INFORMATICA, 1997, Vol. 8, No. 1, 83–118

CLUSTERING OF PARAMETERS ON THE BASIS OF CORRELATIONS: A COMPARATIVE REVIEW OF DETERMINISTIC APPROACHES

Gintautas DZEMYDA

Institute of Mathematics and Informatics
Akademijos 4, 2600 Vilnius, Lithuania
E-mail: dzemyda@ktl.mii.lt

Abstract. The problem is to discover knowledge in the correlation matrix of parameters (variables) about their groups. Results that deal with deterministic approaches of parameter clustering on the basis of their correlation matrix are reviewed and extended. The conclusions on both theoretical and experimental investigations of various deterministic strategies in solving the problem of extremal parameter grouping are presented. The possibility of finding the optimal number of clusters is considered. The transformation of a general clustering problem into the clustering on the sphere and the relation between clustering of parameters on the basis of their correlation matrix and clustering of vectors (objects, cases) of an n-dimensional unit sphere are analysed.

Key words: parameter clustering, data analysis, knowledge discovery.

1. Introduction. Any set of similar objects (cases, vectors) may be often characterized by common parameters (variables). The term "object" may cover, e.g., people, equipment, or produce of manufacturing. Any parameter may take some values. A combination of values of all parameters characterizes a concrete object from the whole set. The values obtained by any parameter depend on the values of other parameters, i.e., the parameters are correlated. There exist groups (clusters) of parameters characterizing different properties of the object. The correlation matrix of parameters may be calculated during the analysis of objects composing the set. The problem is to discover knowledge in the correlation matrix of the parameters about their groups. The methods investigated here are oriented to the analysis of correlation matrices and, in particular, to the clustering of parameters on the basis of correlations.

Examples of real correlation matrices:

1. The matrix of 8 physical parameters measured on 305 schoolgirls [1, 2].

2. The matrix of 11 parameters characterizing the development of agriculture in two Canadian provinces [5].

3. The matrix of 33 parameters of a tractor driver [13].

4. The matrix of 24 psychological tests on 145 pupils of the 7th and 8th forms in Chicago [1].

5. The matrix of 11 frequencies influencing human mentality [13, 26].

6. The matrix of 10 geological parameters [36].

The problem is to partition the parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting and homogeneous, in a certain sense, groups (clusters) $A_1, \ldots, A_p$ by the correlation matrix $R = \{r_{x_i x_j},\ i, j = \overline{1, n}\}$ characterizing the connections among the parameters ($r_{x_i x_j}$ is a correlation coefficient of parameters x_i and x_j). A covariance matrix may be used instead of the matrix R. However, the parameters with a greater variance will be more significant in the analysis. There is no *a priori* information regarding the number and size of groups.

Algorithms of parameter clustering are widely used to analyze the real data. There are two possibilities of such an analysis. The first one is to analyze the data matrix $\widetilde{X} = \{x_{ij},\ i = \overline{1, m},\ j = \overline{1, n}\}$, where m is the number of objects. The rows of $\widetilde{X}$ correspond to objects, and the columns of $\widetilde{X}$ correspond to parameters. The algorithms for clustering of objects are suitable here, because n columns of the matrix $\widetilde{X}$ may be clustered instead of m rows (see [4, 5]). Algorithms of such type for parameter clustering are included, for example, into the sublibrary of Fortran subroutines CLUSTER, Fortran subprogram library IMSL [43], and the integrated statistical data analysis, graphics, and data base management systems STATISTICA [41] and SYSTAT [39, 40]. The subroutines in CLUSTER are described in [4] and are part of the NIST Core Mathematical Library CMLIB [38]. Programs of the NAG Fortran Library [42] may also be used for such clustering. However, sometimes m may be very large or only the correlation matrix of parameters is known. In this case, the analysis of a set of $(n-1)\cdot n/2$ elements of the correlation matrix (or $(n+1)\cdot n/2$ elements of the covariance matrix) may be made instead of $m\cdot n$ elements of the data matrix $\widetilde{X}$, i.e., compressed information may be used. The modification of Harman's algorithm [1] based on the analysis of correlations is included in SAS [2, 37].

The algorithms of such type are also presented, for example, in [5, 13].

