = 0 \cdot a,$$

and, taking again into account Axiom 1, statement (5) below is proved.

THEOREM 5.2.

(1) *(Cancellation law of addition) If $a + b = a + c$ then $b = c$;*

(2) *(Possibility of subtraction) Given a and b, there is exactly one x such that $a + x = b$ (x is denoted by $b - a$; in particular, $0 - a$ is written $-a$ and is the negative of a);*

(3) $b - a = b + (-a)$*;*

(4) $-(-a) = a$*;*

(5) $0 \cdot a = a \cdot 0 = 0$*;*

(6) *(Cancellation law for multiplication) If $ab = ac$ and $a \neq 0$ then $b = c$;*

(7) *(Possibility of division) Given a and b with $a \neq 0$, there is exactly one x such that $ax = b$ (x is denoted by b/a or by $\frac{b}{a}$; in particular, $1/a$ is also written a^{-1} and is the reciprocal of a);*

(8) *If $a \neq 0$ then $b/a = b \cdot a^{-1}$;*

(9) *If $ab = 0$ then $a = 0$ or $b = 0$;*

(10) $\dfrac{a}{b} + \dfrac{c}{d} = \dfrac{ad + bc}{bd}$ *if $b, d \neq 0$;*

(11) $\dfrac{a}{b} \cdot \dfrac{c}{d} = \dfrac{ac}{bd}$ *if $b, d \neq 0$;*

(12) $\dfrac{a/b}{c/d} = \dfrac{ad}{bc}$ *if $b, c, d \neq 0$.*

Let us characterise the set $\mathbb{N}$ of the positive integers.

DEFINITION (Inductive set). A set of real numbers is called an *inductive set* if

(1) The number 1 is in the set;

(2) For every x in the set, $x + 1$ is also in the set.

DEFINITION (Positive integers). A real number is called a *positive integer* if it belongs to every inductive set.

We can then say that the set $\mathbb{N}$ of the positive integers is the smallest inductive set. Note that the set of the positive integers can also be denoted by $\mathbb{Z}^+$.

The set $\mathbb{Z}^-$ of the *negative integers* is the set of the negatives of the elements of $\mathbb{Z}^+$. Finally, the set $\mathbb{Z}$ of the *integers* is $\mathbb{Z} = \mathbb{Z}^+ \cup \mathbb{Z}^- \cup \{0\}$.

The set $\mathbb{Q}$ of the *rational numbers* is defined by

$$\left\{x \mid x = \frac{a}{b},\ a, b \in \mathbb{Z} \text{ and } b \neq 0\right\}.$$

Notice that the set $\mathbb{Q}$ is also a field.
Assume now that there exists a set $\mathbb{R}^+ \subseteq \mathbb{R}$, called the set of positive numbers, which satisfies a set of axioms which we give below. This set of axioms together with the first 6 axioms define the $\mathbb{R}$ as an *ordered field*.

ORDER AXIOMS.

Axiom 7. If $x, y \in \mathbb{R}^+$ then $x + y \in \mathbb{R}^+$ and $xy \in \mathbb{R}^+$;

Axiom 8. For every $x \neq 0$ either $x \in \mathbb{R}^+$ or $-x \in \mathbb{R}^+$ but not both;

Axiom 9. $0 \notin \mathbb{R}^+$.

We give meaning to the symbols $<$, $>$, $\leq$, and $\geq$:

- $x < y$ means that $y - x \in \mathbb{R}^+$;
- $y > x$ means that $x < y$;
- $x \leq y$ means that either $x < y$ or $x = y$;
- $y \geq x$ means that $x \leq y$.

Notice that

$$x > 0 \Leftrightarrow 0 < x \Leftrightarrow x - 0 \in \mathbb{R}^+ \Leftrightarrow x \in \mathbb{R}^+$$

so that $x > 0$ if and only if $x \in \mathbb{R}^+$. If $x < 0$ we say that x is negative.

THEOREM 5.3.

(1) (Trichotomy law) One and exactly one of the relations holds: $a < b$, $b < a$, $a = b$*;*

(2) (Transitive law) If $a < b$ *and* $b < c$ *then* $a < c$*;*

(3) If $a < b$ *then* $a + c < b + c$*;*

(4) If $a < b$ *and* $c > 0$ *then* $ac < bc$*;*

(5) If $a < b$ *and* $c < 0$ *then* $ac > bc$*;*

(6) If $a \neq 0$ *then* $a^2 > 0$*;*

(7) $1 > 0$*;*

(8) If $a < b$ *then* $-a > -b$*;*

(9) If $ab > 0$ *then either* $a, b > 0$ *or* $a, b < 0$*;*

(10) If $a < c$ *and* $b < d$ *then* $a + b < c + d$*.*

Let us prove, for example, (6) in the above theorem. If $a > 0$, from (4) in Theorem 5.3 and (5) in Theorem 5.2 we obtain

$$a > 0 \Rightarrow a \cdot a > 0 \cdot a \Leftrightarrow a^2 > 0$$

and if $a < 0$, from (5) in Theorem 5.3 and (5) in Theorem 5.2,

$$a < 0 \Rightarrow a \cdot a > 0 \cdot a \Leftrightarrow a^2 > 0,$$

so that (6) Theorem 5.3 in is proved.
Notice also that (7) in Theorem 5.3 is an immediate consequence of statement (6) of the same Theorem.
We can now set the geometric interpretation of the real numbers (see Figure 5.1). We draw a straight line and mark the point 0 freely. Then we mark the point 1 to the right of 0. The length of the segment with endpoints 0 and 1 determines the scale. In general, the points to the right of 0 represent the positive numbers and the points to the left of 0, the negative numbers. Also, given any two numbers x and y, if $x < y$ then y is marked to the right of x.

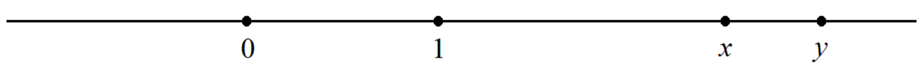


Figure 5.1

The set $\mathbb{Q}$ of the rational numbers satisfies the Axioms $1 - 9$ so that it is an ordered field. We will present a new axiom which the rational numbers do not satisfy: it separates the rational numbers from the real numbers. We first show that there are numbers which are not rational.
We say that an integer n is *even* if $n = 2m$, with m an integer. We say that $n + 1$ is an *odd* integer if n is even.

THEOREM 5.4. *Let m and n be integers.*

(1) If n^2 is even then n is even;

(2) If $m^2 = 2n^2$ then m and n are even.

Using the above theorem we can show that $\sqrt{2}$ is not rational. If is was, it could be written as a quotient a/b, where a and b are integers, $b \neq 0$, and at least one of the integers a and b is odd. Then

$$\sqrt{2} = \frac{a}{b} \Leftrightarrow 2 = \frac{a^2}{b^2} \Leftrightarrow a^2 = 2b^2$$

and from (2) in Theorem 5.4 we conclude that both a and b are even, which is a contradiction. Consequently, $\sqrt{2}$ is not rational: it is *irrational*. We proved the result stated next.

THEOREM 5.5. *$\sqrt{2}$ is an irrational number.*

Thus, we have

$$\mathbb{R} = \mathbb{Q} \cup \{\textit{irrationals}\},$$

where the set of the irrational numbers is not empty.

To move in the direction of the 10th axiom, we introduce a few preliminary notions.

DEFINITION (Upper bound). Let S be a nonempty set. If there is a number b such that $x \leq b$ for every $x \in S$, S is said to be *bounded above* by b and b is said to be an *upper bound* for S. If $b \in S$ than b is called the *maximum* element of S and we write

$$b = \max S.$$

If S has no upper bound, it is said to be *unbounded above*.

DEFINITION (Least upper bound). A number b is called *least upper bound* or *supremum* of a nonempty set S if

(1) b is an upper bound of S;

(2) No number less than b is an upper bound of S.

Suppose that a nonempty set S has two least upper bounds b_1 and b_2. Then, for any upper bound b of S,

$$b_1 \leq b \quad \text{and} \quad b_2 \leq b.$$

In particular,

$$b_1 \leq b_2 \quad \text{and} \quad b_2 \leq b_1.$$

Consequently, $b_1 = b_2$, and we proved the result we state below.

THEOREM 5.6. *If a nonempty set S has a least upper bound b then it is unique, and we write*

$$b = \sup S.$$

The notions of *lower bound* and of *greatest lower bound* (also called the *infimum*) are obtained from the notions of upper bound and least upper bound given above. b' is a lower bound of a nonempty set S if $-b'$ is an upper bound of the set $-S$ of the negatives of the elements of S (if b' is an element of S it is called the *minimum* of S). b' is the greatest lower bound of S if $-b'$ is the least upper bound of $-S$:

$$\inf S = -\sup\{-S\}.$$

Now we state the last axiom.

SUPREMUM AXIOM.

Axiom 10. Every nonempty set S of real numbers which is bounded above has a supremum.

THEOREM 5.7. *Every nonempty set S of real numbers which is bounded below has an infimum.*

Example. Consider the set

$$A = \{x \mid x \in \mathbb{Q} \ \ and \ \ x < \sqrt{2}\}.$$

The set is bounded above in $\mathbb{Q}$ (2, for example, is an upper bound of A). But it has no least upper bound in $\mathbb{Q}$. On the other hand, the set has a least upper bound in $\mathbb{R}$:

$$\sup_{\mathbb{R}} A = \sqrt{2}.$$

We now approach a few topics concerning the real number system.

Roots. We define the square root of a real number $a \geq 0$ as the solution of the equation

$$x^2 = a.$$

THEOREM 5.8. *Every nonnegative real number a as a unique nonnegative square root.*

Note that if $a \geq 0$, $a^{1/2}$ or $\sqrt{a}$ denotes the nonnegative square root of a. If $a > 0$, $-a^{1/2}$ or $-\sqrt{a}$ is the negative square root of a.

In general, the n-th root of a real number a, with $n \geq 2$ an integer, is the solution of the equation

$$x^n = a,$$

if it exists. If n is odd, the solution, denoted by $a^{1/n}$ or $\sqrt[n]{a}$, exists and is unique. If n is even and $a \geq 0$, $a^{1/n}$ or $\sqrt[n]{a}$ denotes the unique nonnegative n-th root of a. If $a > 0$, $-a^{1/n}$ or $-\sqrt[n]{a}$ is the unique negative n-th root of a. If n is even and $a < 0$ the equation $x^n = a$ has no solution.

We consider now the case of powers of rational exponent. Let $r = m/n$, with m and n positive integers. The power a^r, with a a real number, is defined by

$$a^r = a^{\frac{m}{n}} = (a^m)^{\frac{1}{n}} = \sqrt[n]{a^m}$$

if n is odd or if $a^m \geq 0$ if n is even. The power of negative rational exponent exists under the same conditions for $a \neq 0$, and is defined by

$$a^{-r} = \frac{1}{a^r}.$$

Let a and b be real numbers and r and s rational numbers. The following operation rules hold, supposing that all powers involved exist:

- $a^r \cdot a^s = a^{r+s}$;
- $a^r \cdot b^r = (ab)^r$;
- $(a^r)^s = a^{rs}$.

Method of proof by induction.

PRINCIPLE OF MATHEMATICAL INDUCTION.
Let S be a set of positive integers such that

(1) The number 1 is in S;

(2) If an integer k is in S then $k+1$ is also in S.

Then every positive integer is in S.

The following method for proving a result, called *method of proof by induction*, is based on the principle of mathematical induction. The idea is to prove that the set of the solutions of an assertion is the set $\mathbb{N}$ of the positive integers (more generally, the set of all integers greater than or equal to a certain integer n_1).

METHOD OF PROOF BY INDUCTION.
Let $A(n)$ be an assertion depending on an integer n. $A(n)$ is true for all $n \geq n_1$ if

(1) *The basis:* $A(n_1)$ is true;

(2) *The inductive step:* If $A(k)$ is true for $k \geq n_1$ then $A(k+1)$ is also true.

Example. Let us prove that if $x > 1$ then $x^n > x$ for every integer $n \geq 2$.

- The basis: $x > 1 \Rightarrow x^2 > x$;
- The inductive step: If $x^k > x$ for $k \geq 2$ then $x^{k+1} > x^2 > x$.

We can conclude that the assertion holds for all $n \geq 2$.

Absolute values. The absolute value of a real number x is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}.$$

THEOREM 5.9. *If x is a real number then $-|x| \leq x \leq |x|$.*

THEOREM 5.10. *Let a and x be real numbers. Then*

(1) $|x| \geq a \Leftrightarrow x \leq -a$ *or* $x \geq a$;

(2) $|x| \leq a \Leftrightarrow -a \leq x \leq a$;

(3) $|x| = a \Leftrightarrow x = -a$ *or* $x = a$ *if* $a \geq 0$.

THEOREM 5.11. *Let x and y be real numbers. Then*

(1) Positive homogeneity: $|xy| = |x| \cdot |y|$.

(2) Triangle inequality: $|x + y| \leq |x| + |y|$.

THEOREM 5.12 (Cauchy-Schwarz inequality). *Let $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ be real numbers. Then*

$$\left(\sum_{k=1}^{n} x_k y_k\right)^2 \leq \left(\sum_{k=1}^{n} x_k^2\right) \cdot \left(\sum_{k=1}^{n} y_k^2\right).$$

Moreover, the equality holds if and only if there is a real number a such that $x_k = ay_k$, for $k = 1, 2, \ldots, n$.

The metric space $\mathbb{R}$.

Let x and y be real numbers. We define the distance d between x and y as

$$d(x, y) = |x - y|.$$

This distance d satisfies the following properties (x, y, and z are arbitrary be real numbers):

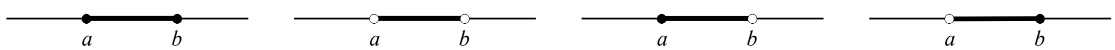
- $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only of $x = y$;
- $d(x, y) = d(y, x)$;
- $d(x, y) \leq d(x, z) + d(z, y)$.

The set $\mathbb{R}$ endowed with d is a *metric space*.
Note that the absolute value of a number represents geometrically its distance to the origin:

$$|x| = |x - 0| = d(x, 0).$$

Some sets of real numbers are very frequently used: the *intervals*. We have the following types of *bounded intervals* (the *endpoints* a and b, with $a < b$, are real numbers) (see Figure 5.2 for their geometric representation as *segment lines*):

- The *closed interval* $[a, b]$, representing the set $\{x \,|\, a \leq x \leq b\}$;
- The *open interval* (a, b) (or $]a, b[$), representing the set $\{x \,|\, a < x < b\}$;
- The *half-open intervals* $[a, b)$ (or $[a, b[$) and $(a, b]$ (or $]a, b]$) representing, respectively, the sets $\{x \,|\, a \leq x < b\}$ and $\{x \,|\, a < x \leq b\}$.



(a) $[a, b]$ (b) (a, b) (c) $[a, b)$ (d) $(a, b]$

Figure 5.2. Bounded intervals.

Independently of the type, the *center* or *midpoint* of a bounded interval with endpoints a and b ($a < b$) is the point

$$c = \frac{a+b}{2}.$$

We consider now the types of *unbounded intervals* (see Figure 5.3 for their geometric representation as semi-straight lines):

- The closed intervals $[a, +\infty)$ (or $[a, +\infty[$) and $(-\infty, b]$ (or $]-\infty, b]$), representing the sets $\{x \,|\, x \geq a\}$ and $\{x \,|\, x \leq b\}$, respectively;
- The open intervals $(a, +\infty)$ (or $]a, +\infty[$) and $(-\infty, b)$ (or $]-\infty, b[$), representing the sets $\{x \,|\, x > a\}$ and $\{x \,|\, x < b\}$, respectively.

(a) $[a, +\infty)$ (b) $(-\infty, b]$ (c) $(a, +\infty)$ (d) $(-\infty, b)$

Figure 5.3. Unbounded intervals.

Finally, the straight line itself can be represented by the unbounded interval $(-\infty, +\infty)$ (or $]-\infty, +\infty[$), that is, the set $\mathbb{R}$ of the real numbers (see Figure 5.4).

Figure 5.4. The set $\mathbb{R}$.

We finish the chapter by presenting basic *topological concepts*.

DEFINITION. Let E be a subset of $\mathbb{R}$.

(1) A *neighbourhood* of $p \in \mathbb{R}$ is a set

$$N_r(p) = \{q \mid q \in \mathbb{R} \ \textit{and} \ d(p,q) < r\}.$$

($r > 0$ is the *radius* of $N_r(p)$).

(2) $p \in \mathbb{R}$ is a *limit point* of the set E if every neighbourhood N of p contains a point $q \neq p$ such that $q \in E$. The set of all limit points of E is denoted by E'.

(3) If $p \in E$ and p is not a limit point of E, p is called an *isolated point* of E.

(4) E is *closed* if every limit point of E is a point of E.

(5) The *closure* $\overline{E}$ of the set E is the set $\overline{E} = E \cup E'$.

(6) A point p is an *interior point* of E if there is a neighbourhood N of p such that $N \subseteq E$. The set of all interior points of E is denoted $\mathring{E}$.

(7) E is *open* if every point of E is an interior point of E.

(8) The *complement* E^c of E is the set of all points $p \in \mathbb{R}$ such that $p \notin E$.

(9) E is *bounded* if there are real numbers q and M such that $d(p,q) < M$ for all $p \in E$.

The geometrical representation of a neighbourhood $N_r(p)$ can be found in Figure 5.5.

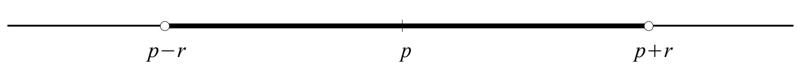


Figure 5.5. Neighbourhood.

Note also that a set E is bounded if and only if there is a real number $M > 0$ such that $d(p, 0) \leq M$ for all $p \in E$.

THEOREM 5.13.

(1) *Every neighbourhood of* $p \in \mathbb{R}$ *is an open set.*

(2) *If* p *is a limit point of a set* E *then every neighbourhood of* p *contains infinitely many points of* E.

(3) *A finite set has no limit points.*

(4) *A set* E *is open if and only if* E^c *is closed.*

(5) *A set* E *is closed if and only if* E^c *is open.*

(6) *The closure* $\overline{E}$ *of a set* E *is closed.*

(7) $E = \overline{E}$ *if and only if* E *is closed.*

Example. Consider the set

$$A = (-3, 0) \cup \left\{ x \,|\, x \in \mathbb{R} \ \textit{and} \ x = 1 + \frac{1}{n}, n \in \mathbb{N} \right\}.$$

Noting that

$$\sup A = \max A = 2 \ \text{ and } \ \inf A = -3,$$

we see that $d(p, 0) \leq 3$ for all $p \in A$. Therefore, the set A is bounded. The set of all limit points of A is

$$A' = [-3, 0] \cup \{1\}$$

and that the closure of A is

$$\overline{A} = A \cup A' = [-3, 0] \cup \{1\} \cup \left\{ 1 + \frac{1}{n}, n \in \mathbb{N} \right\}.$$

Since $A \neq \overline{A}$, A is not closed. The set of all isolated points of A is

$$\left\{ 1 + \frac{1}{n}, n \in \mathbb{N} \right\}.$$

The interior of A is

$$\mathring{A} = (-3, 0),$$

and, since $A \not\subseteq \mathring{A}$, A is not open.

Exercises.

(1) Prove, by using the induction principle:

a) $1+2+3+\ldots+n=\frac{n(n+1)}{2}$ for all integers n such that $n \geq 1$;

b) Partial sum of the arithmetic progression:

$$\sum_{k=1}^{n} a+d(k-1)=\frac{n}{2}(2a+d(n-1))$$

for all positive integers n, with a and d real constants;

c) Partial sum of the geometric progression:

$$\sum_{k=1}^{n} ar^{k-1}=a\frac{1-r^n}{1-r}$$

for all positive integers n, with a and $r \neq 1$ real constants;

d) Bernoulli's inequality: $(1+a)^n \geq 1+na$ if $a>-1$ and $n \in \mathbb{N}$;

e) $2^n < n!$ for all integers n such that $n \geq 4$;

f) $8^n - 3^n$ is divisible by 5 for all positive integers n.

(2) Obtain a law from the equalities

$$1=1; \quad 1-4=-(1+2), \quad 1-4+9=1+2+3, \quad 1-4+9-16=-(1+2+3+4),$$

and prove it by induction.

(3) Interpret geometrically the following sets:

a) $\{x : |x|<1\}$; b) $\{x : |x|<0\}$; c) $\{x : |x-a|<\epsilon\}$ $(\epsilon>0)$;

d) $\{x : |x|>0\}$; e) $\{x : (x-a)(x-b)<0, a<b\}$; f) $\{x : x^3>x\}$;

g) $\{x : |x-1| \geq |x|\}$.

(4) Solve the equations:

a) $x+2=\sqrt{4x+13}$; b) $|x+2|=\sqrt{4-x}$; c) $x^2-2|x|-3=0$.

(5) Let A and B be two subsets of $\mathbb{R}$ such that $A \subset B$, A is nonempty, and B is bounded above by a real number. Explain why $\sup A$ and $\sup B$ exist, and show that $\sup A \leq \sup B$.

(6) Show that, for all real numbers x and y,

a) $|x+y| \leq |x| + |y|$; b) $||x| - |y|| \leq |x-y|$.

(7) Solve the following inequalities, and give their solution sets:

a) $|3-2x| < 1$; b) $|1+2x| \leq 1$; c) $|x-1| > 2$; d) $|x+2| \geq 5$;

e) $|5 - x^{-1}| < 1$; f) $|x-5| < |x+1|$; g) $|x^2 - 2| < 1$; h) $|2-3x| \leq 1$;

i) $|x-3| > 2$; j) $|x-1| > 2$; k) $|3 - x^{-1}| < 1$; l) $|x-4| < |x+2|$;

m) $|x^2 - 5| \leq 2$; n) $x < x^2 - 12 < 4x$; o) $|2x-1| - x \geq 2$; p) $\dfrac{x}{1+|x|} \leq 2$;

q) $x - 2 \geq (|x|-1)^2$; r) $\left|\dfrac{x^2 - x}{1+x}\right| > x$.

(8) Determine in $\mathbb{R}$ the sets of all upper and lower bounds, the supremum and the infimum, and the maximum and the minimum (if they exist) for each one of the following sets:

a) $\left\{1, \sin\dfrac{\pi}{4}, \sin\dfrac{3\pi}{4}\right\}$; b) $\left\{(-1)^n \dfrac{1}{n} : n \in \mathbb{N}\right\}$;

c) $\left\{m + \dfrac{1}{n} : m, n \in \mathbb{N}\right\}$; d) $\left\{\dfrac{1}{n} + \dfrac{1}{m} : m, n \in \mathbb{N}\right\}$.

(9) Determine the interior, the set of all limit points, the closure, the sets of all upper and lower bounds, the supremum and the infimum, and the maximimum and the minimum (if they exist) of each one of the following sets:

a) $A = [2,3[\cup [4,10[$; b) $B =]5,7[\cup \{15\}$; c) $C = [0,1] - \mathbb{Q}$; d) $D = [2,3] \cap \mathbb{Q}$.

(10) Determine the interior, the set of all limit points, and the closure of the following sets:

a) $A = \{x \in \mathbb{R} : x^2 < 49\}$; b) $B = \{x : x \text{ is irrational and } x^2 < 49\}$.

(11) Consider the set

$$A = \left\{ x \in \mathbb{R} : x = 1 + \frac{(-1)^n}{n},\ n \in \mathbb{N} \right\}.$$

a) Determine the interior, the set of all limit points, and the closure of A.

b) Check if the set A is open or closed.

(12) Determine the interior, and the set of all limit points of the set

$$A = \{x \in \mathbb{Q} : |x+3| < 5\} \cup \left\{ x : x \text{ is irrational and } -\sqrt{2} \leq x \leq \sqrt{13} \right\}.$$

(13) Let the set

$$A = \left\{ x \in \mathbb{R} : \left| \frac{x^2}{x-2} \right| \leq 1 \right\}.$$

Determine $\mathring{A}$ and A'.

6. Sequences

Definitions.

DEFINITION (Sequence). A function $a : \mathbb{N} \longrightarrow \mathbb{R}$ is called a *sequence* of real numbers. To designate the sequence value at n it is customary to use the notation a_n instead of $a(n)$; and the sequence itself is in this case denoted by $\{a_n\}_{n\in\mathbb{N}}$ or simply by $\{a_n\}$. The elements a_n are called the *terms* of the sequence and the indices n their *order*.

Example. Consider the following sequence, the arithmetic progression

$$\{a_n\}, \quad a_n = a + (n-1)d,$$

with a and d real constants. The first four terms of the sequence are

$$\begin{aligned} a_1 &= a + (1-1)d = a, \\ a_2 &= a + (2-1)d = a + d, \\ a_3 &= a + (3-1)d = a + 2d, \\ a_4 &= a + (4-1)d = a + 3d. \end{aligned}$$

Alternatively to defining the sequence $\{a_n\}$ by giving its *general term*, that is, the term of order n, the sequence can be defined *by recursion*:

$$\begin{cases} a_1 = a \\ a_{n+1} = a_n + d, \quad \text{if } n \geq 1. \end{cases}$$

Now, the first term of the sequence is given and the remaining terms can be obtained from the recursion formula. We exemplify by determining the next three terms

$$\begin{aligned} a_2 &= a_1 + d = a + d, \\ a_3 &= a_2 + d = a + d + d = a + 2d, \\ a_4 &= a_3 + d = a + 2d + d = a + 3d. \end{aligned}$$

The two ways of defining the sequence are obviously equivalent and this can be proved by induction.

DEFINITION (Convergent sequence). Let $\{a_n\}$ be a sequence of real numbers. We say that $\{a_n\}$ *converges* and has *limit* $L \in \mathbb{R}$, and write

$$\lim_{n \to +\infty} a_n = L \quad \text{or} \quad a_n \to L \ \text{ as } \ n \to +\infty,$$

if for every real number $\varepsilon > 0$ there exists an integer N such that

$$|a_n - L| < \varepsilon \quad \text{whenever} \quad n \geq N.$$

If $\{a_n\}$ does not converge, it is said to *diverge*.

Example. Consider the sequence

$$\{a_n\}, \ \text{ with } \ a_n = \frac{1}{n}.$$

We want to prove that the sequence converges to 0. For this, given an arbitrary real number $\varepsilon > 0$, we need to find an order N such that

$$\left|\frac{1}{n} - 0\right| < \varepsilon \ \text{ if } \ n \geq N.$$

First notice that

$$\left|\frac{1}{n} - 0\right| = \left|\frac{1}{n}\right| = \frac{1}{n}$$

so that

$$\frac{1}{n} < \varepsilon \Rightarrow \left|\frac{1}{n} - 0\right| < \varepsilon.$$

As

$$\frac{1}{n} < \varepsilon \Leftrightarrow n > \frac{1}{\varepsilon},$$

if we chose N such that

$$N > \frac{1}{\varepsilon}$$

then we obtain, as needed,

$$n \geq N \Rightarrow \left|\frac{1}{n} - 0\right| < \varepsilon.$$

We proved that

$$\frac{1}{n} \to 0 \ \text{ as } \ n \to +\infty.$$

DEFINITION (Monotonic sequence). We say that a sequence $\{a_n\}$ is *monotonically increasing* if $a_{n+1} \geq a_n$ for every $n \in \mathbb{N}$, and *monotonically decreasing* if $a_{n+1} \geq a_n$ for every $n \in \mathbb{N}$. The class of the *monotonic* sequences consists of the increasing and the decreasing sequences.

Examples.

(a) Consider, again, the arithmetic progression

$$\{a_n\}, \quad a_n = a + (n-1)d,$$

with a and d real constants. As for a general sequence $\{c_n\}$

$$c_{n+1} \geq c_n \Leftrightarrow c_{n+1} - c_n \geq 0 \quad \text{and} \quad c_{n+1} \leq c_n \Leftrightarrow c_{n+1} - c_n \leq 0,$$

to check the possible monotonicity of the sequence we just need to check the sign of the difference $c_{n+1} - c_n$. In the particular case in study, we have

$$a_{n+1} - a_n = a + (n+1-1)d - (a + (n-1)d) = nd - (n-1)d = d.$$

Then, the monotonicity of $\{a_n\}$ depends on the signal of d: increasing if $d \geq 0$; decreasing if $d \leq 0$.

(b) Consider the sequence

$$\{b_n\}, \text{ with } b_n = 3^n.$$

Notice that for a general sequence $\{c_n\}$ if $c_n > 0$ for all $n \in \mathbb{N}$ we have

$$c_{n+1} \geq c_n \Leftrightarrow \frac{c_{n+1}}{c_n} \geq 1 \quad \text{and} \quad c_{n+1} \leq c_n \Leftrightarrow \frac{c_{n+1}}{c_n} \leq 1.$$

Therefore, we can check the possible monotonicity of a sequence by comparing the quotient c_{n+1}/c_n with 1. As, in our case, $b_n = 3^n > 0$ for all $n \in \mathbb{N}$, and

$$\frac{b_{n+1}}{b_n} \geq 1 = \frac{3^{n+1}}{3^n} = 3 > 1,$$

we conclude that $\{b_n\}$ is an increasing sequence.

Recall that a set $A \subseteq \mathbb{R}$ is bounded if there exists a real number M such that $|a| \leq M$ for every $a \in A$.

DEFINITION (Bounded sequence). We say a sequence $\{a_n\}$ of real numbers is *bounded* if there exists a real number M such that $|a_n| \leq M$ for all $n \in \mathbb{N}$, i.e., if the range of $\{a_n\}$ is a bounded set.

Examples.

(a) Let $\{a_n\}$ be the sequence with general term $a_n = (-1)^n$. As

$$|a_n| = |(-1)^n| = 1,$$

we see that $|a_n|$ is bounded above by any real number M such that $M \geq 1$. Then $\{a_n\}$ is bounded.

(b) Consider the sequence

$$\{b_n\}, \text{ with } b_n = -\frac{1}{2n}.$$

We have that

$$|b_n| = \left|-\frac{1}{2n}\right| = \frac{1}{2n},$$

and also that

$$\left|-\frac{1}{2n}\right| \leq M \Leftrightarrow \frac{1}{2n} \leq M \Leftrightarrow n \geq \frac{1}{2M}.$$

As the inequality

$$n \geq \frac{1}{2M}$$

is satisfied for all n if $M \geq 1/2$, we conclude that $\{b_n\}$ is a bounded sequence.

DEFINITION (Subsequence). Let $\{a_n\}$ be a sequence, and suppose that $\{n_k\}$ is a sequence of positive integers, such that

$$n_1 < n_2 < n_3 < \cdots$$

The sequence $\{a_{n_k}\}$ is called a *subsequence* of $\{a_n\}$. If $\{a_{n_k}\}$ converges, its limit is called a *subsequential limit* of $\{a_n\}$.

Example. Let $\{a_n\}$ be the sequence with general term

$$a_n = (-1)^n.$$

The subsequences of the terms with even and odd indices are, respectively,

$$a_{2n} = (-1)^{2n} = 1 \quad \text{and} \quad a_{2n-1} = (-1)^{2n-1} = -1,$$

for all $n \in \mathbb{N}$. As

$$\lim_{n\to+\infty} a_{2n} = \lim_{n\to+\infty} 1 = 1 \quad \text{and} \quad \lim_{n\to+\infty} a_{2n-1} = \lim_{n\to+\infty} -1 = -1,$$

1 and -1 are subsequential limits of $\{a_n\}$.

Basic results.

THEOREM 6.1. *If L and L' are real numbers, and if $\{x_n\}$ is a sequence in $\mathbb{R}$ converging to L and to L', then $L = L'$.*

THEOREM 6.2. *Suppose that $\{x_n\}$ is a convergent sequence of real numbers. Then for every real number $\varepsilon > 0$ there exists an integer N such that*

$$m, n > N \quad \textit{implies that} \quad |x_n - x_m| < \varepsilon.$$

THEOREM 6.3. *Suppose that $\{x_n\}$ is a convergent sequence of real numbers with $x_n \to L$ as $n \to +\infty$. Then any subsequence $\{x_{n_k}\}$ converges and $x_{n_k} \to L$ as $n \to +\infty$.*

THEOREM 6.4. *If $\{x_n\}$ is a convergent sequence of real numbers, then $\{x_n\}$ is bounded.*

THEOREM 6.5. *Suppose $\{x_n\}$ is a monotonic sequence of real numbers. Then $\{x_n\}$ converges if and only if it is bounded.*

THEOREM 6.6. *Suppose that $\{x_n\}$ and $\{y_n\}$ are sequences in $\mathbb{R}$ such that $x_n \to 0$ as $n \to +\infty$ and y_n is bounded. Then the sequence $\{x_n y_n\}$ converges and $x_n y_n \to 0$ as $n \to +\infty$.*

THEOREM 6.7. *Suppose that $\{x_n\}$ and $\{x_n\}$ are convergent sequences in $\mathbb{R}$ with $L = \lim_{n\to+\infty} x_n$ and $M = \lim_{n\to+\infty} y_n$. Then the following holds:*

(1) The sequence $\{x_n + y_n\}$ converges and

$$\lim_{n\to+\infty}(x_n + y_n) = L + M;$$

(2) For any real number α, the sequences $\{\alpha x_n\}$ and $\{\alpha + x_n\}$ converge and

$$\lim_{n\to+\infty}(\alpha x_n) = \alpha L \quad \text{and} \quad \lim_{n\to+\infty}(\alpha + x_n) = \alpha + L;$$

(3) The sequence $\{x_n y_n\}$ converges and

$$\lim_{n\to+\infty}(x_n y_n) = LM;$$

(4) If $y_n \neq 0$ for all $n \in \mathbb{N}$ and $M \neq 0$, then the sequence $\{x_n/y_n\}$ converges and

$$\lim_{n\to+\infty}\frac{x_n}{y_n} = \frac{L}{M};$$

(5) The sequence $\{|x_n|\}$ converges and

$$\lim_{n\to+\infty}|x_n| = |L|;$$

(6) The sequence $\{(x_n)^p\}$, with p a positive integer, converges and

$$\lim_{n\to+\infty}(x_n)^p = L^p;$$

(7) If $x_n \geq 0$ for all $n \in \mathbb{N}$, then the sequence $\{\sqrt[p]{x_n}\}$, with $p \geq 2$ an integer, converges and

$$\lim_{n\to+\infty}\sqrt[p]{x_n} = \sqrt[p]{L};$$

(8) If $x_n \geq 0$ for all $n \in \mathbb{N}$, then $L \geq 0$;

(9) If $x_n \geq y_n$ for all $n \in \mathbb{N}$, then $L \geq M$.

THEOREM 6.8. *Suppose that* $\{x_n\}$, $\{y_n\}$, *and* $\{z_n\}$ *are sequences in* $\mathbb{R}$ *such that* $y_n \leq x_n \leq z_n$ *for all* $n \in \mathbb{N}$. *Suppose further that the sequences* $\{y_n\}$ *and* $\{z_n\}$ *are convergent, with* $\lim_{n \to +\infty} y_n = \lim_{n \to +\infty} z_n = L$. *Then the sequence* $\{x_n\}$ *converges and* $\lim_{n \to +\infty} x_n = L$.

THEOREM 6.9.

(1) *If* p *is a positive real number, then* $\lim_{n \to +\infty} \frac{1}{n^p} = 0$.

(2) *If* p *is a positive real number, then* $\lim_{n \to +\infty} \sqrt[n]{p} = 1$.

(3) $\lim_{n \to +\infty} \sqrt[n]{n} = 1$.

(4) *If* p *and* α *are real numbers, with* $p > 0$, *then* $\lim_{n \to +\infty} \frac{n^\alpha}{(1+p)^n} = 0$.

(5) *If* x *is a real number such that* $|x| < 1$, *then* $\lim_{n \to +\infty} x^n = 0$.

Examples.

(a) Consider the sequence $\{(-1)^n\}$. As the subsequences $\{(-1)^{2n}\}$ and $\{(-1)^{2n-1}\}$ do not converge to the same number (they converge to 1 and -1, respectively), $\{(-1)^n\}$ diverges.

(b) Consider the bounded sequence $\{(-1)^n\}$, and the sequence $\{1/n\}$ converging to 0 as $n \to +\infty$. Then

$$\lim_{n \to +\infty} (-1)^n \cdot \frac{1}{n} = 0.$$

(c) From the knowledge that

$$\frac{1}{n} \to 0, \qquad \frac{2n+1}{n+2} \to 2 \quad \text{and} \quad \frac{3n+1}{n} \to 3,$$

we can conclude that

$$\lim_{n \to +\infty} \left(5 \cdot \frac{1}{n} \cdot \frac{2n+1}{n+2} - 2 \cdot \left(\frac{3n+1}{n} \right)^2 \right) = 5 \cdot 0 \cdot 2 - 2 \cdot 3^2 = -18.$$

Extended real numbers.

DEFINITION (Infinite limits). Let $\{a_n\}$ be a sequence of real numbers. If for every real number M there exists an integer N such that $a_n > M$ whenever $n \geq N$, then we say the sequence $\{a_n\}$ *diverges to positive infinity*, denoted by

$$\lim_{n\to+\infty} a_n = +\infty.$$

Similarly, if for every real number M there exists an integer N such that $a_n < M$ whenever $n \geq N$, then we say the sequence $\{a_n\}$ *diverges to negative infinity*, denoted by

$$\lim_{n\to+\infty} a_n = -\infty.$$

If $\lim_{n\to+\infty} |a_n| = +\infty$, we say the sequence $\{a_n\}$ *diverges to infinity*, and write

$$\lim_{n\to+\infty} a_n = \infty.$$

THEOREM 6.10. *Let $\{x_n\}$ and $\{y_n\}$ be sequences in $\mathbb{R}$ such that x_n is unbounded and y_n is bounded, with $y_n \neq 0$ for all $n \in \mathbb{N}$. Then the sequence $\{x_n/y_n\}$ is unbounded. In particular,*

(1) If $\lim_{n\to+\infty} x_n = +\infty$ and $\lim_{n\to+\infty} y_n = 0$,

$$\lim_{n\to+\infty} \frac{x_n}{y_n} = +\infty \ \textit{ if } \ y_n > 0 \ \textit{ for all } \ n \in \mathbb{N}$$

$$\lim_{n\to+\infty} \frac{x_n}{y_n} = -\infty \ \textit{ if } \ y_n < 0 \ \textit{ for all } \ n \in \mathbb{N};$$

(2) If $\lim_{n\to+\infty} x_n = -\infty$ and $\lim_{n\to+\infty} y_n = 0$,

$$\lim_{n\to+\infty} \frac{x_n}{y_n} = -\infty \ \textit{ if } \ y_n > 0 \ \textit{ for all } \ n \in \mathbb{N}$$

$$\lim_{n\to+\infty} \frac{x_n}{y_n} = +\infty \ \textit{ if } \ y_n < 0 \ \textit{ for all } \ n \in \mathbb{N}.$$

DEFINITION (Extended real numbers).
The *extended real number system*, denoted $\overline{\mathbb{R}}$, consists of the real field $\mathbb{R}$ and two symbols, $-\infty$ and $+\infty$. We preserve the original order in $\mathbb{R}$, and define

$$-\infty < r < +\infty,$$

for every real number r.

Every subset of $\overline{\mathbb{R}}$ is bounded above by $+\infty$. Moreover, every nonempty subset S of real numbers has always a supremum in $\overline{\mathbb{R}}$: if S is bounded above in $\mathbb{R}$, the supremum is real; if S is not bounded above in $\mathbb{R}$, $\sup(S) = +\infty$ in $\overline{\mathbb{R}}$. The same remarks apply to lower bounds.

The extended real number system does not constitute a field, but it is customary to make the following operation conventions:

(1) For any real number r,

$$r + (+\infty) = r - (-\infty) = r + \infty = +\infty;$$
$$r + (-\infty) = r - (+\infty) = r - \infty = -\infty;$$
$$\frac{r}{\pm\infty} = 0.$$

(2) For any real number $r > 0$, $r \cdot (+\infty) = +\infty$ and $r \cdot (-\infty) = -\infty$.

(3) For any real number $r < 0$, $r \cdot (+\infty) = -\infty$ and $r \cdot (-\infty) = +\infty$.

(4) $\pm\infty \pm \infty = \pm\infty$, $\pm\infty \cdot (\pm\infty) = +\infty$, and $\pm\infty \cdot (\mp\infty) = -\infty$.

(5) If p is an odd positive integer, $(\pm\infty)^p = \pm\infty$ and $\sqrt[p]{\pm\infty} = \pm\infty$.

(6) If p is an even positive integer, $(\pm\infty)^p = +\infty$ and $\sqrt[p]{+\infty} = +\infty$.

Note that the symbols

$$\pm\infty \mp \infty, \quad 0 \cdot (\pm\infty), \quad \frac{\pm\infty}{\pm\infty}, \quad \frac{\pm\infty}{\mp\infty}, \quad \frac{0}{0}, \quad 1^{\pm\infty}, \quad (\pm\infty)^0, \quad 0^0$$

are not defined.

To distinguish the real numbers from $-\infty$ and $+\infty$ we call the former *finite real numbers* and the latter *infinite real numbers*.

Note also that in the extended real number system $\overline{\mathbb{R}}$, a sequence which diverges to $+\infty$ (to $-\infty$) is said to *converge to* $+\infty$ (to $-\infty$).

THEOREM 6.11. *If x_n is a sequence such that $x_n \to +\infty$ as $n \to +\infty$ and a is a real number, then*

$$\lim_{n\to+\infty} \left(1 + \frac{a}{x_n}\right)^{x_n} = e^a.$$

Examples.

(a) By using the limit theorems we can determine the following limit

$$\lim_{n\to+\infty}\frac{n^2}{n^4+1}=\lim_{n\to+\infty}\frac{n^2}{n^4\left(1+\frac{1}{n^4}\right)}=\lim_{n\to+\infty}\frac{1}{n^2\left(1+\frac{1}{n^4}\right)}=\frac{1}{(+\infty)^2\left(1+\frac{1}{(+\infty)^4}\right)}=\frac{1}{+\infty\left(1+\frac{1}{+\infty}\right)}$$
$$=\frac{1}{+\infty(1+0)}=\frac{1}{+\infty}=0.$$

(b) By using Theorem 6.11,

$$\lim_{n\to+\infty}\left(1-\frac{2}{n}\right)^n=e^{-2}.$$

Exercises.

(1) Determine, if it exists, the limit of each one of the following sequences:

a) $\left\{\frac{1}{n}\right\}$; b) $\left\{\frac{1}{n^2}\right\}$; c) $\left\{\frac{n}{n+1}\right\}$; d) $\left\{\frac{2n-1}{3n+2}\right\}$; e) $\left\{\frac{n^2}{n^4+1}\right\}$; f) $\left\{\frac{(-1)^n}{n}\right\}$.

(2) Let $\{x_n\}$ be a sequence of real numbers with $x_n\to+\infty$, and P and Q the polynomial functions

$$P(x)=a_0x^p+\ldots+a_{p-1}x+a_p,\qquad Q(x)=b_0x^q+\ldots+b_{q-1}x+b_q,$$

with real coefficients ($a_0,b_0\neq 0$). Show that

a) $\lim_{n\to+\infty}P(x_n)=\lim_{n\to+\infty}a_0x_n^p$.

b) $$\lim_{n\to+\infty}\frac{P(x_n)}{Q(x_n)}=\lim_{n\to+\infty}\frac{a_0x_n^p}{b_0x_n^q}=\begin{cases}\dfrac{a_0}{b_0} & \text{if } p=q,\\ \frac{a_0}{b_0}\cdot(+\infty) & \text{if } p>q,\\ 0 & \text{if } p<q.\end{cases}$$

(3) Determine, if they exist, the limits of the sequences with general elements:

a) $a_n=\frac{1-n}{4n+3}$; b) $b_n=\frac{2n+3}{3n-1}$; c) $c_n=\frac{n^2+2}{3n+1}$;

d) $d_n=\frac{n^2-1}{n^4+3}$; e) $e_n=\frac{3n}{4n^3+1}$; f) $f_n=\frac{2^n+1}{2^{n+1}-1}$;

g) $g_n=\frac{-n^3+2}{4n^3-7}$; h) $h_n=\frac{n^3+1}{n^2+2n-1}$; i) $i_n=\frac{n^2+3n}{n+2}-\frac{n^2-1}{n}$;

j) $j_n=\frac{(-1)^nn^3}{n^2+2}$; k) $k_n=\frac{n(n-1)(n-2)}{(n+1)(n+2)}$.

(4) Determine, if they exist, the limits:

a) $\lim\limits_{n\to+\infty} \cos^2(n)\sin\left(\frac{1}{n}\right)$;

b) $\lim\limits_{n\to+\infty} \frac{n(n-1)(n-2)(n-3)}{(n+1)(n+2)(n+3)}$;

c) $\lim\limits_{n\to+\infty} (\cos(x))^n, \quad x\in\mathbb{R}$;

d) $\lim\limits_{n\to+\infty} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n+1}} + \cdots + \frac{1}{\sqrt{2n}}\right)$;

e) $\lim\limits_{n\to+\infty} \left(\frac{1}{\sqrt{n^2+1}} + \frac{1}{\sqrt{n^2+2}} + \cdots + \frac{1}{\sqrt{n^2+2n+1}}\right)$;

f) $\lim\limits_{n\to+\infty} \left(\frac{1}{n^2} + \frac{1}{(n+1)^2} + \cdots + \frac{1}{(2n)^2}\right)$;

g) $\lim\limits_{n\to+\infty} \left(\frac{n}{\sqrt{n^4+1}} + \frac{n}{\sqrt{n^4+2}} + \cdots + \frac{n}{\sqrt{n^4+n}}\right)$;

h) $\lim\limits_{n\to+\infty} \left(\sqrt{n+1} - \sqrt{n}\right)$.

