Computer Arithmetic

Master of Science in Electrical Engineering

Erivelton Geraldo Nepomuceno

Department of Electrical Engineering
Federal University of São João del-Rei

August 16, 2018

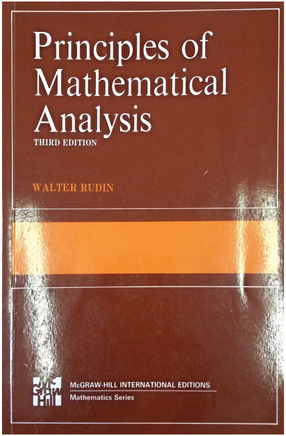
Teaching Plan

Content

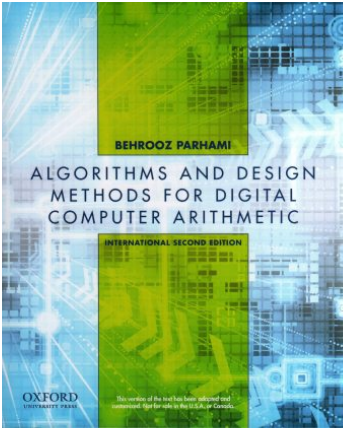
1. The Real and Complex Numbers Systems
2. Basic Topology
3. Numerical Sequences and Series
4. Continuity and Differentiation
5. Sequences and Series of Functions
6. Number Representation
7. IEEE 754-2008: Standard for Floating-Point Arithmetic
8. IEEE 1788-2008: Standard for Interval Arithmetic
9. Programmable Logic Devices (FPGA)
10. Arithmetic Operation in a Computer

References

- Rudin, W. (1976), *Principles of mathematical analysis*, McGraw-Hill New York.

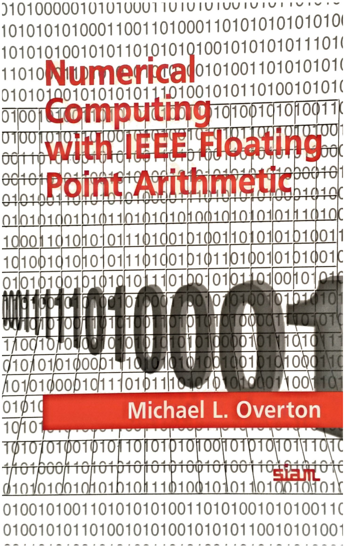


- Parhami, B. (2012). *Computer arithmetic algorithms and hardware architectures*. Oxford University Press, New York.

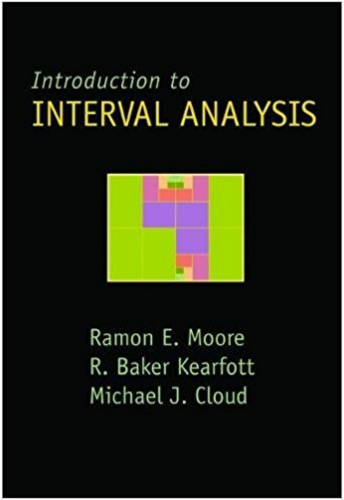


- Tocci, R. J., Widmer, N. S., & Moss, G. L. (2011). *Sistemas Digitais*. Princípios e Aplicações (11th ed.). São Paulo: Pearson Prentice Hall.

- Overton, M. L. (2001), *Numerical Computing with IEEE floating point arithmetic*, SIAM.



- Moore, R. E., Kearfott, R. B., & Cloud, M. J. (2009). *Introduction to Interval Analysis*. SIAM.



Assessment

- $N_1 = 80$ points : Conference paper.
- $N_2 = 20$ points : Activities
- $N = N_1 + N_2$ points
- **If** $N \geq 60$ **then** *Succeed*.
- **If** $N < 60$ **then** *Failed*.
- Second chance.

1. The real and complex number systems

1.1 Introduction

- A discussion of the main concepts of analysis (such as convergence, continuity, differentiation, and integration) must be based on an accurately defined number concept.
- Number: An arithmetical value expressed by a word, symbol,or figure, representing a particular quantity and used in counting and making calculations. (Oxford Dictionary).
- Let us see if we really know what a number is.
- Think about this question:[1]

$$\text{Is} \quad 0.999\ldots = 1? \tag{1}$$

[1] Richman, F. (1999) Is 0.999 ... = 1? *Mathematics Magazine*. 72(5), 386–400.

- The set $\mathbb{N}$ of natural numbers is defined by the Peano Axioms:
  1. There is an injective function $s : \mathbb{N} \to \mathbb{N}$. The image $s(n)$ of each natural number $n \in \mathbb{N}$ is called successor of n.
  2. There is an unique natural number $1 \in \mathbb{N}$ such that $1 \neq s(n)$ for all $n \in \mathbb{N}$.
  3. If a subset $X \subset \mathbb{N}$ is such that $1 \in X$ and $s(X) \subset X$ (that is, $n \in X \Rightarrow s(n) \in X$) then $X = \mathbb{N}$.
- The set $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2 \ldots\}$ of integers is a bijection $f : \mathbb{N} \to \mathbb{Z}$ such that $f(n) = (n - 1)/2$ when n is odd and $f(n) - n/2$ when n is even.
- The set $\mathbb{Q} = \{m/n; m, n \in \mathbb{Z}, n \neq 0\}$ of rational numbers may be written as $f : \mathbb{Z} \times \mathbb{Z}^* \to \mathbb{Q}$ such that $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ and $f(m, n) = m/n$.

- The rational numbers are inadequate for many purposes, both as a field and as an ordered set.
- For instance, there is no rational p such that $p^2 = 2$.
- An irrational number is written as infinite decimal expansion.
- The sequence $1, 1.4, 1.41, 1.414, 1.4142 \ldots$ tends to $\sqrt{2}$.
- What is it that this sequence *tends to*? What is an irrational number?
- This sort of question can be answered as soon as the so-called “real number system” is constructed.

Example 1

We now show that the equation

$$p^2 = 2 \tag{2}$$

is not satisfied by any rational p. If there were such a p, we could write $p = m/n$ where m and n are integers that are not both even. Let us assume this is done. Then (2) implies

$$m^2 = 2n^2. \tag{3}$$

This shows that m^2 is even. Hence m is even (if m were odd, m^2 would be odd), and so m^2 is divisible by 4. It follows that the right side of (3) is divisible by 4, so that n^2 is even, which implies that n is even. Thus the assumption that (2) holds thus leads to the conclusion that both m and n are even, contrary to our choice of m and n. Hence (2) is impossible for rational p.

