# Chapter 2
# Optimization Methods and Their Efficient Use

**Abstract** The main scientific interests of V. S. Mikhalevich were connected with investigations in optimization theory and system analysis and with the development of newest computer technologies and computer complexes, creation of scientific bases for the solution of various problems of informatization in the fields of economy, medicine, and biology, design of complicated processes and objects, environmental research, and control over important objects. An analysis of the results of investigations along these directions is the subject matter of this chapter. Mikhalevich's great interest in the development of system analysis methods is shown. They began to be developed with his active participation at the Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine, International Institute for Applied Systems Analysis (Austria), National Technical University of Ukraine "Kyiv Polytechnic Institute" (NTUU "KPI"), Taras Shevchenko National University of Kyiv, and other scientific and educational establishments of Ukraine. The problematics of this important line of investigations became especially topical in recent decades in connection with investigations of complicated processes of international cooperation in economy, problems of prediction and prevision of possible development of a society under conditions of environmental contaminations, and investigation of other complicated processes from the position of system analysis. In investigating such processes, an important role is played by the Institutes of the Cybernetic Center of NAS of Ukraine and also by the above-mentioned International Institute for Applied Systems Analysis whose creation was directly connected with V. S. Mikhalevich. At present, many scientists from Ukraine and many other countries cooperate with this institute. This chapter elucidates the history of creation of this institute, principles of its work, and the method of financing of important projects. Some important investigations of Ukrainian scientists are also considered that were performed during recent 15–20 years and that consisted of the development of new methods of stochastic and discrete optimization and the further development of the scheme of the method of sequential analysis of variants.

I.V. Sergienko, *Methods of Optimization and Systems Analysis for Problems of Transcomputational Complexity*, Springer Optimization and Its Applications 72,
DOI 10.1007/978-1-4614-4211-0_2, 

## 2.1 General Algorithm of Sequential Analysis of Variants and Its Application

Since the late twentieth century, achievements in exact sciences have been applied more and more widely to support managerial (economic, design and engineering, or technological) decisions. They help experts to understand the deep interrelation of events and to make the analysis and prediction more reliable. This allows public authorities, design and engineering organizations, and individual economic entities to find efficient decisions in complicated and ambiguous situations, which are so numerous nowadays. These applications became most popular in the 1940s. That was promoted by the rampant development of applied mathematics and computer science; nevertheless, the successful use of such tools required developing formal (including mathematical) decision-support methods, among which optimization methods occupy a highly important place. Since the late 1950s, the Institute of Cybernetics has played a leading role in the development of such methods. In many respects, credit for this goes to V. S. Mikhalevich, one of the most prominent scientists of the institute, an academician, and the director of the institute (from 1982 to 1994) who headed the studies in this field.

Mikhalevich's first scientific studies [24, 25, 100] were concerned with empirical distribution functions. The studies he carried out later, in Moscow, dealt with sequential procedures of statistical analysis and decision making [98, 103], which was a new field initiated by A. Wald [13], who proposed sequential models for making statistical decisions, taking the cost of experiment into account. Mikhalevich applied Wald's ideas to practically important acceptance sampling problems in Bayesian formulation where the probability distribution of rejects is known a priori. Different results of experiments lead to different a posteriori distributions, and it is possible to compare, at each step, the average risk of continuing experiments considering their cost and the risk of making a certain decision at the same step. The assessment of the average risk involves complex recurrent relations similar to the Bellman–Isaacs dynamic programming equation [11] (note that there was hardly any relevant American scientific literature at that time). Mikhalevich succeeded in using the concept of "reduced" (in the number of steps) procedures to derive pioneering results in acceptance sampling and reported these results in the paper [103], which actually outlined the key results of his Ph.D. thesis [104]. At the Kyiv State University, he wrote the theoretically interesting paper [99] where he treated continuous-time stopping problems at a high mathematical level. These problems were later developed by A. N. Shiryaev [182] and his disciples.

At the university, Mikhalevich continued his investigations into acceptance sampling problems. Since determining the decision-making limits in a significant number of sequential steps required labor-intensive calculations, he took an interest in the capabilities of the MESM computer, which had recently been launched in Feofaniya (a Kyiv suburb). To test his method of reduced Bayesian strategies [104], Mikhalevich offered several graduate students to develop, as a practical training, MESM programs that would implement his sequential decision-making algorithm in acceptance sampling problems. Some of these students were O. M. Sharkovskii and N. Z. Shor. This training helped Shor to get a job assignment to the Computer Center of the Academy of Sciences of the USSR after graduating from the Kyiv University in 1958.

In 1958, V. M. Glushkov, a new director of the Computer Center, offered Mikhalevich to head a group of experts in probability theory and mathematical statistics to be engaged in studying the reliability of electronic devices and operations research. That was the beginning of the Kyiv optimization school. At first, Mikhalevich worked in this field together with Bernardo del Rio (a senior lecturer of the Rostov Institute of Railway Transport Engineers, invited by V. M. Glushkov to work in the automation of processes in railway transport) and three young experts Yu. Ermoliev, V. Shkurba, and N. Shor.

From its establishment by Mikhalevich, the Department of Economic Cybernetics was a source of experts in optimal planning, management of the national economy, operations research, design of complex structures and systems, simulation and automation of transport processes, etc. In 1960–1962, more than 100 experts from different regions of the USSR (including Yakutia, Irkutsk, Transcaucasia, Central Asia, and Moldova) were trained at the department; some of them remained to work there.

From the outset, the theoretical developments of the department were motivated by the necessity of solving problems in optimal planning and design.

In developing numerical algorithms to solve some technical and economic extremum problems, Mikhalevich noticed that the ideas of the theory of sequential statistical decisions were useful in analyzing a rather wide class of multivariate optimization problems and in selecting optimal and suboptimal algorithms to solve them. The theory of statistical decisions, including the sequential-search algorithm [101, 103] developed by Mikhalevich, introduces the concept of realization of a trial as a random event statistically related to a set of individual objects with respect to which a decision is sought and its quality is assessed. The realization of a trial results in the redistribution of the probability measures of quantities and attributes that characterize individual objects. The selection of a rational algorithm to search for an acceptable decision is based on sequential design of experiments depending on the previous realizations of trials. In considering deterministic multivariate problems, one may encounter the degenerate yet rather important case where, after realization of trials, some subset of individual objects takes on the measure 0, that is, it can no longer be considered an acceptable variant. This special case was called "algorithm of sequential analysis of variants." It was presented for the first time at the fourth All-Union Mathematical Congress in July 1961 [119] and later in the joint publication of Mikhalevich and Shor [120]. This algorithm quickly became commonly accepted and was widely used by academician N. N. Moiseev and his followers at the Computing Center of the Academy of Sciences of the USSR [126] and by Belarussian Professors V. S. Tanaev and V. A. Emelichev [43, 168].

Mikhalevich's closest followers N. Z. Shor and V.V. Shkurba and their colleagues L. A. Galustova, G. A. Donets, A. N. Sibirko, A. I. Kuksa, V. A. Trubin, and others contributed significantly to the theoretical development of the algorithm of sequential analysis of variants and its application to solve economic and design and engineering problems. Sequential optimization algorithms are also described in the joint monograph by V. S. Mikhalevich and V. L. Volkovich [105]. The paper [139] describes sequential algorithms for solving mixed linear-programming problems with preliminary rejection of redundant constraints and variables that take only zero values in the optimal solution.

Professor V. L. Volkovich and academician V. S. Mikhalevich

A rather comprehensive and concentrated description of the basic ideas behind sequential decision making can be found in [101, 102], which underlie Mikhalevich's Dr.Sci. thesis and in [17] he edited.

From the standpoint of formal logic, the algorithm of sequential analysis of variants can be reduced to repeating the following sequence of operations:

– Divide the sets of possible variants of problem solutions into a family of subsets, each having additional specific properties.
– Use these properties in search for logical contradictions in the description of individual subsets.
– Omit from consideration the subsets whose description is logically contradictory.

The procedure of sequential analysis of variants implies generating variants and selecting operators for their analysis in such a way that parts of variants are rejected once they have been discovered to be unpromising, even if not completely generated. Rejecting unpromising parts of variants also rejects the sets of their continuation and thus saves much computing resources. The more specific properties are used, the greater the savings.

The algorithm of sequential analysis of variants includes the above-mentioned generalization of the well-known Bellman's optimality principle to a wider class of optimization problems.

In what follows, we will formulate this principle and give a methodological procedure free from some constraints inherent in dynamic programming.

Consider some basic set $X$. Denote by $P(X)$ a set of finite sequences of the form $p = \{x_1, \ldots, x_{K_p}\}, \quad x_i \in X, \quad 1 \le i \le K_p$. Choose some subset of admissible sequences $W(X) \subseteq P(X)$ in this set and a subset of complete admissible sequences $\overline{W} \subseteq W(X)$ in $W(X)$.

Consider a sequence $p$. Let its $l$-initial interval be a sequence of the form $p_l(x_1, \ldots, x_l), \ 1 \le l \le K_p$, and its $q$-final interval be a sequence of the form $p^q = (x_q, x_{q+1}, \ldots, x_{K_p}), \ 1 \le q \le K_p$. If $q = l + 1$, then the respective parts of $p$ are called conjugate.

Consider two arbitrary admissible sequences $p_1$ and $p_2$. In $p_1$, choose the $l_1$-initial interval $p_{1l_1}$ and conjugate final interval $p_1^{l_1+1}$; in $p_2$, choose the $l_2$-initial interval $p_{2l_2}$ and conjugate final interval $p_2^{l_2+1}$.

A functional $\Phi$ defined on the set $W(X)$ is called monotonically recursive if the truth of $p_{1l_1} \in W, \ p_{2l_2} \in W, \ p_1^{l_1+1} \equiv p_2^{l_2+1}$, and $\Phi(p_{1l_1}) < \Phi(p_{2l_2})$ implies the fulfillment of $\Phi(p_1) < \Phi(p_2)$. Denote $\Phi^* = \sup\{\Phi(p) | p \in \overline{W}\}$. A sequence $p^* \in \overline{W}$ is called maximum if $\Phi(p^*) = \Phi^*$.

Consider an admissible sequence $p$. The subset consisting of elements for which $p$ is the initial interval is called a $p$-generic set. The set of all final intervals of elements of the $p$-generic set conjugate to $p$ is called the set of continuations $P(p)$. The statement below is true [17].

**Theorem 2.1.** *Given a monotonically recursive functional $\Phi$ and two admissible sequences $p_1$ and $p_2$ such that $\Phi(p_1) < \Phi(p_2)$ and $P(p_1) \subseteq P(p_2)$, the elements of the set $P(p_1)$ cannot be maximum sequences.*

Most applications of the algorithm of sequential analysis of variants to be discussed below are based on the generalized optimality principle.

In the 1960–1970s, the Soviet Union actively developed the transport infrastructure, including pipeline networks to export oil and natural gas, created the Unified Energy System, built large industrial facilities (primarily, for the manufacturing and defense industries), and greatly developed air transport. This rapidly expanded the scope of design work and substantially raised the design quality standards, which stimulated the introduction of mathematical design methods. The Institute of Cybernetics, as one of the leading scientific centers, could not keep out of these research efforts. The algorithm of sequential analysis of variants and the generalized optimality principle were applied to solve optimal-design problems for roads [118, 140] and electric and gas networks [121] to find the shortest network routes [6] and critical routes in network diagrams, to solve industrial location problems [118], scheduling problems [117], and some other discrete problems [120].

Let us discuss in more detail some of the classes of optimization models that arose in these applications and the peculiarities of using the mathematics mentioned above.

In the early 1960s, the USSR carried out large-scale projects related to the development of various transportation networks, including power transmission lines and pipelines. Because of the large-scale and limited resources for such projects, it became important to enhance the efficiency of network designs. The complexity of design problems, the necessity of allowing for numerous aspects, and high-quality requirements necessitated the use of mathematical methods to solve such problems.

Problems of designing transport networks implied choosing the network configuration, network components, and their characteristics from a finite set of admissible elements defined by operating engineering standards. The network should be designed to transport a certain product from suppliers to consumers so as to keep the balance between production and consumption and to minimize the total costs of creating and maintaining the network. These requirements can be formalized as an integer optimization problem with recurrent constraints. The objective function and the right-hand sides of the constraints in such a problem are separable functions with linear and nonlinear components. Problems of such type (but less complex) were considered in the fundamental work [11] of R. Bellman, the founder of dynamic programming, which made it expedient to apply the algorithm of sequential analysis of variants to solve them.

Optimization models began to be used to design electric networks in the USSR in the early 1960s. In 1961, a representative of the Ministry of Energy of the USSR organized a relevant meeting and invited representatives of the Institute of Cybernetics (then the Computer Center), Institute of Automation, Polytechnic Institute, and VNDPI "Sil'elektro." Each of the institutes was assigned to develop a technique for the optimization of rural electric networks. The Institute of Cybernetics achieved the most success by focusing on solving the following tasks:

1. Calculation of optimal parameters for 10-kV networks using digital computing machines
2. Computation of optimal parameters for 35-, 10-, and 0.4-kV networks
3. Selection of an optimal configuration of electric networks
4. Selection of optimal parameters considering an optimal arrangement of transformer substations

Computational methods for the first two problems were based on the sequential analysis of variants. For the third problem, L. A. Galustova proposed a heuristic method, which also employed some ideas of the sequential analysis of variants but was not strongly valid mathematically. Nevertheless, this method demonstrated its efficiency during computations based on real data.

These subjects were developed for several years. A series of experimental computations were performed in cooperation with "Sil'elektro," the savings being 10–20 %. The report, together with an implementation and savings certificate, was delivered to the customer and to the Presidium of the Academy of Sciences of the USSR. The above-mentioned methods were also used to design Latvian electric networks (with participation of Ionas Motskus). The results of this work were summarized in the book *A Technique for Optimal Design of Rural Electric Networks* edited by Mikhalevich and in popular science publications.

Almost simultaneously with the above-mentioned research, Gazprom of the USSR together with the Ministry of Energy initiated and guided the automation of gas pipeline design with the participation of the Institute of Cybernetics, VNDI “Gazprom” (Moscow), and VNDPI “Transgaz” (Kyiv).

Several tasks were assigned to the executors.

The first task was to perform a feasibility study for large-diameter gas pipelines, namely, to determine their parameters such as length and diameter and to choose the type and operating modes of engines for gas-compressor stations. The modification of the existing gas pipelines was also an issue: what pipelines should be added to those already laid, what operating conditions of gas pipelines should be, etc. A databank of all the gas pipelines of the Soviet Union was created at the Institute of Cybernetics. Other organizations were also engaged in the design of gas pipelines; however, they tested their algorithms against tentative examples, while the Institute of Cybernetics used real data, which made it possible to refine algorithms by making better allowance for the features of problems.

The second task was to choose optimal parameters of a gas pipeline, taking into account the dynamics of its development. Precedently, the dynamics of investment in the development of gas pipelines and installation of compressors and the possibility of equipment conversion were disregarded. For example, in accomplishing this task, the possibility of using operable overage aircraft engines as compressors was analyzed. A great many engines were expected to arrive after the planned reequipment of the Soviet Army in the 1970s. Not only current but also 5- to 10-year investments were determined. The studies carried out at the Institute of Cybernetics justified that the diameter of gas pipes should be increased from 70 to 100 cm.

These developments were discussed at a top-level meeting in the Kremlin, Moscow. Directors of branch institutes were present. Institute of Cybernetics was represented by L. A. Galustova who demonstrated calculations made at the institute and compared them with American assessment of the development of gas pipelines in the Midwest of the USA. The results appeared similar.

The third major task was as follows: Given the powers of different distributions of gas flows among branches of a gas pipeline, find the optimal distribution for a single gas supply network. The objective function in this problem was nonlinear; to find its extremum, subgradient methods to be discussed in the next section were applied. The calculations were performed for VNDPI “Transgaz” (Kyiv), VNDI “Gazprom” (Moscow), and “Soyuzgazavtomatika.” Comparing this project with the previous ones showed that the introduction of optimization modeling saved 50–60 million rubles at 1960 values since the development of the gas-supplying system of the entire Soviet Union was considered. Gazprom refused to sign an act for such an amount since they were afraid that the Academy of Sciences would take away some of the money. As a compromise, an act for 20 million rubles was signed.

This approach was also followed in the 1960–1970s to design oil and product pipelines. As N. I. Rosina (a former employee of the Institute of Cybernetics who took an active part in the above projects) recollects, the algorithm of sequential analysis of variants made it possible to obtain the following results.

1. Mathematical models and computational algorithms were developed to optimize the arrangement of oil pumping stations along oil trunk pipelines and to determine the optimal pipe diameter for separate segments of an oil pipeline, considering the planned (or predicted) volume of oil pumped and the prospects for the development of oil transportation system. These models and algorithms were used to design the Anzhero-Sudzhensk–Krasnoyarsk–Irkutsk trunk pipeline and the oil pipelines that connected Ukrainian oil refineries with the existing oil transportation system.
2. The algorithms were modified to solve problems of arrangement of auxiliary facilities such as cathodic protection stations (with selection of the type of station), equipment of radio relay communication lines, heat tracing stations for high-paraffin oil pipelines, etc. Problems of the arrangement of these objects were solved in combination with the problem of the arrangement of oil pumping stations, considering the possibility of arranging several objects on one site. This resulted in the following synergetic effect: The gain from solving a set of problems exceeded the sum of gains from solving individual problems. The problem of simultaneous arrangement of stations and auxiliary objects had a high dimension, which increased directly proportional to the number of alternative arrangements of each object as new types of objects were added. The total number of alternative arrangements of pipeline elements could reach many billions, that is, a problem of transcomputational complexity arose. Applying the algorithm of sequential analysis of variants reduced the number of alternatives to be analyzed by many orders of magnitude. This made it possible to develop algorithms that took low-performance Soviet computers used in the 1960–1970s a reasonable time to solve such a problem. These algorithms were used, for example, to design pipelines for oil transportation from the Prikaspiiskii region to the European part of the USSR.
3. The algorithm of sequential analysis of variants was used to develop mathematical models and algorithms to design networks of oil pipelines and bulk plants. These algorithms assumed two-stage design. At the first stage, linear optimization problems similar to the transportation problem were solved to assign consumers to bulk plants and to determine their capacity. The solution was then corrected, if possible, to allow for weakly structured aspects disregarded by the model (such as permanent connections between consumers and suppliers). The second stage was to determine the configuration of the oil pipeline network and locations of pumping stations and other infrastructure components as well as oil pipeline diameters. The algorithms used for this purpose were based on the algorithm of sequential analysis of variants. This approach was followed to design oil pipeline networks in Ukraine, Kazakhstan, Byelorussia, and Uzbekistan. The studies were carried out at the “Gipronaftoprovid” design

institute of national standing located in Kyiv in the 1960–1980s. As N. I. Rosina, who worked at this institute, recollects, the software created during the research was successfully used until the early 1990s.

4. Methods of sequential analysis of variants were used at this institute to formulate recommendations on pipeline cleaning frequency and relevant equipment.

Thus, applying the algorithm of sequential analysis of variants made it possible to resolve challenges that arose in designing the transport infrastructure and in identifying the ways of its development.

Along with methods of sequential analysis of variants, decision-support methods were rapidly developed and implemented at the Institute of Cybernetics in the 1960s. They leaned upon the solution of lattice problems, a special class of discrete optimization problems such as the search for critical (shortest and longest) paths, minimum cut, and maximum flow. Solution algorithms for such problems employed Bellman's optimality principle whose relation to the algorithm of sequential analysis of variants was discussed above. It is no accident that network optimization problems drew the close attention of researchers of the Institute of Cybernetics.

In 1963, Mikhalevich was appointed a USSR-wide coordinator of the introduction of network planning and management systems in the field of mechanical engineering, defense, and construction. The results of these studies promoted the implementation of network design systems for managing many large projects and construction of important complex objects.

To search for critical paths, use was made of an algorithm similar to search for the shortest path on a lattice [6] and based on the ideas of sequential analysis of variants.

In 1966, Mikhalevich organized the first All-Union conference on mathematical problems of network planning and management, which initiated many new studies in this field. The conference stimulated large practical developments in the USSR, in which Mikhalevich participated as a scientific adviser. These were applications in the defensive industry (e.g., in planning the launch of the first Soviet nuclear-powered submarine), to support managerial decision making during the construction of petrochemical integrated plants, blast furnaces, sections of Baikal-Amur Mainline (BAM), in stockpile production of shipbuilding yards, etc. Chapter 5 will detail these applications and further development of the methods to solve extremum problems on graphs.

V. V. Shkurba applied the ideas of sequential analysis of variants to solve ordering problems such as optimal arrangement of machine tools in processing centers. These studies gave an impetus to the cooperation with Byelorussian experts in scheduling theory, which resulted in the monograph [168], a classical manual in this field. V. V. Shkurba, who supervised the development of software for solving scheduling problems under limited resources, participated, together with his disciples, in creation of automated production management systems ("Lvov," "Kuntsevo," etc.) [135, 136, 183, 194].

The book [183] analyzes scheduling theory as a division of discrete optimization, classifies relevant problems, and develops exact methods to solve some special problems in scheduling theory and approximate methods to solve the general scheduling problem. Most of these methods were based on the algorithm of sequential analysis of variants. This study was the first to combine methods of applied mathematics and computer technology to solve industrial scheduling and management problems. It also proposed methods for exact solution of some classes of scheduling problems (such as "one machine," "two machine," and "three machine" problems) and methods for approximate solution of more general classes of scheduling problems.

Note that exact methods to solve scheduling problems are efficiently applicable only to some special classes of problems depending on the constraints imposed on the ways of performing elementary operations and on the availability of resources, whose number is no greater than three. It was proved that for $l > 3$, the mathematical complexity of scheduling problems qualitatively increases. The main approach to their solution is to develop approximate methods based on the information simulation of production processes and the wide use of various heuristic methods. Efficient means for the solution of optimization production scheduling problems based on heuristic algorithms, which substantially enhance the quality of production management, were analyzed in the monograph [194]. It was the first to systematize and propose original approaches to estimating the efficiency of heuristic algorithms, which supplement exact methods, including those based on the algorithm of sequential analysis of variants. The ideas of scheduling, modeling of production processes, and management of a large modern industrial enterprise were developed in [136]. This study argued the necessity of implementing systems of management of a modern production association (large industrial enterprise) as multilevel hierarchical systems and outlined the main principles of their development. It systematized the results of improving production planning based on modern economic and mathematical methods, interactive technologies, and principles of system optimization and presented methods and procedure of solving control problems for all the considered levels of hierarchy.

A number of theoretical results related to the quantitative analysis of the complexity of methods of sequential analysis of variants, including network problems of allocation of limited resources, scheduling, and related problems, were published in the monograph [109].

The paper [151] was the first to analyze the class of scheduling problems that arise in the automation of complex production processes such as metal working during galvanization of parts. Approaches to the formalization of such problems and methods of their solution were proposed [152, 154].

These results were used to develop software for the "Galvanik" domestic computer-aided system, which is still widely used in the industry. This system was first launched at the "Arsenal" factory in Kyiv. It was shown many times at international exhibitions and earned gold medals and other high awards. By the way, its authors were awarded with the State Prize of Ukraine in Science and Technology. These results were also used to develop automated production management systems [135, 136, 183, 194]. In the 1970s, the studies were continued by the followers of V. S. Mikhalevich and V. V. Shkurba.

## 2.2 Solution of Production–Transportation Problems and Development of Methods of Nondifferential Optimization

In 1961, V. S. Mikhalevich came into contact with the Department of Transportation of the Gosplan (State Planning Committee) of the UkrSSR at which A. A. Bakaev (a leading specialist at that time) developed a number of models of integrated transportation of bulk cargoes (e.g., raw material from beet reception centers to sugar mills) whose totality was called *production–transportation problems* and which were of the following form.

