DEVISING METHODS FOR PLANNING A MULTIFACTORIAL MULTILEVEL EXPERIMENT WITH HIGH DIMENSIONALITY

This paper considers the task of planning a multifactorial multilevel experiment for problems with high dimensionality. Planning an experiment is a combinatorial task. At the same time, the catastrophically rapid growth in the number of possible variants of experiment plans with an increase in the dimensionality of the problem excludes the possibility of solving it using accurate algorithms. On the other hand, approximate methods of finding the optimal plan have fundamental drawbacks. Of these, the main one is the lack of the capability to assess the proximity of the resulting solution to the optimal one. In these circumstances, searching for methods to obtain an accurate solution to the problem remains a relevant task.

Two different approaches to obtaining the optimal plan for a multifactorial multilevel experiment have been considered. The first of these is based on the idea of decomposition. In this case, the initial problem with high dimensionality is reduced to a sequence of problems of smaller dimensionality, solving each of which is possible by using precise algorithms. The decomposition procedure, which is usually implemented empirically, in the considered problem of planning the experiment is solved by employing a strictly formally justified technique. The exact solutions to the problems obtained during the decomposition are combined into the desired solution to the original problem. The second approach directly leads to an accurate solution to the task of planning a multifactorial multilevel experiment for an important special case where the costs of implementing the experiment plan are proportional to the total number of single-level transitions performed by all factors. At the same time, it has been proven that the proposed procedure for forming a route that implements the experiment plan minimizes the total number of one-level changes in the values of factors. Examples of problem solving are given

Keywords: planning of multifactorial multilevel experiment, problem with high dimensionality, decomposition plan, exact plan

Lev Raskin
Doctor of Technical Sciences,
Professor, Head of Department*
Oksana Sira
Corresponding author
Doctor of Technical Sciences, Professor*
E-mail: topology@ukr.net
*Department of Distributed Information
Systems and Cloud Technologies
National Technical University «Kharkiv
Polytechnic Institute»
Kyrpychova str., 2, Kharkiv, Ukraine, 61002

1. Introduction

The complete factorial orthogonal multilevel experiment is represented in the coordinate space of the corresponding dimensionality by the vertices of the hypercube [1, 2]. The task of planning such an experiment is to find a route to traverse these vertices at minimal (cost or time) costs. Assume a k-factor experiment in which each factor can take a value at one of the m levels. The number of vertices of the corresponding hypercube is $N=m^k$, and the number of «correct» routes to traverse these vertices is determined from the ratio $M=N!=(m^k)!$ (The «correct» route is a Hamiltonian path of traversing all N vertices without loops, containing $N-1$ transitions). The plan of the experiment is usually represented by a matrix in which the number of columns is equal to the number of factors, and the number of rows is equal to the number of vertices. The corresponding graph for the two-factor (F_1, F_2) three-level (-1; 0; 1) experiment is shown in Fig. 1. A possible option for traversing the vertices of the graph is given in Table 1. In this case, the transition from one of the vertices to any other corresponds is matched by a change in the factor and (or) a change in the level.

Table 1

An option to traverse vertices for a two-factor three-level experiment

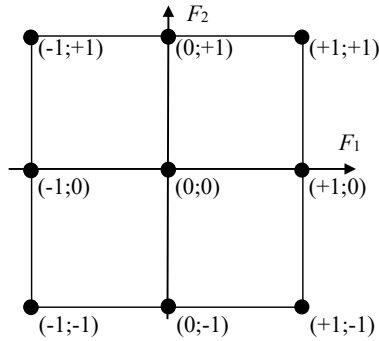| Experiment No.	Factor	
	F_1	F_2
1	−1	−1
2	−1	0
3	−1	+1
4	0	−1
5	0	0
6	0	+1
7	+1	−1
8	+1	0
9	+1	+1

Fig. 1. Experiment graph

For example, in the above plan of a two-factor three-level experiment, the transition from vertex 1 to vertex 2 corresponds to a change in the level of only the F_2 factor, and the transition from vertex 3 to vertex 4 corresponds to a change in the levels of both factors. The costs associated with moving from one vertex to another are determined by the corresponding changes in the levels of factors.

Assign the sets of values of k-factor levels for the i-th experiment $\left(S_1^{(i)}, S_2^{(i)}, ..., S_k^{(i)}\right)$ and for the j-th experiment $\left(S_1^{(j)}, S_2^{(j)}, ..., S_k^{(j)}\right)$. Then the costs of changing the levels of factors in the transition from the i-th experiment to the j-th would be determined from the following formula:

$$C_{ij} = \sum_{e=1}^{k} C\left(S_e^{(i)}, S_e^{(j)}\right). \tag{1}$$

Here, $C\left(S_e^{(i)}, S_e^{(j)}\right)$ is the cost of moving from the $S_e^{(i)}$ level for the e-th factor in the i-th experiment to the $S_e^{(j)}$ level for the same factor in the j-th experiment. At the same time, if one specifies a certain sequence of vertices $\left(i_1, i_2, ..., i_N\right)$, which determines the «correct» route of transitions from the initial vertex i_1 to the final i_N, then the costs of traveling this route would equal:

$$C = \sum_{q=1}^{N-1} C_{q,q+1}. \tag{2}$$

It should be noted that the problem of finding the best-cost route for traversing vertices belongs to the class of combinatorial ones and its computational complexity grows very quickly with an increase in the dimensionality of the problem (Table 2).