- Let us examine more closely the Example 1.
- Let A be the set of all positive rationals p such that $p^2 < 2$ and let B consist of all positive rationals p such that $p^2 > 2$.
- We shall show that A contains no largest number and B contains no smallest.
- In other words, for every $p \in A$ we can find a rational $q \in A$ such that $p < q$, and for every $p \in B$ we can find a rational $q \in B$ such that $q < p$.
- Let each rational $p > 0$ be associated to the number

$$q = p - \frac{p^2 - 2}{p + 2} = \frac{2p + 2}{p + 2}. \tag{4}$$

and

$$q^2 = \frac{(2p + 2)^2}{(p + 2)^2}. \tag{5}$$

- Let us rewrite
$$q = p - \frac{p^2 - 2}{p+2} \qquad (6)$$
- Let us subtract 2 from both sides of (6)
$$\begin{aligned} q^2 - 2 &= \frac{(2p+2)^2}{(p+2)^2} - \frac{2(p+2)^2}{(p+2)^2} \\ q^2 - 2 &= \frac{(4p^2+8p+4) - (2p^2+8p+8)}{(p+2)^2} \\ q^2 - 2 &= \frac{2(p^2-2)}{(p+2)^2}. \end{aligned} \qquad (7)$$
- If $p \in A$ then $p^2 - 2 < 0$, (6) shows that $q > p$, and (7) shows that $q^2 < 2$. Thus $q \in A$.
- If $p \in B$ then $p^2 - 2 > 0$, (6) shows that $0 < q < p$, and (7) shows that $q^2 > 2$. Thus $q \in B$.

- In this slide we show two ways to approach $\sqrt{2}$.
- Newton's method

$$\sqrt{2} = \lim_{n\to\infty} x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n} \tag{8}$$

which produces the sequence for $x_0 = 1$

Table 1: Sequence of x_n of (8)

n	x_n (fraction)	x_n (decimal)
0	1	1
1	$\frac{3}{2}$	1.5
2	$\frac{17}{12}$	$1.41\bar{6}$
3	$\frac{577}{408}$	$1.4142\ldots$

- Now let us consider the continued fraction given by

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \ddots}}} \tag{9}$$

represented by $[1; 2, 2, 2, \ldots]$, which produces the following sequence

Table 2: Sequence of x_n of (9)

n	x_n (fraction)	x_n (decimal)
0	1	1
1	$3/2$	1.5
2	$7/5$	1.4
3	$17/12$	$1.41\bar{6}\ldots$

Remark 1

The rational number system has certain gaps, in spite the fact that between any two rational there is another: if $r < s$ then $r < (r+s)/2 < s$. The real number system fill these gaps.

Definition 1

If A is any set, we write $x \in A$ to indicate that x is a member of A. If x is not a member of A, we write: $x \notin A$.

Definition 2

The set which contains no element will be called the empty set. If a set has at least one element, it is called nonempty.

Definition 3

If every element of A is an element of B, we say that A is a subset of B. and write $A \subset B$, or $B \supset A$. If, in addition, there is an element of B which is not in A, then A is said to be a proper subset of B.

1.2 Ordered Sets

Definition 4

Let S be a set. An order on S is a relation, denote by $<$, with the following two properties:

1. If $x \in S$ and $y \in S$ then one and only one of the statements

$$x < y, \quad x = y, \quad y < x$$

   is true.
2. If $x, y, z \in S$, if $x < y$ and $y < z$, then $x < z$.

- The notation $x \leq y$ indicates that $x < y$ or $x = y$, without specifying which of these two is to hold.

Definition 5

An ordered set is a set S in which an order is defined.

Definition 6

Suppose S is an ordered set, and $E \subset S$. If there exists a $\beta \in S$ such that $x \leq \beta$ for every $x \in E$, we say that E is bounded above, and call β an upper bound of E. Lower bound are defined in the same way (with $\geq$ in place of $\leq$).

Definition 7

Suppose S is an ordered set, $E \subset S$, and E is bounded above. Suppose there exists an $\alpha \in S$ with the following properties:

1. α is an upper bound of E.
2. If $\gamma < \alpha$ then γ is not an upper bound of E.

Then α is called the least upper bound of E or the supremum of E, and we write

$$\alpha = \sup E.$$

Definition 8

The greatest lower bound, or infimum, of a set E which is bounded below is defined in the same manner of Definition 7: The statement

$$\alpha = \inf E.$$

means that α is a lower bound of E and that no β with $\beta > \alpha$ is a lower bound of E.

Example 2

If $\alpha = \sup E$ exists, then α may or may not be a member of E. For instance, let E_1 be the set of all $r \in Q$ with $r < 0$. Let E_2 be the set of of all $r \in Q$ with $r \leq 0$. Then

$$\sup E_1 = \sup E_2 = 0,$$

and $0 \notin E_1$, $0\ in\ E_2$.

Definition 9

An ordered set S is said to have the least-upper-bound property if the following is true: If $E \subset S$, E is not empty, and E is bounded above, then $\sup E$ exists in$S$.

Theorem 1

Suppose S is an ordered set with the least-upper-bound property, $B \subset S$, B is not empty, and B is bounded below. Let L be the set of all lower bounds of B. Then

$$\alpha = \sup L$$

exists in S and $\alpha = \inf B$.

1.3 Fields

Definition 10

A field is a set F with two operations, called addition and multiplication, which satisfy the following so-called "field axioms" (A), (M) and (D):

(A) Axioms for addition

- (A1) If $x \in F$ and $y \in F$, then their sum $x + y$ is in F.
- (A2) Addition is commutative: $x + y = y + x$ for all $x, y \in F$.
- (A3) Addition is associative: $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$.
- (A4) F contains an element 0 such that $0 + x = x$ for every $x \in F$.
- (A5) To every $x \in F$ corresponds an element $-x \in F$ such that $x + (-x) = 0$.

(M) Axioms for multiplication

- (M1) If $x \in F$ and $y \in F$, then their product xy is in F.
- (M2) Multiplication is commutative: $xy = yx$ for all $x, y \in F$.

(M3) Multiplicative is associative: $(xy)z = x(yz)$ for all $x, y, z \in F$.

(M4) F contains an element $1 \neq 0$ such that $1x = x$ for every $x \in F$.

(M5) If $x \in F$ and $x \neq 0$ then there exists an element $1/x \in F$ such that

$$x \cdot (1/x) = 1.$$

(D) The distributive law

$$x(y+z) = xy + xz$$

holds for all $x, y, z \in F$.

Definition 11

An ordered field is a field F which is also an ordered set, such that

1. $x + y < x + z$ if $x, y, z \in F$ and $y < z$.
2. $xy > 0$ if $x \in F$, $y \in F$, $x > 0$, and $y > 0$.

1.4 The real field

Theorem 2

There exists an ordered field R which has the least-upper-bound property. Moreover, R contains Q as a subfield.

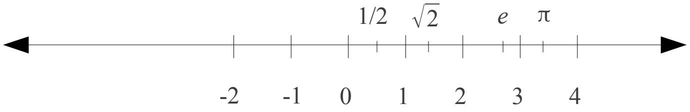


Figure 1: Real Line

Theorem 3

(a) *If $x \in R$, and $x > 0$, then there is a positive integer n such that $nx > y$.*

(b) *If $x \in R$, and $x < y$, then there exists a $p \in Q$ such that $x < p < y$.*

Theorem 4

For every real $x > 0$ and every integer $n > 0$ there is one and only one real y such that $y^n = x$.