Let there be $m$ points of production of some homogeneous product and $n$ points of its consumption. The production of the product unit at each point $i$ $(i = \overline{1,\ m})$ *requires* $a_{ki}\,(k = \overline{1,\ K})$ *resources of kind* $k$, where $K$ is the total amount of resources. The manufacturing cost of the product unit at the $i$th point equals $C_i$. The total amount of the $k$th resource $B_k$ of each kind $k$, the need for the $j$th product $b_j$ for each point of consumption $j\,(j = \overline{1,\ n})$, and the expenditures $C_{ij}$ for the transportation of the product unit from the $i$th point of production to the $j$th point of consumption are also known. It is necessary to determine the volume of production $x_i$ at each point $i$ and the volume of product traffic $y_{ij}$ from points of production to points of consumption so that the total cost of production and transportation is as small as possible

$$F = \sum_{i=1}^{m} C_i x_i + \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} y_{ij} \to \min \tag{2.1}$$

under the resource constraints

$$\sum_{i=1}^{m} a_{ki} x_i \le B_k,\ \ k = \overline{1,\ K}, \tag{2.2}$$

and the conditions

$$\sum_{i=1}^{m} y_{ij} \ge b_j,\ \ j = \overline{1,\ n}, \tag{2.3}$$

of satisfaction of all needs for the product and nonexceeding of the volume of removal of the product over the volume of its production

$$\sum_{j=1}^{n} y_{ij} \le x_i,\ \ i = \overline{1,\ m}, \tag{2.4}$$

to which the following conditions of the positiveness of variables of the problem are added:

$$x_i \geq 0, \quad i = \overline{1,\ m}, \tag{2.5}$$

$$y_{ij} \geq 0, \quad i = \overline{1,\ m},\ j = \overline{1,\ n}. \tag{2.6}$$

Problem (2.1)–(2.6) is a high-dimensional linear-programming problem with weakly filled matrix of constraints. This complicated the use of linear-programming methods, which were well known at that time, in particular, the simplex method, and required the creation of specialized algorithms to solve it. These algorithms were developed proceeding from the following considerations.

Let us consider the optimization problem

$$F_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} y_{ij} \to \min, \tag{2.7}$$

$$\sum_{i=1}^{m} y_{ij} \geq b_j,\ j = \overline{1,\ n}, \tag{2.8}$$

$$\sum_{j=1}^{n} y_{ij} \leq x_i, \quad i = \overline{1,\ m}, \tag{2.9}$$

$$y_{ij} \geq 0, \quad i = \overline{1,\ m},\ j = \overline{1,\ n}, \tag{2.10}$$

which is solved for a fixed $x_i$. It is a transport problem for which efficient algorithms were developed even in the 1940s; in particular, the method of potentials made it possible to simultaneously find solutions to the direct and dual problems. We denote by $v_j^*$ optimal values of dual variables corresponding to constraints (2.8) and by $u_i^*$ values of dual variables corresponding to constraints (2.9). According to the duality theorem, the optimal value of objective function (2.7) is equal to

$$F_1 = \sum_{i=1}^{m} u_i^* x_i - \sum_{i=1}^{n} v_j^* b_j.$$

Thus, instead of problem (2.1)–(2.6), the following problem can be considered:

$$F = \sum_{i=1}^{m} C_i x_i + \sum_{i=1}^{m} u_i^*(x) x_i - \sum_{i=1}^{n} v_j^*(x) b_j$$

under constraints (2.2) and (2.5), where $u_j^*$, $v_j^*$ is the optimal solution of the problem dual to problem (2.7)–(2.10); the solution is obtained for fixed $x_i$, $i = \overline{1,\ m}$.

The dimensionality of this problem is essentially smaller; nevertheless, its objective function is nonlinear and non-everywhere differentiable albeit convex. The need for the solution of such problems stimulated works on nonsmooth optimization at the Institute of Cybernetics. In particular, to solve production–transportation problems in network form, N. Z. Shor proposed the method of generalized gradient descent (called subgradient descent later on) in 1961, which is a simple algorithm that makes it possible to minimize convex functions with a discontinuous gradient.

Let us consider an arbitrary convex function $f(x)$ defined on a Euclidean space $E^n$. Let $X^*$ be a set of minima (it may be empty), and let $x^* \in X^*$ be an arbitrary minimum point; $\inf f(x) = f^*$. A subgradient (a generalized gradient) of the function $f(\bar{x})$ at a point $x \in E^n$ is understood to be a vector $g_f(\bar{x})$ such that the following inequality is fulfilled for all $x \in E^n$ :

$$f(x) - f(\bar{x}) \geq (g_f(\bar{x}),\ x - \bar{x}).$$

It follows from the definition of a subgradient that the following condition is satisfied when $f(x) < f(\bar{x})$ :

$$(-g_f(\bar{x}),\ x - \bar{x}) > 0.$$

Geometrically, the latter formula means that the antisubgradient at the point $\bar{x}$ makes an acute angle with any straight line drawn from the point $\bar{x}$ in the direction of the point $x$ with a smaller value of $f(x)$. Hence, if the set $X^*$ is nonempty and $x \notin X^*$, then the distance to $X^*$ decreases with the movement from the point $\bar{x}$ in the direction of $-g_f(\bar{x})$ with a rather small step. This simple fact underlies the subgradient method or the method of generalized gradient descent.

The method of generalized gradient descent (GGD) is understood to be the procedure of construction of a sequence $\{x_k\}_{k=0}^{\infty}$, where $x_0$ is an initial approximation and $x_k$ are computed by the following recursive formula:

$$x_{k+1} = x_k - h_{k+1} \frac{g_f(x_k)}{\|g_f(x_k)\|}, \quad k = 0,\ 1,\ 2, \ldots; \tag{2.11}$$

where $g_f(x_k)$ is an arbitrary subgradient of the function $f(x)$ at a point $x_k$ and $h_{k+1}$ is a step multiplier. If $g_f(x_k) = 0$, then $x_k$ is a minimum point of the function $f(x)$, and the process comes to an end.

The most general result on the convergence of GGD is contained in the following theorem [186].

**Theorem 2.2.** *Let $f(x)$ be a convex function defined on $E^n$ with a bounded region of minima $X^*$, and let $\{h_k\}$ $(k = 1,\ 2, \ldots)$ be a sequence of numbers that satisfies the conditions*

$$h_k > 0; \quad \lim_{k \to \infty} h_k = 0; \quad \sum_{k=1}^{\infty} h_k = +\infty.$$

Then, for a sequence $\{x_k\}$ $(k = 1, 2, \ldots)$ constructed according to formula (2.11) and for an arbitrary $x_0 \in E^n$, there are the following two possibilities: Some $k = \bar{k}$ can be found out such that we have $x_{\bar{k}} \in X^*$ or

$$\lim_{k\to\infty} \min_{y\in X^*} \|x_k - y\| = 0, \quad \lim_{k\to\infty} f(x_k) = \min_{x\in E^n} f(x) = f^*.$$

The conditions imposed on the sequence $\{h_k\}$ $(k = 1, 2, \ldots)$ in the theorem are well known as classical conditions of step regulation in the GGD. Under definite additional assumptions, GGD variants were obtained that converge at the rate of a geometric progression. For them, a significant part was played by upper bounds of angles between the antigradient direction at a given point and the direction of the straight line drawn from this point to the minimum point.

Further investigations on the development of subgradient methods of nonsmooth optimization were pursued at the Department of Economic Cybernetics of the Institute of Cybernetics. Questions of minimization of non-everywhere differentiable functions [186] arise in solving various problems of mathematical programming when decomposition schemes are used, maximum functions are minimized, exact methods of penalty functions with nonsmooth "penalties" are used, and dual estimates in Boolean and multiextremal problems are constructed. They also arise in solving optimal planning and design problems in which technical and economic characteristics are specified in the form of piecewise-smooth functions.

The method of generalized gradient descent was first used for the minimization of piecewise-smooth convex functions that are used in solving transport and production–transportation problems [189] and then for the class of arbitrary convex functions [116] and convex programming problems. A substantial contribution to the substantiation and development of this method was made by Yu. M. Ermoliev [52] and B. T. Polyak [138], a Moscow mathematician. Ideas of nonsmooth convex analysis were developed in works of B. N. Pshenichnyi, V. F. Demyanov, I. I. Eremin, T. Rokafellar (USA), and other scientists [34, 45, 46, 144, 145, 148]. It should be noted that the interest in the development of subgradient methods in the West was deepened approximately starting with 1974 when investigators saw the key to solving high-dimensional problems in them [213].

However, it turned out that there is a wide range of functions for which the convergence of the subgradient method is weak. This is especially true for the so-called ravine functions. The reason is that the antigradient direction makes an angle with $-(x - x^*)$ that is close to $\frac{\pi}{2}$ and the distance to $X^*$ changes insignificantly with movement in this direction. "Ravine" functions occur rather often in applied problems.

Let us consider, for example, the "ravine" function $F(x) = x_1^2 + 100x_2^2$. After the replacement of the variables $y_1 = x_1$ and $y_2 = 10x_2$, we obtain the new function $F(y) = y_1^2 + y_2^2$, which is very handy for applying the gradient method (the above-mentioned angle is always equal to zero for it). Note that, in most cases, such an obvious replacement is absent. Moreover, some replacement that can improve properties of the function in the vicinity of a point $x$ can worsen them outside of this vicinity. But, in solving a problem with the help of the gradient method, there may be some points at which some replacement will worsen such a function. Thus, it is necessary to develop a method that would allow one to construct a new coordinate system for each iteration and, in case of need, to quickly come back to the initial system. To this end, linear nonorthogonal transformations can be used such as the operation of dilation of an $n$-dimensional vector space.

Let $x \in E^n$ be an arbitrary vector, and let $\xi$ be an arbitrary direction, $||\xi|| = 1$. Then, the vector $x$ can always be represented in the form

$$x = \gamma(x,\ \xi)\xi + d(x,\ \xi),$$

where $\gamma(x,\ \xi)$ is some scalar quantity and $(d(x,\ \xi),\ \xi) = 0$.

The operator of dilation of a vector $x$ in the direction $\xi$ with a coefficient $\alpha > 0$ is understood to be as follows:

$$R_\alpha(\xi)x = \alpha\gamma(x,\ \xi)\xi + d(x,\ \xi) = \alpha(x,\ \xi)\xi + x - (x,\ \xi)\xi = x + (\alpha - 1)(x,\ \xi)\xi.$$

This operation changes the part $\gamma(x,\ \xi)\xi$ (collinear with $\xi$) of the vector $x$ by a factor of $\alpha$ and does not change the part $d(x,\ \xi)$ (orthogonal to $\xi$) of this vector. If this operation is applied to all the vectors that form the space $E^n$, then we say that the space is extended with a coefficient $\alpha$ in the direction $\xi$.

For "ravine" functions, the direction of the generalized gradient is almost perpendicular to the axis of the "ravine" at the majority of points. Therefore, the space dilation in the direction of the generalized gradient decreases the "raviness" of the objective function. This implies the following general idea of the method: Its own new coordinate system is used at each iteration. In this system, the generalized gradient of the objective function is calculated, one iteration of the method of generalized gradient descent is performed, and the coordinate system is changed by space dilation in the direction of the generalized gradient. This idea was embodied in numerous space dilation algorithms, some of which will be considered below.

If an algorithm is designed for the solution of a wide circle of problems, then the researcher can choose a mathematical model that most adequately and economically reflects an actual process. With a view to extending the circle of efficiently solved nonsmooth optimization problems, N. Z. Shor proposed in the 1969–1971s and, together with his disciples, experimentally investigated [184, 190] methods of subgradient type with space dilation that are used to solve essentially ravine problems. An algorithm that was developed at that time and dilated the space in the direction of the difference of two successive subgradients (the so-called $r$-algorithm) remains one of the most practically efficient procedures of nonsmooth optimization even at the present time.

Let us consider this method in more detail. We illustrate it by the example of the problem of minimization of a convex function $f(x)$ defined on $E^n$. We assume that $f(x)$ has a bounded region of minima $X^*$ so that we have $\lim_{x\to\infty} f(x) = +\infty$.

We choose an initial approximation $x_0 \in E^n$ and a nonsingular matrix $B_0$ (a unit matrix $I_n$ or a diagonal matrix $D_n$ that has nonnegative elements on its diagonal and with the help of which variables are scaled are mostly chosen).

The first step of the algorithm is made according to the formula $x_1 = x_0 - h_0\eta_0$, where $\eta_0 = B_0B_0^{\mathrm{T}}g_f(x_0)$, $h_0$ is some step multiplier chosen under the condition that there exists a subgradient $g_f(x_1)$ at the point $x_1$, and it is such that we have $(g_f(x_1),\ \eta_0) \le 0$. When $B_0 = I_n$, we have $\eta_0 = g_f(x_0)$, and the first step coincides with the iteration of the subgradient method with a constant metric.

Let the values of $x_k \in E^n$ and matrices $B_k$ of dimension $n \times n$ be obtained as a result of $k$ $(k = 1,\ 2, \ldots)$ steps of the process of computations. We describe the $(k+1)$th step of the extremum searching process:

1. Compute the following quantities: $g_f(x_k)$, that is, the subgradient of the function $f(x)$ at the point $x_k$, and $r_k = B_k^T(g_f(x_k) - g_f(x_{k-1}))$, that is, the vector of the difference between two successive subgradients in the transformed space.
   The passage from the initial space to the transformed one is specified by the formula $y = A_kx$, where $A_k = B_k^{-1}$. We define the function $\varphi_k(y) = f(B_ky)$ and obtain $g_{\varphi_k}(y) = B_k^Tg_f(x)$. Thus, $r_k$ is the difference of two subgradients of the function $\varphi_k(y)$ that are computed at the points $y_k = A_kx_k$ and $\tilde{y}_k = A_kx_{k-1}$.
2. Compute $\xi_k = r_k/||r_k||$.
3. Specify the quantity $\beta_k$ inverse to the space dilation coefficient $\alpha_k$.
4. Compute $B_{k+1} = B_kR_{\beta_k}(\xi_k)$, where $R_{\beta_k}(\xi_k)$ is the space dilation operator at the $(k+1)$th step. We note that $B_{k+1} = A_{k+1}^{-1}$.
5. Find $\tilde{g}_k = B_{k+1}^Tg_f(x_k)$, that is, the subgradient of the function $\varphi_{k+1} = f(B_{k+1}y)$ at the point $y_{k+1} = A_{k+1}x_k$.
6. Determine $x_{k+1} = x_k - h_kB_{k+1}\tilde{g}_k/||\tilde{g}_k||$. This step of the algorithm corresponds to the step of the generalized gradient descent in the space transformed under the action of the operator $A_{k+1}$.
7. Pass to the next step or terminate the algorithm if some termination conditions are fulfilled.

The practical efficiency of the algorithm depends in many respects on the choice of the step multiplier $h_k$. In the $r$-algorithm, $h_k$ is chosen under the condition of approximate search for the minimum of $f(x)$ along the descent direction, and in this case, during the minimization of convex functions, the condition $h_k \ge h_k^*$ should be satisfied, where $h_k^*$ is the value of the multiplier corresponding to the minimum along the direction. Hence, it is necessary that the angle made by the subgradient direction at the point $x_{k+1}$ with the descent direction at the point $x_k$ be not blunt.

In minimizing nonsmooth convex functions, the following variants of the algorithm turned out to be most successful. Space dilation coefficients $\alpha_k$ are chosen equal to 2–3, and an adaptive method of regulation is used for the multiplier step $h_k$. Some natural number $m$ and constants $q > 1$ and $t_0^0 > 0$ are specified (after $k$ steps, the latter quantity will be accordingly denoted by $t_k^0$). The algorithm moves from the point $x_k$ in the descent direction with the step $t_k^0$ until the termination condition for the descent along the direction is fulfilled or the number of steps becomes equal to $m$. The descent termination condition can be as follows: The value of the function at the next point is no smaller than its value at the previous point; another variant of this condition is as follows: The derivative along the descent direction at a given point is negative. If $m$ steps have been executed and the descent termination condition is not fulfilled, then $t_k^1 = qt_k^0$ is stored instead of $t_k^0$, where $q > 1$ and the descent continues in the same direction with a larger step. If, after the next $m$ steps, the descent termination condition is not fulfilled, then we take $t_k^2 = qt_k^1$ instead of $t_k^1$, etc. By virtue of the assumption $\lim_{||x|| \to \infty} f(x) = +\infty$, after a finite number of steps in the fixed direction, the descent termination condition will be necessarily fulfilled. The step constant $t_k^{p_k} = q^{p_k} t_k^0$ $(p \in \{0, 1, 2, \ldots\})$ that has been used at the last step is considered to be initial for the descent from the point $x_{k+1}$ along a new direction, that is, we have $t_k^0 = t_k^{p_k}$.

As is shown by numerous computational experiments and practical calculations, in most cases when $\alpha \in [2, 3]$, $m = 3$, and $h$ is controlled with the help of the above-mentioned method, the number of steps along a direction, on the average, seldom exceeds 2, and in this case, during $n$ steps of the $r$-algorithm, the accuracy with respect to the corresponding function is improved, as a rule, by a factor of 3–5.

In 1976–1977, A. S. Nemirovskii and D. B. Yudin [128] and N. Z. Shor [185] independently proposed the ellipsoid method that combines ideas of the cutting-plane and space transformation methods. In fact, this method is a special case of an algorithm of subgradient type with space dilation, which was proposed by N. Z. Shor as long ago as 1970 and in which the following parameters are used: The space dilation coefficient is chosen invariable and equal to

$$\alpha_{k+1} = \alpha = \sqrt{\frac{n+1}{n-1}},$$

and the step control is performed according to the following rule:

$$h_1 = \frac{r}{n+1}; \quad h_{k+1} = h_k \frac{n}{\sqrt{n^2 - 1}}; \quad k = 1, 2, \ldots,$$

where $n$ is the space dimension and $r$ is the radius of a sphere whose center is at the point $x_0$ and that contains the point $x^*$. The convergence rate of the ellipsoid method coincides with that of a geometric progression as to the deviation of the best reached value of $f(x)$ from optimal and, at the same time, the denominator of the geometric progression asymptotically depends only on the space dimension $n$,

$$q_n \approx 1 - \frac{1}{2n^2}.$$

The ellipsoid method received wide popularity all over the world after L. G. Khachiyan [175] proposed (in 1979) a polynomial algorithm based on this method for the solution of linear-programming problems with rational coefficients, which conditioned its wide applications in the complexity theory for optimization algorithms. A number of modifications of the ellipsoid method were proposed in [19]. V. S. Mikhalevich supported the development of the line of investigation into nonsmooth optimization in every possible way and propagandized the results obtained at the department. These works received world recognition mainly owing to him. V. S. Mikhalevich and N. Z. Shor participated in the work of the 11th International Symposium on Mathematical Programming in 1982 at which they presented their plenary reports concerning the development of methods and a technology of solution of high-dimensional optimization problems.

V. M. Glushkov, L. V. Kantorovich, and V. S. Mikhalevich made many efforts for embedding the system of optimum workload of pipe-rolling enterprises of the USSR and shop order management that was created together with the All-Union Research and Development Institute (VNII) for the Pipe Industry (the laboratory headed by A. I. Vainzof). The mathematical core of this system consisted of production–transportation models and nonsmooth optimization methods [116] developed at the Department of Economic Cybernetics at the Cybernetics Institute of the Academy of Sciences of the UkrSSR. V. S. Mikhalevich made a substantial contribution to the embedding of methods of solution of production–transportation problems for optimal planning in civil aviation [15, 122]. Algorithms for the solution of production–transportation problems were realized in the form of application packages (APs), in particular, in the form of the AP PLANER [113].

Numerous applications of nonsmooth optimization in production–transportation problems are considered in [116, 186, 193, 220]. Based on them, the software support was developed for some models of optimal planning, design, and control. For example, high-dimensional production–transportation problems were solved with a view to distributing ships among inland waterways. Problems of a particularly high dimensionality arose in the workload planning for rolling mills. In these problems, the number of orders (to which corresponded consumers of products in models) reached 10,000, the number of states (manufacturers of products) exceeded 50, and the nomenclature consisted of about 1,000 names of kinds of products. Thus, the total number of variables of a multi-index analogue of model (2.1)–(2.6) was about hundreds of millions. Nevertheless, the approach developed above to decomposition made it possible to solve such problems of transcomputational complexity on second-generation computers. (On a computer of the type M-220 available at that time, the process of solution required about 12–15 h to ensure the accuracy about $10^{-3}$). Algorithms proposed on this basis were used in planning the workload of rolling mills in Soyuzglavmetal of the USSR Gossnab (State Supplies) with considerable economic effect. Programs that realized methods of nondifferential optimization were also used in creating a system of automated planning of production and shop order management in ferrous metallurgy and for solving problems of distribution of coked coal among by-product coke plants. Together with the State Research and Development Institute of Civil Aviation of the Ministry of Civil Aviation of the USSR, a system of economic–mathematical models and optimization algorithms was developed that used subgradient methods for systems of short-term and long-term planning in civil aviation. They were supported by software tools, and calculations of the prospects of development of the fleet of civil aviation were performed with their help.

A meaningful experience had been accumulated before 1970 in solving other structured problems (similar to the production–transportation problem) of linear and nonlinear programming with the use of subgradient methods, for example, multicommodity network flow problems with bounded capacities, problems of choice of an optimal structure of machine and tractor fleets in individual farms and in a branch on the whole, etc. This experience had shown that methods of nondifferential optimization are most efficient in combination with various decomposition schemes in solving coordination problems of the form (2.7)–(2.10).

It should be mentioned that groundworks on production–transportation problems can be efficiently used now in logistics systems of large mining and smelting and machine-building holdings.

Works on the application of nonsmooth optimization methods to the solution of economic problems also continued in the Cybernetics Institute of AS of UkrSSR throughout the 1970s and 1980s. In particular, under the scientific leadership of V. S. Mikhalevich, new numerical methods were developed that were used for the solution of special classes of nonlinear-programming problems [191, 193, 221].

The majority of such problems are multiextremal and can be exactly solved only with the use of exhaustive search methods (e.g., the branch-and-bound method). At the same time, to increase the practical efficiency of such methods, it is expedient to construct more exact lower estimates of objective functions in minimization problems. N. Z. Shor suggested to use dual quadratic estimates for some multiextremal and combinatorial optimization problems that, as a rule, are more exact in comparison with corresponding linear estimates and can be obtained using nonsmooth optimization methods. Moreover, he developed the methodology of generation of functionally redundant quadratic constraints whose introduction into a model does not change the essence of a problem but nevertheless allows one to obtain more accurate dual estimates in some cases. Examples of functionally redundant constraints can be as follows:

(a) Quadratic consequences of linear constraints, for example, a quadratic constraint in the form $(b_i^T x + c_i)(b_j^T x + c_j) \geq 0$ is a consequence of the following two linear inequality constraints: $b_i^T x + c_i \geq 0$ and $b_j^T x + c_j \geq 0$.
(b) Quadratic constraints that characterize the ambiguity of representation of the product of three, four, or larger numbers of variables of a problem. As a rule, they take place in reducing a polynomial problem to a quadratic one. For example, we have variables $x_1, x_2 = x_1^2,$ and $x_3 = x_1^3$. Then, the quadratic constraint in the form of the equality $x_2^2 - x_1 x_3 = 0$ is a consequence of an ambiguous representation of $x_1^4,$ namely, $x_1^4 = (x_1^2)^2 = (x_1^3)(x_1)$.
(c) Quadratic constraints that are consequences of the Boolean or binary nature of variables of a problem. For example, for binary variables $x_i^2 = 1,$ $x_j^2 = 1,$ and $x_k^2 = 1,$ the following quadratic inequality is always satisfied $x_i x_j + x_i x_k + x_j x_k \geq -1$.