(5) Determine, if they exist, the limits of the sequences with general elements:

a) $u_n = \left(\frac{n+3}{n+1}\right)^{2n}$; b) $v_n = \left(\frac{n+5}{2n+1}\right)^{n}$; c) $w_n = \left(1 - \frac{3}{n^2}\right)^{n}$.

7. Series

Definitions and basic results.

DEFINITION (Series). Given a sequence $\{a_n\}$ of real numbers, we use the notation

$$\sum_{n=p}^{q} a_n \quad \text{for} \quad p \le q$$

to denote $a_p + a_{p+1} + \cdots + a_q$. With $\{a_n\}$ we associate a sequence $\{s_n\}$ where

$$s_n = \sum_{k=1}^{n} a_k.$$

For $\{s_n\}$ we use also the symbolic expression

$$a_1 + a_2 + a_3 + \cdots$$

or, more concisely,

$$\sum_{n=1}^{+\infty} a_n,$$

which we call an *infinite series*, or just a *series*. The numbers s_n are called the *partial sums* of the series. If $\{s_n\}$ converges to a number s, called the *sum* of the series, we say that the series *converges*, and write

$$\sum_{n=1}^{+\infty} a_n = s.$$

Otherwise, the series is said to *diverge*.

When convenient, we shall consider series of the form $\sum_{n=n^*}^{+\infty} a_n$, with $n^* \in \mathbb{N}_0$. Also, when there is no possible ambiguity, or when the distinction is immaterial, we shall simply write $\sum a_n$.

Example. Consider the series $\sum a_n$, where $\{a_n\}$ is the arithmetic progression with general term

$$a_n = a + (n-1)d.$$

As $a_1 = a$, the sequence's general term can be written

$$a_n = a_1 + (n-1)d.$$

The partial sums s_n can be written in the following two different ways

$$s_n = a_1 + (a_1 + d) + (a_1 + 2d) + \cdots + (a_1 + (n-2)d) + (a_1 + (n-1)d)$$
$$s_n = (a_n - (n-1)d) + (a_n - (n-2)d) + \cdots + (a_n - 2d) + (a_n - d) + a_n.$$

Adding both sides of the two equations, the terms involving d cancel, and we obtain

$$2s_n = n(a_1 + a_n).$$

Dividing both sides by 2 gives

$$s_n = \frac{n}{2}(a_1 + a_n).$$

As $a_n = a_1 + (n-1)d$, the partial sums can also be written

$$s_n = \frac{n}{2}\left(2a_1 + (n-1)d\right) = \frac{d}{2}n^2 + \left(a_1 - \frac{d}{2}\right)n.$$

We can see that s_n converges if and only if $a_1 = d = 0$, in which case $a_n = 0$ for all n. We then have that $\sum(a_1 + (n-1)d)$ diverges if $a_1 \neq 0$ or $d \neq 0$ and converges if $a_1 = d = 0$ (with $\sum(a_1 + (n-1)d) = 0$).

THEOREM 7.1. *If* $\sum a_n$ *converges then* $\lim_{n\to+\infty} a_n = 0$.

Note that Theorem 7.1 gives just a necessary condition for convergence of a series. In fact, only from $a_n \to 0$ nothing can be concluded about the convergence of $\sum a_n$.

Examples.

(a) Consider the series

$$\sum a_n, \quad \text{with } a_n = \frac{n}{2n+1}.$$

As

$$\lim_{n\to+\infty} a_n = \lim_{n\to+\infty} \frac{n}{2n+1} = \lim_{n\to+\infty} \frac{n}{2n} = \frac{1}{2} \neq 0,$$

we conclude that $\sum a_n$ diverges.

(b) Consider again the arithmetic progression $\{b_n\}$, with

$$b_n = b + (n-1)d.$$

As $b_n \to 0$ if and only if $b = d = 0$ we can conclude that for $b \neq 0$ or $d \neq 0$ the series $\sum(b + (n-1)d)$ diverges.

THEOREM 7.2. *Let* $\sum a_n$ *and* $\sum b_n$ *be convergent series such that*

$$\sum a_n = a \quad \textit{and} \quad \sum b_n = b.$$

Then the series $\sum(a_n + b_n)$ *and* $\sum c a_n$*, with* c *a fixed real number, converge and*

$$\sum(a_n + b_n) = a + b, \qquad \sum c a_n = ca.$$

THEOREM 7.3. *If* $\sum a_n$ *converges and* $\sum b_n$ *diverges then* $\sum(a_n + b_n)$ *diverges.*

THEOREM 7.4. *If* $\sum a_n$ *diverges and* $c \neq 0$ *then* $\sum c a_n$ *diverges.*

DEFINITION (Absolute convergence). The series $\sum a_n$ is said to be *absolutely convergent if* $\sum |a_n|$ *converges.*

THEOREM 7.5. *If* $\sum a_n$ *converges absolutely then* $\sum a_n$ *converges.*

Geometric series.

DEFINITION (Geometric series). The series generated by successive addition of the terms of a geometric progression, $\sum_{n=0}^{+\infty} x^n$, with x a fixed real number, is called a *geometric series*.

THEOREM 7.6. *Consider the geometric series* $\sum_{n=0}^{+\infty} x^n$*. If* $|x| < 1$*, the series converges and*

$$\sum_{n=0}^{+\infty} x^n = \frac{1}{1-x}.$$

If $|x| \geq 1$*, the series diverges.*

Examples.

(a) The geometric series $\sum(-1)^n$, $\sum 2^n$, and $\sum(-2)^n$ diverge.

(b) The geometric series $\sum_{n=0}^{+\infty} \left(\frac{1}{2}\right)^n$ converges, and

$$\sum_{n=0}^{+\infty} \left(\frac{1}{2}\right)^n = \frac{1}{1-\frac{1}{2}} = 2.$$

As $\left|\left(-\frac{1}{2}\right)^n\right| = \left(\frac{1}{2}\right)^n$, the series $\sum_{n=0}^{+\infty} \left(-\frac{1}{2}\right)^n$ converges (absolutely). Its sum is

$$\sum_{n=0}^{+\infty} \left(-\frac{1}{2}\right)^n = \frac{1}{1+\frac{1}{2}} = \frac{2}{3}.$$

(c) The series $\sum \frac{n}{n+1}$ diverges since $\frac{n}{n+1} \nrightarrow 0$ as $n \to +\infty$, and $\sum \left(\frac{1}{3}\right)^n$ is a convergent geometric series. Thus the series $\sum \left(\frac{n}{n+1} + \left(\frac{1}{3}\right)^n\right)$ diverges.

(d) The series $\sum_{n=0}^{+\infty} \left(\left(-\frac{2}{3}\right)^n - 2\left(\frac{1}{3}\right)^n\right)$ converges, since both $\sum_{n=0}^{+\infty} \left(-\frac{2}{3}\right)^n$ and $\sum_{n=0}^{+\infty} \left(\frac{1}{3}\right)^n$ are convergent geometric series. The sum is

$$\sum_{n=0}^{+\infty} \left(\left(-\frac{2}{3}\right)^n - 2\left(\frac{1}{3}\right)^n\right) = \sum_{n=0}^{+\infty} \left(-\frac{2}{3}\right)^n - 2\sum_{n=0}^{+\infty} \left(\frac{1}{3}\right)^n = \frac{1}{1+\frac{2}{3}} - 2\frac{1}{1-\frac{1}{3}} = \frac{3}{5} - 2\frac{3}{2} = -\frac{12}{5}.$$

Power series.

DEFINITION (Power series). Given a sequence $\{c_n\}$ of real numbers, the series

$$\sum_{n=0}^{+\infty} c_n x^n,$$

where x is real, is called a (real) *power series*. The numbers c_n are called the *coefficients* of the series.

Example. The following series are power series:

$$\sum_{n=0}^{+\infty} n! x^n, \qquad \sum_{n=0}^{+\infty} \frac{1}{n} x^{2n}, \qquad \text{and} \qquad \sum_{n=0}^{+\infty} 2^n (x-1)^n.$$

With every power series there is an interval associated, the *interval of convergence*, such that the series converges absolutely if x is in the interior of the interval and diverges if x is in the exterior (we consider the real line as the interior of an interval of infinite length, and a point as an interval of length zero). The behaviour on the boundary of the interval of convergence is varied.

THEOREM 7.7. *Given the power series $\sum c_n x^n$, suppose that $\lim_{n\to+\infty} \sqrt[n]{|c_n|}$ exists, and put*

$$\alpha = \lim_{n\to+\infty} \sqrt[n]{|c_n|}, \qquad R = \frac{1}{\alpha}.$$

(If $\alpha = 0$, $R = +\infty$; if $\alpha = +\infty$, $R = 0$). Then

(1) $\sum c_n x^n$ converges absolutely if $|x| < R$;

(2) $\sum c_n x^n$ diverges if $|x| > R$.

THEOREM 7.8. *Given the power series* $\sum c_n x^n$*, suppose that* $\lim_{n\to+\infty}\left|\frac{c_{n+1}}{c_n}\right|$ *exists, and put*

$$\alpha = \lim_{n\to+\infty}\left|\frac{c_{n+1}}{c_n}\right|, \qquad R = \frac{1}{\alpha}.$$

(If $\alpha = 0$*,* $R = +\infty$*; if* $\alpha = +\infty$*,* $R = 0$*). Then*

(1) $\sum c_n x^n$ *converges absolutely if* $|x| < R$*;*

(2) $\sum c_n x^n$ *diverges if* $|x| > R$*.*

Examples.

(a) Consider the series $\sum n^n x^n$. As

$$\lim_{n\to+\infty}\sqrt[n]{|n^n|} = \lim_{n\to+\infty} n = +\infty, \text{ and then } R = 0,$$

the interior of the interval of convergence is empty, so nothing can be concluded with the help of Theorem 7.7. But, observing that for $x = 0$ the series is just the zero series, we obtain that the series converges if and only if $x = 0$.

(b) Consider the series $\sum \frac{x^n}{n!}$. As

$$\lim_{n\to+\infty}\left|\frac{\frac{1}{(n+1)!}}{\frac{1}{n!}}\right| = \lim_{n\to+\infty}\left|\frac{n!}{(n+1)!}\right| = \lim_{n\to+\infty}\frac{1}{n+1} = 0, \text{ and then } R = +\infty,$$

the series converges (absolutely) for all real x.

(c) Consider now the series $\sum \frac{x^n}{n^2}$. As

$$\lim_{n\to+\infty}\left|\frac{\frac{1}{(n+1)^2}}{\frac{1}{n^2}}\right| = \lim_{n\to+\infty}\left|\frac{n^2}{(n+1)^2}\right| = \lim_{n\to+\infty}\frac{n^2}{n^2} = 1, \text{ and then } R = 1,$$

the series converges (absolutely) for all x such that

$$|x| < 1 \Leftrightarrow -1 < x < 1.$$

Exercises.

(1) Determine if following series are convergent, and when they are, compute the sum:

$$\text{a) } \sum_{n\geq 0}\left(\frac{1}{2}\right)^n; \qquad \text{b) } \sum_{n\geq 1} 3^n; \qquad \text{c) } \sum_{n\geq 0}\left(\frac{2}{3}\right)^{n+2}; \qquad \text{d) } \sum_{n\geq 3}\left(\frac{1}{4}\right)^n; \qquad \text{e) } \sum_{n\geq 0} 5^n.$$

(2) Determine the values of x for which the following series converge, and, when possible, compute their sum:

a) $\displaystyle\sum_{n=0}^{+\infty}\left(\frac{x}{x+1}\right)^n$; b) $\displaystyle\sum_{n=0}^{+\infty}\left(\frac{2}{x}\right)^n$; c) $\displaystyle\sum_{n=0}^{+\infty}(1-|x|)^n$; d) $\displaystyle\sum_{n=0}^{+\infty}(x+1)^{2n}$;

e) $\displaystyle\sum_{n=0}^{+\infty}\frac{x^n}{n!}$; f) $\displaystyle\sum_{n=0}^{+\infty}\left(\frac{x^n}{n!}+\left(\frac{2}{x}\right)^{n+2}\right)$; g) $\displaystyle\sum_{n=0}^{+\infty}\frac{(1-|x|)^n}{n!}$; h) $\displaystyle\sum_{n=0}^{+\infty}\frac{(x+1)^{2n+6}}{(n+3)}$;

i) $\displaystyle\sum_{n=2}^{+\infty}\frac{(x-2)^{n+3}}{n}$.

(3) Use the theory of geometric series to determine the rational numbers corresponding to the following decimals:

a) $3.666\ldots$

b) $1.571428571428571428\ldots$

c) $1.181818\ldots$

d) $0.999\ldots$

(4) Determine the radius of convergence, and the largest open set where the following power series are absolutely convergent:

a) $\displaystyle\sum_{n=1}^{+\infty}\frac{x^n}{n(n+1)}$; b) $\displaystyle\sum_{n=1}^{+\infty}\frac{(2x+1)^{2n+1}}{\sqrt{n}}$; c) $\displaystyle\sum_{n=1}^{+\infty}\frac{n(x+1)^{2n}}{3^n}$; d) $\displaystyle\sum_{n=1}^{+\infty}n!n^{-n}x^n$.

8. One-variable functions

Generalities.

DEFINITION (Function (informal)). Given two sets X and Y, we call *function* (or *mapping*) the correspondence f of X in Y, symbolically

$$f : X \longrightarrow Y,$$

which associates to each element of X one and only one element of Y, and we say that f *maps* X *into* Y. The set X is called the *domain* of the function. The elements of Y associated to the elements of X form a set called the *range* of the function.

Functions are usually represented by lower and upper case letters of the English and Greek alphabets. The lower case letters f, g and h, the upper case letters F, G and H, and the lower case Greek letters φ (phi) and ψ (psi) are particularly used. It is also usual to denote the domain by the letter D, and the range by R (with $R \subseteq Y$).

Given the function f, if x is an element of its domain, the notation $f(x)$ is used to designate the element in the range associated to x by the function f. This element $f(x)$ in the range of f is called the *value* of f at x or the *image* of x under f, and is read "f of x". To represent the correspondence of x and its image $f(x)$ sometimes we write

$$x \longmapsto f(x).$$

By abuse of language, we say indistinctly "the function f" or "the function $f(x)$". Note that the range R is the set of the images of all elements of the domain D under f, so that we write $R = f(D)$.

In Figure 8.1 shows different ways to represent schematically a function: with the use of Venn diagrams (Figure 1(a)); and with an *input/output* scheme (1(b)).

We also observe that x may be seen as a variable taking values in the domain of the function, and $f(x)$ as a variable taking the corresponding values in the range of the function (sometimes this variable $f(x)$ is represented by another letter, for example y). As the values of the variable $f(x)$ depend on the values assigned to the variable x, x is called the *independent variable*, and $y = f(x)$ the *dependent variable*.

A function can be classified with respect to *injectivity* and to *surjectivity*.

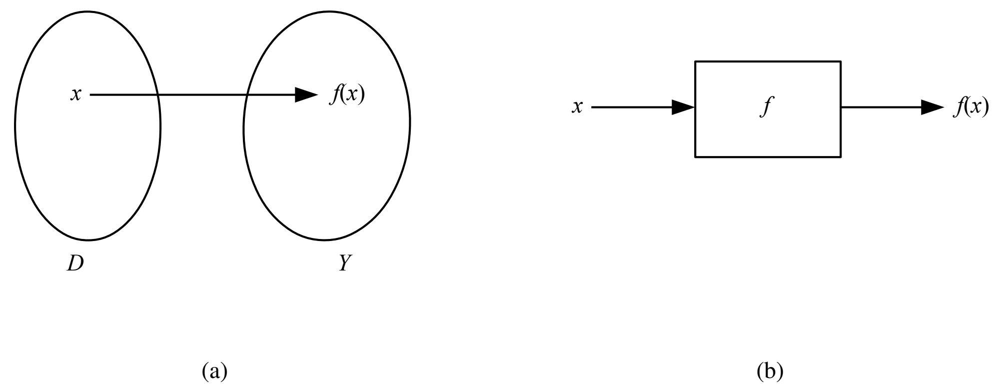


Figure 8.1. Schematic representation of a function.

DEFINITION (Injective and surjective function). Let f be a function

$$f : D \longrightarrow Y,$$

and set $y = f(x)$ for all $x \in D$.

(1) Function f is said to be *injective* (or *one-to-one* or an *injection*) if any two distinct elements in D have distinct images. Symbolically,

$$\forall x_1, x_2 \in D, x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2).$$

(2) Function f is said to be *surjective* (or a *surjection*) if the range coincides with Y $(R = Y)$. Symbolically,

$$\forall y \in Y \; \exists x \in D : y = f(x),$$

and we say that f *maps* D *onto* Y.

(3) If f is injective and surjective it is called *bijective* (or *one-to-one and onto* or a *one-to-one correspondence* or a *bijection*) and, symbolically, we write

$$\forall y \in Y \; \exists^1 x \in D : y = f(x).$$

In Figure 8.2 we represent examples for the different possibilities of classification of a function with respect to injectivity and surjectivity.

DEFINITION (Inverse). Let f be a bijective function defined by $f : D \longrightarrow Y$, and set $y = f(x)$ for all $x \in D$. The *inverse* of f is the function, denoted by f^{-1},

$$f^{-1} : Y \longrightarrow D$$

such that

$$x = f^{-1}(y) \Leftrightarrow y = f(x),\ \forall x \in D\ \forall y \in Y.$$

If f is injective but not surjective there exist elements in Y (the elements in $Y - R$) which are not images of any element in D. In this case, the one-to-one correspondence exists but only between the elements in D and the elements in $R \subset Y$, and the inverse of f is the function

$$f^{-1} : R \longrightarrow D$$

such that

$$x = f^{-1}(y) \Leftrightarrow y = f(x),\ \forall x \in D\ \forall y \in R.$$

Examples.

(a) In Figure 2(d) we represent the bijection $j : \{a, b, c\} \longrightarrow \{u, v, w\}$ such that $j(a) = u$, $j(b) = v$, and $j(c) = w$. The inverse of j is the function $j^{-1} : \{u, v, w\} \longrightarrow \{a, b, c\}$ such that $j^{-1}(u) = a$, $j^{-1}(v) = b$, and $j^{-1}(w) = c$.

(b) Figure 2(c) represents the injection $h : \{a, b\} \longrightarrow \{u, v, w\}$ such that $h(a) = u$ and $h(b) = v$. the range of h is the set $R = \{u, v\}$. The inverse of h is the function $h^{-1} : \{u, v\} \longrightarrow \{a, b\}$ such that $h^{-1}(u) = a$ and $h^{-1}(v) = b$.

DEFINITION (Composite function). Consider the functions $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$. The *composition* of g and f is the function, denoted by $g \circ f$,

$$g \circ f : X \longrightarrow Z$$

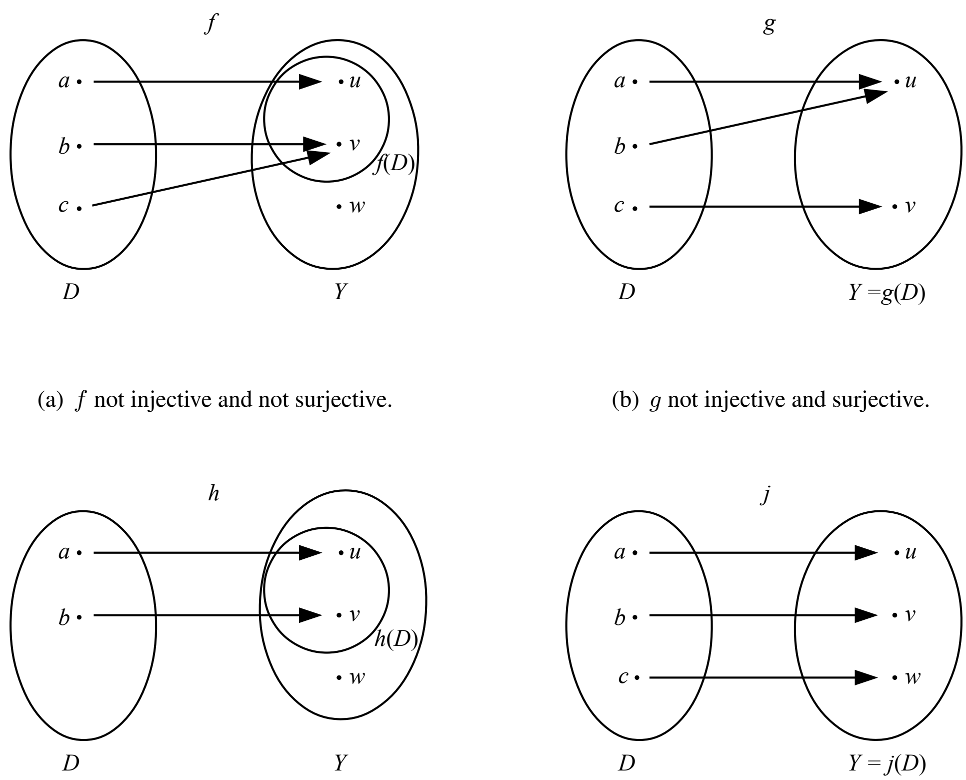
such that $(g \circ f)(x) = g(f(x))$ for all $x \in X$.

Example. Consider the real functions $f(x) = 5x - 1$ and $g(x) = x^2 + 1$, with domain $\mathbb{R}$. The composition of f and g is the function $f \circ g$ defined by

$$(f \circ g)(x) = f(g(x)) = 5(x^2 + 1) - 1 = 5x^2 - 4$$

for all $x \in \mathbb{R}$.

The following definition of function does not make use of the undefined notion of correspondence.



(a) f not injective and not surjective.

(b) g not injective and surjective.

(c) h injective and not surjective.

(d) j bijective.

Figure 8.2. Injectivity and surjectivity.

DEFINITION (Function). A *function* f is a set of ordered pairs (x, y) such that there are not two distinct pairs with the same first element. The set X of all first elements x of the pairs (x, y) of f is called the *domain* of f. The set Y of the second elements y is called the *range* of f. The notation $y = f(x)$ is customarily used to indicate that (x, y) is an element of f.

A function is here understood as the subset of the Cartesian product $X \times Y$ which includes all distinct ordered pairs such that the first element is not the same.

Example. Let us consider the sets $X = \{1, 2, 3\}$ and $Y = \{1, 2\}$. The Cartesian product of X and Y is

$$X \times Y = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2)\}.$$

The set $X \times Y$ is not a function since, for example, to the first element 1 there correspond two distinct second elements, 1 and 2. For the same reason, the subset of $X \times Y$

$$f = \{(1,1), (1,2), (2,2), (3,1)\}$$

is not a function. But the following subsets of $X \times Y$ are functions

$$g = \{(1,1), (2,2), (3,1)\}, \quad h = \{(1,1), (2,2), (3,2)\}, \quad \text{and} \quad j = \{(1,2), (2,1), (3,1)\}.$$

Geometric representation of real functions.

We are interested in a particular class of functions: the *real functions of a real variable*, that is, functions f with domain $D \subseteq \mathbb{R}$ and range $R \subseteq \mathbb{R}$, symbolically,

$$f : D \subseteq \mathbb{R} \longrightarrow \mathbb{R}.$$

Example. The functions g, h, and j in the previous example are real functions of a real variable, with domain $\{1,2,3\} \subseteq \mathbb{R}$ and range $\{1,2\} \subseteq \mathbb{R}$.

For the geometric representation of a function f, we mark on the Cartesian plane the points corresponding to the ordered pairs $(x, f(x))$ (the values of the independent variable x are marked on the first axis, and the values of the dependent variable $y = f(x)$ on the second axis). The geometric representation of f is the set of all those points, and we call it the *graph* of f.

Examples.

(a) The function $g = \{(1,1), (2,2), (3,1)\}$ is represented in Figure 3(a).

(b) Consider the function $i : \mathbb{R} \longrightarrow \mathbb{R}$, with $i(x) = x$ for all $x \in \mathbb{R}$. This function, is designated the *identity function*. The respective graph is sketched in Figure 3(b).

(c) Consider the function $k : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $k(x) = k$ for all $x \in \mathbb{R}$, with $k \in \mathbb{R}$ a constant. A function of this kind is called *constant function* (see Figure 3(c)).

(d) Consider also the function $u : \mathbb{N} \longrightarrow \mathbb{R}$, with $u(n) = 2n - 6$ for all $n \in \mathbb{N}$. Note that it is a particular case of the real-valued functions of a real variable, where the function's domain is the set of all positive integers $\mathbb{N}$. It is the sequence with geometric representation sketched in Figure 3(d).

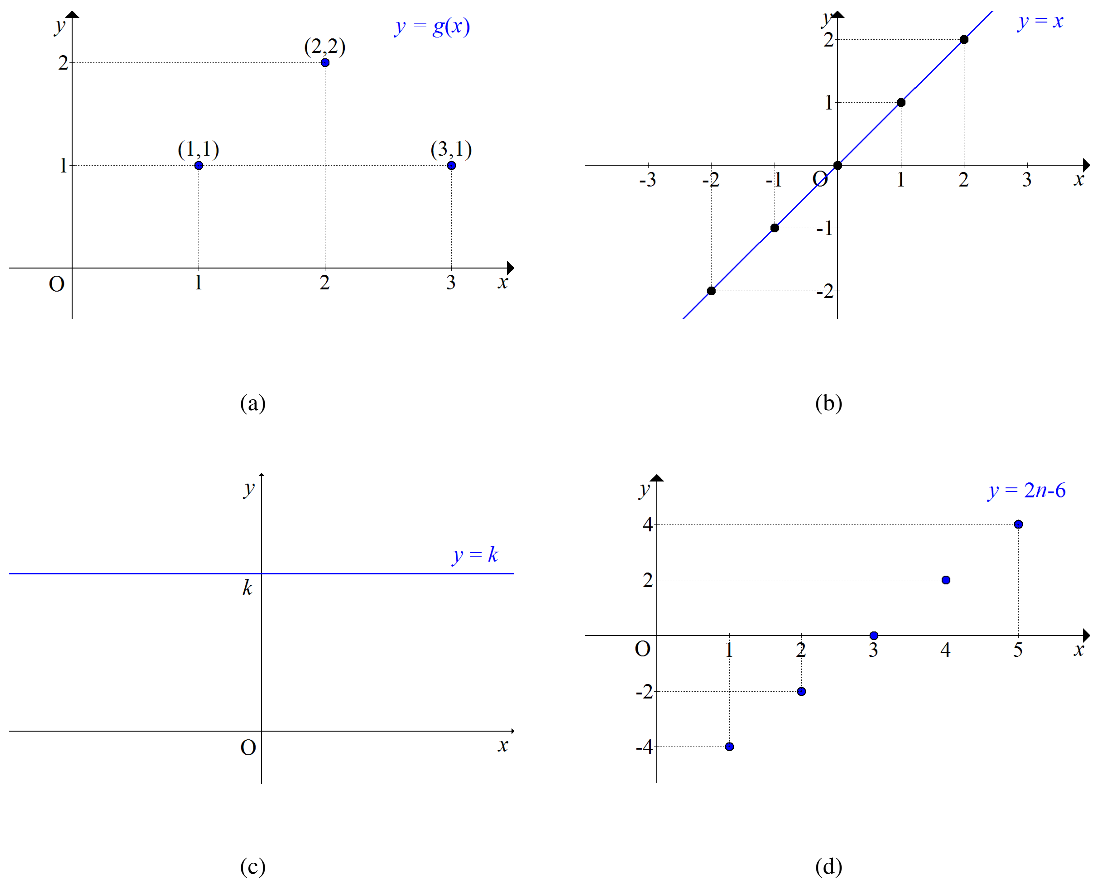


Figure 8.3

Operations on functions.

From given functions new functions can be obtained by using elementary *operations*.

DEFINITION (Operations on functions). Let f and g be real functions with the domain $D \subseteq \mathbb{R}$, k a real number, and n a positive integer. We define the following operations on functions:

(1) Multiplication by a real number: The function kf, the product of k and f, is defined by $(kf)(x) = kf(x)$ if $x \in D$.

(2) Addition: The function $f + g$, the sum of f and g, is defined by $(f + g)(x) = f(x) + g(x)$ if $x \in D$.

(3) Multiplication: The function fg, the product of f and g, is defined by $(fg)(x) = f(x)g(x)$ if $x \in D$.

(4) Division: The function f/g, the quotient obtained by dividing f by g, is defined by $(f/g)(x) = f(x)/g(x)$ if $x \in D$ and $g(x) \neq 0$.

(5) Exponentiation: The function f^n, f raised to the n-th power, is defined by $(f^n)(x) = (f(x))^n$ if $x \in D$.

(6) Radiciation: The function $\sqrt[n]{f}$, the n-th root of f, is defined by $(\sqrt[n]{f})(x) = \sqrt[n]{f(x)}$ if $x \in D$, in the case n is odd, and if $x \in D$ and $f(x) \geq 0\ \forall x \in D$, in the case n is even.

Examples. Consider the real functions $f(x) = 2x$ and $g(x) = x^2 + 1$, with domain $\mathbb{R}$. We are going to construct new functions by using the operations above mentioned.

(a) $(5f)(x) = 5f(x) = 10x$ if $x \in \mathbb{R}$.

(b) $(f + g)(x) = f(x) + g(x) = 2x + x^2 + 1$ if $x \in \mathbb{R}$.

(c) $(fg)(x) = f(x)g(x) = 2x(x^2 + 1) = 2x^3 + 2x$ if $x \in \mathbb{R}$.

(d) $(f/g)(x) = f(x)/g(x) = 2x/(x^2 + 1)$ if $x \in \mathbb{R}$. Note that $g(x) \neq 0\ \forall x \in \mathbb{R}$.

(e) $(f^4)(x) = (f(x))^4 = (2x)^4 = 16x^4$ if $x \in \mathbb{R}$.

(f) $(\sqrt{g})(x) = \sqrt{g(x)} = \sqrt{x^2 + 1}$ if $x \in \mathbb{R}$. Observe that $g(x) \geq 0\ \forall x \in \mathbb{R}$.

Boundedness. Monotonicity. Parity. Zeros.

In the behaviour of a function there are certain properties which are particularly relevant.

DEFINITION (Boundedness). Let f be a real functions with the domain $D \subseteq \mathbb{R}$ and range $R \subseteq \mathbb{R}$. The function f is said to be *bounded* if its range R is a bounded set.

DEFINITION (Monotonicity). Let f be a real functions with the domain $D \subseteq \mathbb{R}$. The function f is said to be an *increasing* function if

$$\forall x_1, x_2 \in D,\, x_2 > x_1 \Rightarrow f(x_2) \geq f(x_1),$$

and *strictly increasing* if, additionally,

$$\forall x_1, x_2 \in D,\, x_2 > x_1 \Rightarrow f(x_2) > f(x_1).$$

If

$$\forall x_1, x_2 \in D,\, x_2 > x_1 \Rightarrow f(x_2) \leq f(x_1),$$

the function is said to be a *decreasing* function, and if, additionally,

$$\forall x_1, x_2 \in D,\, x_2 > x_1 \Rightarrow f(x_2) < f(x_1),$$

the function is called *strictly decreasing*. If f is an increasing or decreasing function, it is said to be *monotonic*; if f is strictly increasing or strictly, it is said to be *strictly monotonic*.

Examples. Consider the real functions of real variable f, g, h, and j, with the same domain $D = \{-3, -1, 1, 3\}$.

(a) If f is defined by $f = \{(-3,-2), (-1,-1), (0,0), (1,1), (3,2)\}$ (see Figure 4(a)) then it is strictly increasing, since $f(3) > f(1) > f(0) > f(-1) > f(-3)$.

(b) If g is defined by $g = \{(-3,2), (-1,1), (0,0), (1,-1), (3,-2)\}$ (see Figure 4(b)) then g is strictly decreasing, as $g(3) < g(1) < g(0) < g(-1) < g(-3)$.

(c) If function h is defined by $h = \{(-3,-2), (-1,0), (0,0), (1,1), (3,1)\}$ (see Figure 4(c)) then it is increasing (but not in the strict sense), since $h(3) = h(1) > h(0) = h(-1) > h(-3)$.

(d) If j is defined by $j = \{(-3,2), (-1,-1), (0,1), (1,-1), (3,2)\}$ (see Figure 4(d)) then j is not monotonic, since, for example, $j(3) > j(1)$ and $j(1) < j(0)$.

Another important property concerning a function respects its *parity*.

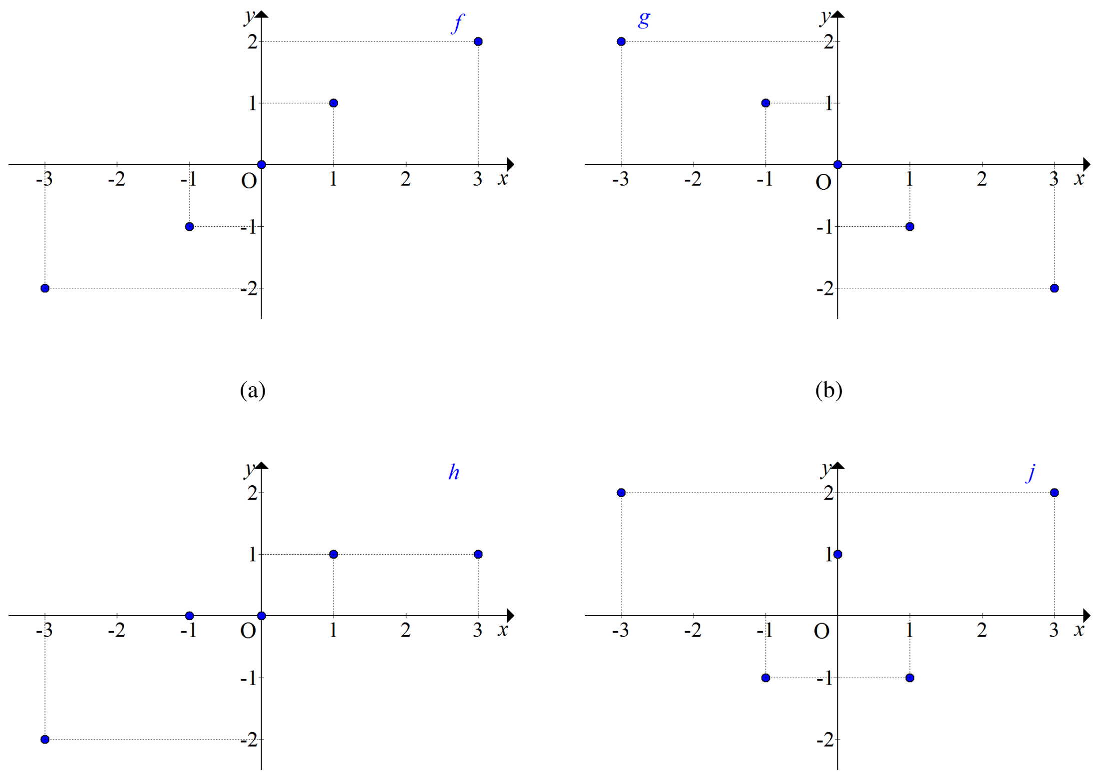
DEFINITION (Parity). A real function f with domain $D \subseteq \mathbb{R}$ is said to be *even* if

$$\forall x \in D,\, -x \in D \wedge f(x) = f(-x),$$

and *odd* if

$$\forall x \in D,\, -x \in D \wedge f(x) = -f(-x).$$

Note that the graph corresponding to an even function is symmetric about the y-axis, and the graph of an odd function is symmetric about the origin of the axes.



(a) (b) (c) (d)

Figure 8.4

Example. Consider the functions f, g, h, and j of the previous example. The functions f and g are odd (see as Figures 4(a) and 4(b)), j is even (Figure 4(d)), and h is neither even nor odd (Figure 4(c)).

Finally, we refer to the existence of *zeros*.

DEFINITION (Zeros). A real function f with domain $D \subseteq \mathbb{R}$ is said to have a *zero* at the point $x_0 \in D$ if $f(x_0) = 0$.

Example. The functions f and g of the previous examples (Figures 4(a) and 4(b)) have only one zero for $x = 0$, the function h (Figure 4(c)) has zero for $x = -1$ and $x = 0$, and j (Figure 4(d)) has no zeros.

Main types of real functions.

We shall present the main types of real functions of a real variable, focusing on some very common elementary functions.

Polynomial functions. Consider the function $p : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$p(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + a_n x^n,$$

where n is a positive integer, and a_k, with $k = 0, 1, 2, \ldots, n$, are constants (called *coefficients*). the expression

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + a_n x^n$$

is called a *polynomial*, and the function p, whose analytic expression is a polynomial, is called a *polynomial function*. If $a_n \neq 0$ the polynomial is said to be of *degree* n, and p a *polynomial function of degree* n. The polynomial functions are the simplest type of so-called *algebraic functions*.

Examples.

(a) The function $q : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $q(x) = x^4 - 5x^2 - 10x + \sqrt{2}$, is polynomial of degree 4.

(b) The real functions of a real variable r and s, defined, respectively, by $r(x) = -2x + 1$ and $s(x) = 3x$, are polynomial functions of degree 1.

(c) The function $t : \mathbb{R} \longrightarrow \mathbb{R}$, with $t(x) = 2x^2 - 2x + 1$, is a polynomial function of degree 2.

(d) The constant function $k : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $k(x) = 7$, is a polynomial function of degree 0.

We shall consider first the particular case of the polynomial functions of degree 0 or 1, called *linear functions*. We begin by the notion of *slope* of a line segment. Choose two distinct points in the plane, (x_1, y_1) and (x_2, y_2), with $x_1 \neq x_2$, (follow the explanation with Figure 8.5). The *slope* (or *gradient* or *angular coefficient*) of the line segment with endpoints (x_1, y_1) and (x_2, y_2) is given by

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

As m is obtained by dividing the lengths of the segments $\overline{CB}$ and $\overline{AC}$, we see that the slope is the tangent of angle α, where α is the *inclination* of segment $\overline{AB}$):

$$m = \tan\alpha = \frac{y_2 - y_1}{x_2 - x_1}.$$

Note that the slope does not depend on the order in which the coordinates are taken, as long as the order is the same for both terms of the fraction:

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2}.$$

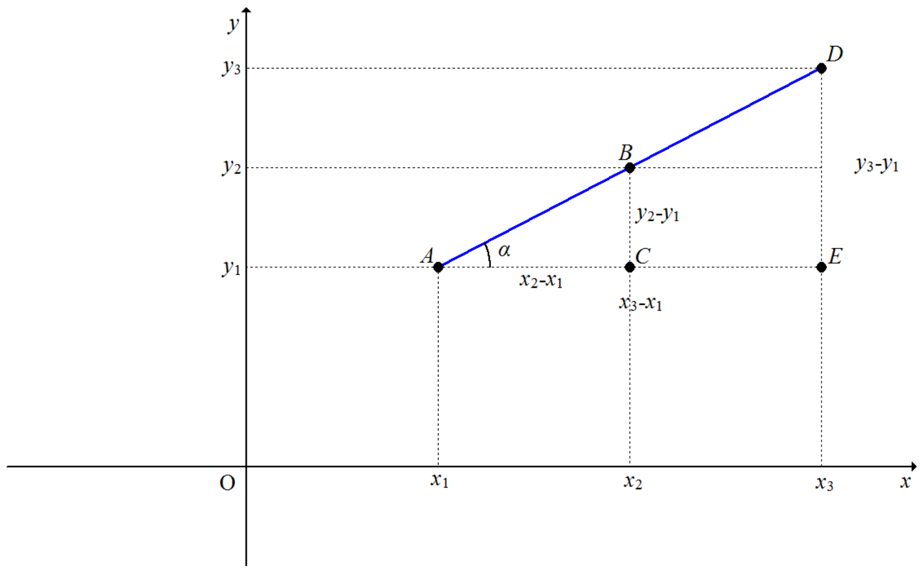


Figure 8.5. Angular coefficient.

Consider also the segment $\overline{AD}$, containing the segment $\overline{AB}$ (see Figure 8.5). This new segment $\overline{AD}$ has the same slope as segment $\overline{AB}$

$$m = \tan\alpha = \frac{y_3 - y_1}{x_3 - x_1} = \frac{y_2 - y_1}{x_2 - x_1},$$

as the numerators and the denominators of the fractions are the lengths of the corresponding sides of the similar triangles ABC, ADE.

In fact, the slope is the same for any segment containing or being contained in segment $\overline{AB}$. Thus, the notion of slope is also meaningful for a straight line: it is the slope of any segment contained in the line.

Take now the following points of the straight line: (x_1, y_1), a fixed point, and (x, y), a moving point taking any position on the straight line. As the slope m is constant for any two distinct points of the line, we have

$$m = \frac{y - y_1}{x - x_1} \Rightarrow y - y_1 = m(x - x_1) \Leftrightarrow y = m(x - x_1) + y_1.$$

The last equation above,

$$y = m(x - x_1) + y_1,$$

is the equation of the straight line *with slope* m *and containing the point* (x_1, y_1).

Note that a straight line is not defined uniquely by its slope: we need also to know a point of the line. In Figure 8.6 there are represented distinct straight line with the same slope: the lines are *parallel*.

Then, we can characterise a straight line as the set of points such that

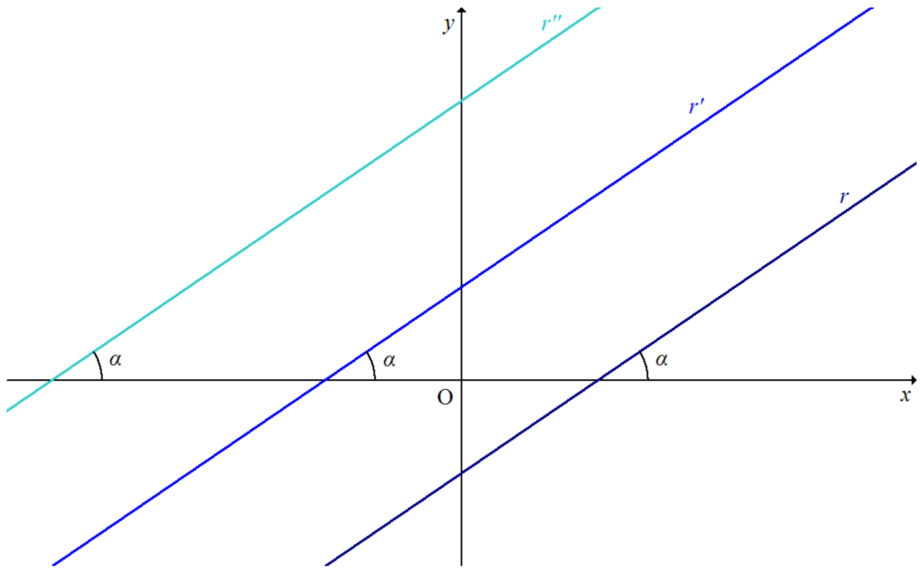


Figure 8.6. Parallel straight lines.

- the slope m of the segments defined by any two distinct points of the set is constant;
- the set contains a given point (x_1, y_1).

For a given straight line, the choice of the point (x_1, y_1) is arbitrary. Nevertheless, it is customary to choose the point $(0, b)$, the intersection of the straight line and the y-axis (b is called the *y-intercept*). We than have

$$y = m(x - 0) + b \Leftrightarrow y = mx + b,$$

where the last equation,

$$y = mx + b,$$

is called the *slope-intercept equation* of the straight line with slope m and y-intercept b (Figure 8.7).

Summarising, the linear functions have analytic expressions of the form $f(x) = mx + b$ with m and b real constants, and are geometrically represented by *slant* (or *oblique*) straight lines if $m \neq 0$ or by *horizontal* straight lines if $m = 0$. The slope m depends on the inclination of the straight line, and b is the y-intercept.