Table 2

Dependence of the number of traverse routes
on the problem dimensionality

Number of factors (k)	Number of levels (m)	Number of vertices (N)	Number of routes (M)
2		4	24
3	2	8	40320
4		16	$2,09 \cdot 10^3$
2		9	362880
3	3	27	$\sim 10^{35}$
4		81	$\sim 5 \cdot 10^{120}$

It is clear that this problem cannot be solved by simple sorting in the case of actual dimensionality.

2. Literature review and problem statement

For problems whose dimensionality is not too large ($N<25$), the possibility to obtain an accurate solution to the problem of planning an experiment was shown as early as the beginning of 1970s, by using any known algorithm developed for the traveling salesman problem [3]. Subsequently, approximate decomposition algorithms for solving this problem were proposed, providing acceptable accuracy for problems whose dimensionality excludes the possibility of obtaining a solution by exact methods [4] but the accuracy of the solution is not discussed [5]. The same idea was implemented in work [6] where the restructuring procedure was carefully substantiated and described in detail for problems with low dimensionality. In [7], an approximate method is proposed that implements the iterative procedure for obtaining a plan. The complexity of the problem is discussed in [8] for problems in which high dimensionality arises due to their multi-level nature. In [9], the absence of a general approach to solving the task of forming an optimal plan for problems with arbitrary dimensionality is stated. In work [10], the recommendation on using block algorithms in these problems for obtaining plans is formulated. At the same time, it should be borne in mind that all known approximate algorithms for solving the routing problem have the following irreparable drawback – structural incompleteness. Specifically, the procedure for obtaining a solution for all optimization algorithms has the same type of three-stage structure: the first stage is to obtain the initial solution to the problem; the second stage is to check the optimality of the solution; the third stage implies that once the tested solution is not optimal, then the transition to a new solution is performed, not worse than the previous one. Known approximate algorithms for solving the routing problem lack the second and third stages. This means that when deriving the next solution, there is fundamentally no possibility of assessing the degree of its proximity to the optimal solution. In addition, in the process of solving the problem, situations really arise where quite a lot of iterations in a row do not lead to the improvement of the existing solution but that, however, does not give grounds for terminating the solution. Therefore, actual running programs terminate the solving procedure either after a predetermined solving time has elapsed or after a specified number of iterations have been executed. The claims in many well-known publications that the approximate algorithm did lead to a solution close to optimal have no evidential value. They typically refer to specially designed test tasks whose optimal solution was known or obtained in advance. It follows that the task of finding accurate algorithms for solving the routing problem remains relevant. However, the direction of the search for precise methods of solving routing problems may need to be changed. Known algorithms for the exact solution to combinatorial problems, in order to reduce the volume of sorting, are focused on finding and using various techniques for screening out unpromising options. At the same time, in many cases, the number of analyzed variants is significantly reduced, which ensures that an accurate solution is obtained if the dimensionality of the problem is not too large. At the same time, one should point to the insufficient knowledge of the possibilities of rational use of an alternative resource of computing systems – memory involved in the implementation of sorting operations.

The above defines the purpose and objectives of the current study.

3. The aim and objectives of the study

The goal is to devise methods for planning a multifactorial multilevel experiment with high dimensionality, minimizing the total costs of implementing the experiment plan.

To accomplish the aim, the following tasks have been set:
– to devise an accurate routing method that increases the value of the dimensionality of the problems being solved;
– to construct a fast approximate method for solving the problem of planning an experiment with high dimensionality;
– to build an accurate routing method for arbitrarily high dimensionality, minimizing the total cost of implementing the plan for the case where these costs are determined by the number of single-level transitions.

4. The study materials and methods

The methods of combinatorics and graph theory were used to resolve our tasks.

When solving the problems, we shall assume that the following is set:
– k – the number of factors;
– m – the number of levels for each factor;
– $C\left(S_e^{(i)}, S_e^{(j)}\right)$ – the cost of moving from the $S_e^{(i)}$ level for the e-th factor in the i-th experiment to the $S_e^{(j)}$ level for the same e-th factor in the j-th experiment;
– $C_{ij} = \sum_{e=1}^{k} C\left(S_e^{(i)}, S_e^{(j)}\right)$ – the total costs of transition from the i-th experiment to j-th;
– $N = m^k$ – the total number of experiments;
– i_q – the number of the q-th experiment in the route of traversing all the set of experiments $q = 1, 2, ..., N$.

In this case, the sequence of numbers $I = (i_1, i_2, ..., i_q, ..., i_N)$ sets some route to traverse the entire set of experiments. Then the total cost of implementing plan I would be determined from ratio (2). The task is to find a route that minimizes these costs.

