MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
National University “Zaporizhzhia Polytechnic”

OPTIMIZATION OF TECHNICAL SOLUTIONS IN MECHANICAL ENGINEERING AND NUMERICAL MATHEMATICAL METHODS

SUMMARY OF LECTURES

for students of specialties 133 Industrial engineering and 131 Applied mechanics of all forms of education

2023

Optimization of technical solutions in mechanical engineering and numerical mathematical methods. Summary of lectures for students of specialties: 133 Industrial engineering, 131 Applied mechanics of all forms of education / Comp.: Mykhaylo FROLOV. – Zaporizhzhia: National University “Zaporizhzhia Polytechnic”, 2023. – 88 p.

Конспект лекцій з дисципліни «Оптимізація технічних рішень в машинобудуванні та чисельні математичні методи» для студентів спеціальностей: 133 Галузеве машинобудування, 131 Прикладна механіка всіх форм навчання (англ.) / Укл.: Михайло Фролов. – Запоріжжя: НУ «Запорізька політехніка», 2023. – 88 с.

Compiler: Mykhaylo FROLOV, associate professor, Ph.D.

Reviewer: Volodimir TSYGANOV, associate professor, Doctor of Technical Sciences

Responsible for the release: Valery ZAGORODNY

Approved at a sitting of the Department of Metal Cutting Machines and Tools. Protocol No. 12 dated 16/06/23

Recommended to the publication by Scientific Methodical Commission of the Mechanical Engineering Faculty. Protocol No. 2 dated 14/09/23

CONTENT

1. STATEMENT OF THE OPTIMIZATION PROBLEM

In mechanical engineering, as in any field of technology, as well as in science in general, economics, and other spheres of activity, there are often met problems with multiple solutions. If these solutions are compared, it is done on the basis of a certain criterion and the one that best meets this criterion is chosen. Such a solution is called optimal, the task of determining it on a given set of solutions is called the optimization problem, and the criterion is the criterion of optimality.

The need for optimization arises when performing the most complex procedures in designing and operating technical objects. At the same time, this set of solutions can consist of both the characteristics of the object and the conditions of its operation.

The solution of optimization problems with the help of computer technology has some characteristic features:

• **Statement of the problem** - involves a well-formalized description of a set of solutions and the appointment of a relevant criterion. When setting the optimization problem, the following prerequisites are necessary:

1. Availability of the object of optimization and the goal of optimization. At the same time, the formulation of each optimization problem should require the extreme value of only one value, that is, the system should not be assigned two or more optimization criteria at the same time since the extremum of one criterion does not always correspond to the extremum of another. A typical example of an incorrect formulation of the optimization problem: "Get maximum performance at minimum cost." The mistake lies in the fact that the task of finding the optimum of two quantities that contradict each other in their essence is set. The correct statement of the task can be:

a) to obtain maximum productivity at a given cost.

b) to obtain the minimum cost at a given productivity

In the first case, the optimization criterion is productivity, and in the second case, the cost price. In general, the two criteria should have their own priorities, which will be discussed below.

2. The availability of optimization resources, which means the possibility of choosing the values of some parameters of the object being

optimized. The object must have certain degrees of freedom - controlling influences.

3. The possibility of a quantitative assessment of the values being optimized, since only in this case it is possible to compare the effects of choosing certain control actions.

4. Consideration of constraints.

• **Mathematical description** - consists in defining the type of optimization problem, based on its formulation, as well as studying its specific features and building a mathematical model that connects the quantitative characteristics of the solution components with the criterion of optimality. On this basis, the method of its numerical solution is selected, studied, and possibly modified

• **Compilation of the algorithm** - consists of the development of a general block diagram with further detailing of the optimization program, developed based on a programming language, and provides the presence of an interface for displaying output data and optimization results, as well as their storage [2].

Thus, the formulation of the optimization problem can be formulated as follows:

Given is a set of solutions or areas of their possible values and a criterion by which these solutions can be compared

It is necessary to find out the best (optimal) solution amid a given set of solutions using a given criterion.

QUESTIONS FOR SELF-TESTING

1. What solution is called optimal?
2. What can a set of solutions consist of when optimizing a technical object?
3. What features of optimization exist when using computer technology?
4. What does each optimization problem require?
5. What prerequisites are necessary for setting the optimization problem?
6. What is the mathematical description of the optimization problem?
7. What is the composition of the optimization algorithm?

2. GENERAL INFORMATION ABOUT OPTIMIZATION MODELS, THEIR STRUCTURE AND DESIGN CONCEPTS

Mathematical modeling based on the principle of optimization implies the possibility and necessity of purposeful regulation of the operating conditions of an object or system to find those in which the system will function in the best way - optimal conditions. The combination of these conditions will represent the optimal solution. Hence, we have the constituent elements of such models, which include [1]:

The criterion of optimality is an accepted indicator of the efficiency of the system under investigation. The maximum or minimum value of the criterion of optimality determines the conditions under which the system works in the best way, that is, optimal conditions. It should be emphasized that usually, the criteria for optimality are not physical values, such as cutting force, temperature, tool life, etc., but indicators in one way or another related to money: cost of machining, profit, quality, production time, etc., which reflects the measure of system efficiency. As an example of an optimality criterion, let's take the cost of machining C, which depends on the cutting conditions:

$$C = f(t, S, V) \tag{2.1}$$

Independent variables (factors, controlling factors) are variables, the optimal values of which are to be found. In this case, these will be cutting conditions t, S, and V, because it is with their help the operating conditions of the system can be purposefully changed. The set of found optimal values will be that very optimal solution mentioned in Section 1.

Parameters are constant values of the model characterizing the constant conditions in which the model was obtained. They are reflected in the values of the constant coefficients of the model.

The goal function is a function connecting the optimality criterion, factors, and parameters. The very name of the function shows that it has a goal. In our case, this is the minimum cost, i.e.

$$C = f(t, S, V) \rightarrow min \tag{2.2}$$

Please note that we are primarily interested in the conditions under which the cost will be minimal, and not the value of the cost as such. This means that in order to find the best operating conditions of the system, the goal function can be simplified to not include cost elements that do not depend on cutting conditions in this case, or on factors if we speak in general.

Constraints are areas of possible values of factors within which the search for optimal solutions takes place. When solving real production problems, limitations always exist and they can be related to the technical capabilities of the equipment (limitations on cutting speed, machine power); the quality of the machined surface (feed limitation); conditions of labor safety, etc. Constraints can be:

- ✓ *Hard constraints* - exceeding the limits of which is unacceptable under any conditions. These may be limitations related to physical constants, safety, technical characteristics of the equipment, etc.
- ✓ *Non-hard constraints* - allow them to go beyond their limits, but it is necessary to understand that the quality of the functioning of the system will significantly deteriorate in this case.

As mentioned above, the goal of optimization modeling is to find the conditions in which the system functions in the best way according to the accepted criterion of optimality. The process of obtaining the goal function cannot be formalized - it is the result of a process analysis that uses previous experience and common sense. But to effectively search for optimal conditions, the optimization model in the form of a goal function must comply with some basic principles.

1. **The principle of unambiguity**. It consists of the fact that only one goal function should be optimized. If, according to the conditions of the problem, there should be several such functions, then they should still be reduced to one, for example, using linear convolution:

$$F = \sum_{i=1}^{k} a_i f_i(X_1, X_2, \dots, X_n) \qquad (2.3)$$

Here a_i is a scaling or weighting factor that reflects the degree of importance of each of the functions included in the convolution and must meet the following conditions:

$$a_i \geq 0 \tag{2.4}$$

$$\sum_{i=1}^{k} a_i = 1$$

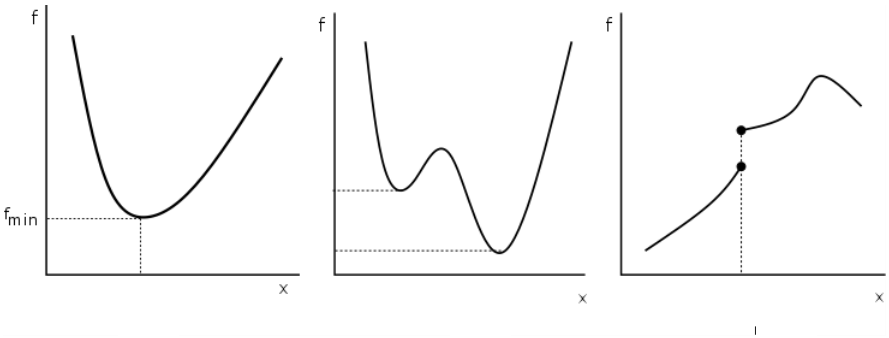
There is a great deal of subjectivity in the selection of weighting factors, but with the help of some approaches, they can be given an objective character: expert evaluations, desirability function, etc.

The convolution according to equation (2.3) can be used when all functions f_i have the same physical nature, that is, dimension. And if the dimensions are different, then instead of the functions in their natural form, their normalized - dimensionless version $\widetilde{f_i}$ should be used. For example, the following method of natural normalization can be used for normalization within [0;1]:

$$\widetilde{f_i} = \frac{f_i - f_{i\,max}}{f_{i\,max} - f_{i\,min}} \tag{2.5}$$

2. **The principle of conformity**. It is desirable that the goal function has one extremum - such a function will be unimodal (the condition of unimodality will be shown below). Functions with gaps and ambiguities should not be used as goals. The presence of local extrema complicates the search for an optimal solution. On Fig. 2.3 function a) has a suitable form, but it is not desirable to use functions b) and c) as targets.

Goal functions that do not have extrema at all, such as linear ones, stand out in particular. At the same time, they must be supplemented with constraints. Otherwise, the optimization problem does not make sense in this case.



a - is a function with one extremum; b – a function with a local extremum;
c - is a function with a gap
Figure 2.1 — Some types of functions

3. **The principle of controllability**. The function must be expressed in terms of variables that can be influenced and controlled.

QUESTIONS FOR SELF-TESTING

1. What is modeling based on the principle of optimization?
2. What elements are present in the optimization model?
3. What are factors?
4. Define the term "parameters".
5. What is an optimality criterion?
6. What is a goal function?
7. According to what principles should the goal function be built?
8. Which function is called unimodal?
9. What types of constraints are there?
10. What is the principle of unambiguity?
11. What is the principle of conformity?
12. What is the principle of controllability?
13. What are "hard constraints"
14. What are "non-hard constraints"?

3. GENERAL INFORMATION ON OPTIMALITY CRITERIA AND GOAL FUNCTIONS

Any physical characteristic of a technical object or process can be optimized - temperature, force, tool life, durability, etc. But non-physical characteristics that are most often optimized - those that are one way or another related to money: cost, productivity, profit, economic efficiency, quality characteristics, etc. In any case, the optimized factor must be evaluated in some quantitative way.

On the basis of the selected criterion of optimality, a goal function is created, which represents the dependence of the criterion of optimality on the factors affecting its value. The type of optimality criterion or goal function is determined by a specific optimization problem. Thus, the optimization problem is reduced to finding the extremum of the goal function.

As mentioned above, the most general formulation of the optimization problem is the expression of the criterion of optimality in the form of an economic assessment (productivity, cost of production, profit, profitability). However, in individual optimization problems, when the object is part of a technological process, it is not always possible or not always advisable to single out a direct economic indicator that would fully characterize the efficiency of the object under consideration. In such cases, the criterion of optimality can serve as a technological characteristic that indirectly evaluates the economic efficiency of the unit (contact time, product yield, degree of conversion, temperature). For example, the optimal temperature profile, cycle duration, etc. are set.

Let us consider in more detail the requirements that must be met for the criterion of optimality.

1. The optimality criterion must be expressed quantitatively.

2. The criterion of optimality must be the only one.

3. The optimality criterion should reflect the most significant aspects of the process.

4. It is desirable that the criterion of optimality has a clear meaning and is easily calculated.

Any object to be optimized can be schematically represented in accordance with Fig. 3.1.

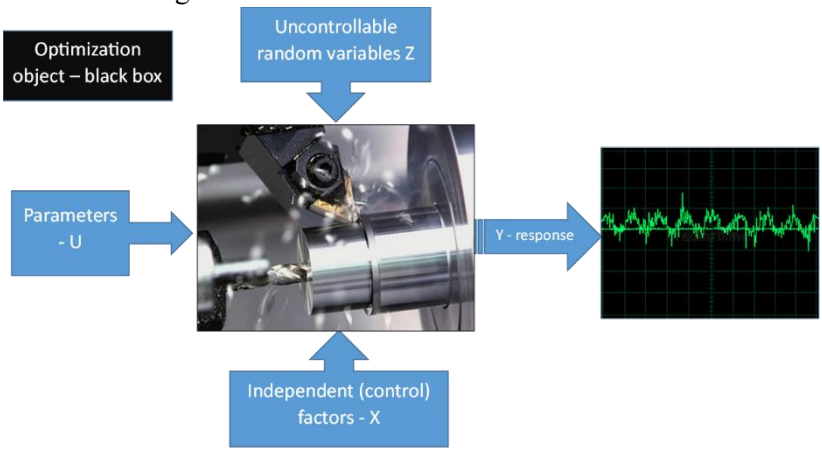


Figure 3.1 – General scheme of the optimization object

When setting specific optimization problems, it is desirable to write down the optimality criterion in the form of an analytical expression.

In the case when the influence of uncontrollable random variables is small and can be ignored, the optimality criterion can be presented as a function of response, controlling factors and parameters:

$$R = R(X_1, X_2, \ldots, X_N, Y_1, Y_2, \ldots, Y_N, U_1, U_2, \ldots, U_N \quad (3.1)$$

Since $\bar{Y} = F(\bar{X})$, then with fixed U we can write:

$$R = R(\bar{X}) \quad (3.2)$$

If the random effects are large enough and must be taken into account, then experimental and statistical methods should be used, which will allow obtaining a model of the object in the form of a function that is valid only for the studied local area.

In principle, the object itself can be used for optimization instead of a mathematical model, but experimental optimization has a number of significant disadvantages:

a) a real object is required;

b) it is necessary to change the technological regime within significant limits, which is not always possible;

c) duration of tests and complexity of data processing. The presence of a mathematical model (provided that it sufficiently reliably describes the process) makes it much easier to solve optimization problems by analytical or numerical methods.

In optimization problems, *simple* and *complex* optimization criteria are distinguished.

The criterion of optimality is called *simple*, and the optimization itself is *unconstrained* if it is necessary to determine the extremum of the goal function without setting conditions for any other values. Such criteria are usually used when solving individual optimization problems (for example, determining the maximum concentration of the target product, the optimal time of the reaction mixture in the apparatus, etc.).

The criterion of optimality is called *complex*, and the optimization itself is *constrained*, if it is necessary to establish the extremum of the goal function under certain conditions that are imposed on a number of other values (for example, determining the maximum productivity at a given cost, determining the optimal temperature under restrictions on the thermal stability of the catalyst, etc.).

The procedure for solving the optimization problem necessarily includes, in addition to the selection of controlling factors, also the establishment of constraints on them (temperature resistance, explosiveness, machine power, etc.). Constraints can be imposed both for technological and economic reasons.

Therefore, to solve the optimization problem, it is necessary:

a) make a mathematical model of the optimization object

$$\overline{Y} = F(\overline{U}, X)$$

b) choose the criterion of optimality and compose the goal function

$$R = R(\overline{Y}, X, \overline{U})$$

c) establish possible constrains that should be imposed on variables;

d) choose an optimization method that will allow you to find the values of the factors that ensure the extremum of the goal function.

It is accepted to distinguish between *static* optimization problems for processes occurring in stable modes and *dynamic* optimization problems. In the first case, the issues of creating and implementing an optimal process model are solved, in the second - the task of creating and implementing an optimal control process under unspecified operating modes.

QUESTIONS FOR SELF-TESTING

1. What is the task of optimizing the goal function?
2. What characteristics are usually accepted as optimality criteria?
3. Describe the requirements for optimality criteria.
4. Is it possible to carry out optimization experimentally, and if so, what are the advantages and disadvantages of this process?
5. Which optimality criterion is called simple?
6. Which criterion of optimality is called complex?
7. What is unconstrained optimization?
8. What is constrained optimization?
9. What is the difference between static and dynamic optimization?

4. SOME TYPICAL OPTIMALITY CRITERIA AND GOAL FUNCTIONS IN MACHINING

Most often, when determining rational cutting conditions during machining, the cost of processing is taken as a criterion of optimality. At the same time, the share of the cost price when cutting with one tool is determined by the formula [3]:

$$C = \left(t_0 + \frac{t_c}{Q}\right)E + \frac{E_i}{Q} \rightarrow min \quad (4.1)$$

Where: t_0 is the main technological (machine time); tc – time for restoration of the cutting properties of the tool and sub-adjustment of the machine; E is the cost of one minute of operation of the machine and the operator; Ei - costs associated with the operation of the tool during its life-time period; Q is the number of parts processed during the tool.