Linear functions are monotonic: strictly increasing if $m > 0$, strictly decreasing if $m < 0$, and constant if $m = 0$. This can be be easily checked. For example, for the case where $m > 0$, with x_1 and x_2 arbitrary real numbers, we have

$$x_2 > x_1 \Rightarrow mx_2 > mx_1 \Rightarrow mx_2 + b > mx_1 + b$$

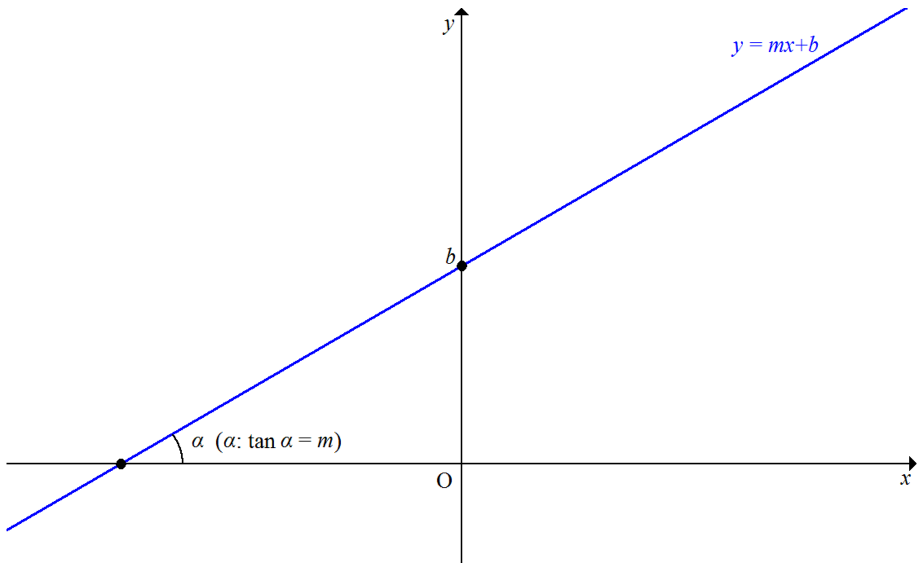


Figure 8.7. Straight line: slope-intercept equation.

and, consequently, $x_2 > x_1 \Rightarrow f(x_2) > f(x_1)$, what sows that f is strictly increasing. The absolute value of m determines "how steep" the straight line is. In fact, from

$$|f(x_2) - f(x_1)| = |(mx_2 + b) - (mx_1 + b)| = |m| \cdot |x_1 - x_2|$$

we see that the distance between the images of x_1 and x_2 (fixed) under function f increases with $|m|$.
The relation between the inclination of a straight line and both the sign and the absolute value of its slope can be observed in Figure 8.8.
With respect to the range and the existence of zeros, we have two distinct situations. If $m \neq 0$, the linear function linear has range $\mathbb{R}$ and only one zero:

$$f(x) = 0 \Leftrightarrow mx + b = 0 \Leftrightarrow x = -\frac{b}{m}.$$

If $m = 0$, that is, if the function is constant, its range is the singular set $\{b\}$. There exist zeros only if $b = 0$ and, in this case, all real values of x are zeros of f.
Finally, a linear function is odd if its analytic expression is of the form $f(x) = mx$, that is, if $b = 0$:

$$f(-x) = m(-x) = -mx = -f(x), \ \ \forall x \in \mathbb{R}.$$

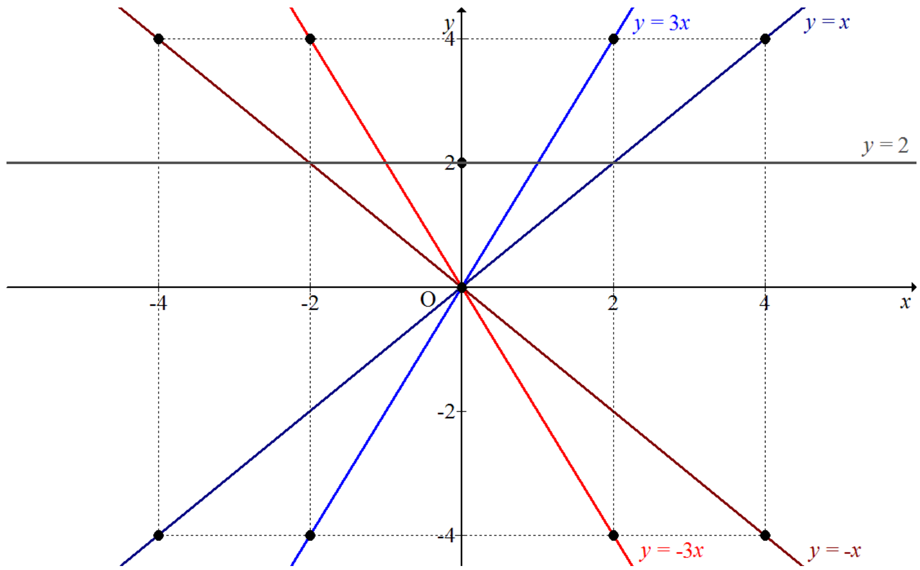


Figure 8.8. Straight lines with various slopes.

The function is even if $m = 0$:

$$f(-x) = b = f(x), \ \ \forall x \in \mathbb{R}.$$

In the study above, we did not find *vertical straight lines* as geometric representation of linear functions. In fact, vertical straight lines do not represent functions. These lines are the representation of formulas of the type $x = a$, with $a \in \mathbb{R}$, where to the value $x = a$ there correspond an infinite number of images.

Examples. The illustration for the following examples can be found in Figure 8.9.

(a) The function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x + 2$ has range $\mathbb{R}$. Solving $x + 2 = 0 \Leftrightarrow x = -2$, we obtain the zero: $x = -2$. The y-intercept is $b = 2$ (it is the function value for $x = 0$: $f(0) = 0 + 2 = 2$). As $m = 2 > 0$, the line slants uphill.

(b) The function $g : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $g(x) = -2x - 4$ has range $\mathbb{R}$ and the zero $x = -2$. The y-intercept is $b = -4$. As the slope is negative ($m = -4 < 0$), the straight line slants downhill.

(c) The functions of a real variable h and j defined by the expressions $h(x) = 4$ and $j(x) = 0$ are constant functions with range $\{4\}$ and $\{0\}$, respectively. The function h has no zeros, and the function j has an infinite number of zeros: $j(x) = 0 \ \forall x \in \mathbb{R}$). The y-intercepts of h and j are $b = 4$ and $b = 0$, respectively. Note that the geometrical representation of the function j is the x-axis.

(d) The formulas $x = -6$ and $x = 0$ are geometrically represented by vertical straight lines. The formula $x = 0$ corresponds to the y-axis.

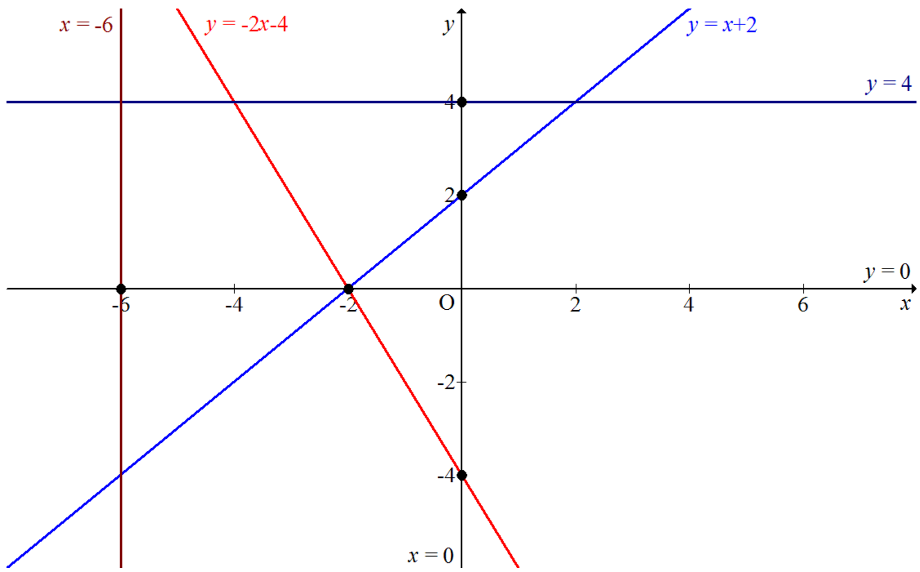


Figure 8.9

We now review briefly the particular case of the polynomial functions of degree 2, designated *quadratic functions*, and corresponding geometric representation.
Firstly, we consider the particular case of the quadratic functions $f : \mathbb{R} \longrightarrow \mathbb{R}$, with analytic expression

$$f(x) = ax^2,$$

where $a \neq 0$ is a constant.
The function is represented geometrically by a *parabola*, with graphs of the type shown in Figure 8.10, depending on $a > 0$ or $a < 0$. The vertical straight line e marked in the figure is the axis of symmetry (the function f is even since $f(x) = f(-x)$ for all real value of x), and the point V is the vertex of the parabola. The graph of f opens upward if $a > 0$ and downward if $a < 0$. Also, the parabola is thinner if the absolute value of a is larger. Notice that the parabola differs from the straight line one fundamental aspect:

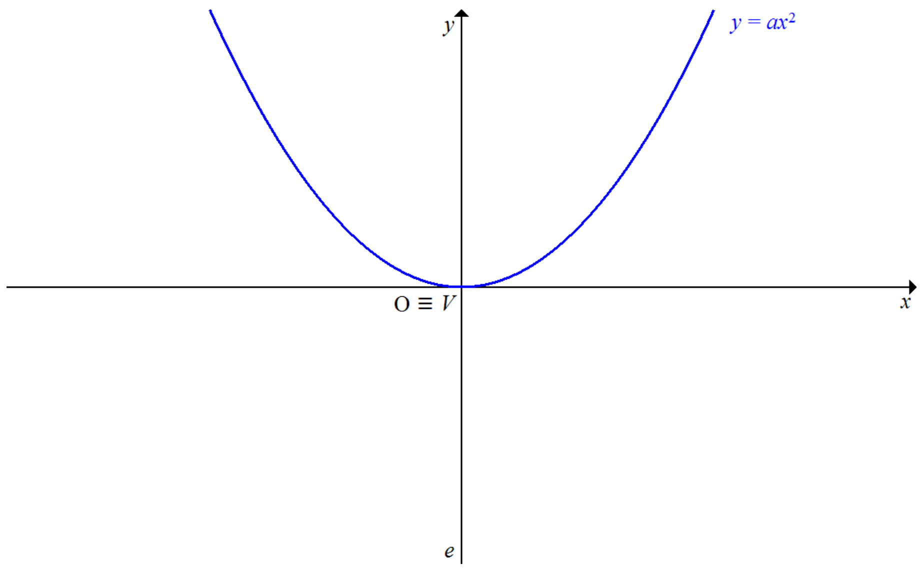


Figure 8.10. Parabola.

- In a straight line the slope is constant, so that to a variation of the independent variable x there corresponds a proportional variation of the dependent variable y (with constant of proportionality m). In absolute value, we have

$$|y_2 - y_1| = |(mx_2 + b) - (mx_1 + b)| = |m| \cdot |x_1 - x_2|.$$

- For a parabola, the absolute value of the variation

$$|y_2 - y_1| = |(a(x_2)^2 - (a(x_1)^2)| = |a| \cdot |(x_1)^2 - (x_2)^2| = |a| \cdot |x_1 + x_2| \cdot |x_1 - x_2|$$

depends not only on the coefficient a but also on the factor $|x_1 + x_2|$, that is, on the values of x.

The function f is not monotonic: it as an increasing section and a decreasing one. For example, if $a > 0$, for $x_1, x_2 > 0$ we have

$$x_2 > x_1 \Rightarrow (x_2)^2 > (x_1)^2 \Rightarrow a(x_2)^2 > a(x_1)^2 \Rightarrow f(x_2) > f(x_1),$$

what shows that f is strictly increasing in the interval $(0, +\infty)$. If $x_1, x_2 < 0$,

$$x_2 > x_1 \Rightarrow (x_2)^2 < (x_1)^2 \Rightarrow a(x_2)^2 < a(x_1)^2 \Rightarrow f(x_2) < f(x_1)$$

and we have that f is strictly decreasing in the interval $(-\infty, 0)$. If $a < 0$, the conclusions with respect to monotonicity are the opposite to the ones above. In Figure 8.11 we represent some parabolas with equation $y = ax^2$ for various values of a.

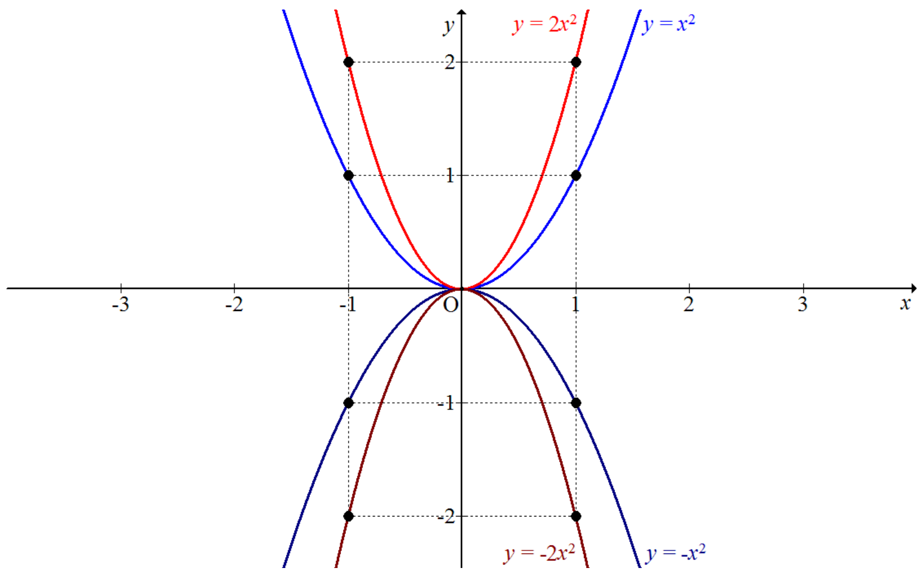


Figure 8.11

So we have that quadratic functions of the form $f(x) = ax^2$ are geometrically represented by parabolas. In fact, any quadratic function $f(x) = ax^2 + bx + c$ is represented by a parabola (the equation $y = ax^2 + bx + c$ is called the *standard form equation of the parabola*).

Example. Consider the equation

$$y = x^2 - 12x + 38.$$

We have

$$y = x^2 - 12x + 38 \Leftrightarrow y = \left(x^2 - 12x + 36\right) - 36 + 38 \Leftrightarrow y = (x-6)^2 + 2 \Leftrightarrow y - 2 = (x-6)^2,$$

and, after the change of variables $X = x - 6$, $Y = y - 2$, we obtain the equation

$$Y = X^2.$$

Geometrically, this means that the parabola equation $y = x^2 - 12x + 38$ can be obtained from the parabola $y = x^2$ by translation: 6 units to the right and 2 units upward (see the Figura 8.12). Note also the symmetry axis of the parabola $y = x^2 - 12x + 38$ is the vertical straight line with equation $x = 6$, and that the vertex of this parabola is the point $(6, 2)$.

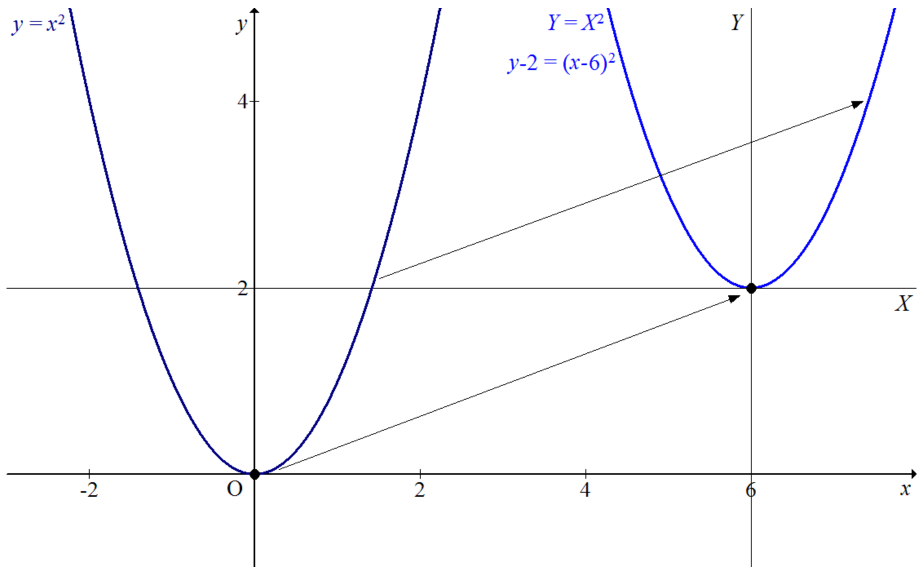


Figure 8.12

Generically, for a parabola with equation $y = ax^2 + bx + c$, the vertex is the point defined as

$$V = \left(-\frac{b}{2a}, -\frac{b^2 - 4ac}{4a}\right).$$

The range of the function $f(x) = ax^2+bx+c$ can be visualised immediately from its geometric representation: $[-(b^2 - 4ac)/(4a), +\infty)$ if $a > 0$ or $(-\infty, -(b^2 - 4ac)/(4a)]$ if $a < 0$.

Finally, we study the intersections with the axes. The possible zeros of the quadratic equation are obtained by solving the equation $ax^2 + bx + c = 0$: the function has 2 zeros, 1 zero or none depending on the coefficient's values.

- If $b^2 - 4ac > 0$, there are 2 zeros:

$$x = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad x = \frac{-b + \sqrt{b^2 - 4ac}}{2a};$$

- If $b^2 - 4ac = 0$, it has one zero:

$$x = -\frac{b}{2a};$$

- If $b^2 - 4ac < 0$ the function has no real zeros.

The y-intercept is c (it is obtained from $f(0) = a \times 0 + b \times 0 + c = c$).

Examples. The following examples are represented in Figure 8.13.

(a) The function $f : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $f(x) = x^2 + 1$, has for geometric representation a parabola opening upward (since $a = 1 > 0$), with vertex

$$V = \left(-\frac{0}{2 \times 1}, \frac{4 \times 1 \times 1 - 0}{4 \times 1} \right) = (0, 1),$$

and y-intercept $c = 1$ (it is the value of the function at $x = 0$: $f(0) = 0 + 1 = 1$). The range of the function is $[1, +\infty)$, and there are no zeros as

$$b^2 - 4ac = 0 - 4 \times 1 \times 1 = -4 < 0.$$

Note that the y-intercept could obtained directly from the vertex coordinates $(0, 1)$,

(b) The function $g : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $g(x) = -2x^2$, is geometrically represented ny a parabola opening downward ($a = -2 < 0$), with vertex

$$V = \left(-\frac{0}{2 \times (-2)}, \frac{4 \times (-2) \times 0 - 0}{4 \times (-2)} \right) = (0, 0),$$

and y-intercept $c = 0$. The function has range $(-\infty, 0]$, and a unique zero:

$$x = \frac{0 \pm \sqrt{0 - 4 \times (-2) \times 0}}{2 \times (-2)} = 0.$$

Notice that the values for the zero and the y-intercept can be obtained from the coordinates of the vertex: $(0, 0)$,

(c) Consider the function $h : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $h(x) = -x^2 + 10x - 21$. The function is geometrically represented by a parabola opening downward ($a = -1 < 0$), with vertex

$$V = \left(-\frac{10}{2 \times (-1)}, \frac{4 \times (-1) \times (-21) - 10^2}{4 \times (-1)} \right) = (5, 4),$$

and y-intercept $c = h(0) = 0 + 0 - 21 = -21$. The function's range is $(-\infty, 4]$, and there are 2 zeros:

$$\begin{aligned} -x^2 + 10x - 21 = 0 \Leftrightarrow x &= \frac{-10 \pm \sqrt{10^2 - 4 \times (-1) \times (-21)}}{2 \times (-1)} \\ \Leftrightarrow x &= \frac{-10 \pm \sqrt{16}}{-2} \Leftrightarrow x = \frac{-10 \pm 4}{-2} \Leftrightarrow x = 3 \vee x = 7. \end{aligned}$$

Note that, by rearranging the function's analytic expression, we have

$$\begin{aligned} h(x) &= -x^2 + 10x - 21 = -(x^2 - 10x + 21) = -(x^2 - 2 \times 1 \times 5x + 25 - 25 + 21) \\ &= -(x-5)^2 + 4, \end{aligned}$$

what makes clear that the graphic of the function h can be obtained from the graphic of $y = -x^2$ by translation: 5 units to the right and 4 upward. The parabola's vertex can be obtained directly from this last form of the function's analytic expression.

(d) Consider the function $j : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $j(x) = (1/2)x^2 - 5x + 21/2$. The graphic of the function is a parabola opening upwards ($a = 1/2 > 0$), with vertex

$$V = \left(-\frac{-5}{2 \times \frac{1}{2}}, \frac{4 \times \frac{1}{2} \times \frac{21}{2} - (-5)^2}{4 \times \frac{1}{2}} \right) = (5, -2)\,,$$

and y-intercept $c = j(0) = 0 + 0 + 21 = 21$. The function has range $[-2, +\infty)$, and 2 zeros:

$$(1/2)x^2 - 5x + 21/2 = 0 \Leftrightarrow x = \frac{5 \pm \sqrt{(-5)^2 - 4 \times \frac{1}{2} \times \frac{21}{2}}}{2 \times \frac{1}{2}} \Leftrightarrow x = 3 \vee x = 7.$$

Following the same steps as for the previous example, we obtain a different form for the function's analytic expression

$$\begin{aligned} j(x) &= \frac{1}{2}x^2 - 5x + \frac{21}{2} = \frac{1}{2}(x^2 - 10x + 21) = \frac{1}{2}(x^2 - 2 \times 1 \times 5x + 25 - 25 + 21) \\ &= \frac{1}{2}(x-5)^2 - 2, \end{aligned}$$

what makes clear that the graphic of j can be obtained from the graphic of $y = (1/2)x^2$ by translation: 5 units to the right and and 2 units downward.

Finally, we refer briefly to the so called *power functions*, that is, functions $f\mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$f(x) = x^p,$$

with p a positive integer. As particular cases, we have the linear function $f(x) = x$ and the quadratic function $f(x) = x^2$.

Power functions have a unique zero zero for $x = 0$. If p is odd, the range is $\mathbb{R}$, and the function is odd:

$$f(-x) = (-x)^p = -x^p = -f(x), \;\; \forall x \in \mathbb{R}.$$

If p is even, the range is the set $[0, +\infty)$, and the function is even:

$$f(-x) = (-x)^p = x^p = f(x), \;\; \forall x \in \mathbb{R}.$$

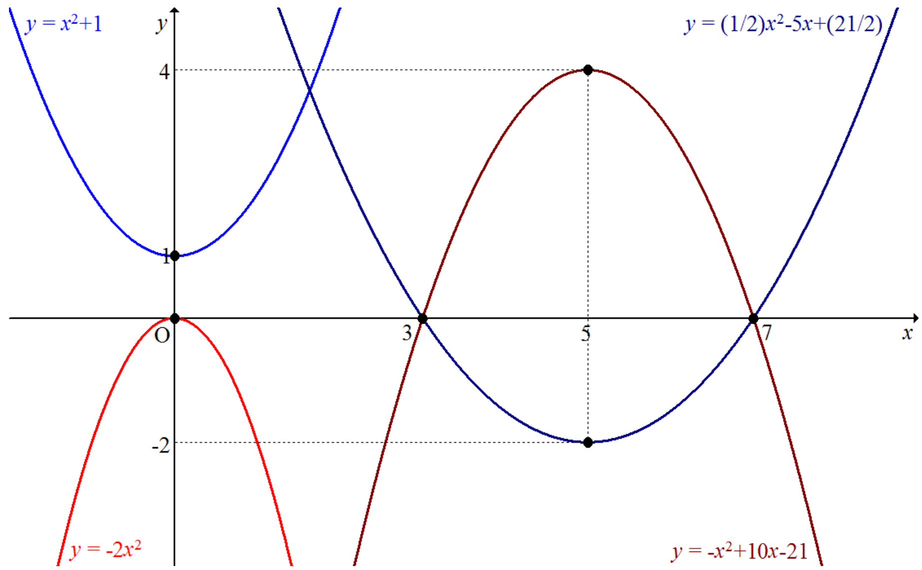


Figure 8.13

The graphic of the power functions depends on the value of the parameter p. If $p = 1$, the function is geometrically represented by a straight line bisecting the odd quadrants. If $p > 1$, the graphics can by grouped in two sets by generic similitude of shape, depending on p being even or odd. In Figure 8.14 we can find the graphics of power functions for a few values of p.

Rational functions. Consider the real function of a real variable r, defined by

$$r(x) = \frac{p(x)}{q(x)},$$

where $p(x)$ and $q(x)$ are polynomials, and $q(x)$ is not the zero polynomial. A function of this type, that is, a function whose analytic expression can be written as a quotient of polynomials, is called *rational*. The domain of r is the set

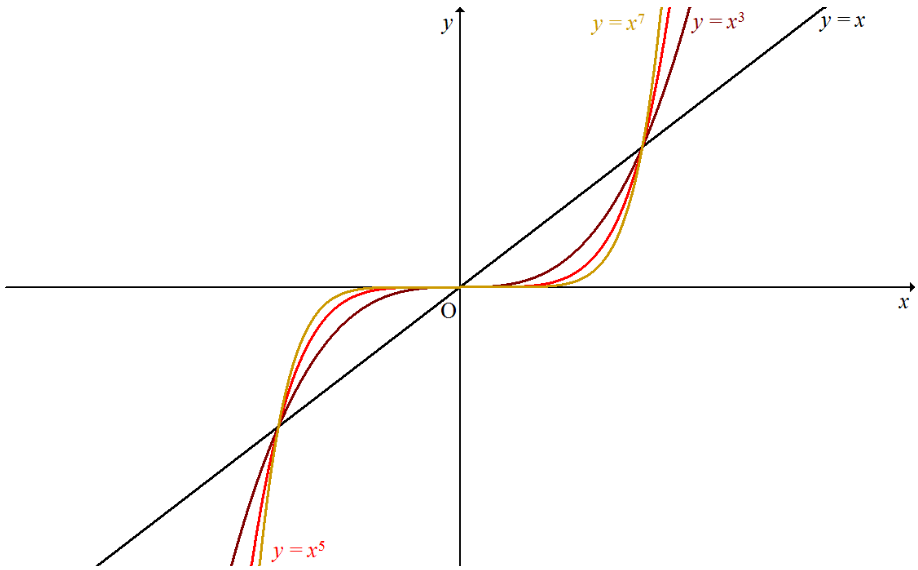
$$D = \{x \in \mathbb{R} : q(x) \neq 0\}.$$

Note that polynomial functions are a particular case of the rational functions: they are obtained from r taking q as a zero degree polynomial ($q(x) = k$, with k a real constant).

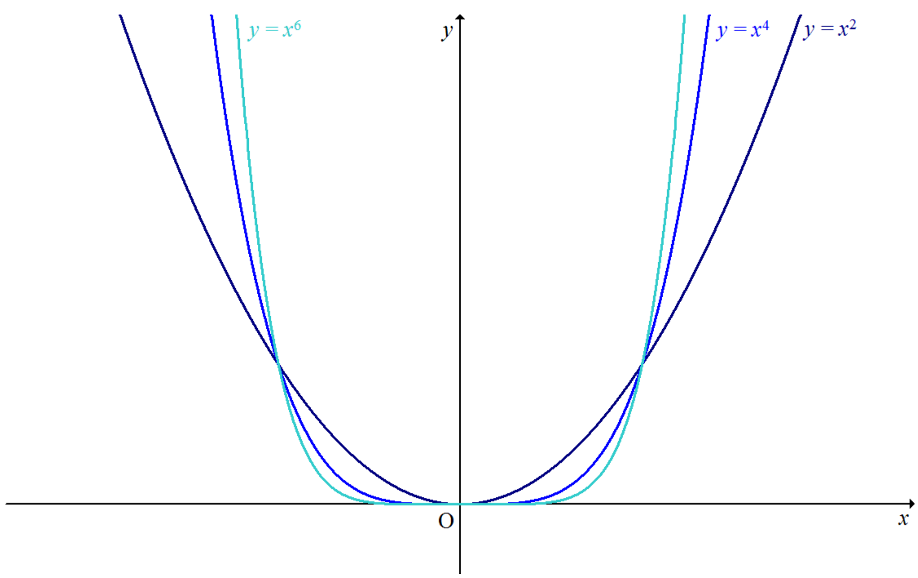
Examples. The following functions are rational:

(a) The function $r : \mathbb{R} - \{1\} \longrightarrow \mathbb{R}$, defined by $r(x) = (x^4 - 10x + 2)/(x - 1)$.

(b) The function $s : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $s(x) = (2x^2 - 2x + 1)/(x^2 + 1)$.



(a) $y = x^p$ with p odd.



(b) $y = x^p$ with p even.

Figure 8.14. Power functions.

(c) The function $t : \mathbb{R} - \{-1, 1\} \longrightarrow \mathbb{R}$, defined by $t(x) = 3/(1 - x^2)$.

(d) The polynomial function $u : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $u(x) = 7x - 1$.

(e) The function $v : \mathbb{R} - \{2\} \longrightarrow \mathbb{R}$, with analytic expression $v(x) = (x + 1)^2 - 2/(2 - x)$. Note that $v(x)$ can be written as the quotient of polynomials:

$$v(x) = (x+1)^2 - \frac{2}{2-x} = \frac{(x+1)^2(2-x) - 2}{2-x} = \frac{3x - x^3}{2-x}.$$

The shape of the graphic of a rational function varies very much with the particular analytic expression of the function. We study two very simple cases.

Firstly, we consider the real function of a real variable r, defined by

$$r(x) = \frac{1}{x}.$$

This function is not defined for $x = 0$: its domain is the set $D = \mathbb{R} - \{0\}$. The function is odd, since

$$r(-x) = \frac{1}{-x} = -\frac{1}{x} = -r(x), \;\; \forall x \in D,$$

has no zeros, and takes positive values if $x > 0$ and negative values if $x < 0$.

The function r is not monotonic: it suffices to see that the images of -1, 1, and 2 are, respectively, $r(-1) = -1$, $r(1) = 1$, and $r(2) = 1/2$, and that $r(-1) < r(1) > r(2)$. Nevertheless, r is strictly decreasing in the intervals $(-\infty, 0)$ and $(0, +\infty)$. This is very simple to check: for $x_1, x_2 > 0$, for example, we have

$$x_2 > x_1 \Rightarrow \frac{1}{x_2} < \frac{1}{x_1} \Rightarrow r(x_2) < r(x_1),$$

what shows that r is strictly decreasing in $(0, +\infty)$.

We can determine the behaviour of the function when x approaches zero and infinity. If $x > 0$, the function approaches zero when x grows to infinity, and grows to infinity when x approaches zero. If $x < 0$, the function approaches zero when x moves away from zero and decreases without limit when x approaches zero.

The geometric representation of r can be found in Figure 8.15(a). Note that the graphic of the function approaches the y-axis (the vertical straight line with equation $x = 0$) when x approaches zero: the straight line $x = 0$ is said to be a vertical *asymptote* of the graph of the function at $x = 0$. Finally, we can see what is the range of the function by simple observation of its graph: $r(D) = \mathbb{R} - \{0\}$.

Note also that the graph of the real function of a real variable s, defined by

$$s(x) = \frac{1}{x-1}$$

can be obtained by translation of the graph of r one unit to the right (see Figure 8.15(b)). The domain of s is the set $\mathbb{R} - \{1\}$, and the function's graph has the vertical asymptote vertical $x = 1$. The function s is neither even nor odd.

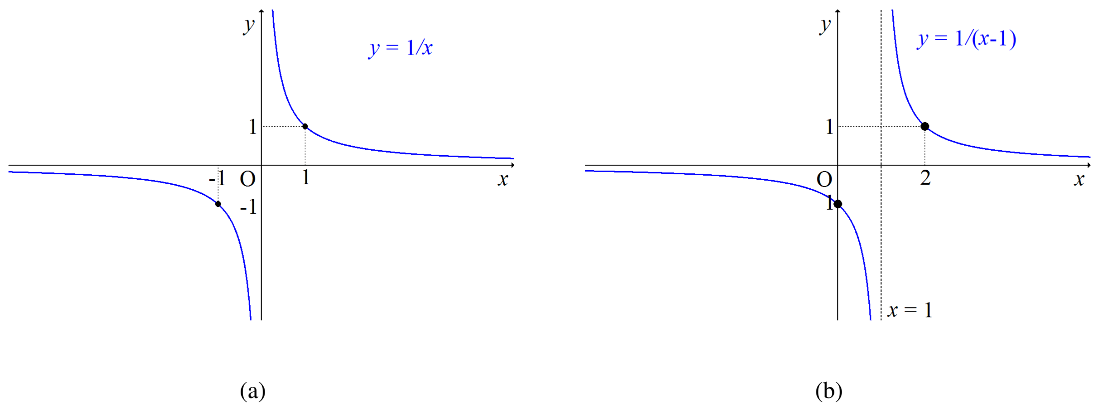


Figure 8.15

Consider now the real function of a real variable t, defined by

$$t(x) = \frac{1}{x^2}.$$

The function has domain $D = \mathbb{R} - \{0\}$, is even since

$$t(-x) = \frac{1}{(-x)^2} = \frac{1}{x^2} = t(x), \quad \forall x \in D,$$

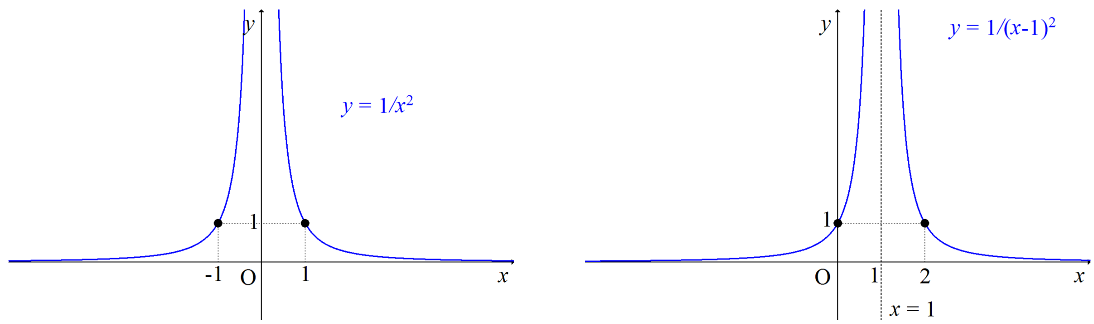
has no zeros, and takes only positive values. The function t is not monotonic: it is strictly increasing in $(-\infty, 0)$, and strictly decreasing in $(0, +\infty)$. For example, for $x_1, x_2 > 0$,

$$x_2 > x_1 \Rightarrow \frac{1}{x_2} < \frac{1}{x_1} \Rightarrow \left(\frac{1}{x_2}\right)^2 < \left(\frac{1}{x_1}\right)^2 \Rightarrow \frac{1}{(x_2)^2} < \frac{1}{(x_1)^2} \Rightarrow t(x_2) < t(x_1),$$

what proves that t is strictly decreasing in $(0, +\infty)$. t approaches zero when x grows in absolute value, and increases without limit when x approaches zero: the y-axis is an asymptote of the function's graph.
The geometric representation of t can be found in Figure 8.16(a). The Figure 8.16(b) represents geometrically the function with analytic expression

$$u(x) = \frac{1}{(x-1)^2}.$$

The graph of u can be obtained from the graph of t translating it to the right by one unit.



(a) (b)

Figure 8.16

Irrational functions. Executing *elementary operations* (addition, subtraction, multiplication, division, taking powers roots) on polynomial functions we obtain new functions called *algebraic*. Naturally, rational function are algebraic. Algebraic functions which are not rational are designated *irrational*.

Examples.
The following functions are algebraic.

(a) The function $a(x) = \sqrt{x^2 - 1}$, with domain

$$D = \{x \in \mathbb{R} : x^2 - 1 \geq 0\} = (-\infty, -1) \cup (1, +\infty).$$

(b) The functions $b(x) = (x + 2)/\sqrt{x - 1}$, with domain

$$D = \{x \in \mathbb{R} : \sqrt{x - 1} \neq 0 \wedge x - 1 \geq 0\} = (1, +\infty).$$

(c) The function $c(x) = \sqrt[3]{(x^4 - 10x + 2)/(x - 1)}$, whose domain is the set

$$D = \{x \in \mathbb{R} : x - 1 \neq 0\} = \mathbb{R} - \{1\}.$$

(d) The function $d(x) = \sqrt[2]{(x^2 + 1)/(1 - x^2)}$, with domain

$$D = \left\{x \in \mathbb{R} : \frac{x^2 + 1}{1 - x^2} \geq 0 \wedge 1 - x^2 \neq 0\right\} = (-1, 1).$$

(e) The function $e(x) = \sqrt[3]{1 - \sqrt{x + 1}}$, whose domain is the set

$$D = \{x \in \mathbb{R} : x + 1 \geq 0\} = [-1, +\infty).$$

(f) The polynomial function $f(x) = x^4 - x + 3$, with $D = \mathbb{R}$.

Next, we shall consider some very simple algebraic functions.
We begin by considering the *absolute value function*, defined by

$$f(x) = |x|.$$

It is an irrational function as $|x| = \sqrt{x^2}$.
The brief study of the function can be done easily, given that

$$f(x) = |x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0. \end{cases}$$

Observe that the absolute value function has domain $\mathbb{R}$, and also that it is an even function:

$$f(-x) = |-x| = |x| = f(x), \ \forall x \in \mathbb{R}.$$

Figure 8.17 shows the geometric representation of the function. This figure shows also the graphs of some other functions whose graphs can be obtained form the graph of $y = |x|$ by translation and, in the case of $y = -|x-4|$, additionally by symmetry about the x-axis.

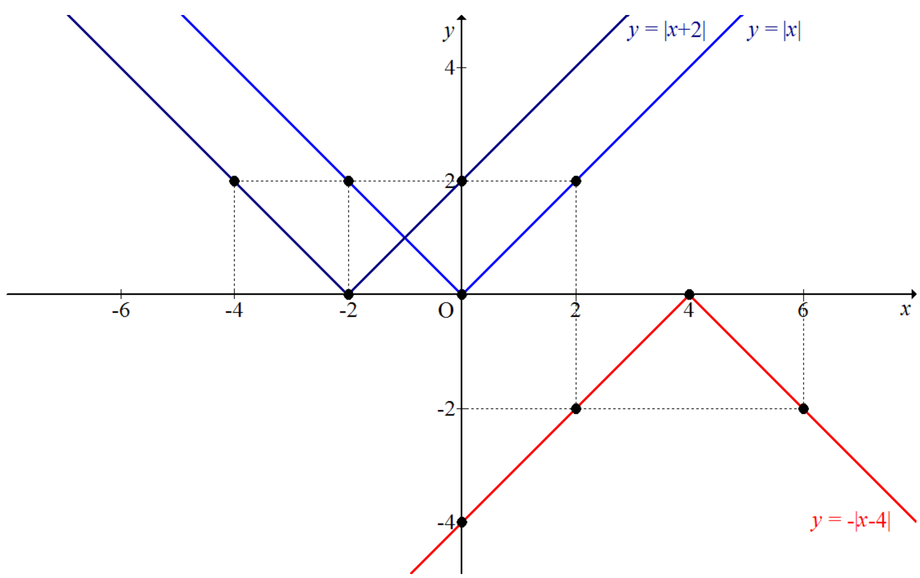


Figure 8.17. Functions involving the absolute value.

We can find irrational algebraic functions in a current situation. The locus of all points (x, y) in the plane whose distance to a fixed point (a, b) is constant and equal to $r > 0$ is defined by the condition:

$$d((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2} = r.$$

This is nothing elese that the well-known equation of the *circle* with center $C = (a, b)$ and radius r, which write, equivalently,

$$(x-a)^2 + (y-b)^2 = r^2.$$

The corresponding geometric representation can be found in Figure 8.18(a).
This formula establishes a correspondence between the variables y and x, but does not define y as a function of x since the correspondence is not univoque. Observe that, for example, the vertical straight line $x = a$ intersects the circle at two distinct points (see Figure 8.18(b)): $x = a$ has two distinct images ($b-r$ and $b+r$).

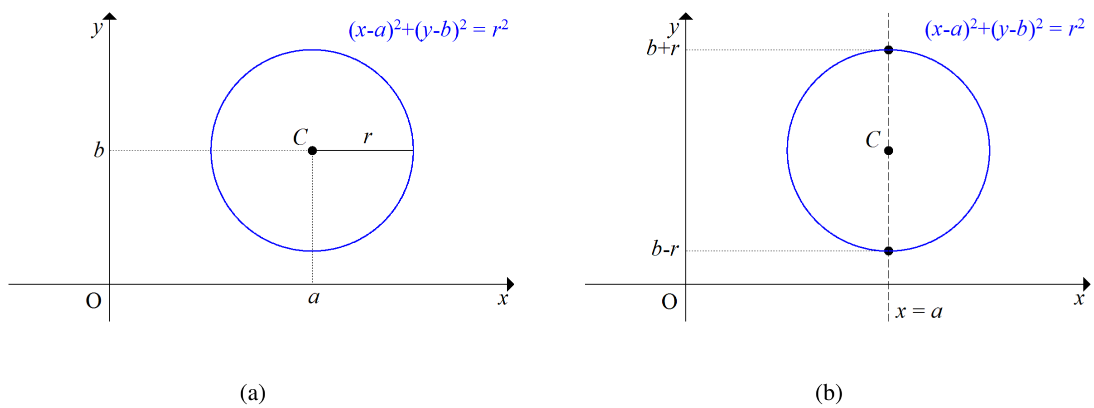


Figure 8.18. Circle.

But, if we consider separately the two semicircles defined by the straight line $y = b$ then y can bow be expressed as a function of x. We illustrate the procedure with an example.
Consider the circle (represented in Figure 8.19) with center $C = (6, 2)$ and radius 4, so that its equation is

$$(x-6)^2 + (y-2)^2 = 16.$$

To isolate y in the first member of the equation, we follow the usual procedures:

$$\begin{aligned}(x-6)^2 + (y-2)^2 = 16 &\Leftrightarrow (y-2)^2 = 16 - (x-6)^2 \\ &\Leftrightarrow y - 2 = -\sqrt{16-(x-6)^2} \vee y - 2 = \sqrt{16-(x-6)^2} \\ &\Leftrightarrow y = 2 - \sqrt{16-(x-6)^2} - 3 \vee y = 2 + \sqrt{16-(x-6)^2}.\end{aligned}$$

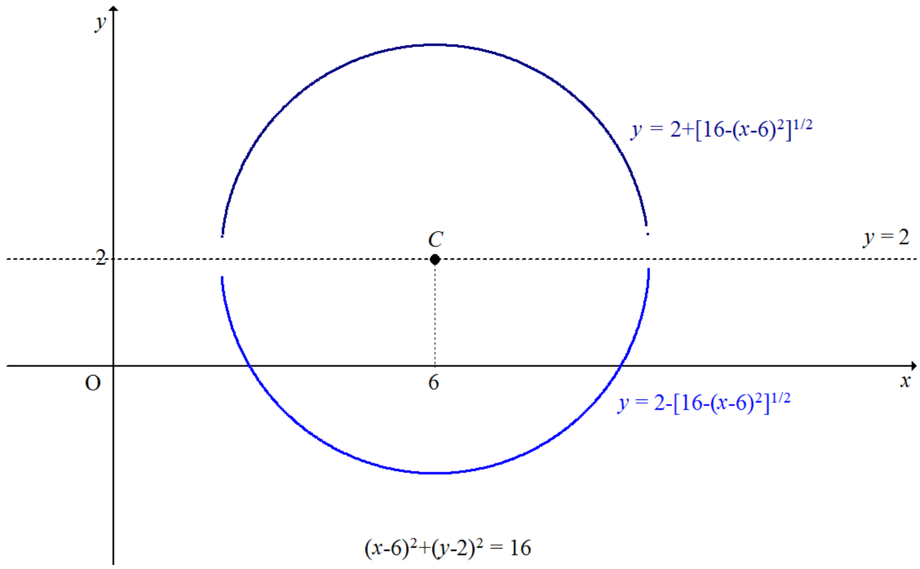


Figure 8.19

Thus, the semicircles below and above the straight line $y = 2$ are defined, respectively, by the irrational algebraic functions with analytic expressions

$$f(x) = 2 - \sqrt{16 - (x-6)^2} - 3 \quad \text{and} \quad g(x) = 2 + \sqrt{16 - (x-6)^2},$$

as shown in Figure 8.19.

Finally, we consider the case of parabolas with horizontal symmetry axis. These lines correspond, generically, to equations of the form

$$x = ay^2 + by + c,$$

with a, b, $andc$ constants, and $a \neq 0$. If $a > 0$ the parabola opens to the right, and if $a < 0$ it opens to the left (see Figure 8.20).

As for the circles, these equations do not define the variable y as a function of x: there are values of x to which there correspond two different values of y. But each one of the “half-parabolas” defined by the symmetry axis is the graph of a function. We illustrate with an example.

Consider the line represented in Figure 8.21, corresponding to the equation

$$x = 2y^2.$$

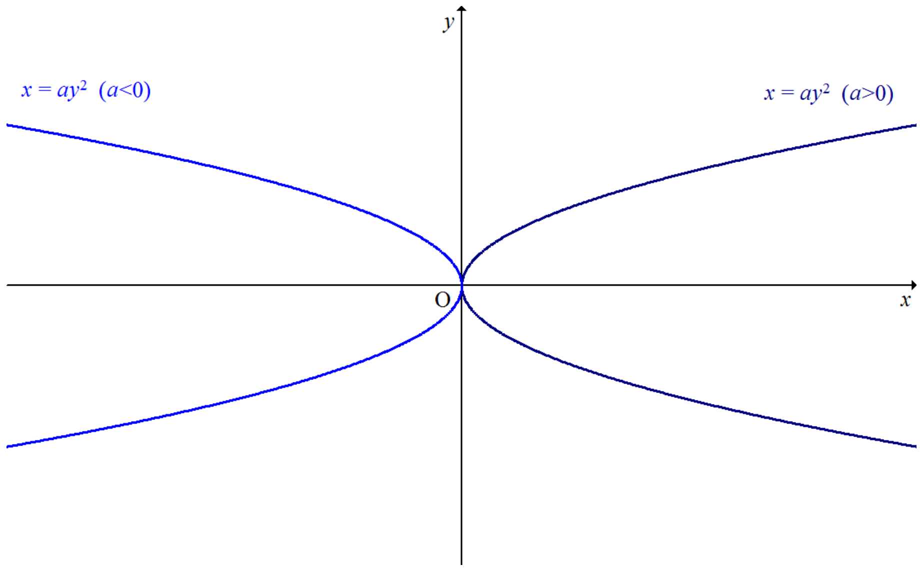


Figure 8.20

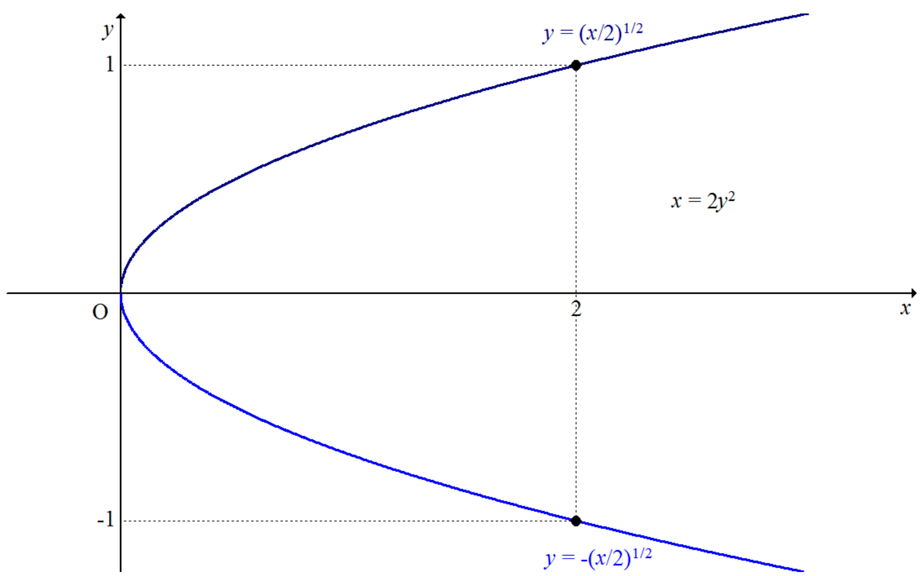


Figure 8.21

Isolating y in the left-hand side of the equation, we obtain

$$x = 2y^2 \Leftrightarrow y^2 = \frac{x}{2} \Leftrightarrow y = -\sqrt{\frac{x}{2}} \vee y = \sqrt{\frac{x}{2}}.$$

The "half-parabolas" below and above the symmetry axis $y = 0$ represent, respectively, the irrational algebraic functions with analytic expressions

$$f(x) = -\sqrt{\frac{x}{2}} \quad \text{and} \quad g(x) = \sqrt{\frac{x}{2}},$$

as shown in Figure 8.21.

