

MINISTRY OF EDUCATION AND SCIENCES OF UKRAINE

NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
“IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE”

NATALIA REMEZ, VADYM BRONYTSKYI

**NUMERICAL METHODS OF THE
SOLUTION OF APPLIED TASKS
FOR FOREIGN STUDENTS**

THEORETICAL MATERIAL AND PRACTICAL

*Recommended by the methodical advice of KPI. Igor Sikorsky
as a textbook for applicants for a master's degree
according to the educational program "Engineering ecology and resource
saving" specialty 101 "Ecology"*

Kiev
Igor Sikorsky Kyiv Polytechnic Institute
2020

Editor-in-chief *Meish V.F.*, doctor of physical and mathematical sciences,
professor, leading researcher, S. P. Timoshenko Institute of
Mechanics The National Academy of Sciences of Ukraine

Reviewer Tkachuk K.K., doctor of technical sciences, professor

*Stamp provided by Methodical Council Igor Sikorsky Kyiv Polytechnic Institute (protocol No.
10 of 18.06.2020) on the proposal of the Scientific Council
Institute of energy saving and energy management (protocol No. 13 of 28.05.2020)*

Electronic network/online educational edition/publication

Remez Natalia Serhiivna, doctor of technical science, professor

Bronytskyi Vadym Olehovych, assistant

NUMERICAL METHODS OF THE SOLUTION OF APPLIED TASKS FOR FOREIGN STUDENTS THEORETICAL MATERIAL AND PRACTICAL

Numerical methods of the solution of applied tasks for foreign students: Theoretical material and practical [Electronic resource] : textbook for students of 101 "Ecology"/ Igor Sikorsky Kyiv Polytechnic Institute; authors.: Natalia Remez, Vadym Bronytskyi (1 file: 2,80 Mbyte). – Kiev: Igor Sikorsky Kyiv Polytechnic Institute, 2020. – 179 p.

In the presented grant basic provisions on performance individual semestrial a task - settlement work which subject occupies sections of a course of the higher mathematics on studying of differential notation of functions of two variables are stated. Educational the edition contains the main theoretical data, samples to performance of tasks on separate subjects, the list of references, and a model of execution of the report with rozrakhunovo ï works.

It is considered the basic concepts of the theory of functions of many variables and existing methods rozv "to a yazk of tasks on findings of a range of definition of functions; calculations of double border of function; findings of private derivatives of function of two variables, and functions which are made or set implicitly; definitions of extrema of function and so forth.

The educational edition is intended for applicants for a master's degree in specialty 101 "Ecology", the educational program "Engineering Ecology and resource-saving".

© *N. S. Remez, V. O. Bronytskyi*, 2020

© Igor Sikorsky Kyiv Polytechnic Institute, 2020

CONTENT

PREFACE	5
TITLE 1. NUMERICAL SOLUTION OF SYSTEMS OF LINEAR ALGEBRAIC EQUATIONS	7
§ 1.1. Short theoretical information	7
§ 1.2. Gauss method of solution of the systems of linear equations.....	8
§ 1.3 LU method of calculation of the systems of linear equations	15
§ 1.5. Iteration methods of untiing of the systems linear equations of algebra	29
A task for independent implementation	44
TITLE 2. APPROXIMATE SOLUTION OF NONLINEAR EQUATIONS.....	47
§ 2.2. Separation of roots of equation	48
2.2.1. Root separation conditions.....	48
2.2.2 Graphic method of separation of root	49
2.2.3 Method of tests.....	50
§ 2.3. Method of half-note division.....	52
§ 2.4. Method of chords	56
§ 2.5. Method of simple iteration.....	59
§ 2.6. Method of Newton	63
§ 2.7. Method of simple iteration.....	70
§ 2.8. Method of Newton	74
§ 2.9. Method of secant.....	80
§ 2.10. A method of simple iteration is for the systems of two equations.....	82
§ 2.11. A method of Newton is for the systems of two equations	91
§ 2.12. Iteration methods of decision of the systems of nonlinear equations..	98
A task is for independent implementation	111

TITLE 3. NUMERICAL APPROACH OF FUNCTIONS.....	117
§ 3.1. Formulation of the problem	117
§ 3.2. Lagrange interpolation polynomial.....	122
§ 3.3. Error estimation of Lagrange interpolation formula.....	130
TITLE 4. NUMERICAL APPROACH OF FUNCTIONS.....	134
§ 4.1. Formulation of the problem	134
§ 4.2. Lagrange interpolation polynomial.....	139
§ 4.3. Error estimation of Lagrange interpolation formula.....	147
TITLE 5. EXPERIMENTAL DATA PROCESSING METHODS.....	151
§ 5.1. The least squares method	151
TITLE 6. APPROXIMATE SOLUTION OF COMMON DIFFERENTIAL EQUATIONS	158
§ 6.1. Problem statement.....	158
§ 6.2. Euler's method and its modifications	161
§ 6.3. Runge-Kutta method	168
Tasks for self-fulfillment.....	174
REFERENCES	178

PREFACE

Mathematical modeling of processes and phenomena in various fields of science and technology is one of the main ways to obtain new knowledge and develop new technologies. When performing mathematical modeling, a modern engineer, regardless of his specialization, must have a minimum set of algorithms for computational mathematics, as well as methods of their software implementation on modern personal computers (PCs).

The main purpose of the presented manual is to present the basics of modern methods of computational mathematics on the basis of the general course "Higher Mathematics" for technical universities in order to implement numerical methods using modern computer technology.

The textbook consists of six sections. The first and second sections discuss the basic numerical methods for solving linear and nonlinear equations. For systems of linear algebraic equations - this is the Gaussian method, the method of LU - decomposition, iterative methods. For nonlinear equations, the method of simple iteration and Newton's method with conditions of their convergence are considered. The third and fourth sections are devoted to the issues of numerical interpolation and numerical integration of functions. The fifth section outlines the basic approaches to constructing numerical algorithms for solving ordinary differential equations. One-step and multi-step methods for solving differential equations are considered. The application of numerical methods for solving ordinary differential equations for problems that arise in modern technology is shown. The sixth section is devoted to methods of processing experimental data.

In presenting the material in all sections of the manual are detailed examples and approaches to the application of numerical methods for solving specific problems with the presentation of appropriate graphic material.

The manual can also be used when conducting a computing workshop on modern PCs.

TITLE 1. NUMERICAL SOLUTION OF SYSTEMS OF LINEAR ALGEBRAIC EQUATIONS

§ 1.1. Short theoretical information

Will consider the system of linear equations in a matrix kind:

$$A\bar{x} = \bar{b}, \quad (1)$$

where A is a matrix of size $n \times n$ with permanent coefficients; \bar{b} is n - dimensional vector of the known constants; \bar{x} - n - dimensional vector of unknown values.

In matrix-vector form this system can be written as

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

In practice for the solution of the systems of not high order ($n = 2 \div 4$) the method of Cramer is used. In obedience to this rule, at the solution of the system (2), k -th component x_k of vector \bar{x} it is determined in obedience to formulas

$$x_k = \frac{\det(A_k)}{\det(A)}, \quad (3)$$

where $\det(A)$ - визначник matrices A and, $\det(A_k)$ – determinant of matrix A , in what k -th column is replaced by a vector \bar{b} .

In particular, for a case $n = 3$ formulas (3) have a next kind

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad x_3 = \frac{\det(A_3)}{\det(A)},$$

where

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad \det(A_1) = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix},$$

$$\det(A_2) = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}, \quad \det(A_3) = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}.$$

As specified already, the rule of Cramer was used for the decision of the systems of equations of not high order and, mainly, at theoretical researches. At the increase of order of the systems this method requires the considerable expenses of machine time, and in calculable practice is used seldom. It is also just for being of decision of the systems of linear equations by means of inverse matrix

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

where A^{-1} it is an inverse matrix to initial.

§ 1.2. Gauss method of solution of the systems of linear equations

Will consider system of equations (2) in a next kind

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \end{cases} \quad (4)$$

Will divide the first equation of the system (4) into a_{11} and will write down him in a kind

$$x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \dots = b_1^{(1)},$$

where next denotations are entered: $a_{12}^{(1)} = a_{12} / a_{11}, \dots, a_{1n}^{(1)} = a_{1n} / a_{11}, b_1^{(1)} = b_1 / a_{11}.$

An overhead index specifies in brackets, that coefficients were one time changed. Will multiply this equation on $-a_{21}$ and will add him to the second equation. The coefficients of the new got second equation look like $a_{2j}^{(1)} = a_{2j} - a_{21}a_{1j}^{(1)}, j = \overline{1, n}; b_2^{(1)} = b_2 - a_{21}b_1^{(1)}$. Such approach at the choice of multiplier provides equality to the zero of coefficient $a_{21}^{(1)}$. Like for other equations next substitution

$$a_{ij}^{(1)} = a_{ij} - a_{i1}a_{1j}^{(1)}, b_i^{(1)} = b_i - a_{i1}b_1^{(1)}, i = \overline{2, n}; j = \overline{1, n};$$

provides equality to the zero of all coefficients in the first column of matrix A, after an exception $a_{11}^{(1)}$, what equals 1. Actually it is not needed to calculate an element that becomes zero. Elements a_{i1} now does not occupy memory the personal COMPUTER, and calculations are executed, beginning from $j = 2$.

As a result of such transformation of initial matrix equations take the form

$$\begin{aligned}
 x_1 + a_{12}^{(1)} x_2 + a_{13}^{(1)} x_3 + \dots + a_{1n}^{(1)} x_n &= b_1^{(1)} \\
 a_{22}^{(1)} x_2 + a_{23}^{(1)} x_3 + \dots + a_{2n}^{(1)} x_n &= b_2^{(1)} \\
 &\dots\dots\dots \\
 a_{n2}^{(1)} x_2 + a_{n3}^{(1)} x_3 + \dots + a_{nn}^{(1)} x_n &= b_n^{(1)}.
 \end{aligned}
 \tag{5}$$

On a next step will exclude from consideration the first line and first column of the system (5) and will apply analogical foregoing procedure to equations from the second to n -th. Will write down formulas for the calculation of new values of coefficients:

$$\begin{aligned}
 a_{2j}^{(2)} &= a_{2j}^{(1)} / a_{22}^{(1)}; \quad j = \overline{3, n}; \quad b_2^{(2)} = b_2^{(1)} / a_{22}^{(1)}; \\
 a_{ij}^{(2)} &= a_{ij}^{(1)} - a_{i2}^{(1)} a_{2j}^{(2)}; \quad b_i^{(2)} = b_i^{(1)} - a_{i2}^{(1)} b_2^{(2)}, \quad i = \overline{3, n}; \quad j = \overline{3, n}.
 \end{aligned}$$

Repeat this procedure for all lines of regenerate matrix. If to designate $a_{ij}^{(0)} = a_{ij}$, then the general formula of **Gauss method of exception** be written down as follows:

$$\begin{aligned}
 a_{kj}^{(k)} &= a_{kj}^{(k-1)} / a_{kk}^{(k-1)}; \quad a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{ik}^{(k-1)} a_{kj}^{(k)}; \\
 b_k^{(k)} &= b_k^{(k-1)} / a_{kk}^{(k-1)}; \quad b_i^{(k)} = b_i^{(k-1)} - a_{ik}^{(k-1)} b_k^{(k)}; \\
 k &= \overline{1, n}; \quad i = \overline{k+1, n}; \quad j = \overline{k+1, n}.
 \end{aligned}
 \tag{6}$$

As a result the system of equations take the form

$$\begin{aligned}
x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \dots + a_{1n}^{(1)}x_n &= b_1^{(1)} \\
x_2 + a_{23}^{(2)}x_3 + \dots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\
x_3 + \dots + a_{3n}^{(3)}x_n &= b_3^{(3)} \\
&\dots\dots\dots \\
&x_n = b_n^{(n)}.
\end{aligned} \tag{7}$$

Formulas are presented higher show by itself ***direct motion*** of method of **Gauss method of exception**. The system of kind (7) has a three-cornered structure, that allows consistently to calculate the value of unknown, beginning from the last

$$\begin{aligned}
x_n &= \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}} = b_n^{(n)}; \\
&\dots\dots\dots \\
x_2 &= \frac{b_2^{(1)} - a_{23}^{(1)}x_3 - \dots - a_{2n}^{(1)}x_n}{a_{22}^{(1)}}; \\
x_1 &= \frac{b_1 - a_{12}x_2 - \dots - a_{1n}x_n}{a_{11}}.
\end{aligned} \tag{8}$$

This process of successive calculation of values unknown is named ***the reverse calculation of Gauss method of exception***.

An algorithmic process is described higher shows by itself ***the method of exception of Gauss***. In the case when $a_{kk} = 0$, then it is impossible to use k -th line for the exception of elements of k -th column. In this case it is needed to change a k -th line placed with other line under a diagonal thus, that main element $a_{kk} \neq 0$. If it not maybe to execute this condition, the initial matrix of coefficients of the system of linear equations means is degenerate and the system does not have an only decision.

The algorithm of method of exception *of Gauss* is realized by two groups of formulas is direct motion and countermove. Direct motion of method of exception consists of n steps that is set forth below.

1-st step. Will divide the first equation of the initial system (4) into a_{11} ($a_{11} \neq 0$) and bring the system (4) over to the kind (5).

2-nd step. Eliminate from consideration the first line and first column of regenerate matrix and will apply analogical foregoing procedure to equations from the second to n th taking into account the formulas of type (6).

n- th step. The system over of equations is brought to the kind (7). The countermove of method of exception of *of Gauss* will be realized in obedience to formulas (8).

For example: will consider the system of kind

$$\begin{cases} x_1 + 2x_2 + x_3 + 4x_4 = 13 \\ 2x_1 + 4x_3 + 3x_4 = 28 \\ 4x_1 + 2x_2 + 2x_3 + x_4 = 20 \\ -3x_1 + x_2 + 3x_3 + 2x_4 = 6. \end{cases}$$

1-st step of exception. In the first equation coefficient at $a_{11} = 1$ coming from it, will increase the elements of the first line accordingly on $a_{21} = 2$, $a_{31} = 4$, $a_{41} = -3$ and will take away them from the second, third and fourth equations. This is provide equality to the zero of coefficients at x_1 beginning from the second equation. Thus, the system of equations assumes a next type

$$\begin{cases} x_1 + 2x_2 + x_3 + 4x_4 = 13 \\ -4x_2 + 2x_3 - 5x_4 = 2 \\ -6x_2 - 2x_3 - 15x_4 = -32 \\ 7x_2 + 6x_3 + 14x_4 = 45. \end{cases}$$

2–nd step of exception. The last three equations of the previous system are examined. The coefficients of the second equation are divided by $a_{22}^{(1)} = -4$ then multiplied accordingly by $a_{31}^{(1)} = -6$ and $a_{41}^{(1)} = 7$ and subtracted from the third and fourth equations of the system. The second step of exception results in the system

$$\left\{ \begin{array}{l} x_1 + 2x_2 + x_3 + 4x_4 = 13 \\ -4x_2 + 2x_3 - 5x_4 = 2 \\ -5x_3 - 7,5x_4 = -35 \\ 9,5x_3 + 5,25x_4 = 48,5. \end{array} \right.$$

3–the step of exception results in the system of three-cornered kind (7)

$$\left\{ \begin{array}{l} x_1 + 2x_2 + x_3 + 4x_4 = 13 \\ -4x_2 + 2x_3 - 5x_4 = 2 \\ -5x_3 - 7,5x_4 = -35 \\ -9x_4 = -18. \end{array} \right.$$

Implementing *ounermove* of method of Gauss, from the last equation $-9x_4 = -18$ find $x_4 = 2$. Put x_4 in the third equation get $x_3 = 4$. From the second equation have $x_2 = -1$ and from the first $x_1 = 3$. As a result, the decision of the initial system is got $\bar{x} = (3, -1, 4, 2)^T$.

At application of direct motion of Gauss method initial system over of equations is brought to the three-cornered kind.

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1 \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \dots + a_{2n}^{(1)}x_n = b_2^{(1)} \\ \dots\dots\dots \\ a_{nn}^{(n-1)}x_n = b_n^{(n-1)}. \end{array} \right. \quad (9)$$

Determinant of initial matrix A, in obedience to representation (9), it is determined after a formula

$$\det A = a_{11} \cdot a_{22}^{(1)} \dots a_{nn}^{(n-1)}. \quad (10)$$

Thus, the determinant $\det A$ is equal to the product of all elements on the main diagonal of the system (9).

For example: will calculate determinant for the system of equations, that was examined in a previous example

$$\begin{cases} x_1 + 2x_2 + x_3 + 4x_4 = 13 \\ 2x_1 + 4x_3 + 3x_4 = 28 \\ 4x_1 + 2x_2 + 2x_3 + x_4 = 20 \\ -3x_1 + x_2 + 3x_3 + 2x_4 = 6. \end{cases}$$

After application of direct motion of Gauss exception method the initial system of equations has a next kind

$$\begin{cases} x_1 + 2x_2 + x_3 + 4x_4 = 13 \\ -4x_2 + 2x_3 - 5x_4 = 2 \\ -5x_3 - 7,5x_4 = -35 \\ -9x_4 = -18. \end{cases}$$

Thus, the determinant of the system is calculated on a formula (2.10)

$$\det A = a_{11} \cdot a_{22}^{(1)} \cdot a_{33}^{(2)} \cdot a_{44}^{(3)} = 1 \cdot (-4) \cdot (-5) \cdot (-9) = -180.$$

§ 1.3 LU method of calculation of the systems of linear equations

One of widespread modern methods of calculation of the systems of linear algebraic equations of *there is a* triangular matrix decomposition method, or *LU of* – factorization.

The algorithms of this method are near to the method of exception of Гаусса. Main advantage of method *of LU of -факторизації* as compared to the method of exception of Gauss is possibility of receipt of more effective decisions for different vectors \bar{b} in right part of the system (2.1) at an unchanging initial matrix A .

Possibly, that the matrix of the system of equations (1) can be decomposed on two factors:

$$A = LU, \quad (11)$$

where a matrix is L it is bottom three-cornered, and matrix U - overhead three-cornered (denotation of these matrices originates from the first letters of the English words **of Low** – lower and **Upper** – overhead, that becomes clear from presentation of kind (2.12)). Will mark, that on the main diagonal of matrix U there are units. Coming from it, determinant of matrix A equals the product of diagonal elements l_{ii} matrices L .

Structure of matrices L and U it is determined by next representation

$$L = \begin{bmatrix} l_{11} & & & & \\ l_{21} & l_{22} & & & \\ l_{31} & l_{32} & l_{33} & & \\ \dots & \dots & \dots & \dots & \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{bmatrix}, \quad U = \begin{bmatrix} 1 & u_{12} & u_{13} & \dots & u_{1n} \\ & 1 & u_{23} & \dots & u_{2n} \\ & & 1 & \dots & u_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ & & & & 1 \end{bmatrix}. \quad (12)$$

It is known that for every undegenerate matrix and curriculum of kind (11 - 12) it exists and only. Will present the system of equations in a next kind:

$$LU\bar{x} = \bar{b}. \tag{13}$$

Will define an auxiliary vector \bar{z} as

$$U\bar{x} = \bar{z}. \tag{14}$$

From this equation vector \bar{z} finding is impossible, as unknown is a vector \bar{x} . But, if to put \bar{z} in (2.13), will get

$$L\bar{z} = \bar{b}. \tag{15}$$

Due to the nospread function of matrix L vector \bar{z} it is possible easily to define. For this purpose will write (2.15) down as a system of equations

$$\begin{aligned} l_{11}z_1 &= b_1 \\ l_{21}z_1 + l_{22}z_2 &= b_2 \\ l_{31}z_1 + l_{32}z_2 + l_{33}z_3 &= b_3 \\ &\dots\dots\dots \\ l_{n1}z_1 + l_{n2}z_2 + l_{n3}z_3 + \dots + l_{nn}z_n &= b_n, \end{aligned} \tag{16}$$

from where get

$$\begin{aligned} z_1 &= b_1 / l_{11}, \\ z_2 &= (b_2 - l_{21}z_1) / l_{22}, \\ z_3 &= (b_3 - l_{31}z_1 - l_{32}z_2) / l_{33}, \\ &\dots\dots\dots \end{aligned} \tag{17}$$

or in a general view

$$z_1 = b_1 / l_{11}$$

$$z_i = \left(b_i - \sum_{j=1}^{i-1} l_{ij} z_j \right) / l_{ii}, \quad i = \overline{2, n}. \quad (18)$$

This process is named *a direct exception (by a direct substitution or direct motion)*. That equation (2.18) made sense, diagonal elements of matrix L must not zero.

As a vector \bar{z} it is found, will come back to (2.14) and will find the vector of unknown \bar{x} . For this purpose will write (2.14) down in a co-ordinate form

$$\begin{aligned} x_1 + u_{12}x_2 + u_{13}x_3 + \dots + u_{1n}x_n &= z_1 \\ x_2 + u_{23}x_3 + \dots + u_{2n}x_n &= z_2 \\ \dots & \\ x_{n-1} + u_{n-1,n}x_n &= z_{n-1} \\ x_n &= z_n. \end{aligned} \quad (19)$$

Beginning from the last equation, it is possible consistently to find the components of vector \bar{x} . In a general view *reverse substitution* (or countermove) determined after formulas

$$x_n = z_n,$$

$$x_i = z_i - \sum_{j=i+1}^n u_{ij}x_j, \quad i = \overline{n-1, 1}. \quad (20)$$

Thus, the decision of CJAP can be found by means of foregoing algorithm, if the curriculum of matrix is known and on the corresponding three-cornered matrices of L and U, in obedience to formulas (12).

For example: to untie the system of equations, using L - U time-table. Will consider the system of kind

$$\begin{cases} 3x_1 + x_2 - x_3 + 2x_4 = 6 \\ -5x_1 + x_2 + 3x_3 - 4x_4 = -12 \\ 2x_1 + x_3 - x_4 = 1 \\ x_1 - 5x_2 + 3x_3 - 3x_4 = 3. \end{cases}$$

Find the coefficients of matrices L and U coming from presentation of their product (11 - 12) for the case of matrices of size 4×4

$$\begin{bmatrix} l_{11} & l_{11}u_{12} & l_{11}u_{13} & l_{11}u_{14} \\ l_{21} & l_{21}u_{12} + l_{22} & l_{21}u_{13} + l_{22}u_{23} & l_{21}u_{14} + l_{22}u_{24} \\ l_{31} & l_{31}u_{12} + l_{32} & l_{31}u_{13} + l_{32}u_{23} + l_{33} & l_{31}u_{14} + l_{32}u_{24} + l_{33}u_{34} \\ l_{41} & l_{41}u_{12} + l_{42} & l_{41}u_{13} + l_{42}u_{23} + l_{43} & l_{41}u_{14} + l_{42}u_{24} + l_{43}u_{34} + l_{44} \end{bmatrix}.$$

In obedience to the last presentation of product of matrices calculate the elements of matrices L and U in a next sequence:

- 1) directly equating the elements of the first column with the corresponding elements of initial matrix A get

$$l_{11} = a_{11}, l_{21} = a_{21}, l_{31} = a_{31}, l_{41} = a_{41}, \text{ or}$$

$$l_{11} = 3, l_{21} = -5, l_{31} = 2, l_{41} = 1;$$

- 2) find unknown u_{12}, u_{13}, u_{14} first to the line

$$u_{12} = \frac{a_{12}}{l_{11}} = \frac{1}{3}, u_{13} = \frac{a_{13}}{l_{11}} = -\frac{1}{3}, u_{14} = \frac{a_{14}}{l_{11}} = \frac{2}{3};$$

- 3) find l_{22}, l_{32}, l_{42} from the second column of matrix

$$l_{22} = a_{22} - l_{21}u_{12}, \quad l_{32} = a_{32} - l_{31}u_{12}, \quad l_{42} = a_{42} - l_{41}u_{12}, \quad \text{or}$$

$$l_{22} = 1 - (-5)\frac{1}{3} = \frac{8}{3}, \quad l_{32} = 0 - 2 \cdot \frac{1}{3} = -\frac{2}{3}, \quad l_{42} = -5 - 1 \cdot \frac{1}{3} = -\frac{16}{3};$$

4) find u_{23}, u_{24} in obedience to formulas

$$u_{23} = \frac{a_{23} - l_{21}u_{13}}{l_{22}}, \quad u_{24} = \frac{a_{24} - l_{21}u_{14}}{l_{22}}, \quad \text{or}$$

$$u_{23} = \frac{(3 - (-5) \cdot (-\frac{1}{3}))}{\frac{8}{3}} = \frac{1}{2}, \quad u_{24} = \frac{(-4 - (-5) \cdot \frac{2}{3})}{\frac{8}{3}} = -\frac{1}{4};$$

5) sizes l_{33}, l_{43} calculate matrices coming from the values of the third column

A

$$l_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23}, \quad l_{43} = a_{43} - l_{41}u_{13} - l_{42}u_{23}, \quad \text{or}$$

$$l_{33} = 1 - 2 \cdot (-\frac{1}{3}) - (-\frac{2}{3}) \cdot \frac{1}{2} = 2, \quad l_{43} = 3 - 1 \cdot (-\frac{1}{3}) - (-\frac{16}{3}) \cdot \frac{1}{2} = 6;$$

6) size u_{34} calculated on a formula

$$u_{34} = \frac{a_{34} - l_{31}u_{14} - l_{32}u_{24}}{l_{33}}, \quad u_{34} = \frac{-1 - 2 \cdot \frac{2}{3} - (-\frac{2}{3}) \cdot (-\frac{1}{4})}{2} = -\frac{5}{4};$$

7) find the last element of matrix L

$$l_{44} = a_{44} - l_{41}u_{14} - l_{42}u_{24} - l_{43}u_{34}, \text{ or}$$

$$l_{44} = -3 - 1 \cdot \frac{2}{3} - \left(-\frac{16}{3}\right) \cdot \left(-\frac{1}{4}\right) - 6 \cdot \left(-\frac{5}{4}\right) = \frac{5}{2}.$$

Thus matrices L and U have a next kind

$$L = \begin{pmatrix} 3 & & & \\ -5 & 2,6667 & & \\ 2 & -0,6667 & 2 & \\ 1 & -5,3333 & 6 & 2,5 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 0,3333 & -0,3333 & 0,6667 \\ & 1 & 0,5 & -0,25 \\ & & 1 & -1,25 \\ & & & 1 \end{pmatrix}.$$

Coming from equations (2.15), or in the unfolded kind are formulas (16), will get upshots for a vector \bar{z}

$$z_1 = \frac{b_1}{l_{11}}, \text{ or } z_1 = \frac{6}{3} = 2;$$

$$z_2 = \frac{b_2 - l_{21}z_1}{l_{22}}, \text{ or } z_2 = \frac{-12 - (-5) \cdot 2}{\frac{8}{3}} = -0,75;$$

$$z_3 = \frac{b_3 - l_{31}z_1 - l_{32}z_2}{l_{33}}, \text{ or } z_3 = \frac{1 - 2 \cdot 2 - \left(-\frac{2}{3}\right) \cdot \left(-\frac{3}{4}\right)}{2} = -1,75;$$

$$z_4 = \frac{b_4 - l_{41}z_1 - l_{42}z_2 - l_{43}z_3}{l_{44}}, \text{ or}$$

$$z_4 = \frac{3 - 1 \cdot 2 - \left(-\frac{16}{3}\right) \cdot \left(-\frac{3}{4}\right) - 6 \cdot \left(-\frac{7}{4}\right)}{\frac{5}{2}} = 3.$$

Going back to the system (19) it is possible to define the components of vector \bar{x}

$$x_4 = z_4, \text{ or } \underline{x_4 = 3};$$

$$x_3 = z_3 - u_{34} \cdot x_4, \text{ or } x_3 = -1,75 - (-1,25) \cdot 3, \quad \underline{x_3 = 2};$$

$$x_2 = z_2 - u_{23} \cdot x_3 - u_{24} \cdot x_4, \text{ or}$$

$$x_2 = -0,75 - 0,5 \cdot 2 - (-0,25) \cdot 3, \quad \underline{x_2 = -1};$$

$$x_1 = z_1 - u_{12} \cdot x_2 - u_{13} \cdot x_3 - u_{14} \cdot x_4, \text{ or}$$

$$x_1 = 2 - \frac{1}{3} \cdot (-1) - \left(-\frac{1}{3}\right) \cdot 2 - \frac{2}{3} \cdot 3, \quad \underline{x_1 = 1}.$$

§ 1.4. Error of decision of the system of linear equations

Will enter the concept of norm of vector and matrix, and also concept of conditionality of matrix, with that there are the constrained questions of estimation of error of untuing of the systems of linear equations.

By the norm of vector $\bar{x} = (x_1, x_2, \dots, x_n)$ name a material number that is designated - $\|\bar{x}\|$ and satisfies next terms:

- 1) $\|\bar{x}\| \geq 0$, thus $\|\bar{x}\| = 0$ then and only after, if $\bar{x} = 0$;
- 2) $\|c\bar{x}\| = |c| \cdot \|\bar{x}\|$, where c is a scalar size;
- 3) $\|\bar{x} + \bar{y}\| \leq \|\bar{x}\| + \|\bar{y}\|$.

Function of kind

$$\|\bar{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

at arbitrary $p \geq 1$ satisfies the indicated axioms of norm. The norm of such to the type is named *the norm of Gelder with an index*. Among Gelder norms most widespread are following:

$$\|\bar{x}\|_1 = \sum_{i=1}^n |x_i| \quad - \quad 1\text{-norm},$$

$$\|\bar{x}\|_k = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} \quad - \quad k\text{-norm}$$

$$\|\bar{x}\|_m = \max_{1 \leq i \leq n} |x_i| \quad - \quad m\text{-norm}.$$

Norm $\|\bar{x}\|_k$ named also euclidean and designated $\|\bar{x}\|_E$.

For example: let a vector is set $\bar{x} = (1, 2, 3)$. In obedience to the brought norms over have

$$\|\bar{x}\|_1 = \sum_{i=1}^3 |x_i| = 1 + 2 + 3 = 6;$$

$$\|\bar{x}\|_k = \left(\sum_{i=1}^3 |x_i|^2 \right)^{1/2} = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{1+4+9} = \sqrt{14} \approx 3,74;$$

$$\|\bar{x}\|_m = \max_{1 \leq i \leq 3} |x_i| = \max(1, 2, 3) = 3.$$

All norms of vectors are equivalent in the that understanding, that if sequence of vectors $\{\bar{x}^j = (x_1^j, x_2^j, \dots, x_n^j)\}$ gathers on some norm to the vector $\bar{x}^0 = (x_1^0, x_2^0, \dots, x_n^0)$ id est

$$\lim_{j \rightarrow \infty} \|\bar{x}^j - \bar{x}^0\| = 0,$$

then she gathers to the vector \bar{x}^0 and on arbitrary to other norm. In finite-dimensional the rationed space from convergence on a norm coordinate-wise convergence swims out and vice versa. It is thus assumed that sequence of vectors $\{\bar{x}^j\}$ coordinate-wise gathers to the vector \bar{x}^0 if for all $i = 1, 2, \dots, n$ correlations are executed

$$\lim_{j \rightarrow \infty} x_i^j = x_i^0.$$

By the norm of matrix A a material number is named $\|A\|$ that satisfies next axioms:

- 1) $\|A\| \geq 0$, $\|A\| = 0$ then and only after, if $A = 0$ (0 is a zero matrix);
- 2) $\|\alpha A\| = |\alpha| \cdot \|A\|$, where α it is a scalar size;
- 3) $\|A + B\| \leq \|A\| + \|B\|$;
- 4) $\|AB\| \leq \|A\| \cdot \|B\|$,

where $\|A\|$ is a matrix the dimension of that coincides with the dimension of matrix \bar{x} .
And.

Norm of matrix $\|A\|$ named *concorded* with the norm of vector $\|\bar{x}\|$, if for arbitrary \bar{x} correlation is executed

$$\|A\bar{x}\| \leq \|A\| \cdot \|\bar{x}\|.$$

Will mark, that with the same norm of vector can be concorded different norms of matrices.

Let $\|A\|$ it is the norm of matrix, concorded with the set norm of vector $\|\bar{x}\|$. Norm $\|A\|$ named *inferior* to the norm of vector $\|\bar{x}\|$ if there will be such vector $\bar{x} \neq 0$ that

$$\|A\bar{x}\| = \|A\| \cdot \|\bar{x}\|.$$

Thus, among all norms concorded with the set vectorial norm, an inferior norm is minimum. Will mark, that for the arbitrary norm of vector exists even one inferior norm of matrix.

The arbitrary norm of matrix satisfies inequalities:

$$\|E\| \geq 1, \text{ where } E \text{ is an unit matrix;}$$

$$\|A\| \cdot \|A^{-1}\| > 1, \text{ if } A \text{ is an undegenerate matrix.}$$

For an inferior norm correlations are executed

$$\|E\| = 1; \quad \|E \pm A^{-1}\| \leq \frac{1}{1 - \|A\|}; \quad \|AB\| \leq \|A\| \cdot \|B\|.$$

It is thus assumed that corresponding operations have maintenance.

The next norms of matrices have the most use:

$$\|A\|_m = \max_i \sum_{j=1}^n |a_{ij}| \quad - \quad (m \text{ is a norm});$$

$$\|A\|_l = \max_j \sum_{i=1}^n |a_{ij}| \quad - \quad (l \text{ is a norm});$$

$$\|A\|_k = \sqrt{\sum_{i,j=1}^n |a_{ij}|^2} \quad - \quad (k \text{ is a norm}).$$

For example: let a matrix is set

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

In obedience to the brought norms over have

$$\|A\|_m = \max_i \sum_{j=1}^3 |a_{ij}| = \max(1+2+3, 4+5+6, 7+8+9) = \max(6, 15, 24) = 24;$$

$$\|A\|_l = \max_j \sum_{i=1}^3 |a_{ij}| = \max(1+4+7, 2+5+8, 3+6+9) = \max(12, 15, 18) = 18;$$

$$\|A\|_k = \sqrt{\sum_{i,j=1}^3 |a_{ij}|^2} = \sqrt{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2} =$$

$$= \sqrt{1 + 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81} = \sqrt{285} \approx 16,8.$$

In particular, norm of matrix

$$\|A\|_m = \max_i \sum_{j=1}^n |a_{ij}|$$

inferior to the norm of vector $\|\bar{x}\|_m$.

Norm of matrix

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$$

inferior to the norm of vector $\|\bar{x}\|_1$.

Euclidean norm of matrix

$$\|A\|_k = \sqrt{\sum_{i,j=1}^n |a_{ij}|^2}$$

concerted with the norm of vector $\|\bar{x}\|_k$. In general case a euclidean norm is not inferior.

In practical problems elements of matrix But also vector \bar{b} in the system $A\bar{x} = \bar{b}$ are close numbers. At untiing of the system the errors of rounding appear an arbitrary exact method. Will set dependence between the error of decision and properties of matrix And.