The idea of functionally redundant constraints has played the key role in analyzing problems of searching for the global minimum of a polynomial $P(z_1, \ldots, z_n)$ in several variables. It turned out that, in using a specially constructed equivalent quadratic problem, the corresponding dual estimates coincide with the minimal value $P^*$ of such a polynomial if and only if the nonnegative polynomial $\bar{P}(z) = P(z) - P^*$ is presented in the form of the sum of squares of other polynomials [187, 188]. This result turned out to be directly connected with the 17th classical Hilbert problem on the representation of a nonnegative rational function in the form of the sum of squares of rational functions. The mentioned investigations extend the domain of efficiently solved global polynomial minimization problems including convex minimization problems as a particular case. Algorithms for construction of dual quadratic estimates were proposed for extremum problems on graphs such as problems of searching for maximal weighted independent set of graphs, minimum graph-coloring problems [193, 219, 220], the combinatorial optimal graph partition problem [193, 220], the weighted maximum cut problem on graphs, etc. Based of nonsmooth optimization methods, new algorithms were developed for the construction of volume-optimal inscribed and circumscribed ellipsoids [221], algorithms for the solution of some problems of the stability theory for control over dynamic systems, etc. These investigations are rather completely presented in [220].

Nondifferential optimization methods played an important role in creating packages of applied programs. Theoretical investigations in the field of construction of automated software tools began in the 1960s and had led to the development of complicated interactive software systems destined for the solution of different classes of optimization problems. In this line of investigations, important results were obtained in different years by E. L. Yushchenko, P. I. Andon, A. A. Letichevskii, V. N. Redko, I. N. Parasyuk, O. L. Perevozchikova, and other researchers. These results are presented, in particular, in [23, 113–115, 134]. Programs that use subgradient methods are included in the APs PLANER, DISPRO, and DISNEL created at the Institute of Cybernetics in the 1980s for ES EVM computers. These packages realized a wide spectrum of methods for the solution and investigation of problems of optimal planning, design and management, distribution and renewal of enterprises, design of technical devices and machines, and work scheduling under bounded resources. In [114], a review of optimization methods is presented that were developed in the Institute of Cybernetics and became central in realizing the mentioned packages of applied programs. In [113], mathematical models of applied problems and the system software of the AP PLANER are described. The majority of these models were included in an improved form in the AP DISNEL. In [115], the destination, classes of solved problems, and system and algorithmic supports for the AP DISPRO-3 are described. Here, nonsmooth optimization methods were mainly used in solving "evaluative" problems using the branch-and-bound method for special classes of discrete problems.

Important results connected with the investigation and solution of some special problems of discrete optimization were obtained by Professor V. A. Trubin who worked many years under the leadership of V. S. Mikhalevich at the Department of Economic Cybernetics. He studied properties of the polytope $M$ of the bounded combinatorial partition problem

$$M = \{Ax = 1,\ x \leq 0\},$$

where $A$ is a (0,1) matrix and 1 and 0 are the unit and zero vectors of the corresponding dimension.

Let $X$ be the set of vertices of the polytope $M$, and let $R$ be the set of its edges. Together with the polytope $M = M(X, R)$, we consider a polytope $\overline{W} = M(\overline{X}, \overline{R})$ generated by only integer-valued vertices $\overline{X} \subset X$ with the set of edges $\overline{R}$. In 1969, V. A. Trubin [171] proved the following property of $\overline{M}$: $\overline{R}$ is a subset of $R$. This property allows one to basically modify the simplex method of linear programming to solve problems of integer-valued optimization on the polytope $M$. Its modification lies in the prohibition of any movement from the current feasible solution $x$ to the adjacent (along an edge from $R$) non-integer-valued solution $y$. In other words, transitions only to adjacent integer-valued vertices are allowable if the value of the criterion of the problem is improved in this case. The mentioned property directly implies the proof of the well-known Hirsch hypothesis on the considered class of integer-valued optimization problems. Thus, this property allows one to construct an algorithm for the solution of the partition problem with a rather small number of iterations. However, as was almost simultaneously noted by many researchers, the amount of computations required for the execution of one iteration of this algorithm can turn out to be rather computationally intensive, which is connected with a high degeneracy of the polytope $\overline{M}$.

Many optimization problems on graphs and networks belong to the class of partition problems such as the search for a maximum independent set, minimal covering, coloring, problem of distribution and synthesis of communication networks, and problems of cargo transportation and scheduling.

Investigations in the field of minimization of concave functions over a transportation polytope [116, 173] began as long ago as 1968. This class of problems includes numerous extremal problems that arise in the design of communication and information networks, distribution of production, choice of the configuration of the equipment for multiproduct manufactures, and unification and standardization. One of the most promising approaches to the solution of these problems is their reduction to problems of discrete optimization of a special structure and the development of decomposition methods that maximally take into account the structural properties of the latter problems during solving them. A general decomposition scheme was developed for the construction of double-ended estimates of functionals in the mentioned problems. It includes a general oracle algorithm of linear optimization (based on the scheme of the simplex method or on nonsmooth optimization methods) in which the role of an oracle that determines the direction of movement at a given point (the subgradient or its projection) and step size is played by minimum-cut and shortest path procedures applied to some graph constructed for a given step. The construction of the decomposition approach allows for the introduction of a partial order over the columns of the constraint matrix of the problem being considered, which, in turn, allows one to replace an algorithm of exact construction of estimates by its approximate variant of gradient type. This replacement dramatically reduces the computational intensity of construction of estimates but, unfortunately, can worsen them. However, numerous experiments with various practical problems showed that the gap between these estimates remains small, and the time required for the solution of problems is completely acceptable [116]. An approximate approach is often unique in connection with high dimensions of problems that arise in design practice. It was used for the solution of problems of distribution of locomotive and car repair bases, design of water supply systems in the Zaporizhia and Zakarpattia oblasts of Ukraine, municipal communication and electric networks, ship's pipeline networks, problems of choosing configurations of tube-rolling shops, etc.

Over the past 10 years, methods of nondifferential optimization have found a number of important practical applications. To solve special classes of problems of two-stage stochastic programming with simple and fixed recursions, the programs Shor1 (whose authors are N. Z. Shor and A. P. Lykhovyd) and Shor2 (whose authors are N. Z. Shor and N. G. Zhurbenko) were developed. The programs are based on the use of the scheme of decomposition with respect to variables and solution of a nonsmooth coordination problem with the help of the $r$-algorithm (which is one of the most well-known subgradient methods) with adaptive step regulation. The Shor1 and Shor2 programs are introduced into the system SLP-IOR developed at the Institute for Operations Research (Zurich, Switzerland) for modeling stochastic linear-programming problems. A software support was also developed for some problems of optimum design and network routing with allowance for possible failures of some components of a network and a modification in requirements on flows. Its description is included in [192]. Methods of nonsmooth optimization were used to solve some problems of design of power installations [89]. These works were performed by Yu. P. Laptin and M. G. Zhurbenko, and their results were introduced into the Kharkiv Central Design Bureau "Energoprogres."

The mentioned methods were also used by Professor E. M. Kiseleva who (together with her disciples) developed the mathematical theory of continuous problems of optimal partition of sets in an $n$-dimensional Euclidean space that, first, are nonclassical problems of infinitely dimensional mathematical programming with Boolean variables and, second, form one more source of generation of nonsmooth optimization problems. Based on the proposed and theoretically substantiated methods for solving problems of optimal partition of sets [77], algorithms are constructed whose component is the $r$-algorithm of N. Z. Shor and its modifications.

In [116, 169, 170, 172, 173], the results of investigations are presented that are devoted to the development of polynomial algorithms for the solution of some network analysis and synthesis problems. Here, we are dealing with the problem of distribution of production over treelike networks, Weber problems in a space with the $L_1$-metric, problems of determination of the strength of a network and its optimal strengthening, and also problems of packing and covering of a network by spanning trees and branchings in continuous and integer-valued statements.

The further advancement of works on the mentioned subjects took place in the 1990s when the Institute of Cybernetics performed joint investigations with several leading companies, in particular, with the well-known Japanese Wacom Co., Ltd., on design methods for telecommunication networks. A distinctive feature of solving these problems was the simultaneous application of the methods of sequential analysis of variants (SAV) and nondifferential optimization algorithms mentioned above.

In 1993–1994, the method of sequential analysis of variants was used in solving problems of planning optimal fueling regimes of thermal power stations over the daily and week hourly scheduled period. This project was headed by Masaki Ito from Wacom Co., Ltd. and its research management was carried out by N. Z. Shor and V. A. Trubin from the Institute of Cybernetics. The mentioned problem was reduced to the following integer-valued problem: It is required to find

$$\min \sum_{i=1}^{n} q_i(x_i), \tag{2.12}$$

under the following constraints:

$$\sum_{i=1}^{n} x_i = b, \tag{2.13}$$

$$x_i \in \{0 \vee [l_i,\ u_i]\},\quad i = 1,\ 2, \ldots,\ n. \tag{2.14}$$

Here, $b, l_i,\ u_i,\quad i = 1,\ 2, \ldots,\ n$, are integers such that $b > 0$, $l_i \geq 0$, and $u_i > l_i$ and $q_i(x_i)$ are nonnegative bounded functions defined over an interval $[l_i,\ u_i]$, $i = 1, \ldots,\ n$.

In the general case, for large $b$, problem (2.12)–(2.14) is rather complicated, and it is inexpedient to use classical methods (e.g., necessary extremum conditions) to solve it. Nevertheless, the fact that the objective function is separable allows one to apply the method of sequential analysis of variants that generalizes the idea of dynamic programming [11, 17].

For problem (2.12)–(2.14), the system of functional Bellman equations is of the form

$$f_1(y) = \begin{cases} 0, & y = 0, \\ q_1(y), & l_1 \leq y \leq u_1, \\ \infty, & 0 < y < l_1,\quad u_1 < y \leq b, \end{cases} \tag{2.15}$$

$$f_k(y) = \min_{l_k \leq x \leq u_k} \{f_{k-1}(y),\ q_k(x) + f_{k-1}(y - x)\},\quad k = 2, \ldots,\ n, \tag{2.16}$$

where $0 \leq y \leq b$ and $f_k(y)$ are Bellman functions, $k = 1, \ldots,\ n$.

In what follows, we call the algorithm constructed according to relationships (2.15) and (2.16) Algorithm 1. Its distinctive feature is that it does not depend on the form of functions $q_i(x_i),\ i = 1, \ldots,\ n$. The basic operation for Algorithm 1 is the addition operation. It determines its complexity (the number of additions in Algorithm 1 is bounded by the value of $b\sum_{i=1}^{n}(u_i - l_i)$). In the case of piecewise concave functions, the speed of Algorithm 1 can be increased with simultaneously decreasing the number of additions. To this end, the lemma presented below is used.

**Lemma 2.1.** *Let* $x^* = (x_1^*, \ldots, x_n^*)$ *be an optimum solution of problem (2.12)–(2.14) for pieces–concave functions* $q_i(x_i),\ i = 1, \ldots, n$. *Then, one of the following statements is true*:

*(a) If an optimum solution* $x^*$ *is unique, then it contains no more than one variable that assumes its value in the interval of concavity. All the other variables assume their values at the end points of intervals of concavity.*

*(b) If an optimum solution is not unique, then there is an optimum solution that contains no more than one variable whose values are within the interval of concavity.*

Let us pay attention to analogies between the problem being considered and linear-programming problems. A distinctive feature of the structure of optimum solutions (the fact that there is at least one basic solution with a relatively small amount of nonzero components is among such solutions) in the latter problems made it possible to create efficient exhaustive algorithms among which the simplex method is most well known. The lemma presented above also allows one to apply a similar approach to the solution of the problem of planning optimal regimes of fueling power-generating units.

Using properties of an optimum solution, one can "accelerate" Algorithm 1 if functions $q_i(x_i),\ i = 1, \ldots, n,$ are piecewise concave. For the variables $x$ for which a conditionally optimum solution of the $(k-1)$th function $f_{k-1}(y)$ already contains a variable whose values belong to the interval of concavity, it is sufficient to take into account only the ends of intervals of concavity of the $k$th function $q_k(x_k),\ l_k \le x_k \le u_k,$ in constructing the function $f_k(y)$.

We call the algorithm modified in this manner Algorithm 2. Its computational scheme practically repeats that of Algorithm 1. A distinction is that the modified algorithm retains an array of flags that indicate whether the current value of $f_{k-1}(y)$ is reached at the ends of intervals of concavity for all $1, \ldots, k-1$ variables or some of these variables have already assumed a value in the interval of concavity. Depending on the value of such a flag, the loop for recomputation of $f_k(y)$ is executed for all points of the interval $[l_k,\ u_k]$ or only for the end points of intervals of concavity of functions $q_k$.

We estimate the computational complexity of Algorithm 2. Here, the number of addition operations is about $nbN$, where $N$ is the maximal number of intervals of concavity, which provides a significant computational effect, especially when $N << \max_i (u_i - l_i)$. Moreover, the additional array of flags makes it possible to find out whether the minimum of the objective function is reached or not reached at the ends of intervals of concavity. This provides an optimum solution to problem (2.12)–(2.14) with an additional useful characteristic in the case of piecewise concave functions.

It should be noted that computational experiments performed on some test problems have shown doubtless advantages of Algorithm 2. There is reason to hope that the time required for the solution of real-world problems of large sizes will be acceptable if this algorithm is used together with decomposition schemes.

The operating speed of the SAV method plays here a decisive role since it is used in computing subgradients of nonsmooth functions at every step of the subgradient process.

The methods of nondifferential optimization that are developed at the V. M. Glushkov Institute of Cybernetics were also used to solve problems of optimal design of structures of reliable networks within the framework of investigations pursued according to a project of the Science and Technology Center in Ukraine (STCU). In particular, the developers of the project investigated problems of designing a minimum-cost network under failure conditions of its separate units, finding throughputs of edges in a reliable directed network, designing an optimal logical structure of a reliable network, modernizing a reliable network, and optimizing networks with allowance for incomplete information and also problems of forward planning of shipping operations and finding of an optimal nomenclature for a rolling stock. To search for optimum solutions under these conditions, they used methods of nondifferential optimization, methods of local discrete search with elements of sequential analysis of variants, and decomposition schemes [192].

As an example, let us consider two mathematical models presented in [192], namely, the problem of finding edges throughputs in a reliable directed network with transmission of flows along arbitrary paths and with transmission of flows along a given set of admissible paths. In both cases, it is necessary to provide the adequate functioning of the network under failure conditions (its separate components, namely, edges and nodes, can fail).

A network failure is understood to be a network state that decreases the throughput of one or several of its edges. In formulating a mathematical model of the problem of finding the throughput of edges of a reliable directed network with transmission of flows along arbitrary paths, we use the following data:

1. A directed network $N(V, A)$ is specified by a set of nodes $V$ and a set of edges $A$. For an edge $(i, j) \in A$, we denote by $c_{ij}$ the cost of creation of unit throughput and by $y_{ij}^0$ the available throughput resource.
2. A set of requests $D$ for volumes of flows between pairs of nodes from some subset $V_0 \subset V$ is given. Each element of the set $D$ is specified by the following three numerical quantities: a pair $(r, s)$ and the corresponding volume of the flow $d_{rs}$ that should be allowed to pass through the network from its node $r \in V_0$ (source) to its node $s \in V_0$ (drain). For all pairs $(r, s)$, the corresponding values of volumes $d_{rs}$ are given in the same units as throughput capacities of edges.
3. A set $T$ of possible failures of the network $N(V, A)$ is given. A failure $t \in T$ of the network is characterized by a collection of coefficients $0 \le \mu_{ijt} \le 1$ for all $(i, j) \in A$, where a concrete coefficient $\mu_{ijt}$ testifies to the fact that the throughput of an edge $(i, j)$ in the network decreases by a factor of $1/\mu_{ijt}$ as a result of the failure $t$. For the convenience of description of mathematical models, we will conditionally consider that the index $t = 0$ as the "zero" failure of the network $N(V, A)$ to which corresponds the collection of coefficients $\mu_{ij0} = 1$ for all $(i, j) \in A$. In fact, the "zero" failure of the network means that the network $N(V, A)$ functions as in the case when failures are absent.

Let us make the following assumption on the input data for items (1)–(3).

**Assumption 2.1.** *For the directed network $N(V,A)$, all the requests $D$ for the transmission of volumes of flows between pairs of nodes can be satisfied for unbounded values of throughputs of edges of this network $N(V,A)$.*

Let $Y = \{y_{ij},\ (i,\ j) \in A\}$ be the set of unknown values of throughputs of edges $(i,j) \in A$ that should be added to the already existing values of throughputs of edges $Y^0 = \{y^0_{ij},\ (i,\ j) \in A\}$ of the network, and let $x^{rs}_{ijt}$ be an unknown value of the portion of the volume of the flow $d_{rs},\ (r,\ s) \in D$ that will pass through the edge $(i,j) \in A$ after the $t$th network failure. Then the mathematical model of the problem of finding values of $Y$ with "optimal total cost" for the reliable directed network $N(V,A)$ can be formulated as the following problem of mathematical programming: It is necessary to minimize the linear function

$$\sum_{(i,\ j)\in A} c_{ij} y_{ij} \tag{2.17}$$

under the constraints

$$\sum_{(r,\ s)\in D} x^{rs}_{ijt} \le \mu_{ijt}(y^0_{ij} + y_{ij}),\quad t \in (0 \cup T),\quad (i,\ j) \in A, \tag{2.18}$$

$$\sum_{j:(i,j)\in A} x^{rs}_{ijt} - \sum_{j:(j,i)\in A} x^{rs}_{jit} = \begin{cases} d_{rs}, & i = r, \\ 0, & i \neq r,\ s, \\ -d_{rs}, & i = s,\quad t \in (0 \cup T),\quad (r,\ s) \in D, \end{cases} \tag{2.19}$$

$$x^{rs}_{ijt} \ge 0,\quad t \in (0 \cup T),\quad (i,\ j) \in A,\quad (r,\ s) \in D, \tag{2.20}$$

$$y_{ij} \ge 0,\quad (i,\ j) \in A. \tag{2.21}$$

Here, function (2.17) being minimized determines the total expenditures with allowance for the cost of the increase in the throughput of edges that should be added to the existing edges with a view to ensuring the reliability of the network $N(V,A)$.

If the transmission of flows should be restricted to a given set of admissible paths, then the following item is added to conditions (1)–(3):

4. A set of admissible paths in a network $P = \bigcup P(r,\ s)$is given, where $P(r,\ s)$ is the subset of paths in the network $N(V,A)$ that connect a source $r$ with drains $s$, $(r,\ s) \in D$ and through which (and only through which) a flow can be transmitted from the node $r$ to the node $s$. We assume that such collections of paths are given for all pairs $(r,\ s) \in D$. We will specify a concrete path $P_k(r,\ s) \subseteq P(r,\ s)$, that is, the path with a number $k$ for transmission of a volume of the flow from the node $r$ to the node $s$, by a vector $a^{rs}_k$ whose length equals $\mid A \mid$, which consists

of zeros and unities, and in which unities correspond to the edges of the network $N(V, A)$ through which this path passes and zeros correspond to the edges that do not belong to this path.

**Assumption 2.2.** *The set $P$ contains nonempty subsets $P(r, s)$ for all pairs $(r, s) \in D$.*

For the directed network $N(V, A)$, a new problem of finding values of throughputs of edges $Y = \{y_{ij}, (i,j) \in A\}$ with minimum total cost that should be added to the already existing edges $(Y_0)$ in order that the network $N(V, A)$ become reliable can be formulated as follows: minimize the linear function

$$\sum_{(i,j)\in A} c_{ij} y_{ij} \tag{2.22}$$

under the constraints

$$\sum_{(r,s)\in D} \sum_{k\in P(r,s)} a_k^{rs} x_{kt}^{rs} \le \mu_{ijt}(y_{ij}^0 + y_{ij}), \quad t \in (0 \cup T), \quad (i, j) \in A, \tag{2.23}$$

$$\sum_{k\in P(r,s)} x_{kt}^{rs} = d_{rs}, \quad t \in (0 \cup T), \quad (r, s) \in D, \tag{2.24}$$

$$x_{kt}^{rs} \ge 0, \quad t \in (0 \cup T), \quad (r, s) \in D, \quad k \in P(r, s), \tag{2.25}$$

$$y_{ij} \ge 0, \quad (i,j) \in A, \tag{2.26}$$

where $x_{kt}^{rs}$ is an unknown value of the subvolume of the flow $d_{rs}$ that will pass through the $k$th path, $k \in P(r, s)$, in the case of the $t$th failure of the network.

A model destined for the solution of problems of forward transportation planning and rational allocation of appropriations for the reconstruction of transport (railway, automobile, and aviation) networks was also considered. The realization of the model allows one to obtain numerical information on the following questions of forward planning: nodes whose throughputs are critical and rational allocation of appropriations for their reconstruction, a rational scheme of flows of transport facilities with allowance for the minimization of operating costs and volumes of empty flows, and the determination of an optimal structure of a fleet of transport facilities and rational use of appropriations for its extension.

At an informal level, the considered problem of forward planning is described as follows. A scheduled period is given that consists of separate intervals. Each interval of the scheduled period is characterized by a rather stable freight traffic flows on a railway network. As a rule, the choice of intervals is determined by a seasonal factor. For the scheduled period, the following main parameters of capabilities of a transport system are determined: its fleet of transport facilities and cargo-handling capacities of stations. A plan of reconstruction of a transport system lies in the determination of an expedient allocation of a given amount of financial investments destined for the increase in cargo-handling capacities of stations and replenishment of the rolling stock.

According to the optimization criterion, optimization consists of the maximization of the total profit obtained during the scheduled period owing to freight traffic. As a result of calculations, it is possible to obtain numerical information on main questions of forward planning such as:

– Determination of railway stations whose throughput capacity is critical and provision of recommendations on the allocation of appropriations for their reconstruction
– Determination of a rational structure of the fleet of transport facilities and provision of recommendations on the allocation of appropriations for its replenishment
– Development of a rational scheme of railcar traffic with allowance for the minimization of operating costs and volumes of empty flows
– Rational choice of orders for transportation to places of realization

These data can be used to make decisions on promising lines of development of a transport system and to objectively substantiate critical factors of its functioning. A scheme of railcar traffic provides important information for solving the problem of determination of routing of freight traffic and train operation scheduling.

The software support developed for the solution of such problems uses decomposition schemes, which take into account block structures, together with the $r$-algorithm. To solve an internal subproblem, a scheme of decomposition under constraints and interior-point methods [35] are used that make it possible to take into account the block structure of the subproblem. A software implementation of the algorithm in the first problem also requires a subroutine for finding shortest paths in a directed network, and this problem can be efficiently solved with the help of the method of sequential analysis of variants.

The considered examples of works on modeling demonstratively illustrate that, even under difficult economic conditions at the beginning of the 1990s, the Institute of Cybernetics retained a powerful scientific potential that made it possible to realize large-scale projects with creating complicated optimization models, to develop computational algorithms, and to apply this support tools to the solution of complicated applied problems.

The enthusiasm inherent in young collectives and forward-looking policy of V. M. Glushkov and V. S. Mikhalevich promoted the formation of groups of scientists with strong creative potential who efficiently worked in topical lines of investigation in the field of optimization. A striking example can be the school of B. N. Pshenichnyi who became an academician of the National Academy of Sciences of Ukraine in due time. The subject matter of his works was connected with the investigation of quasi-Newtonian methods (together with Yu. M. Danilin) [145], convex analysis in Banach spaces [144], methods of linearization and nonlinear programming [141, 142], the theory of multivalued mappings, and guaranteed estimates. Most powerful results were obtained in the field of differential games [63, 146, 177, 178].

In the 1960s, at the initiative of B. N. Pshenichnyi and with the assistance of V. S. Mikhalevich, investigations connected with pursuit-evasion games including group games were initiated, which have applications in economy and technical systems, in particular, in defense systems (air-defense systems, etc.). During these investigations, problems of nondifferential optimization also arose. Main results on these subjects are contained in [63, 177].

The development of nonsmooth optimization methods created preconditions for the research and development of optimization models with random parameters that were called stochastic optimization models. There is a wide class of applied problems that can be appropriately formulated as the mentioned models. Stochastic optimization is particularly widely applied in systems of making decisions on the choice of lines of long-term development of an economy on the whole and its separate branches (power engineering, agriculture, transport, etc.) and also in ecological–economic modeling and investigation of social processes.

