обложка

Petrov Yu.P.

Introduction to the fields of mathematics, taking into account coefficients of the finite precision

("Introduction to Mathematics–2")

Innovational text-book

The calculation of a number of examples, as well as dialing

and preparation of the manuscript made graduate student M. V. Voloshin

at Baltic State Technical University "VOENMEH" named after D. F. Ustinov

scientific leader, Doctor of Technical Sciences Professor Sharovatov V.T.

preprint

Translation into English by Vlasova T.A.

St.-Petersburg 2010

Петров Ю.П.

Введение в разделы прикладной математики, учитывающие конечную точность коэффициентов уравнений

("Введение в математику–2")

Инновационное учебное пособие

Вычисление ряда примеров, а также набор и оформление рукописи выполнены аспирантом Балтийского Государственного Технического Университета "ВОЕНМЕХ" имени Д. Ф. Устинова Волошиным М. В.

научный руководитель доктор технических наук профессор Шароватов В. Т.

препринт

Перевод на английский язык – Власова Т.А.

Санкт-Петерург,
2010

Введение

Настоящая книга является инновационным учебным пособием. Это означает, что в ней излагаются существенные для учебного процесса новые научные результаты (в основном полученные автором), но излагаются на уровне, доступном для студентов.

Под названием "Математика-2" автор предлагает объединить те разделы математики, в которых учитываются неточности и погрешности в коэффициентах и параметрах исследуемых математических моделей. "Математикой-1" автор предла-гает называть те разделы математики, в которых коэффициенты и параметры математических моделей (или же законы их изменения) предполагаются известными и заданными.

Так, например, отыскание корней уравнения

$$x^2 + bx + c = 0 \tag{1}$$

при заданных значениях b и c – например, при $b = 2$; $c = 1$ – является задачей "математики-1" . Если же о коэффициентах b и c известно лишь, что они заключены в пределах

$$\begin{array}{l} 1,98 \le b \le 2,02 \\ 0,99 \le c \le 1,01 \end{array} \tag{2}$$

и нужно вычислить (или хотя бы дать оценку) всем возможным корням уравнения (1) при коэффициентах b и c, подчинённым условиям (2), то эта задача относится уже к типичным задачам "математики-2".

Точно так же вычисление решения уравнения Матьё

$$\frac{d^2U}{dz^2} + (a + b\cos z)U = 0 \tag{3}$$

(где функция $a + b\cos z$ с параметрами a и b описывает закон изменения коэффициента при переменной u) при заданных относительно параметров a и b известно лишь то, что они удовлетворяют неравенствам, подобным неравенствам (2), то в этом случае задача оценки свойств возможных решений уравнения (3) (и, в частности, известная задача оценки их устойчивости) относится к "математике-2".

Очевидно, что "математика-2" лучше чем "математика-1" описывает реальный окружающий нас мир, поскольку коэффициенты и параметры математических моделей реальных объектов и явлений почти всегда известны лишь с конечной, ограниченной точностью, а кроме того почти всегда не могут оставаться идеально постоянными и с течением времени испытывают малые отклонения, вариации. Поэтому чаще всего относительно коэффициентов и параметров математических моделей известны лишь интервалы, внутри которых они находятся, – интервалы, задаваемые, например, неравенствами, подобными неравенствам (2). Точное задание коэффициентов и параметров, используемое в "математике-1" – это почти всегда только идеализация.

Однако правильный учёт неточных значений коэффициентов и параметров, учёт возможных интервалов, внутри которых точные значения находятся, является (как увидим далее) во много раз более сложной задачей, чем просто отыскание решения при заданных коэффициентах. Методы решения этой задачи разработаны далеко не для всех математических моделей. Но это направление исследований интенсивно развивается, получены многие важные результаты и автор считает, что настало время выделить это направление исследований и назвать его "математикой-2" . Хорошо известными разделами "математики-2" являются методы приближённых вычислений (с этого раздела обычно начинаются курсы вычислительной математики – см., например, [1], [2]), интервальный анализ – [3], [4] и некоторые другие разделы. В настоящей книге будут приведены новые результаты в области "математики-2" полученные автором.

Дополнительным доводом в пользу выделения "математики-2" служит недавно обнаруженное [5] её отличие от "математики-1" не только в предмете исследования, но и в методологии. В "математике-1" очень широко используются эквивалентные (их называют ещё равносильными) преобразования – т.е. преобразования, упрощающие уравнения (или системы уравнений), но не изменяющие их решений. Неожиданно было обнаружено (см. публикацию [5]), что эти преобразования, не изменяющие самих решений как таковых, могут изменять многие важные свойства решений – такие, как непрерывная зависимость решений от параметров и ряд других свойств, а самое главное – могут коренным образом изменять степень зависимости решений от вариаций параметров. Это означает, что в "математике-2" эквивалентные (равносиль-ные) преобразования можно применять лишь с большой осторожностью, что, конечно, затрудняет исследование.

Тем не менее "математику-2" нужно обязательно знать, поскольку для достижения практических целей методы "математики-1" чаще всего недостаточны. Совершенно недостаточно ограничиться вычислением решений той или иной математической модели. Необходимо удостовериться в надёжности этих решений, а для проверки надёжности необходимо вычислить величину неизбежной и неустранимой погрешнос-ти решений, происходящей из-за почти всегда неизбежной ограниченной точности знания коэффициентов и параметров реальных объектов, из-за почти всегда неизбеж-ного различия между истинными значениями коэффициентов и значениями, приня-тыми при расчёте. В прежние времена этими различиями часто пренебрегали, из-за чего происходило немало аварий и катастроф. В настоящее время, при широком применении вычислительных машин, настала пора перейти к более точным и надёжным (пусть и более сложным) методам расчёта, перейти к использованию методов "математики-2" . Время для этого пришло.

Отметим, что погрешности расчёты имеют разное происхождение. Есть погрешности, связанные с методами расчёта – это погрешности, связанные с конечностью числа итераций в итерационных методах, с ошибками округления и т.п. Эти погрешности хорошо исследованы и их можно уменьшить до приемлемого уровня за счёт совершенствования методов расчёта. В настоящей книге основное внимание уделено другим погрешностям, причиной которых являются неточное знание коэффициентов и параметров математической модели и непредсказуемые их вариации. Погрешности, зависящие от этих причин никакими методами расчёта нельзя ни уменьшить, ни устранить (поэтому их и называют неустранимыми погрешностями). Их

можно только вычислить (или по крайней мере оценить), и эта оценка очень важна, поскольку без неё результаты расчёта не будут надёжными.

Книга состоит из двух частей. В первой части рассматриваются математические модели реальных объектов, имеющие форму систем линейных алгебраических уравнений (СЛАУ), коэффициенты которых известны с ограниченной точностью и заданы системами неравенств.

Основной результат первой части: приведён и обоснован алгоритм точного вычисления неустранимой погрешности каждой из составляющих вектора решений СЛАУ что коренным образом улучшает надёжность результатов расчёта (до последнего времени использовались только приближённые оценки этой погрешности).

Одно из приложений – увеличение надёжности вычисления электростатического потенциала при испускании электронов с катодов различной формы и смежных задач электронной эмиссии, исследуемых под руководством профессора Егорова Н.В. на кафедре моделирования электромеханических и компьютерных систем в Санкт-Петербургском государственном университете.

Во второй части книги рассматриваются объекты, математическими моделями которых являются системы обыкновенных дифференциальных уравнений.

Основной результат второй части: показано, что широко используемые эквивалентные (равносильные) преобразования систем уравнений, не изменяя самих решений как таковых, могут в ряде случаев изменять многие важные свойства решений – такие, как устойчивость, корректность, непрерывная зависимость решений от параметров, степень зависимости решений от вариаций параметров и т.п.

До последних лет эти опасные свойства эквивалентных преобразований не замечались и это послужило причиной многих аварий и катастроф. Использование приведённых в книге методов позволяет повысить надёжность расчётов, уменьшить вероятность аварий и катастроф.

Автор благодарит М. В. Волошина за помощь в работе над учебным пособием, за вычисление ряда примеров и за оформление рукописи.

Preface

This book is an innovation text–book. This means that in it new scientific results (mainly obtained by the author) that are essential for a training process are given. But these results are given on a level accessible for a undergraduates.

Under the name "Mathematics – 2- the author unites such sections in mathematics in which unexactnesses and errors in coefficients and parameters are taken into account. They are investigated in the form of mathematical models.

By "Mathematics–1- the author proposes to name all these sections of mathematics in which coefficients and parameters of mathematical models (or laws of their change) are supposed to be known and assumed.

So, for example, the search of roots in an equation:

$$x^2 + bx + c = 0 \tag{1}$$

if values b and c are given, for example, when $b = 2$; $c = 1$ is a problem of "Mathematics –1". If about coefficients b and c is only known that they are in the limits:

$$\begin{array}{l} 1,98 \leq b \leq 2,02 \\ 0,99 \leq c \leq 1,01 \end{array} \tag{2}$$

and it is necessary to calculate (or if only to give an estimate) all possible roots of equation (1) with coefficients b and c (subjected to conditions (2) then this problem is already typical to problems of "Mathematics–2".

Just the computation of this solution of Matieu equation:

$$\frac{d^2U}{dz^2} + (a + b\cos z)U = 0 \tag{3}$$

(where function $a + b\cos z$ with c and parameters a and b describes a law of coefficients change with a variable U) if values of parameters a and b are given – is a problem of "Mathematics –1".

If in relation to parameters a and b is only known that they satisfy inequalities analogous to inequalities (2) then in this case a problem of estimating problems of possible solutions of an equation (3) (and, in particular, a known problem of estimating their stability) can be attributed to "Mathematics–2".

It is evident that "Mathematics–2- better than "Mathematics–1- describes a real world that surrounds us since a real world that surrounds us since coefficients and parameters of a mathematical model in real objects and phenomena almost always are only known with a finite limited exactness and besides almost always they cannot remain ideally constant and in the course of time they undergo small deviations of variations. Therefore most often only intervals in the interrior of which they are situated are known. These intervals are given, for example, by inequalities similar to (2). An exact assumption of coefficients

and parameters used in "Mathematics–1" – this is almost always only an idealization.

But a correct account of unexact values of coefficients and parameters, an account of intervals in the interior of which exact values are situated (as we shall see later) is a much more complex problem than a problem of searching a solution while we have assumed coefficients. The method of solving this problem has been developed not for all mathematical models. But this branch of investigation is intensely developing. Many important results have been obtained. The author thinks that time has come to isolate this branch of investigation and to name it "Mathematics–2". Methods of approximated computations are well–known sections in "Mathematics–2". From this section usually start the teaching of courses in computing mathematics – see, for example, [1], [2], an interval analysis can be seen in [3], [4] and in some others sections. In this book we shall give new results in the field of "Mathematics–2- by the author.

Additional arguments for the benefit of isolating "Mathematics–2- are its difference from "Mathematics –1- that has been found recently in [5]. There is difference not only in the investigation subject but in methodology as well. In "Mathematics–1- equivalent transformationes are widely used. They simplify equations (or systems of equations) but they do not change their solutions. Unexpectedly it has been found (see [5]) that these transformations that do not change solutions as such they can change many important properties of solutions such as a conditions dependence of solutions on parameters and a series of other properties. And the most important is that they can greatly change a degree of dependence of solutions on parameters variations. This means that in "Mathematics–2- equivalent transformations can be changed only very carefully. This fact, surely, makes the investigation more difficult.

Inspite of this it is necessary by all means to know "Mathematics–2- since in order to achieve practical aims methods of "Mathematics–1"are most often insufficient. It is not sufficient to limit outselves by computing solutions of some mathematical model. It is necessary to be sure in the reliability of these solutions. And in order to test the reliability it is necessary to compute the value of inevitable and unmovable error in solutions that is due to almost always a limit in exactness of coefficients and limits in exactness of coefficients and parameters knowledge of real objects since almost always there exists difference between true values of coefficients and values admitted during computation. In earlier years these differences were often ignored. Due to this cause a lot of wreckages and catastrophes occurred. These phenomena were ignored not because of malicious intent but because there were no methods of solving as yet when parameters were set only by inequalities systems. Recently as computing machines are widely applied it is necessary to pass to more exact and reliable computation methods (although rather complex ones). It is necessary to pass to applying methods of "Mathematics–2". The time has come.

Note that computation errors are of different origin. There are errors that are connected with finity of a number of iterations in iteration methods, with rounding off errors etc. These errors have been well investigated and they can be decreased up to an acceptable level due to modernizing computation methods. In this book the main attention is paid to other errors whose cause is an inexact knowledge of coefficients and parameters of a mathematical model and their unpredicted variations. Such errors that depend on these causes and not be decreased by any computation methods. They also cannot be removed (therefore they are also called unremovable errors). They can only be computed (or at

the worst estimated). And this estimate is very important since without it computation results will not be reliable.

The book consists of two parts. In the first part important mathematical models are considered of different objects that have the form of linear algebraic equations (LAE) whose coefficients are known with a limited exactness. They are set by systems of inequalities.

The main result of the first part is given and based by an algorithm of an unavoidable error in each of components of a vector of solutions in LAE. This fact greatly improves the reliability of computation results. Up to now only approximated estimates of this error were used.

One of applications is the increase of computation reliability is the computation of electronic potential during the emission of electrons from a catode of different form and adjacent problems of electron emission that are investigated at the chair of modeling electromechanical and computer systems (a chair MECS) of ST.Petersburg state university. The head of the chair is professor Yegorov N.V.

In the second part objects whose mathematical models are systems of ordinary differential equations are examined.

The main result of the second part is to show that equivalent transformations of equations systems that are widely used in a series of cases change many important properties such as stability, correctness, continuous dependence of solutions on parameters, the degree of solutions dependence on parameters variations etc.

Up to recent years these dangerous properties of equivalent transformations were not paid attention at. This fact caused a lot of catastrophes and wreckages. The application of methods given in the book allows to increase the reliability of computation, to decrease the probability of computation, to decrease the probably of wreckages and catastrophes.

The author thanks M.V. Voloshin for help in working on a text-book for the calculation of a series of examples and for preparing the manuscript.

Part 1. Investigation of uneliminated errors in solutions of linear algebraic equations systems

§1. Rules for approximated calculations. Intervals analysis.

As it has been indicate in the "Introduction"to "Applied mathematics – 2" – enter such (as for example) rules for approximated solutions, intervals analysis as well as less investigated sections that are well-known and well investigated ones. In this section we shall deal (in quite a short way) with rules for approximated calculations and intervals analysis. Then we shall start analysing the following theme that has not been earlier examined – the analysis of uneliminated errors in solutions of algebraic equations systems. They are due to inexactness of knowing their coefficients, coefficients variations.

Rules for approximated calculations are the most old and well-known section in "Mathematics – 2". They are in details given in the first chapters of a text–books on computation mathematics. There conceptions of estimates in absolute limited errors of an approximated number "a" and estimate its relative error are introduced. Since a difference between an approximated number "a" and its exact value "a_0" as a rule is unknown then under an estimate of an absolute error they mean obtaining the least number "Δ_a" at which the following inequality is fulfilled:

$$|a - a_0| \leq \Delta_a \tag{4}$$

If $\Delta_a = 2\varepsilon_{abc}$ then inequality (4) can be written in the form:

$$a - \varepsilon_{abc} \leq a_0 \leq a + \varepsilon_{abc}$$

where a – a number that is used in later calculation. An unknown to us exact value of this number lies in the interior of an interval $[2\varepsilon a]$. Most often an estimate of a relative error is used. It is written in the form:

$$a(1 - \varepsilon) \leq a_0 \leq a(1 + \varepsilon), \tag{5}$$

where ε – number that is small in comparision with a unity. In this case an exact value of this number lies in the interval $[2\varepsilon a]$. While making approximated calculations it is taken into account that a relative error in a product is near a sum of relative errors if errors in a fraction are surely not dependent between themselves. If factors are dependent between themselves then a relative error in a product can be much less than a sum of factors errors. The same can be said about a quotient. If a dividend error is independent

on a divisor a relative error of a quotient is near a sum of their relative errors.

If there is a dependence between errors of a dividend and a divisor a relative error can be much less than their sum. A relative error in a difference of two numbers can be larger than a sum of their relative errors by many times. Just the substraction of near numbers served the main cause for an error in calculations. Therefore it is necessary to avoid substractions between near numbers.

If rules for approximated calculations have been known for a long period of time intervals analysis has been developed only in the sixty of the XXth century Works by American mathematician Moore R.E. are considered to be the first ones in this theme. Later mathematicians of different countries took part in the development of intervals analysis including scientists of the USSR and Russia. In a book [4] in a bibliography there are presented 740 publications on the subject of intervals analysis envelopping a period of only 1965-82.

The investigations in intervals analysis include intervals that are called intervals intervals that are called intervals numbers. An interval (or "an interval number") A is called a set of real numbers X that satisfy an inequality:

$$\underline{a} \leq x \leq \bar{a}, \tag{6}$$

where a – a left end of the interval Δ and "a" – its right end. An interval "Δ"is written as $[a; a]$.

Analogically we call B an interval (an interval number) that is a set of numbers that satisfy an inequality:

$$\underline{b} \leq x \leq \bar{b} \tag{7}$$

and it is written as $B = [b; b]$.

An interval with coinciding ends when $a = \bar{a} = a$ is called a degenerated interval. It is identified with a usual real number "a". Also a conception of "a zero containing an interval" is introduced if $\underline{a} < 0 < \bar{a}$.

While using these conceptions an interval arithmetics is formed. i.e. operations of addition, substruction, multiplication and division are determined. They are determined by the following equations:

$$A + B = [\underline{a}; \bar{a}] + [\underline{b}; \bar{b}] = [\underline{a} + \underline{b}; \bar{a} + \bar{b}] \tag{8}$$

$$A - B = [\underline{a}; \bar{a}] - [\underline{b}; \bar{b}] = [\underline{a} - \underline{b}; \bar{a} - \bar{b}] \tag{9}$$

$$A \cdot B = [\underline{a}; \bar{a}] \cdot [\underline{b}; \bar{b}] = [min(\underline{a} \cdot \underline{b}; \bar{a} \cdot \underline{b}; \underline{a} \cdot \bar{b}; \bar{a} \cdot \bar{b}); max(\underline{a} \cdot \underline{b}; \bar{a} \cdot \underline{b}; \underline{a} \cdot \bar{b}; \bar{a} \cdot \bar{b})]; \tag{10}$$

$$\frac{A}{B} = \frac{[\underline{a}; \bar{a}]}{[\underline{b}; \bar{b}]} = [\underline{a}; \bar{a}] \cdot [\frac{1}{\bar{b}}; \frac{1}{\underline{b}}] \tag{11}$$

An operation of division can not be carried out if interval B is an interval that contains a zero.

If A and B are degenerated intervals then equations (8)–(11) coincide with convenient arithmetic operations over real numbers. Thus an interval number is a generalization of a real number.

Let us at once note that an interval arithmetics (not to say anything about interval analysis) is much more complex than a conventional arithmetics. Besides we must not say anything about the fact that in interval arithmetics an important law of distribution is not satisfied, i.e. in it an equality

$$A(B + C) = AB = AC \tag{12}$$

is not always true. All this leads to the conclusion that interval analysis turns out to be a rather complex section of mathematics. Besides it is not taught in all mathematical departments of Universities. So at St.Petersburg state University it has been taught at a department of applied mathematics and control processes (AM–CP) but from 2004 it has become not a required subject. At the AM–CP department it was made an optional subject which is not obligatory.

The cause of difficulties that has arisen during the study of an interval analysis is a very wide and difficult problem. In an interval analysis a smallness of an interval $[\underline{a}; \bar{a}]$ in comparision with a number "a" is not supposed. At the same time in the great majority of technical problems an interval in the interior of which are included unknown to us true values of coefficients and parameters of a mathematical model are almost always small in comparison with values themselves. We can apply the smallness of an interval for an essential simplification of the solution.

Just the smallness of intervals indeterminess while we set coefficients and parameters in a mathematical model will be applied later, in our statement. We shall observe such systems of equations about whose coefficients it is only known that they are situated in

some intervals. But in contract to problems set about interval analysis they are situated in intervals that are small in comparison with coefficients themselves. We shall study in which intervals are contained solutions of investigated problems.

§2. Systems of linear algebraic equations (SLAE).

We call systems of linear algebraic equations (in short - SLAE) the following systems:

$$\left.\begin{array}{c} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2 \\ ..... \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n = b_n \end{array}\right\} \tag{13}$$

which as a rule are written in a vector-matrix form:

$$AX = B \tag{14}$$

where "A" – is a square matrix of coefficients a_{ij} of a measure nxn, B–vector – column of right sides, $x - n$ – measured vector – column of solutions x_1; x_2; ...; x_n. Numbers x_i we shall call components of a vector of solutions X sometimes (for short) we shall call them solutions although a set of numbers x_1; x_2; ...; x_n is their solution that turns equation (14) into an identity.

In quite different sphere of physics and technique we meet with the necessity of computing solutions of SLAE. So, for example, the computation of different building constructions directly leads to systems of equations of the type (14).

The calculation of electric chains by Kirchhoff method also leads to the following systems:

$$\left.\begin{array}{c} r_{11}i_1 + r_{12}i_2 + ... + r_{1n} = E_1 \\ r_{21}i_1 + r_{22}i_2 + ... + r_{2n} = E_2 \\ ..... \\ r_{n1}i_1 + r_{n2}i_2 + ... + r_{nn} = E_n \end{array}\right\} \tag{15}$$

where v_{ij} – resistance, i_i contour currents, E_0 – contour electrodriving forces.

Besides subjects whose mathematical models from the begining are SLAE they appear during a numerical computation of objects whose mathematical models are partial differential or integral equations.

So, for example, the calculation of electrons emission from a cathode leads to partial differential equations. The most popular method of their solution is a method of nets or a method of finite differences based on the reduction of an initial equation to a system of linear algebraic equations.

So, for example, in Fredholm integral equation

$$\int_a^b K(x;s)y(s)ds = k(x)$$

a function $y(x)$ will be a sought function. A function $b(x)$ is a known function (it's the right side of the equation), function $K(x;1)$ is a function of two variables "x" and "s" is called a kernel.

A main method of solving integral equations is based on the change of an integral by a finite sum. For this homogenous nets of knots with a step Δ_s by a variable s and a step Δx by variable x. By changing a continuous function $K(x;1)$ by its value in knots we turn an integral equation into a system of algebraic equations:

$$A_{ij}y_i = k_i$$

where $i = 1, 2, ..., m,\ j = 1, 2, ..., n$.

By solving this system we shall obtain a table of values of "y" in the sought function $y(x)$.

There exists a lot of other problems (some of them are considered in [6]) a stage of solving which is to solve some kind of SLAE. The author of a known book [7] in general thinks that at present more than 70% of mathematical calculations is the computation of solutions of some SLAE.

Note that solutions themselves of SLAE are computed by rather simple ways. The most often an algorithm of a successive exclusion of variables have been developed already by Gauss. In order to compute all components of a solution from x_1 up to x_n this algorithm requires approximately n^3 operations of multiplications (where "n" – an order of a system). Iteration methods are also widely applied.

The main difficulty is not in computation itself of a solution but in an estimate of some unremoved error that arises because all coefficients in SLAE (as we have already mentioned) are known almost always only with a finite limited exactness.

Examples of such SLAE are known in which even the smallest (and thus – inevitable) errors in coefficients lead to large errors in a solution.

Here is a very simple example (№1):
a system

$$\left.\begin{array}{r} 10,02x_1 + x_2 = 11,02 \\ 10x_1 + x_2 = 11 \end{array}\right\} \tag{16}$$

has solutions $x_1 = x_2 = 1$.

If a coefficient in x_1 in the first equation will change by $\frac{1}{1000}$ from an initial value then system (16) will become:

$$\left.\begin{array}{c} 10,01x_1 + x_2 = 11,02 \\ 10x_1 + x_2 = 11 \end{array}\right\} \tag{17}$$

and it will obtain solutions $x_1 = 2$; $x_2 = -9$ i.e. it will change by two times and x_2 by nine times.

Even the existence of such systems means that it is not at all sufficient to calculate a solution in SLAE. It is necessary (by all means) to check and to see that this solution will not essentially change or even substantionally change during inevitable small errors in system coefficients.

Without such a check a solution in SLAE cannot be considered to be reliable and authentical. An uncritical attitude to solutions absence during a careful check of a value of unavoidable errors will lead to (and have already led to) wreckages and even catastrophes. In [6], [10] examples of catastrophes are given.

Depending on sensibility of solutions of coefficients variations in equations systems we traditionally divide these systems into well – conditioned ones in which small changes in coefficients lead to small changes in solutions and systems that are "ill–conditioned- in which during same small changes in coefficients changes in solutions can be large.

In order that such division of systems can be really realized it must be determined (in addition) what changes of coefficients and solutions can be considered "small".

Let us propose the following definition. We shall call such coefficients "small- (or variations) that do not exceed a real inevitable indifference in values of coefficients in a real object. It arises due to a finite exactness of "small- variations of coefficients in the causes of exploitation and due to other such causes. We shall call such changes in solutions "small- that do not lead to the violation of a normal work of an examined construction. If during "small- changes in coefficients that have been defined in such a way changes in solutions will be "small- (in a sense of an above given definition) such solutions (numbers x_i) will be called reliable. But if during the same "small- changes of coefficients changes of solutions will not be "small- (in a sense of an above definition) then such solutions will be considered unreliable. Introduced definitions make more precise widely applied but rather dim definitions of "well–conditioned- and "ill–conditioned- systems of equations and their solutions.

Later we shall say that there exist such SLAE in which during small changes of coefficients solutions can become as positive as negative and they also can be infinitely

large in absolute values. Such systems will be called very "ill–conditioned". Solutions of such SLAE are very unreliable. It is necessary to find such SLAE and it is necessary to avoid them.

§3. Estimates of errors in solutions by means of "number of condition".

In a theory of systems of linear algebraic equations by means of a value in matrix "A" – determinant and well – known and are widely applied. Let us consider (in short) the worth and backgrounds of these estimates. And later in next sections we shall turn to a more detailed statement of a methodics of an exact calculation of a possible error. This methodics has been earlier proposed by the author which was given [6] (in short).

Due to inevitable inexactness in setting coefficients in a system (13) in relation to its coefficients we can only state that they are in the interior of some intervals:

$$a_{ij}(1 - |\varepsilon_{ij}| \leq \bar{a}_{ij} \leq a_{ij}(1 + \varepsilon_{ij}|) \tag{18}$$

$$b_i(1 - |\delta_i|) \leq \bar{b}_i \leq b_i(1 + \delta_i|) \tag{19}$$

where $\bar{a}_{ij}$ and $\bar{b}_i$ – true not known to us coefficients in system (13). But a_{ij} and b_i – values that were admitted during the calculation (usually they are called rating values), $|\varepsilon_{ij}|$ and $|\delta_i|$ – numbers that are small in comparison with 1. They characterize a relative error of rating values.

If there is an estimate of absolute but not a relative error then in this case true values of coefficients are in the interior of intervals which are determined by the following inequalities.

$$a_{ij} - |\varepsilon_{ij}| \leq \bar{a}_{ij} \leq a_{ij} + |\varepsilon_{ij}| \tag{20}$$

$$b_i - |\delta_i| \leq \bar{b}_i \leq b_i + |\delta_{ij}| \tag{21}$$

where in this case $|\varepsilon_{ij}|$ and $|\delta_i|$ – measured values whose measure coincides with the measure of coefficients themselves. More often we have to deal with inequalities (18)–(19), i.e. relative errors.

Estimates of solutions errors in SLAE are naturally carried out by two steps. The first one is an estimate of the largest possible coefficients errors, i.e. – an estimate of values ε_{ij} and δ_i in equalities (18)–(21). On the second stage on the base of estimates ε_{ij} and δ_i errors in solutions are calculated.

The first stage is a purely engineering problem. Value ε_{ij} and δ_i are estimated for each separate case – object – on the basis of a deep knowledge of its properties and its particularies. Therefore the first stage of this general problem of estimating errors it this general problem of estimating errors it this book will not be considered. We shall suppose that values ε_{ij} and δ_i are known and are assumed.

The second step – with the help of known ε_{ij} and δ_i – is to estimate errors in solutions – is a typical problem of applied mathematics. The author has developed a methodics of its solution (which will later be stated). It can be used in computing any objects whose mathematical model is SLAE. The second stage has the following difficulty: a solution error depends not only on absolute values ε_{ij} and b_i but on their signs and a number of possible combinations of signs in coefficients variations is very large. In matrix A of a measure nxn a number of possible combinations is equal to 2^{n^2}. Even if $n = 10$ it is equal to $2^{100} > 10^{30}$. Therefore a direct computation of errors in all possible combinations of signs most often can not be carried out even for the most modern quick operating machines. It is necessary to apply indirect methods.

The most simple method of a primary estimate of a possible error in solving SLAE $AX = B$ is a computation of a determinant in matrix A:

$$detA = \begin{vmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{vmatrix} \tag{22}$$

If $detA$ is much less than coefficients a_{ij} this means that even during small variations of coefficients in a solution errors can be large. If we return to earlier considered system (16) we can note that for it

$$detA = \begin{vmatrix} 10,02 & 1 \\ 10 & 1 \end{vmatrix} = 0,02 \tag{23}$$

is by 50 times less than the least of coefficients in matrix A. Then we shall not wonder that a change of a coefficient 10,02 leads to only $\frac{1}{1000}$ leads to a serious changes in solutions of system (16).

The computation of determinants is not a difficult operation. In order to compute determinants of an order equal to higher than three usually the reduction of a determinant is used – to a "triangular"form – since such a determinant is equal to a product of its elements that are on a main diagonal. In order to reducing to a "triangular"form by means of successive multiplications and additions all elements that are situated lower than the main diagonal turn to zero. Such a reduction requires approximately n^3 multiplications, i.e. we can compare the calculation of a determinant of order "n" and a calculation of a solution in SLAE by their labove consuming character.

But the smallness of a determinant allows us to speak only about possible large errors in solutions. But in order to estimate their value we must use a methodics based on applying "numbers of condition". Primarily a conception of such a matrix that is increase to matrix A (it is denoted by A^{-1}) is introduced. Besides a determination of a multiplication of matrix A by an inverse matrix A^{-1} is carried out. It is equal to a unity matrix E, i.e.:

$$A \cdot A^{-1} = E \tag{24}$$

where E – matrix in which a unity is on the main diagonal. And all other elements are zeros, i.e.:

$$E = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 & 0 \\ 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix} \tag{25}$$

From equality (24) in linear algebraic equations a formula for elements of an inverse matrix is formed:

$$A^{-1} = \frac{1}{detA} \begin{pmatrix} A_{11} & A_{21} & ... & A_{n1} \\ A_{12} & A_{22} & ... & A_{n2} \\ ... & ... & ... & ... \\ A_{1n} & A_{2n} & ... & A_{nn} \end{pmatrix} \tag{26}$$

where A_{ij} – algebraic additions of elements a_{ij} matrix A.

An inverse matrix can be applied for the calculation of solution X in a system $AX = B$ according to the formula

$$X = A^{-1}B \tag{27}$$

i.e. in order to calculate X it is sufficient to multiply matrix A^{-1} by a vector of a right side B. But formula (27) is ravely used. More often a method of Gauss is applied or iteration methods since a calculation of an inverse matrix is labour consuming. Really, while, for example, we apply a formula (26) in order to compute A^{-1} it is necessary to compute n^2 algebraic additions, i.e. determinants of the order $n-1$. The calculation of each determinant requires approximately $(n-1)^3$ multiplications. In order to calculate all determinants it is required $n^2(n-1)^3$ multiplications. And this is more than it is required during the application of a Gauss method.

For matrix of the second order having a measure $2x2$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{28}$$

There will be an inverse matrix:

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \tag{29}$$

i.e. elements a_{11} and a_{22} that are on a main diagonal are shifted and elements a_{12} and a_{21} remain in their places but they change their signs and then they all divid by $detA = a_{11} + a_{22} + a_{12} + a_{21}$.

An important conception of linear algebra is a conception of matrix A norm and of a vector. A norm of a matrix (vector) is a number calculated according to a certain rule by means of elements in a matrix or vector. Most often Euclid (or spherical) norm calculated by a rule is applied:

$$||B|| = \sqrt{b_1^2 + b_2^2 + ... + b_n^2} = \sqrt{\sum_{i=1}^{n} b_i^2} \tag{30}$$

for a vector and according to a rule:

$$||A|| = \sqrt{\sum_{i=1;\, j=1}^{n} a_{ij}^2} \tag{31}$$

for a matrix.

Other norms are used. So, for example, a cubical norm of a vector (according to a terminology in a text-book [8] is determined by a formula

$$||B||_{kub} = max|b_i| \tag{32}$$

and its octahedric norm – by a formula:

$$||B||_{okt} = \sum_{i=1}^{n} |b_i| \tag{33}$$

Note

We must especially note that the designation of matrixes and their norm has not been stated finally. In different books and articles we can meet with different designation. Let us recall that we shall design matrixes by vertical lines with roundings from above and from below (formulas (25)–(29)). A determinant of matrix A is designed by $detA$ or vertical lines from the left and from the right (formula (22)). Norms of a matrix and a vector are designated by double vertical lines from the right and the left (formula (30) and (31)). In the names of Euclid, cubical and octahedrical norms we follow a known text-book [8].

Properties of Euclid norm:

$$||AX|| \leq ||A|| \cdot ||X|| \tag{34}$$

i.e. a norm of a product is not larger than products of norms similarily.

$$||A|| + ||B|| \leq ||A|| + ||B|| \tag{35}$$

i.e. a norm of a sum is not larger than a sum of items norm.

From a property (34) it follows:

$$||A|| \cdot ||A^{-1}|| \geq ||E|| = 1 \tag{36}$$

i.e. products of norms in direct and inverse matrixes is always more than 1. By using properties of matrixes a norm of a solution error is calculated. We must take into account that if, for example, in equation $AX = B$ coefficients in the right side have changed and instead of vector B a vector $B + \Delta B$ appeared then a solution has also changed and instead of equation

$$AX = B \tag{37}$$

an equality will appear:

$$A(X + \Delta X) = B + \Delta B \tag{38}$$

or –that is the same:

$$AX + A\Delta X = B + \Delta B \tag{39}$$

If we substract from (39) an equality (37) by members we shall obtain

$$A\Delta X = \Delta B \tag{40}$$

Hence it follows

$$\Delta X = A^{-1}\Delta B \tag{41}$$

If we turn to norms we have

$$||\Delta X|| \leq ||A^{-1}|| \cdot ||\Delta B|| \tag{42}$$

Hence it follows that

$$\frac{||\Delta X||}{||X||} \leq ||A|| \cdot ||A^{-1}|| \frac{||\Delta B||}{||B||} \tag{43}$$

A product $||A|| \cdot ||A^{-1}||$ is called "a number of condition". And formula (43) shows that a relative change of the right side satisfies inequalities (19) then formula (43) will become:

$$\frac{||\Delta X||}{||X||} \leq ||A|| \cdot ||A^{-1}|| \cdot \delta \tag{44}$$

Example 2. Let us consider a simple system:

$$\begin{array}{c} 2x_1 + x_2 = 2 \\ x_1 + x_2 = 1 \end{array} \tag{45}$$

whose solution will be: $x_1 = 1$; $x_2 = 0$. For this system $||A|| = \sqrt{2^2 + 1 + 1 + 1} = \sqrt{7}$,

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

And thus, $||A^{-1}|| = \sqrt{1 + 1 + 1 + 2^2} = \sqrt{7}$ and a number of condition

$$||A|| \cdot ||A^{-1}|| = 7$$

On the basis of formula (44) we can state that since $||X|| = 1$, then

$$||\Delta X|| = 7 \cdot \delta$$

and if, for example, $\delta = 0,1$ then

$$||\Delta X|| = \sqrt{\Delta x_1^2 + \Delta x_2^2} \leq 0,7 \tag{46}$$

If the right side in the first of equations (45) has changed by 10% and has become equal to 2,2 and the right side of the second of equations (45) has become 1,1 then the solutions of systems (45) will become $x_1 = 1,1$, $x_2 = 0$. This means that $Ax_1 = 0,1$, $x_j = 0$, $||\Delta X|| = 0,1$ and inequality (46) is satisfied.

If $\delta = 0,1$ to the largest changes in solutions will a lead a value of the right side in equations (45) equal to 2,2 and 0,9 as it was shown in [6]. System (45) will become

$$\left.\begin{array}{c} 2x_1 + x_2 = 2,2 \\ x_2 + x_3 = 0,9 \end{array}\right\}$$

and it will obtain solutions: $x_1 = 1,3$; $x_2 = -0,4$.

In this case a value $|\Delta X|$ becomes the maximal possible one and equal to $\sqrt{0,3^2 + 0,4^2} = 0,5$ and inequality (46) is satisfied.

This simple example at once shows that an estimate by means of a "number of condition"most often is not exact even for maximally possible solution errors. Although in formula (43) these is a sign "$\leq$"that means that an equality of the left and right sides of formula (43) is possible (that corresponds to an exact estimate) but in practice sign of equality can appear only for some matrixes A and a vector B. For a majority of SLAE a number of condition gives only an approximated estimate for a norm of errors in solutions.

If it is necessary to take into account as variations of coefficients in the right side of a system of equations $\Delta X = B$ as variations of coefficients in matrix A then instead of formula in matrix A then instead of formula (43) a following formula – its generalization is:

$$\frac{||\Delta X||}{||X||} \leq ||A|| \cdot ||A^{-1}|| \cdot \left(\frac{||\Delta B||}{||B||} + \frac{||\Delta A||}{||A||} \right). \tag{47}$$

If coefficients in matrix A and vector B satisfy inequalities (18) and (19) then formula (47) becomes:

$$\frac{||\Delta X||}{||X||} \leq ||A|| \cdot ||A^{-1}||(\delta + \varepsilon) \tag{48}$$

A detailed analysis of formulas (43), (44), (47), (48) is given, for example, in [9], [10] and in many other text–book and handbooks.

Note. In publications [13], [14] estimates by a number of condition that make more precise a formula (48) are given. But they do not change the essence. Estimates by a "number of condition"remain approximated estimates and they possess a series of drawbacks about which we shall speak in next section.

§4. Drawbacks in estimates by means of a "number of condition".

The main drawback of estimates of SLAE solutions by "a number of condition"is that only a norm $||\Delta X||$ of all components in solutions X is estimated – from x_1 up to x_n. This can be compared with the expression – "something is middling in a hospital". In practice most often it is necessary to estimate an error of a certain component x_i since often a large error in errors of other components of a solution X and with it a norm $||\Delta X||$ can at this time remain in admissible limits.

Example №3.

Let us consider a simple system:

$$\left.\begin{aligned} 3x_1 + 2x_2 = 23 \\ x_1 + x_2 = 11 \end{aligned}\right\} \tag{49}$$

with a solution $x_1 = 1$; $x_2 = 10$. For it a matrix

$$A = \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}$$

and an inverse matrix

$$A^{-1} = \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix}$$

Therefore $||A|| = \sqrt{3^2 + 2^2 + 1 + 1} = \sqrt{15}$; $||A^{-1}|| = \sqrt{1 + 2^2 + 1 + 3^2} = \sqrt{15}$ and a "number of condition"is $||A|| \cdot ||A^{-1}|| = 15$.

Now let coefficients in matrix A in system (49) underwent variations and system (49) has turned into

$$\begin{aligned} 3(1-\varepsilon)x_1 + 2(1+\varepsilon)x_2 = 23 \\ (1+\varepsilon)x_1 + (1-\varepsilon)x_2 = 11 \end{aligned}$$

By using known Cramer formulas according to which systems of equations $AX = B$ will become:

$$x_i = \frac{D_i}{D} \tag{50}$$

where D – determinant of matrix A (i.e. – $detA$) and D_1 – a determinant of the same matrix. But in it the i-th line has been changed by a vector–line B of the right side. It is not difficult to compute:

$$D = \begin{vmatrix} 3(1-\varepsilon) & 2(1+\varepsilon) \\ (1+\varepsilon) & (1-\varepsilon) \end{vmatrix} = 1 - 10\varepsilon + \varepsilon^2$$

$$D_1 = \begin{vmatrix} 23 & 2(1+\varepsilon) \\ 11 & (1-\varepsilon) \end{vmatrix} = 1 - 45\varepsilon$$

$$D_2 = \begin{vmatrix} 3(1-\varepsilon) & 23 \\ (1+\varepsilon) & 11 \end{vmatrix} = 10 - 56\varepsilon$$

$$x_1 = \frac{D_1}{D} = \frac{1 - 45\varepsilon}{1 - 10\varepsilon + \varepsilon^2}$$

$$x_2 = \frac{D_2}{D} = \frac{10 - 56\varepsilon}{1 - 10\varepsilon + \varepsilon^2}$$

Formulas for x_1 and x_2 at once show that even in the most simple systems consisting of two equations of dependence x_1; x_2 and thus – and $||\Delta X||$ on ε are mainly nonlinear.

Now let us form a table of dependence of x_1 and x_2 on ε.
Table 1.

1	ε	0,001	0,01	0,02	0,05	0,1	0,10102	0,15	0,2
2	D	0,99	0,9001	0,3004	0,5025	0,01	0	-0,4775	-0,96
3	x_1	0,965	0,6111	0,125	-2,49	-350	$\mp\infty$	12,04	8,333
4	x_2	10,05	10,499	11,1	14,32	440	$\pm\infty$	-3,35	1,25
5	Δx_1	0,035	0,389	0,875	-3,49	-349	$\mp\infty$	11,04	7,333
6	Δx_2	0,044	0,499	1,1	4,32	430	$\pm\infty$	13,35	8,75
7	$\|\|\Delta X\|\|$	0,0565	0,635	1,405	5,55	555	∞	18,12	1,19
8	$\frac{\|\|\Delta X\|\|}{\|\|X\|\|}$	0,00562	0,0625	0,1398	0,552	55,1	∞	1,7237	1,136
9	$\|\|A\|\| \cdot \|\|A^{-1}\|\| \cdot \varepsilon$	0,015	0,15	0,3	0,75	1,5	1,1515	2,25	3

A considered simple example at once shows the main drawback of estimating errors in solutions by a number of condition. Let us take the second line of a table 1. If $\varepsilon = 0,01$ an estimate by a number of condition says that a relative norm in an error of solutions x_1 and x_2 does not exceed $15 \cdot 0,001 = 15\%$ form a norm of x_1 and x_2. In fact this estimate is very rough and a real norm of an error is much less and it is equal to only 6,3% from a norm of x_1 and x_2. It seems that everything is well. In fact a relative error x_1 is equal to $38,9\%$ and a relative error in x_2 is equal to $4,99\%$.

This simple example already shows that an estimate by a "number of condition"is not perfect and its application can become a cause of catastrophes and wreckages.

Note that a hope of these who think that although a "number of condition"gives a rough estimate (rather excessive) but that is a (guaranteering) estimate of a relative norm of an error in a vector of solutions X is justifved. The comparison of the eighth and the nineth lines in table 1 shows that if $\varepsilon \leq 0,05$ an estimate by a number of condition really gives for a relative error an estimate from above. But when a value ε approaches to a such one at which $D = 0$ a value of errors x_1 and x_2 swiftly increases and when $\varepsilon = 0,1$ an estimate by a "number of condition"already does not guarantee anything.

In [6] additional drawbacks of estimates by a "number of condition- have been noted that were not earlier taken into account. Let us enumerate them:

1. a dependence of a "number of condition- on equivalent transformations of equations.

Let us consider a quite simple system:

$$2x_1 + x_2 = 1 \tag{51}$$

$$x_1 + x_2 = 1 \tag{52}$$

with a solution $x_1 = 0$; $x_2 = 1$. Let us multiply all members of the second equation by a number K. After multiplication it will become

$$kx_1 + kx_2 = k \tag{53}$$

and we shall have to deal with a system of equations (51)–(53). This system has the same solution $x_1 = 0$; $x_2 = 1$ as system (51)–(52). This must be so since a multiplication of all members by a constant K which is equal to zero is an equivalent transformation. Note that in this case this equivalent transformation does not change a degree of conditioning of solutions. Really, equation (53) for any K remains an equation of the same straight line on a plane with axes Ox_1; Ox_2. And an angle between straight lines (51) and (53) for any $K \neq 0$ remains the same. This means that a degree of conditioning of solutions during multiplications by $K \neq 0$ does not change.

Now let us compute for a system (51)–(53) "number of condition". For this example matrix A is equal

$$A = \begin{pmatrix} 2 & 1 \\ K & K \end{pmatrix} \tag{54}$$

a matrix determinant is:

$$detA = \begin{vmatrix} 2 & 1 \\ K & K \end{vmatrix} \tag{55}$$

an inverse matrix is:

$$A^{-1} = \frac{1}{K} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \tag{56}$$

Euclid norm of matrix A is:

$$||A|| = \sqrt{5 + 2K^2}$$

The same norm for an inverse matrix

$$||A^{-1}|| = \frac{1}{K}\sqrt{5 + 2K^2}$$

and a number of condition:

$$||A|| \cdot ||A^{-1}|| = \frac{5}{K} + 2K$$

turns out to be dependent on K.

We have following dependence of a number condition on K (table 2).

Table 2. The dependence of a "number of condition"on K.

K	0,1	1	1,5	1	10	100
$\|\|A\|\| \cdot \|\|A^{-1}\|\|$	50,2	7	0,33	8,5	20,5	200,05

This means that while dealing with equation (53) for different coefficients K we must make quite different conclusions about a "number of condition- of solutions of system (51)–(53). If $K = 1,5$ solutions will seem to be well – conditioned but when $K = 100$ they will seem to be ill-conditioned. In fact the dependence of systems (51)–(53) solutions on K does not depend – Just the same conditioning of solutions in any system $AX = B$ will not depend on the multiplication of all members in any of equations of a system by any number $K \neq 0$.

The methodics of estimating an error in solutions by "a number of condition"can lead us to a wrong way and – to incorrect conclusions.

The dependence of the product $(||A|| \cdot ||A^{-1}||$ on multiplication of all numbers in any of system equations by a constant has been noted in ([9],p. 212). The influence of this dependence on the correctness of estimating an error by means of a "number of condition"in [9] was not considered and it was considered in [6] and [11].

Besides a "number of condition- a determinant of a system equations $AX - B$ also greatly depends on such an equivalent transformation as a multiplication of all members by a number $K \neq 0$. So a determinant in system (51)–(55) is equal to $detA = \begin{vmatrix} 2 & 1 \\ K & K \end{vmatrix} = K$ and in dependence on number K it can become large and small as well. This fact stresses the nonreliability of estimating a degree of condition in solutions by a determinant value of a system. Often we can hear the following statement: if a system determinant is small (in comparison with a norm of matrix A) then a system is ill-conditioned. If a determinant is not small the system is well–conditioned. In fact such a statement is very unreliable since $detA$ depends on equivalent transformations. In [9] about it was already spoken.

Note that in [6] and earlier in [5] and (11) examples were given when equivalent transformations that did not change solutions themselves really had changed correctness and conditioning of solutions and it was shown that these phenomena (that rarely occur) but a possible change of a series of properties of solutions played a very important role during equivalent transformations were causes as for wreckages and catastrophes and also in their prevention. Now we see that an inverse case can occur when some of calculation methods for the errors in solutions lead to false conclusions about the dependence of these errors on equivalent transformations.

2. A false dependence of "number of condition- on the scale of measuring equations coefficients.

One more important drawback of "numbers of condition- is their dependence on the choice of units of measure of rated coefficients in equations.

Example №4.

Let us examine a simple construction. A beam AB lies on two supports at it is seen on figure 1.

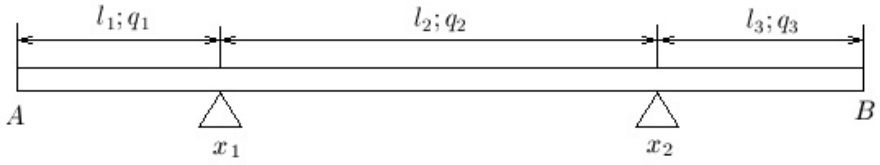

fig. 1

The length of a beam is more left than the left support and it is equal to l_1 meters with a specific loading on a unit of a length $q_1 kN/m$ (kilonewtons/metre). The length of a place between supports is equal to l_2 meters with specific loading $q_2 kN/m$. The length of a beam that is more right than a right support is equal to l_3 meters with specific loading $q_3 kN/m$. A loading on a left support we denote by x_1 – on a right one – by x_2 . Supports are supported to be not holded. This means that if, for example, a loading x_1 turned out to be negative than a beam looses static stability, it will slide from the right support and fall. In order to find x_1 and x_2 we can form equations of equilibrium of forces moments in relation to points A and B. In relation to:

$$l_1 x_1 + (l_1 + l_2)x_2 = (l_1 q_1 + l_2 q_2 + l_3 q_3) \cdot (l_1 + l_2 + l_3) \cdot \frac{1}{2} \tag{57}$$

In relation to B we have:

$$(l_2 + l_3)x_1 + l_3 x_2 = (l_1 q_1 + l_2 q_2 + l_3 q_3) \cdot (l_1 + l_2 + l_3) \cdot \frac{1}{2} \tag{58}$$

Any of equations of equilibrium of moments can be changed by an equation of forces equilibrium in order to compute x_1 and x_2:

$$x_1 + x_2 = l_1 q_1 + l_2 q_2 + l_3 q_3 \tag{59}$$

If lengths are measured in meters, special loadings q_1; q_2; q_3 in kilonewtons (meter, x_1 and x_2 – in kilonewtons and $l_1 = l_2 = 2$ meters, $l_3 = 3,8$ in $q_1 = q_2 = q_3 = 1kN/m$ then equations (58) and (59) became:

$$\left.\begin{aligned} 5,8x_1 + 3,8x_2 = 30,42 \\ x_1 + x_2 = 7,8 \end{aligned}\right\} \tag{60}$$

If lengths are measured in millimeters (as in a technical draft) and loadings – correspondingly – 0,001 kN/m them equations (2) and (12) became:

$$\left.\begin{array}{r} 5800x_1 + 3800x_2 = 304200 \\ x_1 + x_2 = 7,8 \end{array}\right\} \tag{61}$$

We see that the change of measurement units turned out to be equivalent to multiplying all members of the first of equations (60) by 1000. Therefore solutions x_1 and x_2 in systems (60) and (61) will be the same but "numbers of condition- change, as we have already seen, while multiplying all members of one of equations of a system by a constant value "numbers of condition- can change. This circumstance, certainly, speaks about serious (as it was shown in [6] can be removed) drawbacks in estimates by "numbers of condition". But it is necessary to note that the dependence of a "number of condition- on a choice of measurement units will appear only when equations of an examined system have different measurement. Thus all members of equations (59) in a system (59)-(60) have a moment measurement (a force multiplied by length) and all members of equation (60) have force measurement.

3. False opinions about an influence of system parameters on condition of solutions.

Simplicity and universality of computing "numbers of condition- allows us to apply them for an estimate of influencing some parameters of an object on the condition of solutions and thus – on the reliability of work of an object itself as well. But here false opinions can be possible. This is especially important in the course of projecting when it is necessary to quickly estimate many possible projection variants and to choose from them such one that will secure the least change of solutions (and thus – the least change of properties in a projected object) and with inevitable successive changes of parameters in an examined object in the course of exploitation.

Example №5

Let us consider a simple system.

$$\begin{array}{c} (1+m)x_1 + x_2 = 2 \\ x_1 + x_2 = 1 \end{array} \tag{62}$$

with solutions $x_1 = \dfrac{1}{m}$; $x_2 = \dfrac{m-1}{m}$. Let us suppose that coefficients in a system (62) can undergo variations and let us follow a dependence of a condition degree of solution on parameter m.

For system (62) we have:

$$A = \begin{pmatrix} 1+m & 1 \\ 1 & 1 \end{pmatrix};\ detA = m;\ ||A|| = sqrt(1+m)^2 + 3 \tag{63}$$

$$||A|| \cdot ||A^{-1}|| = \frac{4}{m} + m + 2$$

The dependence of "number of condition"on m is seen on table:

m	0,1	1	2	10	100
$\|\|A\|\| \cdot \|\|A^{-1}\|\|$	42,1	7	6	12,4	102,04

According to formula (63) there follows that during the growth of parameter m when $m = 2$ conditioning of a system worsenes and, for example, if m changes from $m = 2$ up to $m = 10$ also worsen more than by two times.

In fact, surely, nothing occurs of the kind. Really, equation in a system (62) – an equation of a straight line on a plane ε with axes Ox_1; Ox_2. A tangent of an angle between straight lines can be computed and is equal to:

$$tg\varphi = \frac{m}{m+2} \tag{64}$$

From formula (64) it follows that an angle between straight lines monotonously increases with the growth of m and thus a degree of conditioning of solutions also increases with the increase of m. A methodics based on "numbers of condition"gives a false answer.

This can be affirmed by a direct calculation as well. Let us consider the most unfavourable combination of signs in coefficients of system (62) when it turns into a system:

$$\left.\begin{array}{r} (1+m)(1-\varepsilon)x_1 + (1+\varepsilon)x_2 = 2 \\ (1+\varepsilon)x_1 + (1-\varepsilon)x_2 = 1 \end{array}\right\} \tag{65}$$

when $m = 2$ and $\varepsilon = 0$ solutions of a system (65) will be $x_1 = x_2 = 0,5$ and if $\varepsilon = 0,1$ solutions will be $x_1 = 0,574$; $x_2 = 0,41$. A relative variation of x_1 is:

$$\frac{\Delta x_1}{x_1} = \frac{0,574 - 0,5}{0,5} = 0,148 = 1,48\varepsilon$$

.

For x_2 we shall have $\dfrac{\Delta x_2}{x_2} = \dfrac{0,41 - 0,5}{0,5} = -0,18 = -1,8\varepsilon$.

If $m = 4$ we obtain:

for $\varepsilon = 0$ a solution will be $x_1 = 0,25$; $x_2 = 0,75$. For $\varepsilon = 0,1$ a solution is: $x_1 = 0,247$; $x_2 = 0,81$. Thus $\dfrac{\Delta x_1}{x_1} = \dfrac{0,247 - 0,25}{0,25} = -0,12 = -1,2\varepsilon$; $\dfrac{\Delta x_2}{x_2} = \dfrac{0,8 - 0,75}{0,75} = 0,08 = 0,8\varepsilon$.

This calculation affirms that when $m = 4$ the influence of coefficients variations on a value of solutions variations is really less than when $m = 2$. Besides this calculation shows in what way a rough approximation secures the calculation by means of "numbers of condition" in relation to real changes in solutions during coefficients variations of a system.

Example 6. Let us consider a system of three equations:

$$\begin{array}{l} x_1 + 2x_2 + 3x_3 = 1 \\ 2x_1 + x_2 + 3x_3 = 2 \\ 3x_1 + x_2 + 2x_3 = 3 \end{array} \tag{66}$$

with solution: $x_1 = 1$; $x_2 = 0$; $x_3 = 0$.

For a system (66) we have:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 1 & 2 \end{pmatrix};\ detA = 6;\ A^{-1} = \frac{1}{6}\begin{pmatrix} -1 & -1 & 3 \\ 2 & -7 & 3 \\ -1 & 5 & -3 \end{pmatrix}$$

$$||A|| = 6,48;\ ||A^{-1}|| = 1,732;\ ||A|| \cdot ||A^{-1}|| = 11,225$$

Let us add to the right side, to a vector B, a variation δB. We shall increase each component of vector B by 0,1 from an initial value. Here we shall have:

$$\frac{||\Delta B||}{||B||} = 0,1$$

According to the main formula (35) the following inequality must be fulfilled:

$$\frac{||\Delta X||}{||X||} \leq 11,225;\ \frac{||\Delta B||}{||B||} = 1,1125$$

At the same time by directly solving a system (66) while taking into account variations in the right side, i.e. while solving a system

$$\begin{array}{l} x_1 + 2x_2 + 3x_3 = 1,1 \\ 2x_1 + x_2 + 3x_3 = 2,2 \\ 3x_1 + x_2 + 2x_3 = 3,3 \end{array} \tag{67}$$

we shall obtain a solution: $x_1 = 1,1;\ x_2 = x_3 = 0$ and thus

$$\frac{||\Delta X||}{||X||} = 0,1 \tag{68}$$

Therefore an estimate by a "number of condition"turned out in this case to be by 11,225 times more than a true error.

§5. The calculation of solutions errors during variations of a right side.

The most simple error in solving equations system $\Delta X = B$ is calculated in such a way when not only coefficients of the right side are changed, coefficients of vector B and coefficients of matrix A remain unchanged or their change is so small that we can ignore them. Such cases can be often met in practice. In construction mechanics it is such a case when a construction mechanics it is such a case when a construction remains unchanged but loading are changed while computing electric chains – it is such a case when electromotive forces change but resistance does not change etc.

On an example of a system of the third order it is convenient to explain the methodics of computing solutions errors.

Example №7. Let us consider system:

$$\begin{array}{l} x_1 + 2x_2 + 3x_3 = 7 \\ 2x_1 + x_2 + 3x_3 = 8 \\ 3x_1 + x_2 + 2x_3 = 9 \end{array} \tag{69}$$

and let us estimate a value of solutions errors during variations of a right side when system (69) becomes:

$$\begin{array}{l} x_1 + 2x_2 + 3x_3 = 7(1 + \delta_1) \\ 2x_1 + x_2 + 3x_3 = 8(1 + \delta_2) \\ 3x_1 + x_2 + 2x_3 = 9(1 + \delta_3) \end{array} \tag{70}$$

A value and numbers signs δ_1; δ_2; δ_3 can be any. It is clear that a value of solutions errors depends on a combination of signs is coefficients variations, on combinations of values signs δ_1; δ_2; δ_3. For a system of the third order (70) a number of a possible combination of signs is equal to $2^3 = 8$, for a system of the nth order a number of signs combination is equal to 2^n and it quickly increases with the increases of n.

A number of necessary computations can be greatly lessened if we apply Cramer formula (50) reasonably.

For a system (69) a determinant D is equal to

$$D = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 1 & 2 \end{vmatrix} = 6 \tag{71}$$

a determinant D_1 is equal to:

$$D_1 = \begin{vmatrix} 7 & 2 & 3 \\ 8 & 1 & 3 \\ 9 & 1 & 2 \end{vmatrix} = 7 \begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} - 8 \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} + 9 \begin{vmatrix} 2 & 3 \\ 1 & 3 \end{vmatrix} = 7 \cdot (-1) - 8 \cdot 1 + 9 \cdot 3 = 12 \tag{72}$$

(we have decomposed a determinant D_1 by the first column) and therefore $x_1 = \dfrac{D_1}{D} = 2$

For system (70) a determinant D does not depend on δ_i and it remains the same: $D = 6$ but determinant D_1 depends on δ_i and is equal to:

$$D_{1\delta} = \begin{vmatrix} 7 + 7\delta_1 \\ 8 + 8\delta_2 \\ 9 + 9\delta_3 \end{vmatrix} = 12 + (-7\delta_1 - 9\delta_2 + 27\delta_3) \tag{73}$$

From formula (73) it is at once seen that the change of determinant D_1 (and thus – a change of solution x_1) will be the largest. If δ_3 is positive and δ_1 and δ_2 are negative. If δ_1; δ_2; δ_3 are equal to each other by an absolute value, i.e. $|\delta_1| = |\delta_2| = |\delta_3| = \delta$ then $D_{1\delta} = 12 + 42\delta$ and thus

$$x_{1\delta} = \frac{D_{1\delta}}{D} = 2 + 7\delta \tag{74}$$

If variations δ_1; δ_2; δ_3 have inverse signs, i.e. if δ_1 and δ_2 – positive and δ_3 – negative then in this case if $|\delta_i| = \delta$ will be $D_{1\delta} = 12 - 42\delta$ and thus

$$x_{1\delta} = \frac{D_{1\delta}}{D} = 2 - 7\delta$$

Similarly determinants $D_{2\delta}$ and $D_{3\delta}$ are computed and by them – solutions $x_{2\delta}$ and $x_{3\delta}$. We obtain:

$$D_{2\delta} = \begin{vmatrix} 1 & (7 + 7\delta_1) & 3 \\ 2 & (8 + 8\delta_2) & 3 \\ 3 & (9 + 9\delta_3) & 2 \end{vmatrix} = 6 + 35\delta_1 - 56\delta_2 + 27\delta_3 \tag{75}$$

and thus: $x_{2\delta} = 1 + (\frac{35}{6}\delta_1 - \frac{56}{6}\delta_2 + \frac{27}{6}\delta_3)$

Here variations in solution x_2 will be the largest if δ_1 and δ_3 are positive and δ_2 – negative. If δ_1; δ_2 and δ_3 are equal to each other by an absolute value, i.e. $|\delta_1| = |\delta_2| = |\delta|$, then $x_{2\delta} = 1 + 18\delta$. If signs of variations of δ are inverse, i.e. if δ_1 and δ_3 are negative and δ_2 is positive we shall have $x_{2\delta} = 1 - 18\delta$.

Analogically we obtain:

$$D_{3\delta} = \begin{vmatrix} 1 & 3 & (7 + 7\delta_1) \\ 2 & 3 & (8 + 8\delta_2) \\ 3 & 2 & (9 + 9\delta_3) \end{vmatrix} = 6 - 7\delta_1 + 40\delta_2 - 27\delta_3 \tag{76}$$

and thus

$$x_{3\delta} = \frac{D_{3\delta}}{D} = 1 - \frac{7}{6}\delta_1 + \frac{40}{6}\delta_2 - \frac{27}{6}\delta_3 \tag{77}$$

if $|\delta_1| = |\delta_2| = |\delta_3| = \delta$ then $x_{3\delta} = 1 + 12,33\delta$ but if signs δ_i be inverse we shall have $x_{3\delta} = 1 - 12,33\delta$.