Will examine the errors of calculations, that by визвані indignation of right part of the initial system of equations. Let $\bar{x} = \bar{x} + \Delta\bar{x}$ it is a decision of the indignant system $A(\bar{x} + \Delta\bar{x}) = \bar{b} + \Delta\bar{b}$ where $\Delta\bar{b}$ it is indignation of vector \bar{b} , $\Delta\bar{x}$ it is a corresponding error of exact decision \bar{x} . Then

$$A\bar{x} + A\Delta\bar{x} = \bar{b} + \Delta\bar{b}; \quad A\Delta\bar{x} = \Delta\bar{b}; \quad \Delta\bar{x} = A^{-1}\Delta\bar{b},$$

$$\|\Delta\bar{x}\| \leq \|A^{-1}\| \cdot \|\Delta\bar{b}\|.$$

Taking into account, that $\bar{b} = A\bar{x}$ and $\|\bar{b}\| \leq \|A\| \cdot \|\bar{x}\|$, it is possible to write down $\|\Delta\bar{x}\| \cdot \|\bar{b}\| \leq \|A\| \cdot \|A^{-1}\| \cdot \|\bar{x}\| \cdot \|\Delta\bar{b}\|$. From where swims out

$$\frac{\|\Delta\bar{x}\|}{\|\bar{x}\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\Delta\bar{b}\|}{\|\bar{b}\|}.$$

Thus, relative error of decision $\delta_x = \frac{\|\Delta\bar{x}\|}{\|\bar{x}\|}$ estimated through the relative error of right part $\delta_b = \frac{\|\Delta\bar{b}\|}{\|\bar{b}\|}$ by means of inequality

$$\delta_x \leq \|A\| \cdot \|A^{-1}\| \cdot \delta_b.$$

(2.21)

Size

$$\text{cond}(A) = \|A\| \cdot \|A^{-1}\| \quad (2.22)$$

named *a measure or number of conditionality of matrix And*. She is a maximally possible amplification factor ($\text{cond}(A) \geq 1$) relative error of untiing at indignat right part. Analogical results take place at the change of coefficients of matrix.

In a number of cases the number of conditionality of matrix is related to the estimation of own numbers.

By the own values of matrix And numbers are named λ that satisfy equation

$$\det(A - \lambda E) = 0,$$

where E is an unit matrix.

Thus, the number of conditionality of matrix is determined in obedience to a formula

$$\text{cond}(A) = \frac{\max_i |\lambda_i(A)|}{\min_i |\lambda_i(A)|}. \quad (2.23)$$

In case of symmetric matrix of value of sizes $\text{cond}(A)$ got for to the formulas (2.22), (2.23) coincide at the choice of norm $\|A\|_k$.

As follows from inequality (2.21), error of untiing of the system $A\bar{x} = \bar{b}$ it can appear considerable, if the matrix of the system is characterized by great variation of own numbers.

For example: system of equations

$$\begin{cases} 3.0x + 4.0y = 7.0 \\ 1.0x + 1.33y = 2.33, \end{cases}$$

has a decision $x = 1, y = 1$. Will consider the system of equations, that can be got from the initial system to small indignations of coefficients of right part of the system

$$\begin{cases} 3.0x + 4.0y = 7.0 \\ 1.0x + 1.33y = 2.32. \end{cases}$$

A decision of the last system will be $x = -3$, $y = 4$. In this case the small change of coefficients of equation (less than one percent) causes the considerable change of decision (hundreds of percents).

Will conduct research of conditionality of matrix A and in relation to different norms.

In obedience to the conducted calculations follows, that $\text{cond}(A)$ on a norm $\|A\|_1$ equals $3.7310e+003$, on a norm $\|A\|_k$ - $2.7769e+003$. Size $\text{cond}(A)$ in obedience to a formula (2.23) equals $1.8769e+003$. At the absolute error of right parts $\Delta b_1 = 0.01$ the relative error of right parts of the indignant system in relation to the initial system is determined after a formula $\delta_b = \frac{\|\Delta \bar{b}\|}{\|\bar{b}\|}$ and error of decision - $\delta_x \approx \delta_b \text{cond}(A)$. On a norm $\|A\|_1$ relative error of decision (of $x=3.9989$, and on a norm $\|A\|_k$ - ($x=3.7639$, that comports with the got results.

§ 1.5. Iteration methods of untiing of the systems linear equations of algebra

In a number of cases direct methods of decision of the systems of linear equations of algebra are effective not enough. Iteration methods are used in these cases. These methods will play an important role in calculable mathematics and will meet farther in next parts of manual.

System of equations $A\bar{x} = \bar{b}$ can be regenerate to the equivalent system of kind

$$\bar{x} = B\bar{x} + \bar{c}, \quad (24)$$

where \bar{x} it is a vector of unknown, and B and \bar{c} - a matrix and vector are some new accordingly. Setting some zero approaching $\bar{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^T$ and

$$\max_i \left| \frac{x_i^{(k+1)} - x_i^{(k)}}{x_i^{(k+1)}} \right| < \varepsilon, \quad i = \overline{1, n}; \quad \varepsilon = \text{const}, \quad (28)$$

where ε it is the set exactness of decision.

The method of simple iteration gathers to the sought after decision, if sufficient terms are executed convergences of iteration process, that can be written down in a kind

$$\max_i \sum_{j=1}^n \left| \frac{a_{ij}}{a_{ii}} \right| < 1. \quad (29)$$

Id est, maximal sum of the modules of relations of coefficients of any line to the diagonal coefficient less unit. This inequality means that the diagonal elements of the system must meet a condition

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|. \quad (30)$$

Taking into account, that by the decision of the system on k th to the iteration there is a vector $\bar{x}^{(k)}$ then for determination of condition of completion of convergence of iteration process it is expedient to apply the concept of distance between two vectors. There are a few methods of distance-finding (or metrics) between two vectors \bar{x}, \bar{y} . It is possible to determine distance between vectors a next formula

$$\rho(\bar{x}, \bar{y}) = \sum_{i=1}^n |x_i - y_i|. \quad (31)$$

Maybe application and another way of determination of metric

$$\rho(\bar{x}, \bar{y}) = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2}. \quad (32)$$

The condition of completion of iteration process looks like

$$\rho(\bar{x}^{(k+1)}, \bar{x}^{(k)}) < \varepsilon, \quad (33)$$

where a metric is $\rho(\bar{x}, \bar{y})$ it is determined in obedience to formulas (31) or (32).

Summarizing, will point the algorithm of decision of the system of linear equations (2.1) of algebra *the method of simple iterations* :

1. In the system of equations (2.1) check up implementation of condition (3.30). If a condition is not executed, then the initial system of equations transforms to the equivalent system a condition (2.30) is executed in that.
2. The equivalent system of equations appears in a kind (2.24).
3. The initial approaching is set $\bar{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^T$ and size of exactness ε . At implementation of condition (2.30) an iteration process gathers at any initial approaching. In practice a vector sets to the initial approaching \bar{c} in obedience to a formula (2.24).
4. Calculate the next approaching on a formula (2.27).
5. Estimate the "closeness" of two progressive approximations $\bar{x}^{(k+1)}$ and $\bar{x}^{(k)}$ by means of formulas of metric (2.31) or (2.32). Check up a condition $\rho(\bar{x}^{(k+1)}, \bar{x}^{(k)}) < \varepsilon$. If a condition is not executed - go back to a point 4. If a condition is executed - a decision is got.

For example: let the system of equations is set

$$\begin{aligned} \text{(A)} \quad & 9,7x_1 - 10x_2 + 0,6x_3 - 1,6x_4 = -12,3; \\ \text{(B)} \quad & 1,2x_1 + 11,2x_2 + 1,5x_3 + 2,5x_4 = 5,3; \\ \text{(C)} \quad & 1,2x_1 - x_2 + 8,5x_3 - 10,8x_4 = -13,7; \\ \text{(D)} \quad & 0,9x_1 + 2,5x_2 + 1,3x_3 + 12,1x_4 = 24,6. \end{aligned}$$

The coefficients of the initial system dissatisfy to the necessary condition of convergence of iteration process in obedience to formulas (3.30). Will conduct the series of equivalent transformations.

In equation (B) coefficient at x_2 there is a more sum of the modules of other coefficients on the module, that is why this equation can be left for the second equation of the new system. Coefficient at x_4 in equation (D) also anymore than sum of the modules of other coefficients of equation (D), that is why this equation can be taken for fourth equation of the new system. Thus, the new system has a next kind:

$$\begin{aligned}
 \text{(I)} & \quad \dots\dots\dots \\
 \text{(II)} & \quad 1,2x_1 + 11,2x_2 + 1,5x_3 + 2,5x_4 = 5,3; \\
 \text{(III)} & \quad \dots\dots\dots \\
 \text{(IV)} & \quad 0,9x_1 + 2,5x_2 + 1,3x_3 + 12,1x_4 = 24,6.
 \end{aligned}$$

Analysing the set system, see that for the receipt of equation (I) with a maximal on the module coefficient at x_1 it is enough to take the sum of equations (A)+ (B) :

$$\text{(I)} \quad 10,9x_1 + 1,2x_2 + 2,1x_3 + 0,9x_4 = -7.$$

For the receipt of equation (III) with a maximal on the module coefficient at x_3 it is enough to take the sum of equations (C)+ (D) :

$$\text{(III)} \quad 2,1x_1 + 1,5x_2 + 9,8x_3 + 1,3x_4 = 10,3.$$

Finally get the regenerate system of equations And - IV, that equivalent to the initial system and meets the condition of convergence of iteration process (2.30)

$$\begin{cases} 10,9x_1 + 1,2x_2 + 2,1x_3 + 0,9x_4 = -7; \\ 1,2x_1 + 11,2x_2 + 1,5x_3 + 2,5x_4 = 5,3; \\ 2,1x_1 + 1,5x_2 + 9,8x_3 + 1,3x_4 = 10,3; \\ 0,9x_1 + 2,5x_2 + 1,3x_3 + 12,1x_4 = 24,6. \end{cases}$$

For application of method of iterations will write down the system in a kind

$$\begin{cases} x_1 = \frac{1}{10,9}(-1,2x_2 - 2,1x_3 - 0,9x_4 - 7); \\ x_2 = \frac{1}{11,2}(-1,2x_1 - 1,5x_3 - 2,5x_4 + 5,3); \\ x_3 = \frac{1}{9,8}(-2,1x_1 - 1,5x_2 - 1,3x_4 + 10,3); \\ x_4 = \frac{1}{12,1}(-0,9x_1 - 2,5x_2 - 1,3x_3 + 24,6). \end{cases}$$

This system of equations in matrix - vectorial kind written down
 $\bar{x} = B\bar{x} + \bar{c}$ where

B =

$$\begin{bmatrix} 0 & -0.1101 & -0.1927 & -0.0826 \\ -0.1071 & 0 & -0.1339 & -0.2232 \\ -0.2143 & -0.1531 & 0 & -0.1327 \\ -0.0744 & -0.2066 & -0.1074 & 0 \end{bmatrix}$$

c =

$$\begin{bmatrix} -0.6422 \\ 0.4732 \\ 1.0510 \\ 2.0331 \end{bmatrix}$$

For a zero approaching of initial vector $\bar{x}^{(0)}$ accept the column of free members of the system $\bar{x}^{(0)} = \bar{c}$. The results of calculations are driven to the next table 2.1

Table 2.1

Number Iterations	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$x_4^{(k)}$
1	-1.0647	-0.0525	0.8465	1.8701
2	-0.9539	0.0565	1.0391	2.0322
3	-1.0164	-0.0174	0.9772	1.9807
4	-0.9921	0.0091	1.0087	2.0073
5	-1.0033	-0.0036	0.9960	1.9966
6	-0.9985	0.0017	1.0017	2.0014
7	-1.0006	-0.0007	0.9992	1.9994
8	-0.9997	0.0003	1.0003	2.0003
9	-1.0001	-0.0001	0.9999	1.9999
10	-0.9999	0.0001	1.0001	2.0001
11	-1.0000	-0.0000	1.0000	2.0000

§ 1.6. Theoretical ground of iteration methods of untiing of the systems of linear equations of algebra

Convergence of iteration processes can be led to going out theoretical generals, in particular, on principle squeezing reflections. Will enter the series of theoretical suppositions.

The plural of X is named *metrical space*, if to every pair of elements $\bar{x}, \bar{y} \in X$ an inalienable material number is put in correspondence $\rho(\bar{x}, \bar{y})$ (distance) that satisfies next axioms:

- 1) $\rho(\bar{x}, \bar{y}) > 0$, $\rho(\bar{x}, \bar{y}) = 0$ then and only after, if $x = y$;
- 2) $\rho(\bar{x}, \bar{y}) = \rho(\bar{y}, \bar{x})$ (axiom of symmetry);
- 3) $\rho(\bar{x}, \bar{y}) \leq \rho(\bar{x}, \bar{z}) + \rho(\bar{z}, \bar{y})$ for arbitrary elements $\bar{x}, \bar{y}, \bar{z} \in X$ (inequality of triangle).

The elements of metrical space are named *points*.

Element $\bar{x}^{(0)}$ metrical space of X named границей sequences $\{\bar{x}^{(k)}\}$ points $\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(k)}, \dots$, what X belong, if sequence of distances $\rho(\bar{x}^{(0)}, \bar{x}^{(k)})$ gathers to the zero at $k \rightarrow \infty$ id est

$$\lim_{k \rightarrow \infty} \rho(\bar{x}^{(0)}, \bar{x}^{(k)}) = 0.$$

Sequence $\{\bar{x}^{(k)}\}$ on a plural X can gather or scatter depending on the choice of metric $\rho(\bar{x}, \bar{y})$. Sequence $\{\bar{x}^{(k)}\}$ named *fundamental*, if for an arbitrary number $\varepsilon > 0$ there will be such number $N(\varepsilon)$ that $\rho(\bar{x}^{(k)}, \bar{x}^{(m)}) < \varepsilon$ at $k, m > N(\varepsilon)$. If in metrical space of X every fundamental sequence gathers to some границі that is the element of the same space, then space of X is named *complete*.

Let X and Y are two arbitrary plurals. If to every element $\bar{x} \in X$ it is put in correspondence one and only one element $\bar{y} \in Y$ then it is said that on X a *reflection* (operator) is set And plurals of X in Y and write down $\bar{y} = A\bar{x}$.

Reflection And metrical space of X for itself named *squeezing, or by a clench*, if for arbitrary two points $\bar{x}, \bar{y} \in X$ inequality is executed

$$\rho(A\bar{x}, A\bar{y}) \leq \alpha \rho(\bar{x}, \bar{y}). \quad (2.34)$$

Point \bar{x} named *immobile point of reflection* And, if $A\bar{x} = \bar{x}$. For equations of kind $A\bar{x} = \bar{x}$ a next *theorem* takes place *about an immobile point* (or *principle of squeezing reflections*).

Theorem. Any squeezing reflection certain in complete metrical space of X has one and only one immobile point \bar{x}^* . Sequence $\{\bar{x}^{(k)}\}$ that is determined by equality

$$\bar{x}^{(k+1)} = A\bar{x}^{(k)}, \quad k = 0, 1, 2, \dots, \quad (2.35)$$

gathers to the point \bar{x}^* at the arbitrary choice of the initial approaching $\bar{x}^{(0)} \in X$. Thus, an estimation takes place

$$\rho(\bar{x}^{(*)}, \bar{x}^{(k)}) \leq \frac{\alpha^k}{1-\alpha} \rho(\bar{x}^{(0)}, \bar{x}^{(1)}). \quad (36)$$

The method of simple iteration is based on transformations of the system of equations of algebra $A\bar{x} = \bar{b}$ to the kind

$$\bar{x} = B\bar{x} + \bar{c}. \quad (37)$$

Possibly, that the initial approaching is chosen $\bar{x}^{(0)} = (\bar{x}_1^{(0)}, \bar{x}_2^{(0)}, \dots, \bar{x}_n^{(0)})$ to the exact decision \bar{x} . Certainly, at calculations lay $\bar{x}^{(0)} = \bar{c}$. Will calculate progressive approximations in the method of simple iteration

$$\bar{x}^{(k)} = B\bar{x}^{(k-1)} + \bar{c}, \quad k = 0, 1, 2, \dots. \quad (38)$$

An iteration process (3.38) is named ***consilient*** to the decision \bar{x} systems (2.37), if at the arbitrary choice of the initial approaching $\bar{x}^{(0)}$ a condition is executed

$$\lim_{k \rightarrow \infty} \|\bar{x}^{(k)} - \bar{x}\| = 0.$$

As marked higher, from convergence on a norm покоординатна convergence of progressive approximations swims out also.

Will consider the terms of convergence of метода of simple iteration. Will take (3.37) away from (3.38), will get

$$\bar{x}^{(k)} - \bar{x} = \mathbf{B}(\bar{x}^{(k-1)} - \bar{x}). \quad (39)$$

From a recurrent formula swims out

$$\bar{x}^{(k)} - \bar{x} = \mathbf{B}(\bar{x}^{(k-1)} - \bar{x}) = \mathbf{B}^2(\bar{x}^{(k-2)} - \bar{x}) = \dots = \mathbf{B}^k(\bar{x}^{(0)} - \bar{x}).$$

Thus, vector $\bar{x}^{(k)} \rightarrow \bar{x}$ at $k \rightarrow \infty$ then and only after, if degree of matrix \mathbf{B}^k heads for a zero matrix Θ at $k \rightarrow \infty$.

It is known that for an arbitrary square matrix In matrix $\mathbf{B}^k \rightarrow \Theta$ at $k \rightarrow \infty$ then and only after, when all her own numbers on the module less unit. From this statement swims out necessary and sufficient terms of convergence of метода of simple iteration.

Theorem. Let the system (2.37) have a decision. The method of simple iteration (3.38) gathers at the arbitrary initial approaching $\bar{x}^{(0)}$ to the decision \bar{x} then and only after, if all own numbers of matrix In on the module less unit.

In practice such criterion it is difficult to take advantage of, so as a problem being of own numbers more difficult, than untiing of the linear system. More comfortable than criterion of convergence to use the norm of matrix. Taking into account, that own numbers of matrix In and her norm bound by inequality

$$|\lambda| \leq \|\mathbf{B}\|,$$

will set forth the sufficient terms of convergence : the method of simple iteration (3.38) gathers to the decision of the system (3.37), if the arbitrary concerted norm of matrix is In less unit

$$\| \mathbf{B} \| < 1. \quad (40)$$

If to choose in space \mathbf{R}^n norm of vector $\| \bar{x} \|$ and to enter a metric $\rho(\bar{x}, \bar{y}) = \| \bar{x} - \bar{y} \|$ then a reflection (2.37) will squeeze, if the arbitrary concerted norm of matrix is In less unit, id est a condition (2.40) swims out on principle the compressed reflections.

In obedience to determination of matrix norms that is presented in a paragraph 1.4, the sufficient terms of convergence of метода of simple iteration (38) can be presented as follows:

$$\| \mathbf{B} \|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n | b_{ij} | < 1, \quad (41)$$

$$\| \mathbf{B} \|_m = \max_{1 \leq i \leq n} \sum_{j=1}^n | b_{ij} | < 1,$$

$$\| \mathbf{B} \|_k = \left(\sum_{i=1}^n \sum_{j=1}^n | a_{ij} |^2 \right)^{1/2} < 1.$$

If initial system of equations of algebra $A\bar{x} = \bar{b}$ to erect to to the kind (37) :

$$x_i = - \sum_{j=1, j \neq i}^n \frac{a_{ij}}{a_{ii}} x_j + \frac{b_i}{a_{ii}}, \quad a_{ii} \neq 0, \quad i = 1, 2, \dots, n,$$

then at presence of in a matrix And diagonal prevailing

$$\sum_{j=1, j \neq i}^n |a_{ij}| < |a_{ij}|, \quad i = 1, 2, \dots, n, \quad (42)$$

the condition of convergence of метода of simple iteration is executed, so as $\| \mathbf{B} \|_m < 1$.

For the error of метода of simple iteration concordantly (36) will get an estimation

$$\| \bar{\mathbf{x}} - \bar{\mathbf{x}}^{(k)} \| \leq \frac{\| \mathbf{B} \|}{1 - \| \mathbf{B} \|} \| \bar{\mathbf{x}}^{(k)} - \bar{\mathbf{x}}^{(k-1)} \|. \quad (44)$$

A method coincides with speed of geometrical progression, the denominator of that equals $\| \mathbf{B} \|$.

For the achievement of the set exactness ε id est for implementation of inequalities

$$|x_i - x_i^{(k)}| < \varepsilon, \quad i = 1, 2, \dots, n,$$

an iteration proceeds until terms will not be executed

$$| \bar{\mathbf{x}} - \bar{\mathbf{x}}^{(k)} | \leq \frac{1 - \| \mathbf{B} \|}{\| \mathbf{B} \|} \varepsilon, \quad i = 1, 2, \dots, n. \quad (45)$$

For description of speed of convergence of iteration methods the concept of order of метода is entered. Consider that a method has p th order, if it exists $c_1 > 0$ and $c_2 < \infty$ such, that

$$\rho(\bar{\mathbf{x}}^{(k+1)}, \bar{\mathbf{x}}) \leq c_2 (\rho(\bar{\mathbf{x}}^{(k)}, \bar{\mathbf{x}}))^p$$

$$B_3 = (E - B_1)^{-1} B_2, \quad c_1 = (E - B_1)^{-1} c.$$

Therefore condition of convergence of process of iteration it easily reformulate for this case: the method of Seidel gathers, if arbitrary norm of matrix B_3 less unit. Area of convergence of method of simple iteration and Seidel does not coincide, but intersect. In particular, at implementation of condition (3.42) the method of Seidel gathers. Certainly, the method of Seidel gives more rapid convergence, than method of simple iteration, although so it is not always.

An iteration process is completed in practice, if two progressive approximations differ less than beforehand set ε in obedience to the chosen norm

$$\| \bar{x}^{(k)} - \bar{x}^{(k-1)} \| < \varepsilon.$$

For example: by the method of Seidel to untie the system of equations (example that was examined in a paragraph 1.5)

$$\begin{cases} 9,7x_1 - 10x_2 + 0,6x_3 - 1,6x_4 = -12,3; \\ 1,2x_1 + 11,2x_2 + 1,5x_3 + 2,5x_4 = 5,3; \\ 1,2x_1 - x_2 + 8,5x_3 - 10,8x_4 = -13,7; \\ 0,9x_1 + 2,5x_2 + 1,3x_3 + 12,1x_4 = 24,6. \end{cases}$$

This system over is brought equivalent transformations to the kind

$$\begin{cases} 10,9x_1 + 1,2x_2 + 2,1x_3 + 0,9x_4 = -7; \\ 1,2x_1 + 11,2x_2 + 1,5x_3 + 2,5x_4 = 5,3; \\ 2,1x_1 + 1,5x_2 + 9,8x_3 + 1,3x_4 = 10,3; \\ 0,9x_1 + 2,5x_2 + 1,3x_3 + 12,1x_4 = 24,6. \end{cases}$$

For application of iteration process will present this system as follows

$$\begin{cases} x_1 = \frac{1}{10,9}(-1,2x_2 - 2,1x_3 - 0,9x_4 - 7); \\ x_2 = \frac{1}{11,2}(-1,2x_1 - 1,5x_3 - 2,5x_4 + 5,3); \\ x_3 = \frac{1}{9,8}(-2,1x_1 - 1,5x_2 - 1,3x_4 + 10,3); \\ x_4 = \frac{1}{12,1}(-0,9x_1 - 2,5x_2 - 1,3x_3 + 24,6). \end{cases}$$

In a vectorial-matrix kind the last system can be written down

$$\bar{x} = B\bar{x} + \bar{c},$$

where type of matrix B and vector \bar{c} it is presented in a paragraph 2.5.

In obedience to a theory, equation $\bar{x} = B\bar{x} + \bar{c}$ will present in a kind

$$\bar{x} = \bar{c}_1 + B_3\bar{x},$$

where

$$B_3 = (E - B_1)^{-1}B_2, \quad \bar{c}_1 = (E - B_1)^{-1}\bar{c}, \quad B_1 + B_2 = B.$$

The results of calculations in obedience to the brought equations over on the method of Seidel are driven to the next table 2.2/

Table 2.2

Number Iterations	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$x_4^{(k)}$
1	-1.0678	0.0266	1.0302	1.9964
2	-1.0086	-0.0023	1.0026	2.0009
3	-1.0004	-0.0005	0.9999	2.0002

4	-1.0001	-0.0000	0.9999	2.0001
5	-1.0001	-0.0000	0.9999	2.0001

A task for independent implementation

To this division 30 variants of tasks are driven for independent implementation by students.

Every student elects a variant after the number in the list of group.

A task consists in the decision of the system of four linear equations of algebra with four unknown. Every system of equations needs to be untied by a few methods, what the stated in this methodical manual, and to compare the got results inter se.

To solve the system of equations:

$$1. \begin{cases} 1,7x_1 - 2x_2 + x_3 + 0,9x_4 = 1 \\ x_1 - 3,1x_2 + x_3 - 1,2x_4 = -1,3 \\ x_1 - 2x_2 + x_3 + 5x_4 = 4,9 \\ 3x_2 + 2x_3 - x_4 = 4,2 \end{cases} \quad 2. \begin{cases} 2,1x_1 - 0,9x_2 + x_3 + x_4 = 1,1 \\ 1,5x_1 + 2x_2 - 1,1x_3 + 4x_4 = 2 \\ x_1 + 7x_2 - 4x_3 + 11x_4 = 3 \\ 2x_1 + x_3 - x_4 = 4,1 \end{cases}$$

$$3. \begin{cases} 1,9x_1 + 1,3x_2 - x_3 + x_4 = 1,4 \\ 2,3x_1 - 2,4x_2 + 2x_3 - 3x_4 = 2 \\ 5x_1 + x_2 - x_3 + 2x_4 = -1,1 \\ 2x_1 - x_2 + x_3 - 3x_4 = 4 \end{cases} \quad 4. \begin{cases} 2,3x_1 - 0,9x_2 + 1,1x_3 - 1,7x_4 = 1 \\ 2,8x_1 - x_2 - 3,2x_4 = 2,7 \\ 3x_1 - 1,7x_3 + x_4 = -3,1 \\ 1,9x_1 + 2x_2 - 2,3x_3 + 5,4x_4 = -6,5 \end{cases}$$

$$5. \begin{cases} x_1 - 2,6x_2 + 3,1x_3 - 4,2x_4 = 4 \\ x_2 - 1,3x_3 + x_4 = -3,5 \\ 0,9x_1 + 3x_2 - 3,7x_4 = 1 \\ x_1 - 7x_2 + 3x_3 + 1,9x_4 = -2,9 \end{cases} \quad 6. \begin{cases} 0,9x_1 + 2,1x_2 + 3,7x_3 + 4,2x_4 = 11,5 \\ 2x_1 + 3,6x_2 + 4,4x_3 + 1,2x_4 = 12,4 \\ 3,7x_1 + 4x_2 + 1,8x_3 + 2x_4 = 13,1 \\ 4,1x_1 + 1,2x_2 + 2x_3 + 3,5x_4 = 14,4 \end{cases}$$

$$7. \begin{cases} 1,9x_1 + 2,1x_2 - 1,1x_3 + 5,2x_4 = -1,1 \\ 3,1x_1 - 1,3x_2 + 2,6x_3 - 6,7x_4 = 1,3 \\ 4,5x_1 + 1,1x_2 - 3,3x_3 + 6,6x_4 = 3,1 \\ 0,7x_1 - 2,4x_2 + 4,3x_3 - 6,9x_4 = 4 \end{cases} \quad 8. \begin{cases} 2x_1 + 4x_2 - 5x_3 + 7x_4 = 1 \\ 2x_1 - 5x_2 + 3,3x_3 - 2x_4 = 0 \\ 3,09x_1 + 11x_2 - 13x_3 + 15x_4 = -1 \\ 4x_1 - x_2 + x_3 - 3,1x_4 = 2 \end{cases}$$

$$9. \begin{cases} -1,61x_2 + x_3 - 3,2x_4 = 1,1 \\ 2,1x_1 + x_2 - 1,4x_3 - x_4 = 2,3 \\ 6,8x_1 + 4,3x_2 - 2,1x_3 + 3x_4 = 3,2 \\ -2x_1 + 2,2x_2 + 4x_3 + 4,4x_4 = 0,8 \end{cases}$$

$$10. \begin{cases} 0,9x_1 - x_2 + 3,3x_4 = 8,1 \\ x_1 + x_2 + 2,1x_3 - x_4 = 2,3 \\ 3,8x_1 - 2,6x_2 + 6,3x_3 + 3,9x_4 = 1,6 \\ 2,4x_1 + 4x_2 - 2,6x_3 - 7,3x_4 = 0,5 \end{cases}$$

$$11. \begin{cases} 1,5x_1 + x_2 + 1,7x_3 + x_4 = 7,4 \\ 3x_1 + 2,3x_2 + x_3 + x_4 = -2 \\ x_2 + 2x_3 + 2,6x_4 = 23 \\ 5x_1 + 4x_2 + 3x_3 + 3,7x_4 = 12 \end{cases}$$

$$12. \begin{cases} 1,3x_1 - 2x_2 + x_3 - 1,8x_4 = -1 \\ 2,4x_1 + x_2 - 1,9x_3 + 2x_4 = 3 \\ 3x_1 - 2x_2 - 1,11x_3 + x_4 = 2 \\ 2x_1 - 5,4x_2 + x_3 - 2x_4 = -2,7 \end{cases}$$

$$13. \begin{cases} x_1 - 2x_2 + x_3 + x_4 = 2,2 \\ 2,6x_1 + x_2 - x_3 - 1,3x_4 = -2 \\ x_1 + 7x_2 - 5x_3 - 5x_4 = -10,7 \\ 3x_1 - x_2 - 2x_3 + x_4 = 2 \end{cases}$$

$$14. \begin{cases} x_1 - 1,1x_2 - 1,3x_3 + x_4 = 1,4 \\ -1,8x_1 + x_2 + x_3 - 2,2x_4 = 0 \\ 3,3x_1 - 3x_2 - 3x_3 + 4,5x_4 = 2 \\ 4,9x_1 - 5x_2 - 5x_3 + 7,7x_4 = 3 \end{cases}$$

$$15. \begin{cases} 2,6x_1 - 2,1x_2 + x_3 - x_4 = 1 \\ x_1 + 2,4x_2 - 1,7x_3 + x_4 = 1,2 \\ 4x_1 - 10x_2 + 5,8x_3 - 5,1x_4 = 1 \\ 2x_1 - 14x_2 + 7x_3 - 7,3x_4 = -1 \end{cases}$$

$$16. \begin{cases} 3,3x_1 + x_2 - 2x_3 + x_4 = 1 \\ 2x_1 - 1,1x_2 + 7x_3 - 3x_4 = 2 \\ x_1 + 3x_2 - 2,2x_3 + 5x_4 = 3,3 \\ 3x_1 - 2x_2 + 7x_3 - 5,5x_4 = 3 \end{cases}$$

$$17. \begin{cases} 3,3x_1 + x_2 - 2x_3 + x_4 = 2,4 \\ 2x_1 - 1,1x_2 + 7x_3 - 3x_4 = -3 \\ x_1 + 3x_2 - 2,2x_3 + 5x_4 = 10 \\ 3x_1 - 2x_2 + 7x_3 - 5,5x_4 = -5 \end{cases}$$

$$18. \begin{cases} x_1 + 2,76x_2 - 3x_4 = 1,8 \\ 1,9x_1 - x_2 - 3x_3 + x_4 = 2 \\ 2x_1 - 3x_2 + 4x_3 - 5,6x_4 = 7 \\ 9x_1 - 9x_2 + 6,9x_3 - 16x_4 = 25 \end{cases}$$

$$19. \begin{cases} 1,3x_1 - 2,1x_2 + 3x_3 - 4x_4 = -2 \\ x_1 + 2,4x_2 - 1,7x_3 = -3,47 \\ x_1 - x_2 + 2,4x_3 - 3,3x_4 = 10 \\ 1,5x_2 - x_3 + 1,1x_4 = -5,3 \end{cases}$$

$$20. \begin{cases} 1,9x_2 - 1,6x_3 + x_4 = -1 \\ 2,7x_1 + 3x_2 - x_3 + 1,8x_4 = -3 \\ x_1 + 2,2x_2 - 1,7x_3 = -4,33 \\ 1,3x_1 - x_2 + 2x_3 - 3,7x_4 = 10 \end{cases}$$

$$21. \begin{cases} 4x_1 + 3x_2 + 3,7x_3 + 5x_4 = 10 \\ x_1 + 2x_2 + 2,6x_3 = 12,9 \\ 2,3x_1 + x_2 + x_3 + 3x_4 = -2 \\ x_1 + 1,7x_2 + x_3 + 1,5x_4 = 7,1 \end{cases}$$

$$22. \begin{cases} x_1 - 1,9x_2 + 2x_3 + 2,4x_4 = 2,8 \\ -5,3x_1 + x_2 - 2x_3 + 2x_4 = -2,1 \\ -2x_1 + x_2 - 1,8x_3 + 1,3x_4 = -0,9 \\ -2x_1 - 1,1x_2 + x_3 + 3x_4 = 2,44 \end{cases}$$

$$23. \begin{cases} 1,8x_1 - x_2 + 3x_3 - 2x_4 = 3 \\ x_1 - 2,4x_2 + x_3 + x_4 = 3,2 \\ -5x_1 + 7x_2 + x_3 - 5x_4 = -15 \\ -x_1 + x_2 + 2x_3 - 1,43x_4 = -3 \end{cases}$$

$$24. \begin{cases} 2,7x_1 + x_2 - x_3 - 1,3x_4 = 1 \\ x_1 - 1,9x_2 + 1,6x_3 + x_4 = 0 \\ 3x_1 + 3,3x_2 - 3x_3 - 3,9x_4 = 2 \\ 4,2x_1 + 5,5x_2 - 5x_3 - 5x_4 = 3 \end{cases}$$

$$25. \begin{cases} -2,3x_2 - x_3 + x_4 = -1,77 \\ -2,7x_1 + x_3 - x_4 = 0 \\ 7x_1 - 10x_2 + 5,4x_4 = -3,2 \\ 11x_1 - 14,1x_2 - 7x_3 = -3,8 \end{cases}$$

$$26. \begin{cases} 1,9x_1 - 2x_2 + x_3 - 1,1x_4 = -4 \\ -x_1 + 7,3x_2 - 3x_3 + 7,5x_4 = 0 \\ 3,8x_1 - 2,6x_2 + 5x_3 - 7x_4 = 2,4 \\ -2x_1 + 7x_2 - 5,8x_3 + 8,5x_4 = 0 \end{cases}$$

$$27. \begin{cases} 2,8x_1 - 2,1x_2 + x_3 - x_4 = 0 \\ 2x_1 + 7x_2 - 3,5x_3 + 5,3x_4 = 3 \\ x_1 - 2,8x_2 + 5,5x_3 - 7x_4 = 0,04 \\ 3,3x_1 + 7x_2 - 5x_3 + 8,8x_4 = 5 \end{cases}$$

$$28. \begin{cases} -x_1 - 3x_2 + x_3 - 3,7x_4 = 1 \\ 2x_1 - 3,1x_3 + 2x_4 = 0,67 \\ -3x_1 + 4,9x_2 - 5x_3 + 2x_4 = 5 \\ -9,1x_1 + 6x_2 - 16x_3 + 2x_4 = 16 \end{cases}$$

$$29. \begin{cases} 1,5x_1 - 2,3x_2 + 3x_3 - 4x_4 = -4 \\ x_1 + 2,4x_2 - 1,7x_3 = -2,18 \\ x_1 - x_2 + 2,8x_3 - 3x_4 = 10 \\ 2,6x_1 + 3x_2 - x_3 + 1,3x_4 = -3 \end{cases}$$

$$30. \begin{cases} x_1 + 2,6x_2 + 3x_3 - 1,6x_4 = 1 \\ 3,2x_1 + 2,4x_2 + x_3 - x_4 = 1 \\ 2x_1 + 3,6x_2 + x_3 + 1,3x_4 = 1 \\ 2x_1 + 2x_2 + 2,5x_3 - 1,4x_4 = 1 \end{cases}$$

TITLE 2. APPROXIMATE SOLUTION OF NONLINEAR EQUATIONS

§ 2.2. Introduction

The problem of solution of algebra and transcendent equations often meets at a study of technical and special disciplines, in engineering practice. To find the exact value of the root of the equation is possible only in some cases. Moreover, formulas are so cumbersome that it is very difficult to use them. Therefore, numerical methods are widely used that make it possible to obtain an approximate solution with an arbitrary given accuracy.