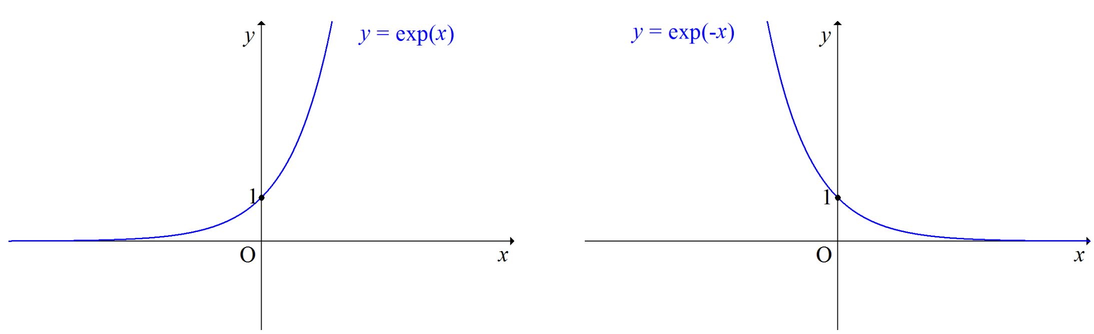
Transcendental functions. As seen above, a function is algebraic if it is obtained by elementary operations on polynomial functions. The real functions of a real variable which are not algebraic are called *transcendental*. The following are examples of transcendental functions:

- The exponential function $\exp x$;
- The logarithmic function $\ln x$;
- The trigonometric functions $\sin x$, $\cos x$, $\tan x$, and $\cot x$;
- The inverse trigonometric functions $\arcsin x$, $\arccos x$, $\arctan x$ and $\operatorname{arccot} x$.

We shall make very brief comments on a few basic types of transcendental functions.
Consider the *exponential function* f, defined by

$$f(x) = \exp x = e^x,$$

where $e \simeq 2.718281828459045235360287$ is the Euler number. The corresponding geometric representation can be found in Figure 8.22(a).



(a) (b)

Figure 8.22

The function has domain $\mathbb{R}$, and range $(0, +\infty)$ (therefore, having no zeros). It is a strictly increasing function, with y-intercept $f(0) = e^0 = 1$.
From its graph, we van observe the behaviour of the function when x approaches infinity: the function grows to infinity if $x > 0$, and approaches zero if $x < 0$.
The geometric representation of the function g, related to f, and defined by

$$g(x) = \exp(-x) = e^{-x},$$

can be found in Figure 8.22(b).
Functions with analytic expression of the form a^x, with $a > 0$ and $a \neq 1$, have geometric representation with shape similar to those of the functions $f(x) = \exp x$ or $g(x) = \exp(-x)$, respectively if $a > 1$ or $0 < a < 1$ (in Figure 8.23 we can find the geometric representation of functions in each of these cases).

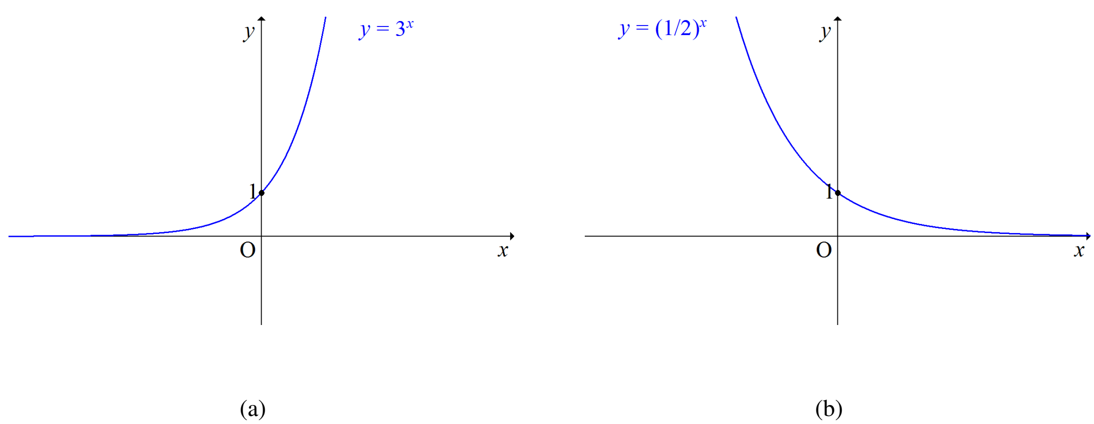


Figure 8.23

Note that

$$g(x) = \exp(-x) = e^{-x} = \frac{1}{e^x} = \left(\frac{1}{e}\right)^x,$$

so that the analytic expression of g is just a particular case of a^x with $0 < a < 1$.
Exponential functions with basis a, with $a > 0$ and $a \neq 1$, are strictly monotonic, thus injective. Then, we can define their inverse functions, designated *logarithmic with basis* a. In general, if $f(x) = a^x$ we write

$$f^{-1}(x) = \log_a x,$$

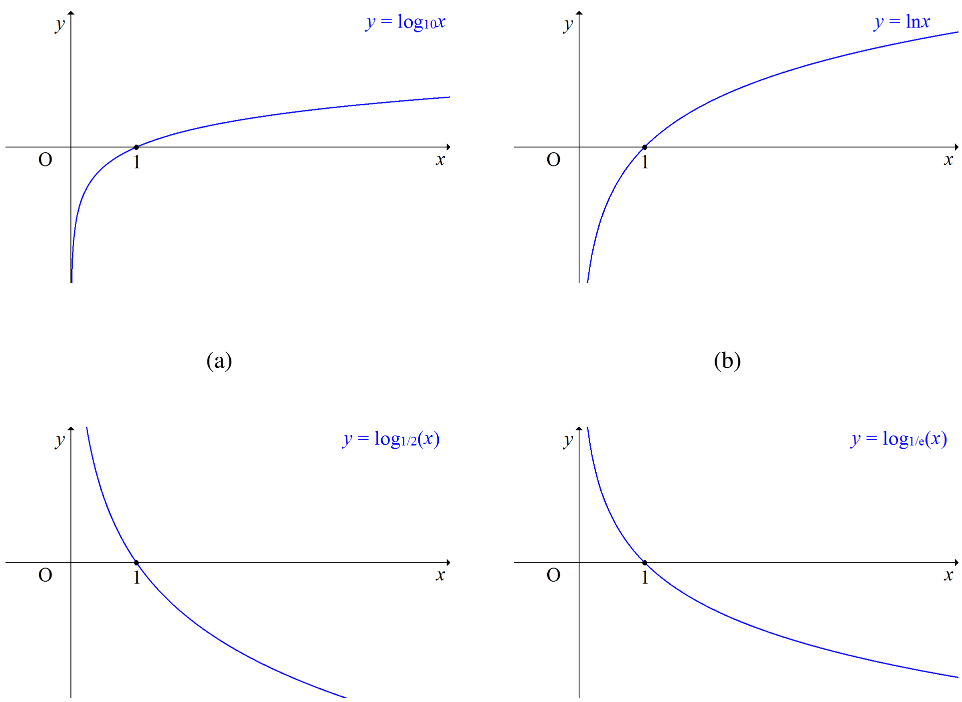
having the logarithmic function domain $(0, +\infty)$, and range $\mathbb{R}$. One particular case is the logarithmic function with basis e (the logarithms with basis e are called *natural*, and are denoted by $\log_e x = \ln x$). Then, for all $x \in \mathbb{R}$ and $y \in (0, +\infty)$,

$$y = a^x \Leftrightarrow x = \log_a y,$$

with $a > 0, a \neq 1$. For example,

$$8 = 2^3 \Leftrightarrow 3 = \log_2 8.$$

In Figure 8.24, we present the geometric representation of a few logarithmic functions, for various bases.



(a) (b) (c) (d)

Figure 8.24

We let to the reader to see what is the behaviour of the the logarithmic function when x approaches zero and infinity, depending on the value of the basis a: $a > 1$ or $0 < a < 1$.

Finally, we recall the rules for operating with logarithms. Let x, y, z, and a be arbitrary real numbers, with x, y, and a positive, and $a \neq 1$. Then

- $\log_a(xy) = \log_a x + \log_a y$,
- $\log_a\left(\frac{x}{y}\right) = \log_a x - \log_a y$,
- $\log_a(x^z) = z \log_a x$.

The following equation establishes the correspondence between logarithms with distinct bases. Let x, a, and b be arbitrary real numbers, with x, a, and b positive, and $a, b \neq 1$. Then

- $\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$.

Consider now the trigonometric functions and their inverses.
The values and the behaviour of the trigonometric functions can be learnt by using the *unit circle* (with radius 1) (see Figure 8.25).

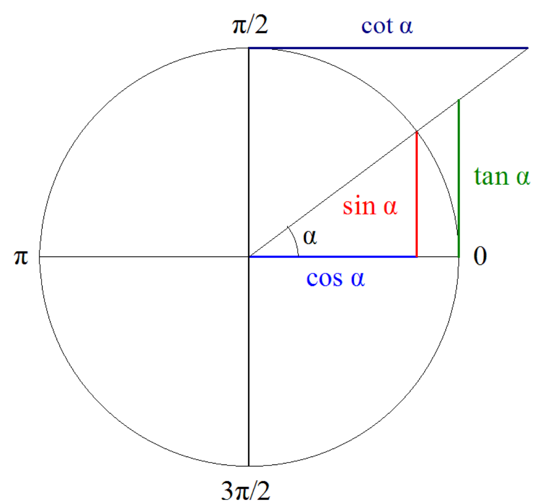


Figure 8.25. The unit circle.

In particular, we give the values of the trigonometric functions for some particular values of the variable x:

x	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$
$\sin x$	0	$1/2$	$\sqrt{2}/2$	$\sqrt{3}/2$	1
$\cos x$	1	$\sqrt{3}/2$	$\sqrt{2}/2$	$1/2$	0
$\tan x$	0	$\sqrt{3}/3$	1	$\sqrt{3}$	∞
$\cot x$	∞	$\sqrt{3}$	1	$\sqrt{3}/3$	0

The functions $\sin x$ and $\cos x$ have domain $\mathbb{R}$, and range $[-1, 1]$. The functions $\tan x$ and $\cot x$ are not defined for $x = \pi/2 + k\pi$ and $x = k\pi$, respectively, with $k \in \mathbb{Z}$, both having range $\mathbb{R}$. With respect to parity, $\cos$ is even, and the remaining trigonometric functions are odd.
It is, however, one other attribute that distinguishes the trigonometric functions from the previous ones we considered: their *periodicity*. In fact, These functions satisfy the condition

$$f(x + p) = f(x),$$

for any value of x in the domain of the function, and for a fixed real number p. p is called the function's *period*, and has value 2π for the functions $\sin x$ and $\cos x$, and π for the functions $\tan x$ and $\cot x$.
In Figure 8.26 we can find the geometric representation of the trigonometric functions.

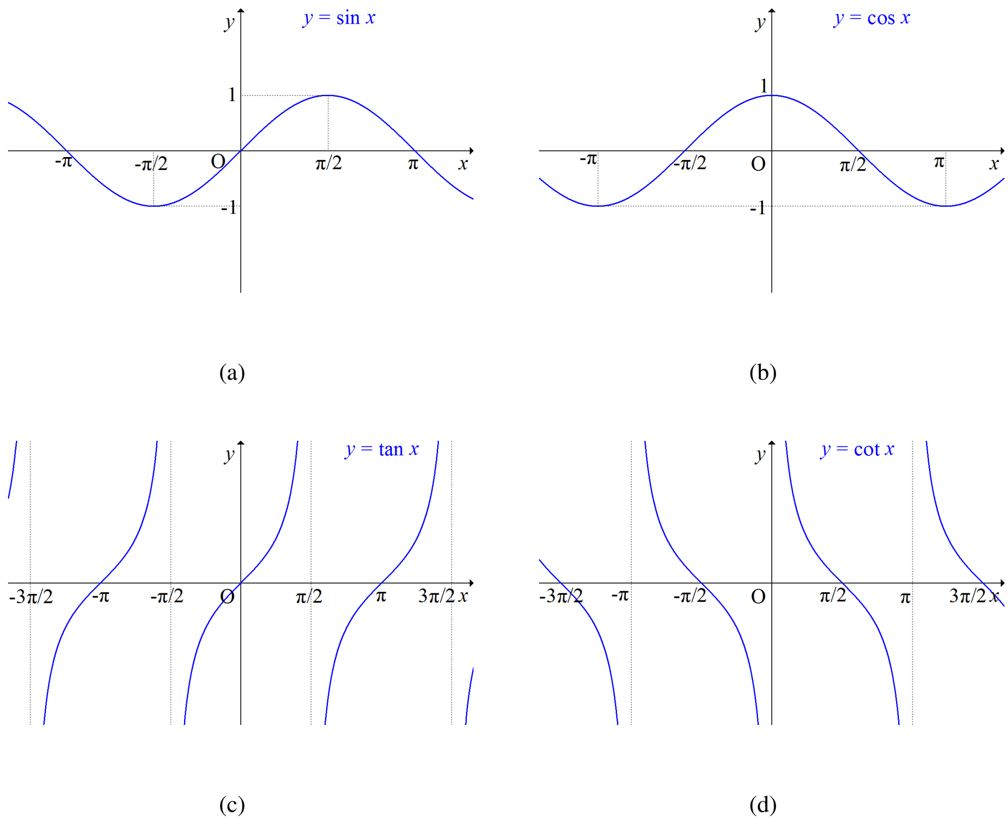


Figure 8.26. Trigonometric functions.

The trigonometric functions are not injective. Therefore, we can only obtain inverses by restricting the functions to subsets of their domains where there is injectivity. The *standard domain restrictions* are

- $[-\pi/2, \pi/2]$ for the function $\sin x$,
- $[0, \pi]$ for the function $\cos x$,
- $(-\pi/2, \pi/2)$ for the function $\tan x$,
- $(0, \pi)$ for the function $\cot x$.

Inverting the trigonometric functions restricted to the above sets we obtain the so called *inverse trigonometric functions*, respectively, the functions $\arcsin x$, $\arccos x$, $\arctan x$, and $\operatorname{arccot} x$. For example, for the function $\sin x$, if $x \in [-\pi/2, \pi/2]$ and $y \in [-1, 1]$,

$$y = \sin x \Leftrightarrow x = \arcsin y.$$

If, for example, $x = \pi/4$ and $y = \sqrt{2}/2$, we obtain

$$\frac{\sqrt{2}}{2} = \sin\frac{\pi}{4} \Leftrightarrow \frac{\pi}{4} = \arcsin\frac{\sqrt{2}}{2}.$$

The geometric representation of the inverse trigonometric functions can be found in Figure 8.27.

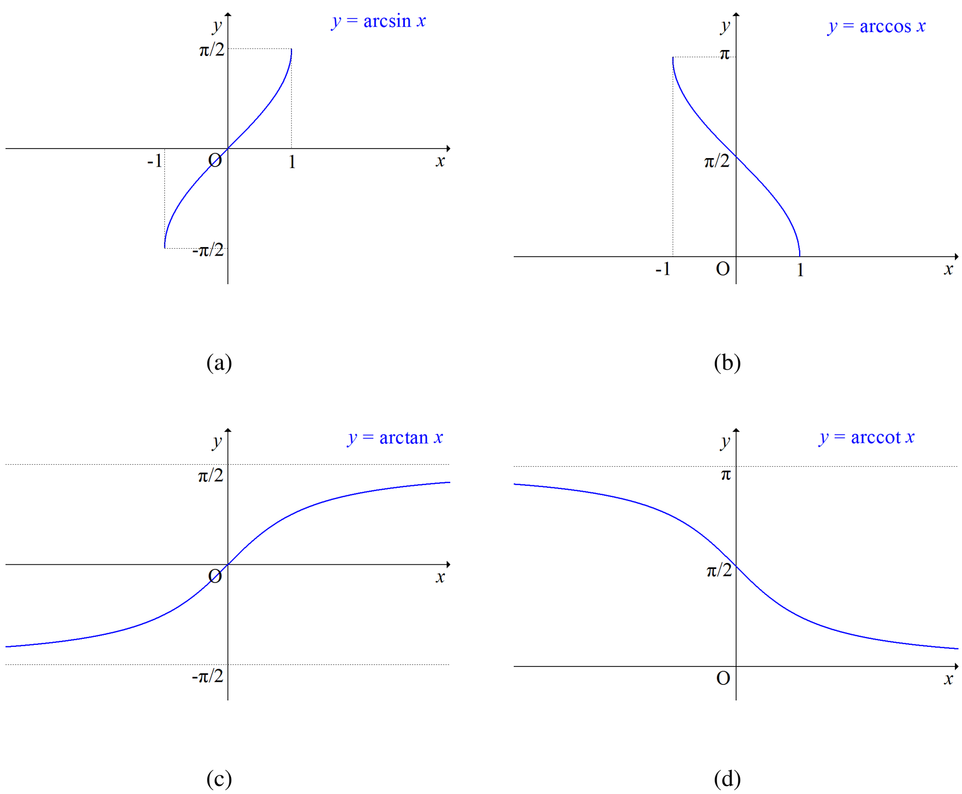


Figure 8.27. Inverse trigonometric functions.

Exercises.

(1) Determine the domain of the functions with analytic expressions:

a) $\sqrt{x+1}$; b) $\ln(1-x)$; c) $\dfrac{1}{\sqrt{x^2-4}}$; d) $\sqrt{9-x^2}$;

e) $\sqrt{e^{x^2}-1}$; f) $\ln(x^2-16)$; g) $\sqrt{|x-2|-4}$; h) $\ln(\ln x)$;

i) $\dfrac{1}{\ln(\ln x)-1}$; j) $\dfrac{2}{\sqrt{2-|x-1|}}$; k) $\sqrt[4]{|1-\sqrt{x}|}$; l) $\dfrac{1}{1-\tan\frac{x}{2}}$;

m) $\dfrac{1}{1-\cot x}$; n) $\log_2\left(4-\sqrt{25-x}\right)$; o) $\dfrac{\sqrt[3]{x^2-2x-3}}{\ln|x+5|}$.

(2) Determine the domain, range, zeros, monotonicity, and parity of the following real functions, and sketch their geometric representation:

a) $a(x)=x+2$; b) $b(x)=-2x+1$; c) $c(x)=-\dfrac{1}{2}x-1$;

d) $d(x)=x^2-1$; e) $e(x)=x^2-4x+3$; f) $f(x)=-2x^2+\dfrac{3}{2}x-1$;

g) $g(x)=x^2+1$; h) $h(x)=-x^2-2$; i) $i(x)=\dfrac{1}{x-2}$;

j) $j(x)=-\dfrac{1}{x+2}$; k) $k(x)=\dfrac{1}{(x-2)^2}$; l) $l(x)=1+\dfrac{1}{x}$;

m) $m(x)=|x-2|$; n) $n(x)=-|x-2|$; o) $o(x)=2+|x+1|$.

(3) Determine the zeros and the y-intercepts of the real functions:

a) $f(x)=-x+5$; b) $g(x)=2-x^2+x$;

c) $h(x)=-x^2(3x-1)(1-x^2)$; d) $i(x)=\dfrac{x^2-16}{(x^2-5x+4)(x^2+3)}$;

e) $j(x)=\log_2(2x-4)$; f) $k(x)=\dfrac{4e^{2x}-4e^x-3}{e^x+5}$.

(4) Solve the following equations:

a) $2\cos(2x)=1$; b) $\tan\dfrac{x}{2}=\sqrt{3}$; c) $\arcsin(3x)=\pi$; d) $2^x=16$;

e) $2^{2x}=4$; f) $e^{3x+1}=1$; g) $e^{3x+1}=e^{x-1}$.

(5) Solve the following inequalities:

a) $2^x \geq 16$;

b) $\left(\frac{1}{2}\right)^{x+1} < 16$;

c) $e^{3x+1} > 1$;

d) $\ln(3x+1) > 1$;

e) $\log_{10}(4x-3) > \log_{10}\left(x^2\right)$;

f) $\frac{\log_{\frac{1}{2}}\left(x^2+x\right)}{\log_{\frac{1}{2}} x} < 0$.

9. Limits and continuity

The notion of limit.

The notion of *limit* of a function plays a central role in Calculus.

Informally, the notion of limit regards the behaviour of a function f when the independent variable x "approaches" a certain point a. To make the notion meaningful we will assume that the point a is a limit point of the domain of f (so that any neighbourhood of a contains points in the domain of f, distinct from a).

DEFINITION (Limit of a function). Let f be a real function of a real variable, a a limit point of D, the domain of f, and L a real number. We say that the *limit of f as x approaches a is L* (or that "f goes to L when x goes to a") and write

$$\lim_{x \to a} f(x) = L$$

if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - L| < \varepsilon$$

for all points $x \in D$ with $x \neq a$ for which

$$|x - a| < \delta.$$

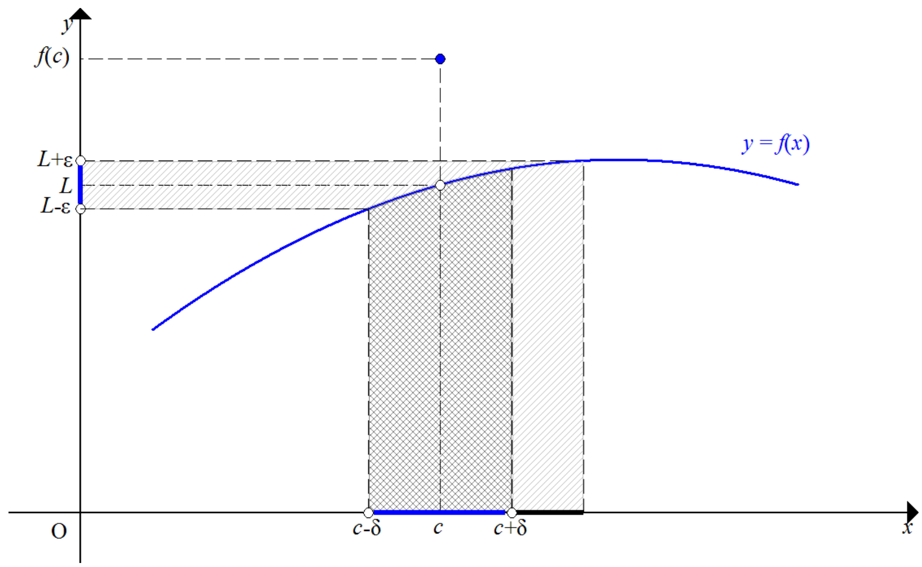
Note that, in the definition above,

$$|x - a| < \delta \Leftrightarrow x \in N_\delta(a) \quad \text{and} \quad |f(x) - L| < \varepsilon \Leftrightarrow f(x) \in N_\varepsilon(L).$$

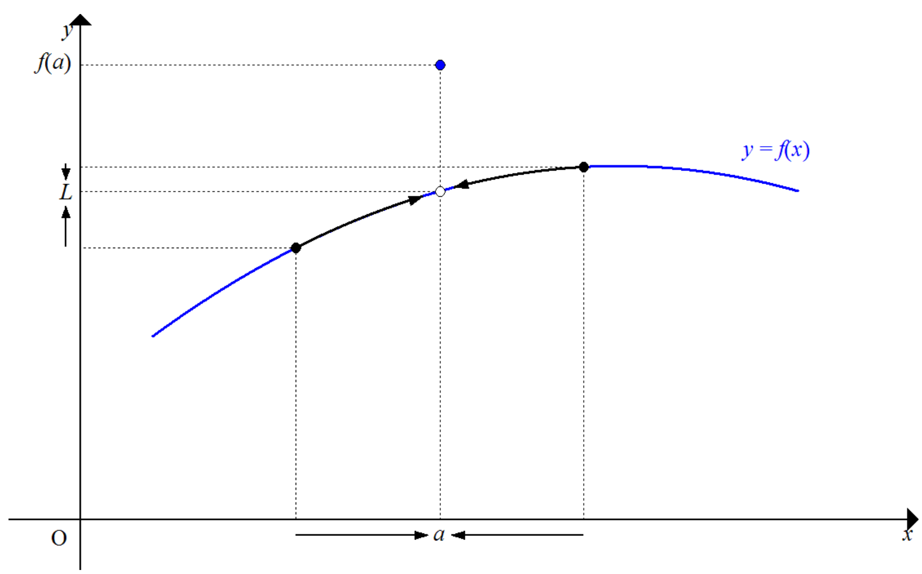
Figure 9.1(a) illustrates the definition of limit. Notice that whatever is the neighbourhood of L we consider (in the sense that however small is the radius ε of the neighbourhood) it is always possible to find a neighbourhood of a with radius δ such that for points of the domain of f,

$$f\left(N_\delta(a) - \{a\}\right) \subseteq N_\varepsilon(L).$$

In Figure 9.1(b) we can find a different illustration of the notion of limit: the second coordinate of a point moving along the graph of f approaches L when its first coordinate approaches a. This interpretation is related to the following theorem.



(a)



(b)

Figure 9.1. Limit of a function.

THEOREM 9.1. *Let f be a real function of a real variable, a a limit point of D, the domain of f, and L a real number. Then*

$$\lim_{x\to a} f(x) = L$$

if and only if

$$\lim_{n\to+\infty} f(a_n) = L$$

for every sequence $\{a_n\}$ in D such that

$$a_n \neq a \ \forall n \in \mathbb{N} \quad \textit{and} \quad \lim_{n\to+\infty} a_n = a.$$

An important property of the limit is that it is unique. We can show this by reductio ad absurdum. Assume that a function f has two distinct limits, L_1 and L_2, when x approaches a. According to the definition of limit, for each $\varepsilon > 0$ there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$|f(x) - L_1| < \varepsilon \quad \text{and} \quad |f(x) - L_2| < \varepsilon$$

for all points $x \in D$ with $x \neq a$ for which, respectively,

$$|x - a| < \delta_1 \quad \text{and} \quad |x - a| < \delta_2.$$

Since the radius ε of the neighbourhoods of L_1 and L_2 is arbitrary, we can choose it so that the neighbourhoods $N_\varepsilon(L_1)$ and $N_\varepsilon(L_2)$ are disjoint:

$$N_\varepsilon(L_1) \cap N_\varepsilon(L_2) = \emptyset.$$

But this means that a point x (in D and distinct from a) belonging both to $N_{\delta_1}(a)$ and $N_{\delta_2}(a)$ has its image $f(x)$ contained in both the sets $N_\varepsilon(L_1)$ and $N_\varepsilon(L_2)$, what is impossible (see Figure 9.2). Consequently, the limit, if it exists, is unique.

THEOREM 9.2. *Let f be a real function of a real variable, a a limit point of the domain of f, and L_1 and L_2 real numbers. If*

$$\lim_{x\to a} f(x) = L_1 \quad \textit{and} \quad \lim_{x\to a} f(x) = L_2$$

then

$$L_1 = L_2.$$

Example. Consider the real function of a real variable f defined by

$$f(x) = \begin{cases} \frac{1}{2}x + 1, & x \neq 4 \\ 2, & x = 4. \end{cases}$$

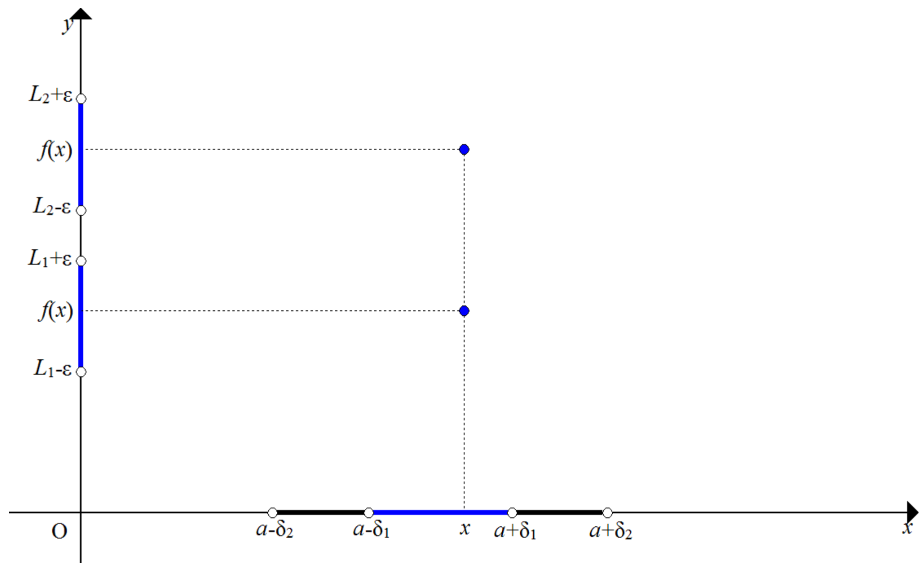


Figure 9.2

The function (with graph sketched in Figure 9.3) has domain $D = \mathbb{R}$. We want to prove that $\lim_{x\to 4} f(x) = 3$. According to definition of limit, we need to establish a relation between the radiuses ε and δ of the neighbourhoods $N_\varepsilon(3)$ and $N_\delta(4)$ so that for $x \neq 4$,

$$x \in N_\delta(4) \Rightarrow f(x) \in N_\varepsilon(3)$$

or, equivalently,

$$|x-4| < \delta \Rightarrow |f(x)-3| < \varepsilon.$$

Since, for $x \neq 4$,

$$|f(x)-3| = \left|\left(\frac{1}{2}x+1\right)-3\right| = \left|\frac{1}{2}x-2\right| = \frac{1}{2}\cdot|x-4|$$

and

$$\frac{1}{2}\cdot|x-4| < \varepsilon \Leftrightarrow |x-4| < 2\varepsilon$$

we have that for each $\varepsilon > 0$ if $\delta \leq 2\varepsilon$,

$$|f(x)-3| < \varepsilon$$

for all points $x \neq 4$ for which

$$|x-4| < \delta.$$

We proved that $\lim_{x\to 4} f(x) = 3$.

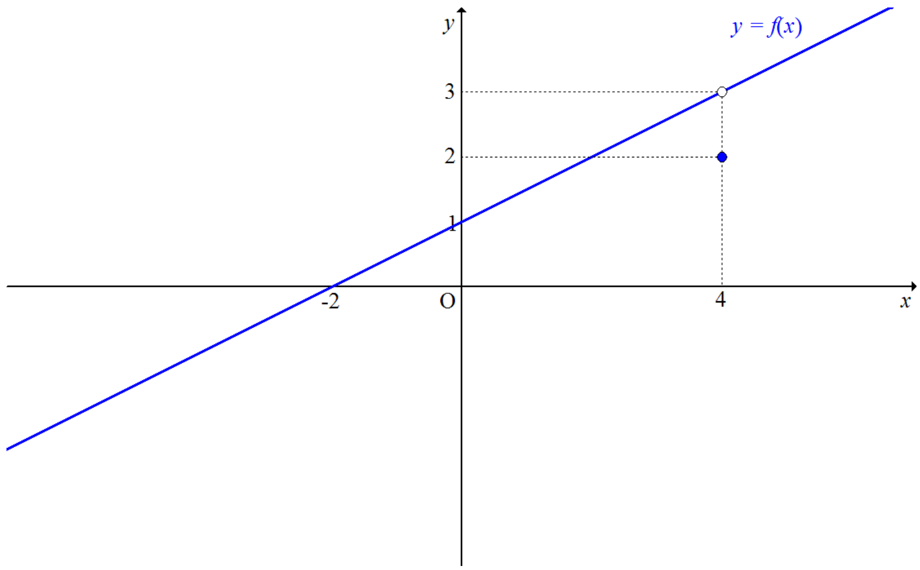


Figure 9.3

Consider now the case of the piecewise function

$$f(x) = \begin{cases} x+1, & x \leq 1 \\ x^2, & x > 1. \end{cases}$$

To study the limit of f when x approaches 1, we have to consider both the analytic expressions $x+1$ and x^2 (depending on the variable x to be, respectively, to the left or to right of point 1). Cases like this one motivate the following definition.

DEFINITION (One-sided limits). Let f be a real function of a real variable with domain D, and L and M real numbers. If a is a limit point of $D \cap (-\infty, a)$, we say that the limit of f as x approaches a *from the left* (or *from below*) is L, and write

$$\lim_{x \to a^-} f(x) = L$$

if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - L| < \varepsilon$$

for all points $x \in D$ for which

$$a - \delta < x < a.$$

If a is a limit point of $D \cap (a, +\infty)$, we say that the limit of f as x approaches a *from the right* (or *from above*) is M, and write

$$\lim_{x \to a^+} f(x) = M$$

if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - L| < \varepsilon$$

for all points $x \in D$ for which

$$a < x < a + \delta.$$

The following theorem establishes the connection between the limit of a function and the one-sided limits.

THEOREM 9.3. *Let f be a real function of real variable with domain D, and L a real number. Assume that a is a limit point of both $D \cap (-\infty, a)$ and $D \cap (a, +\infty)$. Then*

$$\lim_{x \to a} f(x) = L$$

if and only if both the one-sided limits of f as x approaches a exist and

$$\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x) = L.$$

Algebra of limits.

We can, by using the definition and with some computations, show if the limit of a function at a certain point is a certain real number. However, this procedure does not allow the computation of limit: it supposes a previous "hint" about the possible value of the limit. The following theorems, which can be deduced from the definition of the limit of a function, give the instruments needed for the computation of limits.

THEOREM 9.4.

(1) *If f is a real function with domain $\mathbb{R}$ and analytic expression $f(x) = c$, with c a constant, and a is a real number, then*

$$\lim_{x \to a} f(x) = c,$$

also written: $\lim_{x\to a} c = c$.

(2) *If g is a real function with domain $\mathbb{R}$ and analytic expression $g(x) = x$, and a is a real number, then*

$$\lim_{x \to a} g(x) = a,$$

also written: $\lim_{x\to a} x = a$.

THEOREM 9.5. *Let f and g be real functions with the same domain $D \subseteq \mathbb{R}$, a a limit point of D, k a real number, and p a positive integer. If there exist real numbers F and G such that $\lim_{x\to a} f(x) = F$ and $\lim_{x\to a} g(x) = G$ then*

(1) $\lim_{x\to a}(f + g)(x) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = F + G$;

(2) $\lim_{x\to a}(kf)(x) = k \lim_{x\to a} f(x) = kF$;

(3) $\lim_{x\to a}(f \cdot g)(x) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = F \cdot G$;

(4) $\lim_{x\to a}(f/g)(x) = \lim_{x\to a} f(x)/\lim_{x\to a} g(x) = F/G$ *if* $G \neq 0$;

(5) $\lim_{x\to a}(f^p)(x) = (\lim_{x\to a} f(x))^p = F^p$.

If, additionally, a is a limit point of the domain of $\sqrt[p]{f}$ then

(6) $\lim_{x\to a}(\sqrt[p]{f})(x) = \sqrt[p]{\lim_{x\to a} f(x)} = \sqrt[p]{F}$.

We have now the instruments to determine the limits of the constant function $f(x) = c$, the identity function $g(x) = x$, and the functions obtained from these by addition, multiplication by a real number, multiplication, division, exponentiation (with natural exponent), and radiciation. Notice that this class of functions is the class of the algebraic functions, including the polynomial, rational and irrational functions.

Examples. Consider the real functions of a real variable f, g, h, and i, with analytic expressions

$$f(x) = 2x^3 - 2x + 1, \qquad g(x) = \frac{1 - 2x}{2 + x^2}, \qquad h(x) = \left(1 - \sqrt[3]{2x - 1}\right)^2,$$

and

$$i(x) = \begin{cases} x^2 + 1, & x < 2 \\ 3, & x = 2 \\ 1 - x, & x > 2. \end{cases}$$

(a) We want to compute $\lim_{x\to -1} f(x)$. Using (1)-(2) in Theorem 9.4, and (1)-(3) and (5) in Theorem 9.5, we have

$$\lim_{x\to -1} f(x) = \lim_{x\to -1}(2x^3 - 2x + 1) = \lim_{x\to -1} 2 \cdot (\lim_{x\to -1} x)^3 - \lim_{x\to -1} 2 \cdot (\lim_{x\to -1} x) + \lim_{x\to -1} 1$$
$$= 2 \cdot (-1)^3 - 2 \cdot (-1) + 1 = -2 + 2 + 1 = 1.$$

(b) To compute $\lim_{x\to 0} g(x)$, additionally using (4) in Theorem 9.5, we obtain

$$\lim_{x\to 0} g(x) = \lim_{x\to 0} \frac{1-x}{2+x^2} = \frac{1 - 2\cdot 0}{2 + 0^2} = \frac{1}{2}.$$

(c) To determine $\lim_{x\to 1} h(x)$ we use (6) in Theorem 9.5:

$$\lim_{x\to 1} h(x) = \lim_{x\to 1}\left(1 - \sqrt[3]{2x-1}\right)^2 = \left(1 - \sqrt[3]{2\cdot 1 - 1}\right)^2 = \left(1 - \sqrt[3]{1}\right)^2 = 0.$$

(d) We now compute $\lim_{x\to 1} i(x)$. Close to $x = 1$, the function i is given by $i(x) = x^2 + 1$. Then

$$\lim_{x\to 1} i(x) = \lim_{x\to 1}(x^2 + 1) = 1^2 + 1 = 2.$$

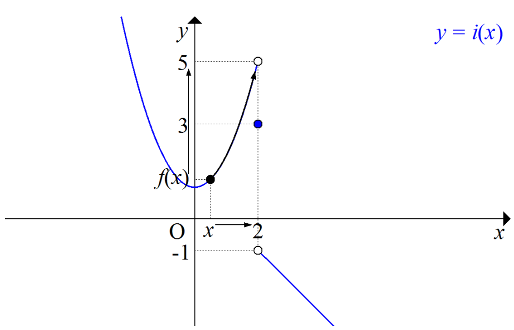
(e) If we want to determine $\lim_{x\to 2} i(x)$, we are in a situation different from the point right above: the function i is defined by distinct expressions to the left and to right of $x = 2$ (see the Figure 9.4). We then have to compute the one-sided limits:

$$\lim_{x\to 2^-} i(x) = \lim_{x\to 2^-}(x^2 + 1) = 2^2 + 1 = 5$$

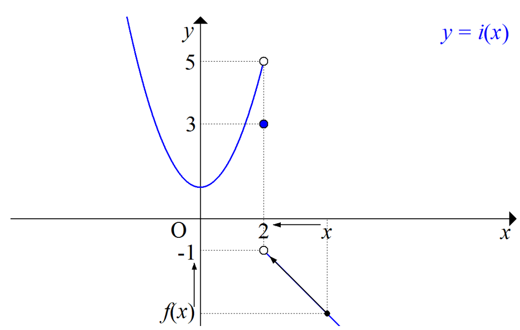
and

$$\lim_{x\to 2^+} i(x) = \lim_{x\to 2^+}(1 - x) = 1 - 2 = -1.$$

Since $\lim_{x\to 2^-} i(x) \neq \lim_{x\to 2^+} i(x)$, we conclude that $\lim_{x\to 2} i(x)$ does not exist.



(a)



(b)

Figure 9.4

Infinite limits and limits at infinity.

We are interested in extending the notion of limit to the cases where a or L (or both) in

$$\lim_{x \to a} f(x) = L$$

take the values $-\infty$ or $+\infty$. This extension arises very naturally.
We begin with the case where $a = +\infty$.

DEFINITION (Limit at infinity). Let f be a real function, with domain $D \subseteq \mathbb{R}$, such that any interval $(l, +\infty)$ contains points in D, and L a real number. We say the the limit of f as x goes to $+\infty$ is L, and write

$$\lim_{x \to +\infty} f(x) = L,$$

if for each $\varepsilon > 0$ there exists an $A \in \mathbb{R}$ such that

$$|f(x) - L| < \varepsilon$$

for all points $x \in D$ for which

$$x > A.$$

The limit of f as x tends to $-\infty$, $\lim_{x \to -\infty} f(x)$, is defined similarly. The case where $L = +\infty$ is considered in the following definition.

DEFINITION (Infinite limit). Let f be a real function with domain $D \subseteq \mathbb{R}$, and a a limit point of D. We say that the limit of f as x approaches a is $+\infty$, and write

$$\lim_{x \to a} f(x) = +\infty,$$

if for each $B \in \mathbb{R}$ there exists a $\delta > 0$ such that

$$f(x) > B$$

for all points $x \in D$ with $x \neq a$ for which

$$|x - a| < \delta.$$

The limit $-\infty$ of f as x approaches a is defined in a similar way. The cases where both a and L are infinite ($-\infty$ or $+\infty$) are defined combining the two previous definitions.

Example. Consider the real function, with domain $D = \mathbb{R} - \{0\}$, defined by $f(x) = 2/x$. We want to show that $\lim_{x\to+\infty} f(x) = 0$. We have that

$$|f(x) - 0| = \left|\frac{2}{x} - 0\right| = \frac{2}{|x|},$$

and since

$$\frac{2}{|x|} < \varepsilon \Leftrightarrow |x| > \frac{2}{\varepsilon} \Leftrightarrow x < -\frac{2}{\varepsilon} \vee x > \frac{2}{\varepsilon}\,,$$

we have that for each $\varepsilon > 0$ if $A \geq \frac{2}{\varepsilon}$

$$|f(x) - 0| < \varepsilon$$

for all points $x \in D$ for which

$$x > A.$$

We showed that $\lim_{x\to+\infty} f(x) = 0$.

The notion of one-sided limit, and the result relating the limit of a function with the one-sided limits are extended naturally to the case of infinite limits. Also, the algebra of limits when the limits are infinite makes use of results extending partially the results presented for finite limits. At the outset, we have that

$$\lim_{x\to+\infty} x = +\infty, \quad \lim_{x\to-\infty} x = -\infty, \quad \lim_{x\to+\infty} c = c, \quad \text{and} \quad \lim_{x\to-\infty} c = c,$$

with c a real constant. Let a be a real number, and p a positive integer. The following rules for operating with limits hold:

Addition

- $a + \infty = +\infty + a = +\infty;$
- $a - \infty = -\infty + a = -\infty;$
- $+\infty + \infty = +\infty$ and $-\infty - \infty = -\infty;$

In the above expressions, $a + \infty$ includes also the case $a - (-\infty)$, $a - \infty$ includes also the case $a + (-\infty)$, $+\infty + \infty$ includes also the case $+\infty - (-\infty)$, and $-\infty - \infty$ includes also the case $-\infty + (-\infty)$.

Multiplication

- $a \cdot (+\infty) = +\infty \cdot a = +\infty$ and $a \cdot (-\infty) = -\infty \cdot a = -\infty$ if $a > 0;$
- $a \cdot (+\infty) = +\infty \cdot a = -\infty$ and $a \cdot (-\infty) = -\infty \cdot a = \infty$ if $a < 0;$
- $+\infty \cdot (+\infty) = -\infty \cdot (-\infty) = +\infty$ and $+\infty \cdot (-\infty) = -\infty \cdot (+\infty) = -\infty;$

Division

- $\dfrac{a}{+\infty} = \dfrac{a}{-\infty} = 0;$

- $\dfrac{+\infty}{a} = +\infty$ and $\dfrac{-\infty}{a} = -\infty$ if $a > 0$;
- $\dfrac{+\infty}{a} = -\infty$ and $\dfrac{-\infty}{a} = +\infty$ if $a < 0$;

Exponentiation

- $(+\infty)^p = +\infty$
- $(-\infty)^p = +\infty$ if p is even;
- $(-\infty)^p = -\infty$ if p is odd;

p-th Roots

- $\sqrt[p]{+\infty} = +\infty$;
- $\sqrt[p]{-\infty} = -\infty$ if p is odd.

The expressions $+\infty - \infty$, $-\infty + \infty$, $0 \cdot (+\infty)$, $0 \cdot (-\infty)$, $+\infty \cdot 0$, $-\infty \cdot 0$, $(+\infty)/(+\infty)$, $(+\infty)/(-\infty)$, $(-\infty)/(+\infty)$, and $(-\infty)/(-\infty)$ are undefined, and are called *indeterminate forms*.