This example at once shows an order of computing solutions errors of a system of equations $AX = B$ of an arbitrary order when a vector–column of the right side is of the form:

$$B_\delta = \begin{pmatrix} b_1 + \delta_1 b_1 \\ b_2 + \delta_2 b_2 \\ \dots\dots\dots\dots \\ b_n + \delta_n b_n \end{pmatrix} \tag{78}$$

In this case in order to calculate, for example, values x_1 we use Cramer formulas and decompose a determinant D_1 by elements of the first column, i.e. in this case – a column (78). We obtain

$$x_1 = \frac{D_{1\delta}}{D} = \frac{(b_1 + \delta_1 b_1)M_1 - (b_2 + \delta_2 b_2) + ... + (-1)^n (b_n + \delta_n b_n)M_n}{D} \tag{79}$$

where M_1; M_2; ...; M_n – minors that correspond to elements of the first column. From (79) it follows that:

$$x_1 = x_{1,\delta=0} + \Delta x_1 \tag{80}$$

where $x_{\delta=0}$ – value x_1; corresponding to $\delta = 0$ and computed with nominal values of coefficients b_i, and Δx_1 – an increase x_1, that has occurred due to variations of these coefficients. From (79) it follows that

$$\Delta x_1 = \frac{1}{D}\sum_{i=1}^{n}(-1)^i b_i M_i \delta_1 = \frac{1}{D}\sum_{i=1}^{n} b_i A_i \delta_i \tag{81}$$

From formula (81) at once it is seen that variations of solution x_1 will be the largest in case when signs of variations $\delta_i b_i$ coincide with signs of a product $(-1)^i M_i$. Note that variations of solutions linearely depend on variations of the right side. This allows us – if, for example, variation of solution x_i (if $|\delta_i| = 0,01$) easily to compute a value of solution variation for any other value $|\delta_i|$.

By applying formula (81) we ca, for example, casily compute the largest possible change of solution x_i in system of equations (66) from example №6 for a case when $|\delta| = 0,1|$. Since for system (66) we have

$$M_1 = \begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} = -1;\ M_2 = \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = 1;\ M_3 = \begin{vmatrix} 2 & 3 \\ 1 & 3 \end{vmatrix} = 3 \tag{82}$$

then the largest change in x_1 will be if $\delta_i = -0,1$; $\delta_2 = -0,1$; $\delta_3 = 0,1$.
And here we have

$$\Delta x_i = \frac{1}{6}(0,1 + 0,2 + 0,9) = 0,2 \tag{83}$$

by directly computing a determinant D_1 while coefficients of the right side have changed and a new value x_1 has occurred:

$$D_{1\delta} = \begin{vmatrix} 1(1-0,1) & 2 & 3 \\ 2(1-0,1) & 1 & 3 \\ 3(1+0,1) & 1 & 2 \end{vmatrix} = 7,2 \tag{84}$$

$$x_1 = \frac{D_{1\delta}}{D} = \frac{7,2}{6} = 1,2 \tag{85}$$

we see that the change of x_1 really is equal to 20% from an initial value. If all variations of the right side have similar signs then a change of x_1 as it has already been calculated in §7 will be by two times less if there are the same $|\delta_i|$.

We see that if only coefficients of the right side in a system of equations $AX = B$ change then changes of a solution x_i and if there are maximally possible changes it is not

difficult to compute. And what is the most important – they are computed exactly (if δ_i are known). Much more difficult it is to compute changes of solutions during variations of coefficients in matrix A. Here a quite new approach is necessary about which we shall speak in the next section

§6. A new approach to the problem of estimating errors: an approach by means of differentials of determinants and by "table of signs".

In this section we shall pass to the solution of the most general (and the most complex) problem of estimating errors of each component of a vector of solutions X in system $AX = B$ during possible variations of coefficients in a matrix A and vector of the right side B.

In difference to an approach by "numbers of condition"$||A|| \cdot ||A^{-1}||$ we shall apply Cramer formulas:

$$x_i = \frac{D_i}{D} \tag{86}$$

and we shall start from the estimate of possible variations of determinants D and D_i that took place due to variations of their coefficients.

Let us consider a determinant of matrix A:

$$D = \det A = \begin{vmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{vmatrix} \tag{87}$$

and let us take into account that about each of coefficients only an interval (18) is known in the interior of which are its possible values. Taking into account inequalities (18) a determinant (87) can be written in the form:

$$D = \det A = \begin{vmatrix} a_{11}(1+\varepsilon_{11}) & a_{12}(1+\varepsilon_{12}) & ... & a_{1n}(1+\varepsilon_{1n}) \\ a_{21}(1+\varepsilon_{21}) & a_{22}(1+\varepsilon_{22}) & ... & a_{2n}(1+\varepsilon_{2n}) \\ ... & ... & ... & ... \\ a_{n1}(1+\varepsilon_{n1}) & a_{n2}(1+\varepsilon_{n2}) & ... & a_{nn}(1+\varepsilon_{nn}) \end{vmatrix} \tag{88}$$

and consider it as a function n^2 of variables ε_{ij} that are subjected to (88). It occurs due to variations of its coefficients. First of all let us compute a partial derivative of a determinant (88), i.e. a value

$$\frac{\partial D}{\partial \varepsilon_{ij}} \tag{89}$$

and then – a total differential, i.e. the main linear part of a determinant change.

If we decompose determinant (88) by minors of such a line in which enters a member $a_{ij} + \varepsilon_{ij} a_{ij}$ we shall see that the only member of the decomposition that depends on ε_{ij} is a member

$$\varepsilon_{ij} a_{ij} A_{ij} \tag{90}$$

where A_{ij} – an algebraic addition of element a_{ij}, i.e. a determinant of an order $n-1$ which is formed from a determinant A by excluding the ith line and jth column and by multiplying by $(-1)^{i+j}$. Hence it follows that a partial derivative (89) will be equal

to a product $A_{ij}a_{ij}$ and a total differential – i.e. the main linear part of a determinant $D = \det A$ will be equal to

$$\Delta_{lin} = \Sigma_{i=1;j=1}^{n} a_{ij} A_{ij} \Delta\varepsilon_{ij} \tag{91}$$

where $\Delta\varepsilon_{ij}$ – a change of a variable ε_{ij} that does not go out of the limits given by inequalities (18).

From formula (91) it follows that the largest change of determinant (88) in the direction of an increase (in a linear approach) will be in a case when for all i and j we shall have $|\varepsilon_{ij}| = \varepsilon_{ijmax}$ and a sign of ε_{ij} coincides with a sign of a product of element a_{ij} by its algebraic addition A_{ij}.

For a convenient computation of determinant variations it is useful to form the so called "table of signs". It can be also called "a matrix of signs- whose elements are signs – "plus-, "minus- and "zero". "A table of signs"is formed according to the following rule. Each its element ("plus- or "minus") coincides with a sign of a product $A_{ij}a_{ij}$ and is equal to zero if this product is equal to zero. These "table of signs- have been formed for the first time in [6].

Example №8

For a determinant

$$\begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} \tag{92}$$

algebraic additions are equal to $A_{11} = a$; $A_{12} = -1$; $A_{21} = -1$; A_{22} and thus "a table of signs"is of the form:

$$\begin{vmatrix} + & - \\ - & + \end{vmatrix} \tag{93}$$

Note that for any determinant of the second order in which all elements are positive it will always be that: $A_{11} > 0$; $A_{12} < 0$; $A_{21} < 0$; $A_{22} > 0$ and therefore a table of signs always has a form (93).

Example №9

For a determinant

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{vmatrix} = 8 \tag{94}$$

we have:

$$A_{11} = \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = -3; \; - \begin{vmatrix} 4 & 2 \\ 3 & 5 \end{vmatrix} = -14 \tag{95}$$

and similarly

$$\left.\begin{array}{c} A_{13} = 13;\ A_{21} = 2;\ A_{22} = -1 \\ A_{23} = 2;\ A_{31} = 1;\ A_{32} = 10;\ A_{33} = -7 \end{array}\right\} \tag{96}$$

a table of signs in determinant (94) follows from inequalities (95)–(96) and is of the form:

$$\begin{vmatrix} - & - & + \\ + & - & + \\ + & + & - \end{vmatrix} \tag{97}$$

If algebraic additions have been calculated then it is not difficult to compute the largest (in a linear approach) increase of a determinant as well. It is equal to:

$$\Delta_{lin\,max} = \sum_{i=1;j=1}^{n} |a_{ij} A_{ij} \Delta\varepsilon_{ij}| \tag{98}$$

If for all i and j there is $|\Delta\varepsilon_{ij}| = \varepsilon_0$ then a formula (98) can be simplified:

$$\Delta_{lin\,max} = \sum_{i=1;j=1}^{n} |a_{ij} A_{ij}| \varepsilon_0 \tag{99}$$

so for a determinant (92) we shall have:

$$\Delta_{lin\,max} = (2 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 2)\varepsilon_0 = 6\varepsilon_0 \tag{100}$$

and for a determinant (94):

$$\Delta_{lin\,max} = 1 \cdot 3 + 2 \cdot 14 + 3 \cdot 13 + 4 \cdot 2 + 1 \cdot 1 + 2 \cdot 2 + 3 \cdot 1 + 4 \cdot 10 + 5 \cdot 7)\varepsilon_0 = 161\varepsilon_0 \tag{101}$$

A relative increase of a determinant (94) is:

$$\frac{\Delta_{lin\,max}}{\det A} = \frac{161}{8}\varepsilon_0 = 20,125\varepsilon_0 \tag{102}$$

or more than by 20 times more than an increase of each of coefficients. The largest possible decrease of a determinant (in a linear approach) can also be easily computed. It will be the largest if signs in $\Delta\varepsilon_{ij}$ be inverse to signs of products $a_{ij}A_{ij}$. Thus to the largest decrease of a determinant, to the largest decrease of a determinant to a the largest change in the direction of decrease will lead the so called "an inverse table of signs- whose signs are inverse to signs of a main table. For determinant (92) "an inverse table of signs- is of the form:

$$\begin{vmatrix} - & + \\ + & - \end{vmatrix} \tag{103}$$

and the same form it will have for all determinants of the second order with positive elements. And for determinant (94) "an inverse table of signs"will be:

$$\begin{vmatrix} + & + & - \\ - & + & - \\ - & - & + \end{vmatrix} \tag{104}$$

The value of the largest (in a linear approach) change of a determinant and in the direction of increase and decrease is similar and it is computed according to formula (99).

"Tables of signs" can be applied not only for computing differentials of determinants but also for the computation of determinants in a linear approximation. Besides they can be applied for exact estimates of determinants changes that occur due to the change of their coefficients. In order to pass to exact estimates it is necessary to take into account that in practical problems.

1. Algebraic additions equal to zero occur very seldom.

2. During variations of coefficients ε_{ij} that are small in comparison with 1 signs in algebraic additions change rarely.

Therefore we shall apply two approaches to the estimate of possible errors of solutions.

The first approach is: we suppose that for our system that we are examining the following conditions are fulfilled:

1. Not any of algebraic additions is equal to zero.

2. Not any of algebraic additions does not change its sign for variations of coefficients of an examined determinant when numbers ε_{ij} are small in comparison with 1.

While fulfilling these conditions we can easily calculate (as it will be shown later) a variation of a determinant not only in a linear approach but exactly as well – for any values of variations of coefficients ε_{ij} that interest us.

If any of these two conditions is not fulfilled then an examined determinant is named a special one. And in order to compute its variations it is necessary to apply more complex algorithm which will be given in next sections.

If these conditions are fulfilled then for an exact computation the largest possible variation of a determinant it is sufficient to compute a determinant with changed coefficients $a_{ij} \pm \varepsilon_{ij} a_{ij}$. And here signs ε_{ij} are chosen in accordance with "table of signs"which has been earlier computed – of the type of table (97).

Example №10. Let us again consider a simple determinant (94) for which the most unfavourable combination of signs ε_{ij} corresponds to "table of signs"(97) as has been already shown. If for all i and j $|\varepsilon_{ij}| = 0,01$ then if there is the most unfavourable combination of signs ε_{ij} determinant (94) will become:

$$\begin{vmatrix} 0,99 & 1,98 & 3,03 \\ 4,04 & 0,99 & 2,02 \\ 3,03 & 4,04 & 4,95 \end{vmatrix} = 9,6604$$

Thus variations of elements in determinant (94) if each value is $\pm 1\%$ and if there is the most unfavourable combination of their signs can lead to a change of determinant value by $\Delta_+ = 1,6604$ or by $+20,76\%$. We can also compute the largest possible change of

determinant value to a negative direction. For this it is necessary to change signs "plus"by "minus"and "minus"by "plus"in a "table of signs"(97). Then – if for all i and j we have $|\varepsilon_{ij}| = 0,01$ – determinant (94) will become:

$$\begin{vmatrix} 1,01 & 2,02 & 2,97 \\ 3,96 & 1,01 & 1,98 \\ 2,97 & 3,96 & 5,05 \end{vmatrix} = 6,3804$$

and in this case the determinant will decrease by $\Delta_- = -1,6196$ or by -20,24%. A small diversion between Δ_+ and Δ_- can be explained by a nonlinear character of a dependence of determinant value on δ_{ij}.

Thus possible changes of determinant (94) (we shall denote them by Δ_{det}) are during variations of its elements by $\pm 0,01$ are subjected to the inequality:

$$-1,6196 \leq \Delta_{det} \leq 1,66 \tag{105}$$

or in relative units:

$$-0,2024 \leq \frac{\Delta_{det}}{det} \leq 0,2076$$

and here estimate of value (105) is exact – i.e. there exist such combinations of elements variations at which inequalities (105) turn into exact equalities.

So in this example the largest variation of determinant (94) turns out to be by 20,76 times more than the largest variation of each of its elements.

A given mathodics can be used when some of elements in a determinant are exactly known (their variations $\varepsilon_{ij}a_{ij}$ are equal to zero) or when signs of variations of some elements are known. In this case only a part of "table of signs- is used. For example, let it be known about a determinant (94) that only variations of elements of its first line are different from zero also for the first line $|\varepsilon_{ij}| = 0,01$ and for other lines $\varepsilon_{ij} = 0$. In this case let us apply only the first line of "table of signs"(97). If there is the most infavourable combination of signs of variations in elements of the first line the determinant becomes:

$$\begin{vmatrix} 0,99 & 1,98 & 3,03 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{vmatrix} = 8,7$$

i.e. in this case variations of three elements in a determinant change its value by 0,7 or by 8,75% from its rating value. This methodics can be applied during the investigation of an influencs of not only relative variations of determinant elements but also "absolute"variations when after variations element a_{ij} turns into element $a_{ij} + \varepsilon_{ij}$. And here a sign in number ε_{ij} can be any.

So a determinant of a general form (87) will turn into the following determinant after such variations:

$$\begin{vmatrix} a_{11}+\varepsilon_{11} & a_{12}+\varepsilon_{12} & ... & a_{1n}+\varepsilon_{1n} \\ a_{21}+\varepsilon_{21} & a_{22}+\varepsilon_{22} & ... & a_{2n}+\varepsilon_{2n} \\ ... & ... & ... & ... \\ a_{n1}+\varepsilon_{n1} & a_{n2}+\varepsilon_{n2} & ... & a_{nn}+\varepsilon_{nn} \end{vmatrix} \tag{106}$$

and determinant (94) will turn into

$$\begin{vmatrix} 1+\varepsilon_{11} & 2+\varepsilon_{12} & 3+\varepsilon_{13} \\ 4+\varepsilon_{21} & 1+\varepsilon_{22} & 2+\varepsilon_{23} \\ 3+\varepsilon_{31} & 4+\varepsilon_{32} & 5+\varepsilon_{33} \end{vmatrix} \tag{107}$$

For "absolute- variations as well as for relative variations its own "table of signs- is formed – for the most unfavorable combination of signs of coefficients ε_{ij} variations that lead to the largest variation of a determinant.

We, as earlier, decompose determinant (106) by minors of a line containing element $a_{ij}+\varepsilon_{ij}$ and then if we compute, as before, a partial derivative and a total differential. We see that the main linear part of the increase of a determinant (107) is its total differential. In this case it is

$$\Delta_{lin} = \sum_{i=1;j=1}^{n} A_{ij}\Delta\varepsilon_{ij} \tag{108}$$

and it reminds us (as expected) of formula (91). It differs from it only by an absence of a multiplier a_{ij}. If all ε_{ij} by an absolute value are equal to the same number ε_{ij} then formula (108) will become

$$\Delta_{lin\,max} = \sum_{i=1;j=1}^{n} = |A_{ij}|\varepsilon_0 \tag{109}$$

(it is an analogue of the formula (99)).

For a determinant (107) "a table of signs- remains as (97). If for all ε_{ij} we shall have $|\varepsilon_{ij}| = 0,01$ then if there is the most unfavorable combination of signs a determinant (107) will become:

$$\begin{vmatrix} 0,99 & 1,99 & 3,01 \\ 4,01 & 0,99 & 2,01 \\ 3,01 & 4,01 & 4,99 \end{vmatrix} = 8,5628$$

i.e. a value of a determinant will increase by 0,5628 or by 7,06%.

Here it is not difficult to compute the largest deviation of a determinant to a negative direction. If we change "pluses- into "minuses- in "table of signs- (97) we shall have the following "table of signs":

$$\begin{vmatrix} + & + & - \\ - & + & - \\ - & - & + \end{vmatrix}$$

and – a determinant:

$$\begin{vmatrix} 1,01 & 2,01 & 2,99 \\ 3,99 & 1,01 & 1,99 \\ 2,99 & 3,99 & 5,01 \end{vmatrix} = 7,4428$$

i.e. a determinant (94) will decrease by 0,5572 or by 6,96%. For a determinant (107) when $|\varepsilon_{ij}| \leq 0,01$ the following inequalities are satisfied:

$$det_{nom} - 0,5572 \leq det_{ver} \leq det_{nom} + 0,5628$$

It is important to note that in the theory of determinants and in the theory of linear algebraic equations as well we can consider variations that are different from zero coefficients by a unique methodics . Besides by this methodics we can investigate variations (absolute variations) whose rating value is equal to zero, thus – to examine "variations of a zero". Recollect that in a theory of differential equations an investigation of "variations of a zero- in higher coefficients requires a special mathematical apparatus that has been developed in a theory of "singularly perturbing equations"(also see [5], pp.67–68 and [11], pp.18–19).

For systems of algebraic equations everything is more simple although here "variations of zero- can also lead to different results. Sometimes a small "variation of a zero- can lead to great changes in solutions.

Let us examine a simple equations system:

$$\begin{vmatrix} a_{11}x_1 + a_{21}x_2 = 1 \\ a_{21}x_1 + a_{22}x_2 = 1 \end{vmatrix} \tag{110}$$

when $a_{11} = 1$; $a_{12} = 0$; $a_{21} = 0$; $a_{22} = 0$. In this case $x_1 = 1$ and x_2 does not exist. Let a zero coefficient a_{22} undergo a variation and became $a_{22} = \varepsilon$ where $0 \leq \varepsilon \leq 0,01$. In this case

$$D = \begin{vmatrix} 1 & 0 \\ 0 & \varepsilon \end{vmatrix} = \varepsilon;\ D_1 = \begin{vmatrix} 1 & 0 \\ 1 & \varepsilon \end{vmatrix};\ D_2 = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1;\ x_1 = \frac{\varepsilon}{\varepsilon} = 1;\ x_2 = \frac{1}{\varepsilon} \tag{111}$$

Thus "variations of a zero- have led in this case to the fact that determinants D and D_1 have changed by small values (from a value zero up to ε). A solution x_1 has not changed but now solution x_2 is now included in the limits of from $x_2 = \dfrac{1}{\varepsilon}$ up to $x_2 \to +\infty$, or in other words, a solution x_2 is now included into an unlimited open interval $(\dfrac{1}{\varepsilon} + \infty)$.

Oftener small changes of a zero element of a determinant do not lead to essential changes of its value. So, for example, a determinant

$$\begin{vmatrix} 2 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 1 \tag{112}$$

during variations of its zero element turns a into a determinant

$$\begin{vmatrix} 2 & \varepsilon & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 1 \end{vmatrix} = 1 - \varepsilon \tag{113}$$

and a variation of a determinant value if $\varepsilon \to 0$ is infinitely small.

If an estimate of determinants variations that enter into Cramer formulas have been carried out it is already not difficult to estimate possible variations of each of solutions x_1; x_2; ...; x_n

The appearance of a methodics of an exact estimate of a possible variation of each of solutions in system $AX = B$ that are due to coefficients variations, of an exact estimate of an interval in which a solution can appear due to variations of coefficients is a large step forward. Now as we surely possess estimates of exactness in coefficients, estimates of their possible variations – can be exactly estimated – in which systems of equations solutions are reliable, in which they are apriori not reliable. They can be unreliable during any calculation methods.

Variations of coefficients, errors in their definition is certainly the only cause of errors in the results of computation. There exist some errors in computation methods such as errors from rounding off in intermediate results, errors from a finite number of iterations in iteration methods etc.

Often the most attention is paid to the decrease of these errors. But nobody paid any attention to the fact that an error that is due to parameters variations can not be removed during any modernizing of computation methods. If, for example, when $|\varepsilon_{ij}| \leq 0,01$ in determinant (94) its possible error only due to coefficients variations is subjected to inequalities (105) then it is not possible to improve these inequalities during any modernizing computation methods.

If,for example, due to coefficients variations an interval in the interior of which there is a solution x_1 of system $AX = B$ is equal to $\pm 0,1$ it is not rational to compute a solution with essentially larger exactness.

Note that during these same estimates with relative (or absolute) variations of coefficients $|\varepsilon_{ij}| \leq \varepsilon_{ij}$ a variation of a determinant value increases with the growth of its order.

This circumstance can be easily explained on an example of the so called "triangle determinants"

$$\begin{vmatrix} a_{11} & 0 & ... & 0 \\ a_{21} & a_{22} & ... & 0 \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{vmatrix}$$

in which all elements that lie higher than a main diagonal (or vice verse – lower than a main diagonal) are equal to zero. Such determinants are equal to a product of their diagonal elements. If we ignore degrees that are higher than the first of a number ε_{ij} that us small in comparison with 1 we obtain the following estimate:

$$det_{|\varepsilon_{ij}|=\varepsilon_0} - det_{\varepsilon_{ij}=0} = (a_{11}+\varepsilon_0 a_{11})(a_{22}+\varepsilon_0 a_{22})\cdot ...\cdot (a_{nn}+\varepsilon_0 a_{nn}) - a_1 a_{22}...a_{nn} > n\varepsilon_0 det_{\varepsilon_{ij}=0}$$

which increases with the growth of an order in a determinant n.

An interesting question arises. The largest possible variation of a determinant value and with it – variation of solutions x_i in a system of equations $AX = B$ if there is the largest unfavourable combination of signs in variations of determinant elements increase with the growth of n. But possibility of realizing just the largest unfavourable combination of signs in elements variations of a determinant is small and it is equal to $\frac{1}{2^{n^2}}$ and it quickly

decreases with the growth of n. It is cvident that it is necessary to take into account a value of determinant variations with not only one and the smallest probable combination of signs in variations of elements but during all such combinations of signs that lead to variations of a value in a determinant that is near to maximal values. This question will be later discussed in §7.

§7.Results of a numerical experiment.

Besides computing the largest variations of determinants due to variations of their elements it is interesting to estimate with what probability one or another variation can appear.

In order to understand what probabilites there are of different variations it is convenient to use a numerical experiment that has carried out M.V.Voloshin. A determinant of the third order contains nine elements. Therefore for it there are possible $2^9 = 512$ combinations of positive and negative variations $\pm\varepsilon_{ij}$. By posing $|\varepsilon_{ij}| = 0,01$ all 512 determinants of the form (114) have been computed for $a_{11} = 1$; $a_{12} = 2$; a_{13}; $a_{21} = 4$; $a_{22} = 1$; $a_{23} = 2$; $a_{31} = 3$; $a_{32} = 4$; $a_{33} = 5$ and all possible combinations of signs in numbers ε_{ij}. If $\varepsilon_{ij} = 0$ we shall have $det_{nom} = 8$. For this determinant the most unfavourable will be the following combination of variations signs: $\varepsilon_{11} = -0,01$; $\varepsilon_{12} = -0,01$; $\varepsilon_{13} = +0,01$; $\varepsilon_{21} = +0,01$; $\varepsilon_{22} = -0,01$; $\varepsilon_{23} = +0,01$; $\varepsilon_{31} = +0,01$; $\varepsilon_{32} = +0,01$; $\varepsilon_{33} = -0,01$; . During this combination of signs in elements variations an examined determinant will become:

$$det = \begin{vmatrix} 0,99 & 1,98 & 3,03 \\ 4,04 & 0,99 & 2,02 \\ 3,03 & 4,04 & 4,95 \end{vmatrix} \tag{114}$$

If we denote by + ((conditionally) elements with a positive variation and elements with a negative variation by − then a determinant can be conditionally denoted in this way:

$$\begin{vmatrix} - & - & + \\ + & - & + \\ + & + & - \end{vmatrix} \tag{115}$$

Such a picture obviously shows what combination of signs in variations in thus case is the most dangerous.

A direct computation while computing between themselves all 512 determinants assertains that the worst combination of variations signs in elements really corresponds to determinant (114) and the largest value of variations value of an initial determinant is really equal to 1,6604.

The analysis of all 512 computed variations (i.e. all possible variants) shows that 16 from them – i.e. 3,12% lie in the limits of from Δ_{max} up to $0,9\Delta_{max}$.

For a determinant with variations of elements a determinant

$$det_{\varepsilon=0} = \begin{vmatrix} 1 & 2 & 4 \\ 1 & 3 & 4 \\ 2 & 3 & 5 \end{vmatrix} = -3;\ det_{\varepsilon} = \begin{vmatrix} 1+\varepsilon_{11} & 2+2\varepsilon_{12} & 4+4\varepsilon_{13} \\ 1+\varepsilon_{21} & 3+3\varepsilon_{22} & 4+4\varepsilon_{23} \\ 2+2\varepsilon_{31} & 3+3\varepsilon_{32} & 5+5\varepsilon_{33} \end{vmatrix}$$

an analogous numerical experiment has shown that if $|\varepsilon_{ij}| = 0,01$ the largest determinant variation is equal to $|\Delta_{max}| = 0,503$. It really is achieved if there is the following combination of signs of variations in determinant elements:

$$\begin{vmatrix} - & - & + \\ - & + & - \\ + & + & - \end{vmatrix}$$

which earlier was predicted on the base of a theory given in a previous section. Here a direct computations has shown that only in two determinants from 512 their variations are in the limits from Δ_{max} up to $0,9\Delta_{max}$ and in eight determinants their variations are from Δ_{max} up to $0,8\Delta_{max}$.

If we take into account that not always a variation of any element of a determinant reaches its higher (by modulus) value it must be said that a variation of a determinant very rarely reaches values that are close to maximal ones. But if such a rare combination of values and signs in variations of a determinant elements is possible it is necessary to take it into account.

From this circumstance there are important consequences. For example, let efforts in some knot of our objects is equal numerically to a solution x_1 of a system with three equations and each of coefficients in this system is measured with the exactness up to $\pm 0,01$. Let 999 objects be manufactured and they all work well. Sorry, this fact does not guarantee that the 1000-th manufactured object will not break and it will not lead to a wreckage. It can break not at once, not during tests but after some period of exploitation in the course of which parameters of an object and coefficients in its mathematical model undergo inevitable small variations. The cause of a wreckage can turn out to be the following: in previous manufactured 999 objects not once a dangerous combination of values and variations signs was realized but on the 1000-th it was realized.

A guarantee from the most dangerous wreckages and breakages, a guarantee from people's death can only give a good reliable and truthful computation.

For illustration let us give some calculated values of a determinant:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{vmatrix} = 8$$

during variations of its elements by $\dfrac{1}{100}$ from rating values, i.e. if $|\varepsilon_{ij}| = 0,01$. We shall give "tables of signs"and corresponding to these tables values of determinants:

$$\begin{vmatrix} + & + & + \\ + & + & + \\ + & + & + \end{vmatrix} \to 8,242408; \quad \begin{vmatrix} + & + & + \\ + & + & - \\ + & + & + \end{vmatrix} \to 8,1608; \quad \begin{vmatrix} + & + & + \\ + & - & + \\ + & + & + \end{vmatrix} \to 8,32401;$$

$$\begin{vmatrix} + & + & \\ + & - & + \\ + & + & - \end{vmatrix} \to 9,0401; \quad \begin{vmatrix} + & + & + \\ + & - & + \\ - & - & + \end{vmatrix} \to 7,44309; \quad \begin{vmatrix} + & + & + \\ + & + & - \\ + & - & + \end{vmatrix} \to 7,28513;$$

$$\begin{vmatrix} + & + & + \\ - & + & + \\ - & + & + \end{vmatrix} \to 8,01798; \quad \begin{vmatrix} + & + & + \\ - & + & - \\ + & + & + \end{vmatrix} \to 7,99758; \quad \begin{vmatrix} + & + & - \\ + & + & + \\ + & - & + \end{vmatrix} \to 6,65004;$$

$$\begin{vmatrix} + & + & + \\ - & - & - \\ - & - & + \end{vmatrix} \to 7,21928; \quad \begin{vmatrix} + & + & - \\ - & + & + \\ + & + & - \end{vmatrix} \to 8,00081; \quad \begin{vmatrix} + & + & - \\ - & + & - \\ - & - & + \end{vmatrix} \to 6,38039;$$

$$\begin{vmatrix} + & - & + \\ + & - & + \\ + & + & - \end{vmatrix} \to 9,5952; \quad \begin{vmatrix} + & + & - \\ - & - & - \\ - & - & + \end{vmatrix} \to 6,4548; \quad \begin{vmatrix} + & - & + \\ + & - & + \\ - & - & + \end{vmatrix} \to 8,0192;$$

$$\begin{vmatrix} + & - & + \\ - & - & + \\ + & + & - \end{vmatrix} \to 9,4; \quad \begin{vmatrix} + & - & - \\ + & - & + \\ + & - & + \end{vmatrix} \to 7,29927; \quad \begin{vmatrix} + & - & - \\ + & - & + \\ + & - & - \end{vmatrix} \to 7,9992;$$

$$\begin{vmatrix} - & + & - \\ - & + & + \\ - & + & - \end{vmatrix} \to 7,9992; \quad \begin{vmatrix} - & + & - \\ - & + & + \\ - & - & + \end{vmatrix} \to 6,51835; \quad \begin{vmatrix} - & - & + \\ + & - & + \\ + & + & - \end{vmatrix} \to 9,6604;$$

These examples show that a value of a determinant can greatly change if there occur quite small changes in "tables of signs" , i.e. during changes of variations signs of only several elements of a determinant.

It would be very desirable to be able to compute probabilities to obtain variations into intervals from Δ_{max} up to $0,9\Delta_{min}$, from $0,9\Delta_{max}$ up to $0,8\Delta_{max}$ etc. This would greatly facilitate technical computations. Sorry to say, computation methods of these probabilities has not been as get developed. Let us present (in addition) on fig.2 and fig.3 computed dependences of a determinant (94) values by ε_0 if we made a choice of signs variations of their coefficients in correspondence with a table of signs (97) (it's a higher curve) and in accordance with an "inverse table of signs"(104) (the lower curve). We can conclude that up to $\varepsilon_0 = 0,07$ the dependence of the largest and the smallest values of a determinant by ε_0 almost cannot be distanguished from linear ones (fig.2) and only if ε are large (fig.3) there appears nonlinearity.

Note that if absolute and relative values of variations in all elements of a determinant of nth order are equal between themselves and are equal to ε_0 then variations of determinant values in any "table of signs"are polynomials of the nth degree on variable ε_0. This at once follows from the conception of a determinant as a sum of $n!$ products formed (each) from n elements. Since into an each element of determinants of a general type (88) and (106) enters ε_0 in the first degree then each from $n!$ products – and thus their sum as well–will be a polynomial of the nth degree on ε_0. If ε_0 are small in comparison withs for relative variations that correspond to determinant (88) or if ε_0 are small in comparison a_{ij} that correspond to determinant (106) in a variation of a determinant value a member with a_{ij} that correspond to determinant (106) in a variation of a determinant value a member with a_{ij} that correspond to determinant (106) in a variation of a determinant value a member with the first degree ε_0 will dominate and therefore it will approximally proportional to ε_0. This will allow us to easily estimate the value of a determinant variations value when $\varepsilon_0 = 0,001$, $\varepsilon_0 = 0,005$ etc if, for example, we have computed a variation of a determinant when $\varepsilon_0 = 0,01$.

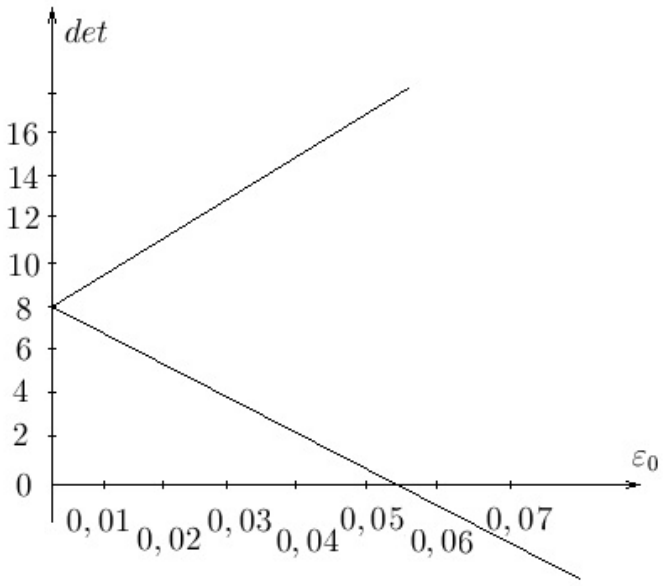

fig. 2

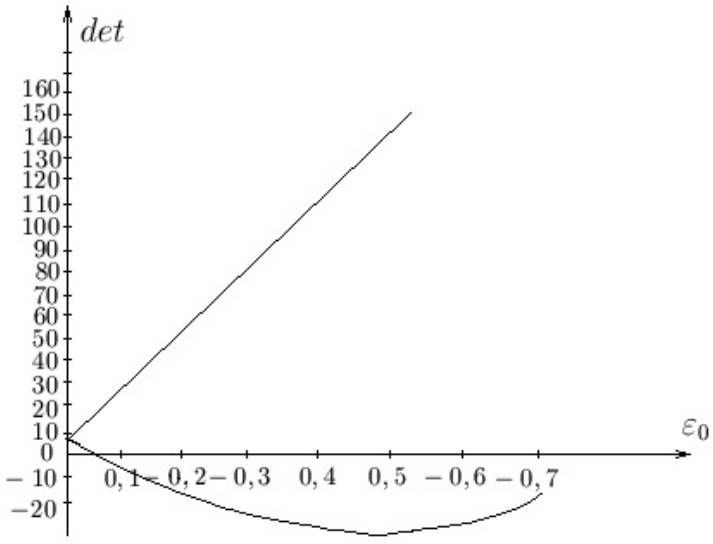

fig. 3

Note that fig.3 shows that a dependence of a determinant variation on ε can be not only nonlinear but not monotonous as well.

As to "table of signs"for determinants of higher orders they can be rather odd. So for a determinant of the fourth order it can be:

$$\begin{vmatrix} 1,8 & -3,8 & 0,7 & -3,7 \\ 0,7 & 2,1 & -2,6 & -2,8 \\ 7,3 & 8,1 & 1,7 & -4,9 \\ 1,9 & -4,3 & -4,9 & -4,7 \end{vmatrix} = 616,9496 \tag{116}$$

it is of a form:

$$\begin{vmatrix} - & + & + & + \\ - & + & - & + \\ + & + & - & - \\ + & + & + & - \end{vmatrix}$$

If for all i and j we shall have $|\varepsilon_{ij} = 0,01$ then a determinant (116) will turn into the following determinant if there is the most unfavourable combination of signs in variations of its elements:

$$\begin{vmatrix} 1,8(1-0,01) & -3,8(1+0,01) & 0,7(1+0,01) & -3,7(1+0,01) \\ 0,7(1-0,01) & 2,1(1+0,01) & -2,6(1-0,01) & -2,8(1+0,01) \\ 7,3(1+0,01) & 8,1(1+0,01) & 1,7(1-0,01) & -4,9(1-0,01) \\ 1,9(1+0,01) & -4,3(1+0,01) & -4,9(1+0,01) & -4,7(1-0,01) \end{vmatrix} = 666,9333$$

whose value is equal to 108,1% from a value of a determinant if $\varepsilon_{ij} = 0$

§8. Applications in practice. How to find unreliable and dangerous objects by means of their mathematical models.

We have obtained computation methods of the largest possible variations of determinants in the direction of the increase and decrease. They can be applied for finding unreliable and dangerous objects on the basis of investigating their mathematical models. Surely, for unreliable and dangerous objects that are able to essentially change their properties and thus lead to wreckaging situation. These objects by all means will be met in the course of exploitation.

As we have shown before the first step in solving a problem of a dependence in coefficients variations on properties of solutions is an estimate of possible values of coefficients variations. This is a purely engineering problem since a possible value of variations wholly depends on properties of some certain object of investigation. Therefore we shall not examine this problem. But we shall give examples of computing for values $\varepsilon_{ij} = \pm 0,01$ and $\delta_i = \pm 0,01$ as we know it is not difficult to compute for any other values of ε_{ij} and δ_i.

The most unreliable are objects in whose mathematical models in the form of a system $AX = B$ a determinant of matrix A can turn into zero in the course of normal exploitation during variations of their coefficients.

The main step in reliable discovery of dangerous systems is the forming of an "inverse table signs"for matrix A. If this "table of signs"is formed it is not difficult to calculate at which value of ε a determinant of a matrix will turn into zero. For this it is sufficient to calculate $detA$ for a series of values ε.

Let us return to an example in §4 example №3 and let us once more consider a system of equations (49) for which matrix A is equal to:

$$A = \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}, \tag{117}$$

$$detA = \begin{vmatrix} 3 & 2 \\ 1 & 1 \end{vmatrix} = 1 \tag{118}$$

since in this case an inverse matrix is:

$$A^{-1} = \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix} \tag{119}$$

then a "table of signs"for determinant (118) is equal to:

$$\begin{vmatrix} + & - \\ - & + \end{vmatrix} \tag{120}$$

and "an inverse table of signs correspondingly:

$$\begin{vmatrix} - & + \\ + & - \end{vmatrix} \tag{121}$$

Hence it follows that determinant (118) will decrease most quickly in such a case if signs of variations of its coefficients corresponds to "table of signs"(121) and determinant (118) will become:

$$detA_{\varepsilon} = \begin{vmatrix} 3(1-\varepsilon) & 2(1+\varepsilon) \\ 1(1+\varepsilon) & 1(1-\varepsilon) \end{vmatrix} = 1 - 10\varepsilon + \varepsilon^2 \tag{122}$$

and it will turn into zero if

$$\varepsilon_{1,2} = 5 \pm \sqrt{25-1} = 5 \pm 4,89898 \tag{123}$$

i.e. already when $\varepsilon = 0,10102$.

When variations have come up to a critical value $\varepsilon = 0,10102$ solutions x_1 and x_2 quickly increase (in an absolute value). So if when $\varepsilon = 0$ system (49) had solutions $x_1 = 1$; $x_2 = 10$ then when $\varepsilon = 0,1$ we shall have $x_1 = -350$, $x_2 = 440$, i.e. x_1 and x_2 will already have nothing in common with their values if $\varepsilon = 0$.

Value ε at which a determinant of matrix A turns into zero we shall call "a natural boundary of variations".

Another source of unreliability in solutions is a change of a sign in any of components of x_i in a vector of solutions X. According to Cramer formulas a change of sign in x_i occurs first of all due to the change of a sign in determinant D_i.

When we return to system (49) and supposing that variations undergo not only coefficients of matrix A and right sides remain unchanged we shell obtain in this case:

$$D_1 = \begin{vmatrix} 23 & 2(1 \pm \varepsilon) \\ 11 & 1 \pm \varepsilon \end{vmatrix} \tag{124}$$

By putting signs in ε according to an inverse table of signs of a determinant (124) we obtain

$$D_1 = \begin{vmatrix} 23 & 2(1+\varepsilon) \\ 11 & 1-\varepsilon \end{vmatrix} = 1 - 45\varepsilon \tag{125}$$

This means that already when $\varepsilon = 0,0222$ determinant D_1 and with it a solution x_1 changes a sign. Thus solution x_1 is not reliable.

Got a preliminary choosing dangerous systems we can apply a formula for the main linear part of the increase in a determinant – formula (91).

Examples of using this formula will be given in next section.

§9. Analysis of computing one of constructions.

In this section we shall aplly a mathodics of disclosing unreliable and dangerous objects for the analysis of computing one of certain constructions loaded by a concentrated force.

Let us consider an example (example №11) of calculating efforts in one of simple constructions considered in a well-known text–book [12], p.205. There a loaded frame shown on fog.4 where $l_1 = l_2 = l_3 = l$ is observed:

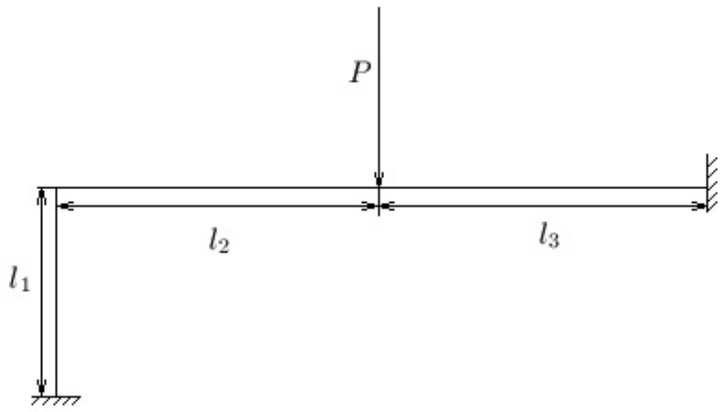

fig. 4

when the ends of a frame are closed up there is a force P that is applied to the middle of a horizontal part. It is necessary to compute a horizontal force x_1 that acts in the lower closure, a vertical force x_2 and a bending moment x_3.

We have chosen this example as it has been given in a known text–book [12] that has been many times republished surely because its author and many teachers that have used this test–book considered the calculation of forces x_1 and x_2 and of moment x_3 – sufficiently trustworthy and demonstrative.

In a text–book [12] on p.205-206 in order to define $x_1;\ x_2;\ x_3$ a system of three equations was formed:

$$\begin{array}{l} 14lx_1 + 12lx_2 + 15x_3 = 3Pl \\ 12lx_1 + 16lx_2 + 12x_3 = 5Pl = Pl \\ \qquad 5lx_1 + 4lx_2 + 6x_3 \end{array} \tag{126}$$

that has the following solution: $x_1 = -\frac{1}{4}P$; $x_2 = \frac{7}{16}$; $x_3 = \frac{1}{12}P$.

The trustworthiness and reliability of this solution in [12] was not checked although all coefficients entering into system (126) can undergo some changes due to inexact accordance of real lengths l_1; l_2; l_3 with projected ones, due to inexact knowledge of an elasticity module on different parts of a frame.

The estimate of reliability of solutions of a system (126) by means of "a number of condition"does not make us cautions for system (126) if we suppose that $P = l$ and $l = 1$ we have.

$$\begin{pmatrix} 14 & 12 & 15 \\ 12 & 16 & 12 \\ 5 & 4 & 6 \end{pmatrix}; \; detA = 48; \; A^{-1} = \frac{1}{48}\begin{pmatrix} 48 & -12 & -96 \\ -12 & 9 & 12 \\ -32 & 4 & 80 \end{pmatrix}. \tag{127}$$

Euclid norm of matrix A is equal to $||A|| = 34,438$ and the same norm of an inverse matrix is $||A^{-1}|| = 2,907$ and a number of condition $||A|| \cdot ||A^{-1}|| = 100,111$ means that system (126) is sufficiently well–conditioned.

Let us make a preliminary deck while using formulas for the variations of a determinant in a linear approximation – formulas (98) and (99) that have been earlier published in [6].

For a system – of equations (126) we suppose that for convenience of further computations $P = 1$ and $l = 1$ it is not difficult to compute algebraic additions for all elements of determinants:

$$D = \begin{pmatrix} 14 & 12 & 15 \\ 12 & 16 & 12 \\ 5 & 4 & 6 \end{pmatrix} = 48 \tag{128}$$

$$D_1 = \begin{pmatrix} 3 & 12 & 15 \\ 5 & 16 & 12 \\ 1 & 4 & 6 \end{pmatrix} = 12 \tag{129}$$

$$D_2 = \begin{pmatrix} 14 & 3 & 15 \\ 12 & 5 & 12 \\ 5 & 1 & 6 \end{pmatrix} = 21 \tag{130}$$

$$D_3 = \begin{pmatrix} 14 & 12 & 3 \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{pmatrix} = 4 \tag{131}$$

and their differentials. If we apply formulas (98) and (99) shall obtain that for a determinant (128) we shall have

$$\Delta_{lin} = 2630\varepsilon_0 \tag{132}$$

that in relation to a rating value of a determinant will be $55\varepsilon_0$.

For a determinant (129)

$$\Delta_{lin} = 1048\varepsilon_0 \tag{133}$$

For a determinant (130)

$$\Delta_{lin} = 984\varepsilon_0 \tag{134}$$

for a determinant (131)

$$\Delta_{lin} = -618\varepsilon_0 \tag{135}$$

or in relation to a nominal value of a determinant $\dfrac{\Delta_{lin}}{D_3} = 154,5\varepsilon_0$.

Now it is at once seen that the most sensible to a value of an unremovable error is a component x_3 of vector in solutions X. Really if we compute according to a linear approximation then when already $\varepsilon_0 = \dfrac{1}{154,5} = 0,0065$ a determinant D_3 and with it a solution x_3 will turn into zero but if $\varepsilon_0 > 0,0065$ a solution x_3 will change a sign (but, for example, if determinant D turns into zero only when $\varepsilon_0 = \dfrac{1}{55} = 0,0182$). Investigation of determinants D, D_1 and D_2 will be continued in §11.

A combination of values of solutions (x_1 and x_2 that is more than x_3) and their errors that has been met in this example just leads us to the conclusion that a norm of relative errors of all solutions (x_1; x_2; x_3) is rather small although in a solution x_3 an error is large. We have already said – that "a number of condition"allows us to estimate only a norm for all components from x_1 up to x_n – of a vector of solutions but not an error in a certain solution x_i. Just because of this a test by "a number of condition"does not allow us (in this example) to disclose a bad condition for a solution x_3.

Passing from a preliminary estimate (in a linear approach) to an exact solution we shall compute "an inverse table of signs"of this form a determinant (131):

$$\begin{vmatrix} + & - & + \\ - & + & - \\ - & + & - \end{vmatrix}$$

and computing in accordance with it values of a determinant (131) during the variation of its elements a determinant (131) turns into:

$$\begin{vmatrix} 14(1+\varepsilon_0) & 12(1-\varepsilon_0) & 3(1+\varepsilon_0) \\ 12(1-\varepsilon_0) & 16(1+\varepsilon_0) & 5(1-\varepsilon_0) \\ 5(1-\varepsilon_0) & 4(1+\varepsilon_0) & 1(1-\varepsilon_0) \end{vmatrix} \tag{136}$$

we obtain (by making precise a linear approximation) a value ε_0 when $\varepsilon_0 = 0,0075$ at which a determinant (131) and with it but a solution x_3 turns into zero but when $\varepsilon_0 = 0,0075$ a solution x_3 will change its sign. The fact that a sign in a solution x_3 can change while there are infinitely small variations of a system coefficients mean that a solution x_3 is very unreliable.

An investigation that has taken place earlier in a work [11] has shown that even if not all twelve coefficients of a system of equations (126) but only three coefficients of a determinant D_3 change and it will become

$$D_3 = \begin{vmatrix} 14(1-\varepsilon) & 12(1+\varepsilon) & 3(1-\varepsilon) \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix} = 4 - 3008\varepsilon \qquad (137)$$

then when even $\varepsilon > 1,3\%$ a determinant D_3 and with it a moment x_3 can essentially change. They can change a sign and become negative. Therefore solution x_3 is not at all reliable and has no practical sense. Inevitable in the course of exploitation small changes of parameters in a construction can lead to its destruction if a sign of a moment x_3 changes.

It is interesting that a text–book [12] only up to 1979 has undergone right editions and had a large circulation. An investigated example with a simple construction shown in fig.4 and given in [12],p.205 have read and solved tens of thousands of undergraduates. Thousands of teachers checked their solutions. But nobody paid any attention to the fact that solution x_3 given in [12] is not reliable.

§10. An investigation of particular special cases.

In previous sections we have examined the most often occuring general case when during variations of coefficients signs of algebraic additions do not change. In this section we shall investigate rather seldom occuring but at the same time possible special cases when some algebraic additions can change their sign during variations of a determinant coefficients. And we shall also investigate a special case when some algebraic additions are equal to zero.

During the forming of "table of signs"a special particular case is an equality to zero of one or several algebraic additions A_{ij}. In this special case a methodics described in a previous section does not allow us to choose in a corresponding place of a "table of signs"a sign "plus"or "minus".

Example №12.

The following simple determinant can serve as an example:

$$\begin{vmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{vmatrix} \tag{138}$$

in which $A_{11} = 1$; $A_{12} = 0$; $A_{13} = -1$; $A_{21} = -1$; $A_{22} = 1$; $A_{23} = -1$; $A_{31} = -1$; $A_{32} = -1$; $A_{33} = 3$.

Let us suggest that in places corresponding to zero algebraic additions in "table of signs"is put a number "zero". Then a "table of signs"will become:

$$\begin{vmatrix} + & 0 & - \\ - & + & - \\ - & - & + \end{vmatrix} \tag{139}$$

The main linear part of a determinant (138) increase will not depend on value of variations of element a_{12} and on its sign since in this case (we examine "absolute" variations $a_{ij\varepsilon} = a_{ij} + \varepsilon_{ij}$)
we have:

$$\Delta_{lin} = \varepsilon_{11}\cdot 1+\varepsilon\cdot 0+\varepsilon_{13}\cdot(-1)+\varepsilon_{21}\cdot(-1)+\varepsilon_{22}\cdot 1+\varepsilon_{23}\cdot(-1)+\varepsilon_{31}\cdot(-1)+\varepsilon_{32}\cdot(-1)+\varepsilon_{33}\cdot 3 \tag{140}$$

If all $|\varepsilon_{ij} \leq 0,01$ then the largest growth of a determinant (139) in a linear approximation is equal to:

$$\Delta_{linmax} = 0,01(1+0+1+1+1+1+1+1+3) = 0,11 \tag{141}$$

and doest not depend on ε_{12}. But a total growth of a determinant owing to nonlinear effects can depend on ε_{12}. If in "table of signs"(139) instead of a zero we put "plus" then if $|\varepsilon_{ij}| = 0,01$ and instead of determinant (138) we obtain the following determinant:

$$\begin{vmatrix} 2,01 & 1,01 & 0,99 \\ 0,99 & 2,01 & 0,99 \\ 0,99 & 0,99 & 1,01 \end{vmatrix} = 1,1308 \tag{142}$$

and thus $\Delta_+ = 0,1308$ but if in "table of signs"(139) instead of a zero we put a sign "minus"then we obtain a determinant:

$$\begin{vmatrix} 2,01 & 0,99 & 0,99 \\ 0,99 & 2,01 & 0,99 \\ 0,99 & 0,99 & 1,01 \end{vmatrix} = 1,1204 \tag{143}$$

and $\Delta_+ = 0,1204$.

Thus if we take into account nonlinear but a change of a determinant depends on a sign corresponding to $A_{ij} = 0$ and in a "table of sign"instead of a zero it is make correct to put a double signs – i.e., for example, a table (139) is written in the form:

$$\begin{vmatrix} + & \pm & - \\ - & + & - \\ - & - & + \end{vmatrix} \tag{144}$$

and to compute a determinant as for one and so for the other sign as well that lies in the table.

Example №13. Let us examine a determinant:

$$\begin{vmatrix} 1 & 2 & 3 \\ -2 & -4 & -5 \\ 3 & 5 & 6 \end{vmatrix} = 1 \tag{145}$$

at which $A_{11} = 1$; $A_{12} = -3$; $A_{13} = 3$; $A_{21} = 3$; $A_{22} = -3$; $A_{23} = 1$; $A_{31} = 2$; $A_{32} = -1$; $A_{33} = 0$.

If for all i and j we shall have $|\varepsilon_{ij}| \leq 0,01$ then the largest growth of a determinant in a linear approximation is:

$$\Delta_{linmax} = 0,01(1 + 2 \cdot 3 + 3 \cdot 2 + 2 \cdot 3 + 4 \cdot 3 + 5 \cdot 1 + 3 \cdot 2) + 5 \cdot 1 + 6 \cdot 0 = 0,47 \tag{146}$$

A "table of signs"for determinant (145) is of the form.

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & \pm \end{vmatrix} \tag{147}$$

Let us compute a variated determinant during a choice of a sign "plus"in a "table of signs"(147) when it becomes:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix} \tag{148}$$

and a variated determinant is equal to:

$$\begin{vmatrix} 1,01 & 1,98 & 3,03 \\ -1,98 & -4,04 & -4,95 \\ 3,03 & 4,95 & 6,06 \end{vmatrix} = 1,4749$$

We see that the growth of a determinant in this case will achieve 0,4749 or 47,49% from a rating one. If in a "table of signs"(147) a sign "minus"is chosen it will become

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & - \end{vmatrix}, \tag{149}$$

then a determinant (145) will become

$$\begin{vmatrix} 1,01 & 1,98 & 3,03 \\ -1,98 & -4,04 & -4,95 \\ 3,03 & 4,95 & 5,94 \end{vmatrix} = 1,4934$$

and a growth will become equal to 0,4903 or 49,34% from its nominal value and 105% of its maximal change in a linear approximations.

A choice of a sign "minus"in a "table of signs"(147) leads to the largest change of a determinant in this case and a variation of a determinant turns out to be in this case by 49,34 times more than the largest variation of its each elements.

With other combinations of variations in determinant elements its variation can be much less. So if $|\varepsilon_{ij}| = 0,01$ and when a combination of variations signs that corresponds to "table of signs":

$$\begin{vmatrix} + & - & + \\ + & - & + \\ + & - & - \end{vmatrix} \tag{150}$$

when determinant (145) turns into

$$\begin{vmatrix} 1,01 & 1,98 & 3,03 \\ -2,02 & -3,96 & -5,05 \\ 3,03 & 4,95 & 5,94 \end{vmatrix} = 1,099 \tag{151}$$

a change of a determinant will only be 9,9% from its rating values.

Note that if in a determinant there are many additions equal to zero (or close to zero) then a value of variations will be in general and as a whole less than during not equal to zero algebraic additions.

Example №14. As an example let us examine a determinant

$$\begin{vmatrix} 2 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = -1 \tag{152}$$

in which $A_{11} = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0$; $A_{12} = -\begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 0$ and similarly $A_{13} = 1$; $A_{21} = -1$; $A_{22} = 1$; $A_{23} = 0$; $A_{31} = 1$; $A_{32} = 0$; $A_{33} = -2$.

If for all i and j we have $\varepsilon_{ij} = \varepsilon_0 = \mp 0,01$ then the largest variation of a determinant in a linear approximation will be equal to

$$\Delta_{linmax} = 0,01(0+2+1+2+1+0+1+0+2) = 0,09 \tag{153}$$

i.e. variation in a linear approximation will be only by 9 times more than variations of its elements.

"Tables of signs"with possible variations in this case are of the form:

$$\begin{vmatrix} \pm & - & + \\ - & + & \pm \\ + & \pm & - \end{vmatrix} \tag{154}$$

A direct computation of a determinant for all $2^9 = 512$ combinations of positive and negative variations of its nine elements leads to an interesting result: to a similar (with the exactness of the sixth sign) value of a determinant equal to -0,911897 (and thus – to the largest variation of a determinant in the direction of an increase $\Delta_+ = 0,088103$) they lead to different "tables of signs" , i.e.

$$\begin{array}{lll} the\ first: \begin{vmatrix} - & - & + \\ - & + & - \\ + & + & - \end{vmatrix}, & the\ second \begin{vmatrix} - & - & + \\ - & + & + \\ + & - & - \end{vmatrix}, & the\ third: \begin{vmatrix} - & - & + \\ - & + & + \\ + & + & - \end{vmatrix}, \\ the\ fourth \begin{vmatrix} + & - & + \\ - & + & + \\ + & - & - \end{vmatrix}, & the\ fifth \begin{vmatrix} + & - & + \\ - & + & - \\ + & - & - \end{vmatrix}, & the\ sixth \begin{vmatrix} + & - & + \\ - & + & - \\ + & + & - \end{vmatrix}. \end{array} \tag{155}$$

They all are variations of table (154) and lead to the one and the same value of a determinant: -0,911897.

It is curious to note that to the largest negative value of a determinant. to a value -1,092695 and thus – to the largest variation in the direction of a decrease, to $\Delta_- = -0,092695$ leads only one "table of signs" , i.e. to a table:

$$\begin{vmatrix} - & + & - \\ + & - & - \\ - & - & + \end{vmatrix} \tag{156}$$

which is inverse in relation to one of variations of table (154) but (what is surprising) are not inverse to any of tables (155).

For determinant (152) if $|\varepsilon_{ij}| \leq 0,01$ the largest variation of a determinant is by 9,2695 times more than each of its elements.

For order "table of signs"different from tables (155) and (156) determinant variations will be less.

So for "table of signs"

$$\begin{vmatrix} + & + & + \\ + & + & + \\ + & + & + \end{vmatrix}$$

we shall have $det = -1,030301$ and a variation of a determinant is equal to 0,030301.

For tables

$$\begin{vmatrix} + & + & + \\ + & - & + \\ - & + & + \end{vmatrix} \quad and \quad \begin{vmatrix} + & + & + \\ + & - & + \\ - & - & + \end{vmatrix}$$

we shall equally have $det = -1,071509$.

For a table

$$\begin{vmatrix} - & - & - \\ + & + & + \\ + & + & + \end{vmatrix}$$

we have $det = -1,009899$, i.e. determinant variation in this case will be very small – even more small than variation of each of elements of a determinant. More less variation of a determinant will be with "tables of signs":

$$\begin{vmatrix} - & - & - \\ + & - & + \\ - & + & - \end{vmatrix} \quad and \quad \begin{vmatrix} - & - & - \\ + & - & + \\ - & - & - \end{vmatrix}$$

for which analogically $det = -1,0095$.

To tables

$$\begin{vmatrix} - & - & - \\ - & + & + \\ + & - & + \end{vmatrix} \quad and \quad \begin{vmatrix} - & - & - \\ - & + & - \\ + & - & + \end{vmatrix}$$

corresponds $det = -0,98109$ and a similar small variation of a determinant in the direction of an increase, so $\Delta_+ = 0,010891$.

Example №15. Now let us examine the most exotic object – a determinant

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0 \tag{157}$$

in which all algebraic additions are equal to zero. In such determinants during variations of their elements the main linear part of a determinant growth is equal to zero and a value of the growth will depend only on members of higher orders.

For such an exotic object as determinant (157) a general theory of forming "tables of signs"for the calculation of the largest deviation of a determinant from its rating value does not work. A direct consideration has shown that in this case there exists not one but several "tables of sings"that lead (if for all i and j we shall have $|\varepsilon_{ij}| = 0,01$) to similar maximal variations of a determinant with the exactness of up to the fourth sign.

So to the largest deviation in the direction of an increase equal (with exactness up to the fourth sign we have 0,0012) leads the following "table of sings"

$$\begin{vmatrix} + & + & - \\ - & + & + \\ + & - & + \end{vmatrix} \quad and \quad \begin{vmatrix} + & - & + \\ + & + & - \\ - & + & + \end{vmatrix} \tag{158}$$

and to the largest deviation in the direction of a decrease equal to -0,0012 lead "tables of signs"

$$\begin{vmatrix} + & - & + \\ - & + & + \\ + & + & - \end{vmatrix} \quad and \quad \begin{vmatrix} + & + & - \\ + & - & + \\ - & + & + \end{vmatrix}$$

Thus for determinant (157) its largest variation will not exceed 0,12 from the largest variation of each of elements of a determinant.

Let us consider the first of "table of signs"(158) when determinant (157) if $|\varepsilon_{ij} = \varepsilon_0$ will be of the form:

$$\begin{vmatrix} 1+\varepsilon_0 & 1+\varepsilon_0 & 1-\varepsilon_0 \\ 1-\varepsilon_0 & 1+\varepsilon_0 & 1+\varepsilon_0 \\ 1+\varepsilon_0 & 1-\varepsilon_0 & 1+\varepsilon_0 \end{vmatrix} = 12\varepsilon_0^2 + 4\varepsilon_0^3.$$

This formula quite obviously shows that a variation of a determinant wholly depends on nonlinear members.

If we investigate the second of "tables of signs"when determinant (157) after variations of its elements has become:

$$\begin{vmatrix} 1+\varepsilon_0 & 1-\varepsilon_0 & 1+\varepsilon_0 \\ 1+\varepsilon_0 & 1+\varepsilon_0 & 1-\varepsilon_0 \\ 1-\varepsilon_0 & 1+\varepsilon_0 & 1+\varepsilon_0 \end{vmatrix} = 12\varepsilon_0^2 + 4\varepsilon_0^3$$

we see that for the first and the second of "tables of signs"(158) as well variations of determinants are exactly equal to each other. The dependence of determinants value on ε_0 is shown in figure 5.