Let equation is given

$$f(x) = 0, \quad (1)$$

where $f(x)$ is an algebra or transcendent function with one unknown. The calculation of the real roots of the equation (1) is reduced to finding the set of its roots in the interval at which the equation is transformed into the identity.

If $f(x^*) = 0$ then x^* named *the root of equation* (1). Roots of the given equation are the zeros of function $y = f(x)$ geometrically represent the points of intersection of its graph with the abscissa axis.

For example, will consider equation $f(x) = x^2 - 2 = 0$. On figure 1 the graph of this function is presented. Intersections of curve $y = x^2 - 2$ with an axis OX, in that $f(x) = 0$, are the roots of equation. As we can see, roots are on segments $[-1,5; -1]$ and $[1; 1,5]$ thus the values of roots can be defined only approximately.

The problem of finding the approximate roots of an equation with arbitrary given accuracy consists of two stages:

- 1) separation (isolation) of root, that is, finding a segment $[a; b]$ that belongs to the domain of definition of the function $y = f(x)$ and on which there is one and only one root of equation $f(x) = 0$;
- 2) a calculation or clarification of value of root with the set exactness.

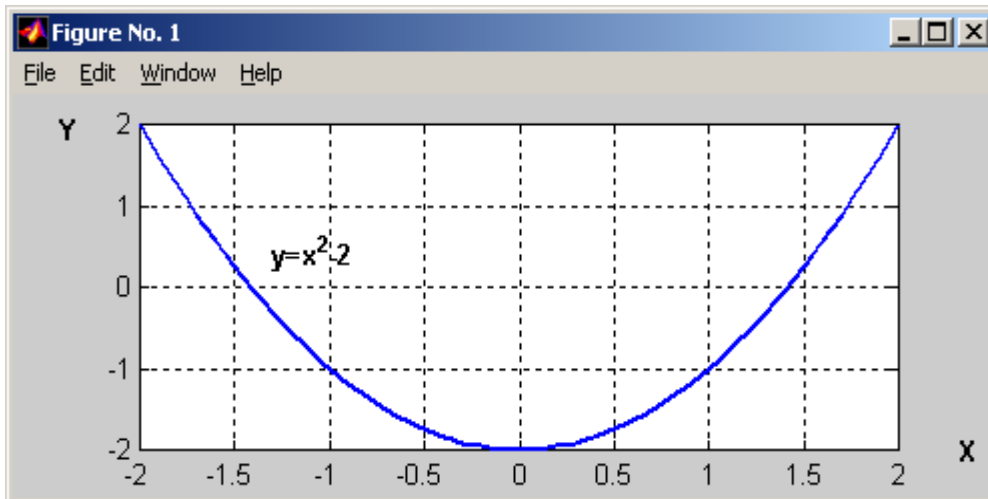


Figure 1.

§ 2.2. Separation of roots of equation

2.2.1. Root separation conditions

Let function $f(x)$ in equation (1) certain and continuous on some interval $(\alpha; \beta)$ and has continuous first $f'(x)$ and second $f''(x)$ derivatives.

The task is reduced to finding a segment of the domain of definition of the function on which three conditions are satisfied

The problem of separation of root of equation (1) is reduced to finding a segment $[a; b]$ of the domain of definition of the function on which three conditions are satisfied :

- 1) function $f(x)$ at the borders of segment $[a; b]$ has different signs $f(a) \cdot f(b) < 0$;

2) derivative $f'(x)$ does not change a sign on $[a; b]$ it means the function $f(x)$ is monotonous on $[a; b]$;

second derivative of function $f''(x)$ does not change a sign on $[a; b]$ it is a function $f(x)$ keeps a curvature or bulge

3) second derivative of function $f''(x)$ does not change a sign on $[a; b]$ it is a function $f(x)$ keeps a curvature or bulge on $[a; b]$.

Segment $[a; b]$ at implementation of terms 1-3 for a function $f(x)$ named a segment that separates the root of the given function.

In general case there is not an algorithm for the separation of root of equation $f(x) = 0$. For the separation of real root use a graphic method or make the table of values of function $f(x)$ on some interval (the change of signs in two nearest lines of table testifies to the presence even one root). For application of modern computer packages the special commands of graphic representation are used, that allows it easily enough to separate the roots of initial equation.

2.2.2 Graphic method of separation of root

Graphicly roots of equation $f(x) = 0$ it is possible to separate, if to build the graph of function $y = f(x)$ and approximately to define the points of it crossing with an axis OX. But a problem of construction of graphic is not always simple. Usually equation $f(x) = 0$ is replaced by equivalent equation $\varphi_1(x) = \varphi_2(x)$ ($f(x) = \varphi_1(x) - \varphi_2(x)$), and select functions $y_1 = \varphi_1(x)$ and $y_2 = \varphi_2(x)$ so that it is easier to build their graphic than graphic of function $y = f(x)$. Abscissas of intersections of graphics $y_1 = \varphi_1(x)$ and $y_2 = \varphi_2(x)$ are the required roots of initial equation.

For example: by a graphic method to separate the roots of equation

$$e^{-x} + x^2 - 2 = 0.$$

Will present the set equation in a type $e^{-x} = 2 - x^2$ and will consider two functions $\varphi_1(x) = e^{-x}$ and $\varphi_2(x) = 2 - x^2$. Intersections of graphic of these functions are the roots of the set equation.

As evidently from picture 2, the given equation has two real root (graphics intersect in two points), thus one of root is negative, and second is positive. These roots are in intervals $x_1 \in (-\sqrt{2}; 0)$ but $x_2 \in (0; \sqrt{2})$.

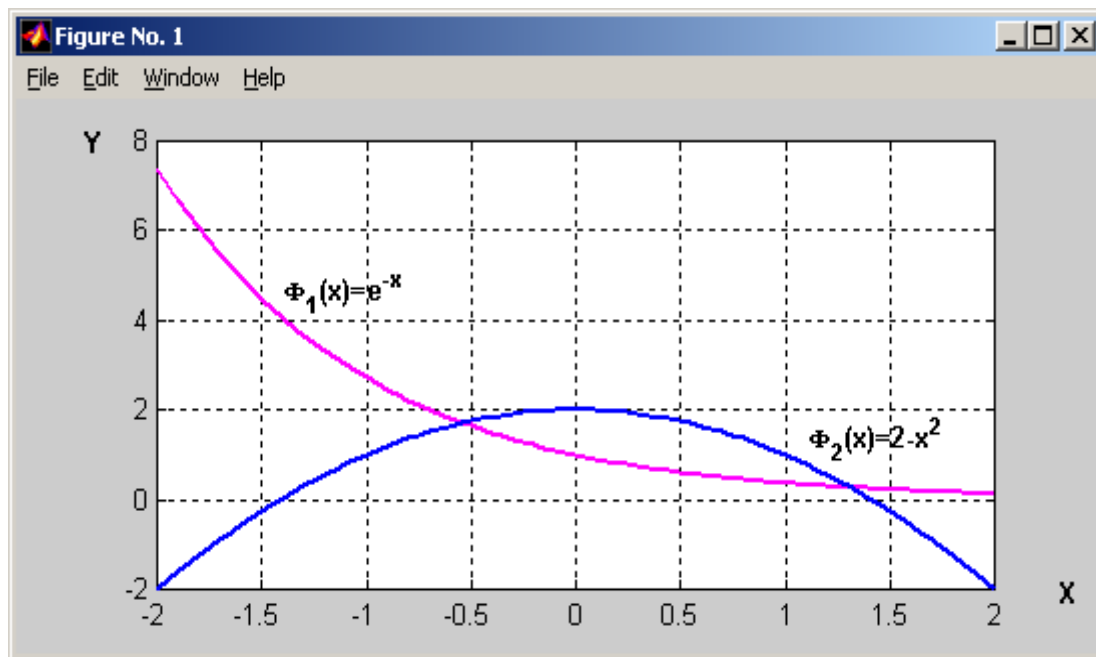


Figure 2

2.2.3 Method of tests

The method of tests consists in that at random gets out point $x = a$ from an area determination of function (or from more narrow area), there is a sign of function $f(a)$ and then the point of b sneaks up thus, that value of function $f(b)$ small sign opposite to the sign $f(a)$. A sign is farther determined $f'(x)$ on a segment $[a; b]$. If $f'(x)$ does not change a sign on $[a; b]$ then root separated, there

is a segment in another case $[a; b]$ narrow, taking the point of c , that lies inwardly відрізка $[a; b]$. A sign is determined $f(c)$ and for a new segment examined or $[a; c]$ (if $f(a) \cdot f(c) < 0$) or $[c; b]$ (if $f(c) \cdot f(b) < 0$). Designating a new segment through $[a_1; b_1]$ repeat the same operations, that and on a segment $[a; b]$ and т. д. The indicated operation is closed, if the executed terms of separation of root on a corresponding segment - $[a_n; b_n]$.

For example: by the method of tests to separate the root of equation to positive

$$x^4 + x^3 - 36x - 20 = 0.$$

Function $f(x) = x^4 + x^3 - 36x - 20$ certain on all numerical line. As it is needed to separate the root of equation to positive, will consider півінтервал $[0; \infty)$.

1. Find $f(0) = -20 < 0$. Then choose an arbitrary point, for example $x = 1$ and calculate $f(1) = -54 < 0$. So as, $f(0) \cdot f(1) > 0$ then pass to the next point. Pick up a point $x = b$ thus, that a condition was executed $f(b) > 0$. Let $x = 4$ then $f(4) = 156 > 0$ from where swims out, that on a segment $[1; 4]$ there is a root ($f(1) \cdot f(4) < 0$).

2. As $f'(x) = 4x^3 + 3x^2 - 36 = 4(x^3 - 9) + 3x^2$ then make sure direct verification, that on a segment $[1; 4]$ derivative $f'(x)$ changes a sign ($f'(1) = -29 < 0$; $f'(4) = 268 > 0$).

Narrow a segment $[1; 4]$. Will take, for example point $x = 3$. Then $f(3) = -20 < 0$ and $f(3) \cdot f(4) < 0$ from it swims out, that on a segment $[3; 4]$ there is a root. Check up a sign $f'(x)$. Have $f'(3) = 69 > 0$ and for $x > 3$ obviously, a derivative grows, that is why it remains positive. Thus, root separated. On a

segment $[3; 4]$ it is positive actual root of the set equation. Will mark, that $f''(x) = 12x^2 + 6x > 0$ for $x \in [3; 4]$. Chart $y = f(x)$ it is presented on rice. 3.

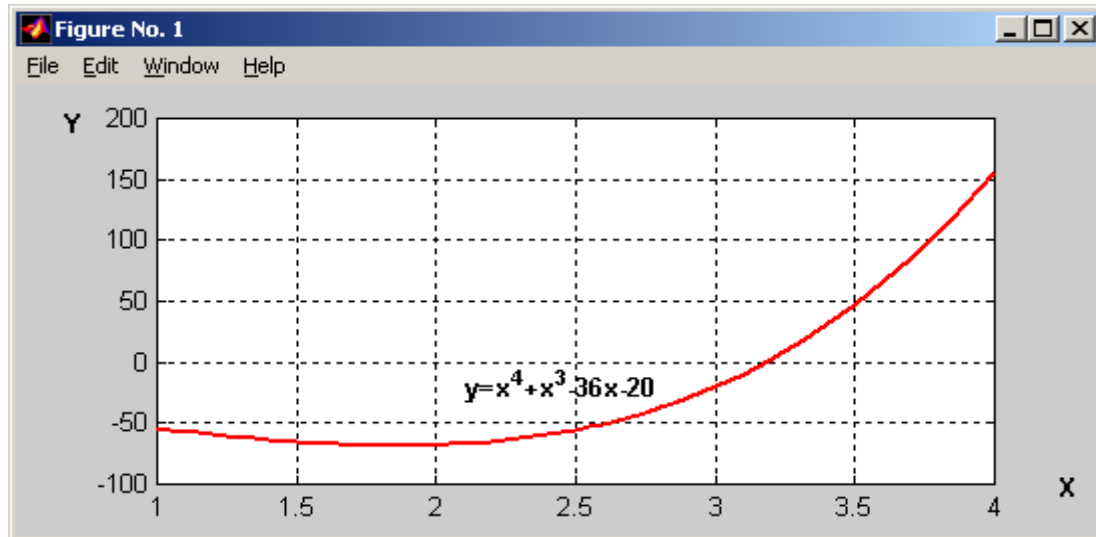


Figure 3.

§ 2.3. Method of half-note division

Let function $f(x)$ certain and continuous at all $x \in [a; b]$ and on $[a; b]$ changes a sign, id est $f(a) \cdot f(b) < 0$. From it swims out, that equation $f(x) = 0$ it is had on $(a; b)$ even one root. Will take an arbitrary point $c \in (a; b)$. In this case will name a segment $[a; b]$ **by the interval of existence of root, and point of c - by a trial point**. As here the question is only about the actual functions of the real variable, then calculation of value $f(c)$ will result in that - **небудь** one of next mutually exceptional situations :

$$\text{a) } f(a) \cdot f(c) < 0; \quad \text{б) } f(c) \cdot f(b) < 0; \quad \text{в) } f(c) = 0.$$

In relation to a problem that is examined they can be interpreted thus :

- a) root is on an interval (a; c);
- b) root is on an interval (c; b);
- c) point of c is the sought after root.

The most used case of part of метода dichotomy (from the Greek word that means a bipartioning) is *a method of half-note division*, that will realize the simplest method of choice of trial point - division of interval of existence of root in half. To execute the close calculation of root of equation $f(x) = 0$ with exactness ε by the method of half-note division on condition that $f(x)$ continuous on $[a; b]$ and $f(a) \cdot f(b) < 0$ it is possible for example, on a next chart:

Step 0. To set the ends of відрізка but also b , function of f , small number $\varepsilon > 0$ (admissible absolute error of root or півдовжину of his interval of vagueness), small number $\delta > 0$ (admittance related to the real exactness of calculation of values of the set function).

Step 1. To calculate $c := 0.5(a + b)$.

Step 2. If $b - a < 2\varepsilon$ to put $\xi := c$ (ξ it is a root) and to stop.

Step 3. To calculate $f(c)$.

Step 4. If $f(c) < \delta$ to put $\xi \approx c$ and to stop.

Step 5. If $f(a) \cdot f(c) < 0$ to put $b := c$ and to go back to a step 1; otherwise to put $a := c$, $f(a) := f(c)$ and to go back to a step 1.

For one step of method of half-note division the interval of existence of root grows short exactly twice. To Tom, if after k th approaching by this method to Cornu ξ equation $f(x) = 0$ will take a point x_k that is a middle got on k th step of відрізка $[a_k; b_k]$ as a result of the successive narrowing of this відрізка $[a; b]$ will put $a_1 := a$, $b_1 := b$ then will come to inequality

$$|\xi - x_k| < \frac{b - a}{2^k} \quad \forall k \in \mathbb{N} \quad (2)$$

(a priori ξ it is an arbitrary point of interval $(a_k; b_k)$ and distance from her to the middle of this interval does not exceed the half of his length. It is visible from (4.2) at $k = 1$).

Inequality (2), from one side, allows to assert that sequence (x_k) has a limit - sought after root ξ equation $f(x)=0$; on the other hand, being a priori estimation of absolute error of close equality $x_k \approx \xi$ gives an opportunity to count up the number of steps (iterations) of метода of half-note division, sufficient for the receipt of root ξ with the set exactness ε for what it is needed only to find least natural k , that satisfies inequalities

$$\frac{b-a}{2^k} < \varepsilon, \quad (3)$$

namely:

$$k > \frac{\lg |b-a|}{\lg 2} + 1, \quad \text{or} \quad k = \left\lceil \frac{\lg |b-a|}{\lg 2} + 1 \right\rceil. \quad (4)$$

For example: to find the root of equation

$$x^3 + 3x^2 - 1 = 0$$

with the set exactness. Initial function can be presented in a kind $\varphi_1(x) = \varphi_2(x)$ where $\varphi_1(x) = x^3$, $\varphi_2(x) = 1 - 3x^2$. From the construction of charts $y_1 = \varphi_1(x)$ but $y_2 = \varphi_2(x)$ swims out, that the sought after root is on a segment $[0, 1]$ it is figure 4.

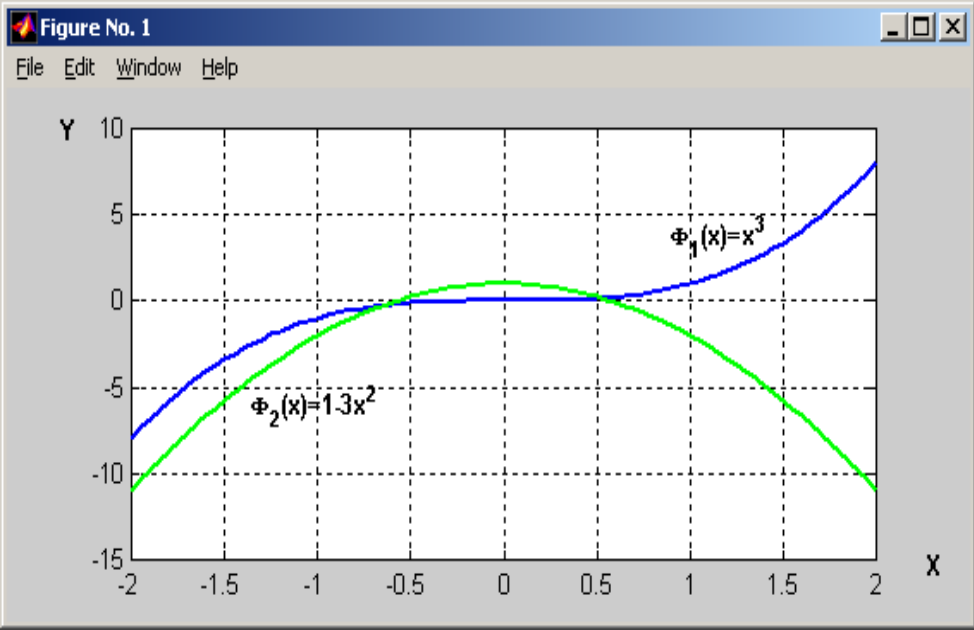


Figure 4.

For the calculation of number k progressive approximations will take advantage of formula (4) in obedience to that it is necessary to conduct for the receipt of the set exactness $k = 21$ progressive approximations. For the receipt of results with exactness $\varepsilon = 10^{-5}$ it is necessary to conduct $k = 18$ approaching, with exactness $\varepsilon = 10^{-4}$ - $k = 14$ approaching, with exactness $\varepsilon = 10^{-3}$ - $k = 11$ approaching.

The results of calculations are driven to the table 2.1.

Table 2.1

n	x_n	$f(x_n)$
1	0,5	-0,125

2	0,75	1,1094
3	0,625	0,4160
4	0,5625	0,1272
5	0,5313	-0,0034
6	0,5469	0,0608
7	0,5391	0,0284
8	0,5352	0,0124
9	0,5332	0,0045
10	0,5322	5,5656e- 4
11	0,5317	-0,0014
12	0,5320	-4,3027e- 4
13	0,5321	6,3079e- 5
14	0,5320	-1,8361e- 4
15	0,5321	-6,0270e- 5
16	0,5321	1,4032e- 6
17	0,5321	-2,9434e- 5
18	0,5321	-1,4015e- 5
19	0,5321	-6,3061e- 6
20	0,5321	-2,4515e- 6
21	0,5321	-5,2414e- 7

§ 2.4. Method of chords

The method of jiggling of trial point is used in the method of half-note division it is possible to describe as passive, as he comes true after the beforehand set hard plan and in any way does not take into account the values of function calculated at every step. Logically to assume that in family of methods of

dichotomy it is possible to attain the best results, if segment $[a; b]$ to divide the point of c to pieces not in half, but proportionally to the sizes of ordinates $f(a)$ and $f(b)$ graphic arts of the set function $f(x)$. It means that the point of c can be found as an abscissa of intersection to the landmark OH straight-in, that passes through points $A(a; f(a))$ and $B(b; f(b))$ otherwise, with the chord of AB of arc.

The idea of метода chords consists in that on a segment $[a; b]$ the chord of AB, that tightens the ends of arc of chart of function, is built $y = f(x)$ and for the close value of root x_0 a number gets out $c = c_1$ that is the abscissa of intersection of this chord with an axis OH. For determination of number c_1 will lay down equation of chord as a line that passes through two points $A(a; f(a))$ and $B(b; f(b))$:

$$\frac{x - a}{b - a} = \frac{y - f(a)}{f(b) - f(a)}.$$

At $y = 0$; $x = c_1$ get

$$c_1 = a - \frac{f(a)(b - a)}{f(b) - f(a)} \quad \text{or} \quad c_1 = b - \frac{f(b)(b - a)}{f(b) - f(a)} \quad (5)$$

Number c_1 accept for the first approaching to the sought after Cornu. The schematically indicated procedure is presented on rice. 4.6.

Obviously, that at the accepted suppositions about the signs of derivatives $f'(x)$ and $f''(x)$ on $[a; b]$ point $(c_1; 0)$ it will be from the side of вгнутости crooked and will divide $[a; b]$ on two segments $[a; c_1]$ and $[c_1; b]$ there is a root in one of that x_0 (see rice. 4.6). A new segment that separates a root can be defined comparing signs $f(a), f(c_1), f(b)$. From the analysis of lines. 4.6

evidently, that point c_1 nearer to the point and, than x_0 if $y'y'' > 0$ then a segment that separates a root will be $[c_1; b]$; in another case, if $y'y'' < 0$ that separates a root a segment, it will be $[a; c_1]$.

Farther repeat the same procedure on a new segment that separates a root, and determine a number c_2 (second approaching) after the formulas got from (5) :

$$c_2 = c_1 - \frac{f(c_1)(b - c_1)}{f(b) - f(c_1)} \quad (y'y'' > 0), \tag{6}$$

$$c_2 = c_1 - \frac{f(c_1)(c_1 - a)}{f(c_1) - f(a)} \quad (y'y'' < 0).$$

After being C_2 find C_3 and so on (Figure 5).

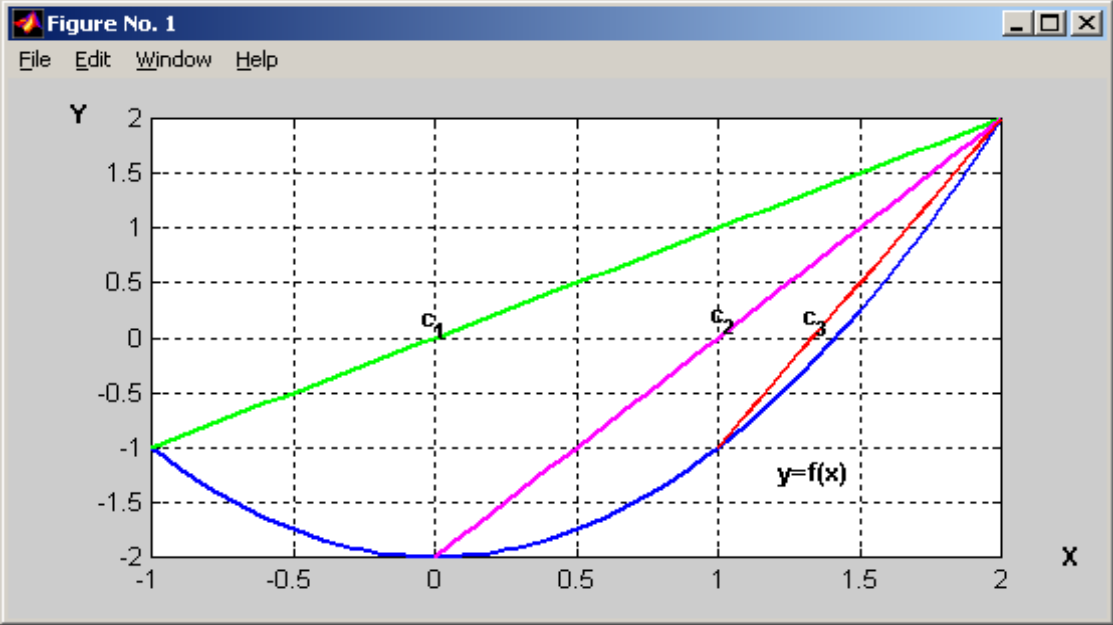


Figure 5.

End a process then, when the estimation of the got approaching satisfies the set exactness.

For simplification of calculations usually set some small enough number $\varepsilon > 0$ (no more set exactness). A process is closed then, when an absolute value of difference is between two next approaching c_{n-1} and c_n less than ε :

$$|c_{n-1} - c_n| < \varepsilon.$$

Number c_n accept for the close value of root, id est $\bar{x} = c_n$.

§ 2.5. Method of simple iteration

Equation is examined $f(x) = 0$ on a segment $[a; b]$. Laid, that on a segment $[a; b]$ there is one and only one root.

Will replace equation $f(x) = 0$ by equivalent to him equation

$$x = \varphi(x). \tag{7}$$

Will mark that equation $f(x) = 0$ it is possible to replace equivalent to him equation (7), for example, putting many methods $\varphi(x) = x + \psi(x)f(x)$ where $\psi(x)$ it is an arbitrary continuous знакостала function.

Set by some initial approaching x_0 next approaching to Cornu x^* find after a formula

$$x_{n+1} = \varphi(x_n), \quad n = 0, 1, \dots \tag{8}$$

Formulas (8) is *a method of simple iteration (or method of progressive approximations)*.

If sequence of approaching $\{x_n\}$ coincides then there is границя

$\lim_{n \rightarrow \infty} x_{n+1} = x^*$, coming from what for a continuous function $\varphi(x)$ it is possible to write down

$$\lim_{n \rightarrow \infty} x_{n+1} = \varphi(\lim_{n \rightarrow \infty} x_n), \quad \text{or} \quad x^* = \varphi(x^*).$$

Decision x^* equation (4.7) is the immobile point of mapping $\varphi(x)$. Thus, the terms of convergence of iteration process (3.8) can be got on principle squeezing mapping (see the division of III).

Will formulate the sufficient terms of convergence of method of simple iteration.

Theorem. Let function $\varphi(x)$ it is certain and differentiated on a segment $[a, b]$ thus all her values $\varphi(x) \in [a, b]$. Then, if there is such number of q , that on this segment

$$|\varphi'(x)| \leq q < 1, \quad (9)$$

then a sequence (3.8) coincides to only on a segment $[a, b]$ upshot of equation (3.7) for any initial approaching $x_0 \in [a, b]$. Thus, if on the noted segment derivative $\varphi'(x) > 0$,

$$|x_n - x_*| < \frac{q}{1-q} |x_n - x_{n-1}|; \quad (10)$$

if derivative $\varphi'(x) < 0$ then

$$|x_n - x_*| < |x_n - x_{n-1}|. \quad (11)$$

These conclusions are summarized on more wide class of functions, which satisfy the condition of Lipschitz with a constant, less from unit. If a condition (8) is not executed, then iterations (3.9) can scatter.

At every step iterations calculate a value $y = \varphi(x_n)$ and, if $|y - x_n| \geq \varepsilon$, then, putting $x_n = y$ pass to the next iteration. If $|y - x_n| < \varepsilon$, then, calculations stop and as a root is accepted by approaching x_n . The error of the found result depends on a derivative sign $\varphi'(x)$. At $\varphi'(x) > 0$ the error of determination of root presents $q\varepsilon/(1 - q)$; if $\varphi'(x) < 0$ then an error does not exceed ε .

At application of method of simple iteration one of complexity there is bringing equation over $f(x) = 0$ to the kind $x = \varphi(x)$ thus, that the terms of convergence of iteration process were executed. Will consider one of general approaches of bringing initial equation over $f(x) = 0$ to the equivalent kind $x = \varphi(x)$. Let the sought after root ξ initial equation is on a segment $[a, b]$ thus

$$0 < m_1 \leq f'(x) \leq M_1, \quad \text{at } x \in [a, b]. \quad (12)$$

In particular, after m_1 it is possible to take on a the least value of derivative $f'(x)$ on $[a, b]$ which must be positive. After M_1 it is possible to take on a most value of derivative $f'(x)$ on $[a, b]$. Will replace equation $f(x) = 0$ by equivalent to him equation

$$x = x - \lambda f(x), \quad \lambda > 0.$$

It is thus possible to put $\varphi(x) = x - \lambda f(x)$. Parameter λ sneaks up thus, that in околі $[a, b]$ root ξ inequality was executed

$$0 \leq \varphi'(x) = 1 - \lambda f'(x) \leq q < 1.$$

Last inequality with the use of formula (4.12) it is possible to write down :

$$0 \leq 1 - \lambda M_1 \leq 1 - \lambda m_1 \leq q.$$

From where swims out :

$$\lambda = \frac{1}{M_1} \quad \text{and} \quad q = 1 - \frac{m_1}{M_1} < 1.$$

In practical calculations number M_1 gets out thus, that $|M_1| > Q/2$, where $Q = \max |f'(x)|$ for all $x \in [a, b]$. Thus an iteration process coincides subject to condition $|\varphi'(x)| < 1$ on $[a, b]$.

Will consider the example of application of method of simple iteration for solution of equation

$$f(x) \equiv x^3 - x - 1 = 0.$$

An initial function can be presented in a kind $\varphi_1(x) = \varphi_2(x)$ where $\varphi_1(x) = x^3$, $\varphi_2(x) = x + 1$. From the construction of charts $y_1 = \varphi_1(x)$ but $y_2 = \varphi_2(x)$ swims out, that the sought after root is on a segment $[1, 2]$ it is figure 6.

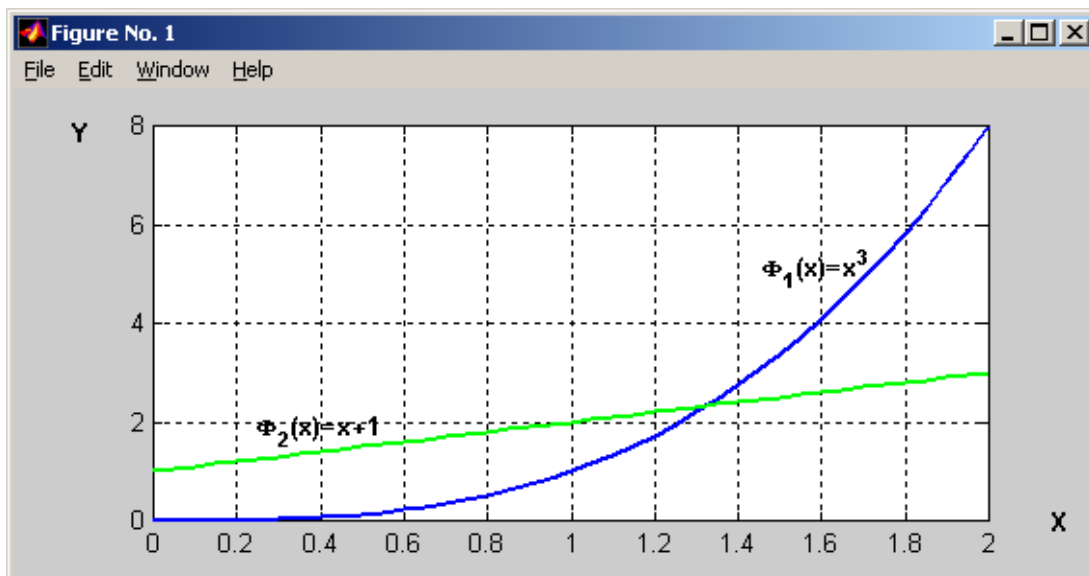


Figure 6.

Initial equation $x^3 - x - 1 = 0$ it is possible to present in a kind $x = x^3 - 1$ or $x = \varphi(x)$ where $\varphi(x) = x^3 - 1$. In the presented kind function $\varphi(x)$ dissatisfies to the terms of convergence, so as $\varphi'(x) = 3x^2$ and $\varphi'(x) \geq 3$ on a segment $[1, 2]$.

Will present initial equation in a kind $x = \sqrt[3]{x+1}$. In this case function

$$\varphi(x) = \sqrt[3]{x+1}, \quad \text{and} \quad \varphi'(x) = \frac{1}{3\sqrt[3]{(x+1)^2}}.$$

From swims out here, that $\frac{1}{3\sqrt[3]{9}} \leq \varphi'(x) \leq \frac{1}{3\sqrt[3]{4}}$ on a segment $[1, 2]$ or

$\varphi'(x) < \frac{1}{4}$, what satisfies to the terms of convergence of iteration process.

The results of calculations are driven to the table 2.2.

Table 2.2

n	0	1	2	3	4	5	6	7
x_n	1	1,2599	1,3123	1,3224	1,3243	1,3246	1,3247	1,3247

§ 2.6. Method of Newton

Let in equation $f(x) = 0$ function $f(x)$ has continuous second derivative $f''(x)$ on a segment $[a; b]$ which the separated root is on x^* . It is assumed that derivatives $f'(x)$ but $f''(x)$ different from a zero, знакосталі and initial approaching of root x_0 it is belonged to the segment $[a; b]$. In obedience to the

indicated requirements on a segment $[a; b]$ absent extremums and in flection points of initial function, that allows in any point of segment $[a; b]$ to build a tangent to the curve. Will consider arbitrary point-on-wave $M_0(x_0, y(x_0))$, $x_0 \in [a; b]$ and will conduct a tangent to the curve in this point. Equation of tangent looks like

$$y - f(x_0) = f'(x_0)(x - x_0).$$

A tangent crosses abscise axis in some point $(x_1; 0)$. Taking into account the last, it is possible to write down

$$-f(x_0) = f'(x_0)(x - x_0),$$

from where get the first approaching of root

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Through a point $M_1(x_1, f(x_1))$ again will conduct a tangent to the curve, the intersection of which with abscise axis gives the second approaching of root and т. д.

Described process of construction of tangents a calculation of points of their crossing with abscise axis is an iteration process of Newton - Рафсона.

The construction of iteration sequence of approaching of chums takes place in obedience to next formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots \quad (13)$$

Geometrical maintenance of метода Newton – Rafson consists in substituting of arc crooked by every iteration a tangent to her in a point x_n .