5. Results of devising methods for solving problems of planning multifactorial multilevel experiments

5. 1. Building an accurate method for constructing the Hamiltonian path

To solve this problem, we shall compile a matrix of costs for all possible one-step transitions from one experiment to another:

$$C_1 = \begin{pmatrix} \infty & C_{12} & C_{13} & ... & C_{1q} & ... & C_{1N} \\ C_{21} & \infty & C_{23} & ... & C_{2q} & ... & C_{2N} \\ ... & ... & ... & ... & ... & ... & ... \\ C_{q1} & C_{q2} & C_{q3} & ... & \infty & ... & C_{qN} \\ ... & ... & ... & ... & ... & ... & ... \\ C_{N1} & C_{N2} & C_{N3} & ... & C_{Nq} & ... & \infty \end{pmatrix}. \tag{3}$$

It is assumed that the transition graph is fully accessible, that is, from any vertex a one-step transition to any other is possible. Next, for two randomly selected experiments with numbers i, j, we shall find the least costly transition from i to j using the intermediate experiment q:

$$C_{ij} = \min_{q} \left\{ C_{iq} + C_{q1} \right\}. \tag{4}$$

Performing this operation for all pairs $i = 1, 2, ..., N$, $j = 1, 2, ..., N$, we obtain a matrix of the least expensive two-step transitions from all possible experiments to all others. Operation (4) is conveniently represented in the following matrix form:

$$C_2 = C_1 \oplus C_1, \tag{5}$$

where the $\oplus$ operation is performed in line with (4) for all elements of the matrix C_2. At the same time, based on the results of calculations of this matrix, for its each element (i, j), it is necessary to remember the number q of the intermediate experiment, which provides the least costly two-step transition. Further, according to the same scheme, the matrix $C_3 = C_1 \oplus C_2$ is calculated. Here, the numbers of the intermediate experiments are also to be remembered.

The following procedure:

$$C_p = C_1 \oplus C_{p-1}, \quad p = 1, 2, ..., N - 1, \tag{6}$$

continues until the C_{N-1} matrix is obtained, the last column of which contains the values of the least expensive $(N-1)$-step transitions, starting from some initial experiment. From these values, one needs to choose the smallest. The position of this element in the column determines the numbers of the initial, final, and intermediate experiments.

Let us analyze the proposed procedure. Merits:
a) the method provides an accurate solution to the problem;
b) to implement the method, the simplest matrix operations are used, the total number of which polynomially depends on N. The disadvantage of the method is obvious: the need to memorize the numbers of intermediate experiments, which complicates the procedure, and, more significantly, the higher the dimensionality of the problem. In this regard, to solve problems with high dimensionality, it is advisable to find an approximately optimal route, using for this purpose, for example, the decomposition procedure.

5. 2. Devising an approximate method for constructing a Hamiltonian path for problems with high dimensionality

The essence of the proposed approach is to implement the following five-step procedure.

Step 1. All the set of experiments is broken down into subsets. In the task of planning the experiment, splitting is natural to perform by fixing for each subset the level of a factor (or group of factors).

Step 2. Find a rational way to traverse the subsets corresponding to different values of the factor selected for decomposition.

Step 3. According to the chosen route for traversing subsets, for each pair of neighboring subsets, the least costly approach is found from one subset to another. In this case, for the first of these subsets, the experiment output is fixed, and, for the second subset, the experiment input.

Step 4. In each of the subsets, a local optimal route between the input and output experiments is found.

Step 5. All received local routes are connected in the order of passage of subsets into a single resulting route.

Consider an example. Let the optimal traverse route be searched for in a three-factor three-level experiment. The corresponding graph with vertex numbering is shown in Fig. 2. In this graph, the abscissa axis displays the change in the levels of the F_1 factor, the ordinal axis – the F_2 factor, the applicate axis – the F_3 factor.
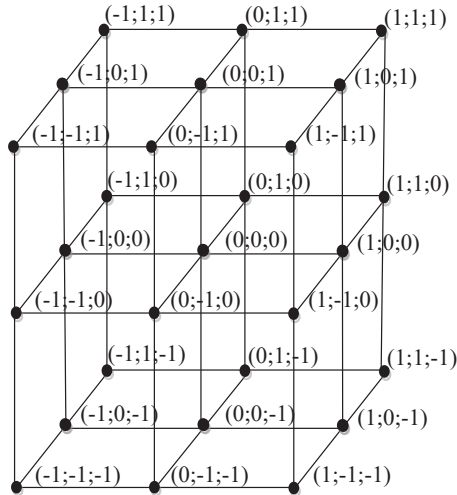
Fig. 2. Graph of a three-factor three-level experiment

We introduce a table of cost values for the transition from one level to another for each factor.

Table 3

Cost value (a.u.)