In investigating problems of stochastic optimization in the Institute of Cybernetics, significant attention was given to the development of numerical methods of solution of stochastic problems and to the practical application of these methods. In particular, in the 1970s, stochastic quasigradient methods for the solution of general optimization problems with nondifferential and nonconvex functions were developed at the initiative of V. S. Mikhalevich. The possibility of application of these methods to problems with nondifferential functions is rather significant for important applied problems with a fast and unforeseen behavior of objects being modeled. Stochastic quasigradient methods can be considered as a generalization of stochastic approximation methods to constrained problems and also as a development of random search methods. The following important distinctive feature of stochastic quasigradient methods should be emphasized: They do not require exact values of objective functions and constraints. This provides ample opportunities of using them for the optimization of complex systems under uncertainty.

It should be noted that subgradient methods of nondifferential optimization that were developed in the Cybernetics Institute under the guidance of V. S. Mikhalevich and N. Z. Shor exerted influence on the development of theory and practice in many directions of investigations of mathematical programming. They have gained acceptance of leading specialists of the world's scientific community in the field of optimization. The achievements of the domestic school of nondifferential optimization were also highly appreciated in the former USSR and in Ukraine. For their fruitful work on the development of numerical optimization methods and their applications, N. Z. Shor and his disciples were awarded the State Prizes of the USSR (1981) and Ukraine (1973, 1993, and 2000) in science and engineering and the V. M. Glushkov Prize (1987) and V. S. Mikhalevich Prize (1999) of the National Academy of Sciences of Ukraine.

At the present time, the active development of nondifferential optimization methods and their applications continues. As early as 1982, N. Z. Shor and V. I. Gershovich wrote in [19] that the theory of the entire class of space dilation algorithms was still a long way off perfection and that the construction of an algorithm whose practical efficiency would not concede the efficiency of the

$r$-algorithm and that would be as well substantiated as the ellipsoid method seemed to be a rather realistic objective. Although more than 25 years have passed since that time, the problem of rigorous substantiation of $r$-algorithms still remains topical until now. As a step in this direction, we can mention [167] in which the transformation of a special ellipsoid into a sphere is based on a space transformation close to that used in $r$-algorithms. However, space is dilated here in the direction of the difference between two normalized subgradients. It is close to the direction of the difference between two subgradients only when the norms of the subgradients are close. If the norms of the subgradients are considerably different, then the result will be essentially distinct from the space dilation in the direction of the difference of two subgradients. A similar transformation of space is also used in new modifications of ε-subgradient space dilation algorithms developed by N. G. Zhurbenko [192, p. 29–39]. Investigations along these lines are still being actively developed.

## 2.3 Problems in the Development and Use of Systems Analysis Methods

Studying systems of different nature and adapting methods intended for one class of problems to other classes was always one of the main trends in the development of cybernetics. The Glushkov Institute of Cybernetics always took an active part in such studies, which were later called "systems analysis."

Systems analysis means studying an object as a set of elements that form a system. In scientific research, it is intended to study the behavior of an object as a system with all the factors that influence it. This method is widely applied in integrated scientific research of the activity of production associations and the industry as a whole, in determining the proportions of the development of economic branches, etc.

There is no unified technique for systems analysis in scientific research yet. In practical studies, it is applied in combination with operations research theory, which allows quantitative estimation of the objects of study, and analysis of systems of study of objects under uncertainty, using system engineering to design and synthesize complex systems during the study of their operation, for example, to design and estimate the economic efficiency of computer-aided control systems for technological processes. Systems analysis is important in human resource management.

As far back as in the early 1960s, Mikhalevich drew the attention of employees of the Institute of Cybernetics to recognition and identification theory for stochastic lumped- and distributed-parameter systems. The great applied importance of these studies stimulated the development of new subject areas in the statistics of random functions. New approaches to the analysis of problems in nonlinear and nonparametric regression analysis were proposed, new classes of estimates and their asymptotic properties were investigated, and stochastic optimization and estimation theories were related. In this field, noteworthy are the studies [78, 80, 81].

In a number of studies, Mikhalevich developed modern notions of informatics and its interaction with allied sciences such as cybernetics, mathematics, economy, etc.

He investigated the trends in the formation of informatics as a complex scientific discipline that studies all aspects of the development, design, creation, operation of complex systems, their application in various fields of social practice, ways of formalization of information for computer processing, and development of information culture of the society [97, 107]. By virtue of the aforesaid, informatics should widely use methods of systems analysis.

The breadth of interests of Mikhalevich was huge. In particular, by his initiative, Yu. M. Onopchuk studied regulation processes in the major functional systems of living organisms: circulatory, respiratory, and immune [137].

The original scientific work [110], which is a vivid example of the versatility of interests of Mikhalevich, was carried out jointly with V. M. Kuntsevich. It is devoted to the extremely vital (at the late 1980s) problem of identifying the causes of the arms race (primarily nuclear) between the superpowers going on at that time. The model was a discrete game with two terminal sets. Such problems are distinguished in dynamic game theory as extremely difficult; nevertheless, the use of the methods of general control theory and systems analysis allowed conducting a qualitative analysis of this phenomenon and formulating relevant recommendations. Indirectly, the authors estimated the efficiency of investment strategies for conflict interaction projects. Important studies into discrete optimization, in particular, solution of discrete–continuous problems, were conducted at the department headed by I. V. Sergienko, in close cooperation with Mikhalevich's department. In addition to creating specialized approximate local algorithms according to the descent vector method, they considered solution stability, general problems in global optimization, sequential search, multicriterion optimization, the asymptotic complexity of simple approximate algorithms for solving extremum problems on graphs (together with Prof. V. O. Perepelytsya), problems on permutations (together with Belarusian mathematicians V. A. Emelichev and V. S. Tanaev) and others. These studies were accompanied by large-scale computing experiments and development of application software.

In the 1950–1960s, the rapid development of science and technology made it necessary to use a computer to solve complex scientific and engineering problems unsolvable by traditional methods and to which systems analysis methods were applied. To solve such problems, the power of available computing facilities had to be used to the greatest possible extent. This resulted in a new scientific field, computation optimization theory, Mikhalevich being one of the founders.

Mikhalevich was actively engaged in training and professional development of specialists in optimization, systems analysis, and operations research. He was a cochairman of the majority of Winter Schools on Mathematical Programming in Drohobych and Summer Schools on Optimization, which were conducted by academician N. N. Moiseev. Scientists of the Institute of Cybernetics participated in Mathematical Schools in Sverdlovsk (organized by academician N. N. Krasovskii) and at the Baikal lake (organized by Irkutsk scientists), in All-Union Seminars on graph theory, etc.

Let us now dwell on the cooperation with the International Institute for Applied Systems Analysis. The Academy of Sciences of Ukraine placed strong emphasis on this institute, which was launched in 1972 with active participation of

Glushkov and Mikhalevich. The International Institute for Applied Systems Analysis (IIASA) is a unique nongovernmental institution engaged in scientific study of global problems that can be solved only at the international level. The number of such problems and their scope are growing, which is demonstrated by the current global crisis of financial and economic systems of different countries. It shows that the lack of systemic analysis and coordination at the international level can easily lead the global economy to a megacatastrophe. It also demonstrates the need for an independent scientific institute, common to all countries and engaged in research and solutions to global interdisciplinary problems. The issues IIASA deals with include safe, in the broadest sense, development of global power engineering based on new energy-saving and environmentally friendly technologies and related climate change, sustainable development of agriculture and food production, and water and demographic problems, including the possibility of large-scale migration of population, animals, and insects that can adversely affect the health system. All this requires new models and methods to be developed for robust (stable) solutions under uncertainty and impossible accurate forecast of various risks.

An attractive feature of IIASA in terms of scientific research is its private nature. IIASA activities are coordinated by national systems analysis committees of member countries, which (by tradition, beginning with the USA and USSR and other countries, including Ukraine) are created by national academies of sciences. Since the global issues affect the interests of various organizations, national committees include representatives of leading organizations that are directly related to specific IIASA programs.

International Institute for Applied Systems Analysis (Laxenburg, Austria)

Obviously, the search for solutions to these problems taking into account the specifics of individual countries is only possible through cooperative efforts of researchers from different countries, which is the main task of the IIASA. In this sense, the IIASA can be considered a Ukrainian Research Institute, Institute of the USA, Russia, Germany, Japan, Poland, and all other countries that are its members. Along with these countries, Austria, Egypt, India, China, the Netherlands, Norway, Pakistan, Sweden, South Africa, and South Korea are currently members of the IIASA.

The IIASA is located 16 km from the Austrian capital Vienna, in the picturesque town of Laxenburg. From the USSR, the chairman of the Council of Ministers of the USSR A. M. Kosygin and academician D. M. Gvishiani and, from the USA, President L. Johnson and his adviser M. Bundy had a great influence. Academician V. M. Glushkov and Professor H. Rayffa from Harvard University (USA), widely known for numerous monographs in optimal solutions, provided substantial assistance. Professor Rayffa was appointed the first director of the IIASA, and Gvishiani became the Chairman of the Council, which manages the IIASA.

Extensive experience in solving complex multidisciplinary problems was accumulated at the Institute of Cybernetics prior to 1972. In fact, the management of the Academy of Sciences of the UkrSSR deeply understood the need for comprehensive cybernetic or, as it is often said nowadays, systemic approaches long before the formulation of the basic ideas of the IIASA. Much of it corresponded to the Soviet approach to the coordination of various areas on the basis of interrelated planning. It was obvious that the practical implementation of these approaches to solve complex scientific problems is only possible based on computers. The first in the USSR MESM, available at the Institute of Cybernetics, attracted a wide range of researchers from different organizations in the former Soviet Union. The Institute of Cybernetics was deluged with new mathematical problems, which often arose at the junction of various sciences and could not be solved by traditional methods.

As already mentioned, the faculty of cybernetics was established at the Shevchenko State University of Kyiv to develop cybernetic approaches. The Institute of Cybernetics included a Department of Moscow Institute of Physics and Technology and a Training Center to teach high-ranking managers responsible for making governmental decisions. After the creation of the Institute of Cybernetics, cybernetic approaches based on the use of computers and mathematical models of new type became the primary activities of many organizations and educational institutions of the Soviet Union.

From the formal mathematical point of view, these problems were similar to those that arose in different divisions of the IIASA since its creation. The need for considering the diversity of system solutions often led to new optimization problems characterized by extremely high dimension and substantial uncertainty.

By the time of formation of the IIASA, the Institute of Cybernetics accumulated considerable experience in solving such problems. First of all, fundamentally new methods of nondifferential and stochastic optimization that significantly differed from the Western approaches were developed. Since the original problems that require systemic or cybernetic approaches were similar, it was decided to compare the methods being evolved and to create more efficient hybrids that can be applied to various IIASA applied problems.

This idea was proposed to the IIASA by Western scientists and was actively supported by Glushkov and Mikhalevich. While the initial cooperation of the Institute of Cybernetics and IIASA began with the development of computer networks, which provided an opportunity for our organizations to use Western computer networks, in 1976 the cooperation center moved to new models and methods for solving important problems of systems analysis. In particular, great attention was paid to the development of approaches to the analysis of robust solutions for high-dimensional optimization problems under substantial uncertainty. Two types of problems usually arise in this case. They correspond to the cases of over- and underestimated risk and result in optimization models with nonsmooth and discontinuous functions. In particular, overestimated capabilities of a system can have large-scale consequences (like the Chernobyl disaster) and involve a large number of variables for modeling. For example, the IIASA's stochastic model for the analysis of robust fields of the development of the world's power engineering had up to 50,000 variables and took into account critical nonconvex effects of increasing returns (profits) of new technologies and the uncertainty of their appearance on the world market, which required the use of powerful Cray T3E-900 supercomputer at the Computing Center of the U.S. Department of Energy.

Since 1976, Mikhalevich and his student Ermoliev started permanent cooperation with the IIASA employing leading Western experts in optimization such as G. Dantzig, R. Wets, and T. Rockafellar. Many Nobel laureates such as Arrow, Koopmans, Prigogine, Klein, Kantorovich, Crutzen, and Schelling came with short- and long-term visits.

The far-sighted policy of Glushkov and Mikhalevich directed to work closely with the IIASA had important consequences. The relations with leading Western organizations and scientists were established, which made it possible to quickly evaluate and review the experience gathered at the Institute of Cybernetics in modeling and solution of complex problems. The essential feature of the IIASA, which has a small permanent staff, is its extensive network of relationships with leading research organizations and scientists all over the world. This makes it possible to organize the world experience, find relevant information, formulate a problem adequately, and find solutions that reflect the experience of many countries, rather than one-sided recommendations of a single, even leading country.

The illustrative example that demonstrates the effectiveness of the international cooperation at the IIASA may be the project of development of models and stochastic-programming methods for the analysis of optimal strategic decisions under uncertainty. Considering the time it took well-known teams of researchers from different countries to develop models and methods, acquire and analyze data for various applied problems, and create software that appeared in the final project report, it was calculated that the project completed at the IIASA approximately in 2 years from 1982 required about 200 person-years for its implementation. From the

IIASA, only two scientists (R. Wets from the USA and Yu. Ermoliev from the Institute of Cybernetics) worked on the project. Through a network of researchers from IIASA member countries, they managed to bring together theoretical and practical developments of various scientific schools, gained practical experience and contemporary software. This was published as a joint monograph and became public. This, one of the numerous IIASA projects, also led to the international stochastic-programming community, which now includes a wide range of researchers from different countries and organizations. The community regularly holds conferences and organizes joint projects, publications, and training.

The nongovernmental status and the research orientation of the institute allow selecting candidates for vacant research positions not according to quotas but by matching their research interests and qualification to the main areas in the project where a vacancy appeared. Each applicant is interviewed by institute program managers and typically must make a scientific report. The final decision is made by the program leader who completely manages the budget allocated to the institute. Noteworthy is that the employees of the All-Union Institute of Systems Studies headed by D. M. Gvishiani, Chairman of the Council of the IIASA, worked at the IIASA only for the first years, while employees of the Institute of Cybernetics worked at the IIASA all the time.

The influence of Mikhalevich at the IIASA began to grow significantly since the mid-1980s. Previously, the status of the IIASA was shaken by the Republicans who come to power in the USA and did not support multilateral cooperation and refused to pay membership fees. In addition, Gvishiani voluntarily resigned from the position of the chairman of the IIASA Council and necessitated chairman reelection. A critical issue was the choice of a new chairman of the council, who could convince decision makers in the Soviet Union to continue paying membership fees, which equally with the former US fees made major contributions to the IIASA budget. Although Nobel laureates' letters in support of the IIASA could partially compensate for the US fee from other sources, the situation actually questioned the existence of IIASA.

In June 1987, the IIASA Council elected academician Mikhalevich as the new chairman of the council, considering that he was the director of the Institute of Cybernetics, which then had wide relationships in the USSR and successfully cooperated with almost all parts of the IIASA, clearly demonstrating the benefits of international scientific cooperation.

It should be noted that the council of the IIASA and thus the position of the chairman are very important for that institution. The main objective of the council is a strategic plan, including the selection of promising areas of research, the institute's budget allocation among its departments, admission of new and exclusion of insufficiently active member states, election of the director who was to follow the decisions of the council. The council meets twice a year; its decisions are made by voting of representatives of the member states, one representative from each state. Preparing and making decisions is a rather complicated and long-term process that requires proposals to be approved by all the member countries between meetings, the chairman playing an active role and providing permanent consultations.

Mikhalevich successfully coped with the difficult task of consolidating IIASA members with a strong financial and scientific support from the USSR. To coordinate the work with the IIASA, a new National Committee in Systems Analysis was created at the Presidium of the Academy of Sciences of the USSR, which included representatives of the relevant organizations of the Academy of Sciences and other organizations directly related to the IIASA studies. A Systems Research Laboratory was established at the Institute of Cybernetics at Ermoliev's department. Its tasks included the analysis and distribution of IIASA's results among stakeholders in the USSR in the form of short statements, analytical notes, and speeches with overview reports at various conferences.

Unfortunately, the declaration of independence caused almost unresolvable situation with the further participation of Ukraine in the IIASA because of the necessity to pay membership fees. In this situation, B. E. Paton, president of the National Academy of Sciences, and Mikhalevich applied to the council to allow Ukraine to be admitted as a member country at the IIASA as an exception, while bound to organize groups of employees at the institutions of the Academy of Sciences to perform tasks in IIASA programs equivalent to the fee of Ukraine. This letter was signed by Paton, and the council made favorable decision due to the efforts of Mikhalevich and Shpak who became a representative of Ukraine to the IIASA Council after the death of Mikhalevich. Naturally, the services of the National Academy of Sciences of Ukraine to the IIASA since its establishment were taken into account.

The work of employees of leading institutes of the National Academy of Sciences (NAS) for IIASA programs in power engineering, economy, agriculture, control of catastrophic risks, demography, and environmental control enrich their understanding of global problems and of the approaches to their analysis dominating over the world. In many respects, this prevents one-point views in Ukraine, which reflect interests of isolated countries, say, from Europe or the USA.

Noteworthy is that different models developed in the IIASA are intended to evaluate the contribution of individual countries and regions to the global situation and to analyze the prospects of joint mutually beneficial development. These models include massive bases of mutually coordinated initial data, predictive information, expert estimates, and scenarios of uncertainties, which can affect the sustainable development of systems under study. The models are refined and developed, specific formal and informal methods of their analysis, including the analysis of obtained multidimensional solutions, are accumulated. All this is very rich material for member countries (including Ukraine), which can use original IIASA models directly or create their modifications and simplified versions on their own account. Along with global power engineering models, the IIASA has a global model of manufacturing agricultural products and meal, a model of interstate (interregional) transmission of air pollution and their influence on the environment and the human, which is constantly used in negotiations about the reduction of pollution emissions by countries and individual enterprises, and a model of estimating global demographic tendencies in different countries. Leading institutes of the NAS of Ukraine and other Ukrainian organizations participate in the development of individual modules of these models and use the results in their study during the preparation of strategic documents related to the development of power engineering in Ukraine, demographic and social issues, safe development of rural territories and food stuff manufacture, prospects of the development of biofuel, trade, and greenhouse gases emission. Together with the IIASA, the V. M. Glushkov Institute of Cybernetics and a number of other institutes of the Cybernetics Center of the NAS of Ukraine continue the active development of new models and methods that adequately reflect current global changes. For example, new approaches to the management of catastrophic risks related to large territories and a great number of people are developed. In systems analysis of processes with possible catastrophic consequences, the central issue is the search for robust solutions based on stochastic optimization methods being developed at the V. M. Glushkov Institute of Cybernetics and IIASA, in scientific organizations of the USA, Norway, and Great Britain.

Thus, the participation of Ukraine in the activity of the International Institute for Applied Systems Analysis, on the one hand, bears witness to the fact that the international scientific community recognizes the high level of studies conducted by Ukrainian research institutions, primarily by those of the NAS of Ukraine, and on the other hand, it is a priority in the international cooperation of Ukraine since it opens ample opportunities for the participation of Ukrainian scientists in important scientific international projects.

Undoubtedly, the recognized authority of the Ukrainian science at the IIASA, active participation of scientists from Ukraine in programs of the institute is in many respects a great merit of academician Mikhalevich. His name is cherished at the International Institute for Applied Systems Analysis. The council of the institute established the annual Mikhalevich Award to encourage the best scientific study performed by young scientists from different countries during their participation in the IIASA Summer School. It is symbolic that many young scientists from Ukrainian research institutes take part in this Summer School.

## 2.4 Stochastic Optimization Methods

In the first chapter, we briefly reviewed the first research projects conducted by V. S. Mikhalevich under the supervision of B. V. Gnedenko, academician of the National Academy of Sciences of Ukraine, while he studied at the Shevchenko State University in Kyiv and under the supervision of academician A. M. Kolmogorov during the postgraduate study at the Lomonosov Moscow State University. These works were published in authoritative scientific journals and at once drew the attention of experts in the statistical decision theory, which made him one of the most cited authors in statistical analysis. Mikhalevich obtained outstanding results on sequential Bayesian solutions and optimal methods of acceptance sampling. He developed the technique of recurrent relations for risks, which made it possible to solve new problems arising in various areas of science and technology.

Mikhalevich was the first to justify Wald's problem of comparing two simple hypotheses on the mean value of a discrete-time Wiener process and to prove that a Bayesian solution can be obtained by solving the Stefan problem with moving boundaries, which is well known in the theory of differential equations. Note that in the case of discrete time, it is in many cases possible to qualitatively describe Bayesian solutions and the domains in which the observation should be terminated to accept a hypothesis. However, it is often very difficult to describe such domains, and a specific form of solution cannot always be obtained; if it can, it is very awkward. Mikhalevich proposed a method to solve the problem by passing to the limit in a discrete-time model to make it continuous time, which allows obtaining optimal solutions. As he proved, if it is possible to pass to the limit in the discrete-time model, the risk satisfies some parabolic differential equation and finds the bounds reaching which requires terminating the observation process and making some decision. This involves difficulties related to finding the optimal solution, control, and objective function for stochastic continuous-time systems. Note that the results are far beyond the scope of optimal statistical control problems and, in fact, form the basis of optimal control theory for stochastic discrete- and continuous-time systems.

Later on, the methods of reduced Bayesian strategies developed by Mikhalevich were used as the basis for solving important economic problems related to optimal planning and design.

It should be emphasized that Mikhalevich paid much attention to new fields in optimization theory, he tried, in every possible way, to encourage talented young people and unveil their creativity. Among such fields, we will dwell on stochastic optimization and identification theory, optimal control theory, and risk assessment. The problems that were studied lacked complete information on objective functions, constraint functions, and their derivatives.

A comprehensive system analysis of the future development of economy, power engineering, and agriculture and their influence on the environment and a human being requires stochastic optimization models and methods that would make explicit allowance for the probabilistic nature of the processes and the risk due to the uncertainty inherent in the decision-making process. Moreover, there is a wide class of applied problems that cannot be formulated and solved within the deterministic framework. Being unable to list all such problems, we will nevertheless mention problems related to inventory theory, maximum average network time, generalized Polya scheme with increments that have a random number of balls (these problems necessitate analyzing processes with discontinuous urn functions). These problems are widely applied in economics, biology, chemistry, etc., and are discussed in [3, 49, 53].

For illustration, let us present two examples of such problems.

*Example 2.1.* Consider a one-product inventory management model with one storehouse and random consumption. Given a storehouse of finite capacity, create a homogeneous stock $x$ of the product for which demand is a random variable. The costs of storage of non-sold product and pent-up demand can be described by the function

$$f^0(x,\ \omega) = \begin{cases} \alpha(x-\omega), & \text{if } x \geq \omega, \\ \beta(\omega - x), & \text{if } x < \omega, \end{cases}$$

where $\alpha$ is the cost of storage of a product unit and $\beta$ is pent-up demand penalty. It is obvious that $f^0(x, \varpi)$ is a convex and nonsmooth function. The task is to minimize $f^0(x, \varpi)$.

*Example 2.2 (knapsack problem).* The problem is formulated as follows: minimize

$$F(x) = E\sum_{i=1}^{n} d_i \ max\ [0, (\theta_i - x_i)]$$

subject to the constraints

$$\sum_{i=1}^{n} c_i x_i = a,\ \ x_i \geq 0,\ \ i = 1,\ 2, \ldots,\ n,$$

where $x_i$ is the amount of the $i$th product ordered; $d_i$ is the loss because of incomplete delivery of a unit of the $i$th product; $\theta_i$ is the demand for the $i$th product, which is a random variable with known (preset or obtained from analytically processed observation results) distribution; $c_i$ is the cost of a unit of the $i$th product; and $a$ is the product inventory.

Sometimes, this problem is subject to the constraint

$$0 \leq x_i \leq a_i,\ \ i = 1,\ 2, \ldots,\ n,$$

where $a_i$ is the maximum order volume depending on the capacity of the storehouse occupied by the $i$th product. In our case, $F(x)$ is a nonsmooth function.

**Stochastic Optimization.** This field of research was headed by Yu. M. Ermoliev, Academician of the National Academy of Sciences of Ukraine, a cofounder of the well-known Ukrainian school of optimization. The direct stochastic-programming methods developed by Ermoliev made him recognized worldwide, became classical, appeared in almost all tutorials on stochastic programming, and referred to in many scientific papers and monographs. The international community highly appreciated his achievements and awarded him (as one of the world-famous founders of stochastic-programming theory) a medal and an international prize for prominent services to the cause of developing the theory of financial and insurance mathematics.

