

**Corresponding Member of
the Academy of Sciences
of the USSR I. M.
GELFAND and M. L.
TSETLIN**

1961

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196101.15123>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

CYBERNETICS AND CONTROL THEORY

Corresponding Member of the Academy of Sciences of the USSR I. M. GELFAND and M. L. TSETLIN

THE PRINCIPLE OF NONLOCAL SEARCH IN SYSTEMS OF AUTOMATIC OPTIMIZATION

In many practical problems there arises the need to control complex systems with a large number of degrees of freedom. The consideration of processes of control of many physiological mechanisms (the construction of movements, the analysis of afferentations, etc.) also leads to problems of this kind. Attempts to solve these problems by means of classical mathematics often prove unsuccessful. Even when an algorithm can readily be constructed that makes it possible to solve the problem in all cases, the implementation of this algorithm does not seem possible because of the limitations of the speed of computing machinery and of the time within which the problem must be solved.

In these situations a solution can be achieved by making use of the **organization** which, to one degree or another, is possessed by problems encountered in human practical activity or, possibly, in physiology. In other words, the solution is achieved by abandoning consideration of the most labor-intensive (chaotic) situations. Let us note that in a number of problems these unfavorable (chaotic) cases are, in the formal mathematical sense, the most probable.

This note is an attempt at such a solution of one of the problems of the theory of automatic control—the problem of automatic optimization with a large number of working parameters.

As is known (see, for example, ^(1,2)), self-adjusting systems are systems with feedback in which the required value of the output quantity is obtained as a result of an automatic search. In a number of important cases the aim of the search is to attain an extremal value of the output quantity. Systems of this kind are called systems of automatic optimization. Below we shall describe a principle of automatic optimization based on one special method of nonlocal search, proposed by I. M. Gelfand and found to be very effective in solving a number of computational problems involving the search for a minimum.

Thus, let $F(x_1, \dots, x_n, y_1, \dots, y_m)$ be the output function of an automatic-optimization device. We shall call the group of arguments x_1, \dots, x_n the **working** group. The values of these arguments are changed as a result of the automatic search. The arguments y_1, \dots, y_m are hidden parameters of the system,

generally speaking dependent on time, so that the output function may be written in the form $F(x_1, \dots, x_n, y_1, \dots, y_m) = \Phi(x_1, \dots, x_n, t)$.

Let us note that in self-adjusting systems the function $\Phi(x_1, \dots, x_n, t)$ is most often not specified analytically or in any other way, so that the selection of the required (optimizing) values of the working parameters can be carried out only experimentally. The dependence of Φ on time leads to the necessity of a continuous search for the required values of the arguments. This dependence is not assumed to be known. In this connection, an important characteristic of the search is its speed; roughly speaking, it is necessary that “satisfactory” values of the optimized function be reached over time intervals during which this function “does not have time to change substantially.”

An estimate of the performance of an automatic-optimization system may be not the closeness of the instantaneous value of the function to the extremal one, but the function-

functionals of the form

$$T^{-1} \int_0^T \Psi(\Phi(x_1, \dots, x_n, t)) dt. \quad (1)$$

The function Ψ may be chosen in different ways. Thus, for example, when $\Psi(\Phi) = \Phi$, the value of this functional corresponds to the so-called “payment for the search,” or “cost of the search,” defined for the simplest automatic minimization systems with one working parameter and one minimum of the function Φ ⁽¹⁾.

In problems where it is sufficient to attain values of Φ not exceeding a certain level C (“minimization by level”), a convenient criterion is obtained from (1) if one sets

$$\Psi(\Phi) = 0 \quad \text{for } \Phi \leq C; \quad \Psi(\Phi) = 1 \quad \text{for } \Phi > C. \quad (2)$$

Automatic search may be carried out by various methods, which it is convenient to divide into three groups.

The first group comprises the so-called methods of **blind search**, in which experiments for selecting the required parameter values are carried out independently of one another. In this case, either all points of the space of working parameters are examined in a definite order (scanning), or these points are chosen at random (the homeostat principle) ^(3,4) and are not changed as long as the values of Φ remain satisfactory. In continuous search by these methods, the values of Φ from experiment to experiment do not systematically improve, so that the “cost of the search” proves to be high.