Proof of Theorem 4:

- That there is at most one such y is clear, since $0 < y_1 < y_2$, implies $y_1^n < y_2^n$.
- Let E be the set consisting of all positive real numbers t such that $t^n < x$.
- If $t = x/(1+x)$ then $0 < t < 1$. Hence $t^n < t < x$. Thus $t \in E$, and E is not empty. Thus $1 + x$ is an upper bound of E.
- If $t > 1 + x$ then $t^n > t > x$, so that $t \notin E$. Thus $1 + x$ is an upper bound of E and there is $y = \sup E$.
- To prove that $y^n = x$ we will show that each of the inequalities $y^n < x$ and $y^n > x$ leads to contradiction.

- The identity $b^n - a^n = (b-a)(b^{n-1} + b^{n-2}a + \cdots a^{n-1})$ yields the inequality
$$b^n - a^n < (b-a)nb^{n-1}$$
when $0 < a < b$.
- Assume $y^n < x$. Choose h so that $0 < h < 1$ and
$$h < \frac{x - y^n}{n(y+1)^{n-1}}.$$
- Put $a = y$, $b = y + h$. Then
$$(y+h)^n - y^n < hn(y+h)^{n-1} < hn(y+1)^{n-1} < x - y^n.$$
- Thus $(y+h)^n < x$, and $y + h \in E$. Since $y + h > y$, this contradicts the fact that y is an upper bound of E.
- Assume $y^n > x$. Put
$$k = \frac{y^n - x}{ny^{n-1}}.$$
Then $0 < k < y$. If $t \geq y - k$, we conclude that
$$y^n - t^n \geq y^n - (y-k)^n < kny^{n-1} = y^n - x.$$

- Thus $t^n > x$, and $t \notin E$. It follows that $y - k$ is an upper bound of E. But $y - k < y$, which contradicts the fact that y is the least upper bound of E.
- Hence $y^n = x$, and the proof is complete.

Definition 12

Let $x > 0$ be real. Let n_o be the largest integer such that $n_0 \leq x$. Having chosen $n_0, n_1, \ldots, n_{k-1}$, let n_k be the largest integer such that

$$n_0 + \frac{n_1}{10} + \cdots + \frac{n_k}{10^k} \leq x.$$

Let E be the set of these numbers

$$n_0 + \frac{n_1}{10} + \cdots + \frac{n_k}{10^k} \quad (k = 0, 1, 2, \ldots). \tag{10}$$

Then $x = \sup E$. The decimal expansion of x is

$$n_0 \cdot n_1 n_2 n_3 \cdots . \tag{11}$$

1.5 The extended real number system

Definition 13

The extended real number system consists of the real field R and two symbols: $+\infty$ and $-\infty$. We preserve the original order in R, and define

$$-\infty < x < +\infty$$

for every $x \in R$. A symbol for the extended real number system is $\bar{R}$.

- $+\infty$ is an upper bound of every subset of the extended real number system, and that every nonempty subset has a least upper bound.
- The same remarks apply to lower bounds.
- The extended real number system does not form a field.
- It is customary to make the following conventions:
  - (a) If x is real then
    $x + \infty = \infty, \quad x - \infty = -\infty, \quad \frac{x}{+\infty} = \frac{x}{-\infty} = 0.$
  - (b) If $x > 0$ then $x \cdot (+\infty) = +\infty, x \cdot (-\infty) = -\infty.$

1.6 The complex field

(c) If $x < 0$ then $x \cdot (+\infty) = -\infty, x \cdot (-\infty) = +\infty$.

Definition 14

A complex number is an ordered pair (a, b) of real numbers. Let $x = (a, b), y = (c, d)$ be two complex numbers. We define

$$x + y = (a + c, b + d),$$

$$xy = (ac - bd, ad + bc).$$

- $i = (0, 1)$.
- $i^2 = -1$.
- If a and b are real, then $(a, b) = a + bi$.

1.7 Euclidean Space

Definition 15

For each positive integer k, let R^k be the set of all ordered k-tuples

$$\mathbf{x} = (x_1, x_2, \ldots, x_k),$$

where $x_1, \ldots, x_k$ are real numbers called the coordinates of $\mathbf{x}$.

- Addition of vectors: $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_k + y_k)$.
- Multiplication of a vector by a real number (scalar): $\alpha\mathbf{x} = (\alpha x_1, \ldots, \alpha x_k)$.
- Inner product: $x \cdot y = \sum_{i=1}^{k} x_i y_i$.
- Norm: $|x| = (x \cdot x)^{1/2} = \left(\sum_1^k x_i^2\right)^{1/2}$.
- The structure now defined (the vector space R^k with the above product and norm) is called Euclidean k-space.

Theorem 5

Suppose $\mathbf{x}, \mathbf{y}, \mathbf{z} \in R^k$ *and* α *is real. Then*

1. $|\mathbf{x}| \geq 0$;
2. $|\mathbf{x}| = 0$ *if and only if* $|\mathbf{x} = 0|$;
3. $|\alpha \mathbf{x}| = |\alpha||\mathbf{x}|$;
4. $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|;$
5. $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$;
6. $|\mathbf{x} - \mathbf{z}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{x} - \mathbf{z}|$.

- Items 1,2 and 6 of Theorem 5 will allow us to regard R^k as a metric space.

Exercises Chapter 1

(1) Let the sequence of numbers $1/n$ where $n \in \mathbb{N}$. Does this sequence have an infimum? If it has, what is it? Explain your result and show if it is necessary any other condition.

(2) Comment the assumption: Every irrational number is the limit of monotonic increasing sequence of rational numbers (Ferrar, 1938, p.20).

(3) Prove Theorem 1.

(4) Prove the following statements

a) If $x + y = x + z$ then $y = z$.
b) If $x + y = x$ then $y = 0$.
c) If $x + y = 0$ then $y = -x$.
d) $-(-x) = x$.

(5) Prove the following statements

a) If $x > 0$ then $-x < 0$, and vice versa.
b) If $x > 0$ and $y < z$ then $xy < xz$.
c) If $x < 0$ and $y < z$ then $xy > xz$.
d) If $x \neq 0$ then $x^2 > 0$.
e) If $0 < x < y$ then $0 < 1/y < 1/x$.

(6) Prove the Theorem 2. (Optional)

(7) Prove the Theorem 3.

(8) Write addition, multiplication and distribution law in the same manner of Definition 22 for the complex field.

(9) What is the difference between R and $\bar{R}$?

(10) Prove the reverse triangle inequality: $||a| - |b|| \leq |a - b|$.

2. Basic Topology

2.1 Finite, Countable, and Uncountable Sets

Definition 16

Consider two sets A and B , whose elements may be any objects whatsoever, and suppose that with each element x of A there is associated, in some manner, an element of B, which we denote by $f(x)$. Then f is said to be a function from A to B (or a mapping of A into B). The set A is called the domain of f (we also say f is defined on A), and the elements of $f(x)$ are called the values of f. The set of all values of f is called the range of f.