Any optimization model, such as

$$\min \{f^0(x, \varpi) | f^i(x, \varpi) \leq 0, \ i = 1, \ldots, \ m; \ x \in X \subset R^n\},$$

includes parameters $\varpi \in \Omega$, which may generally be random either because of incomplete information about their values due to, for example, measurement errors, the use of their statistical estimates or because of the stochastic nature of the parameters such as weather forecasts, variations of price, demand, productivity, share prices, etc. Then, the constraint $f^i(x, \varpi) \leq 0$ can fail for any fixed $x \in X$ and some realizations of the random parameter $\varpi$. The set of functions,

$$f(x, \ \omega) = \{f^0(x, \ \omega), \ f^1(x, \ \omega), ..., f^m(x, \ \omega)\}, \ \ \omega \in \Omega,$$

can be considered as a vector characteristic of a solution $x \in X$, and finding the optimal solution can generally be regarded as a vector optimization problem with an infinite number of criteria.

Optimization problems with random parameters are formalized in stochastic-programming theory. The stochastic quasigradient methods proposed by Ermoliev underlie direct stochastic-programming methods and can be considered a generalization of stochastic approximation methods on a constrained problem, the Monte Carlo method on an optimization problem, and a development of random search methods. The key feature of quasigradient methods is that they use statistical estimates (rather than exact values) of objective functions and constraints obtained from realizations of subintegral functions and their generalized gradients. This opens ample opportunities for these methods to be applied to optimize complex stochastic systems by simulation modeling.

Among the pioneering studies that underlie stochastic quasigradient methods, noteworthy are [53, 59, 62] and some others written in the late 1960s and early 1970s and discussing the solutions of both nonlinear and convex stochastic-programming problems with general constraints. It should be underlined that objective functions and constraints that are not necessarily differentiable were considered. The general idea of stochastic quasigradient methods is as follows.

Let us minimize a convex objective function $F^0(x)$ subject to the convex constraints $F^i(x) \leq 0, \ i = 1, \ldots, \ m; \ x \in X \subset R^n$, where the functions may have the form of expectations. When deterministic optimization methods are used, the sequence of approximate solutions $\{x^k, k = 0, \ 1, \ldots\}$ that minimizes the function $F^0(x)$ is usually found from the exact values of the functions $F^i(x), \ i = 0, \ldots, \ m,$ and their gradients (generalized gradients if the functions are nonsmooth). Nevertheless, in stochastic quasigradient methods, such a sequence is constructed based on statistically biased and unbiased estimates of the functions $F^i(x)$ and their generalized gradients $F^i_x(x)$:

$$\eta^i(k) = F^i(x^k) + a_i(k),$$

$$\xi^i(k) = F^i_x(x^k) + b_i(k), \ \ i = 0, \ldots, \ m,$$

where the errors $a_i(k)$, $b_i(k)$, $i = 0, \ldots, m$, can depend on the path $\{x^0, \ldots, x^k\}$ for the current iteration $k$ and tend to zero in some probabilistic sense as $k \to \infty$. The random vector $\xi^i(k)$ is called stochastic quasigradient of the function $F^i(x)$ at the point $x^k$ for $b_i(k) \neq 0$. If $b_i(k) = 0$, then $\xi^i(k)$ is called the stochastic gradient or stochastic generalized gradient (stochastic subgradient) of the function $F^i(x)$ depending on whether it is smooth or undifferentiable. The main way of constructing stochastic generalized gradients is to introduce the differentiation or generalized-differentiation sign under the expectation sign. Assume that

$$F^i(x) = Eg^i(x, \varpi) = \int g^i(x, \varpi) P(d\varpi).$$

Then the continuously differentiable function $F^i(x)$ has the following gradient:

$$F^i_x(x) = Eg^i_x(x, \varpi) = \int g^i_x(x, \varpi) P(d\varpi),$$

and thus, the random vector $\xi^i(x, \varpi) = g^i_x(x, \varpi)$ is the stochastic generalized gradient of the function $F^i(x)$ at the point $x$. If $F^i(x)$ and $g^i(x, \varpi)$ are nonsmooth functions, then the stochastic generalized gradient $\xi^i(x, \varpi)$ should be taken as an intersection (measurable in a set of variables) of the subdifferential $g^i_x(x, \varpi)$ and subintegral function $g^i_x(x, \varpi)$. Another way to construct stochastic quasigradients is based on the finite-difference method. For example [27], the stochastic quasigradient of the function $F^i(x)$ at the point $x^k$ can be calculated by the formula

$$\xi^i(k) = (3/2)[g^i(x^k + \delta_k h_k, \varpi^k) - g^i(x^k, \varpi^k)]h^k/\delta_k,$$

where $h^k$ is independent observation of the random vector $h = (h_1, \ldots, h_n)$ with independent components uniformly distributed over the interval $[-1, 1]$. Note that to find the vector $\xi^i(k)$ in this case, it is necessary to calculate the values of the function $g^i(x, \varpi)$ only at two points, irrespective of the space dimension $n$. If the functions $g^i(x, \varpi)$ have complex implicit structure and are defined, for example, on the solutions of differential equations (which is typical of ecological applications) or by simulation models, then the use of the random finite-difference direction $\xi^i(k)$ is advantageous over the standard finite-difference approximation of the gradient.

To solve the problem

$$\min_{x \in X} [F^0(x) = Eg^0(x, \varpi)] = F^*,$$

the stochastic quasigradient method suggests constructing a sequence of approximations $\{x^k, k = 0, 1, \ldots\}$ as follows [53]:

$$x^{k+1} = \Pi_X\{x^k - \rho_k \xi^0(k)\},\ k = 0,\ 1, \ldots,$$

where $\Pi_X\{\cdot\}$ is the operator of projecting onto the convex compact set $X$,

$$E\{\xi^0(k)|\ x^0, \ldots,\ x^k\} \in F_x^0(x),\ \ E\|\xi^0(k)\|^2 \leq \text{const},$$

the step factors satisfying the following conditions with probability one:

$$\rho_k \geq 0,\ \ \sum\nolimits_k E\rho_k = +\infty,\ \ \sum\nolimits_k E\rho_k^2 < \infty.$$

Under these assumptions, the sequence $\{x^k\}$ coincides with the optimal solution $x^*$ of the problem. The optimal value $F^*$ can be estimated from the independent observations $f^0(x^k, \varpi^k)$ of values of the subintegral function by the formula

$$F^k = (1/(k+1)) \sum\nolimits_0^k f^0(x^i, \varpi^i) \to F^* \text{ as } k \to \infty.$$

Let us dwell on one possible application of stochastic quasigradients where stochastic decomposition methods are used to solve multistage problems. The basic two-stage stochastic-programming model can be formulated as follows. Two stages of planning can be distinguished: the current state (first stage) and the future state (second stage). A decision $x \in R^n$ at the current instant of time is made under inexact information on the future state $w$. For example, the next year's production volume $x$ is planned under unknown demand and prices.

The task is to choose a long-term plan $x$ stable in some sense against possible variations of unknown parameters $w$ (demand, prices, resources, weather conditions, possible sociopolitical situations, etc.). The concept of stability of the solution $x$ is associated with its possible correction during the observation of a specific state $w$. To this end, correction mechanisms $y(x,\ w)$ that optimally react to the situation $w$ for any given $x$ are assumed. For example, for a random demand, it is assumed that absent products can be purchased to form the inventory. In hazardous production, there should be services to mitigate the consequences of possible emergencies. Then, a two-stage stochastic-programming problem can be represented as the following linear-programming problem: minimize the function

$$cx + \sum_{k=1}^{N} p_k d_k y_k \tag{2.27}$$

subject to the constraints

$$\begin{gathered} A_k x + D_k y_k = b_k,\ \ k = 1,\ 2, \ldots,\ N, \\ x \geq 0,\ y_k \geq 0,\ \ k = 1,\ 2, \ldots, N, \end{gathered} \tag{2.28}$$

where $x$ and $y_k$ are vectors of variables; $c, d_k, b_k, A_k$, and $D_k$ are vectors and matrices of corresponding dimensions; and $p_k$ are numbers such that $p_k \geq 0,\ \sum_{k=1}^{N} p_k = 1.$ To select such numbers (weight coefficients), many methods can be used to determine, in each specific case, the stochastic properties of the decomposition procedure. For example, in global models of interacting countries, the matrix $D_k$ describes the production capacity of an individual country, and the matrix $A_k$ defines its relationships with the outside world (common resources, migratory processes, environmental standards). The number $N$ can be very large. For example, if only the components of the vector of right-hand sides in the constraints of the second stage take two values independently, then $N = 2^m$, where $m$ is the number of such constraints. Therefore, if the number $m$ of constraints is relatively small, standard linear-programming methods become inapplicable. The stochastic approach to decomposing problems (2.27) and (2.28) into independent subproblems is as follows.

Fix $x \geq 0$ and let $y_k(x) \geq 0$ be a solution of the $k$th subproblem that has the following form: min $\{d_k y_k / D_k y_k = b_k - A_k x,\ y_k \geq 0\}$. Denote by $u_k(x)$ the dual variables that correspond to $y_k(x)$. Let $x^s$ be the current value of $x$ obtained after the $s$th iteration. In accordance with the probabilities $p_1, \ldots, p_N$, select the block (subproblem) $k_s$ and determine $u_k(x^s)$. In this case, the vector $\xi(s) = c - u_{k_s}(x^s) A_{k_s}$ is the stochastic generalized gradient of the function $F(x) = cx + \sum_{k=1}^{N} p_k d_k y_k(x)$.

Such an approach makes it possible to develop easily implementable stochastic decomposition procedures stable against random perturbations. They can be applied to optimize complex stochastic systems. These procedures do not need all the $N$ subproblems to be solved simultaneously at each iteration followed by complicated coordination of connecting variables $x$, which is due to the use of traditional deterministic approaches.

Formally, an elementary two-stage programming model is reduced to minimizing the function

$$F^0(x) = \mathrm{E} f^0(x,\ w),\ x \geq 0,$$

where $f^0(x,\ w) = cx + py(x,\ w) = cx + \min\{\ py / Dy = b - Ax,\ y \geq 0\}$. Note that all the coefficients $w = \{p,\ b,\ A,\ D\}$ can be random variables. The equations $Ax + Dy = b$, where $A$ and $D$ are some matrices, $x \in R^n$, $y \in R^r$, and $b \in R^m$, reflect the stochastic balances between cost and production output in the system being modeled. The quantity $py$ determines the cost of correcting the above-mentioned balance using the vector $y$. The objective function $F^0(x)$ of the model is convex but may generally be nondifferentiable since there is minimization operation under the expectation sign in the definition of $F^0(x)$. The values of the function $F^0(x)$ can be found analytically only in exceptional cases. If the state $w$ has a finite number of possible values, then the model formulated has a block structure, as discussed above.

The stochastic subgradient $\xi^0(s)$ of the objective function of a two-stage problem can be written as $\xi^0(s) = c + u(x^s, w^s) A(w^s)$, where $w^s$, $s = 0, 1, \ldots,$ are independent observations of random parameters $w$; $u(x^s, w^s)$ are dual variables that correspond to the optimal plan $y(x^s, w^s)$; $x^s$ is the current approximation to the optimal solution; and $A(w^s)$ is a random realization of the matrix $A$ that corresponds to the observation $w^s$. Thus, the use of a random vector $\xi^0(s)$ and one of the versions of the stochastic quasigradient procedure allows us to avoid virtually insuperable difficulties in evaluating multidimensional integrals (expectations).

Later on, Ermoliev assembled a team of graduates of the Mechanics and Mathematics Faculty and the Faculty of Cybernetics of the Taras Shevchenko National University of Kyiv and Moscow Institute of Physics and Technology. They did fundamental research in stochastic-programming theory, which was recognized worldwide. One of the main problems in stochastic optimization theory is to extend the concept of stochastic gradient to wider classes of stochastic extremum problems. The results obtained by Yu. M. Ermoliev, A. A. Gaivoronskii, A. M. Gupal, A. N. Golodnikov, V. I. Norkin, S. P. Uryas'ev, and others considerably extended the practical capabilities of direct stochastic-programming methods. Let us briefly discuss some of them. The stochastic quasigradient method was generalized to nonconvex stochastic-programming problems with weakly convex [131], generalized differentiable [60], almost differentiable [106], and locally Lipschitz [27] functions. The modifications of the method related to averaging of stochastic generalized gradients [27], solution of nonstationary stochastic problems [18, 131], analysis of limiting extreme stochastic-programming problems [18], and adaptive step control [174] were developed and analyzed. The concept of stochastic quasigradient was used in [48, 53, 106] to develop stochastic analogues of the following constrained optimization methods: methods of penalty functions, linearization, reduced gradient, Lagrangian multipliers, methods for solving stochastic minimax problems, etc. Mikhalevich analyzed random search procedures for solving optimization problems formulated in terms of binary preference relations. These relations are a mathematical model for the pairwise comparison of alternatives in interactive decision-making algorithms. The interpretation of such procedures as a special case of stochastic quasigradient methods made it possible to propose algorithms stable against random errors of decision makers. Due to the wide use of cluster computers, methods of parallel stochastic optimization developed in [50] were widely applied. It should be emphasized that stochastic quasigradient methods were widely used to solve general optimization problems with nondifferentiable and nonconvex functions. The possibility of applying these methods to problems with nondifferentiable functions is very important for the solution of applied problems with fast and unpredictable behavior of objects being modeled, for the optimization of complex systems under uncertainty. The scope of application was also substantially extended after the analysis of non-Markov (with dependent observations) stochastic optimization procedures whose convergence in some probabilistic sense was proved [61, 75].

The other series of studies was concerned with the asymptotic properties of stochastic quasigradient methods (limiting theorems, convergence rate) and confidence intervals for iterative stochastic optimization procedures. The results were reported in the monograph [75].

Of importance is the development of stochastic discrete optimization methods. To solve stochastic global and stochastic discrete optimization problems, Yu. M. Ermoliev, V. I. Norkin, B. O. Onishchenko, G. Pflug, A. Rushchinskii, and others developed the stochastic branch-and-bound algorithm.

To estimate the optimal values, use was made of the permuted relaxation of stochastic-programming problems, which implies permutation of the operations of minimization and expectation (or probability):

$$F^*(x) = \min_{x \in X} Ef^0(x, \varpi) \geq E \min_{x \in X} f^0(x, \varpi) = F_*(x),$$
$$P_* = \max_{x \in X} P\{f(x, \varpi) \in B\} \leq P\{\exists x(\varpi) \in X : f(x'(\varpi), \varpi) \in B\} = P^*(x).$$

Thus, to obtain the lower-bound estimate $F_*(X)$ of the optimal value $F^*(X)$, it is sufficient to efficiently solve the problem $\min_{x \in X} f(x,\ \theta)$. To obtain the upper-bound estimate $P^*(X)$ of $P(X)$, it is sufficient to test the conditions $f(x,\ \theta) \leq 0$ for compatibility for fixed $\theta$, which is possible for many applications (see [198] for some of them).

It should be emphasized that along with direct stochastic optimization, other stochastic-programming methods are also developed; they are outlined, for example, in the monographs [198, 211, 212]. One of such methods is the so-called method of empirical means. It approximates the expectation functions $F^i(x) = Eg^i(x, \varpi)$ by their empirical estimates

$$F^i(x,\ N) = (1/N) \sum_0^N g^i\left(x,\ \varpi^k\right)$$

and passes to approximate deterministic optimization problems of the form

$$\min \{F^0(x,\ N) |\ F^i(x,\ N) \leq 0,\ i = 1, \ldots,\ m;\ x \in X \subset R^n\},$$

to which well-known deterministic optimization methods can then be applied. Various aspects of the convergence of the method of empirical means (almost sure convergence with respect to functional or solution, convergence rate, and asymptotic convergence principle) were examined in the monograph [206] and in papers by Yu. M. Ermoliev, P. S. Knopov, V. I. Norkin, and M. A. Kaizer. The traditional and best-known approach to the analysis of this method was elaborated by R. Wets and his colleagues, and was widely developed in studying many classes of stochastic-programming problems. However, there is a wide range of problems that need new, alternative approaches to the analysis of the method of empirical means to be developed. In particular, stochastic optimization methods are widely applied to solve identification problems [54], where it is important to analyze the behavior of solutions with probability one. New approaches are widely used in modern asymptotic estimation theory. It is the method of empirical means that allows deeper understanding of how the stochastic optimization theory is related to the statistical estimation theory. It made it possible to describe all the main classes of statistical estimates and, moreover, to study new classes of robust estimates if there is a priori information on the unknown parameters and observations are dependent random variables. This, in turn, allowed developing new nonlinear and nonparametric methods to estimate the parameters of stochastic systems based on incomplete observations and a priori constraints for unknown parameters and studying their asymptotic properties. Such methods are often used in the theory of estimating parameters of random variables, processes, and fields with special (risk) functionals, in classification and recognition problems. The asymptotic properties of estimates and the general conditions for the convergence (in some probabilistic sense) of an approximation problem to the original one were established. This problem was also studied in the case of dependent observations and in the case where the unknown parameters were elements of some functional space. The results are outlined, for example, in [78, 206].

Along with establishing the convergence conditions for estimates, it is important to analyze their asymptotic distribution. A priori constraints for unknown parameters may cause estimates not to be asymptotically normal. The key result is that finding the limiting distribution reduces to some quadratic-programming problem [206]. Estimating the convergence rate of the method of empirical means is also very important. Theorems on large deviations of the difference in certain measure between an approximate value of an optimum point and its exact value were proved.

*Stochastic Optimization and Robust Bayesian Estimates*. The studies in this field involved finding Bayesian estimates under incomplete information on their a priori distribution that, however, is known to belong to some class of functions. Such a situation arises, for example, in estimating the unknown reliability parameters of high-technology systems where failure statistics for system elements is very poor, which makes it impossible to adequately choose a priori distribution. Under these conditions, calculating the Bayesian estimates becomes a nontrivial problem.

The idea of the so-called minimax approach is to find an estimate that would minimize the supremum of the objective functional with respect to a given class $\Gamma$ of a priori distribution functions.

A. N. Golodnikov, P. S. Knopov, P. Pardalos, V. A. Pepelyaev, and S. P. Uryas'ev (see, e.g., [207]) studied the choice of the following types of objective functionals:

1. Bayesian risk

$$r_H(\delta(x)) = \int\limits_{\Theta}\int\limits_{X} L(\theta,\ \delta(x))f(x|\theta)dxdH(\theta).$$

2. A posteriori risk

$$\phi_H(\delta(x)) = \int_\Theta L(\theta,\ \delta(x))dG(\theta|x) = \frac{\int_\Theta L(\theta,\ \delta)f(x|\theta)dH(\theta)}{\int_\Theta f(x|\theta)dH(\theta)}.$$

3. A posteriori expectation of the parameter $\theta$ for fixed sampled data $x$

$$\widehat{\theta}_H = \frac{\int_\Theta \theta f(x|\theta)dH(\theta)}{\int_\Theta f(x|\theta)dH(\theta)}.$$

To analyze the sensitivity of Bayesian estimates to the choice of the a priori distribution function from the class $\Gamma$, it is necessary to find the lower and upper bounds of the range of possible values of the objective functional. According to the Bayesian approach, if $H(\theta)$ is a "true" a priori distribution function, the quality of the Bayesian point estimate $\widehat{\theta}_H(x)$ is measured in terms of the Bayesian risk $r_H(\widehat{\theta}_H(x))$ or a posteriori risk $\phi_H(\widehat{\theta}_H(x))$. Let not one a priori distribution function $H(\theta)$ but rather a class $\Gamma$ of such distribution functions be available. In this case, the values of $r_H(\widehat{\theta}_H(x))$ or $\phi_H(\widehat{\theta}_H(x))$ can no longer be used to characterize the quality of the Bayesian point estimate $\widehat{\theta}_H(x)$. The range $(r_*(\widehat{\theta}_H(x)),\ r^*(\widehat{\theta}_H(x)))$ of possible values of the Bayesian risk or the range $(\phi_*(\widehat{\theta}_H(x)),\ \phi^*(\widehat{\theta}_H(x)))$ of possible values of a posteriori risk, which are defined for all the a priori distribution functions from the class $\Gamma$, are more adequate for this purpose.

For example, in estimating the reliability parameters for probabilistic safety analysis, it is important to know how wide the range $(\theta_*,\ \theta^*)$of possible values of the a posteriori expectation of the parameter $\theta$ is for fixed sample data $x$ obtained for any a priori distribution function from the class $\Gamma$.

To find the lower $r_*(\widehat{\theta}_H(x))$ and upper $r^*(\widehat{\theta}_H(x))$ bounds of the range $(r_*(\widehat{\theta}_H(x)),\ r^*(\widehat{\theta}_H(x)))$ of possible values of the Bayesian risk for the estimate $\widehat{\theta}_H(x)$ obtained on the assumption that each distribution function $G(\theta) \in \Gamma$ is a true a priori distribution function, it is necessary to solve the following optimization problems in the space of distribution functions:

$$r_*(\widehat{\theta}_H(x)) = \inf_{G\in\Gamma} \int_\Theta \int_X L(\theta,\ \widehat{\theta}_H(x))f(x|\theta)dx\ dG(\theta),$$

$$r^*(\widehat{\theta}_H(x)) = \sup_{G\in\Gamma} \int_\Theta \int_X L(\theta,\ \widehat{\theta}_H(x))f(x|\theta)dx\ dG(\theta).$$

They are stochastic-programming problems, where the optimization is in the space of distribution functions and the objective functionals are linear in the distribution functions.

To find the lower $\phi_*(\widehat{\theta}_H(x))$ and upper $\phi^*(\widehat{\theta}_H(x))$ bounds of the range $(\phi_*(\widehat{\theta}_H(x)),$ $\phi^*(\widehat{\theta}_H(x)))$ of possible values of the a posteriori Bayesian risk for the estimate $\widehat{\theta}_H$ obtained on the assumption that each distribution function $G(\theta) \in \Gamma$ is a true a priori distribution function, it is necessary to solve the following optimization problems in the space of distribution functions:

$$\phi_*(\widehat{\theta}_H(x)) = \inf_{G \in \Gamma} \frac{\int_\Theta L(\theta, \widehat{\theta}_H(x)) f(x|\theta) dG(\theta)}{\int_\Theta f(x|\theta) dG(\theta)},$$

$$\phi^*(\widehat{\theta}_H(x)) = \sup_{G \in \Gamma} \frac{\int_\Theta L(\theta, \widehat{\theta}_H(x)) f(x|\theta) dG(\theta)}{\int_\Theta f(x|\theta) dG(\theta)}.$$

To find the lower $\theta_*$ and upper $\theta^*$ bounds of the range $(\theta_*, \ \theta^*)$ of possible values of the a posteriori expectation of the parameter $\theta$ for fixed sample data $x$ obtained on the assumption that each distribution function $G(\theta) \in \Gamma$ is a true a priori distribution function, it is necessary to solve the following optimization problems in the space of distribution functions:

$$\theta_* = \inf_{G \in \Gamma} \frac{\int_\Theta \theta f(x|\theta) dG(\theta)}{\int_\Theta f(x|\theta) dG(\theta)}, \quad \theta^* = \sup_{G \in \Gamma} \frac{\int_\Theta \theta f(x|\theta) dG(\theta)}{\int_\Theta f(x|\theta) dG(\theta)}.$$

The above-mentioned problems are stochastic-programming problems where optimization is in the space of distribution functions and the objective functionals are linear in the distribution functions.

The methods to solve these rather complex stochastic optimization problems in functional spaces and algorithms of their numerical implementation were developed and computational experiments were conducted. The results for classes of distribution functions with bounded moments or given quantiles turned out to be of special interest. For these cases, a stochastic optimization problem was proved to be reducible to a linear-programming problem. Robust Bayesian estimates, the lower and upper bounds for the estimates, which are a measure of indeterminacy due to a limited a priori information were found. The proposed methods were shown to substantially improve the quality of these estimates in estimating reliability parameters.