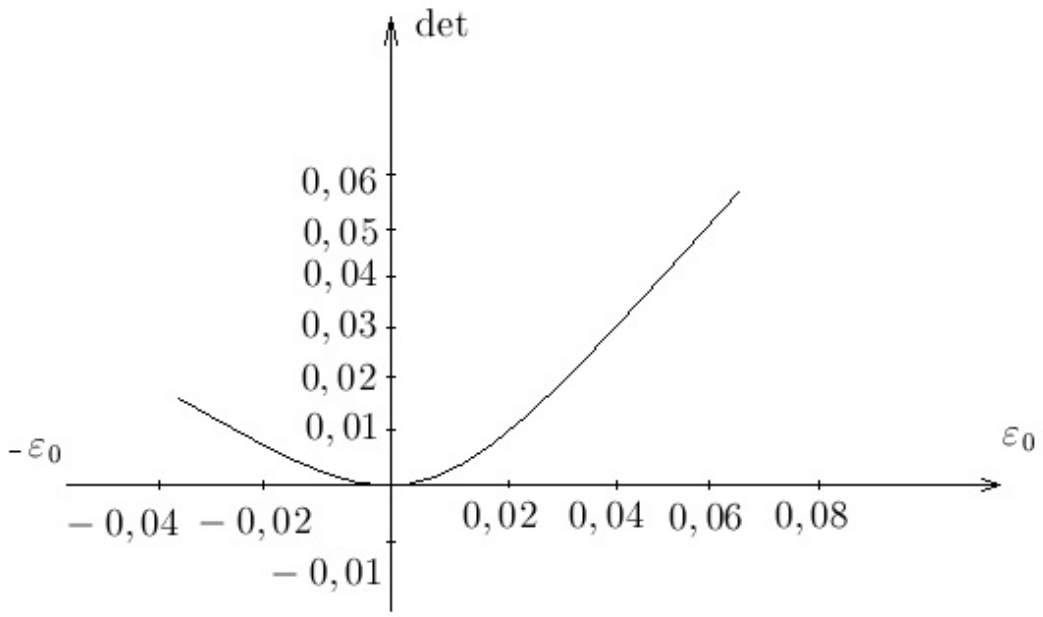

fig. 5

Similarly while forming variations of elements of elements in determinant (157) in accordance with the first and the second of "tables of signs"we see that

$$\begin{vmatrix} 1+\varepsilon_0 & 1-\varepsilon_0 & 1+\varepsilon_0 \\ 1-\varepsilon_0 & 1+\varepsilon_0 & 1+\varepsilon_0 \\ 1+\varepsilon_0 & 1+\varepsilon_0 & 1-\varepsilon_0 \end{vmatrix} = \begin{vmatrix} 1+\varepsilon_0 & 1+\varepsilon_0 & 1-\varepsilon_0 \\ 1+\varepsilon_0 & 1-\varepsilon_0 & 1+\varepsilon_0 \\ 1-\varepsilon_0 & 1+\varepsilon_0 & 1+\varepsilon_0 \end{vmatrix} = -12\varepsilon_0^2 - 4\varepsilon_0^3.$$

Surely, determinants in which all algebraic additions are equal to zero in practice almost never occur but as a special an exotic object they are interesting.

The above material allows as to give additional basement of trustworthiness in estimating a determinant during small but final variations of its elements. Strictly speaking a "table of signs"is formed for infinitely small variations ε_{ij}. Will it not change if variations ε_{ij} are small but finite? A sign that is in a "table of signs"on the crossing of the ith line and jth column can change in such a case if an algebraic addition A_{ij} is very small and can change a sign during the pass from rating values of determinants elements to variated ones. In order to obtain an exact value of a maximal variation of determinant it is necessary to watch possible change of signs in algebraic additions.

Example №16. Let us examine instead of determinant (145) a determinant that is close to it:

$$\begin{vmatrix} 1 & 2+m & 3 \\ -2 & -4 & -5 \\ 3 & 5 & 6 \end{vmatrix} = 1 - 3m$$

where m – positive number that is small in comparison with 1. Algebraic additions of elements of this determinant are equal to: $A_{11} = 1$; $A_{12} = -3$; $A_{13} = 2$; $A_{21} = 3 - 6m$; $A_{22} = -3$; $A_{23} = 1 + 3m$; $A_{31} = 2 - 5m$; $A_{32} = -1$; $A_{33} = 2m > 0$ and therefore its "table of signs"is of the form:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}.$$

If all elements of a determinant have obtained small changes $\pm\varepsilon$ whose signs correspond to a written table then a determinant will become:

$$\begin{vmatrix} 1+\varepsilon & (2+m)(1-\varepsilon) & 3(1+\varepsilon) \\ -2(1-\varepsilon) & -4(1+\varepsilon) & -5(1-\varepsilon) \\ 3(1+\varepsilon) & 5(1-\varepsilon) & 6(1+\varepsilon) \end{vmatrix}.$$

For ε that are small in comparison with 1 signs in algebraic additions from A_{11} up to A_{32} do not change but an addition A_{33} that is equal to the following determinant

$$A_{33} = \begin{vmatrix} 1+\varepsilon & (2+m)(1-\varepsilon) \\ -2(1+\varepsilon) & -4(1+\varepsilon) \end{vmatrix} = 2m - 16\varepsilon - 4m\varepsilon + 2m\varepsilon^2$$

can easily change a sign if ε increases from an initial value $\varepsilon = 0$. If, for example, $m = 0,01$ then a sign of A_{33} will change when $\varepsilon > 0,0124$ but when $m = 0,1$ a sign of A_{33} will change when $\varepsilon > 0,122$. Therefore if, for example, $m = 0,01$ then during the calculation of a determinant variations if $\varepsilon > 0,0124$ it is necessary to take into account that "table of signs"has changed and corresponds to $A_{33} < 0$.

It is also necessary to note that small algebraic additions hardly influence the variation of an examined determinant and therefore we can consider signs of additions unchangable with a sufficient for practical aims exactness and in order to compute variations of determinants it is necessary to apply a simple algorithm given in §6 as values of variations in coefficients in equations systems which is a mathematical model of a real object almost always can be estimated only approximately.

§11. Computation of exact values of variations of each of components in solutions vector.

Methods of computing determinants variations given in previous sections allows us to obtain exact variations of all components x_1; x_2; ...; x_n in solutions vector based on the following Cramer formula:

$$x_i = \frac{D_i}{D} \ (i = 1, 2, ..., n)$$

But here there are some special traits. In general and as a whole variation x_i will be the largest (in the direction of the increase) in such a case if a determinant D obtains the largest possible variation in the direction of a decrease of determinant value. But determinant D_i undergoes the largest possible variation in the direction of an increase.

In the first approximation variation x_i, a variation of a quotidien due to the division of determinant D_i by a determinant D is a sum of variations D and D_i. But in fact variation of x_i is less than a sum of variations of D and D_i since it is necessary to take into account that variations of elements in determinants D and D_i are not independent since $n-1$ columns in determinants D and D_i coincide.

Example №17. Let us explain the above problems on a simple example of a system:

$$\left.\begin{array}{r} 2x_1 - x_2 = 1 \\ x_1 + x_2 = 2 \end{array}\right\} \tag{159}$$

For this system

$$D = \begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} = 3;\ D_1 = \begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix};\ D_2 = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3 \tag{160}$$

and thus $x_1 = \frac{3}{3} = 1$; $x_2 = \frac{3}{3} = 1$.

Let us speak about the computation of a variation of x_2 supposing that relative variations of all six coefficients in system (159) do not exceed (by a module) a value 0,01 but their signs can be any.

For determinant D the largest variation in the direction of a decrease will be when signs of elements variations defined by its inverse "table of signs"which for a determinant D is of the form:

$$\begin{vmatrix} - & + \\ - & - \end{vmatrix} \tag{161}$$

and here determinant itself turns into a determinant:

$$\begin{vmatrix} 2(1-0,01) & -1(1+0,01) \\ 1(1-0,01) & 1(1-0,01) \end{vmatrix} = 2,96 \tag{162}$$

(here and later computations have been carried out with the exactness of the third sign). For determinant D_2 the largest variation in the direction of an increase will be with the following "table of signs"

$$\begin{vmatrix} + & - \\ - & + \end{vmatrix} \tag{163}$$

when determinant D_2 becomes:

$$\begin{vmatrix} 2(1+0,01) & 1(1-0,01) \\ 1(1-0,01) & 2(1+0,01) \end{vmatrix} = 3,06 \tag{164}$$

variation of determinant D_2 is equal to 0,06 or 2% from its rating value.

Variation of x_2 in the first approach is equal to a sum of variations D_2 and D is equal in the first approach to three percents.

But elements variations of the first column of determinants D and D_2 are the same (these are variations of elements a_{11} and a_{12}. They are the same in D and D_2). Therefore two variants of computation are possible.

The first variant: we shall base ourselves in general on "table of signs"of determinant D. In it and as well in the first column of determinant D_2 we choose signs in coordination with "tables of signs"(161). And only in the second column of determinant D_2 we put signs of variations in accordance with its "table of signs" , i.e. in accordance with table (163). We obtain:

$$x_2 = \frac{\begin{vmatrix} 2(1-0,01) & 1(1-0,01) \\ 1(1-0,01) & 2(1+0,01) \end{vmatrix}}{\begin{vmatrix} 2(1-0,01) & -1(1+0,01) \\ 1(1-0,01) & 1(1-0,01) \end{vmatrix}} = \frac{3,02}{2,96} = 1,020270 \tag{165}$$

i.e. variation of x_2 (in the direction of an increase) is equal to 0,02.

The second variant. Let us base ourselves on a numerator, on determinant D_2. In it and in the first column of a determinant D as well we shall choose signs in accordance with "tables of signs"for a determinant D_2, i.e. with a table (163). And only in the second column of determinant D we choose signs in accordance with an inverse "table of signs"of determinant D, i.e. accordance with table (161).

We obtain:

$$x_2 = \frac{\begin{vmatrix} 2(1+0,01) & 1(1-0,01) \\ 1(1-0,01) & 2(1+0,01) \end{vmatrix}}{\begin{vmatrix} 2(1+0,01) & -1(1+0,01) \\ 1(1-0,01) & 1(1-0,01) \end{vmatrix}} = \frac{3,06}{3} = 1,02 \tag{166}$$

i.e. variation of x_2 in the second variant in a combination of signs of variations of matrix A elements and vector B in system (159) is equal to 0,02 or 2% of its rating values.

In this case the first and the second variants of computation have led to the same result in practice.

Let us explain. In the first variant we take into consideration the largest possible (in the direction of a decrease) variation of determinant D. Variation of determinant D_i can turn out to be not maximally possible in this variant.

In the second variant we take into consideration a computation of maximally possible (in the direction of an increase) variation of determinant D_2. Here variation of determinant D can turn out to be not maximally possible. Now we can form a general rule of computing the largest variation of any component of x_i in a vector of solutions x.

A general rule. An initial material for the calculation serve estimates of maximal absolute values of variations of elements a_i and b_i in the form of the following inequalities:

$$|\varepsilon_i| \leq \varepsilon_{ij0};\ |b_i| \leq b_{i0} \tag{167}$$

and computed "tables of signs"in determinants D and D_i in Cramer formulas.

Let us take two variants of computations. The first variant is based on determinant D in a formula $x_i = \dfrac{D_i}{D}$. We must compute the least possible value of determinant D by means of its "inverse table of signs". After this we must calculate a value of determinant D_i while taking into account variations of its elements. Here signs of elements variations in all columns of a determinant excluding the ith column of a determinant (recall that this column coincides with a column b_1; b_2; ...; b_n in the right side) then we must choose a determinant D in accordance with a "table of signs". Later we divide values of D and D_i by each other that have been obtained while taking into account variations and as a result we obtain – a value x_i.

The second variant. It is oriented on determinant D_i. Signs in variations of all elements of determinant are chosen in accordance with its "table of signs". After this we must calculate a determinant D_i and obtain its largest possible value (while setting this value of coefficients variations). After this we start calculating determinant D while taking into account variations of its elements. Here signs in elements variations of all columns of determinant D except the ith column we choose in accordance with "table of signs"of a determinant D_i. Later when determinants have been calculated we calculate $x_i = \dfrac{D_i}{D}$ and compare it with the result of computing x_i by the first variant.

Note that by applying this rule we compute the largest possible variation of x_i in the direction of an increase – in a positive direction. If it is necessary to compute the largest possible variation of x_i in the negative direction, in the direction of a decrease then it is necessary to take into account that this variation will be the largest if variation of determinant D_i is largest in the direction of a decrease (i.e. signs in elements variations corresponds to "inverse table of signs"of determinant D_i). But variation of determinant D will be the largest in the direction of an increase (i.e. signs in elements variations will correspond to a direct "table of signs"of determinant D). Since determinants D and D_i have $n-1$ general columns it is impossible to combine these contradictable requirements and it is necessary (as earlier) to apply two variants of computing for the least possible value of x_i.

The first variant. As earlier it is oriented on determinant D. In determinant D we

choose signs for all its "table of signs"after which we compute a determinant while we take into account variations. Later we put (in determinant D) signs in variations of elements of all columns except the ith column in correspondence with "table of signs"of determinant D. And we put signs of variations of the ith column into correspondence with "an inverse table of signs"of determinant D_i. Then we compute a determinant D_i while taking into account variations of its elements and divide D_i by D.

The second variant. As before it is oriented on determinant D_i. We choose signs of variations in all elements according to "inverse table of signs"and then we calculate determinant D while taking into account variations of its elements. In determinant D signs of variations of all its columns except the ith we put in accordance with "inverse table of signs"of determinant D_i and we put signs of variations of the ith column in accordance with "table of signs"of determinant D. Then we compute determinant D taking into account variations of its elements and we divide D_i by D. Later we compare in what variant of computing variation of x_i is larger.

As a whole a general rule of computing variations of x_i in the direction of an increase and a decrease while taking into account two variants of computing has become rather cumbersome. For a real application of this rule it is necessary of form a program for computing on an electronic machine.

the continuation of example №7.

Let us consider the same system (159) and calculate the largest possible variation of x_2 in the direction of its decrease.

The first variant. While taking into account "table of signs"(161) and (162) we shall obtain that for the first variant when a "table of signs"is a base in determinant D and only in the second column of determinant D_2 signs of variations are given in an accordance with its "inverse table of signs"will become:

$$x_2 = \frac{\begin{vmatrix} 2(1+0,01) & 1(1+0,01) \\ 1(1+0,01) & 2(1-0,01) \end{vmatrix}}{\begin{vmatrix} 2(1+0,01) & -1(1-0,01) \\ 1(1+0,01) & 1(1+0,01) \end{vmatrix}} = \frac{2,98}{3,03}$$

The second variant. In the second variant for a basis a determinant D_2 is taken in which signs of variations are put in accordance with its "inverse table of signs". In correspondence with this same table we have to put variations signs in the first column of determinant D as well.

And only in its second column signs of variations are put in accordance with "table of signs"of determinant D, i.e. the table:

$$\begin{vmatrix} + & - \\ + & + \end{vmatrix} \tag{168}$$

Thus

$$x_2 = \frac{\begin{vmatrix} 2(1-0,01) & 1(1+0,01) \\ 1(1+0,01) & 2(1-0,01) \end{vmatrix}}{\begin{vmatrix} 2(1-0,01) & -1(1-0,01) \\ 1(1+0,01) & 1(1+0,01) \end{vmatrix}} = 0,967 \tag{169}$$

In this case to the largest value of variation of x_2 in the direction of a decrease leads the second variant and a combination of variations signs in elements a_{11}; a_{12}; a_{21}; a_{22}; b_1; b_2 shown in formula (169), i.e. variation a_{11} has a sign "minus a_{12} – a sign "plus a_{22} – a sign "plus b_1 – a sign "plus b_2 – a sign "minus".

Finally we conclude that a component x_2 in a vector of solutions of a simple system of equations (159) (due to variations in system coefficients do not exceed 0,01 from its rating values) is subjected to the inequalities

$$0,967 \leq x_2 \leq 1,02 \tag{170}$$

Here an estimate (170) is exact since we can show such a certain combination of signs of variations in coefficients of system (159) at which inequalities (170) turn into exact equalities.

This example shows that an exact estimate of an error in solutions of a system of linear algebraic equations is in some degree a more complex problem than a computation of a solution itself.

Example №18.

Let us return once more to a system of equations (126). Earlier it has been shown that a component in solution x_3 is not at all reliable. This fact has already become evident during the investigation of determinant D_3 since during vary small variations in coefficients of determinant D_3 if $\varepsilon_0 \geq 0,0065$ determinant D_3 and with it x_3 as well (i.e. a moment applied to the end of a frame) changes a sign.

Now let us examine x_1 and x_2.

Components of a solution x_1 and x_2 defined by Cramer formulas by means of determinants D_1; D_2 and D are equal to:

$$x_{1N} = \frac{D_1}{D} = \frac{\begin{vmatrix} 3 & 12 & 15 \\ 5 & 16 & 12 \\ 1 & 4 & 6 \end{vmatrix}}{\begin{vmatrix} 14 & 12 & 15 \\ 12 & 16 & 12 \\ 5 & 4 & 6 \end{vmatrix}} = -\frac{12}{48} = -0,25;\ x_2 = \frac{D_2}{D} = \frac{\begin{vmatrix} 14 & 3 & 15 \\ 12 & 5 & 12 \\ 5 & 1 & 6 \end{vmatrix}}{\begin{vmatrix} 14 & 12 & 15 \\ 12 & 16 & 12 \\ 5 & 4 & 6 \end{vmatrix}} = \frac{21}{48} = 0,4375 \tag{171}$$

For preliminary estimate of possible variations of a component of a vector in solutions x_1 and x_2 let us compute variations of determinants that enter into formula (171) in a linear approach.

While computing algebraic additions for determinant D we shall obtain:

$$A_{11} = \begin{vmatrix} 16 & 12 \\ 4 & 6 \end{vmatrix} = 48;\ A_{12} = -\begin{vmatrix} 12 & 11 \\ 5 & 6 \end{vmatrix} = -12;\ A_{13} = \begin{vmatrix} 12 & 16 \\ 5 & 4 \end{vmatrix} = -32 \tag{172}$$

and similarly $A_{12} = -12;\ A_{22} = 9;\ A_{23} = 4;\ A_{31} = -96;\ A_{32} = 4;\ A_{33} = 80.$

"Table of signs"of determinant D becomes:

$$\begin{vmatrix} + & - & - \\ - & + & + \\ - & + & + \end{vmatrix} \tag{173}$$

If absolute values in variations of all elements in a determinant D are equal to $|\varepsilon_{ij}| \leq \varepsilon_0$ then the largest variation of a determinant if combinations of variations signs of its elements are the most unfavourable (in a linear approach) is equal to:

$$\Delta_{lin} = \varepsilon_0(14{\cdot}48+12{\cdot}12+15{\cdot}32+12{\cdot}12+16{\cdot}9+12{\cdot}4+5{\cdot}96+4{\cdot}12+6{\cdot}80) = 2630\varepsilon_0 \tag{174}$$

and determinant D satisfies (in a linear approximation) inequalities:

$$D_N - 2630\varepsilon_0 \leq D \leq D_N + 2630\varepsilon_0$$

or in relative units:

$$1 - 54,79\varepsilon_0 \leq \frac{D}{D_N} \leq 1 + 54,79\varepsilon_0, \tag{175}$$

where D_N – a rating value of determinant D.

In order to take into account a weak nonlinear dependence of determinant variation on a variation of its elements it is necessary to compute determinant D with variated elements. Here signs of variations correspond to "table of signs"(173). If for all i and j we have $|\varepsilon_{ij}| = 0,01$ then determinant D turns into:

$$\begin{aligned} D_{max} &= \begin{vmatrix} 14(1+0,01) & 12(1-0,01) & 15(1-0,01 \\ 12(1-0,01) & 16(1+0,01) & 12(1+0,01) \\ 5(1-0,01) & 4(1+0,01) & 6(1+0,01) \end{vmatrix} = \\ &= \begin{vmatrix} 14,14 & 11,88 & 14,85 \\ 11,88 & 16,16 & 12,12 \\ 4,95 & 4,04 & 6,06 \end{vmatrix} = 74,665 = D_N + 26,605 \end{aligned} \tag{176}$$

Formula (176) shows that for determinant D we shall have $\Delta_+ = 26,665$ and it differs very little from Δ_{lin} since if $\varepsilon_0 = 0,01$ we shall have $\Delta_{lin} = 26,3$.

In order to compute the largest variation of determinant D in the direction of a decrease signs in elements variations must be chosen in accordance with such "table of signs"that is inverse in relation to table (173), i.e. table

$$\begin{vmatrix} + & + & - \\ - & - & - \\ + & - & - \end{vmatrix} \tag{177}$$

when determinant D turns into a determinant

$$D_{min} = \begin{vmatrix} 13,86 & 12,12 & 15,15 \\ 12,12 & 15,84 & 11,88 \\ 5,05 & 3,96 & 5,94 \end{vmatrix} = 21,863 \tag{178}$$

We again see that $\Delta_+ = 48 - 21,863 = 27,173$. It little differs from Δ_{lin}.

The same computations must be carried out for determinant D_2 for which

$$A_{11} = \begin{vmatrix} 5 & 12 \\ 1 & 6 \end{vmatrix} = 18;\ A_{12} = -\begin{vmatrix} 12 & 12 \\ 5 & 6 \end{vmatrix} = -12;\ A_{13} = \begin{vmatrix} 12 & 5 \\ 5 & 1 \end{vmatrix} = -13; \tag{179}$$
$$A_{21} = -3;\ A_{22} = 9;\ A_{23} = 1;\ A_{31} = -39;\ A_{32} = 12;\ A_{33} = 34$$

"table of signs"is of the form:

$$\begin{vmatrix} + & - & - \\ - & + & + \\ - & + & + \end{vmatrix}$$

and thus if for all i and j we shall have $|\varepsilon_{ij}| \leq \varepsilon_0$ then the largest possible variation of determinant D_2 in a linear approach is equal to

$$\Delta_{lin} = \varepsilon_0(14 \cdot 18 + 3 \cdot 12 + 15 \cdot 13 + 12 \cdot 3 + 5 \cdot 9 + 12 \cdot 1 + 5 \cdot 39 + 1 \cdot 9 + 6 \cdot 34) = 984\varepsilon_0 \tag{180}$$

and thus determinant D_2 in a linear approach satisfies the following inequalities:

$$D_{2N} - 984\varepsilon_0 \leq D_2 \leq D_{2N} + 984\varepsilon_0 \tag{181}$$

or in relative units:

$$1 - 47\varepsilon_0 \leq \frac{D_2}{D_{2N}} \leq 1 + 47\varepsilon_0 \tag{182}$$

From formulas (175) and (182) a simple "estimate from above"follows in a linear approach for the variations of x_2:

$$1 - 47\varepsilon_0 - 49,7\varepsilon_0 \leq \frac{x_2}{x_{2N}} \leq 1 + 47\varepsilon_0 + 49,7\varepsilon_0 \tag{183}$$

and thus:

$$1 - 96,7\varepsilon_0 \leq \frac{x_2}{x_{2N}} \leq 1 + 96,7\varepsilon_0$$

but to achieve the highest and lowest borders of this estimate a relation $\frac{x_2}{x_{2N}}$ can be only if variations of determinants are independent.

In fact these variations are dependent and this allows us to give a more exact estimate.

Inequalities (175) and (182) show that the investigation of determinant D_2 does not allow to male a final conclusion about the reliability or nonreliability of a solution component x_2 and it is necessary to make a more detailed investigation of variation of

x_2 while taking into account possible combinations of variations in determinants D and D_2.

Variation of x_2 in the direction of an increase will be the largest in such a case if determinant D_2 that stands in a numerator will greatly increase and determinant D in a denominator will in a greater degree decrease. A value of a component of a solution x_2 in a greater degree will decrease if determinant D increases in a possibly large degree during variations of its elements. But variations of D and D_2 are not independent and therefore it is necessary (as we have already said before) it is necessary to take into account two variants of combining elements variations of determinants D and D_2.

The first variant of computation.

We orient ourselve on a denominator, on the largest (in the direction of an increase and decrease) variation of determinant D. In a linear approximation (as formula (175) shows) is equal to $\pm 2630\varepsilon_0$ and it corresponds to either direct or inverse "table of signs"of determinant D, i.e. either to table (173) or inverse – to the table (177).

In determinant D_2 the first and the third columns in the first variant of computing coincide with corresponding columns of determinant D. Therefore signs of their variations are not arbitrary and they must correspond to "table of signs"of determinant D. Signs of elements variations of the second column of determinant D_2 are arbitrary and to the largest increase of D_2 will lead such variations whose signs correspond to "table of signs"(180). As whole "table of signs"of determinant D_2 that secures its largest increase during the choice of the first variant of computing will become:

$$\begin{vmatrix} - & - & + \\ + & + & - \\ + & + & - \end{vmatrix} \tag{184}$$

while computing determinant D_2 if $|\varepsilon_0| = 0,01$ and "table of signs"(184) in this case we obtain:

$$D_2 = \begin{vmatrix} 13,86 & 2,97 & 15,15 \\ 12,12 & 5,05 & 11,88 \\ 5,05 & 1,01 & 5,94 \end{vmatrix} = 12,91 \tag{185}$$

By dividing D_2 by determinant (178) we obtain:

$$x_{2max} = \frac{12,91}{21,863} = 0,5905 \tag{186}$$

It is a maximal value of x_2 that is obtained by combining signs of variations corresponding to the firest variant of computing.

Now let us start computing the change of x_2 in the direction of a decrease. This change will be the largest if determinant D_2 becomes possibly very small and determinant D that is in a denominator – as much as larger. Thus the change of x_2 in the direction a decrease will be the largest if signs of elements variations in determinant D and the first and the third columns of determinant D_2 as well will correspond to "table of signs"(173), and signs in variations of the second column of determinant D_2 will correspond to "inverse table of signs"for D_2. As a whole "tables of signs"for a fraction are:

$$x_2 = \frac{D_2}{D} \tag{187}$$

that secure the largest change of x_2 in the direction of a decrease are of the form:

$$x_2 = \frac{\begin{vmatrix} + & + & - \\ - & - & + \\ - & - & + \end{vmatrix}}{\begin{vmatrix} + & - & - \\ - & + & + \\ - & + & + \end{vmatrix}} \tag{188}$$

while calculating determinants D and D_2 if variations are $|\varepsilon_{ij}| = 0,01$ and if we use "tables of signs"of variations given in formula (188) we obtain:

$$x_{2min} = \frac{\begin{vmatrix} 14,14 & 3,03 & 14,85 \\ 11,88 & 4,95 & 12,12 \\ 4,95 & 0,99 & 6,06 \end{vmatrix}}{\begin{vmatrix} 14,14 & 11,88 & 14,85 \\ 11,88 & 16,16 & 12,12 \\ 4,95 & 4,04 & 6,06 \end{vmatrix}} = \frac{28,93}{74,665} = 0,38746 \tag{189}$$

It is the least value of x_2 that is achieved while combining signs in variations that corresponds to the one shown in formula (188). Now let us pass to the second variant of computing.

The second variant of computing

In this variant we base ourselves on a numerator, on determinant D_2 and taking into account that variation of x_2 in the direction of an increase will be largest when D_2 takes the largest value but determinant D – the least in the possible ones. Determinant D_2 – the largest if there are signs of variations in its elements that correspond to "table of signs"(180). In determinant D signs of variations of the first and third columns must correspond to that same table (180) and only in the second column signs of variations are not connected with this condition and to a minimal value of D lead signs corresponding to "inverse table of signs"for D, i.e. – to table (177).

As a whole "table of signs"for computing x_{2max} can be written in the form:

$$x_2 = \frac{\begin{vmatrix} + & - & - \\ - & + & + \\ - & + & + \end{vmatrix}}{\begin{vmatrix} + & + & - \\ - & - & + \\ - & - & + \end{vmatrix}}, \tag{190}$$

and then

$$x_{2max} = \frac{\begin{vmatrix} 14,14 & 2,97 & 14,85 \\ 11,88 & 5,05 & 12,12 \end{vmatrix}}{\begin{vmatrix} 14,14 & 12,12 & 14,85 \\ 11,88 & 15,84 & 12,12 \\ 4,95 & 3,96 & 6,06 \end{vmatrix}} \tag{191}$$

while calculating a determinant we obtain:

$$x_{2max} = \frac{30,878}{67,484} = 0,45756 \tag{192}$$

while passing to calculate the least possible value of x_2 (during the second variation of computation) we must note that this least value will be achieved if a nominator D_2 be minimal and denominator D – maximal. Since during the second variant of computation we start from a nominator, D then for D_2 we take into account its "inverse table of signs i.e. a table:

$$\begin{vmatrix} - & + & + \\ + & - & - \\ + & - & - \end{vmatrix} \tag{193}$$

that is inverse to table (180).

In a "table of signs" for a denominator, for determinant D the first and the third columns coincide with corresponding columns of table (193) and only the second column can be chosen as such that secures the largest possible value of D. This value will secure a column that coincides with the second column of table (173). As a whole "table of signs"for a denominator, for determinant D is of the form:

$$\begin{vmatrix} - & - & + \\ + & + & - \\ + & + & - \end{vmatrix} \tag{194}$$

By computing determinants D and D_2 with variations in which $|\varepsilon_{ij}| = 0,01$ and having such signs that correspond to tables (193) and (194) we obtain:

$$x_{2min} = \frac{\begin{vmatrix} 13,86 & 3,03 & 15,15 \\ 12,12 & 4,95 & 11,88 \\ 5,05 & 0,99 & 5,94 \end{vmatrix}}{\begin{vmatrix} 13,86 & 11,88 & 15,15 \\ 12,12 & 16,16 & 11,88 \\ 5,05 & 4,04 & 5,94 \end{vmatrix}} = \frac{11,23}{28,124} = 0,39766 \tag{195}$$

By computing inequalities (186), (189), (193) and (195) finally we find an interval in the interior of which a component of x_2 of a solutions X vector in a system of equations (126) can be:

$$0,38746 \leq x_2 \leq 0,5905 \tag{196}$$

or in relative units:

$$0,886 \leq \frac{x_2}{x_{2N}} \leq 1,3497 \tag{197}$$

and here estimates (196) and (197) are exact estimates, i.e. we can always indicate such a combination of coefficients variations in an examined system of equations at which the left or the right inequality of a form (196) is satisfied with a sign of an equality.

So, for example, an equality:

$$x_2 = 0,38746 \tag{198}$$

is satisfied if variations of a matrix in coefficients of equations system (126) (if $|\varepsilon_{ij}| = 0,01$) correspond to the following table of signs:

$$\begin{vmatrix} + & - & - \\ - & + & + \\ - & + & + \end{vmatrix} \tag{199}$$

and variations of coefficients in a vector–column of the right side satisfy "table of signs"

$$\begin{vmatrix} + \\ - \\ - \end{vmatrix} \tag{200}$$

Note. Formulas (179) show that in D_2 algebraic additions A_{21} and A_{23} are small therefore in order to take into account the change of their signs if $\varepsilon_0 = 0,01$ (about which we spoke in §10) we can slightly make more exact values of x_{2max} and x_{2min}.

§12. A general algorithm for an exact estimate of errors in each of components of solutions vector.

The material of a previous section shows that an exact estimate of a possible error of each of components x_1; x_2; ...; x_n of vector X requires rather cumbersome computations.

First of all in order to form a table of signs it is necessary to compute algebraic additions of determinants D and D_i in Cramer formulas. Each of algebraic additions in a system consisting of n equations is a determinant of the $n-1$th degree. As is known from [9, 10] its computation requires approximately $(n-1)^3$ multiplications. In all computation of all n^2 algebraic additions requires approximately n^5 multiplications. But in order to form a "table of signs"we can manage without direct computation of algebraic additions but to apply an inverse matrix A^{-1} for whose computation there exists convenient and well–develop programs since each of elements in an inverse matrix is a quotient of the division of a corresponding algebraic addition by matrix determinant.

Besides not in all cases it is necessary to use all computations described in a previous section. We can start with computing variations of determinants D and D_i in a linear approximation. If these variations are large in comparison with their rating values this means that an examined system is ill–conditioned if they are small – they are well–conditioned.

We can also compute "a natural limit of variations i.e. variations of matrix A whose signs correspond to an "inverse table of signs"and their absolute values are such that a determinant of matrix A turns into zero.

Later if in the course of computations of determinants D_1; D_2; ...; D_n it turns out that for a component (that interests as) of a solution x_i vector a determinant D_i (while there are such variations of its elements which can be really met during its exploitation) changes its sign and determinant D does not change a sign then x_i will change a sign. This means that a solution is not reliable and we can finish our calculations at this moment. So, in §9 during the examination of example №11 the investigation of a variation in determinant D_3 has already shown that the computation of a component x_3 of a vector of solution X in a system of equations (126) is apriori not reliable.

If the examination of determinants D_i does not at once mean the reliability or nonreliability of a solution then it is necessary to carry out an investigation of exact intervals in the interior of which are components that interest us: x_1; x_2; ...; x_n of a vector of solution X by applying a methodics given in §11.

Note that although in examples investigated in §11 we have limited ourselves by computing variations of components of vector of solution X when $\varepsilon_0 = 0,01$ this fact does not lower the universality of an investigation. Really earlier it has been shown that if ε_0 is small the dependence of determinant variations on ε_0 with a very good degree of exactness is close to a linear one (figures №2 and №3 can serve as examples). Therefore when, for example, it has been computed that when $\varepsilon_0 = 0$ we shall have $x_2 = 0,4375$ but when $\varepsilon_0 = 0,01$ we have $x_2 = 0,6824 = 0,4375 + 0,2449$ then we can state that with good exactness if $\varepsilon_0 = 0,01$ we shall have $x_2 \leq 0,4375 + 0,02449 = 0,462$ etc...

Later it must be noted that although in these examples during variations of coefficients a_{ij} satisfying inequalities $|\varepsilon_{ij}| \leq \varepsilon_0$ the calculation has been carried out for a limited the worst variant when for all i and j we had $\varepsilon_{ij} = \pm\varepsilon_0$. It is not difficult to compute variations of determinants for any certain values ε_{ij} surely if these values are known to us. The main difficulty is in forming "table of signs". If it is formed then a variation of a determinant for any ε_{ij} can be easily computed.

Let us return to example №17 where a system (159) was considered with a determinant

$$D = \begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} = 3 \tag{201}$$

for which "table of signs"is of the form:

$$\begin{vmatrix} - & + \\ - & - \end{vmatrix} \tag{202}$$

For a determinant (201) and $|\varepsilon_{ij}| = 0,01$ according to formula (162) the least possible value equal to 2,96 has been computed. If it is known that for element a_{11} will be $|\varepsilon_{ij}| = 0$ and for other i and j remains $|\varepsilon_{ij}| = 0,01$ then the least possible value for determinant D can be computed according to formula (similar to formula (162)):

$$D_\varepsilon = \begin{vmatrix} 2 & -1,01 \\ 0,99 & 0,99 \end{vmatrix} = 2,98. \tag{203}$$

Known additional difficulties are the presence of "agreed" signs of variations of some coefficients. Let us suppose that a beam has moved a little to the right in comparison with its projected accomodation. In this case parameter l_3 (a length of a beam that is more right than a right support) will obtain a variation with a positive sign but then parameter l_1 (a length of a beam that is more left than the left support) will have a variation with (by all means) a negative sign.

The presence of "agreed"variations decreases a value of determinants variations and containing x_1; x_2; ...; x_n vector of solutions X in comparison with a case of completely independent variations.

Let us return to example №17 in which a system of equations (159) is examined and let us consider determinant D_2:

$$D_2 = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3 \tag{204}$$

for which "table of signs"is of a form:

$$\begin{vmatrix} + & - \\ - & + \end{vmatrix} \tag{205}$$

If for all i and j we have $|\varepsilon_{ij}| = 0,01$ but signs of variations ε_{ij} do not depend on each other and can be any then the largest possible value of a determinant (as it has already been calculated in example №17) is equal to:

$$\begin{vmatrix} 2,02 & 0,99 \\ 0,99 & 2,02 \end{vmatrix} = 3,06 \tag{206}$$

Now let us suppose that coefficients variations of the second column depend on each other and can be either both positive or both – negative (recall that coefficients of the second column in determinant (204) are coefficients of a right side of a system of equations (159) so that such dependence between variations is possible). If both variations of elements of the second column are positive then a determinant becomes:

$$\begin{vmatrix} 2,02 & 1,01 \\ 0,99 & 2,02 \end{vmatrix} = 3,04 \tag{207}$$

If both variations are negative then a determinant will be equal to

$$\begin{vmatrix} 2,02 & 0,99 \\ 0,99 & 1,98 \end{vmatrix} = 3,02 \tag{208}$$

In both cases variations of a determinant (as it will be expected) are smaller than during independent variations of elements.

Thus if we take into account the dependence between themselves of variations in a system of equations coefficients then such variations can decrease that enter into Cramer formulas determinants and at the end they can narrow intervals in the interior of which are intervals of examined equations systems.

But there is an important special case that must be investigated – it is a case of symmetric matrixes A in equations systems $AX - B$. Very often we had to meet with problems of construction mechanics but it does not bring any new difficulties into a general algorithm.

A special case of symmetrical matrixes.

In construction mechanics a lot of computation problems can be reduced to the computation of solutions of such systems of algebraic equations $AX = B$ in which a matrix of coefficients A is symmetrical, i.e. $a_{ij} = a_{ji}$.

Example №19. A system of equations

$$\begin{aligned} 3x_1 + x_2 + x_3 &= 4 \\ x_1 + 2x_2 + x_3 &= 2 \\ x_1 + x_2 + 2x_3 &= 1 \end{aligned}$$

has a symmetrical matrix since $a_{12} = a_{21}$; $a_{13} = a_{31}$; $a_{23} = a_{32}$.

A determinant of a system

$$D = \begin{vmatrix} 3 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 7$$

naturally is also symmetrical but, for example, a determinant D_1 is:

$$D_1 = \begin{vmatrix} 4 & 1 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 9$$

already is not symmetrical as $a_{12} \neq a_{21}$.

A component of x_1 of a vector of solutions X is equal to

$$x_1 = \frac{D_1}{D} = \frac{9}{7}$$

In order to compute variations of x_1 that occur due to variations of coefficients a_{ij} and b_{ij} we can apply a methodics given earlier. Its first step is the forming of "tables of signs"for determinants D and D_1. In determinant D the formation of "tables of signs"faciliates the symmetry of a determinant. It is sufficient to compute the most unfavourable signs of variations not for nine but here only for six elements – for elements that are on a main diagonal and that lie higher than it. For elements that are lower than a main diagonal signs are put according to symmetry principle – really if, for example, $a_{12} = a_{21}$ then $\varepsilon_{12} = \varepsilon_{21}$.

As in determinant D we shall have:

$$\begin{array}{ccc} A_{11} = 3; & A_{12} = -1; & A_{13} = -1 \\ & A_{22} = 5; & A_{23} = -2; \\ & & A_{33} = 5 \end{array}$$

then its "table of signs"will be.

$$D \to \begin{vmatrix} + & - & - \\ - & + & - \\ - & - & + \end{vmatrix}$$

For determinant D_1 we have:

$$\begin{array}{ccc} A_{11} = 3; & A_{12} = -3; & A_{13} = 0 \\ A_{21} = -1; & A_{22} = 7; & A_{23} = -3 \\ A_{31} = -1; & A_{32} = -2; & A_{33} = 6 \end{array}$$

and its "table of signs"is

$$D_1 \to \begin{vmatrix} + & - & \pm \\ - & + & - \\ - & - & + \end{vmatrix}$$

By using "table of signs"we can find an interval in the interior of which is x_1 when variations of ε_{ij} and b_i are any. Here we must apply the above methodics. In order not to repeat rather cumbersome computing variations of D and D_1 (in a linear approximation) for $|\varepsilon_{ij}| = \varepsilon_{ij} = 0,01$. For determinant D we shall have:

$$\Delta_{lin} = \Sigma |a_{ij} A_{ij}| \varepsilon_0 = 0,37$$

For D_1 we have:

$$\Delta_{lin} = 0,49$$

These data are sufficient to conclude that variations of x_1 will be small and that an equations system examined in example №19 is well–conditioned. Surely we can compute an exact value of the largest variations of x_1 and find an interval in the interior of which is x_1 during some variations of coefficients.

§13. The application of estimating variations during the computation of solutions of ordinary differential equations.

In previous sections estimates of solving variations in systems of linear algebraic equations were illustrated by examples from construction mechanics, from resistance of materials. Here everything is simple. Problems of computing enforcements and loads in some constructions can be directly reduced to the calculation of components of x_1; x_2; ...; x_n in a vector of solution X of algebraic equations systems of the for m, $AX = B$.

But the necessity to solve such equations also arises in other problems of physics and technique as one of necessary steps of solving a problem as a whole. And in these cases the knowledge of possible variations of each of components of vector X in a system of the form $AX = B$ is a necessary precondition of reliability during the solving of the problem as a whole.

An important example is a problem of computing solutions of ordinary differential equations of different degrees or systems of such equations. One of steps in computing a solution often is a definition of integration constants that enter into a general solution of a differential equation or into the solution of equations system. But in order to define integration constants we usually have to form and solve a system of algebraic equations.

Example №20. It is necessary to find a solution $x(t)$ of a differential equation:

$$\ddot{x} - 3\dot{x} + 2 = 0 \tag{209}$$

that satisfies initial conditions: $x(0) = 1$; $\dot{x}(0) = 0$. A characteristic polynomial of equation (209) is:

$$\lambda^2 - 3\lambda + 2 \tag{210}$$

has roots: $\lambda_1 = 1$; $\lambda_2 = 2$. Therefore a general solution is of the form:

$$x(t) = C_1 e^t + C_2 e^{2t} \tag{211}$$

and thus $\dot{x}(t) = C_1 e^t + C_2 e^{2t}$.

From an initial condition $x(0) = 1$ it follows that

$$C_1 + C_2 = 1 \tag{212}$$

From the second initial condition $\dot{x} = 0$ it follows that

$$C_1 + 2C_2 = 0 \tag{213}$$

Equalities (212) and (213) form a system of two equations for the definition of two integration constants C_1 and C_2.

For large systems of differential equations a degree of a system of algebraic equations which satisfy integration constants C_i can be very large.

So in a text–book [11], p.83 equations of a system of control of a frequency of rotation in an electrodrive were examined. The behaviour of two variables x_1 and x_2 are described by two systems of equations one of which is of a third and another – of a first (in relation to x_2) order. Here x_1 is a deviation of rotation frequency from a rating one and x_2 – deviation of a rating moment from a rating value. General solutions of such a system, x_1 and x_2 are of the form:

$$x_1(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + C_3 e^{\lambda_3 t} + C_4 e^{\lambda_4 t} \tag{214}$$

where λ_1; λ_2; λ_3; λ_4 – roots of a characteristic polynomial – of the fourth degree. Similar form has a solution $x_2(t)$ as well.

In order to define arbitrary constants from initial conditions here we must find a solution of a system consisting of four equations for the unknowns: C_1; C_2; C_3; C_4. It is necessary to take into account that roots of a characteristic polynomial in any system can be calculated only with finite limited exactness and therefore all coefficients of the system of equations from which integration constants have in themselves inevitable errors. In order that a computed solution of differential equations system be reliable it is necessary to conclude (by all means) in what way inevitable errors of computation of integration constants are connected with errors of system coefficients to which these constants satisfy.

This can be done by methods described in previous sections. Since there exist many examples (given in previous sections) when small errors in coefficients lead to large errors in components of a solution x_1; x_2; ...; x_n (and thus – in integration constants as well – C_1; C_2; ...; C_n) then in order to obtain a reliable solution of a differential equations system it is necessary to apply a computations a computation methodics of passible variations in components of a solution stated in previous sections.

Note that if an examined system is turned into a normal Cauchy form – i.e. a system of n equations of the first order

$$\left.\begin{array}{l} \dot{x}_1 = f_1(x_1;\, x_2;\, ...;\, x_n;\, t) \\ \dot{x}_2 = f_2(x_1;\, x_2;\, ...;\, x_n;\, t) \\ ..... \\ \dot{x}_n = f_n(x_1;\, x_2;\, ...;\, x_n;\, t) \end{array}\right\} \tag{215}$$

then on the first sight it seems that we can avoid difficulties in the definition of integration constants. Really, if initial conditions are known for all variables in system (215) values $x(0)$ are known then solutions of system (215) can be directly obtained by one of numerical methods including the use of computational technique. But as it has been shown in [5] if initial differential equations of a mathematical model equations of a mathematical model of an examined object do not have forms of a system consisting of n equations of the first degree but consisted of a system of several equations of different degrees then a fact of turning of such a system to a rating form of Cauchy, to a form of equations (215) can completely misrepresent a real dependence of solutions in an object mathematical model on coefficients variations, to misrepresent a real dependence of examined object behaviour on variations of its parameters. Therefore in order to obtain a reliable solution that correctly reflects a real behaviour of an examined object, to transformations of a mathematical model including to a mathematical model including to a transformation it to a normal form of Cauchy, it is necessary to be careful, to check

its lawfulness, to check whether the influence of variations of object parameters has not been distorted on its true behaviour. About all this it is in details given in [5]. Let us also note that the necessity of composing systems of algebraic equations, to solve and estimate solution errors arises during solving boundary problems when conditions for sought functions and their derivatives are set not in one but in several points (boundary conditions).

When we must form and solve a system of algebraic equations and during the solution of differential equations it is necessary to apply methods of operational computation.

§14. Application to the solution of integral equations.

It is well–known (see,for example, [2]) that a solution of many problems in technique and physics can be reduced to computing solutions of interval equations of different types (equations of Fredholm of the first and second type, Volterra equations of the first and second type, singular equations, equations with a degenerated kernel etc.)

In interval equations a sought function is under an interval. So, in Fredholm equations of the second type

$$y(x) - \lambda \int_a^b K(x;s)y(s)ds = k(x) \tag{216}$$

a sought function will be $y(x)$. A function $f(x)$ is a known function (the right side), function $K(x;1)$ of two variables x and s is called a kernel but a constant λ – parameter of an equation.

The main method of solving integral equations is based on the change of the integral by a finite sum with the help of one of quadrature formulas (formulas of an approximated integration), i.e. on the change

$$\int_a^b F(x)dx \approx \sum_{i=1}^{n} A_j F(x_j) \tag{217}$$

where x_j – abscissas of points in a segment [a; b], $A_j\,(j = 1, 2, ..., n)$ – coefficients of a quadrature formula (rectangle, trapeziums or others). By approximately changing an integral in equation (216) according to formula (217) we obtain:

$$y_i - \lambda \cdot \sum_{j=1}^{n} A_j K_{ij} y_j = k_j \tag{218}$$

where $y_i = y(x_i)$; $K(x_i; y_j)$; $f_i = f(x_i)$ – i.e. we shall obtain a system of linear algebraic equations in relation to y_i. By solving this system we shall obtain a table of approximated values y_i in points x_i. This will allow to write an approximated solution of equation (216) in the form:

$$y(x) = k(x) + \lambda \sum_{j=1}^{n} A_j K(x; x_j) \cdot y_j \tag{219}$$

Let us give an example (taken from [2], page 294) by using Simpson quadrature formula when $n = 2$ we shall find an approximated solution for an equation:

$$y(x) + \int_0^1 xe^{xs} y(s)ds = e^x. \tag{220}$$

Solution: since for Simpson quadrature formula $A_6 = A_2 = \frac{1}{6}$; $A_1 = \frac{2}{3}$; $k_0 = 0$; $x_1 = 0,5$; $x_2 = 1$ then for equation (220) (while using Simpson formula) we can write:

$$y(x) + \frac{1}{6}(xe^{0\cdot x} y_0 + 4xe^{0,5x} y_1 + xe^{1\cdot x} y_2) = e^x \tag{221}$$

Supposing that in this equality in succession $x = x_1 = 0,5; x = x_2 = 1$ we obtain a system of three equations for y_0; y_1; y_2

$$y_0 = 1$$
$$y_1 + \frac{0,5}{6}(y_0 + 4e^{0,25}y_1 + e^{0,5}y_2) = e^{-0,5}$$
$$y_2 + \frac{1}{6}(y_0 + 4e^{0,5}y_1 + ey_2 = e$$

when we have solved it we find that: $y_0 = 1; y_1 = 1,0002;\ y_2 = 0,9995$. After this an approximated solution of equation (220) can be written in the form:

$$y(x) = e^x - \frac{x}{6}(1 + 4,001e^{\frac{x}{2}} + e^x)$$

Other examples of turning integral equations into systems of algebraic equations are given in [2], p.205–303.

So the reliability of solutions in integral equations depends on the reliability of solutions of algebraic equations systems whose all coefficients due to approximation of a change of an integral by a finite sum are known only with finite limited exactness. As now we possess a methodics of an exact estimate of an error (variation) of each of components x_1; x_2; ...; x_n of a vector in solutions X of system $AX = B$ in dependence on errors (variations) of coefficients this allows us to greatly increase the reliability of computing solutions of integral equations and thus – the reliability of solutions of many problems in technique and physics that can be reduced to integral equations.

Not more often numerical methods of solving partial differential equations lead to systems of linear algebraic equations. To them a separate section will be dedicated.

§15. Other criteria of estimating condition degree in systems of linear algebraic equations.

Besides of an algorithm of computing an exact value of an inavoidable error in solving systems of linear algebraic equations stated in previous sections it is useful to have simple criteria for the distinguishing well-conditioned and ill-conditioned systems from each other.

Up to recent years as such a oriterimn "a number of condition" has been used $||A|| \cdot ||A^{-1}||$, i.e. a product of norms of a direct and inverse matrix of a system $AX = B$. A lot of drawbacks of this criterium were considered in §4 and – in more details – in [6].

Therefore let us examine other criteria of a condition degree and, in particular a value of "a natural boundary of variations", i.e. such a value of variations in elements of a matrix A at which its determinant is equal to zero and thus – a value of any components of x_i of solutions X vector can be (in correspondence to a module) infinitely large.

If "a natural boundary of variations" is less than such variations that can appear in the course of exploitation of an object whose mathematical model (in the form of a system $AX = B$) we investigate this means that an examined system is ill–conditioned. Values of any x_i in such a system can be any and an object whose mathematical model in this system is highly unreliable. If there is a possible combination of signs in coefficients variations (if there is a combination of corresponding to "inverse table of signs" matrix A when $detA > 0$ and "a direct table of signs" if $detA < 0$) values x_i can be any. This means that an examined object can break, warp etc. i.e. it can form a dangerous wreckage situation.

If "natural boundary of variations" is a little bigger than such variations that can appear in the course of exploitation such system can be considered ill–conditioned and dangerous since estimates of coefficients variations of a system which can occur in the course of exploitation of an examined object are approximated.

And at last if "a natural boundary of variations" is greatly (by a degree) larger than variations that can occur in the course of exploitation then such a system can be usually considered well–conditioned especially if components of x_i in a vector of solutions X differ from each other in a small degree. A particular case when one or several of components in solutions X vector in much less than others. It requires a separate investigation which will be later given.

The computation of "a natural boundary of variations" is rather cumbersome. First of all it is necessary to form "an inverse table of signs" for matrix A if $detA > 0$ and "a direct table of signs" if $detA \leq 0$ and then we must find values ε that turn a matrix determinant into zero by using a corresponding to a sign A a table.

For an approximated estimate of "a natural boundary of variations" (estimates in a linear approach) we can apply a formula for the largest value of a main linear part of the growth (decrease) of a determinant:

$$\Delta_{linmax} = \sum_{i=1;j=1}^{i=n;j=n} |a_{ij}A_{ij}|\varepsilon_0 \tag{222}$$

In paragraph 6 you can see how formula (222) can be found. From formula (222) it follows that an examined determinant will turn into zero when

$$\varepsilon_{0lin} = \frac{detA}{\sum_{i=1;j=1}^{i=n;j=n} |a_{ij}A_{ij}|} \tag{223}$$

Example №21 Let us compute ε_{0lin} for an earlier examined determinant (94) for which $A_{11} = -3$; $A_{13} = -14$; $A_{13} = 13$; $A_{21} = 2$; $A_{22} = -1$; $A_{23} = 2$; $A_{31} = 1$; $A_{32} = 10$; $A_{33} = -7$ since

$$\sum_{i=1;j=1}^{i=3;j=3} |a_{ij}A_{ij}| = 161$$

and thus,

$$\varepsilon_{0lin} = \frac{8}{161} = 0,0496.$$

An earlier computed value ε_0 is equal to $0,0525$ if we take into account nonlinear members (with the exactness of three significant numbers). The divergence is small.

This computation shows that for all equations $AX = B$ in which a determinant of matrix A coincides with determinant (94) (if for all i and j we have $(\varepsilon_{ij} \leq \varepsilon_0)$ a number $0,0525$ will be "a natural boundary of variations".

When $\varepsilon_0 = 0,01$ the most possible decrease of determinant (94) (as it has been computed earlier – in §6) will be equal to 20,29% of a rating value.

An approximated value of "a natural boundary of variations" can be easily computed according to formula (223). It can be applied for a preliminary estimate of condition degree of different equations systems in order to compare them between themselves etc. – for the same aims for which up to now "a number of condition" was used.

In order to obtain the best analogy with a usual "number of condition" we can apply a number $\frac{1}{\varepsilon_{0lin}}$ that is inverse to "a natural boundary of variations" in order to estimate the conditioning of systems. Then everything will be as usual. The less is a number the better is the conditioning of an examined system. It is just the same when we use usual numbers of condition: the less they are then better is the conditioning.

But an estimate by "a natural boundary of variations" better reflects real qualities of the system than when we apply a usual estimate by means of "a number of condition".

Let us return to system, earlier examined system (62) for which in §4 it has been shown that "numbers of condition" falsely reflect a real dependence of a system on a coefficient m.

For a determinant of matrix A in system (62), i.e. for a determinant

$$\begin{vmatrix} 1+m & 1 \\ 1 & 1 \end{vmatrix} = m \tag{224}$$

we have: $A_{11} = 1; A_{12} = -11; A_{21} = -1; A_{22} = 1 + m$.

Thus,

$$\sum_{i=1;j=1}^{i=2;j=2} |a_{ij}A_{ij}| = 4 + 2m \tag{225}$$

$$\varepsilon_{0lin} = \frac{1}{2} \cdot mm + 2 \tag{226}$$

i.e. "a natural boundary of variations" in a linear approximation, a value ε_{0lin} is equal to a half of a tangent of an angle φ between straight lines whose equations have been written in the form of a system (62). While there is the growth of an angle a degree of condition monotonously grows. In the same way ε_{0lin} also monotonously grows. Therefore a criterium ε_{0lin} (in difference from "number of condition") it correctly reflects a real dependence of a degree of conditioning on coefficient m. And as it has been shown in §4 estimates by a number of condition gives for $m > 2$ an incorrect answer.

For system (62) it is not difficult to compute an exact value of ε_{0b}. It is

$$\varepsilon_{0b} = \frac{1}{m}(m + 2 - 2\sqrt{m+1})$$

and it monotonously increases as well with the increase of m. The dependence of ε_{0lin} and ε_{0gr} are reflected in table 4.

Table 4.

m	0,5	1	2	4	10
ε_{0b}	0,101	0,172	0,268	0,382	0,5366
ε_{0lin}	0,1	0,166	0,25	0,333	0,418
$\frac{\varepsilon_{0lin}}{\varepsilon_{0b}}$	0,99	0,972	0,934	0,875	0,774

One more drawback of "numbers of condition" is their dependence on a multiplication of any of equations in a system by a constant number. Such a multiplication does not change solutions and is often applied for the simplification of the system. In §4 on an example of a simple system

$$\left.\begin{aligned} 2x_1 + x_2 &= 1 \\ kx_1 + kx_2 &= k \end{aligned}\right\} \tag{227}$$

it has been shown that a "number of condition" for it is

$$||A|| \cdot ||A^{-1}|| = \frac{5}{k} + 2k$$

for it greatly depends on k.

At the same time "a natural boundary of variations" – as an exact one or – in a linear approximation – does not depend on k and therefore it is much a better way characterizes

a degree of condition of real systems.

Really, for systems (227) we shall have: $detA = k; A_{11} = k; A_{12} = -k; A_{21} = -1; A_{22} = 2$ and thus

$$\varepsilon_{0lin} = \frac{k}{2k + k + k + 2k} = \frac{1}{6} = 0,166$$

an exact value of "natural boundary of variations" is equal to $\varepsilon_{0b} = 3 - \sqrt{8} = 0,172$ for all k.

Neither an exact nor an approximated (in linear approximation) values in "natural boundary of variations" (in difference with "number of condition") do not depend on equivalent transformations of any equations in an examined system. Neither they depend on the choice of measurement units. It has already been said in §4 about this dependence for "numbers of condition" which can lead to the delusion.

As a whole we can draw the following conclusion. "A natural boundary of variations" – it is a number which is a good simple criterium for estimating conditioning of linear algebraic equations systems (SLAE). This criterium is free from many drawbacks characteristic of known "members of condition".

Although we need not refuse from computing an exact value of errors of solutions in linear algebraic equations systems we can use this criterium for preliminary estimates of conditioning of SLAE. It is convenient.

In order to compute "a natural boundary of variations" in a linear approximation it is sufficient to compute algebraic additions of elements in matrix A vector and then we can use formula (223). In order to carry out an exact computation of this boundary (in a linear approach) it is sufficient to form "a table of signs" of determinant D on the basis of computed values of algebraic additions A_{ij} of elements of a determinant D if $D > 0$ and "an inverse table of signs"if $D > 0$. With the help of these tables an exact value of "a natural boundary of variations" is computed, i.e. a value ε_0 at which for the first time determinant D turns into zero.

"A natural boundary of variations" can be computed (as in previous sections) as for independent from each other elements of variations in determinant as for variations in determinant as for variations that are connected by dependencies as well – for example, for symmetrical matrixes A in equations $AX = B$.

Addition

Besides "natural boundary of variations" it is useful to compute a value of variations that change a sign of a component x_i of solutions X vector in a system $AX = B$. Surely you can do this if these variations are less than "a natural boundary of variations". In many SLAE sensibility to variations of elements of matrix A in system $AX = B$ for different x_i is different. Usually such x_i are especially sensible which are smaller than others by an absolute value. It is not difficult also to compute a value of variations by the same method as the "natural boundary of variations" has been computed. Then a solution that interests us x_i changes in such a way that it turns into zero. And if variations of coefficients in system $AX = B$ later grows a solution x_i changes its sign. For this it is

sufficient to compute a "table of signs" for determinant D_i in Cramer formulas (if $D_i < 0$) or – to compute "an inverse table of signs" if $D_i > 0$. After this we can easily compute variations that turn D_i into zero.

If we return to system (49) that has been considered earlier in §4 we see that although "a natural boundary of variations" for system (49) is equal to $\varepsilon_{gr} = 0,10102$ but when $\varepsilon = 0,0222$ solution x_1 turns out into zero. But if $\varepsilon > 0,0222$ it changes a sign. A solution x_2 in comparison with x_1 is less sensible to variations of matrix A elements. At the same time if $0,0222 \leq \varepsilon \leq 0,10102$ we see that solution x_1 will greatly change by more than 100%. Therefore all characteristics of an examined object that depend on x_1 can turn out to be quite different than if $\varepsilon = 0$.

Taking into account the above said we can recommend to use as the first step the investigation of "a natural boundary of variations" of determinants D and D_i. This method will allow us to depict very ill–conditioned dangerous systems and then it would not be necessary to take unnecessary computations.

At the same time it is necessary to note that if components of x_i in solutions A vector greatly differ from each other then "a natural boundary of variations" (as well as "a number of condition") do not allow to estimate "a degree of condition" of each of components of x_i in solutions X vector and especially – for such x_i that are smaller than others.

Let us consider a system of equations:

$$\begin{array}{c} 2x_1 + x_2 = 2 \\ x_1 + x_2 = 1,5 \end{array} \tag{228}$$

with solutions $x_1 = 0,5; x_2 = 1$.

In this system a determinant of matrix A is equal to:

$$\begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 1$$

its "inverse table of signs" is:

$$\begin{vmatrix} - & + \\ + & - \end{vmatrix}$$

and in order to compute "a natural boundary of variations" it is sufficient to compute at what value ε the following determinant will turn into zero:

$$\begin{vmatrix} 2(1-\varepsilon) & 1+\varepsilon \\ 1+\varepsilon & 1-\varepsilon \end{vmatrix} = 1 - 6\varepsilon + \varepsilon^2 \tag{229}$$

By computing the smallest of solutions $\varepsilon_1; \varepsilon_2$ of an obtaibed square equation we shall find that "a natural boundary of variations" is equal to: $\varepsilon_{0b} = 3 - \sqrt{8} = 0,172$.

If we examine solutions of an investigated system when $|\varepsilon| = 0,01$ that is by 17,2 times less than "a natural boundary of variations" we can expect that components of x_1 and

x_2 in solutions X vector will change little. In fact if $\varepsilon = 0,01$ an examined system will become:

$$\left.\begin{array}{r} 1,98x_1 + 1,01x_2 = 2 \\ 1,01x_1 + 0,99x_2 = 1,5 \end{array}\right\} \tag{230}$$

and it has the following solutions:

$$x_1 = 0,4947;\ x_2 = 1,009564$$

Thus a solution x_1 has decreased by 1,06% and a solution x_2 has increased by 0,9564%.

Now let us examine (for comparison) an analogous system of equations (the same matrix A but right sides are different) – a system

$$\left.\begin{array}{r} 2x_1 + x_2 = 1,02 \\ x_1 + x_2 = 1,01 \end{array}\right\} \tag{231}$$

with solutions $x_1 = 0,01; x_2 = 1$.

With the same value of $\varepsilon = 0,01$ this system turns into a system:

$$\left.\begin{array}{r} 1,98x_1 + 1,01x_2 \\ 1,01x_1 + 0,99x_2 \end{array}\right\} \tag{232}$$

Thus if $\varepsilon = 0,01$ already a solution x_1 has changed by 214% and has even changed a sign. If for an object whose mathematical mode; is system (231) variations of coefficients corresponding to $\varepsilon = \pm 0,01$ are possible such an object (and a system (230) that describes it as well) must be attributed to an il–conditioned one.

Note that "a number of condition" $||A|| \cdot ||A^{-1}||$ in system (231) is equal to seven.