In turn, the costs associated with the operation of the tool during the period of tool life can be determined:

For a durable tooling:

$$E_i = t_c E_н + C_z + \frac{(C_t - C_w)k_z}{i+1} \quad (4.2)$$

Where: $E_н$ is the nominal salary of the adjuster; C_z – sharpening cost; C_t – cost of the tool; Cw - cost of tool waste; i – the number of resharpening of the tool before its complete wear; k_z is a coefficient that takes into account the accidental loss of the tool.

For a tool with a mechanical attachment of replaceable polyhedral plates, taking into account the long service life of the holders, the costs of which are more likely to be capital:

$$E_i = t_c E_н + \frac{\left(C_{\Sigma p} - C_{\sum w}\right)k_z}{n_p} \quad (4.3)$$

Where: $C_{\Sigma p}$– the total cost of plates in the tool; $C_{\sum w}$ – the total cost of plate waste; n_p - the number of working faces of the plate.

For calculations according to formulas (4.1), (4.2), (4.3) and in general when designing a manufacturing process, the cutting tool life is of great importance. Depending on the specific conditions and problems of production, different tool life periods can be used: T_{max} – maximum tool life

period; T_{opt} - optimal tool life, which corresponds to the optimal cutting speed; Tek is the economic period of tool life, which corresponds to the lowest processing cost; Tpr is the tool life period corresponding to the maximum productivity of the process. In practice, the tool life periods Tek and Tpr are most often used [3]:

$$T_{\text{ек}} = \frac{1-m}{m}\left(t_c + \frac{E_i}{E}\right) \tag{4.4}$$

$$T_{\text{пр}} = \frac{1-m}{m} t_c \tag{4.5}$$

Where m is a relative tool life factor.

In the future, using the known dependence for cutting speed calculation, it is possible to optimize the cutting conditions:

$$V = \frac{C_v}{t^{x_v} S^{y_v} T^m} K_v \tag{4.6}$$

Where: t is the cutting depth; S - feed; T – tool life; V – cutting speed; Cv, Kv – coefficients; x_v, y_v are degree indicators.

QUESTIONS FOR SELF-TESTING

1. What indicator is most often accepted as a criterion for optimality in mechanical processing?
2. Name the existing varieties of tool life periods.
3. What tool life periods are more often used in practice?

5. ONE-DIMENSIONAL UNCONSTRAINED OPTIMIZATION

5.1 Terms

Optimization of a function of one variable under the condition of no constraints, or one-dimensional unconditional optimization, aka one-parameter optimization, is the simplest type of optimization problem. Nevertheless, it occupies an important place in the theory of optimization. This is due to the fact that single-parameter optimization problems are quite common in engineering practice and, in addition, are used in the implementation of more complex interactive multi-parameter optimization procedures.

Let the function *f (x)* be continuous on the interval [a, b]. Consider the problem of minimizing this function on this segment (the task of finding the smallest value on the segment).

Note that if the function f (x) has a minimum at the point x, then the function -f (x) has a maximum at x*. Therefore, the same methods are used to find the maximum as for finding the minimum. Next, we will, as a rule, talk only about the task of finding the minimum.*

For those who studied the course of mathematical analysis, this task may seem elementary. So, if the function *f (x)* is piecewise differentiable on the segment [a, b], then to find its minimum, it is enough to find its derivative and solve the equation of the form:

$$f'(x) = 0 \tag{5.1}$$

The value of x that corresponds to (5.1) determines the stationary (critical) point, that is, the point at which the gradient vector of the function is zero:

$$\nabla f(x) = 0 \tag{5.2}$$

We will assume that there is only one stationary point on the segment [a, b], which can be a point of both a local and a global minimum, in contrast to other non-stationary points of the function, which can be taken as a minimum, based on the conditions of the problem (for example, taking into account the constraints which will be discussed later).

Note that at the stationary point of the minimum, the following condition must also be fulfilled:

$$f''(x) > 0 \tag{5.3}$$

However, as practice shows, for a wide class of functions, the analytical solution of the optimization problem is not so simple. On the one hand, even if the function *f (x)* is continuous and differentiable on [a, b] and the calculation of its derivative $f'(x)$ is not difficult, the solution of equation (5.1) for finding critical points may turns out to be no less difficult than initial minimization problem. On the other hand, in practical problems it is often unknown whether the function *f(x)* is differentiable, moreover, even if we know that it is differentiable, we may not know the analytical representation of its derivative $f'(x)$ (for example, if the function is not given in analytical form). In this regard, minimization methods that do not require the calculation of the derivative are of significant importance. Therefore, we come to the use of numerical optimization methods, which involve the use of mainly calculation algorithms. Such algorithms involve repeated application of some mathematical operations, which is called *iteration*, and the whole process is called *iterative*.

The application of any of the numerical methods begins with the determination of the localization interval or segment - the segment in which the minimum of the function is located, within which the uncertainty interval [a, b] is located. At the same time, a trio of numbers on the localization segment is determined: x_1, x_2, x_4. Such a trio of numbers is called *best-case* if the requirement is met $f(x_2) < \min [f(x_1), f(x_4)]$ see Fig. 5.1. Here we write the condition of unimodality of the function:

$$f(x_2) < f(x_1), \;\; f(x_3) < f(x_4) \tag{5.4}$$

During the iterative process, the calculated solution gets closer and closer to the true minimum, and in this context, we are talking about the accuracy of the solution ε, which can be defined as the length of the uncertainty interval at iteration *k,* as the absolute difference of the found values of *x* or goal function value in two consecutive iterations *k* and *k-1* (See formula (5.5). The higher the accuracy, the more iterations should be done.

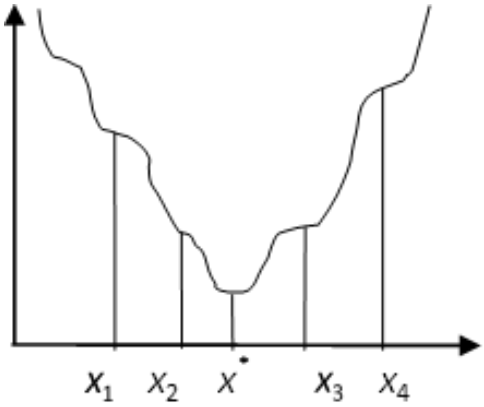


Figure 5.1 – Best-case numbers trio

$$|b_k - a_k| \le \varepsilon$$
$$or \ |x_k^* - x_{k-1}^*| \le \varepsilon \qquad (5.5)$$

A large number of one-dimensional optimization methods have been developed, and we will consider two groups of such methods:

- methods of narrowing the uncertainty interval;
- methods of derivatives.

5.2 Methods of narrowing the uncertainty interval

Let it be necessary to find the minimum of the function *f (x)* on some interval [a, b]. The task of approximate finding of the minimum by the methods of narrowing the uncertainty interval is to find a set of abscissas x_1, x_2,..., x_k in which the function is calculated, such that the minimum value of *f** lies at some *i* in the interval $x_{i-1} < x^* < x_i$. Such an interval is called the uncertainty interval D. It is obvious that initially the uncertainty interval D coincides with the segment [a, b].

There are several ways to systematically narrow the uncertainty interval. Let's consider three of them.

5.2.1 General search

Let it be necessary to find the minimum of the function *f (x)* on some interval [a, b]. If no additional information is known about the function *f (x)* on this interval, then to find the minimum on [a, b], you can apply the simplest sorting method, or, otherwise, a general search.

In this method, the interval [a, b] is divided into several equal parts *N*, followed by the calculation of the function values in *n = N+1* nodes of the resulting grid. The abscissa with the minimum calculated value of the function is taken as the minimum (Fig. 5.2).

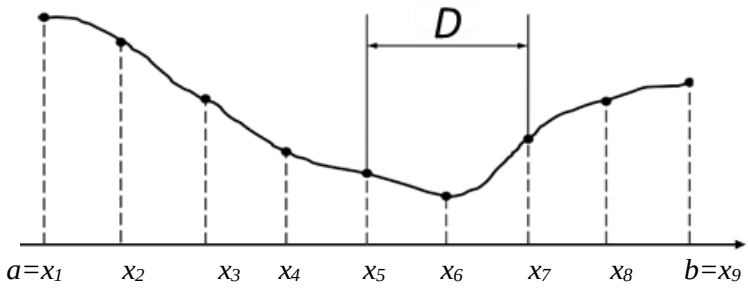


Figure 5.2 – General search method

As a result, the uncertainty interval is narrowed to two grid steps D. The narrowed interval is again divided into *N* equal parts. The procedure is repeated until the required accuracy ε is obtained, which is defined as the difference of the found values of *x* in two consecutive iterations.

Let's denote a_i, b_i– limits of the uncertainty interval for iteration *i.* Accordingly, for the first iteration (i = 1): $a_1 = a, b_1 = b$. Following iterative process is performed cyclically until the specified accuracy is obtained:

- The length of the uncertainty interval:

$$D_i = b_i - a_i \tag{5.6}$$

- Division of uncertainty interval in iteration i:

$$\delta_i = \frac{D_i}{N} \tag{5.7}$$

- Determination of the value of the factor x_{ij}:

$$x_{ij} = a_i + (j-1) \cdot \ \delta_i, \qquad j = \overline{1, N+1} \tag{5.8}$$

- Calculation of the values of the function $f(x_{ij})$ and selection of the minimal value:

$$x_i^* = x_{ij} \ for \ which \ f(x_i^*) = \min\big(f(x_{ij})\big), \ \ j = \overline{1, N+1} \tag{5.9}$$

- The length of the uncertainty interval after the first iteration:

$$D_{i+1} = 2\delta_i \tag{5.10}$$

- Checking the accuracy and repeating the cycle if the condition (5.11) or (5.5) is not fulfilled:

$$|x_i^* - x_{i-1}^*| \le \varepsilon, \quad i = 2, 3, \ldots \tag{5.11}$$

- The limits of the new interval relatively point x_i^*:

$$a_{i+1} = x_i^* - \delta_i, \quad b_{i+1} = x_i^* + \delta_i \tag{5.12}$$

5.2.2 Bisection (dichotomy) method

Let's assign the uncertainty interval [a, b] (see Fig. 5.3)
At the first stage (for the first iteration), we denote:

$$x_{1(1)} = a, \qquad x_{2(1)} = b \tag{5.13}$$

Here, the index in parentheses is the iteration number.

Let's define the middle of the uncertainty interval:

$$x_{3(i)} = \frac{x_{1(i)} + x_{2(i)}}{2} \tag{5.14}$$

Let's make sure that $x_{1(1)}$, $x_{2(1)}$, $x_{3(1)}$ is a best-chance trio of numbers and determine a small value $\delta \le \varepsilon$, where ε is the specified accuracy of the calculation.

Further, the calculation takes place in the following sequence:

- Calculate the values of the function:

$$f\left(x_{3(i)}-\frac{\delta}{2}\right),\qquad f\left(x_{3(i)}\right),\qquad f\left(x_{3(i)}+\frac{\delta}{2}\right)$$

- If $f\left(x_{3(i)}+\frac{\delta}{2}\right) < f\left(x_{3(i)}-\frac{\delta}{2}\right)$ then define the new limits of the interval as:

$$x_{1(i+1)}=x_{3(i)}-\frac{\delta}{2},\quad x_{2(i+1)}=x_{2(i)} \tag{5.15}$$

Such a case is shown in Fig. 5.3

- If $f\left(x_{3(i)}-\frac{\delta}{2}\right) < f\left(x_{3(i)}+\frac{\delta}{2}\right)$ then define the new limits of the interval as:

$$x_{1(i+1)}=x_{1(i)},\quad x_{2(i+1)}=x_{3(i)}+\frac{\delta}{2} \tag{5.16}$$

The new interval D will be almost 2 times smaller than the previous one. After that, the middle of the interval is determined according to equation (5.14), and the actions are repeated for the new interval. Iterations are performed until the specified accuracy is obtained, which is defined as the absolute difference between the middle of uncertainty intervals in two consecutive iterations:

$$\left|x_{3(i)}-x_{3(i-1)}\right|\le\varepsilon \tag{5.17}$$

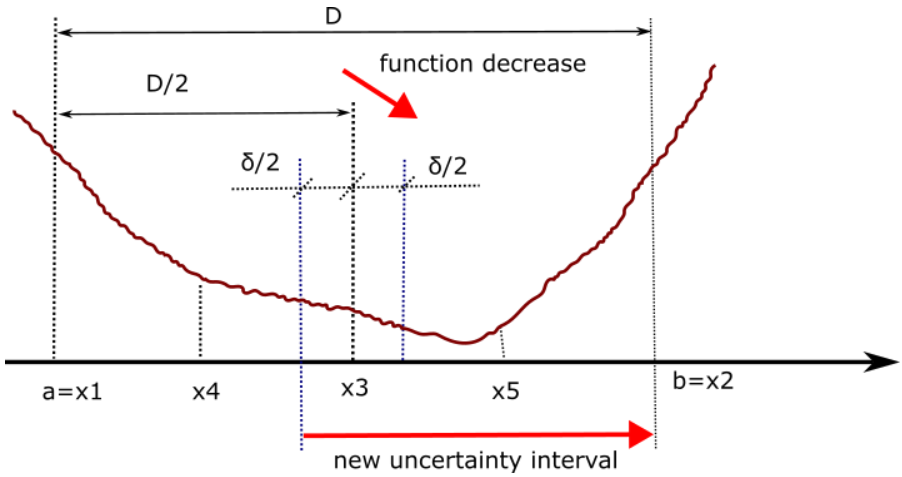


Figure 5.3 - An illustration of the bisection method

5.2.3 Golden ratio method

Dividing the interval into unequal parts allows to find an even more effective method. Let's calculate the function at the ends of the segment [a,b] and put $a = x_1, b = x_2$, and also calculate the function at the two interior points x_3, x_4. We compare all four values of the function and choose the smallest among them (Fig. 5.4). For example, let *f (x3)* turn out to be the smallest. Obviously, the minimum is to be in one of the segments adjacent to it. Therefore, the segment [x_4, b] can be discarded and the segment [a, x_4] left.

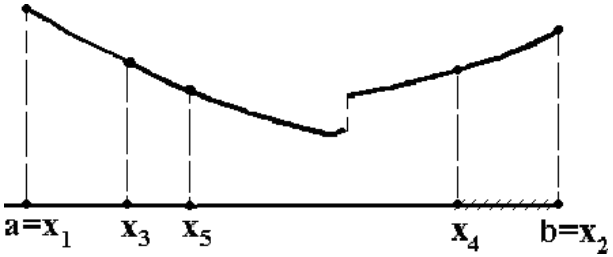


Figure 5.4 - An illustration of the golden ratio method

The first step has been taken. On the segment [*a, x4*], it is necessary to select two internal points again, calculating the value of the function in them and at the ends, and take the next step. But in the previous step of the calculations, we already found the function at the ends of the new segment [*a, x4*] and at one of its interior points *x3*. Therefore, it is enough to choose another point *x5* inside [*a, x4*], determine the value of the function in it and make the necessary comparisons. These reduce to one-quarter the number of calculations in one step of the process. What is the best way to place points? Each time the segment that remains is divided into three parts and then one of the outer segments is discarded.

Let us denote the initial uncertainty interval by D (Fig. 5.5). Since in the general case any of the segments [x_1x_3] or [x_4x_2] can be discarded, we will choose the points x_3 and x_4 so that the lengths of these segments to be the same: $x_3 - x_1 = x_2 - x_4$. After discarding, a new uncertainty interval of length D' will be obtained.