*Some Problems of Controlled Discrete Stochastic Systems*. The fundamental studies by V. S. Mikhalevich pioneered the research of control of random processes at the Institute of Cybernetics. They initiated a new area in the analysis of controlled random processes based on the generalization of the well-known Bayesian methods of stochastic decision theory. It is these studies that stimulated the development of various areas in the theory of controlled random processes and fields. The general theory of controlled Markov processes with discrete time crystallized in Wald's studies on sequential analysis and was then developed by R. Bellman, R. Howard, A. N. Shiryaev, A. V. Skorokhod, and others. Scientists from the V. M. Glushkov Institute of Cybernetics also greatly contributed to the theory of controlled random processes.

Let us formulate the control problem for stochastic discrete-time systems.

Let $X$ and $A$ be separable spaces, and $\Phi$ and $\Lambda$ be the $\sigma$-algebras of Borel subsets of $X$ and $A$. Let there be given a mapping $F$ (multiple-valued function) that associates each $x \in X$ with a nonempty closed set $A_x \subset A$ so that the set $\Delta = \{x \in X,\ a \in A_x\}$ is Borel measurable in $X \times A$. The random evolution of the system is controlled by the set of transition probabilities $P\{B/x_n, a_n\} = P\{X_{n+1} \in B/X_0 = x_0,\ D_0 = a_0,\ \ldots,\ X_n = x_n,\ D_n = a_n\}$, where $B \in \Phi$; $(x_k,\ a_k) \in \Delta$; $X_k$ is the state of the system at the time $k$, $D_k$ is the decision made at the time $k$, and $A_k$ is the control chosen at the time $k$, $k \leq n$.

Denote by $r(x,\ a)$ the expected losses over one period if the system is in the state $x$ at the beginning of the period and a decision $a \in A_x$ is made. The function $r(x,\ a)$ is assumed to be bounded and measurable on $\Delta$, $|\ r(x,\ a)\ | \leq C < \infty,\ (x,\ a) \in \Delta$.

The general feasible system control strategy is a sequence $\delta = \{\delta_1, .., \delta_n ...\}$ such that the probabilistic measure $\delta_n(\cdot/x_0, \ldots,\ x_n)$ on $(A,\ \Lambda)$ is concentrated on $A_{x_n}$. A strategy is called stationary Markov if $\delta_n(\cdot/x_0, \ldots, x_n) = \delta(\cdot/x_n),\ n = 0,\ 1, \ldots$. A stationary Markov strategy is called deterministic if the measure $\delta(\cdot/x_n)$ is concentrated at one point for any $x$.

Denote the class of all feasible strategies by $R$ and the class of stationary Markov deterministic strategies by $R_1$ and consider two strategy optimality criteria:

1. Average cost for the chosen strategy $\delta$

$$\varphi(x,\ \delta) = \lim_{n\to\infty} sup \frac{1}{n+1} \mathrm{E}_x^\delta \sum_{k=0}^{n} r(x_k,\ D_k).$$

2. Expected (discounted) total cost with $\beta \in (0,\ 1)$ and strategy $\delta$

$$\Psi_\beta(x,\ \delta) = \mathrm{E}_x^\delta \sum_{k=0}^{n} \beta^k r(x_k,\ D_k),$$

where $\mathrm{E}_x^\delta$ is the conditional expectation corresponding to the process following the strategy $\delta$ under the condition $x_0 = x$.

A strategy $\delta^*$ is assumed optimal ($\varphi$-optimal or $\Psi_\beta$-optimal) with respect to these criteria if $\varphi(x,\ \delta^*) = \inf_{\delta\in R} \varphi(x,\ \delta)$ or $\Psi_\beta(x,\ \delta^*) = \inf_{\delta\in R} \Psi_\beta(x,\ \delta)$, respectively. The existence and uniqueness theorems for $\varphi$ and $\Psi_\beta$ strategies in $R_1$ were proved.

Let us consider an example of applying the optimal stochastic control theory to a modern inventory problem. Consider an inventory control system for one type of product, which can be replenished in continuous units. The maximum storehouse capacity is $Q$; thus, the inventory level takes values from the interval $[0, Q]$. The inventory level is checked periodically at discrete instants of time $n$, and a decision is then made to order an additional amount of goods. If the inventory level is $X_n = x \in [0, Q]$ at time $n \in N$, amount $D_n \in A_x = [0, Q - x]$ is ordered. The order is delivered instantaneously with probability $p \in (0, 1]$ and is lost with probability $1 - p \in [0, 1)$.

Let $\eta = (\eta_n : n \in N)$ be a Bernoulli sequence, where $\eta_n = 1$ means that the order made at the time $n$ has been delivered. We assume that $\eta_n$ does not depend on the history of the system until the time $n$ inclusive.

There is a random demand $\xi_n$ at the time $n$; $\xi = (\xi_n : n \in N)$ is a sequence of equally distributed independent random variables with continuous distribution $G(x)$, $x \geq 0$. These quantities are assumed to be independent of the history of the system until the time $n$ inclusive and, moreover, $G(Q) < 1$.

The demand at the time $n$ is satisfied by the quantity $X_n + D_n\eta_n$ at the end of the time interval $[n, n + 1)$ if possible. In the case of full or partial shortage of the goods, orders are not reserved but lost. The evolution equation for the inventory level has the form

$$X_{n+1} = (X_n + D_n\eta_n - \xi_n)_+, \ n \in N,$$

where $(a)_+ = max\,\{a, \ 0\}$ is the positive part of $a \in R$.

The assumptions guarantee that for the above Markov strategy, the sequence $X = (X_n : n \in N)$ is a homogeneous Markov chain for which the transition probabilities can easily be determined. Thus, if we assume that all the random variables for the model are defined on the common basic probabilistic space $(\Omega, \Im, P)$, we have a controlled random process with discrete time, phase space $X = [0, Q]$, and control space $A = \{d_a, \ a \in [0, Q]\}$, where $d_a$ is a decision to order goods in amount $a$. In the state $x$, the set of feasible controls takes the form $A_x = \{d_a, \ a \in [0, \ Q - x]\}$.

Let the function $c(x)$ related to inventory supply be linearly dependent on the amount of the order to be delivered, and the loss function $f(x)$ related to inventory storage and deficiency be convex.

As was proved, $\varphi$-optimal and $\Psi_\beta$-optimal deterministic strategies exist for the model; moreover, the conditions are established whereby the optimal strategies have two-level structure, that is, there exists a point $x^* \in [0, Q]$ such that the optimal strategy has the form

$$\delta^* = \begin{cases} d_{Q-x}, & x \leq x^*, \\ d_0, & x > x^*. \end{cases}$$

Such inventory strategies are called $(s, S)$-strategies.

Similar strategies will also be optimal for some other models from queuing, reliability, and inventory theories. These results were obtained by P. S. Knopov, V. A. Pepelyaev, and their colleagues. Thus, finding optimal strategies reduces to some optimization problem; to this end, well-known methods can be applied. The scientists of the Institute of Cybernetics pay much attention to the development of numerical methods to find optimal strategies for stochastic dynamic systems. The monograph [55] outlines the finite-difference methods for deterministic and stochastic lumped- and distributed-parameter systems.

Currently, the theory of stochastic systems with local interaction is widely used to solve complex biological, engineering, physical, and economic problems. An example of such systems is Gibbs random fields and Ising fields, which are widely used to solve many problems in statistical physics. Under some natural conditions, these fields are Markov, which makes them useful for the description and solution of real-world problems in order to model them and make optimal decisions. Creating the optimal control theory for stochastic systems with local interaction, developing methods to estimate unknown parameters, and finding optimal solutions during their operation are important tasks. The monograph [217] outlines the fundamentals of this theory. It has numerous applications in recognition theory, communication network theory, economics, sociology, etc. For example, a Markov model that describes the behavior of companies with competing technologies is analyzed in [217]. It is required to develop a recognition strategy that makes a clever (in a sense) decision based on known observation. One of the best-known approaches in recognition theory is that based on modeling the behavior of unknown objects using Gibbs random fields, which are Markov fields with distribution function of known form dependent on unknown parameters. A recognition problem reduces to estimating these parameters. One more example of applying Markov random fields will be given below to describe catastrophe risk models.

*Methods of Catastrophe Risk Insurance.* The insurance of catastrophe risks is closely related to the previous subject. The question is not only technogenic and ecological catastrophes but also the insurance of events related to financial turmoil. The number of natural disasters has considerably increased for the last two decades, and relevant economic losses have increased fivefold. A series of interesting and important studies in this field is outlined, for example, in [56, 161].

As an example, we will discuss the activity of insurance companies intended to earn the maximum profit under a priori given constraints. The real expenses of such companies are uncertain. Actually, the cost of insurance contracts remains unknown for a long time, the total cost of insurance contracts or a portfolio not being additive.

As a rule, traditional insurance deals with independent risks of comparatively high frequency and small losses, which allows using the law of large numbers to formulate individual insurance strategies. The situation will drastically change if there are accidents with significant human and financial losses. Insurance companies usually allow for possible risks, each being characterized by the distribution of damage, which is mostly expressed in monetary terms. In other words, each risk is associated with possible losses $w_i$, with $w_i$ and $w_j$, $i \neq j$, being interdependent random variables.

For example, an earthquake causes damage on a wide area, destructing buildings and roads, causing fire, etc. The risk portfolio is characterized by a solution vector $x = (x_1, \ldots, x_n)$, and its total value is defined by

$$c(x, w) = \sum_{i=1}^{n} c_i(x_i, w_i),$$

where the random insurance function $c_i(x_i, w_i)$ is generally nonconvex and nonsmooth with respect to the solution vector $c$.

Note that the central limit theorem is true for $c(x, w)$ under rather general conditions in the case of independent values. This allows using approaches that are based on the method of empirical means. Nevertheless, as indicated above, catastrophes cause dependent risks $w_1, \ldots, w_n$, and there is no certainty that summing risks will result in normal distribution. One more feature of such problems is the excessive sensitivity of the constraints appearing in the optimization problem to the joining of distribution. This may result in discontinuous functions and make it impossible to apply statistical approximations based on sample averages.

Noteworthy is another feature of catastrophe risk portfolio construction problems. The dependence of damage on the space and time coordinates is of crucial importance in many cases. The concurrence of two rare events in time and at a certain point has extremely small probability. The explicit form of insurance portfolio in the dynamic case can be obtained by introducing appropriate variables.

The function $c(t, x, w)$ can be introduced by the formula

$$c(t, x, w) = \sum_{i=1}^{n} \sum_{k=1}^{N_i(t)} c_i(t_i, x_i, w_{ik}),$$

where $N_i(t)$ is a random point process, $w_{ik}$ is the realization of the random variables $w_i$ of $i$th-type damage, $x = (x_1, \ldots, x_n)$ is the full vector of variables of the problem with values in some set $X$.

If we denote by $R(T, x)$ the profit of an insurance company and by $\{0, T\}$ the time horizon of insurance business planning, then to find the optimal insurance contract portfolio, the insurant should solve the optimization problem

$$F_T(x) = \mathrm{E}R(T, x) I_{\{\tau(x) > T\}} \to \max_{x \in X}$$

subject to the constraints imposed on the probability of ruin in time $T$

$$\Psi_T(x) = P\{v(x) \leq T\} \leq \alpha,$$

where $\tau(x)$ is a stopping time such that $\min \mathrm{R(T,\ x)} \geq 0$ and $R(v(x), x) \leq 0$; $I_A$ is the indicator function of the set $A$.

The functions $F_T(x)$ and $\Psi_T(x)$ can be discontinuous for continuous distributions of the functions $R(T, x)$, which makes traditional stochastic-programming methods unacceptable to solve the problem posed. Thus, new methods are necessary. Some of them are outlined in [27] and those for insurance problems in [56].

One of such methods consists in stochastic smoothing of a risk process by replacing the original function with a sufficiently smooth approximation.

Modeling of accidents was discussed in [217].

*Some New Approaches to Estimating Financial Risks*. Risk management is one of the key problems in asset allocation by banks, insurance and investment companies, and other risk assessment financial institutions.

Credit risk is due to a trading partner who defaults on his duties. There are many approaches to the assessment of such risk. The best-known approach is based on estimating so-called Value-at-Risk (VaR for short), which has become a standard of risk assessment and management. VaR is known to be defined as the $d$-quantile of an investment portfolio loss function. Constructing a portfolio with predetermined constraints imposed on VaR or with minimum admissible VaR is an important problem.

In his Dr.Sci. thesis, V. S. Kirilyuk developed a new approach to risk assessment. It consists in estimating conditional average loss CVaR, which exceeds $\alpha$-VaR. In many cases, CVaR appeared better than VaR. Though CVaR has not become a standard risk measure in the financial industry yet, it plays an important role in financial and insurance mathematics.

## 2.5 Discrete Optimization Methods

An important field of studies focused on the development of scientific foundation in modern computer technologies [163] is the development of optimization (in particular, discrete optimization) models and methods [17, 29, 105, 109, 116, 124, 153, 156, 158, 159]. As of today, mathematical models of discrete optimization, which has evolved for more than 40 years, cover a wide range of applied problems related to the efficient solution of numerous problems of the analysis and optimization of decision making in various scientific and practical fields. It is thus of importance to develop a mathematical theory, to improve the available mathematical models and methods as well as software and algorithms, and to develop new ones to solve complex discrete optimization problems. Nevertheless, the majority of discrete optimization problems are usually universal (NP-hard), which makes it hardly possible to create linear and convex optimization methods for them, which are as efficient as the available ones. The computational difficulties that arise in solving discrete optimization problems are often due to their high dimension, multiextremality, specific structure, multicriteriality, weak structure, and incomplete and uncertain input information on their parameters. This obviously makes it impossible to develop acceptable exact methods for most classes of discrete problems. Moreover, mathematical models (including discrete ones) of applied problems of finding best solutions do not usually represent real situations but rather approximate them. Therefore, it is expedient to solve discrete optimization problems with appropriate degree of approximation to the optimum taking into account the possible inaccuracy and incorrectness of the mathematical models. It is thus important to develop and analyze various approximate methods for solving these problems (these are usually methods that do not guarantee the optimal solution of the problem). It is this way that could lead to a significant effect of applying mathematical methods and computers to solve complex applied problems of discrete optimization.

The development and active use of various modifications of the method of sequential analysis of variants [17] in analyzing important economic problems gave rise to other general and special algorithms of solving multiple-choice problems.

In the early 1960s, in parallel with the method of sequential analysis of variants, I. V. Sergienko proposed [164] a new approach to develop algorithms for the approximate solution of discrete optimization problems of general form:

$$\min_{x \in G} f(x), \tag{2.29}$$

where $G$ is a finite or countable set of feasible solutions of the problem and $f(x)$ is the objective function defined on this set. A new algorithm, so-called descent vector algorithm, was created. It makes it possible to develop a wide range of new local search algorithms and to select the most suitable one after the analysis of a specific problem.

The descent vector method may be classed among probabilistic local search methods due to the random choice of the initial approximation and enumeration of points in the neighborhoods.

Let us detail the descent vector method since it was well developed by Sergienko, his disciples, and various experts (e.g., [26, 124, 153, 156, 158, 159]) and was used to create new methods and various computer-aided technologies to solve many classes of discrete programming problems.

A point $x \in G$ will be called the point of local minimum of a function $f(x)$ with respect to a neighborhood $N(x)$ if this neighborhood is not degenerate (is not empty) and any of its points satisfies the condition $f(x) \le f(y)$.

Let for $k = 1, \dots, r,\ r \ge 1$, systems of neighborhoods $N^k(x)$ be given, such that $\forall x \in G\ N^j(x) \subseteq N^i(x)$ if $i, j \in \{1, \dots, r\},\ j \le i$.

A vector function $\Delta^k(x)$ specified for each point $x \in G$ and neighborhood $N^k(x)$, $k = 1, \dots, r$, is called the descent vector of the function $f$ with respect to the neighborhood $N^k(x)$ if the following holds:

1. At each point $x \in G$, the value of the function $\Delta^k(x)$ is a $q$-measurable ($q = q(k, x) = |N^k(x)|$) vector with the coordinates $\Delta_1^k(x), \dots, \Delta_q^k(x)$, which are real numbers, where $|N^k(x)|$ is the cardinality of the set $N^k(x)$.
2. $x \in G$ is a point of local minimum of the function $f$ with respect to the neighborhood $N^k(x)$ if and only if $\Delta_i^k(x) \ge 0$ for all $i = 1, \dots, q$.

3. If $x \in G$ is not a point of local minimum of the function $f$ with respect to the neighborhood $N^k(x)$, then the descent vector can be used to find a point $y \in N^k(x)$ such that $f(y) < f(x)$.

By definition, vector $\Delta^k(x)$ allows finding the direction of decrease ("descent") of values of the function $f$ in the neighborhood $N^k(x)$ for each point $x \in G$. The idea of using the descent vector $\Delta^k(x)$ to find such a direction underlies the method being considered.

In most cases, it is much easier to calculate the coordinates $\Delta_1^k(x), ..., \Delta_n^k(x)$ of the descent vector than the values of the function $f$ at points of the neighborhood $N^k(x)$. Moreover, to find a point $y$ in $N^k(x)$ such that $f(y) < f(x)$ or to establish that $f(x)$ is locally minimal with respect to this neighborhood, it will suffice to restrict the consideration to only some coordinates of the vector $\Delta^k(x)$.

The general scheme of the descent vector method can be presented as follows:

procedure **descent vector method**
**initial_setting** $(s,\ x^s, s_{max},\ \{I_k\}), 1 \le k \le r$
*while* $(s \le s_{max})$ do
$k = 1$
**calculate_descent_vector** $(\Delta^{I_k}(x^s))$
**while** $(k \le r)$ do
**find_correction** $(\Delta^{I_k}(x^s), found)$
if *found* = "FALSE" then
**comment** the local minimum of the function $f$ with respect to the neighborhood $N^k(x^s)$ has been found
$k = k + 1$
**calculate_descent_vector** $(\Delta^{I_k}(x^s))$
else
$x^{s+1} =$ **new_point** $(N^k(x^s))$
$s = s + 1$
end if
end while
if $(k > r)$ break
end while
end

The procedure **initial_setting** specifies the initial data: the number of iterations $s = 0$, point $x^s \in G$, maximum number of iterations $s_{\max}$, and a sequence $\{I_k\} = = I_1, ..., I_r$ of neighborhood numbers such that $I_1 < ... < I_r$.

The procedure **find_correction** searches for a negative coordinate of the descent vector $\Delta^{I_k}(x^s)$, $k \in \{1, ..., r\}$, $s \in \{0, ..., s_{max}\}$. A solution $y \in N^k(x^s)$ such that $f(y) < f(x^s)$ corresponds to this coordinate. After the procedure is executed, the variable *found* is assigned "TRUE" if a negative coordinate is found and "FALSE" otherwise.

The procedure **calculate_descent_vector** calculates (recalculates) the descent vector, and the procedure **new_point** uses some rule to choose a new solution $x^{s+1} \in N^k(x^s)$ such that $f(x^{s+1}) < f(x^s)$.

The algorithm of descent vector converges, which follows from the theorem proved in [153].

Only few steps of the algorithm require considering the neighborhood with the maximum number $I_r$. For most steps, it is sufficient to calculate the coordinates of the descent vector of the function $f(x)$ in the neighborhood $N^{I_1}(x^s)$. If neighborhoods with numbers larger than $I_1$ should be considered at some step, it will suffice to consider the set $N^{I_j}(x^s) \backslash N^{I_{j-1}}(x^s)$ instead of any such neighborhood $N^{I_j}(x^s)$, $j \in \{2, ..., r\}$, and to analyze only the coordinates of the descent vector $\Delta^{I_j}(x^s)$ that correspond to points of this set.

The algorithm of the descent vector is terminated once the number of steps has exceeded a preset value. It is also possible to assess how the value $f(x^s)$ obtained at each $s$th step meets practical requirements and to terminate the computational process if they are satisfied. The algorithm can also be terminated after some preset period $t$ if the corresponding extremum problem is solved in a real-time decision-making system. In this case, naturally, the accuracy of the solution is strongly dependent on $t$.

In decision-support systems with participation of decision makers, they can stop the computational process based on informal considerations.

To complete the description of the descent vector method, we will dwell on the possible case where the local solution $x^*$ found by the algorithm does not satisfy the researcher (user). To find other approximate solutions, the following ways are proposed:

1. Increase the neighborhood number $I_r$ and continue the computational process according to the algorithm, based on the obtained point $x^*$ as a new initial approximation. This procedure can be repeated several times. The feasible solutions produced by such procedures will obviously be better than $x^*$. Nevertheless, the computational process can be significantly complicated for large neighborhoods.
2. Increase the neighborhood number $I_r$ and perform only one step of the algorithm to obtain point $x'$ such that $f(x') < f(x^*)$. It is then recommended to come back to the initial neighborhood number $I_1$ and continue the computational process according to the algorithm.
3. Choose another initial approximation $x^{01} \in G$ and repeat computations based on this new approximation but with the same neighborhood number. The algorithm will generally result in another local solution. Repeated search for local minimum with a fixed neighborhood number and different initial approximations will obviously yield several local solutions; the solution is one that minimizes the function $f(x)$.
4. Use the following way: If there is a point $x'$ in the neighborhood $N^{I_r}(x^*)$ such that $f(x') \neq f(x^*)$, then the algorithm can be continued by passing to the next step based on the point $x'$. Such a supplement to the basic algorithm does not usually involve intensive additional computations and, at the same time, can result in a more accurate solution in some cases.

Noteworthy is that the descent vector method was successfully used to solve many practical problems [124, 153, 156, 158, 159] formulated in terms of integer and combinatorial optimization. It is easy to use and combine it with other approximate algorithms.

The idea of regular change of the neighborhood of local search in the descent vector scheme was used in [216] to develop an efficient local-type algorithm with variable neighborhood. This algorithm and its modifications were successfully used to solve the traveling salesman and median problems and to find extremum graphs.

The GRASP algorithm [199], which fits well into the general scheme of the descent vector method, was first used to solve a complex graph covering problem and then other problems as well.

A school known to the scientific community was established at the V. M. Glushkov Institute of Cybernetics; it covers a wide range of modern problems in discrete optimization and systems analysis. Scientists of the Institute obtained fundamental scientific results in this field and analyzed applied problems for which real economic, engineering, social, and other characteristics and constraints are known. To solve such real problems, mathematical models and methods were developed; they allow obtaining solutions with a practicable accuracy and spending allowable amount of resources such as time and number of iterations.

These methods were incorporated into application software packages (ASP) created (and mentioned above in Chap. 1) at the Institute of Cybernetics and widely used to solve applied problems. They include DISPRO ASP for solving various discrete optimization problems [153], the PLANER ASP intended for high-dimensional manufacturing and transportation scheduling problems [116], etc. The experience gained in applying these packages underlies modern computer technologies.

As the application range widens, discrete programming models become more and more complicated [159, 160]. The dimension of actual applied discrete optimization problems that arise at the current stage exceeds by an order of magnitude the dimension of such problems solved earlier. This means that even the best discrete optimization methods fail to solve them, and the need arises to create essentially new mathematical methods that would be a combination, a system of interacting algorithms that purposefully collect and use information on the problem being solved and adjust the algorithm to the specific problem.

The recent studies carried out at the Institute of Cybernetics to develop new advanced probabilistic discrete optimization methods and compare them with the available ones based on a large-scale computational experiment on solving various classes of discrete optimization problems show the domination of the methods that combine various ideas developing within the framework of the local optimization theory. They helped to resolve a great many real-world challenges (pattern recognition, object classification and arrangement, design, scheduling, real-time control of processes, etc.).

One of such methods is the probabilistic global equilibrium search (GES) intended to solve complex discrete optimization problems [181]. It develops the descent vector method and uses the ideas of simulated annealing method. The GES method is based on repeated application of a "temperature" loop, where a series of iterations is carried out to choose points from the set of elite solutions and improve the obtained solutions using the tabu search algorithm. After finding a local optimum, the new method obtains a new initial point in a random way, the probabilities of its generation being determined by formulas similar to those used in the simulated annealing method.