Examples.

(a) Consider the real function f, with domain $\mathbb{R} - \{1\}$, and analytic expression $f(x) = 2 - |x-1|$ (see the Figure 9.5(a)). Let us compute some limits:

$$\lim_{x\to 1} f(x) = \lim_{x\to 1}(2 - |x-1|) = 2 - |1-1| = 2,$$

$$\lim_{x\to -\infty} f(x) = \lim_{x\to -\infty}(2 - |x-1|) = 2 - |-\infty - 1| = 2 - |-\infty| = 2 - \infty = -\infty,$$

$$\lim_{x\to +\infty} f(x) = \lim_{x\to +\infty}(2 - |x-1|) = 2 - |+\infty - 1| = 2 - |+\infty| = 2 - \infty = -\infty.$$

(b) Consider the real function g, with domain $\mathbb{R}$, and defined by

$$g(x) = \begin{cases} x - 2, & x < 1 \\ -(x-1)^2 + 2, & x \geq 1. \end{cases}$$

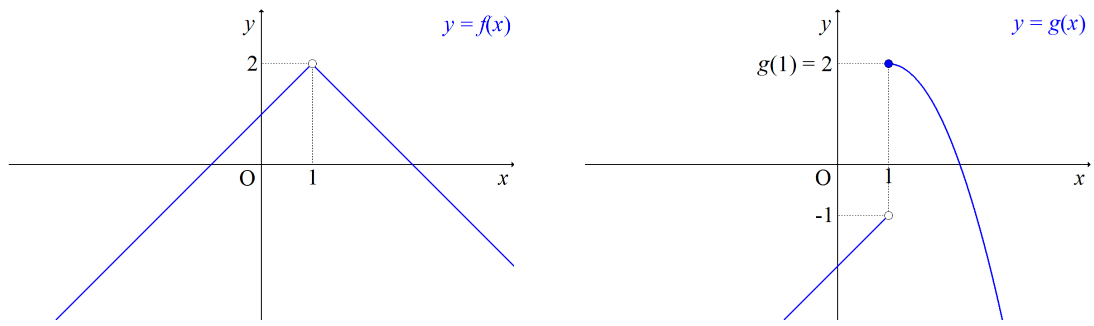
(See the Figure 9.5(b)). To study the limit of g as x approaches 1 we have to study the one-sided limits. Since

$$\lim_{x\to 1^-} g(x) = \lim_{x\to 1^-}(x-2) = 1 - 2 = -1,$$

and

$$\lim_{x\to 1^+} g(x) = \lim_{x\to 1^+}(-(x-1)^2 + 2) = -(1-1)^2 + 2 = 2,$$

we obtain $\lim_{x\to 1^-} g(x) \neq \lim_{x\to 1^+} g(x)$, and conclude that $\lim_{x\to 1} g(x)$ does not exist.



(a) (b)

Figure 9.5

(c) Consider the real function h, with domain $\mathbb{R}$, given by the rational expression

$$h(x) = \frac{2x^2 - 2x + 2}{x^2 + 1}.$$

(See the Figure 9.6). When we compute the limits $\lim_{x\to+\infty} h(x)$ e $\lim_{x\to-\infty} h(x)$ we obtain

$$\lim_{x\to+\infty} h(x) = \lim_{x\to+\infty} \frac{2x^2 - 2x + 2}{x^2 + 1} = \frac{2\cdot(+\infty)^2 - 2\cdot(+\infty) + 2}{(+\infty)^2 + 1} = \frac{2\cdot(+\infty) - 2\cdot(+\infty) + 2}{+\infty + 1} = \frac{+\infty - \infty}{+\infty},$$

and

$$\lim_{x\to-\infty} h(x) = \lim_{x\to-\infty} \frac{2x^2 - 2x + 2}{x^2 + 1} = \frac{2\cdot(-\infty)^2 - 2\cdot(-\infty) + 2}{(-\infty)^2 + 1} = \frac{2\cdot(+\infty) - 2\cdot(-\infty) + 2}{+\infty + 1}$$
$$= \frac{+\infty + \infty}{+\infty} = \frac{+\infty}{+\infty},$$

expressions which involve indeterminate forms. In either of the cases above, the limit can be found by following the procedure:

- To factor in both the numerator and the denominator the terms of higher degree;
- To simplify the fraction;
- To take the simplified fraction to the limit.

Thus,

$$\lim_{x\to+\infty} h(x) = \lim_{x\to+\infty} \frac{2x^2 - 2x + 2}{x^2 + 1} = \lim_{x\to+\infty} \frac{2x^2\left(1 + \frac{-2x}{2x^2} + \frac{2}{2x^2}\right)}{x^2\left(1 + \frac{1}{x^2}\right)}$$
$$= \lim_{x\to+\infty} \frac{2\cdot\left(1 + \frac{-2}{2x} + \frac{2}{2x^2}\right)}{1 + \frac{1}{x^2}} = \frac{2\cdot\left(1 + \frac{-2}{+\infty} + \frac{2}{+\infty}\right)}{1 + \frac{1}{+\infty}} = \frac{2\cdot(1 + 0 + 0)}{1 + 0} = 2,$$

and

$$\lim_{x\to-\infty} h(x) = \lim_{x\to-\infty} \frac{2x^2-2x+2}{x^2+1} = \lim_{x\to-\infty} \frac{2\cdot\left(1+\frac{-2}{2x}+\frac{2}{2x^2}\right)}{1+\frac{1}{x^2}} = \frac{2\cdot\left(1+\frac{-2}{-\infty}+\frac{2}{+\infty}\right)}{1+\frac{1}{+\infty}}$$
$$= \frac{2\cdot(1+0+0)}{1+0} = 2.$$

We then have that the horizontal straight line with equation $y = 2$ is an asymptote to the graph of the function close to both $+\infty$ and $-\infty$.

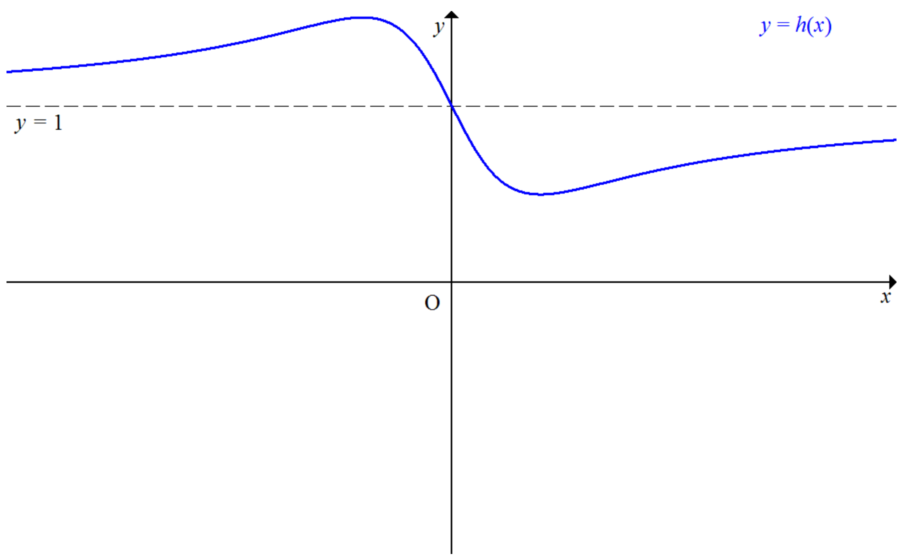


Figure 9.6

We have not yet considered the case where the computation of the limit involves a division by zero. If the result obtained is of the form $0/0$, it is considered undefined ($0/0$ is a new indeterminate form). To a result of the type $a/0$, with $a \in \mathbb{R} - \{0\}$ or $a \in \{-\infty, +\infty\}$ it seems natural to assign an infinite value: when the denominator goes to zero the quotient goes to infinity. But what should be sign of infinity?
We illustrate the matter with an example. Consider the real functions of a real variable f, g, and h, defined, respectively, by $f(x) = 1/x^2$, $g(x) = 1/(-x^2)$, and $h(x) = 1/x$ (see the Figures 9.7(a), 9.7(b), and 9.7(c)). If we compute the limits of these functions as x approaches zero we obtain, in every case, the expression $1/0$. But the behaviour of each one of the functions when x is close to zero is quite distinct.
The function f goes to $+\infty$ when x goes to zero, since the denominator of $1/x^2$ goes to zero assuming only positive values. We express this fact by writing $1/0^+ = +\infty$.
The denominator of $1/(-x^2)$ goes to zero taking only negative values when x goes to zero. Hence, the function g goes to $-\infty$. We write, in this case, $1/0^- = -\infty$.

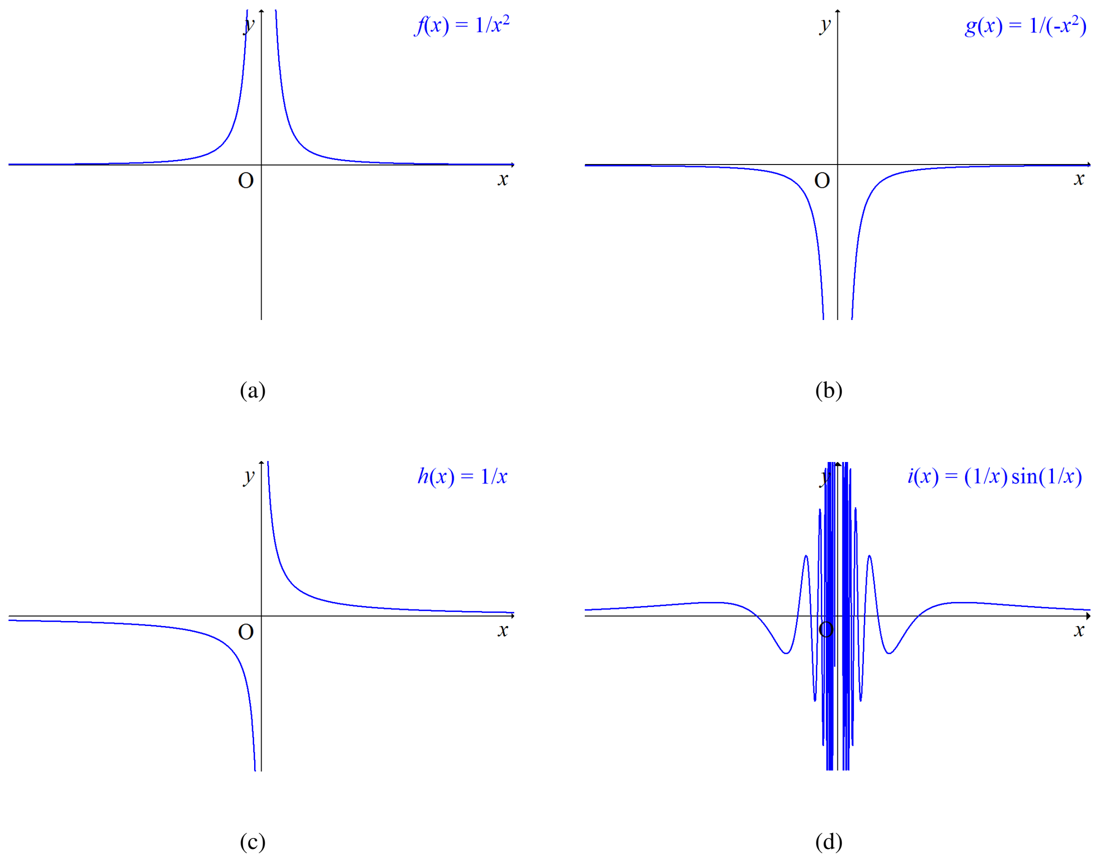


Figure 9.7

In general, we have (a is a real number):

- $\dfrac{a}{0^+} = +\infty$ and $\dfrac{a}{0^-} = -\infty$ if $a > 0$;
- $\dfrac{a}{0^+} = -\infty$ and $\dfrac{a}{0^-} = +\infty$ if $a < 0$;
- $\dfrac{+\infty}{0^+} = +\infty$, $\dfrac{-\infty}{0^+} = -\infty$, $\dfrac{+\infty}{0^-} = -\infty$, and $\dfrac{-\infty}{0^-} = +\infty$.

The behaviour of function h when x is close to zero differs from the one of the functions f and g: h approaches "simultaneously" $+\infty$ and $-\infty$. Observe that the situation where a function is unbounded above and below close to a point does not occur only because of a division by zero, as illustrated in Figure 9.7(d).

Examples.

(a) Consider the real function f, with domain $\mathbb{R}$, and defined by

$$f(x) = \begin{cases} x-1, & x \leq 2 \\ \dfrac{1}{(x-2)^2}, & x > 2. \end{cases}$$

(See Figure 9.8(a)). We want to determine the limit of f as x approaches 2. For this, let us determine the one-sided limits:

$$\lim_{x \to 2^-} f(x) = \lim_{x \to 2^-} (x-1) = 2 - 1 = 1$$

and

$$\lim_{x \to 2^+} f(x) = \lim_{x \to 2^+} \frac{1}{(x-2)^2} = \frac{1}{(2^+-2)^2} = \frac{1}{(0^+)^2} = \frac{1}{0^+} = +\infty.$$

Since these limits are distinct, the limit of f as x approaches 2 does not exist.

(b) Let g be a real function, with domain $\mathbb{R} - \{-1\}$, and analytic expression

$$g(x) = \frac{(2x^2+1)(x+1)}{x^2+2x+1}.$$

(see Figure 9.8(b)). We compute $\lim_{x \to -1} g$, and obtain

$$\lim_{x \to -1} g(x) = \lim_{x \to -1} \frac{(2x^2+1)(x+1)}{x^2+2x+1} = \frac{0}{0},$$

what is an indeterminate form. But, as both the polynomials in the numerator and the denominator of g are null at $x = -1$, they are divisible by $x+1$. Thus,

$$\lim_{x \to -1} g(x) = \lim_{x \to -1} \frac{(2x^2+1)(x+1)}{x^2+2x+1} = \lim_{x \to -1} \frac{(2x^2+1)(x+1)}{(x+1)^2} = \lim_{x \to -1} \frac{2x^2+1}{x+1} = \frac{3}{0}.$$

To conclude whether the limit exists in $\{-\infty, +\infty\}$, we have to determine the one-sided limits

$$\lim_{x \to -1^-} g(x) = \lim_{x \to -1^-} \frac{(2x^2+1)(x+1)}{x^2+2x+1} = \lim_{x \to -1^-} \frac{2x^2+1}{x+1} = \frac{3}{-1^-+1} = \frac{3}{0^-} = -\infty.$$

and

$$\lim_{x \to -1^+} g(x) = \lim_{x \to -1^+} \frac{(2x^2+1)(x+1)}{x^2+2x+1} = \lim_{x \to -1^+} \frac{2x^2+1}{x+1} = \frac{3}{-1^++1} = \frac{3}{0^+} = +\infty.$$

As the one-sided limits do not coincide, g has no limit when x goes to -1. We can, however, conclude that the graphic of the function has a vertical asymptote with equation $x = -1$.

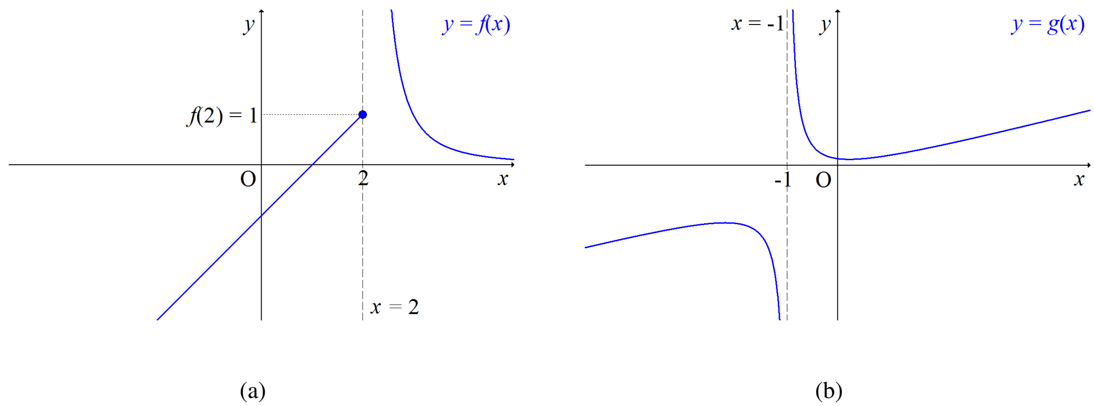


Figure 9.8

Continuous functions.

We begin by defining continuity of a function.

DEFINITION (Continuity). Let f be a real function of a real variable, with domain D, and $a \in D$. f is said to be *continuous* at the point a if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon$$

for all points $x \in D$ for which

$$|x - a| < \delta.$$

If f is not continuous at $a \in D$, we say that it is *discontinuous at* a. If f is continuous at every point of $E \subseteq D$, we say that f is *continuous in* E.

Note that, according to the above definition, a function is always continuous at isolated points of the domain.

THEOREM 9.6. *Let f be a real function of a real variable, with domain D, and $a \in D$. If a is a limit point of D then f is continuous at a if and only if*

$$\lim_{x \to a} f(x) = f(a).$$

The notion of continuity is illustrated in Figure 9.9

Example. Consider the real function of real variable f, defined by

$$f(x) = \begin{cases} x - 1, & 2 < x < 4 \vee x > 4 \\ 2, & x = 4. \end{cases}$$

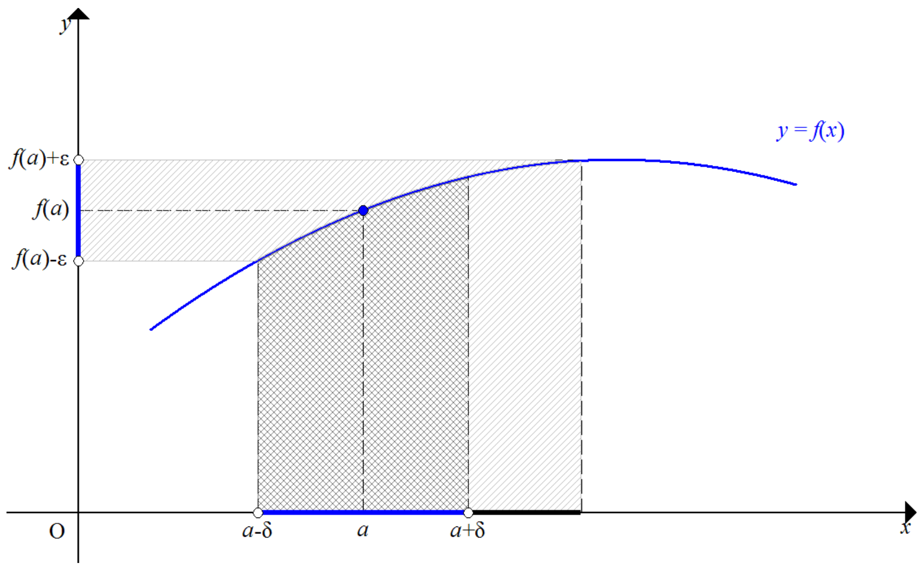


Figure 9.9. Continuity.

The function (with graph sketched in Figure 9.10) has domain $D = (2, +\infty)$. We study the continuity of the function at certain points of D. f is continuous at $x = 3$ since $\lim_{x\to 3} f(x) = f(3) = 2$. At $x = 4$ the function is discontinuous since $\lim_{x\to 4} f(x) \neq f(4)$ ($\lim_{x\to 4} f(x) = 3$ and $f(4) = 2$).

The notion of one-sided continuity is particularly relevant for the case of piecewise functions.

DEFINITION (One-sided continuity). Let f be a real function of a real variable, with domain D, and $a \in D$. We say that f is *left-continuous at* a if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon$$

for all points $x \in D$ for which

$$a - \delta < x \leq a.$$

We say that f is *right-continuous at* a if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon$$

for all points $x \in D$ for which

$$a \leq x < a + \delta.$$

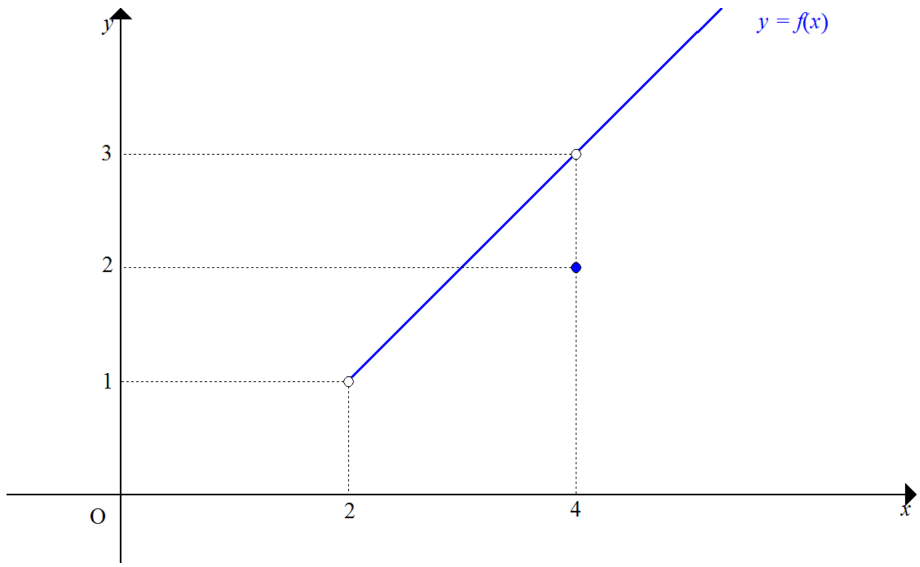


Figure 9.10

THEOREM 9.7. *Let f be a real function of a real variable, with domain D, and $a \in D$. If a is a limit point of $D \cap (-\infty, a]$ then f is left-continuous at a if and only if*

$$\lim_{x \to a^-} f(x) = f(a).$$

If a is a limit point of $D \cap [a, +\infty)$ then f is right-continuous at a if and only if

$$\lim_{x \to a^+} f(x) = f(a).$$

THEOREM 9.8. *A real function of a real variable f is continuous at a point a of its domain if and only if f is left- and right-continuous at a.*

Examples.

(a) Consider the real function f, with domain $D = \mathbb{R}$, defined by

$$f(x) = \begin{cases} x^2 - 4, & x < 2 \\ x, & x \geq 2. \end{cases}$$

The graph of the function is sketched in Figure 9.11(a). To determine the continuity of f at 2, which is a limit point of the domain, we study the one-sided continuity at 2. The function is right-continuous, since

$$\lim_{x\to 2^+} f(x) = \lim_{x\to 2^+} x = 2 = f(2).$$

As

$$\lim_{x\to 2^-} f(x) = \lim_{x\to 2^+} (x^2-4) = 0 \neq f(2),$$

we conclude that the function is not left-continuous at $x = 2$, thus it is discontinuous at the point.

(b) Consider now the real function of a real variable g, with domain $D = \mathbb{R}$, defined by

$$g(x) = \begin{cases} x^2-4, & x < 2 \\ x-2, & x \geq 2. \end{cases}$$

In Figure 9.11(b) we can find the sketch of function's graph. Let us study the continuity of g at the point $x = 2$. The function is right-continuous, since

$$\lim_{x\to 2^+} g(x) = \lim_{x\to 2^+} (x-2) = 0 = g(2),$$

and left-continuous

$$\lim_{x\to 2^-} g(x) = \lim_{x\to 2^+} (x^2-4) = 0 = g(2).$$

Consequently, the function g is continuous at $x = 2$.

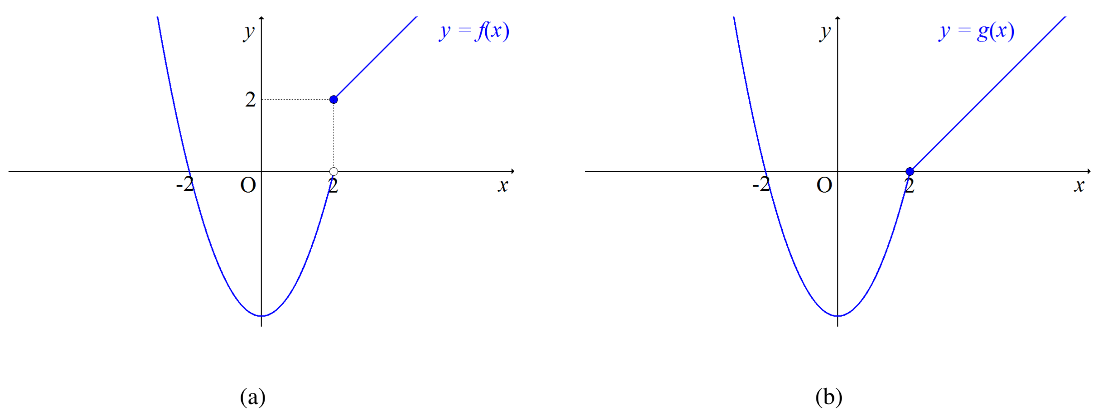


Figure 9.11

From the knowledge that some given functions are continuous at a fixed point, we can deduce the continuity at that point of new functions, obtained by executing elementary operations.

THEOREM 9.9. *Let f and g be real functions of a real variable, with the same domain D, $a \in D$, k a real number, and p a positive integer. If f and g are continuous at a then the following functions are also continuous at a:*

(1) $f + g$;

(2) kf;

(3) $f \cdot g$;

(4) f/g *if* $g(a) \neq 0$;

(5) f^p;

(6) $\sqrt[p]{f}$ *if* $f(a) \geq 0$ *when* p *even.*

Next, we present a result on the continuity of the composite function.

THEOREM 9.10. *Let f and g be real functions of a real variable, with domains D and E, respectively, such that $a \in D$ and $f(a) \in E$. If f is continuous at the point a and g is continuous at the point $f(a)$ then the function $g \circ f$ is continuous at a.*

The previous two results together with the next one give the instruments to determine the continuity of a function in a practical manner.

THEOREM 9.11.

(1) If f is a real function with domain $\mathbb{R}$, and analytic expression $f(x) = c$, with c a constant, then f is continuous in $\mathbb{R}$.

(2) If g is a real function with domain $\mathbb{R}$, and analytic expression $g(x) = x$, then g is continuous in $\mathbb{R}$.

So we have that the constant function $f(x) = c$ and the identity function $g(x) = x$ are continuous in their domains, and so are the functions obtained from these by using elementary operations. Therefore, the polynomial functions, the rational functions, and, more generally, the algebraic functions are continuous functions.

Examples.

(a) Consider the real function f, with domain $\mathbb{R}$, defined by $f(x) = x^3 - 4x$ (see Figure 9.12(a)). It is a polynomial function, thus continuous in its domain.

(b) The real function g, with domain $\mathbb{R} - \{1\}$, and defined by $g(x) = x/(x-1)$ (see Figure 9.12(b)), is continuous in its domain since it is a rational function.

(c) Let h be the real function, with domain $\mathbb{R}$, defined by $h(x) = |x| - 2$ (see Figure 9.13(a)). h is an irrational function, since it can be written $h(x) = \sqrt{x^2} - 2$. Consequently, h is continuous in its domain.

(d) Consider the function i with domain $\mathbb{R}$, defined by

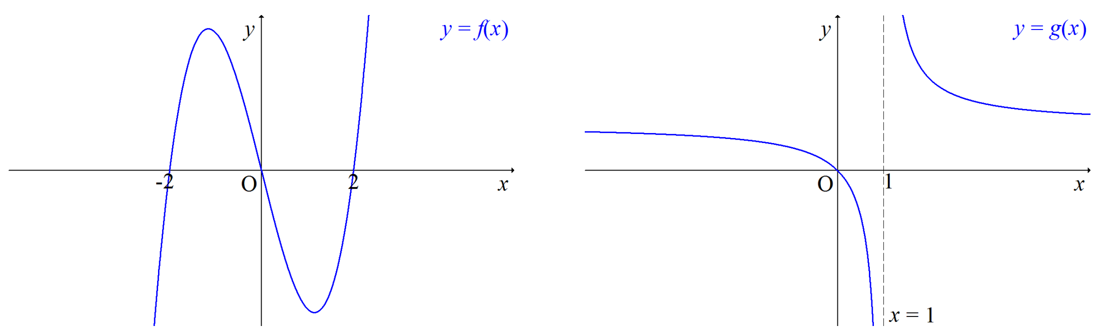
$$i(x) = \begin{cases} x-1, & x \leq 2 \\ \dfrac{1}{(x-2)^2}, & x > 2. \end{cases}$$

(See Figure 9.13(b)). The function is continuous in the intervals $(-\infty, 2)$ and $(2, +\infty)$, since it is defined for the first interval by a polynomial, and for the second by a quotient of polynomials. Let us study the continuity of i at $x = 2$. We have

$$\lim_{x\to 2^+} i(x) = \lim_{x\to 2^+} \frac{1}{(x-2)^2} = \frac{1}{(2^+-2)^2} = \frac{1}{(0^+)^2} = \frac{1}{0^+} = +\infty,$$

so that i does not have a finite limit as x approaches 2 and, consequently, i is discontinuous at the point. However, it is left-continuous at $x = 2$:

$$\lim_{x\to 2^-} i(x) = \lim_{x\to 2^-} (x-1) = 2-1 = 1 = j(2).$$



(a) (b)

Figure 9.12

Together with the algebraic functions, the exponential and the logarithmic functions, and the trigonometric and the inverse trigonometric functions, are also continuous functions. As they are the functions obtained from the above by composition.

THEOREM 9.12. *The following functions are continuous in their domains:*

(1) The exponential and the logarithmic functions, with analytic expressions a^x *e* $\log_a x$*, respectively* ($a > 0$, $a \neq 1$)*;*

(2) The trigonometric functions, with analytic expressions $\sin x$, $\cos x$, $\tan x$*, and* $\cot x$*;*

(3) The inverse trigonometric functions, with analytic expressions $\arcsin x$, $\arccos x$, $\arctan x$*, and* $\operatorname{arccot} x$*.*

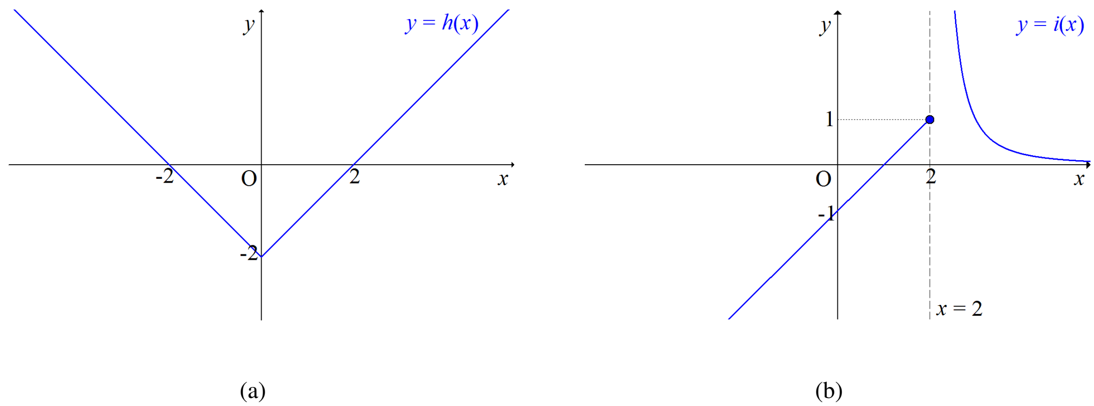


Figure 9.13

Theorems about continuous functions.

We conclude the current topic with the presentation of a few important results on continuous functions.

THEOREM 9.13 (Bolzano's Theorem). *Let f be a real function of a real variable with domain D, $[a, b] \subseteq D$, and $f(a) \cdot f(b) < 0$. If f is continuous in $[a, b]$ then there is at least a number c in the open interval (a, b) such that $f(c) = 0$.*

Next result in an immediate consequence of Theorem 9.13.

THEOREM 9.14 (Intermediate-Value Theorem). *Let f be a real function of a real variable with domain D, $[a, b] \subseteq D$, and C a number such that $f(a) < C < f(b)$ or $f(b) < C < f(a)$. If f is continuous in $[a, b]$ then there is at least a number $c \in (a, b)$ such that $f(c) = C$.*

Example. Consider the real function of a real variable f, with analytic expression

$$f(x) = x^5 - 2x^4 - x^3 + 3x^2 - x - 1$$

The sketch of the graph of f can be found in Figure 9.14. Even for simple functions as this one, the exact computation of the zeros may be difficult. Theorem 9.13 gives us instruments to both conclude about the existence of zeros, and to approximate them if they exist. We exemplify by searching for possible zeros of f in the interval $(-1, 1)$. The function is continuous in its domain $D = \mathbb{R}$. As

$$f(-1) = 1 > 0 \quad \text{and} \quad f(1) = -1 < 0,$$

we have that the function has at least one zero in the interval $(-1,1)$. We can improve the approximation. The value of f at the midpoint of the interval $(-1,1)$ is

$$f(0) = -1 < 0,$$

and, since $f(-1)$ and $f(0)$ have opposite signs, we conclude that there exists a zero in the interval $(-1,0)$. We repeat the procedure and evaluate the function now at the midpoint of the interval $(-1,0)$:

$$f(-0.5) = 0.2188 > 0.$$

As $f(-0.5)$ and $f(0)$ have opposite signs, we know that there is a zero in the interval $(-0.5,0)$. By repeating the procedure, we can approximate the zero with any level of accuracy. For example, after 10 further iterations we obtain the interval

$$(-0.43603515625, -0.435546875).$$

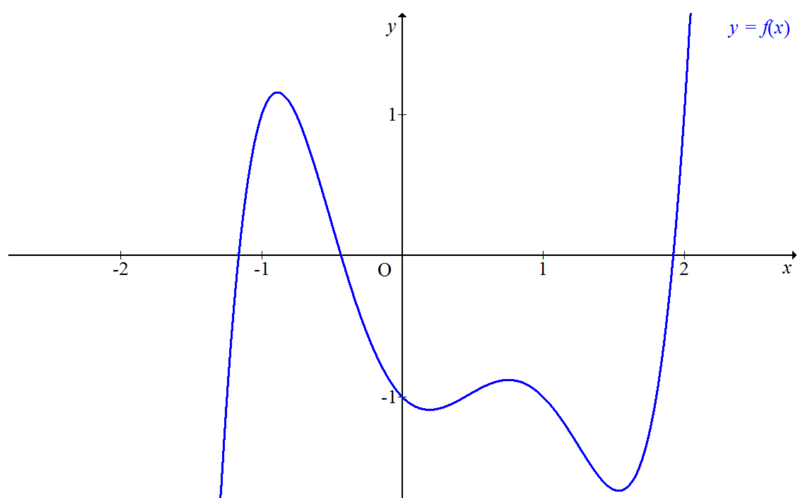


Figure 9.14

THEOREM 9.15 (Bolzano-Weierstrass Theorem). *Let f be a real function of a real variable with domain D, and $[a,b] \subseteq D$. If f is continuous in $[a,b]$ then f attains a maximum and a minimum, each at least once, that is,there exist numbers c and d in $[a,b]$ such that $f(c) \geq f(x) \geq f(d)$ for all $x \in [a,b]$.*

Examples.

(a) The function $f(x) = x^3 + x - 1$ is continuous in the whole domain $\mathbb{R}$. Thus, it attains a maximum and a minimum in any interval $[a, b]$, with a and b real numbers.

(b) The function $g(x) = x$ defined in $(-\infty, +\infty)$ is not bounded;

(c) The function $h(x) = 1/(1 + x)$ defined in $[0, +\infty)$ is bounded but does not attain a minimum;

(d) The function $i(x) = 1/x$ defined in $(0, 1]$ is not bounded from above;

(e) The following function is bounded, but does not attain a maximum (it is not continuous at $x = 0$):

$$j(x) = \begin{cases} 1 - x, & 0 < x \leq 1 \\ 0, & x = 0 \end{cases} .$$

Exercises.

(1) Determine the following limits, if they exist:

a) $\lim_{x\to -1}(x^2 - 3x + 1)$;

b) $\lim_{x\to -\infty}(x^2 - 3x + 1)$;

c) $\lim_{x\to -\infty} \frac{x^2-16}{(x^2-5x+4)(x^2+3)}$;

d) $\lim_{x\to +\infty} \frac{x^2+x-1}{x-3x^2+4}$;

e) $\lim_{x\to -\infty} \frac{4x^3-5x+1}{2x^2-3x+5}$;

f) $\lim_{x\to +\infty} \frac{x(x^2-1)}{x^2(2x+3)}$;

g) $\lim_{x\to -3} \frac{x^2+3x}{x^2+6x+9}$;

h) $\lim_{x\to 2} \frac{x^2-4}{x^3-8}$;

i) $\lim_{x\to 2} \frac{\sqrt{x+2}-\sqrt{3x-2}}{x-2}$;

j) $\lim_{x\to 1} \frac{\sqrt{x-1}}{x-1}$;

k) $\lim_{x\to +\infty} \log_2(2x - 4)$;

l) $\lim_{x\to -1} \frac{x-1}{\ln(x+1)}$;

m) $\lim_{x\to +\infty} \frac{4e^x-4e^{-x}-3}{e^x+5}$;

n) $\lim_{x\to 0^+}(x + \ln x)$;

o) $\lim_{x\to 0} x\cos x$;

p) $\lim_{x\to +\infty} x \arctan x$.

(2) Study the continuity of the following functions at the indicated points:

a) $a(x) = 2 - x^2 + x$, at $x = 0$;

b) $b(x) = \frac{x^2-16}{(x^2-5x+4)(x^2+3)}$, at $x = 1$ and $x = 4$;

c) $c(x) = \log_2(2x - 4)$, at $x = 3$ and $x = 0$.

(3) Study the continuity of the following functions in their domains:

a) $f(x) = x^2 - 3x + 1$;

b) $g(x) = \frac{x^2-16}{(x^2-5x+4)(x^2+3)}$;

c) $h(x) = \frac{x-1}{\ln(x+1)}$;

d) $i(x) = \frac{4e^x-4e^{-x}-3}{e^x+5}$;

e) $j(x) = (x + \ln x)$;

f) $k(x) = \cos x$;

g) $l(x) = x \arctan x$.

(4) Compute, if they exist:

a) $\lim_{x\to 0} \sin \frac{1}{x}$;

b) $\lim_{x\to 0} \left(x \sin \frac{1}{x}\right)$;

c) $\lim_{x\to +\infty} \left(x \sin \frac{1}{x}\right)$;

d) $\lim_{x\to 0} \frac{e^{x^2}-1}{x}$,

(5) Consider the real function of a real variable defined by:

$$h(x) = \begin{cases} 2x + \arccos(x), & 0 \leq x < 1 \\ 2, & x = 1 \\ \dfrac{x+5}{3}, & 1 < x \leq 4 \end{cases}$$

a) Show that h is continuous in its domain.

b) By using Bolzano's theorem, show that: $\exists c \in (2, 4) : h(c) = c$.

(6) Consider in $\mathbb{R} - \{0\}$ the function

$$f(x) = \frac{1 - e^{3x}}{5x}.$$

Let g be a continuous extension of f to $\mathbb{R}$. What is value of g at $x = 0$?

(7) Consider the real function $f(x) = 1 - x\sin\left(\frac{1}{x}\right)$ defined in $\mathbb{R} - \{0\}$. Let g be an extension of f to $\mathbb{R}$. Determine $g(0)$ such that g is continuous at $x = 0$.

(8) Determine the values of a and b for each one of the following function such that they are continuous at the indicated points:

a) $f_1(x) = \begin{cases} 3x - 7, & x \geq 3 \\ ax + 3, & x < 3 \end{cases}$, at $x = 3$;

b) $f_2(x) = \begin{cases} x + a, & x < -2 \\ 3ax + b, & -2 \leq x \leq 1 \\ ax + 3, & x > 1 \end{cases}$, at $x = -2$ and $x = 1$;

c) $f_3(x) = \begin{cases} \sin x, & x \leq 0 \\ ax + b, & x > 0 \end{cases}$, at $x = 0$.

(9) Let f be a continuous mapping of $[a, b]$ into $[a, b]$. Show that there exists $c \in [a, b]$ such that $f(c) = c$.

(10) Show that any polynomial of odd degree is null at at least one point.

10. Differentiation

The notion of derivative. Geometric interpretation.

Let f be a real function function defined in an open interval (a, b). For a fixed point x in the interval, define the difference quotient

$$\frac{f(x+h)-f(x)}{h},$$

where h is a nonzero real number such that $x + h$ lies also in the interval (a, b). The difference quotient measures the *average rate of change* of f when x changes from x to $x + h$. If the limit of the difference quotient when h approaches zero exists, we call it the *derivative of f at the point x*.

DEFINITION (Derivative). Suppose that f is a real function function defined in an open interval (a, b), and x a point in the interval. The *derivative* $f'(x)$ is defined by

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h},$$

provided that the limit exists in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. If $f'(x)$ is finite we say that the function f is *differentiable* at the point x.

Geometrically, the difference quotient

$$\frac{f(x+h)-f(x)}{h}$$

is the slope of the secant to the graph of the function through the points $(x, f(x))$ and $(x + h, f(x + h))$ (see Figure 10.1). As h goes to zero, the point $(x + h, f(x + h))$ moves along the curve in the direction of $(x, f(x))$, and the corresponding secant lines approach a line we call the *tangent to the graph of the function at the point* $(x, f(x))$. Therefore, we can interpret the derivative as the slope of the tangent to the graph at the point $(x, f(x))$, that is, the trigonometric tangent of the angle α that the tangent line makes with the horizontal. This slope is referred to as the *slope of the graph of the function at* $(x, f(x))$. The equation of the tangent line containing the point $(x_0, f(x_0))$ is, if $f'(x_0)$ is finite,

$$y = f'(x_0)(x - x_0) + f(x_0),$$

and, if it is infinite,

$$x = x_0.$$

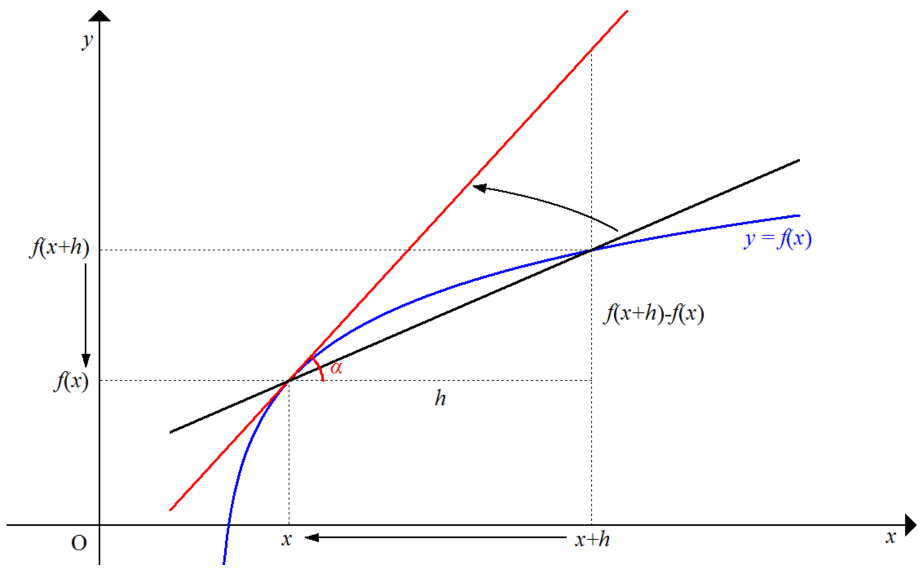


Figure 10.1. Derivative.

The limit process which gives the (finite) derivative $f'(x)$ from $f(x)$ allows us to obtain a new function f' from the given function f. This new function f' is called the (first) *derivative function*, and the process for producing it, *differentiation*. If the function f' is defined in an open interval we can still differentiate it, and obtain the *second derivative* f''; the *third derivative* f''' is obtained from the differentiation of f''. Sometimes, we write $f^{(1)}$ instead of f', $f^{(2)}$ instead of f'', $f^{(3)}$ instead of f'''. Using this notation, we have that, in general, $f^{(n)}$ is obtained from the differentiation of $f^{(n-1)}$ (by convention, $f^{(0)} = f$).
We mention that other notations are frequently used for derivatives, such as:

$$\frac{df}{dx}, \ \frac{d^2 f}{dx^2}, \ \dots \ , \ \frac{d^n f}{dx^n} \qquad \text{or} \qquad Df, \ D^2 f, \ \dots \ , \ D^n f.$$

Examples.

(a) Let us determine the derivative of the real function of a real variable f, defined by $f(x) = x^2$, at $x = 1$. We have

$$f'(1) = \lim_{h\to 0} \frac{f(1+h) - f(1)}{h} = \lim_{h\to 0} \frac{(1+h)^2 - 1^2}{h} = \lim_{h\to 0} \frac{h^2 + 2h}{h} = \lim_{h\to 0}(h+2) = 2.$$

More generally, the derivative function f' is given by

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} \frac{h^2 + 2hx}{h} = \lim_{h\to 0}(h+2x) = 2x,$$

for all $x \in \mathbb{R}$.

(b) To differentiate the real function g, with domain $\mathbb{R} - \{0\}$, defined by $g(x) = 1/x$ we compute the limit

$$g'(x) = \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} = \lim_{h\to 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \lim_{h\to 0} \frac{x - (x+h)}{x(x+h)h} = \lim_{h\to 0} \frac{-1}{x(x+h)} = -\frac{1}{x^2},$$

for all $x \in \mathbb{R} - \{0\}$.