Thus a traditional estimate by means of "a number of condition" would attribute system (231) to a well-conditioned system. Surely it is not correct.

Therefore although an estimate by "a natural boundary of variations" is free from drawbacks of "numbers of condition" it shares with "numbers of condition" such a drawback as the possibility of incorrect estimate of some components of x_1 in a vector of solutions X. Therefore an algorithm for computing an inavoidable error in each of components of x_i of vector X (given in §12 and earlier in publications [6] and [22]) is so important.

One more method of estimating a degree of condition is a "method of modular determinants" (see [6] and [21]). It is known that any determinant of an order n can be presented in the form of a sum $n!$ of products of n elements in a determinant taken in a certain order. So a determinant of the second order

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

is a sum $a! = 2$ of products of two elements, and a determinant of the third order

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

is equal to a sum of six products (since $3! = 6$) from the elements. A determinant of the fourth order is equal to a sum of 24 products (since $4! = 24$) from four elements each etc.

Naturally signs of products depend on signs of elements. If these signs have not been taken into account and if we think that a determinant is a sum of products modules then such a determinant has been proposed to be called a modular determinant (see [6]). So a modular determinant of the second order is of the form:

$$|a_{11}a_{22}| + |a_{12} + a_{21}|$$

and a modular determinant of the third order is equal to

$$|a_{11}a_{22}a_{33}| + |a_{12}a_{23}a_{31}| + |a_{13}a_{21}a_{32}| + |a_{13}a_{22}a_{31}| + |a_{11}a_{23}a_{32}| + |a_{12}a_{21}a_{33}| \qquad (233)$$

In an anologous way modular determinants of the fourth, fifth and other degrees are written.

With the help of modular determinant conditioning of equations systems and a value of inavoidable errors has been carried out (see [6] and [22]). For systems of linear equations of a moderated order a method of modular determinants works well but for systems consisting of a large number of equation computing difficulties occur. A value of $n!$ very quickly (as n^n) grows with the increase of n. If $n = 10$ we shall already have $10! = 3628800 > 3 \cdot 10^6$ and when $n = 20$ we shall have $20! > 2 \cdot 10^{18}$.

Therefore in order to compute determinants (if $n > 7$) we do not say that a determinant is a sum of $n!$ products consisting of n elements but Gauss method is applied. It is similar to computation method of solving SLAE, i.e. a method of successive multiplications and additions leads to a determinant of "a triangular type" after which it is computed as a product of elements that are on a main diagonal. In order to compute a determinant of an order n it approximately requires n^3 multiplications.

But to modular determinants Gauss method cannot be directly applied and a direct computation of $n!$ is too cumbersome. While making the first acquaintance with "modular determinants" in 2007 mathematicians – people that make computations suppose that they can quickly develop an analogue of Gauss method for their computation. But these problems have turned up to be more difficult than it was before supposed. And up to now this problem has not been solved.

Therefore in this book we do not give a methodics for modular determinants. If a convenient method of their computation will be developed then they will find a wide application. And then it will be possible to make acquaintance with the method of modular determinants in publications [6] and [21].

Now let us examine an estimate of a degree of condition for such systems $AX = B$ in which it is necessary to take into account only variations of a right side coefficients. And variations of coefficients in matrix A can be thought to be equal to zero. Then we can write the system in the form:

$$\left.\begin{array}{r} a_{11}a_1 + a_{12}a_2 + ... + a_{1n}x_n = b_1(1 \pm \delta_1) \\ a_{21}a_1 + a_{22}a_2 + ... + a_{2n}x_n = b_2(1 \pm \delta_2) \\ ..... \\ a_{n1}a_1 + a_{n2}a_2 + ... + a_{nn}x_n = b_n(1 \pm \delta_n) \end{array}\right\} \tag{234}$$

As it has been shown in §5 in such systems any component of x_1 of solutions X vector can be computed by means of a formula

$$x_i = \sum_{j=1}^{n} K_j b_j, \tag{235}$$

where K_j is a quotient as a result of dividing algebraic addition A_{ij} by a determinant of matrix A. At the same time we can compute a maximally possible value of an inavoidable error by means of a formula:

$$\Delta x_i = \sum_{j=1}^{n} ||K_i b_j \delta_j| \tag{236}$$

and in relation to δ_i only general estimate is known:

$$|\delta_j| \leq \delta_0$$

then according to the formula we have:

$$\Delta x_i = \sum_{j=1}^{n} ||K_i b_j| \delta_0 \tag{237}$$

Since there are possible combinations of coefficients signs and algebraic additions A_{ij} then in a sum (235) a compensation of positive and negative members is possible. Therefore an absolute value x_i can be very small–small in such a way that an inavoidable error Δx (even if δ_i is small) can be compared with and it can be even larger than x_i. Here a probability of such a combination of signs will not be a very small value.

Hence an important conclusion can be drawn. Even if we take into account only variations (or errors) in the right side of system $AX = B$ errors δ_i of coefficients b_i then there is always a possibility that we meet with very ill–conditioned systems in which an inavidable error of separate components in x_i of solutions X vector can be compared (or even it can be larger) with a component itself of x_i.

It is clear that a conventional estimate by means of "a number of condition" $||A|| \cdot ||A^{-1}||$ will not point nothing about the meeting of such type of ill–conditioned systems. The test by means of "number of condition" can note that the system is well–conditioned, that "everything is well". But at the same time some of x_i can turn out to be completely unreliable.

At the same time an ill–condition of solutions in system $AX = B$ that is due to variations (errors) of their right side can be easily found as a value itself x_i and a value of its inavotable error as well are computed by simple formulas. It is sufficient to test whether there exists in an examined system $AX = B$ such x_i for which a value $|\delta x_i|$ can be compared (or larger than) with $|x_i|$. We can compute it while computing x_i and δx_i.

A generalization to a general case. We have considered a particular case of checking a degree of condition while supposing that variations (errors) of coefficients in matrix A of a system $AX = B$ are equal to zero. But since the account of coefficients variations in matrix A can only worsen a degree of condition this conclusion remains effective in a general case as well. If among components of x_i in vectors of solutions X we shall find such x_i for which a value $|\Delta x_i|$ that is calculated by means of formulas (236) or (237) turns out to be comparable or larger than $|x_i|$ itself then such x_i is ill–conditioned and is quite unreliable.

§16. An estimate of difficulties in computing algorithms for an exact value of an inavoidable error in SLAE. Examples of computations.

In this section we shall estimate a difficulty in computing of a proposed algorithm for the computing of an inavoidable error, i.e. we shall estimate a quantity of operations necessary for computing an error in SLAE in dependence on n where n is a number of equations in a system. The first step in an algorithm of computing an error in solution x_i is the forming of "a table of signs" for determinants D and D_i in formulas by Cramer. In order to form "a table of signs" it is necessary to compute (or in the least – to estimate signs) of all algebraic additions A_{ij} in two determinants – D and D_i. A number of these additions is equal to $2n^2$ and each of them is a determinant of the order $n-1$.

It is well–known that when $n \leq 4$ determinants are most often computed by means of reducing to a triangular form. This computation requires (approximately, we drop a numerical coefficient) n^3 operations of multiplication (where n is an order of a determinant). Therefore the formation of "table of signs" for determinants D and D_i in Cramer formulas requires $2n^2(n-1)^3$ operations.

Later in order to noice the change of x_i (if ε is set) in the direction of an increase it is necessary to twice compute variated determinants D and D_i (according to the first and the second variant of computation) and then it is necessary to carry out the same operation for the estimate of a change in x_i in the direction of a decrease. In all it is necessary to carry out approximately

$$2n^2(n-1)^3 + 8n^3 \tag{238}$$

operations. If it is necessary to carry out estimates of all components of x_i in vector of solutions X from $i = 1$ up to $i = n$ then a number of operations will increase up to

$$2n^2(n-1)^3 + 8n^4 \tag{239}$$

Formulas (238) and (239) show that a number of computation operations with the increase of a number of equations in a system increase not very quickly. It is proportional to a polynomial in a variable n. It is a very favourable circumstance for a practical application of a proposed algorithm, especially, in comparison with many other algorithms in which a number of computing operations increases exponentially, proportional to 2^n, 4^m and even 2^{n^2+n}. About these algorithms see [16], p.106.

Note that for a particular case when variations of matrix A elements in system $AX = B$ are equal to zero or are so small in such a way that can be ignored then an important role play variations in the right side and variations of vector B and a number of necessary computing operations greatly reduces. In this particular case it is necessary to compute only n algebraic additions (or minors) that will approximately require

$$n(n-1)^3 \tag{240}$$

operations.

A number of necessary computing operations also decreases when variations of coefficients in SLAE are not independent (symmetrical matrixes A that are considered in §11) etc.

Note that from formulas (239) and (240) it follows that the computing of exact values of errors in all components of a vector solutions X in system $AX = B$ requires much more computing operations than the computation of a vector itself X if values of coefficients in matrix A and vector B are nominal (if we do not take into account their errors). But we must not be surprised by this fact since the computation of errors in a solution almost always turns out to be more complex than the computation of a solution itself. There is a following general rule. Problems in "mathematics–2" are always more complex than analogical problems in "mathematics–1" . Just because of this "mathematics–2" has developed later and more slowly than "mathematics–1" . A lot of its sections up to now has not been developed or were not sufficiently developed.

Examples of computing an inavoidable error

Example №22

For one of known systems by A.Newmair i.e. a system $AX = B$ in which

$$A = \begin{pmatrix} 8,5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 8,5 & 1 & 1 & 1 & 1 \\ 1 & 1 & 8,5 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 8,5 & 1 \\ 1 & 1 & 1 & 1 & 1 & 8,5 \end{pmatrix} \tag{241}$$

$$B = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \tag{242}$$

the computation of an error of a solution during variations of all coefficients in the limits of $\pm 0,01$ from their rating values has been carried out by V.V.Lapitzki on a personal computer, a language C++ in a medium Visual Studio 2010 Express. Note that A.Neumair systems of different orders has been repeatedly used as tests problems. They have been examined in [16], p.105.

It has been stated that due to a system symmetry all x_{iN} are equal to each other and we have if there are rating values: between themselves and when we have rating values of coefficients we have $x_{iN} = 0,0741$. But at the same time when $\varepsilon_0 = \pm 0,01$ we shall have $x_{imax} = 0,0741 + 0,0032; x_{min} = 0,0741 - 0,0032$ or

$$\frac{x_{imax}}{x_{iN}} = 1 + 0,0432; \ \frac{x_{imin}}{x_{iN}} = 1 - 0,0432 \tag{243}$$

i.e. if coefficients have changed by 1% solutions have changed by 4,32%. The time of computing x_{min} and x_{max} is less than 1 second.

"Table of signs" of matrix (241) determinant is diagonal one:

$$\begin{vmatrix} + & - & - & - & - & - \\ - & + & - & - & - & - \\ - & - & + & - & - & - \\ - & - & - & + & - & - \\ - & - & - & - & + & - \\ - & - & - & - & - & + \end{vmatrix}. \tag{244}$$

A determinant of matrix A is equal to $detA = 320361$. Algebraic additions of matrix A are equal to: $A_{ij} = 39550$ if $i = j$ and $A_{ij} = -3164$ if $i \neq j$. A natural boundary of variations (in a linear approximation) is equal to:

$$\varepsilon_{0b} = \frac{detA}{\sum_{i=1;j=1}^{i=6;j=6} a_{ij}A_{ij}} = \frac{320361}{334220} = 0,959$$

This means that system (241)–(242) is well–conditioned during variations of coefficients that do not exceed $\varepsilon_0 = 0,01 = 0,01043\varepsilon_{ob}$ variations of solutions will not be large (although they are by 4,32 times more than coefficients variations)

By computing Euclid norms of matrix A and an inverse matrix A^{-1} we shall obtain "a number of condition" $||A|| \cdot ||A^{-1}|| = 20,4$. By using formula (48) from §3 we shall obtain an estimate by "a number of condition"

$$\frac{||\Delta X||}{||X||} = 20,4 \cdot (0,01 + 0,01) = 0,408 \tag{245}$$

But in fact $\dfrac{||\Delta X||}{||X||} \leq 0,0432$ so that an estimate by a "number of condition" turns out to be rough.

Example №23

As an example №23 Neumair system of the 10th order has been considered, i.e. system $AX = B$ in which

$$A = \begin{pmatrix} 11,7 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 11,7 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 11,7 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 11,7 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 11,7 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 11,7 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 11,7 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 11,7 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 11,7 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 11,7 \end{pmatrix} \tag{246}$$

$$B = \begin{pmatrix} 0,8 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0,8 \\ 0,8 \end{pmatrix} \tag{247}$$

With the help of the same computer V.V.Lapitzki has computed a solution x_i for rating values of coefficient in matrix (246) and vector (247) for system (246)–(247). Due to system symmetry all x_i were equal to $x_{iN} = 0,0386$. If there were relative variations of all system coefficients equal to $\pm 0,01$ a computed maximal error in solutions in the direction of an increase has been 0,018 and in the direction of a decrease – 0,0019. Thus

$$\begin{gathered} 0,0386 - 0,0019 \leq x_i \leq 0,0386 + 0,0018 \\ 1 - 0,0494 \leq \frac{x_i}{x_i N} \leq 1 + 0,0467 \end{gathered} \tag{248}$$

Computations have taken not less than three minutes. "A table of signs" in system (246)–(247) turned out to be a diagonal one:

$$\begin{vmatrix} + & - & - & - & - & - & - & - & - & - \\ - & + & - & - & - & - & - & - & - & - \\ - & - & + & - & - & - & - & - & - & - \\ - & - & - & + & - & - & - & - & - & - \\ - & - & - & - & + & - & - & - & - & - \\ - & - & - & - & - & + & - & - & - & - \\ - & - & - & - & - & - & + & - & - & - \\ - & - & - & - & - & - & - & + & - & - \\ - & - & - & - & - & - & - & - & + & - \\ - & - & - & - & - & - & - & - & - & + \end{vmatrix} \tag{249}$$

Algebraic additions to matrix (246) turned out to be equal $A_{ij} = 3,38 \cdot 10^9$ if $i = j$; $A_{ij} = -0,171 \cdot 10^9$ if $i \neq j$. Determinant of matrix A has turned out to be equal to $detA = 38,056 \cdot 10^9$.

A natural boundary of variations (in a linear approach) is

$$\varepsilon_{0b} = \frac{detA}{\sum_{i=1;j=1}^{i=10;j=10} |a_{ij} A_{ij}|} = \frac{38,056 \cdot 10^9}{49,2 \cdot 10^9} = 0,773 \tag{250}$$

Examined variations of coefficients in system (246)–(247) that are equal to $\frac{1}{100}$ of their rating values ($\varepsilon = 0,01$) have turned out to be by almost a degree less than a natural boundary of variations. Therefore system (246)–(247) in relation to variations $\varepsilon = 0,01$ turns out to be well–conditioned. As formula (248) shows that errors in solutions are not large (but at the same time variations of solutions turn out to be much larger than variations of system coefficients by 4,5 times).

By computing Euclid norms of matrix A:
$||A|| = \sqrt{10 \cdot 11,82 + 90 \cdot 1} = 38,2$ and an inverse to it matrix $A^{-1} : ||A^{-1}|| = 2,82$ we find a "number of condition" is $||A|| \cdot ||A^{-1}|| = 108$. By using formula (48) from §3 we obtain an estimate:

$$\frac{||\Delta X||}{||X||} \leq 108 \cdot (0,01 + 0,01) = 2,16 \tag{251}$$

In fact if we take into account inequality (248) we have:

$$\frac{||\delta X||}{||X||} \leq 0,0494, \tag{252}$$

i.e. an estimate by a "number of condition" turns out (as expected) to be very rough.

Example №24

A system is considered:

$$\left.\begin{array}{r} 20,9x_1 + 1,2x_2 + 2,1x_3 + 0,9x_4 = 21,7 \\ 1,2x_1 + 21,2x_2 + 1,5x_3 + 2,5x_4 = 27,46 \\ 2,1x_1 + 1,5x_2 + 19,8x_3 + 1,3x_4 = 28,76 \\ 0,9x_1 + 2,5x_2 + 1,3x_3 + 32,1x_4 = 49,72 \end{array}\right\} \tag{253}$$

The solution of this system by iteration method has been examined earlier in [2], p. 80 (without computing an inavoidable error).

Here are components of solutions vector in system (243): $x_1 = 0,8047; x_2 = 1,0171; x_3 = 1,2082; x_4 = 1,2478$.

Determinant of matrix A in system (253) is:

$$detA = 271554,18$$

An inverse matrix A^{-1} in system (253) is:

$$\begin{pmatrix} 0,0488 & -0,0023 & -0,0049 & -0,0010 \\ -0,0012 & 0,0479 & -0,0033 & -0,0036 \\ -0,0102 & -0,0031 & -0,0008 & 0,0317 \end{pmatrix} \tag{254}$$

Euclid norm of matrix A is equal to:

$$||A|| = \sqrt{a_{11}^2 + a_{12}^2 + ... + a_{nn}^2} = 48,88 \tag{255}$$

Euclid norm of an inverse matrix is:

$$||A^{-1}|| = 0,0923 \tag{256}$$

A number of condition is:

$$||A|| \cdot ||A^{-1}|| = 4,51 \tag{257}$$

(computations have been carried out by Voloshin M.V.)
Table of signs in matrix A is:

$$\begin{vmatrix} + & - & - & - \\ - & + & - & - \\ - & - & + & - \\ - & - & - & + \end{vmatrix} \tag{258}$$

An inverse table of signs is:

$$\begin{vmatrix} - & + & + & + \\ + & - & + & + \\ + & + & - & + \\ + & + & + & - \end{vmatrix} \tag{259}$$

Basing ourselves on "tables of signs" (258) and (259) and while using an algorithm for computing an exact value of an inavoidable error given in §12 and supposing that relative variations of all coefficients in system (253) do not exceed values of $\pm\varepsilon$ it has been computed that

for $\varepsilon = 0,01$:

According to the first variant of computation we have obtained: $x_{1min} = 0,8052; x_{2min} = 1,0179; x_{3min} = 1,2079; x_{4min} = 1,2451$.

According to the second variant of computing we have: $x_{1min} = 0,8044; x_{2min} = 1,0169; x_{3min} = 1,2079; x_{4min} = 1,2441$.

Therefore for a system (243) if $\varepsilon = 0,001$ minimal values of all four components of solutions vector are achieved in the second variant of computing which finally defines minimal values of all x_i.

In a similar way maximal values of x_i are computed as well if $\varepsilon = 0,001$ (they again are achieved in the second variant of computation).

If we repeat computations for $\varepsilon = 0,001; \varepsilon = 0,002; \varepsilon = 0,01; \varepsilon = 0,02$ we finally obtain a summary table of value x_{imin} and x_{imax} for different ε:

Table 5.

ε	0	0,001	0,002	0,005	0,01	0,02
x_{1min}	0,8047	0,8044	0,8041	0,8030	0,8012	0,7977
x_{2min}	1,0171	1,0169	1,0168	1,0163	1,0155	1,0139
x_{3min}	1,2082	1,2079	1,2076	1,2067	1,2053	1,2033
x_{4min}	1,2478	1,2441	1,2404	1,2293	1,2109	1,1746
x_{1max}	0,8047	0,8051	0,8054	0,8065	0,8082	0,8116
x_{2max}	1,0171	1,0173	1,0174	1,0179	1,0187	1,0203
x_{3max}	1,2082	1,2085	1,2088	1,2096	1,2111	1,2139
x_{4max}	1,2478	1,2515	1,2552	1,2664	1,2852	1,3232

Let us also give values of absolute and relative errors of different components in solutions vector, i.e. values $x_{max} - x_{min}$ and $\dfrac{x_{imax} - x_{imin}}{x_{i\varepsilon=0}}$ for $\varepsilon = 0,001$:

$$x_{1max} - x_{1min} = 0,007; \frac{x_{1max} - x_{1min}}{x_{1\varepsilon=0}} = \frac{0,007}{0,8047} = 0,87\% \tag{260}$$

$$x_{2max} - x_{2min} = 0,0032; \frac{x_{2max} - x_{2min}}{x_{2\varepsilon=0}} = \frac{0,0032}{1,0171} = 0,315\% \quad (261)$$

$$x_{3max} - x_{3min} = 0,0058; \frac{x_{3max} - x_{3min}}{x_{3\varepsilon=0}} = \frac{0,0058}{1,2082} = 0,47\% \quad (262)$$

$$x_{4max} - x_{4min} = 0,0743; \frac{x_{4max} - x_{4min}}{x_{4\varepsilon=0}} = \frac{0,0743}{1,2478} = 6\% \quad (263)$$

Given numbers once more show that absolute and relative variations of different components x_i of solutions X vector can greatly differ from each other and therefore often an applied estimate by means of "a number of condition" and according to a value in relation $\frac{||\Delta X||}{||X||}$ can give incorrect recommendations. So for system (243) if $\varepsilon = 0,01$ a variation of x_1 is by 6,92 times more than the largest of variations of $x_1; x_2$ or x_3.

All this is clearly seen in fig.6 where the dependences $x_{1max}; x_{1min}$ and also $x_{4max}; x_{4min}$ on coefficients variations in system ε are shown.

For any system $AX = B$ dependences of x_{imax} and x_{imin} on ε functions of a "rational" fraction – i.e. a quotient of two polynomials. And here a numeral and a denominator as well are polynomials of a degree n of a variable ε. But this is true if all coefficients of matrix A and vector B have relative or absolute variations equal to $\pm\varepsilon$. If variations of some coefficients are equal to zero then degrees of polynomials in variable ε in a numerator and denominator can be less than n. Dependences of x_{max} and x_{min} on ε are continuous functions except such values of ε at which a denominator – i.e. a determinant of matrix A if we take into account variations of its coefficients – turns out into zero.

If a maximal value of ε for which dependences x_{max} and x_{min} on ε are formed is small in comparison with "a natural boundary of variations" then dependences x_{max} and x_{min} on ε are near to straight lines (Fig. 6 reflects it):

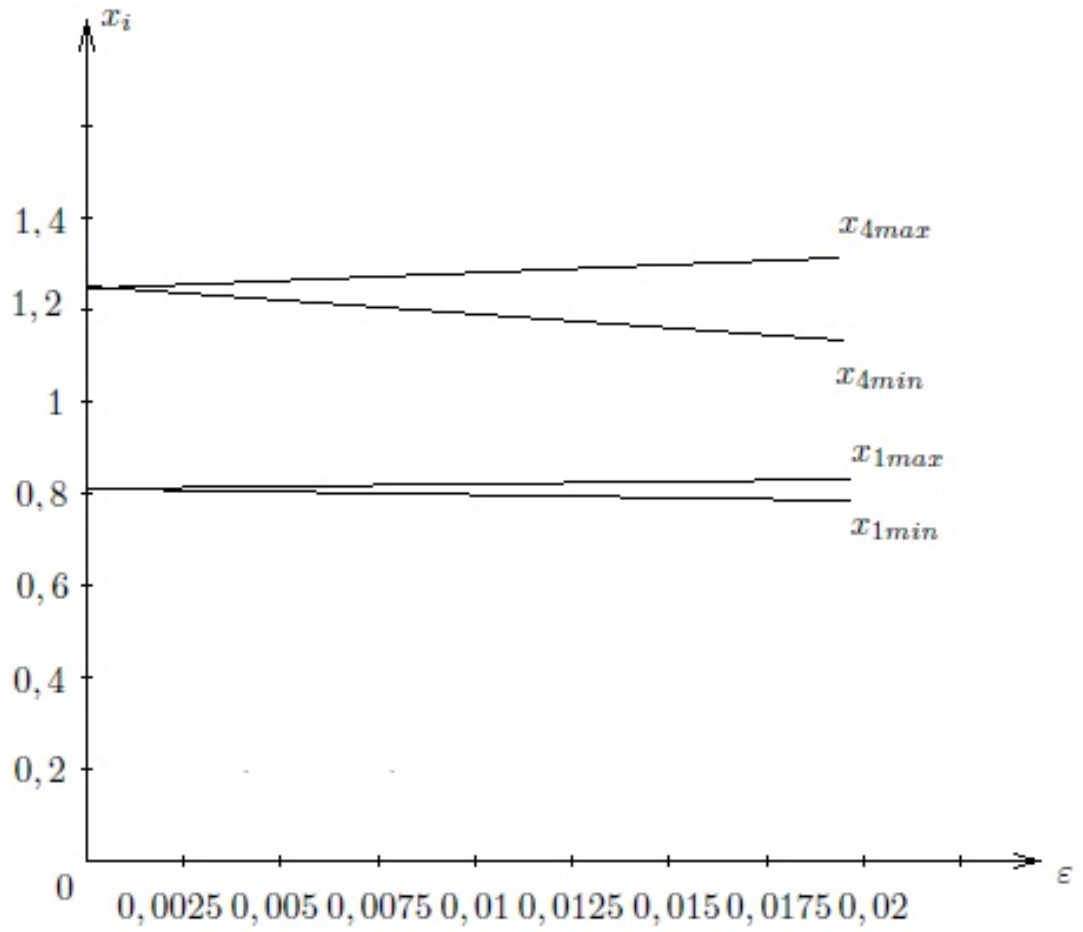

fig. 6

If a maximal value of ε for which dependences x_{max} and x_{min} on ε are formed, i.e. it can be in accordance with "a natural boundary of variations" (and this means that the system is ill–conditioned) then dependence of x_{max} and x_{min} on ε can be more complex.

Example №25

Let us consider a system:

$$\left.\begin{array}{r} 14x_1 + 12x_2 + 15x_3 = 3 \\ 12x_1 + 16x_2 + 12x_3 = 5 \\ 5x_1 + 4x_2 + 6x_3 = 1 \end{array}\right\} \tag{264}$$

that earlier has been examined in §9 where it is stated that for it a number of condition $||A||\cdot||A^{-1}|| = 100,12$ is a natural boundary of variations $\varepsilon_{ob} = 0,018525$ and $\dfrac{1}{\varepsilon_{0b}} = 53,98$.

Dependences of x_{3max} and x_{3min} on ε are shown on figure 7. Here it is at once seen that already for $\varepsilon = 0,01$ system (264) is very ill–conditioned in relation to a component of solutions x_2 vector. x_{3max} (if $\varepsilon = 0,01$) is by six times more than an initial value of x_{3max} if $\varepsilon = 0$ and x_{3min} has altogether changed its sign and has become negative. The computation of values of x_{3max} and x_{3min} when $\varepsilon > 0,01$ has already no practical sense.

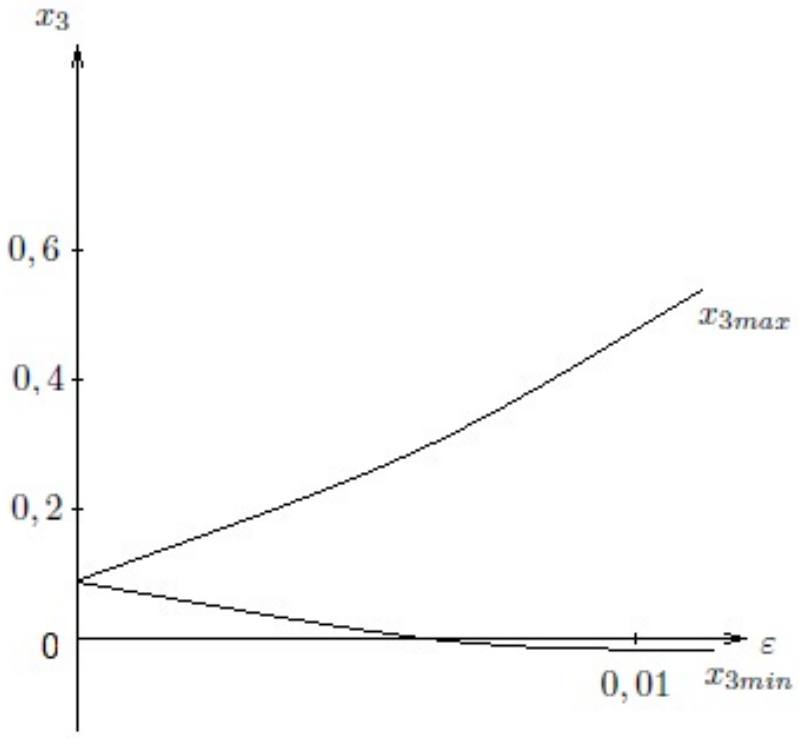

fig. 7

At the same time a component of a vector of solutions x_2 (if $\varepsilon = 0,01$) is conditioned much better. As shows the computation that has been earlier carried out in [6], p. 122–126 (when $\varepsilon = 0,01$) a value of x_{2max} is only by 34,97% more than an initial value if $\varepsilon = 0$ and a value of x_{2min} is only by 11,4% less than an initial value computed for rating values of coefficients in (264).

§17. Comparison with a methodics of interval analysis.

In books dedicated to interval analysis (see, for example, [3,4,15]) it is stated that there is a possibility of applying this methodics in order to find an inavoidable error in solutions of linear algebraic equations systems that occurs due to variations of system coefficients. The statement of solving a problem little differs from the one examined by us in previous sections. In interval analysis such intervals are set in the interior of which are coefficients of matrix A and vector B of system $AX = B$ and such intervals are searched in the interior of which are components of x_i of solutions X vector. If these intervals are computed then thus a value of an inavoidable error is also computed.

But in practice of engineering computations of errors in SLAE by methods of interval analysis are not applied in practice. The main cause of this is the complexity in methods of interval analysis, its inaccessability for a wide circle of engineers and those who applies computers. Inspite of this although methods of interval analysis have started developing at the sixties of the 20th century it has not been included into the program of teaching not only in technical high schools but in the majority of mathematical departments of universities. So, for example, at St. Petersburg state university, department of mathematics – mechanics an interval analysis is not taught and at the department of applied mathematics – control processes an interval analysis has been taught up to 2006 but then it has been excluded from an obligatory program and it was transferred into a subject to be investigated by undergraduates as an elective course – if they wished to known the problem. An exception is, probably, university in Novosibirsk (Russia). But as we know an exception confirms the rule.

Sometimes there is such an opinion that since programs for computers have been developed that allow us to compute intervals (on the basis of interval analysis) intervals in the interior of which are solutions of SLAE then we can apply these programs without knowing anything about the interval analysis. We cannot agree with such an opinion. Certainly for the most simple systems of equations we can obtain a correct answer without knowing an interval analysis. But in more complex cases a formal application of ready programmes if there is no understanding the essence of a sought problem and methods of its solution can lead to an incorrect results and errors in computing real objects inevitably lead to wreckages and catastrophes. The possibility of a fatal mistake is quite probable since in an interval analysis serious drawbacks have been found. One of them was found already in 1979 by Raihman and was stated in [4], p.29–30 as a Raihman example. It is an example of linear algebraic equations system for which methods of interval analysis do not lead to obtaining a correct solution. The existence of such an example means that there exists a series of equations system that is up to the end undefined. For them interval analysis does not allow to obtain a solution. And since a sphere of such systems has not been determined a person who uses routines for interval analysis can meet with impossibility of a solution and will not know what to do.

One more drawback has been found quite recently in the course of investigating "a natural boundary for variations" about which we spoke in precious sections. From these investigations it follows that if a system of equations $AX = B$ is investigated in which coefficients of matrix A are set by its intervals and these intervals are out of "natural boundaries for variations" then there are no finite intervals for solutions, they do not

exist. Solutions can be (in an absolute value) infinitely large. But this is not all. If for such a system we seek intervals in solutions by methods of intervals analysis we can obtain for them finite values (even small ones) but these values will be apriori false ones. As a result a false conclusion will be made about a system conditioning. And such a conclusion can become a cause for catastrophes during the realization of examined system "in metal".

Note that in very seldom examples which are given in courses on interval analysis [3,4,15] only such intervals are considered that do not "go" out of "natural boundarics for variations". Therefore in these examples everything is correct. But if a person that uses these programs written on the basis of interval analysis meets with such a system of equations in which "a natural boundary for variations" lies in the interior of an examined interval of coefficients (and there is a lot of such systems) then such a person can obtain a quite incorrect, false answer which can lead to wreckages if computation results will be applied during the projection of a real object of industry or transport. We can avoid this mistake if before calculating intervals in the interior of which lie solutions of an examined system of equations by given intervals of coefficients in a system if we compute beforehand a "natural boundary for variations" and see whether an examined interval of coefficients values goes out of its limits. But for this investigation it is necessary to apply methods described in §15 and first of all it is necessary to compute "an inverse table of signs" of matrix A in system $AX = B$. In interval analysis of such methods that allow to compute there is probably a "natural boundary for variations" or "an inverse table of signs".

As to methods and algorithms for computations such as "a natural boundary for variations" and an exact value of inavoidable error in each of components of x_i of solutions X vector in system $AX = B$ then for their application it is not necessary to know even matrixes theory.

In fact only theory of determinants is applied. As it is known they are investigated in all technical high schools. This allows different circles of engineers to easily use methods described in the book and thus – to heighten the reliability of a lot of computations used – as a necessary step – the solution of systems of linear algebraic equations.

One more serious defect of methods of computing an inavoidable error based on interval analysis is a very great volume of necessary computations. In interval analysis (as applied to linear algebraic equations systems) a very wide, extremely wide problem is investigated. Variations of solutions are examined – not only when coefficients variations are small but even when large. Therefore the possibility of simplifying computations is not applied. It is connected with the smallness of coefficients variations in a system in real practical problems.

If coefficients variations are not small (if, for example, variations are not small and with coefficients themselves) then we can at once say without any computations that in this case solutions will essentially change and properties of the object will also essentially change. Mathematical model of an object is an examined system of equations. Therefore almost always an investigated object will not carry out its aim.

So in a book example of computing analysis [4], p.25, an example of computing intervals of solutions for system $AX = B$ is given in which

$$A = \begin{pmatrix} [2;3] & [0;1] \\ [1;2] & [2;3] \end{pmatrix}; B = \begin{pmatrix} [0;120] \\ [60;240] \end{pmatrix} \tag{265}$$

It is a known example by Hansen. We can write system (265) in a from:

$$\left.\begin{array}{r} (2,5 \pm \varepsilon)x_1 + (0,5 \pm \varepsilon)x_2 = (60 \pm \delta_1) \\ (1,5 \pm \varepsilon)x_1 + (2,5 \pm \varepsilon)x_2 = (150 \pm \delta_2) \end{array}\right\} \tag{266}$$

where $\varepsilon = 0,5; \delta_1 = 60; \delta_2 = 90$.

For this system a rating vector of solutions (if $\varepsilon = 0; \delta_1 = \delta_2 = 0$) will be equal to $x_1 = 13,6; x_2 = 51,8$.

If we even do not take into account variations of right sides (if $\delta_1 = \delta_2 = 0$) but only take into account variations of coefficients in matrix A then we shall obtain: $x_{imin} = -6; x_{imax} = 45; x_{2min} = -10; x_{2max} = 90$.

It is clear that an object with such variations of values in x_1 and x_2 by no means is not able to be applied in practice.

In answer to such examples we hear such objections. If an interval analysis is able to compute solutions variations then it is able to compute during small variations. And an increase of computations volume in comparison with a methodics based on "table of signs" for modern quick operating computers is not dangerous.

In fact it is not so. A volume of computations often turns out to be a critical factor and an investigation of this question was devnoted in a large article by an outstanding expect on interval analysis S.P.Shariy [15,16].

Conclusions to which S.P.Shariy comes are: the search of exact intervals for solutions of linear algebraic equations systems (according to terminology of [15] of optimal estimating interval vector) is a difficult problem for solutions. Its labour consuming factor increase proportional to an exponent in a number of equations. Often it exceeds the possibilities of quickly operating computers. Thus (says S.P.Shariy in [15]) a theoretical basement has been obtained that during recent thirty – fourty years (during which an interval analysis has developed rather in the direction of a width but not into the (depth) achievements in creating algorithms of exact computing of searched intervals have been modest.

Inspite of a lot of fruitful applications of interval methods in modern natural sciences and in mathematics algorithms for optimal (i.e. "exact" – Petrov Yu.P.) solutions of many interval problems are either not found or they are extremely labour consuming. And they are hardly better than complete excessive ones (see [15].p.95).

In an article [16] S.P.Shairy continues his conclusions – "all approaches developed up to now for optimal (i.e. – "exact") estimates of united a set of solutions in interval systems of linear algebraic equations have exponential difficulty in the worst as it has been already stated this fact is not a consequence of "bad" proposed algorithms but it reflects deep properties of united and other sets of interval linear systems.

Therefore a difficulty with exponents of all examined algorithms is essential and cannot be removed (see [16], p.102–103). "Thus, says S.P.Shary – if a measure (i.e. – a number of equations) of an interval linear system is sufficiently large (more than several dozens) then a number of arithmetics and logical operations exceeds a number of operations carried out on the most powerful computer during any reasonable period of time (an hour a year or even a century) – see [16], p.103.

As an example in [16], p.102 an estimate of complexity (a number of necessary operations) for one of algorithms equal to a number for one of algorithms equal to a number 2^{n^2+n} where n is a number of equations in a system. Already if $n = 10$ this number is equal to $2^{110} > 10^{36}$ that surely exceeds possibilities not of all contemporary but other possible ones in near ten years period computers.

Note that all this discussion does not deny possibilities of an interval analysis. Its methods are constantly modernized and will be modernized in future.

But an interval analysis has inavoidable drawbacks:
1. The complexity of statement and investigation. Everybody can be convinced, everybody – if he got to know with rather simply stated works (see [3],[4],[15],[16] and [17],[18],[13] and others). The complexity of statement in a great degree is connected with the following. Interval analysis bases itself on sufficiently complex interval arithmetics in which (this at once makes more complex everything for the future) a law of distribution – which is especially important for conventional arithmetics – is not fulfilled.

2. An interval analysis for linear algebraic equations systems computes intervals in the interior of which are solutions for any (not, surely, small) intervals in the interior of which are system coefficients. Here we do not apply such simplifications that bring the smallness of these intervals that is characteristic of almost all practical problems. As a result for algorithms based on interval analysis a very quick exponential growth is characteristic of a number of necessary operations with the increase of a number of equations in a system. At the same time for a method of computing an inavoidable error based on "tables of signs" a less slow gradual increase of solution complexity is characteristic (as in §16). This increase is dependent on a number of equations in a system. This essentially simplifys all computations.

3. While applying methods of interval analysis it is necessary to check whether investigated intervals of matrix A coefficients in equations systems $AX = B$ exceed the limits of "natural boundaries of variations" . If they do it is useless to search intervals of solutions since in this case solutions intervals are infinitely large. At the same time all the same we have to compute "a natural boundary of variations" all the same by means of forming an "inverse table of signs" for a matrix by using an algorithm proposed earlier in a monograph [6] and stated in more detouls in this book.

§18. Recommedations for practical application.

Let us at once note that the first step in the solution of a problem concerning the computation of an inavoidable error in SLAE is surely an estimate of a value in possible changes (variations) of object parameters whose mathematics model is an examined SLAE and an estimate of depending on the scope of changing these parameters during a change of coefficients in SLAE.

Note that the first step is a typical engineering problem and for its solution it is necessary to know quite well properties and characteristics of investigated objects. We cannot give any general recommendations suitable for any objects. Therefore in this book this first step of investigations is not considered. And variations of coefficients in SLAE are supposed to be known. But it is necessary to always remember that the first step is not only the first but it is the most important one. Without computing depending on parameters objects of coefficients variations in SLAE or even without a good estimate of possible limits of their change all the most important computations have no sense.

Subsequent steps of computing an inavoidable error is the forming of "tables of signs" , the computation of a natural boundary of variations of matrix A and so on – these are typical problems of applied mathematics. They can be solved without the knowledge of a peculiarity of an examined object. They can be solved without the knowledge of a peculiarity of an examined object. They can be solved by a unique methodics for all SLAE. But during the computing of determinant s variations in Cramer formulas it is necessary to take into account possible dependences between determinants elements. In §12 we have already spoken about it. It is necessary to take into account that the presence of dependences between elements of determinant (for example, in a case of symmetric matrixes) decreases the value of determinants variations. And thus it decreases a value of an inavoidable error and increases "a natural boundary of variations" . Therefore in a majority of investigated cases of examples, in cases of similar (in absolute value) of coefficients variations give the most possible limited values of variations in possible components of x_i of solutions X vector of system $AX = B$.

As we have noted earlier the possibility of realization of this the most dangerous combinations of variations of system coefficients is very small. Computed by an algorithm (described in the book) values x_{imax} and x_{imin}must be considered as an exact higher order of x_i – a order that is achieved very seldom, but sometimes it can be achieved. At the same time values x_i are less than x_{imax} and a little more than x_{imin} and there is a more essential probability as it is shown in an example in [6], p.86.

For practical aims it would be very useful to be able to have computed probabilities of different variations of x_i that are less than x_{imax} and more than x_{imin}.

For example, it is useful to have computed probabilities of getting x_i into an interval from $x_i = 0,9x_{imax}$ up to x_{imax} and into any other interval.

The first steps in this direction of investigations have been made in [6], p.85–89 and also in a report by Shariy S.P. [20].

But up to our days we have not secceeded in solving the main problem – to compute or estimate the law of distributing probabilities of different value of inavoidable error – from x_{iN} up to x_{imax} and from x_{imin} up to x_{iN}. Here there is a wide field for a good interesting scientific work.

A small probability of dangerous combinations of signs in variations of different construction elements (and thus – a small probability of dangerous combinations of signs in coefficients variations of equations system that is a mathematical model of such a construction) leads to important consequences. Let us suppose that some constauction has been computed and projected without computing an inavoidable error. Suppose that a thousand of products that apply such a construction has been manufactured. And during ten years they have worked in good working order. This means that inevitable in the course of exploitation variations of parameters in articles of manufacture have not gone above admissible limits. Are we sure that at the eleventh year of work one of these articles will not lead to dangerous wreckage even if the article is perfectly exploited? Sorry to say, we have no such a guarantee. The absence of wreckages during ten years only means that during these years not one of articles has realized a dangerous combination of variations in signs. But on the eleventh year it can realize. A true guarantee of a reliable work of any article gives only a computation, only such a computation that contains in itself as a step – the computation of a value of an inavoidable error.

The computation of an inavoidable error is advised to begin from the computation of "a natural boundary of variations" . If this boundary is less than such variations that are characteristic of an examined article that means that there is a possibility of very large variations of solutions. Further computation is of no sense as an examined article is not reliable and must be replaced.

Before computing "a natural boundary of variations" it is useful to check whether some of components of x_i in solutions X vector in system $AX = B$ are unreliable only due to variations of right sides, variations of coefficients b_i with unchangeable matrix A, i.e. when there is an absence of variations in coefficients a_{ij}.

Such a primary check is not at all complex. As it is said in §15 it is only sufficient to check – has it not turned out that an inavoidable error x_i that is computed in this case with the help of formulas (236) or (237) can be compared with (or larger) than a solution itself x_i computed by a formula (235).

Note that we have spoken all the time about computing an inavoidable error in solutions of linear algebraic equations system. This does not mean that we have not estimated different mathods of decreasing avoidable errors that are due, for example, to errors during rounding off, a finite number of iterations during applying iteration methods of computing solutions etc. A lot of publications – such as [1,2,29,30,31] and recent works – [32] (to which the reader can apply) are devoted to methods of decreasing inavoidable errors.

We advise to start computing inavoidable errors. If it, for example, is such that we can guarantee only the third sign of a solution then it is useless to apply methods of decreasing avoidable errors in a solution then it is useless to apply methods of a decrease avoidable errors for the computation of the fourth and successive signs of solution.

§19.Application to computation of an inavoidable error in solutions of partial differential equations.

The investigation of solutions error of partial differential equations has its peculiarity, therefore we speak about it in a separate section.

As it is well–known many problems of physics and technique lead to the possibility of computing solutions of partial differential equations such as, for example, Laplace equations:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0 \tag{267}$$

Poisson equations:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = F(x; y) \tag{268}$$

and many others that usually are called equations of mathematical physics. One of modern examples of practical problems that leads to the necessity of computing such equations and give an estimate of inavoidable error in solutions is a problem investigated at the chair of $MECS^{x)}$ of St.Petersurg state University under the head of professor Yegorov N.N. a problem of computing electrostatic potential during the emission of electrons from cathodes of different forms (the computation of "electron guns" , emission in electro-optical systems, systems of forming electron and ionnic pencils etc. [23,24,25,26]. Since an analytical solution of partial differential equations can be obtained only in separate exclusive cases then numerical methods are applied. One of basic methods is a method of nets or a method of finite difference that uses an approximated change of derivatives by finite – differences relations and turning the solution of partial differential equations to the solution of systems of linear algebraic equations (see the book "How to make reliable solutions of equations systems" by Yu. Petrov).

If equations with two independent variables x and y are considered then the solution is a function $u(x; y)$ which in the interior of some sphere G with a boundary Γ on a space xOy (fig.7) corresponds to partial differential equations and on a boundary Γ it satisfies boundary conditions.

$^{x)}$ – MECS is a chair of modeling electromechanical and computing systems

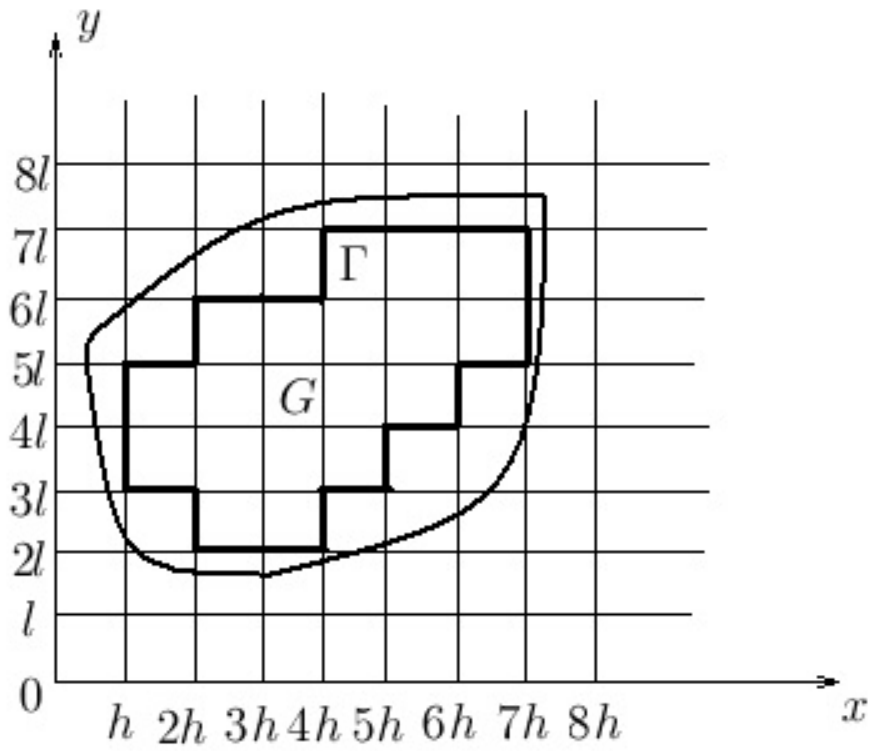

fig. 8

In a method of nets on a space xOy two families of parallel straight lines are formed.

$$\begin{array}{l} x = x_0 \pm ih (i = 0, 1, 2, ...) \\ y = y_0 \pm kl (k = 0, 1, 2, ...) \end{array} \tag{269}$$

Points of crossing these straight lines are called knots. Values of a searched functions in knots of a net are denoted by $U_{iK} = u(x_0 + iH; y_0 + kl)$ and in each interior knot partial derivatives are approximately changed by difference relations:

$$(\frac{\partial^2 U}{\partial x^2}_{iK} \approx \frac{U_{i+1;K} - 2U_{iK} + U_{i-1;K}}{h^2} \tag{270}$$

$$(\frac{\partial^2 U}{\partial y^2}_{iK} \approx \frac{U_{i+1;K} - 2U_{iK} + U_{i-1;K}}{l^2} \tag{271}$$

After such a change an initial partial differential equation turns into a system of linear algebraic equations for values U_{iK} – i.e. values of a function $U(x; y)$ in knots of nets (here they play the role of components of x_i in solutions X vector in systems that we have earlier considered).

Later we first of all shall limit ourselves by Laplace equations set on an isolated square of the size $[0,1]x[0,1]$ cut by a net with a discretization step $h = l\dfrac{1}{m+1}$.

Laplace differential equation in interior knots of a net are approximated by finite differences:

$$U_{i+1;j} + U_{i-1;j} + U_{i;j+1} + U_{i;j-1} - 4U_{ij} = h^2 F_{ij} \tag{272}$$

where F_{ij} depends on boundary conditions.

There are m^2 interior knots in a square on the whole. Therefore an investigated Laplace equation on a square will be approximated by a system of m^2 equations for m^2 unknowns (values of a function $U(x;y)$ in knots) from U_{11} up to U_{mm}. Usually this system is written in a matrix form:

$$AU = F \tag{273}$$

where U – vector–column of unknowns that must be computed of a measure $1\text{x}m^2$, F – vector–column of right sides, A – matrix of a measure $x^2\text{x}x^2$. It is a five diagonals undegenerated symmetric matrix. The majority of coefficients of a matrix are equal to zero. See properties of a matrix are equal to zero. And properties of this matrix in details in [13].

For $m = 2$ matrix A is of the form:

$$A = \begin{pmatrix} 4 & -1 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & 0 \\ -1 & 0 & -1 & 4 \end{pmatrix} \tag{274}$$

For $m = 3$ this matrix becomes:

$$A = \begin{pmatrix} 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 4 \end{pmatrix} \tag{275}$$

Matrix A in a net is chosen and a chosen value of m^2 turns out to be unchanged values of knots U_{ij} only depend on right sides, on values of functions $U(x;y)$, on a boundary of a sphere.

An inavoidable error in computing a value of knots depends on errors obtained during the setting of these values. They cannot be set idially exact. Later we shall estimate a value of this inavoidable error.

By later applying Cramer formulas and decomposing determinants D in numerators of Cramer formulas by elements of the ith column we obtain a formula for the computing of a value of any knots. So for U_{11} we have

$$U_{11} = \sum_{i=1}^{i=m^2} \frac{A_{1;m}}{D} F_m \tag{276}$$

where $A_{i;m}$ – algebraic additions of elements of the column.

Since algebraic additions of elements in any column depend on right sides we can write:

$$U_{ij} = \sum_{i=1}^{i=m^2} K_i F_i \tag{277}$$

where $K_i = \dfrac{A_{ij}}{D_i}$.

So, for example, for $m = 2$ a determinant D in Cramer formulas will be equal to

$$D = \begin{vmatrix} 4 & -1 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & 0 \\ -1 & 0 & -1 & 4 \end{vmatrix} = 75 \tag{278}$$

determinant D_1:

$$D_1 = \begin{vmatrix} F_1 & -1 & 0 & -1 \\ F_2 & 4 & -1 & 0 \\ F_3 & -1 & 4 & 0 \\ F_4 & 0 & -1 & 4 \end{vmatrix} \tag{279}$$

and thus

$$U_{11} = \frac{1}{D}(F_1 A_{11} + F_2 A_{12} F_3 A_{13} F_4 A_{14}) \tag{280}$$

where

$$A_{11} = 12; A_{12} = 12; A_{13} = 8; A_{15}$$

and thus

$$U_{11} = \sum_{i=1}^{4} F_i F_i = 0,16F_1 + 0,16F_2 + 0,107F_3 + 0,2F_4 \tag{281}$$

Therefore for a given measure of a net a value of any of knots U_{ij} only depends on right sides.

A value of right sides can be known to us only with inevitable error:

$$F_i = F_{iN}(1 \pm \delta_i) \tag{282}$$

where F_{iN} – a rating value – applied for the computation of knots values.

The presence of an error can either increase or decrease a value of U_{ij} in comparison with its rating value corresponding to $\delta_i = 0$. An increase of U_{ij} will be the largest if a sign ε_i coincides with a sign of a product F_{iN} by $\dfrac{D_i}{D} = K_i$. The change of value U_{ij} in the direction of a decrease will be the largest of a sign ε_i is opposite to a sign of a product F_{iN} by $\dfrac{D_i}{D} = K_i$.

Now it is possible to answer the main question: by how many times can a value of any knot U_{ij} change due to the presence of an error in a right side?

Answer: by any times. All depends on right sides. If they are with variable signs than even if algebraic additions A_{ij} have one sign then a value U_{ij} can be near to zero and then an inavoidable error is equal to

$$\Delta U_{ij} = \sum_{i=1}^{i=m^2} |K_i F_i \delta_i| \tag{283}$$

even if δ_i is small can be by any times more than a rating value of a knot equal to:

$$U_{ijN} = \sum_{i=1}^{i=m^2} K_i F_{iN} \tag{284}$$

Example: for $m = 2$ computed values k_i for knot U_{11} (formula (281)) are: $K_1 = 0,16; K_2 = 0,16; K_3 = 0,0107; K_4 = 0,2$. If $F_1 = 5; F_2 = 0; F_3 = 0; F_4 = -3,99$ then $U_{11} = 0,16 \cdot 5 - 0,2 \cdot 3,99 = 0,002$. If $\delta_1 = 0,01; \delta_4 = -0,01$ then $\Delta U_{11} = 0,01(0,16 \cdot 5 + 0,2 \cdot 3,99) = 0,01598$, i.e. even if an error in value of a right side does not exceed $\dfrac{1}{100}$ then an inavoidable error in the value of a knot U_{11} is by 8 times more than a rating value U_{11}.

Hence it follows an important conclusion: during a numerical computation of partial differential equations (while reducing them to systems of linear algebraic equations) it is by all means necessary that a step of solutions must be computation of inavoidable errors in a solution that depends on inavitable errors while right sides are set. Without computing a possible value of inavoidable error the solution cannot by no means be reliable since there exist such right sides for which an inavoidable error can be larger than a solution itself by any number of times.

All this, surely, is true for problems of computing electrostatic potential while emissing electons from cathods of different forms, systems of forming electronic and ionnic pencils, emission electron – optical systems etc. investigated at a chair MECS in St.Petersburg state university (a chair of modelling of electromechanical and computer systems).

At the same time the existence of such boundary conditions for which their small error leads to large inavoidable errors in solutions, surely does not mean that for all other boundary conditions we shall have the same picture. There exist such boundary conditions for which an error while setting them do not greatly influence the solutions. Everything depends on certain boundary conditions.

Therefore the computation of an inavoidable error in value of knots U_{ij} for certain boundary conditions by a methodics given in this section makes computing results reliable.

There is not an essential difference between computing a maximal value of inavoidable error for a case when it depends on variations of coefficients in matrix A and vector B in a system of equations $AX = B$ and a particular case considered in this section when an inavoidable error only depends on variations of boundary conditions, i.e. finally – on variations of components of vector B. While taking into account variations of matrix A a combination of variations leading not to the largest value of inavoidable error of solutions but to values that are near to maximal one has a very small probability. Probabilities of variations combinations that lead not to a maximal value of inavoidable error but to values near to a maximal one already have a more essential probability but computing methods of probability of different values of an inavoidable error of solutions in SLAE up to now have not been developed.

For a peculiar case when an inavoidable error only depends on variations of a vector B in a system everything is simplier. A comparison of formulas (283) and (284) at once shows that if F_{iN} are such that a value of a knot U_0 is small when F_{iN} is not very small then many such combinations of signs ε_i be found in which an inavitable error in a value of knot U_{ij} will be very large even if $|\varepsilon_i|$ is small in absolute value of variations in boundary conditions and right sides in system $AX = B$. Therefore a probability of a substantial change in a value of any of knots U_{ij} will not be small even if variations of boundary conditions are small.

Surely, it would be advisable to compute an exact value of this probability but this is a rather difficult problem and it has not as yet been solved. Now let us consider a question of labour consuming computation of an inavoidable error of solutions of partial differential equations. We shall use known relations: for the computing by Gauss method of all components of vector X in solving system $AX = B$ whose matrix A is of a measure nxn require approximately n^3 multiplications. How many multiplications it is approximately required for computing a determinant in matrix A if we apply Gauss method as well.

During the computation of an inavoidable error for each of knots U_{ij} in system of algebraic equations that approximate (in a square net) a solution of Laplace equation it is required (according to formula (283)) to compute a determinant D and its all algebraic additions that enter into formula (276) and (277). In all it is required (if we take into account formula (283)) approximately $n^3 + n^2(n-1)^3 + n^2$ multiplications. Thus a number of necessary multiplications (with the growth of a number of equations) increases approximately proportional to the fifth degree from a number of equations n. And since $n = m^2$ (where $m^2 - a$ a number of knots) then if, for example, $m = 30$ (i.e. when a net is 30x30) a number of equations will reach 900 but a necessary number of multiplications will be near $6 \cdot 10^{15}$.

But a number of necessary computing operations can be greatly reduced. We can apply the fact that coefficients K_i in formulas (282) and (283) do not depend on boundary conditions, we can compute them beforehand and introduce them into a machine memory.

Then for computing an inavoidable error during a partial differential equation approximation from a system of n algebraic equations a necessary number of multiplications

for computing all n knots U_{ij} will be equal to only n^2 (according to formula (283)) after computing coefficients k_i. But while using Gauss method in a general case it would be necessary to have n^3 multiplications.

Therefore we can draw the following conclusions of a general character. They are true during the computation of solutions of partial differential equations by means of approximating them by systems of linear algebraic equations systems:

1.Without computing an inavoidable error which is due to inevitable inexactnesses and variations of boundary conditions solutions cannot be considered reliable (by no means) since such (not known earlier) combinations of boundary conditions and their variations are possible in which even small (in absolute value) errors can lead to large (and even essential) changes in solutions;

2. The calculation of an inavoidable error can be carried out if there is a moderate number of necessary computing operations – even for systems with a large number of equations if there is a sufficiently useful net of approximation.

Since conclusions about possible large inaboidable error in solutions of SLAE even if right sides have small errors (and thus – for SLAE that approximate solutions of partial differential equations considered in [23;24;25]) have been published in 2009 [6] we can hope that works presented, for example, in 2011 that contain numerical solutions of partial differential equations will be rejected. Here an inavoidable error in solutions is not computed.

At the same time it is not at all difficult to compute an avoidable error and the presence of such a computation will restore the reliability of an obtained result.

§20. Wreckages that occur due to inexactnesses in methods of computing and projection.

As all previous sections have shown conventional and widely applied computation methods are far from being always perfect.

So up to 2009 methods of computing an inavoidable error in solutions of systems of linear algebraic equations were far from being perfect. This position has been partly improved by publication [6].

But as it was shown in [7] computations connected with systems of algebraic equations contained up to 70% of all computations.

Therefore drawbacks and inexactnesses in computations have inevitably become a source of wreckages and catastrophes. We know little about these wreckages and catastrophes since organizations that realize projection and computation are not interested in making known drawbacks of applied by them conventional computation methods. And when wreckages are investigated they are not in a hurry to apply new more advanced methods and represent causes for wreckages such as bad manufacture, "people's errors" etc. So they mention every cause but the true one – they do not aknowledge the necessity of improving computing and projection methods.

The most known and dreadful catastrophe whose cause we do not succeeded in completely discovering was a catastrophe of aquapark ("a water" park) "Transvaal" that occured on the 14th of February, 2004 in Moscow when twenty seven persons perished including children and 113 persons received different traumas.

Aquapark "Transvaal" situated in the suburbs of Moscow was one of the favourite places of Moscovites rest. Swimming pools, different shows, water small hills, ponds with arificial sea waves etc. – everything made people comfortable, be healthy and light – hearted. The building of the aquapark was beautiful and original. The roof of a pool was supported by a series of columns situated as a half circle between which there was a place for large windows.

Therefore on that fatal day – the 14th of February 2004 the aquapark was full of light – hearted and gay grown ups and children.

And suddenly one of columns on which a roof was based crushed due to its weight and then other columns due to the influence of additional weight also began crushing. The roof crushed on the heads of swimming people. Cries of crushed people arose, the electric light went out and in a complete darkness these who survived tried to get out from fragments that had crushed on them.

Then rescuers appeared. For several hours they dissembled obstructions they saved those who were alive. Then bitter conclusions must be drawn: twenty seven persons perished, 113 persons were injured among those who perished and was injured were children.

Surely such a catastrophes was carefully investigated. The first version was – as usual – about terrorists who supposedly have put an explosive into one of columns on which a roof stood. But on a place where an explosive must have small marks of an exposive or products of their decomposition must be found. They are closely absorbed into the pores of concrete by a large pressure of an explosion. This reaction can be easily found by methods of modern exact chemical analysis. These exact analysis have ascertained that there are no traces of explosives.

Later a version about a terrorist act has been wholly rejected. A quality of the construction has been checked – whether non qualitative and substructive materials has been applied, whether a real construction corresponded to projection requirements etc. The check has shown that there was no deviation from the project.

After all investigations and experts have drawn the conclusion that a single cause of a catastrophe is errors during a computations and projection of aquapark. A well-known architects burean – ZAO "K" has projected it at the head of an experienced and respected architect Nodar Vahtangovitch Kancheli. Before this catastrophe he was the head in projecting many unique buildings in Moscow. On the first of April, 2005 N. Kancheli was charged with two serious articles of Criminal Code of the Russian Federation: article 109, part 3 (to cause death due to imprudence that has led to death to two or more persons) and article 118 part 2 (to cause a gross harm to the health of people due to unreliable relization of professional duties). Moscow procurator undertook a criminal affair against N.V.Kanchelli.

But N.V.Kanchelli and achitects that have worked under his guidance are experienced experts. Therefore the possibility of elementary errors in projection and computation (not necessary coefficient was taken, not a necessary formula was applied etc.) were by all means excluded. Another problem is the computation of an inavoidable error for a mathematical model of a building. But in 2004 and earlier when "Transvaal" was projected there did not exist reliable methods of computing an inavoidable error. Only approximated estimates were applied by means of "a number of condition" $||A||\cdot||A^{-1}||$. But these estimates (as we know from previous section) can easily lead to false projection solutions and catastrophes.