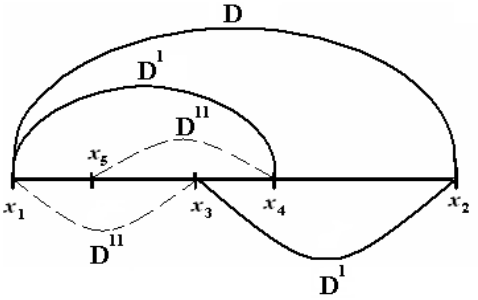


Figure 5.5 - Division of segments by the method of the golden section

Let us denote $D/D' = \varphi$. Next, we will continue in the same way - for this, we will divide the interval D^1 in the same way as the interval D, i.e.:

$$D'/D'' = D/D' = \varphi \qquad (5.18)$$

Where $D''$is the next uncertainty interval. With this division, the entire interval refers to the larger part, as the larger part to the smaller and **φ = 1.618** - which is called the "golden ratio". This division of segments is known as the "golden ratio". Therefore, this method of search is called the "golden

ratio". The ratio φ shows how many times the initial size of the uncertainty interval is reduced at each iteration. With the number of iterations k→∞, the length of the uncertainty interval will tend to 0 as a geometric progression with the denominator 1/φ, that is, the method of the golden section always converges. Obviously, this method is more efficient than the method of bisection

Having marked the coordinates of the calculation points $x_{1(i)}, x_{2(i)}, x_{3(i)}, x_{4(i)}$ - where the index in brackets is the iteration number, and the uncertainty interval in iteration i - Di, we write the following calculation algorithm: (see Fig. 5.5):

- $x_{1(1)} = a,\ x_{2(1)} = b$
- $D_1 = x_{2(1)} - x_{1(1)}$
- $D_{i+1} = D_i / 1.618$
- $x_{3(i)} = x_{2(i)} - D_{i+1}$
- $x_{4(i)} = x_{1(i)} + D_{i+1}$
- Calculate function value *f(x)* at points $x_{3(i)}, x_{4(i)}$
- If $f(x_{3(i)}) < f(x_{4(i)})$ – dismiss section $x_4 x_2$

  $x_{1(i+1)} = x_{1(i)}, \quad x_{2(i+1)} = x_{4(i)}, \quad x^*_{(i)} = x_{3(i)}$
- If $f(x_{4(i)}) < f(x_{3(i)})$ – dismiss section $x_1 x_3$

  $x_{1(i+1)} = x_{3(i)}, \quad x_{2(i+1)} = x_{2(i)}, \quad x^*_{(i)} = x_{4(i)}$
- Continue iterations until obtaining the specified accuracy, which will be evaluated as:

$$\left|x^*_{(i)} - x^*_{(i-1)}\right| \le \varepsilon \qquad (5.19)$$

- If condition (5.19) is fulfilled, then the following is accepted as a solution:

  $x^* = x_{3(i)}\ if\ f(x_{3(i)}) < f(x_{4(i)})$

$$x^* = x_{4(i)} \; if \; f\left(x_{4(i)}\right) < f\left(x_{3(i)}\right)$$

5.2.4 Setting the localization segment (initial uncertainty interval)

The methods of narrowing the uncertainty interval discussed above include two stages:

• the stage of setting the interval limits;

• the stage of reducing the interval.

We have already considered ways to reduce the interval. Let's now consider the stage of setting the interval limits. Usually a heuristic method is used, for example, Svenn's method.

So, let it be necessary to find the minimum of the function *f (x)* not on the segment, but on the entire x axis. Assume again that the function *f(x)* is unimodal. Let's choose some initial approximation x_0, and make a step of some length h - $x_1 = x_0 + h$ from it (Fig. 5.6):

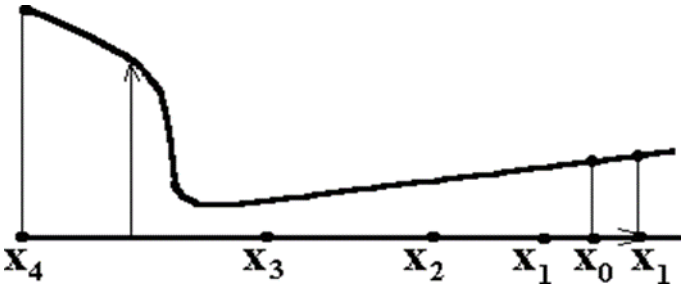


Figure 5.6 – Illustration to Sven's method

If $f(x_1)$ turns out to be greater than $f(x_0)$, then we change the direction of the step and define $x_1 = x_0 - h$. Now let $f(x_1) < f(x_0)$. Let's double the step $h' = 2h$ and define $x_2 = x_1 - h'$, etc., until the condition $f(x_n) > f(x_{n-1})$ is fulfilled at a certain stage.

Now it is clear that the minimum of the unimodal function lies on the segment $[x_4, x_3]$ and it can be found by one of the considered methods.

The main advantage of search methods is that they are based on the calculation of only the values of the function and, therefore, do not require

the fulfillment of the differentiability condition and the recording of the goal function in analytical form. The last property is especially valuable in simulation modeling. However, this advantage of search methods is also their disadvantage - the speed of their convergence is low.

QUESTIONS FOR SELF-TESTING

1. What optimization is called one-dimensional?
2. What is a stationary point?
3. What is one-dimensional unconditional optimization?
4. Why does the chapter consider function minimization?
5. Why is it necessary to use numerical methods?
6. What is a localization segment?
7. What three numbers are called best-chance?
8. What is iteration?
9. How is the accuracy of the numerical method determined?
10. What groups of methods are used in one-dimensional optimization.
11. What interval is called the interval of uncertainty?
12. What is the general minimum search method?
13. What is the method of bisection?
14. What is the golden ratio method?
15. What is the "golden ratio"?
16. Name the method and algorithm for establishing the initial uncertainty interval.
17. How is the condition of unimodality of a function written?

5.3. Methods using derivatives: the Newton-Raphson method

Let the function *f(x)* be differentiated twice. As is known from mathematical analysis, the condition for the minimum of a function is equality:

$$f'(x^*) = 0 \tag{5.20}$$

This is a necessary condition. However, in order for the point x* to be a minimum, the sufficient condition must also be fulfilled:

$$f''(x^*) > 0 \tag{5.21}$$

So, let's numerically solve the equation:

$$f'(x) = 0 \tag{5.22}$$

For this, some initial approximation x_k is given and the function itself is decomposed into a Taylor series limited by terms up to and including the second order, that is, a quadratic model of the function is built:

$$\hat{f}(x) \approx f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f'(x_k)(x - x_k)^2 \tag{5.23}$$

The solution of this equation with respect to *x* gives the following *k+1* approximation to the minimum:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \tag{5.24}$$

The algorithm expressed by equation (5.24) with some further clarifications is called "Newton's method" or "Newton-Raphson's method". Algorithm (5.24) has two disadvantages:

First, equation (5.20) can determine not only the minimum, but also the maximum.

Secondly, the model function (5.23) may differ greatly from the optimized function *f (x)* and the step $x_{k+1} - x_k$ may be too large (Fig. 5.7):

Therefore, the strategy (5.24) should be clarified. To be sure that we are moving to the minimum, we will check the ratio $f(x_{k+1}) < f(x_k)$ at each step. If it is executed, then we proceed to the next step, etc. If $f(x_{k+1}) > f(x_k)$, and $f'(x_k)(x_{k+1} - x_k) < 0$, then the function *f(x)* must

first decrease in the direction from x_k to x_{k+1}, so the next acceptable point can be found, by splitting a step in the same direction, for example by setting

$$x'_{k+1} = \frac{x_k + x_{k+1}}{2} \tag{5.25}$$

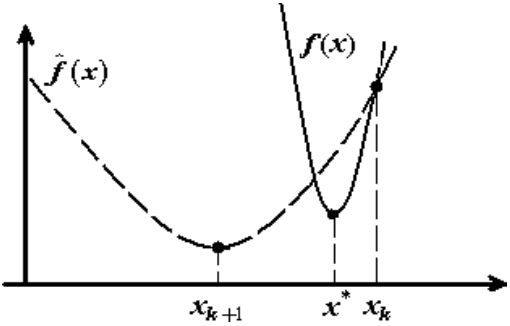


Figure 5.7 – Illustration to the Newton-Raphson's method

It can be seen from formula (5.24) that the expression $f'(x_k)(x_{k+1} - x_k)$ has a negative value if and only if $f''(x_k) > 0$. This means that if the local model used to obtain the Newtonian step has a minimum and not a maximum, then the existence of the corresponding step direction is guaranteed. On the other hand, if $f''(x_k) < 0$ and $f'(x_k)(x_{k+1} - x_k) > 0$, then on shifting from x_k to x_{k+1}, $f(x)$ increases, therefore, a step must be taken in the opposite direction:

$$x'_{k+1} = \frac{x_k + x_{k-1}}{2} \tag{5.26}$$

The criterion for terminating iterations during optimization can be selected as:

$$\left|\frac{f(x_k)}{f'(x_k)}\right| < \varepsilon \tag{5.27}$$

where ε is the predetermined accuracy.

In some problems, the derivative functions of *f(x)* are not available and Newton's method can be modified. Let's choose an initial approximation x_k and a small step h. Then:

$$x_{k+1} = x_k - h\frac{f(x_k + h) - f(x_k - h)}{f(x_k + h) - 2f(x_k) + f(x_k - h)} \qquad (5.28)$$

Formula (5.28) gives the main step of the algorithm, which is called the quasi-Newton method or the modified Newton method. All considerations regarding the step $x_{k+1} - x_k$ given for the Newton-Raphson method remain valid.

QUESTIONS FOR SELF-TESTING

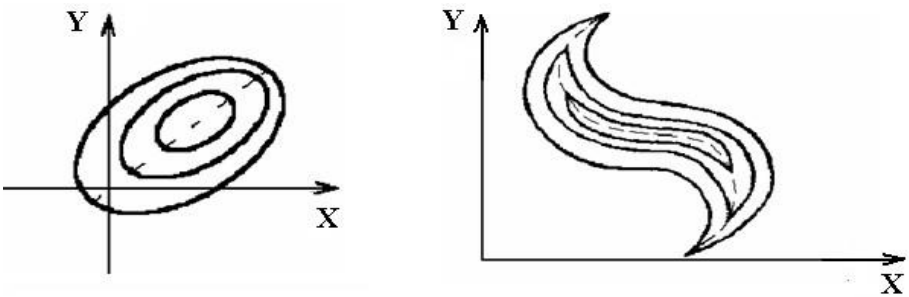
1. How is a quadratic function model constructed?
2. What is the algorithm of the Newton-Raphson method?
3. Name the shortcomings of Newton's method and how they are corrected.
4. What check should be done at each step of Newton's method?
5. Under what conditions should the step be split and in which direction?
6. How is the criterion for terminating iterations according to Newton's method written?
7. In what cases is Newton's modified method used?
8. How is the value of the argument determined in iterations according to the modified Newton method?

6. MINIMUM OF FUNCTIONS OF MANY VARIABLES

6.1. Relief of the function

It is convenient to consider the concept of "relief of a function" using the example of a function of two variables $z = F(x, y)$. This function describes some surface in three-dimensional space with coordinates x, y, z. The task F (x, y)→min means finding the lowest point of this surface.

As in topography, let's depict the relief of this surface with level lines. Let's draw equidistant planes $z = const$ and find the lines of their intersection with the surface $F(x, y)$. Projections of these lines on the x, y plane are called level lines or equal response lines. The direction of the decrease of the function will be indicated by dashes next to the level lines. According to the type of level lines, three types of relief can be conventionally distinguished: *valley, gully*, and *disordered* (Fig. 6.1a, b; - 6.5).



a b

a – valley; b – gully

Figure 6.1 – Relief of functions

In valley relief, the level lines look like ellipses (Fig. 6.1a).

Let's consider the gully type of relief. If the level lines are piecewise-smooth (Fig. 6.1b), then we select a breaking point on each of them. The geometric location of the breaking points will be called a *real gully* if the angle is directed in the direction of increasing function, and a *true or real ridge* if it is in the direction of decreasing.

Most often, the level lines are smooth everywhere, but there are areas with a large curvature on them. The geometric locations of the points with

the greatest curvature are called *solvable gullies or ridges* - depicted by a dotted line in Fig. 6.2

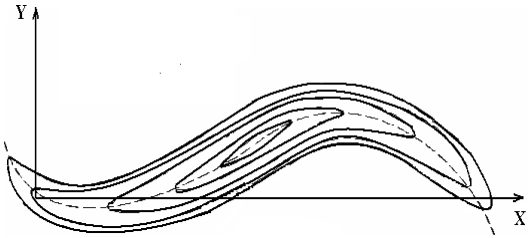


Figure 6.2 – Solvable gully

For example, the relief of the function $F(x,y) = 10[y - sin(x)]^2 + 0.1x^2$ (Fig. 6.3) has a pronounced tortuous gully, the "bottom" of which is a sinusoid, and the lowest point - origin of coordinates.

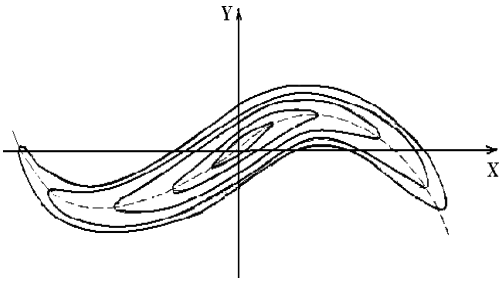


Figure 6.3 – Relief of function $F(x,y) = 10[y - sin(x)]^2 + 0.1x^2$

The disordered type of relief is characterized by the presence of many maxima and minima. For example, the function
$F(x,y) = (1 + sin^2x)(1 + sin^2y)$ - Fig. 6.4, has minima at the points $x_k = \pi k, y_k = \pi l$ and maxima at points shifted relative to the minima by $\pi/2$ along each coordinate.

All effective methods of finding a minimum are reduced to the construction of trajectories along which the function decreases; different methods differ in the ways of constructing such trajectories. A method adapted to one type of relief may not perform well on another one.

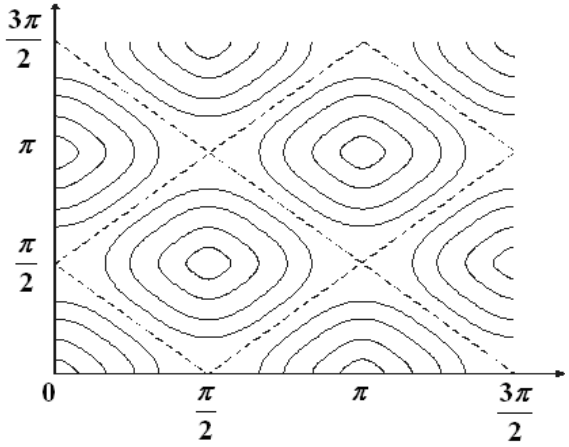


Figure 6.4 – Disordered relief

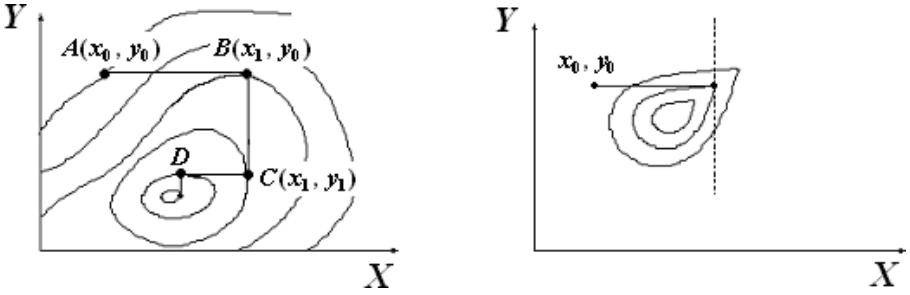
6.2. Method of coordinate descent - Gaussian method

Let's explain this method using the example of a function of two variables F(x, y). Let's choose the initial approximation x_0, y_0. We fix the value of one of the coordinates $y = y_0$. Then the function will depend only on one variable - x; denote it by $f_1(x) = F(x, y_0)$. Using any method of one-dimensional optimization, let's find the minimum of the function $f_1(x)$ and denote it by x_1. We took a step from the point (x_0, y_0) to the goal (x_1, y_0) in the direction parallel to the x axis; at this step the value of the function has decreased.

Now, from the new point, let's make a decent in the direction parallel to the y axis, that is, consider the function $f_2(y) = F(x_1, y)$, find its minimum and denote it by y_1. The second step brings us to the point (x_1, y_1), which completes the cycle of descents or the first iteration.

Next, we will repeat the cycles. At each descent, the function does not increase, and the value of the function is bounded from below by its minimum value $F^* = F(x^*, y^*)$. Therefore, the iterations converge to some limit $F \geq F^*$. Whether will be equality here, that is, will the descents converge to a minimum and how quickly? It depends on the function and the choice of

initial approximation. Using the example of a function of two variables, it is easy to make sure that there are cases of convergence of the descent along the coordinates to a minimum and cases when this descent does not converge to a minimum. Indeed, consider the geometric interpretation of this method (Fig. 6.5):



a b

a – valley relief; b – real gully relief

Figure 6.5 – Illustration to Gaussian method

We will move in the chosen direction, that is, along some straight line in the *x, y* plane. In those areas where the straight line crosses the level lines, we move from one level line to another, so that the function changes (increases or decreases, depending on the direction of movement). Only at the point where the given line is tangent to the level line (Fig. 6.5a), the function has an extremum along this direction. Having found such a point, we finish the descent in the first direction and must start the descent in the second.