Transition cost	Factor		
	F_1	F_2	F_3
from «−1» to «0»	2	3	4
from «0» to «1»	2	3	4
from «1» to «0»	2	3	4
from «0» to «−1»	2	3	4
from «−1» to «1»	4	6	8
from «1» to «−1»	4	6	8

Differences in the values of the cost of moving from one level to another for different factors determine the rational way of decomposition. In the problem under consideration, decomposition by any factor reduces the initial problem of finding a route to traverse a three-factor set containing 27 experiments to three simpler subtasks for finding routes in two-factor subsets containing 9 experiments. Let us evaluate the computational advantages that arise when using decomposition. In the experiment with n experiments, the total number of «correct» routes is $M = n! \approx \left(\dfrac{n}{e}\right)^n \cdot \sqrt{2\pi n}$. For the original problem, $n=27$, and

$$M_0 \cong \left(\frac{27}{2.71}\right)^{27} \cdot \sqrt{2\pi \cdot 27} \cong 1.2 \cdot 10^{28}.$$

For each subtask, $n=9$, and

$$M_1 \cong \left(\frac{9}{2.71}\right)^9 \cdot \sqrt{2\pi \cdot 9} \cong 3.8 \cdot 10^5.$$

The difference is oh so huge but the amount of total sorting is still excessively large. This difference remains quite impressive for many effective brute force methods such as the method of branches and boundaries, ant colonies, taboo search, genetic algorithm [4, 11–13] but not for all. An estimate of the amount of search for one of the fastest heuristic algorithms (elastic network method [10]) is as fol-

lows: $M \cong n \cdot 2^n$. At the same time, if, in the initial problem, $M_0 = 27 \cdot 2^{27} \cong 3.6 \cdot 10^{10}$, then, for each subtask after decomposition, $M_1 = 9 \cdot 2^9 \cong 4608$. The expediency of decomposition is obvious. Let us move on to solving the problem.

Step 1. The question of the decomposing factor is solved by choosing the one for which the transition from one level to another is the costliest. At the same time, the number of the most expensive transitions would be minimal. In a given example, this is the F_3 factor. As a result, the initial set of experiments is divided into three subsets $A^{(-1)}$, $A^{(0)}$, $A^{(1)}$.
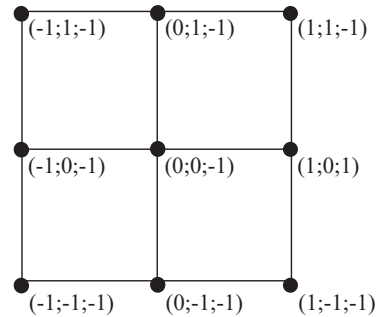

Fig. 3. Graph of a two-factor subset $A^{(-1)}$ at a fixed level of the F_3 factor value, $F_3 = -1$
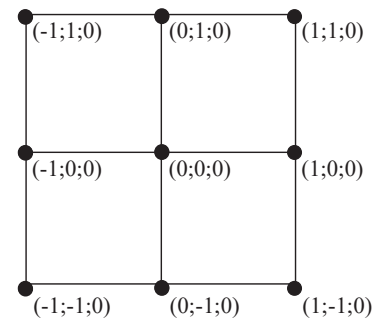

Fig. 4. Graph of a two-factor subset $A^{(0)}$ at a fixed level of the F_3 factor value, $F_3 = 0$
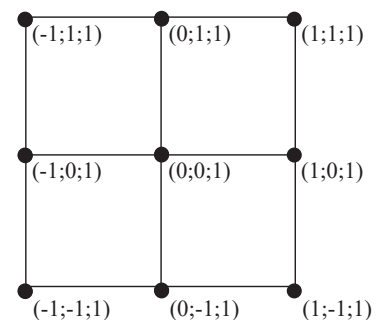

Fig. 5. Graph of a two-factor subset $A^{(1)}$ at a fixed level of the F_1 factor value, $F_3 = 1$

Step 2. As the initial one, let us choose an experiment from the subset $A^{(-1)}$, in which all three factors take a value equal to −1. The natural order of traversing the subsets is as follows: $A^{(-1)}$, $A^{(0)}$, $A^{(1)}$.

Step 3. In the problem to be solved, in accordance with the cost table, transitions from $A^{(-1)}$ to $A^{(0)}$ and from $A^{(0)}$ to $A^{(1)}$ are naturally performed by changing the level of only the F_3 factor with the fixed values of the levels of factors F_1 and F_2. At the same time, the costs of both transitions are

equal to 4 regardless of what the values of the levels of factors F_1 and F_2 are in the experiments selected for the transitions. Taking this circumstance into consideration, we shall choose a pair $((1, 1, -1), (1, 1, 0))$ for the transition from $A^{(-1)}$ to $A^{(0)}$, and a pair $((-1, -1, 0), (-1, -1, 1))$ to move from $A^{(0)}$ to $A^{(1)}$.

Step 4. As a result of step 3, the initial and final experiments for traversing in subsets are defined. Let us list them: in the $A^{(-1)}$ subset, the $((-1, -1, -1), (1, 1, -1))$ pair; in the $A^{(0)}$ subset, the $((1, -1, 0), (-1, -1, 0))$ pair; in the $A^{(1)}$ subset, the $((-1, -1, 1), (1, 1, 1))$ pair.

Step 5. In each of the obtained subsets, the problem of finding a locally optimal route for traversing the corresponding set of experiments is independently solved. These three tasks have some common features to consider when choosing a route:

1. In all cases, the same type of problem of finding a two-factor route between the known initial and final experiments with seven intermediate experiments is solved. Thus, each route contains eight transitions.