Despite the simulated annealing method is successfully used to solve many complex optimization problems, its asymptotic efficiency is even lower than that of the trivial repeated random local search. The global equilibrium search has all the advantages, but not disadvantages, of the simulated annealing method. The results of numerous computational experiments on solving various classes of optimization discrete programming problems by the GES method [159, 160] suggest that this trend in the development of discrete optimization methods is promising.

To describe the global equilibrium search method, we will use the following formulation of a discrete optimization problem:

$$\min \{f(x) \mid x \in G \cap B^n\}, \tag{2.30}$$

where $G \subset R^n$, $R^n$ is the set of $n$-measurable real vectors, and $B^n$ is the set of $n$-measurable vectors whose coordinates take the values 0 or 1.

Denote $S = G \cap B^n$. Suppose $Z(\mu) = \sum_{x \in S} exp\,(-\mu f(x))$ for any $\mu \geq 0$. Define a random vector $\xi(\mu, \varpi)$ :

$$\pi\,(x,\ \mu) = P\{\xi(\mu,\ \omega) = x\} = exp\,(-\mu f(x))/Z(\mu),\ \ x \in S. \tag{2.31}$$

Distribution (2.31) is the Boltzmann distribution well known in statistical physics. It is related to simulated annealing method: Expression (2.31) determines the stationary probability of the fact that a Markov chain generated by this method is at the point $x$.

Let $S_j^1 = \{x \mid x \in S,\ x_j = 1\}$ and $S_j^0 = \{x \mid x \in S, x_j = 0\}$ be subsets of the set of feasible solutions of problem (2.30), where the $j$th, $j = 1, \ldots,\ n,$ coordinate is equal to 1 and 0, respectively.

For $j = 1, \ldots,\ n,$ determine

$$Z_j^1(\mu) = \sum_{x \in S_j^1} exp\,(-\mu f(x)),\ \ Z_j^0(\mu) = \sum_{x \in S_j^0} exp\,(-\mu f(x)),$$

$$p_j(\mu) = P\{\xi_j(\mu,\ \omega) = 1\} = \frac{Z_j^1(\mu)}{Z(\mu)}.$$

Suppose that the set $\tilde{S}$ is a subset of the set of feasible solutions of problem (2.30) found by the GES method and

$$\tilde{S}_j^1 = \{x|\ x \in \tilde{S}, x_j = 1\},\ \tilde{S}_j^0 = \{x|\ x \in \tilde{S}, x_j = 0\},\ j = 1, ...,\ n.$$

Let us present the general GES scheme.

procedure GES
1. **initialize_algorithm_parameters** $(maxiter, ngen, maxnfail, \{\ \mu_k\}, 0 \leq k \leq K)$
2. while (stoppage_criterion = FALSE) do
3. if $(\tilde{S} = \emptyset)$ then $\{\tilde{S}$ is the set of known solutions$\}$
4. $x \leftarrow$ **construct_solution**
   **comment** constructing a random feasible solution $x$
5. $g(x) \leftarrow$ **construct_decay_vector**$(x)$
6. $x_{max} = x$; $x_{best} = x$; $g_{max} = g(x)$
7. $\tilde{S} \leftarrow x_{max}$
8. $Elite \leftarrow x_{max}$ $\{Elite$ is the set of elite solutions$\}$
9. end if
10. $nfail \leftarrow 0$ $\{nfail$ is the number of full loops without correction$\}$
11. while $(nfail < maxnfail)$ do
12. $k \leftarrow 0$
13. $old\_x_{max} \leftarrow x_{max}$
14. while $(k < K)$ do
15. **calculate_probabilities_of_generation** $(pr^k, \tilde{S})$
16. $t \leftarrow 0$
17. while $(t < ngen)$ do
18. $x \leftarrow$ **generate_solution** $(x_{max}, pr^k)$
19. $R \leftarrow$ **search_method** $(x)$
    **comment** $R$ is the set of found feasible solutions
20. $R \leftarrow R \backslash P$ $\{P$ is the set of infeasible solutions$\}$
21. $\tilde{S} \leftarrow \tilde{S} \cup R$
22. $x_{max} \leftarrow argmin f(x)$
23. $g(x_{max}) \leftarrow^{x \in \tilde{S}}$ **calculate_descent_vector** $(x_{max})$
24. if $(f(x_{max}) < f(x_{best}))$ then
25. $x_{best} \leftarrow x_{max}$
26. end if
27. **form_elite** $(Elite,\ R)$
28. $t \leftarrow t + 1$
29. end while
30. $k \leftarrow k + 1$
31. end while
32. if $(f(old\_x_{max}) = f(x_{max}))$ then $nfail \leftarrow nfail + 1$
33. else $nfail \leftarrow 0$
34. $\tilde{S} \leftarrow Elite$
35. end while
36. $P = P \cup N(x_{best}, d_p)$
37. $Elite = Elite \backslash P$
38. if (RESTART-criterion = TRUE) then $Elite \leftarrow \oslash$

39. $\tilde{S} \leftarrow Elite$
40. $x_{max} = \underset{x \in \tilde{S}}{argmin} f(x)$
41. $g(x_{max}) \leftarrow$ **calculate_descent_vector** $(x_{max})$
42. end while
end GES

In developing GES-based algorithms, use is made of rules that allow:

- Using the information obtained in the problem solution
- Intensifying and diversifying the search of an optimum
- Using the principles of RESTART technology [159]
- Combining different metaheuristics and forming a "team" of algorithms for the optimal use of the advantages of the algorithms and the specific features of the problem

Let us detail the design of GES algorithms based on these rules. The outer loop (iteration loop, lines 2–42) is intended for repeated search for the optimal solution. Together with the operators in lines 37–39, it allows using the RESTART technology. This loop either performs a prescribed number of repetitions or is executed until some stopping criterion. The stopping criterion may be, for example, a solution with a value of the objective function better than a known record. An essential element of the GES algorithm is the "temperature" loop (lines 14–31), which needs the values of $K$ and "temperature" $\mu_k,\ k = 0, \ldots, K,$ to be specified. This loop starts a series of searches for the optimal solution for increasing "temperatures." The "temperature" loop and its repetitions allow flexible alternation of narrowing and expanding of the search zone, which eventually contributes to the high efficiency of the GES method.

If the set $\tilde{S}$ of solutions known at some moment is empty, the operators of lines 4–9 allow finding the first solution and initializing the necessary data. Note that the solutions found by the GES algorithm are not saved (except for elite solutions, which appear in the set *Elite*) and are used to calculate the quantities

$$\tilde{Z}_k = \sum_{x \in \tilde{S}} exp\{-\mu_k f(x)\},\ \ \tilde{F}_k = \sum_{x \in \tilde{S}} f(x) exp\{-\mu_k f(x)\},\ \ \tilde{E}_k = \frac{\tilde{F}_k}{\tilde{Z}_k},$$

$$\tilde{Z}^l_{kj} = \sum_{x \in \tilde{S}^1_j} exp\{-\mu_k f(x)\},\ \ \tilde{F}^l_{kj} = \sum_{x \in \tilde{S}^1_j} f(x) exp\{-\mu_k f(x)\},\ \ \tilde{E}^l_{kj} = \frac{\tilde{F}^l_{kj}}{\tilde{Z}^l_{kj}},$$

$$k = 0, \ldots, K,\ \ l = 0,\ 1,\ \ j = 1, \ldots, n. \tag{2.32}$$

They allow determining the probabilities $\tilde{p}^k = (\tilde{p}_1{}^k, \ldots, \tilde{p}_n{}^k)$ according to one of the following formulas (see, e.g., [159]):

$$\tilde{p}^k_j = \tilde{p}^k_0 exp\left\{\frac{1}{2}\sum_{i=0}^{k-1}(\tilde{E}^0_i + \tilde{E}^0_{i+1} - \tilde{E}^1_{ij} - \tilde{E}^1_{i+1j})(\mu_{i+1} - \mu_i)\right\}, \tag{2.33}$$

$$\tilde{p}_j^k = \frac{1}{1 + \frac{1-\tilde{p}_0^k}{\tilde{p}_0^k} exp\left\{-0.5 \sum_{i=0}^{k-1} (\tilde{E}_{ij}^0 + \tilde{E}_{i+1j}^0 - \tilde{E}_{ij}^1 - \tilde{E}_{i+1j}^1)\,(\mu_{i+1} - \mu_i)\right\}}, \tag{2.34}$$

$$\tilde{p}_j^k = \tilde{Z}_{kj}^1 / \tilde{Z}_k. \tag{2.35}$$

The probabilities $\tilde{p}^k$ of generating the initial solutions (line 18) for the **search_method**($x$) procedure (line 19) are estimated based on the concepts taken from the simulated annealing method and depend on the current "temperature" $\mu_k$ and the set $\tilde{S}$ of feasible solutions found earlier.

The values of $\mu_k$, $k = 0, \ldots, K$, determine the annealing curve and are usually estimated by the formulas $\mu_0 = 0$, $\mu_{k+1} = \sigma\mu_k$, $k = 1, \ldots, K-1$. The values of $\mu_1$ and $\sigma > 1$ are chosen so that the probability $\tilde{p}^K$ is approximately equal to $x_{max}\|\| x_{max} - \tilde{p}^K \| \cong 0$. Note that the annealing curve is universal and is used to solve all problems rather than a specific one. Only the coefficients of the objective function are scaled so that the record is equal to a prescribed value. The initial probability $\tilde{p}^0$ can be specified in different ways: All the coordinates of the vector $\tilde{p}^0$ are identical (e.g., 0.5), this vector is a solution to problem (2.30) with the relaxation of Boolean conditions or it is obtained by statistical modeling of a Markov chain, which corresponds to the simulated annealing method for problem (2.30) with infinite temperature (all the feasible solutions are equiprobable).

The correcting loop (lines 11–34) is carried out until the solution $x_{max}$ is improved *maxnfail* times. The procedure **calculate_probabilities_of_generation** (line 15) is used to estimate the probabilities of random perturbation of the solution by one of the formulas (2.33)–(2.35). The loop of finding new solutions (lines 17–29) repeats a prescribed number of times *ngen*.

Let us present the **generate_solution** procedure.

procedure **generate_solution** ($x_{max}$, $pr^k$)

$x \leftarrow x_{max}$; $dist = 0$; $j \leftarrow 1$
while ($j \leq n$)
if $x_j = 1$ then
if ($\tilde{p}_j^k < random[0,\ 1]$) then
$x_j \leftarrow 0$; $dist \leftarrow dist + 1$
$g(x) \leftarrow$ **recalculate_descent_vector** ($x$)
end if
end if
else
if ($\tilde{p}_j^k \geq random[0,\ 1]$) then
$x_j \leftarrow 1$; $dist \leftarrow dist + 1$
$g(x) \leftarrow$ **recalculate_descent_vector** ($x$)
end if
end else
$j \leftarrow j + 1$
if ($dist = max\_dist$) then break
end while

This procedure randomly generates the solution $x$, which is initial for the **search_method** procedure (line 19). For low "temperatures" $\mu_k$, the perturbation of the solution $x_{max}$ is purely random and equiprobable; for high ones, the values of the coordinates that are identical in the best solutions vary slightly. The "temperature" loop allows diversifying the search (for low "temperatures") and intensifying it (for high "temperatures"). Thus, the solutions generated as the "temperature" increases get features inherent in the best solutions, and as a result, they coincide with the solution $x_{max}$.

In line 1, the initial values are assigned to the vector $x$ and variables $dist$ and $j$. The vector $x_{max}$ is perturbed in the loop in lines 2–17. The quantity $random[0, 1]$ uniformly distributed on the interval [0, 1] (lines 4 and 10) is used to model random events that change the solution $x_{max}$. Noteworthy is that the descent vector $g(x)$ is recalculated after each perturbation of $x_{max}$. It is specified by an $n$-measurable vector defined on the neighborhood of unit radius with the center at the point $x$: $g_j(x) = f(x_1, \ldots, x_{j-1}, 1 - x_j, x_{j+1}, \ldots, x_n) - f(x_1, \ldots, x_j, \ldots, x_n)$.

Recalculating the vector $g(x)$ corresponds to the idea of the descent vector method and saves much computational resources for many problems. For example, for an unconstrained quadratic-programming problem with Boolean variables, calculating the objective function requires $O(n^2)$ arithmetic operations, and recalculating the vector $g(x)$ needs only $O(n)$ such operations.

The perturbation loop (lines 2–17) is terminated if the Hamming distance between the point $x_{max}$ and the perturbed solution $x$ is equal to a prescribed parameter $max\_dist$ (line 16). This parameter can be specified dynamically and depend on a "temperature" index $k$. Usually, $max\_dist$ is chosen on the interval $\left[\frac{n}{10}, \frac{n}{2}\right]$.

In solving different problems, the **search_method** procedure employs randomized (probabilistic) algorithms of local search, descent vector, tabu search, and simulated annealing. When an algorithm is chosen for the procedure, specific features of the problem being solved are taken into account. The algorithm should intensify the search and allow the efficient study of the neighborhoods of the initial solution in order to correct it. Moreover, the algorithm can be chosen dynamically, that is, one algorithm can be used for low "temperatures" and another for high ones.

The **search_method** procedure is used to find the set of local (in the sense of a neighborhood of certain radius) solutions that form the set $R$. Lines 20–21 formally specify that the set $R$ has no currently infeasible solutions and that quantities (2.32) are recalculated. In the GES algorithm, the search for infeasible solutions is blocked, and recalculation is carried out as new solutions are found.

The solution $x_{max}$ and the descent vector at this point are found in lines 22 and 23, respectively. The **form_elite** procedure specifies the set $Elite$ of elite solutions, which consists of solutions with the best values of the objective function: $min f(x)_{x \in Elite} \geq max\, f(x)_{x \in \tilde{S} \setminus Elite}$. The maximum cardinality of the set is determined by the parameter max_elite_size and is usually chosen from the interval $[1, n]$.

After the correction loop (lines 11–35) is terminated, the set $P$ of infeasible points is supplemented with points from a neighborhood centered at $x_{max}$ specified by the parameter $d_P$ (line 36), and infeasible points are removed from the set $Elite$ (line 37).

If the analysis of the optimization process detects the so-called RESTART distribution of the time of finding solutions, the set *Elite* is made empty, which allows restarting a new independent solution process.

The GES method was extended to the following classes of discrete optimization problems: a multidimensional knapsack problem with Boolean variables, problems on the maximum cut of an oriented graph, maximum feasibility, scheduling, finding a clique, graph coloring, quadratic programming with Boolean variables, etc. The efficiency of the developed GES algorithms was compared with that of known algorithms [159, 160]. An analysis of the algorithms and the results they produced shows that for all the problems solved, the global equilibrium search has significant advantages over the other methods, which demonstrate high efficiency for one problems and very low efficiency for others. The GES algorithm performed the best for all the subsets of problems, including the above classes of complex high-dimensional discrete programming problems, and successfully competed with the approximate algorithms that are currently the best.

Noteworthy is that the main advantage of the GES method is not its computational efficiency but rather its team properties (the ability to make efficient use of the solutions obtained by other algorithms). For example, the classical local search method has zero team index. The team properties allow the GES method to be efficiently used for parallel computing, which is especially important for the optimization in computer networks and the Internet. In our opinion, in developing optimization methods of new generation in the future, optimization should be interpreted not in the commonly accepted sense but as a team of algorithms that work in parallel, acquiring and exchanging information.

Recent developments also include a theoretical basis for accelerated solution of complex discrete optimization problems. In particular, the above-mentioned RESTART technology [159] was developed; it is based on the new concepts of RESTART distribution and RESTART stopping criterion. RESTART distribution is the distribution of time necessary to solve the problem such that the average time can be reduced by using a restarting procedure. The idea of the RESTART technology is to randomize and modify an optimization algorithm so as to make the distribution of the time it solves the problem a RESTART distribution for which the use of the optimal RESTART stopping criterion considerably accelerates the problem solution.

Procedures for choosing optimization algorithms based on the method of probabilistic decomposition and the Wald criteria of sequential analysis were developed and analyzed.

Supercomputers created in the recent years due to the rapid development of computer and communication facilities determined one of the possible ways to accelerate the solution of complex and labor-intensive discrete programming problems and required new algorithms for parallelizing the optimization process.

The new approach to parallelizing the optimization process proposed in [159] is promising. Instead of the operations executed by the algorithm, it parallelizes its interacting copies.

Theoretical studies and operation paralleling were used to develop an information technology to solve problems of constructing maximum-size barred codes.

Such problems have attracted interest because of the development of information technologies, the Internet, computer networks, and modern communication facilities. The developed information technology allowed obtaining new record results: Using the SCIT-3 multiprocessor complex, maximum-size barred codes were constructed for the first time or the records existing for them were improved [159, 160]. To get acquainted with them, see http://www.research.att.com/~njas/doc/graphs.html (research center of computer science AT&T Labs, New Jersey, USA), where the global level of achievements in this field is shown.

Problems of correctness and stability of multicriterion (vector) problems [90, 157, 159] occupy an important place in the studies carried out in the last decades by experts in discrete optimization, including scientists at the Institute of Cybernetics. The concept of a correct mathematical problem is closely related to the following properties: The problem should be solvable and have a unique solution, and this solution should continuously depend on the variations in input data. The third property is usually called stability. Close attention to the correctness of optimization problems is largely due to the fact that in solving many applied problems that can be formalized by mathematical optimization models, it is necessary to take into account uncertainty and random factors such as the inaccuracy of input data, mathematical models inadequate to real-world processes, roundoff errors, errors of numerical methods, etc. Studies carried out by scientists at the Institute were mainly intended to establish the conditions whereby the set of optimal solutions (Pareto, Slater, or Smale set) of a vector discrete optimization problem has any of the five properties, each characterizing in a certain way the stability of the problem against small perturbations of input data. Omitting the necessary definitions and formalized description of the results obtained, we will specify only the main ones. Five types of problem stability were considered. The regularization of unstable problems was investigated. The stability of a vector integer optimization problem on a finite set of feasible solutions was related to the stability of optimal and non-optimal solutions. The concepts that may underlie the description of different types of stability to formulate necessary and sufficient stability conditions were defined. It was proved that different types of stability of the problem can be ascertained on a common basis by studying two subsets of a feasible set, namely, all the points that stably belong to the set of optimal solutions and all the points that do not stably belong to it. A common approach to the analysis of different types of stability of a vector integer optimization problem was proposed. The obtained results are related to the stability of both the perturbations of all input data of the problem and the perturbations of the input data necessary to represent its vector criterion or constraints, which is important since a problem stable against perturbations of some input data may be unstable against perturbations of the remaining data.

The theoretical and applied studies in discrete optimization carried out at the Institute of Cybernetics provide a basis for various information technologies.

## 2.6 Results in Applied Systems Analysis and Professional Training in this Field

V. S. Mikhalevich paid special attention to methods of systems analysis. In the last 15–20 years, they have been substantially developed by scientists from many organizations, including the Institute of Applied Systems Analysis of the National Academy of Sciences of Ukraine and Ministry of Education and Science of Ukraine (IASA). The ideological fundamentals of interdisciplinary studies and scientific developments combined with the educational process turned out to be important for both scientific research and training of students who have come to be often engaged in real research. More complex interdisciplinary problems to be solved necessitate the creation of a system methodology and the development of the theory and practice of systems analysis. This necessity is due to not only the intensive development of science and technology but also constant accumulation of threats of ecological, technogenic, natural, and other disasters.

Fundamental and applied research intended to develop a methodology of systems analysis of complex interrelated social, economic, ecological, and technological objects and processes go to the front burner more than ever today. A methodology to solve this class of problems is a new field of studies called system mathematics. It is a complex of interrelated subdisciplines (classical and new ones) of mathematics that allow solving various modern interdisciplinary problems. For example, elements of system mathematics are a formalization of the interrelations between continuous and discrete mathematics, transformation of some optimization methods into fuzzy mathematics and development of appropriate optimization methods in fuzzy formulation, description of interrelated processes that develop in different time scales (with different rates), analysis of distributed- and lumped-parameter systems on a unified platform, and combination of methods of quantitative and qualitative analysis in unified computational processes in constructing man–machine systems, etc.

The department of mathematical methods of systems analysis (MMSA), which is a part of the IASA, was created, as already mentioned, at the NTUU “KPI” in 1988 on Mikhalevich’s initiative. Five years later, Academician Kukhtenko and Mikhalevich participated in creating the Research Institute of Interdisciplinary Studies (RIIS) on the basis of the research sector of MMSA department at the Kyiv Polytechnic Institute. It was this department and the RIIS and two research departments of the V. M. Glushkov Institute of Cybernetics supervised, respectively, by B. N. Pshenichnyi, academician of the National Academy of Sciences of Ukraine, and V. S. Mel’nik, corresponding member of the National Academy of Sciences of Ukraine, that became the basis for the Scientific and Training Complex “Institute for Applied System Analysis.”

M. Z. Zgurovsky, academician of NAS of Ukraine

In what follows, we will dwell on some important developments in Applied Systems Analysis carried out at the IASA under the guidance of M. Z. Zgurovsky. These scientific studies prove that it is expedient to use system methodology to solve modern interdisciplinary problems and substantiate the role and place of systems analysis as a universal scientific methodology in science and practice and its interrelation with other fundamental disciplines. The key concepts, axioms, and definitions of systems analysis were formulated. The concept of complexity as a fundamental property of problems of systems analysis and principles and methods of solving problems of transcomputational complexity were formulated within the framework of the general problem of systems analysis [70, 72, 225]. A constructive and convenient way of representing the initial information on the object as conceptual spaces of conditions and properties of the object was proposed to solve complex interdisciplinary problems. Under modern conditions, these spaces should provide new vision of element interaction in the structure "system analyst $\Leftrightarrow$ the human $\Leftrightarrow$ object $\Leftrightarrow$ environment." This will provide more clear presentation of the coordination of major factors: properties of the object under study, input information, and operating conditions in view of various uncertainties and multifactorial risks. The coordination should be systemic and take into account the objectives, problems, expected results of operation of the object, the complexity of situations in which it operates, and the shortage of information on the complexities related to the objectives and operation conditions.

In order to create tools to solve complex interdisciplinary problems, approaches are developed, which help to formalize the procedures of conceptual–functional spaces of conditions and properties of system operation. Under these conditions, problems of evaluating indeterminate forms of objectives, situations, and conflicts in problems of interaction and counteraction of coalitions are formulated and methods are proposed to solve them. Methods of evaluating indeterminate forms of objectives are developed: using technical constraints, reducing to a system of nonlinear incompatible equations, and reducing to a Chebyshev approximation problem. The problems of evaluating natural and situational indeterminate forms were considered: evaluating indeterminate forms for known characteristics of random factors, evaluating indeterminate forms in the case of incomplete information on random factors. Problems of evaluating indeterminate forms in conflict situations were analyzed: evaluation of indeterminate forms of active interaction of partners and counteraction of opposite parties, multi-objective interaction of partners under situational uncertainty, and multi-objective active counteraction of opposite parties under situational uncertainty.

The approaches to searching for a rational compromise in evaluating conceptual indeterminate forms and an approach to evaluating functional dependences in these problems were proposed. The problem of evaluating functional dependences is essentially more complex than the typical problem, which is due not only to heterogeneous input information but also heterogeneous properties of the groups of factors considered. To overcome the transcomputational complexity, it is expedient to generate approximation functions as a hierarchical multilevel system of models. Within the conceptual–functional space of properties, it is expedient to analyze the issues related to constructing the structure and functions of complex multilevel hierarchical systems and to formulate the principles and methods of structuring the formalized description of properties, structures, and functions of such class of systems. To give the mathematical formulation of the problem of systems analysis of a complex multilevel hierarchical system, the general solution strategy and structure of the generalized algorithm of the structural–functional analysis, as well as the method of its solution are proposed [70, 72, 225]. An approach to solving the problem of system structural optimization of complex structural elements of modern technology is proposed, which is based on a purposeful choice of functional elements of each hierarchical level. Problems of system parameter optimization are solved, which provides a rational compromise of inconsistent requirements to the strength, reliability, manufacturability, and technical and economic efficiency of the structure.