Definition 17

Let A and B be two sets and let f be a mapping of A into B. If $E \subset A$, $f(E)$ is defined to be the set of all elements $f(x)$, for $x \in E$. We call $f(E)$ the image of E under f. In this notation, $f(A)$ is the range of f. It is clear that $f(A) \subset B$. If $f(A) = B$, we say that f maps A onto B.

Definition 18

If $E \subset B$, f^{-1} denotes the set of all $x \in A$ such that $f(x) \in E$. We call $f^{-1}(E)$ the inverse image of E under f.

- f is a 1-1 mapping of A into B provided that $f(x_1) \neq f(x_2)$ whenever $x_1 \neq x_2, x_1 \in A, x_2 \in A$.

Definition 19

If there exists a 1-1 mapping of A onto B, we say that A and B, can be put in 1-1 correspondence, or that A and B have the same cardinal number, or A and B are equivalent, and we write $A \sim B$.

- Properties of equivalence
  - It is reflexive: $A \sim A$.
  - It is symmetric: If $A \sim B$, then $B \sim A$.
  - It is transitive: If $A \sim B$ and $B \sim C$, then $A \sim C$.

Definition 20

Let $n \in N$ and J_n be the set whose elements are the integers $1, 2, \ldots, n$; let J be the set consisting of all positive integers. For any set A, we say:

(a) A is finite if $A \sim J_n$ for some n.

(b) A is infinite if A is not finite.

(c) A is countable if $A \sim J$.

(d) A is uncountable if A is neither finite nor countable.

(e) A is at most countable if A is finite or countable.

Remark 2

A is infinite if A is equivalent to one of its proper subsets.

Definition 21

By a sequence, we mean a function f defined on the set J of all positive integers. If $f(n) = x_n$, for $n \in J$, it is customary to denote the sequence f by the symbol $\{x_n\}$, or sometimes $x_1, x_2, x_3, \ldots$. The values of f are called terms of the sequence. If A is a set and if $x_n \in A$ for all $n \in J$, then $\{x_n\}$ is said to be a sequence in A, or a sequence of elements of A.

- Every infinite subset of a countable set A is countable.
- Countable sets represent the “smallest infinity.

Definition 22

Let A and Ω be sets, and suppose that with each element of α of A is associated a subset of Ω which denote by E_α. A collection of sets is denoted by $\{E_\alpha\}$.

Definition 23

The union of the sets E_α is defined to be the set S such that $x \in S$ if and only if $x \in E_\alpha$ for at least one $\alpha \in A$. It is denoted by

$$S = \bigcup_{\alpha \in A} E_\alpha. \tag{12}$$

- If A consists of the integers $1, 2, \ldots, n$, one usually writes

$$S = \bigcup_{m=1}^{n} E_m = E_1 \cup E_2 \cup \cdots \cup E_n. \tag{13}$$

- If A is the set of all positive integers, the usual notations is

$$S = \bigcup_{m=1}^{\infty} E_m. \tag{14}$$

- The symbol ∞ indicates that the union of a countable collection of sets is taken. It should not be confused with symbols $+\infty$ and $-\infty$ introduced in Definition 13.

Definition 24

The intersection of the sets E_α is defined to be the set P such that $x \in P$ if and only if $x \in E_\alpha$ for every $\alpha \in A$. It is denoted by

$$P = \bigcap_{\alpha \in A} E_\alpha. \tag{15}$$

- P is also written such as

$$P = \bigcap_{m=1}^{n} = E_1 \cap E_2 \cap \cdots E_n. \tag{16}$$

- If A is the set of all positive integers, we have

$$P = \bigcap_{m=1}^{\infty} E_m. \tag{17}$$

Theorem 6

Let $\{E_n\}, n = 1, 2, 3, \ldots,$ *be a sequence of countable sets, and put*

$$S = \bigcup_{n=1}^{\infty} E_n. \tag{18}$$

Then S is countable.

- The set of all rational numbers is countable.
- The set of all real numbers is uncountable.

2.2 Metric Spaces

Definition 25

A set X, whose elements we shall call points, is said to be a metric space if with any two points p and q of X there is associated a real number $d(p, q)$ the distance from p to q, such that

(a) $d(p, q) > 0$ if $p \neq q$; $d(p, p) = 0$.

(b) $d(p, q) = d(q, p)$;

(c) $d(p, q) \leq d(p, r) + d(r, q)$, for any $r \in X$.

Definition 26

By the segment (a, b) we mean the set of all real numbers x such that $a < x < b$.

Definition 27

By the interval $[a, b]$ we mean the set of all real number x such that $a \leq x \leq b$.

Definition 28

If $\mathbf{x} \in R^k$ and $r > 0$, the open (or closed) ball B with center at $\mathbf{x}$ and radius r is defined to be the set of all $y \in R^k$ such that $|\mathbf{y} - \mathbf{x}| < r$ (or $|\mathbf{y} - \mathbf{x}| \leq r$).

Definition 29

We call a set $E \subset R^k$ convex if $(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \in E$ whenever $\mathbf{x} \in E$, $\mathbf{y} \in E$ and $0 < \lambda < 1$.

Example 3

Balls are convex. For if $|\mathbf{y} - \mathbf{x}| < r$, $|\mathbf{z} - \mathbf{x}| < r$, and $0 < \lambda < 1$, we have

$$\begin{aligned} |\lambda \mathbf{y} + (1 - \lambda)\mathbf{z} - \mathbf{x}| &= |\lambda(\mathbf{y} - \mathbf{x}) + (1 - \lambda)(\mathbf{z} - \mathbf{x})| \\ &\leq \lambda|\mathbf{y} - \mathbf{x}| + (1 - \lambda)|\mathbf{z} - \mathbf{x}| < \lambda r + (1 - \lambda) r \\ &= r. \end{aligned}$$

Definition 30

Let X be a metric space. All points and sets are elements and subsets of X.

(a) A neighbourhood of a point p is a set $N_r(p)$ consisting of all points q such that $d(p, q) < r$.

(b) A point p is a limit point of the set E if every neighbourhood of p contains a point $q \neq p$ such that $q \in E$.

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

(d) E is closed is very limit point of E is a point of E.

(e) A point p is an interior point of E if there is a neighbourhood N of p such that $N \subset E$.

(f) E is open is every point of E is an interior point of E.

(g) The complement of E (denoted by E^c) is the set of all points $p \in X$ such that $p \notin E$.

Definition 30

(h) E is perfect if E is closed and if every point of E is a limit point of E.

(i) E is bounded if there is a real number M and a point $q \in X$ such that $d(p, q) < M$ for all $p \in E$.

(j) E is dense in X if every point of X is a limit point of E, or a point of E (or both).

- If p is a limit point of a set E, then every neighbourhood of p contains infinitely many points of E.
- A set E is open if and only if its complement is closed.

Definition 31

If X is a metric space, if $E \subset X$, and if E' denotes the set of all limit points of E in X, then the closure of E is the set $\bar{E} = E \cup E'$.

Theorem 7

If X is a metric space and $E \subset X$, then

(a) $\bar{E}$ *is closed.*

(b) $E = \bar{E}$ *if and only if* E *is closed.*

(c) $\bar{E} \subset F$ *for every closed set* $F \subset X$ *such that* $E \subset F$.