(c) Consider the function $j : \mathbb{R} \to \mathbb{R}$ defined by

$$j(x) = |x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}.$$

The function is not differentiable at $x = 0$ as the limit

$$\lim_{h\to 0} \frac{j(h) - j(0)}{h}$$

does not exist. In fact, the one-sided limits at $x = 0$ are distinct:

$$\lim_{h\to 0^-} \frac{j(h) - j(0)}{h} = \lim_{h\to 0} \frac{-h}{h} = -1 \qquad \text{and} \qquad \lim_{h\to 0^+} \frac{j(h) - j(0)}{h} = \lim_{h\to 0} \frac{h}{h} = 1.$$

These limits are the *one-sided derivatives* at $x = 0$ denoted, respectively, $j'(0^-)$ and $j'(0^+)$.

Now note that $f(x+h)$ can be written

$$f(x+h) = f(x) + h\frac{f(x+h) - f(x)}{h},$$

with $h \neq 0$. If we assume that $f'(x)$ exists and is finite, and let $h \to 0$, we obtain

$$\lim_{x\to 0} f(x+h) = \lim_{x\to 0} \left(f(x) + h\frac{f(x+h) - f(x)}{h} \right) = f(x) + 0 \cdot f'(x) = f(x).$$

This shows that f is continuous at x. We proved the theorem stated next.

THEOREM 10.1. *Let f be a real function defined in an open interval (a, b), and x a point in (a, b). If f is differentiable at x then it is continuous at x.*

The algebra of derivatives. Derivatives of elementary functions.

We can, by using the definition, determine the derivative of a function at a given point or even the derivative function. However, this procedure, involving the computation of a limit, takes some effort and is, sometimes, difficult. The following results, which can be deduced from the definition of derivative, give practical rules for the computation of derivatives.

THEOREM 10.2 (Chain rule). *Let f and g be real functions of a real variable. Suppose that g is differentiable at the point x and that f is differentiable at $y = g(x)$. Then the composition $f \circ g$ is also differentiable at the point x and*

$$(f \circ g)'(x) = f'(y) \cdot g'(x) = f'(g(x)) \cdot g'(x).$$

Example. We want to determine the derivative of the function $f(x) = 1/x^2$. We have that

$$f = u \circ v,$$

with $u(x) = 1/x$ and $v(x) = x^2$. As $u'(x) = -1/x^2$ and $v'(x) = 2x$, we obtain

$$f'(x) = (u \circ v)'(x) = u'(v(x)) \cdot v'(x) = -\frac{1}{(x^2)^2} \cdot 2x = -\frac{2}{x^3}.$$

THEOREM 10.3. *Let f and g be real functions defined in the same interval. If f and g have a (finite) derivative at a point x in the interval, then the same is true for the functions $f + g$, $f - g$, $f \cdot g$, and f/g if $g(x) \neq 0$. The derivatives of these functions are given by*

(1) $(f + g)'(x) = f'(x) + g'(x)$;

(2) $(f - g)'(x) = f'(x) - g'(x)$;

(3) $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$;

(4) $\left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x) - g'(x)f(x)}{(g(x))^2}$.

These and other rules are listed in the following table.

Derivation rules

Notation: u and v designate differentiable functions, a and α real numbers, and n a positive integer.

(1) $a' = 0$	(2) $x' = 1$	(3) $(au)' = au'$
(4) $(u+v)' = u' + v'$	(5) $(uv)' = u'v + uv'$	(6) $(u^\alpha)' = \alpha u^{\alpha-1}u'$ $(\alpha \neq 0)$
(7) $\left(\frac{u}{v}\right)' = \frac{vu' - v'u}{v^2}$	(8) $\left(\frac{1}{u}\right)' = -\frac{u'}{u^2}$	(9) $(e^u)' = e^u u'$
(10) $(a^u)' = a^u \ln a\, u'$ $(a > 0, a \neq 1)$	(11) $(u^v)' = vu^{v-1}u' + u^v \ln u\, v'$	(12) $(\ln u)' = \frac{u'}{u}$
(13) $(\log_a u)' = \frac{u'}{u \ln a}$ $(a > 0, a \neq 1)$	(14) $(\sin u)' = \cos u\, u'$	(15) $(\cos u)' = -\sin u\, u'$
(16) $(\tan u)' = \frac{u'}{\cos^2 u}$	(17) $(\cot u)' = -\frac{u'}{\sin^2 u}$	(18) $(\arcsin u)' = \frac{u'}{\sqrt{1-u^2}}$
(19) $(\arccos u)' = -\frac{u'}{\sqrt{1-u^2}}$	(20) $(\arctan u)' = \frac{u'}{1+u^2}$	(21) $(\operatorname{arccot} u)' = -\frac{u'}{1+u^2}$

Examples. Consider the real functions of a real variable

$$f(x) = \left(\frac{2x-1}{3-x^3}\right)^5; \qquad g(x) = \sqrt[3]{e^{x^2+1} - \log_2 x}; \qquad h(x) = \frac{3}{\ln x}.$$

(a) We want to determine the derivative of f. Using rules $(1) - (4)$, (6), and (7), we obtain

$$\begin{aligned} f'(x) = \left(\left(\frac{2x-1}{3-x^3}\right)^5\right)' &= 5\left(\frac{2x-1}{3-x^3}\right)^4 \cdot \left(\frac{2x-1}{3-x^3}\right)' \\ &= 5\left(\frac{2x-1}{x-x^3}\right)^4 \cdot \frac{\left(3-x^3\right)\cdot 2 - \left(-3x^2\right)(2x-1)}{\left(3-x^3\right)^2} \\ &= 5\left(\frac{2x-1}{3-x^3}\right)^4 \cdot \frac{4x^3-3x^2+6}{\left(3-x^3\right)^2}. \end{aligned}$$

(b) The derivative of g can be determined by using rules $(1)-(4)$, (6), (9), and (13),

$$g'(x) = \left(\sqrt[3]{e^{x^2+1} - \log_2 x}\right)' = \left(\left(e^{x^2+1} - \log_2 x\right)^{\frac{1}{3}}\right)'$$
$$= \frac{1}{3}\left(e^{x^2+1} - \log_2 x\right)^{-\frac{2}{3}} \cdot \left(e^{x^2+1} - \log_2 x\right)'$$
$$= \frac{1}{3}\left(e^{x^2+1} - \log_2 x\right)^{-\frac{2}{3}} \cdot \left(e^{x^2+1} \cdot 2x - \frac{1}{x \ln 2}\right).$$

(c) For the derivative of h, by using rules (3), (8), and (12),

$$h'(x) = \left(\frac{3}{\ln x}\right)' = 3 \cdot \left(-\frac{\frac{1}{x}}{(\ln x)^2}\right) = -\frac{3}{x(\ln x)^2}.$$

Theorems about differentiable functions.

THEOREM 10.4 (Rolle's theorem). *Let f be a real function which is continuous in the closed interval $[a,b]$ and differentiable in the open interval (a,b). Suppose that $f(a) = f(b)$. Then there exists a point $c \in (a,b)$ such that $f'(c) = 0$.*

Figure 10.2 illustrates the above theorem.

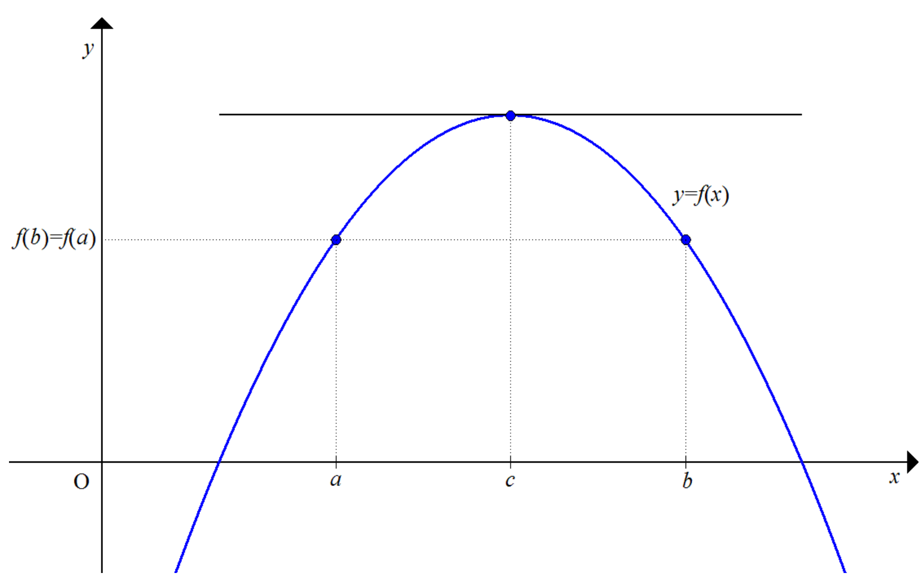


Figure 10.2

Example. Consider the real function of real variable f defined by $f(x) = 4 - x^2$. The function is differentiable in $\mathbb{R}$ and is null for $x = -2$ and $x = 2$. We can then guarantee that the derivative f' has a zero in the interval $(-2, 2)$. In fact, as in this case the derivative function is very simple, $f'(x) = -2x$, we can easily see that f' it is null for $x = 0$.

We next state the Lagrange's mean-value theorem, a generalisation of the Rolle's theorem.

THEOREM 10.5 (Lagrange's theorem). *Let f be a real function which is continuous in the closed interval $[a, b]$ and differentiable in the open interval (a, b). Then there exists a point $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Figure 10.3 illustrates Lagrange's theorem.

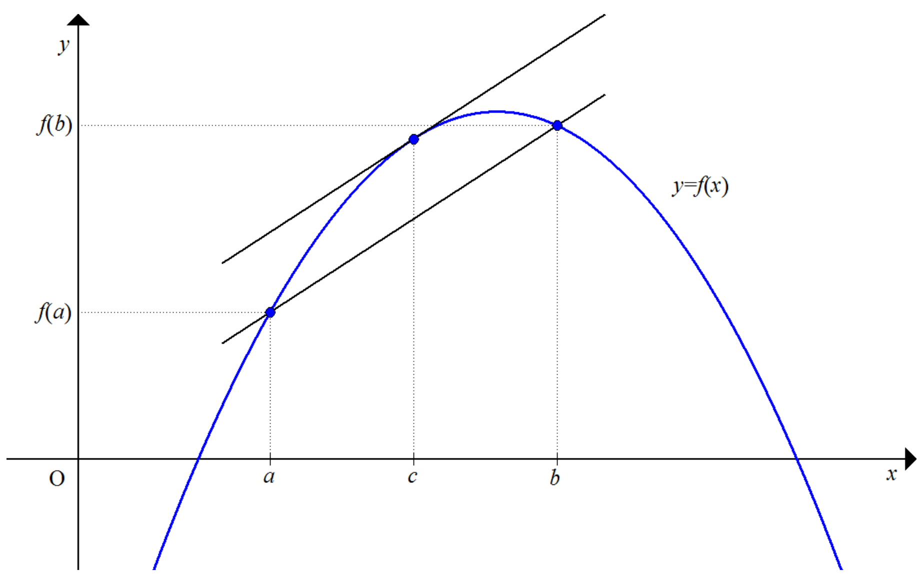


Figure 10.3

Example. Consider the real function of a real variable f, differentiable in $\mathbb{R}$, defined by $f(x) = -x^2 + 4x - 3$. The values of f at $x = 0$ and at $x = 2$ are $f(0) = -3$ and $f(2) = 1$, respectively. Then there exists a point c in the interval $(0, 2)$ such that

$$f'(c) = \frac{f(2) - f(0)}{2 - 0} = \frac{1 + 3}{2} = 2.$$

As for the previous example, we can determine the value of c: from $f'(x) = -2x + 4 = 2$ we obtain $c = 1$.

The following two results are useful for the evaluation of indeterminate forms. The first considers the case where the limit is taken when x approaches an endpoint of an open interval, whereas in the second x approaches an interior point of an open interval.

THEOREM 10.6 (Cauchy's rule). *Let f and g be differentiable real functions defined in the interval (a, b) (with $a, b \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and $a < b$). Suppose that the following conditions are satisfied*

(1) $g'(x) \neq 0$ *for all* $x \in (a, b)$*;*

(2) $\lim_{x\to a} f = \lim_{x\to a} g = 0$ *or* $\lim_{x\to a} f = \lim_{x\to a} g = +\infty$.

Then, if the limit

$$\lim_{x\to a} \frac{f'(x)}{g'(x)}$$

exists in $\overline{\mathbb{R}}$ so exists the limit

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

and

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Note that the cases where the limits of f and g are both $-\infty$ or infinity of opposite signs can be reduced to the one above by a simple change of signs. Note also that the result still holds when the limits are taken as x goes to the upper endpoint b. Finally, the the same result is obtained if we replace condition (2) simply by $\lim_{x\to a} g(x) = +\infty$ or by $\lim_{x\to a} g(x) = -\infty$.

COROLLARY 10.7. *Let I be an open interval, c a point of I, and f and g two differentiable real functions in $I - \{c\}$. Suppose that the following conditions are satisfied*

(1) $g'(x) \neq 0$ *for all* $x \in I - \{c\}$*;*

(2) $\lim_{x\to c} f = \lim_{x\to c} g = 0$ *or* $\lim_{x\to c} f = \lim_{x\to c} g = +\infty$.

Then

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)}$$

if the second limit exists in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.

As for Theorem 10.6, Corollary 10.7 can be easily adapted to the cases where the limits of f and g are both $-\infty$ or infinity of opposite signs.

The Cauchy's rule can still be applied for the evaluation of the indeterminate forms $0 \cdot (+\infty)$, $0 \cdot (-\infty)$, $+\infty - \infty$, and also $1^{+\infty}$, $1^{-\infty}$, 0^0, and $(+\infty)^0$. The transformation of these forms into indeterminate forms of the types considered in Theorem 10.6 and Corollary 10.7 can be obtained by using the equalities:

- $f(x)g(x) = \dfrac{f(x)}{\dfrac{1}{g(x)}}$ or $f(x)g(x) = \dfrac{g(x)}{\dfrac{1}{f(x)}}$ if the limit of $f(x)g(x)$ gives $0 \cdot (+\infty)$ or $0 \cdot (-\infty)$;

- $f(x) + g(x) = \dfrac{\dfrac{1}{f(x)} + \dfrac{1}{g(x)}}{\dfrac{1}{f(x)g(x)}}$ if the limit of $f(x) + g(x)$ gives $+\infty - \infty$;

- $f(x)^{g(x)} = e^{g(x) \ln f(x)}$ if the limit of $f(x)^{g(x)}$ gives $1^{+\infty}$, $1^{-\infty}$, 0^0, or $(+\infty)^0$.

Examples.

(a) When computing the limit

$$\lim_{x \to 0} \frac{e^x - 1}{x}$$

we obtain the indeterminate form $0/0$. Observe also that both the numerator and the denominator of the fraction, $f(x) = e^x - 1$ and $g(x) = x$, respectively, are differentiable in $\mathbb{R}$, with $g \neq 0$ for all $x \neq 0$. As

$$\lim_{x \to 0} \frac{e^x}{1} = e^0 = 1,$$

we conclude that

$$\lim_{x \to 0} \frac{e^x - 1}{x} = 1.$$

(b) The computation of the limit

$$\lim_{x \to +\infty} \frac{\ln x}{x}$$

produces the indeterminate form $+\infty / +\infty$. As

$$\lim_{x \to +\infty} \frac{\frac{1}{x}}{1} = \frac{1}{+\infty} = 0,$$

we have that

$$\lim_{x \to +\infty} \frac{\ln x}{x} = 0.$$

(c) The computation of

$$\lim_{x \to +\infty} x \ln\left(1 + \frac{1}{x}\right)$$

leads to the indeterminate form $(+\infty) \cdot 0$. As

$$x \ln\left(1 + \frac{1}{x}\right) = \frac{\ln\left(1 + \frac{1}{x}\right)}{\frac{1}{x}},$$

and

$$\lim_{x \to +\infty} \frac{\ln\left(1 + \frac{1}{x}\right)}{\frac{1}{x}}$$

gives the indeterminate form $0/0$, we can now apply the Cauchy's rule to this second limit. We obtain

$$\lim_{x\to+\infty} \frac{\frac{-\frac{1}{x^2}}{1+\frac{1}{x}}}{-\frac{1}{x^2}} = \lim_{x\to+\infty} \frac{1}{1+\frac{1}{x}} = \frac{1}{1+\frac{1}{+\infty}} = \frac{1}{1+0} = 1,$$

and, finally,

$$\lim_{x\to+\infty} x \ln\left(1+\frac{1}{x}\right) = 1.$$

(d) We want to determine

$$\lim_{x\to+\infty} \left(1+\frac{1}{x}\right)^x.$$

The computation of the limit produces the indeterminate form $1^{+\infty}$. As

$$\left(1+\frac{1}{x}\right)^x = e^{\ln\left(1+\frac{1}{x}\right)^x} = e^{x\ln\left(1+\frac{1}{x}\right)},$$

and, by item c),

$$\lim_{x\to+\infty} x \ln\left(1+\frac{1}{x}\right) = 1,$$

we have that

$$\lim_{x\to+\infty} \left(1+\frac{1}{x}\right)^x = \lim_{x\to+\infty} e^{\ln\left(1+\frac{1}{x}\right)^x} = e^{\lim_{x\to+\infty} \ln\left(1+\frac{1}{x}\right)^x} = e^1 = e.$$

Polynomial approximations.

Consider a real function of a real variable f, differentiable up to the order n at the point $x = 0$ ($n \geq 1$). We want to a find a polynomial P approximating f in a neighbourhood of $x = 0$. More precisely, we want to a find a polynomial P such that P and f agree at $x = 0$ as well as their first n derivatives, that is,

$$P(0) = f(0), \quad P'(0) = f'(0), \quad P''(0) = f''(0), \quad \ldots, \quad P^{(n-1)}(0) = f^{(n-1)}(0), \quad P^{(n)}(0) = f^{(n)}(0).$$

If P is a polynomial of degree not greater than n,

$$P(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_2x^2 + c_1x + c_0,$$

the derivatives of P up to the order n evaluated at $x = 0$ are

$$P(0) = c_0, \quad P'(0) = c_1, \quad P''(0) = 2c_2, \quad \ldots, \quad P^{(n-1)}(0) = (n-1)!c_{n-1}, \quad P^{(n)}(0) = n!c_n.$$

Then, the matching conditions give

$$c_0 = f(0), \quad c_1 = f'(0), \quad 2c_2 = f''(0), \quad \ldots, \quad (n-1)!c_{n-1} = f^{(n-1)}(0), \quad n!c_n = f^{(n)}(0),$$

and then the polynomial's coefficients are

$$c_0 = f(0), \quad c_1 = f'(0), \quad c_2 = \frac{f''(0)}{2}, \quad \ldots, \quad c_{n-1} = \frac{f^{(n-1)}(0)}{(n-1)!}, \quad c_n = \frac{f^{(n)}(0)}{n!}.$$

Summarising, we have

$$c_k = \frac{f^{(k)}(0)}{k!},$$

with $k = 0, 1, \ldots, n$.

THEOREM 10.8. *Let f be a real function, differentiable up to the order n at $x = 0$ ($n \geq 1$). Then there exists a unique polynomial P of degree not greater than n satisfying the conditions*

$$P(0) = f(0), \quad P'(0) = f'(0), \quad P''(0) = f''(0), \ \ldots, \ P^{(n-1)}(0) = f^{(n-1)}(0), \quad P^{(n)}(0) = f^{(n)}(0).$$

This polynomial is called the Taylor polynomial of degree n generated by f at the point 0, and is given by

$$P(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k.$$

More generally, the Taylor polynomial of degree n generated by f at the point a is

$$P(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

Notice that if $n = 1$, $P(x)$ is just the equation of the tangent to the graph of f at $x = a$.

Example. Consider the exponential function $f(x) = e^x$. As $f^{(n)}(0) = e^0 = 1$ for any nonnegative integer n, the Taylor polynomial of degree n generated by f at $x = 0$ is

$$P(x) = \sum_{k=0}^{n} \frac{1}{k!} x^k.$$

For example, for $n = 1, 2, 3$ we obtain, respectively,

$$P_1(x) = 1 + x, \quad P_2(x) = 1 + x + \frac{1}{2}x^2, \quad P_3(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3.$$

The approximation of f by these Taylor polynomials is illustrated in Figure 10.4. Notice that the approximation is more accurate if the degree of the polynomial is higher.

The following theorem, which is a generalisation of the Lagrange's mean-value theorem, gives an estimate for the approximation error.

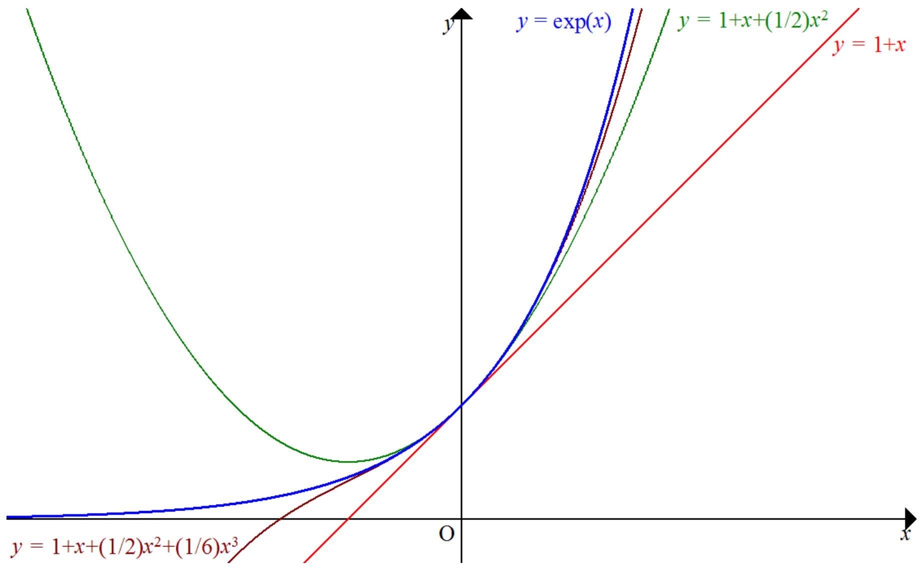


Figure 10.4

THEOREM 10.9 (Taylor's theorem). *Suppose that f is a real function in $[\alpha,\beta]$, n a nonnegative integer, $f^{(n)}$ is continuous in $[\alpha,\beta]$, and $f^{(n+1)}(t)$ exists and is finite for every $t\in(\alpha,\beta)$. Let a and x be distinct points of $[\alpha,\beta]$. Then there exists a point c between x and a, such that*

$$f(x)=\sum_{k=0}^{n}\frac{f^{(k)}(a)}{k!}(x-a)^k+R_n(x),\qquad \text{(Taylor's formula)}$$

where the remainder $R_n(x)$ is given by

$$R_n(x)=\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.\qquad \text{(Lagrange form of the remainder)}$$

Example. Consider the logarithmic function $f(x)=\ln x$. The n-th order derivative is

$$f^{(n)}(x)=(-1)^{n-1}\frac{(n-1)!}{x^n}$$

for any positive integer n, and, in particular,

$$f^{(n)}(1)=(-1)^{n-1}(n-1)!.$$

The Taylor formula of order n of f about $x = 1$, with the remainder in the Lagrange form, is given by

$$\begin{aligned} f(x) &= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k}(x-1)^k + \frac{(-1)^n}{(n+1)c^{n+1}}(x-1)^{n+1} \\ &= (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 + \cdots + \frac{(-1)^{n-1}}{n}(x-1)^n + \frac{(-1)^n}{(n+1)c^{n+1}}(x-1)^{n+1}, \end{aligned}$$

where c is a number between x and 1. If, for example, $n = 3$, we obtain

$$f(x) = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4c^4}(x-1)^4.$$

The approximation of f by the Taylor polynomial of degree 3 generated by f at $x = 1$ is represented in Figure 10.5.

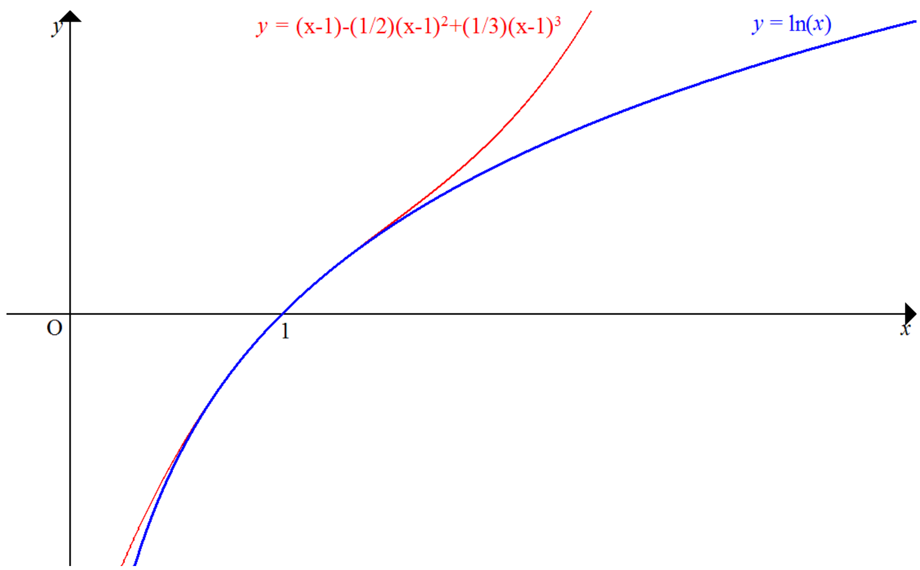


Figure 10.5

Exercises.

(1) Determine the derivatives of the following real functions of a real variable:

a) $f(x) = x^2$;

b) $f(x) = 2x + 2$;

c) $f(x) = \frac{1}{2}x^2$;

d) $f(x) = 2x^2 + 4x + 4$;

e) $f(x) = c \quad (c \in \mathbb{R})$;

f) $f(x) = \frac{1}{x^2} + 3x^{\frac{1}{3}}$;

g) $f(x) = \frac{3}{x^4} - \sqrt[4]{x} + x$;

h) $f(x) = 6x^{1/3} - x^{0.4} + \frac{9}{x^2}$;

i) $f(x) = \frac{1}{\sqrt[3]{x}} + \sqrt{x}$;

j) $f(x) = \left(x^4 + 4x + 2\right)(2x + 3)$;

k) $f(x) = (2x - 1)\left(3x^2 + 2\right)$;

l) $f(x) = \left(x^3 - 12x\right)\left(3x^2 + 2x\right)$;

m) $f(x) = \left(2x^5 - x\right)(3x + 1)$;

n) $f(x) = \frac{2x+1}{x+5}$;

o) $f(x) = \frac{3x^4+2x+2}{3x^2+1}$;

p) $f(x) = \frac{x^{\frac{3}{2}}+1}{x+2}$;

q) $f(x) = \frac{x^2+x}{2x-1}$;

r) $f(x) = (x + 5)^2$;

s) $f(x) = \left(x^3 - 2x + 5\right)^2$;

t) $f(x) = \sqrt{1 - x^2}$;

u) $f(x) = \frac{(2x+4)^3}{4x^3+1}$;

v) $f(x) = (2x + 1)\sqrt{2x + 2}$;

w) $f(x) = \frac{2x+1}{\sqrt{2x+2}}$;

x) $f(x) = \sqrt{2x^2 + 1}\left(3x^4 + 2x\right)^2$;

y) $f(x) = \frac{2x+3}{(x^4+4x+2)^2}$;

z) $f(x) = \sqrt{x^3 + 1}\left(x^2 - 1\right)$;

a′) $f(x) = \left((2x + 3)^4 + 4(2x + 3) + 2\right)^2$;

b′) $f(x) = \sqrt{1 + x^2}$;

c′) $f(x) = \left(3x^2 + e\right)e^{2x}$;

d′) $f(x) = e^{2x^2+3x}$;

e′) $f(x) = e^{e^{2x^2}+1}$;

f′) $f(x) = 2^{x-3}\sqrt{x^3 - 2} + \ln x$;

g′) $f(x) = \ln x - 2e^x + \sqrt{x}$;

h′) $f(x) = \ln\left(\ln\left(x^3(x + 1)\right)\right)$;

i′) $f(x) = \ln\left(2x^2 + 3x\right)$;

j′) $f(x) = \ln^4 x + \ln x^4 + 4\ln x$;

k′) $f(x) = \ln(\sin x)$;

l′) $f(x) = \ln\frac{1+x}{1-x}$;

m′) $f(x) = \ln\sqrt{x^2 + 1}$;

n′) $f(x) = x^2 \log_4 x$;

o′) $f(x) = 3e^x - 4\cos x - \frac{1}{4}\ln x$;

p′) $f(x) = \arctan x + \tan x$;

q′) $f(x) = \arcsin\frac{x}{2}$;

r′) $f(x) = \arccos\left(2x^2\right)$;

s′) $f(x) = \sin(2x) - \sin^2 x + 2\cot x$;

t′) $f(x) = \frac{\sin x+\cos x}{\sin x-\cos x}$;

u′) $f(x) = e^x(\sin x + \cos x)$;

v′) $f(x) = e^{ax}\sin(ax) \quad (a \in \mathbb{R})$.

(2) Determine:

a) $\frac{d}{dx}e^{x^2}$;

b) $\frac{d}{dx}e^{2^x}$;

c) $\frac{d}{dx}\arctan\left(x^4\right)$;

d) $\frac{d}{dx}\arctan(2x + 4)$;

e) $\frac{d}{dx}\ln\left(x^4\right)$;

f) $\frac{d}{dx}\ln(2x + 4)$;

g) $\frac{d}{dx}\frac{1}{1+x^4}$;

h) $\frac{d}{dx}\frac{1}{2x+4}$;

i) $\frac{d}{dx}\arctan e^x$.

(3) Consider the real functions of a real variable and the points in $\mathbb{R}^2$:

$$f(x) = \frac{x^3}{3} + x^2 + 5, \qquad (3, 23)$$
$$g(x) = x^3 - 3x + 1, \qquad (1, -1)$$
$$h(x) = (x^2 + 1)(2 - x), \qquad (2, 0)$$

a) For each one of the above functions, determine the values of x for which the tangent to its graph is horizontal.

b) For each one of the functions f, g, and h, write the equation of the tangent to its graph at the given point.

(4) Check whether the Rolle's theorem is applicable to:

a) $f(x) = x^2 - 3x + 2$ in the interval $[1, 2]$;

b) $g(x) = |x - 1|$ in the interval $[0, 2]$.

(5) Let f and g be real functions of a real variable, differentiable in $[a, b]$, such that $f(a) = g(a)$ and $f(b) = g(b)$. Show that there exists a point $c \in (a, b)$ such that $f'(c) = g'(c)$.

(6) Prove that $e^x \geq 1 + x$ by using the Lagrange's mean value theorem.

(7) Compute the following limits:

a) $\displaystyle\lim_{x\to 0} x^x$; b) $\displaystyle\lim_{x\to 0} \frac{e^x - e^{-x} - 2x}{x - \sin x}$; c) $\displaystyle\lim_{x\to +\infty} \frac{\ln x}{\sqrt[3]{x}}$;

d) $\displaystyle\lim_{x\to 0} \left(\frac{1}{x} - \frac{1}{\sin x}\right)$; e) $\displaystyle\lim_{x\to +\infty} x \sin\frac{1}{x}$; f) $\displaystyle\lim_{x\to 3} \left(\frac{1}{x-3} - \frac{5}{x^2 - x - 6}\right)$.

(8) Write the Taylor's formula with Lagrange's remainder of order n for each one of the following functions:

a) $f(x) = e^x$ about $x = 0$;

b) $g(x) = \ln x$ about $x = 1$;

c) $h(x) = \frac{1}{1-x}$ about $x = 0$;

d) $i(x) = \ln(1 + x)$ about $x = 0$.

e) $j(x) = \frac{1}{x^2}$ about $x = -1$.

11. Optimisation

Extreme values of functions.

DEFINITION (Maximum of a function). Let f be a real function defined in a set $S \subseteq \mathbb{R}$. f is said to have an *absolute maximum* in S if there is a point $c \in S$ such that

$$f(x) \leq f(c)$$

for all $x \in S$. f is said to have a *relative maximum* at a point $c' \in S$ if there is an open interval (a, b) containing c' such that

$$f(x) \leq f(c')$$

for all $x \in (a, b) \cap S$.

Absolute and *relative minima* of f can be defined, respectively, as the negatives of absolute and relative maxima of $-f$. Maxima and minima are called *extrema*.

THEOREM 11.1. *Let f be a real function defined in an open interval (a, b), and suppose that f has a relative maximum or a relative minimum at a point $c \in (a, b)$. If the derivative $f'(c)$ exists and is finite, then $f'(c) = 0$.*

Examples.

(a) Consider the real function of a real variable f defined by $f(x) = 1 - x^2$ (see Figure 11.1). The function has a relative (and absolute) maximum for $x = 0$. Since f is differentiable at 0, the derivative f' is null for $x = 0$. In fact, $f'(x) = -2x$, and $f'(0) = 0$.

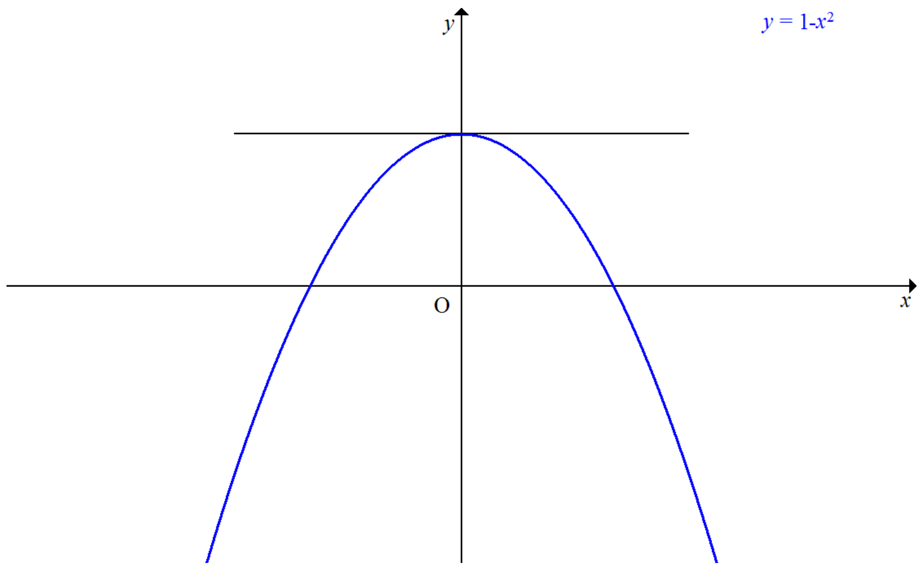
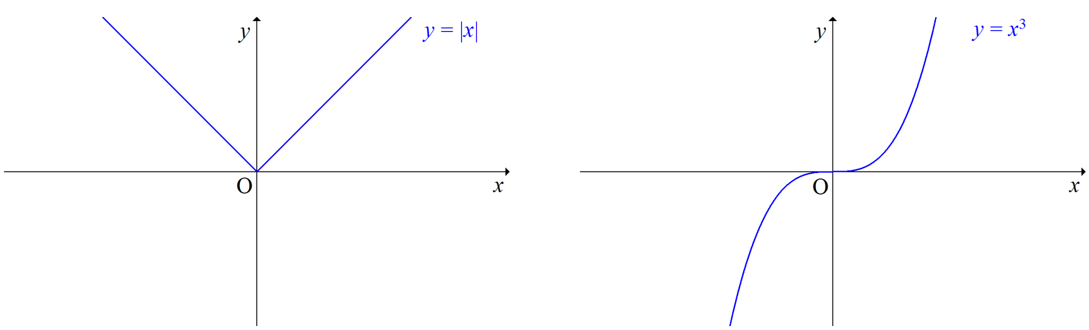


Figure 11.1

(b) Let g be the real function of a real variable given by $g(x) = |x|$ (see Figure 11.2(a)). The derivative of g is never null but g has a relative (and absolute) minimum at 0. Note that the derivative of g for $x = 0$ does not exist.

(c) Consider the real function of a real variable h defined $h(x) = x^3$ (see Figure 11.2(b)). The derivative of h, $h'(x) = 3x^2$, is null for $x = 0$. But h has no extremum at 0.



(a) (b)

Figure 11.2

Geometric properties of functions.

The following theorem is a consequence of the Lagrange's mean-value theorem.

THEOREM 11.2. *Let f be a real function which is continuous in a closed interval $[a, b]$ and differentiable at every point of the open interval (a, b). Then*

(1) If $f'(x) > 0$ for every $x \in (a, b)$, f is strictly increasing in $[a, b]$;

(2) If $f'(x) < 0$ for every $x \in (a, b)$, f is strictly decreasing in $[a, b]$;

(3) If $f'(x) = 0$ for every $x \in (a, b)$, f is constant in $[a, b]$.

THEOREM 11.3. *Let f be a real function which is continuous in a closed interval $[a, b]$, and differentiable at every point of the open interval (a, b), except possibly at a point c in the interval.*

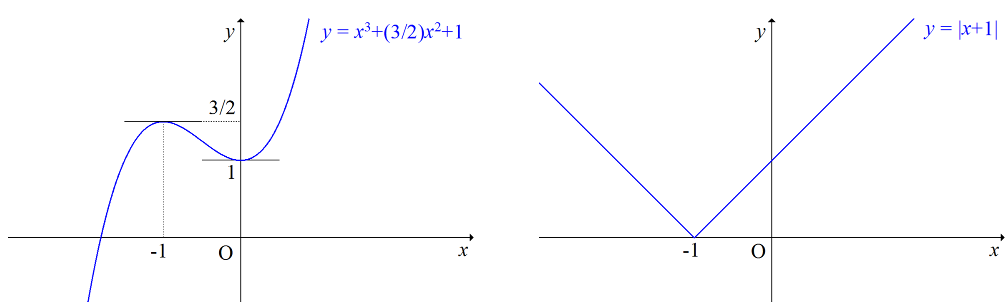
(1) If $f'(x) > 0$ for all $x < c$ and $f'(x) < 0$ for all $x > c$, then f has a relative maximum at c;

(2) If $f'(x) < 0$ for all $x < c$ and $f'(x) > 0$ for all $x > c$, then f has a relative minimum at c.

Examples.

(a) Consider the real function of a real variable f defined by $f(x) = x^3 + \frac{3}{2}x^2 + 1$ (see Figure 11.3(a)). The derivative $f'(x) = 3x^2 + 3x$ is positive in $(-\infty, -1) \cup (0, +\infty)$, negative in $(-1, 0)$, and null for $x = -1$ or $x = 0$. We than have that f is strictly increasing in $(-\infty, -1)$ and in $(0, +\infty)$, strictly decreasing in $(-1, 0)$, and attains a relative maximum at -1 and a relative minimum at 0.

(b) Consider the real function of a real variable g defined by $g(x) = |x + 1|$ (see Figure 11.3(b)). The function is continuous in $\mathbb{R}$ and although is not differentiable for $x = -1$, the derivative changes sign at this point. Therefore g has an extremum (a minimum) for $x = -1$.



(a) (b)

Figure 11.3

Second derivative test for extrema.

Let f be a real function which is continuous in a closed interval $[a, b]$. It is clear that f attains an absolute maximum and an absolute minimum in the interval. Moreover, if f is differentiable at every point of (a, b), relative and absolute extrema can only occur at:

- The endpoints of the interval;
- The points of (a, b) for which the derivative is null.

These latter points for which the derivative is null are called *critical points*.

THEOREM 11.4. *Let f be a real function which is continuous in a closed interval $[a, b]$, and differentiable at every point of the open interval (a, b). Assume that $c \in (a, b)$ is a critical point of f. Assume also that f'' exists and is finite in (a, b). Then*

(1) If $f''(x) < 0$ for all $x \in (a, b)$, f has a relative maximum at c;

(2) If $f''(x) > 0$ for all $x \in (a, b)$, f has a relative minimum at c.

In the above theorem, if f'' is continuous in a neighbourhood of the critical point c it suffices to have $f''(c) < 0$ ($f''(c) > 0$) to conclude that f has a relative maximum (minimum) at c.
Recall that a function is *convex* if the chord joining any two points of its graph lies above the graph, and *concave* if it lies below the graph.

THEOREM 11.5. *Let f be a real function which is continuous in a closed interval $[a, b]$, and differentiable at every point of the open interval (a, b). If f' is increasing in (a, b) then f is convex in $[a, b]$. In particular, f is convex if f'' exists, is finite and nonnegative in (a, b).*

The discussion of the concave case can be reduced to the convex case by noting that a function f is concave if $-f$ is convex.

Examples.

(a) Consider the real function of a real variable f defined by $f(x) = \frac{1}{6}x^3$ (see Figure 11.4(a)). Its second derivative is the function $f(x) = x$. Since $f'' < 0$ for $x < 0$ and $f'' > 0$ for $x > 0$, f is concave in $(-\infty, 0]$ and convex in $[0, +\infty)$. The point 0, at which the function f is continuous and the second derivative changes sign is called a *point of inflection*.

(b) Consider the real function of a real variable g defined by $g(x) = 1 - (x - 1)^2$ (see Figure 11.4(b)). The function has only one critical point: the point 1. As $g''(1) < 0$, g attains a maximum for $x = 1$.

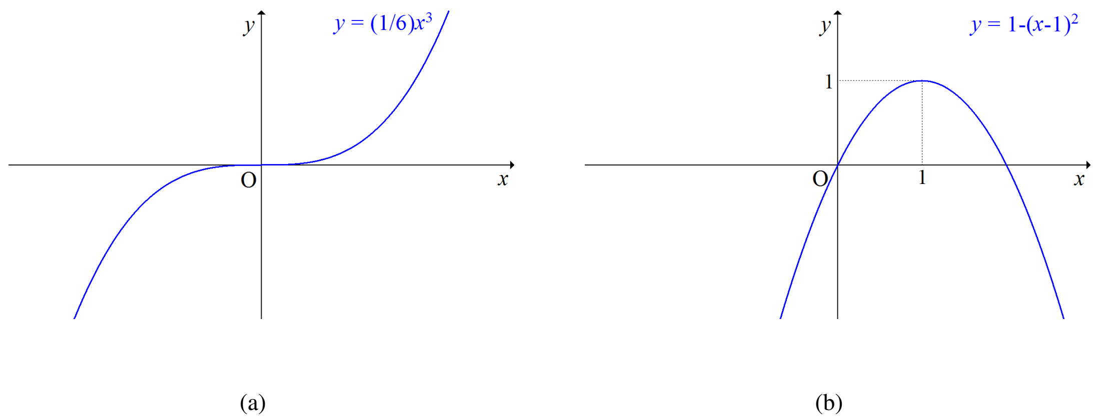


Figure 11.4

Sketching the graph of a function.

In order to sketch the graph of a function, we have to study the following aspects:

- Domain, parity, and intercepts;
- Continuity and asymptotes;
- Monotonicity and local extrema;
- Convexity and points of inflection;
- Range.

Example. Consider the real function of a real variable f defined by $f(x) = x - \frac{1}{x}$. The function has domain $\mathbb{R} - \{0\}$, and is odd as $f(-x) = -f(x)$ for all $x \neq 0$. Also, the function f is continuous in the whole domain since it is the difference and the quotient of continuous functions. It does not intercept the y-axis and, to determine the zeros, we solve the equation

$$f(x) = 0 \Leftrightarrow x - \frac{1}{x} = 0 \Leftrightarrow \frac{x^2 - 1}{x} = 0 \Leftrightarrow x^2 - 1 = 0 \wedge x \neq 0 \Leftrightarrow (x = -1 \vee x = 1) \wedge x \neq 0$$
$$\Leftrightarrow x = -1 \vee x = 1.$$

To look for an asymptote near $+\infty$, we compute the limit

$$\lim_{x \to +\infty} \frac{f(x)}{x} = \lim_{x \to +\infty} \frac{x - \frac{1}{x}}{x} = \lim_{x \to +\infty} \left(1 - \frac{1}{x^2}\right) = 1.$$

We found the slope $m = 1$ of the possible asymptote. Now we compute

$$\lim_{x \to +\infty} (f(x) - mx) = \lim_{x \to +\infty} \left(x - \frac{1}{x} - x\right) = \lim_{x \to +\infty} \left(-\frac{1}{x}\right) = 0.$$

We found the intercept of the asymptote, so that the asymptote's equation is $y = x$. Near $-\infty$, and repeating the same computations we obtain the same asymptote $y = x$. We now check the existence of a vertical asymptote at $x = 0$. We determine

$$\lim_{x\to 0^-} f(x) = \lim_{x\to 0^-}\left(x - \frac{1}{x}\right) = \left(0 - \frac{1}{0^-}\right) = +\infty.$$

By performing the same computations we obtain

$$\lim_{x\to 0^+} f(x) = -\infty.$$

The derivative function

$$f'(x) = \left(x - \frac{1}{x}\right)' = 1 + \frac{1}{x^2} = \frac{x^2+1}{x^2}$$

is positive for all $x \neq 0$. Thus, f is strictly increasing in $(-\infty, 0)$ and in $(0, +\infty)$, and has no extreme values. The second derivative

$$f''(x) = \left(1 + \frac{1}{x^2}\right)' = -\frac{2}{x^3}$$

is positive if $x < 0$ and negative if $x > 0$. Therefore, f is convex for $x < 0$ and concave for $x > 0$. The function has no points of inflection. The graph of f is sketched in Figure 11.5. From the graph, it is clear that the range of the function is $\mathbb{R}$.