It is quite possible and probable that a cause of a catastrophe in the aquapark "Transvaal" has become an inexact estimate of an inavoidable error during the computation of a mathematical model of a building. Sorry to say, we cannot state this with an absolute certainity. The investigation of this criminal affair in the course of which a complete answer would have been given – was on the fifth of August, 2006 was stopped by amnisty. As in 1906 in Russia the first "Duma" has started working (as a nundred years clasped) an amnesty for all convicted and accused whose age is more than 60 years was declared. N.V.Kancheli received the amnesty and this fact finally prevented to be sure in finding a cause for the catastrophe of aquapark "Transvaal" . If this cause (asseredly – almost) is an inexactness in the estimate of inavoidable error then N.V.Kancheli is not guilty since in these years when an aquapark was projected there were no reliable methods of computing an inavoidable error. In more details about the catastrophes of "Transvaal" and other technogene catastrophes you can read in the book [39].

Much worse is a position that appeared later, in 2010. In 2009 in a monograph [6] a methodics of an exact computation of a value of inavoidable error was published that

allowed to substentially decrease the wreckages probability. This methodics has been stated together with associated questions. The book has been reviewed by doctors of physico-mathematics sciences, professors Blekhman J.J., Ignatiev M.B., Ushakov A.V. A short story of problems discussed in the book was given in an article [21] published in a respectable scientific magazine "Vestnik grazdanski enzinierov" .

It seems that if a methodics of computing an exact value of inavoidable error that allows to essentially decrease the probability of wreckages has been published and recognized by authorative experts then it must be quickly applied by organizations that carry out computation and by projecting developments in order to increase the reliability of computation results, for the decrease of a probability of catastrophes.

Not to say that the application of more well–grounded new computation methods – "coefficients of a reserve" can greatly reduce the cost of construction – and this is a very large sum of money. Since a methodics of computing an inavoidable error has been developed at St.Petersburg state university (St.PSU), the chair of modelling electromechanical and computer systems (at the head of profeccor N.V.Yegorov) St.P.S.U. – addressed itself to a series of firms that carry out computations and projections to apply published results and programs developed at SPbSU in order to work with electronic machines. In 2009 they have obtained a state registration certificate N2009613251 on the 23th of June, 2009.

Sorry to say once again words by our President Miedvedev D.A. have been conferred that Russian companies, Russian businessmen are extremely slow in applying any scientific advancements and they do not like to use something new and modern.

Then St.P.S.U. (by a pro-vector for scientific work N.G.Skvortzov) addressed a service of state construction inspection and Expertize to the head of this service A.S.Ort (a letter №01–115 on the 19.07.2010). The address of St.P.S.U. supported by the Russian academy of sciences (RAS) under the head of the scientific union of RAS Ya.B.Danielievitch. After he has become acquainted with results of St.P.S.U. investigations Ya.Danilievitch wrote to A.J.Ort urgently recommended to take into consideration and further to realize an innovation methodics and algorithms of computations developed at St.P.S.U. under the head of professor Petrov Yu.P. since this methodics is formed on mathematical analysis that allows to guarantee the reliability and safery of building constructions and prevent wreckages that occur in modern practice of construction. He stresses that this methodics possessed great exactness in comparison with contemprary estimates methods.

To this adress of St.P.S.U. and the Russian academy of sciences a service of a State construction Inspection and Expertize received the answer that the Service is not authorized to carry out an examination for the estimating of construction objects condition. It is interesting to know to what is the Service authorised?

It is clear that to take part in innovations, in applying more advanced methods this Service does not wish. It is sad since the main problem of the Service is, surely, to secure safety of construction by means of revealing dangerous and ill–computed projects and the Service can carry out this aim only if advancements of science be applied but the Service does not want to apply them.

Too much wreckages and catastrophes occur due to errors and inexactnesses during the projection and computation. Exact numbers were given in the "Conclusion" of one of scientific instructions (TZNII Psk, N.P.Melnikov). This conclusion was formed with Moscow "City centre of expertizes". In this "Conclusion" (published in 2006) it is said that 9,8% of buildings crushes that have occured during recent years were due to inexactnesses and errors of projection and computation. It is a very large number. Science can decrease a number of wreckages but in order to achieve this aim it is necessary to use its advancements.

Part II. Systems of differential equations and equivalent transformations

§21. Examples of equations systems and equivalent transformations.

One more interesting section of "Mathematics–2"turns out to be a theory of differential equations. Here quite recently (published in [36] then [37], [38], [5]) it has turned out that during small changes of coefficients such an important property of equations as stability of solutions has changed. This change itself is closely connected with everywhere applied equivalent transformations of equations.

Let us recall that such equations are called differential into which derivative of searched functions enter. And we call solutions of different equations such functions that turn an equations into an identity. The most simple and at the same time the most often occuring in practice are linear differential equations with constant coefficients, such as, for example,

$$a_1\frac{dx}{dt} + a_0x = sint$$

equations of the first order or

$$a_2\frac{d^2x}{dt^2} + a_1\frac{dx}{dt} + a_0 = 0$$

Later we shall examine a change in solutions in differential equations during variations of their coefficients. i.e. we shall take into account that real coefficients almost always are known only with limited exactness and they satisfy inequalities:

$$a_i(1-\varepsilon_i) \le \bar{a}_i \le a_i(1+\varepsilon_i), \tag{1}$$

where $\bar{a}_i$ – unknown to us true values, a_i – rating values that are used during computations, $\varepsilon_i a_i$ – coefficients variations.

It is important that in differential equations (as earlier shown in [6], [11]) we consider only relative variations if $a_i = 0$ then $\varepsilon_i a_i = 0$, i.e. "a zero is not variated". Thus automatically the so called "singularly perturbing equations"are excluded from examination. These equations have small coefficients at higher derivatives as equations:

$$\varepsilon\frac{d^2x}{dt^2} + a_1\frac{dx}{dt} + a_0 = f(t) \tag{2}$$

where ε – small parameter or systems of similar equations.

To singularly perturbing equations a wide of literature is devoted (see, for example, [45]) but this is quite another area of examination. And in this book singularly perturbing equations are not considered (i.e. in the second part of the book – where in the second part – in difference with the first – a zero is not variated). The cause of this is that in singularly perturbing problems the change of coefficients in higher derivative from a zero

up to infinitely small value surely leads to essential changes in solutions. It is at once clear. In difference of this we consider less evident but earlier not noted such cases when small changes of large coefficients can lead to substantial changes of solutions.

If differentiation operator $\frac{d}{dt}$ is denoted by D (i.e. $\frac{dx}{dt} = Dx$ and, correspondingly. $\frac{d^n x}{dt^n} = D^n x$) then a linear differential equation with constant coefficients of any order – an equation

$$a_n \frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} + ... + a_1 \frac{dx}{dt} + a_0 x = f(t) \tag{3}$$

can be written in the form:

$$(a_n D^n + a_{n-1} D^{n-1} + ... + a_1 D + a_0)x = f(t)$$

or

$$A(D)x = f(t) \tag{4}$$

i.e. in the form of a product of a sought function $x(t)$ by polynomial $A(D) = a_n D^n + ... + a_0$ of differential operator $\frac{d}{dt} = D$

Systems of linear differential equations with constant coefficients are called the following systems:

$$\left.\begin{array}{c} A_{11}(D)x_1 + A_{12}(D)x_2 + ... + A_{1n}(D)x_n = f_1(t) \\ ..... \\ A_{n1}(D)x_1 + A_{n2}(D)x_2 + ... + A_{nn}(D)x_n = f_n(t) \end{array}\right\} \tag{5}$$

i.e. systems that consist (in a general case) of n equations that include in themselves n unknown functions $x_1; x_2; ...; x_n$ and n^2 polynomial A_{ij} of different degrees of a differention operator $D = \frac{d}{dt}$.

Thus systems of linear differential equations are a generalization of systems of linear algebraic equations (SLAE) examined in the first part. If in polynomials $A_{ij}(D)$ all members except members with a zero degree in differentiation operator D are equal to zero and functions $F_i(t)$ in the right sides of system (5) turn to constant numbers then system (5) transforms into a system of linear algebraic equations (SLAE) that is one of such systems that have been examined in the first part of the book.

Thus problems considered in the second part of the book that have been more complex than in the first part but while solving them we shall base our selveves (mainly) on results obtained in the first part. See in publications [5], [11], [22] and others.

One of examples of systems of differential equations that is a mathematical model of a real object is a system that is describing transient processes that take place in a system of regulating a frequency of rotation in an electrodrive of a constant current that work on a mechanism with a variable resistance moment. This system is of a form:

$$(D^3 + 4D^2 + 5D + 2)x_1 - (D^2 + 2D + 1)x_2 = 0 \tag{6}$$

$$(D^2 + 4D + 5)x_1 - (D + 1)x_2 = 0 \tag{7}$$

and in it a variable x_1 – a frequency of rotation of an electrodrive (to be more exact – its deviation from a rating value) variable x_2 – current of an anchor that plays the role of the control equation (6) - an equation of the control object – electrodrive equations (7) – an equation for a regulator reflecting processes that occur in a chain of an inverse connection. The later we shall speak in more details how to obtain equations (6) and (7).

Now let us speak about equivalent transformations. We call equivalent transformations of equations such transformations that do not change solutions of equations (or systems of equations). A set of solutions in a transformed equations system must be identical with each other. Here are examples of equivalent transformations:

1. the reduction of similar members
2. the transfer of members from the left side of an equation into a right side and vice versa with the change of a sign
3. the multiplication of all members in an equation by one and the same number that is not equal zero.
4. substitution – i.e. a change of any of equation members by a member equal to it.
5. Differentiation (by members) of all members of equation.

The first four examples of equivalent transformations are related with transformations studied even in secondary school and they do not arise any doubts.

It is clear that equation

$$3x + 4x = 14$$

after the reduction of similar members will become equivalent to an equation

$$7x = 14,$$

which in its turn is equivalent to equation (after have been multiplied by $\frac{1}{7}$:

$$x = 2.$$

The fifth example of equivalent transformations (differentiation by members), possibly, will require explanations. So, for example, equation of the first order

$$\dot{x} + x = 0 \tag{8}$$

with an initial condition:

$$x(0) = 0 \tag{9}$$

has a solution $x = 0$. After a differentiation by members equation (8) turns out into an equation of the second order:

$$\ddot{x} + \dot{x} = 0 \tag{10}$$

and in order that it has a definite solution it is necessary to one initial condition (9) to add the second initial condition for $\dot{x}$ when $t = 0$. But this second initial condition is not

arbitrary but it wholly results from initial equation (8) with initial condition (9). From (8) and (9) it really follows that when $t = 0$ we shall have:

$$\dot{x}(0) = 0 \tag{11}$$

and just this equality is a necessary second condition for equation (10). If we take in to account equation (10) we shall have a solution $x = 0$. Therefore if we set additional initial conditions the differentiation by members will not change solutions and it is an equivalent transformation. Hence it follows that the multiplication of all equation members by any operator polynomial $A(D)$ is also (if we have set additional initial conditions correctly) an equivalent transformation.

Sometimes we hear an objection against this statement connected with a conception of a general solution of a differential equation – i.e. a solution that contains integration constants that later will be defined from initial (or boundary conditions).

A general solution of equation (8) is of the form:

$$x = C_1 e^{-t}$$

with one integration constant C_1 and a general solution of equation (10) obtained from equation (8) by means of members differentiation is of the form:

$$x = C_3 + C_2 e^{-t} \tag{12}$$

and it does not coincide with a general solution of (8). Hence a false conclusion is sometimes drawn: since a general solution has changed the members differentiation is not an equivalent transformation. A mistake occurs from a false understanding of a term "a general solution" that has been created even since the XVIIIth century [46], [47]. In the essence instead "a general solutions" , it is more correct to call a family of functions that depend on integration constants which in their turn depend on initial conditions. Therefore there is nothing surprising that during equivalent transformations of families of solutions (conventionally called "general solutions") can change its form but certain solutions that satisfy certain initial conditions during members differentiation do not change.

So if we should like to define integration constant in solution (11) from an initial condition $x(0) = 0$ we shall obtain: $C_1 = 0$ and $x(t) = 0$. If we define integration constants C_1 and C_2 from initial conditions $x(0) = 0$ and $\dot{x}(0) = 0$ we shall obtain $C_1 = 0; C_2 = 0$ and therefore we again have $x(t) = 0$. Equivalentness of such a transformation as members differentiation in this example is confirmed. The same is true in a general case.

By using equivalent transformations linear differential equations with constant coefficients of any order (equation (4)) and also any system of such equations (system (5)) can be reduced to the so called Cauchy normal form (Cauchy, 1789–1857), i.e. to a system of n equations of the first order:

$$\left.\begin{array}{c} \dot{x}_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ ..... \\ \dot{x}_n = a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n \end{array}\right\} \tag{13}$$

(later we shall limit ourselves by a case $f(t) = 0$) while considering the transformation of a system of a general type (5) into a form (13).

If into a normal Cauchy form an equation of the n-th order is transformed then a number of equations in system (13) is equal to an order of an equation and the system itself (13) for, example, equation

$$a_2\ddot{x} + a_1\dot{x} + a_0 x = 0 \tag{14}$$

becomes (if we suppose that $x = x_1$):

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{a_0}{a_2}x_1 - \frac{a_1}{a_2}x_2 \end{aligned} \tag{15}$$

If a system of several differential equations of different orders (of a type (5)) is reduced to a normal Cauchy form then a number of equations of the first order in a normal form in a general care is equal to a sum of orders of differential equations for each of variables.

In special cases (if there is the so called degenerated system) a number of equations of the first order in an equivalent normal Cauchy form can be even less.

Example: a system of equations (6) and (7) has in equation (6) the third order in relation to x_1 but in equation (7) in relation to variable x_2 it has the first order. Therefore a normal Cauchy form for a system (6)-(7) must contain four equations of the first order for four variables $x_1; x_2; x_3; x_4$. Variables x_1 and x_2 are such ones that already are present in equations (6) and (7) but x_3 and x_4 – are additional variables. Since, really, a system (6)-(7) is degenerated (as we shall show it in the next section) then for it one of four differential equations degenerates into a finite equation that does not contain derivatives (or – just the same – it contains derivatives of a zero order, i.e. functions x_1, x_2, etc.). And a normal Cauchy form for system (6)–(7) consists of three equations of the first order.

For systems of linear differential equations with constant coefficients equations in the normal Cauchy form (13) can be usually written in a vector-matrix form:

$$\dot{X} = AX \tag{16}$$

where X – n-measured vector whose components are sought functions $x_1(t); x_2(t); ...; x_n(t)$, A – matrix of coefficients a_{ij}.

Note that since transformations that have turned equations system of a general type (5) into a normal Cauchy form is an equivalent transformation then in the limits of "Mathematics – 1"(that supposes that we exactly know equations coefficients and that they are unchanged) mathematical models of examined objects and processes in the form of equations systems (5) and (16) are similar.

And since a form (16) is more simple then just form (16) prevales in discussions in modern text-book on the theory of differential equations (see, for example, [33], [34] and in computation practice).

In particular package of applied programs used during a numerical solution of differential equations systems and while checking stability of solutions (packages MATLAB, MathCad etc.) have been formed for systems in a normal form while admitting that transformation of an initial system into a normal form does not change anything. In fact – as we shall see in next section – it is not always so. There exist special cases when equivalent transformations of equations systems of different orders into a normal form – can change some important properties of equations.

These special cases must be attentively taken into account and investigated thoroughly since they can become (and have not once become! See, for example, [39]) – a cause for error in computation and thus a cause for wreckages and catastrophes. It is not (at all) permissible to apply programs from packages MATLAB, MathCad and any others without taking into account assumptions applied during the formation of these programs that are not always clearly formulated.

This simple but important notice stated already in 1994 [37] and for many times repeated later in [38], [5], [52] and others – was not at once introduced into the practice of technical computations.

In fact it was for the first time applied in 2007–2010 at Moscow state technical University (MSTU) named after Bauman N.E. and at this university scientifico-production unions. The account of special cases in equivalent transformations allowed scientifico-production union at MSTU to sufficiently reduce a number of wreckages. They have developed a complex equipment.

In 2011 on the site of MSTU on the 30.01.11 blog V.B.Manichev applied to all high schools and scientific organizations of Russia to widely apply in their computations and projections to carry out a check for having "special" systems and equivalent transformations that change parametric stability. In order to realize such checks that would essentially reduce wreckages and catastrophes at MSTU good programs provision (a programs complex FMS PA10 and programs blocks SADEL as seen in publication [32]) have been developed.

§22. Characteristic polynomials and a check of stability.

It is known that a solution of a linear differential equations with constant coefficients (4) is a sum of a quotient solution that depends on the right side of equation (4), on function $f(t)$ and a general solution of a homogeneous equation:

$$(a_n D^n + a_{n-1} D^{n-1} + ... + a_0)x = A(D)x = 0 \tag{17}$$

whose solution is of the form:

$$x(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + ... + C_n e^{\lambda_n t} \tag{18}$$

where $\lambda_1; \lambda_2; ...; \lambda_n$ – roots of a characteristic polynomial in equation (17), a polynomial

$$a_n \lambda^n + a_{n-1} \lambda^{n-1} + ... + a_0 = A(\lambda) \tag{19}$$

A characteristic polynomial (19) is formed from equation (17) by a simple exchange of differentiation operator $D = \dfrac{d}{dt}$ by any variated value which conventionally is denoted as λ.

If among roots of a characteristic polynomial (19) there are complex ones then they can enter only by connected with each other pairs:

$$\lambda_{i;i+1} = \alpha \pm j\beta \tag{20}$$

In formula (20) and later we shall adopt that $j = \sqrt{-1}$, i.e. an imaginary unity we shall denote by a letter j (often a designation $\sqrt{-1} = i$ is used) but in electrotechnique electric current is denoted by i. In order to avoid errors it is better to use a designation $\sqrt{-1} = j$.

To each pair of complex roots in a general solution of equation (17) (to be more exact – in a family of its solutions) will correspond a member of a form:

$$e^{-\alpha t}(C_1 sin\beta t + C_2 cos\beta t) \tag{21}$$

If all roots in a characteristic polynomial have negative real parts then any solution of a homogeneous equation during any initial conditions in the course of time will speed to zero. On this conclusion does not influence the presence of multiple roots – their preseace leads to the fact that in a general solution can appear the following members:

$$C_i t^m e^{\lambda_i t} \tag{22}$$

But if only all roots λ_i have negative real parts then any solutions – including solutions of the form (22) – tends to zero when $t \to \infty$.

For systems of linear differential equations their characteristic polynomial is equal to a determinant of equations system in which a differentiation operator $D = \dfrac{d}{dt}$ is changed by a letter λ (certainly, you can change operator D by any other letter as well but conventionally – a letter λ is used).

So, for example, for system of equations (6)-(7) its characteristic polynomial is equal to a determinant:

$$\begin{vmatrix} \lambda^3+4\lambda^2+5\lambda+2 & -(\lambda^2+2\lambda+1) \\ \lambda^2+4\lambda+5 & -(\lambda+1) \end{vmatrix} = \\ = \lambda^4+6\lambda^3+14\lambda^2+14\lambda+5-\lambda^4-5\lambda^3-9\lambda^2-7\lambda-2 = \\ = \lambda^3+5\lambda^2+7\lambda+3 = (\lambda+3)(\lambda+1)^2 \tag{23}$$

(members with the fourth degree λ are mutually reduced).

A characteristic polynomial (23) has roots $\lambda_1 = -3; \lambda_2 = \lambda_3 = -1$ and therefore a general solution of system (6)-(7) is of the form:

$$x(t) = C_1 e^{-3t} + (C_2 t + C_3) e^{-t} \tag{24}$$

Equality (23) shows that members with the fourth degree λ in a characteristic polynomial mutually are reduced. Therefore its degree turns out to be not the fourth but the third degree. This means that system (6)-(7) is degenerated.

For systems of differential equations with constant coefficients the same real negative partes preserves force. Then all solutions of a system tends to zero if $t \to \infty$ and such solutions are called stable (to be more exact – asymptotically stable).

If only one root a characteristic polynomial in a system (or only one pair of complexly connected roots) has a positive real root then solutions of a system can increase without any boundaries when $t \to \infty$.

If all roots of a characteristic polynomial in system with constant coefficients have negative real parts then all solutions of a system are stable. Therefore we often hear not about stability of solutions but about stability of some system. So, for example, a system of equations (6)-(7) is called a stable system since all its solutions are stable (and all solutions that form its family (24) are stable. Conventionally this family of solutions are called "general solutions").

Let us note that for nonlinear differential equations and systems of such equations everything is more complex. There one solution can be stable and another one – unstable. Therefore there we cannot speak about stability of a system of equations but only – about stability of some solution.

In an a theory of linear systems of differential equations criteria that allow us to judge about stability or instability directly by coefficients of a characteristic polynomials without computing its roots has for a long period of time been developed. The most simple – and the most important – criterium is Stodola criterium (Stodola, 1859-1942) for the stability of a system it is necessary (but not sufficient) that all coefficients of a characteristic polynomial have one and the same (for example, positive) sign and that among them there were no zero coefficients.

Sufficient conditions for stability have been found by a German mathematician *A* Hurwitz (Hurwitz, 1859-1919) already in 1895. In honour of *A*. Hurwitz polynomials in

which real parts of all roots are negative are called Hurwitz polynomials. As it is known a necessary and sufficient condition of negation in real parts of all roots in polynomial:

$$a_n D^n + a_{n-1} D^{n-1} + ... + a_0 \tag{25}$$

consist in positivity of all diagonal minors of the next matrix (Hurwitz matrix) cousisting of coefficients of polynomial (25) and containing n lines and n columns:

$$\begin{vmatrix} a_{n-1} & a_{n-3} & a_{n-5} & ... & 0 & 0 \\ a_n & a_{n-2} & a_{n-4} & ... & 0 & 0 \\ 0 & a_{n-1} & a_{n-3} & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & a_1 & 0 \\ 0 & 0 & 0 & ... & a_2 & 0 \end{vmatrix} \tag{26}$$

Here is a rule for forming Hurwitz matrix on a main diagonal: coefficients of a polynomials (25) from a_{n-1} up to a_0 are written out. Each column is then added in such a way that indexes of coefficients decrease by a unity from below to up during the transfer from one line to the other line. Coefficients whose indexes are less than a zero and that are larger than n are changed by zeros.

While consequently writing out diagonal minors of matrix (26) we see that for Hurwitz polynomials of the first and the second degree positivity of their all coefficients (coefficients a_1 and a_2) is necessary and sufficient.

For Hurwitz polynomial of the third degree

$$a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$$

besides positivity of all coefficients the fulfillment of the following additional conditions are necessary and sufficient:

$$a_2 a_1 > a_3 a_0$$

i.e. a product of middle coefficients must be more large than a product of extreme coefficients.

On the whole a theory of stability of linear systems with constant coefficients (in the limits of "Mathematics – 1") is for a long period of time been well developed. And in previous statement we have only recallected these main positions that will later be used.

But in practice it is not at all sufficient that an examined object be stable. It is also necessary that its stability was preserved during if only small – and thus inevitable in practice – deviations of object parameters from computed values occur.

Definition

The property of preserving stability during parameters deviations of an object from computed values during infinitely small changes of parameters and generated by parameters variations changes of coefficients in mathematical model of an object can be in short called parametric stability.

Quite recently during the investigation of parametric stability unexpected and paradoxical phenomena have be found about which we shall speak more details in next sections.

Note that only parametric stability can be considered a true real stability. Systems that are stable during rating values of parameters but which loose stability during infinitely small values and thus – inevitable in practice – parameters variations are not at all better than – and even they are more dangerous if the loss of stability occurs – as we shall see later – occurs only during variations of a certain sign.

§23. The change of parametric stability during equivalent transformations.

A system of differential equations (6)-(7) is a stable system. Its characteristic polynomial equal to determinant (23) has all roots with negative real parts. Really, as shows formula (24) all solutions of system (6)-(7) tends to zero if $t \to \infty$.

But system (6)-(7) is not parametrically stable. In fact if we change by infinitely small value ε some coefficients in the system, for example, a coefficient in Dx_2 in equation (7) – and system (6)-(7) will become:

$$(D^3 + 4D^2 + 5D + 2)x_1 - (D^2 + 2D + 1)x_2 = 0 \tag{27}$$

$$(D^2 + 4D + 5)x_1 - [(1 + \varepsilon)D + 1]x_2 = 0 \tag{28}$$

then its characteristic polynomial will be equal to:

$$\begin{vmatrix} \lambda^3 + 4\lambda^2 + 5\lambda + 2 & -(\lambda^2 + 2\lambda + 1) \\ \lambda^2 + 4\lambda + 5 & -[(1 + \varepsilon)\lambda + 1] \end{vmatrix} = \\ = -\varepsilon\lambda^4 + (1 - 4\varepsilon)\lambda^3 + (5 - 5\varepsilon)\lambda^2 + (7 - 2\varepsilon)\lambda + 3 \tag{29}$$

and thus during infinitely small positive ε polynomial (29) will not already be Hurwitz since a necessary Stodola condition has been broken. And during infinitely infinitely small $\varepsilon > 0$ system (26)–(27) will not already be stable.

In a family of its solutions besides three exponentially decreasing members reflected in formula (24) the fourth exponentially increasing members (during small ε) approximately equal to $C_4 e$ and increasing more quickly than ε is smaller.

It is important that during small negative ε system (27)-(28) remains stable.

Thus system of equations (6)-(7) does not possess parametric stability.

Now let us transform system (6)-(7) into a normal Cauchy form by introducing, for example, additional variables x_3 and x_4 determined by equalities:

$$\left.\begin{array}{c} x_3 = \dot{x}_1 - 2x_1 - x_2 \\ x_4 = \dot{x}_3 \end{array}\right\} \tag{30}$$

As to new variables equation (6) will turn into the following system of three equations of the first order:

$$\left.\begin{array}{c} \dot{x}_1 = -2x_1 + x_2 + x_3 \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = -x_3 - 2x_4 \end{array}\right\} \tag{31}$$

and equation (7) in new variables will become degenerated, it will turn into an equation of a zero order that does not contain derivatives. Really, equation (7) can be written in the form:

$$[(D^2 + 2D)x_1 - Dx_2] + [(2D + 4)x_1 - 2x_2] + x_1 + x_2 = 0 \tag{32}$$

Now while comparing (32) with equalities (30) we see that to the first square bracket corresponds variable x_4, to the second square bracket corresponds $2x_3$ and as a whole equation (32), i.e. equation (7) introduce an expression x_2 by means of $a_1; x_2; x_4$ into (31) and we can obtain a system of three equations of the first order in relation to three variables $x_1; x_3$ and x_4:

$$x_2 = -x_1 - 2x_3 - x_4 \tag{33}$$

i.e. in new variables it degenerates, it stops to be a differential equation (or – which is the same) – it becomes a differential equation of a zero order, an equation that does not have derivatives.

We can easily see that a characteristic polynomial in system (31)-(33) equal to a determinant

$$\begin{vmatrix} 2+\lambda & -1 & -1 & 0 \\ 0 & 0 & \lambda & -1 \\ 0 & 0 & 1 & \lambda \\ 1 & 1 & 2 & 1 \end{vmatrix} = \lambda^3 + 5\lambda^2 + 7\lambda + 3 \tag{34}$$

coincides with polynomial (23) and polynomial (34) and a general solution of system (31)-(33) in relation to, for example, variable x_1 preserves a form of (24). Thus systems (6)-(7) and (31)-(33) are equivalent between themselves.

It can surely put an expression x_2 by means of $x_1; x_3; x_4$ into (31) while using equation (33). Then we shall obtain a system of three equations of the first order in relation to three variables $x_1; x_3; x_4$:

$$\left.\begin{array}{c} \dot{x}_1 = -3x_1 - x_3 - x_4; \\ \bar{x}_3 = x_4; \\ \bar{x}_4 = -x_3 - 2x_4. \end{array}\right\} \qquad 35$$

We can easily see that a characteristic polynomial of system (35) equal to a determinant

$$\begin{vmatrix} \lambda+3 & 1 & 1 \\ 0 & \lambda & -1 \\ 0 & 1 & \lambda+2 \end{vmatrix} = (\lambda+3) \cdot \begin{vmatrix} \lambda & -1 \\ 1 & \lambda+2 \end{vmatrix} = (\lambda+3)(\lambda+1)^2 = \lambda^3 + 5\lambda^2 + 7\lambda + 3 \tag{36}$$

coincides with polynomial (23) and polynomial (34) and a general solution of system (35), in example, relative with variable x_1 preserves a form (24). Thus systems (6)–(7), (31)–(33) and (35) are equivalent between themselves.

But (it can be easily checked) systems (31)-(33) and (35) in difference from system (6)-(7) are parametrically stable, i.e. they remain stable during sufficiently small changes of their any coefficients.

Thus a transformation of system (6)-(7) into a system (31)-(33) or into a system (35) is an example of equivalent transformation that changes parametric stability.

Inverse transformations – transformations of system (31)-(33) or system (35) into a system (6)-(7) are also examples of equivalent transformations that change parametric

stability (in initial systems it has but in system (6)-(7) it has not).

A lot of example of such equivalent transformations that change parametric stability and other properties of transformed systems have been earlier given in [5].

Note that the existence of such examples is quite natural although for a long period of time nobody paid any attention to it. Really in determination itself of equivalent transformation only contains a requirement of preserving unchanged solutions of transformed systems but nothing is said that unchanged remain all properties of solutions (including such an important property as parametric stability).

But the appearance of publications [36], [37], [38] where for the first time it has been shown that equivalent transformations can change many important properties of transformed systems at first led to amazment and sharp discussions. A lot of mathematicians and engineers became accustomed that "equivalent transformations are those that "do not change anything" and it was very difficult to refuse this presentation at once as has become a custom. Examples given in [36], [37], [38] seemed to be strange to manu people and they (seemed) even – paradoxial. They often arose bewilderment and even unacception of published results. Recognition appeared not at once. Partly the cause has been that according to conventional "Mathematics – 1" that dealt only with fixed values of equations coefficients equivalent transformations really "do not change anything". Besides this they simplify equations and make more easy their investigation. In the limits of "Mathematics – 2" that takes into account inevitable small variations of coefficients and parameters (and thus – it is better than in "Mathematics – 1"), taking into account specification of applied problems everything has changed essentially. In "Mathematics – 2" we can apply equivalent transformations with great caution. Incautions use of equivalent transformation can lead (and has led many times) to wreckages and catastrophes some of which were described, for example, in publication [39].

The appearance in publications [36], [37], [38] and in [5] examples of equivalent transformations that change many important properties of solutions helped of depicting "Mathematics – 2" into a separate direction of investigations, into a separate subregion of since. It has become clear that "Mathematics –2" differs from "Mathematics – 1" not only in sphere of examined objects but in methoddogy. In particular it differs in a more cautious applications of equivalent transformations which in "Mathematics – 1"are widely used and without any limits. Somelimes it leads to mistakes during the solution of practical problems which usually can be described by "Mathematics – 2"than "Mathematics – 1".

In order to avoid mistakes and inadmission of new results of investigations (and sometimes – even hostile to them) it is necessary to – once more – stress the existence of "Mathematics – 2"and its difference from conventional "Mathematics – 1". It is necessary to stress that "Mathematics –2" better reflect peculiarities of computing certain objects in technique and physics.

§24. Changes during equivalent transformations in Lyapunov stability.

Note that during equivalent transformation not only parametric stability can change but also – Lyapunov classical stability formulated by a great Russian scientist A.M.Lyapunov (1856-1918) already in 1892. Recollect that a solution of a system of differential equations:

$$\frac{dx_i}{dt} = f_i(t; x_1; x_2; ...; x_n), i = 1, 2, ..., n$$

with initial conditions: $y_i(0) = y_{i0}$ is called stability by Lyapunov if small changes o initial conditions can not arise large changes of solutions. A strict definition of Lyapunov stability is given, for example, in [33], pp. 316-327; in [34], pp. 138-142 and in other publications.

Let us return to equation (8) wuth an initial condition $x(0) = 0$ and a solution $x(t) = 0$. For an initial condition $x(0) = \delta$ a solution will become $x(t) = \delta e^{-i}$ which means that there is Lyapunov stability of a solution $x(t) = 0$. Let us multiply the left and right sides of equation (8) by an operator polynomial $D = 1$. After this multiplication equation (8) will turn into equation.

$$(D + 1)(D - 1)x = \ddot{x} - x = 0 \tag{37}$$

i.e. – into an equation of the second order. In order to obtain a definite solution of an equation we must take into account the second initial condition for equation (37) – i.e. besides already posed the first initial condition x(0)=0 it is necessary to find the second initial condition, a condition for $\dot{x}(0)$, From already earlier obtained solution of equation (8) solution $x(t) = 0$ it follows that the second initial condition for equation (37) can only be a condition $\dot{x}(0) = 0$.

A general solution of equation (37) – if we use a more precise definition – not a "general solution" but a family of solutions that depends on integration constants – will be of the form:

$$x(t) = C_1 e^{-t} + C_2 e^{t} \tag{38}$$

While defining integration constants C_1 and C_2 from initial conditions $x(0) = 0; \dot{x}(0) = 0$ we obtain that $C_1 = C_2 = 0$ and a single solution that satisfies initial conditions is a solution $x(t) = 0$ that coincides with earlier obtained solution of equation (8). Thus equation (8) and equation (37) are equivalent but a solution $x = 0$ of equation (8) is stable by Lyaounov but the same solution $x = 0$ of equation (37) is unstable (unstable in a classical sense, by Lyapunov). Really, if there are zero initial conditions $x(0) - 0$ then $\dot{x}(0) = 0$ will be $x(t) = 0$ but if initial conditions will deviate from zero ones even by infinitely small values δ_1 and δ_2 and instead of $x(0) = 0$ we shall have $x(0) = \delta_1$ instead of $\dot{x}(0) = 0$ we shall have $\dot{x}(0) = \delta_1$ then

$$x(t) = 0,5(\delta_1 - \delta_2)e^{-t} + 0,5(\delta_1 + \delta_2)e^{t} \tag{39}$$

and a difference between solution (39) and solution $x(t) = 0$ will increase without any limits in the course of time.

Thus some of equivalent transformations can change not only parametric stability but a classical stability by Lyapunov.

But this phenomenon (found already in the 30-ths of the XX-th century) did not pick up one's care from investigators. Multiplication by non Hurwitz operator polynomial was recommended not to apply. It was called an unequivalent transformation (although in fact it is not so) and everybody continued to think that equivalent transformations "do not change anything". This old prejudice later has lead – already in the 60ths of the XXth century - to a series of wreckages during the realization of a methodics of "analytical construction of optimal regulators" that has been proposed by a well-known Russian scientist A.M.Lyetov (1911-1974). It was published in a series of articles [48] and then in a monograph [49]. In more details in publication [39] about these wreckages (and well-known other wreckages and catastrophes) you can find in a publication (39). One of sections of which is called "a tragedy of A.M.Lyetov".

Note that a possible change of stability by Lyapunov during equivalent transformations is connected with the fact that A.M.Lyapunov in his investigations considered the behaviour of solutions of differential equations not only during rigidly put initial conditions but during small changes in these conditions as well, i.e. we can consider Lyapunov A.M. one of those who in fact has already come to problems of "Mathematics – 2". But at that time this fact was not realized. And later findings out of examples of changing stability by A.M.Lyapunov during equivalent transformations has not picked up the ears of investigators as we have already noted.

A calm relation to the change of Lyapunov stability while the right and left sides of an equation are multiplied by non Hurwitz polynomial was due to the fact that this change can be easily found out and do not lead to hostiles. They can be easily avoided by a simple forbiddence of multiplying by non Hurwitz operator polynomial. At the same time changes in parametric stability during equivalent transformations about which we have spoken in previous section is much more dangerous and it has led (not once) to wreckages and catastrophes (see [5], pp. 27-29; [11], p.8892 and pp. 139-148; [39], pp. 30-48 etc.)

§25. The change of correctness during equivalent transformations. The third class of mathematical models – that are intermediate between correct and incorrect ones.

Earlier (in the first part of the book) ill-conditioned systems of equations have been considered. Their solutions were essentially changed during small (but by all means-finite!) changes of coefficients.

But system of equations (6)-(7) can greatly change their solutions (as we have seen in this in recent section) – not only during small finite but also during infinitely small variations of some coefficients.

Systems of differential equations that are able during infinitely small relative variations to change (variations) of coefficients into solutions by finite values (or even to change them essentially) we shall call incorrect (recall that "variations of zero" is excluded). In such systems a continuous dependence of solutions on coefficients can be absent.

Systems of equations in which solutions depend on coefficients (or on parameters on which in their turn depend coefficients) and depend continuously are called correct correctly posed systems.

The existence of a class of correct and a class of incorrect systems of equations has been known for a long period of time. First of all French mathematician Hadamar (Hadamar, 1865–1963) introduced a conception of correct and incorrect equations in mathematical physics even in 1902 and later it turned out to be a very important mathematical conception. See, for, example, a known publication [40] and a large bibliography to it.

In 1998 (publication [41]) an existence of one more – third class of problems in physics and technique was found. It is an intermediate class between these problems third class in physics and technique. Really, a system of equations (6)-(7) is incorrect. But if a solution of this system is solved by preliminary transformation to a normal Cauchy form – to a system (35) it turns out that this transformed system is correct.

Thus a problem of computing a solution of a system of equations (6)–(7) can be attributed to the third class – to a class of problems that changes its correctness in the course of equivalent transformations applied during their solution.

The discovery of the third class in problems of physics and technique (see in more details [42]) that is an intermediate between earlier known classes – correct and incorrect – besides theoretical interest turned out to have a large practical value since it allowed to find a method of discovering very dangerous "special" objects which repeatedly become (and, sorry to say has became) causes of wreckages and catastrophes.

Besides incorrect systems of equations naturally there exist incorrect objects (whose mathematical models are these systems).

A solid body in which a mass center lies exactly above an edge of a support is an example of an incorrect object. During infinitely small change of coordinates in a center

of masses such a body can fall.

Here is another example: a system of automatic control that is described by means of a system of differential equations whose characteristic polynomial has besides roots with negative real parts if only one zero root. Such system is theoretically stable by Lyapunov but it is incorrect as infinitely small change of coefficients can make it unstable. This is due to an infinitely small and thus inevitable – in practice change of some coefficients can turn a zero root of a characteristics can turn a zero root of a characteristic polynomial into a root with positive real part.

Surely, incorrect objects are dangerous and they must, not be allowed to realize in order to avoid almost inevitable wreckages and catastrophes. On a step of projection and computation incorrect objects are eliminated on the basis of analizing their mathematical model of a projected object is incorrect then it is necessary to eliminate this object and we must not permit to realize it "in metal" by no means. If on a step of projecting an incorrect object has been allowed to use this fact almost inevitably leads to later wreckage since already "realized in metal" incorrect (or "ill-conditioned") object is very difficult to pick out an object during tests. And in the course of exploitation (later) it will bring a lot of disorders and troubles.

But if an incorrect object is of the third order then an analysis of its mathematical model is not able to discover its incorrectness and then later a wreckage is almost inevitable.

A characteristic example: a system of controlling an electrodrive whose mathematical models can serve both–equations (6)–(7) and equations (31)–(33) that have one and the same characteristic polunomial. If we judge by equations (31)–(33) then a system of control is parametrically stable and correct but if we judge by equations (6)–(7) it is incorrect and parametrically instable. Will an object be correct whose mathematical models can have both-equations (6)–(7) and equations (31)–(33). We cannot determine by these equations. It is necessary to examine more delicate properties of a projected object later we shall speak about their account.

On the basis of such examples already in 1994 in an article [37] the main conclusion was published: neither any investigation of a characteristic polynomial in equations systems nor examining of coefficients matrix if we write of an electrodrive whose mathematical models are equations (31) a controlling interaction x_2 can be formed by good transient processes in correspondence with recommendations from "analytical constructioning of optimal regulators" – in the form of a linear combination of variables $x_1; x_2; x_3$ with constant coefficients of strengthening, i.e. – in the form:

$$x_2 = -k_1 x_2 - k_2 x_2 - k_3 x_3 \tag{40}$$

If we pose that $k_1 = 1; k_2 = 2; k_3 = 1$ then equation (40) becomes (33) and in this case a structural scheme of a control system will become (shown on figure 11) a real behavior of a projected system of control in this case reflects a parametrically stable mathematical model in the form of equations (31)–(33). "A realized in model" projected system with a structural scheme (shown in fig.11) will also be parametrically stable.

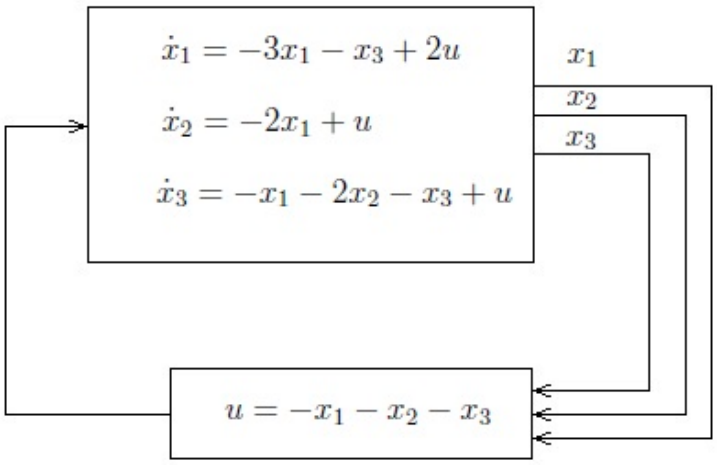

fig. 11

But the forming of an inverse connection in the form of a combination of all variables is often difficult (as it is well-known not all variables can be easily measured). And therefore often an inverse connection is modified by means of changing variables that are difficult to measure by combinations of variables that are easily measured – variables and their derivatives (by means of surely equivalent transformation).

Especially, for an examined electrodrive it is difficult to measure variables x_3 and x_4. By excluding them (with the help of equivalent transformations) we shall obtain a dependence between a control x_2 and a rotation frequency x_1 in the form of equation (7). And equations of an electrodrive (31) in variables x_1 and x_2 will turn into equation (6). A structural scheme of the system in this case will turn into a form given in figure 12 (and different from a scheme shown in figure 11). And a real behavior of a control system during variations of its parameters will better reflect equations (6)–(7) than equations (31)–(33). Although equations (6)–(7) have the same characteristic polynome as equations (31)–(33) they have the same solutions but parametric stability of these solutions (as it has been earlier shown) is different. This means that in real conditions of exploitation during inevitable small variations of parameters the behavior of objects described by equations (6)–(7) and (31)–(33) will be quite different. It can be easily checked during practical tests.

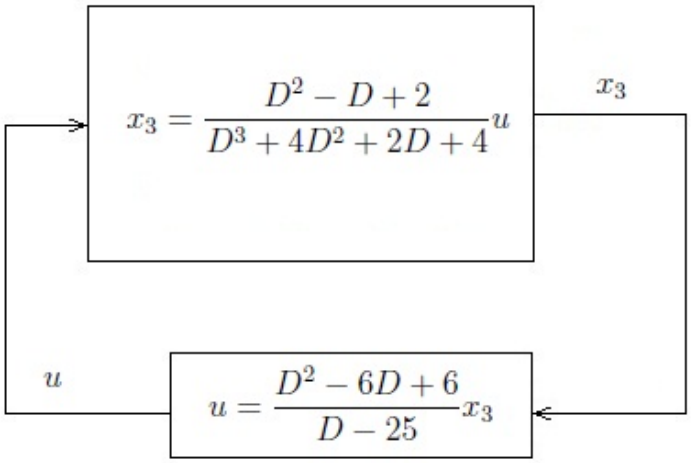

fig. 12

But from the paint of vien of "Mathematics–1" that defines the behavior of objects during inevitable set values of coefficients and parameters (or posed laws of their change) mathematical models in the form of equations (6)–(7) and (31)–(33) cannot be distinguished.

We once more see that "Mathematics–2" better reflects a real behavior of investigated objects and practical requirements to them than "Mathematics–1".

As a conclusion let us note that is a known publication [40] and in many others a somewhat other approach to a problem of incorrectness is used. There the foundation of it lies not in the definition of incorrect mathematical model but in the definition of an incorrect problem.

In these publications incorrect (or – incorrectly posed) problems are called such ones which do not satisfy even one of three conditions:

1. a solution exists
2. the solution is a single one
3. the solution depends on initial data in a conditions way.

This definition cannot be called successful since it "dissolves" real incorrect problems in an immense sea of problems that do not have solutions or that have not a single solution. Equation $x^2 + 1 = 0$ has no solution in the field of real numbers. Equation $\sin x = 0$ has a countless set of solutions $x_0 = n\pi$, where $n = 0; \pm 1; ...; \pm n$ but a problem of computing its solutions cannot be attributed as an incorrect one. In order to avoid contradictions with already a conventional definition applied in [40] we base ourselves – as already we have shown – not on the definition of an incorrect problem but on the definition of an incorrect

mathematical model. In more details you can find this material and its foundation in publication [11].

§26. The existence of Lyapunov function does not guarantee stability.

In this section stability and parametrical stability of both linear and nonlinear equations systems will be examined.

In difference with linear equations with constant coefficients that has been considered in previous sections it is more difficult to check stability (or instability) of solutions of nonlinear equations systems.

The most reliable method of checking stability of solutions in nonlinear equations systems is the forming of the so called "Lyapunov functions". The construction of these functions is very complex, it requiers a large investigation work. But earlier it has been thought that if after great work Lyapunov function is constructed it guarantees stability of an examined solution. In the limits of "Mathematics–1"this point of veiw is quite true. But investigations that has been carried out in the limits of "Mathematics–2"that in a more complete way take into account real requirements to examined objects has shown that in fact everything is not so. In publication [5] it was shown that a system for which Lyapunov function had been constructed could loose stability during infinitely small (and thus – inevitable in practice) variations of coefficients and parameters. Surely, for practical applications stability which looses stability during infinitely small parameters variations is identical to instability. Therefore the existence of Lyapunov function does not guarantee real stability.

Since Lyapunov function (called in nonour of a great Russian mathematican A.M.Lyapunov) who introduced them in his known publication in 1892 and reedited in 1959 (see [50]) is devoted to a very large quanties of investigations and publications. A lot of people are sure that the existence of Lyapunov function guarantees stability. Let us give you some explanations.

Let us consider a system of nonlinear differential equations of the form:

$$\left.\begin{array}{c} \dfrac{dx_1}{dt} = f(x_1; x_2; ...; x_n) \\ ..... \\ \dfrac{dx_n}{dt} = f(x_1; x_2; ...; x_n) \end{array}\right\} \tag{41}$$

where $f(x_1; x_2; ...x_n)$ – functions of arguments $x_1; x_2; ...; x_n$ in a general case they are nonlinear. It is known that by corresponding the change of variables we can reduce a question of stability of any solution of system (41) to the investigation of a zero solution: $x_1 = x_2 = ...; x_n(t) = 0$.

Let us introduce into consideration function V of variables $x_1; x_2; ...; x_n$ which is equal to zero only when all variables are equal to zero. It is positive if variables have all other values of variables. For example, such a function can serve as:

$$V_1 = x_1^2 + x_2^2 + ... + x_n^2 \tag{42}$$

Now let us compute a total derivative of function V in time in relation to solutions of system (41). In order to compute this derivative (due to system (41)) a known formula

for total derivative is used:

$$\frac{dV}{dt} = \frac{\partial V}{\partial x_1} \cdot \frac{dx_1}{dt} + \frac{\partial V}{\partial x_2} \cdot \frac{dx_2}{dt} + ... + \frac{\partial V}{\partial x_n} \cdot \frac{dx_n}{dt} \tag{43}$$

And instead of each of derivatives $\frac{dx}{dt}$ we introduce their values from system (41). After this introduction for a derivative "due" to a system a formula is obtained:

$$\frac{dV}{dt} = \frac{\partial V}{\partial x_1} \cdot f_1(x_1; ...; x_n) + \frac{\partial V}{\partial x_2} \cdot f_2(x_1; ...; x_n) + ... + \frac{\partial V}{\partial x_n} \cdot f_n(x_1; ...x_n) \tag{44}$$

Now let function V be such that derivative (44) for all $x_i \neq 0$ is negative. Such a function is accepted to call Lyapunov function. In 1892 A.M.Lyapunov had proved that if such a function existed then a zero solution of system (41) is stable.

The proof by A.M.Lyapunov admits an obvious interpretation i if a derivative (due to system (44)) is negative then function V can only decrease tending to its smallest value equal to zero and it achieves it if $x_1 = x_2 = ...x_n = 0$. But this means that all variables $x_1(t)$ if $t \to \infty$ will tend to zero. And thus a difference between any solution of system (41) and its zero solution (if $t \to \infty$) will tend to zero and this fact proves that a zero solution is stable.

Here is a simple example for a system of equations:

$$\left.\begin{array}{l} \dot{x}_1 = -x_1 \\ \dot{x}_2 = -x_2 \end{array}\right\} \tag{45}$$

as Lyapunov function we can try to apply the function:

$$V = \frac{1}{2}(x_1^2 + x_2^2) \tag{46}$$

(it is a particular case of square form of variables x_1 and x_2)

Its total derivative is:

$$\frac{dV}{dt} = x_1 \cdot \frac{dx_1}{dt} + x_2 \cdot \frac{dx_2}{dt} \tag{47}$$

"due" to system (45) it becomes:

$$\frac{dV}{dt} = -(x_1^2 + x_2^2) \tag{48}$$

and for any variables x_1 and x_2 besides value $x_1 = 0; x_2 = 0$ it is negative. So it is proved that function (46) is really Lyapunov for system (45). And thus its zero solution is stable.

For such a simple system as (45) we can directly find solutions: $x_1 = C_1 e^{-t}; x_2 = C_2 e^{-t}$ and see that solution $x_1 = 0; x_2 = 0$ is really stable. And this fact coincides with the conclusion obtained on the basis of Lyapunov function.

This material is the most simple part of A.M.Lyapunov theory. More delicate methods have been developed, for example, using Lyapunov function whose derivatives "due to the

system" not without fault will be negative – but only not positive. All these – and many other questions have been considered in a lot of hundred of works devoted to Lyapunov theory (see, for example, publications [43;44] and a wide bibliography in these works).

Note that it is most often very difficult to find Lyapunov function for a certain system. But efforts of hundreds of investigators for a period of tens of years have been directed to finding these functions since it was thought that if Lyapunov function was found then the problem of stability would be solved.

Sorry, all this is only in the limits of "Mathematics–1". More detailed investigations that apply results of "Mathematics–2" at once have shown (see publication [5]) that the existence of Lyapunov function does not guarantee from loosing stability during infinitely small variations of parameters. Really it has been proved for a long period of time that any system of linear differential equations with constant coefficients and Hurwitz characteristic polynomial has Lyapunov function. Thus system (35) also has it. It has been obtained by equivalent transformations from system (6)–(7) which does not possess parametric stability. But if the existence of Lyapunov function does not possess parametric stability. But if the existence of Lyapunov function does not guarantee a real (i.e. parametric) stability even for linear systems and the more its existence cannot guarantee anything for systems of nonlinear differential equations.

Surely, the above said does not discredit outstanding results of A.M.Lyapunov. If coefficients values are not changed in considered systems of equations (i.e. in the limits of "Mathematics–1"), all his results are by all means correct. But if we wish to apply Lyapunov results for the solution of practical problems where small (and also infinitely small, really) variations of coefficients are inevitable then for a reliable judgement about real stability it is necessary to additionally check such equivalent transformations that were used during the finding of Lyapunov function. If we have not carried out such check then while constructing Lyapunov function we obtain false judgements about a real stability of an examined object. Here is an example: during the investigation of an electrodrive whose structural scheme is given in fig. 12. And a mathematical model is a system of equations (6)–(7) we can reduce this system to a system of equations (31) by means of equivalent transformations and we can construct for it Lyapunov function – for system (31) then (by all means) Lyapunov function exists. But the judgement about stability of an examined electrodrive on the basis of existence of this function will be false – as it was shown – system (6)–(7) could loose stability during inevitable in practice infinitely during inevitable in practice infinitely small variations of some its coefficients and such a system is equivalent to instable system – and there is not real stability. The judgement about stability on the basis of Lyapunov function existence in this case is a mistake. The cause is that equivalent transformations of system (6)–(7) into system (31) have changed a parametrical stability of an examined system.

§27. Is a theorem about continuous dependence of solutions of differential equations systems on parameters always true?

A theorem on a continuous dependence of solutions of differential equations and systems of differential equations on parameters is considered to be one of the most important theorems in the theory of differential equations. It is on the basis of all practical applications of the theory since, surely, coefficients and parameters of their different objects and thus parameters of their mathematical models as well–cannot in details correspond to calculated values. They almost always do not remain ideally canstant in the course of exploitation. Variations of parameters, their small changes are inevitable and if there is no continuous dependence of solutions on parameters then their infinitely small variations can lead to large change in solutions and calculation results can turn out to be not at all authentic and unreliable.

In text-book on ordinary differential equations proofs of a theorem about continuous dependence of solutions on parameters are given. Therefore the theorem is considered proved and just for all systems of equations satisfying conditions used during its proof. These conditions in [33] and in other text–books are formulated in the following way. A system of differential equations in a normal Cauchy form (in a form of n equations of the first order) is considered and it is written as:

$$\frac{dx_i}{dt} = f_i(t; x_1; x_2; ...; x_n), (i = 1, 2, ..., n) \tag{49}$$

where λ – parameter, and it is proved that if functions $f_i(t; x_1; x_2; ...x_n; \lambda)$ are continuous and limited – i.e.

$$|f_i(t; x_1; ...; x_n; \lambda)| \leq M \tag{50}$$

where M does not depend on parameter λ and satisfies known Lipshitz conditions (Lipshitz 1832–1903) then all solutions $x(t)$ depend on parameters λ in a continuous way. There also exists a proof for one differential equation of the nth order.

It is important to note that for systems of equations consisting of equations of different orders proofs on continuous dependences of solutions on parameters of solutions on parameters was not given and was not published.

Probably the authors of such text–books thought that such proofs were excessif since a system consisting of equations of different orders can be turned (by equivalent transformations) to normal Cauchy form (without changing solutions) for which everything had been already proved.

But in [5] it was shown that equivalent transformations although they did not change solutions themselves as such could change many properties of investigated systems including a property of a continuous dependence of solutions on parameters.

For example, let us examine a differential equation for a regulated electrodrive. A main equation for an electrodrive – an equation of moments equilibrium on a roller – can be written in the form:

$$T_m \frac{dv}{dt} = M_m - M_i \tag{51}$$

where $M \cdot dV$ – moment in an electrodrive, M_c – a moment of resistance in an excecutive mechanism, v – rotation frequency, T_m – mechanical constant of time in an electrodrive equal to the time of speeding up of an engine from a zero rotation frequency up to a rating one when an engine moment is equal to a rating one and when a resistance moment is equal to zero.

Usually an equation for regulated electrodrives are written in deviations from rating values and time is measured in fractions from T_m. Then equation (51) will become:

$$m\dot{x}_1 = -k_0 x_1 + x_2 + x_3 \tag{52}$$

where m – parameter that is a proportional mechanical time constant for electrodrive and in a rating regime it is equiqual to 1. But in the course of exploitation due to small tensuon oscillation and other causes this parameter can deviate from value $m = 1$ and it can be equal to $m = 1 + \varepsilon$ where ε – can be a small number, x_1 – a deviation of oscillation rotation from rating one, x_2 – a deviation of engine moment, it plays the role of control, x_3 – deviation of resistance moment from rating value, k_0 – coefficient of viscons friction.

If resistance moment is a stationary arbitrary process and a spectral density of power of variable x_3 can be written in the form:

$$S_{x_3} = \frac{1}{(\omega^2 - \alpha^2 - \beta^2)^2 - 4\alpha^2\omega^2} \tag{53}$$

where ω – variable that has a frequency measure 1/sec and α and β – parameters of a stationary processes then equation (52), as in known, can be supplemented by equations:

$$\dot{x}_3 = x_4 \tag{54}$$

$$\dot{x}_4 = -(\alpha^2 + \beta^2)x_3 - 2\alpha x_4 \tag{55}$$

but in order to decrease oscillation in rotation frequency a regulator that works, for example, according to the law is set:

$$x_2 = -k_1 x_1 - k_3 x_3 - k_4 x_4 \tag{56}$$

At the expense of a choice of coefficients $k_1; k_3; k_4$ – regulator "intesity coefficients" a good quality of regulation is acheived.

A system of four equations (52), (54), (55), (56) which (by excluding one of variables by means of relation (56)) can be reduced to three differential equations of the first order – describes transient processes that occur in an electrodrive.

Solutions of systems of equations (52), (54), (55), (56) depend on any of parameters in a continuous way since the limit of right sides and Lipshitz conditions were satisfied.

In order to simply further work let us consider a special case when $k_0 = 2; x = 1; \beta = 0; k_1 = 1; k_2 = 2; k_4 = 1$ and a system of equations becomes:

$$\left.\begin{array}{c} m\dot{x}_1 = -2x_1 + x_2 + x_3 \\ x_2 = -x_1 - 2x_3 - x_4 \\ \dot{x}_3 = x_4 \\ \dot{x_4} = -x_3 - 2x_4 \end{array}\right\} \tag{57}$$

In system (57) the second equations is a differential equation of a zero order that does not contain derivatives.

A characteristic polynomial (57) is equal to determinant:

$$\begin{vmatrix} m\lambda + 2 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 0 & 0 & \lambda & -1 \\ 0 & 0 & 1 & \lambda + 2 \end{vmatrix} = (m\lambda + 3)(\lambda + 1)^2 \tag{58}$$

and when a rating value of parameter $m = 1$ a characteristic polynomial is equal to:

$$\lambda^3 + 5\lambda^2 + 7\lambda + 3 = (\lambda + 3)(\lambda + 1)^2 \tag{59}$$

and a general solution of system (57) becomes:

$$x_i(t)C_1 e^{-\frac{3}{m}t} + C_2 t e^{-t} + C_3 e^{-t} \tag{60}$$

where $C_1; C_2; C_3$ – integration constants that depend on initial conditions.

Equations (56) shows that a controlling interaction of a regulator (variable x_2) is formed in the functions of variables $x_1; x_2; x_4$. But the application of variables x_3 and x_4 in a regulator is difficult and therefore by means of equivalent transformations usually it is transformed (system (57)) into a system consisting of two equations:

$$[mD^3 + (2 + 2m)D^2 + (4 + m)D + 2]x_1 = (D + 1)^2 x_2 \tag{61}$$

$$(D + 1)x_2 = (D^2 + 4D + 5)x_1 \tag{62}$$

Equation (62) shows that now a controlling interaction x_2 is formed in relation to a function of one available variable x_1. It is formed as a solution of equation (62). And this solution is easily realized technically x_1. A characteristic polynomial of system (61)–(62) is equal to a determinant:

$$\begin{vmatrix} m\lambda^3 + (2 + 2m)\lambda^2 + (4 + m)\lambda + 2 & -(\lambda + 1)^2 \\ \lambda^2 + 4\lambda + 5 & -(\lambda + 1) \end{vmatrix} = \\ = (1 - m)\lambda^4 + (4 - 3m)\lambda^3 + (8 - 3m)\lambda^2 + (8 - m)\lambda + 3 \tag{63}$$

when $m = 1$, i.e. if a parameter has a rating value m a characteristic polynomial (63) will be equal to polynomial:

$$\lambda^3 + 5\lambda^2 + 7\lambda + 3 \tag{64}$$

and it will coincide with polynomial (59). This means that a transformation of system (57) into system (61)–(62) was an equivalent transformation. A general solution of system (61)–(62) if $m = 1$ will become:

$$x_i(t) = C_1 e^{-3t} + C_2 t e^{-t} + C_3 e^{-t} \tag{65}$$

Surely, we can propose that $m = 1$ in system (57) and by means of equivalent transformations turn it into a system of two equations with two variable x_1 and x_2. We shall obtain a system:

$$\left.\begin{array}{c}(D^3 + 4D^2 + 5D + 2)x_1 = (D+1)^2 x_2 \\ (D+1)x_2 = (D^2 + 4D + 5)x_1\end{array}\right\} \tag{66}$$

with a characteristic polynomial (59) that once more will ascertain the equivalence of a transformation.

Now let us put a main question: if parameter m deviates from a rating value $m = 1$ by infinitely small value ε and has become equal to $1 + \varepsilon$ then will solutions of a system (61)–(62) depend on parameter m in a continuous way everywhere including if $m = 1$ as well? If we introduce a value $m = 1 + \varepsilon$ into equations (61)–(62) and after having computed a characteristic polynomial we shall obtain:

$$-\varepsilon\lambda^4 + (1 - 3\varepsilon)\lambda^3 + (5 - 3\varepsilon)\lambda^2 + (7 - 3)\lambda + 3. \tag{67}$$

A polynomial (67) if ε is small, first of all, will have three negative roots that differ from roots $\lambda_1 = -3; \lambda_2 = \lambda_3 = -1$ not much of polynomial (64) and secondly, it will have a very large positive (if $\varepsilon > 0$) and negative (when $\varepsilon < 0$) root that approximately (for small ε) equal to $\frac{1}{\varepsilon}$. If it is equal to $C_4 e^{\frac{1}{\varepsilon}}$ where C_4 – the fourth integration constant but if $\varepsilon < 0$ this fourth member will very quickly decrease.

A a result system (61)–(62) will have a break if $m = 1$ in dependence of solutions on parameter m. Then there will be no continuous dependence of solutions on parameters.

We have carried out a detailed conclusion: systems of equations (61)–(62) that describe processes that occur in a regulated electrodrive so as to show that systems of differential equations in which there is no continuous dependence of solutions on parameters are not exceptional artifically invented systems. There is a lot of such systems and they describe a real behavior of many quite real objects.

A possible characteristic simpton in systems of differential equations that have no continuous dependence of solutions on parameters is in the following. Some of equations from a system are the so called noncanonic, i.e. – equations in which an order of derivatives in the right side is equal to or is more large than an order of derivatives in the left side. In system (61)–(62) equation (61) is canonic and equation (62) – noncanonic.