Theorem 8

Let E be a nonempty set of real numbers which is bounded above. Let $y = \sup E$. Then $y \in \bar{E}$. Hence $y \in E$ if E is closed.

2.3 Compact Sets

Definition 32

By an open cover of a set E in a metric space X we mean a collection $\{G_\alpha\}$ of open subsets of X such that $E \subset \bigcup_\alpha G_\alpha$.

Definition 33

A subset K of a metric space X is said to be compact if every open cover of K contains a finite subcover.

Definition 34

A set $X \subset R$ is compact if X is closed and bounded[a].

[a] Lima, E. L. (2006) *Análise Real v. 1.*. RJ: IMPA, 2006.

Definition 35

If $\{K_n\}$ is a sequence of nonempty compact sets such that $K_n \supset K_{n+1} \quad (n = 1, 2, 3 \ldots)$,, then $\bigcap_1^\infty K_n$ is not empty.

Definition 36

If $\{I_n\}$ is a sequence of intervals in R^1, such that $I_n \supset I_{n+1} \quad (n = 1, 2, 3 \ldots)$,, then $\bigcap_1^\infty I_n$ is not empty.

Theorem 9

If a set E in R^k has one of the following three properties, then it has the other two:

1. *E is closed and bounded.*
2. *E is compact.*
3. *Every infinite subset of E has a limit point in E.*

Theorem 10

(Weierstrass) *Every bounded subset of R^k has a limit point in R^k.*

2.4 Perfect Sets

Theorem 11

Let P be a nonempty perfect set in R^k. Then P is uncountable.

- Every interval $[a, b] (a < b)$ is uncountable. In particular, the set of all real numbers in uncountable.
- The Cantor ternary set is created by repeatedly deleting the open middle thirds of a set of line segments. One starts by deleting the open middle third $(1/3, 2/3)$ from the interval $[0, 1]$, leaving two line segments: $[0, 1/3] \cup [2/3, 1]$. Next, the open middle third of each of these remaining segments is deleted, leaving four line segments: $[0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$. This process is continued ad infinitum, where the nth set is

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right) . C_0 = [0, 1].$$

- The first six steps of this process are illustrated in Figure 50.

Figure 2: Cantor Set. Source: Wikipedia.

2.5 Connected Sets

Definition 37

Two subsets A and B of a metric space X are said to be separated if both $A \cap \bar{B}$ and $\bar{A} \cap B$ are empty, i.e., if no point of A lies in the closure of B and no point of B lies in the closure of A.
A set $E \subset X$ is said to be connected if E is not a union of two nonempty separated sets.

Theorem 12

A subset E of the real line R^1 is connected if and only if it has the following property: If $x \in E$, $y \in E$, and $x < z < y$, then $z \in E$.

Exercises Chapter 2

(1) Let A be the set of real numbers x such that $0 < x \leq 1$. For every $x \in A$, be the set of real numbers y, such that $0 < y < x$. Complete the following statements

   (a) $E_x \subset E_z$ if and only if $0 < x \leq z \leq 1$.

   (b) $\bigcup_{x\in A} E_x = E_1$.

   (c) $\bigcap_{x\in A} E_x$ is empty.

(2) Prove Theorem 6. Hint: put the elements of E_n in a matrix and count the diagonals.

(3) Prove that the set of all real numbers is uncountable.

(4) The most important examples of metric spaces are euclidean spaces R^k. Show that a Euclidean space is a metric space.

(5) For $x \in R^1$ and $y \in R^1$, define

$$\begin{aligned} d_1(x,y) &= (x-y)^2, \\ d_2(x,y) &= \sqrt{|x-y|}, \\ d_3(x,y) &= |x^2-y^2|, \\ d_4(x,y) &= |x-2y|, \\ d_5(x,y) &= \frac{|x-y|}{1+|x-y|}. \end{aligned}$$

Determine for each of these, whether it is a metric or not.

Work 1

To find the square root of a positive number a, we start with some approximation, $x_0 > 0$ and then recursively define:

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right). \tag{19}$$

Compute the square root using (19) for

(a) $a = 2$;

(b) $a = 2 \times 10^{-300}$

(c) $a = 2 \times 10^{-310}$

(d) $a = 2 \times 10^{-322}$

(e) $a = 2 \times 10^{-324}$

Check your results by $x_n \times x_n$, after defining a suitable stop criteria for n.

3. Numerical Sequences and Series

3.1 Convergent Sequences

Definition 38

A sequence $\{p_n\}$ in a metric space X is said to converge if there is point $p \in X$ with the following property: For every $\varepsilon > 0$ there is an integer N such that $n \geq N$ implies that $d(p_n, p) < \varepsilon$. In this case we also say that p_n converges to p, or that p is the limit of $\{p_n\}$, and we write $p_n \to p$, or

$$\lim_{n\to\infty} p_n = p.$$

- If $\{p_n\}$ does not converge, it is said to diverge.
- It might be well to point out that our definition of convergent sequence depends not only on $\{p_n\}$ but also on X.
- It is more precise to say convergent in X.
- The set of all points p_n $(n = 1, 2, 3, \dots)$ is the range of $\{p_n\}$.
- The sequence $\{p_n\}$ is said to be bounded if its range is bounded.

Example 4

Let $s \in R$. If $s_n = 1/n$, then

$$\lim_{n \to \infty} s_n = 0.$$

The range is infinite, and the sequence is bounded.

Example 5

Let $s \in R$. If $s_n = n^2$, the sequence $\{s_n\}$ is unbounded, is divergent, and has infinite range.

Example 6

Let $s \in R$. If $s_n = 1$ $(n = 1, 2, 3, \ldots)$, then the sequence $\{s_n\}$ converges to 1, is bounded, and has finite range.

Theorem 13

Let $\{p_n\}$ be a sequence in a metric space X.

(a) *$\{p_n\}$ converges to $p \in X$ if and only if every neighbourhood of p contains all but finitely many of the terms of $\{p_n\}$.*

(b) *If $p \in X$, $p' \in X$, and if $\{p_n\}$ converges to p and to p', then $p' = p$.*

(c) *If $\{p_n\}$ converges, then $\{p_n\}$ is bounded.*

(d) *If $E \subset X$ and if p is a limit point of E, then there is a sequence $\{p_n\}$ in E such that $p = \lim_{n\to\infty} p_n$.*

Theorem 14

Suppose $\{s_n\}$, $\{t_n\}$ *are complex sequences, and* $lim_{n\to\infty} s_n = s$ *and* $lim_{n\to\infty} t_n = t$. *Then*

(a) $\lim_{n\to\infty}(s_n + t_n) = s + t$;

(b) $\lim_{n\to\infty} cs_n = cs$, $\lim_{n\to\infty}(c + s_n) = c + s$, *for any number* c;

(c) $\lim_{n\to\infty}(s_n t_n) = st$;

(d) $\lim_{n\to\infty} \frac{1}{s_n} = \frac{1}{s}$;

3.2 Subsequences

Definition 39

Given a sequence $\{p_n\}$, consider a sequence $\{n_k\}$ of positive integers, such that $n_1 < n_2 < n_3 < \cdots$. Then the sequence $\{p_{n_i}\}$ is called a subsequence of $\{p_n\}$. If $\{p_{n_i}\}$, its limit is called a subsequential limit of $\{p_n\}$. It is clear that $\{p_n\}$ converges to p if and only if every subsequence of $\{p_n\}$ converges to p.