Method of Newton - Rafson can be examined as a partial case of method of simple iteration (3.8), if to put $\varphi(x) = x - \frac{f(x)}{f'(x)}$. As

$$\varphi'(x) = \frac{f(x)f''(x)}{(f'(x))^2},$$

from the condition of convergence of method of simple iteration swims out, that method of Newton - Rafson coincides, when the initial approaching is chosen by near enough to simple root x^* ($\varphi'(x^*) = 0$).

Thus, if initial approaching x_0 it is chosen close enough to root x^* then the method of Newton always coincides. At the arbitrary initial approaching an iteration method (3.13) coincides, if for all $x \in [a; b]$ inequality is executed

$$\frac{|f(x)f''(x)|}{[f'(x)]^2} < 1. \quad (14)$$

A next theorem takes place thus.

Theorem. Let $f(x) \in C_2 [a; b]$ (function $f(x)$, derivatives $f'(x)$ i $f''(x)$ continuous on $[a; b]$), $f(a)f(b) < 0$ and derivatives $f'(x)$, $f''(x)$ keep a sign on a segment $[a; b]$. Then, if initial approaching $x_0 \in [a; b]$ satisfies inequality

$$f(x_0)f''(x_0) > 0, \quad (15)$$

a sequence (4.13) coincides (thus droningly) to only on a segment $[a; b]$ root x^* equation $f(x) = 0$.

For the estimation of speed of convergence of method around the root point will take advantage of formula of Taylor

$$f(x^*) = f(x_n) + f'(x_n)(x^* - x_n) + \frac{1}{2}f''(c)(x^* - x_n)^2,$$

or
$$f(x_n) + f'(x_n)(x^* - x_n) + \frac{1}{2}f''(c)(x^* - x_n)^2 = 0,$$

where a point is $x = c$ lies between x_n i x^* . From the last equation have

$$x^* - x_n + \frac{f(x_n)}{f'(x_n)} = -\frac{1}{2} \frac{f''(c)}{f'(x_n)} (x^* - x_n)^2. \quad (16)$$

In obedience to a formula (4.13)

$$x_n - \frac{f(x_n)}{f'(x_n)} = x_{n+1},$$

to the volume

$$x^* - x_{n+1} = -\frac{1}{2} \frac{f''(c)}{f'(x_n)} (x^* - x_n)^2. \quad (17)$$

If to designate a most value through $M |f''(x)|$ on $[a; b]$ and through m is the least value $|f'(x)|$ on a segment $[a; b]$ then it is possible to write down next inequality for the estimation of error of two progressive approximations x_n but x_{n+1}

$$|x^* - x_{n+1}| < \frac{M}{2m} (x^* - x_n)^2. \quad (18)$$

In obedience to an estimation (4.18) swims out, that the error of the next new approaching diminishes proportionally to the square of error previous, id est convergence of method of Newton - Rafson is quadratic.

If two progressive approximations are known x_n but x_{n+1} in obedience to the method of Newton - Rafson, then it is possible to write down on the basis of correlation (4.18)

$$|x^* - x_{n+1}| < \frac{M}{2m} (x_{n+1} - x_n)^2. \quad (19)$$

Thus, for determination of root of equation (4.1) in obedience to the method of Newton - Рафсона with the set exactness ε an iteration process will be executed, while

$$|x_{n+1} - x_n| \leq \sqrt{2m\varepsilon / M}. \quad (20)$$

Remark. If derivative $f'(x)$ poorly changes on $[a; b]$ and her calculation is bulky enough, then an iteration process can be conducted on a formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}, \quad n = 0, 1, 2, \dots. \quad (21)$$

The iteration process of kind (3.21) has the name *of the modified method of Newton*.

Will consider the example of application of method of Newton for untiing of equation

$$y \equiv \operatorname{tg}(0.3x + 0.4) - x^2 = 0.$$

An initial function can be presented in a kind $\varphi_1(x) = \varphi_2(x)$ where $\varphi_1(x) = \operatorname{tg}(0.3x + 0.4)$, $\varphi_2(x) = x^2$. From the brought graphic material over evidently, that graphic arts of functions $y_1 = \varphi_1(x)$ but $y_2 = \varphi_2(x)$ intersect in two points, roots belong to the intervals $[-1, 0)$ but $(0, 1]$ it is figure 7.

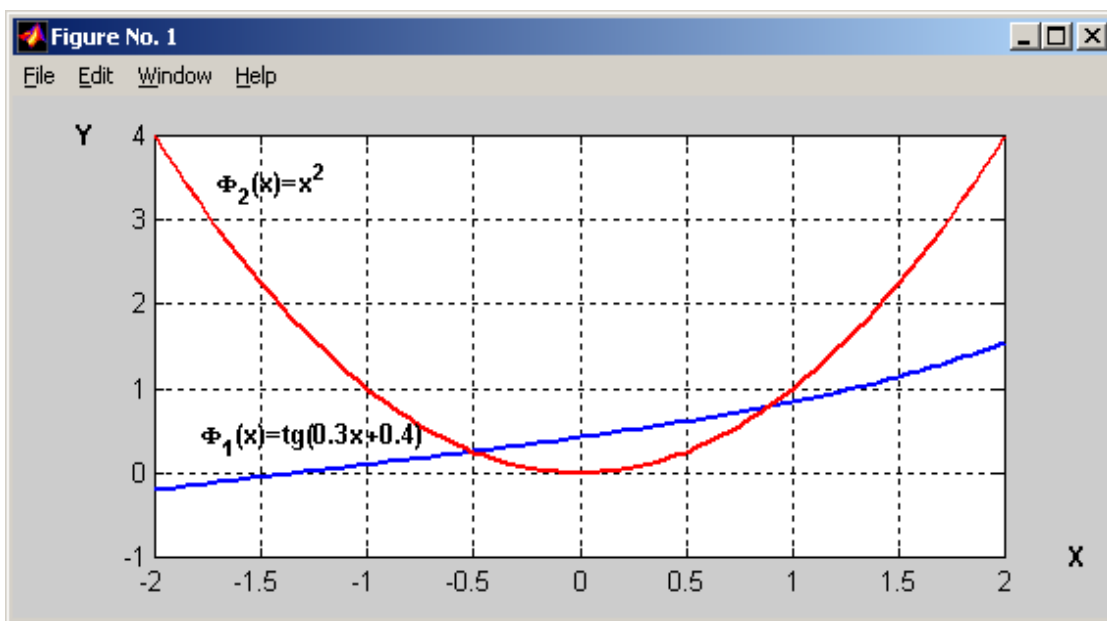


Figure 7.

Stopped for being of positive root.

$$y'(x) = \frac{0.3}{\cos^2(0.3x + 0.4)} - 2x, \quad y''(x) = \frac{0.18 \sin(0.3x + 0.4)}{\cos^3(0.3x + 0.4)} - 2.$$

Coming from the presented chart it is expedient to accept for the initial approaching $x_0 = 1$. Thus $y(x_0) = -0,1577$; $y''(x_0) = -1,7408$ and

accordingly $y(x_0) \cdot y''(x_0) > 0$. The results of calculations are driven to the next table 2.3

Table 2.3

n	0	1	2	3	4
x_n	1	0,8940	0,8864	0,8863	0,8863

Will consider application of method of Newton to the decision of next equation $f(x) \equiv x^3 - x - 1 = 0$ (this problem was examined at application of simple iteration method in a previous paragraph). A root of equation is on a segment $[1, 2]$ it is figure 6. First and second the derivatives of initial function are written down

$$f'(x) = 3x^2 - 1; \quad f''(x) = 6x.$$

A point sets to the initial approaching $x_0 = 2$. Thus $f(x_0) \cdot f''(x_0) > 0$. The results of calculations are driven to the table 2.4.

Table 2.4

n	0	1	2	3	4
x_n	2	1,5455	1,3258	1,3247	1,3247

The brought results over in a table 3.4 and comparing to the corresponding results according to calculations on the method of simple iteration (table 4.2) allow to draw conclusion about efficiency of application of метода Newton.

§ 2.7. Method of simple iteration

Equation is examined $f(x) = 0$ on a segment $[a; b]$. Laid, that on a segment $[a; b]$ there is one and only one root.

Will replace equation $f(x) = 0$ by equivalent to him equation

$$x = \varphi(x). \quad (7)$$

Will mark that equation $f(x) = 0$ it is possible to replace equivalent to him equation (7), for example, putting many methods $\varphi(x) = x + \psi(x)f(x)$ where $\psi(x)$ it is an arbitrary continuous знакостала function.

Set by some initial approaching x_0 next approaching to Cornu x^* find after a formula

$$x_{n+1} = \varphi(x_n), \quad n = 0, 1, \dots \quad (8)$$

Formulas (8) is *a method of simple iteration (or method of progressive approximations)*.

If sequence of approaching $\{x_n\}$ coincides then there is границя $\lim_{n \rightarrow \infty} x_{n+1} = x^*$, coming from what for a continuous function $\varphi(x)$ it is possible to write down

$$\lim_{n \rightarrow \infty} x_{n+1} = \varphi(\lim_{n \rightarrow \infty} x_n), \quad \text{or} \quad x^* = \varphi(x^*).$$

Decision x^* equation (4.7) is the immobile point of mapping $\varphi(x)$. Thus, the terms of convergence of iteration process (8) can be got on principle squeezing mapping.

Will formulate the sufficient terms of convergence of method of simple iteration.

Theorem. Let function $\varphi(x)$ it is certain and differentiated on a segment $[a, b]$ thus all her values $\varphi(x) \in [a, b]$. Then, if there is such number of q , that on this segment

$$|\varphi'(x)| \leq q < 1, \quad (9)$$

then a sequence (3.8) coincides to only on a segment $[a, b]$ upshot of equation (7)

for any initial approaching $x_0 \in [a, b]$. Thus, if on the noted segment derivative $\varphi'(x) > 0$,

$$|x_n - x_*| < \frac{q}{1-q} |x_n - x_{n-1}|; \quad (10)$$

if derivative $\varphi'(x) < 0$ then

$$|x_n - x_*| < |x_n - x_{n-1}|. \quad (11)$$

These conclusions are summarized on more wide class of functions, which satisfy the condition of Lipschitz with a constant, less from unit. If a condition (4.8) is not executed, then iterations (9) can scatter.

At every step iterations calculate a value $y = \varphi(x_n)$ and, if $|y - x_n| \geq \varepsilon$, then, putting $x_n = y$ pass to the next iteration. If $|y - x_n| < \varepsilon$, then, calculations stop and as a root is accepted by approaching x_n . The error of the found result depends on a derivative sign $\varphi'(x)$. At $\varphi'(x) > 0$ the error of determination of root presents $q\varepsilon/(1-q)$; if $\varphi'(x) < 0$ then an error does not exceed ε .

At application of method of simple iteration one of складностей there is bringing equation over $f(x) = 0$ to the kind $x = \varphi(x)$ thus, that the terms of

convergence of iteration process were executed. Will consider one of general approaches of bringing initial equation over $f(x)=0$ to the equivalent kind $x = \varphi(x)$. Let the sought after root ξ initial equation is on a segment $[a, b]$ thus

$$0 < m_1 \leq f'(x) \leq M_1, \quad \text{at } x \in [a, b]. \quad (12)$$

In particular, after m_1 it is possible to take on a the least value of derivative $f'(x)$ on $[a, b]$ which must be positive. After M_1 it is possible to take on a most value of derivative $f'(x)$ on $[a, b]$. Will replace equation $f(x)=0$ by equivalent to him equation

$$x = x - \lambda f(x), \quad \lambda > 0.$$

It is thus possible to put $\varphi(x) = x - \lambda f(x)$. Parameter λ sneaks up thus, that in околі $[a, b]$ root ξ inequality was executed

$$0 \leq \varphi'(x) = 1 - \lambda f'(x) \leq q < 1.$$

Last inequality with the use of formula (4.12) it is possible to write down :

$$0 \leq 1 - \lambda M_1 \leq 1 - \lambda m_1 \leq q.$$

From where swims out :

$$\lambda = \frac{1}{M_1} \quad \text{and} \quad q = 1 - \frac{m_1}{M_1} < 1.$$

In practical calculations number M_1 gets out thus, that $|M_1| > Q/2$, where $Q = \max |f'(x)|$ for all $x \in [a, b]$. Thus an iteration process coincides subject to condition $|\varphi'(x)| < 1$ on $[a, b]$.

Will consider the example of application of method of simple iteration for untiing of equation

$$f(x) \equiv x^3 - x - 1 = 0.$$

An initial function can be presented in a kind $\varphi_1(x) = \varphi_2(x)$ where $\varphi_1(x) = x^3$, $\varphi_2(x) = x + 1$. From the construction of charts $y_1 = \varphi_1(x)$ but $y_2 = \varphi_2(x)$ swims out, that the sought after root is on a segment $[1, 2]$ it is figure 8.

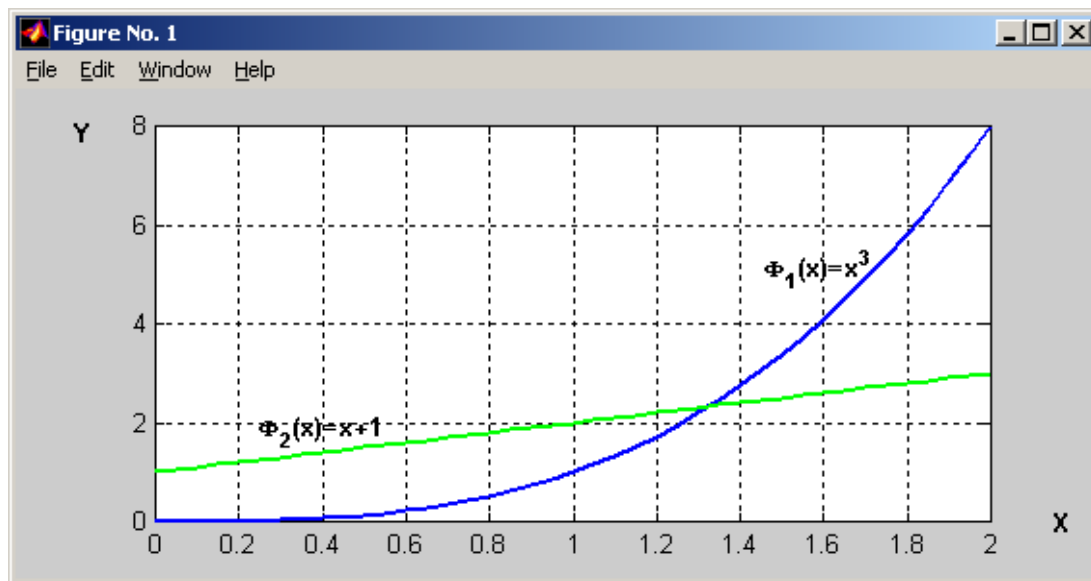


Figure 8.

Initial equation $x^3 - x - 1 = 0$ it is possible to present in a kind $x = x^3 - 1$ or $x = \varphi(x)$ where $\varphi(x) = x^3 - 1$. In the presented kind function $\varphi(x)$ dissatisfies to the terms of convergence, so as $\varphi'(x) = 3x^2$ and $\varphi'(x) \geq 3$ on a segment $[1, 2]$.

Will present initial equation in a kind $x = \sqrt[3]{x+1}$. In this case function

$$\varphi(x) = \sqrt[3]{x+1}, \quad \text{and} \quad \varphi'(x) = \frac{1}{3\sqrt[3]{(x+1)^2}}.$$

From swims out here, that $\frac{1}{3\sqrt[3]{9}} \leq \varphi'(x) \leq \frac{1}{3\sqrt[3]{4}}$ on a segment $[1, 2]$ or

$\varphi'(x) < \frac{1}{4}$, what satisfies to the terms of convergence of iteration process.

The results of calculations are driven to the table 2.5.

Table 2.5.

n	0	1	2	3	4	5	6	7
x_n	1	1,2599	1,3123	1,3224	1,3243	1,3246	1,3247	1,3247

§ 2.8. Method of Newton

Let in equation $f(x) = 0$ function $f(x)$ has continuous second derivative $f''(x)$ on a segment $[a; b]$ which the separated root is on x^* . It is assumed that derivatives $f'(x)$ but $f''(x)$ different from a zero, знакосталі and initial approaching of root x_0 it is belonged to the segment $[a; b]$. In obedience to the indicated requirements on a segment $[a; b]$ absent extremums and inflectionpoints of initial function, that allows in any point of segment $[a; b]$ to build a tangent to the curve. Will consider arbitrary point-on-wave $M_0(x_0, y(x_0))$, $x_0 \in [a; b]$ and will conduct a tangent to the curve in this point. Equation of tangent looks like

$$y - f(x_0) = f'(x_0)(x - x_0).$$

A tangent crosses abscise axis in some point $(x_1; 0)$. Taking into account the last, it is possible to write down

$$-f(x_0) = f'(x_0)(x - x_0),$$

from where get the first approaching of root

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Through a point $M_1(x_1, f(x_1))$ again will conduct a tangent to the curve, the intersection of which with abscise axis gives the second approaching of root and т. д.

Described process of construction of tangents a calculation of points of their crossing with abscise axis is an iteration process of Newton - Рафсона.

The construction of iteration sequence of approaching of chums takes place in obedience to next formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots \quad (13)$$

Geometrical maintenance of метода Newton - Рафсона consists in substituting of arc crooked by every iteration a tangent to her in a point x_n .

Method of Newton - Rafson can be examined as a partial case of method of simple iteration (3.8), if to put $\varphi(x) = x - \frac{f(x)}{f'(x)}$. As

$$\varphi'(x) = \frac{f(x)f''(x)}{(f'(x))^2},$$

from the condition of convergence of method of simple iteration swims out, that method of Newton - Рафсона coincides, when the initial approaching is chosen by near enough to simple Cornu x^* ($\varphi'(x^*) = 0$).

Thus, if initial approaching x_0 it is chosen close enough to Cornu x^* then the method of Newton always coincides. At the arbitrary initial approaching an iteration method (13) coincides, if for all $x \in [a; b]$ inequality is executed

$$\frac{|f(x)f''(x)|}{[f'(x)]^2} < 1. \quad (14)$$

A next theorem takes place thus.

Theorem. Let $f(x) \in C_2 [a; b]$ (function $f(x)$, derivatives $f'(x)$ i $f''(x)$ continuous on $[a; b]$), $f(a)f(b) < 0$ and derivatives $f'(x)$, $f''(x)$ keep a sign on a segment $[a; b]$. Then, if initial approaching $x_0 \in [a; b]$ satisfies inequality

$$f(x_0)f''(x_0) > 0, \quad (15)$$

a sequence (4.13) coincides (thus droningly) to only on a segment $[a; b]$ root x^* equation $f(x) = 0$.

For the estimation of speed of convergence of method around root will take advantage of formula of Taylor

$$f(x^*) = f(x_n) + f'(x_n)(x^* - x_n) + \frac{1}{2}f''(c)(x^* - x_n)^2,$$

or
$$f(x_n) + f'(x_n)(x^* - x_n) + \frac{1}{2}f''(c)(x^* - x_n)^2 = 0,$$

where a point $x = c$ lies between x_n i x^* . From the last equation have

$$x^* - x_n + \frac{f(x_n)}{f'(x_n)} = -\frac{1}{2} \frac{f''(c)}{f'(x_n)} (x^* - x_n)^2. \quad (16)$$

In obedience to a formula (13)

$$x_n - \frac{f(x_n)}{f'(x_n)} = x_{n+1},$$

to the volume

$$x^* - x_{n+1} = -\frac{1}{2} \frac{f''(c)}{f'(x_n)} (x^* - x_n)^2. \quad (17)$$

If to designate a most value through $M |f''(x)|$ on $[a; b]$ and through m is the least value $|f'(x)|$ on a segment $[a; b]$ then it is possible to write down next inequality for the estimation of error of two progressive approximations x_n but x_{n+1}

$$|x^* - x_{n+1}| < \frac{M}{2m} (x^* - x_n)^2. \quad (18)$$

In obedience to an estimation (18) swims out, that the error of the next new approaching diminishes proportionally to the square of error previous, id est convergence of Newton - Rafson method is quadratic.

If two progressive approximations are known x_n but x_{n+1} in obedience to the method of Newton - Rafson, then it is possible to write down on the basis of correlation (18)

$$|x^* - x_{n+1}| < \frac{M}{2m} (x_{n+1} - x_n)^2. \quad (19)$$

Thus, for determination of root of equation (4.1) in obedience to the method of Newton - Rafson with the set exactness ε an iteration process will be executed, while

$$|x_{n+1} - x_n| \leq \sqrt{2m\varepsilon/M}. \quad (20)$$

Remark. If derivative $f'(x)$ poorly changes on $[a; b]$ and her calculation is bulky enough, then an iteration process can be conducted on a formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}, \quad n = 0, 1, 2, \dots \quad (21)$$

The iteration process of kind (3.21) has the name *of the modified method of Newton*.

Will consider the example of application of method of Newton for untiing of equation

$$y \equiv \text{tg}(0.3x + 0.4) - x^2 = 0.$$

An initial function can be presented in a kind $\varphi_1(x) = \varphi_2(x)$ where $\varphi_1(x) = \text{tg}(0.3x + 0.4)$, $\varphi_2(x) = x^2$. From the brought graphic material over evidently, that graphic arts of functions $y_1 = \varphi_1(x)$ but $y_2 = \varphi_2(x)$ intersect in two points, roots belong to the intervals $[-1, 0)$ but $(0, 1]$ it is figure 9.

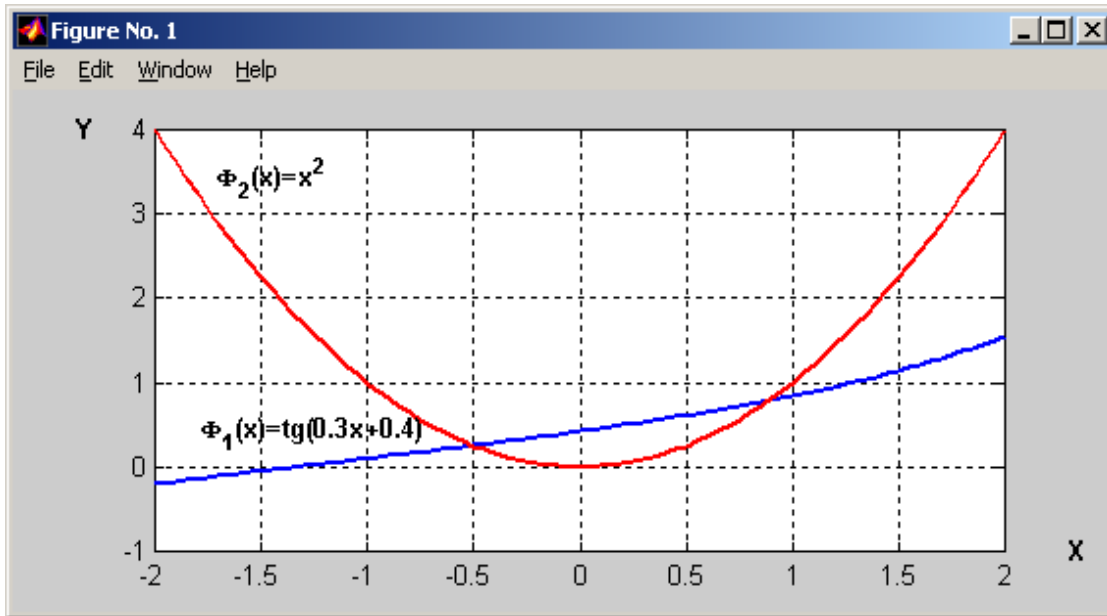


Figure 9.

Stopped for being of positive root.

$$y'(x) = \frac{0.3}{\cos^2(0.3x + 0.4)} - 2x, \quad y''(x) = \frac{0.18 \sin(0.3x + 0.4)}{\cos^3(0.3x + 0.4)} - 2.$$

Coming from the presented chart it is expedient to accept for the initial approaching $x_0 = 1$. Thus $y(x_0) = -0,1577$; $y''(x_0) = -1,7408$ and accordingly $y(x_0) \cdot y''(x_0) > 0$. The results of calculations are driven to the next table 2.6.

Table 2.6

n	0	1	2	3	4
x_n	1	0,8940	0,8864	0,8863	0,8863

Will consider application of method of Newton to the decision of next equation $f(x) \equiv x^3 - x - 1 = 0$ (this problem was examined at application of simple iteration method in a previous paragraph). A root of equation is on a segment $[1, 2]$ it is figure 7. First and second the derivatives of initial function are written down

$$f'(x) = 3x^2 - 1; \quad f''(x) = 6x.$$

A point sets to the initial approaching $x_0 = 2$. Thus $f(x_0) \cdot f''(x_0) > 0$. The results of calculations are driven to the table 2.7.

Table 2.7

n	0	1	2	3	4
x_n	2	1,5455	1,3258	1,3247	1,3247

The brought results over in a table 2.7 and comparing to the corresponding results according to calculations on the method of simple iteration (table 2.6) allow to draw conclusion about efficiency of application of Newton method.

§ 2.9. Method of secant

If at the calculation of derivative $f'(x)$ there are some difficulties, then more comfortable is application *of method of secant*. In this case derivative $f'(x_n)$ replaced by the first up-diffused difference, found for to two last iterations

$$f'(x_n) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}.$$

If to put expression $f'(x_n)$ in a formula (13), then will get iteration process of method of secant

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})}, \quad n = 1, 2, \dots \quad (22)$$

It is needed to set for the beginning of iteration process x_0 but x_1 . From the geometrical point of view at every step iteration method of secant part of curve is replaced by secant which passes through points with abscissas x_n but x_{n-1} .

It is known that if ξ it is a root of equation $f(x) = 0$ and $f'(\xi) \neq 0$, $f''(\xi) \neq 0$ and $f''(\xi)$ it is a continuous function, then there is such окіл points ξ that if x_0 but x_1 are different points of this околу, then the method of secant (22) coincides to root $x = \xi$. A next estimation takes place thus

$$|x_{n+1} - \xi| \leq C |x_n - \xi|^p,$$

where $C \neq 0$ – const, $p \approx 1,6$.

Thus, the method of secant coincides more slowly as compared to the method of Newton, but here at every step iterations are calculated only value of function in the set points.

Will consider application of method of secant to the decision of next equation $f(x) \equiv x^3 - x - 1 = 0$ (this problem was examined for evidentness at application of метода of simple iteration and method of Newton in previous paragraphs). A root of equation is on a segment $[1, 2]$ it is figure 7. Points set to the initial approaching $x_0 = 2$ but $x_1 = 1,8$. The results of calculations are driven to the table 2.8.

Table 2.8

n	1	2	3	4	5	6
x_n	1,8	1,4919	1,3760	1,3317	1,325	1,3247
x_{n-1}	2	1,8	1,4919	1,3760	1,3317	1,325

As see, in obedience to the brought results over, there is more slow convergence of iteration process of secant as compared to the iteration method of Newton (table 2.7).

§ 2.10. A method of simple iteration is for the systems of two equations

Let the set system of two equations with two unknown

$$\begin{cases} F_1(x, y) = 0, \\ F_2(x, y) = 0. \end{cases} \quad (23)$$

The problem of decision of the initial system of equations consists in being of actual chums with the set exactness.

Laid, that the system (23) assumes the only isolated roots. Number of these chums and them close values can be set, if to build curves $F_1(x, y)$ but $F_2(x, y)$ and to define the co-ordinates of their intersections.

For application of method of simple iteration the system (3.23) over is brought to the kind

$$\begin{cases} x = \varphi_1(x, y), \\ y = \varphi_2(x, y). \end{cases} \quad (24)$$

The algorithm of decision is set by formulas

$$\begin{cases} x_{n+1} = \varphi_1(x_n, y_n), \\ y_{n+1} = \varphi_2(x_n, y_n), \end{cases} \quad (n = 0, 1, 2, \dots), \quad (25)$$

where x_0, y_0 it is some initial approaching.

A next theorem takes place thus.

Theorem. Let in some limited area R ($a \leq x \leq A, b \leq y \leq B$) there is one and only one decision $x = \xi, y = \eta$ systems (24). If

- 1) functions $\varphi_1(x, y), \varphi_2(x, y)$ certain and continuously differentiated in R ;
- 2) initial approaching x_0, y_0 and all next approaching x_n, y_n ($n = 0, 1, 2, \dots$) areas belong R ;
- 3) in area of R inequalities are executed

$$\begin{cases} \left| \frac{\partial \varphi_1}{\partial x} \right| + \left| \frac{\partial \varphi_2}{\partial x} \right| \leq q_1 < 1, \\ \left| \frac{\partial \varphi_1}{\partial y} \right| + \left| \frac{\partial \varphi_2}{\partial y} \right| \leq q_2 < 1; \end{cases} \quad (26)$$

then the process of progressive approximations (25) coincides to the decision $x = \xi, y = \eta$ if there are

$$\lim_{n \rightarrow \infty} x_n = \xi \quad \text{and} \quad \lim_{n \rightarrow \infty} y_n = \eta.$$

Consequence. The indicated theorem remains faithful, if to replace terms (26) terms

$$\begin{cases} \left| \frac{\partial \varphi_1}{\partial x} \right| + \left| \frac{\partial \varphi_1}{\partial y} \right| \leq q_1 < 1, \\ \left| \frac{\partial \varphi_2}{\partial x} \right| + \left| \frac{\partial \varphi_2}{\partial y} \right| \leq q_2 < 1. \end{cases} \quad (27)$$

Estimation of error n – the n th approaching is determined by inequality

$$|\xi - x_n| + |\eta - y_n| \leq \frac{M}{1-M} (|x_n - x_{n-1}| + |y_n - y_{n-1}|), \quad (28)$$

where M - most from numbers q_1, q_2 in inequalities (26) or (27). Convergence of method iterations is considered satisfactory, if $M < 1/2$ thus $M/(1-M) < 1$.

For example, will consider being of decision for the next system of equations

$$\begin{cases} \frac{(x-3)^2}{9} + \frac{(y-2)^2}{4} = 1, \\ y = x^3, \end{cases}$$

with an error $\varepsilon = 10^{-3}$.

For determination of number of chums and them close values will present the initial system of equations graphically. The first equation of the indicated system is equation of ellipse, semi-axes which $a_{\text{el}} = 3$, $b_{\text{el}} = 2$. A center of ellipse is in a point $x = 3$, $y = 2$. The second equation of the system is equation of cube parabola. The graphic image of the indicated equations is presented on figure 10.

Intersections of charts of equations of the system are the roots of equations, upshots of which are. As evidently from the brought graphic representation over, the indicated system has two pair of chums.

For application of method of simple iteration will write down the initial system of equations in a next kind

$$x = y^{1/3} \equiv \varphi_1(x, y),$$

$$y = 2 \pm \sqrt{4 - \frac{4}{9}(x - 3)^2} \equiv \varphi_2(x, y).$$

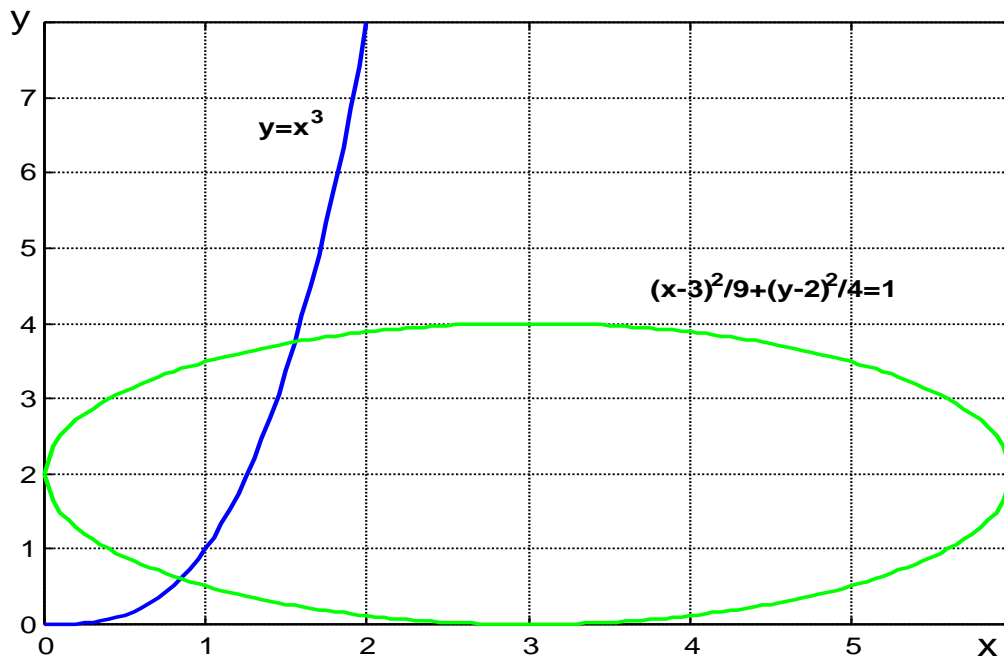


Figure 10.

Coming from graphic presentation of the system of equations – lines 10, for the separation of chums of the system will consider areas

$$\Omega_1 = \{0,6 \leq x \leq 1; \quad 0,5 \leq y \leq 1\} \text{ and } \Omega_2 = \{1,3 \leq x \leq 1,7; \quad 3,5 \leq y \leq 4\}.$$

Will consider being of the first pair of chums in area of Ω_1 . It is undifficult to show that at any choice (x_0, y_0) that belong Ω_1 value of sequence (x_n, y_n) , $(n = 1, 2, \dots)$ will belong also Ω_1 . In particular, at $0,5 \leq y \leq 1 - 0,794 \leq \varphi_1(x_n, y_n) \leq 1$ and at $0 \leq x \leq 1 - 0,509 \leq \varphi_2(x_n, y_n) \leq 0,894$.

For the points of area Ω_1 have

$$\left| \frac{\partial \varphi_1}{\partial x} \right| + \left| \frac{\partial \varphi_1}{\partial y} \right| = \frac{1}{3} \frac{1}{y^{2/3}} < 0,529,$$

$$\left| \frac{\partial \varphi_2}{\partial x} \right| + \left| \frac{\partial \varphi_2}{\partial y} \right| = \frac{\frac{4}{9}(3-x)}{\sqrt{4 - \frac{4}{9}(x-3)^2}} < 0,889.$$

Coming from the brought inequalities over it is possible to draw conclusion, that in area of Ω_1 there are an only decision of the initial system of equations and this decision can be found the method of iterations in obedience to formulas

$$x_{n+1} = y_n^{1/3},$$

$$y_{n+1} = 2 \pm \sqrt{4 - \frac{4}{9}(x_n - 3)^2}.$$

For a case $x_0 = 0,6$ i $y_0 = 0,5$ the results of calculations are driven to the table 2.10.

Table 2.10.