In text-books published in Russian noncanonic equations were considered probably even in 1927 in a book V.A.Stekhlov [51] (Stekhlov V.A. 1864-1926). Later almost exclusively systems of equations in a normal. Cauchy form, in the form of n canonic equations of the first order in which in the left side are derivatives of the first order and in the right side – derivatives of a zero order that is by a unity less were considered. Noncanonic equations were in the most forgotten and probably, in vain. They appear in applications and for nothing an outstanding Russian mathematician A. Stekhlov (who

was an academician) paid them serious attention.

Note that these facts do not refute any proved theorems. They sooner indicate prejudicies that exist oven in mathematics, sorry to say. The fact that in a system consisting of n differential equations of the first order (and also in differential equations of the nth order) solutions depend on parameters in a continuous way – is a proved theorem and its proof does not arise any dobts. But a wide belief (it is not rejected in text-book) that in any system consisting of equations of different orders (if, surely, limitation conditions and Lipshitz conditions are satisfied) solutions by all means depend on parameters in a continuous way – this belief is not ascertained by a proof. It is not proved and it is false. It is only a widely admitted prejudice (a false statement) based on other prejudice – on a statement that equivalent transformations of equations systems as if "do not change anything". In fact these transformations do not change solutions of equations but they can change many important properties of solutions – such as stability, correctness, continuous dependence of solutions on coefficients and parameters. In [5] we spoke about this earlier.

Note that a prejudice about a continuous dependence of solutions on parameters in any systems of differential equations (as many other prejudices) are not at all harmless. An engineer or a man using computers that has met with a mathematical model of an object in the course of his work in which these is no continuous dependence of solutions on parameters and who has not checked this fact this man can obtain false results in his computations that can later become a cause of wreckages and even catastrophes. We must also note that popular packages of applied programs that include the solution of differential equations – such as MATLAB, Mathcad etc. – do not isolate such systems in which there is no continuous dependence of solutions on parameters. And therefore a practical application of these programs without additional investigation can lead to mistaken results in computations with all cosequences that follow. They are gross. Already in [52] and [39] we have spoken about them.

Often mathematics is considered (and without sufficient base) consider it erreproachably a strict science all assertions of which are strictly proved and true in fact this is not so. There were many false statements and falsely proved theorems in the history of mathematics. A lot of examples, can be found in, for example, in [47], [54].

In the course of historic development its conceprions were made more precise. Such theorems that were said (falsely) to be proved were rejected. It is, surely, so. But we cannot (by no means) confirm that up to now this process is finished and there no prejudices remained in mathematics and that there is no mistaken statements. The process of development including making it more precise of mathematics continues and we must not be surprised at it. Rather we are surprised at something different – that not sufficient attention is paid during last tens of years to "contre-examples", i.e. examples that are rejecting theorems that were considered proved but in fact that they are false.

Here is an example from history of mathematics. In 1821 in his "course of analysis" a great French mathematics O.Cauchy has given a proof of a theorem: "A sum of connected row of continuous functions is continuous". With the Cauchy proof his colleages mathematics agreed. But in 1826 N.Abel (N.Abel, 1802-1829) formed a contre-example. He gave the following example:

$$S = \sin\alpha - \frac{1}{2}\sin 2\alpha + \frac{1}{3}\sin 3\alpha - ... \tag{68}$$

which was a meeting one and consisted of continuous functions but its sum had breakings. Note that N.Able did not try to find a mistake in Cauchy-proof. He only constructed a contre example and besides all authority of great O.Cauchy his theorem was at once recognized to be false.

And only after many years clasped the theorem has received a more precised (and this time a correct one) formulation. "A sum of an uniformly meeting rows of continuous functions is continuous". In this formulation this theorem is given in text-books see, for example, [53], pp.442-449. If there is no uniform meeting a sum of the row can be continuous but it can also not be. Contre-examples play a very large role in applied mathematics that has allowed us to clear it from false and thus dangerous theorems for their practical usage. It is very difficult to find a mistake in the proof of some theorem. Really mathematics clears itself from false and thus dangerous most often by means of contre-examples. Not for nothing a Hungarian mathematician D.Polia (D.Polia 1887-1985) thought that mathematics consists of two parts – theorems and contre-examples. In more details about contre-examples and their role your can see in [47], pp. 123-125 and in [54].

But contre-example can play their important (and very important!) role where after the publication of a contre-example a theorem is recognized as a false one and it is changed into a correct theorem. Up to recent ten years of the XXth century in Russia it was a way – and it was one of necessary components of its successful development. But starting from approximately 1990 much has changed. Financing of science was reduced, many outstanding scientists went abroad to foreign countries and a hard material position of those who remained in Russian - all this led to a wish to make easier for a part of scientists their life. And as a result – not to attentively follow now current publications, not to react to published contre-examples and even-not to discuss them. As a result – association has been broken, many scientific achievements were not applied and the development of science stopped (in details – see in [55]).

A characteristic example – there are a lot of many statements given in this book that have already been published eartier even in 1991-99 it was stated and published that:

1. There are no such investigations of a characteristic polynomial or coefficients matrixes in equations systems that cannot guarantee a correct conclusion about stability of a system without the analysis of equivalent transformations by the help of which they were obtained.

2. The presence of Lyapunov function does not guarantee its stability (a real stability that is inevitable – during infinitely small deviations of their parameters from, calculated values)

3. Not in all systems of differential equations (that even satisfy boundary conditions and Lipshitz conditions) solutions depend in a continuous way on parameters (see publications [5], [36], [37], [38]). A little later the same problems were discussed (in details) in [41], [42], [55], [56], [11], [57], [58].

But up to 2011 only very few firms have applied published results, others continued to make computations and to project "as of old" without taking into account these (publications and as a result wreckages (and even catastrophes!) account these publications and as a result wreckages occured which could be easily avoided. These wreckages have been considered in more details in [39].

This sad fact in great degree is due to a system of scientific information has been broken in Russia beginning from 1990. The circulation of scientific magazines has lowered by ten and more times. Therefore even undisputed successes of Russian scientists are in practice almost is out of reach for the majoruty of engineers and thus they remain not realiazed.

In this book we once more tried to return to problems which had been partly considered in previous publications. But now we consider them according to a unique methodics, as components of "Mathematics–2".

By presenting contre-examples that prove faultness of some popular theorems or conventional computation methods it is advisable to indicate how it is necessary to change incorrect theorems or computation methods in order that they become correct and reliable. This will be carried out in next sections.

Besides we must pick out a theorem about a continuous dependence of solutions of systems of differential equations on parameters. In text-book [33], [34] etc. it is stated that solutions depend on parameters in a continuous way if right sides are continuous, limited and satisfied Lipshitz conditions. Our contre-example (earlier published in [5]) has shown that in such a formulation this theorem is false and therefore it is necessary to introduce additional conditions to which examined systems of differential equations must satisfy that in a general case consist of equations of different orders.

Probably, a theorem will become correct if to conditions of continnity and limitation and Lipshitz conditions we add the following condition: a system of equations must be canonic in a sense of academician V.A.Stekhlov given in [51], i.e. – in an examined system a degree of derivatives that are in a right side of each equation in a system must be lower than a degree in derivatives of the left side.

But in such a question as an exact formulation of one of the most important theorem in the theory of differential equations the author refrains from unquestionable judgement. It is advisable if a final formulation of the theorem be coordinated with specialists after the discussion among many interested experts.

§28. The dependence between object parameters and coefficients in its mathematical model.

In previous sections we have considered equations systems that are mathematical models of real objects. We also picked out such systems of equations that loose stability during infinitely small variations of some of coefficients.

It real techniqual objects in the course of their exploitation undergo variations of parameters in an object and as a result of parameters variations appear variations of coefficients in a mathematical model.

The dependences of coefficients variations on variations of parameters can be rather strange. In a previous section on an example of equations (57) and (61)–(62) we can see dependences of coefficients in mathematic models of an electrodrive on one of parameters of an examined object – on a mechanical time constant m. We can follow coefficients of a mathematical model if we take into account other parameters of a regulated electrodrive – their dependence on coefficients of a viscous friction k_0 in equation (52) – coefficients $k_1; k_3; k_4$ in equation (56).

If we take into account these parameters a mathematical model of an electrodrive has become equations (52), (54), (55), (56). Supposing (as earlier) that $\alpha = 1; \beta = 0$ we obtain that a system characteristic polynomial has become equal to a determinant

$$\begin{vmatrix} m\lambda + k_0 & -1 & -1 & 0 \\ 0 & 0 & \lambda & -1 \\ 0 & 0 & 1 & \lambda + 2 \\ k_1 & 1 & k_2 & k_3 \end{vmatrix} = (m\lambda + k_0 + k_1)(\lambda + 1)^2 \tag{69}$$

From (69) it is at once seen that if coefficients m, k_{ij}, k_i are any and positive all roots of a characteristic polynomial are real and negative. This means that a system is stable and preserves stability during variations of parameters – a mechanical time constant m, coefficients k_{ij}, k_i. Thus the system possesses a parametric stability (recall that we call parametrically stable systems of differential equations – and objects that they describe – which preserve stability during variations of parameters). Note that coefficients k_2 and k_3 do not influence system stability.

Quite another picture is obtained after equivalent transformations when there is unmeasurable variables x_3 and x_4 they are changed by variables x_1, x_2 and their derivatives. The change of variables is carried out in this way from equation (52) follows (where $D = \dfrac{d}{dt}$ is a differentiation operator):

$$x_3 = (mD + k_0)x_1 - x_2 \tag{70}$$

If we take into account equation (54) we have:

$$x_4 = (mD^2 + k_0 D)x_1 - Dx_2 \tag{71}$$

If we put (70) and (71) into equations (52) and (56) we shall obtain an equivalent to system (57) system of equations:

$$[mD^3 + (2m + k_0)D^2 + (m + 2k_0)D + k_0]x_1 = (D^2 + 2D + 1)x_2 \tag{72}$$

$$[k_3mD^2 + (k_2m + k_3k_0)D + k_2k_0 + k_1]x_1 = (k_3D + k_2 - 1)x_2 \tag{73}$$

In system (72)–(73) equation (72) is an equation of a control object, an equation of an electrodrive and equation (73) is an equation of a regulator.

It is important that parameters of electrodrive in the course of its exploitation can change indipendenly from a regulator parameters since they are two different techniqual mechanisms.

It is advised to start from considering the most simple case. Let us suppose that parameters of a regulator have remained unchanged and are equal to their rating values (for rating values of parameters let us suppose that: $m = 1; k_0 = 2; k_1 = 1; k_2 = 2; k_3 = 1$) and parameters of an electrodrive have changed and have become $m = 1 + \varepsilon; k_0 = 2 + \delta$ where ε and δ – number that are small in comparison with a unity. If for this the simpliest but possible combination of parameters variations of a control object and a regulator stability disappears then this means that system (72)–(73) apriori does not possess a parametric stability. Recall that parametrically stable is called such a system in which any possible combination of coefficients and parameters variations does not lead to the loss of stability.

We can check a parametric stability for a particular case by computing a characteristic polynomial of system (72)–(73) if parameters $m; k_0; k_1; k_2; k_3$ in equation (73) remain equal to their rating values. In this case a characteristic polynomial will be equal to a determinant

$$\begin{vmatrix} m\lambda^3 + (2m + k_0)\lambda^2 + (m + k_0)\lambda + k_0 & \lambda^2 + 2\lambda + 1 \\ \lambda^2 + 4\lambda + 5 & \lambda + 1 \end{vmatrix} = \tag{74}$$
$$= (1 - m)\lambda^4 + (6 - 3m - k_0)\lambda^3 + (14 - 3m - 3k_0)\lambda^2 + (9 - m - 3k_0)\lambda + 5 - k_0$$

If $m = 1 + \varepsilon$ and $k_0 = 2 + \delta$ then a characteristic polynomial (74) becomes:

$$-\lambda^4 + (1 - 3\varepsilon - \delta)\lambda^3 + (5 - 3\varepsilon - 3\delta)\lambda^2 + (7 - \varepsilon - 3\delta)\lambda + 3 - \delta \tag{75}$$

and this ascertains that during infinitely small variations of parameter m stability of the system can disappear since if $\varepsilon > 0$ a necessary condition of stability is broken – a positivity of all coefficients of a characteristic polynomial (Stodola condition).

When $\varepsilon > 0$ (i.e. if $m > 1$) in solutions of equations (72)–(73) there appeared swiftly increasing exponential members of the form $C_1e^{fi\varepsilon}$. Deviations of rotation frequences and engine movement from rating value (variables x_1 and x_2) increase very quickly.

At the same time if $\varepsilon = \delta = 0$ a polynomial (75) is Hurwitz and it coincides with polynomial (69). This fact once more ascertains that systems of equations (52), (54), (55), (56) and (72)–(73) are equivalent between themselves (in a classical sense) and they are obtained one from the other by equivalent transformations.

At the same time a problem of checking stability for systems (52), (54), (55), (56) is correct. But for equivalent to it system (72)–(73) it is incorrect. Really if in system (72)–(74) will be $m = 1 + \varepsilon$ then if $\varepsilon < 0$ the system is stable but during infinitely small $\varepsilon > 0$ it is unstable. Thus equivalent transformations have changed the correctness of a solved problem.

This example (and in publication [41] many other similar examples are given) at once has shown that an existing up to 1998 division of all problems in mathematics, physics and technique into two classes – into a class of correct and a class of incorrect problems is insufficient. There exists one more (a very cruel class) – third class – a class of problems that changes its correctness in the course of equivalent transformations during their solution.

Cruelty (and a practical importance) of the third class of problems is that for them conventional methods of solution that do not take into account recently found at St.Petersburg State University new properties of equivalent transformations almost always lead to mistaken results. And due to mistakes in computations wreckages and even catastrophes can become (and have become).

Really, let us consider a system of control for an electrodrive whose mathematical model are equations (72)–(73). In order to investigate stability (and parametric stability) of this system it is recommended (and is realized in packages of applied programs MATLAB, Mathcad and others) to use the following approach: to reduce a system to a normal form and investigate its parametric stability by checking signs of real parts of roots of a characteristic polynomial (69) during the "swimming" of parameters of an examined object or – coefficients of a characteristic polynomial. During such checking method (we recommend it!) during any investigation of polynomial (69) inevitably a conclusion will be made about a good parametric stability of a computed system and it will be recommended to make it "in metal". In fact a reserve of stability of this system (a reserve in parameters variations) will be every small. It will be determined only by small deviations of real values of parameters from computed ones (if there are no deviations then a reserve of stability turns into zero). In the course of exploitation if there is inevitable small wear out of all details that have led to small changes of coefficients in a mathematical model an electrodrive can loose stability in an unexpected time moment and it can break out such an object and can lead to wreckages and even catastrophes an which it has been mounted. As we know electrodrives are mounted on quite different, sometimes very responsible objects – on aeroplanes, vessels, atomic electrostations etc. Therefore the security of reliability in calculations on computers and reliability of technical computations as a whole is an important practical problem if we take into account new properties of equivalent transformations that have been recently found at St.Petersburg State University. This fact will help avoid dangerous wreckages and catastrophes.

Note at once that technological objects during whose computation conventional computation methods (that do not take into account recently found new properties of equivalent transformations) lead to mistakes. They rarely occur. They were proposed to be called "peculiar" objects (not for nothing). For the majority of objects conventional computations give true results. This fact is ascertained by practice during many years. Just the same can be said about equivalent transformations that are everywhere applied during computations. Only in rare cases they change correctness of a solved problem.

Therefore "peculiar" objects and new properties of equivalent transformations have been found so late – only at the end of the XXth century. The first book on this question is in [35], pp. 220-230 published in 1987.

Rarity of "peculiar" objects makes it difficult to install modernized computation methods which take into account new recently found properties of equivalent transformations and do not lead to mistakes while meeting with "peculiar" objects.

The majority of engineers and investigates have not met with "particular" objects during their whole life. They sincerely do not understand what for they must investigate modernized calculation methods that take into account the possibility of such a meeting. But during recent tens of years optimization methods of technical objects are applied more widely. While using them "peculiar" objects occur more and more often. In [39] it was already said about many wreckages that had occured during the application of one the first optimization methods – "analytical construction of optimal regulators" proposed by A.M.Lietov in 1960 (articles [48]).

A little later wonderful optimal regulators that allow to substantially improve the quality of work of many objects were proposed in works by V.B.Larin, V.J.Naumenko, V.N.Suntzev (see [58]). But a direct realization of these regulators turned out to be impossible since they often led to the appearance of "peculiar" systems with unexpected loss of stability and other drawbacks. And thus – to wreckages and in themselves they have for a long period of time harmed the reputation of optimal control in the eyes of engineers.

I think that it is not necessary to refuse from optimizing technical objects by no means. Optimization allows us to create technical objects with the best possible quality of their work. Difficulties that are connected with the fact that among optimal systems most often occur "peculiar" objects we must overcome on the basis of methods proposed in this book and in earlier publications [5; 11; 35-38; 41; 42; 55; 57]. Note that if earlier intuition helped to discover "peculiar" objects for experienced engineers then after calculations started by means of computers meeting with "peculiar" objects became the most dangerous. "Peculiar" objects occur rarely but almost each unexpected meeting with such object leads to wreckage or even to catastrophes. Wreckages and catastrophes also occur not very rarely, not each day. But we must not reconcile with them. And especially – when the cause has been found and it can be easily removed. Just in this way it is with wreckages that occur due to errors in projection and in computation.

In order to remove them it is sufficient to apply modernized computation methods. They do not require substantial financial losses. But modernized methods at present are applied much more rare than it is necessary. Although at Moscow technical university named after N.E.Bauman (MTU) they are widely applied and they are succesful. We hope that other universities and scientific unions will take an example from them.

The possibility of correctness checking by coefficients of a mathematical model

The most reliable method of checking correctness is to analysize the influence of parameters in an examined object on a solution. For an examined example with control system for an electrodrive parameters are a mechanical time constant m and coefficients of intensifying a regulator $k_1; k_2; k_4$. But coefficients of a finally choiced mathematical

model model calculated on a computer can in a rather complex way depend on these parameters. Here is an example: the dependence of coefficients in equations (72)–(73) of a mathematical model for an electrodrive on parameters m, k_0, k_1, k_3, k_4.

Therefore the following question arises: whether it is not possible to carry out the check of correctness simply – to check the influence on a solution by already not by parameters variations of an object but by a variation of coefficients of a mathematical model. Let the behavior of an examined object depend on l parameters $(m_1; m_2; ...; m_l)$ and equations for a mathematical model are reduced to the form that depends on k coefficients $(n_1; n_2; ...; n_k)$. An example: the behavior of a control system considered in previous section of an electrodrive depends on five parameters: m, k_0, k_1, k_2, k_4 but in an equation of its mathematical model – equation (72)–(73) thirteen coefficients enter. Each of coefficients $n_1, n_2, ..., n_k$ can turn out to be a function – in relation to all object parameters:

$$n_i = f_k(m_1, m_2, ..., m_l) \tag{76}$$

By differentiating any of equalities (76) we obtain an equality for the first differentials

$$dn_i = \sum_{j=1}^{l} \frac{df_k}{dm_j} \cdot dm_j. \tag{77}$$

If all functions (76) are continuous in relation to all variables then from equations (76) it follows that to infinitely small variations of parameters correspond infinitely small variations of coefficients n_i (the same is true for a particular case when $\frac{df_i}{dm_i} = 0$ and coefficient n does not depend on parameters. In this case its variation and a differential dn is equal to zero. Therefore if during infinitely small change of only one of coefficients a solution will change by a finite (or even by infinitely large) value then a solution is almost (for certain) incorrect. In order to secure the more reliability, surely, it is useful to check the behavior of the solution during variations of parameters of an object. If only one of function f_i is not continuous and during infinitely small changes of an argument changes by a finite or infinitely large value then this means that to infinitely small changes of object parameters correspond finite (or infinitely large) changes of coefficients in a mathematical model. This once again speaks about an incorrectness of a solution.

Thus correctness of a solution can be checked by means of mathematical model coefficients. It is more difficult to check the good or bad stiputaion of a solution since relative changes of coefficients can be greater (or less) of changes in object parameters. Here is an example. During the change of coefficients m in a control system for an electrodrive from a value $m = 1$ up to $m = 1,01$ – i.e. by one percent – the first coefficient in equation (73) will change from value k_3 up to $k_3 + 0,01k_3$, i.e. by $k_3\%$ or by k_3 times more than the change of parameter m.

§29. Wreckages and catastrophes connected with the unperfection of computation methods. Their peculiarities.

There are lots of causes for wreckages. One of causes (as it was already said) are mistakes and inexactnesses during projection and computation. What is the share of wreckages that occur due to this cause in a general quantity of all wreckages? Sorry to say, this most important question has been investigated in a bad way. Only recently experts from TZNII named after N.P.Mielnikov and Moscow "City centre of experts" have carried out such investigation in the sphere of construction. They have established that 9,3% of all fall-down buildings were due to mistakes in projection and computations (this fact was published in "Izvestiya" , its application, №192, 17.10.06) 9,3% from all number of wreckages is not at all little. Note that this takes place in the field of civil construction where all constructions are comparatively simple and computation methods have been developed long ago. In the field of automatics and also in aviation a share of wreckages that occur due to mistakes and inexactnesses of computation methods is much higher. Since now the majority of calculations is carried out by computers then in this book we first of all speak about mistakes and errors in computer calculations and about securing their reliability.

One of the source of mistakes during computer calculations is that a possibility of an essential change in reserves of solutions stability has not been valued during equivalent transformations. Up to now a prendice has been widely distributed that "equivalent transformations do not change anything". It is not based on any proofs.

Therefore a mathematical model of a projected object is calmly reduced (by means of equivalent transformations) to the most convenient for investigation form and thus by it they compute stability reserves and secure work of an object during inevitable in the course of exploitation small deviations of its parameters from computed ones. During such an approach mistaken estimates of admissible deviations of parameters appeared. We have already spoken about this in previous sections And later these mistaken estimates have led to wreckages and catastrophes.

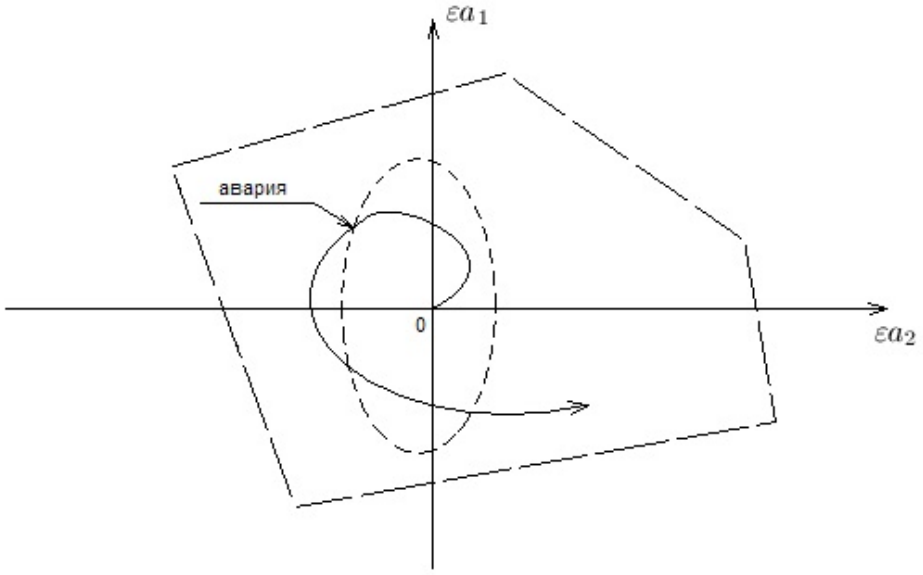

fig. 13

Let us analyze characteristic traits of such wreckages. For example, let us consider a particular case – the influence of two parameters on the work of an examined object – a_1 and a_2. For an electrodrive it can, for example, be a mechanical time constant m and a coefficients of intensifying a regulator k_1 in equations (72) and (73). Let us consider a plane of coordinates where on axises Ox and Oy are put deviations of these parameters εa_1 and εa_2 from their rating values (figure 13). Since deviations of real parameters of rating ones are inevitable and in the course of exploitation they in the course of time usually gradually increase then the behavior of an object on fig.13 will be expressed by a trajectory that will go out from the beginning of coordinates – from a point $\varepsilon a_1 = 0; \varepsilon a_2 = 0$ but which in the course of time it will gradually "untwist" and more and more it will move away from the beginning of coordinates. Therefore during the projection and computation of any responsible object a space of admissible deviations of parameters a_1 and a_2 are calculated. These deviations do not still lead to violation of a normal work of the object during a standard time of its exploitation. On figure 13 this space is outlined by a hatch-dotted curve. The magnitude of this space is computed in such a way that during standard time of the object work (for, example, during thirty years) deviations $a_1 - a_{1nom}$ and $a_2 - a_{norm}$ apriori have not gone away from the limits of this space. But if recently found new properties of equivalent transformations have not been taken into account the computation can turn out to be mistaken i(n previous sections see examples) and a real space of a reliable work of a projected object can turn out to be much less than a computed one. On figure 13 a real space is outlined by a dotted line. Since a trajectory representling in figure 13 a behavior of an object in the course of time will come out of the boundaries of real space during a normal work outlined by a dotted line a wreckage will occur. If this wreckage will not outgrow into a catasrophe (for example, a defence will switch it off) an object continues to work then at the moment of a check of an object after the wreckage it can turn out that the trajectory during this time will again return to a safe space outlined

(on figure 4) by a dotted line and then a check will show that an object is in good repair and works well?

Thus there exists a very characteristic peculiarity that help to isolate such wreckages that occur due to the fact that recently found new properties of equivalent transformations were not taken into account from wreckages that occur due to other causes.

Another characteristic peculiarity in the loss of stability of "peculiar" objecs is accomplished by a very quick swift deviation of regulated values from their rating values. During the analysis of equations (72)–(73) that are mathematical models of one of "peculiar" objects we have already noted that after the loss of stability in solutions x_1 and x_2 appeared members of the form $C_1 e^{\frac{2}{\pi}}$ that are swiftly increasing.

While taking into account these peculiarities let us consider known wreckages in aerobuses of the type $A-310$. These are large passangers aeroplanes manufactured at Franco-German united company for constructing aeroplanes whose office is situated in Tuluza (France).

One of the most known wreckages of aeroplanes accurred on the 22nd of March, 1994 near Miezduriechinsk (Russia) where all passangers and the crew perished. The so called "black boxes" (a self recording board mechanism) in which parameters of a flight and talks of the crew are registered were found and deciphered. This allowed to find out that before a wreckage and during it the plane flew in an automatic regime under the control of an autopilot. Suddenly very quickly deviation of a crepe and tangage from rating values occurred. While the crew tried to turn from an automatic regime to a hand control the deviation have increased in such a way that it was impossible to return them into normal forms at all. The aerobus has fallen and perished.

After several months passed another aerobus $A-310$ flew near Buhkarest also in an automatic regime under the control of an autopilot. But in this case a pilot was able to correctly react. He quickly switched off an autopilot and in a regime of a hand control succeeded in levelling the plane. When after happy landing autopilot and control system has been checked then it turned out that they were in perfect order and that they work well.

Comparison of these facts allows us to make the following conclusion. A system of automatic control at the flight of aerobus – $A-310$ turns out to be a "peailiar" system that possesses small stability resaurces in variations of parameter of an autopilot. And these variations became a cause of two losses of stability one of which (near Miezduriechinsk) finished by the perish of passengers and the crew. It seems to me that the computation of autopilot and system of control has been carried out on computers with transformations of equations in its mathematical model to a normal Cauchy form. This fact has not allowed to disclose dangerous properties of a projected system.

The investigation of a cause of a catastrophe (near Miezduriechinsk) was carried out by Interstate aviation commetee (MAC) commission since the aeroplane had been constructed by France-German company but in that fatal flight a Russian crew led it. On the conclusion of MAC depended who would pay a large financial sum of compensations

to relatives of the perished (that is – approximately, 70 millions of dollars). If Mac would acknowledged that the cause of a wreckage were mistakes in the aeroplane systems then Franco-German firm must pay. But if the cause of the wreckage will be declazed mistakes of the crew then Russia would have to pay. Therefore the investigation of this catastrophe (and the investigation of all other catastrophes and wreckages – see the book [39]) has been complicated by a mercenary interest of influenced organizations and persons who had made a serious influence on those who investigated – including even large bribes. Therefore in the course of investigations often no truth is sought but – wham it is more convenient to accure. Most often they try to put an accusation on pilots especially if they had perished and could not object to it. In such a way has been carried out the investigation of the catastrophe of $A-310$ near Miezduriechinsk. The investigators have found falt with the information obtained from the "black box". It was known that the crew had permitted children to come into the cabin and supposedly "their playing with the steering control" had become the cause of the catastrophe. But it can be known from board selfrecorders that the plane before the wreckage and during it was flying under the autopilot and only due to this fact the crew had let children be at the cabin and allowed them "to play" with the inactive steering control. This fact could not be the cause of the catastrophe by no means though the instructions for flying rules have been broken. But MAC has made the conclusion that the crew is to blame as they have allowed to come children into the cabin. So Russia had to pay more than 70 millions of dollars. The main problem does not lie in finances but that methods of modernizing computations of aeroplanes systems were neglected not using already found new properties of equivalent transformations and all the guilt for the wreckage of a plane had been placed on perished pilots. This has led to a series of wreckages of aeroplanes $A-310$ and $A-320$. These wreckages and catastrophes are in detail described in [5], p. 28-29. These were catastrophes that could have been avoided, much people could be saved.

Later after in 1999 has been published the first edition of a book [5] in which a true cause of the catastrophe near Miezduriechinsk has been in details and with proofs given MAC have changed its conclusion. As they could not dispute with the argumentation in the book and at the same time they do not wish to be a too partial organization MAC took off the guilt from pilots. MAC has recognized that the cause of the catastrophe was defects in aeroplane systems, small resources of stability in planes. But this recognition MAC have secreted from everybody and made it a mystery up to 2006. This mystery has been made known by journalists (I don't known how they have dove this) and the statement about a changed solution of MAC has been published in a newspaper "Izvestya" №139, 03, VIII – 2006. But since a new "conclusion" of MAC has been thoroughly hided at their time it has not shown any influence on the computation and projection methodics and catastrophes of aeroplanes continued.

The author has spoken so much about the investigation of catastrophes (and in many details) because those who wold hear over the television about different "conclusions" (not once) it is necessary to remember about their partiarity , "one-sidedness" of their conclusions about causes of catastrophes. It is necessary to remember what powerful forces presses on experts that investigate wreckages, how difficult it is for experienced and litterate specialists to preserve impartiality. Therefore usually true causes of wreckages and catastrophes became known only after several years have passed, and sometimes– even during many years.

There is a characteristic example – the perish of a supersonic passager aeroplane Tu–144 on the 3d of June, 1973. This plane has been a pride of USSR leaders. It had to be compared with French plane "Concorde" and be acknowledged as the best one. The plane Tu-144 was sent to demonstrate its possibilities to the largest aviasalon near Paris. But during the first demonstrative flight from an aerodrom in Le Burge it has lost its stability and crushed. How many causes for this catastrophe has been given in these years! And only after 34 years elapsed N. Uprov and A. Burtzev who have worked at KB of Tupolev for many years told us about a true cause – the losses in the work of an aeroplane automatics. As a plane $A-310$ (near Miezduriechinsl) has come into a sharp peak and while coming out of it has obtained a critical overload and has crushed in the air. Their text has been published in the newspaper "Izvestiya" on the 6th of July, 2002. Again the cause of the crush was the knock down (a false work) in control systems. It is probable that for Tu-144 these systems were projected on the basis of a methodics of "analytical constructioning" by A. M. Lyetov that had been very popular in these years. As we have already said in publication [39] especially often "peculiar" systems occured with small reserves of stability that could not be discovered during computations since new properties of equivalent transformations, and in partial ar, their ability for changing stability reseves of solutions and reserves of stability in aeroplanes systems were not found at that time.

§30. The explanation of difficulties during the exposing new properties of equivalent transformations and the existence of "peculiar" systems.

It seems strange to understand why new properties of equivalent transformations, the possibilities of changing important properties of solutions during these transformations were discovered so late – only at the end of the XXth century. The main cause of this phenomenon although mathematicians had been using equivalent transformations since the IXth century (yes, already since the IXth century, since the time of Al Horesnei (787–850) it was for a long period of time only stressed that at that time it was considered that equivalent transformations "did not change anything". Therefore you can apply them without any restrictions. For a long period of time – up to recent time – up to recent decade of the XXth century nobody noticed that equivalent transformations although they do not change solutions as such they can change a lot of important properties of solutions – such as correctness, parametric stability, reserves of stability etc.

The cause of the fact that for so a long period of time nobody noticed this because most often equivalent transformations really "do not change" anything. To put together the facts is difficult and there arose a belief in this. And as we known transformations that change correctness, parametric stability etc. occur rarely. Just because of this they have not been noticed for such a long period of time.

Later dangerous wreckages occured in the 60th of the XXth century – with objects "analytically constructed" regulators were set. These facts made it necessary to investigate more attentively phenomena that occurred during such equivalent transformations as a change of unmeasured variation in control systems by a combination of measured variables. Difficulties that have occured during these transformations can be explained by the following example of a control system of the third order:

$$\left.\begin{array}{c} \dot{x}_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + b_1u \\ \dot{x}_2 = a_{12}x_1 + a_{22}x_2 + a_{23}x_3 + b_2u \\ \dot{x}_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + b_3u \\ u = -k_1x_1 - k_2x_2 - k_3x_3 \end{array}\right\} \tag{78}$$

In system (78) the first three equations are equations for a control object, the favith equation is an equation for a regulator that was chosen due to the method of "analytical construction" as a linear function of all three variables $x_1; x_2; x_3$ that describe a state of an object of control. Certain values of coefficients $k_1; k_2; k_3$ have been given according to a methodics proposed by A.M.Lietov in his work [49].

If not all variables $x_1; x_2; x_3$ could be available for measuring and were applied in a regulator they were (as have already said) excluded from equations (78) by means of equivalent transformations and we received such regulators that used only available variables.

Unexpectebly it turns out that different equivalent transformations, different ways of excluding variables lead to such control systems that differ by stability reserves during parameters variations. In it there is no contradictions as different equivalent transformations lead to equivalent control systems, to systems with the same characteristic

polynomial, with the same solutiones $x_1(t); x_2(t); x_3(t); u(t)$. This is indisupable. But these equivalent systems that have been obtained by different equivalent transformations can differ between themselves in properties of solutions and in particular – in parametric stability in stability reserves during parameters variations. This circumstance greatly made it difficult the investigation of properties of equivalent transformations and the acknownledgement of these investigations resilts.

Often the following situation occured: when one investigator applied some method of excluding inmeasurable variables and has come to such a control system that does not possess a parametric stability. The second investigator while examining the same technical object has applied another exclusion method of unmeasured variables and has obtained such a control system that possessed a wonderful parametric stability and he said that the first one has made a mistake. And the first investigator said that the second investigator had made a mistake but not he himself. A confusion has arisen and to make everything clear was rather difficult. Besides it turns out that there exist such control systems of a type (78) in which most different methods of excluding variables have not influenced a final result and always have led to closed systems. They (at the same time) do not possess a parametric stability as everything depends on a value of coefficients a_{ij} and b_i in equations (78). In book [5], pp. 88-111 are given (in more details) these delicate questions. They for a long period of time hampered in obtaining clear and not difficult in understanding results of investigations.

One more essential difficulty has arisen during the analysis of the influence of parameters changes of different mathematical models on their stability. In §8 during the examining of parameters influence in electrodrives on its parametric stability has been evidently established that this formal mathematical investigation of equivalent transformation (excluding variables x_3 and x_4) and its influence on parametric stability leads to a completely false conclusion. In order to obtain a correct result it is necessary to take into account (truely) mutucal relations in possible coefficients variations of mathematical models in different parts and elements of a considered technical object. In a particular case (see §8) it must be taken into account a possible independence of variations of such elements as an executive electrodrive and its regulator.

Thus although the investigation of equivalent transformations properties is a theme of an applied mathematics. Its solution first of all requires a mathematical the knowledge. Besides it also requires the knowledge of technique. Usually mathematicians do not possess knownledge in technique. Probably this is due to such a late discovery of new properties of equivalent transformations.

One more difficult problem is a confusion with the so called "singular perturbing" equations. Let us return (once more) to equations (72)–(73). Recall that equation (72) describes an electrodrive and an equation (73) is its regulator. As earlier let us consider a quite possible case when parameters in a regulator remained equal to its calculated values (i.e. – $k_1 = 1; m = 1; k_3 = 2; k_4 = 1$) but in electrodrive equation (72) a coefficient k_0 remained unchanged ($k_0 = 2$) but a parameter m (a mechanical time constant) has changed by a small value ε and it has become equal to $1 + \varepsilon$. By excluding variable x_2 from system (72)–(73) we shall obtain the following differential equation of the fourth order:

$$[-\varepsilon D^4 + (1-3\varepsilon)D^3 + (5-3\varepsilon)D^2 + (7+\varepsilon)D + 3]x_1 = 0 \tag{79}$$

i.e. we shall obtain an equation with a small coefficient in the higher derivative. In the course of intermediate transformations and for many other "peculiar" objects we often come. But equations with small coefficients (small parameters) if derivatives are higher have been examined for a long period of time and they have obtained a special name – "singular-perturbing differential equations".

This circumstance has led a lot of investigators to a false way. Solutions of singular-perturbing equations that correspond to $\varepsilon = 0$ and $\varepsilon \neq 0$ (however small be ε) often differ and very much. This is clear since the transfer from a zero value of a coefficient in a higher derivative to value $\varepsilon \neq 0$ will change the order of an equation of different orders and solutions of equations of different orders most often differ from one another sufficiently. This fact can be understood and does not arise any doubts.

The most simple example: an equation:

$$\varepsilon\dot{x} - x = 0 \tag{80}$$

is an equation of the first order and it has a solution:

$$x = C_0 e^{\frac{t}{\varepsilon}} \tag{81}$$

If ε is small this solution is a quickly increasing function. At the same time if $\varepsilon = 0$ equation (80) transfers into a differential equation of a zero order. It does not contain derivations and it can be named an equation of a zero order ant it transfers into equation

$$-x = 0 \tag{82}$$

and naturally solution $x = 0$ in equation (82) has nothing in common with a solution (81) of equation (80) if parameter ε has the most small values. The same can be also said about a solution of a solution of an equation of the second order:

$$\varepsilon\ddot{x} + \dot{x} + x = 0 \tag{83}$$

and when $\varepsilon \neq 0$ they have little in common with solutions of an equation of the first order:

$$\dot{x} + x = 0 \tag{84}$$

however small is a value ε.

At the same time solutions of an equation

$$(1+\varepsilon)\dot{x} + x = 0$$

that is not "singular-perturbing" are of the form:

$$x = C_0 e^{\frac{t}{1+\varepsilon}}$$

and when ε are small they differ very little (for any t) they differ from solutions that correspond to $\varepsilon = 0$.

In order to evade misunderstanding it is necessary to clearly differentiate small absolute variations and small relative variations, i.e. variations that are small in relation to an initial value of a coefficient or a parameter. So, for example, variation $\varepsilon = -9 \cdot 10^{-8}$ is small in comparison with a unity but if an initial value of a unity but if an initial value of a coefficient is $a_0 = 10^{-7}$ then variation $\varepsilon = -9 \cdot 10^{-8}$ will change this coefficient by ten times in relation to a_0. This will be not a small but a large change.

If an initial value of a coefficient is equal to zero then an addition to it a small (by an absolute value) variation in ε would change a coefficient (not in a literal sense) by an infinitely large number of times.

The absence of an accurate difference between relative and absolute variations leads to a confusion (which has clearely manifested itself during the first discussions of a publication [5]). Therefore because of this from the very beginning – in §1 of the second part it was strictly noted that in a later statement "variations of a zero"would not be considered, i.e. such objects in which some coefficient of an initial mathematical model was equal to zero but after variations it would take although small but not equal to zero value. Such objects exist but in this book they are not considered. Only relative variations are considered. Only relative variations are considered, only such objects in a mathematical model if an initial value of some coefficient or a parameter in which is equal to, for example, its value after variation is equal to $a_i(1+\varepsilon_i$ where ε_i – a number that is small in comparison with a unity. Such a limitation in an investigation object is connected with a fact when some zero coefficient of a mathematical model of an object will become nonzero (let it be even small) then properties and a behavior of an object (as already we have said) can change greatly essentially. In this there is nothing surprising. This fact has been known long ago and it has been investigated (in particular) on the basis of a theory of singularly perturbing equations as well. At the same time during a relative variation while we transfer from values of coefficients a_i to values $a_i(1+\varepsilon_i)$ we can expect that during small ε_i properties of an object will change little. Most often it is so. Strictly speaking all modern technique is based on this since a transfer from object values and of parameters a_i to value $a_i(1+\varepsilon_i)$ in the course of exploitation of an object is almost always inevitable. If during such transfers properties of many objects change substantially all modern technique would be pulled down. But dangerous objects that during small relative variations of parameters (namely, relative but not absolute!) during a transfer from a value a_i to a value a_i (if ε_i) change substantially their properties but they exist although we can meet them not often. In fact their behavior that us similar to the behavior of "singularly perturbing" systems although "singularly perturbing" they are not. And the fact that similar to "singularly perturbing" equations appear in the course of intermediate transformations of mathematical models of "special" objects must not mislead us.

Besides in such objects their properties that has been described cannot be revealed by conventional computation methods. Such objects that first of all can substantionally change their behavior during small relative variations and secondly, this dangerous property cannot be revealed by conventional computation methods that do not take into account new properties of equivalent transformations (that have been recently discovered) – were called "special" objects.

These objects do not occur very often. Therefore they were discovered very late. But they must be very seriously examined and investigated since each unexpected meeting with such an object can become – and has already become – a cause of wreckages and catastrophes. In [5] we can find a more detailed investigation of the problem.

A detailed investigation and a clear understanding of essential differences between relative and absolute variations of coefficients and parameters is also important because even in serious and argumented books and test–books devoted to the exactness and reliability of computations relative and abstract variations are sometimes not distinguished (and thus we have a confusion) they must be by all means distinguished. Mathematical models in which "variations of a zero" are possible and models where "variations of a zero" are not possible describe different classes of objects. These classes of objects possess essentially different properties. Therefore investigation methods do not coincide and cannot coincide. The mixing of these quite different objects and their mathematical models has delayed a discovery of "special" objects greatly and it has delayed the discovery of new properties of equivalent transformations for a long period of time.

§31. Inexactnesses in stability computations in relation to a part of variables.

By completing the analysis of errors and inexactnesses in computing stability let us consider a stability problem in relation to a part of variables. Many investigators devoted their works to this problem [59], [60] etc. The importance of this problem is due to the fact that not always stability is required in relation to all variables. And in these cases a possibility appears to project a more simple system that is stable not in relation to all variables but only in relation to a part of them. As an example a movement of a rocket that is symmetrical in relation to a longtudinal axis can serve. A vast movement of a rocket as well as any other solid body is described in relation to six variables – to three coordinates of a mass centre and to three angles of a mutual perpendicular axises that of a rocket in relation to three mutually perpendicular axises that pass through a mass centre. One of these axises can be combined with a longitudinal axis of symmetry. Then stability of one of variables – an angle of a turn in relation to this axis is not important while the rocket has hit a target.

Let us show a method of checking stability in relation to a part of variables on an example of equations system:

$$\left.\begin{array}{r} \dot{x}_1 = -x_1 + x_2 - 2x_3 \\ \dot{x}_2 = 4x_1 + x_2 \\ \dot{x}_3 = 2x_1 + x_2 - x_3 \end{array}\right\} \tag{85}$$

which has been earlier examined in a work [59].

A characteristic polynomial of system (85)

$$\begin{vmatrix} \lambda+1 & -2 & 2 \\ -4 & \lambda-1 & 0 \\ -2 & -1 & \lambda+1 \end{vmatrix} = \lambda^3 + \lambda^2 - \lambda - 1 = (\lambda+1)^2(\lambda-1) \tag{86}$$

It has both positive an negative roots. Therefore all solutions of system (85) can not be stable. In order to judge about, for example, variable x_i we can apply a known method of "μ – transformations". This method is: a part of old variable in a system is changed by new ones (by using equivalent transformations) which according to a tradition are denoted by "μ". Hence – the name of the method. We try to make this change in such a way that a new system consisting of a part of old variables x_i and new variables μ_i be stable in relation to all variables. This will mean that variables x_i have been stable even in an initial system.

For system (85) in a well known monograph [59] a new variable $\mu = x_2 - 2x_3$ was proposed to introduce. And then since $\dot{\mu} = \dot{x}_2 - 2\dot{x}_3$ then from the second and a third equation from system (85) it follows that $\dot{\mu} = -\mu$. Finilly for variables x_i and μ we obtain equations:

$$\left.\begin{array}{r} \dot{x}_1 + x_1 = \mu \\ \dot{\mu} + \mu = 0 \end{array}\right\} \tag{87}$$

system (87) has the following characteristic polynomial:

$$\begin{vmatrix} \lambda+1 & 1 \\ 0 & \lambda+1 \end{vmatrix} = (\lambda+1)^2 \tag{88}$$

with roots $\lambda_1 = \lambda_2 = -1$. Therefore solutions $x_1(t)$ and $\mu(t)$ are stable for any initial conditions. And since transformations that have turn system (85) into system (87) are equivalent (in a classical sense) in relation to variable x_1 this means that variable x_1 is stable in an initial system (85) as well.

In this case this conclusion can be directly tested by the integration of system (85) with initial conditions $x_1(0) = x_{10}; x_2(0) = x_{20}; x_3(0) = x_{30}$. We shall obtain:

$$x_1(t) = x_{10}e^{-t} + (x_{20} - 2x_{30})te^{-t} \tag{89}$$

$$x_2(t) = 2(x_{10} + x_{20} - 2x_{30})e^t + (4x_{30} - 2x_{20})te^{-t} + (2x_{30} - x_{20} - 2x_{10})e^{-t} \tag{90}$$

$$x_3(t) = (x_{10} + x_{20} - x_{30}e^t + (2x_{30} - x_{20})te^{-t} + (2x_{30} - x_{10} - x_{20})e^{-t} \tag{91}$$

$$\mu(t) = x_2 - 2x_3 = (x_{20} - 2x_{30})e^{-t} \tag{92}$$

Formulas (89) and (92) confirm that solutions x_1 and $\mu(t)$ are stable but $x_2(t)$ and $x_3(t)$ are not stable.

If a direct integration of some system is difficult then a conclusion about stability of some variable can be carried out on the basis of methodics of "μ – transformations".

But this conclusion and its particular case – a conclusion about stability of variable x_1 in system (85) on the basis of an investigation of system (87) stability will be incomplete. In fact variable x_1 stability in system (85) can be lost if there are infinitely small – and thus – inevitable in practice – variations of parameters in system (85). Variable x_1 in system (85) does not possess parametric stability although in system (87) solution x_1 that is equivalent in relation to variable x_1 in system (87) possesses parametric stability. Note that on monograph [59] – from which an example with system (85) has been taken – the loss of stability in x_1 during infinitely small variations of parameters the cause of this phenomenon has not been explained correctly. The author of a monograph [59] explained it in such away: "a property of asymptotic stability in relation to a part of variables possesses an increased sensibility in relation to variations of coefficients". In fact "an increased sensibility" here has no senses. Simply speaking μ – transformation as well as any other transformation that is equivalent in a classical sense (but not in a widened sense) to transformation can change a property of parametric stability of solutions both in all variables and in a part of them.

Since the system is stable but can loose it even when infinitely small (and thus inevitable in practice) variations of parameters occur the system is not better than an unstable system. In practice only such systems that are equivalent not only in classical but in a widened sense are of any practical sense. More details about equivalent transformations in a widened sense can be found in the next section. If there is no equivalence in a widened sense then a judgement about stability of some system of

equations in a part of variables on the base of "μ – transformations" can lead to mistakes.

The investigation of stability problem "in relation to a part of variables" once more stresses differences of approaches and results in "Mathematics – 1" and "Mathematics – 2".

In the frame of "Mathematics – 1" that supposes that exact assumed coefficients are unchanged a conclusion about a stability of a solution x_1 in system (85) has any sense and it is of importance.

But in "Mathematics – 2" that reflects practical requirements to results of mathematical investigations more exactly at once discovers that a conclusion about stability of x_1 has no practical sense as stability that looses during infinitely small – and thus inevitable in practice – variations of parameters is not (by all means) better than instability.

But this means that the majority of stability investigations in relation to a part of variables has no sense in practice. But so much labour has been spent on these investigations! (see publications [59] and [60]). It is only a small part of a lot of works devoted to stability in relation to a part of variables!

An example of system (85) has shown – besides other things – that a change of stability during equivalent transformations can take place not only in degenerated systems. And this fact makes us especially cautious to the question of a security of technical computations and computational algorithms to which the next section will be devoted.

Among many systems that are stable by a part of variables it is necessary to depict such systems in which stability by a part of variables is preserved only during infinitely small coefficients variations. Only such systems (if they exist) have a practical sense. But in systems in which stability by a part of variables can disappear during infinitely small coefficients variations (as it occurs in system (85)) have no practical sense. They are not better than systems that are unstable by all variables.

§32. The provision of security of computing algorithms.

The discovery of equivalent transformations that are able of change many important properties of transformed systems – such as parametric stability, correctness etc. make us to be especially attention to providing security of different computing algorithms.

Without applying equivalent transformations we cannot get without them. Therefore it is necessary to watch that applied equivalent transformations would not lead to calculation mistakes.

In a work [5] for the first time has been made an attempt to pick out two classes of equivalent transformation:

1. Transformations that are equivalent in a classical sense as they do not change solutions of transformed systems.

2. Transformations that are equivalent in a widened sense – which first of all are equivalent in a classical sense and secondly, do not change correctness of a solved problem.

If we have succeeded in picking out the second class of equivalent transformations and applying only them then we could have made a great step to the problem of security of computations.

But it turned out that depending on solved problems and considered mathematical models even the most simple and "innocent" transformations can turn out to be equivalent in a classical but not in a widened sense.

Here is a simple example: a system of equations:

$$(D^3 + 4D^2 + 5D + 2)x_1 = (D^2 + 2D + 1)x_2 \tag{93}$$

$$(D^2 + 4D + 5)X_1 = (D + 1)x_1 \tag{94}$$

describes (as it has been shown earlier) transient processes in a system of control rotation frequency of an electrodrive. Here x_1 is a deviation of frequency of rotation from a rating one and x_2 – a deviation of a rotating moment from a nonrating one.

Now let us introduce new variables defining them by equalities:

$$x_3 = \dot{x}_1 + 2x_1 - x_2 \tag{95}$$

$$x_4 = \dot{x}_3 \tag{96}$$

In relation to new variables equation (93) will turn into a system of three equations of the first order:

$$\left.\begin{array}{c} \dot{x}_1 = -2x_1 + x_2 + x_3 \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = -x_3 - 2x_4 \end{array}\right\} \tag{97}$$

Now let us transform equation (94). Transformations will only be breaking into a sum: $2Dx_i + 2Dx_i$. And by transform to members from one part of equality into another with the change of a sign.

Equivalence of these most simple transformations does not lead to any doubts. After having done them we shall transform equations (94) into:

$$[(D^2 + 2D)x_1 - Dx_2] + [2D + 4 - 2x_2] + x_1 = -x_2 \tag{98}$$

Now while comparing (98) with equalities (95) and (96) we see that those that stand in the first square bracket corresponds to variable x_1 and the second square bracket corresponds to $2x_1$ and as a whole equation (98) can be written in the form:

$$x_2 = -x_1 - 2x_3 - x_4 \tag{99}$$

i.e. it transfer into a differential equation of a zero order, into a relation between variables that do not contain derivatives.

If we put a value x_1 from (99) into the first of equation (97) we obtain for variables x_1, x_3 and x_4 a normal Cauchy form:

$$\left.\begin{array}{c} \dot{x}_1 = -3x_1 - x_3 - x_4 \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = -x_3 - 2x_4 \end{array}\right\} \tag{100}$$

i.e. a system of three equations of the first order for three variables x_1, x_3 and x_4.

Thus equations of a system controlling an electrodrive can be written either in the form of a system of equations (93)–(94) or in the form of equivalent to it system of equations (97)–(98) or in the form of system (100) that is equivalent to any at these systems. All three systems of equations have the same characteristic polynomial. Really, for system (93)–(94) it will be a polynomial

$$\begin{vmatrix} \lambda^3 + 4\lambda^2 + 5\lambda + 2 & \lambda^2 + 2\lambda + 1 \\ \lambda^2 + 4\lambda + 2 & \lambda + 1 \end{vmatrix} = \lambda^3 + 5\lambda^2 + 7\lambda + 3 = (\lambda + 3) \cdot (\lambda + 1)^2$$

For system (97)–(99) it will be the same polynomial:

$$\begin{vmatrix} \lambda + 2 & -1 & -1 & - \\ 0 & 0 & \lambda & -1 \\ 0 & 0 & 1 & \lambda + 2 \\ 1 & 1 & 2 & 1 \end{vmatrix} = \lambda^3 + 5\lambda^2 + 7\lambda + 3 = (\lambda + 3)(\lambda + 1)^2$$

but for system (100) we again have the same characteristic polynomial:

$$\begin{vmatrix} \lambda + 3 & 1 & 1 \\ 0 & \lambda & -1 \\ 0 & 1 & \lambda + 2 \end{vmatrix} = (\lambda + 3)(\lambda + 1)^2 = \lambda^3 + 5\lambda^2 + 7\lambda + 3$$

We can also directly check that, for example solution $x_1(t)$ for system (93)–(94) and of system (97)–(99) and of system (100) all are of the similar form:

$$x_1 = C_1 e^{-3t} + (C_2 t + C_3) e^{-t},$$

where integration constants C_1, C_2, C_3 depend on initial conditions and are defined by them. But solutions of system (93)–(94) – as has already been shown – are incorrect but solutions of system (97)-(99) – correct.

Therefore we can say that very elementary, by all, means, equivalent (in a classical sense) and such that do not present any doubts –the most simple transformations (the division a member $4Dx_1$ into a sum $2Dx_1+2Dx_1$ the transfer of members from the left side of an equation to the right side with the change of a sign) can change the correctness of a solution and they can be attributed to transformations that are equivalent in a classical sense but not in a widened one.

Therefore we doubt that could find a simple criterium for the discovery of transformations that are equivalent in a widened sense. Great efforts that have been undertaken in 1994-2003 have not led to success.

In order to guarantee security of computer calculation it is necessary to check the correctness of solutions in a mathematical model that has been reduced to the most correspondent (to a physical sense of a solved problem. The correctness check be means of a mathematical model transfered (by equivalent transformations) to the most convenient for investigation form can give a false answer even if this form has been obtained from an initial one by quite equivalent (in a classical sense) transformations.

So, for example, the correctness of a problem solution about stability of a frequency of solution $x_1(t)$ in system of control of an electrodrive must be investigated by means of a mathematical model in the form of a system of equations (93)–(94). The investigation by means of a mathematical model in the form of a system of equations (100) does not give a correct answer although systems (93)–(94) and (100) are equivalent (in a classical sense) in relation to variable $x_1(t)$ and a solution $x(t)$ in systems (93)–(94) and (100) – are identical.

This circumstance is important and it must be stressed that up to now it was not reserved in text-books devoted to engineering computing and calculations on computers. It is also not applied while constructing computing programs and packages of programs. The correctness of solutions are checked (if they in general check them) according to the most convenient for investigation form of a mathematic model. And this method can lead to false conclusions.

A direct check of correctness of a solution of a certain system of equations can require (as we have already mentioed in chapter 1) a large volume of calculations. Therefore it is advisable to pick out such classes of problems for which correctness has already been tested and therefore correctness of each separate certain problem need not require testing.

This can be carried out in the form of "triads". A "triade" will be named a combination of the following three elements (as it was proposed earlier in [11])

1. Investigated mathematical
2. A posed problem during its investigation
3. The used method of solution.

We shall call "a diad" such a combination of a mathematical model and a problem

posed for its solution (for such when a solution method can not be taken into account)

The first triade of testing stability

Elements of a triade:
1. An investigated mathematical model: a system of linear differential equations of different order with constant variables.

2. A problem is posed: to check stability of this system

3. A method of solving: a characteristic polynomial of a system and its roots is computed. If all roots of real parts are negative then conventionally we make a conclusion about the system of a stability.

In this triad the conclusion will be reliable and trustworthy not always, not for all systems since there exist (as it has already been mentioned) special systems in which all roots have negative real parts and which besides this have became unstable during infinitely small, inevitable (in practice) deviations of coefficients from ealculated values.

Note that in a formal way from a purely mathematical point of view conclusion about stability will be true always: negative real parts all roots tells us that if coefficients in a system idially and exactly are equal to their computed values then solutions are stable. But a system can loose stability during infinitely small and thus – inevitable in practice variations of parameters is not at all better than unstable one and it is even more dangerous than it.

The second triad the check of stability by another method

Mathematical model and a posed problem: the same as in the first triade.

Method of solution: computation of a characteristic polynomial and its roots must be added by checks: 1. either a degree of a characteristic polynomial is lower than an order of a system (its a check of degeneration). 2. whether some coefficients of a characteristic polynomial a small difference between large numbers – small in such a way that during inevitable in practice values of variations of these large numbers a coefficient of a characteristic polynomial can change its sign. Only in such a way if both checks have given a negative answer then a computer calculation of a characteristic polynomial leads to a secure conclusion about the stability of a system. Conventional methods of calculating stability without the above additional checks do not secure trustworthiness of computer calculations.

The third triade: the application of Lyapunov functions

A mathematical model: system of nonlinear differential equations of different orders.

A posed problem: to define whether solutions of this system are stable? Solution method: by transforming a system to a normal form of Cauchy we try to construct Lyapunov function for this system. If Lyapunov function has been constructed then conventionally we make a conclusion about stability of a zero solution of a system and

the investigation of any solution can be reduced to the investigation of stability in a zero solution.

In this triad a conventional conclusion about stability of a zero solution (as it has been earlier shown) is secure for all systems and not always since there exist systems for which Lyapunov function has been constructed but which has become unstable during infinitely small, inevitable in practice variations of coefficients in equation systems or an object parameters whose mathematical model is investigated.

The fourth triade: the calculation of principal value.

A problem of computing principal values is one of important problems in computer mathematics. Its solution has been given in many text-book but remain inattended such systems for which a problem of computing principal values is incorrect.

So, during the calculation of principal values (as well as roots of a characteristic polynomial) for a system of differential equations are limited by a particular case when an investigated system is reduced to a system of n equations of the first order. In this case principal values can be determined from an equation:

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

where A is a matrix of coefficients in an investigated system, E – a unity matrix and a solution of this equation will be correct.

But for systems consisting of equations of different orders, for example, for system (93)–(94) principal values will be roots of a determinant:

$$\begin{vmatrix} \lambda^3 + 4\lambda^2 + 5\lambda + 2 & \lambda^2 + 2\lambda + 1 \\ \lambda^2 + 4\lambda + 5 & \lambda + 1 \end{vmatrix}$$

but a problem of computing its roots is incorrect.

The reduction of an investigated system to a system of n equations of the first order camouflages a possible incorrectness but it does not abolish it and unnoticed incorrectness leads to mistakes in computations.

One more example of incorrectness is a case that often meet with – when a part of equations contains a parameter λ and a part of them – does not. Here an investigated mathematical model is of the form:

$$\left.\begin{array}{c} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = \lambda x_2 \\ ..... \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n = 0 \end{array}\right\} \qquad (101)$$

Note that in system (101) enters $n - r$ equations with right sides containing parameter λ and it also contains r equations that do not contain parameter λ and their right sides are equal to zero.

While investigating differential equations such systems appear when investigated variables satisfy both linear differential equations and algebraic equations that do not contain derivatives.

In vector-matrix forms we can write a system of equalities (101) in the following form:

$$(A - \lambda\bar{E})X = 0 \tag{102}$$

where X – vector of variables $x_1, x_2, ..., x_n$, A – a square matrix of coefficients, $\bar{E}$ – a quasi unity matrix, i.e. a matrix in which all elements except those that lay on a main diagonal – zeroes but on the main diagonal there are $n - r$ unities and r zeroes. If $r = 0$ matrix $\bar{E}$ turns into a known unity matrix $\bar{E}$.

A posed problem: is to find principal values of a parameter λ, i.e. values at which system (101) has nonzero solutions x_1.

Method of solution: a successive exclusion of variables from system (101) up to when the last equation for variable x_n remains:

$$M(\lambda)x_n = 0 \tag{103}$$

where $M(\lambda)$ – polynomial among whose roots are principal values.

This triad leads (as shown in [51]) to incorrect conclusions about principal values in system (101).

If we change the third point of a triad, change a method of solution, i.e. by using equations that do not contain λ, if we express some variables by means of others and come as a result to a system that has less number of equations but in which already each of equations contains parameter λ and only then start successive exclusion of variables then in this case principal values will be (as seen in [5]) correct.

This example shows that it is necessary to making precise used algorithms during the transfer to computer calculations. During the area of "hand calculation" during the meeting with systems of the type (101) we started, surely, from equations that did not contain parameter λ. By applying them we decreased a number of variables, we came to a system of lesser number of equation in which entered parameter λ and only then we started to successively exclude variables. Such a method of calculation was convenient during a hand calculation and it did not lead to incorrectnesses. Therefore a problem of computing principal values conventionally was considered correct.

But during machine computation this method is not very convenient since it requires the application of two different programs. For a computer it is more convenient to exclude variables one by one from an initial system by one program. Not at once it was noted that this convenient for a computer algorithm led to incorrect principal values as it was in details considered in [5].