Professor N. D. Pankratova

Based on methods of system mathematics and probability theory, functional analysis, function approximation theory, and discrete mathematics, Pankratova proposed [133] a new principle of before-the-fact prevention of the causes of possible transition of an object from the operable to a faulty state, based on systems analysis of multifactor risks of contingency situations, reliable estimation of the resources of admissible risk of different operation modes of a complex engineering object, and prediction of the key “survivability” parameters of the object during its operation life. An apparatus is proposed for a system-coordinated solution of problems of detecting, recognizing, predicting, and minimizing risks of contingency, critical, and emergency situations, accidents, and catastrophes.

Let us formulate the problem of detecting a contingency situation in the dynamics of operation of an ecologically dangerous object [133].

*Given*: for each situation $S_k^\tau \in S_\tau$, a set $M_k^\tau \in M_\tau$ of risk factors is formed as $M_k^\tau = \{\rho_{q_k}^\tau | q_k = \overline{1,\ n_k^\tau}\}$. For each risk factor of the set $M_k^\tau$, fuzzy information vector $I_q^\tau = \{I_{q_k}^\tau | q_k = \overline{1, n_k^\tau};\ k = \overline{1,\ K_\tau}\}$ is known and its components are known to have the form

$$I_{qk}^\tau = \{\tilde{x}_{q_k j_k p_k}^\tau | q_k = \overline{1,\ n_k^\tau};\ j_k = \overline{1,\ n_{q_k}^\tau}; p_k = \overline{1,\ n_{q_k j_k}^\tau}\},$$

$$\tilde{x}_{q_k j_k p_k}^\tau = \langle x_{q_k j_k p_k}^\tau, \mu_{H_{q_k j_k p_k}}(x_{q_k j_k p_k}^\tau)\rangle,\ \ x_{q_k j_k p_k}^\tau \in H_{q_k j_k p_k}^\tau,\ \ \mu_{H_{q_k j_k p_k}} \in [0,\ 1],$$

$$H_{q_k j_k p_k}^\tau = \langle x_{q_k j_k p_k}^\tau | x_{q_k j_k p_k}^- \le x_{q_k j_k p_k}^\tau \le x_{q_k j_k p_k}^+ \rangle.$$

*Required*: for each situation $S_k^\tau \in S_\tau$ and each risk factor $M_k^\tau \in M_\tau$, recognize a contingency situation in the dynamics of operation of an ecologically dangerous object and ensure the “survivability” of a complex engineering system during its operation.

The problem of diagnostics of the operation of a deep pumping facility is used to show that it is expedient to apply methods that allow before-the-fact prevention of contingency situations, that is, ensure the “survivability” of the complex system. The main purpose of the pumping facility is to maintain a prescribed level of water delivery for various needs, cooling a technological ecologically dangerous installation being of priority.

A systemic approach is employed to solve problems that are of current importance for the economy and society. A system methodology of developing scenarios of future events in various spheres of human activity with the use of unique man–machine tools of technological forecast as an information platform of scenario analysis is developed [71]. This technology is a complex of mathematical, program, logic, and technical–organizational means and tools to provide an integral process of forecasting based on the interaction of man and specially created software-engineering environment that allows monitoring the decision-making process and choosing a rational compromise based on proposed alternative scenarios.

Based on the developed approach to the solution of problems of technological forecast, a number of projects were implemented for Ministry of Economics, Ministry of the European Integration of Ukraine, Ministry of Education and Science of Ukraine, National Space Agency of Ukraine, and Scientific Center for Aerospace Research of the Earth at the Institute of Geological Sciences of the National

Academy of Sciences of Ukraine. In particular, by the order of the National Space Agency of Ukraine, priority fields of the consumption of space information provided by Earth remote sensing systems were identified. For the integrated iron-and-steel plant in Krivoi Rog, alternative scenarios for increasing the efficiency of the logistic system of the plant as to maintaining all the stages of metallurgical production were created at the stage of short-term forecast. By the order of the State Agency of Ukraine on Investments and Innovations, a strategic plan for the development of the Autonomous Republic of Crimea and Sevastopol were developed.

The methodological and mathematical principles and approaches to implementing the strategy of technological forecast can be used to predict the technological development of the society. This approach can be applied to develop scenarios for the innovation development of large enterprises and individual branches of industry and to form the technological policy of the society.

Based on the methods of systems analysis, Zgurovsky obtained fundamental results in global modeling of processes of sustainable development in the context of the quality and safety of human life and the analysis of global threats [69]. A system of indices and indicators was generated and a new metrics was proposed to measure processes of sustainable development in three spaces: economic, ecological, and social (institutional). This metrics was used to perform the global modeling of processes of sustainable development for a large group of countries. An important division of modeling deals with the analysis of the pattern of world conflicts as a fundamental property of the global society. An attempt was made to predict the next world conflict called "the conflict of the twenty-first century," and its nature and key characteristics (duration, main phases, and intensity) were analyzed. The set of main global threats that generate this conflict was presented. The method of cluster analysis was used to determine how these threats influence different countries and eight large groups of countries (civilizations by Huntington) with common culture. The assumptions on possible scenarios of the "conflict of the twenty-first century" were suggested.

The reliability of any forecast "resonating" with other global or local tendencies, hypotheses, and patterns is known to substantially increase. For such additional conditions, Zgurovsky took the modern concept of the acceleration of historical time and the hypothesis that the duration of large $K$-cycles decreases with scientific and technical progress [67]. Starting from these and considering the evolutionary development of civilization as an integral process determined by the harmonic interaction of its components, the Kondratieff cycles of the development of the global economy were compared with the $C$-waves of global systemic conflicts, and an attempt was made to predict these periodic processes for the twenty-first century. Analyzing these two processes on a common time axis reveals their agreement (synchronism), which is formulated as the following two principles.

1. *Quantization Principle*. Time intervals $\Delta(C_n)$, $n \geq 5$, during which the wave $C_n$ passes five evolutionary phases "origin $\rightarrow$ growth $\rightarrow$ culmination $\rightarrow$ decrease $\rightarrow$ decay" contain an integer number $n_k(\Delta(C_n))$ of full $K$-cycles of a modified sequence of Kondratieff cycles (KCs) $\{K_n\}_{n\geq 1}$.

2. *Monotonicity Principle*. The average duration $T_k(\Delta(C_n))$ of one full $K$-cycle of the KC $\{K_n\}_{n\geq 1}$ on the time intervals $\Delta(C_n)$ decreases as $n$ grows.

   Denote by

$$G(C_k;\ \{\ K_n\}_{n\geq 1})\ \triangleq \{K_{s(k)};\ K_{s(k)+1};\ \ldots\ ;\ K_{s(k)+m(k)}\},\quad k\geq 5,$$

   a group (quantum) of $K$-cycles separated by the $C_k$-wave from the KC $\{K_n\}_{n\geq 1}$. Then,

$$n_k(\Delta(C_k)) = m(k)+1;\ \ T_k(\Delta(C_k)) = (m(k)+1)^{-1}\times\sum_{r=0}^{m(k)} T(K_{s(k)+r}),$$

   where $T(K_j)$ is the duration of one full $K$-cycle $K_j$.

   In our case,

$$G(C_5;\ \{K_n\}_{n\geq 1}) = \{K_1;K_2;K_3\},\quad G(C_6;\ \{\ K_n\}_{n\geq 1}) = \{\ K_4;K_5\},$$

$$T_k(\Delta(C_5)) = 3^{-1}\sum_{i=1}^{3} T(K_i) = 56.6 \text{ years},\quad n_k(\Delta(C_5) = 3,$$

$$T_k(\Delta(C_6)) = 2^{-1}\sum_{i=4}^{5} T(K_i) = 43.5 \text{ years},\quad n_k(\Delta(C_6) = 2.$$

   The pattern revealed allows formulating the hypothesis on the next quantization step, that is, separating the next group $G(C_7;\ \{K_n\}_{n\geq 1})$ of $K$-cycles from the KC $\{K_n\}_{n\geq 1}$ by the $C_7$-wave.

   Since the development of the global economy and the course of systemic global conflicts are interdependent components of the same process of evolution of the globalized society, the agreement (synchronism) of these processes revealed on the time intervals $\Delta(C_5)$ and $\Delta(C_6)$ in the sense of quantization and monotonicity principles remains on the time interval $\Delta(C_7)$.

   Based on the basic hypothesis, the course (in the metric aspect) of $K$-cycles in the twenty-first century is forecasted as follows:

   (a) The time interval $\Delta(C_7)$ contains no less than two full cycles of the KC $\{K_n\}_{n\geq 1}$.
   (b) The average duration of one full $K$-cycle on the time interval $T_k(\Delta(C_7))$ is shorter than $T_k(\Delta(C_6)) = 43.5$ years.

   Thus, two cases are possible, which correspond to two scenarios of the course of Kondratieff cycles in the twenty-first century.

1. The time interval 2008–2092 contains two full Kondratieff cycles:

$$G(C_7;\ \{K_n\}_{n\geq 1}) = \{K_6;\ K_7\},\quad n_k(\Delta(C_7) = 2,$$

$$T_k(\Delta(C_7)) = 2^{-1}\sum_{i=6}^{7} T(K_i) = 42.5 \text{ years} < T_k(\Delta(C_6)) = 43.5 \text{ years}.$$

2. The time interval 2008–2092 contains three full Kondratieff cycles:

$$G(C_7;\ \{K_n\}_{n\geq 1}) = \{K_6;\ K_7;\ K_8\},\quad n_k(\Delta(C_7) = 3,$$

$$T_k(\Delta(C_7)) = 3^{-1}\sum_{i=6}^{8} T(K_i) = 28.3 \text{ years} < T_k(\Delta(C_6)) = 43.5 \text{ years}.$$

The results of the study confirm the refined hypothesis that the duration of large Kondratieff cycles tends to decrease with the scientific and technical progress: It is most probable that in the twenty-first century, there will appear three large $K$-cycles with about 30-year average duration of one full cycle, which is much shorter than the average duration of one of the previous five Kondratieff cycles ($\approx$50 years). Moreover, the revealed synchronization of the development of the global economy and of the course of global systemic conflicts may be interpreted as an indirect confirmation that the models of Kondratieff cycles and $C$-waves are correct.

Methods of the optimal control of nonlinear infinite-dimensional systems developed at the IASA were used to develop a fundamentally new approach to the analysis of singular limiting problems for equations of mathematical physics. This approach is based on reducing an ill-posed problem to an auxiliary (associated) optimal control problem. Following this way, it is possible to develop stable solution algorithms for three-dimensional Navier–Stokes equations, Bénard systems, etc. Thus, methods of optimal control theory are a mathematical tool for efficient analysis of ill-conditioned equations of mathematical physics.

A number of results were obtained in nonlinear and multivalued analysis, solution of nonlinear differential-operator equations, analysis of inclusions and variational inequalities as an important field of the development of methods of infinite-dimensional analysis, theory and methods of optimization, game theory, etc. [91, 92, 205, 214]. A new unified approach was developed to solve optimal control problems using systems of nonlinear processes and fields. New results were obtained in the theory of control of complex nonlinear dynamic systems that contain modules with distributed and lumped parameters.

Studies on constructing and analyzing models of social systems were performed recently at the IIASA. One of the fields was constructing network-type models with associative memory (similar to Hopfield neural networks) for a wide class of processes: social, economic, political, and natural. We may say that revealing the multiple-valued behavior of the solutions of such models is an achievement in this field. This opens up a way to the wide application of methods of system mathematics. For example, there arise problems of the analysis of model behavior depending on parameter variations in the case of multivaluedness, limit cycles, and especially in the case of multivalued mappings (the groundwork for their analysis was laid earlier). Solving these problems requires almost the whole arsenal of means of system mathematics: nonlinear functional analysis, dynamic systems theory, and theory of differential and discrete equations.

One more field of studies is related to modeling the motion of a great number of objects. The cellular automata method was chosen as one of the approaches. It analyzes the motion within the framework of discrete approximation (the space being divided into regular cell unions) and considers the passages of objects from one cell to another. Research of cellular automata with prediction was launched (O. S. Makarenko). Mathematically, this problem looks like studying discrete dynamic systems (deterministic or stochastic). Traffics were modeled under some conditions with the use of models adapted to real geometry of transportation flows and obstacles, and problems on the allowance for motion prediction by the participants were also considered. Such problems again necessitate the mathematical analysis based on optimization and multivalued mapping theories.

Dynamic and time characteristics of some macroeconomic and social processes as fundamentals for their analysis and prediction were found out and investigated. The following problems were formulated and solved based on statistical information: qualitative analysis of the stages of development, growth, decrease, and decay of social processes; analysis of conceptual relationships among the dynamic and time characteristics of each individual stage of a process; formalization of the dynamic and time characteristics of processes; creation of computation algorithms based on statistical data; and testing the results of the analysis. Specific features of "living" systems are the conceptual basis of the analysis. The theoretical basis is the dynamic systems theory, nonequilibrium states theory, self-organization theory, and systems analysis (M. V. Andreev, G. P. Poveshchenko). Applied mathematical models for short- and intermediate-term forecast of the key macroeconomic indices of Ukraine (rate of gross domestic product, rate of employment of the population, rate of consolidated budget, rate of expenses, etc.), a mathematical model of labor market, and a mathematical model of the structural evolution of productive forces were created and tested.

The system methodology in optimization, in particular, system optimization as a purposeful modification of the feasible domain of complex engineering and economic optimization problems, is developed at the department of numerical optimization methods of the IASA. The subject of research is mathematical models that describe flows in networks with generalized conservation law, where the difference of inflows and outflows at a node may be not equal exactly to the volume of consumption at the node but belong to some interval. Such a model occurs, for example, in control problems of water transfer in irrigation channels. Flow calculation problems consider not only arc but also nodal variables. Flow transportation costs significantly decrease in this case due to optimal redistribution of sources (drains). Nonlinear complex models are constructed, which describe resource transportation and distribution among consumers in view of the possibility of creating inventory in certain tanks (storehouses) such as gas flow distribution models under active operation of gasholders [132].

In his scientific studies, Mikhalevich paid much attention to modeling and prediction. During the last two decades, scientists of the IASA fruitfully worked to develop methods of mathematical and statistical modeling and forecast of arbitrary processes. Methods of regression data analysis received further development and original structures of models were obtained to describe nonlinear nonstationary processes (e.g., models of heteroscedastic and cointegrable processes, which are popular in the finance and economy). An original technique was created to choose the type and to evaluate the structure of a model and is successfully used in creating decision-support systems. Computational procedures were proposed for data preprocessing in order to reduce them to a form that is most convenient to estimate the parameters of mathematical, including statistical, models. To overcome difficulties due to different types of uncertainties, probabilistic modeling and prediction methods were developed, which use the Bayesian approach to data analysis, in particular, Bayesian networks [68]. The advantages of models in the form of Bayesian networks are their clearness (directed graph), the possibilities of creating highly dimensional models (in the sense of a large number of nodal variables), the use of continuous and discrete variables in one model, allowance for uncertainties of structural and stochastic types, and possible use of methods of exact and approximate probabilistic inference. Bayesian networks can generally be characterized as a rather complex resource-intensive yet highly effective probabilistic method to model and predict the development of various processes.

Processes with uncertainties were predicted in the best way with the use of probabilistic methods and fuzzy logic. Inherently, these methods are close to situation modeling and decision making by man; therefore, their application in control and decision-support systems may produce a significant beneficial effect.

To preserve the quality of forecasts under the nonstationarity of the process under study and to improve the quality of forecast estimates for processes with arbitrary statistical characteristics, adaptive forecast estimation schemes are applied. The input data for the analysis of forecast quality and for the formation of adaptive schemes of their estimation are the values of forecast errors and their statistical characteristics (variance, standard deviation, and mean values).

P. I. Bidyuk and his disciples proposed methods to adapt forecast estimation schemes to statistical data characteristics in order to improve the quality of forecast estimates. Which scheme is applied depends on specific problem formulation, the quality and amount of experimental (statistical) data, formulated requirements to the quality of forecast estimates and the time interval for the computations. Each method of the adaptive formation of a forecast estimate has specific features, which are taken into account in creating an adaptive prediction system [12].

An error feedback in an adaptive prediction system improves the quality of the model to a level necessary for a high-quality forecast. Error feedback also improves the accuracy of forecast estimates due to enhancing the quality (informativeness) of data and refining the structure of the model. It allows avoiding overlearning, which improves data approximation accuracy but deteriorates the quality of forecast estimates.

A technique was developed to construct a Bayesian network as a directed acyclic graph, intended to model and visualize the information on a specific problem of network training based on available information and to make statistical inference. A Bayesian network is a model that represents probabilistic interrelations among nodes (variables of the process). Formally, it is a triple $N = \langle V, G, J \rangle$ whose first component is a set of variables $V$, the second is a directed acyclic graph $G$ whose nodes correspond to random variables of the process being modeled, and $J$ is the general distribution of the probabilities of variables $V = \{X_1, X_2, ..., X_n\}$. The set of variables obeys the Markovian condition, that is, each variable of the network depends only on the parents of this variable.

First, the problem of computing the values of mutual information among all nodes (variables) of the network is formulated. Then it is necessary to generate the optimal network structure based on the chosen optimization criterion, which is analyzed and updated at each iteration of the training algorithm.

Given a sequence of $n$ observations $x^n = (d_1, ..., d_n)$, the function of formation of the structure $g \in G$ has the form:

$$\log\left(P(g, x^n)\right) = \log\left(P(g) \cdot \prod_{j \in J}\left(\prod_{s \in S(j,g)} \frac{(\alpha^{(j)} - 1)! \cdot \prod_{q \in A^{(j)}} (n[q, s, j, g]!)}{(n[s, j, g] + \alpha^{(j)} - 1)!}\right)\right)$$

$$= \log\left(P(g)\right) + \sum_{j \in J}\left(\sum_{s \in S(j,g)}\left(\sum_{i=1}^{\alpha^{(j)}-1} i + \sum_{q \in A^{(j)}}\left(\sum_{i=1}^{n[q, s, j, g]} i\right) - \sum_{i=1}^{n[s, j, g]+\alpha^{(j)}-1} i\right)\right)$$

$$= \log\left(P(g)\right) + \sum_{j \in J}\left(\sum_{s \in S(j,g)}\left(\sum_{q \in A^{(j)}}\left(\sum_{i=1}^{n[q, s, j, g]} i\right) - \sum_{i=\alpha^{(j)}}^{n[s, j, g]+\alpha^{(j)}-1} i\right)\right),$$

where $P(g)$ is a priori probability of the structure $g \in G$; $j \in J = \{1, ..., N\}$ denotes the enumeration of the nodes of the structure of the network $g$; $s \in S(j, g)$ denotes the enumeration of the set of all sets of values taken by the parent nodes of the $j$th node:

$$n(s, j, g) = \sum_{i=1}^{n} I(\pi_i^{(j)} = s), \; n[q, s, j, g] = \sum_{i=1}^{n} I(x_i = q, \pi_i^{(j)} = s),$$

where $\pi^{(j)} = \Pi^{(j)}, X^{(k)} = x^{(k)}, \forall k \in \phi^{(j)}$, and the function $I(E) = 1$ if the predicate $E = true$, otherwise $I(E) = 0$.

The created network is used as a basis to form a probabilistic inference with the use of training data (two-step procedure).

**Step 1.** On the set of training data, compute the matrix of empirical values of the general probability distribution of the whole network $P(X^{(1)}, ..., X^{(N)})$:

$$P_{\text{matrix}}(X^{(1)} = x^{(1)}, ..., X^{(N)} = x^{(N)}) = \frac{n[X^{(1)} = x^{(1)}, ..., X^{(N)} = x^{(N)}]}{n},$$

where $n$ is the number of training observations, $x^{(j)} \in A^{(j)}$;

**Step 2.** Sequentially analyze the state of nodes of the Bayesian network. If a node is uninstantiated, then compute the probabilities of all its possible states. If the values of the nodes of a row coincide with the values of uninstantiated nodes and the state of the node being analyzed, then add the corresponding value to the value of the probability of the corresponding state of the node. Then normalize the probabilities of states of the node being analyzed.

Academicians I. V. Sergienko, M. Z. Zgurovsky (at the center), and Professor V. P. Shilo (second from the left) among students of US universities who underwent practical training in Kyiv

The models were used to perform short- and long-term predictions of the dynamics of financial (bank) and industrial enterprises, to form a strategy of the development of small and medium business using Bayesian networks, to estimate and predict the credit status of clients of a bank, to analyze trends in indices when performing commercial operations at stock exchange, to form the optimal portfolio of securities, and the method of analysis and prediction of risks of portfolio investments.

The IASA participates in the international programs TACIS, EDNEC, INTAS, etc. The achievements of the institute include applied research in socioeconomic problems of the state, common with the IIASA (Austria). The studies in planning and creating a national computer network of educational institutions and sciences (URAN system) are performed under active participation of the institute and are supported by the NATO technical committee. The Research Institute “INFORES” participated in projects of the European Commission CALMAPTACIS and Copernicus-INCO, projects of the Ukrainian Science and Technology Center

(STCU), in the joint project of the Intel company and the Ukrainian and South Korean company SeoduLogic, etc. The state scientific program "Technological forecast as a system methodology of the innovation-based development of Ukraine" was initiated by the institute and was carried out under active cooperation with UNIDO (Austria).

The UNESCO department "Higher education in the sciences, applied systems analysis, and computer science," the World Data Center "Geoinformatics and sustainable development," a base of the UNESCO Institute in information technologies in education, the national committee of the international organization "CODATA," the national committee in information and data of the International Council for Science (ICSU) work under the aegis of the institute.

The IASA cooperates with scientific institutes of the branches of computer science, geosciences, physics and astronomy of the NAS of Ukraine and with higher educational institutions and takes part in solving problems of Kyiv and AR of Crimea.

These data predetermine the urgency and practical necessity of training experts who are able to solve complex system problems of timely prediction, fair forecast and systems analysis both of available socioeconomic, scientific and technical, and other problems, tasks, and situations and of possible technogenic, ecological, natural, and other disasters. Noteworthy is that in practice, much of the efficiency and reliability of timely prediction, objective forecast, systems analysis of different alternatives of complex solutions and strategies depend on how the system researcher can learn in due time and rationally use the capabilities of the methodology of systems analysis.

Fundamental and applied studies at the IASA are carried out in close relation with the educational process. The results of scientific developments are incorporated directly into the educational process. The academic training of students is combined with their practical activities in scientific subdivisions of the institute.

The doctorate and postgraduate studies of the institute are open for researchers in physics and mathematics and in engineering sciences with a major in "Systems analysis and optimal decision theory," "Mathematical modeling and computational methods," "Control systems and processes," "Information technologies," "Systems of automation of design processes," etc.

The target training of senior students involves their obligatory participation in scientific research at the institute and at the place of future work. As experience shows, the graduates successfully work as systems analysts, systems engineers, mathematicians, programmers, LAN administrators, economists at the institutions of the NAS of Ukraine, state authorities of different levels, branch research institutes, and commercial firms and banks.

The up-to-date computer laboratories provide favorable conditions for practical training such as training courses of CISCO network academy in design, construction, and administration of local and global networks, and in the fields related to design of modern integrated circuits and analytic support of banking.

There are faculties of system studies, preuniversity training, course training, second higher and postgraduate education. Experienced scientists and educators from the NTUU “KPI,” V. M. Glushkov Institute of Cybernetics, Institute for Information Recording of the NAS of Ukraine, Institute of Mathematics of the NAS of Ukraine, and Institute of Space Research of the NAS of Ukraine and National Space Agency of Ukraine are invited for lecturing, practical training, and supervising thesis works.

Combining the educational process with research studies where senior students take part yields positive results: Foreign universities have become interested in this experience. Not in vain, senior students from the USA who specialize in developing computer-aided technologies win annual competition to go to Kyiv to the Institute of Cybernetics and Institute of Applied Systems Analysis for practical training during which common scientific seminars and conferences are held.

http://www.springer.com/978-1-4614-4210-3

Methods of Optimization and Systems Analysis for
Problems of Transcomputational Complexity
Sergienko, I.V.
2012, XIV, 226 p., Hardcover
ISBN: 978-1-4614-4210-3