And root		
N	x_n	y_n
0	0,6	0,5
1	0,7937	0,8
2	0,9283	0,6448
3	0,8634	0,5535
4	0,8210	0,5957
5	0,8414	0,6253
6	0,8551	0,6111
7	0,8486	0,6017
8	0,8442	0,6061
9	0,8463	0,6091
10	0,8477	0,6077
11	0,8470	0,6067
12	0,8466	0,6072
13	0,8468	0,6075
14	0,8469	0,6074
15	0,8469	0,6073
16	0,8468	0,6073
17	0,8468	0,6073

II root	
x_n	y_n
1,3	3,5
1,5183	3,6479
1,5394	3,7390
1,5521	3,7469
1,5532	3,7516
1,5538	3,7520
1,5539	3,7523
1,5539	3,7523

Will consider being of decision of the system for the case of area $\Omega_2 = \{1,3 \leq x \leq 1,7; 3,5 \leq y \leq 4\}$. In obedience to formulas

$$x = y^{1/3} \equiv \varphi_1(x, y),$$

$$y = 2 + \sqrt{4 - \frac{4}{9}(x - 3)^2} \equiv \varphi_2(x, y).$$

As well as in previous case, will show that at be - what choice (x_0, y_0) that belong Ω_2 value of sequence (x_n, y_n) , $(n = 1, 2, \dots)$ areas will belong also Ω_2 . At

$$1,3 \leq x \leq 1,7 - 3,65 \leq \varphi_2(x, y) \leq 3,89\}.$$

For an area Ω_2 next inequalities are executed

$$\left| \frac{\partial \varphi_1}{\partial x} \right| + \left| \frac{\partial \varphi_1}{\partial y} \right| = \frac{1}{3} \frac{1}{y^{2/3}} < 0,145,$$

$$\left| \frac{\partial \varphi_2}{\partial x} \right| + \left| \frac{\partial \varphi_2}{\partial y} \right| = \frac{\frac{4}{9}(x - 3)}{\sqrt{4 - \frac{4}{9}(x - 3)^2}} < 0,459.$$

For a case $x_0 = 1,3$ i $y_0 = 3,5$ the results of calculation are driven to the table 2.10.

From the brought numeral results over see that for being of decision $x = 0,8469$ i $y = 0,6073$ in area of Ω_1 it is necessary to conduct 15 iterations, and for being of decision $x = 1,5538$ i $y = 3,7520$ it is necessary to conduct 5 iterations. Speed of convergence of method of iterations depends on a size M in correlation (28). For the case of decision in Ω_1 $M = \max(0,529; 0,889) = 0,889$. For a decision in area of Ω_2 - $M = \max(0,145; 0,459) = 0,459$. In practical calculations convergence of метода iterations is considered satisfactory, if $M < 0,5$.

In a number of cases, for the decision of the system of kind (3.23) *the iteration method of Seidel*, which is modification of метода of simple iteration,

is used. Basic essence of this method consists in that at a calculation $(n + 1)$ – the approaching is for a size y_{n+1} the already calculated value of size is taken into account x_{n+1} .

For the decision of the system (24) the algorithm of Seidel method is set by formulas

$$\begin{cases} x_{n+1} = \varphi_1(x_n, y_n), \\ y_{n+1} = \varphi_2(x_{n+1}, y_n), \end{cases} \quad (n = 0, 1, 2, \dots). \quad (29)$$

The theorem of convergence and consequence of theorem of convergence of метода of simple iteration take place thus.

For example, to untie the next system of equations the iteration method of Seidel

$$\begin{cases} \frac{(x-3)^2}{9} + \frac{(y-2)^2}{4} = 1, \\ y = x^3. \end{cases}$$

This system was examined higher at application of метода of simple iteration. Coming from graphic presentation of the system of equations - lines 10, for the separation of chums of the system two areas are examined

$$\Omega_1 = \{0,6 \leq x \leq 1; \quad 0,5 \leq y \leq 1\} \text{ and } \Omega_2 = \{1,3 \leq x \leq 1,7; \quad 3,5 \leq y \leq 4\}.$$

There is an iteration process in area of $\Omega_1 = \{0,6 \leq x \leq 1; \quad 0,5 \leq y \leq 1\}$ and

$\Omega_2 = \{1,3 \leq x \leq 1,7; \quad 3,5 \leq y \leq 4\}$ set by formulas

$$x_{n+1} = y_n^{1/3}, \quad y_{n+1} = 2 \pm \sqrt{4 - \frac{4}{9}(x_{n+1} - 3)^2}.$$

In the last formula sign minus before підкореневим expression answers the value of root from an area Ω_1 and sign plus - root from an area Ω_2 .

The results of calculations are driven to the table 2.11.

As evidently from the brought calculations over, for being of розв'язкув area Ω_1 it is necessary to conduct 8 iterations (for comparison in the case of method of simple iteration - 15 iterations), and for being of decision in area of Ω_2 it is necessary to conduct 4 iterations (in the case of method of simple iteration - 5 iterations).

Table 2.11

And root			II root	
n	x_n	y_n	x_n	y_n
0	0,6	0,5	1,3	3,5
1	0,7937	0,6448	1,5183	3,7390
2	0,8639	0,5957	1,5521	3,7516
3	0,8414	0,6111	1,5538	3,7523
4	0,8486	0,6061	1,5539	3,7523
5	0,8463	0,6077		
6	0,8470	0,6072		
7	0,8468	0,6074		
8	0,8469	0,6073		
9	0,8468	0,6073		

§ 2.11. A method of Newton is for the systems of two equations

Will consider the system of two equations with two unknown

$$\begin{cases} F_1(x, y) = 0, \\ F_2(x, y) = 0, \end{cases} \quad (30)$$

where F_1, F_2 are the continuously differentiated functions for to the variables of x and y .

Will consider that the system of equations has the separated decision and will designate this decision through (x^*, y^*) . Every function in the system (30) it is possible to decompose in the series of Taylor around decision. Ignoring the elements of high order in this time-table, system of equations in the linearized form it is possible to write down in a kind

$$F_1(x^*, y^*) \approx F_1(x, y) + \frac{\partial F_1(x^*, y^*)}{\partial x} (x - x^*) + \frac{\partial F_1(x^*, y^*)}{\partial y} (y - y^*), \quad (31)$$

$$F_2(x^*, y^*) \approx F_2(x, y) + \frac{\partial F_2(x^*, y^*)}{\partial x} (x - x^*) + \frac{\partial F_2(x^*, y^*)}{\partial y} (y - y^*).$$

If in the system (31) to equate right parts with a zero, then a decision of the system will be different from (x^*, y^*) in connection with that in equations (31) cast aside the elements of higher orders. As a result will get some new value (x, y) . Coming from it, iteration procedure can be written down in a kind

$$\bar{F}(\bar{x}_k) + M(\bar{x}_{k+1} + \bar{x}_k) = 0, \quad (32)$$

where vectors are $\bar{F} = (F_1, F_2)$, $\bar{x} = (x, y)$ and the matrix of M looks like

$$M = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{bmatrix}. \quad (33)$$

The matrix of kind (33) is named *the matrix of Jacobi*. Formally the decision of the system (32) is determined in obedience to next formulas

$$\bar{x}_{k+1} = \bar{x}_k - M^{-1}\bar{F}(\bar{x}_k). \quad (34)$$

In practice being of matrix of Jacobi M^{-1} it is inadvisable. If to designate

$$\Delta\bar{x}_k = \bar{x}_{k+1} - \bar{x}_k, \quad (35)$$

then the system (34) can be presented in a kind

$$M\Delta\bar{x}_k = -\bar{F}(\bar{x}_k). \quad (36)$$

The system (36) is the system of two equations in relation to unknown $\Delta x_k, \Delta y_k$. Untiing the system (36), determine sizes

$$\bar{x}_{k+1} = \bar{x}_k + \Delta\bar{x}_k. \quad (37)$$

Equation (36 - 37) are *an iteration method of Newton* for the initial system of equations (3.30).

The algorithmic chart of Newton method consists in the following.

It is considered that it is set some k of the n -re approaching $\bar{x}_k = (x_k, y_k)$.

I-й step. The value of component of vector is calculated

$$\bar{F}(\bar{x}_k) = \begin{bmatrix} F_1(x_k, y_k) \\ F_2(x_k, y_k) \end{bmatrix}.$$

II-й step. The matrix of Jacobi calculates at a point (x_k, y_k)

$$M(\bar{x}_k) = \begin{bmatrix} \frac{\partial F_1(x_k, y_k)}{\partial x} & \frac{\partial F_1(x_k, y_k)}{\partial y} \\ \frac{\partial F_2(x_k, y_k)}{\partial x} & \frac{\partial F_2(x_k, y_k)}{\partial y} \end{bmatrix}.$$

III-й step. The linear system of equations gets untied relatively $\Delta\bar{x}_k$

$$M(\bar{x}_k)\Delta\bar{x}_k = -\bar{F}(\Delta\bar{x}_k),$$

or

$$\begin{bmatrix} \frac{\partial F_1(x_k, y_k)}{\partial x} & \frac{\partial F_1(x_k, y_k)}{\partial y} \\ \frac{\partial F_2(x_k, y_k)}{\partial x} & \frac{\partial F_2(x_k, y_k)}{\partial y} \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta y_k \end{bmatrix} = - \begin{bmatrix} F_1(x_k, y_k) \\ F_2(x_k, y_k) \end{bmatrix}.$$

IV step. There is a value of next point

$$\bar{x}_{k+1} = \bar{x}_k + \Delta\bar{x}_k, \text{ or}$$

$$x_{k+1} = x_k + \Delta x_k, \quad y_{k+1} = y_k + \Delta y_k.$$

Farther an iteration process recurs beginning from the first step.

For example, will consider the system of equations of kind

$$\begin{cases} \frac{(x-3)^2}{9} + \frac{(y-2)^2}{4} = 1, \\ y = x^3. \end{cases}$$

This system was examined in a previous paragraph. The graphic image of equations is presented on figure 9. It is considered that the roots of the system of equations are separated and are in areas

$$\Omega_1 = \{0,6 \leq x \leq 1; \quad 0,5 \leq y \leq 1\} \text{ and } \Omega_2 = \{1,3 \leq x \leq 1,7; \quad 3,5 \leq y \leq 4\}.$$

Will apply the iteration process of Newton for being of decisions of the initial system.

Will write down the initial system of equations in a kind

$$F_1(x, y) = 0, \quad F_2(x, y) = 0,$$

where

$$F_1(x, y) = \frac{(x-3)^2}{9} + \frac{(y-2)^2}{4} - 1,$$

$$F_2(x, y) = y - x^3.$$

The matrix of Jacobi looks like

$$M = \begin{bmatrix} \frac{2}{9}(x-3) & \frac{1}{2}(y-2) \\ -3x^2 & 1 \end{bmatrix}.$$

Will consider being of decisions in area of $\Omega_1 = \{0,6 \leq x \leq 1; 0,5 \leq y \leq 1\}$. For the initial approaching will accept $x_0 = 0,6; y_0 = 0,5$. In obedience to the algorithmic chart of Newton have

And root

$$x_0 = 0,6; y_0 = 0,5.$$

And step.

$$F_1(x_0, y_0) = 0,2025,$$

$$F_2(x_0, y_0) = 0,2840,$$

$$M(x_0, y_0) = \begin{pmatrix} -0,533 & -0,75 \\ -1,08 & 1 \end{pmatrix},$$

$$\Delta x = 0,3093; \quad x_1 = 0,9093;$$

$$\Delta y = 0,05; \quad y_1 = 0,550.$$

II step.

$$x_1 = 0,9093; \quad y_1 = 0,550.$$

$$F_1(x_1, y_1) = 0,0113,$$

$$F_2(x_1, y_1) = 0,2018,$$

$$M(x_1, y_1) = \begin{pmatrix} -0,4646 & -0,725 \\ -2,4805 & 1 \end{pmatrix},$$

$$\begin{aligned}\Delta x &= -0,0597; & x_2 &= 0,8496; \\ \Delta y &= 0,0538; & y_2 &= 0,6038.\end{aligned}$$

III step.

$$x_2 = 0,8496; \quad y_2 = 0,6038.$$

$$F_1(x_2, y_2) = 0,0011,$$

$$F_2(x_2, y_2) = 0,0095,$$

$$M(x_2, y_2) = \begin{pmatrix} -0,4779 & -0,6981 \\ -2,1656 & 1 \end{pmatrix},$$

$$\Delta x = -0,0028; \quad x_3 = 0,8469;$$

$$\Delta y = 0,0035; \quad y_3 = 0,6073.$$

For being of the second root will consider an area
 $\Omega_2 = \{1,3 \leq x \leq 1,7; \quad 3,5 \leq y \leq 4\}.$

II root.

$$x_0 = 1,3; \quad y_0 = 3,5.$$

And step.

$$F_1(x_0, y_0) = -0,1164,$$

$$F_2(x_0, y_0) = 1,303,$$

$$M(x_0, y_0) = \begin{pmatrix} -0,3778 & 0,75 \\ -5,07 & 1,0 \end{pmatrix},$$

$$\begin{aligned}\Delta x &= 0,3193; & x_1 &= 1,6193; \\ \Delta y &= 0,316; & y_1 &= 3,816.\end{aligned}$$

II step.

$$x_1 = 1,6193; \quad y_1 = 3,816.$$

$$F_1(x_1, y_1) = 0,0363,$$

$$F_2(x_1, y_1) = -0,4303,$$

$$M(x_1, y_1) = \begin{pmatrix} -0,3068 & 0,9080 \\ -7,8668 & 1 \end{pmatrix},$$

$$\Delta x = -0,0625; \quad x_2 = 1,5569;$$

$$\Delta y = -0,0611; \quad y_2 = 3,755.$$

III step.

$$x_2 = 1,5569; \quad y_2 = 3,755.$$

$$F_1(x_2, y_2) = 0,0014,$$

$$F_2(x_2, y_2) = -0,0187,$$

$$M(x_2, y_2) = \begin{pmatrix} -0,3207 & 0,8775 \\ -7,2716 & 1 \end{pmatrix},$$

$$\Delta x = -0,0029; \quad x_3 = 1,5539;$$

$$\Delta y = -0,0026; \quad y_3 = 1,7523.$$

§ 2.12. Iteration methods of decision of the systems of nonlinear equations

Will consider the system n nonlinear equations from n unknown

$$f_i(x_1, x_2, \dots, x_n) = 0, \quad i = \overline{1, n}, \quad (38)$$

where f_i are some algebra or transcendent functions. Will designate $\bar{x} = (x_1, x_2, \dots, x_n)$ but $\bar{f} = (f_1, f_2, \dots, f_n)$ the system (38) will write down in a vectorial form

$$\bar{f}(\bar{x}) = 0. \quad (39)$$

A decision of the system (39) is more intricate problem, than decision of one equation. For solution of such systems iteration methods are used. Will generalize the method of progressive approximations and method of Newton- Rafson for the case of the system (39).

Method of simple iteration. Will replace the nonlinear system (39) the equivalent system of the special kind

$$\bar{x} = \varphi(\bar{x}), \quad (40)$$

where $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_n)$.

Possibly, that the system (3.40) has in the limited protuberant closed area of D of n -size of space of X only decision $\bar{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ and components \bar{x}_i^0 vector $\bar{x}^0 = (x_1^0, x_2^0, \dots, x_n^0)$ are scalar sizes, accordingly near to \bar{x}_i^* , $i = \overline{1, n}$.

Will find the next approaching to the exact decision by means of method of simple iteration (progressive approximations) after formulas:

$$\bar{x}^{k+1} = \varphi(\bar{x}^k), \quad (41)$$

or in a co-ordinate form

$$x_i^{k+1} = \varphi_i(x_1^k, x_2^k, \dots, x_n^k), \quad i = \overline{1, n}, \quad k = 0, 1, 2, \dots$$

Let $\rho(\bar{x}^1, \bar{x}^2) = \|\bar{x}^1 - \bar{x}^2\|$ it is distance between elements \bar{x}^1 i \bar{x}^2 in space of X , where for the norm of vector it is possible to choose an arbitrary canonical norm.

In obedience to principle of the squeezing mapping the system of equations (40) has an only decision $\bar{x}^* \in D$ which can be found the method of iterations (41) at any choice of the initial approaching $\bar{x}^0 \in D$ if all progressive approximations $\bar{x}^k \in D$, ($k = 1, 2, \dots$) and mapping $\bar{\varphi}(\bar{x})$ squeezes in D .

Will consider the sufficient terms of convergence of method of iterations which are comfortable during realization of practical calculations.

Possibly, that in some protuberant closed area of D of function $\varphi_i(\bar{x})$ have continuous derivatives of part $\frac{\partial \varphi_i}{\partial x_j}$ and in area of D the system (3.40) has an only

decision \bar{x}^* . Let for the arbitrary initial approaching $\bar{x}^0 \in D$ all next approaching $\bar{x}^k \in D$.

Arround of decision \bar{x}^* after the generalized formula of Lagrange

$$x_i^{k+1} - x_i^* = \varphi_i(\bar{x}^k) - \varphi_i(\bar{x}^*) = \sum_{j=1}^n \frac{\partial \varphi_i(\theta_i^k)}{\partial x_j} (x_j^k - x_j^*), \quad i = \overline{1, n}; \quad (42)$$

where θ_i^k it is some point of segment of line which connects points

$$\bar{x}^k = (x_1^k, x_2^k, \dots, x_n^k) \text{ and } \bar{x}^* = (x_1^*, x_2^*, \dots, x_n^*).$$

To Tom

$$\rho(\bar{x}^{k+1}, \bar{x}^*) = \rho(\varphi(\bar{x}^k), \varphi(\bar{x}^*)) = \|\varphi(\bar{x}^k) - \varphi(\bar{x}^*)\| = \quad (43)$$

$$= \|J(\bar{\theta}^k) \cdot (\bar{x}^k - \bar{x}^*)\| \leq \|J(\bar{\theta}^k)\| \cdot \rho(\bar{x}^k - \bar{x}^*),$$

where $J(x)$ it is a matrix of Jacobi of the system (40)

$$J(\bar{x}) = \begin{pmatrix} \frac{\partial \varphi_1(\bar{x})}{\partial x_1} & \frac{\partial \varphi_1(\bar{x})}{\partial x_2} & \dots & \frac{\partial \varphi_1(\bar{x})}{\partial x_n} \\ \frac{\partial \varphi_2(\bar{x})}{\partial x_1} & \frac{\partial \varphi_2(\bar{x})}{\partial x_2} & \dots & \frac{\partial \varphi_2(\bar{x})}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial \varphi_n(\bar{x})}{\partial x_1} & \frac{\partial \varphi_n(\bar{x})}{\partial x_2} & \dots & \frac{\partial \varphi_n(\bar{x})}{\partial x_n} \end{pmatrix}. \quad (3.44)$$

In obedience to principle of the squeezing mapping (see the division of III) the method of progressive approximations (3.41) coincides to the decision \bar{x}^* systems (3.40), if the arbitrary concerted norm of matrix of Jacobi $\|J(\bar{\theta}^k)\|$ there will be less unit.

Will write down inequality (3.43) in a kind

$$\rho(\bar{x}^k, \varphi(\bar{x}^*)) \leq \|J(\bar{\theta}^k)\| \cdot \rho(\bar{x}^k, \bar{x}^*) \leq \|M\| \cdot \rho(\bar{x}^k, \bar{x}^*). \quad (45)$$

In practice comfortably to examine the matrix of M with elements

$$M_{ij} = \max_D \left| \frac{\partial \varphi_i}{\partial x_j} \right|,$$

the norm of which majorizes a norm $\|J(x)\|$.

Mapping (40) will squeeze in D , if for the arbitrary concerted norm of matrix of M a condition is executed:

$$\|M\| < 1. \quad (46)$$

A next theorem takes place.

Theorem. If functions $\varphi_i(x_1, x_2, \dots, x_n)$, $i = \overline{1, n}$ in some protuberant area of D , which contains a decision $\bar{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ systems (3.40), continuous and have the continuous first derivatives, then for convergence of method of iterations sufficiently, that in the matrix of M with elements

$$M_{ij} = \max_D \left| \frac{\partial \varphi_i}{\partial x_j} \right| \text{ all own values were on the module less unit, and initial}$$

approaching $\bar{x}^0 = (x_1^0, x_2^0, \dots, x_n^0)$ it is near enough to the decision $\bar{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$.

Approaching of vector is to the decision \bar{x}^k to the decision \bar{x}^* will characterize одною from norms $\|\bar{x}\|_1$, $\|\bar{x}\|_m$ or $\|\bar{x}\|_k$.

Going out the generally accepted standard norms for convergence of iterations method (41) enough implementation of one of terms:

$$\max_{1 \leq i \leq n} \sum_{j=1}^m M_{ij} = q_i < 1; \quad (47)$$

$$\max_{1 \leq j \leq n} \sum_{i=1}^n M_{ij} = q_i < 1; \quad \left(\sum_{i,j=1}^n M_{ij}^2 \right)^{1/2} = q_i < 1.$$

Will consider an example. Let the set system of equations

$$\begin{cases} x^2 + x - 2yz = 0,1 \\ -y^2 + y + 3xz = 0,2 . \\ z^2 + z - 2xy = 0 \end{cases}$$

It is necessary to get a decision the method of simple iteration with the set exactness ε .

Will convert the initial system to the kind

$$\begin{cases} \left(x + \frac{1}{2} \right)^2 = 0,35 + 2yz \\ \left(y - \frac{1}{2} \right)^2 = 0,05 + 3xz , \\ \left(z + \frac{1}{2} \right)^2 = 0,25 + 2xy \end{cases}$$

or

$$\begin{cases} x = -\frac{1}{2} \pm \sqrt{0,35 + 2yz} \\ y = \frac{1}{2} \pm \sqrt{0,05 + 3xz} . \\ z = -\frac{1}{2} \pm \sqrt{0,25 + 2xy} \end{cases}$$

Will build the graphic image of initial equations, which is presented on figure

11.

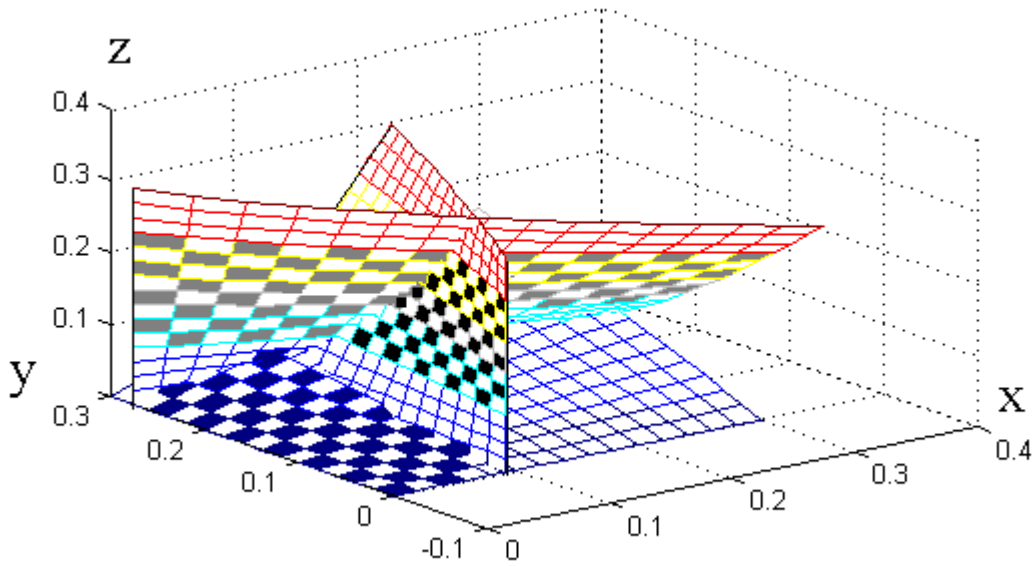


Figure 11.

In obedience to the presented graphic material - lines. 11, roots of equations of the initial system are in intervals $\{0 \leq x \leq 0,2; \quad 0 \leq y \leq 0,3; \quad 0 \leq z \leq 0,1\}$.

An iteration process for being of chums of the system build in obedience to next formulas

$$x_{n+1} = \varphi_1(x_n, y_n, z_n),$$

$$y_{n+1} = \varphi_2(x_n, y_n, z_n),$$

$$z_{n+1} = \varphi_3(x_n, y_n, z_n),$$

where

$$\varphi_1(x_n, y_n, z_n) = -\frac{1}{2} + \sqrt{0,35 + 2y_n z_n},$$

$$\varphi_2(x_n, y_n, z_n) = \frac{1}{2} - \sqrt{0,05 + 3x_n z_n},$$

$$\varphi_3(x_n, y_n, z_n) = -\frac{1}{2} + \sqrt{0,25 + 2x_n y_n}.$$

At application of method of simple iteration it is necessary to adhere to the terms of convergence of iteration process. Will go out a condition in obedience to which for convergence of method of simple iteration sufficiently, that own values of matrix of M (elements of which $M_{ij} = \max \left| \frac{\partial \varphi_i}{\partial x_j} \right|$) there was less unit on the module in area of values of initial equations. In our case an area is certain in intervals $\{0 \leq x \leq 0,2; 0 \leq y \leq 0,3; 0 \leq z \leq 0,1\}$. The matrix of M takes on next values (maximal values of derivatives of part $\left| \frac{\partial \varphi_i}{\partial x_j} \right|$ arrive at in points $x = 0,2; y = 0,3; z = 0,1$)

$$M = \begin{pmatrix} 0 & 0,1562 & 0,4685 \\ 0,4523 & 0 & 0,9045 \\ 0,4932 & 0,3288 & 0 \end{pmatrix}.$$

The own numbers of matrix of M have next values

$$d = 0.8712, - 0.2629, - 0.6083.$$

As see that value of own numbers of matrix of M on the module less unit, that allows to talk about convergence of iteration process.

For a zero approaching accepted $x_0 = 0; y_0 = 0; z_0 = 0$. The results of calculations are driven to the next table 2.12.

Table 2.12

n	x_n	y_n	z_n
0	0	0	0
1	0.0916	0.2764	0
2	0.0916	0.2764	0.0483
3	0.1138	0.2485	0.0483
4	0.1116	0.2422	0.0537
5	0.1132	0.2393	0.0514
6	0.1120	0.2403	0.0515
7	0.1122	0.2405	0.0512
8	0.1121	0.2407	0.0513
9	0.1121	0.2407	0.0513

Method of Newton - Rafson. The system of nonlinear equations of kind (4.39) is examined - $\bar{f}(\bar{x}) = 0$. Will consider that the system (4.39) has a decision and will designate him \bar{x}^* . Will decompose a function $\bar{f}(\bar{x})$ in the series of Taylor limited to in a time-table the elements of a zero and first degree, that is, we linearize initial function $\bar{f}(\bar{x})$

$$\bar{f}(\bar{x}^*) \approx \bar{f}(\bar{x}) + J(\bar{x}^* - \bar{x}), \quad (48)$$

where a matrix of J is a matrix of Jacobi of function $\bar{f}(\bar{x})$

$$J(\bar{x}) = \begin{pmatrix} \frac{\partial f_1(\bar{x})}{\partial x_1} & \frac{\partial f_1(\bar{x})}{\partial x_2} & \dots & \frac{\partial f_1(\bar{x})}{\partial x_n} \\ \frac{\partial f_2(\bar{x})}{\partial x_1} & \frac{\partial f_2(\bar{x})}{\partial x_2} & \dots & \frac{\partial f_2(\bar{x})}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n(\bar{x})}{\partial x_1} & \frac{\partial f_n(\bar{x})}{\partial x_2} & \dots & \frac{\partial f_n(\bar{x})}{\partial x_n} \end{pmatrix}. \quad (49)$$

If to choose some initial approaching \bar{x}_0 in near enough околі root \bar{x}^* recognition (39), will get a next iteration process

$$\bar{f}(\bar{x}^k) + J(\bar{x}^{k+1} - \bar{x}^k) = 0, \quad (k = 0, 1, 2, \dots). \quad (50)$$

In particular, if there is an inverse matrix of Jacobi $J^{-1}(\bar{x}^k)$ then iteration process of Newton - Rafson can be written down in a kind

$$\bar{x}^{k+1} = \bar{x}^k - J^{-1}(\bar{x}^k) \cdot \bar{f}(\bar{x}^k). \quad (51)$$

In practice being of matrix of Jacobi $J^{-1}(\bar{x}^k)$ it is inadvisable. If to designate

$$\Delta \bar{x}^k = \bar{x}^{k+1} - \bar{x}^k, \quad (52)$$

then the system (50) can be presented in a kind

$$J(\bar{x}^k) \Delta \bar{x}^k = -\bar{f}(\bar{x}^k). \quad (53)$$

The system (53) is the system of equations in relation to an unknown vector $\Delta \bar{x}^k$. Unting the system (53), determine sizes

$$\bar{x}^{k+1} = \bar{x}^k + \Delta \bar{x}^k . \quad (54)$$

Equation (3.53), (3.54) are *an iteration method of Newton - Rafson* for the initial system of equations (3.39).

Algorithmic chart of method of Newton - Rafson for the system of n equations consists in the following.

It is considered that it is set some k of the -re approaching $\bar{x}^k = (x_1^k, x_2^k, \dots, x_n^k)$.

I-й step. The value of КОМПОНЕНТ of vector is calculated

$$\bar{f}(\bar{x}^k) = \begin{bmatrix} f_1(\bar{x}^k) \\ f_2(\bar{x}^k) \\ \dots \\ f_n(\bar{x}^k) \end{bmatrix} .$$

II-й step. The matrix of Jacobi calculates at a point $\bar{x}^k = (x_1^k, x_2^k, \dots, x_n^k)$

$$J(\bar{x}^k) = \begin{pmatrix} \frac{\partial f_1(\bar{x}^k)}{\partial x_1} & \frac{\partial f_1(\bar{x}^k)}{\partial x_2} & \dots & \frac{\partial f_1(\bar{x}^k)}{\partial x_n} \\ \frac{\partial f_2(\bar{x}^k)}{\partial x_1} & \frac{\partial f_2(\bar{x}^k)}{\partial x_2} & \dots & \frac{\partial f_2(\bar{x}^k)}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n(\bar{x}^k)}{\partial x_1} & \frac{\partial f_n(\bar{x}^k)}{\partial x_2} & \dots & \frac{\partial f_n(\bar{x}^k)}{\partial x_n} \end{pmatrix} .$$

III-й step. The linear system of equations gets untied relatively $\Delta\bar{x}^k$

$$J(\bar{x}^k)\Delta\bar{x}^k = -\bar{f}(\Delta\bar{x}^k),$$

or

$$\begin{pmatrix} \frac{\partial f_1(\bar{x}^k)}{\partial x_1} & \frac{\partial f_1(\bar{x}^k)}{\partial x_2} & \dots & \frac{\partial f_1(\bar{x}^k)}{\partial x_n} \\ \frac{\partial f_2(\bar{x}^k)}{\partial x_1} & \frac{\partial f_2(\bar{x}^k)}{\partial x_2} & \dots & \frac{\partial f_2(\bar{x}^k)}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n(\bar{x}^k)}{\partial x_1} & \frac{\partial f_n(\bar{x}^k)}{\partial x_2} & \dots & \frac{\partial f_n(\bar{x}^k)}{\partial x_n} \end{pmatrix} \begin{bmatrix} \Delta x_1^k \\ \Delta x_2^k \\ \cdot \\ \cdot \\ \Delta x_n^k \end{bmatrix} = - \begin{bmatrix} f_1(\bar{x}^k) \\ f_2(\bar{x}^k) \\ \cdot \\ \cdot \\ f_n(\bar{x}^k) \end{bmatrix}.$$

IV step. There is a value of next point

$$\bar{x}^{k+1} = \bar{x}^k + \Delta\bar{x}^k,$$

or

$$x_1^{k+1} = x_1^k + \Delta x_1^k, \quad x_2^{k+1} = x_2^k + \Delta x_2^k, \quad \dots, \quad x_n^{k+1} = x_n^k + \Delta x_n^k.$$

Farther an iteration process recurs beginning from the first step.

Will consider application of метода Newton for the decision of the next nonlinear system of the third order

$$\begin{cases} x^2 + x - 2yz = 0,1 \\ -y^2 + y + 3xz = 0,2 . \\ z^2 + z - 2xy = 0 \end{cases}$$

Will present the initial system of equations in a kind

$$\begin{cases} f_1(x, y, z) = 0, \\ f_2(x, y, z) = 0, \\ f_3(x, y, z) = 0, \end{cases}$$

where

$$\begin{aligned} f_1(x, y, z) &= x^2 + x - 2yz + 0.1, \\ f_2(x, y, z) &= y^2 - y - 3xz + 0.2, \\ f_3(x, y, z) &= z^2 + z - 2xy. \end{aligned}$$

The matrix of Jacobi of the initial system of equations looks like

$$J(x, y, z) = \begin{pmatrix} 2x + 1 & -2x & -2y \\ -3z & 2y - 1 & -3x \\ -2y & -2x & 2z + 1 \end{pmatrix}.$$

A graphic system of equations image is presented on rice.4.10, coming from what for the initial approaching lay $x_0 = 0$; $y_0 = 0$; $z_0 = 0$.

And step.

$$\begin{aligned} f_1(x_0, y_0, z_0) &= -0.1; \\ f_2(x_0, y_0, z_0) &= 0.2; \\ f_3(x_0, y_0, z_0) &= 0; \end{aligned} \quad J(x_0, y_0, z_0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

$$\Delta x^1 = 0.1; \quad x_1 = x_0 + \Delta x^1 = 0.1;$$

$$\Delta y^1 = 0.2; \quad y_1 = y_0 + \Delta y^1 = 0.2;$$

$$\Delta z^1 = 0; \quad z_1 = z_0 + \Delta z^1 = 0.$$

II step.

$$\begin{aligned} f_1(x_1, y_1, z_1) &= 0.01; \\ f_2(x_1, y_1, z_1) &= 0.04; \\ f_3(x_1, y_1, z_1) &= -0.04; \end{aligned} \quad J(x_1, y_1, z_1) = \begin{pmatrix} 1.2 & 0 & -0.4 \\ 0 & -0.6 & -0.3 \\ -0.4 & -0.2 & 1 \end{pmatrix};$$

$$\begin{aligned} \Delta x^2 &= 0.0089; & x_2 &= x_1 + \Delta x^2 = 0.1089; \\ \Delta y^2 &= 0.0408; & y_2 &= y_1 + \Delta y^2 = 0.2408; \\ \Delta z^2 &= 0.0517; & z_2 &= z_1 + \Delta z^2 = 0.0517. \end{aligned}$$

III step.

$$\begin{aligned} f_1(x_2, y_2, z_2) &= -0.0041; \\ f_2(x_2, y_2, z_2) &= 0.0003; \\ f_3(x_2, y_2, z_2) &= 0.0019; \end{aligned}$$

$$J(x_2, y_2, z_2) = \begin{pmatrix} 1.2178 & -0.1034 & -0.4816 \\ -0.1552 & -0.5184 & -0.3267 \\ -0.4816 & -0.2178 & 1.1034 \end{pmatrix};$$

$$\begin{aligned} \Delta x^3 &= 0.0032; & x_3 &= x_2 + \Delta x^3 = 1.1121; \\ \Delta y^3 &= -0.0002; & y_3 &= y_2 + \Delta y^3 = 0.2406; \\ \Delta z^3 &= -0.0004; & z_3 &= z_2 + \Delta z^3 = 0.0513. \end{aligned}$$

If to compare the brought results over with corresponding results, which are got the method of simple iteration (table 4.8), then see that for achievement of necessary exactness in the method of Newton it is needed to execute three

iterations (concordantly метода of simple iteration it is necessary it was to execute eight iterations).

A task is for independent implementation

I. To find the least on the module actual root of equation with exactness, here: а) to separate a root a graphic method; б) to calculate a root by means of method of simple iteration; в) to calculate a root by means of method of Newton – Rafson and compare the amount of iterations which are needed for the receipt of root with the set exactness in obedience to the method of simple iteration.