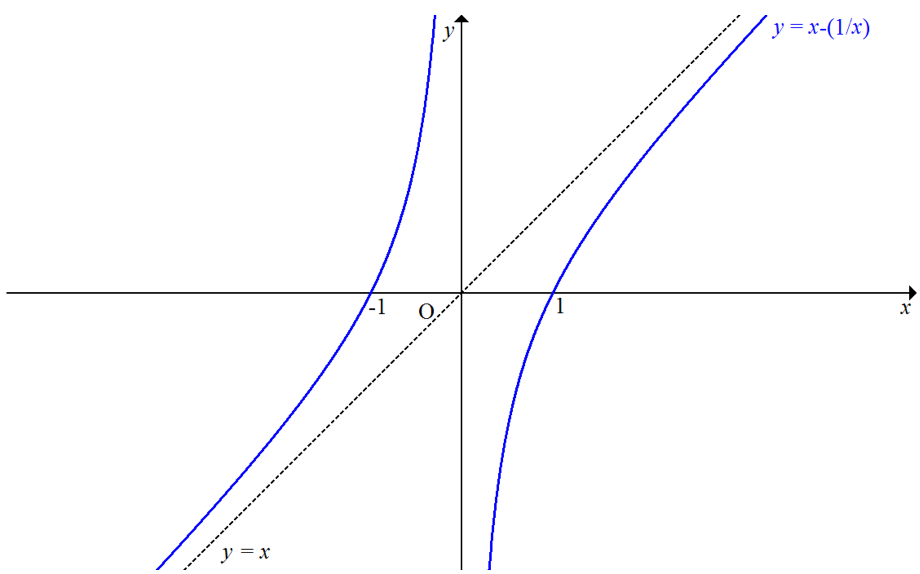


Figure 11.5

Exercises.

(1) Find the local extrema of the following functions:

a) $f(x) = x^4 - 2x^2 + 3$;

b) $g(x) = -4x^3 + 36x^2 + 3000$;

c) $h(x) = 4x^3 - 4x$;

d) $i(x) = \frac{x^2}{x^2+1}$;

e) $j(x) = (2x-1)^5$;

f) $k(x) = \frac{x}{x+2}$;

g) $l(x) = \sin^2 x,\ \text{for}\ 0 < x < 2\pi$;

h) $m(x) = \frac{1}{2}x - \sin x,\ \text{for}\ 0 < x < 2\pi$.

(2) Study each one of the following functions (domain, continuity, parity, intercepts, asymptotes, monotonicity and local extrema, convexity and points of inflection, and range), and sketch the corresponding graphs:

a) $a(x) = x^4 + x^2 - 2$;

b) $b(x) = x^4 - 5x^2 + 4$;

c) $c(x) = \frac{x-1}{x+1}$;

d) $d(x) = \sqrt{x^2 - 1}$;

e) $e(x) = e^{\frac{1}{\ln x}}$;

f) $f(x) = \frac{x}{\ln x}$;

g) $g(x) = e^{-\frac{1}{x}}$;

h) $h(x) = e^{-x^2}$;

i) $i(x) = x^2 \ln x$.

12. Antidifferentiation

Definitions and basic results.

The present topic can be motivated by the problem: "Which are the functions whose derivative is $f(x) = 2x$?".

DEFINITION (Antidifferentiable function). Let f be a real function defined in an interval I. f is said to be *antidifferentiable* if there exists a real differentiable function g in I such that

$$g' = f.$$

If this is the case, we say that g is an *antiderivative* (or an *indefinite integral* or a *primitive*) of f in I. The process of solving for antiderivatives is called *antidifferentiation* (or *indefinite integration* or *primitivation*).

For the above definition to be meaningful when the interval I is not open, we have to make clear what is the meaning of g to be differentiable at an endpoint of the interval where it is closed. If we consider, for example, the interval $I = [a, b)$, we define the derivative of g at a as $g'(a) = g'(a^+)$. The same is assumed if the interval is closed at the right endpoint.

Note also that g is obviously continuous in I.

It should be clear that if g is an antiderivative of f then any function $g + c$, with c a constant, is also an antiderivative of f, since

$$(g + c)' = g' + c' = f.$$

Moreover, all antiderivatives of f are written as $g + c$. To see this, assume that h is also an antiderivative of f. Then

$$(h - g)' = h' - g' = f - f = 0,$$

which allows us to conclude that $h - g$ is a constant function, that is,

$$h - g = c \Leftrightarrow h = g + c.$$

THEOREM 12.1. *Let f be a real function defined in an interval I, $x_0 \in I$, and $y_0 \in \mathbb{R}$. If f is antidifferentiable in I there exists a unique antiderivative g such that*

$$g(x_0) = y_0.$$

The problem for which the above theorem guaranties the existence of a unique solution can be written

$$y' = f(x) \quad \text{in} \quad I, \qquad y(x_0) = y_0,$$

and is called the *Cauchy problem* or the *initial value problem* for the *differential equation* $y' = f(x)$.
We now fix some notation. If f is an antidifferentiable function, we denote the family of antiderivatives of f by

$$\int f, \qquad \int f(x)\,dx \qquad \text{or} \qquad \int f\,dx.$$

If g is an antiderivative of f we then write

$$\int f = g + c, \qquad \int f(x)\,dx = g(x) + c \qquad \text{or} \qquad \int f\,dx = g + c,$$

with c a constant.

Examples.

(a) Let us determine the antiderivatives of the real function f in $\mathbb{R}$ defined by $f(x) = 2x$ (see Figure 12.1(a)). The function $F(x) = x^2$ is an antiderivative of $f(x)$ since

$$F'(x) = \left(x^2\right)' = 2x.$$

Thus,

$$\int f(x)\,dx = x^2 + c,$$

with c a real constant.

(b) Consider the problem of determining the antiderivative $F(x)$ of $f(x) = 2x$ satisfying the condition $F(2) = 3$ (see Figure 12.1(b)). We have

$$\begin{cases} F(x) &= x^2 + c \\ F(2) &= 3 \end{cases} \Leftrightarrow \begin{cases} F(x) &= x^2 + c \\ 2^2 + c &= 3 \end{cases} \Leftrightarrow \begin{cases} F(x) &= x^2 + c \\ c &= -1 \end{cases} \Rightarrow F(x) = x^2 - 1.$$

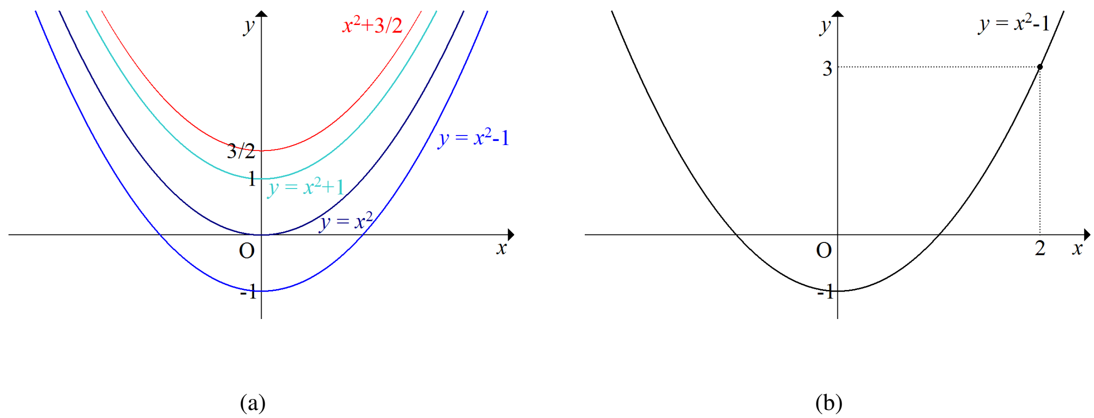


Figure 12.1

The following theorem gives some properties of algebra of antiderivatives.

THEOREM 12.2. *Let u and v be antidifferentiable real functions in an interval I, and a a real number. Then*

(1) The function au is antidifferentiable in I, and

$$\int au = a\int u;$$

(2) The function $u+v$ is antidifferentiable in I, and

$$\int u+v = \int u + \int v.$$

By simple inversion of the derivative rules, the following antidifferentiation formulas can be obtained.

(1) $\int 0 = c$	(2) $\int 1 = x + c$	(3) $\int a = ax + c$
(4) $\int u^\alpha u' = \dfrac{u^{\alpha+1}}{\alpha+1} + c \ (\alpha \neq -1)$	(5) $\int \dfrac{u'}{u} = \ln\|u\| + c$	(6) $\int e^u u' = e^u + c$
(7) $\int a^u u' = \dfrac{a^u}{\ln a} + c \ (a > 0, a \neq 1)$	(8) $\int \sin u\, u' = -\cos u + c$	(9) $\int \cos u\, u' = \sin u + c$
(10) $\int \dfrac{u'}{\cos^2 u} = \tan u + c$	(11) $\int \dfrac{u'}{\sin^2 u} = -\cot u + c$	(12) $\int \dfrac{u'}{\sqrt{1-u^2}} = \arcsin u + c$
(13) $\int \dfrac{u'}{\sqrt{a^2-u^2}} = \arcsin \dfrac{u}{a} + c \ (a > 0)$	(14) $\int \dfrac{u'}{1+u^2} = \arctan u + c$	(15) $\int \dfrac{u'}{a^2+u^2} = \dfrac{1}{a}\arctan \dfrac{u}{a} + c \ \ (a > 0)$

Notation: u designates a differentiable function, a, α, and c real numbers.

Since $\arcsin x = \pi/2 - \arccos x$ and $\arctan x = \pi/2 - \operatorname{arccot} x$ for all real x, the formulas (12), (13), (14), and (15) can also be written

$(12')$ $\int \frac{u'}{\sqrt{1-u^2}} = -\arccos u + c,$

$(13')$ $\int \frac{u'}{\sqrt{a^2-u^2}} = -\arccos \frac{u}{a} + c,$

$(14')$ $\int \frac{u'}{1+u^2} = -\operatorname{arccot} u + c,$

$(15')$ $\int \frac{u'}{a^2+u^2} = -\frac{1}{a} \operatorname{arccot} \frac{u}{a} + c.$

Examples.

(a) $\int 3\,dx = 3x + c$. (Using formula (2) and Theorem 12.2(1)).

(b) $\int x^3\,dx = \frac{x^4}{4} + c$. (Using formula (4)).

(c) Using formula (4) and Theorem 12.2(1),

$$\int (3x+1)^4\,dx = \int (3x+1)^4 \cdot 3 \cdot \frac{1}{3}\,dx = \frac{1}{3}\int (3x+1)^4 \cdot 3\,dx = \frac{1}{3} \cdot \frac{(3x+1)^5}{5} + c = \frac{(3x+1)^5}{15} + c.$$

(d) Using formula (4) and Theorem 12.2(1),

$$\int \sqrt{3x+1}\,dx = \frac{1}{3}\int (3x+1)^{\frac{1}{2}} \cdot 3\,dx = \frac{1}{3} \cdot \frac{(3x+1)^{\frac{3}{2}}}{\frac{3}{2}} + c = \frac{2}{9}(3x+1)^{\frac{3}{2}} + c.$$

(e) $\int \sin x \cos x\,dx = \frac{\sin^2 x}{2} + c$. (Using formula (4)).

(f) $\int \frac{x}{1+x^2}\,dx = \frac{1}{2}\int \frac{2x}{1+x^2}\,dx = \frac{1}{2}\ln|1+x^2| + c = \ln\sqrt{1+x^2} + c$. (Using formula (5) and Theorem 12.2(1)).

(g) $\int \tan x\,dx = -\int \frac{-\sin x}{\cos x}\,dx = -\ln|\cos x| + c$. (Using formula (5) and Theorem 12.2(1)).

(h) $\int e^{x^2} x\,dx = \frac{1}{2}\int e^{x^2} 2x\,dx = \frac{1}{2}e^{x^2} + c$. (Using formula (6) and Theorem 12.2(1)).

(i) $\int 2^x\,dx = \frac{2^x}{\ln 2} + c$. (Using formula (7)).

(j) $\displaystyle\int \sin(3x)\,dx = \frac{1}{3}\int \sin(3x)\cdot 3\,dx = \frac{1}{3}(-\cos(3x)) + c = -\frac{1}{3}\cos(3x) + c$. (Using formula (8) and Theorem 12.2(1)).

(k) $\displaystyle\int \frac{e^x}{\cos^2(e^x)}\,dx = \tan e^x + c$. (Using formula (10)).

(l) $\displaystyle\int \frac{x}{\sqrt{1-x^4}}\,dx = \frac{1}{2}\int \frac{2x}{\sqrt{1-(x^2)^2}}\,dx = \frac{1}{2}\arcsin x^2 + c$. (Using formula (12) and Theorem 12.2(1)).

(m) $\displaystyle\int \frac{x}{1+x^4}\,dx = \frac{1}{2}\int \frac{2x}{1+(x^2)^2}\,dx = \frac{1}{2}\arctan x^2 + c$. (Using formula (14) and Theorem 12.2(1)).

(n) $\displaystyle\int \frac{1}{5+x^2}\,dx = \int \frac{1}{(\sqrt{5})^2+x^2}\,dx = \frac{1}{\sqrt{5}}\arctan\frac{x}{\sqrt{5}} + c$. (Using formula (25)).

Methods of antidifferentiation.

In general, the antiderivative of a given function cannot be found by using the basic antiderivative formulas presented above. This is the case, for example, of the function $f(x) = xe^x$.
We introduce the so-called *methods of antidifferentiation*. We begin with the *method of antidifferentiation by parts*, which is derived from the rule for differentiating of the product of two functions

$$(uv)' = u'v + uv' \Leftrightarrow u'v = (uv)' - uv'.$$

Now, if we antidifferentiate both members of the second equation above,

$$\int u'v = uv - \int uv'.$$

Note that $\int u'v$ exists if and only if $\int uv'$ does. Note also that the method is useful if the second integral is easier to compute than the first.

THEOREM 12.3 (Antidifferentiation by parts). *If u and v are differentiable real functions in an interval I, the product $u'v$ is antidifferentiable if and only if the product uv' is. If this is the case, we have*

$$\int u'v = uv - \int uv'.$$

Examples.
(a) $\displaystyle\int (e^x x)\ dx = e^x x - \int (e^x\cdot 1)\ dx = e^x x - e^x + c = e^x(x-1) + c.$

(b) $\displaystyle\int \ln x\,dx = \int (1\cdot\ln x)\,dx = x\ln x - \int \left(x\cdot\frac{1}{x}\right) fx = x\ln x - \int 1\,dx = x\ln x - x + c = x(\ln x - 1) + c.$

(c) To determine $\int \sin^2 x\, dx$ we antidifferentiate by parts,

$$\begin{aligned}\int \sin^2 x\, dx = \int \sin x \cdot \sin x\, dx &= -\cos x \cdot \sin x - \int -\cos x \cdot \cos x\, dx \\ &= -\cos x \cdot \sin x + \int \cos^2 x\, dx \\ &= -\cos x \cdot \sin x + \int 1 - \sin^2 x\, dx \\ &= -\cos x \cdot \sin x + x - \int \sin^2 x\, dx.\end{aligned}$$

We then have

$$\begin{aligned}\int \sin^2 x\, dx = -\cos x \cdot \sin x + x - \int \sin^2 x\, dx + c &\Leftrightarrow 2\int \sin^2 x\, dx = -\cos x \cdot \sin x + x + c \\ &\Leftrightarrow \int \sin^2 x\, dx = \frac{-\cos x \cdot \sin x + x}{2} + c.\end{aligned}$$

As Theorem 12.3 states the existence of the first antiderivative in $\int u'v = uv - \int uv'$ only in the case the second exists, we have, for the present problem, to check that the functions $F(x) = \frac{-\cos x \cdot \sin x + x}{2} + c$ are, in fact, the antiderivatives of $f(x) = \sin^2 x$:

$$\left(\frac{-\cos x \cdot \sin x + x}{2} + c\right)' = \frac{\sin x \cdot \sin x - \cos x \cdot \cos x + 1}{2} = \frac{2\sin^2 x}{2} = \sin^2 x.$$

To motivate the method of *antidifferentiation by substitution* or *by change of variable* consider the following. Let I and J be real intervals, and assume that $f : I \to \mathbb{R}$ is an antidifferentiable function, and $\varphi : J \to I$ a differentiable bijection. If g is an antiderivative of f in I, then

$$\psi(t) = (g \circ \varphi)(t) = g(\varphi(t))$$

is differentiable in J and

$$\psi'(t) = g'(\varphi(t)) \cdot \varphi'(t) = f(\varphi(t)) \cdot \varphi'(t).$$

Thus, the family of the antiderivatives of ψ' can be written

$$\int f(\varphi(t)) \cdot \varphi'(t)\, dt$$

and the aimed antiderivatives of f can be obtained from ψ by simply undoing the change of variable

$$\int f(x)\, dx = \left[\int f(\varphi(t)) \cdot \varphi'(t)\, dt\right]_{t=\varphi^{-1}(x)}.$$

THEOREM 12.4 (Antidifferentiation by substitution). *Let I and J be intervals in $\mathbb{R}$, $f : I \to \mathbb{R}$ an antidifferentiable function, and $\varphi : J \to I$ a differentiable bijection. Then $(f \circ \varphi)\varphi'$ is antidifferentiable and, denoting an antiderivative by ψ, $\psi \circ \varphi^{-1}$ is an antiderivative of f.*

The above theorem gives the previously mentioned formula for antidifferentiation by substitution

$$\int f(x)\,dx = \left[\int f(\varphi(t))\varphi'(t)\,dt\right]_{t=\varphi^{-1}(x)}.$$

Examples.

(a) We want to antidifferentiate the real function $f(x) = \frac{e^{3x}}{e^{2x}+1}$ in $\mathbb{R}$. Let us consider the real differentiable function $\varphi(t) = \ln t$, which maps one-to-one $(0,+\infty)$ onto $\mathbb{R}$. Changing variable $x = \varphi(t) = \ln t$, and noting that

$$x' = \varphi'(t) = \frac{1}{t} \qquad \text{and} \qquad t = \varphi^{-1}(x) = e^x,$$

we have

$$\int f(x)\,dx = \left[\int f(\varphi(t))\varphi'(t)\,dt\right]_{t=\varphi^{-1}(x)} = \left[\int f(\ln t)\cdot(\ln t)'\,dt\right]_{t=e^x} = \left[\int \frac{t^3}{t^2+1}\cdot\frac{1}{t}\,dt\right]_{t=e^x}$$

$$= \left[\int \frac{t^2}{t^2+1}\,dt\right]_{t=e^x} = \left[\int \frac{t^2+1-1}{t^2+1}\,dt\right]_{t=e^x} = \left[\int \left(1-\frac{1}{t^2+1}\,dt\right)\right]_{t=e^x}$$

$$= [t - \arctan t + c]_{t=e^x} = e^x - \arctan e^x + c.$$

(b) The antiderivatives of the real function $f(x) = \frac{\ln x}{x(1+\ln^2 x)}$ in $(0,+\infty)$ can be determined by considering the change of variable $x = \varphi(t) = e^t$, a one-to-one differentiable mapping from $\mathbb{R}$ onto $(0,+\infty)$. We have

$$x' = \varphi'(t) = e^t \qquad \text{and} \qquad t = \varphi^{-1}(x) = \ln x,$$

and

$$\int f(x)\,dx = \left[\int f(\varphi(t))\varphi'(t)\,dt\right]_{t=\varphi^{-1}(x)} = \left[\int f\left(e^t\right)\cdot\left(e^t\right)'\,dt\right]_{t=\ln x} = \left[\int \frac{t}{e^t(1+t^2)}\cdot e^t\,dt\right]_{t=\ln x}$$

$$= \left[\int \frac{t}{1+t^2}\,dt\right]_{t=\ln x} = \left[\frac{1}{2}\int \frac{2t}{1+t^2}\,dt\right]_{t=\ln x} = \left[\frac{1}{2}\ln\left|1+t^2\right| + c\right]_{t=\ln x}$$

$$= \frac{1}{2}\ln\left|1+\ln^2 x\right| + c = \ln\sqrt{1+\ln^2 x} + c.$$

(c) To determine the antiderivatives of the real function f in $(0,+\infty)$, defined by $f(x) = \frac{1}{\sqrt{x}(1+x)}$, we consider the change of variable $x = \varphi(t) = t^2$, a one-to-one differentiable mapping from $(0,+\infty)$ onto $(0,+\infty)$. We have

$$x' = \varphi'(t) = 2t \qquad \text{and} \qquad t = \varphi^{-1}(x) = \sqrt{x}.$$

Thus,

$$\int f(x)\,dx = \left[\int f(\varphi(t))\varphi'(t)\,dt\right]_{t=\varphi^{-1}(x)} = \left[\int f\left(t^2\right)\cdot\left(t^2\right)'\,dt\right]_{t=\sqrt{x}} = \left[\int \frac{1}{t(1+t^2)}\cdot 2t\,dt\right]_{t=\sqrt{x}}$$

$$= \left[2\int \frac{1}{1+t^2}\,dt\right]_{t=\sqrt{x}} = [2\arctan t + c]_{t=\sqrt{x}} = 2\arctan\sqrt{x} + c.$$

(d) We want to find the antiderivatives of the real function f in $(0,+\infty)$, defined by $f(x) = \frac{1}{\sqrt{x}+\sqrt[3]{x}}$. Consider the differentiable bijection $\varphi : (0,+\infty) \to (0,+\infty)$ defined by $\varphi(t) = t^6$, and the change of variable $x = \varphi(t)$. We have

$$x' = \varphi'(t) = 6t^5 \qquad \text{and} \qquad t = \varphi^{-1}(x) = \sqrt[6]{x}.$$

Thus,

$$\int f(x)\,dx = \left[\int f(\varphi(t))\varphi'(t)\,dt\right]_{t=\varphi^{-1}(x)} = \left[\int f\left(t^6\right)\cdot\left(t^6\right)'\,dt\right]_{t=\sqrt[6]{x}} = \left[\int \frac{1}{t^3+t^2}\cdot 6t^5\,dt\right]_{t=\sqrt[6]{x}}$$

$$= \left[6\int \frac{t^3}{t+1}\,dt\right]_{t=\sqrt[6]{x}} = \left[6\int t^2 - t + 1 - \frac{1}{t+1}\,dt\right]_{t=\sqrt[6]{x}}$$

$$= \left[6\left(\frac{t^3}{3} - \frac{t^2}{2} + t - \ln|t+1|\right) + c\right]_{t=\sqrt[6]{x}} = 2\sqrt{x} - 3\sqrt[3]{x} + 6\sqrt[6]{x} - \ln\left(\sqrt[6]{x}+1\right)^6 + c.$$

Partial fractions.

In the examples we gave above for the substitution method, the antidifferentiation problem was reduced the antidifferentiation of a rational function, that is, a function whose analytical expression is a quotient of polynomials. In fact, this is a frequent result obtained with the use of the substitution method. This is a strong reason to study the antidifferentiation of rational functions. Consider the following:

- A rational expression can always be written as a sum

$$Q(x) + \frac{N(x)}{D(x)},$$

  where Q, N, and D are polynomials, with the degree of N less than the degree of D (the fraction $\frac{N(x)}{D(x)}$ is called a *proper rational fraction*);
- The coefficient of the term of the highest degree in D can be assumed to be 1, with no loss of generality (this can be easily obtained if we factor the coefficient).

It can be shown that a proper rational fraction $\frac{N(x)}{D(x)}$, where the coefficient of the term of the highest degree in D is 1, can be written as a sum of *partial fractions* of the types

- $\dfrac{A}{(x-a)^r}$, with $r = 1, 2, \ldots, m$,
- $\dfrac{Bx + C}{((x-p)^2 + q^2)^s}$, with $s = 1, 2, \ldots, n$,

where, A, B, C, a, p, and q are real constants, and m and n positive integers. The types of partial fraction in the sum depend on the factorisation of the polynomial D. The main point is that partial fractions are easily antidifferentiated.

Examples.

(a) We want to antidifferentiate the real function of a real variable f defined by

$$f(x) = \frac{4x^2 + x + 1}{x^3 - x}.$$

As the denominator of the rational fraction can be decomposed

$$x^3 - x = x(x+1)(x-1),$$

the fraction can be written as s sum

$$\frac{4x^2 + x + 1}{x^3 - x} = \frac{A}{x} + \frac{B}{x+1} + \frac{C}{x-1}.$$

The constants A, B, and C can be determined by solving the equation

$$4x^2 + x + 1 = A(x^2 - 1) + Bx(x-1) + Cx(x+1)$$

$$\Leftrightarrow 4x^2 + x + 1 = (A + B + C)x^2 + (-B + C)x - A \Leftrightarrow \begin{cases} A + B + C &= 4 \\ -B + C &= 1 \\ -A &= 1 \end{cases} \Leftrightarrow \begin{cases} A &= -1 \\ B &= 2 \\ C &= 3 \end{cases}.$$

We then have

$$\frac{4x^2 + x + 1}{x^3 - x} = -\frac{1}{x} + \frac{2}{x+1} + \frac{3}{x-1},$$

so that

$$\int \frac{4x^2 + x + 1}{x^3 - x}\,dx = \int -\frac{1}{x} + \frac{2}{x+1} + \frac{3}{x-1}\,dx = -\ln|x| + 2\ln|x+1| + 3\ln|x-1| + c$$

$$= \ln\left|\frac{(x+1)^2(x-1)^3}{x}\right| + c.$$

(b) To find the antiderivatives of the real function of a real variable g defined by

$$g(x) = \frac{2x^3 + 5x^2 + 6x + 2}{x(x+1)^3}$$

we write the rational fraction a the sum

$$\frac{2x^3+5x^2+6x+2}{x(x+1)^3} = \frac{A}{x} + \frac{B}{x+1} + \frac{C}{(x+1)^2} + \frac{D}{(x+1)^3}.$$

Note that in the above decomposition there are as many terms with denominator $(x+1)^r$ as the multiplicity of the root -1 of the polynomial $x(x+1)^3$. The constants A, B, C, and D are determined by solving

$$2x^3+5x^2+6x+2 = A(x+1)^3 + Bx(x+1)^2 + Cx(x+1) + Dx$$

$$\Leftrightarrow 2x^3+5x^2+6x+2 = (A+B)x^3 + (3A+2B+C)x^2 + (3A+B+C+D)x + A$$

$$\Leftrightarrow \begin{cases} A+B &= 2 \\ 3A+2B+C &= 5 \\ 3A+B+C+D &= 6 \\ A &= 2 \end{cases} \Leftrightarrow \begin{cases} A &= 2 \\ B &= 0 \\ C &= -1 \\ D &= 1 \end{cases}.$$

We then have

$$\frac{2x^3+5x^2+6x+2}{x(x+1)^3} = \frac{2}{x} - \frac{1}{(x+1)^2} + \frac{1}{(x+1)^3},$$

and

$$\int \frac{2x^3+5x^2+6x+2}{x(x+1)^3}\,dx = \int \frac{2}{x} - \frac{1}{(x+1)^2} + \frac{1}{(x+1)^3}\,dx = 2\ln|x| - \frac{(x+1)^{-1}}{-1} + \frac{(x+1)^{-2}}{-2} + c$$

$$= \ln x^2 + \frac{1}{x+1} - \frac{1}{2(x+1)^2} + c.$$

(c) To determine the antiderivatives of the real function of a real variable

$$h(x) = \frac{x+2}{(x-1)(x^2+1)}$$

we consider de decomposition

$$\frac{x+2}{(x-1)(x^2+1)} = \frac{A}{x-1} + \frac{Bx+C}{x^2+1}.$$

Note the denominator of the last partial fraction is an irreducible polynomial of second degree. The correspondent numerator is then of degree up to 1. The constants A, B, and C are determined in the usual way

$$x+2 = A(x^2+1) + (Bx+C)(x-1)$$

$$\Leftrightarrow x+2 = (A+B)x^2 + (2A-B+C)x + (A-C) \Leftrightarrow \begin{cases} A+B &= 0 \\ 2A-B+C &= 1 \\ A-C &= 2 \end{cases} \Leftrightarrow \begin{cases} A &= \frac{3}{2} \\ B &= -\frac{3}{2} \\ C &= -\frac{1}{2} \end{cases}.$$

The decomposition we have obtained is

$$\frac{x+2}{(x-1)(x^2+1)} = \frac{\frac{3}{2}}{x-1} + \frac{-\frac{3}{2}x - \frac{1}{2}}{x^2+1},$$

and

$$\int \frac{x+2}{(x-1)(x^2+1)}\,dx = \int \frac{\frac{3}{2}}{x-1} + \frac{-\frac{3}{2}x - \frac{1}{2}}{x^2+1}\,dx = \frac{3}{2}\int \frac{1}{x-1}\,dx - \frac{3}{4}\int \frac{2x}{x^2+1}\,dx - \frac{1}{2}\int \frac{1}{x^2+1}\,dx$$

$$= \frac{3}{2}\ln|x-1| - \frac{3}{4}\ln\left|x^2+1\right| - \frac{1}{2}\arctan x + c$$

$$= \ln\sqrt[4]{\frac{(x-1)^6}{(x^2+1)^3}} - \frac{1}{2}\arctan x + c.$$

Exercises.

(1) Determine the family of antiderivatives of each one of the following real functions of a real variable:

a) $f(x) = x^2$;

b) $f(x) = 2x + 2$;

c) $f(x) = \frac{1}{2}x^2$;

d) $f(x) = 2x^2 + 4x + 4$;

e) $f(x) = c$, with c a real constant;

f) $f(x) = 2x^2 + 4$;

g) $f(x) = 2x^5 + 8x^2 + x - 78$;

h) $f(x) = \frac{1}{x^2} + 3x^{\frac{1}{3}}$;

i) $f(x) = \frac{3}{x^4} - \sqrt[4]{x} + x$;

j) $f(x) = 6x^{1/3} - x^{0.4} + \frac{9}{x^2}$;

k) $f(x) = \frac{1}{\sqrt[3]{x}} + \sqrt{x}$;

l) $f(x) = (x^4 + 4x + 2)(2x + 3)$;

m) $f(x) = (2x - 1)(3x^2 + 2)$;

n) $f(x) = (x^3 - 12x)(3x^2 + 2x)$;

o) $f(x) = (a + bx^3)^2$, with a and b real constants;

p) $f(x) = \sqrt{2ax}$, with a a real constant;

q) $f(x) = \frac{1}{\sqrt{x}}$;

r) $f(x) = \cos 5x \sin 5x$;

s) $f(x) = \sin^5 4x \cos 4x$;

t) $f(x) = 4e^{5x}$;

u) $f(x) = xe^{4x^2}$;

v) $f(x) = (x+5)^2 e^{(x+5)^3}$;

w) $f(x) = \frac{1}{1+x}$; $f(x) = \frac{1}{1+x^2}$; $f(x) = \frac{x}{1+x^2}$; $f(x) = \frac{x}{(1+x^2)^2}$;

x) $f(x) = \frac{e^x}{1+e^x}$; $f(x) = \frac{e^x}{1+e^{2x}}$; $f(x) = \frac{e^x}{(1+e^x)^2}$;

y) $f(x) = \frac{\cos x}{1+\sin x}$; $f(x) = \frac{\cos x}{1+\sin^2 x}$; $f(x) = \frac{\cos x}{(1+\sin x)^2}$; $f(x) = \cos x(1+\sin x)^2$;

z) $f(x) = \frac{\ln x}{x}$; $f(x) = \frac{\ln^5 x}{x}$; $f(x) = \frac{1}{x(1+\ln x)}$; $f(x) = \frac{1}{x(1+\ln^2 x)}$.

(2) Antidifferentiate the following rational functions:

a) $f(x) = \frac{1}{(x+1)(x+2)}$;

b) $f(x) = \frac{x}{x+1}$;

c) $f(x) = \frac{1}{x(x+1)}$;

d) $f(x) = \frac{x^2-5x+1}{x^2-5x+8}$;

e) $f(x) = \frac{x^2}{x^2+1}$;

f) $f(x) = \frac{x^2+2x}{x^2-1}$;

g) $f(x) = \frac{x}{x^4+4}$;

h) $f(x) = \frac{2x^3}{x^4-1}$.

(3) Antidifferentiate by parts the functions:

a) $f(x) = xe^x$; $f(x) = x^2e^x$; $f(x) = x^2e^{3x}$;

b) $f(x) = \ln x$; $f(x) = \arctan x$;

c) $f(x) = x \sin x$;

d) $f(x) = x \cos 3x$;

e) $f(x) = \frac{x}{e^x}$;

f) $f(x) = x^2 \ln x$;

g) $f(x) = x \arctan x$;

h) $f(x) = \sin 2x \cos 3x$;

i) $f(x) = x \sin x \cos x$.

(4) Antidifferentiate by substitution the following functions:

a) $f(x) = \frac{x+\sqrt{x}}{x-\sqrt{x}}$;

b) $f(x) = \frac{x^2}{\sqrt{2-x^2}}$;

c) $f(x) = \frac{\sqrt{x}-1}{\sqrt[3]{x}+1}$;

d) $f(x) = \frac{e^{3x}}{1-e^{2x}}$;

e) $f(x) = \frac{\cos x}{\sin^2 x-2}$;

f) $f(x) = 2 + \sqrt{1-x^2}$;

g) $f(x) = \frac{e^{2x}}{\sqrt{1+e^x}}$.

(5) Antidifferentiate the following functions:

a) $f(x) = (-2x+5)e^{-x}$;

b) $f(x) = \frac{x}{\sqrt{x+1}}$;

c) $f(x) = e^{\sqrt{x}}$;

d) $f(x) = xe^{-x^2}$;

e) $f(x) = x(x^2+1)^{20}$;

f) $f(x) = x \cos x$;

g) $f(x) = \frac{1}{e^{2x}-3e^x}$;

h) $f(x) = \frac{\sqrt{1+\sqrt{x}}}{\sqrt{x}}$;

i) $f(x) = \frac{x^6+1}{x+1}$;

j) $f(x) = \frac{\sin x}{1+\sin x}$.

13. Integration

Definite integral.

We give the notion of *definite integral*. Consider a closed interval $[a, b]$, with a and b real numbers such that $a \leq b$, and points

$$a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b.$$

The set $P = \{x_0, x_1, \dots, x_{n-1}, x_n\}$ is called a *partition* of the interval $[a, b]$. Denote

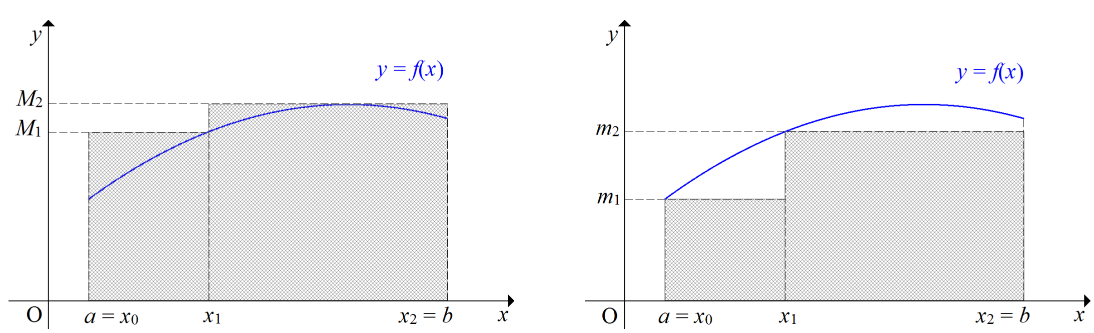
$$\Delta x_i = x_i - x_{i-1} \quad \text{for} \quad i = 1, 2, \dots, n.$$

Let f be a bounded real function in $[a, b]$. For a given partition P of $[a, b]$ we set

$$M_i = \sup f(x), \quad m_i = \inf f(x), \quad \text{with } x_{i-1} \leq x \leq x_i,$$

and define the *upper and lower Darboux sums* (see Figure 13.1), respectively,

$$U(P, f) = \sum_{i=1}^{n} M_i \Delta x_i \quad \text{and} \quad L(P, f) = \sum_{i=1}^{n} m_i \Delta x_i.$$



(a) Upper Darboux sum. (b) Lower Darboux sum.

Figure 13.1. Darboux sums.

Now, the *upper and lower Riemann integrals* are defined, respectively, as,

$$\overline{\int_a^b} f(x)\,dx = \inf U(P,f) \quad \text{and} \quad \underline{\int_a^b} f(x)\,dx = \sup L(P,f),$$

where $\inf U(P,f)$ and $\sup L(P,f)$ are taken over all partitions P of $[a,b]$.
Finally, if

$$\overline{\int_a^b} f(x)\,dx = \underline{\int_a^b} f(x)\,dx,$$

the function f is said to be *Riemann-integrable* (or just *integrable*) in $[a,b]$, and we denote the common value of the upper and lower integrals by

$$\int_a^b f(x)\,dx \text{ or just by } \int_a^b f\,dx,$$

the *Riemann-integral* (or simply the *integral*) of f over the interval $[a,b]$.
There is some terminology we give now: the symbol $\int$ is called the *integral sign*; the notation dx is used to indicate that the integration is taken over the variable x, and x is called the *variable of integration*; the numbers a and b are called, respectively, the *lower and the upper limits of the integral*, and $[a,b]$ is called *interval of integration*; finally, the function $f(x)$ to be integrated is called the *integrand*.
We gave the notion of Riemann integral for bounded real functions in a closed and bounded interval. An important issue is to know what types of functions are integrable. The three theorems we state next give some important classes of integrable functions. Unless otherwise stated, we shall assume in the sequel that integrands are bounded in the interval of integration.

THEOREM 13.1. *If the real function f is continuous in $[a,b]$ then f is integrable in $[a,b]$.*

THEOREM 13.2. *If the real function f has only finitely many points of discontinuity in $[a,b]$ then f is integrable in $[a,b]$.*

THEOREM 13.3. *If the real function f is monotonic in $[a,b]$ then f is integrable in $[a,b]$.*

We then have that the continuous functions, the monotonic functions, and also the functions which are discontinuous at a finite number of points are integrable. Next theorem states that the composition of a continuous function and an integrable function is integrable.

THEOREM 13.4. *Let f be a real function of a real variable in an interval $[a, b]$. Suppose that*

(1) f is integrable in $[a, b]$;

(2) There exist numbers m and M such that $m \leq f \leq M$ for all $x \in [a, b]$;

(3) φ is a real continuous function in $[m, M]$.

Then the composite function $h(x) = (\varphi \circ f)(x) = \varphi(f(x))$ is integrable in $[a, b]$.

Examples. Consider the real functions of a real variable

$$f(x) = 2x^2 - 1, \qquad g(x) = \frac{x-1}{x+1}, \qquad h(x) = \begin{cases} x, & x < \frac{1}{2} \\ 1, & x \geq \frac{1}{2} \end{cases}, \qquad i(x) = \begin{cases} x^2 + 1, & x \neq \frac{1}{2} \\ 1, & x = \frac{1}{2} \end{cases}.$$

Let us evaluate the integrability of the above functions, and also of the function $f \circ h$ in the interval $[0, 1]$.

(a) f and g are continuous functions in $[0, 1]$, thus integrable in $[0, 1]$.

(b) h is a monotonically increasing function in $[0, 1]$, thus integrable in $[0, 1]$.

(c) i is not monotonic in $[0, 1]$ but it is continuous in the interval except for $x = \frac{1}{2}$. Therefore it is integrable in $[0, 1]$.

(d) Since h is integrable in $[0, 1]$ with $0 \leq h(x) \leq 1$ for all $x \in [0, 1]$ and f is continuous in $\mathbb{R}$, the composition $f \circ h$ defined by

$$(f \circ h)(x) = f(h(x)) = \begin{cases} 2x^2 - 1, & x < \frac{1}{2} \\ 1, & x \geq \frac{1}{2} \end{cases}$$

is integrable in $[0, 1]$.

The following theorems give some practical properties for computing integrals.

THEOREM 13.5. *If $f(x) = M$ in $[a, b]$, with M a constant, then*

$$\int_a^b f\,dx = M(b - a).$$

In particular, if $f(x) = 0$ in $[a, b]$ then

$$\int_a^b f\,dx = 0.$$

Examples.

(a) $\displaystyle\int_2^4 3\,dx = 3 \cdot (4 - 2) = 6.$

(b) $\displaystyle\int_2^4 1\,dx = 4 - 2 = 2.$

(c) $\displaystyle\int_1^5 0\,dx = 0.$

It is clear that if the upper and lower limits of the integral coincide,

$$\int_a^a f\,dx = 0,$$

for any function f. Also, by convention,

$$\int_a^b f\,dx = -\int_b^a f\,dx.$$

Example. We saw that

$$\int_2^4 3\,dx = 3 \cdot (4-2) = 6.$$

Then

$$\int_4^2 3\,dx = -\int_2^4 3\,dx = -6.$$

Note that the integral $\int_4^2 3\,dx$ could have been computed by extending Theorem 13.5 to the case where the integral upper and lower limits are reversed

$$\int_4^2 3\,dx = 3 \cdot (2-4) = -6.$$

THEOREM 13.6. *Let f and g be integrable functions in $[a,b]$. Then*

(1) The function $f+g$ is integrable in $[a,b]$ and

$$\int_a^b f+g\,dx = \int_a^b f\,dx + \int_a^b g\,dx;$$

(2) The function cf, with c a real constant, is integrable in $[a,b]$ and

$$\int_a^b cf\,dx = c\int_a^b f\,dx;$$

(3) If $f(x) \leq g(x)$ for each $x \in [a,b]$ then

$$\int_a^b f\,dx \leq \int_a^b g\,dx;$$

(4) The function $|f|$ is integrable in $[a,b]$ and

$$\left|\int_a^b f\,dx\right| \leq \int_a^b |f|\,dx;$$

(5) If $a < c < b$, then f is integrable in $[a,c]$ and in $[c,b]$ and

$$\int_a^b f\,dx = \int_a^c f\,dx + \int_c^b f\,dx;$$

(6) The function fg is integrable in $[a,b]$.

Note that, owing to the above mentioned convention

$$\int_b^a f\,dx = -\int_a^b f\,dx,$$

(5) in Theorem 13.6 can be rewritten, for example,

$$\int_a^b f\,dx = \int_a^c f\,dx - \int_b^c f\,dx.$$

The fundamental theorem of calculus.

We now consider the integral as a function of the upper limit

$$\int_a^x f(t)\,dt \quad \text{for } a \leq x \leq b.$$

Next results establish the connection between integration and differentiation. In particular, they give the conditions under which the technique of antidifferentiation can be used to compute integrals.

THEOREM 13.7. *Assume that f is an integrable real function in $[a,b]$. Put*

$$F(x) = \int_a^x f(t)\,dt \quad \textit{for } a \leq x \leq b.$$

Then F is continuous in $[a,b]$. Furthermore, if f is continuous at a point $x_0 \in [a,b]$ then F is differentiable at x_0 and

$$F'(x_0) = f(x_0).$$

Note that from the above theorem we learn that if f is continuous in $[a,b]$ then F is an antiderivative of f in $[a,b]$. Since

$$F(a) = \int_a^a f\,dx = 0,$$

F is the unique antiderivative of f which is null for $x = a$.

THEOREM 13.8 (Fundamental theorem of calculus). *Suppose that the real function f is integrable in $[a,b]$, and that there exists a differentiable real function F in $[a,b]$ such that $F' = f$. Then*

$$\int_a^b f\,dx = F(b) - F(a).$$

We shall use the notation

$$[F(x)]_a^b = F(b) - F(a).$$

Example. As the real function f defined by $f(x) = x$ is continuous in the interval $[1, 2]$, to integrate f over $[1, 2]$ we just have to find an antiderivative F of f and compute the difference $F(2) - F(1)$:

$$\int_1^2 x\,dx = \left[\frac{x^2}{2}\right]_1^2 = \frac{2^2}{2} - \frac{1^2}{2} = \frac{3}{2}.$$

Methods of integration.

THEOREM 13.9 (Integration by parts). *Suppose that u and v are real differentiable functions in $[a, b]$, and that u' and v' are integrable functions in $[a, b]$. Then*

$$\int_a^b u'v\,dx = [uv]_a^b - \int_a^b uv'\,dx.$$

Example. We want to compute

$$\int_1^2 e^x x\,dx.$$

We set $u' = e^x$ and $v = x$ so that $u = e^x$ and $v' = 1$. Note that u' and v' are continuous functions, therefore integrable. By integrating by parts we obtain

$$\int_1^2 e^x x\,dx = [e^x x]_1^2 - \int_1^2 e^x \cdot 1\,dx = [e^x x]_1^2 - [e^x]_1^2 = 2e^2 - e - \big(e^2 - e\big) = e^2.$$

Note that, alternatively, we can antidifferentiate by parts

$$\int e^x x\,dx = e^x x - \int e^x \cdot 1\,dx = e^x x - e^x + c = e^x(x-1) + c,$$

and then compute the definite integral

$$\int_1^2 e^x x\,dx = [e^x(x-1)]_1^2 = e^2(2-1) - e(1-1) = e^2.$$

THEOREM 13.10 (Integration by substitution). *Let f be an integrable function in $[a, b]$, and φ a strictly increasing differentiable function mapping $[A, B]$ onto $[a, b]$ such that φ' is integrable in $[a, b]$. Then*

$$\int_a^b f(x)\,dx = \int_A^B f(\varphi(t))\varphi'(t)\,dt.$$

Note that in the above theorem, $A = \varphi^{-1}(a)$ and $B = \varphi^{-1}(b)$.