2. In accordance with the values of costs when passing routes, it is advisable to use those transitions whose execution leads to a change to the neighboring level only by one of the factors:

3. The cost of changing a level for the F_1 factor is less than the corresponding costs of change for the F_2 factor.

Taking into consideration these features, as well as the input and output experiments known for each subset, we obtain the following routes.

For subset $A^{(-1)}$:

$(-1, -1, -1), (0, -1, -1), (1, -1, -1), (1, 0, 1), (0, 0, -1),$
$(-1, 0, -1), (-1, 1, -1), (0, 1, -1), (1, 1, -1);$

for subset $A^{(0)}$:

$(1, 1, 0), (0, 1, 0), (-1, 1, 0), (-1, 0, 0), (0, 0, 0), (1, 0, 0),$
$(1, -1, 0), (0, -1, 0), (-1, -1, 0);$

for subset $A^{(1)}$:

$(-1, -1, 1), (0, -1, 1), (1, -1, 1), (1, 0, 1), (0, 0, 1),$
$(-1, 0, 1), (-1, 1, 1), (0, 1, 1), (1, 1, 1).$

Each of these routes contains six transitions out of eight with a cost value of 2 a.u. and two transitions with a cost value of 3 a.u.

Step 6. Combining these three locally optimal routes with the addition of transitions between subsets produces the desired optimal route, shown in Fig. 6. The problem is solved.

The effectiveness of using decomposition in the planning of an experiment is not in doubt. The most important advantage of decomposition is the reduction of the original complex (perhaps unsolvable) problem to a set of simpler, solvable ones. This benefit of the method in the problem of planning an experiment is further supported by the following structural features of the factor experiment.

Let us consider first a more general variant of the statement of a routing problem – the traveling salesman problem [3–5]. In this problem, *n* points and a matrix of costs for moving from one point to another are specified. It is required to find the «right» route to traverse this set of points, minimizing the total costs. When solving this problem, all the necessary stages are sequentially performed in accordance with the decomposition technique: the partitioning of the

original set into subsets; finding the order of traversing subsets and transitions between them; the calculation of locally optimal routes in subsets, and combining these routes into the desired route. It is clear that this approach does not warrant finding an accurate solution to the problem.
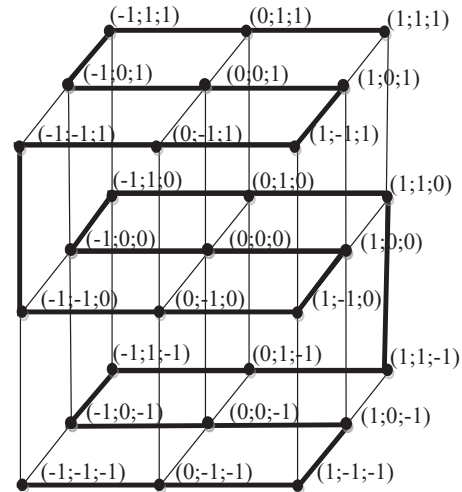

Fig. 6. Optimal route to traverse 27 experiments in a three-factor three-level experiment

Possible inaccuracies arise already at the initial stage – during clustering. The point is in the structural features of this procedure. It is not properly formalized. Indeed, it is not known how many subsets the original set should be broken down into; the clustering procedure itself is not canonized; transitions between subsets are determined by a «greedy» algorithm. It is clear that the mistakes of the initial stages lead to erroneous solutions on all the others.

The use of the same decomposition technique in the task of planning an experiment, of course, does not guarantee an accurate solution. However, the situation here is different. Experiments in the task of planning an experiment are marked by the values of the levels of factors. The partitioning of their set into subsets is performed according to a set of formal features and their fixation gives the corresponding unambiguous result. The costs of transitioning from one experiment to another are also tied to the identifying numerical characteristics of those experiments. Therefore, the choice of a solution at each stage of the procedure relies heavily on the initial data of the problem. Due to these circumstances, the use of decomposition in the task of planning an experiment yields an expected good result.

However, for multifactorial and multi-level tasks of planning an experiment with high dimensionality, the problem of finding optimal or close to them solutions remains relevant. At the same time, the indisputable futility of finding strictly optimal solutions in a general case still leaves this problem open to many specific special cases.

5. 3. Devising an accurate method for finding the Hamiltonian path for problems with equal-cost single-level transitions

Consider the problem of planning experiments for a special case where the cost of moving from one experiment to another is proportional to the number of changes in the levels of factors. For example, it is natural to assume that the cost of a two-level transition from level «−1» to level «1» for any

of the factors in a three-level task is equivalent to the cost of moving first from level «−1» to level «0» with the subsequent transition from level «0» to level «1». In this case, if the costs for all single-level transitions for all factors are approximately the same, then the total costs would be proportional to the total number of single-level transitions. It is clear that the optimal route for traversing all the experiments of such an experiment corresponds to the minimum number of one-level changes. The total number of experiments of the k-factor s-level experiment is $N=S^k$. At the same time, if there is a plan for which during the transition from one experiment to another there is a change in only one level by one of the factors, then the minimum possible number of level changes is $M=N-1=S^k-1$.

