ON SYSTEM OPTIMIZATION
[Cybernetics.—1980.—No 5]

Modern theory and practice of optimization are based on the classical formulation of optimization problems. As is known, the essence of this formulation consists in finding a point (or a set of points) p in a predetermined invariable allowable region P, in which a given scalar objective function $f(p)$ takes an extreme value.

For a large number of planning and economic and industrial design tasks, such a formulation is unsatisfactory in at least two respects. First, the objective function $f(x)$ in such problems is not scalar, but vector. At the same time, it turns out to be practically irreducible to a scalar function by some a priori procedure (for example, the procedure for weighing various components of the original vector function). Secondly, the permissible area of P may change during the optimization process. Moreover, its purposeful change is precisely the main content essence of the optimization process for the class of problems under consideration.

Since the laws of possible changes in the permissible domain P are usually set by a system of models, it is natural to call the described approach to optimization problems a system approach. Note that with the system approach, changes in the constraints that define the permissible domain in the space of certain parameters occur, as a rule, as a result of a sequence of solutions selected from a discrete set of possible ones. Moreover, this set itself is usually completely undefined at the beginning of the optimization process and is replenished in the process of dialogue with people (planners or designers) who do not have fully formalized techniques for developing new solutions.

Here is one of the characteristic formalized statements of the system optimization problem. In order to better understand the idea, illustrating it graphically, let's consider a two-criterion case. Let's also assume that the choice of the values of these criteria uniquely determines the corresponding solution. In other words, the desired solution is searched directly in the space of K optimization criteria, which we will denote x_1 and x_2 (Fig. 1).

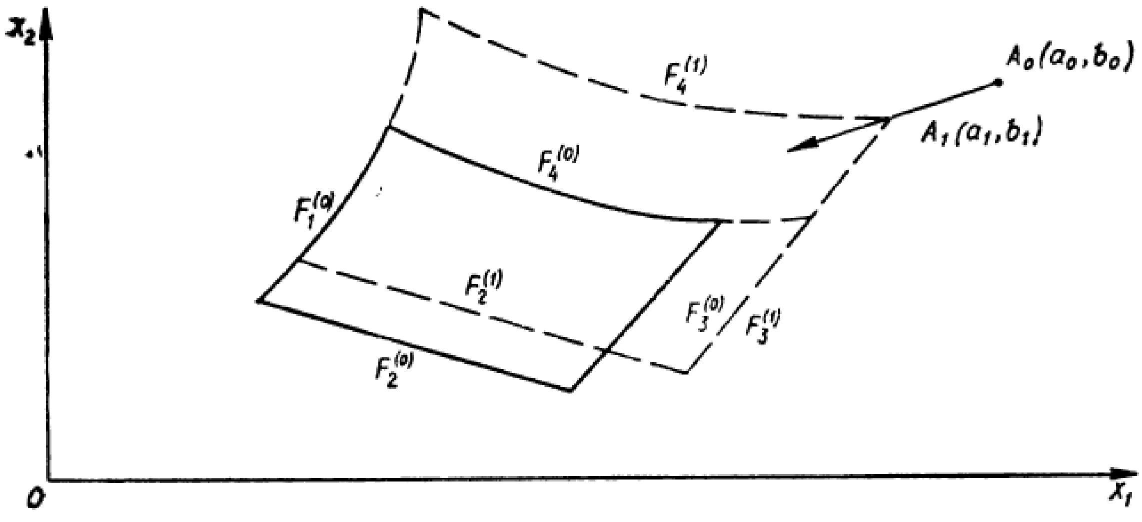

Рис. 1.

The solution process begins with that in a given space K, a certain point A_0 with coordinates a_0 and b_0 is selected — the desired solution to the problem. Next, the initial constraints $F_1^{(0)}(x_1,x_2)\geqslant 0,\ldots,F_n^{(0)}(x_1,x_2)\geqslant 0$ are constructed, which specifies the initial allowable region P_0. A direct check establishes the fact that point A_0 belongs or does not belong to the region of P_0. In the first case, in principle, the usual (classical) optimization procedure can be applied either according to one of the criteria x_1, x_2 or according to one or another combination of them. However, in the system approach, a completely different technique is usually used, namely: in accordance with the higher-level model M, which controls the selection of criteria, the point A_0 is taken out of the limits of the permissible region P_0, as shown in Fig. 1.