Let the level lines form a real gully. Then a case is possible (Fig. 6.5b) when a descent along one coordinate leads us to the "bottom of the gully", and any movement along such a coordinate (dashed line) leads us to an ascent. No further descent along the coordinates is possible, although the minimum has not yet been reached. In this case, the coordinate descent process does not converge to a minimum.

On the contrary, if the function is sufficiently smooth, then in some vicinity of the minimum, the process of descent along the coordinates converges to this minimum. However, the rate of convergence strongly

depends on the shape of the level lines. So, if the topography of the function has the type of «solvable gully", then when the descent trajectory hits such a gully, the convergence becomes so slow that it is practically impossible to carry out the calculation.

Usually, the method of ordinate descent is used as the first attempt when finding the minimum.

QUESTIONS FOR SELF-TESTING

1. What are the level lines formed?
2. What types of relief features are there?
3. How is the relief of the function depicted on the plane?
4. What is a real gully?
5. What is a real ridge?
6. What is a solvable gully or ridge?
7. What characterizes the disordered type of relief?
8. What is the Gaussian method?
9. Is it always possible to find the minimum using the Gaussian method?
10. For what types of relief is the Gaussian method most suitable?
11. What is the Gaussian method most often used for?

7. CONSTRAINED OPTIMIZATION

7.1. Terms

A number of engineering problems are related to optimization in the presence of a certain number of constraints on controlled variables. Such constraints significantly reduce the size of the area in which the optimum is sought. At first glance, it may seem that reducing the dimensions of the admissible region should simplify the procedure for finding the optimum. However, on the contrary, the optimization process becomes more complicated, because in the presence of constraint, it is not even possible to use the optimality conditions we applied above. At the same time, even the basic condition, according to which the optimum must be reached at a stationary point characterized by a zero gradient, may be violated. In this way, we distinguish between an *unconditional minimum*, as a minimum without taking into account constraints at a stationary point, and a *conditional minimum* - a minimum with constraints taken into account, even if it is not located at a stationary point.

For example, the unconditional minimum of the function $f(x) = (x-2)^2$ takes place at the stationary point $x = 2$. But if the minimization task is solved taking into account the constraint $x > 4$, then the conditional minimum corresponding to the point $x = 4$ will be found This point is not a stationary point of the function $f(x)$, since $f'(x)=4$. Therefore, it is necessary to study the necessary and sufficient conditions for the optimum in tasks with constraints, which are otherwise called *conditional optimization problems*.

In general, the problem of conditional optimization can be written:

$$f(x) \to min, x \in R^n \tag{7.1}$$

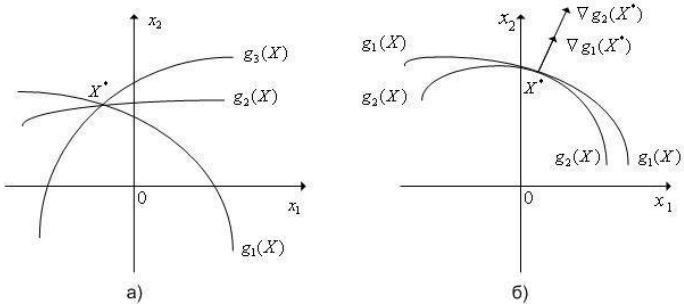
$$h_j(x) = 0, j = 1, 2, \dots, K \tag{7.2}$$

$$g_j(x) \le 0, j = 1, 2, \dots, J \tag{7.3}$$

At the point of the local minimum x^* of problem (7.1), each of the constraints (7.2), (7.3) is satisfied either in the form of the equality $h_j(x) = 0$, $g_j(x^*) = 0$, or in the form of the inequality $g_j(x^*) < 0$.

Constraints in the form of equality are called *active constraints*. Other constraints are called *inactive constraints*.

If the point $x^* \in D$ and the constraints are active, then the condition of linear independence of the gradients of the functions g$g_j(x^*)$ (або $h_j(x) = 0$) of the active constraints at the point x^* is called *the condition of regularity* of the limiting functions at the point x^*. This condition means that, for example, with the number of independent factors $n = 2$, the number of limiting functions passing through the point x^* should not exceed 2 and at the point x^* vectors $\nabla g_1(x)$, $\nabla g_2(x)$ (або $\nabla h_1(x)$, $\nabla h_2(x)$) should not be collinear. For example, in Fig. 7.1 in situation (a) the number of limiting functions passing through the point x^* exceeds the dimension of the vector of varied factors, in situation (b) at the point x^* gradients $\nabla g_1(x)$, $\nabla g_2(x)$ of the limiting functions are collinear.



a – exceeding the number of limiting functions; b - collinearity of vectors

Figure 7.1 - Situations in which the condition of regularity of limiting functions is not fulfilled

QUESTIONS FOR SELF-TESTING

1. What is conditional optimization?
2. What is a stationary point?
3. What is the absolute optimum?
4. What is a conditional optimum?
5. What constraints are called active?

6. What constraints are called inactive?
7. What does the condition of regularity of the limiting functions at the point x* mean?
8. In what situations is the condition of regularity of limiting functions not fulfilled?

7.2. Problems with constraints in the form of equalities. Lagrange multipliers

With the help of the method of Lagrange multipliers, in essence, the necessary conditions are established that allow the identification of optimum points in optimization problems with constraints-equalities. In this case, the problem with constraints will turn into an equivalent problem of unconditional optimization, in which some unknown parameters, called Lagrange multipliers, appear.

Consider a problem with one equality constraint:

$$f(x) \to min, x \in R^n \tag{7.4}$$

$$h_1(x) = 0 \tag{7.5}$$

According to the method of Lagrange multipliers, this problem turns into the following problem of unconditional minimization:

$$L(x.\lambda) = f(x) - \lambda h_1(x) \to min, x \in R^n \tag{7.6}$$

The function $L(x.\lambda)$ is called the Lagrange function. Here, λ is the Lagrange multiplier. Let for a given value of $\lambda = \lambda^0$, the unconditional minimum of the function $L(x,\lambda)$ with respect to the variable x is reached at the point $x = x_0$ and x_0 satisfies the equation (7.5). Then, as it is not difficult to see, x_0 minimizes (7.4) taking into account (7.5), because for all values of x that satisfy (7.5), $minL(x, \lambda) = f(x)$.

Of course, it is necessary to choose the value $\lambda = \lambda^0$ in such a way that the coordinate of the point of the unconditional minimum x_0 satisfies equality (7.5). This can be done if, considering λ as a variable, find the unconditional minimum of the Lagrange function (7.6) in the form of a function of λ, and then choose the value of λ at which equality (7.5) is fulfilled.

Example.

Solve the problem

$f(x_1, x_2) = x_1^2 + x_2^2$

If there is a constraint:

$h_1(x_1, x_2) = 2x_1 + x_2 - 2 = 0$

Let's construct the Lagrange function:

$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 - \lambda(2x_1 + x_2 - 2)$

Let's define its absolute minimum (stationary point) by equating its partial derivatives to zero

$$\frac{\partial L}{\partial x_1} = 2x_1 - 2\lambda = 0 \Longrightarrow x_1^0 = \lambda$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - \lambda = 0 \Longrightarrow x_2^0 = \frac{\lambda}{2}$$

In order to check whether the stationary point with coordinates (x_1^0, x_2^0) *corresponds to the global minimum, we calculate the Hessian matrix of the Lagrange function - this matrix, consisting of second partial derivatives, must be positive definite in order to have a global minimum in the indicated points. In our case, the Hessian matrix will have the form:*

$$\begin{pmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

A matrix is positive definite when all its angle minors are positive. In our case – the 2x2 matrix has one corner minor equal to 2 – i.e., the matrix is positively defined:

$$\begin{pmatrix} 2 & \cancel{0} \\ \cancel{0} & \cancel{2} \end{pmatrix}$$

The optimal value of λ will be found by substituting x_1^0, x_2^0 into the equation for $h_1(x_1, x_2)$, from which we get λ=4/5. Thus, the constrained minimum is reached at: $x_1^0 = \frac{4}{5}, x_2^0 = \frac{2}{5}$ and is equal to $f(x_1, x_2) = \frac{4}{5}$. This method can be extended to the case where there are several constraints in the form of equalities. It is not considered in this course.

QUESTIONS FOR SELF-TESTING

1. How are constraints in the form of equalities taken into account?
2. How is the Lagrange function written?
3. Name the condition of the global minimum of the Lagrange function?
4. Which matrix is positive definite?

7.3. Problems with constraints in the form of inequalities.

Consider a problem with constraints-inequalities

$$f(x) \to min, x \in R^n \tag{7.7}$$

$$g_j(x) \leq 0, j = 1, 2, \dots, J \tag{7.8}$$

The process of optimizing function (7.7) in the presence of conditions (7.8) may turn out to be not very different from those considered in sections 5 and 6. That is, it makes sense to start with an analytical solution of the problem of unconditional optimization of function (7.7) and find its stationary points, after which to check the found solution for compliance with the conditions (7.8) or, if there are several solutions, select those that meet them. You can formalize this process or find solutions that are not stationary points and meet the conditions (7.8) using the Lagrange method discussed above. But let's first consider a few general provisions.

The Kuhn-Tucker theorem is of great importance in the theory and computational practice for the problem of conditional optimization with constraints of the type of inequalities. Without considering the theorem itself, we note that if the function (7.7) under constraints (7.8) has a local minimum at the point x, which meets the requirement of the regularity of the limiting functions, then instead of the conditional optimization problem, the problem of unconditional optimization of the Lagrange function (7.6) can be solved – See—fig. 7.2.*

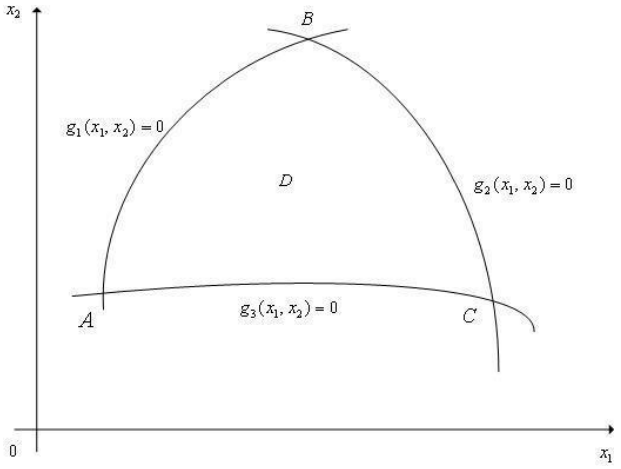


Figure 7.2 - To Kuhn-Tucker theorem

The Lagrange method can be extended to the case of constraints of the type (7.8), transforming them into constraints in the form of equalities by adding a weakening variable u_j^2 to each of them:

$$g_j(x) + u_j^2 = 0 \qquad (7.9)$$

Thus, the problem is reduced to the minimization of the function in the presence of constraints in the form of equalities, and the Lagrange function will have the form:

$$L(x,u,\lambda) = f(x) - \sum_{j=1}^{J} \lambda_j \left[g_j(x) + u_j^2 \right] \qquad (7.10)$$

The necessary conditions that must be fulfilled at a stationary point (if it is a minimum point) are the following:

$$\begin{cases} \dfrac{\partial L}{\partial x} = \dfrac{\partial f}{\partial x} - \sum \lambda_j \dfrac{\partial g_j}{\partial x} \\ \dfrac{\partial L}{\partial \lambda_j} = g_j(x) + u_j^2 \\ \lambda_j \geq 0 \\ \dfrac{\partial L}{\partial u_j} = 2\lambda_j u_j = 0 \end{cases} \qquad (7.11)$$

We transform the last equation taking into account (7.9) and multiply it by $u_j/2$. We will get:

$$\lambda_j u_j^2 = 0, \quad -\lambda_j g_j(x) = 0 \text{ або } \lambda_j g_j(x) = 0 \qquad (7.12)$$

The obtained equation (7.12) is quite fundamental, it means:

- If $\lambda j \neq 0$, then $g_j(x) = 0$ and the constraint is active
- If the constraint is inactive, i.e. $g_j(x) < 0$, then the corresponding Lagrange multiplier $\lambda j = 0$

Example:

Find the conditional minimum of the following problem:

$f(x_1, x_2) = x_1^2 + x_2^2 \rightarrow min$

$g(x_1, x_2) = 2x_1 + x_2 - 2 \leq 0$

Let's add the Lagrange function:

$L = x_1^2 + x_2^2 - \lambda(2x_1 + x_2 - 2 + u^2)$

Conditions for the existence of a conditional minimum:

$$\begin{cases} (x_1 - \lambda) = 0 \\ 2x_2 - \lambda = 0 \\ \lambda(2x_1 + x_2 - 2) = 0 \\ \lambda \geq 0 \end{cases}$$

If λ=0, then $x_1 = 0$ та $x_2 = 0$which satisfies the condition $2x_1 + x_2 - 2 \leq 0$. The value of the function at the point is *f(0,0)=0*

If $\lambda \neq 0$ then $2x_1 + x_2 - 2 = 0$ and then we solve the system of 3 equations:

$$\begin{cases} (x_1 + \lambda) = 0 \\ 2x_2 + \lambda = 0 \\ \lambda(2x_1 + x_2 - 2) = 0 \\ \lambda \geq 0 \end{cases}$$

From which $x_1 = \frac{4}{5}, x_2 = \frac{2}{5}, \lambda = \frac{4}{5}$ but the indicated point is not stationary and a local minimum. The value of the function at the point is *f(x)=4/5*.

QUESTIONS FOR SELF-TESTING

1. What actions make sense to start solving the problem of conditional optimization with constraints in the form of inequalities?
2. How are constraints in the form of inequalities transformed into constraints of equalities?
3. What is the essence of the Kuhn-Tucker theorem?
4. What conditions for the Lagrange function must be fulfilled at a stationary point?

7.4. Penalty functions methods. Non-hard constraints

Consider the conditional optimization problem:

$$f(x) \to min, x \in R^n \quad (7.13)$$
$$h_j(x) = 0, j = 1, 2, \dots, K \quad (7.14)$$
$$g_j(x) \leq 0, j = 1, 2, \dots, J \quad (7.15)$$

$$x_i^{(l)} < x_i < x_i^{(u)} \tag{7.16}$$

Such a problem is also called a nonlinear programming problem. It is said that the point x corresponds to an admissible solution to the problem of nonlinear programming, if all constraints are fulfilled for it, that is, relations (7.13-7.16). It is assumed that for the vector x^*, which is the solution of the problem of nonlinear programming, some initial approximation x_0 is known, possibly inadmissible. In penalty function methods, a sequence of points x_m, $m = 0,1, ..., M$ is constructed, which starts from a given point x_0 and ends at the point x_M, which gives the best approximation to x^* among all the points of the constructed sequence. As x_m, the decision points of the auxiliary problem of unconditional minimization, obtained by transforming the original goal function using the so-called *penalty functions*, are taken. In these methods, the original conditional optimization problem is transformed into a sequence of unconditional optimization problems.

Penalty function methods are classified according to the methods of accounting for constraints-inequalities. Depending on whether the elements of the sequence x_m are valid or invalid points, we call the *internal* and *external* point methods, respectively. if the sequence x_m contains points of both types, the method is called *mixed*.

Thus, to solve the conditional problem, an auxiliary function of the following form is constructed:

$$Q(x, r, l) = f(x) + \sum_{j=1}^{M} r_j G_j\left[g_j(x)\right] + \sum_{j=1}^{K} l_j H_j\left[h_j(x)\right] \tag{7.17}$$

Or

$$Q(x, r, l) = f(x) + r\sum_{j=1}^{M} G_j\left[g_j(x)\right] + l\sum_{j=1}^{K} H_j\left[h_j(x)\right] \tag{7.18}$$

and solve for it a sequence of unconditional optimization problems:

$$Q(x, r, l) \to min, x \in R^n \tag{7.19}$$

In the external point method, which is also called the method of external penalty functions, the functions *G* and *H* are chosen in such a way that they become nonzero (positive) when the corresponding constraint is violated (see Fig. 7.3). And since we minimize (7.13), the movement towards the violation becomes unprofitable. Inside the admissible region, the functions H and G must be zero.