Example. We want to determine

$$\int_1^2 \frac{\ln x}{x(1 + \ln x)}\,dx.$$

We consider the strictly increasing differentiable function $\varphi(t) = e^t$ mapping $[0, \ln 2]$ onto $[1, 2]$, and change variable $x = \varphi(t) = e^t$ (consequently, $t = \varphi^{-1}(x) = \ln x$, and $x' = \varphi'(t) = e^t$). Note that f and φ' are continuous functions, therefore integrable. Integrating by substitution,

$$\int_1^2 \frac{\ln x}{x(1+\ln x)}\,dx = \int_0^{\ln 2} \frac{t}{e^t(1+t)} \cdot e^t\,dt = \int_0^{\ln 2} \frac{t}{1+t}\,dt = \int_0^{\ln 2} 1 - \frac{1}{1+t}\,dt$$

$$= \int_0^{\ln 2} 1\,dt - \int_0^{\ln 2} \frac{1}{1+t}\,dt = [t]_0^{\ln 2} - [\ln|1+t|]_0^{\ln 2}$$

$$= \ln 2 - 0 - (\ln|1+\ln 2| - \ln|1+0|) = \ln 2 - \ln(1+\ln 2) = \ln \frac{2}{1+\ln 2}.$$

Alternatively, since $\varphi(t)$ maps one-to-one $[0, \ln 2]$ onto $[1, 2]$, we can antidifferentiate by substitution

$$\int \frac{\ln x}{x(1+\ln x)}\,dx = \left[\int \frac{t}{e^t(1+t)} \cdot e^t\,dt\right]_{t=\ln x} = \left[\int \frac{t}{1+t}\,dt\right]_{t=\ln x} = \left[\int 1 - \frac{1}{1+t}\,dt\right]_{t=\ln x}$$

$$= [t - \ln|1+t| + c]_{t=\ln x} = \ln x - \ln|1+\ln x| + c,$$

and then compute the definite integral

$$\int_1^2 \frac{\ln x}{x(1+\ln x)}\,dx = [\ln x - \ln|1+\ln x|]_1^2 = \ln 2 - \ln|1+\ln 2| - (\ln 1 - \ln|1+\ln 1|)$$

$$= \ln 2 - \ln(1+\ln 2) = \ln \frac{2}{1+\ln 2}.$$

Improper integrals.

We introduced the definite integral

$$\int_a^b f\,dx$$

under the hypotheses

- The interval $[a, b]$ is bounded;
- The function f is bounded in $[a, b]$.

We extend the notion of definite integral by relaxing each one of the hypotheses above, giving rise to the so-called *improper integrals*.

Improper integrals of the first kind. The improper integrals of the *first kind* are obtained by relaxing the hypothesis concerning the boundedness of $[a, b]$.

Let f be a real function in $[a, +\infty)$, and suppose that f is integrable in every bounded interval $[a, x]$, with $x > a$. We define

$$\int_a^{+\infty} f(x)\,dx = \lim_{x\to+\infty} \int_a^x f(t)\,dt$$

if the limit exists and is finite. In this case, we say that the integral $\int_a^{+\infty} f(x)\,dx$ *converges*. If not, we say that it *diverges*.

The case where it is the lower limit of the integral that is infinite is defined similarly:

$$\int_{-\infty}^a f(x)\,dx = \lim_{x\to-\infty} \int_x^a f(t)\,dt$$

if the limit exists and is finite.

Examples.

(a) In order to evaluate the improper integral

$$\int_0^{+\infty} e^{-x}\,dx$$

we compute

$$\lim_{x\to+\infty} \int_0^x e^{-t}\,dt = \lim_{x\to+\infty} \left(-\int_0^x e^{-t}\cdot(-1)dt\right) = -\lim_{x\to+\infty} \left[e^{-t}\right]_0^x = -\lim_{x\to+\infty} \left(e^{-x} - e^0\right) = -(0-1) = 1.$$

Since the above limit exists and is a finite number, we conclude that the improper integral converges and

$$\int_0^{+\infty} e^{-x}\,dx = 1.$$

(b) We want to evaluate the improper integral

$$\int_1^{+\infty} \frac{1}{x}\,dx.$$

We determine

$$\lim_{x\to+\infty} \int_1^x \frac{1}{t}\,dt = \lim_{x\to+\infty} [\ln|t|]_1^x = \lim_{x\to+\infty} (\ln|x| - \ln|1|) = +\infty - 0 = +\infty.$$

As the above limit is not finite, we conclude that the improper integral diverges.

(c) We evaluate the improper integral

$$\int_{-\infty}^{-1} \frac{1}{x^2}\,dx$$

by computing

$$\lim_{x\to-\infty}\int_x^{-1}\frac{1}{t^2}\,dt=\lim_{x\to-\infty}\int_x^{-1}t^{-2}\,dt=\lim_{x\to-\infty}\left[\frac{t^{-1}}{-1}\right]_x^{-1}=\lim_{x\to-\infty}\left[-\frac{1}{t}\right]_x^{-1}=\lim_{x\to-\infty}\left(-\frac{1}{-1}-\left(-\frac{1}{x}\right)\right)$$

$$=\lim_{x\to-\infty}\left(1+\frac{1}{x}\right)=1+0=1.$$

We then conclude that the improper integral converges and

$$\int_{-\infty}^{-1}\frac{1}{x^2}\,dx=1.$$

We now consider the case where both the lower and the upper limits of the integral are infinite. Let f be a real function in $\mathbb{R}$, and suppose that f is integrable in every bounded interval $[x_1,x_2]$, with $x_2>x_1$. We say that the improper integral

$$\int_{-\infty}^{+\infty}f(x)\,dx$$

converges if for $a\in\mathbb{R}$ both $\int_a^{+\infty}f(x)\,dx$ and $\int_{-\infty}^{a}f(x)\,dx$ converge. In this case, we define

$$\int_{-\infty}^{+\infty}f(x)\,dx=\int_a^{+\infty}f(x)\,dx+\int_{-\infty}^{a}f(x)\,dx.$$

The improper integral *diverges* if at least one of the integrals $\int_a^{+\infty}f(x)\,dx$ and $\int_{-\infty}^{a}f(x)\,dx$ diverges.

Example. We want to evaluate the improper integral

$$\int_{-\infty}^{+\infty}e^x\,dx.$$

Since

$$\lim_{x\to+\infty}\int_0^x e^t\,dt=\lim_{x\to+\infty}\left[e^t\right]_0^x=\lim_{x\to+\infty}\left(e^x-e^0\right)=+\infty-1=+\infty,$$

the integral $\int_0^{+\infty}e^x\,dx$ diverges. Consequently, the integral $\int_{-\infty}^{+\infty}e^x\,dx$ also diverges.

Improper integrals of the second kind. An improper integral is said to be of the *second kind* if the integrand is unbounded at a point of the bounded interval $[a,b]$.

Suppose that the real function f is integrable in every interval $[x,b]$, with $a<x\le b$, but not bounded in $(a,b]$. We define

$$\int_a^b f(x)\,dx=\lim_{x\to a^+}\int_x^b f(t)\,dt$$

if the limit exists and is finite. In this case, we say that the integral $\int_a^b f(x)\,dx$ *converges*. Otherwise, we say that it *diverges*.

The case where the unboundedness of f occurs near the upper limit of the integral is defined similarly:

$$\int_a^b f(x)\,dx = \lim_{x\to b^-} \int_a^x f(t)\,dt$$

if the limit exists and is finite.
If f is bounded in $[a,b]$ except near a point $c \in (a,b)$ we say that the improper integral

$$\int_a^b f(x)\,dx$$

converges if both $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ converge. In this case, we define

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

The improper integral *diverges* if at least one of the integrals $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ diverges.

Examples.

(a) We want to evaluate the improper integral

$$\int_0^1 \frac{1}{x}\,dx.$$

We determine

$$\lim_{x\to 0^+} \int_x^1 \frac{1}{t}\,dt = \lim_{x\to 0^+} [\ln|t|]_x^1 = \lim_{x\to 0^+} (\ln|1| - \ln|x|) = 0 - (-\infty) = +\infty.$$

As the above limit is infinite, we conclude that the improper integral diverges.

(b) In order to evaluate the improper integral

$$\int_{-1}^1 \frac{1}{x}\,dx$$

we begin by computing

$$\lim_{x\to 0^+} \int_x^1 \frac{1}{t}\,dt = +\infty,$$

so that the integral $\int_0^1 \frac{1}{x}\,dx$ diverges. Consequently, the integral $\int_{-1}^1 \frac{1}{x}\,dx$ also diverges.

The improper integrals where both the integrand is unbounded near a point of the interval of integration and the interval is unbounded are sometimes referred to as improper integrals of the *third kind*. Their evaluation is obtained by combining the procedures given for the improper integrals of the first and second kinds.

Application to the computation of areas.

One interesting application of the integration concerns the computation of the area of a region of the plane.

Suppose that f and g are integrable real functions in $[a,b]$, and that $g(x) \le f(x)$ for each $x \in [a,b]$. Then the area of the region

$$Z = \{(x,y) : a \le x \le b,\ g(x) \le y \le f(x)\}$$

is given by

$$\alpha(Z) = \int_a^b f(x) - g(x)\,dx.$$

Examples.

(a) We want to determine the area of the region bounded by the curves $y = x^2 + 2x + 1$, $y = x^2 - 2$, $y = 0$, $x = 0$, and $x = 2$ (see Figure 13.2(a)). The area is given by

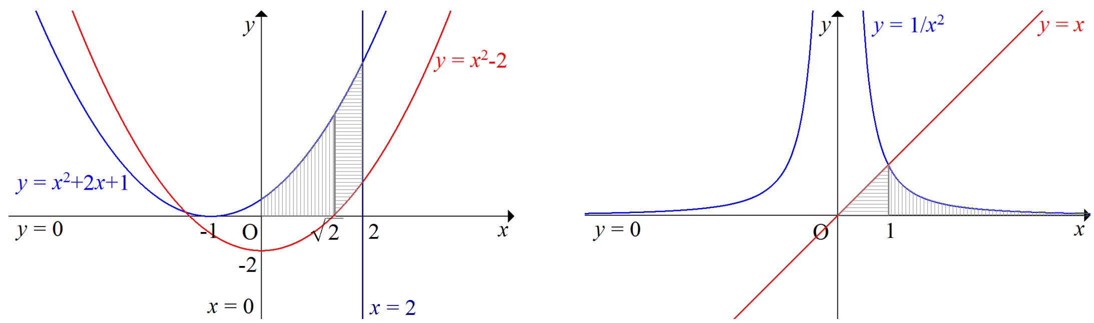
$$\begin{aligned}\int_0^{\sqrt{2}} x^2 + 2x + 1 - 0\,dx + \int_{\sqrt{2}}^2 x^2 + 2x + 1 - \left(x^2 - 2\right)dx &= \int_0^{\sqrt{2}} x^2 + 2x + 1\,dx + \int_{\sqrt{2}}^2 2x + 3\,dx \\ &= \left[\frac{x^3}{3} + x^2 + x\right]_0^{\sqrt{2}} + \left[x^2 + 3x\right]_{\sqrt{2}}^2 \\ &= \frac{30 - 4\sqrt{2}}{3}.\end{aligned}$$

(b) The area of the region defined by the conditions

$$\begin{cases} y \le \frac{1}{x^2} \\ y \le x \\ y \ge 0 \end{cases}$$

is (see Figure 13.2(b))

$$\begin{aligned}\int_0^1 x - 0\,dx + \int_1^{+\infty} \frac{1}{x^2} - 0\,dx &= \int_0^1 x\,dx + \lim_{x\to+\infty} \int_1^x \frac{1}{t^2}\,dt \\ &= \left[\frac{x^2}{2}\right]_0^1 + \lim_{x\to+\infty}\left[\frac{t^{-1}}{-1}\right]_1^x = \frac{1}{2} - \lim_{x\to+\infty}\left(\frac{1}{x} - 1\right) = \frac{1}{2} - (0 - 1) = \frac{3}{2}.\end{aligned}$$



(a) (b)

Figure 13.2

Exercises.

(1) Determine the value of the integrals:
 a) $\int_1^2 x^2 - 2x + 3\, dx$;
 b) $\int_0^8 \sqrt{2x} + \sqrt[3]{x}\, dx$;
 c) $\int_1^4 \frac{1+\sqrt{x}}{x^2}\, dx$;
 d) $\int_0^{\pi/4} \cos^2 x\, dx$;
 e) $\int_e^{e^2} \frac{1}{x \ln x}\, dx$;
 f) $\int_0^{-3} \frac{1}{\sqrt{25+3x}}\, dx$.

(2) Determine the area of the region bounded by the parabola with equation $y = \frac{x^2}{2}$ and the straight lines defined by $x = 1,$ $x = 3,$ and $y = 0$.

(3) Determine the area of the region bounded by the curves with equations:
 a) $y = x^3 + 1$ and $y = 2x^2 + x - 1$;
 b) $y = 2 - x^2$ and $y^3 = x^2$.

(4) Study the convergence of each one of the following improper integrals, and determine the value of the integral in the case of convergence:
 a) $\int_0^{+\infty} \frac{1}{1+x^2}\, dx$;
 b) $\int_1^{+\infty} \frac{1}{x^2}\, dx$;

c) $\int_{-\infty}^{-1} \frac{1}{x^2}\,dx$;

d) $\int_0^1 \frac{1}{\sqrt{x}}\,dx$;

e) $\int_0^{+\infty} \frac{\arctan x}{x^2+1}\,dx$;

f) $\int_0^1 \ln x\,dx$.

(5) Compute the derivatives of the following functions:

a) $f(x) = \int_0^x t^4\,dt$;

b) $g(x) = \int_0^{x^2} e^{t^2}\,dt$.

(6) Compute the following limits:

a) $\lim_{x\to 0} \frac{\int_0^x \cos t\,dt}{x}$;

b) $\lim_{x\to 0} \frac{\int_0^x \sin^2 t\,dt}{x^3}$.

Solutions to exercises

Chapter 1. Vectors: 2a) $\bar{u}\cdot\bar{v}=2$, $\|\bar{u}\|=\sqrt{3}$, $\|\bar{v}\|=\sqrt{6}$; 2b) No & No & Yes; 2c) $\sphericalangle(\bar{v},\bar{y})=\arccos\left(\frac{\sqrt{2}}{3}\right)$, $\sphericalangle(\bar{w},\bar{z})=\frac{3\pi}{4}$; 2d) $\mathrm{d}(\bar{u},\bar{v})=\sqrt{3}$, $\mathrm{d}(\bar{y},\bar{z})=\sqrt{29}$; 3a) Linearly independent, 3b) Linearly dependent, 3c) Linearly independent, 3d) Linearly dependent, 3e) Linearly dependent, 3f) Linearly dependent, 3g) Linearly dependent; 4a) $\bar{y}$ cannot be written as a linear combination of $\bar{v},\bar{w}$ and $\bar{x}$, $\bar{z}=\bar{v}+\bar{w}+3\bar{x}$ for example, the set $\{\bar{v},\bar{w},\bar{x}\}$ is linearly dependent, 4b) $\bar{z}=-\frac{1}{2}\bar{v}+\bar{w}$, the set $\{\bar{v},\bar{w}\}$ is linearly independent; 5) The set is linearly independent if and only if $\alpha\neq\frac{1}{2}$.

Chapter 2. Matrices: 1a) $A_{3\times3}$ square matrix, $B_{3\times4}$ rectangular matrix, $C_{1\times3}$ row matrix, $D_{2\times1}$ column matrix, $E_{1\times3}$ null row matrix, $F_{3\times3}$ square matrix, $G_{3\times3}$ diagonal matrix, $H_{3\times3}$ diagonal matrix, $I_{2\times2}$ identity matrix, $J_{3\times3}$ identity matrix; 1b) $-B=\begin{bmatrix}0&0&0&-1\\-1&0&2&5\\-1&-3&-2&0\end{bmatrix}$ 1c) Diagonal elements of F: $1,3,0$; 2a) $\begin{bmatrix}4&-3&3&5\\2&-5&-1&-4\end{bmatrix}$, 2b) Not possible, 2c) $\begin{bmatrix}-2&4&-4\\0&10&-\frac{6}{5}\end{bmatrix}$, 2d) $\begin{bmatrix}-3&-3&-3\end{bmatrix}$; 3a) $0A-F+2G=\begin{bmatrix}1&1&-2\\0&-1&0\end{bmatrix}$, 3b) Not possible, 3c) $2(B+H)=\begin{bmatrix}4&0&0\\2&4&-4\\-2&6&8\end{bmatrix}$, 3d) $3(2C+E)-C=\begin{bmatrix}5&0&-5\end{bmatrix}$; 4) $X=\frac{B-C+3A}{5}$; 5a) $\begin{bmatrix}4&-5\\-1&0\\0&0\end{bmatrix}$, 5b) $\begin{bmatrix}0\\-4\end{bmatrix}$, 5c) Not possible, 5d) $\begin{bmatrix}2\end{bmatrix}$, 5e) $\begin{bmatrix}0&-2&2\\0&2&-2\\0&0&0\end{bmatrix}$; 6a) $A+B=\begin{bmatrix}4&1&-1\\9&2&7\\3&-1&4\end{bmatrix}$, 6b) $A-B=\begin{bmatrix}-2&2&-5\\1&-2&-3\\-1&-1&-2\end{bmatrix}$, 6c) $AB=\begin{bmatrix}5&3&3\\19&-5&16\\1&-3&0\end{bmatrix}$, 6d) $BA=\begin{bmatrix}0&4&-9\\19&3&-3\\5&1&-3\end{bmatrix}$, 6e) $(AB)C=\begin{bmatrix}23&8&25\\92&-28&76\\4&-8&-4\end{bmatrix}$;

7) $C(A+3IB)D = \begin{bmatrix} -6 & 0 & -6 \end{bmatrix}$; 8) $(A+B)(A-B) = A^2 - B^2$ if A and B are commuting matrices; 10) $a = 2$; 13) $a = -\frac{3}{4}$ and $= \frac{3}{4}$; 14b) Notice that $CD \neq I$, 14c) No since $|D| = 0$; 16) $X = A^{-1}DC^{-1} - B$; 17) $X = \frac{1}{2}(A')^{-1}DA' + B'$; 19a) $2A_1 - A_2 + 3A_3 = \begin{bmatrix} 2 & 5 & 5 & 3 \end{bmatrix}$, 19b) $a_1 + a_2 - 3a_4 = \begin{bmatrix} 3 \\ -1 \\ -3 \end{bmatrix}$, 19c) Yes & no, 19d) $\mathrm{r}(A) = 3$; 21a) If $x = -1 \vee x = 2$ then $\mathrm{r}(A) = 2$, if $x \neq -1 \wedge x \neq 2$ then $\mathrm{r}(A) = 3$, 21b) If $t = \pm 2 \vee t = -4$ then $\mathrm{r}(A) = 2$, if $t \neq -4 \wedge t \neq 2 \wedge t \neq -2)$ then $\mathrm{r}(A) = 3$, 21c) If $z = 0 \wedge w = 0$ then $\mathrm{r}(A) = 2$, if $z \neq 0 \vee w \neq 0$ then $\mathrm{r}(A) = 3$.

Chapter 3. Determinants: 1a) 0, 1b) 0, 1c) -10, 1d) 0, 1e) -3, 1f) 0, 1g) 0, 1h) 0, 1i) 0, 1j) -6, 1k) 3, 1l) 0, 1m) $-abc$, 1n) $abcd$, 1o) 4, 1p) 360; 2a) 42, 2b) 0, 2c) $abcd$; 5) $|A'B| = 0$; 6) $|C| = 0$; 7) $\mathrm{adj}(A) = \begin{bmatrix} 3 & -2 \\ -1 & -1 \end{bmatrix}$, $A^{-1} = -\frac{1}{5}\begin{bmatrix} 3 & -2 \\ -1 & -1 \end{bmatrix}$, $\mathrm{adj}(B) = \begin{bmatrix} -2 & 2 \\ -1 & 1 \end{bmatrix}$, B^{-1} does not exist, $\mathrm{adj}(C) = \begin{bmatrix} -7 & 6 & -1 \\ 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix}$, $C^{-1} = -\frac{1}{2}\begin{bmatrix} -7 & 6 & -1 \\ 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix}$, $\mathrm{adj}(D) = D^{-1} = D$, $\mathrm{adj}(E) = \begin{bmatrix} 0 & -3 & 2 \\ 1 & 3 & -2 \\ 0 & 2 & -1 \end{bmatrix}$, $E^{-1} = \begin{bmatrix} 0 & -3 & 2 \\ 1 & 3 & -2 \\ 0 & 2 & -1 \end{bmatrix}$, $\mathrm{adj}(F) = \begin{bmatrix} 6 & 0 & -3 \\ -4 & 0 & 2 \\ -6 & 0 & 3 \end{bmatrix}$, F^{-1} does not exist, $\mathrm{adj}(G) = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -6 \end{bmatrix}$, $G^{-1} = -\frac{1}{6}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -6 \end{bmatrix}$; 8) A is invertible if $a \neq -\frac{1}{3}$ & for $a = 1$, $A^{-1} = \frac{1}{4}\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$.

Chapter 4: Systems of linear equations: 4) $X = \begin{bmatrix} 5 & 7 & -3 \\ 4 & 7 & -4 \\ -\frac{3}{2} & -2 & \frac{3}{2} \end{bmatrix}$; 5) $M^{-1} = \begin{bmatrix} 5 & 2 & -1 \\ -4 & -2 & 1 \\ -3 & -1 & 1 \end{bmatrix}$ and $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ -5 \\ -2 \end{bmatrix}$; 6a) $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\frac{2}{5} \\ -\frac{1}{5} \\ -\frac{9}{5} \end{bmatrix}$, 6b) $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$, 6c) $\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} -2-y \\ y \\ 3+u \\ u \end{bmatrix}$, $y, u \in \mathbb{R}$, 6d) $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$, 6e) $\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} -3 \\ 6 \\ 5 \\ -5 \end{bmatrix}$; 7) Unique solution if $|M| = -10$ and $(x_1, x_2, x_3) = \left(\frac{1}{2}b_1 - \frac{1}{10}b_2 - \frac{1}{5}b_3, \frac{3}{10}b_2 - \frac{1}{2}b_1 + \frac{3}{5}b_3, \frac{2}{5}b_3 - \frac{1}{2}b_1 + \frac{7}{10}b_2\right), b_1, b_2, b_3 \in \mathbb{R}$; 9) If $a = 0 \wedge b = \frac{9}{2}$ the system is consistent and dependent, if $a = 0 \wedge b \neq \frac{9}{2}$ the system is inconsistent, if $a = 7 \wedge b = \frac{10}{3}$ the system is consistent and dependent, if $a = 7 \wedge b \neq \frac{10}{3}$ the system is inconsistent, if $a \neq 0 \wedge a \neq 7$ the system is consistent and independent $\forall b$; 10a) $p \neq 3 \forall q$, 10b) $p = 3 \wedge q = 0$, $p = 3 \wedge q \neq 0$; 11) $c^* = \frac{3}{2}$, $k^* = 2$; 12) $a = \frac{7}{5} \wedge b = 3$; 13) If $\alpha = 1$ the system is consistent and dependent, otherwise the system is inconsistent.

Chapter 5. The real number system: 4a) $x = 3$, 4b) $x = 0 \vee x = -5$, 4c) $x = \pm 3$; 7a) $1 < x < 2$, 7b) $-1 \leq x \leq 0$ 7c) $x < -1 \vee x > 3$, 7d) $x \leq -7 \vee x \geq 3$, 7e) $\frac{1}{6} < x < \frac{1}{4}$, 7f) $x > 2$, 7g) $\left(-\sqrt{3} < x < -1\right) \vee \left(1 < x < \sqrt{3}\right)$, 7h) $\frac{1}{3} \leq x \leq 1$, 7i) $x < 1 \vee x > 5$, 7j) $x < -1 \vee x > 3$, 7k) $\frac{1}{4} < x < \frac{1}{2}$, 7l) $x > 1$, 7m) $\left(-\sqrt{7} \leq x \leq -\sqrt{3}\right) \vee \left(\sqrt{3} \leq x \leq \sqrt{7}\right)$, 7n) $4 < x < 6$, 7o) $-\frac{1}{2} \leq x \vee x \geq -\frac{3}{2}$, 7p) $x \in \mathbb{R}$, 7q) $x = \emptyset$, 7r) $x < -1 \vee -1 < x < 0$; 8a) $UB = [1, +\infty)$, $LB = (-\infty, \frac{\sqrt{2}}{2}]$, $\sup = 1$, $\inf = \frac{\sqrt{2}}{2}$, $\max = 1$, $\min = \frac{\sqrt{2}}{2}$, 8b) $UB = [\frac{1}{2}, +\infty)$, $LB = (-\infty, -1]$, $\sup = \frac{1}{2}$, $\inf = -1$, $\max = \frac{1}{2}$, $\min = -1$, 8c) $UB = \emptyset$, $LB = (-\infty, 1]$, the supremum does not exist, $\inf = 1$, the maximum does not exist, the minimum does not exist, 8d) $UB = (2, +\infty)$, $LB = (-\infty, 0]$, $\sup = 2$, $\inf = 0$, $\max = 2$; the minumum does not exist; 9a) $\mathring{A} = (2,3) \cup (4,10)$, $A' = [2,3] \cup [4,10]$, $\overline{A} = [2,3] \cup [4,10]$, $UB = [10, +\infty)$, $LB = (-\infty, 2]$, $\sup_A = 10$, $\inf_A = 2$, the maximum does not exist, $\min_A = 2$, 9b) $\mathring{B} = (5,7)$, $B' = [5,7]$, $\overline{B} = [5,7] \cup \{15\}$, $UB = [15, +\infty)$, $LB = (-\infty, 5]$, $\sup_B = 15$, $\inf_B = 5$, $\max_B = 15$, the minimum does not exist, 9c) $\mathring{C} = \emptyset$, $C' = [0,1]$, $\overline{C} = [0,1]$, $UB = [1, +\infty)$, $LB = (-\infty, 0]$, $\sup_C = 1$, $\inf_C = 0$, the maximum does not exist, the minimum does not exist, 9d) $\mathring{D} = \emptyset$, $D' = [2,3]$, $\overline{D} = [2,3]$, $UB = [3, +\infty)$, $LB = (-\infty, 2]$, $\sup_D = 3$, $\inf_D = 2$, $\max_D = 3$, $\min_D = 2$; 10a) $\mathring{A} = (-7,7)$, $A' = [-7,7]$, $\overline{A} = [-7,7]$, 10b) $\mathring{B} = \emptyset$, $B' = [-7,7]$, $\overline{B} = [-7,7]$; 11a) $\mathring{A} = \emptyset$, $A' = \{1\}$, $\overline{A} = A \cup \{1\}$, 11b) A is neither open nor closed; 12) $\mathring{A} = (-\sqrt{2}, 2)$, $A' = [-8, \sqrt{13}]$; 13) $\mathring{A} = (-2,1)$, $A' = A$.

Chapter 6: Sequences: 1a) 0, 1b) 0, 1c) 1, 1d) $\frac{2}{3}$, 1e) 0, 1f) 0; 3a) $-\frac{1}{4}$, 3b) $\frac{2}{3}$, 3c) ∞, 3d) 0, 3e) 0, 3f) $\frac{1}{2}$, 3g) $-\frac{1}{4}$, 3h) ∞, 3i) 1, 3j) Does not exit 3k) ∞; 4a) 0, 4b) ∞, 4c) Does not exist 4d) ∞, 4e) 2, 4f) 0, 4g) 0, 4h) 0; 5a) e^4, 5b) 0, 5c) 1.

Chapter 7. Series: 1a) 2, 1b) Diverges, 1c) $\frac{4}{3}$, 1d) $\frac{1}{48}$, 1e) Diverges; 2a) $x \in (-\frac{1}{2}, +\infty)$, $S = x + 1$, 2b) $x \in (-\infty, -2) \cup (2, +\infty)$, $S = \frac{x}{x-2}$, 2c) $x \in (-2, 0) \cup (0, 2)$, $S = \frac{1}{|x|}$, 2d) $x \in (-2, 0)$, $S = -\frac{1}{x(x+2)}$, 2e) $x \in \mathbb{R}$, 2f) $x \in (-\infty, -2) \cup (2, +\infty)$, 2g) $x \in \mathbb{R}$, 2h) $x \in \mathbb{R}$, 2i) $x \in \mathbb{R}$; 3a) $\frac{11}{3}$, 3b) $\frac{1571427}{999999}$, 3c) $\frac{13}{11}$, 3d) 1; 4a) $R = 1$, $-1 < x < 1$, 4b) $R = 1$, $-1 < x < 0$, 4c) $R = 3$, $-1 - \sqrt{3} < x < \sqrt{3} - 1$, 4d) $R = e$, $-e < x < e$.

Chapter 8: One-variable functions: 1a) $[-1, +\infty)$, 1b) $(-\infty, 1)$, 1c) $(-\infty, -2) \cup (2, +\infty)$, 1d) $[-3, 3]$, 1e) $\mathbb{R}$, 1f) $(-\infty, -4) \cup (4, +\infty)$, 1g) $(-\infty, -2] \cup [6, +\infty)$, 1h) $(1, +\infty)$, 1i) $(1, +\infty) \setminus \{e^e\}$, 1j) $(-1, 3)$, 1k) $\mathbb{R}^+$, 1l) $\mathbb{R} \setminus \left\{\frac{\pi}{2} + 2k\pi \in \mathbb{Z}\right\}$, 1m) $\mathbb{R} \setminus \left\{\frac{\pi}{4} + k \in \mathbb{Z}\right\}$, 1n) $(9, 25]$, 1o) $\mathbb{R} \setminus \{-6, -4\}$; 3a) $x \in \{5\}$, $y \in \{5\}$, 3b) $x \in \{-1, 2\}$, $y \in \{2\}$, 3c) $x \in \left\{0, \frac{1}{3}, -1, 1\right\}$, $y \in \{0\}$, 3d) $x \in \{-4\}$, $y \in \left\{-\frac{4}{3}\right\}$, 3e) $x \in \left\{\frac{5}{2}\right\}$, $y \in \{\}$, 3f) $x \in \left\{\ln \frac{3}{2}\right\}$, $y \in \left\{-\frac{1}{2}\right\}$; 4a) $x \in \left\{\pm\frac{\pi}{6} + k\pi,\ \ k \in \mathbb{Z}\right\}$, 4b) $x \in \left\{\frac{2\pi}{3} + 2k\pi,\ \ k \in \mathbb{Z}\right\}$, 4c) $x = \emptyset$, 4d) $x = 4$, 4e) $x = 1$, 4f) $x = -\frac{1}{3}$, 4g) $x = -1$; 5a) $x \in [4, +\infty)$, 5b) $x \in (-5, +\infty)$, 5c) $x \in (-\frac{1}{3}, +\infty)$, 5d) $x \in (\frac{e-1}{3}, +\infty)$, 5e) $x \in (1, 3)$, 5f) $x \in (\frac{\sqrt{5}-1}{2}, 1)$.

Chapter 9. Limits and continuity: 1a) 5, 1b) $+\infty$, 1c) 0, 1d) $-1/3$, 1e) $-\infty$, 1f) $1/2$, 1g) The limit does not exist, 1h) $1/3$, 1i) $-1/2$, 1j) $+\infty$, 1k) $+\infty$, 1l) 0, 1m) 4, 1n) $-\infty$, 1o) 0 1p) $+\infty$; 2a) Continuous at $x = 0$, 2b) Not defined at both $x = 1$ and $x = 4$, 2c) Not defined at $x = 0$ and continuous at $x = 3$; 3a) Continuous in $\mathbb{R}$, 3b) Continuous in $\mathbb{R} - \{1, 4\}$, 3c) Continuous in $(-1, +\infty) - \{0\}$, 3d) Continuous in $\mathbb{R}$, 3e) Continuous in $(0, +\infty)$, 3f) Continuous in $\mathbb{R}$; 3g) Continuous in $\mathbb{R}$; 4a) The limit does not exist, 4b) 0, 4c) 1, 4d) 0; 6) $g(0) = -3/5$; 7) $g(0) = 1$; 8a) $a = -1/3$, 8b) $a = 5/9$, $b = 17/9$, 8c) $a \in \mathbb{R}$, $b = 0$.

Chapter 10: Differentiation: 1a) $2x$, 1b) 2, 1c) x, 1d) $4x + 4$, 1e) 0, 1f) $-\frac{2}{x^3} + \frac{1}{\sqrt[3]{x^2}}$, 1g) $-\frac{12}{x^5} - \frac{1}{4\sqrt[4]{x^3}} + 1$, 1h) $\frac{2}{\sqrt[3]{x^5} - \frac{4}{10\sqrt[10]{x^6}} - \frac{18}{x}}$, 1i) $\frac{1}{3\sqrt[3]{x^2} + \frac{1}{2\sqrt{x}}}$, 1j) $10x^4 + 12x^3 + 16x + 16$, 1k) $18x^2 - 6x + 4$, 1l) $15x^4 + 8x^3 - 108x^2 - 48x$, 1m) $36x^5 + 10x^4 - 6x - 1$, 1n) $\frac{4}{(x+5)^2}$, 1o) $\frac{18x^5+12x^3-6x^-12x+2}{(3x^2+1)^2}$, 1p) $\frac{x^{\frac{3}{2}}+6\sqrt{x}-2}{2(x+2)^2}$, 1q) $\frac{2x^2-2x-1}{(2x-1)^2}$, 1r) $2(x+5)$, 1s) $6x^5-16x^3+30x^2+8x-20$, 1t) $-\frac{x}{\sqrt{1-x^2}}$, 1u) $\frac{24(x+2)^2+(8x^2-1)}{(4x^3+1)^2}$, 1v) $\frac{6x+5}{\sqrt{2}\sqrt{x+1}}$, 1w) $1-\frac{2x+1}{\sqrt{2x+2}}$, 1x) $\frac{4x(3x^4+2x)^2}{2\sqrt{2x^2+1}}+2(3x^4+2x)(12x^3+2)\sqrt{2x^2+1}$, 1y) $\frac{2}{(x^4+4x+2)^2}-\frac{2(2x+3)(4x^3+4)}{(x^4+4x+2)^3}$, 1z) $\frac{7x^4-3x^2+4x}{2\sqrt{x^3+1}}$, 1a′) $16((2x + 3)^4 + 8x + 14)((2x + 3)^3 + 1)$, 1b′) $\frac{x}{\sqrt{1+x^2}}$, 1c′) $2e^{2x}(3x^2 + 6x + e)$, 1d′) $(4x + 3)e^{2x^2+3x}$, 1e′) $4xe^{2x^2+1}$, 1f′) $2^{x-3}\ln 2\sqrt{x^3 - 2} + \frac{2^{x-4}*3x^2}{\sqrt{x^3-2}} + \frac{1}{x}$, 1g′) $\frac{1}{x} - 2e^x + \frac{1}{2\sqrt{x}}$, 1h′) $\frac{4x+3}{(x^2+x)\ln x^3(x+1)}$, 1i′) $\frac{4x+3}{2x^2+3x}$, 1j′) $\frac{4}{x}\left(\ln^3 x + 2\right)$, 1k′) $\frac{1}{\tan x}$, 1l′) $\frac{2}{1-x^2}$, 1m′) $\frac{x}{x^2+1}$, 1n′) $x + 2x\log_4 x$, 1o′) $3e^x + 4\sin x - \frac{1}{4x}$, 1p′) $\frac{1}{x^2+1} + \sec^2 x$, 1q′) $\frac{1}{2\sqrt{1-\frac{x^2}{4}}}$, 1r′) $-\frac{4x}{\sqrt{1-4x^4}}$, 1s′) $2(\cos 2x - \cos x \sin x + \csc^2 x)$, 1t′) $\frac{-2}{(\sin x - \cos x)^2}$, 1u′) $2e^x \cos x$, 1v′) $e^x(\sin ax + a\cos ax)$; 2a) $2xe^{x^2}$, 2b) $\ln(2)2^x e^{2^x}$, 2c) $\frac{4x^3}{1+x^8}$, 2d) $\frac{2}{4x^2+16x+17}$, 2e) $\frac{4}{x}$, 2f) $\frac{1}{x+2}$, 2g) $-\frac{4x^3}{1+2x^4+x^8}$, 2h) $-\frac{1}{2x^2+8x+8}$, 2i) $\frac{e^x}{1+e^{2x}}$; 3a) $f'(x) = 0$ iff $x = 0 \vee x = -2$, $g'(x) = 0$ iff $x = 1 \vee x = -1$, $h'(x) = 0$ iff $x = 1 \vee x = \frac{1}{3}$, 3b) $y = 15x - 22$, $y = -1$, $y = 10 - 5x$; 4a) Yes, 4b) No; 7a) 1, 7b) 2, 7c) 0, 7d) 0, 7e) 1, 7f) $\frac{1}{5}$.

Chapter 11: Optimisation: 1a) Minimum: 2 at $x = -1 \vee x = 1$, 1b) Maximum: 3432 at $x = 6$ & minimum: 3000 at $x = 0$, 1c) Maximum: $\frac{8}{3\sqrt{3}}$ at $x = -\frac{1}{\sqrt{3}}$ & minimum: $-\frac{8}{3\sqrt{3}}$ at $x = \frac{1}{\sqrt{3}}$, 1d) No maximum & minimum: 0 at $x = 0$, 1e) No maximum & no minimum, 1f) No maximum & no minimum, 1g) Maximum: 1 at $x = \frac{\pi}{2} \vee x = \frac{3\pi}{2}$ & minimum: 0 at $x = \pi$, 1h) Maximum: $\frac{5\pi}{6} + \frac{\sqrt{3}}{2}$ at $x = \frac{5\pi}{3}$ & minimum: $\frac{\pi}{6} - \frac{\sqrt{3}}{2}$ at $x = \frac{\pi}{3}$.

Chapter 12: Antidifferentiation: 1a) $F(x) = \frac{x^3}{3} + K$, 1b) $F(x) = x^2 + 2x + K$, 1c) $\frac{x^3}{6} + K$, 1d) $F(x) = \frac{2x^3}{3} + 2x^2 + 4x + K$, 1e) $F(x) = cx + K$, 1f) $\frac{2x^3}{3} + 4x + K$, 1g) $\frac{x^6}{3} + \frac{8x^3}{3} + \frac{x^2}{2} - 78x + K$, 1h) $F(x) = -\frac{1}{x} + \frac{9x^{\frac{4}{3}}}{4} + K$, 1i) $F(x) = -\frac{4x^{\frac{5}{4}}}{5} - \frac{1}{x^3} + \frac{x^2}{2} + K$, 1j) $F(x) = \frac{9x^{\frac{4}{3}}}{2} - \frac{5x^{\frac{7}{5}}}{7} + -\frac{9}{x} + K$, 1k) $F(x) = \frac{2x^{\frac{3}{2}}}{3} + \frac{3x^{\frac{2}{3}}}{2} + K$, 1l) $F(x) = \frac{x^6}{3} + \frac{3x^5}{5} + \frac{8x^3}{3} + 8x^2 + 6x + x + K$, 1m) $F(x) = \frac{3x^4}{2} - x^3 + 2x^2 - 2x + K$, 1n) $F(x) = \frac{x^6}{2} + \frac{2x^5}{5} - 9x^4 - 8x^3 + K$, 1o) $F(x) = ax^2 + \frac{1}{2}abx^4 + \frac{1}{7}b^2x^7 + K$, 1p) $F(x) = \frac{2}{3}\sqrt{2ax^3} + K$, 1q) $F(x) = \sqrt{4x} + K$, 1r) $F(x) = -\frac{1}{20}\cos(10x) + K$, 1s) $F(x) = \frac{\sin^6(4x)}{24} + K$, 1t) $F(x) = \frac{4}{5}e^{5x} + K$, 1u) $F(x) = \frac{1}{8}e^{4x^2} + K$, 1v) $F(x) = \frac{1}{3}e^{(x+5)^3} + K$, 1w) $F(x) = \ln(1+x) + K$, $F(x) = \arctan(x) + K$, $F(x) = \frac{1}{2}\ln(1+x^2) + K$, $F(x) = -\frac{1}{2(1+x^2)} + K$, 1x) $F(x) = \ln(1+e^x) + K$, $F(x) = \frac{1}{2}\ln(1+e^{2x}) + K$, $F(x) = \frac{1}{1+e^x} + K$, 1y) $F(x) = \ln(1+\sin x) + K$, $F(x) = \arctan(\sin x) + K$, $F(x) = -\frac{1}{1+\sin x} + K$, $F(x) = \frac{(1+\sin x)^3}{3} + K$, 1z) $F(x) = \frac{(\ln x)^2}{2} + K$, $F(x) = \frac{(\ln x)^6}{6} + K$, $F(x) = \ln(1+\ln x) + K$, $F(x) = \arctan(\ln x) + K$; 2a) $F(x) = \ln\frac{x+1}{x+2}$, 2b) $F(x) = x - \ln(x+1)$, 2c) $F(x) = \ln\frac{x}{x+1}$, 2d) $F(x) = x - 2\sqrt{7}\arctan\frac{2x-5}{\sqrt{7}}$, 2e) $F(x) = x - \arctan x$, 2f) $F(x) = x + \frac{3}{2}\ln(1-x) - \frac{1}{2}\ln(x+1)$, 2g) $F(x) = \frac{1}{4}\arctan\frac{x^2}{2}$, 2h) $F(x) = \frac{1}{2}\ln(x^4-1)$; 3a) $F(x) = e^x(x-1)$, $F(x) = e^x(x^2-2x+2)$, $F(x) = \frac{1}{27}e^{3x}(9x^2-6x+2)$, 3b) $F(x) = x\ln x - x$, $F(x) = x\arctan x - \frac{1}{2}\ln(x^2+1)$, 3c) $F(x) = \sin x - x\cos x$, 3d) $F(x) = \frac{1}{9}(3x\sin 3x + \cos 3x)$, 3e) $F(x) = e^x(-x-1)$, 3f) $\frac{1}{9}x^3(3\ln x - 1)$, 3g) $F(x) = \frac{1}{2}(x^2+1)\arctan x - \frac{1}{2}x$, 3h) $F(x) = \frac{1}{10}(5\cos x - \cos 5x)$, 3i) $F(x) = \frac{1}{8}(\sin 2x - 2x\cos 2x)$; 4a) $F(x) = x + 4\sqrt{x} + 4\ln(\sqrt{x}-1)$ (make $x = t^2$), 4b) $F(x) = -\frac{1}{3}\sqrt{2-x^2}(x^2+4)$ (make $x^2 = 2-t^2$), 4c) $F(x) = \frac{6x\sqrt[6]{x}}{7} - \frac{6\sqrt[6]{x^5}}{5} - \frac{3\sqrt[3]{x^2}}{2} + 2\sqrt{x} + 3\sqrt[3]{x} - 6\sqrt[6]{x} - 3\ln(\sqrt[3]{x}+1) + 6\arctan(\sqrt[6]{x})$ (make $x = t^6$), 4d) $F(x) = \frac{1}{2}(\ln(e^x+1) - \ln(1-e^x)) - e^x$ (make $e^x = t$), 4e) $F(x) = \frac{\ln(\sqrt{2}-\sin x) - \ln(\sin x + \sqrt{2})}{2\sqrt{2}}$ (make $\sin x = t$), 4f) $F(x) = 2x + \frac{x}{2}\sqrt{1-x^2} + \frac{1}{2}\arcsin x)$ (make $x = \sin t$), 4g) $\frac{2}{3}(e^x-2)\sqrt{e^x+1}$ (make $e^x = t^2-1$); 5a) $F(x) = e^{-x}(2x-3)$, 5b) $F(x) = \frac{2}{3}\sqrt{(x+1)^3} - 2\sqrt{x+1}$, 5c) $F(x) = 2e^{\sqrt{x}}(\sqrt{x}-1)$, 5d) $F(x) = -\frac{e^{-x^2}}{2}$, 5e) $F(x) = \frac{(x^2+1)^{21}}{42}$, 5f) $F(x) = x\sin x + \cos x$, 5g) $F(x) = -\frac{1}{9}x + \frac{1}{3}e^{-x} + \frac{1}{9}\ln(e^x-3)$, 5h) $F(x) = \frac{4}{3}(\sqrt{x}+1)^{3/2}$, 5i) $F(x) = \frac{1}{60}(10x^6 - 12x^5 - 20x^3 + 30x^2 - 60x - 147) + 2\ln(x+1)$, 5j) $F(x) = \arcsin(1+\sin x) + \frac{1}{2}\ln(\sqrt{1-(1+\sin x)^2}+1) - \frac{1}{2}\ln(\sqrt{1-(1+\sin x)^2}-1)$.

Chapter 13: Integration: 1a) $\frac{7}{3}$, 1b) $\frac{100}{3}$, 1c) $\frac{7}{4}$, 1d) $\frac{2+\pi}{8}$, 1e) $\ln 2$, 1f) $-\frac{2}{3}$; 2) $\int_1^3 \frac{x^2}{2}dx = \frac{13}{3}$; 3a) $10 - \frac{4\sqrt{2}}{9}$, 3b) $\frac{37}{12}$; 4a) $\frac{\pi}{2}$, 4b) 1, 4c) 1, 4d) $\frac{\pi^2}{8}$, 4f) -1; 5a) x^4, 5b) $2xe^{x^4}$; 6a) 1, 6b) $\frac{1}{3}$.

Bibliography

[APO] Apostol, T. M. (1969): *Calculus, Vol. 1: One-Variable Calculus, with an Introduction to Linear Algebra*, 2nd ed., John Wiley, USA.

[BLY] Blyth T. S. e E. F. Robertson (1998): *Basic Linear Algebra*, Springer-Verlag, UK.

[MON] Monteiro, A. M. (2001): *Álgebra Linear e Geometria Analítica*, Mcgraw-Hill, Portugal.

[SIM] Simmons, G.F. (1996): *Calculus With Analytic Geometry*, 2nd ed., McGraw-Hill, USA.

[STE] Stewart, J. (2006): *Calculus – Early Transcentals*, 6th ed., Thomson Learning, USA.