1. $x^2 - 2x + \ln x = 0.$

2. $x^4 - 6x^2 + 12x - 8 = 0.$

3. $x^2 - 2\lg(x + 2) = 0.$

4. $2^x + 2x^2 - 3 = 0.$

5. $x^3 + 2x - 13 = 0.$

6. $x^2 + \operatorname{arctg} x - 0.5 = 0.$

7. $xe^{2x} - 4 = 0.$

8. $\operatorname{ctg} 0.8x - 2x^2 = 0.$

9. $x^5 + 5x + 1 = 0.$

10. $x^5 + 18x^3 - 34 = 0.$

11. $(x - 2)^2 - e^x = 0.$

12. $2e^{x^2} - 5 = 0.$

13. $x^3 + 2x^2 - 11 = 0.$

14. $2e^{-x^2} - 3x + 4 = 0.$

15. $x^2 - 1 - \cos 1.2x = 0.$

16. $2x - 3\sin 2x - 1 = 0.$

$$17. (x - 0.5)^2 - \sin \pi x = 0.$$

$$18. x^3 + 3x^2 - 6x - 1 = 0.$$

$$19. x^3 - 2 \cos \pi x = 0.$$

$$20. \operatorname{tg} 0.8x - x - 2 = 0.$$

$$21. \operatorname{tg} 1.2x - 2 + 3x = 0.$$

$$22. x^4 + 3x - 3 = 0.$$

$$23. (x - 1)^2 - 0.5e^x = 0.$$

$$24. 3 - x^3 + \sin \frac{\pi}{2} x = 0.$$

$$25. 1 - \arcsin 0.5x = 0.$$

$$26. (x + 2) \log_2(x) - 1 = 0.$$

$$27. \operatorname{arcctg}(x - 1) + 2x - 3 = 0.$$

$$28. 2x^4 - x^2 - 10 = 0.$$

$$29. 2 \lg x - \frac{x}{2} + 1 = 0.$$

$$30. x^2 \cos 2x + 1 = 0.$$

II. Using the method of iterations, to untie the system of nonlinear equations within 0,0001.

$$1. \begin{cases} \sin(x + 1) - y = 1,2; \\ 2x + \cos y = 2. \end{cases}$$

$$2. \begin{cases} \cos(x - 1) + y = 0,5; \\ x - \cos y = 3. \end{cases}$$

$$3. \begin{cases} \sin x + 2y = 2; \\ \cos(y - 1) + x = 0,7. \end{cases}$$

$$4. \begin{cases} \cos x + y = 1,5; \\ 2x - \sin(y - 0,5) = 1. \end{cases}$$

$$5. \begin{cases} \sin(x + 0,5) - y = 1; \\ \cos(y - 2) + x = 0. \end{cases}$$

$$6. \begin{cases} \cos(x + 0,5) + y = 0,8; \\ \sin y - 2x = 1,6. \end{cases}$$

$$7. \begin{cases} \sin(x-1) = 1,3 - y; \\ x - \sin(y+1) = 0,8. \end{cases}$$

$$8. \begin{cases} 2y - \cos(x+1) = 0; \\ x + \sin y = -0,4. \end{cases}$$

$$9. \begin{cases} \cos(x+0,5) - y = 2; \\ \sin y - 2x = 1. \end{cases}$$

$$10. \begin{cases} \sin(x+2) - y = 1,5; \\ x + \cos(y-2) = 0,5. \end{cases}$$

$$11. \begin{cases} \sin(y+1) - y = 1,2; \\ 2y + \cos x = 2. \end{cases}$$

$$12. \begin{cases} \cos(y-1) + x = 0,5; \\ y - \cos x = 3. \end{cases}$$

$$13. \begin{cases} \sin y + 2x = 2; \\ \cos(x-1) + y = 0,7. \end{cases}$$

$$14. \begin{cases} \cos y + x = 1,5; \\ 2y - \sin(x-0,5) = 1. \end{cases}$$

$$15. \begin{cases} \sin(y+0,5) - x = 1; \\ \cos(x-2) + y = 0. \end{cases}$$

$$16. \begin{cases} \cos(y+0,5) + x = 0,8; \\ \sin x - 2y = 1,6. \end{cases}$$

$$17. \begin{cases} \sin(y-1) + x = 1,3; \\ y - \sin(x+1) = 0,8. \end{cases}$$

$$18. \begin{cases} 2x - \cos(y+1) = 0; \\ y + \sin x = -0,4. \end{cases}$$

$$19. \begin{cases} \cos(y+0,5) - x = 2; \\ \sin x - 2y = 1. \end{cases}$$

$$20. \begin{cases} \sin(y+2) - x = 1,5; \\ y + \cos(x-2) = 0,5. \end{cases}$$

$$21. \begin{cases} \sin(x+1) - y = 1; \\ 2x + \cos y = 2. \end{cases}$$

$$22. \begin{cases} \cos(x-1) + y = 0,8; \\ x - \cos y = 2. \end{cases}$$

$$23. \begin{cases} \sin x + 2y = 1,6; \\ \cos(y-1) + x = 1. \end{cases}$$

$$24. \begin{cases} \cos x + y = 1,2; \\ 2x - \sin(y-0,5) = 2. \end{cases}$$

$$25. \begin{cases} \sin(x + 0,5) - y = 1,2; \\ \cos(y - 2) + x = 0. \end{cases}$$

$$26. \begin{cases} \cos(x + 0,5) + y = 1; \\ \sin y - 2x = 2. \end{cases}$$

$$27. \begin{cases} \sin(x - 1) + y = 1,5; \\ x - \sin(y + 1) = 1. \end{cases}$$

$$28. \begin{cases} \sin(y + 1) - x = 1; \\ 2y + \cos x = 2. \end{cases}$$

$$29. \begin{cases} \cos(y - 1) + x = 0,8; \\ y - \cos x = 2. \end{cases}$$

$$30. \begin{cases} \cos(x - 1) + y = 1; \\ \sin y + 2x = 1,6. \end{cases}$$

III. Using the method of Newton, to untie the system of nonlinear equations within 0,001.

$$1. \begin{cases} \operatorname{tg}(xy + 0,4) = x^2; \\ 0,6x^2 + 2y^2 = 1, \quad x > 0, \quad y > 0. \end{cases}$$

$$2. \begin{cases} \sin(x + y) - 1,6x = 0; \\ x^2 + y^2 = 1, \quad x > 0, \quad y > 0. \end{cases}$$

$$3. \begin{cases} \operatorname{tg}(xy + 0,1) = x^2; \\ x^2 + 2y^2 = 1. \end{cases}$$

$$4. \begin{cases} \sin(x + y) - 1,2x = 0,2; \\ x^2 + y^2 = 1. \end{cases}$$

$$5. \begin{cases} \operatorname{tg}(xy + 0,3) = x^2; \\ 0,9x^2 + 2y^2 = 1. \end{cases}$$

$$6. \begin{cases} \sin(x + y) - 1,3x = 0; \\ x^2 + y^2 = 1. \end{cases}$$

$$7. \begin{cases} \operatorname{tg} xy = x^2; \\ 0,8x^2 + 2y^2 = 1. \end{cases}$$

$$8. \begin{cases} \sin(x + y) - 1,5x = 0,1; \\ x^2 + y^2 = 1. \end{cases}$$

$$9. \begin{cases} \operatorname{tg} xy = x^2; \\ 0,7x^2 + 2y^2 = 1. \end{cases}$$

$$10. \begin{cases} \sin(x + y) - 1,2x = 0,1; \\ x^2 + y^2 = 1. \end{cases}$$

$$11. \begin{cases} \operatorname{tg}(xy + 0,2) = x^2; \\ 0,6x^2 + 2y^2 = 1. \end{cases}$$

$$12. \begin{cases} \sin(x + y) = 1,5x - 0,1; \\ x^2 + y^2 = 1. \end{cases}$$

$$13. \begin{cases} \operatorname{tg}(xy + 0,4) = x^2; \\ 0,8x^2 + 2y^2 = 1. \end{cases}$$

$$14. \begin{cases} \sin(x + y) = 1,2x - 0,1; \\ x^2 + y^2 = 1. \end{cases}$$

$$15. \begin{cases} \operatorname{tg}(xy + 0,1) = x^2; \\ 0,9x^2 + 2y^2 = 1. \end{cases}$$

$$16. \begin{cases} \sin(x + y) - 1,4x = 0; \\ x^2 + y^2 = 1. \end{cases}$$

$$17. \begin{cases} \operatorname{tg}(xy + 0,1) = x^2; \\ 0,5x^2 + 2y^2 = 1. \end{cases}$$

$$18. \begin{cases} \sin(x + y) = 1,1x - 0,1; \\ x^2 + y^2 = 1. \end{cases}$$

$$19. \begin{cases} \operatorname{tg}(x - y) - xy = 0; \\ x^2 + 2y^2 = 1. \end{cases}$$

$$20. \begin{cases} \sin(x - y) - xy = -1; \\ x^2 - y^2 = \frac{3}{4}. \end{cases}$$

$$21. \begin{cases} \operatorname{tg}(xy + 0,2) = x^2; \\ x^2 + y^2 = 1. \end{cases}$$

$$22. \begin{cases} \sin(x + y) - 1,5x = 0; \\ x^2 + y^2 = 1. \end{cases}$$

$$23. \begin{cases} \operatorname{tg} xy = x^2; \\ 0,5x^2 + 2y^2 = 1. \end{cases}$$

$$24. \begin{cases} \sin(x + y) = 1,2x - 0,2; \\ x^2 + y^2 = 1. \end{cases}$$

$$25. \begin{cases} \operatorname{tg}(xy + 0,1) = x^2; \\ 0,7x^2 + 2y^2 = 1. \end{cases}$$

$$26. \begin{cases} \sin(x + y) - 1,5x = 0,2; \\ x^2 + y^2 = 1. \end{cases}$$

$$27. \begin{cases} \operatorname{tg} xy = x^2; \\ 0,6x^2 + 2y^2 = 1. \end{cases}$$

$$28. \begin{cases} \sin(x + y) - 1,2x = 0; \\ x^2 + y^2 = 1. \end{cases}$$

$$29. \begin{cases} \operatorname{tg}(xy + 0,3) = x^2; \\ 0,5x^2 + 2y^2 = 1. \end{cases}$$

$$30. \begin{cases} \sin(2x - y) - 1,2x = 0,4; \\ 0,8x^2 + 1,5y^2 = 1. \end{cases}$$

TITLE 3. NUMERICAL APPROACH OF FUNCTIONS

§ 3.1. Formulation of the problem

The simplest case of function approximation is the interpolation of the function of one variable. Suppose points are given and we need to find a function $f(x)$ that passes through these points, that is

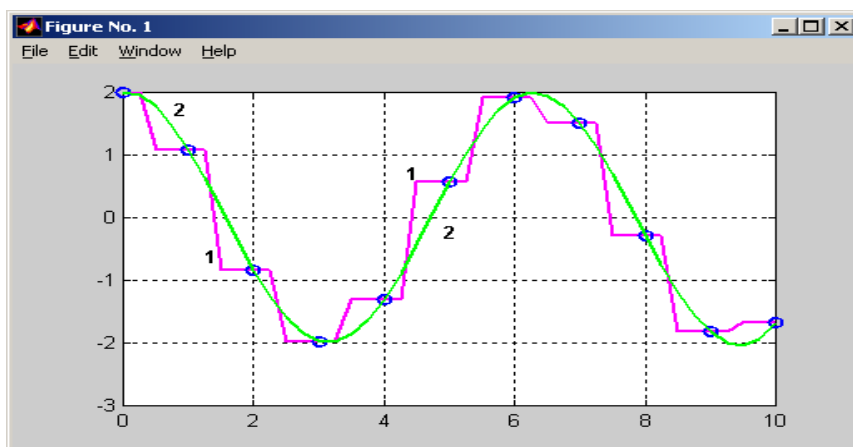
$$f(x_i) = y_i, \quad i = \overline{1, n}.$$

A function that interpolates the source data is called an interpolator or **interpolating function**.

The fixed data set $(x_i, y_i), i = \overline{1, n}$ alone cannot determine the interpolator. Obviously, there are infinitely many interpolators for a fixed data set.

In particular, as an example, in Fig. 1 shows the curves $y = 2\cos x$ that interpolate a discretely given function, according to step interpolation and spline interpolation, respectively denoted by the numbers '1' and '2'.

Figure. 1



We will assume that the set of points $\{x_i\}$ is arranged in ascending order $x_1 < x_2 < x_3 < \dots < x_n$. The problem is finding an interpolant $f(x)$ that gives acceptable values for $x \neq x_i$. This cannot be done completely rigorously because it all depends on the process under study, our understanding of the acceptability of the source data, etc.

In the standard approach, a set of basic functions $b_1(x), b_2(x), \dots, b_n(x)$ is predefined. These functions can be selected for experience, on the basis of mathematical or physical intuition, etc. But in any case, it is relies that these functions are known to be known.

Let the set of points $\{x_i\}$ be arranged in ascending order $x_1 < x_2 < x_3 < \dots < x_n$. The problem is finding an interpolant $f(x)$ that gives acceptable values for $x \neq x_i$. This cannot be done completely rigorously because it all depends on the process under study, the acceptability of the raw data, and so on. In the standard approach, a set of basic functions is predefined.

These functions can be selected from experience, based on mathematical or physical intuition and the so on. In any case, these functions are known. Based on the basic functions, the model is built

$$f(x) = \sum_{j=1}^n \alpha_j b_j(x), \quad (1)$$

in which the numbers α_j are unknown and are determined so that the function $f(x)$ is an interpolator. Therefore, the model must satisfy the interpolation conditions

$$f(x_i) = y_i, \quad \text{or} \quad \sum_{j=1}^n \alpha_j b_j(x) = y_i, \quad i = \overline{1, n}. \quad (2)$$

From here we obtain a system of n linear equations for α_j with the coefficient matrix

$$B = [b_{ij}], \quad b_{ij} = b_j(x_i), \quad i, j = \overline{1, n}.$$

If the functions $b_j(x)$ are chosen well enough, then the system of equations of form (2) can be solved relative to α_j and the interpolator $f(x)$ is found.

Two main approaches to the choice of basic functions are considered in interpolation problems: polynomial and piece polynomial. For each approach, matrix B has its own specificity, which in turn allows us to find effective interpolates.

For historical and practical reasons, when considering the polynomial interpolation the greatest use was obtained the class of basis functions, which is a set of algebraic polynomials. Polynomials have obvious advantages: they can be easily calculated, added, multiplied, integrated and differentiated.

It is clear that a class of functions may satisfy these conditions, but may not be suitable for approximation of functions. On the other hand, it is known that any continuous function $g(x)$ can be approximated on a closed interval by some polynomial. This follows from the Weierstrass approximation theorem.

Although it is theoretically known about the existence of some polynomial $P_n(x)$ that approximates a function $g(x)$ with some precision on $[a, b]$, but there is no guarantee that such a polynomial can be found using a practical algorithm.

If we choose for the basic functions $b_i(x) = x^{i-1}$, $i = \overline{1, n}$; the model (1) will take the form $P_{n-1}(x) = \alpha_1 + \alpha_2 x + \dots + \alpha_n x^{n-1}$ with a matrix B has a kind

$$B = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}.$$

The determinant of the matrix B

$$\det(B) = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{vmatrix} = \prod_{1 \leq j < i \leq n} (x_i - x_j)$$

is called Vandermonde's determinant. If the interpolation nodes are different, that is $\det(B) \neq 0$, then system (2) has a single solution.

Consider the partial cases of the above interpolation approach. In the case of **linear interpolation**, the unique interpolator for points (x_1, y_1) and (x_2, y_2) is determined by the formula

$$P_1(x) = \alpha_1 + \alpha_2 x,$$

where α_1 and α_2 satisfy the system of equations

$$\begin{cases} \alpha_1 + \alpha_2 x_1 = y_1, \\ \alpha_1 + \alpha_2 x_2 = y_2. \end{cases}$$

The matrix $B = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}$ corresponds to the specified system. If $x_1 \neq x_2$, then

matrix B is nondegenerate and one can find α_1 and α_2 . A linear interpolator has the form

$$P_1(x) = \frac{y_1x_2 - y_2x_1}{x_2 - x_1} + \frac{y_2 - y_1}{x_2 - x_1} x. \quad (3)$$

In the case of **quadratic interpolation**, the interpolator for points (x_1, y_1) , (x_2, y_2) , (x_3, y_3) is given by the formula

$$P_2(x) = \alpha_1 + \alpha_2x + \alpha_3x^2, \quad (4)$$

where $\alpha_1, \alpha_2, \alpha_3$ satisfy the system of equations

$$\begin{cases} \alpha_1 + \alpha_2x_1 + \alpha_3x_1^2 = y_1, \\ \alpha_1 + \alpha_2x_2 + \alpha_3x_2^2 = y_2, \\ \alpha_1 + \alpha_2x_3 + \alpha_3x_3^2 = y_3. \end{cases}$$

The matrix $B = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}$ corresponds to the specified system.

If $x_1 \neq x_2 \neq x_3$, then matrix B is nondegenerate and the determinant of the system is $\det(B) = (x_3 - x_2)(x_3 - x_1)(x_2 - x_1) \neq 0$. The values $\alpha_1, \alpha_2, \alpha_3$ can be found by Kramer formulas

$$\alpha_1 = \frac{\det(B_1)}{\det(B)}, \quad \alpha_2 = \frac{\det(B_2)}{\det(B)}, \quad \alpha_3 = \frac{\det(B_3)}{\det(B)}, \quad (5)$$

where the matrix B_i ($i = 1, 2, 3$) is obtained by replacing the corresponding vector -column with the vector of the right-hand sides of the original system of equations. The determinants of these matrices are given by formulas

$$\det(B_1) = y_1 x_2 x_3 (x_3 - x_2) - y_2 x_1 x_3 (x_3 - x_1) + y_3 x_1 x_2 (x_2 - x_1),$$

$$\det(B_2) = -y_1 (x_3^2 - x_2^2) + y_2 (x_3^2 - x_1^2) - y_3 (x_2^2 - x_1^2),$$

$$\det(B_3) = y_1 (x_3 - x_2) - y_2 (x_3 - x_1) + y_3 (x_2 - x_1).$$

In the examples of linear and quadratic interpolations, the coefficient matrix B is nondegenerate, so the systems of equations are solved uniquely. In the general case, as already mentioned, if the interpolation nodes do not coincide, then matrix B has a determinant $\det(B) \neq 0$.

Thus, if no two abscissa of the original data coincide, then the system of equations for polynomial interpolation always has a nondegenerate matrix of coefficients and, accordingly, a single solution, that is, for given n points, there exists a single polynomial of degree not higher than $n - 1$, which passes through all these points.

§ 3.2. Lagrange interpolation polynomial

A linear interpolator of the form (3) is considered. By equivalent transformations $P_1(x)$ can be represented as

$$P_1(x) = y_1 \frac{x - x_2}{x_1 - x_2} + y_2 \frac{x - x_1}{x_2 - x_1}, \quad (6)$$

or

$$P_1(x) = \alpha_1 b_1(x) + \alpha_2 b_2(x), \quad \text{де } b_1(x) = \frac{x - x_2}{x_1 - x_2}; \quad b_2(x) = \frac{x - x_1}{x_2 - x_1}.$$

In this case, the basic functions $b_1(x)$, $b_2(x)$ are accepted for interpolation.

The corresponding matrix B has the form

$$B = \begin{bmatrix} b_1(x_1) & b_2(x_1) \\ b_1(x_2) & b_2(x_2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

that is, the basic functions $b_1(x)$, $b_2(x)$ satisfy the conditions

$$b_j(x_i) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \quad (7)$$

Therefore, matrix B is singular. In this case, the solution of the equation system $B\bar{\alpha} = \bar{y}$ will be $\bar{\alpha} = \bar{y}$ either $\alpha_1 = y_1$, $\alpha_2 = y_2$.

In the case of quadratic interpolation of the form (4) by means of equivalent transformations, we have

$$P_2(x) = y_1 \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + y_2 \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + y_3 \frac{(x-x_3)(x-x_1)}{(x_3-x_1)(x_3-x_2)}, \quad (8)$$

or

$$P_2(x) = \alpha_1 b_1(x) + \alpha_2 b_2(x) + \alpha_3 b_3(x),$$

where

$$b_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}, \quad b_2(x) = \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)},$$

$$b_3(x) = \frac{(x-x_3)(x-x_1)}{(x_3-x_1)(x_3-x_2)}.$$

In this case, during interpolation for the basis functions $b_1(x)$, $b_2(x)$, $b_3(x)$ is taken which satisfy condition (7). The corresponding matrix B is a unit $B = E$. In this case, the solution of the equation system $B\bar{\alpha} = \bar{y}$ is $\bar{\alpha} = \bar{y}$ or $\alpha_1 = y_1$, $\alpha_2 = y_2$, $\alpha_3 = y_3$.

Let us generalize the results obtained when considering the linear and quadratic interpolations of the form (6), (8).

Suppose that we have a set of functions $l_1(x), l_2(x), \dots, l_n(x)$, each of which is a polynomial of degree $n - 1$ and satisfies the condition

$$l_j(x_i) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

That is, functions $l_j(x)$ take values of 1 at points $x = x_j$ and are zero for all other values $x = x_i$ ($i \neq j$). It should be noted that functions $b_j(x)$ with similar properties were used to construct the interpolants of the form (6), (8).

Any linear combination of functions $l_j(x)$ is a polynomial of degree no more $(n - 1)$. In particular, we consider a polynomial

$$P_{n-1}(x) = y_1 l_1(x) + y_2 l_2(x) + \dots + y_n l_n(x). \quad (9)$$

According to the properties of the functions $l_j(x)$ it follows that

$$P_{n-1}(x_i) = y_1 l_1(x_i) + y_2 l_2(x_i) + \dots + y_n l_n(x_i) = y_i l_i(x_i) = y_i,$$

and $P_{n-1}(x)$ is an interpolation polynomial. Functions are defined by formulas

$$l_j(x) = \frac{(x - x_1)(x - x_2) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_n)}{(x_j - x_1)(x_j - x_2) \dots (x_j - x_{j-1})(x_j - x_{j+1}) \dots (x_j - x_n)} = \frac{\prod_{i=1, i \neq j}^n (x - x_i)}{\prod_{i=1, i \neq j}^n (x_j - x_i)}. \quad (10)$$

The polynomial of the form (9), in which the coefficients $l_j(x)$ are determined according to formula (10), is called the **Lagrange interpolation polynomial**.

When finding a Lagrange polynomial, in some cases it is convenient to use the **Aitkin interpolation scheme**, a feature of which is the uniformity of calculations.

If the function f is given at two points x_0 and x_1 (its values are respectively equal y_0 to and y_1), then its value at the point $x \in (x_0; x_1)$ can be calculated by the linear interpolation formula (6). If the value of the function f at x is denoted by $P_{0,1}(x)$, then the linear interpolation formula (.6) can be written in equilateral form

$$P_{0,1}(x) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x \\ y_1 & x_1 - x \end{vmatrix},$$

the right part of which contains the second-order determinant. It's easy to make sure

$$P_{0,1}(x_0) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x_0 \\ y_1 & x_1 - x_0 \end{vmatrix} = y_0, \quad P_{0,1}(x_1) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x_1 \\ y_1 & x_1 - x_1 \end{vmatrix} = y_1.$$

Now let the function f be given at three points x_0, x_1, x_2 (corresponding values y_0, y_1, y_2), and it is necessary to calculate its value at a point $x \in (x_0; x_2), x \neq x_1$. In this case, the values of two linear polynomials are first calculated using the Eitkin scheme at x .

$$P_{0,1}(x) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x \\ y_1 & x_1 - x \end{vmatrix} \quad \text{and} \quad P_{1,2}(x) = \frac{1}{x_2 - x_1} \begin{vmatrix} y_1 & x_1 - x \\ y_2 & x_2 - x \end{vmatrix},$$

and then the quadratic three-term view

$$P_{0,1,2}(x) = \frac{1}{x_2 - x_0} \begin{vmatrix} P_{0,1}(x) & x_0 - x \\ P_{1,2}(x) & x_2 - x \end{vmatrix}. \quad (11)$$

Direct verification convinces you that $P_{0,1}(x_0) = y_0$, $P_{0,1}(x_1) = y_1$, $P_{1,2}(x_1) = y_1$, $P_{1,2}(x_2) = y_2$, $P_{0,1,2}(x_1) = y_1$, $P_{0,1,2}(x_2) = y_2$.

We prove that $P_{0,1,2}(x)$ coincides with the second-order Lagrange interpolation polynomial. Indeed, since

$$P_{0,1}(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1, \quad P_{1,2}(x) = \frac{x - x_2}{x_1 - x_2} y_1 + \frac{x - x_1}{x_2 - x_1} y_2,$$

revealing the determinant of the second order of formula (11), we obtain

$$\begin{aligned} P_{0,1,2}(x) &= \frac{1}{x_2 - x_0} \left[\left(\frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 \right) (x_2 - x) - \right. \\ &\quad \left. - \left(\frac{x - x_2}{x_1 - x_2} y_1 + \frac{x - x_1}{x_2 - x_1} y_2 \right) (x_0 - x) \right] = \\ &= \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} y_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} y_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} y_2. \end{aligned}$$

This scheme generalizes to higher-degree interpolation polynomials. If the function f is set at four nodes, then cubic interpolation is performed by the formula

$$P_{0,1,2,3}(x) = \frac{1}{x_3 - x_0} \begin{vmatrix} P_{0,1,2}(x) & x_0 - x \\ P_{1,2,3}(x) & x_3 - x \end{vmatrix},$$

where $P_{0,1,2}(x)$ и $P_{1,2,3}(x)$ – the value of the quadratic three terms in a point $x \in (x_0; x_3)$, $x \neq x_1$, $x \neq x_2$. In this case, the values of the polynomials $P_{0,1,2}(x)$ are calculated by the formula (11) and the values $P_{1,2,3}(x)$ by the formula:

$$P_{1,2,3}(x) = \frac{1}{x_3 - x_1} \begin{vmatrix} P_{1,2}(x) & x_1 - x \\ P_{2,3}(x) & x_3 - x \end{vmatrix}, \quad P_{2,3}(x) = \frac{1}{x_3 - x_2} \begin{vmatrix} y_2 & x_2 - x \\ y_3 & x_3 - x \end{vmatrix}.$$

By direct verification we make sure that $P_{1,2,3}(x_1) = y_1$, $P_{1,2,3}(x_2) = y_2$, $P_{1,2,3}(x_3) = y_3$, $P_{0,1,2,3}(x_i) = y_i$, ($i = 0, 1, 2, 3$) and $P_{0,1,2,3}(x)$, and coincide with the Lagrangian cubic interpolation polynomial.

Generally, if at $(n + 1)$ interpolation nodes the function f acquires values y_i ($i = 0, 1, \dots, n$), then the value of the interpolation polynomial of degree n at a point $x \in (x_0; x_n)$ that does not coincide with the interpolation nodes can be calculated by the formula

$$P_{0,1,\dots,n}(x) = \frac{1}{x_n - x_0} \begin{vmatrix} P_{0,1,\dots,n-1}(x) & x_0 - x \\ P_{1,2,\dots,n}(x) & x_n - x \end{vmatrix},$$

where $P_{0,1,\dots,n-1}(x)$ and $P_{1,2,\dots,n}(x)$ are the values of the interpolation polynomials of $(n - 1)$ degree calculated at the point x in the previous step of the calculations. It is easy to be sure that $P_{0,1,\dots,n}(x_i) = y_i$, ($i = 0, 1, \dots, n$) and $P_{0,1,\dots,n}(x)$ coincides with the n -th degree Lagrange interpolation polynomial.

Therefore, to calculate at point x the value of the n -th degree interpolation polynomial according to the Aitkin scheme (Table 3.1), it is necessary to calculate at this point the values of n linear, $n - 1$ quadratic, $n - 2$ cubic polynomials, etc.,

two polynomials $(n - 1)$ –th degree and, finally, one n -th degree polynomial. All these polynomials are expressed in terms of the second-order determinant, and this makes the computations homogeneous, cyclic.

Table 3.1.

Interpolation Aitkin scheme

x_i	y_i	$P_{i-1,i}$	$P_{i-2,i-1,i}$	$P_{i-3,i-2,i-1,i}$	$x_i - x$
x_0	y_0	–	–	–	$x_0 - x$
x_1	y_1	$P_{0,1}(x)$	–	–	$x_1 - x$
x_2	y_2	$P_{1,2}(x)$	$P_{0,1,2}(x)$	–	$x_2 - x$
x_3	y_3	$P_{2,3}(x)$	$P_{1,2,3}(x)$	$P_{1,2,3,0}(x)$	$x_3 - x$
...

In calculations according to this scheme, new nodes x_i (corresponding to the transition to higher-degree interpolation polynomials) involve until the calculations themselves show that the required accuracy has already been achieved.

Example: The function $y = \sqrt[3]{x}$ is set as follows:

x	1,0	1,1	1,3	1,5	1,6
y	1,000	1,032	1,091	1,145	1,170

Apply the Aitkin scheme to find the value $\sqrt[3]{1,15}$. We write the output of the values of the function in table. 3.2 and calculate the difference at. Then we find consistently

$$P_{0,1} = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x \\ y_1 & x_1 - x \end{vmatrix} = \frac{1}{0,1} \begin{vmatrix} 1 & -0,15 \\ 1,032 & -0,05 \end{vmatrix} = 1,048,$$

$$P_{1,2} = \frac{1}{x_2 - x_1} \begin{vmatrix} y_1 & x_1 - x \\ y_2 & x_2 - x \end{vmatrix} = \frac{1}{0,2} \begin{vmatrix} 1,032 & -0,05 \\ 1,091 & 0,15 \end{vmatrix} = 1,047,$$

$$P_{2,3} = \frac{1}{x_3 - x_2} \begin{vmatrix} y_2 & x_2 - x \\ y_3 & x_3 - x \end{vmatrix} = \frac{1}{0,2} \begin{vmatrix} 1,091 & 0,15 \\ 1,145 & 0,35 \end{vmatrix} = 1,050,$$

$$P_{3,4} = \frac{1}{x_4 - x_3} \begin{vmatrix} y_3 & x_3 - x \\ y_4 & x_4 - x \end{vmatrix} = \frac{1}{0,1} \begin{vmatrix} 1,145 & 0,35 \\ 1,170 & 0,45 \end{vmatrix} = 1,057.$$

Table 3.2

i	x	y	$x_i - x$	$P_{i-1, i}$	$P_{i-2, i-1, i}$
0	1,0	1,000	-0,15		
1	1,1	1,032	-0,05	1,048	
2	1,3	1,091	0,15	1,047	1,048
3	1,5	1,145	0,35	1,050	
4	1,6	1,170	0,45	1,057	

The calculated values are entered in table. 2, after which we calculate

$$P_{0,1,2} = \frac{1}{0,3} \begin{vmatrix} 1,048 & -0,15 \\ 1,047 & 0,15 \end{vmatrix} = 1,048.$$

The values $P_{0,1}$ and $P_{0,1,2}$ coincide with the third character. This calculation can be completed up to $\varepsilon = 10^{-3}$ and written down $\sqrt[3]{1,15} = 1,048$.

§ 3.3. Error estimation of Lagrange interpolation formula

If the function f on a segment $[a; b]$ is a polynomial of degree less than or equal to n , it follows from the unity of the interpolation polynomial that the interpolation polynomial $P_n(x)$ is also equal to f , that is $f(x) - P_n(x) \equiv 0, x \in [a; b]$.

If f on a segment $[a; b]$ that containing interpolation nodes $x_i (i = \overline{0, n})$ is not a polynomial of degree less than or equal to n , then the difference

$$R_n(f, x) = f(x) - P_n(x) \quad (12)$$

will be zero only at the interpolation nodes $x_i (i = \overline{0, n})$, and at other points of the segment is different from the identity zero. A function $R_n(f, x)$ that characterizes the accuracy of the approximation of a function f to an interpolation polynomial $P_n(x)$ is called a **residual term** of the Lagrangian interpolation formula or an **interpolation error**. If an analytic expression of the function f is known, then it can be estimated. This holds true for this theorem.

Theorem. If the interpolation nodes $x_i (i = \overline{0, n})$ are different and belong to the segment $[a; b]$, then for any point $x \in [a; b]$ there is such a point $\xi \in (a; b)$ that for the error of interpolation equals

$$R_n(f, x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{j=0}^n (x - x_j),$$

where

$$\prod_{j=0}^n (x - x_j) = (x - x_0)(x - x_1) \dots (x - x_n).$$

If $M_{n+1} = \max_{x \in [a; b]} |f^{(n+1)}(x)|$, than for the absolute error of the Lagrange interpolation formula, we obtain the following estimate:

$$|R_n(f, x)| = |f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \left| \prod_{j=0}^n (x - x_j) \right|. \quad (13)$$

From formula (12) it is obvious that the residual term of the linear interpolation formula (6) is equal to

$$R_1(f, x) = \frac{\prod_{j=0}^1 (x - x_j)}{2!} f''(\xi),$$

where

$$\prod_{j=0}^1 (x - x_j) = (x - x_0)(x - x_1), \quad \xi \in (x_0; x_1);$$

and the residual term of the quadratic interpolation formula (8)

$$R_2(f, x) = \frac{\prod_{j=0}^2 (x - x_j)}{3!} f'''(\xi),$$

where

$$\prod_{j=0}^2 (x - x_j) = (x - x_0)(x - x_1)(x - x_2), \quad \xi \in (x_0; x_2).$$

For example, construct a Lagrange interpolation polynomial for a function $f(x) = \sqrt[3]{x}$ with interpolation nodes $x_0 = 1, x_1 = 2, x_2 = 3$. Estimate the error of an interpolation polynomial at a point $x = 2,5$.

The output is written as follows.

x_i	1	2	3
-------	---	---	---

$y_i = \sqrt[3]{x_i}$	1	1,280	1,442
-----------------------	---	-------	-------

The Lagrange interpolation polynomial for the case $n = 2$ is of the form

$$P_2(x) = y_1 \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x - x_3)(x - x_1)}{(x_3 - x_1)(x_3 - x_2)},$$

or taking into account the output

$$P_2(x) = \frac{(x - 2)(x - 3)}{2} \cdot 1 + \frac{(x - 1)(x - 3)}{-1} \cdot 1,260 + \\ + \frac{(x - 1)(x - 2)}{2} \cdot 1,442 = -0,039x^2 + 0,377x + 0,662.$$

To estimate the error, we use formula (13), which in this case has the form

$$|R_2(x)| \leq M_3 \frac{|(x - x_0)(x - x_1)(x - x_2)|}{3!},$$

where $M_3 = \max_{[1; 3]} |f'''(x)|$.

According to the condition $f(x) = \sqrt[3]{x}$, then

$$f'(x) = \frac{1}{3}x^{-2/3}, \quad f''(x) = -\frac{2}{9}x^{-5/3}, \quad f'''(x) = \frac{10}{27}x^{-8/3} = \frac{10}{27x^2 \sqrt[3]{x}}.$$

The function $f'''(x)$ is positive and descending on the interval $[1; 3]$.