For inequality constraints:

$$G_j\left(g_j(x)\right) > 0 \text{ при } g_j(x) > 0 \qquad (7.20)$$

$$G_j\left(g_j(x)\right) = 0 \text{ при } g_j(x) \le 0$$

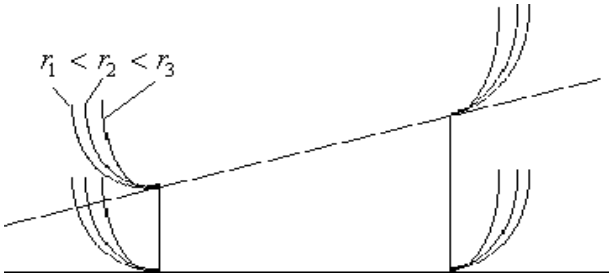


Figure 7.3 - Behavior of penalty functions according to the external point method

One of the forms of the function (7.20) is the quadratic penalty function:

$$G_j\left(g_j(x)\right) = \left\{\frac{1}{2}\left[g_j(x) + \left|g_j(x)\right|\right]\right\}^2 = \frac{1}{4}\left[g_j(x) + \left|g_j(x)\right|\right]^2 \qquad (7.21)$$

In general, the degree exponent in equation (7.21) can be any even.

For restrictions - equalities:

$$H_j\left(h_j(x)\right) \to 0 \text{ при } h_j(x) \to 0 \qquad (7.22)$$

The requirements (7.22) are met by the following penalty functions:

$$H_j\left(h_j(x)\right) = \left|h_j(x)\right| \qquad (7.23)$$

$$H_j\left(h_j(x)\right) = \left(h_j(x)\right)^{\alpha} \tag{7.24}$$

Where α is any even number. For example, α = 2

Constraints-inequalities of the conditional optimization problem to which the external point method can be applied can be interpreted as non-hard, that is, such that can be overcome. But it should be understood that the quality of the process can significantly deteriorate. In this case, the factors are usually limited by the upper - b_j^U and lower - b_j^L limits, and the constraints themselves take the form (7.25) and (7.26), respectively:

$$g_j(x) = x - b_j^U - \varepsilon \leq 0 \tag{7.25}$$

$$g_j(x) = b_j^L - x - \varepsilon \leq 0 \tag{7.26}$$

Where ε is the optimization accuracy

In the method of internal point or barrier functions, the functions *H, G* are chosen to be non-zero in the admissible region and such that when approaching the boundary of the admissible region (from the inside), they grow, preventing exit when searching for the boundary of the region (Fig. 7.4). In this case, these functions must be small (positive or negative) inside the admissible region and large positive near the boundary (inside the region) and are called *barriers*. For example, for inequality constraints:

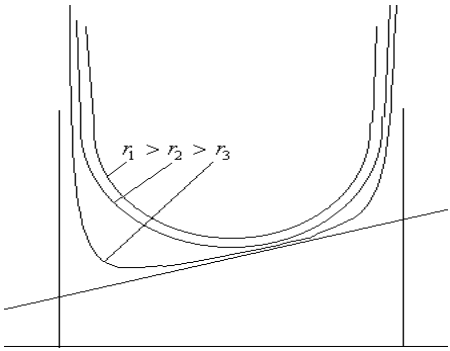


Figure 7.4 - Behavior of penalty functions according to the method of barrier functions

$$G_j\left(g_j(x)\right) \to \infty \text{ при } g_j(x) \to 0^- \tag{7.27}$$

Functions of the bellow form can serve as barrier functions for constraints-inequalities:

$$G_j\left(g_j(x)\right) = \frac{1}{-g_j(x)} \tag{7.28}$$

$$G_j\left(g_j(x)\right) = -\ln\left(-g_j(x)\right) \tag{7.29}$$

The logarithmic penalty (7.29) is a barrier function that is not defined at inadmissible points (that is, for such x that $g(x) > 0$). Therefore, in those cases when it is necessary to deal with inadmissible points (for example, when the given initial approximation x_0 is not admissible), a special procedure is needed that ensures getting into the admissible region.

The penalty given by the function (7.28) does not have negative values in the admissible region. This penalty, like the previous one, is a barrier; in this case, there are also difficulties associated with the possible appearance of inadmissible points.

Penalty functions-equalities for the interior point method can be chosen in the form (7.23), (7.24).

In general, the optimization algorithm by the method of penalty functions looks as follows:

1. On the basis of the problem (7.13 - 7.16), we construct the function (7.17) or (7.18), select the initial approximations x and the initial values of the penalty coefficients r_j, l_k (r, l), the number of iterations, the accuracy of unconditional optimization, the accuracy of compliance with restrictions, etc.

2. We solve the problem (7.19) - the stationary point of the function $Q(x,r,l)$ must correspond to the minimum value of the function $f(x)$ that satisfies the constraint.

3. If the obtained solution does not satisfy the system of constraints in the case of using the method of penalty functions, then:

- For the outer point method, we increase the value of the penalty coefficients
- For the internal point method, we decrease the value of the penalty coefficients

and again, solve the problem (7.18). The process stops if the found solution satisfies the system of constraints with a certain accuracy.

Example:
Using the method of penalty functions, find the conditional minimum of the function:

$$f(x_1, x_2) = 2(x_1 - 2)^2 + x_2^2$$

With restrictions:
$g_1(x_{1,}x_2) = x_1 - 2x_2 + 1 \leq 0$
$g_2(x_{1,}x_2) = -2x_1 - x_2 + 2 \leq 0$

The function *f(x)* has a stationary point of unconditional minimum $x_1=2$, $x_2=0$ which does not satisfy the constraints. We take the penalty function in the form (7.21) and obtain the auxiliary function (7.18), the search for the unconditional minimum of which is carried out using the methods of half-division and coordinate descent (Gaussian) considered earlier, while adjusting the coefficient *r* in the direction of increase starting from $r = 1$ in such a way that the minimum value of the function (7.18) was equal to the value of the function f(x) corresponding to the constraints.
We get the solution*:* $x_1=2$; $x_2=1.25$; $f(x_1, x_2) = 2.5$ $g_1(x_{1,}x_2) = 0$ $g_2(x_{1,}x_2) = -3{,}5; r = 3$. The graphical interpretation of the solution and behavior of the penalty function in the intersection at $x_1=2$ is presented in Fig. 7.5. Here, a green square shows the point of the unconditional minimum of the function *f(x)*, and red triangles show the points of the conditional minimum of *f(x)* and the unconditional minimum of *Q(x,r)*, which is the solution to the problem, as well as the values of the constraints.

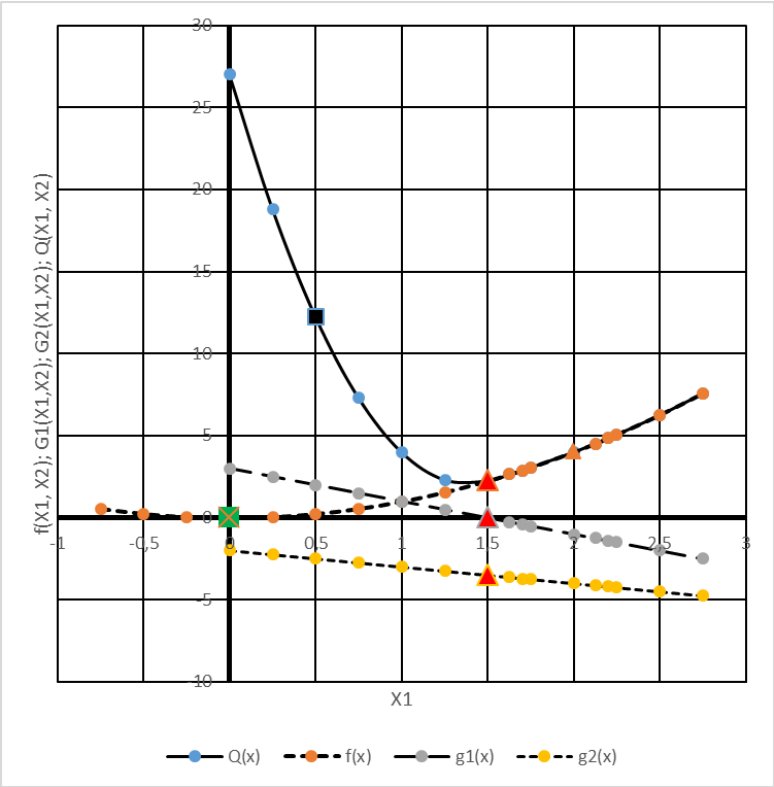


Figure 7.5 – Graphical interpretation of penalty functions method

QUESTIONS FOR SELF-TESTING

1. What are the varieties of the penalty functions method?
2. What is the outer point method?
3. Which constraints can be interpreted as non-hard?
4. What is the form of non-hard constraints of factors?
5. How is the quadratic penalty function written?
6. Give an example of a penalty function for an equality constraint using the interior point method?
7. What is the interior point method?
8. What functions are called barriers?
9. Give an example of a barrier function.
10. What difficulties can arise when applying barrier functions?
11. What is the algorithm for finding the minimum by the method of penalty functions?

8. SOME TYPICAL LIMITATIONS OF MACHINING

Let's consider some limitations that arise when optimizing the processes of machining, in particular, turning.

Limiting the cutting speed, based on the given cutting tool life:

$$V \leq \frac{C_v}{t^{x_v} S^{y_v} T^m} K_v \; or \; n - \frac{318 C_v}{t^{x_v} S^{y_v} T^m D} K_v \leq 0 \qquad (8.1)$$

Where n is the rotation frequency of the machine spindle, D is the diameter of the workpiece

At the same time, the tool life must be no less than the machining time of the one part t_0: $t_0 \leq T$, and the cutting depth must be no more than the allowance A: t≤A

Feed limitation based on surface roughness Rz. In the first approximation, when choosing a feed, you can use the formulas for calculating the geometric component of roughness, bearing in mind that the real roughness will be higher, since the influence of the rigidity of the technological system, individual properties of the machined and tool material, other modes and conditions. The given dependencies are valid for axial turning. For a sharp cutter tip, the condition will look like this:

$$S \leq \frac{R_z (tg\varphi + tg\varphi_1)}{tg\varphi \cdot tg\varphi_1} \qquad (8.2)$$

Where φ and φ_1 are respectively the main and cutting-edge angles.

If there is a radius of rounding of the cutter tip R, which is typical for cutters with mechanical fastening of replaceable polyhedral plates, the condition will look like this:

$$S \leq \sqrt{8R \cdot R_z} \qquad (8.3)$$

Limitation of the estimated rotation frequency of the workpiece *n* and the selected feed *S* due to the kinematic capabilities of the machine tool:

$$n \leq n_{max} \text{ або } n - n_{max} \leq 0 \qquad (8.4)$$

$$S \leq S_{max} \text{ або } S - S_{max} \leq 0 \qquad (8.5)$$

The limitation is due to the power of the machine:

$$N_e \leq N_{em}\eta \tag{8.6}$$

Where N_e is the effective cutting power; N_{em}– power of the electric motor of the machine; η –efficiency factor.

Since the effective power is determined by:

$$N_{\text{e}} = \frac{P_z V}{1000 \cdot 60} \tag{8.7}$$

Where Pz is the main component of the cutting force, which is determined by:

$$P_z = C_P t^{x_p} S^{y_p} V^{z_p} K_p \tag{8.8}$$

From (8.6), (8.7), (8.8) we obtain another dependence that limits the cutting speed:

$$V \leq \left(\frac{1000 \cdot 60 \cdot N_{e\text{д}}\eta}{C_P t^{x_p} S^{y_p} K_p}\right)^{\frac{1}{1+z_p}} \tag{8.9}$$

Limitations related to machining accuracy:

$$P_y \left(\frac{1}{j_{wp}} + \frac{1}{j_m} + \frac{1}{j_t}\right) \leq k_a \frac{\delta_p}{2} \tag{8.10}$$

Where P_y is the radial component of the cutting force; $j_{wp,}\, j_m, j_t$ - rigidity of the workpiece, machine and tool, respectively; k_a - coefficient showing in which part of the tolerance should fit the error caused by the deformation of the technological system elements - is usually taken within 0.7 ... 0.8; δ_p– tolerance on the size being machined. The components of formula (8.10) can be determined by the following dependencies:

$$P_y = C_{Py} t^{x_{Py}} S^{y_{Py}} V^{z_{Py}} K_{Py} \tag{8.11}$$

$$\frac{1}{j_{\text{заг}}} = \frac{L^3}{k_1 J E} \tag{8.12}$$

where L is the free length of the workpiece, mm; k_1 is a coefficient that takes into account the method of fastening the workpiece (in the lathe chuck - $k_1=3$,

in the machining center $k_1=70$, in the lathe chuck with the tightening by machining center of the poppet head - $k_1=100$); E – modulus of elasticity of the workpiece material, MPa; J is the inertia moment of the most dangerous section of the workpiece:

- for a solid section $J = 0{,}05D^4$
- for the ring section: $J = 0{,}05(D^4 - d^4)$

$$\frac{1}{j_в} = 0{,}000025k_2\sqrt[3]{\frac{200}{H_c}} \tag{8.13}$$

Where H_c center-line height of the machine, mm; k_2 is a coefficient that takes into account the type of fastening:

- when fixed in the centers and in the chuck with the tightening by the center of the poppet head $k_2 = 1$
- when fixed in the lathe chuck $k_2 = 0.03L$

$$\frac{1}{j_t} = \frac{l^3}{3J_\text{p}E_\text{p}} \tag{8.14}$$

In the case of turning, here l is the length of the cutter, mm; Ep – modulus of elasticity of the cutter material, MPa; J is the cutter section inertia moment:

- for a round cross-section $J = 0{,}05d_\text{p}^4$
- for a rectangular section: $J = \frac{HB^3}{12}$

In addition, the following additional restrictions may be introduced:

- Cutting temperature limitations based on empirical formulas or theoretical calculations;
- Limitations associated with the study of the kinetics of thermal processes during cutting, that is, with structural transformations in the machined material;

As we can see, there are numerous and quite complex interrelated constraints in cutting, which must be taken into account when optimizing cutting conditions.

QUESTIONS FOR SELF-TESTING

1. By what criteria the cutting speed constraints are determined?
2. By what criteria the feed constraints are determined?
3. Name some additional constraints when cutting.

9. LINEAR PROGRAMMING PROBLEMS

9.1. General provisions

A large class of optimization problems consists of linear programming problems - conditional optimization with linear goal functions and linear constraints. Unlike nonlinear programming problems, where the optimal solution can be found both inside the space delineated by constraints and on the constraints themselves, in linear programming problems, the solution is always on the constraint lines.

The main problem of linear programming can be written in the following form:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \quad (9.1)$$

$$F = c_0 + c_1x_1 + c_2x_2 + \cdots + c_nx_n \to \min \quad (9.2)$$

Here (9.1) is the system of problem constraints; (9.2) is the goal function.

In addition to equality constraints, the main problem also includes inequality constraints:

$$\begin{cases} x_1 \geq 0 \\ x_2 \geq 0 \\ \cdots \\ x_m \geq 0 \end{cases} \quad (9.3)$$

Any non-negative solution of the system (9.1) - $x_i^{(0)} \geq 0$ is called *admissible Totality of admissible solutions called a plan* of a linear programming problem. *Plan that minimizes the function (9.2) is called optimal.* The problem has a solution only if the rank of the matrix of the constraint system (consisting of unknown variables held constants) r is not greater than the number of unknown variables n. If $r = n$, then the system has a unique solution that will be optimal if it is admissible. If not, then the problem has no solutions. Thus, only the case when $n < r$ is of interest, which will be considered later.

Every problem of linear programming can be reduced to the basic form. For this:

- Any maximization problem must be reduced to a minimization problem.
- Be able to move from inequality constraints to equivalent equality constraints.

If the constraints of the problem are given in the form of inequalities:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1 \geq 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2 \geq 0 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_m \geq 0 \end{cases} \quad (9.4)$$

Then (9.4) can always be reduced to the form (9.1) by introducing additional non-negative variables: $x_{n+1}, x_{n+2}, \ldots x_{n+m}$:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1 = x_{n+1} \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2 = x_{n+2} \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_m = x_{n+m} \end{cases} \quad (9.5)$$

It is obvious that the condition of non-negativity of additional variables is equivalent to the fulfillment of inequalities (9.4). Thus, the constraints - inequalities (9.4) are equivalent to the constraints-equalities (9.5), and hence also (9.1).

Let's consider several examples of a linear programming problem.