Note that if parameter λ enter into each of equations of a system then a problem of finding principal value λ is called a classical problem about principal values but if in some of equations parameter λ does not enter then this same problem is called a generalized

problem of finding principal values. A classical as well as generalized problem about principal values can often be found in applications. Even in the XIXth century they have been investigated. An incorrectness of a solution in a generalized problem while directly applying computers in order to successfully exclude variables from an initial system was noted quite recently in [5]. In order to guarantee reliability of computer calculations in this case it is sufficient to reduce a generalized problem about principal values to a classical one. If we investigate a system of n linear homogeneous algebraic equations from which $n - r$ equations contain a parameter λ and r equations do not contain it is necessary to transform this system into a system of $n - r$ equations each of which contains λ. Thus we have come to a classical problem about principal values which can be solved by using successive exclusion of variables. While applying such a method of computing incorrectness of solutions (as is shown in [5]) will not be.

The fifth triade: a numerical solution of a system of differential equations

Mathematical model: a system of ordinary differential equations of different orders and in a general case-nonlinear ones. A posed problem: to find a numerical solution of a system if initial conditions are posed.

Method of solving: after a system has been reduced to a normal Cauchy form ready programs of a numerical solution given in MATLAB, Mathcad or in other packages are applied.

A solution obtained by this method will not always be correct. For some (that apriori have not been known) systems a real behavior of an investigation object can essentially differ from calculated one during infinitely small (that are inevitable in practice) deviations of coefficients of a system from admitted ones during calculations. In order to guarantee security of calculations it is necessary to previously check whether solutions of an investigated system are incorrect or ill–conditioned.

The sixth triad: a solution with additional checks.

A mathematical model and a posed problem remain the same as for the fifth triade. A solution method is added by checks: 1. whether an investigated system is degenerated. 2. whether a determinant consists of higher members of a system equations a small value – is small in such a way that it can change a sign during possible variations of coefficients. Such a check essentially decreases a possibility of appearing unexpected correct solutions.

The seventh triad: differential equations, particular cases.

Mathematical model: systems consisting of n equations of the first order or one equation of the nth order. Right sides are continuous and they satisfy Liepshitz conditions.

A posed problem: to find a numerical solution if initial conditions are posed.

Solution method: applying routine programs for computers.

In this particular case solutions are correct, they depend on coefficients and parameters in a continuous way.

At the same time for a more general mathematical model – a system consisting of n equations of different orders (an example – system (91)–(92)) the application of routine computer programs can lead to white incorrect results in computer calculations.

The eighth triad: Integral equations

Mathematical model: integral Volterra equations of the first type, i.e. an equation:

$$\int_a^x K(x;s)y(s)ds = f(x) \tag{104}$$

As it is known to a mathematical model in the form of equation (104) many important problems of physics and technics can be attributed (sometimes).

Posed problem: to solve equation (104), i.e. to find a sought function $y(s)$ by means of a given right side $f(s)$ and a nucleus $K(x;s)$.

Solution method: a transfer from Volterra equation to a more simple Fredholm equation by equivalent transformations and then – to solve this equation.

Professor V. S. Sizikov has found (for the first time) that in this triad an incorrect solution is obtained and that a cause of incorrectness is that a transformation of Volterra equation into Fredholm equation is a transformation that is equivalent in a classical sense but not in a widened one. This transformation changes correctness of a solution.

This triad – in more details – and a method of finding a secure result of computer calculations during the solution of integral equations has been investigated by V. S. Sizikov in a text–book [42] on page 145–147.

The first diad, real roots of polynomials

For a series of particular cases correctness or incorrectness of a solution does does not depend on applied method. In this case it is sufficient to form the investigation result for a complex of two elements.

1. A mathematical model
2. A problem posed during its solution.

This complex (us was earlier posed in [11]) we shall call a diad.

The first example of a diad:
Mathematical model: a polynomial of n degree with real coefficients.

A posed problem: a calculation of real (and only real!) roots of a polynomial. The significance of this problem is defined in the following way. For many investigation objects physical sense has only real roots.

The presence of only complex roots means in this case that a posed problem has no solution.

It is not difficult to check that if real roots are not multiple the solution is correct. If roots are multiple the solution is not correct and most often it does not have any physical sense during infinitely small variation of coefficients. Any pair of multiple real roots can disappear and roots will become complex.

It is useful not to forget about thus simple and known eircamstance because often people forget to mention about in correctness of a solution in case of multiple roots in even detailed guides on the methodics of computations.

The second diad: complex roots in polynomials

Mathematical model: the same as in the first diad.

A posed problem: the computation of roots that are real or complex.

The solution of this problem is always correct. During small changes of polynomial coefficients the situation of its roots on a complex espace changes little. We must not carry out the check of a solution correctness for each separate polynomial.

The third diad: problems of maximum and minimum

Mathematical model: in this case it can be any.

A posed problem: to find the maximum or minimum.

Different problems on maximum and minimum and methods of their solution during the instruction of undergraduates is paid much attention to (and this is right) but often it is not indicated that traditional formulas of solutions often do not have any physical sense and can lead to disconfiture in practice.

Here is a simple example: it is necessary to find the minimum of a length of a fence in order to enclose a section of an arbitrary form having an areas. It is a known "Didonn problem" whose solution was known in ancient Grece. Really, from all enclosed curved lines of a set length the largest area limits a circumference. Therefore if it is required to enclose a section of an area S then it is best if this section has a form of circumference and the lowest length of a fence in this case is equal to $x_{min} = 2\sqrt{\pi}\sqrt{s}$. If $S = 100m^2$, then $z = 35,448ms$ (if a section is of the form of a square then $z_{min} = 40ms$).

But this does not at all mean that a minimal element is a fence whose length is $35,448ms$ – will give a real solution of a posed problem.

Initial data – in this case $S = 100ms^2$ – as in almost all practical problems are known with a limited exactness. In fact as initial data is a condition $S = (1 + \varepsilon)100ms^2$ where a value ε depends on the exactness of a measured area. It can be a small value. But almost never an exact equality $\varepsilon = 0$. A is satisfied if $\varepsilon > 0$ then lengths $L_{min} = 35,448ms$ are not emough for the enclosure of a fence. And a posed problem will not be solved.

In order to obtain a real solution it is necessary to change the formulation of a problem, for example, in the following way. A section of an arbitrary form is given with an area S

and it is measured with an exactness ε. It is necessary to find a minimal length of a fence that can guarantee to enclose a section with some value of ε. The solution is:

$$L_{gar} = 2\sqrt{\pi} \cdot \sqrt{S}\sqrt{1+\varepsilon} \tag{105}$$

If $\varepsilon = 0,05$ then $L_{zap} = 36,167 meters$ and $L_{rap} - L_{min} = 0,727$ meters.

An examined simple example is typical for a lot of problems on minimum (maximum). Since initial conditions of the problem are known almost always but with a limited exactness then a computed for fixed values of initial data a minimal element in real conditions turns out to be insufficient. In order to obtain a suitable in practice solution it is necessary to take into account possible deviations of real initial data of these computed values. Without taking into accent these possible deviations in problems on minimum and maximum a computed minimal or maximal element usually gives an incorrect solution while using any computation method. The account of possible variations of initial data reconstructs the correctness of a solution and its reliability.

Sometimes we hear: an incorrect problem can be solved by a regularizing method given, for example in [40]. It is not an exact opinion. In fact incorrect problems in principle cannot have reliable solutions if we take into account inevitable in practice small deviations of parameters from calculated values. Regularization is not a method of solution. It is a change (in fact) of incorrect problem into another one – a correct one but which is in some way near at an initial one.

Sometimes it is a direct reformulization of an initial problem when, for example, instead of searching a minimal length of a fence that encloses a section of an area S we search a length that guarantees a barrier of a section while taking into account estimates of maximally possible error ε in the estimate of its area (formula (105)).

Often regularization is reduced to the application of additional information about a solution. But the transfer to a problem with additional information it is a transfer from one problem to another one – from incorrect problem to a correct one.

As a conclusion let us consider an algorithm of a solution of a problem on synthesis of optimal law of control which in its initial phase is incorrect but can become correct if initial data change little.

Synthesis of an optimal control law

A control object is considered whose mathematical model is the following differential equation of an order n

$$ADx = u + \varphi(t) \tag{106}$$

in which is a polynomial from differential operator $D = \dfrac{d}{dt}$; S – a regulated variable, a scalar U – a controlled influence, a scalar $\varphi(t)$ – perturbing influence, a stationary arbitrary time function data about its spectral density in power are known, i.e. about its even function $S_\varphi(\omega)$, variable ω that has a physical sense of frequancy. Later these experimental data are approximated by a rating fraction:

$$S_\varphi = \frac{a_p\omega^{2p} + a_{p-1}\omega^{2p-2} + ... + a_1\omega^2 + a_0}{b_q\omega^{2q} + b_{q-1}\omega^{2q-2} + ... + b_1\omega^2 + b_0} \tag{107}$$

Degrees p and q and coefficients a_i and b_i of polynomials entering into the fraction (107) are picked up in such a way that if p and q are moderate the divergence between analytical approximation (107) and experimental data is not very large. A regulator which first of all must secure stability of a closed system is supposed to be linear. The following equation is its mathematical model:

$$W_1(D)x = W_2(D)u \tag{108}$$

where $W_1(D)$ и $W_2(D)$ are polynomials of a differentiation operator $D = \frac{d}{dt}$. It is advisable to find such degrees and coefficients of polynomial $W_1(D)$ and $W_2(D)$ in formula (49) that besides stability of a closed system a regulator secured the best of a possible quality of the control which is estimated by a value of an integral:

$$J = \lim_{T\to\infty} \frac{1}{T} \int_0^T (m^2x^2 + u^2)dt \tag{109}$$

If polynomials $W_1(D)$ and $W_2(D)$ have been computed then a problem of a technical realization of a regulator by its mathematical model in the form of equation (106) will describe well many important objects in industry and transport.

In order to compute polynomials $W_1(D)$ and $W_2(D)$ a rather simple algorithm has been developed that is in details described in publications [35] and [61]. The main steps of an algorithm are:

1. In analytical approximation (107) a change is carried out: $j\omega = s$ after which a spectral density is factorized:

$$S_\varphi(s) = S_1(s) \cdot S_1(-s) \tag{110}$$

i.e. it is decomposed into a product of two symmetrical multipliers: $S_i(s)$ and $S_i(-s)$ one of which is a function of s and the order – of $-s$. Since a numerator and a decomposer of a fraction (107) are even functions they have $2p$ and $2q$ (correspondindingly) symmetric roots λ_1 and $-\lambda_1$; λ_2 and $-\lambda_2$ etc.

Into a multiplier $S_1(s)$ all roots with negative real parts will enter and into multiplier $S_1(-s)$ will enter all roots with positive real parts. Thus $S_1(s)$ can be computed according to formula:

$$S_1(s) = \frac{\sqrt{a_p} \cdot (s - \lambda_{p_1}) \cdot (s - \lambda_{p_2})...(s - \lambda_{p_p})}{\sqrt{b_q} \cdot (s - \lambda_{q_1}) \cdot (s - \lambda_{q_2})...(s - \lambda_{q_q})} \tag{111}$$

Here numerator and denominator of fraction (111) both turn out to be Hurvitz polynomials.

2. On the second step of an algorithm in equation (106) differential operator D is changed by a variable s and a factorization of a polynomial $A(s)A(-s) + m^2$ is carried out:

$$A(s)A(-s) + m^2 = G(s)G(-s) \tag{112}$$

and here $G(s)$ will be Hurwitz polynomial and $G(-s)$ – not Hurwitz one.

3. On the third step the following decomposition (separation) occurs:

$$\frac{A(-s)}{G(-s)} \cdot S_1(s) = M_0 + M_+ + M_- \tag{113}$$

where M_0 – total polynomial, M_+ – a fraction in which a denominator degree is lager than numerator degree with poles in the left half space of a complex variable s and M_+ – a fraction in which a denominators degree is larger than a numerators degree with poles in the right half space.

4. At the fourth step the following function is formed:

$$\frac{\Phi_1(s)}{\Phi_2(s)} = \frac{M_0 + M_+}{G(s)S_1(s)} \tag{114}$$

with the help of which already directly polynomials $W_1(D)$ and $W_2(D)$ in optimal regulator (108) are posed:

$$\frac{W_1(D)}{W_2(D)} = A(D) - \frac{\Phi_2(D)}{\Phi_1(D)} \tag{115}$$

A similar but a more complex algorithm has been developed for control objects of the form:

$$A(D)x = B(D)u + \varphi(t) \tag{116}$$

where $B(D) = b_m D^m + b_{m-1} D^{m-1} + ... + b_i D + b_0$.

Developed algorithms have provided stability of a closed system but it turned quickly out that this stability was often broken even during infinitely small variations of parameters which led to a series of wreckages at the first steps of applying optimal systems in technique in the 60ths of the XXth century the confidence to optimization methods has been then undermined and their application has been stopped. In order to successfully apply optimal regulators it is necessary to solve the following two problems:

1. To find a criterium that allows to depict – for what control objects and perturbing interactions the loss of stability occurs during parameters variations (in fact – as it turned out later – it was necessary to develop a method of revealing "special" objects in this filed of computations).

2. It is necessary to find what changes it is necessary to introduce into an algorithm of optimal regulators synthesis in order that during parameters variations the loss of stability did not take place.

Both problem have been successfully solved and their solution was published in [61] and [35]. It turned out that the main role played an inequality

$$p \geq m + q - 1 \tag{117}$$

where p and q are taken from formula (107) for am analytical approximation of spectral density of power in perturbing interaction $\varphi(t)$ and m – a degree of a polynomial $B(D)$ in a mathematical model of a control object (116). If inequality (117) is satisfied then a control system is parametrically stable. If it is not fulfilled then a control system can loose stability during infinitely small variations of parameters in control object or in a regulator. Inequality (117) has gradually obtained the name of "inequality of Yu.P.Petrov" or "a criterium of Yu.P.Petrov". It is widely applied for the check of parametric stability of optimal control systems.

Let us give a simple example that has been earlier considered in [35].

A mathematical model of control object is of the form:

$$4Dx = (D + 1)u + \varphi(t) \tag{118}$$

coefficient m^2 in a quality criterium (109) is equal to $m^2 = 9$, a spectral density of power in a perturbing interaction is approximated by a formula:

$$S_\varphi(\omega) = \frac{2}{\pi} \cdot \frac{1}{1 + \omega^2} \tag{119}$$

After computations necessary for the synthesis of an optimal regulator have been carried out we shall seel that its mathematical model is of the form:

$$12(D + 4)x = (3D - 5)u \tag{120}$$

By enclosing a control object (118) by a regulator (120) we shall find an equation for an enclosed system:

$$4(20D + 12)x = (3D - 5)\varphi(t) \tag{121}$$

that confirms that an enclosed system is stable. Regulator (120) secures the quality criterium (109) a value $J_{min} = 0,4336$ is the smallest from all possible ones. But Yu.P.Petrov's criterium (117) for a control object (118) and a spectral density of power in a perturbing interaction (119) is not satisfied. In this case a degree of a polynomial $B(D)$ is equal to a unity. Therefore $m - 1$ from formula (119) it follows that $p = 0; q = 1$ and an inequality (117) is not satisfied. This means that stability can disappear during infinitely small variations of parameters.

Really, if only one of coefficients in control object changes and its mathematical model will become as:

$$4D(1 + \varepsilon)x = (D + 1)u + \varphi(t) \tag{122}$$

then by enclosing an object (122) by a regulator (120) we shall see that a characteristic polynomial of an enclosed system becomes:

$$-3\varepsilon\lambda^2 + (20 + 5\varepsilon)\lambda + 12 \tag{123}$$

Already during infinitely small $\varepsilon > 0$ it stops to be Hurvitz one and stability is lost.

In order to guarantee parametric stability in monograph [35] it has been proposed to change an analytic approximation of a density spectrum of a power in perturbing interaction that is applied during the computation of an optimal regulator. It is necessary to change it in such a way that Yu.P.Petrov inequality (117) turns out to be satisfied. In an examined example in order to satisfy – an inequality (117) it is sufficient to transfer from $p = 0$ to $p = -1$, i.e. to change an approximation (119) into:

$$S_{\varphi}(\omega) = \frac{2}{\pi} \cdot \frac{1 + k\omega^2}{1 + \omega^2} \tag{124}$$

which (if k is moderate) differ in a small way from approximated system (117).

By synthesizing an optimal regulator now we obtain the following regulator for already a control object (122) and a spectral density (124):

$$12[(1 + 3k)D + 4]x = [(3 - 11k)D - (5 + 3k)]u \tag{125}$$

(compare it with a regulator (120)!) and a characteristic polynomial of a limited system turns out into a polynomial:

$$(20k - 3\varepsilon + 11\varepsilon k)\lambda^2 + (20 + 12k + 5\varepsilon + 3\varepsilon k)\lambda + 12 \tag{126}$$

Compare it with a polynomial (123)! While analysing a polynomial (126) we see that it remains Hurwitz not only during infinitely small variations but even in finite values of ε. They are as bigger as k is larger. So even when $k = 0,01$ stability is preserved during all $|\varepsilon| \leq 0,69$ but if $k = 0,1$ when $|\varepsilon \leq 1,05$.

At the same time when we put a additional requirement to a system – its parametric stability – we must pay some kind of "a secrifice" in a value of a quantity criterium (107). So if $k = 0$ we have (as it was already said) $j = 0,4336$ but if $k = 0,1$ we shall have $j = 0,4374$ or by $0,88\%$ more. Surely such a small increase of quantity criterium will not be felt in practice.

Thus a problem of securing reliability of computations during the synthesis of optimal control systems has obtained a full and thorough solution:

1. A simple and easily checked criterium for a possible problem of synthesis incorrectness – Yu.P.Petrov criterium – in the form of inequality (117).

2. A method of basing and approach to incorrect synthesis problem is proposed – a change of analytical approximation of experimental data on a spectral density of power in a perturbing interaction – i.e. only such a change at which inequality (117) has started to be fulfilled. As this example has shown during such an approach an initial incorrect problem that has no practical sense is changed by sequence of correct problems which – in a limit – if $k \to 0$ stop to be correct and coincides with an initial incorrect problem.

The securing of reliability of computation and projection of optimal control systems allowed us to successfully solve a series of practical optimization problems, of improving a quality of functioning different technical objects. About this publications [35], [61], [62] can be found.

Let us at once note that not for all computational algorithms whose unreliability has been foud during the investigation of new properties of equivalent transformations recently found in St.Petersburg State University we had scceeded in thoroughly solving these both problems:

1. The revealing of a possible incorrect solving problems.

2. The reliability of solving results of computation while taking into account possible variations of coefficients and parameters.

§33.Additional examples

In this section additional examples that will allow us to analyse in the methods of reliable solutions that satisfy standards "Mathematics–2" will be considered. Here variations of object parameters and coefficients of its mathematical model, their inevitable deviations from computed values will b considered.

Example №1. A mathematical model of a peculiar object

It is necessary to find a solution of a family (a family of solutions) of the following equations system:

$$(D^2 + 2D + 1)x_1 + (D + 1)x_2 = 0 \tag{127}$$

$$Dx_1 + x_2 = 0 \tag{128}$$

For system (127)–(128) a characteristic polynomial can be computed:

$$\begin{vmatrix} \lambda^2 + 2\lambda + 1 & \lambda + 1 \\ \lambda & 1 \end{vmatrix} = \lambda + 1 \tag{129}$$

that has the only one root: $\lambda_1 = -1$.

From (129) it follows that a general solution is of the form:

$$x_1 = C_0 e^{-t} \tag{130}$$

We can obtain the following equation if we directly introduce into (127) a value $x_2 = -Dx_1$ from equation (128):

$$(D + 1)x_1 = 0 \tag{131}$$

whose general solution will again be a family of functions (130) that depends on one integration constant C_0. Both solution methods lead to one and the same result.

But solution (130) will not be correct. Really if, for example, coefficient in Dx_3 in equation (127) instead of a unity will adopt a value $(1+\varepsilon)$ then a characteristic polynomial will be equal to a determinant:

$$\begin{vmatrix} \lambda^2 + 2\lambda + 1 & (1 + \varepsilon)\lambda + 1 \\ \lambda & 1 \end{vmatrix} = -\varepsilon\lambda^2 + (1 - \varepsilon)\lambda + 1 \tag{132}$$

and it will have two roots λ_1 and λ_2.

If $\varepsilon > 0$ are small then one of roots turns out to be a large positive and the second one – near to minus unity. So if $\varepsilon = 0,01$ then with the exactness of up to the fourth sign after a comma $\lambda_1 = +100, \lambda_2 = -1$ and a general solution will become:

$$x_1 = C_1 e^{-t} + C_2 e^{100t} \tag{133}$$

In general with the exactness up to members of an order ε^3 a general solution will be of the form:

$$x_1 = C_1 e^{-t} + C_2 e^{\frac{1}{\varepsilon}t} \tag{134}$$

In this case everything is simple: system (127)–(128) is generated. Its characteristic polynomial in a general case must have a second order but in certain values of system coefficients with the second order λ is reciprocally reduced and a characteristic polynomial becomes a polynomial of the first order. It is clear that such a lowering of a degree occurs only with coefficients that are exactly equal to calculated ones. Therefore inspite of the simplicity of a system its solution (130) will be incorrect and it will not have a practical sense. Already during inevitable in practice infinitely small deviations of system coefficients from computed ones the solution can change substantially and it can transfer, for examples, into solution (134) which even with the most small ε has nothing in common with solution (130).

Mathematical model (127)–(128) describes a "special" object. In order to find a right approach to finding and investigating a solution it is necessary to check a degeneration of the system. System (127)–(128) is degenerated and its solution has no practical sense. Conventional methods of solving a solution that do not suppose the checking of the degeneration or undegeneration of the system will not secure a right answer about object behavior that has been described by a system (127)–(128).

In such a simple system as (127)–(128) certainly is everything clear and an investigator surely will not make a mistake. This example was given in order to show: even in the most simple systems a change of correctness in the solution is possible while using the most simple equivalent transformation. Really, solutions of system (127)–(128) are incorrect and after equivalent transformation – a change in equation (127) of a value x_2 by a value Dx_1 (that is equal to it) from equation (128) we obtain equation (131) whose solution is correct. An initial incorrectness is a transformed mathematical model disappears. Therefore a man that carries out computation cannot notice incorrectness, he cannot notice that he has met with a dangerous "special" object. Certainly, in relation to such a simple system as (127)–(128) nobody makes any mistakes. But in complex system that consist of many equations it is very easy to make such mistakes and such mistakes often occurred.

Example 2. The check of stability and its preservation during variation of coefficients

For the same system (127)–(128) it is necessary to check stability and preservation of stability during variations of coefficients in an examined system.

A conventional method of investigation of roots in a characteristic polynomial (129) gives the following answer. A system is stable and preserves it not only during small but also during large variations of coefficients in a characteristic polynomial. Surely this answer is not true but the authenthity of an investigation result can be easily reestablished by means of an additional check that has been recommended earlier. But whether an examined system is a degenerated one? This simple check reconstructs security and authenthicy of examined results. A true answer does not possess parametric stability if we take into account the check on a possible degeneration of system (127)–(128). The system can become unstable during infinitely small variations of parameters.

Example 3. The check of stability in a nondegenerated system.

Let it be posed that we check parametric stability of a system of equations:

$$(D^3 + 4D^2 + 5D + 2)x_1 = (D^2 + 2D + 1)x_2, \tag{135}$$

$$(D^2 + 4D + 5)x_1 = (0,96D + 1)x_2. \tag{136}$$

A computation of a characteristic polynomial in a system is a conventional method of checking stability and also – the check of a positivity in diagonal determinant of Hurwirz matrix.

For system (135)–(136) a characteristic polynomial is equal to:

$$\begin{vmatrix} \lambda^3 + 4\lambda^2 + 4D + 2 & -(\lambda^2 + 2\lambda + 1) \\ \lambda^2 + 4\lambda + 5 & -(0,96\lambda + 1) \end{vmatrix} = 0,04\lambda^4 + 1,16\lambda^3 + 5,2\lambda^2 + 7,08\lambda + 3 \tag{137}$$

For polynomial (137) Hurwitz matrix becomes:

$$\begin{pmatrix} 1,16 & 7,08 & 0 & 0 \\ 0,04 & 5,2 & 3 & 0 \\ 0 & 1,16 & 7,08 & 0 \\ 0 & 0,04 & 5,2 & 3 \end{pmatrix} \tag{138}$$

While computing diagonal determinants: $det_1 = 3; det_2 = \begin{vmatrix} 7,08 & 0 \\ 5,2 & 3 \end{vmatrix} = 21,24; det_3 = \begin{vmatrix} 5,2 & 3 & 0 \\ 1,16 & 7,08 & 0 \\ 0,04 & 5,2 & 3 \end{vmatrix} = 3 \cdot \begin{vmatrix} 5,2 & 3 \\ 1,16 & 7,08 \end{vmatrix} = 97,008; det_4 = \begin{vmatrix} 1,16 & 7,07 & 0 & 0 \\ 0,04 & 5,2 & 3 & 0 \\ 0 & 1,16 & 7,08 & 0 \\ 0 & 0,04 & 5,2 & 3 \end{vmatrix} = 106,5135.$

We see that they all are positive and therefore system (135)–(136) is stable.

In order to check a parametric stability conventionally we check a sign of diagonal determinants of Hurwitz matrix that has been formed while taking into account possible variations of polynomial (137) coefficients, i.e. signs of the following diagonal determinants in matrix are checked:

$$\begin{pmatrix} 1,16(1 \pm \varepsilon_2) & 7,07(1 \pm \varepsilon_4) & 0 & 0 \\ 0,04(1 \pm \varepsilon_1) & 5,2(1 \pm \varepsilon_3) & 3 & 0 \\ 0 & 1,16(1 \pm \varepsilon_2) & 7,08(1 \pm \varepsilon_4) & 0 \\ 0 & 0,04(1 \pm \varepsilon_1) & 5,2(1 \pm \varepsilon_3) & 3(1 \pm \varepsilon_5) \end{pmatrix} \tag{139}$$

In matrix (139) an index of variation ε corresponds to an ordinal number of a coefficient in polynomial (137) and a value of variations $\varepsilon_1; \varepsilon_2; ...; \varepsilon_5$ are generally speaking, different and they require a separate investigation. Usually an approximated methodics is applied considering all modules in variations to be limited from above by one number $|\varepsilon_1| \leq m$.

Another approach is possible to check stability and parametric stability: an examined system is reduced to a normal form, to a system of n equations of the first order and then

routine programs for computing roots in a characteristic polynomial are used. They will coincide with principal values of coefficients matrix.

Equation (135) is similar with equation of electrodrive (72) that has been considered in §8 and it corresponds to values $m = 1; k_0 = 2$ in equality (73). In a normal form this equation will be of the form:

$$\left.\begin{array}{c} \dot{x}_1 = -2x_1 + x_2 + x_3 \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = -x_3 - 2x_4 \end{array}\right\} \tag{140}$$

Equation (134) corresponds to regulator equation by its structure (formula (73) from §8) but with changed coefficients. While using in §8 values $m = 1; k_1 = 1; k_2 = 2; k_3 = 1$ a regulator equation (73) will become:

$$(D^2 + 4D + 5)x_1 = (D + 1)x_2. \tag{141}$$

But in a regulator equation (136) coefficient is a unity in D_2 let us change by $0,96$ which makes system (135)–(136) undegenerated. Therefore if we take into account the equalities that determine new variable x_3 and x_4 we have:

$$x_3 = \dot{x}_1 + 2x_1 - x_2 \tag{142}$$

$$x_4 = \dot{x}_3, \tag{143}$$

and equation (136) becomes:

$$[(D^2 + 2D)x_1 - Dx_2] + [(2D + 4)x_1 - 2x_2] + 0,04x_2 + x_2 + x_1 = 0. \tag{144}$$

If we take into account equalities (142) and (143) in the first square bracket from equality (144) we see a variable x_4. To the second square bracket corresponds $2x_3$. Therefore equation (144) can be written in the form:

$$0,04\dot{x}_2 = -x_1 - x_2 - 2x_3 - x_4 \tag{145}$$

As a whole a system of equations (135)–(136) is reduced to the following normal form:

$$\left.\begin{array}{c} \dot{x}_1 = -2x_1 + x_2 + x_3 \\ \dot{x}_2 = -25x_1 - 25x_2 - 50x_3 - 25x_4 \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = -x_3 - 2x_4 \end{array}\right\} \tag{146}$$

A normal form (146) that consists of four equations of the first order proves that an initial system (135)–(136) is not degenerated. While computing a characteristic polynomial in the system in a normal form (146) equal to a determinant:

$$\begin{vmatrix} \lambda + 2 & -1 & -1 & 0 \\ \lambda + 25 & 25 & 50 & 25 \\ 0 & 0 & \lambda & -1 \\ 0 & 0 & 1 & \lambda + 2 \end{vmatrix} = 0,04\lambda^4 + 1,16\lambda^3 + 5,2\lambda^2 + 7,08\lambda + 3 \tag{147}$$

we see that it coincides with polynomial (137). This fact once more confirms that a transformation of system (135)–(137) into system (146) was an equivalent (in a classical

sense) to transformation. Usually about stability and parametric stability of system (146) we judge in relation to coefficients of a characteristic polynomial. It is convenient to apply a condition for stability for the following polynomial of the fourth order:

$$a_4\lambda^4 + a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 \tag{148}$$

which is given in text–books on automatic control and it is formulated as a necessary and sufficient condition of a system that has a characteristic polynomial (148) and in this system all its coefficients are positive and the following inequality is satisfied:

$$\Delta_3 = a_1a_2a_3 - a_2^2a_4 - a_0a_3^2 > 0 \tag{149}$$

For system (146) and a characteristic polynomial (147) we shall have:

$$\Delta_3 = 7,08 \cdot 5,2 \cdot 1,16 - 5,2^2 \cdot 0,04 - 3 \cdot 1,16^2 = 41,88 > 0 \tag{150}$$

For polynomial (147) positive members in inequality (149) are much larger than negative ones. This means that system (146) is stable and it will preserve stability during infinitely small deviations of coefficients from computed values. In order to define more exactly – during what "small" variations the system will preserve stability let us suppose that variations of all coefficients of polynomial (147) do not exceed a number m by a module: $|\varepsilon_i| \leq m$ and its sign can be any.

The most dangerous for the possible loss of stability if a combination of coefficients variations in inequality (150) is as follows: $7,08; 5,2; 1,16$ in coefficients variations are negative but in coefficients $0,04$ and 3 – they are positive.

If we take into account these variations we obtain:

$$\Delta_3 = 7,08\cdot 5,2\cdot 1,16(1-m)^3 - 5,2^2(1-m)^2\cdot 0,04(1+m) - 3(1+m)\cdot 1,16^2(1-m)^2 \tag{151}$$

Even when $m = 0,5$ i.e. when coefficients $7,08; 5,2; 1,16$ decrease by two times we shall have $\Delta_3 > 0$ and stability will be preserved.

We shall come to the same result if we use a methodics based on the results by Kharitonov V. L [63]. They have not been used in this example because for systems of the fourth order the investigation of the influence of coefficients variations of a characteristical polynomial on stability it is more simple directly not to use results published in [63] which are useful if $n > \varphi$.

Thus conventional investigation methods that do not take into account possible change of correctness and conditions during equivalent transformation give the following answer: system (135)–(136) is stable and is parametrically stable. It preserves stability not only during small but also large (more than by 50%) deviations of coefficients of a characteristic polynomial from computed values.

This result is not true. While directly investigating system (135)–(136) without transforming it to a higher member of a characteristic polynomial in system (135)–(136) and while taking into account possible variations of coefficients (we shall denote by points members of lower degree) we have:

$$\begin{vmatrix} (1+\varepsilon)\lambda^3+... & -[(1-\varepsilon)\lambda^2+...] \\ (1-\varepsilon)\lambda^2+... & -[0,96(1+\varepsilon)+...] \end{vmatrix} = [(1-\varepsilon)^2 - 0,96(1+\varepsilon)^2]\cdot\lambda^4.. \qquad (152)$$

we see that already when $\varepsilon \geq 0,011$ the highest member of a characteristic polynomial became negative and a necessary Stodola condition is broken and stability is lost Thus system (135)–(136) possesses a very small reserve of stability: stability is lost during the change of some coefficients even by $1,1\%$. Equivalent transformation of the system to a normal form raises true reserves by 50 times.

If we have not carried out additional check – a check of a possible change of a highest member sign in a characteristic polynomial during variations of coefficients of initial equations – then results of the check in stability by means of coefficients of a characteristic polynomial or by a matrix of coefficients of a normal form can turn out to be quite unrealiable and not trustworthy. It would seem (according to computation) that everything is all right, reserves of stability are sufficient and therefore a projected object will work well and reliably for many years. Then small reserves will quickly exhaust and in the most unexpectable time moment wreckage will occur or even a catastrophe. Such wreckages and catastrophes that have been generated by an insufficient development of a theory of equivalent transformations have repeatedly occurred in the past and sorry to say they occur now although they can be easily prevented with the help of not at all difficult additional checks during our computations. These additional checks restore the security of computer calculations.

Example 4. The computation of solutions of a system of differential equations

A system of equations (135)–(136) is posed. It has already been considered in previous example. It is necessary to find its solution $x_1(t)$ and $x_2(t)$ that satisfy initial method is: to introduce new variables x_3 and x_4 that are determinated by means of formulas (142) and (143). Here system (135)–(136) is reduced to a normal form (146) in order to find a numerical solution of which it is sufficient to apply routine computer programs.

But conventional methods do not lead to a reliable result as it has been shown during the consideration of example 3, during the change of coefficients in a system only by $1,1\%$ in a characteristic polynomial can occur substantional change. Its higher member can transfer from a positive to a negative one and small in an absolute value. This means that in a characteristic polynomial a large positive root will appear λ_4 and in a general solution of system (135)–(136) we have:

$$x_1(t) = C_1e^{\lambda_1 t} + C_2e^{\lambda_2 t} + C_3e^{\lambda_3 t} + C_4e^{\lambda_4 t} \qquad (153)$$

a swiftly increasing fourth member will appear. Thus if $\varepsilon \geq 0,011$ a substantial change of a solution can occur. A solution that can substantionally change during the change of initial data by only $1,1\%$ is quite unreliable. The application of such a solution for some practical aims is very dangerous as it can lead to wreckages and catastrophes. And the most unpleasant is that we shall not see this danger after the transformation of a system (135)–(136) into a normal form – into a form of system (146). In system (146) solutions depend on coefficients continually (according to known theorem on a theory of differential equations). During the change of any coefficients of system (146) by $\pm 2\%$

solutions $x_1(t), x_2(t), x_3(t), x_4(t)$ will change very little (you can easily check this). At the same time an initial system (135)–(136) is one more example of a system in which there is no such continuous dependence.

All this can be easily noted for a case when in system (135)–(136) one coefficient will change. Then it will become:

$$\begin{array}{c}(D^3+4D^2+5D+2)x_1=(D^2+2D+1)x_2\\(D^2+4D+5)x_1=[(1+\varepsilon)\cdot 0,96D-1]x_2\end{array} \tag{154}$$

and it has a characteristic polynomial:

$$\begin{vmatrix}\lambda^3+4\lambda^2+5\lambda+2 & -(\lambda^2+2\lambda+1)\\ \lambda^2+4\lambda+5 & -[(1+\varepsilon)\cdot 0,96D+1]\end{vmatrix}] =$$
$$= [1-0,96(1+\varepsilon)]\lambda^4+[5-3,84(1+\varepsilon)]\lambda^3+[10-4,8(1+\varepsilon)]\lambda^2+[9-1,92(1+\varepsilon)]\lambda+3 \tag{155}$$

Now at once it is seen that as soon as value ε will exceed $\varepsilon = 0,04166$ the higher member of a polynomial (155) will change a sign and the fourth its root λ_4 will become from negative one to a large positive one. And all solution will swiftly and rapturally change.

For illustration in figure 14 a dependence of a value x_1 (if $t=1$) on ε for a case when integration constants in a solution (153) is equal to 1 ($C_1 = C_2 = C_3 = C_4 = 1$) was shown.

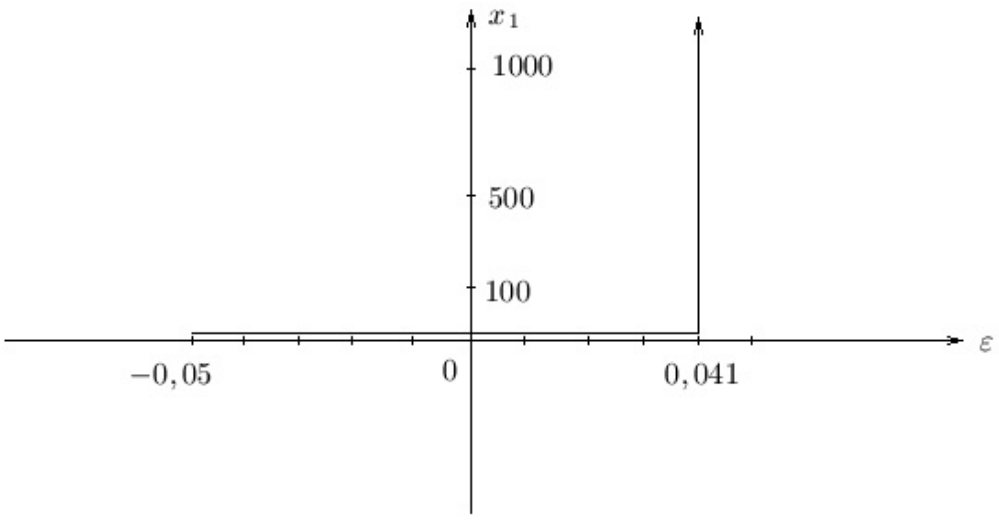

fig. 14

It is at once seen that if $\varepsilon = 0,04166$ a solution has a break and a value $x_1(t)$ when $\varepsilon > 0,04166$ cannot be placed on the diagram (if $\varepsilon = 0,4167$ will be $x_1(1) > 10^5$).

In order to secure reliability and authenticy of results of computation it is necessary (before transformation of equations system into a normal form) to check whether some coefficient of a characteristic polynomial is a small difference of large coefficients in an

initial system and whether it changes a sign during possible changes of these coefficients.

On an example of system (135)–(136) we once more can conclude that:

1. There exist systems of differential equations that have no continuous dependence of solutions on coefficients and parameters. At the same time after the system has been transformed into a normal form by equivalent transformations these same solutions will possess a continuous dependence on coefficients in a normal form (146) and this fact can lead to errors.

2. These exist such systems of differential equations for which a conventional computation method of solving by means of a transformation into a normal form the application of routine programs can lead to erroneous conclusions about the behavior of an object whose mathematical model is an investigated system. In particular a conclusion about a value of stability resourses in investigated system can be mistaken. These mistaken conclusions can become (and not once have become) a cause of wreckages and even catastrophes.

Example 5. Additional checks that restore the reliability and authencity of computer calculations.

A system of differential equations (135)–(136) once more. On its example we present additional checks that allow to increase the basis of our judgment about properties of a system solutions.

1. The first (and the most simple) check consists of: whether the system is degenerated. For this check it is sufficient to compute a characteristic polynomial (137) and see that its degree is equal to a degree of system (135)–(136).

The system is nondegenerated and therefore a conclusion about stability will be true and it will preserve power in the least during an infinitely small variations of coefficients. The reliability and trustworthiness of the conclusion during finite small variations the first check does not guarantee.

2. The second check. The presence of coefficient $0,04$ in the higher member in a polynomial (137) that is by more than a degree less than the least of other coefficients – it is necessary to check. Whether it has become a small difference of not small values and whether it will not change a sign during such variations of coefficients that are inevitable in the course of exploitation.

In order to carry out this check it is sufficient to write only these components that influence the value of a coefficient in a higher member during computation, i.e. it is sufficient to write a determinant

$$\begin{vmatrix} (1 \pm \varepsilon_1)\lambda^3 & -(1 \pm \varepsilon_2)\lambda^2 \\ (1 \pm \varepsilon_3)\lambda^2 & -0,96(1 \pm \varepsilon_4)\lambda \end{vmatrix} \tag{156}$$

Later it sufficient to find the most unfavourable combination of variations signs: $\varepsilon_1; \varepsilon_2; \varepsilon_3; \varepsilon_4$. Here we can use results obtained in the first paret of the book. For determinant (156) it is the most unfavourable when ε_1 and ε_2 are positive and ε_3 and ε_4 – negative

and then – to compute during which value $\varepsilon \geq |\varepsilon_i|$ and during the most unfavourable combination of signs ε_i a determinant (156) will become negative. For this it is sufficient to solve the following equation:

$$(1-\varepsilon)^2 = 0,96(1+\varepsilon)^2$$

After having solved it we shall find $\varepsilon = 0,011$. Hence follows (as it was already shown in example 4) that coefficients variations in the system that exceed $1,1\%$ from rating values can lead to substantional changes of solutions. Since such values of coefficients variations (in the course of exploitation) are quite possible then an initial judgement about stability of the system does not maintain additional changes. After an additional change it is necessary to make a conclusion about small resources of stability. In the course of exploitation stability of a system can be lost at any moment that has not been predicted apriori at a time moment.

This example has shown that an additional check that substantially increases the reliability of computation results is not complex and it can be reduced to computing a simple determinant (156).

Example 6. A possible change of a coefficient sign for lower members of a characteristic polynomial.

A system of equations is posed:

$$(D+1)x_1 + (D+3,98)x_2 = 0 \tag{157}$$

$$Dx_1 + (2D+2)x_2 = 0 \tag{158}$$

It is necessary to check stability and how to preserve stability during possible variations of coefficients in this system.

With the help of equivalent transformations let us denote x_2 by x_1 with the help of equation (158) and while introducing it into equation (157) we shall obtain an equivalent system (157)–(158) in relation to variable x_1 and the following equation of the second order:

$$(D^2 + 0,02D + 2)x_1 = 0 \tag{159}$$

(which means that an initial system (157)–(158) is not degenerated) with the following characteristic polynomial:

$$\lambda^2 + 0,02\lambda + 2 \tag{160}$$

For variable x_2 we shall obtain the same equation:

$$(D^2 + 0,02D + 2)x_2 = 0 \tag{161}$$

By investigating a characteristic polynomial (160) during variations of its coefficients, i.e. by investigating a polynomial

$$(1 \pm \varepsilon_1)\lambda^2 + 0,02(1 \pm \varepsilon_2)\lambda + (1 \pm \varepsilon_3) \cdot 2 \tag{162}$$

it is not difficult to conclude that polynomial (162) remains Hurwitz during not only infinitely small variations of its coefficients but even during large variations – up to $|\varepsilon_1| < 1$.

But this does not at all mean that stability of an initial system (157)–(158) is preserved during small variations of its coefficients since an additional check shows that a coefficient in the first degree λ in polynomial (160) turns out to be a difference of numbers each of which is larger than it by two orders. Therefore if coefficients in an initial system change only by $|\varepsilon_i| \leq 0,0025$ then coefficient in λ in a characteristic polynomial (160) can become negative and the system will loose stability.

A conclusion about parametric stability of system (157)–(158) that has been, for example, obtained on the basis of a known methodics of investigating a characteristic polynomial proposed in [63] will not be reliable and trustworthy. An additional check shows that stability will be preserved only if $|\varepsilon_i| < 0,25\%$. For the majority of practical applications a system with so small stability reserve is equal to an unstable system. To a correct conclusion about stability reserves we can come only on the basis of initial equations (157)–(158). An equivalent transformation of system (157)–(158) into an equivalent to it system of equations (159)–(161) (without changing solutions themselves) strongly changes a value of stability reserves. Really system (157)–(158) looses stability during the change of coefficients only by $0,25\%$. But the investigation of transformated system (159)–(161) speaks about large reserves of stability which is far from reality.

Note that if during coefficients variations of an initial system of equations changes (becomes negative) any of coefficients in a characteristic polynomial then this fact also speaks about the loss of solutions stability.

Note that if during variation of coefficients in an initial system of equations changes (becomes negative) any of coefficients in a characteristic polynomial then this again speaks about loosing stability of solutions. But processes that occur after the loss of stability depends on – which of members in a characteristic polynomial has changed a sign. If it has changed a sign and has become negative and small by an absolute value a higher member of a characteristic polynomial then solutions of the system start in a very quick way to strongly increase (in an absolute value). We have already seen this on an example of system (154) and earlier (in §9) we have spoken about this during the investigation of a system of equations (27)–(28). If any of lower coefficients in a characteristic polynomial becomes small and negative then solutions start to increase without any limits (in an absolute value) but they slowly increase.

So, for example, solutions of equations

$$(D^2 - \varepsilon D + 1)x = 0 \tag{163}$$

with a characteristic polynomial

$$\lambda^2 - \varepsilon\lambda + 1 \tag{164}$$

are of the form:

$$x(t) = e^{\frac{\varepsilon t}{2}}(C_1 sin\sqrt{(1-\frac{\varepsilon^2}{4})}t + C_2 cos\sqrt{(1-\frac{\varepsilon^2}{4})}t) \tag{165}$$

and if ε are small they increase in the course of time without any limit but very slowly.

Example 7. A system of three differential equations with three variables.

A system of three equations is posed:

$$\left.\begin{array}{c}(D^3+2,15D^2+1,23D+0,98)x_1+(1,05D^2+0,95D+1,12)x_2+(1,08D^2+2,7D+3,23)x_3=0\\(1,07D^2+2,12D+3,75)x_1+(1,09D+2,98)x_2+(2,5D+2,08)x_3=0\\(1,16D+3,63)x_1+1,18x_2+1,15x_3=0\end{array}\right\} \tag{166}$$

It is necessary to check stability of this system and the stability preserving during coefficients variations.

Conventionally – a characteristic polynomial is computed:

$$HP=\begin{vmatrix}\lambda^3+2,15\lambda^2+1,23\lambda+0,98 & 1,05\lambda^2+0,95\lambda+1,12 & 1,08\lambda^2+2,7\lambda+3,23\\ 1,07\lambda^2+2,12\lambda+3,75 & 1,09\lambda+2,98 & 2,5\lambda+2,08\\ 1,16\lambda+3,63 & 1,18 & 1,15\end{vmatrix} \tag{167}$$

and then its roots. If in all roots their real parts are negative then a system is stable.

But it is more useful to before hand check whether a coefficient in a higher member of a characteristic polynomial is a zero or a small difference of large numbers. The check is not very complex since a higher member (a member of the fourth order) will depend only on higher members of polynomials that are in a determinant (167) and it will be equal to a determinant:

$$\begin{vmatrix}1 & 1,05 & 1,08\\ 1,07 & 1,09 & 2,5\\ 1,16 & 1,18 & 1,15\end{vmatrix}=0,01 \tag{168}$$

An equality (168) at once has shown that already during small variations of coefficients in system (166) when $\varepsilon < 0,01$ the higher coefficient of a charateristic polynomial can change a sign and at once a necessary stability Stodola condition is broken. Reserves of stabiliy are small and a conclusion about solutions stability is not reliable.

In a rather simple system (166) this conclusion is at once evident but in more complex equations systems it is useful to apply results obtained in the first part of a book "Inverse table of signs" of a determinant that has been there described. Then we shall easily compute at what variations of coefficients a higher member in a characteristic polynomial will change a sign and the system will loose stability.

A numerical integration of system (166), computation of its solutions is also not reliable since during the change of a sign in a higher member a substantional change of all solutions

can occur as it has been already told during the investigation of example 4.

Example of securing reliability of computing technical objects

We have given a series of example of mathematical models for which computation results are not reliable and not trusworthy.

Surely in practice it is important not only to simply certify the unreliability of the computation but to find ways of securing its reliability.

This can be done due to the change of an object mathematical model. It is well known that one and the same technical object can be described by different mathematical models with different degree of exactness and detailability of describing processes that occur in the object. Therefore after the check which has shown that computation results primarily chosen a mathematical model are not reliable then we can transfer to another model that will secure reliability.

If the search of a good model have not given a success another way can be used – to change parameters of a projected technical object. Then automatically its mathematical model will change as well. This change of parameters (and sometimes – a construction) of a projected technical object must be carried out in such a way that a mathematical model of a changed technical object guaranteed the reliability of computations, of coinciding of computation results with a real behavior of an investigated object during possible small changes of its parameters.

Example 8. Optimal control for ships (vessels)

For vessels – tankers of the type "Kazbek" (displacement is equal to 16000 tons, velocity – 14 knots) a mathematical model of movement according to the course under the action of a rudder and perturbing forces due to the wind and sea roughness is the following equation of the second order:

$$(690D^2 + 17,2D)\theta = u + \varphi(t) \tag{169}$$

In equation (169) time t is expressed in seconds, θ – angle of deviation in a vessel from a posed course in degrees, u – angle of deviation in a rudder from a diametral in degrees, $\varphi(t)$ – a perturbing interaction from wind and sea roughness, a stationary arbitrary process, whose spectral density of power (spectrum) can be approximated by different analytical equations.

Usually it is recommended to approximate the spectrum by Rahman–Firssov formula:

$$S_\varphi(\omega) = \varphi^2 \cdot \frac{4\alpha}{\pi} \cdot \frac{\alpha^2 + \beta^2}{\alpha^2 + \beta^2 + \omega^2) - 4\beta^2\omega^2} \tag{170}$$

where φ^2 – a middle square of a perturbing interaction, ω – frequency, its measurement, $rad/s, \alpha, \beta$ – measured coefficients (measure – $1/s$). that depend on the intensity of roughness. For the roughness of a middle intensity usually we pose that $\beta = 1c^{-1}$ and $\alpha = 0,21\beta$.

As it is known the vessel loss of velocity is proportional to an integral

$$\Delta v = \lim_{T\to\infty} \frac{1}{T} \int_0^T (k^2\theta^2 + u^2)dt \tag{171}$$

where k^2 – unmeasured coefficient that depends on water weight above the vessel corps. For a tanker "Kazbek" it is equal to $0,25$.

A law of ruddez plant, i.e. a law of the dependence of a rudder deviation $u(t)$ from a deviation of a vessel on a given course $\theta(t)$ it is advisable to choose in such a way that the loss of velocity be the minimal one. In order to find an optimal law of control for a rudder plant and the projection of optimal autosteering that automatically realizing this law we can apply an optimizations theory of root – mean square functionals that has been given in books [35], [61].

For tanker "Kazbek" an optimal law of a control of steerage is of the form:

$$u = [690D^2 + 17,2D - \frac{690D^2 + 61,2D + 2,5}{0,973 - 0,06D}]\theta \tag{172}$$

In a book [35] a method of giving a formula (172) can be found. Here it is shown that really a control law (172) secures a stable movement of a vessel in its course on condition that vessel parameters are equal to calculated ones. But a control law (172) does not secure parametric stability.

Really, suppose that vessel parameters have deviated from computed ones by small values and its mathematical model (169) has become:

$$[690(1+\varepsilon_1)D^2 + 17,2(1+\varepsilon_2)D]\theta = u + \varphi(t) \tag{173}$$

If we substitute into formula (173) instead of its expression (172) we shall obtain:

$$[-41,4\varepsilon_1 D^3 + (690 + 690\varepsilon_1 + 1,03\varepsilon_2)D + (61,2 + 16,7\varepsilon_2)D + 2,5]\theta = (0,973 - 0,06D)\varphi \tag{174}$$

From formula (174) at once follows that already during infinitely small $\varepsilon_1 > 0$ stability os lost since a necessary Stodola condition of stability. A control law (172) for a practical use is not suitable. The computation result during rating parameters values for which the movement of a vessel according to a course is stable and it will substantially differ from an actual movement. Already during infinitely small $\varepsilon > 0$ the movement will become unstable.

In order to guarantee reliability of results in computationing we can slightly change parameters of a mathematical model or a spectrum of perturbing interaction. Here, naturally we shall obtain another control law (here we shall, surely, obtain another law of control). Such changes are lagimate approximations of experimental data about spectra of a rough sea with an approximately similar degree of approach describe real perturbing interactions on sea vessels.

For optimal control a question of securing parametric stability has been in details developed in [33] and [61] if a control object is posed in the form of the following mathematical model:

$$A(D)x = B(D)u + \varphi(t) \tag{175}$$

where x – controlled value, u – controlling interaction, $A(D)$ and $B(D)$ – arbitrary polynomials from differentiation operator $D = \dfrac{d}{dt}$. Here a degree of polynomial $A(D)$ is equal to n, a degree of polynomial $B(D)$ is equal to m and a spectrum of perturbing interaction $\varphi(t)$ is approximated by an even rational fraction:

$$S_\varphi = \frac{a_p\omega^{2p} + a_{p-1}\omega^{2p-2} + ... + a_0}{b_q\omega^{2p} + b_{q-1}\omega^{2q-1} + ... + b_0} \tag{176}$$

then a control law that secures a minimum of a root mean square functional will guarantee parametric stability only in such a case when Yu.Petrov criterium is satisfied.

$$p \geq m + q - 1 \tag{177}$$

where m – a degree of polynomial $B(D)$ in a mathematical model of a control object (175), p – a half part of a numerator in an analytical approximation of spectrum (176), q – a half part of a denominator degree in approximation (176).

For a mathematical model of tanker "Kazbek" (169) we shall have $m = 0$. For Rakhman-Firsov spectrum (170) we have: $p = 0, q = -2$. Since $0 < 0 + 2 = 1$ Yu. Petrov criterium is not satisfied and we can at once say that a control law (172) that applies a perturbing interaction with spectrum (170) will not apriori secure stability.

In order to secure parametric stability it is conventient to change spectrum in analytical approximation and thus – a control law. Since frequency characteristic of tanker "Kazbek" is almost completely in such a part of spectrum (171) where it is as yet almost constant and where it very slightly depends on frequency ω then it is possible (as it was proposed already in [64]) to approximate a spectrum of perturbing interaction simply by means of a constant value:

$$S_\varphi(\omega) = C. \tag{178}$$

Such an approximation corresponds to $p = 0$ and $q = 0$. For these values p and q Yu. Petrov criterium is satisfied. To spectrum (178) corresponds the following control law:

$$u_2 = -[43, 6D + 2, 5]\theta \tag{179}$$

Surely, the simplifying of analytical approximation in a spectrum can increase the loss of velocity but not much. In a publication [64] on page 137 a computation has been carried out that shows that if a true spectrum of a perturbing interaction exactly corresponds to formula (170) then even in this case during the computation of a regulator the change of spectrum (170) by spectrum (178) will increase velocity loss only by $9, 7\%$.

In order to take into account slowly changing components in perturbing interaction $\varphi(t)$ that have not been taken into account spectrum (178) an autosteering has been added by an integrating link with a small intensivity coefficient. And a control law becomes:

$$u_3 = -[43, 6D + 2, 5 + \frac{0, 005}{D}]\theta \tag{180}$$

On the basis of formula (180) for a long period of time (up to the appearance of digital control systems) a structure of automatic steering was realized. In its foundation lies a parallel connection of differentiated, intensifying and integrating link. This structure has been often used during the projection of automatic steerings. A specific numerical value of intensification coefficients in automatic steerings depend on a displacement of a vessel. Its velocity weight of water above courpses. And they have been computed by a methodics given in [64], pp. 132–149.

Additional materials on projection and computation of automic steerings that secure a small number of rearrangement of steering, the exactness of looking after the movement by a river channel etc. and at the same time invariably preserving parametric stability are given in [62], pp. 215–226 and in [35], pp.243–248. There are also given examples that guarantee parametric stability and reliability of computing algorithms for many other optimal control systems.

It is necessary to note that just in the theory of optimal control systems a problem of securing the reliability of computing algorithms used during projection is sharp. Really, if we try to secure optimal and the best quantity of work in control systems we inevitable approach to the limits of stability and about this important problem it is always necessary not to forget.

The experience in projecting and computing of safely working optimal control systems can be applied during computing other technical objects as well – and they must not be (by all means) optimal.

§34. The check of stability preservation in system of the type $\dot{X} = AX$ during finite variations of elements of coefficients matrix.

In this section we shall turn to the investigation of stability in different objects during finite (not infinitely small) parameters variations of mathematical models for real objects. During the analysis of these variations influence on stability we shall widely apply methods given in the first part of the book and in particular – we shall use "tables of signs" of determinants. They will allow to establish the largest velocity of increasing or decreasing of determinants.

It is well–known (see, for example, [66]) that the investigation of stability of the majority of objects can be reduced to the investigation of their mathematical models in a linear approach.

The most convenient and the most often used form of writing a system of equations for an object in a linear approach is a vector–matrix form:

$$\dot{X} - AX = 0, \tag{181}$$

where X – a vector of investigated variables $x_1(t); x_2(t); ...; x_n(t)$ that describe the behavior of an object, A – a square coefficients matrix of a system of a measure nxn. Also often an equivalent form of writing is applied:

$$\dot{X} = AX \tag{182}.$$

Elements in matrix A depend on object parameters and are determined by them. In §7 as an object a regulated electrodrive was considered (equations (52)-(56)) whose mathematical model in a normal Cauchy form can be written in the form of following three equations:

$$\left.\begin{array}{c} \dot{x}_1 = -\left(\dfrac{k_0 + k_1}{m}\right)x_1 - \dfrac{k_3}{m}x_2 - \dfrac{k_4}{m}x_3 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = -(\alpha^2 + \beta^2)x_2 - 2\alpha x_3 \end{array}\right\} \tag{183}$$

where x_1 – a deviation of rotation frequency in an electrodrive from a rating one, x_2 and x_3 – auxillary variables that were introduced in order to take into account a spectrum of perturbing interaction according to formulas (53)-(55) of the first part of the book, k_0 – a coefficient of a viscous friction, $k_1; k_2$ and k_3 – coefficients of strengthening of a regulator (56). In order to reducing to a normal Cauchy form of equations (52)-(56) a variable x_2 is excluded that is in equation (56) and variables x_3 and x_4 are transformed into x_2 and x_3.

For system (183) a matrix A is of a form:

$$\begin{pmatrix} -\dfrac{k_0 + k_1}{m} & -\dfrac{k_3}{m} & -\dfrac{k_4}{m} \\ 0 & 0 & 1 \\ 0 & -(\alpha^2 + \beta^2) & -2\alpha \end{pmatrix} \tag{184}$$

and here we clearly see the dependence of matrix elements on object parameters.

Another example (in [65], p. 255) it is shown that equations for an electric chain that consists of a successive joining of active resistance R, induction L and capacity C can be written in the form of equations

$$\left.\begin{aligned} \dot{x}_1 &= \frac{1}{LC}x_2 \\ \dot{x}_2 &= -x_1 - \frac{R}{L}x_2 \end{aligned}\right\} \tag{185}$$

where x_i – tension in capacity and x_2 – flow coupling of induction. In this case matrix A will become:

$$A = \begin{pmatrix} 0 & -\frac{1}{LC} \\ 1 & \frac{R}{L} \end{pmatrix} \tag{186}$$

Examples of forming equations systems in a normal form and matrixes A for a more complex systems are given in many other text-books.

As we have already shown in previous sections a normal form of writing equations can be obtained from different initial equations systems by means of equivalent transformations. Therefore it is necessary to note whether during equivalent transformations some important properties of equations have changed and in particular – parametric stability. Earlier it has been shown that if system (183) was obtained by equivalent transformations from equations that corresponded structurally to a scheme shown on figure 12 then these transformations had changed a parametric stability. And an initial system (and thus – a real object) has lost stability during infinitely small deviations of parameters from calculated values. But equations (183) do not possess this property and therefore do not reflect real behavior of an object during infinitely small variations of parameters.

Therefore the first step in investigating a system of the form (181) must become – to check whether during equivalent transformations have changed a property of its parametric stability. In previous sections we have in details said about this necessary step of our investigation.

Now let us suppose that the first step of checking has been carried out and we have seen that during infinitely small deviations of parameters from calculated values do not occur and we can turn to the second step of computing stability – the investigation of preservation (or not preserving) stability of system of the form (181) during small (but finite) variations of coefficients in matrix A.

These small finite variations are inevitable since due to a finite exactness of manufacturing any real technical objects their real parameters will inevitably differ from calculated values by finite values. And in the course of exploitation of an object additional variations will arise.

The resulting parameters variations in an object (and thus – elements in matrix A) are different in values and a sign (they can be either positive or negative). An absolute value of

possible coefficients variations usually can be if only estimated from above while analizing the exactness of manufacturing an object and possible maximal values of variations of its parameters in the course of exploitation. At the same time a sign of coefficients variations most often cannot be at all predicted.

Us a whole we can only state (as a rule) about each of matrix A elements in equations (181) that they are included in intervals:

$$a_{ij}(1-\varepsilon_{ij}) \leq \bar{a}_{ij} \leq a_{ij}(1+\varepsilon_{ij}) \tag{187}$$

where $\bar{a}_{ij}$ – a true value of elements a_{ij} that is not known to us; a_{ij} – a value that was accepted during the calculation ε_{ij} – a number that is small in comparison to a unity.

Let us start the investigation from the most simple (but the most dangerous) particular case when all ε_{ij} are equal in an absolute value (i.e. – for all i and j $|\varepsilon_{ij}| = \varepsilon$) but their signs do not depend on each other.

Later we shall see that if ε is unchanged stability or instability of system (181) it is necessary to investigate it fundamentally during the most unfavourable combination of signs ε_{ij} which must be first of all be found. It is not easy to find these unfavourable combinations since a number of possible combinations of variations signs ε_{ij} increase very quickly with the increase of an order if n system (181) since a number of combinations k is equal to 2^{n^2}. When $n = 2$, k will be $2^4 = 16$, when $n = 3$, $k = 2^9 = 512$; if $n = 4$, $k = 2^{16} = 65536$, if $n = 5$, $k = 2^{25} > 3 \cdot 10^7$ and if $n = 6$, $k = 2^{36} > 10^{12}$.

It is clear that at such orders of n matrix A in system $\dot{X} = AX$ which occur in technical problems a direct sorting out is often cannot be carried out even for the most quickly operating computers.