Therefore,

$$|f'''(x)| \leq f'''(1) = \frac{10}{27} \approx 0,37$$

and

$$|R_2(2,5)| \leq 0,37 \frac{|(2,5-1)(2,5-2)(2,5-3)|}{3!}, \quad |R_2(2,5)| \leq 0,023.$$

It follows from formula (13) that the absolute error of the Lagrange interpolation formula is proportional to the product of two factors M_{n+1} and

$\left| \prod_{j=0}^n (x - x_j) \right|$, of which the value of the first depends only on the function f , and

the value of the second is determined solely by the choice of nodes of interpolation. The magnitude of the absolute error of the Lagrangian interpolation formula can be reduced by the choice of nodes of interpolation, at which the factor

$\left| \prod_{j=0}^n (x - x_j) \right|$ acquires the smallest maximum value per segment $[a; b]$.

TITLE 4. NUMERICAL APPROACH OF FUNCTIONS

§ 4.1. Formulation of the problem

The simplest case of function approximation is the interpolation of the function of one variable. Suppose points are given and we need to find a function $f(x)$ that passes through these points, that is

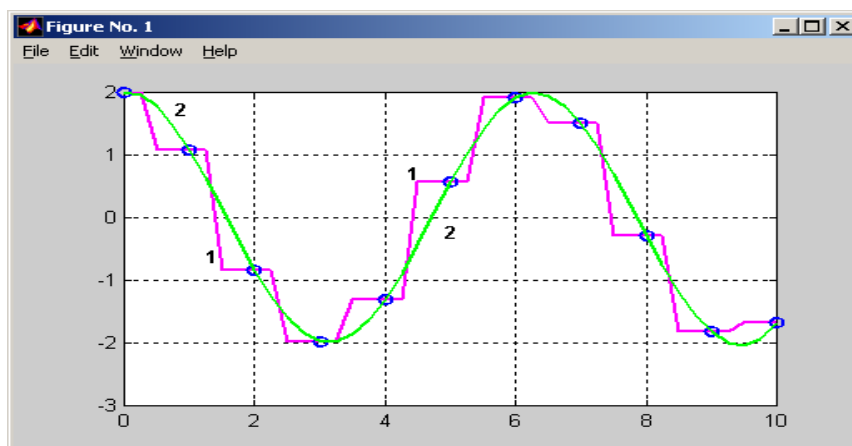
$$f(x_i) = y_i, \quad i = \overline{1, n}.$$

A function that interpolates the source data is called an interpolator or **interpolating function**.

The fixed data set $(x_i, y_i), i = \overline{1, n}$ alone cannot determine the interpolator. Obviously, there are infinitely many interpolators for a fixed data set.

In particular, as an example, in Fig. 1 shows the curves $y = 2 \cos x$ that interpolate a discretely given function, according to step interpolation and spline interpolation, respectively denoted by the numbers '1' and '2'.

Figure. 1



We will assume that the set of points $\{x_i\}$ is arranged in ascending order $x_1 < x_2 < x_3 < \dots < x_n$. The problem is finding an interpolant $f(x)$ that gives acceptable values for $x \neq x_i$. This cannot be done completely rigorously because it all depends on the process under study, our understanding of the acceptability of the source data, etc.

In the standard approach, a set of basic functions $b_1(x), b_2(x), \dots, b_n(x)$ is predefined. These functions can be selected for experience, on the basis of mathematical or physical intuition, etc. But in any case, it is relies that these functions are known to be known.

Let the set of points $\{x_i\}$ be arranged in ascending order $x_1 < x_2 < x_3 < \dots < x_n$. The problem is finding an interpolant $f(x)$ that gives acceptable values for $x \neq x_i$. This cannot be done completely rigorously because it all depends on the process under study, the acceptability of the raw data, and so on. In the standard approach, a set of basic functions is predefined.

These functions can be selected from experience, based on mathematical or physical intuition and the so on. In any case, these functions are known. Based on the basic functions, the model is built

$$f(x) = \sum_{j=1}^n \alpha_j b_j(x), \quad (1)$$

in which the numbers α_j are unknown and are determined so that the function $f(x)$ is an interpolator. Therefore, the model must satisfy the interpolation conditions

$$f(x_i) = y_i, \quad \text{or} \quad \sum_{j=1}^n \alpha_j b_j(x) = y_i, \quad i = \overline{1, n}. \quad (2)$$

From here we obtain a system of n linear equations for α_j with the coefficient matrix

$$B = [b_{ij}], \quad b_{ij} = b_j(x_i), \quad i, j = \overline{1, n}.$$

If the functions $b_j(x)$ are chosen well enough, then the system of equations of form (2) can be solved relative to α_j and the interpolator $f(x)$ is found.

Two main approaches to the choice of basic functions are considered in interpolation problems: polynomial and piece polynomial. For each approach, matrix B has its own specificity, which in turn allows us to find effective interpolates.

For historical and practical reasons, when considering the polynomial interpolation the greatest use was obtained the class of basis functions, which is a set of algebraic polynomials. Polynomials have obvious advantages: they can be easily calculated, added, multiplied, integrated and differentiated.

It is clear that a class of functions may satisfy these conditions, but may not be suitable for approximation of functions. On the other hand, it is known that any continuous function $g(x)$ can be approximated on a closed interval by some polynomial. This follows from the Weierstrass approximation theorem.

Although it is theoretically known about the existence of some polynomial $P_n(x)$ that approximates a function $g(x)$ with some precision on $[a, b]$, but there is no guarantee that such a polynomial can be found using a practical algorithm.

If we choose for the basic functions $b_i(x) = x^{i-1}$, $i = \overline{1, n}$; the model (1) will take the form $P_{n-1}(x) = \alpha_1 + \alpha_2 x + \dots + \alpha_n x^{n-1}$ with a matrix B has a kind

$$B = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}.$$

The determinant of the matrix B

$$\det(B) = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{vmatrix} = \prod_{1 \leq j < i \leq n} (x_i - x_j)$$

is called Vandermonde's determinant. If the interpolation nodes are different, that is $\det(B) \neq 0$, then system (2) has a single solution.

Consider the partial cases of the above interpolation approach. In the case of **linear interpolation**, the unique interpolator for points (x_1, y_1) and (x_2, y_2) is determined by the formula

$$P_1(x) = \alpha_1 + \alpha_2 x,$$

where α_1 and α_2 satisfy the system of equations

$$\begin{cases} \alpha_1 + \alpha_2 x_1 = y_1, \\ \alpha_1 + \alpha_2 x_2 = y_2. \end{cases}$$

The matrix $B = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}$ corresponds to the specified system. If $x_1 \neq x_2$, then

matrix B is nondegenerate and one can find α_1 and α_2 . A linear interpolator has the form

$$P_1(x) = \frac{y_1x_2 - y_2x_1}{x_2 - x_1} + \frac{y_2 - y_1}{x_2 - x_1} x. \quad (3)$$

In the case of **quadratic interpolation**, the interpolator for points (x_1, y_1) , (x_2, y_2) , (x_3, y_3) is given by the formula

$$P_2(x) = \alpha_1 + \alpha_2x + \alpha_3x^2, \quad (4)$$

where $\alpha_1, \alpha_2, \alpha_3$ satisfy the system of equations

$$\begin{cases} \alpha_1 + \alpha_2x_1 + \alpha_3x_1^2 = y_1, \\ \alpha_1 + \alpha_2x_2 + \alpha_3x_2^2 = y_2, \\ \alpha_1 + \alpha_2x_3 + \alpha_3x_3^2 = y_3. \end{cases}$$

The matrix $B = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}$ corresponds to the specified system.

If $x_1 \neq x_2 \neq x_3$, then matrix B is nondegenerate and the determinant of the system is $\det(B) = (x_3 - x_2)(x_3 - x_1)(x_2 - x_1) \neq 0$. The values $\alpha_1, \alpha_2, \alpha_3$ can be found by Kramer formulas

$$\alpha_1 = \frac{\det(B_1)}{\det(B)}, \quad \alpha_2 = \frac{\det(B_2)}{\det(B)}, \quad \alpha_3 = \frac{\det(B_3)}{\det(B)}, \quad (5)$$

where the matrix B_i ($i = 1, 2, 3$) is obtained by replacing the corresponding vector -column with the vector of the right-hand sides of the original system of equations. The determinants of these matrices are given by formulas

$$\det(B_1) = y_1 x_2 x_3 (x_3 - x_2) - y_2 x_1 x_3 (x_3 - x_1) + y_3 x_1 x_2 (x_2 - x_1),$$

$$\det(B_2) = -y_1 (x_3^2 - x_2^2) + y_2 (x_3^2 - x_1^2) - y_3 (x_2^2 - x_1^2),$$

$$\det(B_3) = y_1 (x_3 - x_2) - y_2 (x_3 - x_1) + y_3 (x_2 - x_1).$$

In the examples of linear and quadratic interpolations, the coefficient matrix B is nondegenerate, so the systems of equations are solved uniquely. In the general case, as already mentioned, if the interpolation nodes do not coincide, then matrix B has a determinant $\det(B) \neq 0$.

Thus, if no two abscissa of the original data coincide, then the system of equations for polynomial interpolation always has a nondegenerate matrix of coefficients and, accordingly, a single solution, that is, for given n points, there exists a single polynomial of degree not higher than $n - 1$, which passes through all these points.

§ 4.2. Lagrange interpolation polynomial

A linear interpolator of the form (3) is considered. By equivalent transformations $P_1(x)$ can be represented as

$$P_1(x) = y_1 \frac{x - x_2}{x_1 - x_2} + y_2 \frac{x - x_1}{x_2 - x_1}, \quad (6)$$

or

$$P_1(x) = \alpha_1 b_1(x) + \alpha_2 b_2(x), \quad \text{де } b_1(x) = \frac{x - x_2}{x_1 - x_2}; \quad b_2(x) = \frac{x - x_1}{x_2 - x_1}.$$

In this case, the basic functions $b_1(x)$, $b_2(x)$ are accepted for interpolation.

The corresponding matrix B has the form

$$B = \begin{bmatrix} b_1(x_1) & b_2(x_1) \\ b_1(x_2) & b_2(x_2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

that is, the basic functions $b_1(x)$, $b_2(x)$ satisfy the conditions

$$b_j(x_i) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \quad (7)$$

Therefore, matrix B is singular. In this case, the solution of the equation system $B\bar{\alpha} = \bar{y}$ will be $\bar{\alpha} = \bar{y}$ either $\alpha_1 = y_1$, $\alpha_2 = y_2$.

In the case of quadratic interpolation of the form (4) by means of equivalent transformations, we have

$$P_2(x) = y_1 \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + y_2 \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + y_3 \frac{(x-x_3)(x-x_1)}{(x_3-x_1)(x_3-x_2)}, \quad (8)$$

or

$$P_2(x) = \alpha_1 b_1(x) + \alpha_2 b_2(x) + \alpha_3 b_3(x),$$

where

$$b_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}, \quad b_2(x) = \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)},$$

$$b_3(x) = \frac{(x-x_3)(x-x_1)}{(x_3-x_1)(x_3-x_2)}.$$

In this case, during interpolation for the basis functions $b_1(x)$, $b_2(x)$, $b_3(x)$ is taken which satisfy condition (7). The corresponding matrix B is a unit $B = E$. In this case, the solution of the equation system $B\bar{\alpha} = \bar{y}$ is $\bar{\alpha} = \bar{y}$ or $\alpha_1 = y_1$, $\alpha_2 = y_2$, $\alpha_3 = y_3$.

Let us generalize the results obtained when considering the linear and quadratic interpolations of the form (6), (8).

Suppose that we have a set of functions $l_1(x), l_2(x), \dots, l_n(x)$, each of which is a polynomial of degree $n - 1$ and satisfies the condition

$$l_j(x_i) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

That is, functions $l_j(x)$ take values of 1 at points $x = x_j$ and are zero for all other values $x = x_i$ ($i \neq j$). It should be noted that functions $b_j(x)$ with similar properties were used to construct the interpolants of the form (6), (8).

Any linear combination of functions $l_j(x)$ is a polynomial of degree no more $(n - 1)$. In particular, we consider a polynomial

$$P_{n-1}(x) = y_1 l_1(x) + y_2 l_2(x) + \dots + y_n l_n(x). \quad (9)$$

According to the properties of the functions $l_j(x)$ it follows that

$$P_{n-1}(x_i) = y_1 l_1(x_i) + y_2 l_2(x_i) + \dots + y_n l_n(x_i) = y_i l_i(x_i) = y_i,$$

and $P_{n-1}(x)$ is an interpolation polynomial. Functions are defined by formulas

$$l_j(x) = \frac{(x - x_1)(x - x_2) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_n)}{(x_j - x_1)(x_j - x_2) \dots (x_j - x_{j-1})(x_j - x_{j+1}) \dots (x_j - x_n)} = \frac{\prod_{i=1, i \neq j}^n (x - x_i)}{\prod_{i=1, i \neq j}^n (x_j - x_i)}. \quad (10)$$

The polynomial of the form (9), in which the coefficients $l_j(x)$ are determined according to formula (10), is called the **Lagrange interpolation polynomial**.

When finding a Lagrange polynomial, in some cases it is convenient to use the **Aitkin interpolation scheme**, a feature of which is the uniformity of calculations.

If the function f is given at two points x_0 and x_1 (its values are respectively equal y_0 to and y_1), then its value at the point $x \in (x_0; x_1)$ can be calculated by the linear interpolation formula (6). If the value of the function f at x is denoted by $P_{0,1}(x)$, then the linear interpolation formula (.6) can be written in equilateral form

$$P_{0,1}(x) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x \\ y_1 & x_1 - x \end{vmatrix},$$

the right part of which contains the second-order determinant. It's easy to make sure

$$P_{0,1}(x_0) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x_0 \\ y_1 & x_1 - x_0 \end{vmatrix} = y_0, \quad P_{0,1}(x_1) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x_1 \\ y_1 & x_1 - x_1 \end{vmatrix} = y_1.$$

Now let the function f be given at three points x_0, x_1, x_2 (corresponding values y_0, y_1, y_2), and it is necessary to calculate its value at a point $x \in (x_0; x_2), x \neq x_1$. In this case, the values of two linear polynomials are first calculated using the Eitkin scheme at x .

$$P_{0,1}(x) = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x \\ y_1 & x_1 - x \end{vmatrix} \quad \text{and} \quad P_{1,2}(x) = \frac{1}{x_2 - x_1} \begin{vmatrix} y_1 & x_1 - x \\ y_2 & x_2 - x \end{vmatrix},$$

and then the quadratic three-term view

$$P_{0,1,2}(x) = \frac{1}{x_2 - x_0} \begin{vmatrix} P_{0,1}(x) & x_0 - x \\ P_{1,2}(x) & x_2 - x \end{vmatrix}. \quad (11)$$

Direct verification convinces you that $P_{0,1}(x_0) = y_0$, $P_{0,1}(x_1) = y_1$, $P_{1,2}(x_1) = y_1$, $P_{1,2}(x_2) = y_2$, $P_{0,1,2}(x_1) = y_1$, $P_{0,1,2}(x_2) = y_2$.

We prove that $P_{0,1,2}(x)$ coincides with the second-order Lagrange interpolation polynomial. Indeed, since

$$P_{0,1}(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1, \quad P_{1,2}(x) = \frac{x - x_2}{x_1 - x_2} y_1 + \frac{x - x_1}{x_2 - x_1} y_2,$$

revealing the determinant of the second order of formula (11), we obtain

$$\begin{aligned} P_{0,1,2}(x) &= \frac{1}{x_2 - x_0} \left[\left(\frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 \right) (x_2 - x) - \right. \\ &\quad \left. - \left(\frac{x - x_2}{x_1 - x_2} y_1 + \frac{x - x_1}{x_2 - x_1} y_2 \right) (x_0 - x) \right] = \\ &= \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} y_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} y_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} y_2. \end{aligned}$$

This scheme generalizes to higher-degree interpolation polynomials. If the function f is set at four nodes, then cubic interpolation is performed by the formula

$$P_{0,1,2,3}(x) = \frac{1}{x_3 - x_0} \begin{vmatrix} P_{0,1,2}(x) & x_0 - x \\ P_{1,2,3}(x) & x_3 - x \end{vmatrix},$$

where $P_{0,1,2}(x)$ и $P_{1,2,3}(x)$ – the value of the quadratic three terms in a point $x \in (x_0; x_3)$, $x \neq x_1$, $x \neq x_2$. In this case, the values of the polynomials $P_{0,1,2}(x)$ are calculated by the formula (11) and the values $P_{1,2,3}(x)$ by the formula:

$$P_{1,2,3}(x) = \frac{1}{x_3 - x_1} \begin{vmatrix} P_{1,2}(x) & x_1 - x \\ P_{2,3}(x) & x_3 - x \end{vmatrix}, \quad P_{2,3}(x) = \frac{1}{x_3 - x_2} \begin{vmatrix} y_2 & x_2 - x \\ y_3 & x_3 - x \end{vmatrix}.$$

By direct verification we make sure that $P_{1,2,3}(x_1) = y_1$, $P_{1,2,3}(x_2) = y_2$, $P_{1,2,3}(x_3) = y_3$, $P_{0,1,2,3}(x_i) = y_i$, ($i = 0, 1, 2, 3$) and $P_{0,1,2,3}(x)$, and coincide with the Lagrangian cubic interpolation polynomial.

Generally, if at $(n + 1)$ interpolation nodes the function f acquires values y_i ($i = 0, 1, \dots, n$), then the value of the interpolation polynomial of degree n at a point $x \in (x_0; x_n)$ that does not coincide with the interpolation nodes can be calculated by the formula

$$P_{0,1,\dots,n}(x) = \frac{1}{x_n - x_0} \begin{vmatrix} P_{0,1,\dots,n-1}(x) & x_0 - x \\ P_{1,2,\dots,n}(x) & x_n - x \end{vmatrix},$$

where $P_{0,1,\dots,n-1}(x)$ and $P_{1,2,\dots,n}(x)$ are the values of the interpolation polynomials of $(n - 1)$ degree calculated at the point x in the previous step of the calculations. It is easy to be sure that $P_{0,1,\dots,n}(x_i) = y_i$, ($i = 0, 1, \dots, n$) and $P_{0,1,\dots,n}(x)$ coincides with the n -th degree Lagrange interpolation polynomial.

Therefore, to calculate at point x the value of the n -th degree interpolation polynomial according to the Aitkin scheme (Table 3.1), it is necessary to calculate at this point the values of n linear, $n - 1$ quadratic, $n - 2$ cubic polynomials, etc.,

two polynomials $(n - 1)$ –th degree and, finally, one n -th degree polynomial. All these polynomials are expressed in terms of the second-order determinant, and this makes the computations homogeneous, cyclic.

Table 4.1.

Interpolation Aitkin scheme

x_i	y_i	$P_{i-1,i}$	$P_{i-2,i-1,i}$	$P_{i-3,i-2,i-1,i}$	$x_i - x$
x_0	y_0	–	–	–	$x_0 - x$
x_1	y_1	$P_{0,1}(x)$	–	–	$x_1 - x$
x_2	y_2	$P_{1,2}(x)$	$P_{0,1,2}(x)$	–	$x_2 - x$
x_3	y_3	$P_{2,3}(x)$	$P_{1,2,3}(x)$	$P_{1,2,3,0}(x)$	$x_3 - x$
...

In calculations according to this scheme, new nodes x_i (corresponding to the transition to higher-degree interpolation polynomials) involve until the calculations themselves show that the required accuracy has already been achieved.

Example: The function $y = \sqrt[3]{x}$ is set as follows:

x	1,0	1,1	1,3	1,5	1,6
y	1,000	1,032	1,091	1,145	1,170

Apply the Aitkin scheme to find the value $\sqrt[3]{1,15}$. We write the output of the values of the function in table. 3.2 and calculate the difference at. Then we find consistently

$$P_{0,1} = \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x \\ y_1 & x_1 - x \end{vmatrix} = \frac{1}{0,1} \begin{vmatrix} 1 & -0,15 \\ 1,032 & -0,05 \end{vmatrix} = 1,048,$$

$$P_{1,2} = \frac{1}{x_2 - x_1} \begin{vmatrix} y_1 & x_1 - x \\ y_2 & x_2 - x \end{vmatrix} = \frac{1}{0,2} \begin{vmatrix} 1,032 & -0,05 \\ 1,091 & 0,15 \end{vmatrix} = 1,047,$$

$$P_{2,3} = \frac{1}{x_3 - x_2} \begin{vmatrix} y_2 & x_2 - x \\ y_3 & x_3 - x \end{vmatrix} = \frac{1}{0,2} \begin{vmatrix} 1,091 & 0,15 \\ 1,145 & 0,35 \end{vmatrix} = 1,050,$$

$$P_{3,4} = \frac{1}{x_4 - x_3} \begin{vmatrix} y_3 & x_3 - x \\ y_4 & x_4 - x \end{vmatrix} = \frac{1}{0,1} \begin{vmatrix} 1,145 & 0,35 \\ 1,170 & 0,45 \end{vmatrix} = 1,057.$$

Table 3.2

i	x	y	$x_i - x$	$P_{i-1, i}$	$P_{i-2, i-1, i}$
0	1,0	1,000	-0,15		
1	1,1	1,032	-0,05	1,048	
2	1,3	1,091	0,15	1,047	1,048
3	1,5	1,145	0,35	1,050	
4	1,6	1,170	0,45	1,057	

The calculated values are entered in table. 2, after which we calculate

$$P_{0,1,2} = \frac{1}{0,3} \begin{vmatrix} 1,048 & -0,15 \\ 1,047 & 0,15 \end{vmatrix} = 1,048.$$

The values $P_{0,1}$ and $P_{0,1,2}$ coincide with the third character. This calculation can be completed up to $\varepsilon = 10^{-3}$ and written down $\sqrt[3]{1,15} = 1,048$.

§ 4.3. Error estimation of Lagrange interpolation formula.

If the function f on a segment $[a; b]$ is a polynomial of degree less than or equal to n , it follows from the unity of the interpolation polynomial that the interpolation polynomial $P_n(x)$ is also equal to f , that is $f(x) - P_n(x) \equiv 0, x \in [a; b]$.

If f on a segment $[a; b]$ that containing interpolation nodes $x_i (i = \overline{0, n})$ is not a polynomial of degree less than or equal to n , then the difference

$$R_n(f, x) = f(x) - P_n(x) \quad (12)$$

will be zero only at the interpolation nodes $x_i (i = \overline{0, n})$, and at other points of the segment is different from the identity zero. A function $R_n(f, x)$ that characterizes the accuracy of the approximation of a function f to an interpolation polynomial $P_n(x)$ is called a **residual term** of the Lagrangian interpolation formula or an **interpolation error**. If an analytic expression of the function f is known, then it can be estimated. This holds true for this theorem.

Theorem. If the interpolation nodes $x_i (i = \overline{0, n})$ are different and belong to the segment $[a; b]$, then for any point $x \in [a; b]$ there is such a point $\xi \in (a; b)$ that for the error of interpolation equals

$$R_n(f, x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{j=0}^n (x - x_j),$$

where

$$\prod_{j=0}^n (x - x_j) = (x - x_0)(x - x_1) \dots (x - x_n).$$

If $M_{n+1} = \max_{x \in [a; b]} |f^{(n+1)}(x)|$, than for the absolute error of the Lagrange interpolation formula, we obtain the following estimate:

$$|R_n(f, x)| = |f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \left| \prod_{j=0}^n (x - x_j) \right|. \quad (13)$$

From formula (12) it is obvious that the residual term of the linear interpolation formula (6) is equal to

$$R_1(f, x) = \frac{\prod_{j=0}^1 (x - x_j)}{2!} f''(\xi),$$

where

$$\prod_{j=0}^1 (x - x_j) = (x - x_0)(x - x_1), \quad \xi \in (x_0; x_1);$$

and the residual term of the quadratic interpolation formula (8)

$$R_2(f, x) = \frac{\prod_{j=0}^2 (x - x_j)}{3!} f'''(\xi),$$

where

$$\prod_{j=0}^2 (x - x_j) = (x - x_0)(x - x_1)(x - x_2), \quad \xi \in (x_0; x_2).$$

For example, construct a Lagrange interpolation polynomial for a function $f(x) = \sqrt[3]{x}$ with interpolation nodes $x_0 = 1, x_1 = 2, x_2 = 3$. Estimate the error of an interpolation polynomial at a point $x = 2,5$.

The output is written as follows.

x_i	1	2	3
-------	---	---	---

$y_i = \sqrt[3]{x_i}$	1	1,280	1,442
-----------------------	---	-------	-------

The Lagrange interpolation polynomial for the case $n = 2$ is of the form

$$P_2(x) = y_1 \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x - x_3)(x - x_1)}{(x_3 - x_1)(x_3 - x_2)},$$

or taking into account the output

$$P_2(x) = \frac{(x - 2)(x - 3)}{2} \cdot 1 + \frac{(x - 1)(x - 3)}{-1} \cdot 1,260 + \frac{(x - 1)(x - 2)}{2} \cdot 1,442 = -0,039x^2 + 0,377x + 0,662.$$

To estimate the error, we use formula (13), which in this case has the form

$$|R_2(x)| \leq M_3 \frac{|(x - x_0)(x - x_1)(x - x_2)|}{3!},$$

where $M_3 = \max_{[1; 3]} |f'''(x)|$.

According to the condition $f(x) = \sqrt[3]{x}$, then

$$f'(x) = \frac{1}{3}x^{-2/3}, \quad f''(x) = -\frac{2}{9}x^{-5/3}, \quad f'''(x) = \frac{10}{27}x^{-8/3} = \frac{10}{27x^2 \sqrt[3]{x}}.$$

The function $f'''(x)$ is positive and descending on the interval $[1; 3]$.

Therefore,

$$|f'''(x)| \leq f'''(1) = \frac{10}{27} \approx 0,37$$

and

$$|R_2(2,5)| \leq 0,37 \frac{|(2,5-1)(2,5-2)(2,5-3)|}{3!}, \quad |R_2(2,5)| \leq 0,023.$$

It follows from formula (13) that the absolute error of the Lagrange interpolation formula is proportional to the product of two factors M_{n+1} and

$\left| \prod_{j=0}^n (x - x_j) \right|$, of which the value of the first depends only on the function f , and

the value of the second is determined solely by the choice of nodes of interpolation. The magnitude of the absolute error of the Lagrangian interpolation formula can be reduced by the choice of nodes of interpolation, at which the factor

$\left| \prod_{j=0}^n (x - x_j) \right|$ acquires the smallest maximum value per segment $[a; b]$.

TITLE 5. EXPERIMENTAL DATA PROCESSING METHODS

§ 5.1. The least squares method

It is known that if given values of a function f at $n + 1$ different points x_0, x_1, \dots, x_n , then there exists a unique polynomial of degree n such that

$$p(x_i) = f(x_i), \quad i = 0, 1, \dots, n.$$

Suppose now that the function f itself is a polynomial of n degree, and the problem is to determine its coefficients. Then, according to the above result, if the values of the function f were given accurately, it would be sufficient to know these values at different points.

But in many cases the values f are determined by the measurements and may contain errors. However, they usually perform much more than $n + 1$ measurements, hoping that as a result of "averaging" these errors will disappear. How these errors are "averaged" depends on the measurement processing method used to determine the polynomial coefficients. For statistical reasons, the least-squares method is often chosen for this method. This method is also extremely simple.

Now suppose that there are given m points x_1, x_2, \dots, x_m where $m \geq n + 1$, and at least of $n + 1$ these points are different. Let f_1, f_2, \dots, f_m be approximate values of a function f in points x_1, x_2, \dots, x_m . It is necessary to find such a polynomial $p(x) = a_0 + a_1x + \dots + a_nx^n$ that on it the value

$$\sum_{i=1}^m [f_i - p(x_i)]^2 \tag{1}$$

has reached a minimum among all polynomials of degree n , that is, it is necessary to find such coefficients a_0, a_1, \dots, a_n that the sum of squares of errors $f_i - p(x_i)$ is minimal.

The simplest case of such a problem occurs when $n = 0$ and the polynomial $p(x)$ is simply constant. Suppose, for example, that we have m measurements and w_1, w_2, \dots, w_m are the weight of an object, and this data is obtained on m different scales. All points x_1, x_2, \dots, x_m are identical and are not clearly included in the expression. Using the least squares principle, we come to the problem of minimizing a function

$$g(w) = \sum_{i=1}^m (w_i - w)^2.$$

From mathematical analysis it is known that the function g reaches a minimum (local) at the point \tilde{w} at which $g'(\tilde{w}) = 0$ and $g''(\tilde{w}) \geq 0$. Since

$$g'(w) = -2 \sum_{i=1}^m (w_i - w), \quad g''(w) = 2m,$$

it follows

$$\tilde{w} = \frac{1}{m} \sum_{i=1}^m w_i.$$

Since the equation $g'(w) = 0$ has only one solution, the point \tilde{w} will be the only point of minimum of g . Thus, the least squares approximation for the value of w will be simply the arithmetic mean of the measurements w_1, w_2, \dots, w_m .

The next simple situation is when a linear polynomial $p(x) = a_0 + a_1x$ is used. Such situations often occur in practice when it is assumed that the data is subject to some linear dependence. In this case, function (2) takes the form

$$g(a_0, a_1) = \sum_{i=1}^m (f_i - a_0 - a_1x_i)^2, \quad (2)$$

this requires a minimum of coefficients a_0 and a_1 . It is known from mathematical analysis that the necessary condition for the minimum of the function g is the fulfillment of relations

$$\frac{\partial g}{\partial a_0} = -2 \sum_{i=1}^m (f_i - a_0 - a_1x_i) = 0,$$

$$\frac{\partial g}{\partial a_1} = -2 \sum_{i=1}^m x_i (f_i - a_0 - a_1x_i) = 0.$$

Grouping together the coefficients for a_0 and a_1 , we arrive at a system of two linear equations

$$\begin{cases} ma_0 + \left(\sum_{i=1}^m x_i \right) a_1 = \sum_{i=1}^m f_i, \\ \left(\sum_{i=1}^m x_i \right) a_0 + \left(\sum_{i=1}^m x_i^2 \right) a_1 = \sum_{i=1}^m x_i f_i, \end{cases}$$

relatively unknown a_0 and a_1 whose solution is found by explicit formulas:

$$a_0 = \frac{\sum x_i^2 \sum f_i - \sum x_i f_i \sum x_i}{m \sum x_i^2 - (\sum x_i)^2},$$

$$a_1 = \frac{m \sum x_i f_i - \sum f_i \sum x_i}{m \sum x_i^2 - (\sum x_i)^2},$$

$i = \overline{1, m}$.

In the general case, when considering polynomials of degree n , we arrive at the problem of minimizing the function (1), that is, looking for the minimum of the function

$$g(a_0, a_1, \dots, a_n) = \sum_{i=1}^m (a_0 + a_1 x_i + \dots + a_n x_i^n - f_i)^2. \quad (3)$$

As in the case $n = 2$, the necessary condition for the minimum of the function g is zero for all its partial derivatives of the first order:

$$\frac{\partial g}{\partial a_j}(a_0, a_1, \dots, a_n) = 0, \quad j = \overline{0, n}.$$

Writing out these partial derivatives explicitly, we get the ratio

$$\sum_{i=1}^m x_i^j (a_0 + a_1 x_i + \dots + a_n x_i^n - f_i)^2 = 0, \quad j = \overline{0, n}, \quad (4)$$

which are a system of $n + 1$ linear equations relatively $n + 1$ unknown a_0, a_1, \dots, a_n . These equations are called *normal equations*. Collecting the coefficients for a_i and rewriting equation (4) in vector-matrix form, we arrive at the system

$$\begin{bmatrix} m & \sum x_i & \sum x_i^2 & \dots & \sum x_i^n \\ \sum x_i & \sum x_i^2 & \dots & \dots & \dots \\ \sum x_i^2 & \sum x_i^3 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_i^n & \dots & \dots & \dots & \sum x_i^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum f_i \\ \sum x_i f_i \\ \dots \\ \sum x_i^n f_i \end{bmatrix}, \quad (5)$$

where all sums are taken from 1 to m .

Note that system (4) can be rewritten as

$$E^T E \bar{a} = E^T \bar{f}, \quad (6)$$

where

$$E = \begin{bmatrix} 1 & x_1 & \dots & x_1^n \\ 1 & x_2 & \dots & x_2^n \\ \dots & \dots & \dots & \dots \\ 1 & x_m & \dots & x_m^n \end{bmatrix}, \quad a = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{bmatrix}, \quad f = \begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_m \end{bmatrix}. \quad (7)$$

Matrix E is a Vandermonde type matrix with size $m \times (n + 1)$. Assuming that at least $n + 1$ the points are different, we can show that the rank of the matrix E is equal $n + 1$. Therefore, the matrix is positively defined. It follows that the solution of system (5) gives a single point of minimum of the function g of (4), so that the least-squares approximation problem has a uniform solution.

Consider an example. Suppose

$$\begin{aligned} x_1 &= 0, & x_2 &= 1/4, & x_3 &= 1/2, & x_4 &= 3/4, & x_5 &= 1, \\ y_1 &= 1, & y_2 &= 2, & y_3 &= 1, & y_4 &= 0, & y_5 &= 1. \end{aligned}$$

Five points x_i are considered in this case, so these data uniquely determine the fourth-degree interpolation polynomial. We obtain the linear and quadratic polynomials of the least squares method on the basis of normal equations.

The following quantities are required to calculate a linear polynomial of the least-squares method based on normal equations:

$$\sum_{i=1}^5 x_i = \frac{5}{2}, \quad \sum_{i=1}^5 x_i^2 = \frac{15}{8}, \quad \sum_{i=1}^5 f_i = 5, \quad \sum_{i=1}^5 x_i f_i = 2. \quad (8)$$

Then, by the formulas above for the coefficients a_0 and a_1 , we obtain

$$a_0 = \frac{\left(\frac{15}{8}\right)(5) - (2)\left(\frac{5}{2}\right)}{(5)\left(\frac{15}{8}\right) - \left(\frac{25}{4}\right)} = \frac{7}{5},$$

$$a_1 = \frac{(5)(2) - (5)\left(\frac{5}{2}\right)}{(5)\left(\frac{15}{8}\right) - \left(\frac{25}{4}\right)} = \frac{-4}{5}.$$

Therefore, the linear polynomial of the least squares method constructed from this data looks like this

$$p_1(x) = 7/5 - (4/5)x. \quad (9)$$

To determine the coefficients of a quadratic polynomial using the least squares method from normal equations, it is necessary to solve system (5) with:

$$\begin{bmatrix} m & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_n \end{bmatrix} = \begin{bmatrix} \sum f_i \\ \sum x_i f_i \\ \sum x_i^2 f_i \end{bmatrix}.$$

With given data, this system looks like

$$\begin{bmatrix} 640 & 320 & 240 \\ 320 & 240 & 200 \\ 240 & 200 & 177 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 640 \\ 256 \\ 17 \end{bmatrix}.$$

Finding out of this system

$$a_0 = 7/5, \quad a_1 = -4/5, \quad a_2 = 0. \quad (10)$$

TITLE 6. APPROXIMATE SOLUTION OF COMMON DIFFERENTIAL EQUATIONS

When studying various processes and phenomena containing elements of motion, they often use mathematical models in the form of equations, which, in addition to the independent quantities and dependent functions, also include derivatives of the sought functions. Such equations are called differential equations.