9.2. The task of using raw materials

Manufacturing of two types of products - P1 and P2 requires four types of raw materials - S1, S2, S3, S4. The stock of raw materials is limited and is, respectively, b1, b2, b3, b4. The number of units of raw materials required to produce a unit of given products is known and represented in the form of a table. It is necessary to draw up such a plan for the manufacturing of products that the income from its sale would be maximum. The conditions of the problem are presented in Table. 9.1.

Table 9.1 - Conditions of the task on the use of raw materials

Types of raw material	Stocks of raw materials	Types of products	
		P_1	P_2
S_1	b_1	a_{11}	a_{12}
S_2	b_2	a_{21}	a_{22}
S_3	b_3	a_{31}	a_{32}
S_4	b_4	a_{41}	a_{42}
Income		c_1	c_2

If we denote the quantity of products produced by x_1 and x_2, respectively, the goal function will look like:

$$F = c_1x_1 + c_2x_2 \rightarrow max \tag{9.6}$$

Since any problem must be reduced to a minimization problem, the function (9.6) will take the form:

$$F = -c_1x_1 - c_2x_2 \rightarrow min \tag{9.7}$$

The conditions of the problem can be written in the form of a system:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \leq b_1 \\ a_{21}x_1 + a_{22}x_2 \leq b_2 \\ a_{31}x_1 + a_{32}x_2 \leq b_3 \\ a_{41}x_1 + a_{42}x_2 \leq b_4 \end{cases} \tag{9.8}$$

System (9.8) can be reduced to the conditions of the basic problem in the form:

$$\begin{cases} b_1 - a_{11}x_1 - a_{12}x_2 \geq 0 \\ b_2 - a_{21}x_1 - a_{22}x_2 \geq 0 \\ b_3 - a_{31}x_1 - a_{32}x_2 \geq 0 \\ b_4 - a_{41}x_1 - a_{42}x_2 \geq 0 \end{cases} \tag{9.9}$$

Or, by entering additional variables, lead to the form:

$$\begin{cases} b_1 - a_{11}x_1 - a_{12}x_2 = x_3 \\ b_2 - a_{21}x_1 - a_{22}x_2 = x_4 \\ b_3 - a_{31}x_1 - a_{32}x_2 = x_5 \\ b_4 - a_{41}x_1 - a_{42}x_2 = x_6 \end{cases} \quad (9.10)$$

We will consider the solution using a quantitative example:

Table 9.2 - Given data

Types of raw materials	Stocks of raw materials	Types of products	
		P_1	P_2
S_1	*19*	*2*	*3*
S_2	*13*	*2*	*1*
S_3	*15*	*0*	*3*
S_4	*18*	*3*	*0*
Income		*7*	*5*

Conditions from Table. 9.2 can be written in the following form:

$$\begin{cases} 2x_1 + 3x_2 \le 19 \\ 2x_1 + x_2 \le 13 \\ 3x_2 \le 15 \\ 3x_1 \le 18 \\ x_1 \ge 0 \\ x_2 \ge 0 \end{cases}$$

Or in the forms (9.8), (9.9):

$$\begin{cases} x_1 \geq 0 & (I) \\ x_2 \geq 0 & (II) \\ 19 - 2x_1 - 3x_2 \geq 0 & (III) \\ 13 - 2x_1 - x_2 \geq 0 & (IV) \\ 15 - 3x_2 \geq 0 & (V) \\ 18 - 3x_1 \geq 0 & (VI) \end{cases}$$

$$\begin{cases} x_1 \geq 0 \\ x_2 \geq 0 \\ 19 - 2x_1 - 3x_2 = x_3 \\ 13 - 2x_1 - x_2 = x_4 \\ 15 - 3x_2 = x_5 \\ 18 - 3x_1 = x_6 \end{cases}$$

At the same time, the matrix of the system, composed of unknow variable held constants, will have the form:

$$I = \begin{pmatrix} 2 & 3 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

It is clear that the system cannot be solved unambiguously, because the number of unknown values is greater than the number of equations: the rank of the matrix $r = 4$, the number of unknowns $n = 6$. The number of degrees of freedom $k = 6 - 4 = 2$

The goal function has the form:

$$F = -7x_1 - 5x_2 \rightarrow min$$

To solve the problem, we will use its geometric interpretation. For this, we introduce the x_1Ox_2 coordinate system. It is known that the locus of points on the plane, the coordinates of which satisfy the system of linear inequalities, form a convex polyhedron, which we will call *the polyhedron of solutions* of the system of inequalities. The sides of this polyhedron are located on straight lines, the equations of which are obtained if the inequality signs are replaced by exact equalities in the system of inequalities. This

polyhedron itself is the intersection of the half-planes into which each of the indicated lines divides the plane. In our case, such straight lines are:

$$\begin{cases} x_1 = 0 & (I) \\ x_2 = 0 & (II) \\ 19 - 2x_1 - 3x_2 = 0 & (III) \\ 13 - 2x_1 - x_2 = 0 & (IV) \\ 15 - 3x_2 = 0 & (V) \\ 18 - 3x_1 = 0 & (VI) \end{cases}$$

The lines are shown in Fig. 9.1, the arrows indicate which half-planes form the polyhedron of solutions.

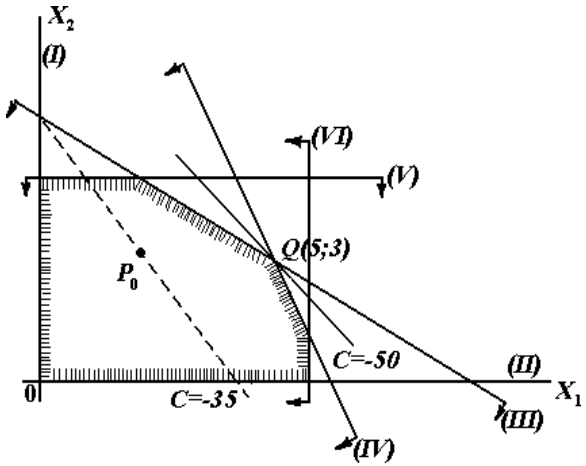


Figure 9.1 - Polyhedron of solutions

Along with this, we will consider the goal function F mentioned above and write it in the following form:

$-7x_1 - 5x_2 = \mathrm{C}$

By changing *C*, we get a set of parallel lines. On Fig. 9.2 shows the straight lines for different *C*, and the vector *g* indicates the direction of movement to decrease the values of C. Therefore, it is obvious that the solution will lie on the side or the vertex of the polyhedron of solutions farthest in the direction of the vector *g*.

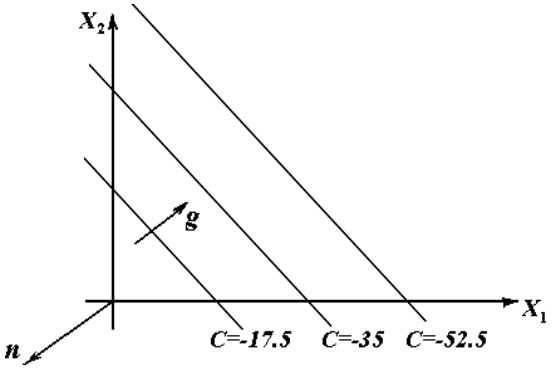


Figure 9.2 - Goal function

Such a vertex in this case is Q(5;3) (See Fig. 9.1) - which can be determined by a simultaneous solution of equations (III) and (IV). At the same time, the value of the goal function will be minimal, which satisfies the specified requirements: F = -50. Therefore, the most profitable solution is the manufacturing of 5 units of P1 and 3 units of P2.

9.3. The task of using productivity of equipment

In time T, the enterprise must produce N_1 units of P_1 and N_2 units of P_2 products. Both types of products can be manufactured on equipment A and B, which have different productivity (number of product units manufactured per unit of time) and production costs per unit of time. The given data are summarized in Table. 9.3

Table 9.3 - General type of given data to the problem

Equipment	Productivity per unit of time		Production costs per unit of time		Total time for the plan production	
Products	P_1	P_2	P_1	P_2	P_1	P_2
A	a_1	a_2	α_1	α_2	x_1	x_2
B	b_1	b_2	β_1	β_2	x_3	x_4

According to the conditions of the task, it is necessary to plan a quantity of products manufacturing on the appropriate equipment in order to minimize production costs.

The goal function will look like this:

$$F = \alpha_1 x_1 + \alpha_2 x_2 + \beta_1 x_3 + \beta_2 x_4 \to \min \tag{9.11}$$

The conditions of the task will be written down:

$$\begin{cases} x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0 \\ x_4 \geq 0 \\ x_1 + x_2 \leq T \\ x_3 + x_4 \leq T \\ a_1 x_1 + b_1 x_3 = N_1 \\ a_2 x_2 + b_2 x_4 = N_2 \end{cases} \text{ або } \begin{cases} x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0 \\ x_4 \geq 0 \\ x_5 \geq 0 \\ x_6 \geq 0 \\ T - x_1 - x_2 = x_5 \\ T - x_3 - x_4 = x_6 \\ a_1 x_1 + b_1 x_3 = N_1 \\ a_2 x_2 + b_2 x_4 = N_2 \end{cases} \tag{9.12}$$

That is, with 4 equations, we have 6 variables, which requires expressing 4 of them in terms of 2 others, since for a graphical solution you need to have 2 coordinates. We will consider the solution using a quantitative example. The initial data are given in Table. 9.4. The total time for the manufacturing of the entire plan is *T=6* hours.

Table 9.4 - Initial data

Equipment	Productivity per unit of time		Production costs per unit of time		Total time for the plan production		Plan	
Products	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
A	*6*	*24*	*4*	*47*	x_1	x_2	*30*	*96*
B	*13*	*13*	*13*	*26*	x_3	x_4		

Assuming by default that all variables are non-negative, we write the system (9.12) and the goal function (9.11):

$$\begin{cases} x_1 + x_2 \le 6 & (I') \\ x_3 + x_4 \le 6 & (II') \\ 6x_1 + 13x_3 = 30 & (III') \\ 24x_2 + 13x_4 = 96 & (IV') \end{cases}$$

$$F = 4x_1 + 13x_3 + 47x_2 + 26x_4$$

The constraint system matrix will look like this:

$$I = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 6 & 13 & 0 & 0 & 0 & 0 \\ 0 & 0 & 24 & 13 & 0 & 0 \end{pmatrix}$$

Let's choose x_1 and x_2 as free variables and express x_3 and x_4 from equations (III') and (IV') through them:

$$x_3 = \frac{1}{13}(30 - 6x_1) = \frac{6}{13}(5 - x_1) \ge 0 \Rightarrow x_1 \le 5$$

$$x_4 = \frac{1}{13}(96 - 24x_2) = \frac{24}{13}(4 - x_2) \ge 0 \Longrightarrow x_2 \le 4$$

By substituting the obtained expressions into equation (II'), we obtain the system and the goal function:

$$\begin{cases} x_1 \ge 0 & (I) \\ x_2 \ge 0 & (II) \\ x_1 \le 5 & (III) \\ x_2 \le 4 & (IV) \\ x_1 + x_2 \le 6 & (V) \\ x_1 + 4x_2 \ge 8 & (VI) \end{cases}$$

$$F = 222 - 2x_1 - x_2 \to min$$

The geometric interpretation of the solution will look like (Fig. 9.3):

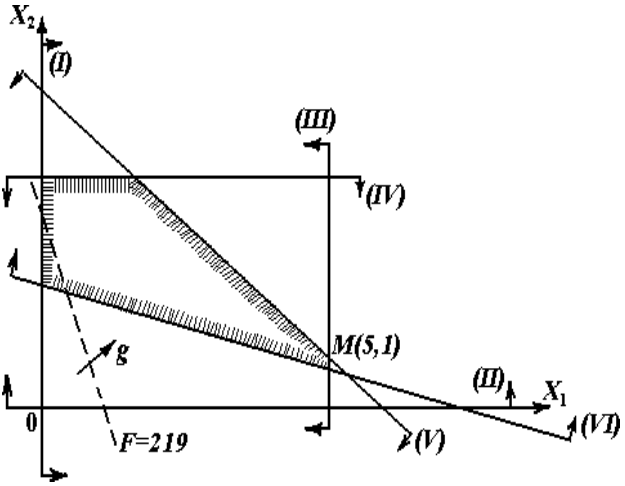


Figure 9.3 - Geometric interpretation of the solution

It can be seen from the figure that the solution will be the point M(5;1), the coordinates of which are determined by the simultaniouse solution of equations (III) and (V):

$x_1 = 5;\ x_2 = 1; F = 211$

Now it is possible to define: $x_3 = 0;\ x_4 = \frac{72}{13}$

So, the task is completely solved. Thus, $x_1 a_1 = 5 \cdot 6 = 30$ units of product P1 and $x_2 a_2 = 1 \cdot 24 = 24$ units of product P2 will be produced on the equipment A for the specified time. Equipment B will produce $x_3 b_1 = 0 \cdot 13 = 0$ units of P1 and $x_4 b_2 = \frac{72}{13} \cdot 13 = 72$ units of P2

9.4. Transportation problem

Two suppliers *A1* and *A2* have some homogeneous products in quantities a_1 and a_2, respectively. These products must be delivered to three consumers *B1*, *B2* and *B3* in quantities respectfully b_1, b_2, b_3. We denote the cost of transportation from point A_i to point B_j by C_{ij}, and the quantity of products to be transported by x_{ij}. It is necessary to draw up a transportation

plan in order to minimize the total cost of transportation. Often, the task is supplemented with data of the cost of products from each of the suppliers, and then the total costs of products and transportation are minimized. Thus, the mathematical formulation of the transport problem will look like this:

$$\begin{cases} x_{11} + x_{21} = b_1 \\ x_{12} + x_{22} = b_2 \\ x_{13} + x_{23} = b_3 \\ x_{11} + x_{12} + x_{13} \le a_1 \\ x_{21} + x_{22} + x_{23} \le a_2 \end{cases} \tag{9.13}$$

$$F = \sum_{i=1}^{2} \sum_{j=1}^{3} x_{ij} C_{ij} \to min \tag{9.14}$$

A simplified version of the transport problem consists in the equality of total needs and stocks:

$$\sum_{i=1}^{2} a_i = \sum_{j=1}^{3} b_j \tag{9.15}$$

In this case, the problem is simplified, the last two equations in the system (9.12) are transformed into equalities, and the problem itself completely takes the form of the main one. Undoubtedly, the number of suppliers and consumers may differ from those indicated in the larger side - which will definitely complicate the solution of the problem.

QUESTIONS FOR SELF-TESTING

1. What are linear programming problems?
2. Where are the solutions to linear programming problems in relation to constraints?
3. What is the form of the main problem of linear programming?
4. What solution is called admissible?
5. What is the outline of a linear programming problem?
6. What solution to the problem of linear programming is called optimal?
7. When there is a single solution to the problem of linear programming and it is optimal?

8. In which case does the linear programming problem have no solution?
9. How can each linear programming problem be represented as a basic one?
10. What is a polyhedron of solutions?
11. What is the essence of the transport problem?
12. What is the feature of the simplified version of the transport problem?

10. OPTIMIZATION ACCORDING TO EXPERIMENTAL RESEARCH

10.1. Terms. A full factorial experiment

The main requirement for a mathematical model is to predict the behavior of the object with a certain accuracy. And if this requirement is specified for optimization models, then it means predicting the direction of movement to those conditions where the object will function in the best way. In any case, we make the following assumptions: the response surface is continuous, smooth, and has a single extremum (at least within the domain of definition).