Let us now proceed to consider the problem of building a method for forming an optimal plan. For the convenience of presenting the material, we shall introduce some special system of designations.

Introduce:

$[1]_k$ – a column containing k ones;

$[2]_k$ – a column containing k twos;

$[q]_s$ – a column containing k qs;

$$A_1^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

– a possible plan for a one-factor two-level experiment;

$$A_1^S = \begin{pmatrix} 1 \\ 2 \\ \dots \\ S-1 \\ S \end{pmatrix}$$

– a possible plan for a one-factor s-level experiment;

$$\overline{A}_1^S = \begin{pmatrix} S \\ S-1 \\ \dots \\ 2 \\ 1 \end{pmatrix}$$

– a plan of a one-factor s-level experiment inverted with respect to A_1^S;

$A_1^1 = \begin{pmatrix} 1 & 1 \end{pmatrix}$ – a possible plan for a two-factor one-level experiment;

$$A_2^2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \\ 2 & 2 \end{pmatrix}$$

– a possible plan for a two-factor two-level experiment;

$$\overline{A}_2^2 = \begin{pmatrix} 2 & 2 \\ 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}$$

– a plan of a possible two-factor two-level experiment inverted in relation to A_2^2;

$$A_3^2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \\ 2 & 2 & 2 \end{pmatrix}$$

– a possible plan for a three-factor two-level experiment.

A more compact record of the last plan, taking into consideration the introduced designations, can be written through smaller plans as follows:

$$A_3^2 = \begin{pmatrix} [1]_4 & A_2^2 \\ [2]_4 & A_2^2 \end{pmatrix},$$

where $A_2^2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} [1]_2 & A_1^2 \\ [2]_2 & A_1^2 \end{pmatrix}$, $A_1^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

Then

$$A_4^2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 2 \\ \dots & \dots & \dots & \dots \\ 2 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 \end{pmatrix}$$

– a possible plan for a four-factor two-level experiment.

This plan, taking into consideration the introduced simplified notation, takes the following form:

$$A_4^2 = \begin{pmatrix} [1]_8 & A_3^2 \\ [2]_8 & A_3^2 \end{pmatrix}.$$

Similarly, plans for experiments are formed, the number of levels of which is more than two:

$$A_2^3 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 2 & 1 \\ 2 & 2 \\ 2 & 3 \\ 3 & 1 \\ 3 & 2 \\ 3 & 3 \end{pmatrix} = \begin{pmatrix} [1]_3 & A_1^3 \\ [2]_3 & A_1^3 \\ [3]_3 & A_1^3 \end{pmatrix}$$

– a possible plan for a two-factor three-level experiment.

The plan of a three-factor three-level experiment using the A_2^3, plan taking into consideration the introduced simplifying notation, takes the following form:

$$A_3^3 = \begin{pmatrix} [1]_9 & A_3^2 \\ [2]_9 & A_3^2 \\ [3]_9 & A_3^2 \end{pmatrix}.$$

Finally, the plan of the k-factor s-level experiment is as follows:

$$A_k^S = \begin{pmatrix} [1]_{S^{k-1}} & A_{k-1}^S \\ [2]_{S^{k-1}} & A_{k-1}^S \\ \dots & \dots \\ [S-1]_{S^{k-1}} & A_{k-1}^S \\ [S]_{S^{k-1}} & A_{k-1}^S \end{pmatrix}.$$

Using the introduced notation, we shall determine the exact method of constructing the optimal plan of an experiment, using the principle of mathematical induction.

Let the optimal route exist and be known for the k-factor s-level experiment, determined by the A_k^S plan. Build now the plan of the $(k+1)$-factor s-level experiment according to the following rule (let us call it rule A):

$$A_{k+1}^S = \begin{pmatrix} [1]_{S^k} & A_k^S \\ [2]_{S^k} & \overline{A}_k^S \\ \dots & \dots \\ [S-1]_{S^k} & A_k^S \\ [S]_{S^k} & \overline{A}_k^S \end{pmatrix}. \tag{7}$$

Let us now show that the optimality of the A_{k+1}^S plan built according to rule A follows from the optimality of the A_k^S plan. The optimal A_k^S plan is implemented using (S^k-1) one-level transitions. Let us calculate the number of such transitions in the A_{k+1}^S plan. It is obviously equal to:

$$M\left[A_{k+1}^S\right] = S\left(S^k-1\right) + S - 1 = S^{k+1} - S + S - 1 = S^{k+1} - 1. \tag{8}$$

Thus, the A_{k+1}^S plan contains the minimum possible number of single-level transitions and, therefore, it is optimal.

Let us now check that rule A produces an optimal plan for $k=1$.