Theorem 15

(a) *If* $\{p_n\}$ *is a sequence in a compact metric space* X, *then some subsequence of* $\{p_n\}$ *converges to a point of* X.

(b) *Every bounded sequence in* R^k *contains a convergent subsequence.*

Theorem 16

The subsequential limits of a sequence $\{p_n\}$ *in a metric space* X *form a closed subset of* X.

3.3 Cauchy Sequence

Definition 40

A sequence $\{p_n\}$ is a metric space X is said to be a Cauchy sequence if for every $\varepsilon > 0$ there is an integer N such that $d(p_n, p_m) < \varepsilon$ if $n \geq N$ and $m \geq N$.

Figure 3: Augustin-Louis Cauchy (1789-1857), French mathematician who was an early pioneer of analysis. Source: Wikipedia.

Definition 41

Let E be a subset of a metric space X, and let S be the set of all real number of the form $d(p,q)$, with $p \in E$ and $q \in E$. The sup of S is called the diameter of E.

- If $\{p_n\}$ is a sequence in X and if E_N consists of the points $p_N, p_{N+1}, p_{N+2}, \ldots,$ it is clear from the two preceding definitions that $\{p_n\}$ is a Cauchy sequence if and only if

$$\lim_{N\to\infty} \operatorname{diam} E_N = 0.$$

Theorem 17

(a) *If $\bar{E}$ is the closure of a set E in a metric space X, then*

$$\operatorname{diam} \bar{E} = \operatorname{diam} E.$$

(b) *If K_a is a sequence of compact sets in X such that $K_n \supset K_{n+1}$ $(n = 1, 2, 3, \ldots)$ and if*

Theorem 18

(a) *In any metric space X, every convergent sequence is a Cauchy sequence.*

(b) *If X is a compact metric space and if $\{p_n\}$ is a Cauchy sequence in X, then $\{p_n\}$ converges to some point X.*

(c) *In R^k, every Cauchy sequence converges.*

- A sequence converges in R^k if and only if it is a Cauchy sequence is usually called the Cauchy criterion for convergence.

Definition 42

A sequence $\{s_n\}$ of real numbers is said to be

(a) monotonically increasing if $s_n \leq s_{n+1}$ $(n = 1, 2, 3, \ldots)$;

(b) monotonically decreasing if $s_n \geq s_{n+1}$ $(n = 1, 2, 3, \ldots)$;

3.4 Upper and Lower Limits

Theorem 19

Suppose $\{s_n\}$ is monotonic. Then $\{s_n\}$ converges if and only if it is bounded.

Definition 43

Let $\{s_n\}$ be a sequence of real numbers with the following property: For every real M there is an integer N such that $n \geq N$ implies $s_n \geq M$. We then write $s_n \to +\infty$.

Definition 44

Let $\{s_n\}$ be a sequence of real numbers. Let E be the set of numbers $x \in \bar{R}$ such that $s_{n_k} \to x$ for some subsequence $\{s_{n_k}\}$. This set E contains all subsequential limits plus possibly the numbers $+\infty$ and $-\infty$. Let $s^* = \sup E$, and $s_* = \inf E$. These numbers are called upper and lower limits of $\{s_n\}$.

- We can also write Definition 44 as

$$\lim_{n\to\infty} \sup s_n = s^*, \quad \lim_{n\to\infty} \inf s_n = s_*.$$

3.5 Some Special Sequences

- If $0 \leq x_n \leq s_n$ for $n \geq N$, where N is some fixed number, and if $s_n \to 0$, then $x_n \to 0$. This property help us to compute the following the limit of the following sequences:

(a) If $p > 0$, then $\lim_{n\to\infty} \frac{1}{n^p} = 0$.

(b) If $p > 0$, then $\lim_{n\to\infty} \sqrt[n]{p} = 1$.

(c) $\lim_{n\to\infty} \sqrt[n]{n} = 1$.

(d) If $p > 0$ and α is real, then $\lim_{n\to\infty} \frac{n^\alpha}{(1+p)^n} = 0$.

(e) If $|x| < 1$, then $\lim_{n\to\infty} x^n = 0$.

3.6 Series

Definition 45

Given a sequence $\{a_n\}$, we use the notation

$$\sum_{n=p}^{q} a_n \quad (p \leq q)$$

to denote the sum $a_p + a_{p+1} + \cdots + a_q$. With $\{a_n\}$ we associate a sequence $\{s_n\}$, where $s_n = \sum_{k=1}^{n} a_k$. For $\{s_n\}$ we also use the symbolic expression $a_1 + a_2 + a_3 + \cdots$ or, more concisely,

$$\sum_{n=1}^{\infty} a_n. \tag{20}$$

The symbol (33) 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 s, we say that the series converges, and we write

$$\sum_{n=1}^{\infty} a_n = s. \tag{21}$$

- s is the limit of a sequence of sums, and is not obtained simply by addition.
- If $\{s_n\}$ diverges, the series is said to diverge.
- Every theorem about sequences can be stated in terms of series (putting $a_1 = s_1$, and $a_n = s_n - s_{n-1}$ for $n > 1$), and vice versa.

- The Cauchy criterion can be restated as the following Theorem.

Theorem 20

$\sum a_n$ converges if and only if for every $\varepsilon > 0$ there is an integer N such that

$$\left| \sum_{k=n}^{m} a_n \right| \leq \varepsilon \tag{22}$$

if $m \geq n \geq N$.

Theorem 21

If $\sum a_n$ converges, then $\lim_{n \to \infty} a_n = 0$.

Theorem 22

A series of nonnegative terms converges if and only if its partial sums form a bounded sequence.

- Comparison test
  (a) If $|a_n| \leq c_n$ for $n \geq N_0$, where N_0 is some fixed integer, and if $\sum c_n$ converges, then $\sum a_n$ converges.
  (b) If $a_n \geq d_n \geq 0$ for $n \geq N_0$, and if $\sum d_n$ diverges, then $\sum a_n$ diverges.
- Geometric series
  - If $0 \leq x < 1$, then
$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$
    If $x \geq 1$, the series diverges.
  - **Proof** If $x \neq 1$, we have
$$s_n = \sum_{k=0}^{n} x^k = 1 + x + x^2 + x^3 \cdots + x^n. \tag{23}$$
    If we multiply (23) by x we have
$$xs_n = x + x^2 + x^4 \cdots x^{n+1}. \tag{24}$$
    Applying (23)$-$(24) we have

$$\begin{aligned} s_n - xs_n &= 1 - x^{n+1} \\ s_n(1 - x) &= 1 - x^{n+1} \\ s_n &= \frac{1 - x^{n+1}}{1 - x}. \end{aligned}$$

The result follows if we let $n \to \infty$.