Then the constraints that are not met at point A_0 are highlighted (in the case under consideration, they will be $F_3^{(0)}$ and $F_4^{(0)}$). Referring to the M_3 and M_4 models that form these constraints, certain solutions are tested in the interactive dialogue mode that change the corresponding constraints in the right direction (if such a change is possible). In this case, the direction that reduces the absolute value of the negative discrepancies $F_i^{(0)}(a_0,b_0)$ (in the case under consideration, $F_3^{(0)}(a_0,b_0)$ and $F_4^{(0)}(a_0,b_0)$) is considered necessary.

It should be borne in mind that in many cases the constraints F_i are interrelated, so that a change in one of them leads to a change in a certain part of the two constraints. The control of the choice of solutions for changing constraints is determined by minimizing a certain penalty function $g_0(a_0,b_0)$. As such a function, the maximum absolute value of negative discrepancies $\lambda_i F_i^{(0)}(a_0,b_0)$ is usually chosen (where λ_i are some positive weighting coefficients). If there are no such discrepancies, then by definition $g_0(a_0,b_0)=0$.

As a result of the control, a number of solutions $R_1,\ldots,R_m$ appear leading to a decrease in the value of the penalty function, which after the $m-th$ solution we denote $g_0(a_0,b_0)$. By changing the constraints, each of the decisions made leads to a corresponding change in the permissible area. Figure 1 shows two such changes. The first of them changes the constraints of $F_3^{(0)}$ and $F_2^{(0)}$, replacing them respectively with the constraints of $F_3^{(1)}$ and $F_2^{(1)}$. The second affects only $F_4^{(0)}$ replacing it with the restriction $F_4^{(1)}$. The permissible area P_2 obtained after these changes is bounded by the lines $F_1^{(0)},F_2^{(1)},F_3^{(1)}$ and $F_4^{(1)}$, and the corresponding value of the penalty function is $g_2(a_0,b_0)$. Note that it is impossible to choose a finite permissible area in advance due to the fact that the sequence of regions $P_0,P_1,\ldots$ may not be ordered. In addition, the huge complexity of forming new constraints does not allow us to perform this work in advance, since it would require a lot of extra work to change insignificant constraints.

If $g_2(a_0,b_0)\neq 0$ (see Fig. 1), and there are no solutions leading to a further decrease in the value of the penalty function, then there is a return to the higher model M, which controls the choice of the desired solution to problem $A(a,b)$. By a series of successive solutions $D_1,D_2,\ldots,D_k$ to change the initial solution of problem $A_0(a_0,b_0)$, it is successively replaced by $A_1(a_1,b_1),\ldots,A_k(a_k,b_k)$ until the next point $A_k(a_k,b_k)$ is in the acceptable area (in Figure $k=1$). Solutions for changes are selected from an acceptable set of solutions in order to minimize the penalty function. This process is close to the classical optimization process, except only for the fact that the steps are chosen not arbitrarily, but in accordance with acceptable (by model M) solutions.

Finally, after the point A_k falls into the final permissible area P_m, an additional optimization procedure can be applied for any combination of criteria x_1 and x_2 within this permissible area. This procedure differs from the classical one only in that the choice of optimization steps is not arbitrary, but is controlled by a higher-level model M. If the further improvement of the selected criterion is hindered by some constraints that can be further changed in the right direction, then the optimization process can be continued by including successive solutions for such changes in it.

Note that an unambiguous definition of the solution of the problem by choosing the values of all optimization criteria is not as rare as it may seem at first glance. It is the case, for example, for planning and economic tasks, where the criterion (vector) is the net output of different types of products, and the solution of the problem is the full output [1]. In the case when there is no such unambiguity, the space in which the solution is being searched, in addition to the coordinates that correspond to the optimization criteria, may have others. The optimization process described above is complicated then by the fact that the points $A_i(a_i, b_i)$ are replaced by hyperplanes. The definition of the penalty function also becomes more complicated: for example, the distance from the selected hyperplane to the next allowable area in space with specified compressions (stretches) along the axes corresponding to the optimization criteria, can be chosen as it.

In the most general case, point sets of arbitrary form can appear instead of hyperplanes. There are possible statements in which the values of the criteria are ambiguously defined on these sets, and in order to distinguish more or less preferred solutions on the sets, the corresponding weight functions are set (by the model of the highest level M). However, an important feature of system optimization that persists with all approaches, in addition to multi-criteria and the possibility of changing the allowable area, is the interaction of models of different levels. In the case of planning and economic tasks, decisions in these models are made by managers of various levels, and in the case of design tasks - by designers working on various parts of the project.

The author has developed one of the specific optimization schemes based on the stated principles implemented in the "Displan" system [1].

BIBLIOGRAPHY

1. Glushkov V. M. Macroeconomic models and principles of construction of the OGAS. M.: Statistics, 1975.