A differential equation is called **ordinary** if an unknown function is a function of one variable, and a differential **equation in partial derivatives** if an unknown function is a function of many variables. In the following we will consider only ordinary differential equations.

§ 6.1. Problem statement

The first order differential equation is called the equation

$$F(x, y, y') = 0, \quad (1)$$

which associates the independent variable x , the unknown function $y = y(x)$ and their derivative y' .

The differential equation (1), which is not solved relative to the derivative y' , is called the implicit differential equation. If equation (1) can be solved relatively y' , then it is written as

$$y' = f(x, y) \quad (2)$$

and is called the first-order equation, solved relative to the derivative, or the equation in the normal form. In the future, we will mainly consider such equations.

Among the applied problems are those tasks for which **the Cauchy problem** is solved, which is to find a partial solution $y = y(x)$ of the ordinary differential equation (2) that satisfies the given initial condition

$$y|_{x=x_0} = y_0 \quad \text{or} \quad y(x_0) = y_0. \quad (3)$$

The following theorem holds.

Theorem (on the existence and uniqueness of a solution). Let the function $f(x, y)$ and its partial derivative $f'_y(x, y)$ be defined and continuous in the open region D of the plane xOy and point $(x_0, y_0) \in D$. Then there is a unique solution $y = \varphi(x)$ of equation (2) which satisfies the condition

$$y = y_0 \quad \text{at} \quad x = x_0, \quad \text{that is} \quad \varphi(x_0) = y_0. \quad (4)$$

This theorem gives sufficient conditions for the existence of a single solution of equation (2).

From the point of view of geometry, to solve the Cauchy problem (2) - (3) means to select the one that passes through the point (x_0, y_0) from the set of integral curves.

In what follows, we assume that the function $f(x, y)$ of equation (2) satisfies the condition of the above theorem.

It is rare to integrate the differential equation in the finite form. In this case, for the most part, the expression to which the desired function is implicit is obtained, and therefore inconvenient to use.

In practice, the approximate integration of differential equations is mainly used. It allows you to find an approximate solution to the Cauchy problem. Solve the Cauchy problem (2) - (3) numerically - this means that for a given sequence of values x_0, x_1, \dots, x_n of the independent variable x and numbers y_0 find a numerical sequence y_0, y_1, \dots, y_n , that is, for a given sequence of values x_k ($k = \overline{0, n}$), construct a table of approximate values of the desired solution of the Cauchy problem y_k ($k = \overline{0, n}$).

If the approximate solution of problem (2) - (3) is known at the point x_k , then integrating equation (.2) from x_k to x_{k+1} , we find its solution at the point x_{k+1} by the formula

$$y(x_{k+1}) = y(x_k) + \int_{x_k}^{x_{k+1}} f(x, y(x)) dx. \quad (5)$$

Formula (5) is the starting point for constructing many numerical methods for solving problem (2), (3).

There are two main approaches to solving the Cauchy problem.

1. **One-step methods** in which to find the values at the next point on the curve, only information about the value of the function in the previous step is required. The Euler method and its modifications and the Runge - Kutta methods are one-step methods.
2. **Multistep methods** (forecasting and correction methods) in which information is required to find values at the next point of the curve $y = f(x)$ than at one of the previous points of the solution. In many cases, iterations are used to obtain a sufficiently accurate numerical solution. Such methods include Milne, Adams - Bashfort and Hamming.

§ 6.2. Euler's method and its modifications

The Euler method is the simplest method of solving a Cauchy problem, which allows to integrate first order differential equations. This method is practiced when precision is not required and the number of steps is small. In general, this method is illustrative. It helps you understand the essence of building more sophisticated and effective methods.

The Euler method is based on the decomposition of a function $y(x)$ into a Taylor series around the point x_0 :

$$y(x_0 + h) = y(x_0) + hy'(x_0) + \frac{1}{2}h^2 y''(x_0) + \dots \quad (6)$$

If the integration step h is small, then additives containing h in the second and higher degrees are small orders of magnitude higher and can be neglected. Then $y(x_0 + h) = y(x_0) + hy'(x_0)$, where $y'(x_0)$ is determined from differential equation (2) under the initial condition (3). Therefore, it is possible to obtain an approximate value of the dependent variable y for small values of h according to the formula

$$y_{n+1} = y_n + hf(x_n, y_n), \quad (n = 0, 1, 2, \dots). \quad (7)$$

The geometric interpretation of the Euler method is as follows. At every n -th step, starting from a point (x_0, y_0) , the solution does not search on an integral curve $y(x)$ but on a segment tangent to that curve passing through the point (x_n, y_n) . The equation of such tangent $y = y_n + y'(x_n)(x - x_n)$. Given that $y'(x_n) = f(x_n, y_n)$ and $x_{n+1} = x_n + h$, then the formula takes the form (7).

The Euler method has the first order of accuracy. With the distance from the starting point (x_0, y_0) , the errors in the determination are accumulated quickly enough. There are many ways to improve this method. Two methods are

considered from these methods: the Euler – Cauchy method and the modified Euler method, which have second order accuracy.

The Euler – Cauchy method. If we calculate the integral in the right-hand side of formula (5) by the formula of the mean rectangles, that is, calculate the value of the integrand function $f(x, y(x))$ at a point $x_{k+1/2} = x_k + \frac{1}{2}h$, then we find

$$y(x_{k+1}) = y(x_k) + hf(x_{k+1/2}, y(x_{k+1/2})) + O(h^2). \quad (8)$$

The value of the unknown value of the function $y(x_{k+1/2})$ is calculated by the formula

$$y(x_{k+1/2}) = y(x_k) + hf(x_k, y(x_k)) + O(h^2)$$

With the step $\frac{1}{2}h$ we will

$$y(x_{k+1/2}) = y(x_k) + \frac{1}{2}hf(x_k, y(x_k)) + O(h^2).$$

Substituting this value $y(x_{k+1/2})$ into (8), we obtain

$$\begin{aligned} y(x_{k+1}) &= y(x_k) + hf(x_{k+1/2}, y(x_k) + \frac{1}{2}hf(x_k, y(x_k)) + O(h^2)) + \\ &+ O(h^2) = y(x_k) + hf\left(x_{k+1/2}, y(x_k) + \frac{h}{2}f(x_{k+1/2}, y(x_k))\right) + O(h^3). \end{aligned}$$

Having rejected the proportional to h^3 addition, we have

$$y(x_{k+1}) = y(x_k) + hf\left(x_{k+1/2}, y(x_k) + \frac{1}{2}hf(x_k, y(x_k))\right).$$

The formulas for Euler's advanced method can be written in the form

$$y_{k+1/2} = y_k + \frac{1}{2}hf(x_k, y_k), \quad (9)$$

$$y_{k+1} = y_k + hf(x_{k+1/2}, y_{k+1/2}). \quad (10)$$

Therefore, in the Euler advanced method, the first by the Euler method (formulas 9) calculated the approximate solution $y_{k+1/2}$ of problem 2 – 3 at the point $x_{k+1/2} = x_k + \frac{1}{2}h$ and then by the formula (10) the approximate solution y_{k+1} at the point x_{k+1} . At each integration step, calculate the right-hand side of equation (2) twice (at points (x_k, y_k) and $(x_{k+1/2}, y_{k+1/2})$).

Geometrically, this implies that in the segment $[x_k; x_{k+1}]$ the graph of the integral curve of problem (2) - (3) is replaced by the segment of the line passing through the point (x_k, y_k) and having an angular coefficient $k = f(x_{k+1/2}, y_{k+1/2})$. In other words, this line forms with the positive direction the axis Ox angle $\varphi = \text{arctg } f(x_{k+1/2}, y_{k+1/2})$.

The point $(x_{k+1/2}, y_{k+1/2})$ is the point of intersection of the problem tangent to the integral curve (6.2) - (6.3) at a point (x_k, y_k) with a line $x = x_k + \frac{1}{2}h$. The error of Euler's advanced method at every step is in order.

Modified Euler-Cauchy method. If the integral in the right part of formula (6.5) is calculated by the trapezoidal formula, then we have

$$y(x_{k+1}) = y(x_k) + \frac{h}{2} [f(x_k, y(x_k)) + f(x_{k+1}, y(x_{k+1}))] + O(h^3). \quad (6.11)$$

The unknown value $y(x_{k+1})$ included in the right-hand side of this equation can be calculated by the formula

$$y_{k+1} = y_k + hf(x_k, y_k), \quad (k = 0, 1, 2, \dots, n-1), \quad h = x_{k+1} - x_k.$$

Substituting the value $y(x_{k+1})$ in the right part (6.11), we obtain

$$\begin{aligned} y(x_{k+1}) &= y(x_k) + \frac{h}{2} [f(x_k, y(x_k)) + \\ &+ f(x_{k+1}, y(x_{k+1})) + hf(x_k, y(x_k)) + O(h^2)] + O(h^3) = \\ &= y(x_k) + \frac{h}{2} [f(x_k, y(x_k)) + (f(x_{k+1}, y(x_{k+1})) + hf(x_k, y(x_k)))] + O(h^3). \end{aligned}$$

Thus, for the improved Euler – Cauchy method we have the following calculation formulas:

$$\tilde{y}_{k+1} = y_k + hf(x_k, y_k), \quad (6.12)$$

$$y_{k+1} = y_k + \frac{h}{2} (f(x_k, y_k) + f(x_{k+1}, \tilde{y}_{k+1})). \quad (6.13)$$

According to this method, at each step of integration, the right-hand side of equation (6.2) is calculated twice: first, using the Euler method (formula (6.12)), calculate the approximate value of the desired solution \tilde{y}_{k+1} at the point x_{k+1} , which

is then refined by formula (6.13). The error of the method at each step is in order $O(h^3)$.

This construction of the approximate solution of the problem (6.2) - (6.3) from the point of view of Homeric means that on the segment $[x_k; x_{k+1}]$ the graph of the integral curve is approached by the segment of the line passing through the point (x_k, y_k) and having an angular coefficient $k = \frac{1}{2}(f(x_k, y_k) + f(x_{k+1}, \tilde{y}_{k+1}))$.

That is, this line forms an angle with the positive direction of the axis Ox $\varphi = \arctg \frac{f(x_k, y_k) + f(x_{k+1}, \tilde{y}_{k+1})}{2}$.

Euler's method with iterations is improved. If in equation (6.11) we reject the term proportional h^3 , then to find the value of the unknown solution y_{k+1} at the point x_{k+1} we obtain the formula:

$$y_{k+1} = y_k + \frac{h}{2}(f(x_k, y_k) + f(x_{k+1}, y_{k+1})). \quad (6.14)$$

Since the unknown y_{k+1} is included in both parts of the equation (6.14), the method defined by formula (6.14) belongs to *the implicit methods* of numerical integration of the problem (6.2) - (6.3). solution of equation (6.14), its solution can always be calculated with a predetermined accuracy $\varepsilon > 0$, if we use the method of iterations.

$$y_{k+1}^{(i+1)} = y_k + \frac{h}{2}(f(x_k, y_k) + f(x_{k+1}, y_{k+1}^{(i)})) \quad (i = 0, 1, 2, \dots). \quad (6.15)$$

For the zero approximation $y_{k+1}^{(0)}$, we can take the value y_k or the value y_{k+1} calculated by Euler's formula (6.7). The process of iterations according to the formula (6.15) is stopped with the fulfillment of the condition $|y_{k+1}^{(i+1)} - y_{k+1}^{(i)}| < \varepsilon$,

ie when the modulus of the difference of two consecutive approximations to the desired value is less than y_{k+1} the predetermined accuracy $\varepsilon > 0$. For the approximate value of the value y_{k+1} at the point x_{k+1} take the value $y_{k+1}^{(i+1)}$.

It is easy to establish the conditions under which the iterative process given by formula (6.15) coincides. To do this, we subtract equality (6.15) from equality (6.14). We will receive

$$y_{k+1} - y_{k+1}^{(i+1)} = \frac{h}{2} (f(x_{k+1}, y_{k+1}) - f(x_{k+1}, y_{k+1}^{(i)})).$$

Hence, using the condition Lipschitsa

$$|f(x_1, y_1) - f(x_2, y_2)| \leq \kappa |y_1 - y_2|, \quad (\kappa = \text{const}),$$

find

$$|y_{k+1} - y_{k+1}^{(i+1)}| \leq \frac{h}{2} |f(x_{k+1}, y_{k+1}) - f(x_{k+1}, y_{k+1}^{(i)})| \leq \frac{h}{2} \kappa |y_{k+1} - y_{k+1}^{(i)}|,$$

or

$$|y_{k+1} - y_{k+1}^{(i+1)}| \leq \left(\frac{h}{2} \kappa\right)^{i+1} |y_{k+1} - y_{k+1}^{(0)}|.$$

Therefore, the iterative process (6.15) coincides, ie $y_{k+1}^{(i+1)} \rightarrow y_{k+1}$, when $i \rightarrow \infty$, and the integration step h is chosen so that the inequality holds $\frac{1}{2} h \kappa < 1$.

The rate of convergence is determined by the value $\frac{1}{2} h \kappa$.

For example, we apply Euler's method and its modifications to solve the equation

$$y' = y - \frac{2x}{y} \text{ with the initial condition } y(0) = 1 \text{ and the integration step } h = 0,2.$$

The analytical solution to this problem is a function $y = \sqrt{2x + 1}$.

Using the above algorithms for solving the Cauchy problem, we calculate the discrete values of the function (Table 6.1).

Table 6.1

х	Метод Ейлера	Метод Ейлера–Коші	Модифікований метод Ейлера	Точний розв'язок
0,2	1,2000	1,1867	1,1836	1,1832
0,4	1,3733	1,3483	1,3426	1,3416
0,6	1,5294	1,4937	1,4850	1,4832
0,8	1,6786	1,6279	1,6152	1,6124
1,0	1,8237	1,7543	1,7362	1,7320

For clarity, these results are shown in Fig. 6.1.

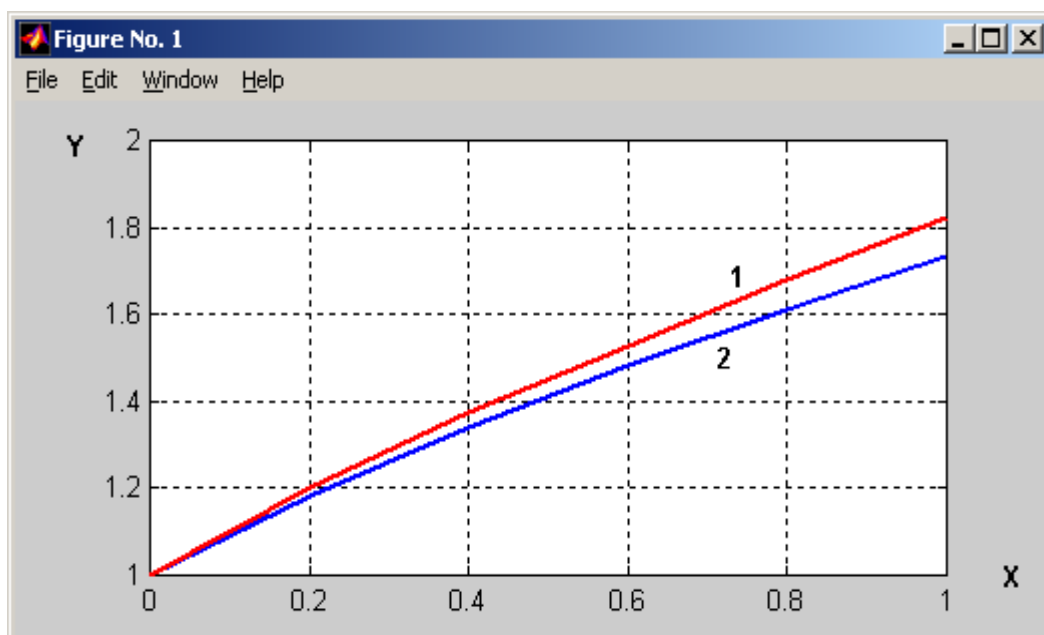


Fig. 6.1

Curve 1 corresponds to the solution according to the Euler method, curve 2 corresponds to the exact solution and the results according to the modified Euler method (it should be noted that the graph data according to the modified Euler method coincide with the data of the exact solution). As can be seen from the table and graph, the approximate numerical solutions are markedly different from the exact ones.

§ 6.3. Runge-Kutta method

The Runge-Kutta method has the widest application among one-step methods of increased accuracy of the approximate solution of the Cauchy problem for the differential equation

$$\frac{dy}{dx} = f(x, y), \quad (6.16)$$

with the initial condition

$$y|_{x=x_0} = y_0. \quad (6.17)$$

The idea of this method has much in common with the idea of Euler's methods and modifications and is to fit the Taylor series.

Let us determine the desired solution $y(x)$ of the Cauchy problem (6.16) - (6.17) in the vicinity of each point $x = x_n$ ($n = 0, 1, 2, \dots$) by the Taylor formula. Calculate the decomposition coefficients directly on the right-hand side of equation (6.16), using condition (6.17). The specified schedule will be written in the form

$$y(x) = y_0 + h \frac{dy}{dx} + \frac{h^2}{2!} \frac{d^2 y}{dx^2} + \frac{h^3}{3!} \frac{d^3 y}{dx^3} + \dots, \quad (6.18)$$

where the values of the derivatives are taken at $x = x_k$. Depending on how many members of the schedule we are satisfied with in formula (7.18), we get one or another accuracy of the approximate solution. In the Runge-Kutta method, we limit ourselves to four or five members of the schedule (members with degrees up to h^3 or h^4 including are retained).

Consider the Runge-Kutta method of the third order of accuracy:

$$y(x) \approx y_n + h \frac{dy}{dx} + \frac{h^2}{2!} \frac{d^2 y}{dx^2} + \frac{h^3}{3!} \frac{d^3 y}{dx^3}. \quad (6.19)$$

Suppose that

$$y_{n+1} = y_n + \lambda_n, \quad (6.20)$$

where

$$\lambda_n = h \frac{dy}{dx} + \frac{h^2}{2} \frac{d^2 y}{dx^2} + \frac{h^3}{6} \frac{d^3 y}{dx^3}. \quad (6.21)$$

Values λ_n are determined using linear combinations of the form

$$\lambda_n = \alpha k_1 + \beta k_2 + \gamma k_3 + \delta k_4, \quad (6.22)$$

where $\alpha, \beta, \gamma, \delta$ - indefinite coefficients, and k_1, k_2, k_3, k_4 - numbers determined by equations

$$k_1 = hf(x_n, y_n); \quad k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right);$$

$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right); k_4 = hf(x_n + h, y_n + k_2). \quad (6.23)$$

To determine the coefficients α , β , γ , δ , we express the derivatives included in equation (6.22) through the right-hand side of equation (6.16)

$$\frac{d^2 y}{dx^2} = \frac{\partial f}{\partial x} + y' \frac{\partial f}{\partial y} = \frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y}.$$

For simplicity of the statement further we will enter the operator

$$D = \frac{\partial}{\partial x} + f \frac{\partial}{\partial y}, \text{ then}$$

$$\frac{d^2 y}{dx^2} = D f;$$

$$\begin{aligned} \frac{d^3 y}{dx^3} &= D(D f) = \frac{\partial^2 f}{\partial x^2} + 2f \frac{\partial^2 f}{\partial x \partial y} + f^2 \frac{\partial^2 f}{\partial y^2} + \frac{\partial f}{\partial y} \left(\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right) = \\ &= D^2 f + \frac{\partial f}{\partial y} D f. \end{aligned}$$

Substituting the found values of derivatives in (6.21), we obtain

$$\lambda_n = hf + \frac{h^2}{2} D f + \frac{h^3}{6} \left(D^2 f + \frac{\partial f}{\partial y} D f \right). \quad (6.24)$$

Express k_2 , k_3 , k_4 as functions of two variables by the Taylor formula.

We have:

$$k_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)h = \left[f + \left(\frac{h}{2} \frac{\partial}{\partial x} + \frac{k_1}{2} \frac{\partial}{\partial y}\right) f + \frac{1}{2} \left(\frac{h}{2} \frac{\partial}{\partial x} + \frac{k_1}{2} \frac{\partial}{\partial y}\right)^2 f + \dots \right] h.$$

We limit ourselves to the third powers of h and, given that $k_1 = hf$, we obtain:

$$k_2 = \left[f + \frac{h}{2} \left(\frac{\partial}{\partial x} + f \frac{\partial}{\partial y}\right) f + \frac{h^3}{8} \left(\frac{\partial}{\partial x} + f \frac{\partial}{\partial y}\right)^2 f \right] = hf + \frac{h^2}{2} Df + \frac{h^3}{8} D^2 f.$$

Similarly for k_3 and k_4 you can write:

$$\begin{aligned} k_3 &= \left[f + \left(\frac{h}{2} \frac{\partial}{\partial x} + \frac{k_2}{2} \frac{\partial}{\partial y}\right) f + \frac{1}{2} \left(\frac{h}{2} \frac{\partial}{\partial x} + \frac{k_2}{2} \frac{\partial}{\partial y}\right)^2 f \right] h = \\ &= \left[f + \frac{h}{2} \left(\frac{\partial}{\partial x} + f \frac{\partial}{\partial y} + \frac{h}{2} Df \frac{\partial}{\partial y}\right) f + \frac{h^3}{8} \left(\frac{\partial}{\partial x} + f \frac{\partial}{\partial y} + \dots\right)^2 f \right] h = \\ &= hf + \frac{h^2}{2} Df + \frac{h^2}{4} \frac{\partial f}{\partial y} Df + \frac{h^3}{8} D^2 f; \end{aligned}$$

$$\begin{aligned} k_4 &= \left[f + \left(h \frac{\partial}{\partial x} + k_2 \frac{\partial}{\partial y}\right) f + \frac{1}{2} \left(h \frac{\partial}{\partial x} + k_2 \frac{\partial}{\partial y}\right)^2 f \right] h = \\ &= \left[f + h \left(\frac{\partial}{\partial x} + f \frac{\partial}{\partial y} + \frac{h}{2} Df \frac{\partial}{\partial y}\right) f + \frac{h^2}{2} \left(\frac{\partial}{\partial x} + f \frac{\partial}{\partial y} + \dots\right)^2 f \right] h = \\ &= hf + h^2 Df + \frac{h^3}{2} \frac{\partial f}{\partial y} Df + \frac{h^3}{2} D^2 f. \end{aligned}$$

Now find the amount $\lambda_n = \alpha k_1 + \beta k_2 + \gamma k_3 + \delta k_4$.

Comparing the coefficients at the same powers h in the last equation and expression (6.24) to determine $\alpha, \beta, \gamma, \delta$, taking into account the expressions for the coefficients k_1, k_2, k_3, k_4 , we obtain a system of equations

$$\alpha + \beta + \gamma + \delta = 1 \quad (\text{at } hf); \quad \frac{\beta}{2} + \frac{\gamma}{2} + \delta = \frac{1}{2} \quad (\text{at } h^2 Df);$$

$$\frac{\beta}{8} + \frac{\gamma}{8} + \frac{\delta}{2} = \frac{1}{6} \quad (\text{at } h^3 D^2 f); \quad \frac{\gamma}{4} + \frac{\delta}{2} = \frac{1}{6} \quad (\text{at } h^3 \frac{\partial f}{\partial y} Df).$$

This system has a solution $\alpha = \delta = \frac{1}{6}, \quad \beta = \gamma = \frac{1}{3}$.

So,

$$\lambda_n = \frac{1}{6}(k_1 + 2k_2 + 3k_3 + k_4). \quad (6.25)$$

To calculate y_{n+1} at a point x_{n+1} we have the formula

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 3k_3 + k_4). \quad (6.26)$$

In the case of the Runge - Kutta method of the fourth order of accuracy, ie when in development (6.18) members with degrees from h to h^4 inclusive are kept, the integration process is carried out similarly, only the numbers change k_1, k_2, k_3, k_4

$$k_1 = hf(x_n, y_n); \quad k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right);$$

$$k_3 = hf \left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2} \right); \quad k_4 = hf(x_n + h, y_n + k_2). \quad (6.27)$$

Similarly, you can build formulas of higher degrees.

For example, solve a linear differential equation $y' = 1,5e^{2x} - x^2 - x + 0,5$.

$$\frac{dy}{dx} = 2x^2 + 2y$$

with the initial condition $y(0) = 1$, $0 \leq x \leq 1$ and step $h = 0,1$. This equation has an exact solution

The results of the calculations are presented in Fig. 6.2. and in table. 6.2.

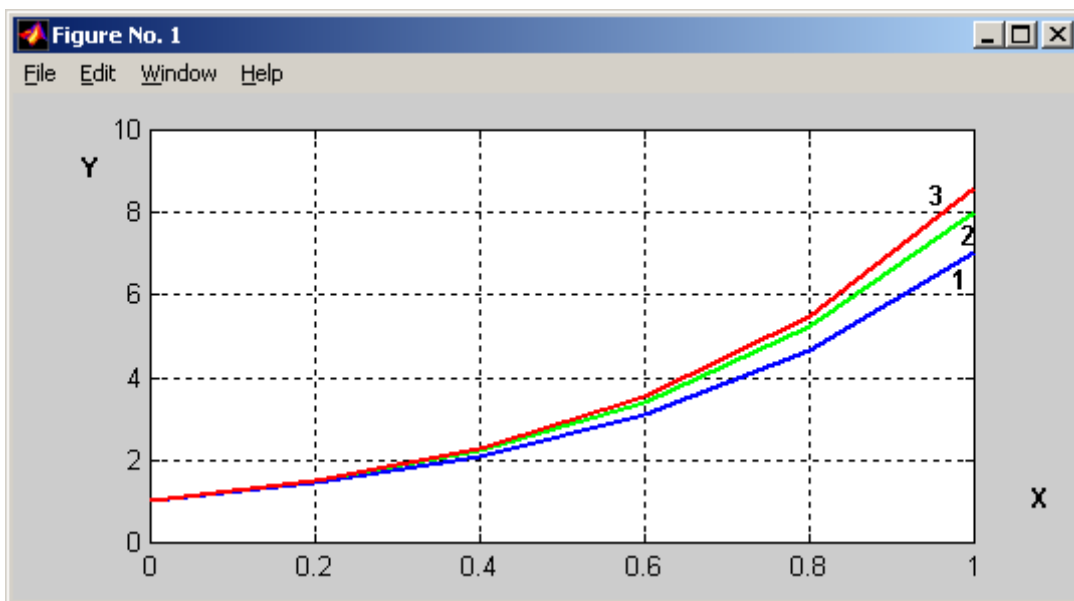


Рис. 6.2

In fig. 6.2 curve 1 corresponds to the numerical solution according to the Euler method, curve 2 - according to the modified Euler method, curve 3 corresponds

to the exact solution and the solution according to the Runge - Kutta method (in the figure the curves coincide).

These calculations allow us to compare the results of these methods.

Table 6.2

x	Euler's method	Modified Euler's method	Method Runge-Kutta	Exact solution
0,0	1,0000	1,0000	1,0000	1,0000
0,1	1,2000	1,2210	1,2221	1,2221
0,2	1,4420	1,4923	1,4977	1,4977
0,3	1,7384	1,8284	1,8432	1,8432
0,4	2,1041	2,2466	2,2783	2,2783
0,5	2,5569	2,7680	2,8274	2,8274
0,6	3,1183	3,4176	3,5201	3,5202
0,7	3,8139	4,2257	4,3927	4,3928
0,8	4,6747	5,2288	5,4894	5,4895
0,9	5,7376	6,4704	6,8643	6,8645
1,0	7,0472	8,0032	8,5834	8,5836

Tasks for self-fulfillment

I. For this Cauchy problem, find the approximate Runge-Kutta solution on the given segment with accuracy $\varepsilon = 10^{-3}$ and present the results in the form of graphs.

1. $y' = y^2 + x^2, \quad y(0) = 0, \quad [0; 1].$
2. $y' = 1 + x + x^2 - 2y^2, \quad y(1) = 1, \quad [1; 2].$

3. $y' = 1 + y\sqrt{x}$, $y(0) = 1$, $[0; 0,5]$.
4. $y' = y^3 + x^2$, $y(0,1) = 0,5$, $[0,1; 1,1]$.
5. $y' = \frac{y}{x} - y^2$, $y(1) = 1$, $[1; 2]$.
6. $y' = x + y^2 + x^2$, $y(0) = 0,8$, $[0; 1]$.
7. $y' = 2x + \cos y$, $y(0) = 0$, $[0; 0,1]$.
8. $y' = y^3 - x$, $y(0) = 0,5$, $[0; 1]$.
9. $(1-x)y' = 1 + x - y$, $y(0) = 0$, $[0; 1,5]$.
10. $y' = y + x^2$, $y(0) = 1$, $[0; 1]$.
11. $y' = 1 + x - y^2$, $y(0) = 1$, $[0; 1]$.
12. $y' = y - \sin x$, $y(0) = 0$, $[0; 0,5]$.
13. $y' = x^2y + e^{2x}$, $y(0) = 0$, $[0; 1]$.
14. $y' = xy^2$, $y(0) = 0$, $[0; 1]$.
15. $y' = x^2 + \sin^2 x + 2y$, $y(0) = 1$, $[0; 1]$.
16. $y' = xy^3 + x^2$, $y(0) = 0,5$, $[0; 1]$.
17. $y' = y^3 - xy + y \cdot \cos x$, $y(0,1) = 1$, $[0,1; 1,1]$.
18. $y' = x^2 + xy + y^2$, $y(0) = 0,5$, $[0; 1]$.
19. $y' = y^2x^2 - 1$, $y(0) = 1$, $[0; 1]$.
20. $y' = \frac{1}{x^2} - y - 2x$, $y(1) = 2$, $[1; 1,8]$.
21. $y' = xy^3 - 1$, $y(0) = 0,5$, $[0; 1]$.
22. $y' = \frac{y}{1+x} - y^2$, $y(0) = 1$, $[0; 1]$.
23. $y' = 0,1(x^2 + y^2)$, $y(1) = 1$, $[0; 5]$.
24. $y' = \frac{1}{x^2 + y^2}$, $y(0,5) = 0,5$, $[0,5; 3,5]$.

25. $y' = y^2 + \frac{2}{x}$, $y(1) = 1$, $[1; 1,5]$.
26. $y' = 1 + y - x^2$, $y(0) = 1$, $[0; 1]$.
27. $y' = x^3 + y^3$, $y(0,1) = 0,5$, $[0,1; 0,6]$.
28. $y' = y^2 e^x - 2y$, $y(0) = -1$, $[0; 1]$.
29. $y' = x^2 - y^2$, $y(0) = 0$, $[0; 1]$.
30. $y' = y^2 + 2x$, $y(1) = 1$, $[1; 1,5]$.

II. Find the approximate solution of the second-order differential equation under the given initial Runge-Kutta conditions on the segment $[0; 1]$ with accuracy $\varepsilon = 10^{-3}$ and present the results in the form of graphs.

1. $y'' + (1 + x^2)y = 0$, $y(0) = 0$, $y'(0) = 1$.
2. $y'' + xy = 0$, $y(0) = 0$, $y'(0) = 1$.
3. $y'' + xy = 0$, $y(0) = 1$, $y'(0) = 0$.
4. $y'' = xy' - y \cos x$, $y(0) = 1$, $y'(0) = 1$.
5. $y'' = \frac{1}{x}y' + y$, $y(1) = 1$, $y'(1) = 0$.
6. $y'' - xy' - y = 0$, $y(0) = 1$, $y'(0) = 0$.
7. $y'' + y \cos x = 0$, $y(0) = 0$, $y'(0) = 0$.
8. $y'' = x^2y - y'$, $y(0) = 1$, $y'(0) = 0$.
9. $y'' = yy' - x^2$, $y(0) = 1$, $y'(0) = 1$.
10. $y'' = 3y^2y' - 1$, $y(0) = 1$, $y'(0) = 0$.
11. $y'' = x^2y$, $y(0) = 1$, $y'(0) = 1$.
12. $y'' = xyy'$, $y(0) = 1$, $y'(0) = 1$.
13. $y'' = xy - y' + \sin x$, $y(0) = 1$, $y'(0) = 0$.

14. $y'' = xe^x + 2yy'$, $y(0) = 0$, $y'(0) = 1$.
15. $y'' = y' + x^2 - y^2$, $y(0) = 1$, $y'(0) = 1$.
16. $y'' = xy' - y + e^x$, $y(0) = 1$, $y'(0) = 0$.
17. $y'' = xy' + y \sin x = 0$, $y(0) = 0$, $y'(0) = 1$.
18. $y'' = x \cos x - y^2 + e^{2x}$, $y(0) = 1$, $y'(0) = 1$.
19. $y''(1 + x^2) + xy' - y = 0$, $y(0) = 1$, $y'(0) = 0$.
20. $y'' = x^2y' + 2y - 2e^{-x} = 0$, $y(0) = 1$, $y'(0) = 1$.
21. $y'' = 2xy - 3y' + y^3$, $y(0) = 0$, $y'(0) = 1$.
22. $y'' - (1 + x^2)y = 0$, $y(0) = 2$, $y'(0) = 2$.
23. $y'' = x^2 + yy' - e^x$, $y(0) = 1$, $y'(0) = 1$.
24. $y'' = x^2 - y^2 + 2y'$, $y(0) = 0$, $y'(0) = 1$.
25. $y'' = y' \sin x - y + 1$, $y(0) = 0$, $y'(0) = 1$.
26. $y'' = \cos x - y^2 + y' + 3x$, $y(0) = 1$, $y'(0) = 0$.
27. $y'' = yy' - \cos x - y$, $y(0) = 0$, $y'(0) = 1$.
28. $y'' = \frac{5}{1+x}y' + y$, $y(0) = 0$, $y'(0) = 1$.
29. $y'' - xy' - 2y = 0$, $y(0) = 1$, $y'(0) = 0$.
30. $y'' + 4y = \frac{1}{\cos 2x}$, $y(0) = 1$, $y'(0) = 1$.

REFERENCES

1. Bugrov JS, Nikolaevsky SM Elements of linear algebra and analytic geometry. M: Higher School, 1981.-368p.
2. Ilyin VA, Poznyak EG Analytical geometry. M.: Nauka, 1981.
3. Bugrov JS, Nikolsky SM Differential and integral calculus. M.: Nauka, 1980.
4. Bugrov JS, Nikolsky SM Differential equations. Multiple integrals. Rows. Functions of a complex variable. M.: Nauka, 1980.
5. Piskunov NS Differential and integral calculus for universities. Volume I.: Science, 1978.
6. Piskunov NS Differential and integral calculus for universities. Volume 2: Science, 1978.
7. Dubovik VP, Yurik II Higher mathematics. Kyiv: Higher School, 1993.
8. Collection of problems in mathematics (for universities). Edited by Efimova AM, Demidovicha BP, vol.1. M.: Nauka, 1981.
9. Collection of problems in mathematics (for universities). Edited by Efimova AM, Demidovicha BP, vol.2. M.: Nauka, 1981.
10. Higher mathematics. Collection of tasks. For ed .. Dubovika VP, Eureka. Kyiv: ASK, 2001.
11. Kuznetsov AV Collection of problems in higher mathematics (type calculations). M.: Higher school. 1973