The second group of methods of automatic search (methods of **local search**) provides for analysis of the results of each experiment and the obtaining, in this

way, of initial data for the next experiment. This group includes the gradient method, the steepest-descent method, the relaxation method, and certain others. Their common feature is locality: the working point moves continuously through the space of working arguments; preparation of the next experiment is carried out on the basis of knowledge of the values of the function Φ in a small neighborhood of the parameter values of the preceding experiment. When local methods are used, the values of Φ improve in the course of the search, which gives them a substantial advantage in comparison with blind-search methods.

At the same time, by admitting only a nearby (local) search, these methods make only slight use of the features of the function being optimized. For small values of the gradient, all local methods become ineffective; they force one to resort repeatedly to changes in the magnitude of the working step in the direction of motion, which leads to a substantial slowing of the search rate.*

Systems of automatic optimization with local search methods are described in detail in the works of A. A. Feldbaum^(2,5,6), in which possible methods for the circuit implementation of automata of this kind are also indicated.

The third group of methods may naturally be called methods of **nonlocal** search. Their characteristic feature is that the curve along which the working point moves through the space of parameters is not continuous. In this case, the volume of the region examined per unit time increases sharply; it becomes possible to use features of the structure of the function Φ ; the optimization process is considerably accelerated.

The simplest nonlocal method is a combination of local methods and the homeostat principle. This method is often used in computational practice and amounts to the following. A random point is chosen, and from it descent is performed in accordance with the selected local method until the change in the function becomes small. Then a random point is again chosen, descent is carried out, and so on. Thus there arises, for example, the method of nonlocal gradient search.

* In addition, there is always a substantial danger that such a search will become “stuck” in some secondary “shallow pit.”

This method of nonlocal search suffers from the drawback that after each descent the value of Φ again (and, generally speaking, substantially) increases; the information about the function acquired during the local descent is not used further in any way. Owing to the systematic departure into regions of large values of Φ and the resulting necessity of prolonged use of local methods, the “cost of the search” proves to be relatively high.

We shall now describe a method of nonlocal search which we shall call the **ravine method**. The method applies to the case in which the working parameters x_1, \dots, x_n can be divided into two groups. The first group—which includes almost all the parameters—consists of those parameters whose change leads to a significant change in the value of the function Φ . Therefore adjustment with respect to these parameters (we shall call them **inessential**) is carried out com-

Fig. 1

Figure 1: Fig. 1

paratively simply and quickly. The second group of variables includes a small (for example, 2 or 3) number of functions from x_1, \dots, x_n , changes in which lead to a relatively small change in the values of the function Φ . We shall call these variables **essential**. Of course, the division of parameters into groups depends on time and therefore must also be carried out automatically. Naturally, such a division is impossible for every function that a mathematician might specify; however, for functions encountered in practical activity (reasonable problems of physics, engineering, etc.), such a division is apparently always possible. Understanding the difficulty of defining these concepts precisely, we shall nevertheless venture here to call such functions **well organized**.

Fig. 1

The automatic search is carried out as follows. First an arbitrary point X_0 is chosen. From this point a descent along the gradient is performed (one may, of course, also use some other local method). This descent should be carried out roughly; if, for example, the next step reduces the values of Φ by less than, say, 5 ÷ 15%, then it should be stopped. The point is that as soon as the descent along the gradient ceases to have a substantial effect on the values of Φ , we enter a zone where the variables of the first and second groups become equal in status, and, without moving noticeably along the essential variables, we continue to wander disorderly, changing the inessential variables. Strictly speaking, this is precisely the reason for the low efficiency of local methods.