Usually, when searching for an optimal solution in the factor space, some point *C* is selected and several points symmetrical with respect to it are considered, which together form some subregion (see Fig. 10.1). An experiment is conducted here, based on the results of which the first model is built. This model is used to predict the behavior of the object at points that were not included in the experiment. If the points are inside the subregion, then such a prediction is called *interpolation*, and if outside the subregion, it is called *extrapolation*. The further the point is from the experimental area, the less confidence this can be done. Therefore, we are forced to extrapolate not far, and use the obtained results for conducting further experiments. If you change only one of the factors - x_1 *or* x_2, then points *A* or *B* (Fig. 10.1) can be taken as the optimal solution, while in fact it lies in point M, and in order to find it, you need to change both factors at the same time - this is one of meanings of a multifactorial experiment which is implemented only in combination with *design of experiment (DOE) - the procedure for choosing the number of tests and the conditions for their implementation, necessary to solve the problem with the required accuracy.*

Of all the variety of possible mathematical models, the most suitable for the purposes of a multifactorial experiment are models in the form of algebraic polynomials of the first and second degrees. This suitability lies in the simplicity, convenience and possibility of building a model in the neighborhood of any point. In addition, such models are the most mathematically developed class of models. Thus, for two factors:

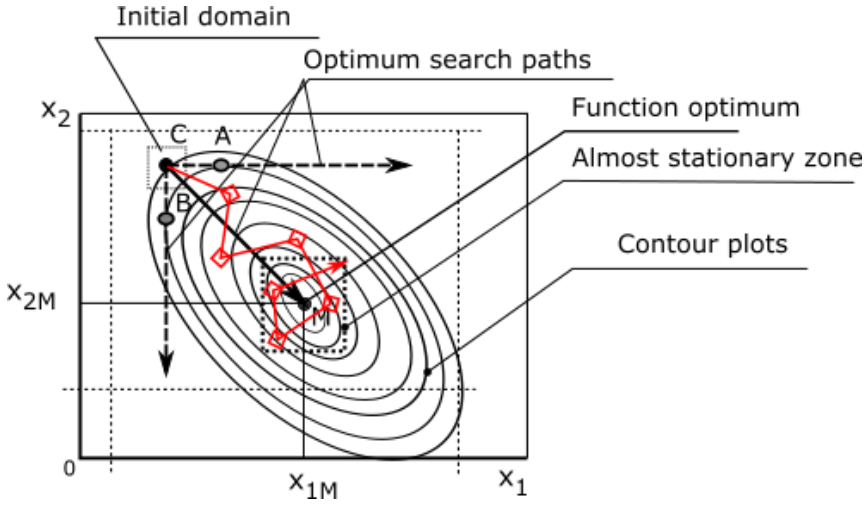


Figure 10.1 – Strategy for finding the optimal solution

- polynomial of the first degree:

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 \qquad (10.1)$$

- polynomial of the second degree:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 \qquad (10.2)$$

The direction of movement to the optimal solution is called the direction of the gradient - $\overline{CM}$ in Fig. 10.1. It is clear that movement in this direction will lead to success sooner than in any other. As can be seen from Fig. 10.1, a polynomial of the first degree satisfies this purpose, having at the same time the minimum number of coefficients. But this will happen until we get to the zone close to the optimum, which is called *"almost stationary"*. In this zone, the use of a linear function in the form of a first-degree polynomial becomes ineffective and does not lead to improved results. In this case, getting into an almost stationary region should be considered the completion of the optimization process or it makes sense to switch to a polynomial of the second degree, because it is a function that has an

extremum and can adequately describe the section of the response surface. Thus, when designing a multifactorial experiment, we will limit ourselves to polynomials of the first and second degrees.

The DOE is characterized by the number of degrees of freedom:

$$f = N - l - 1 \tag{10.3}$$

where N is the number of independent experiments in the design; l is the number of regression coefficients in the equation (without a free coefficient).

Depending on the number of degrees of freedom, the designs are: *unsaturated*, when $f > 0$, that is, the number of experiments is greater than required to determine the coefficients; *saturated* when $f = 0$ and *supersaturated* when $f < 0$.

Before directly planning the experiment, a number of things should be done:

Analysis of a priori information. The study of a technical object usually begins in conditions where such or similar objects have been studied in some way before. The information contained in the results of previous studies is called *a priori*.

Estimation of boundaries of definition areas. For this, the restrictions imposed on the factors must be taken into account.

Selection of the basic level. The best conditions obtained from a priori information correspond to a combination or several factor levels. Each such combination is a multidimensional point in a multidimensional factor space. It can be considered as a starting point for building an experiment design. Let's define it as the *basic or zero level*. The experimental design is reduced to the selection of experimental points symmetrical with respect to the basic (zero) level, which must be within the definition area. Let us denote the real value of the main level of the factor j as $\tilde{x}_{0j}$.

Choice of variability intervals. The next task is to select experimental points symmetrical with respect to the basic level. The position of these points is determined by the *variability interval* - I_j. If the variability interval is added to the zero level, we get the upper level (U), and if it is subtracted from zero level - the lower level (L), i.e., respectively:

$$\tilde{x}_{Uj} = \tilde{x}_{0j} + I_j \tag{10.4}$$

$$\tilde{x}_{Lj} = \tilde{x}_{0j} - I_j$$

It can be graphically depicted as shown in Fig. 10.2

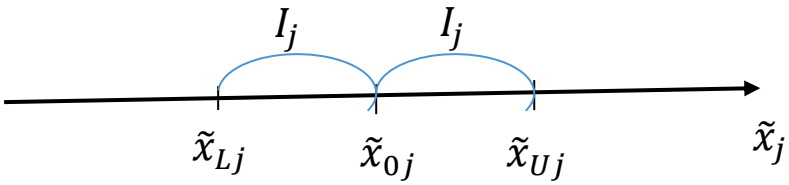


Figure 10.2. — Levels of factors and variability interval

The selection of variability intervals is subject to restrictions from both the bottom and the top. On the one hand, the variability interval cannot be less than the error with which the experimenter fixes the value of the level, otherwise the levels cannot be distinguished. On the other hand, they should not be so large that the values of the levels are outside the definition area. Within these limits, there is definitely considerable uncertainty, the choice of the variation interval is a difficult task because it is related to the informal stage of design the experiment, therefore it should be decided solely due to the experience and intuition of the researcher. The selected value must often be adjusted in the course of experiments. Roughly, if the interval of variation is no more than 10% of the definition area, then it is considered *narrow*; from 10% to 30% - *average*; more than 30% - *wide*.

When designing the experiment, not the actual values of the factors are used, but the coded ones, which are determined as follows:

$$x_j = \frac{\tilde{x}_j - \tilde{x}_{0j}}{I_j} \qquad (10.5)$$

It can be seen from equations (10.4) and (10.5) that the coded value of the basic level is 0 - that is why it is also called the zero level. The coded value of the lower level is -1 (can be marked simply as "-"), and the upper level is +1 (can be marked as "+").

In a *full factorial experiment (FFE)*, the factors vary on two levels - upper and lower. At the same time, all possible combinations of factor values must be implemented. The number of such combinations, and therefore the number of independent tests, is determined by:

$$N = 2^k \tag{10.6}$$

here 2 is the number of levels, k is the number of factors.

The conventional notation FFE - 2^k is also based on formula (10.6). Thus, for two factors, the FFE is denoted by 2^2, and the number of independent tests to be conducted is four. Based on the results of FFE, a model (regression equation) is built in the form of a polynomial of the first degree. For two factors, this model has the form (10.1).

The 2^k design for constructing the regression equation (10.1) is unsaturated because it has $f = 1$ degree of freedom. The conditions of the experiment are recorded in a table called the *DOE matrix* (see Table 10.1). A column corresponding to the fictitious variable x_0 with the free coefficient b_0, which in all tests has a value of +1, is additionally entered into this matrix - it will be used exclusively for calculations, and not as a condition of DOE.

Table 10.1 — 2^2 DOE matrix

N	x_0	Tests conditions		Tests results
		x_1	x_2	y
1	+ 1	+ 1	+ 1	y_1
2	+ 1	- 1	+ 1	y_2
3	+ 1	+ 1	- 1	y_3
4	+ 1	- 1	- 1	y_4

Each column of the DOE matrix is called a *column-vector*, and a row is called a *row-vector*.

A FFE in combination with a regression equation in the form of an algebraic polynomial makes it quite easy to determine the regression coefficients. It should be remembered that the factors in such an equation have coded values, not real ones. In the following formula (10.7), j is the index of the regression coefficient:

$$b_j = \frac{1}{N}\sum_{i=1}^{N}(x_{ji} \cdot y_i) \qquad (10.7)$$

Based on equation (10.7), the regression coefficients from equation (10.1) based on the matrix Table. 10.1 will be defined as the sum of y with the signs of the corresponding x:

$$b_0 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

$$b_1 = \frac{1}{4}(y_1 - y_2 + y_3 - y_4)$$

$$b_2 = \frac{1}{4}(y_1 + y_2 - y_3 - y_4)$$

The presence of one degree of freedom allows one more element to be introduced into equation (10.1) - and this element will be nonlinear. Equation (10.1) will take the form:

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{12} \cdot x_1 \cdot x_2 \qquad (10.8)$$

Here, the element $b_{12} \cdot x_1 \cdot x_2$ characterizes the interaction of factors and is called the *interaction effect*. It reflects the change in the effect of one of the factors when the level of another factor changes. To calculate the coefficient b_{12}, the corresponding column is additionally entered into the DOE matrix (Table 10.2)

Table 10.2 - 2^2 DOE matrix with interaction effect

N	x_0	Tests conditions		x_1x_2	Tests results
		x_1	x_2		y
1	+ 1	+ 1	+ 1	+ 1	y_1
2	+ 1	- 1	+1	- 1	y_2
3	+ 1	+ 1	- 1	- 1	y_3
4	+ 1	- 1	- 1	+ 1	y_4

The coefficient for the interaction effect will be determined accordingly:

$$b_{12} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4)$$

In the presence of two factors, it is not difficult to go through all possible combinations of level values, but when they increase, some rules are needed to formalize this procedure. Consider two:

a) Alternation of signs: the first column - signs one after the other; the second in two; the third in four, etc. And in general, the signs change by 2^{j-1} where *j* is the number of the column.

b) Let's agree that when $k > 2$, for DOE 2^k, DOE 2^{k-1} is the original. When a new factor appears, the original plan is repeated twice: in combination with the lower and upper levels of the original plan.

As the number of factors increases, the number of possible interactions increases rapidly. In the presence of three factors, the element $x_1x_2x_3$ is called the second-order interaction effect. The DOE matrix, taking into account interaction effects, will of course be expanded. A fundamentally important point when designing an experiment and analyzing the obtained results is that, as a rule, *the greater the order of interaction, the less likely it is to occur.*

The resulting model must be checked for adequacy (not considered in this course).

10.2. Orthogonal central composite designs

Once in the almost stationary zone during the solution of the optimization problem, the polynomial of the first degree no longer improves the result. Now it is necessary to either stop and consider the obtained results as the best, or try to describe the response with a more complex model - a polynomial of the second degree, which, unlike a polynomial of the first degree, has an extremum. This approach makes sense if, on the basis of a priori information, including theoretical ideas about the process, there are reasons to believe the presence of such an extremum in the response function.

This situation can be illustrated in a simplified way, as shown in Fig. 10.3. The figure shows a response surface trace in the direction of the gradient. Indeed, the gradient in the form of a polynomial of the first degree correctly indicates the direction to the optimal solution represented by the point M. But moving in this direction, we get to point *P* and expect the corresponding result. While we actually get a result that corresponds to point *Q* and which is worse (or not better) than the result at the initial point *C*. That is, to find the optimal solution within the limits of the lower and upper levels of the factors (represented by points *L* and *U*, respectively), the polynomial of the first degree already it is not enough and it is necessary to have a more accurate description of the response surface with a reflection of the existing extremum. And this description can already give a polynomial of the second degree, which, taking into account the possible interactions of factors for two factors, has the form (10.9):

$$y = \beta_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 \qquad (10.9)$$

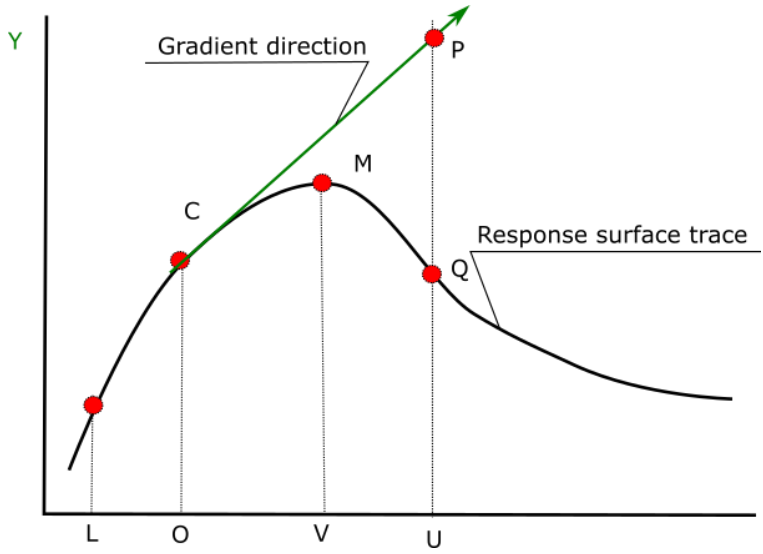


Figure 10.3 – Almost stationary zone

And in general, for *k* factors:

$$y = \beta_0 + \sum_{j=1}^{k} b_j x_j + \sum_{j=1}^{k} b_{jj} x_j^2 + \sum_{j<u}^{k} b_{ju} x_j x_u \qquad (10.10)$$

To obtain the model (10.10), the variables must vary at least on 3 levels, and the number of series of independent experiments, excluding repetitions, *N* must meet the requirement:

$$N \geq \frac{(k+2)(k+1)}{2} \qquad (10.11)$$

The *orthogonality* of any DOE is a property that ensures the independence of regression coefficient estimates, and in this case, it is achieved by using orthogonal central composite designs (OCCD) of the second order - which are among the simplest among a large number of possible ones. The experiment plan includes experiments at the basic level and the FFE matrix (or its fractional replica for k > 5), which make up the core of the design, and several additional points called stars. So, for *k* factors, the design will consist of the following elements:

- Core 2^k or 2^{k-p}
- Center with coordinates (0,0,0, …, 0)
- Star points with coordinates (+ d, 0, 0, …, 0); (- d, 0, 0, ..., 0); (0, + d, 0, …, 0); (0, - d, 0, ..., 0) etc. Here d is the so-called shoulder (encoded value), and d≥1.

The total number of independent tests, excluding repetitions, will be determined by:

$$N = 2^k + 1 + 2k \qquad (10.12)$$

or

$$N = 2^{k-p} + 1 + 2k \qquad (10.13)$$

For example, for two factors we have the plan shown in the table. 10.3, where $N = 2^2 + 1 + 2 \cdot 2 = 9$.

Table 10.3 — Orthogonal central composite design $2^2 + 1 + 2 \cdot 2$

N	**Design elements**	x_1	x_2	y
1	Core	+ 1	+ 1	y_1
2		- 1	+ 1	y_2
3		+ 1	- 1	y_3
4		- 1	- 1	y_4
5	Center	0	0	y_5
6	Star points	+ d	0	y_6
7		- d	0	y_7
8		0	+ d	y_8
9		0	- d	y_9

To ensure the orthogonality of the design (sum of products of corresponding elements of any two columns should be equal to zero), the following substitutions $b_0 = \left(\beta_0 + \lambda \sum_{j=1}^{k} b_{jj}\right)$ and $x_j' = x_j^2 - \lambda$ are introduced into equation (10.10), and it takes the form:

$$y = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{j=1}^{k} b_{jj} x_j' + \sum_{j<u}^{k} b_{ju} x_j x_u \qquad (10.14)$$

The DOE matrix from Table 10.3, taking into account the column vectors for calculation, will take the form shown in Table 10.4.

The unknowns - λ and d are determined from the conditions for ensuring the orthogonality of the columns:

$$d = \sqrt{\frac{1}{2}\left[\sqrt{N \cdot 2^{k-p}} - 2^{k-p}\right]} \qquad (10.15)$$

$$\lambda = \frac{2^{k-p} + 2d^2}{N}$$

Some values of λ and d depending on the number of factors and the core of the experiment are given in Table. 10.5

Table 10.4 — Orthogonal central composite design $2^2 + 1 + 2 \cdot 2$ in the general form with vector-columns for calculations

N	x_0	x_1	x_2	$x_1' = x_1^2 - \lambda$	$x_2' = x_2^2 - \lambda$	$x_1 x_2$	y
1	+ 1	+ 1	+ 1	*1-λ*	*1-λ*	*+ 1*	y_1
2	+ 1	- 1	+ 1	*1-λ*	*1-λ*	*- 1*	y_2
3	+ 1	+ 1	- 1	*1-λ*	*1-λ*	*- 1*	y_3
4	+ 1	- 1	- 1	*1-λ*	*1-λ*	*+ 1*	y_4
5	+ 1	0	0	*-λ*	*-λ*	*0*	y_5
6	+ 1	+ d	0	d^2- λ	*-λ*	*0*	y_6
7	+ 1	- d	0	d^2- λ	*-λ*	*0*	y_7
8	+ 1	0	+ d	*-λ*	d^2- λ	*0*	y_8
9	+ 1	0	- d	*-λ*	d^2- λ	*0*	y_9

The regression coefficients of the transformed equation (10.14) are determined as follows [36]:

$$b_j = \frac{1}{C_j} \sum_{i=1}^{N} (x_{ji} y_i) \tag{10.16}$$

Where C_j is the sum of the squares of the elements of each column vector:

$$C_j = \sum_{i=1}^{N} x_{ji}^2 \tag{10.17}$$

Based on Table. 10.5, the DOE given in Table 10.4 is presented in a general form in Table 10.6, the bottom row of which shows the sum of the squares of the elements of each column vector - C_j.