Therefore instead of investigating matrix A usually we resorted to the investigating of its characteristic polynomial which as it is known is equal to a determinant:

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22}-\lambda & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn}-\lambda \end{vmatrix} \tag{188}$$

A characteristic polynomial of matrix A (at the same time it is a determinant of the form (188)) is a polynomial of an order n whose coefficients are functions of elements in matrix A. In particular a free member of a characteristic polynomial is equal to a determinant in matrix A – i.e. a determinant of the n-th order and coefficients before other members can be expressed by means of determinants of less orders consisting of elements of matrix A. A characteristic polynomial of a measure 2x2, i.e. a determinant

$$\begin{vmatrix} a_{11}-\lambda & a_{12} \\ a_{12} & a_{22}-\lambda \end{vmatrix}$$

can be written in the form:

$$\lambda^2 - (a_{11}+a_{22})\lambda + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

coefficients in λ can be considered as a sum of determinants of the first order – numbers a_{11} and a_{22}.

A characteristic polynomial the nth matrix of a dimension 3x3 – i.e. a determinant

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} \tag{189}$$

can be written in the form:

$$-\lambda^3-(a_{11}+a_{22}+a_{33})\lambda^2+(\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}+\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix})\lambda+ \\ +\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \tag{190}$$

A characteristic polynomial in matrix (measure – 4x4 can be written in the form:

$$\lambda^4-(a_{11}+a_{22}+a_{33}+a_{44})\lambda^3+(\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{14} \\ a_{41} & a_{44} \end{vmatrix}+\begin{vmatrix} a_{22} & a_{24} \\ a_{42} & a_{43} \end{vmatrix})\lambda^2- \\ -(\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{13} & a_{14} \\ a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \end{vmatrix}+\begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix})\lambda+ \\ +\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} \tag{191}$$

Similarilly for matrix of any order (nxn) coefficients in all degrees $\lambda n-r$ can be expressed by determinants. As it is known (see, for example, [70], p. 400) a coefficient in λ^{n-2} is equal to a multiplier $(-1)^N$ taken with a sum of main minors of nth order in determinant of a matrix.

If elements of matrix A undergo variations and they are posed by intervals of their possible values – i.e. by inequalities (187) then even coefficients of its characteristic polynomial

$$(-1)^n\lambda^n+a_{n-1}\lambda^{n-1}+...+a_0 \tag{192}$$

will also be in some intervals:

$$\underline{a}_i \le a_i \le \bar{a}_i \tag{193}$$

where $\underline{a}_i$ – is the least possible value of coefficient a_i in a characteristic polynomial and $\bar{a}_i$ – the largest value from possible ones.

The problem of system (181) stability during elements variations of matrix A can be now reduced to the check of Hurwitz polynomial (192) whose coefficients satisfy conditions (193). Here we can also act by means of a direct sorting out and carry out a check of all possible combinations of coefficients $\underline{a}_i$ and $\bar{a}_i$ to be a Hurwitz ones. A number of all

possible combinations is equal to 2^n. It's less than 2^{n^2} but too much all the same.

A decisive break in simplifying the solution was carried out in 1978 by an employce of an applied mathematics – control processes department of St.Petersburg state university V.L.Kharitonov. In [63] he has shown that in order to check stability of an interval polynomial – and any polynomial of the form $a_n\lambda^n + a_{n-1}\lambda^{n-1} + ... + a_n$ as well whose coefficients satisfy conditions (193) it is sufficient to check that only four polynomials – just polynomials

$$\bar{a}_n\lambda^n + \underline{a}_{n-1}\lambda^{n-1} + \underline{a}_{n-2}\lambda^{n-2} + \bar{a}_{n-3}\lambda^{n-3} + ... \tag{194}$$

$$\underline{a}_n\lambda^n + \bar{a}_{n-1}\lambda^{n-1} + \bar{a}_{n-2}\lambda^{n-2} + \underline{a}_{n-3}\lambda^{n-3} + ... \tag{195}$$

$$\bar{a}_n\lambda^n + \bar{a}_{n-1}\lambda^{n-1} + \underline{a}_{n-2}\lambda^{n-2} + \underline{a}_{n-3}\lambda^{n-3} + ... \tag{196}$$

$$\underline{a}_n\lambda^n + \underline{a}_{n-1}\lambda^{n-1} + \bar{a}_{n-2}\lambda^{n-2} + \bar{a}_{n-3}\lambda^{n-3} + ... \tag{197}$$

are Hurwitz ones. Sometimes they are called "summit" or "angular" polynomials.

An article by V.L.Kharitonov has received a worthy fame. There appeared a lot of works (see [67, 68, 69]) and many others dedicated to the same subjects that were continuing and developing results by V.L.Kharitonov.

Great efforts have been turned to try solving a similar problem of checking stability for systems (181) or (182) during variations of matrix A coefficients (without turning to the investigation of a characteristic polynomial of a matrix). Coefficients in characteristic polynomial are connected by such relations with coefficients in a majority of mathematical models for real objects that are too complex. In systems (181) and (182) these relations are more simple. Therefore the solution of a problem of checking stability directly for these systems (if we take into account variations of their parameters) has been during many years a tempting aim of a lot of investigators. See publication [67; 68; 69] and many others. But the problem has turned very difficult and a methodics applied by V.L.Kharitonov for polynomials in this case has not obtained any success.

In order to solve this difficult problem we shall apply results given in the first part of this book

As it is used in a series of text–books we shall write a characteristic polynomial of matrix A in system (181) equal to a determinant (188) in such a way that a coefficient in its higher member, in λ^n be equal to $+1$. For this it is sufficient to multiply a characteristic polynomial (188) by $(-1)^n$ which will not change neither its roots nor stability conditions. Then a necessary condition of stability of system (181) with such a form of writing a characteristic polynomial (Stodola condition) will become the following simple form. Coefficients in all its members (including its free member – a coefficients in λ in a zero order – equal to a determinant of matrix A multiplied by $(-1)^n$) must be positive. If we use the writing of a characteristic polynomial in the form of a determinant (188) and without multiplying it by $(-1)^n$ then Stodola condition is written in a more complex form

and undergraduates memorize it in a worse way.

If we apply such a form of writing in a Hurwitz characteristic polynomial its free member $(-1)^n a_0$ equal to a determinant:

$$(-1)^n \cdot \begin{vmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{vmatrix} \tag{198}$$

is by all means positive. During variations of determinant (198) elements stability of a characteristic polynomial can be lost the more quickly the more quickly will diminish a free member. As soon as it achieves a value 0 stability will be lost. But in the first part of the book (and earlier in publication [6]) we have established that if absolute values in variations are the same a determinant (198) will diminish with the largest velocity in such a case if signs of variations of its elements correspond to its "inverse table of signs".

Hence at once follows a simple algorithm for checking necessary conditions for stability of systems (181) and (182) during variations of elements in matrix A. By using estimates of absolute values of numbers ε_{ij} ($|\varepsilon_{ij}| \leq \varepsilon_{ij\ max}$) in equalities (187) we put their signs in correspondence with "an inverse table of signs" in a determinant (198) and we shall compute the determinant:

$$(-1)^n \cdot \begin{vmatrix} a_{11}(1 \pm \varepsilon_{11}) & a_{12}(1 \pm \varepsilon_{12}) & ... & a_{1n}(1 \pm \varepsilon_{1n}) \\ ... & ... & ... & ... \\ a_{n1}(1 \pm \varepsilon_{n1}) & a_{n2}(1 \pm \varepsilon_{n2}) & ... & a_{nn}(1 \pm \varepsilon_{nn}) \end{vmatrix} \tag{199}$$

we must take into account signs of ε_{ij} that correspond to "inverse table".

If here a determinant (199) turns out to be not positive then system (182) by all means can loose stability during unfavourable combinations of absolute values (in the limits of $(\varepsilon_{ij} \leq \varepsilon_{ij\ max})$ and signs of elements matrix A variations.

If determinant (199) is positive it can be advised to additionally investigate determinants variations that enter into coefficients of other members in a characteristic polynomial – from a member with the first degree λ up to a member with λ^{n-1}. Theoretically it is possible that in a free positive member of a characteristic polynomial some of its other members (computed while taking into account variations of elements a_{ij}) will become not positive – but this fact is rare – since as it has been said in the first part of the book – during the same absolute values of elements variations a velocity of diminishing a determinant increases with the increase of its order.

Therefore a free member of a characteristic polynomial is sensible to elements variations. And in practice often are content with checking the fact that its sign during parameters variations has not changed.

Examples.

Let us consider a system of equations:

$$\left.\begin{array}{l}\dot{x}_1 = -7x_1 - 6x_2 \\ \dot{x}_2 = -8x_1 - 7x_2\end{array}\right\} \tag{200}$$

with characteristic polynomial:

$$\lambda^2 + (7+7) + \begin{vmatrix} 7 & 6 \\ 8 & 7 \end{vmatrix} = \lambda^2 + 14\lambda + 1 \tag{201}$$

having negative roots $\lambda_{1,2} = -7 \pm \sqrt{49-1}, \lambda_1 = -0,0718, \lambda_2 = -13,422$. If coefficients in system (200) that have rating values are stable but it can loose stability during their variations. If variations of all matrix elements do not exceed $\pm\varepsilon_m$ from their rating values then the most unfavourable combination of their signs will be such that corresponds to "an inverse table of signs" in a determinant

$$\begin{vmatrix} 7 & 6 \\ 8 & 7 = 1 \end{vmatrix} \tag{202}$$

that is of the form:

$$\begin{vmatrix} - & + \\ + & - \end{vmatrix} \tag{203}$$

"An inverse table of signs" for determinant (202) as well as a universal one for all determinants of the second order with positive elements we have already computed in the first part of own book.

If we compute a determinant with variations

$$\begin{vmatrix} 7(1-\varepsilon_m) & 6(1+\varepsilon_m) \\ 8(1+\varepsilon_m) & 7(1-\varepsilon_m) \end{vmatrix} = 1 - 194\varepsilon_m + \varepsilon_m^2, \tag{204}$$

we can state that it has for the first time turned into zero when $\varepsilon_m = 0,005155$. This means that if $\varepsilon_m \geq 0,005155$ a system (200) can already loose stability.

By computing a value of the second member of a characteristic polynomial during variations of parameters we can state that during the most unfavourable combination of variations signs a coefficient in the second member will be equal to

$$14 - 2\varepsilon_m \tag{205}$$

and it will change a sign only if $\varepsilon_m = 0,1428$ that is much more than a variation that changes a sign of a free member. As said before the most dangerous (in relation to a possible loss of stability) turn out to be variations of elements in a free member of a characteristic polynomial.

Thus when $\varepsilon_m \leq 0,005155$ the system will apriori preserve stability but if $\varepsilon_m \geq 0,005155$ stability can disappear. If variations have obtained only elements of the high line in determinant (202) if the same the most dangerous for stability "inverse table of signs" (203) turns into

$$\begin{vmatrix} 7(1-\varepsilon_m) & 6(1+\varepsilon_m) \\ 8 & 7 \end{vmatrix} = 1 - 97\varepsilon_m \tag{206}$$

then in this case system (200) can loose stability when $\varepsilon_m \geq \frac{1}{97} = 0,0103$ and it will preserve stability if $\varepsilon_m \leq 0,0103$.

Similarilly we can check stability preservation not during relative but during absolute variations of elements a_{ij} in matrix A in system (179) when they, for example, are subjected to inequalities:

$$a_{ij} - |\varepsilon_{ij}| \leq \bar{a}_{ij} \leq a_{ij} + |\varepsilon_{ij}|, \tag{207}$$

where $\bar{a}_{ij}$ are true and not known to us coefficients values, a_{ij} – rating values that are used during calculations and ε_{ij} satisfy the inequalities:

$$|\varepsilon_{ij}| \leq \varepsilon_{ijm} \tag{208}$$

where ε_{ijm} – values that are small in comparison with a_{ij}.

Example 2. For system (200) if inequalities (207) and (208) are taken into account "an inverse table of signs" will preserve a form (203) and if all elements in matrix A have obtained variations $\pm\varepsilon_m$ then if a "table of signs" (203) is taken into account a matrix determinant will become:

$$\begin{vmatrix} 7-\varepsilon_m & 6+\varepsilon_m \\ 8+\varepsilon_m & 7-\varepsilon_m \end{vmatrix} = 1 - 28\varepsilon_m \tag{209}$$

and the system can loose stability if $\varepsilon_m \geq \frac{1}{28} = 0,0357$.

If a combination of signs is not the most unfavourable for preserving stability and it is not corresponding to "an inverse table of signs" then during the same absolute values in variations stability will be preserved. So, for example, if all variations of matrix elements have one and the same negative sign then a matrix determinant will become:

$$\begin{vmatrix} 7-\varepsilon_m & 6-\varepsilon_m \\ 8-\varepsilon_m & 7-\varepsilon_m \end{vmatrix} = 1 \tag{210}$$

and it will not depend on ε_m and coefficient in the second member of a characteristic polynomial equal to $(7 - \varepsilon + 7 - \varepsilon)$ will change a sign in $\varepsilon_m = \frac{1}{7}$.

Example 3. Let us consider system (182) with matrix of a size 3x3:

$$A = -\begin{pmatrix} 1 & 2 & 3 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{pmatrix}. \tag{211}$$

If there are rating values of coefficients a free term of a characteristic polynomial is equal to:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{vmatrix} = 8, \tag{212}$$

but during variations of determinant elements it can change a sign. In example 3 we again consider relative variations subjected to inequalities (187). A determinant (212) has already been considered in part 1, §6. There its "direct" and "inverse" tables of signs have been computed. "An inverse table of signs" for determinant (212) is of the form:

$$\begin{vmatrix} + & + & - \\ - & + & - \\ - & - & + \end{vmatrix} \tag{213}$$

In the same place in the first part of the book, in §7 it has been stated that if all variations of determinant elements are equal in an absolute value, i.e. – $|\varepsilon_{ij}| = \varepsilon_m$ but their signs correspond to "an inverse table of signs" (213) then determinant (212) will for the first time turn into zero when $\varepsilon_m = 0,0515$.

This means that if $\varepsilon_m \geq 0,0515$ stability of system $\dot{X} = AX$ with matrix (211) can disappear.

Note that as it has already been shown in part one when ε_m are the same the velocity of a decrease of a determinant will be the largest if signs of variations in all determinant elements are independent. If signs of variations are connected between themselves by any dependences then a determinant decreases more slowly and losses of stability in system (182) occurs later.

Now let us write (in a complete way) a characteristic polynomial of matrix (211) on the basis of formula (199) but we must reduce the polynomial (by multiplying by $(-1)^n$) to a form that the higher member remains positive.

We shall obtain

$$\begin{aligned} \lambda^3 + (1+1+5)\lambda^2 + (\begin{vmatrix} 1 & 2 \\ 4 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 3 \\ 3 & 5 \end{vmatrix} + \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix})\lambda + \begin{vmatrix} 1 & 2 & 3 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{vmatrix} = \\ = \lambda^3 + 7\lambda^2 + 14\lambda + 8 \end{aligned} \tag{214}$$

Again it is not difficult to see that the change of a sign in coefficients of λ^2 and λ can occur only during much more large (in an absolute value) variations of elements of matrix (211) in comparison with variations that are able to change a coefficient sign if a degree of λ is zero (of a free member).

As it has been shown in the first part of the book if there are the same absolute values in variations of determinant elements it changes the quicker the higher is its order. Therefore (as we have already said) during the check of preserving necessary stability conditions often we limit ourselvesley by checking a free member although surely it is advisable to check coefficients in other degrees of λ as well.

Let us also note that if variations of elements of matrix A in system (182) are independent there is very little probability that a combination of signs in variations corresponding to just "inverse table of signs" is realized. Besides it may also correspond to the largest possible decrease of a determinant.

This probability is equal to $\frac{1}{2^{n^2}}$ and it quickly decrease with the growth of n.

Therefore when, for example, during the investigation of example 3 we have established that if $\varepsilon_m = 0,0515$ a free member of a matrix can be equal to zero and stability is lost but the loss of stability here is improbable (probability is equal to $\frac{1}{2^9} = \frac{1}{512}$). But, for example, when $\varepsilon_m = 2 \cdot 0,0515$ and in table of signs (213) a determinant (212) will already obtain a value that is near – 10. This means that such a value of a free member can be realized not only by one unique combination of possible signs of variations (a combination that corresponds to a table of signs in (213)). But it may be realized by means of very many combinations and thus (it's important!) now the loss of stability will not have a nonsmall possibility at all.

It would be very useful to compute exact probabilities of stability loss in ε_m that are by k times such ε_m in which a free member will for the first time reach a value zero. But this question will be investigated later.

Surely, if during parameters variations any of coefficients in characteristic polynomial (multiplied by $(-1)^n$) of system (182) has become negative or equal to zero this means that a necessary condition of Stodola has been broken and a system is apriori unstable. But system (182) (if $n \geq 3$) can be unstable in all positive coefficients of a characteristic polynomial as well (multiplied by $(-1)^n$).

In order to obtain a more firm in preserving or not preserving stability in system of the form (182) during variations elements of matrix A it is adised to check the fulfilment of necessary and sufficient conditions that a characteristic polynomial is Hurwitz one. During such a check methods of computing help the largest and the smallest values of determinants that enter into formulas (190), (191) (and such ones if $n > 4$) during the variations of determinants elements. These methods have been stated in the first part of our book.

Example 4. Let us again examine system (182) with matrix (211). Let us suppose that all elements of a matrix can undergo variations that are equal to $\frac{1}{100}$ from an initial value (i.i. $\varepsilon_m = 0,01$) and let us put a question whether an investigated system preserves stability during such variations.

By applying a methodics given in the first part and by using there "a direct" and "an inverse" signs table for a determinant

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{vmatrix} \tag{215}$$

and for determinants

$$\begin{vmatrix} 1 & 2 \\ 4 & 1 \end{vmatrix}; \begin{vmatrix} 1 & 3 \\ 3 & 5 \end{vmatrix}; \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} \tag{216}$$

that enter into formula (214) we shall easily compute that when $\varepsilon_m = 0,01$ and we shall obtain:

$$6,3804 \leq \begin{vmatrix} 1(1 \pm 0,01) & 2(1 \pm 0,01) & 3(1 \pm 0,01) \\ 4(1 \pm 0,01) & 1(1 \pm 0,01) & 2(1 \pm 0,01) \\ 3(1 \pm 0,01) & 4(1 \pm 0,01) & 5(1 \pm 0,01) \end{vmatrix} \leq 9,6604 \tag{217}$$

and similarly

$$\begin{aligned} -7,1207 \leq \begin{vmatrix} 1(1 \pm 0,01) & 2(1 \pm 0,01) \\ 4(1 \pm 0,01) & 1(1 \pm 0,01) \end{vmatrix} \leq -0,68007 \\ -4,2804 \leq \begin{vmatrix} 1(1 \pm 0,01) & 3(1 \pm 0,01) \\ 3(1 \pm 0,01) & 5(1 \pm 0,01) \end{vmatrix} \leq -3,7204 \\ -3,2603 \leq \begin{vmatrix} 1(1 \pm 0,01) & 2(1 \pm 0,01) \\ 4(1 \pm 0,01) & 5(1 \pm 0,01) \end{vmatrix} \leq -2,7403 \end{aligned} \tag{218}$$

and thus a coefficient in the third member of a characteristic polynomial during parameters variations will be in the limits:

$$13,2614 \leq (\begin{vmatrix} 1 & 2 \\ 4 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 3 \\ 3 & 5 \end{vmatrix} + \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}) \leq 14,7214. \tag{219}$$

A coefficient in the second member of a characteristic polynomial (in a member with λ^2) will be in the limits of

$$6,93 \leq [1 \cdot (1 \pm 0,01) + 1 \cdot (1 \pm 0,01) + 5 \cdot (1 \pm 0,01)] \leq 7,07 \tag{220}$$

Thus during parameters variations coefficients in all members of a characteristic polynomial (multiplied by $(-1)^n$) remain positive and a necessary Stodola condition has been fulfilled. It remains to check the last condition (which together with Stodola condition will be necessary and sufficient): the product of middle members of a characteristic polynomial must be more than a product of extreme members. It is sufficient to also check such combinations of variations signs at which middle members are the most little from possible ones but the lost ones – the largest, i.e. it is necessary to check the fulfillment of inequality:

$$6,93 \cdot 13,2614 > 1 \cdot 9,6604 \tag{221}$$

Inequality (221) is certainly fulfilled. And this means that system (182) with matrix (211) when variations of its elements are relative that satisfy a condition $|\varepsilon_{ij}| \leq 0,01$ preserves stability.

Note that carried by us a check is very "strict" as we have supposed the possibility of such combination of signs variations at which simultancously the second and the third member of a characteristic polynomial reach minimally possible values and the fourth member reaches a maximally possible value. But this admission (persistent calculations) leads to have a reserve for a reliable conclusion about preserving stability.

Making this note we can propose the following algorithm of checking the preservation of stability in system (182) for matrix A $(3x_3)$ during variations of its elements in a characteristic polynomial (multiplied by $(-1)^3$) and compute during known estimates of $|\varepsilon_{ij}|$ the largest and the smallest values of determinants that enter into formula (190).

If the most small values of coefficients in all members of a polynomial are positive and a product of the least small values of coefficients in middle members of a polynomial is larger than the most large value of a free member then the system will apriori preserve stability.

Similar algorithms can be formed also for systems of the type (182) with matrixes whose measure are more than 3x3.

Let us also note that during the computation of stability it is necessary to take into account a specification given in the first part of our book: "tables of signs" during variations of elements in determinants preserve their form up to the time when algebraic additions of elements do not change their sign. Although for the majority of determinants algebraic additions preserve their sign during elements variations but there exist special cases when even during a small increase of variations algebraic additions change signs and this fact leads to the change in "tables of signs" . Although such special cases occur rarely it is necessary to take them into account in order to secure that computations be correct.

The above material has shown that "tables of signs" in matrix A of system (182) can greatly help us in solving the following two problems:

1. to depict such systems which already due to checking results of sign in a determinant of matrix A can apriori loose stability during some variations of matrix A elements. If during matrix elements variations it turns out that $(-1)^n \cdot \det A > 0$ then the system can be after variations of elements stable and unstable.

2. it is necessary to depict such systems that during given maximal (in their absolute value) variation of matrix A elements preserve stability.

The first problem is easily solved the second one – is more difficult. Therefore we can apply "a table of signs" first of all in order to minimize a value of computations by applying more complex methods of depicting such systems that preserve stability during variations of matrix A elements only for systems that have not endured a simple check in preserving a sign in its determinant.

§35. The synthesis of control systems with good reserves of stability.

Recall that we call an optimal system of control such a system that is the best one from all possible according to some quality criterium. There is a lot of quality criteria. They can be quick-operating systems, a damping velocity of a transient processes, a small reregulation etc. All these questions have been in details considered in text-books and monographs on automatic control, optimal control and in the following books: [35], [42], [49], [58], [61], [62], [64], [65], [67], [68], [69] etc.

Any real control object must, be by all means, satisfy not one but several quality criteria. Since in practice always an object that is the best from all possible ones cannot (as we know) cannot be the best according to one quality criterium and at the same time it cannot be best according to other criterium then usually we act in such a way. The main, the most important for this object quality criterium is chosen the control is computed that is sufficiently good (or optimal) in this criterium. And then we look in what degree this control secure an accepted value of other qualities criteria. If a value of some criterium turned out to be inacceptable then a control is changed by sacrifying a part of a value of the main criterium and try to achieve a good compromise between different quality criteria. In more details this question was examined earlier in [35], p.p. 132-158 and 200-234 and in many other books.

There exist such control objects for which the main (although, surely, not the only one) quality criterium is a value of stability reserves (to them we can, for example, regard objects in which their parameters can greatly change in the course of exploitation). We shall examine such control objects whose mathematical objects are systems of equations (182)–(183) but to which control interactions have been put. Therefore their mathematical models have a well-known form in a control theory:

$$\dot{X} = AX + BU \tag{222}$$

where U – vector-column of control, B – vector-column of coefficients for regulators. These systems in initial state without control can be stable or unstable.

The control must, first of all, secure stability of a system if without control it is unstable and, secondly, it must secure stability during variations of matrix A parameters

Let us consider control that is formed according to the principle of an inverse connection and that linearily depends on X:

$$U = KX, \tag{223}$$

where K — vector-line of amplification coefficients in each of channels of an inverse connection. An object whose mathematical model is equation (223) is called a linear regulator with inverse connection. By closing system (222) by a control (220) we shall obtain an equation of a closed control system

$$\dot{X} = (A + BK)X, \tag{224}$$

where BK – a matrix of dimension nxn (it's a product of vector-column by vector-line).

Due to the choice of vector K elements undertaken by an engineer and who projects a control system we can secure any values of matrix BK elements. Thus let us consider how it is necessary to choose their values in order to secure stability preservation during variations of parameters that are the largest in an absolute value and during the most unfavourable combination of their elements.

Let us return to the investigation of system (200) with a characteristic polynomial (201). In previous section we have already stated that the system is stable but it can loose stability during variations of elements in matrix A that are only from a part of 0,5155% their rating values.

Surely such stability reserves cannot be considered sufficient. An object whose mathematical model is of the form (200) will work unreliably. And it is necessary to greatly increase stability reserves. As we have already said this can be carried out at the expense of a control with an inverse connection and at the expense of a choice of values in elements of matrix BK in equations (224). For a system (200) consisting of two equations this matrix has only four elements:

$$BK = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \tag{225}$$

But already in this the simplest case a choice of values C_{ij} that increase stability is not at all trivial.

By picking out different values of elements $C_{11}; C_{12}; C_{21}; C_{22}$ of matrix (225) at the expense of a choice of coefficients in amplification regulator (223) we can increase stability reserves and also – decrease them. And we can (in general) make U projected object of control unstable.

Example 4

If matrix (225) is chosen in the form:

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{226}$$

then if we take into account a control object (200) that is closed by a regulator that secures matrix (226) this object will be described by equations:

$$\left.\begin{aligned} \dot{x}_1 &= (-7+1)x_1 + (-6-1)x_2 = -6x_1 - 7x_2 \\ \dot{x}_2 &= (-8+1)x_1 + (-7+1)x_2 = -9x_1 - 6x_2 \end{aligned}\right\} \tag{227}$$

with a characteristic polynomial:

$$\lambda^2 + (6+6)\lambda + \begin{vmatrix} 6 & 7 \\ 9 & 6 \end{vmatrix} = \lambda^2 + 12\lambda - 27 \tag{228}$$

and instead of increasing stability it will become unstable.

Surely, for systems that consist of two equations it is rather easy to find a form of matrix (225) even ley means of sorting out. This matrix secures the increase of stability reserves but for systems consisting of three, four or more equations difficulties increase quickly. Apriori no sorting out we can't do.

In order to solve this problem the main help will be (described in the first part of our book) "a direct" and "inverse tables of signs" in determinants that secure the largest velocities of the increase ("a direct" table) and decrease (an inverse table) of determinants while they change its elements.

Example 5

In previous section we have established that the most dangerous in the stability loss factor during parameters variations in control object is the change of a sign in a free member of its characteristic polynomial – a determinant of matrix A in system (182). Let us recall – as it is said in a previous section – that we are examining such characteristic polynomials (that are reduced to a "routine" form) when a coefficient in a higher member is positive and – thus a necessary condition for stability is a positivity of a free member. Hence it follows that in order to increase stability reserve it is useful to increase a free member at the extent of controlling interactions – and from the point of mathematics – to increase a determinant of matrix A in system (182) at the extent of adding to matrix A matrix BK – as reflects formula (224). It is only necessary that an addition of matrix BK in formula (224) by all means has increased a determinant of matrix $A + BK$ in comparison with a determinant of matrix A. The above example 1 has shown that while adding matrix BK a determinant cannot increase but even decrease.

In order that the choice of matrix BK be correct note that "a direct table of signs" for matrix A determinant of system (201)

$$\begin{vmatrix} -7 & -6 \\ -8 & -7 \end{vmatrix} \tag{229}$$

is of the form:

$$\begin{vmatrix} - & + \\ + & - \end{vmatrix} \tag{230}$$

Hence in order to increase a free member it is necessary that in matrix (225) elements C_{11} and C_{22} be positive and elements C_{12} and C_{23} – negative. In order that it will be convenient to compare with the above example 1 let us choose matrix (225) in the form:

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \tag{231}$$

In this case equations in system (200) (if we take in to account a controlling interaction) will become:

$$\left.\begin{array}{l} \dot{x}_1 = (-7-1)x_1 + (-6+1)x_2 = -8x_1 - 5x_2 \\ \dot{x}_2 = (-8+1)x_1 + (-7-1)x_2 = -7x_1 - 8x_2 \end{array}\right\} \tag{232}$$

A characteristic polynomial in system (232) will be equal to:

$$\lambda^2 + (8+8)\lambda + \begin{vmatrix} 8 & 5 \\ 7 & 8 \end{vmatrix} = \lambda^2 + 16\lambda + 29, \tag{233}$$

which means that system (232) is stable.

Now let us compute its stability reserves during variables of matrix A elements of an initial system (200). Suppose that relative variations of all its elements by an absolute value are equal to ε.

If we take into account these variations and control equations of system (200) will become

$$\left.\begin{array}{l} x_1 = [-7(1 \pm \varepsilon) - 1]x_1 + [-6(1 \pm \varepsilon) + 1]x_1 \\ x_2 = [-8(1 \pm \varepsilon) + 1]x_2 + [-7(1 \pm \varepsilon) - 1]x_2 \end{array}\right\} \tag{234}$$

Now in order to calculate stability reserves it is necessary to find the most unfavourable combination of variations signs $\pm\varepsilon$ in system (231) – it corresponds to a table of signs.

$$\begin{vmatrix} - & + \\ + & - \end{vmatrix} \tag{235}$$

at which a free member of a characteristic polynomial in system (234) if there are the most unfavourable signs of variations in its elements becomes:

$$\begin{vmatrix} 8 - 7\varepsilon & 5 + 6\varepsilon \\ 7 + 8\varepsilon & 8 - 7\varepsilon \end{vmatrix} = 29 - 194\varepsilon + \varepsilon^2, \tag{236}$$

From formula (236) it follows that if

$$0 \le \varepsilon \le 87 - \sqrt{87^2 - 29} = 0,1688, \tag{237}$$

a free member will be positive and only when $\varepsilon > 0,1668$ it can become negative and then stability of system (234) can disappear.

In previous section we have already examined stability reserves in system (199) without the control and we have seen that without the control stability is preserved only during coefficients variations satisfy an inequality:

$$0 \le \varepsilon \le 0,005155 \tag{238}$$

Thus a correctly directed controlling interaction of a rather small intensity increased stability reserves by more than 32 times – from $\varepsilon = 0,5155\%$ up to $\varepsilon = 16,68\%$.

By all means it is necessary to check at that ε a sign in the second member of a characteristic polynomial in system (234) can change – at member ε with the first degree λ. But this member is equal to

$$(8 - 7\varepsilon + 8 - 7\varepsilon) = 16 - 14\varepsilon, \tag{239}$$

and it can change a sign only when $\varepsilon \ge 1,14$ that is much more than a variation that can change a sign of a free member of a characteristic polynomial and thus – to change the system stability. In our example – as in all other previous ones the most dangerous source

of stability loss is a change of a sign in a free member during variations of parameters in a control object.

Note that in examined examples we have not taken into account variations of parameters in regulator $U = KX$. Certainly we can take into account even them but parameters variations in a regulator almost always are much less than parameters variations in a control object. And therefore on the first step of investigation we can suppose that they are equal to zero.

Recommendations on control synthesis that increase stability reserve

A given material allows us to give the following recommendations on the synthesis of control that secures the increase of stability reserves and improving the safety of objects work whose parameters can greatly change in the course of exploitation

1. The first step is to form "a direct table of signs" for a determinant of matrix A in a control object whose mathematical model is of the form: $\dot{X} = AX + BU$ and to which we can apply a controlling interaction that a linear regulator has produced with an inverse connection $U = KX$ where K – vector of control coefficients. After this processes in a control system will be described by a system of equations: $\dot{X} = (A + BK)X$;

2. While taking into account the fact that at the extent of picking out coefficients of intensification in regulator $U = KX$ we can realize the most different matrixes BK it is necessary to choose such BK that it will increase elements in matrix A corresponding to signs "plus" of "a direct table of signs" in matrix A determinant and will decrease such elements that will correspond to sings "minus" of this table.

Note that if due to matrix BK realization of such elements matrixes A will be increased that correspond to signs "minus" in "a direct table of signs" in a determinant of matrix A then stability reserves will decrease and a control object can even loose stability (which has been illustrated in example 1 of this section).

Therefore the application of "table of signs" that has been in details examined in the first part of the book plays a big role in control theory as well. These "tables" allow us to correctly dispose with limited resources of controlling interactions in order to increase stability reserves.

Example 3. Let us consider a control object that is described by a system of equations $\dot{X} = AX$ in which a free member of a characteristic polynomial is a determinant of the third degree and after multiplying by $(-1)^3$ is of the form:

$$\begin{vmatrix} 14 & 12 & 3 \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix} = 4, \tag{240}$$

For determinant (237) "an inverse table of signs" is equal to:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ - & + & - \end{vmatrix}. \tag{241}$$

By using a methodics presented in the first part we can calculate that if all elements of a determinant (240) have the same relative variations $\pm\varepsilon$ then during the most unfavourable combination of signs in variations (a combination that corresponds to "inverse table of signs") a determinant (240) can become negative already if $|\varepsilon| \leq 0,0075$. This means that an examined system with a free member of a characteristic polynomial (240) has very small reserves of stability. Already if $\varepsilon \leq 0,75\%$ stability can be lost.

Even if variations undergo not all elements of determinant (240) but only its first line determinant (240) will become:

$$\begin{vmatrix} 14(1+\varepsilon) & 12(1-\varepsilon) & 3(1+\varepsilon) \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix} 4 - 308\varepsilon. \tag{242}$$

To a variation of elements of the first line corresponds "an inverse table of signs" then in this case already when $\varepsilon > 1,3\%$ a determinant can become negative and stability of an examined object will disappear.

We are sure that stability reserves of an examined system are not sufficient (even if variations occur only in elements of the first line) and that it is advisable to increase stability reserves. This can be done if with the help of control – i.e. with the help of matrix BK – we increase such elements of a determinant which help its increase or – to decrease such elements which help its diminishing.

It is quite easy to determine what elements of a determinant increases it and what – decreases. It is sufficient to apply "a direct" or "an inverse table of signs". During the increase of an element of a determinant that is in the place of a sign "plus" in a direct table os signs" or during the decrease of an element that stands in the place of a sign "plus" in an "inverse table of signs" the determinant increases.

Let us use this rule for the increase of stability reserves of a system investigated by us. From "table of signs" (241) it is at once seen that the decrease of element 14 that stands on the first place – in the first line of determinant (240) will increase a value of the determinant and thus – while preserving the same variations of initial elements – will increase stability reserves.

For example, let us choose such intensification coefficients of a regulator with an inverse connection $U = KX$ that it will lead to a matrix BK that is of the form:

$$BK = \begin{pmatrix} -\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{243}$$

In this case a free member of a characteristic polynomial in an examined system be equal to determinant:

$$\begin{vmatrix} 14-\alpha & 12 & 3 \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix} = 4 + 4\alpha, \tag{244}$$

and if we take into account variations of coefficients in the first line of determinant (240) and there is the most unfavourable combination of signs of these variations that correspond to "increase table of signs" in determinant (240) a determinant (244) will become:

$$\begin{vmatrix} 14-\alpha+14\varepsilon & 12-12\varepsilon & 3+3\varepsilon \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix} = 4+4\alpha-308\varepsilon. \qquad (245)$$

In the next table are given values of absolute values of variations ε whose excess can break stability of an examined system in the following table:

Table 1.

$-\alpha$	-1	0	1	10
ε	0	0,013	0,026	0,143

We see that due to the choice it is sufficient to have large negative values a in matrix (243) we can secure any desired value for stability reserve.

At the same time an error in a correct choice of absolute values or signs of matrix BK is rather dangerous. So in an example that we are investigating a choice of matrix (243) in which a is positive and it is equal to $a = +1$ ar once leads an examined system to the boundary of stability. The reserve of stability is equal to zero but in $a \geq +1$ the system looses stability even in the absence of parameters variations in a control object.

A correct choice of signs in elements of matrix BK that secures the increase of stability reserves in control system can be easily accomplished if we use "tables of signs" of a determinant in matrix A in control system $\dot{X} = AX + BU$, $U = KX$.

Note that in a control theory problems of synthesis systems with good stability reserves have been repeatedly considered – see publications [67;68;72;73] and many others.

But as in these years there were no methods that allowed to directly synthesis regulators that were optimal in stability reserves and to correctly choose matrix BK then indirect methods were applied – to estimate "stability reserves by amplitude, by a phase" etc. which surely was not at all convenient. The application of signs tables that has been in details described in the first part of the book allows us in a new and in a much more fruitful way to come up to the problem of synthesising control systems with good stability reserves.

At the same time it is necessary to take into account that an examined method of increasing stability reserves works well only in systems of the form $\dot{X} = AX + BU$ that most often occur. In them the loss of stability during variations of matrix A parameters occurs due to the change of a sign in matrix A determinant. (It is also a free member in a characteristic polynomial of matrix). There are also possible (but they are rare) such matrixes A in which a sign in a determinant of matrix $A + BK$ does not change during examined variations of its elements in matrix A. But all the same stability is lost due to the break of other stability conditions. Therefore it is necessary to additionally check that a characteristic polynomial in matrix $A + BK$ is Hurwitz one.

Tables of algebraic additions

Note that just modernized "tables of signs" can help in the choice of matrix BK that increases the stability reserves in the best way. Recall that in the first part of our book it has been shown that during the change of element a_{ij} in any determinant by value $a \cdot A_{ij}$ where A_{ij} – an algebraic addition of element a_{ij}. Since for the best management of limited stability resources it is important to take into account not only a sign of an algebraic addition but also its value it is useful to apply besides earlier described "tables of signs" a table of algebraic additions of determinants.

So, for a determinant (240) we can easily compute that

$$A_{11} = \begin{vmatrix} 16 & 5 \\ 4 & 1 \end{vmatrix} = -4; A_{12} = -\begin{vmatrix} 12 & 5 \\ 5 & 1 \end{vmatrix} = +13; A_{13} = \begin{vmatrix} 12 & 16 \\ 5 & 4 \end{vmatrix} = -32$$

$$A_{21} = 0; A_{22} = -1; A_{23} = -4; A_{31} = +12; A_{32} = -34; A_{33} = 80$$

and "a table of algebraic additions" will become:

$$\begin{vmatrix} -4 & +13 & -32 \\ 0 & -1 & -4 \\ +12 & -34 & +80 \end{vmatrix}. \tag{246}$$

Hence it follows that if matrix BK is equal to:

$$\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix}, \tag{247}$$

then as value of elements and their quality limited it is necessary first of all to apply element C_{33}. If matrix BK consists of one element and is of the form:

$$BK = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C_{33} \end{pmatrix}, \tag{248}$$

then the growth of determinant (240) from the addition of matrix (248) to an initial matrix

$$\begin{pmatrix} 14 & 12 & 3 \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{pmatrix} \tag{249}$$

will be equal to $80 \cdot C_{33}$ but if

$$BK = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ C_{31} & 0 & 0 \end{pmatrix}, \tag{250}$$

then the growth will be $12 \cdot C_{31}$ – i.e. if values C_{31} and C_{33} are the same the growth of determinant will be by $\frac{12}{80} = \frac{3}{20}$ times less and this means that the growth of stability reserves will be also less.

If matrix BK is such that in it only one of elements of the first line is different from zero then it is better to choose it in the form:

$$BK = \begin{pmatrix} 0 & 0 & C_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{251}$$

Basing ourselves on "a direct table of signs" in determinant (240) we conclude that in order to increase a determinant a sign of an element C_{13} of matrix (251) must be negative. If $C_{13} = -2$ then during the most unfavourable combination of signs in elements variations of the first line of determinant (241) and during such combination of signs that corresponds to its "inverse table of signs" a determinant (240) will become:

$$\begin{vmatrix} 14(1+\varepsilon) & 12(1-\varepsilon) & 3(1+\varepsilon)-\gamma \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix} 4 + 32\gamma - 308\varepsilon. \tag{252}$$

If, for example, $\varepsilon_{13} = 1$ then a determinant (252) is equal to $36 - 308\varepsilon$ and stability reserve in variations of the first line elements will be equal to $\varepsilon = 0,117$, i.e. it will increase by nine times in comparison with $\gamma = 0$. Earlier we have seen that if the only one element in matrix BK has the same absolute value in matrix BK that corresponds to matrix (244) stability reserve will increase only by two times.

Thus "a table of algebraic additions" of a determinant in matrix A of system (222) allows us to manage in the best way by limited control resources in order to increase stability reserves.

But it is necessary to take into account that during the addition to matrix A matrix BK algebraic additions can change and this fact can limit the possibility of increasing a determinant in matrix A (and a corresponding stability reserve) due to the addition to it matrix BK.

Example. For matrix

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \tag{253}$$

with determinant

$$\det A = \begin{vmatrix} 3 & 2 \\ 2 & 3 \end{vmatrix} = 5 \tag{254}$$

a table of algebraic additions is of the form:

$$\begin{vmatrix} +3 & -2 \\ -2 & +3 \end{vmatrix} \tag{255}$$

and it shows that the determinant can be increased, for example, at the expense of adding to matrix A the following matrix

$$BK = \begin{pmatrix} 0 & -\lambda \\ -\lambda & 0 \end{pmatrix} \tag{256}$$

with negative elements on an auxillary diagonal.

Really, if $a = 1$ we shall have

$$\det(A + BK) = \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} = 8,$$

if $a = 2$ we shall have

$$\det(A + BK) = \begin{vmatrix} 3 & 0 \\ 0 & 3 \end{vmatrix} = 9,$$

but if we later increase a, for example, up to $a = 3$ we shall have:

$$\det(A + BK) = \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} = 8$$

i.e. a determinant will decrease. This is due to the fact that when $a = 3$ a table of algebraic additions in matrix $A + BK$ will become:

$$\begin{vmatrix} +3 & +1 \\ +1 & +3 \end{vmatrix}, \tag{257}$$

and it will differ from table (255).

The investigation of dependence signs in algebraic additions of matrix $A + BK$ from BK has shown that signs of algebraic additions A_{12} and A_{23} will change if $a = 2$ but when $a < 2$ they are negative, if $a > 2$ they have become positive.

Hence it follows that we have chosen a regulator $U = KX$ that is such that matrix BK is of the form (256) then it is not necessary to increase absolute values of elements a_{12} and a_{21} in matrix more than up to a value $a = 2$.

At is also necessary to carry out similar checks whether signs of algebraic additions in matrix $A + BK$ in comparison with matrix A and also – for matrixes of such more large dimensions than a simple matrix (253) of dimension 2x2 in an example that we have investigated.

For matrix A of any dimension nxn in control system of the type $\dot{X} = AX + BU$ with a regulator $U = KX$ the application of "tables of algebraic additions" helps us to choose a control $U = KX$ that secures the best increase of stability reserve in control systems.

§36. Rude and robust systems. About terms in mathematics –2.

In the first part of the book we have indicated such publicatiens in which earlier have already investigated systems of linear algebraic equations during variations of their coefficients. In the field of systems that are described by differential equations with a mathematical model that is known not in a complete way it is necessary to note as the first and the most important publication a monograph [74] that was published in 1937 and also – an article [75].

In this monograph beginning from the first pages of its "Introduction" it was stressed that any theoretical investigation of real objects and processes inevitably requized idealization, requized throwing of a series of factors which in some sense influenced "in a small way" on an examined object.

The most simple case is when a mathematic model of an examined object or process in the form of a differential equations or a system of equations sufficiently well reflects a character of processes that occur in an examined object. But even in this most simple case it is necessary to recall that coefficients in equations are determined almost always from experience and because of this they cannot ideally exactly describe an examined object. Besides in the course of time object parameters and thus mathematical model coefficients cannot remain idially exact and unchangable. Small variations of coefficients are inevitable and they must be taken into account the more that even small changes of coefficients can lead to large consequences.

In more complex cases it is necessary to take into account that a chosen mathematical model that has been chosen on the first step of investigation apriori does not exactly determine an examined object. In these cases it is necessary to check in what measure the behavior of a mathematical model changes during small change of a type of differential equations that determine an examined system. Here consequences of "small" changes can be much more striking.

In 1937 in monograph [74] for the first time put a very important question: by what properties a mathematical model must possess in order that it can (at least – can!) correctly reflect the behavior of real object in physics and technics? What mathematical models must be at once thrown off as such that do not reflect the behavior of real objects?

Here is the answer that gave the anthers of monograph [74]. A real physical interest have only "rude" systems, i.e. such systems that essentially change their behavior during that describe the system. Only "rude" systems can serve theoretical models of real systems ([74], p. 33). But in [74] on the same page 33 an important reservation has been made: "these small changes will be supposed such ones which do not change an order of a differential equation (or – just the same – they do not change a number of differential equations of the first order if a system of n equation of the first order is investigated – i.e. a "normal" Cauchy system of differential equations is considered.

As it was indicated in §21 an equation

$$\varepsilon\ddot{x} + \dot{x} + x = 0, \tag{258}$$

in which ε is small a change of equation

$$0 \cdot \ddot{x} + \dot{x} + x = 0 \tag{259}$$

can be considered small.

A coefficients 0 in equation (259) in $\ddot{x}$ has changed by a small value ε. And a system

$$\left.\begin{array}{r} \varepsilon\dot{x}_1 = x_1 + x_2 \\ \dot{x}_2 = x_1 + 2x_2 \end{array}\right\} \tag{260}$$

with small ε can be considered as a small change of system

$$\left.\begin{array}{r} 0 = x_1 + x_2 \\ \dot{x}_2 = x_1 + 2x_2 \end{array}\right\} \tag{261}$$

that is equivalent to one equation:

$$\dot{x}_2 = x_2 \tag{262}$$

But on the example of equations (258)–(262) that are the simplest linear equations with constant coefficients at once it is seen that, for example, solutions of equation (258) even if ε is very small are not at all similar to solutions of same equation if $\varepsilon = 0$ when it turns into equation $\dot{x} + x = 0$.

Equations with small coefficients with higher derivatives are called singular – perturbing equations. Their solutions even if there are small perturbations with small ε essentially differ from solutions of unperturbing equations (or systems). Therefore in this book singular – perturbing equations are not considered.

It is important to stress that during small variations of large coefficients can appear such equations that are like singular – perturbing ones. So, for example, in earlier examined system of equations (27)-(28) while coefficients in Dx_2 change little in equation (28) equal to 1 during the transfer from 1 in $(1 + \varepsilon)$ a system with characteristic polynomial (29) appeared. It is equivalent to a differential equation:

$$[-\varepsilon D^4 + (1 - 4\varepsilon)D^3 + (5 - 5\varepsilon)D^2 + (7 - 2\varepsilon)D + 3]x = 0 \tag{263}$$

that is in fact singular – perturbing in relation to equation

$$(D^3 + 5D^2 + 7D + 3)x = 0 \tag{264}$$

but not to system (27)-(28).

This exterior similarity has led to bewilderments during the discussion of the first publication of the book [5]. Some of participants of discussions thought that answers to all questions that had arisen could be found in a well investigated sphere of singular-perturbing equations. In fact there is no place for misunderstanding and confusions. It is necessary to bear in mind that the subject of examination is not equation (264) but a system of equations (27)-(28) in which there is not "singularities" or "variations of zero" as well. And small changes undergoes only a coefficient in Dx_2 of equation (28), a coefficient that is equal to a unity.

Besides in the theory of singular-perturbing equations surely are investigated only stable solutions of equations – for equation (264) these are solutions corresponding to $\varepsilon \leq 0$. Unstable solutions in the essence do not require examination. In them everything is clear: unstable solutions speedily increase (in an absolute value) and they increase the more quickly the less is a module of a small value ε.

But we are interested just in the possibility of quickly increasing solutions. We investigate at what conditions and for what systems these speedily increasing solutions appear. Such systems are not those that have been thrown away on the step of projecting and "manufactured in metal" lead – and have not once led to wreckages and catastrophes.

The systems that we are investigating can not be called not rude. They are rather called "a stipulation"given in [74], on page 33. But they are (so much) dangerous as "unrude" ones since they must be thrown away already at the step of a projection. But systems that we are considering although they are "not rude" in their time have not become a subject of many investigations devoted to "rude" and "not rude" systems. These important investigations first of all have been developed in the USSR after the publications [74] and [75] and then – they were continued abroad. Investigations from USA instead of terms rudeness "rude systems" prefored to use "robustness" and "robust systems" – from am English word "robust" – i.e. – "strong" . Later even Russian investigators also started applying terms "robustness "robust system". See, for example, a book [68] and in many others. Such a distortion in terms cannot be considered correct. First of all a priority of Russian scientists is forgotten – the authors of publications [74] and [75] who for the first time have found a new field in investigations and, secondly, in the definition of a term "robustness" an important note given in [74] on page 33 has been omitted. The absence of this note does not help to feel the completeness and clearness in the understanding of a field of our phenomena we are examining.

Now let us consider a question of lawfulness in introducing a term "mathematics – 2" which in §1 of the first part of our book has been determined as "such sections of mathematics in which inexactnesses and errors in coefficients and parameters of examined mathematical models were taken into account". Opposing to "mathematics – 1"into which it is useful to unite all such sections in which parameters of mathematical models (or laws of their changes) are known and are unchanged.

Although investigations concerning uncompleteness and errors of mathematical models have been carried out for a long period of time and especially they have developed after publications [74], [75] there were no foundations to depict such investigations into a special section named "mathematics – 2". In mathematics all the time a sphere of investigated objects and processes widened. This is a usual phenomenon. Persons who investigated objects and processes while taking into account errors in their description for long period of time have not understood that the investigation of such objects requires new methods of investigation that are different from those which so successfully were applied during the examination of objects whose parameters (or laws of their change) are known and unchanged.

And only at the end of the 20th century after publications [5], [35], [37] it has been recognized that new investigation objects (whose coefficients are known only with inevitable errors) require new (rather – more precise) investigation methods. It was said

that equivalent transformations that had been so successfully and everywhere applied earlier could lead to mistaken results and that this most important investigation method required specifications and must be grounded in order it was possible to be applied.

Therefore a proposal appeared – to depict such sections of mathematics which differed from other ones not only by a sphere of examined objects but also investigation methods and it was proposed to call them "mathematics – 2" . Surely this proposal must find a wide discussion but later it would be possible that it will be recognized and widely applied.

Literature

1. Diemidovich B.P., Maron I.A. Fundamentals of computing mathematics. The sixth edition, Izdatelstvo "Lan 2007, 664 pp.

2. Kopchienova N.V., Maron I.A. Computing mathematics in examples and problems. The second edition, Izdatelstvo "Lan 2008, 367pp.

3. Shokin Yu.I. Intervals analysis. Novosibirsk, Nauka, 1981, 112p.

4. Kalmikov S.A., Shokin Yu.I., Yuldashiev Z.H. Methods of interval analysis. Novosibirsk, Nauka, 1986, 221p.

5. Petrov Yu.P., Petrov L.Yu. Unexpected phenomena in mathematics and its connection with wreckages and catastrophes., The first edition - 1999; 108p., the second edition - 2000; the third edition - 2002; the fourth edition, with additions and enlarged, 2005; "BHV-Peterburg, 224p. There is a translation into English. See site: www.petrov1930.narod.ru.

6. Petrov Yu.P. How to obtain reliable solutions of equations systems. Izdatelstvo "BHV-Peterburg 2009, 175pp. There is a translation into English. See sight: www.petrov1930.narod.ru.

7. Mahmutov M.M. Lections on numerical methods. Institute Kompiuternih issliedovaniy. Moscow-Izhevsk, 2007, 236p

8. Ilyin V.A., Poznyak E.G. Lineal algebra; M., Nauka, 1978, 302h.

9. Mishkis A.D. Applied mathematics for engineers. Special courses. The third edition, M., Physmat, 2007, 687 pp.

10. Zaliznyak V.E. The basis of scientific computations. Moscow-Izhevsk, 2006, 264p.

11. Petrov Yu.P. The securing of reliability and trustworthiness of computer calculations. Izdatelstvo "BHV-Petersburg 2008. 160p. There is a translation into English. See: site – www.petrov1930.narod.ru

12. V.I.Fiedosiev. Resistance of materials. Courses for high schools. The eighth edition, M. "Nauka 1979, 559 pp.

13. Demmel J.W. Applied numerical linear algebra, M., "Mir 2001, 421pp (Translation into Russian – SIAM, Philodelphia, 1997)

14. Ustinov S.M., Zimnitsky V.A. Computing mathematics, Izdatelstvo "BHV–Peterburg 2008, 100pp.

15. Shariy S.P. Optimal exteriour estimate of a set of solutions of interval equations systems. The first part. Computation technology, vol.7, N6, 2002, pp.90-113

16. Shariy S.P. Optimal exteriour estimate of a set of solutions of interval equations systems. Part II. Computing technology, vol.8, N1, 2003, pp.84–109

17. Hansen E. Global optimization using interval analysis. N.Y. Marcel Dekker, 1992.

18. Moore R.E. Methods and application of interval analysis, SIAM, Philadelphia, 1979.

19. Newmaier A. Interval methods for systems of equations. Cambridge: Cambridge UniPress, 1990

20. Shariy S.P. Interval analysis or methods of Monte Carlo? All Russian conference on intervals analysis. Thesesis of reports, "Interval-06 on the 1-4 of July, 2006, Peterghoff (Russia) pp.140-144.

21. Petrov I.A., Petrov Yu.P. Analysis of different approaches to the estimate of an error in the computing of efforts in construction buildings. Drawbacks in estimating methods by a number of conditions. Vestnik grazdansky inzhenierov 2008, N4(17), 33–38.

22. Petrov Yu.P. How to obtain exact estimates of errors in solutions of linear algebraic equations systems. Vestnik grazhdansky enzhinierov, 2010, N1(22), pp.68–73

23. Yegorov N.V. Ovsyannikov D.A. Mathematical modeling of systems in forming electronic and ion knots. Izdatelstvo St.P.S.U., 1998, 274pp.

24. Vinogradova Ye.M., Yegorov N.V. Krimskaya K.A. Computation of electrostatic field in a system of spherical segments. Zhurnal technichesky physics. 2008, v.78, N8, pp.155–159

25. Vinogradova E.M. Mathematical modelling of electronoptical systems, St.Petersburg, Izdatelstvo St.P.S.U, 2005, 112pp.

26. Ivanova K.F. Estimate of an error in a numerical solution of Poisson equations under the influence of fluctuations of entering parameters. Preprint St.P: St.P.S.U., 2010, 34pp

27. Panov D.Yu. A reference book on numerical integration of partial defferential equations Gostechizdat, 1951.

28. Samarsky A.A. Introduction to a theory of difference scheme. M. Nauka, 1971.

29. Faddieev D.K. Faddieeva V.N. Computation methods in linear algebra. M., Fizmatgiz, 1966.

30. Tiurtishnikov Ye. Ye. Methods of numerical analysis. M. "Academya 2007, 320pp.

31. Verzhbitsky V.N. Fundamentals of numerical methods. M., "Visshaya shkola 2002, 870pp.

32. Zhuk D.M., Manichev V.B. Iitsky A.O. Methods and algorithms of solving differential – algebraic equations for modeling systems and objects in a time space. Informatzionniy technologii 2010, N7 (part 1) and N8 (part 2).

33. Matviyeviev N. M. Ardinaty differential equations. St.-Petersburg, "spetzialnaya literatura" , 1996, 371pp.

34. Arnold V. J. Ordinary differential equations, M., Nauka, 1975, 239 pp.

35. Petrov Yu. P. Synthesis of optimal control systems when perturbing forces are known not completely. Izdatelstvo Leningradskogo Gosudarstvennogo Universiteta, 1989, 289 pp.

36. Petrov Yu. P. Concealed dangers in traditional methods of stability check. Izvestya VUZ, Yelektromekhanika, 1991, №11, p. 106-108

37. Petrov Yu. P. Stability of linear systems during parameters variations. Avtomatica i telemekhanika, 1994, №11, p.186-189

38. Petrov Yu. P. Prevention of wreckages in control systems. Izvestya VUZ. Yelektromekhanika, 1994, №1-2, p. 37-40

39. Petrov Yu. P. The inquest and warning of technogene catastrophes. St.P. Izdatelstvo "BHV-Peterburg" , 2007, 104 pp.

40. Tihonov A. N., Arsienin V. Ya. Solution methods of incorrect problems. M., Nauka, izdaniye tretiye, 1986, 287 pp.

41. Petrov Yu. P. The third class of problems in physics and technique that are intermediate between correct and incorrect problems. N. J. J. Kh., St.Petersburg, 1998, 29 pp.

42. Petrov Yu. P., Sizikov V. S. Correct, incorrect and intermediate problems with applications. St.Pbs.: Izdatielstvo "Polytiekhnica 2003, 261 pp. There is a transformation into English – "Well-posed, ill-posed and intermediate problems will applications izdat. "VSP Boston-Leiden, 2005, 234 pp.

43. Zubov V. J. Lyapunov methods and their application. I., Izdatielstvo Leningradskogo Universiteta, 1957, 241 pp.

44. Barbashin Ye. A. Lyapunov functions. M., Nauka, 1970, 240 pp.

45. Vasiliev A. B., Butussov V. F. Asymptotic decompositions of singular - perturbing equations. M., Nauka, 1973.

46. Stroik D. Ya. A short essay of mathematics history. M., Nauka, 1978, 335 pp. (translation from "Abriss ger geschichte der mathematik". Von Dirk Z. Struik, Berlin, 1963)

47. Petrov Yu. P. History and phylosofy in science. Mathematics, computer technics, informatics, computer technics, informatics. Izdatelstvo "BHV-Peterburg" , 2005, 441 pp.

48. Lietov A. M. Analytical construction of regulators. Automatika i telemekhanika, 1960, №4, №5, 6 and 1961 №4.

49. Lietov A. M. Dynamics of flying and control. M., Nauka, 1969, 359 pp.

50. Lyapunov A. M. A general problem on stability of movement. Fizmat, 1959.

51. Stieklov V. A. Foundations of integration theory of ordinary differential equations. M. - L. Giz, 1927.

52. Sharovatov V. T., Petrov Yu. P. "On mistakes in a package MATLAB"and Chertkov K. G., Petrov Yu. P. "Mistakes found in a package MATLAB. Trudi vtoroi Vsierossiski konfierientzii"Projecting of scientific and engineering applications in MATLAB. Institut problem upravleniya Rossiiski Akademii Nauk, 2004, p. 318-323 and 324-327.

53. Fihtengoltz G. M. A course of differential and integral computation. Volume 2, Ogiz, GITTL. M.-L., 1948, 860 pp.

54. Lakatos J. Proofs and refutations. M. Nauka, 1967, 151 p.

55. Petrov Yu. P. On "grammar" in science. St.Psb., OOP, St.Petersburg University, 2003, 40 pp.

56. Academician Danielevitch Ya. B., Petrov Yu. P. On the necessary of widening the conception of equivalence in mathematical models. Dokladi Akademii Nauk, 200, vol. 371, №4, p. 473-475.

57. Sharovatov V. T. How to secure stability of quality index in automatic systems. L. Yenergoatomizdat, 1987, 176 pp.

58. Larin V. B., Naumenko K. I., Syuntzev V. N. Synthesis of optimal linear systems with inverse connection. Kiev, Naukova dumka, 1973, 150 pp.

59. Vorotnikov V. I. Stability of dynamical systems in relation to a part of variables. M., Nauka, 1991, 287

60. Zhubov V. I. Mathematical methods of investigating systems of automatic regulation. L., Mashinostroyenie, 1974, 335 pp.

61. Petrov Yu. P. Variation methods in the theory of optimal control. Izdaniye vtoroye. L., izdatelstvo "Yenergiya" 1977, 280 pp.

62. Abdulaev N. D., Petrov Yu. P. Theory and methods of projecting optimal regulators. L., Yenergoatomizdat, 1985, 240 pp

63. Khoaritonov V. L. On asymptotic stability in equilibrium position of a family of linear differential equations. Differential equations. 1978, №11, p.2086-2088.

64. Petrov Yu. P. Optimization of control systems undergoing wind interaction and rough sea. L., Sudostroyeniye, 1973, 216 pp.

65. Pervozhanski A. A. A course of automatic control theory. M., Nauka, 1986, 615 pp.

66. Mierkin D. R. Introduction to the theory of movement stability. M., Nauka, 1971.

67. Polyak B. T., Stcherbakov P. S. Robust stability and control. M., Nauka, 2002, 302 pp.

68. Nikiforov V. O., Ushakov A. V. Control in the conditions of indefinity: sensitivity, adaptation, robustness, St. Petersburg, St.PsbITMO(TU), 2002, 232 pp.

69. Zhubov N. V. Mathematical methods of investigation of dynamical safety. M., Vichislitielni tzentr after A. A. Dorodnitzin. – RAN, 2007, 105 pp.

70. G. Korn and T. Kown. A reference book on mathematics. M., Nauka, 1973, 831 pp.

71. Voevodin V. V., Kuznetzov Yu. A. Matrixes and computation. M., Nauka, 1984

72. Alieksandrov A. G. synthesis of regulators of manydimensional systems. M., Mashinostroeniye, 1986, 272 pp.

73. Afanasiev V. N., Kolmanovski V. B., Nessov V. R. Mathematical theory of constructing control systems. M., Visshaya shkola, 1985, 447 pp.

74. Andronov A. A., Vitt A. A., Haikin S. E. Oscillation theory. 1937 (republished in 1959 and in 1981), M., Nauka, 568 pp.

75. Andronov A. A., Pontriyagin L. S. Rude systems. Dokladi Akademiyi Nauk SSSR, 1937, №14, №5, p. 247-250.

Contents.