3.7 The Root and Ratio Tests

Theorem 23

(Root Test) *Given* $\sum a_n$, *put* $\alpha = \lim_{n\to\infty} \sup \sqrt[n]{|a_n|}$. *Then*

(a) *If* $\alpha < 1$, $\sum a_n$ *converges;*

(b) *If* $\alpha > 1$, $\sum a_n$ *diverges;*

(c) *If* $\alpha = 1$, *the test gives no information.*

Theorem 24

(Ratio Test) *The series* $\sum a_n$

(a) *converges if* $\lim\limits_{n\to\infty} \sup \left|\dfrac{a_{n+1}}{a_n}\right| < 1$,

(b) *diverges if* $\left|\frac{a_{n+1}}{a_n}\right| \geq 1$ *for* $n \geq n_0$, *where* n_0 *is some fixed integer.*

- The ratio test is frequently easier to apply than the root test. However, the root test has wider scope.

Exercises Chapter 3

(1) Let $s \in R$. and $s_n = 1 + [(-1)^n/n]$. $\{s_n\}$ is bounded and its range is finite? Which value $\{s_n\}$ converges to?

(2) Write a Definition for $-\infty$ equivalent to Definition 43.

(3) Apply the root and ratio tests in the following series

(a) $\frac{1}{2} + \frac{1}{3} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{2^4} + \frac{1}{3^4} + \cdots$,

(b) $\frac{1}{2} + 1 + \frac{1}{8} + \frac{1}{4} + \frac{1}{32} + \frac{1}{16} + \frac{1}{128} + \frac{1}{64} + \cdots$,

4. Continuity and Differentiation

4.1 Limits of Functions

Definition 46

Let X and Y be metric spaces: suppose $E \subset X$, f maps E into Y, and p is a limit point of E. We write $f(x) \to q$ as $x \to p$, or

$$\lim_{x \to p} f(x) = q \tag{25}$$

if there is a point $q \in Y$ with the following property: For every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$d_Y(f(x), q) < \varepsilon \tag{26}$$

for all points $x \in E$ for which

$$0 < d_X(x, p) < \delta. \tag{27}$$

- Alternative statement for Definition 46 based on (ε, δ) limit definition given by Bernard Bolzano in 1817. Its modern version is due to Karl Weierstrass [2]

Definition 47

The function *f* approaches the limit *L* near *c* means: for every ε there is some $\delta > 0$ such that, for all x, if $0 < |x - c| < \delta$, then $|f(x) - L| < \varepsilon$.

- *f* approaches *L* near *c* has the same meaning as the Equation (28)

$$\lim_{x \to c} f(x) = L. \tag{28}$$

[2] Addapted from Spivak, M. (1967) *Calculus*. Benjamin: New York.

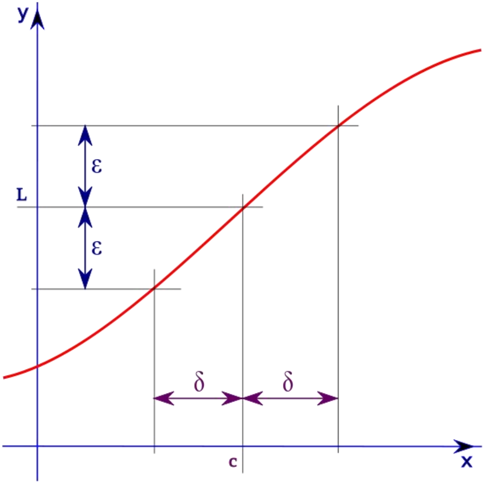


Figure 4: Whenever a point x is within δ of c, $f(x)$ is within ε units of L. Source: Wikipedia.

Theorem 25

Let X, Y, E, f, and p be as in Definition 46. Then

$$\lim_{x \to p} f(x) = q \tag{29}$$

if and only if

$$\lim_{n \to \infty} f(p_n) = q \tag{30}$$

for every sequence $\{p_n\}$ in E such that

$$p_n \neq p, \quad \lim_{n \to \infty} p_n = p. \tag{31}$$

Theorem 26

Suppose $E \subset X$, a metric space, p is a limit point of E, f and g are complex functions on E, and

$$\lim_{x \to p} f(x) = A, \quad \lim_{x \to p} g(x) = B.$$

Then

(a) $\lim_{x \to p} (f + g)(x) = A + B$;

(b) $\lim_{x \to p} (fg)(x) = AB$;

(c) $\lim_{x \to p} \left(\frac{f}{g}\right)(x) = \frac{A}{B}, \quad \text{if} B \neq 0.$

4.2 Continuous Functions

Definition 48

Suppose X and Y are metric spaces, $E \subset X, p \in E$, and f maps E into Y. Then f is said to be continuous at p if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$d_Y(f(x), f(p)) < \varepsilon$$

for all points $x \in E$ for which $d_X(x, p) < \delta$.

- If f is continuous at every point of E, then f is said to be continuous on E.
- f has to be defined at the point p in order to be continuous at p.
- f is continous at p if and only if $\lim_{x \to p} f(x) = f(p)$.

Theorem 27

Suppose X, Y, Z are metric spaces, $E \subset X$, f maps E into Y, g maps the range of f, $f(E)$, into Z, and h is the mapping of E into Z defined by

$$h(x) = g(f(x)) \quad (x \in E).$$

If f is continuous at a point $p \in E$ and if g is continuous at the point $f(p)$, then h is continuous at p. The function $h = f \circ g$ is called the composite of f and g.

4.3 Continuity and Compactness

Definition 49

A mapping $\mathbf{f}$ of a set E into R^k is said to be bounded if there is a real number M such that $|\mathbf{f}(x)| \leq M$ for all $x \in E$.

Theorem 28

Suppose f is a continuous mapping of a compact metric space X into a metric space Y. Then $f(X)$ is compact.

Theorem 29

Suppose f is a continuous real function on a compact metric space X, and

$$M = \sup_{p \in X} f(p), \quad m = \inf_{p \in X} f(p). \tag{32}$$

Then there exist points $p, q \in X$ such that $f(p) = M$ and $f(q) = m$.

- The conclusion may also be stated as follows: There exist points p and q in X such that $f(q) \leq f(x) \leq f(p)$ for all $x \in X$; that is, f attains its maximum (at p) and its minimum (at q).

Definition 50

Let f be a mapping of a metric space X into a metric space Y. We say that f is uniformly continuous on X if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d_Y(f(p), f(q)) < \varepsilon \tag{33}$$

for all p and q in X for which $d_X(p, q) < \delta$.

Theorem 30

Let f be a continuous mapping of a compact metric space X into a metric space Y. Then f is uniformly continuous on X.

4.4 Continuity and Connectedness

Theorem 31

If f is a continuous mapping of a metric space X into a metric space Y, and if E is a connected subset of X, then $f(E)$ is connected.