Table 10.5 — Characteristics of some OCCDs

Number of factors k	p	Core	Number of tests N	Shoulder d	λ	$1-\lambda$	$d^2 - \lambda$
2	0	2^2	9	1	0,667	0,333	0,333
3	0	2^3	15	1,215	0,730	0,270	0,746
4	0	2^4	25	1,414	0,800	0,200	1,199
5	0	2^5	43	1,596	0,863	0,137	1,684
5	1	2^{5-1}	27	1,547	0,770	0,230	1,623
6	0	2^6	77	1,761	0,912	0,088	2,189
6	1	2^{6-1}	45	1,724	0,843	0,157	2,129
6	2	2^{6-2}	29	1,664	0,743	0,257	2,026
7	0	2^7	143	1,909	0,946	0,054	2,698
7	1	2^{7-1}	79	1,885	0,900	0,100	2,653
7	2	2^{7-2}	47	1,841	0,825	0,175	2,564
7	3	2^{7-3}	31	1,771	0,718	0,282	2,418

Table 10.6 — Orthogonal central composite design $2^2 + 1 + 2 \cdot 2$

N	DOE elements	x_0	x_1	x_2	$x_1' = x_1^2 - \lambda$	$x_2' = x_2^2 - \lambda$	$x_1 x_2$	y
1	Core	+ 1	+ 1	+ 1	0,333	0,333	+ 1	y_1
2		+ 1	- 1	+ 1	0,333	0,333	- 1	y_2
3		+ 1	+ 1	- 1	0,333	0,333	- 1	y_3
4		+ 1	- 1	- 1	0,333	0,333	+ 1	y_4
5	Center	+ 1	0	0	-0,667	-0,667	0	y_5
6	Star points	+ 1	+ 1	0	0,333	-0,667	0	y_6
7		+ 1	- 1	0	0,333	-0,667	0	y_7
8		+ 1	0	+ 1	-0,667	0,333	0	y_8
9		+ 1	0	- 1	-0,667	0,333	0	y_9
$C_j = \sum_{i=1}^{N} x_{ji}^2$		9	6	6	*2*	*2*	*4*	

Given the known coefficients of equation (10.14), the free term of the original equation (10.10) can be found:

$$\beta_0 = b_0 - \lambda \sum_{j=1}^{k} b_{jj} \qquad (10.18)$$

10.3. DOE and optimization according to the Taguchi concept

The concept was put forward by Dr. Henichi Taguchi while he was working at the Electrical Communication Laboratory (ECL) established in Japan after World War II. Dealing with the optimization of DOE, he proposed designs in which the number of tests is significantly reduced in relation to traditional designs, especially with a large number of factors. A comparison of the number of necessary independent tests when conducting a full factorial experiment and according to Taguchi's concept is given in Table. 10.7.

Table 10.7 - Comparison of the number of necessary independent tests for a full factorial experiment and according to Taguchi's design

Factors	Levels	Number of tests	
		FFE	Taguchi DOE
2	2	4	4
3	2	8	4
4	2	16	8
7	2	128	8
15	2	32768	16
4	3	81	9

The initial purpose of the plans was to ensure and manage product quality indicators. Taguchi's methodology began to be widely used in the United States in the early 80s of the XX century and brought significant changes to quality engineering and provided impressive results, which gave some reason to consider the proposed approach the most significant contribution to statistical methods of quality management in the last few decades.

In principle, the approach to DOE practically does not differ from the multifactorial experiments discussed above - the sequence of experiments is a set of different combinations of factor levels. Designing is carried in the form of standardized *orthogonal arrays* (OA). The result of the application of the technique is the search for the optimal combination of factor levels that provides the minimum or maximum value of the goal or the resulting indicator. But there are some differences:

- The number of factor levels can be any;
- The number of levels for different factors may be different;
- Factors are divided into those that can be controlled or the *external array* and those that cannot be controlled - *the internal array*. The factors of the internal array are called *noise factors*. The presence of noise leads to the deviation of the system from ideal operation. In turn, the factors of the internal array can be divided into those that can be controlled and measured and random variations. There are three categories of noise factors:

a) External environment (humidity, temperature, etc.);
b) Aging and changing properties of the object over time;
c) Manufacturing variations.

- Some factors from the internal array - those whose values are subject to control and measurement can be entered into the DOE. The number of levels for such factors is usually equal to 2 (for example, above or below some threshold value).

- As a result of the experiment, the object or process model is not built, but only the optimal combination of factor values is determined. The optimal combination is considered to be the one that provides the minimum or maximum value of the resulting indicator. Based on this, increasing the number of levels increases the accuracy of determining the optimal solution, but increases the required number of tests.

- The optimal combination of factor values, determined during the experiment and by means of a simple calculation, as a rule, does not coincide with those used for conducting tests. This is quite understandable, because Taguchi's plans are only a small fragment of a full factorial experiment. Such

a result is predicted, and not what was obtained during the experiment. If so, a mandatory element of Taguchi's concept is the need to conduct a verification experiment on the determined optimal combination of factor levels.

The difference between using the Taguchi approach to quality management and optimizing any physical metrics will be only what is used as the resulting metric. If we are talking about optimization as such, then the resulting indicator will be any physical indicator of the process or product, for example, the amount of wear of the cutting tool or its life period, cutting force, etc. Such feedback can be both minimized and maximized. When evaluating the quality, the resulting indicators are the signal-to-noise ratio - *SN*, which characterize the scattering (dispersion) or distribution of the resulting indicator relative to the nominal value (not considered here).

We will also not consider here exactly how Taguchi orthogonal arrays are generated. For this purpose, you can use special software, tables, etc. Arrays are denoted as follows: $OA_N(s^m)$ where N is the number of independent tests in the array, s is the number of levels, m is the number of factors. If a different number of levels is set for different factors, the notation will look like: $\llbracket OA_N\left(s_1^{m_1} \times s_2^{m_2} \ldots\right)$. Levels are more often denoted by the sequence 1, 2, 3,..., s or sometimes 0, 1, 2, 3, ..., s-1.

Some of the OA with the same number of levels for all factors are shown in Table 10.8. Examples of OA with different number of levels are shown in Table. 10.9, 10.10. It should be noted that any OA that is suitable for the number of levels can be used for a smaller number of factors, discarding the extra columns.

Table 10.8 – Examples of orthogonal arrays with equal number of levels

$OA_4(2^2)$

N	X_1	X_2
1	1	1
2	1	2
3	2	1
4	2	2

$OA_4(2^3)$

N	X_1	X_2	X_3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

$OA_8(2^4)$

N	X_1	X_2	X_3	X_4
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

$OA_9(3^2)$

N	X_1	X_2
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3

$OA_9(3^3)$

N	X_1	X_2	X_3
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

$OA_8(2^4)$

N	X_1	X_2	X_3	X_4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

$OA_{16}(4^2)$

N	X_1	X_2
1	1	1
2	1	2
3	1	3
4	1	4
5	2	1
6	2	2
7	2	3
8	2	4
9	3	1
10	3	2
11	3	3
12	3	4
13	4	1
14	4	2
15	4	3
16	4	4

$OA_{16}(4^3)$

N	X_1	X_2	X_3
1	1	1	1
2	1	2	2
3	1	3	3
4	1	4	4
5	2	1	2
6	2	2	1
7	2	3	4
8	2	4	3
9	3	1	3
10	3	2	4
11	3	3	1
12	3	4	2
13	4	1	4
14	4	2	3
15	4	3	2
16	4	4	1

$OA_{16}(4^4)$

N	X_1	X_2	X_3	X_4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	1	4	4	4
5	2	1	2	3
6	2	2	1	4
7	2	3	4	1
8	2	4	3	2
9	3	1	3	4
10	3	2	4	3
11	3	3	1	2
12	3	4	2	1
13	4	1	4	2
14	4	2	3	1
15	4	3	2	4
16	4	4	1	3

Table 10.9 – Orthogonal Array $OA_{18}(6^1 \times 3^6)$

N	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	1	1	1	1	1	1	1
2	1	2	2	2	2	2	2
3	1	3	3	3	3	3	3
4	2	1	1	2	2	3	3
5	2	2	2	3	3	1	1
6	2	3	3	1	1	2	2
7	3	1	2	1	3	2	3
8	3	2	3	2	1	3	1
9	3	3	1	3	2	1	2
10	4	1	3	3	2	2	1
11	4	2	1	1	3	3	2
12	4	3	2	2	1	1	3
13	5	1	2	3	1	3	2
14	5	2	3	1	2	1	3
15	5	3	1	2	3	2	1
16	6	1	3	2	3	1	2
17	6	2	1	3	1	2	3
18	6	3	2	1	2	3	1

Table 10.10 - Orthogonal Array $OA_{18}(2^1 \times 3^7)$

N	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	3	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1

The purpose of analyzing the results of the Taguchi experiment is primarily to find answers to three key questions:

- What conditions are optimal?
- What factors influence the variability of the results and how much?
- What will be the expected result under optimal conditions and how strongly each factor affects the improvement of results

To answer these questions, based on the experimental results and the orthogonal array itself, the average effect of each factor for each of the levels is calculated. The calculation of the average effect of factor $X_i$at the level of l - $\bar{\bar{X}}_i^l$ is performed as follows:

$$\bar{\bar{X}}_i^l = \frac{1}{r\{l \mid X_i = l\}} \sum_{\substack{j=1 \\ \{j \mid X_{ij}=l\}}}^{N} Y_j \qquad (10.19)$$

Here $r\{l \mid X_i = l\}$is the number of levels l in the corresponding column X_i

$\sum^{N}_{\substack{j=1 \\ \{j \mid X_{ij}=l\}}} Y_j$- the sum of the values of the resulting indicators Y for which the level in the corresponding row j of the column $X_i$is equal to l

The optimal solution will be a combination of levels of factors that provide minimum or maximum effects (depending on the need). Consider an example of calculating average effects.

Example

Consider the plan of the $OA_4(2^2)$ experiment and its results (Table 10.11). The resulting indicator Y is obtained from the results of three parallel tests ($n = 3$) and should be minimized.

Table 10.11 - Taguchi DOE OA4(23) and its results

N	Factors			Results
	X_1	X_2	X_3	Y_j
1	1	1	1	30
2	1	2	2	25
3	2	1	2	34
4	2	2	1	27

Let's calculate the effects of factors for each of the levels according to the formula (10.18):

$$\bar{\bar{X}}_1^1 = \frac{1}{2}(Y_1 + Y_2) = \frac{1}{2}(30 + 25) = 27{,}5$$

$$\bar{\bar{X}}_1^2 = \frac{1}{2}(Y_3 + Y_4) = \frac{1}{2}(34 + 27) = 30{,}5$$

$$\bar{\bar{X}}_2^1 = \frac{1}{2}(Y_1 + Y_3) = \frac{1}{2}(30 + 34) = 32{,}0$$

$$\bar{\bar{X}}_2^2 = \frac{1}{2}(Y_2 + Y_4) = \frac{1}{2}(25 + 27) = 26{,}0$$

$$\bar{\bar{X}}_3^1 = \frac{1}{2}(Y_1 + Y_4) = \frac{1}{2}(30 + 27) = 28{,}5$$

$$\bar{\bar{X}}_3^2 = \frac{1}{2}(Y_2 + Y_3) = \frac{1}{2}(25 + 34) = 29{,}5$$

On the basis of the performed calculation, diagrams of effects are built (Fig. 10.4). Based on the minimization condition, we choose the values of the factor levels that provide the minimum average effect - Table. 10.12

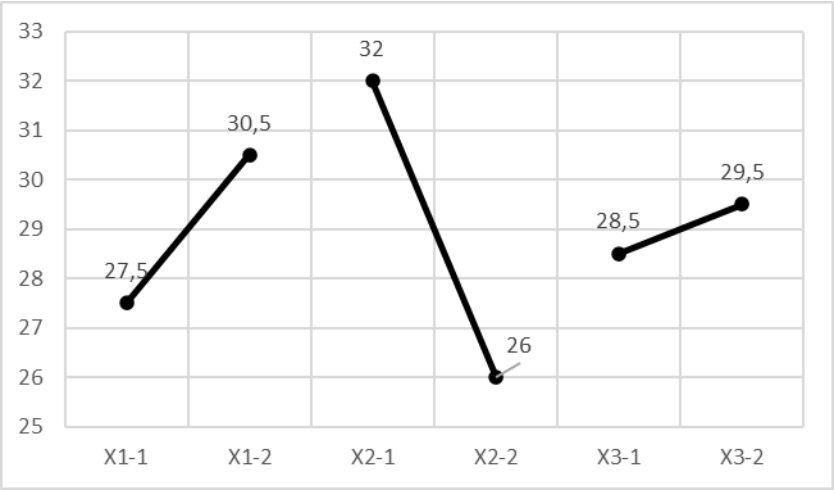


Figure 10.4 – Diagram of effects

Table 10.12 – Levels of factors with minimum average effect

Factor	X_1	X_2	X_3
Level	1	2	1

As can be seen from Table. 10.12 the levels of the factors do not coincide with any combination of conditions of tests defined by OA in Table. 10.11. In this case, according to the Taguchi concept, it is absolutely necessary to conduct a verification experiment under the conditions specified in Table. 10.12 to confirm or refute the obtained result.

In addition, according to the Taguchi method, using the obtained results, a specific dispersion analysis - ANOVA (Analysis of variance) - is not considered here, to assess the degree of influence of factors on the response. In addition, DOE according to the Taguchi concept allows determining the interaction of factors, as well as the size of the tolerance for the factor under investigation, based on the known costs for correcting the defect (not considered here).

QUESTIONS FOR SELF-TESTING

1. What is the sequence of finding the optimal solution in experimental research?
2. What is interpolation?
3. What is extrapolation?
4. What types of models are accepted when planning an experiment?
5. What type of model is used in a full factorial experiment?
6. What is the purpose of the model in the optimization of the full factorial experiment?
7. What is a "level"?
8. What is a variation interval?
9. How do you get the "upper level"?
10. How do you get the "lower level"?
11. What is "basic" or "zero" level?
12. How is the design (matrix) of a full factorial experiment constructed?
13. How are the values of regression coefficients determined in a full factorial experiment?
14. What is a priori information?
15. What expression describes the interaction of factors?
16. How is the order of interaction of factors related to the probability of its occurrence?
17. How is the number of tests determined in a full factorial experiment?
18. What actions of the researcher are expected when entering the almost stationary zone?
19. What is accepted as a model when designing an experiment according to the OCCD?
20. How is the number of experiments determined when designing an experiment under the OCCD?
21. What is the structure of the OCCD?
22. In what case is it necessary to design an experiment according to the OCCD?
23. What ensures the orthogonality of the experimental design?

24. What is the advantage of Taguchi's designs compared to a full factorial experiment?
25. What are the features of Taguchi's designs?
26. What are internal array factors?
27. What are external array factors?
28. What is Taguchi's DOE used for?
29. What is the design of the Taguchi experiment?
30. How are Taguchi's designs designated?
31. What is the purpose of analyzing the results of the Taguchi experiment?
32. How is the average factor effect according to Taguchi determined?
33. Is a model built when designing an experiment according to Taguchi, and if so, in what form?
34. What decision is made based on the results of calculating the average effects of factors?
35. What should be done if the combination of levels of the desired effects does not coincide with any of the designed tests?

LITERATURE

1. Методи одновимірної оптимізації: практикум з дисципліни «Дослідження операцій» [Електронний ресурс]: навч. посіб. для студ. спеціальності 113 «Прикладна математика», спеціалізації «Наука про дані та математичне моделювання» / Т. С. Ладогубець, О. Д. Фіногенов; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 453 Кбайт). – Київ: КПІ ім. Ігоря Сікорського, 2020. – 47 с.
2. Пляскин, И. И. Оптимизация технических решений в машиностроении // И.И. Пляскин. – М.: Машиностроение, 1982. – 176 с.
3. Roy, Ranjit K. A primer on Taguchi method [Текст] // Ranjit K Roj. – Dearborn, MI: SME, 2010 – 336 p.
4. Kaker, Raghu N. Taguchi's orthogonal arrays are classical designs of experiments [Текст] // Raghu N Kaker, Eric S. Lagergren, James L. Filliben / Journal of Research of the National Institute of Standards and Technology. – 1991. – Vol.96, № 5. – P.577 – 591.