Suppose that the gradient descent has brought us to the point A_0 (Fig. 1). After this, some point X_1 is chosen in the neighborhood of the point X_0 at a distance **substantially** exceeding the step of the gradient descent. From the point X_1 a gradient descent is performed to the point A_1 . After the points A_0 and A_1 have been obtained, the so-called “ravine step” is performed. The points A_0 and A_1 are joined by a straight line, and on this straight line a point X_2 is chosen at a distance from A_1 called the length of the **ravine step**. For well-organized functions this length is chosen to be considerably greater than the length of the gradient step. The selection of the ravine step is carried out experimentally (by trials) and is an important characteristic of the function Φ . After the point X_2 has been chosen, gradient descent is again performed to the point A_2 . The point X_3 is chosen from the points A_1 and A_2 in the same way as the point X_2 was chosen from A_0 and A_1 , after which the process is repeated. The points X_i are thus chosen in places where small values of Φ are expected,* or near them, so that the whole search is carried out mainly in the region of small values of the function. In

* Because of the influence of the inessential variables, the value of Φ at the points X_i themselves may fail to be small.

With a reasonable choice of the ravine step, as one advances along the ravine, adaptation to its course takes place, so that the lengths of the gradient descents become much smaller than the magnitude of the ravine step.* As a result, the cost of the search is reduced very substantially in comparison with local methods and with the simplest nonlocal method described above. Thus, with 8–10 variables in problems of phase analysis, the cost of the search decreases hundreds of times.

Let us also note that, for choosing the ravine step, it proves useful to employ the functional (1), with Ψ chosen in accordance with (2). The length of the ravine step must ensure some fixed value of this functional. If the step is too small, the value of the functional becomes small, which indicates a slowing of the search. If the step is too large, the value of the functional increases, indicating the coarseness of the search.

For functions of a large number of variables, determination of the gradient, associated with the need to fix the $(n + 1)$ -st value of Φ , becomes excessively cumbersome. Here the following device may prove useful. Some initial probabilities p_1, \dots, p_n are assigned to the variables x_1, \dots, x_n . Then, in accordance with these probabilities, a certain small number of directions is chosen and the partial gradient is computed. Depending on the results of the motion along the partial gradient, the initial probability distribution is corrected, so that the direction of motion is adapted to the direction of the full gradient.** It was already mentioned above that, instead of the gradient, one may use any trial of nearby situations.

It is important to note that, for functions changing rapidly in time, when moving along a ravine the correlation between the values of the function at points separated by the ravine-step distance becomes small, and in this sense the “ravine method” begins to approach the nonlocal gradient method. A further increase in the dependence of Φ on time leads to a decrease also in local correlation, and in these cases the tactic approaches blind search. In essence, the “ravine method” includes both the nonlocal gradient method and blind search, giving a substantial gain in those cases where the speed of the search exceeds the speed of change of Φ , and not yielding to these methods in the remaining cases. A similar situation also obtains when the degree of organization of the function decreases.

In conclusion, let us note that there seems to us to be a plausible connection between the simplest tracking systems, methods of blind search, local and nonlocal methods of automatic optimization, on the one hand, and the levels of construction of movement in humans and higher animals, first considered by N. A. Bernstein (7), on the other.

We express our gratitude to M. A. Evgrafov, L. N. Ivanova, and I. I. Pyatetskii-Shapiro for numerous useful discussions.

Received
14 XII 1960

REFERENCES

1. Tsian Syo-sen, *Technical Cybernetics*, IL, 1956.
2. A. A. Feldbaum, *Computing Devices in Automatic Systems*, Moscow, 1959.
3. W. R. Ashby, *An Introduction to Cybernetics*, London, 1956; W. R. Ashby, *Introduction to Cybernetics*, IL, 1959.
4. G. V. Savinov, *Problems of Cybernetics*, issue 4, Moscow, 1960.
5. A. A. Feldbaum, *Automation and Remote Control*, 17, No. 11 (1956).
6. A. A. Feldbaum, *Automation and Remote Control*, 19, No. 8 (1958).
7. N. A. Bernstein, *On the Construction of Movements*, Moscow, 1948.

* The choice of the ravine step determines the qualitative characteristics of the search tactic. With a prescribed ravine step we “climb over small ridges” and “go around high mountains.” These scales are determined precisely by the magnitude of the ravine step.

** These considerations are related to the range of questions considered by M. L. Tsetlin, connected with the behavior of automata, which are intended to be set forth separately.

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.