Theorem 32

(Intermediate Vaalue Theorem) *Let f be a continuous real function on the interval $[a, b]$. If $f(a) < f(b)$ and if c is a number such that $f(a) < c < f(b)$, then there exists a point $x \in (a, b)$ such that $f(x) = c$.*

4.5 Discontinuities

- If x is a point in the domain of definition of the function f at which f is not continuous, we say that f is discontinuous at x.

Definition 51

Let f be defined on (a,b). Consider any point x such that $a \leq x < b$. We write $f(x+) = q$ if $f(t_n) \to q$ as $n \to \infty$, for all sequences $\{t_n\}$ in (x,b) such that $t_n \to x$. To obtain the definition of $f(x-)$, for $a < x \leq b$, we restrict ourselves to sequences $\{t_n\}$ in (a,x).

- It is clear that any point x of (a,b), $\lim_{t\to x} f(t)$ exists if and only if

$$f(x+) = f(x-) = \lim_{t\to x} f(t).$$

Definition 52

Let f be defined on (a,b). If f is discontinuous at a point x and if $f(x+)$ and $f(x-)$ exist, then f is said to have a discontinuity of the first kind. Otherwise, it is of the second kind.

4.6 Monotonic Functions

Definition 53

Let f be real on (a, b). Then f is said to be monotonically increasing on (a, b) if $a < x < y < b$ implies $f(x) \leq f(y)$.

Theorem 33

Let f be monotonically increasing on (a, b). Then $f(x+)$ and $f(x-)$ exist at every point of x of (a, b). More precisely

$$\sup_{a<t<x} f(t) = f(x-) \leq f(x) \leq f(x+) = \inf_{x<t<b} f(t). \tag{34}$$

Furthermore, if $a < x < y < b$, then

$$f(x+) \leq f(x-). \tag{35}$$

4.7 Infinite Limits and Limits at Infinity

- For any real number x, we have already defined a neighborhood of x to be any segment $(x-\delta, x+\delta)$.

Definition 54

For any real c, the set of real numbers x such that $x > c$ is called a neighborhood of $+\infty$ and is written $(c, +\infty)$. Similarly, the set $(-\infty, c)$ is a neighborhood of $-\infty$.

Definition 55

Let f be a real function defined on E. We say that

$$f(t) \to A \text{ as } t \to x$$

where A and x are in the extended real number system, if for every neighborhood U of A there is a neighborhood V of x such that $V \cap E$ is not empty, and such that $f(t) \in U$ for all $t \in V \cap E, t \neq x$.

- Three important theorems.

Theorem 34

If f is continuous on $[a, b]$ and $f(a) < 0 < f(b)$, then there is some x in $[a, b]$ such that $f(x) = 0$.

Theorem 35

If f is continuous on $[a, b]$, then f is bounded above on $[a, b]$, that is, there is some number N such that $f(x) \leq N$ for all x in $[a, b]$.

Theorem 36

If f is continuous on $[a, b]$, then there is some number y in $[a, b]$ such that $f(y) \geq f(x)$ for all x in $[a, b]$.

4.8 The Derivative of a Real Function

Definition 56

Let f be defined (and real-valued) on $[a, b]$. For any $x \in [a, b]$ form the quotient

$$\phi(t) = \frac{f(t) - f(x)}{t - x} \quad (a < t < b, t \neq x), \tag{36}$$

and define

$$f'(x) = \lim_{t \to x} \phi(t), \tag{37}$$

provided this limit exists. f' is called the *derivative of f*.

Theorem 37

Let f be defined on $[a, b]$. If f is differentiable at a point $x \in [a, b]$, then f is continuous at x.

Theorem 38

Suppose f and g are defined on $[a, b]$ *and are differentiable at point* $x \in [a, b]$*. Then* $f + g$*,* fg *abd* f/g *are differentiable at* x*, and*

(a) $(f + g)'(x) = f'(x) + g'(x)$;

(b) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$;

(c) $\left(\frac{f}{g}\right)' = \frac{g(x)f'(x) - g'(x)f(x)}{g^2(x)} \quad with g(x) \neq 0.$

Theorem 5.1

Suppose f os continuous on $[a, b]$*,* $f'(x)$ *exists at some point* $x \in [a, b]$*, g is defined on an interval I which contains the range of f, and g is diffrentiable at the point* $f(x)$*. If* $h(t) = g(f(t))$ *and* $(a \leq t \leq b)$*, then h is differentiable at x, and*

$$h'(x) = g'(f(x))f'(x). \tag{38}$$

Example 7

Let f be defined by

$$f(x) = \begin{cases} x \sin \frac{1}{x} & (x \neq 0) \\ 0 & (x = 0) \end{cases} \tag{39}$$

Applying the theorems, we have

$$f'(x) = \sin \frac{1}{x} - \frac{1}{x} \cos \frac{1}{x} \quad (x \neq 0) \tag{40}$$

At $x = 0$ there is no $f'(x)$.

Definition 57

Let f be a real function defined on a metric space X. We say that f has a *local maximum* at a point $p \in X$ if there exists $\delta > 0$ such that $f(q) \leq f(p)$ for all $q \in X$ with $d(p, q) < \delta$.

Theorem 39

Let f be defined on $[a, b]$; if f has a local maximum at a point $x \in (a, b)$, and if $f'(x)$ exists, then $f'(x) = 0$.

Theorem 40

If f is a real continuous function on $[a, b]$ which is differentiable in (a, b), then there is a point $x \in (a, b)$ at which $f(b) - f(a) = (b - a)f(x)$.

Theorem 41

Suppose f is a real differentiable function on $[a, b]$ and suppose $f'(a) < \gamma < f'(b)$. Then there is a point $x \in (a, b)$ such that $f'(x) = \gamma$.

Theorem 42

Suppose f and g are areal and differentiable in (a, b) *and* $g'(x) \neq 0$ *for all* $x \in (a, b)$, *where* $\infty \leq < b \leq +\infty$. *Suppose*

$$\frac{f'(x)}{g'(x)} \to A \quad as \quad x \to a. \tag{41}$$

If

$$f(x) \to 0 \; and \; g(x) \to 0 \; as \; x \to a \tag{42}$$

or if

$$g(x) \to +\infty \; as \; x \to a, \tag{43}$$

then

$$\frac{f(x)}{g(x)} \to A \; as \; x \to a. \tag{44}$$

Definition 58

If f has a derivative f' on a interval, and if f' is itself differentiable, we denote the derivative of f' by f'' the second derivative of f'. Continuing in this manner, we obtain functions

$$f, f', f'', f^{(3)}, \dots, f^{(n)},$$

each of wich is the derivative of the preceding one. $f^{(n)}$ us cakked tge nth derivative, or the derivative of order n, of f.

Theorem 43

Suppose f is a real function on $[a, b]$, n is a positive integer, $f^{(n-1)}$ is continuous on $[a, b]$, $f^{(n)}(t)$ exists for every $t \in (a, b)$. Let α, β be distinct points of $[a, b]$, and define

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)(\alpha)}}{k!}(t-\alpha)^k. \tag{45}$$

Example 8

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad \text{for all } x \tag{46}$$

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \quad \text{for all } x \tag{47}$$