Record a one-factor s-level plan:

$$A_1^S = \begin{pmatrix} 1 \\ 2 \\ \dots \\ S \end{pmatrix}.$$

This plan contains $(S-1)$ one-level transition and, therefore, it is optimal. Now, using rule A, let us build a plan for a two-factor s-level experiment:

$$A_2^S = \begin{pmatrix} [1]_S & A_1 \\ [2]_S & \overline{A}_1 \\ \dots & \dots \\ [S]_S & A_1^S \end{pmatrix}. \tag{9}$$

Calculate the number of single-level transitions in the A_2^S plan. It is equal to:

$$M\left[A_2^S\right] = S\left(S-1\right) + S - 1 = S^2 - 1. \tag{10}$$

Thus, choosing the initial optimal one-factor s-level plan and using the A rule, we obtain the optimal plan of a two-factor s-level experiment. However, if a statement is true for $k=1$ and, given its validity for arbitrary k, it follows that it is also true for $k+1$, then, according to the principle of mathematical induction, this statement would hold for any k.

Thus, the rule of formation of the optimal k-factor s-level experiment for any values of k and s has been obtained.

Here is an example of the application of this rule to build an optimal plan for a three-factor three-level experiment. The sequence of actions is as follows. First, the optimal plan of a one-factor three-level experiment A_1^3 is formed. Then, according to rule A, the optimal plan of a two-factor three-level experiment A_2^3 is determined. Finally, after that, again using rule A, we obtain the desired optimal plan for a three-factor three-level experiment.

For three factors F_1, F_2, F_3, set the values of three levels – (−1, 0, 1). Further relations are given without explanation.

$$A_1^3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad A_3^3 = \begin{pmatrix} [-1]_3 & A_1^3 \\ [0]_3 & \overline{A}_1^3 \\ [1]_3 & A_1^3 \end{pmatrix} = \begin{matrix} F_2 & F_1 \\ \begin{pmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & -1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \end{matrix}.$$

$$A_3^3 = \begin{pmatrix} [-1]_9 & A_2^3 \\ [0]_9 & \overline{A}_2^3 \\ [1]_9 & A_2^3 \end{pmatrix} = \begin{matrix} F_3 & F_2 & F_1 \\ \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 0 \\ -1 & -1 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 0 \\ -1 & 0 & -1 \\ -1 & 1 & -1 \\ -1 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \\ 0 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \end{matrix}.$$

The resulting plan of the A_3^3 experiment is optimal. In this regard, the transition from one experiment to another is accompanied by a change in the level of only one factor to the neighboring level. The total number of single-level transitions is 26, as it should be in the optimal plan. Note, by the way, that this A_3^3 plan determines the route of transitions, exactly coinciding with the route that was obtained earlier by the method of decomposition (Fig. 6).

Let us make an important point. The plan obtained using the rule A is not the only possible one. If such a plan is obtained, then any rearrangement of its columns gives a new plan that defines a new route for traversing the experiments. Moreover, this new plan would still be optimal since rearranging the columns does not change the number of single-level transitions. Let us explain what has been said on the example of an optimal three-factor two-level experiment. Let us obtain a plan for such an experiment and give several options for its restructuring, accompanying each new plan with an appropriate figure (Fig. 7–10):

$$A_1^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \overset{F_1}{}, \quad A_2^3 = \begin{pmatrix} [0]_2 & A_1^2 \\ [1]_2 & \overline{A_1^2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \overset{F_2\ F_1}{},$$

$$A_3^2 = \begin{pmatrix} [0]_4 & A_2^2 \\ [1]_4 & \overline{A_2^2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \overset{F_3\ F_2\ F_1}{}.$$
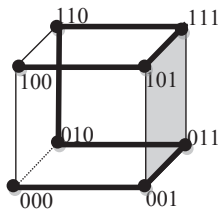
Fig. 7. Optimal route. Original variant

Rearranging the first and third columns in places, we obtain a new plan:

$$\begin{matrix} F_1 & F_2 & F_3 \end{matrix}$$
$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$
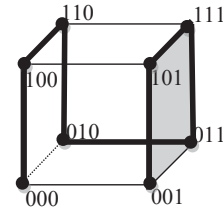
Fig. 8. Optimal route. Option 2

Let us rearrange the second and third columns in places now. In this case, we obtain:

$$\begin{matrix} F_3 & F_1 & F_2 \end{matrix}$$
$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$
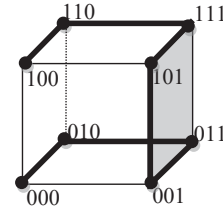
Fig. 9. Optimal route. Option 3

Rearrange the first and second columns in places. The corresponding plan takes the following form:

$$\begin{matrix} F_2 & F_3 & F_1 \end{matrix}$$
$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$
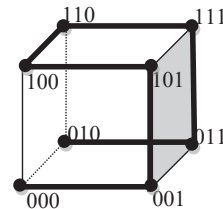
Fig. 10. Optimal route. Option 4

The total number of different optimal routes is determined by the number of possible column permutations and is equal to $k!$. Additional opportunities arise if one selects a different starting vertex in the original one-factor route. The number of such possibilities is exactly s^k. In this case, the total number of optimal routes becomes equal to $s^k \cdot k!$. Note that the existence of some set of routes, equivalent in the number of single-level transitions, can be useful if the cost of transition for different

factors is different. In the example considered, the route options differ from each other in the number of transitions due to different factors. At the same time, the distribution by factors F_1, F_2, F_3 of the number of single-level transitions for the first route takes the form (4; 2; 1), for the second route – (1; 2; 4), for the third route – (2; 4; 1), for the fourth route – (4; 1; 2).

It is clear that it is advisable to use the route in which more expensive transitions are used less often than less expensive ones.

6. Discussion of results of building the methods for planning a multifactor multi-level experiment

A method for solving the problem of planning a multifactorial multilevel experiment that is relevant from the point of view of theory and important for practice has been proposed. The method of solving this problem in a general statement is not known. An approximate approach to solving it based on decomposition has been considered. The method is illustrated by the transition from Fig. 2 through the system of Fig. 3–5 to the final Fig. 6. At the same time, it is shown that taking into consideration the peculiarities of its structure, it is possible to reduce the original NP-complete problem to a set of problems of significantly smaller dimensionality.

A special study was conducted for a practically important case where the costs of implementing an experiment plan are determined by the number of one-level transitions. To solve such a problem, an exact method (ratios (7) to (10)) has been proposed and substantiated. Based on a proven theorem, the computational procedure for forming a plan is extremely simple. Its complexity practically does not depend on the dimensionality of the problem.

A possible area of further research is associated with the construction of a method for solving the problem of planning a multifactorial multilevel experiment for cases where the costs for different single-level transitions are significantly different. In addition, solving this problem is of considerable interest under conditions when the cost values during the transition from one experiment to another are not clearly defined [14] or inaccurately set [15]. In this case, when solving the problem, the approaches proposed in [16, 17] could be used.

7. Conclusions

1. An exact method for solving the routing problem has been devised, the computational complexity of which is polynomial, which significantly increases the dimensionality of the problems actually being solved.

2. A fast approximate method for solving the problem of planning a high-dimensionality experiment has been built. The method is based on the decomposition of the original NP-complex problem into a set of simpler problems, the dimensionality of which makes it possible to use accurate algorithms to solve them.

3. An exact method for solving the problem of planning experiments for the case where the costs of implementing the plan are proportional to the number of single-level transitions have been constructed. The most important advantage of the method is that its computational complexity is not polynomial but linearly depends on the dimensionality of the problem.

References

1. Nalimov, V. V., Chernova, N. A. (1965). Statisticheskie metody planirovaniya ekstremal'nyh eksperimentov. Moscow: Nauka, 340.
2. Adleman, L. (1994). Molecular computation of solutions to combinatorial problems. Science, 266 (5187), 1021–1024. doi: https://doi.org/10.1126/science.7973651
3. Levitin, A. V. (2006). Algoritmy: vvedenie v razrabotku i analiz. Moscow: Vil'yams, 576.
4. Kureychik, V. M., Lebedev, B. K., Lebedev, O. K. (2006). Poiskovaya adaptaciya: teoriya i praktika. Moscow: Fizmatlit, 272.
5. Kormen, T., Leyzerson, Ch., Rivest, R., Shtayn, K. (2006). Algoritmy: postroenie i analiz. Moscow: Vil'yam, 1296.
6. Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. Biometrika, 12 (1-2), 1–85. doi: https://doi.org/10.1093/biomet/12.1-2.1
7. How to Use «Design of Experiments» to Create Robust Designs With High Yield. Available at: https://www.youtube.com/watch?v=hfdZabCVwzc
8. Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science, 22 (11), 1359–1366. doi: https://doi.org/10.1177/0956797611417632
9. Physics envy: Do 'hard' sciences hold the solution to the replication crisis in psychology? Available at: https://www.theguardian.com/science/head-quarters/2014/jun/10/physics-envy-do-hard-sciences-hold-the-solution-to-the-replication-crisis-in-psychology
10. Montgomery, D. C. (2012). Design and analysis of experiments. Wiley, 752.
11. Blum, C. (2005). Ant colony optimization: Introduction and recent trends. Physics of Life Reviews, 2 (4), 353–373. doi: https://doi.org/10.1016/j.plrev.2005.10.001
12. Saymon, D. (2020). Algoritmy evolyucionnoy optimizacii. Moscow: LVR Press, 940.
13. Durbin, R., Willshaw, D. (1987). An analogue approach to the travelling salesman problem using an elastic net method. Nature, 326 (6114), 689–691. doi: https://doi.org/10.1038/326689a0
14. Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8 (3), 338–353. doi: https://doi.org/10.1016/s0019-9958(65)90241-x
15. Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. European Journal of Operational Research, 99 (1), 48–57. doi: https://doi.org/10.1016/s0377-2217(96)00382-7
16. Raskin, L., Sira, O. (2016). Fuzzy models of rough mathematics. Eastern-European Journal of Enterprise Technologies, 6 (4 (84)), 53–60. doi: https://doi.org/10.15587/1729-4061.2016.86739
17. Raskin, L., Sira, O. (2016). Method of solving fuzzy problems of mathematical programming. Eastern-European Journal of Enterprise Technologies, 5 (4 (83)), 23–28. doi: https://doi.org/10.15587/1729-4061.2016.81292