We deal here with algorithms of extremal parameter grouping [5–12, 31] based on the analysis of correlations and maximizing the partitioning quality

$$I_1 = \sum_{L=1}^{p} \sum_{x_i \in A_L} r^2_{x_i F_L}, \tag{1}$$

where F_L is a factor with a unit variance, corresponding to the group A_L; $r_{x_i F_L}$ is a correlation coefficient of the parameter x_i and the factor F_L. The factors F_L, $L = \overline{1,p}$, are selected so that to maximize the sums

$$\sum_{x_i \in A_L} r^2_{x_i F_L}, \quad L = \overline{1,p}. \tag{2}$$

We review and extend here the results referring to the deterministic approaches of parameter clustering published in [6, 8, 9, 10, 31]. Probabilistic approaches are also used to solve the problem above. The results are presented in [11, 12, 31].

2. A theoretical basis for extremal parameter grouping. In [5] it is proved that

$$F_L = \frac{\sum_{x_i \in A_L} a_i^L x_i}{\sqrt{\sum_{x_i, x_j \in A_L} a_i^L a_j^L r_{x_i x_j}}}, \tag{3}$$

$$r_{x_s F_L} = \frac{\sum_{x_i \in A_L} a_i^L r_{x_i x_s}}{\sqrt{\sum_{x_i, x_j \in A_L} a_i^L a_j^L r_{x_i x_j}}}, \tag{4}$$

where a_i^L are components of the eigen-vector of the matrix $R_L = \{r_{x_i x_j}, x_i, x_j \in A_L\}$, corresponding to the maximal eigen-value λ_L of R_L.

PROPOSITION 1 [8]. *The following formulae are equivalent to* (3) *and* (4):

$$F_L = \sum_{x_i \in A_L} \alpha_i^L x_i / \sqrt{\lambda_L}, \tag{5}$$

$$r_{x_s F_L} = \sum_{x_i \in A_L} \alpha_i^L r_{x_i x_s} / \sqrt{\lambda_L},$$

where λ_L *is the maximal eigen-value of the matrix* $R_L = \{r_{x_i x_j},\ x_i, x_j \in A_L\}$, α_i^L *are components of the normalized eigen-vector (i.e., a vector of unit length) of* R_L, *corresponding to* λ_L.

PROPOSITION 2 [8]. $r_{x_s F_L} = \sqrt{\lambda_L}\alpha_s^L$ *as* $x_s \in A_L$.

PROPOSITION 3 [8]. *The following formula is equivalent to* (1):

$$I_1 = \sum_{L=1}^{p} \lambda_L.$$

From (3) and (5) we observe that the factors F_L, $L = \overline{1,p}$, are linear combinations of parameters from the corresponding groups A_L, $L = \overline{1,p}$. The coefficients of the linear combinations are selected so that to maximize the sums (2). The values of coefficients of the linear combination are proportional to the elements of the eigen-vector corresponding to the maximal eigen-value λ_L of R_L, and $\lambda_L = \sum\limits_{x_i \in A_L} r^2_{x_i F_L}$. In this manner the ideas of factor analysis [1, 36] are applied to the formulation of the problem of parameter clustering.

Definition 1 [5, 7]. By the local maximum of the functional I_1 we call its value, corresponding to such a partition, where the squared correlation coefficient of any parameter with the factor, corresponding to the group including this parameter, is greater than or equal to that of the parameter with other factors, i.e., where for any parameter x_s (let $x_s \in A_k$) the following inequality holds:

$$r^2_{x_s F_k} \geqslant r^2_{x_s F_L}, \quad L = \overline{1,p},\ L \neq k.$$

Such a definition is useful in creating a strategy for maximizing I_1.

REMARK 1 [5]. Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given and factors $F_1, \ldots, F_p$ be fixed. Consider a parameter x_s (let $x_s \in A_k$). If we succeed in finding a factor F_L, $L \neq k$, such that $r^2_{x_s F_k} < r^2_{x_s F_L}$, then the transfer of x_s from the group A_k into the group A_L will increase the value of I_1.

Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given and a parameter x_s be in the group A_k. Let, in this case:

- λ_L and λ_k be the maximal eigen-values of matrices $R_L = \{r_{x_i x_j}, x_i, x_j \in A_L\}$ and $R_k = \{r_{x_i x_j}, x_i, x_j \in A_k\}$, respectively;
- λ_k^{-s} be the maximal eigen-value of the matrix $R_k^{-s} = \{r_{x_i x_j}, x_i, x_j \in A_k^{-s}\}$, where the group A_k^{-s} is obtained from the group A_k by eliminating the parameter x_s;
- λ_L^{+s} be the maximal eigen-value of the matrix $R_L^{+s} = \{r_{x_i x_j}, x_i, x_j \in A_L^{+s}\}$, where the group A_L^{+s} is obtained from the group A_L $(L \neq k)$ by adding the parameter x_s.

REMARK 2 [8, 9]. Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given and a parameter x_s be in the group A_k. If we succeed in finding λ_L^{+s} $(L \neq k)$ such that $\lambda_k - \lambda_k^{-s} < \lambda_L^{+s} - \lambda_L$, then the transfer of x_s from the group A_k into the group A_L will increase the value of I_1.

REMARK 3 [8, 9]. Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given. If for any parameter x_s (let $x_s \in A_k$) the following inequality holds: $\lambda_k - \lambda_k^{-s} \geqslant \lambda_L^{+s} - \lambda_L$, $L = \overline{1,p}$, $L \neq k$, then the partition of parameters corresponds to the local maximum of I_1.

Remark 3 defines the partition when the transfer of any parameter from its group to another one does not increase the value of I_1. Remarks 2 and 3 make a basis for algorithms maximizing the functional I_1 without computing separate factors $F_1, \ldots, F_p$ and correlation coefficients between separate parameters and factors. Maximal eigen-values of the symmetric matrices should be computed only in this case.

PROPOSITION 4 [8]. *Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given and factors $F_1, \ldots, F_p$ be fixed. Consider a parameter x_s (let $x_s \in A_k$). If one can find two factors F_{L_1} and F_{L_2} $(L_1 \neq L_2 \neq k)$ such that*

$$r^2_{x_s F_k} < r^2_{x_s F_{L_2}} < r^2_{x_s F_{L_1}},$$

then three opportunities are possible:

$$I_1^{L_1} > I_1^{L_2}, \quad I_1^{L_1} = I_1^{L_2}, \quad I_1^{L_1} < I_1^{L_2},$$

where $I_1^{L_1}$ and $I_1^{L_2}$ are values of the functional I_1 obtained after transferring the parameter x_s from the group A_k into the group A_{L_1} or A_{L_2}, respectively.

PROPOSITION 5 [8]. *Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given and factors $F_1, \ldots, F_p$ be fixed. Consider a parameter x_s (let $x_s \in A_k$). Let there be a factor F_L ($L \neq k$) such that $r^2_{x_s F_k} = r^2_{x_s F_L}$. After transferring the parameter x_s from the group A_k into the group A_L, the value of functional I_1 may be increased.*

PROPOSITION 6 [8]. *Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given and factors $F_1, \ldots, F_p$ be fixed. Consider a parameter x_s (let $x_s \in A_k$). Let there be a factor F_L ($L \neq k$) such that $r^2_{x_s F_k} > r^2_{x_s F_L}$. After transferring the parameter x_s from the group A_k into the group A_L, the value of functional I_1 may be increased.*

REMARK 4 [8]. Let a partition of parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting groups $A_1, \ldots, A_p$ be given and factors $F_1, \ldots, F_p$ be fixed. Consider a parameter x_s (let $x_s \in A_k$). It follows from Propositions 4–6 that using the values of $r_{x_s F_L}$, $L = \overline{1, p}$, it is impossible to determine the group where the value of I_1 increases most after transferring the parameter x_s.

3. Deterministic algorithms for maximizing I_1. Deterministic algorithms [5–10] often find only the local maximum of I_1 which is not global. All the algorithms start from some initial partition selected by some algorithm or by a certain knowledge about the problem. They are based on the analysis of parameters in consecutive order and on the search for a group of transferring a separate parameter with a view to increase the I_1 value. They use different strategies to determine when the parameter must be transferred from its group to another. The algorithms stop when the transfer of any parameter by the chosen strategy does not increase the value of I_1.

Let us introduce two functions used in a formal description of algorithms: $\underline{\arg}\max_q \omega(q)$ and $\underline{\arg}\min_q \omega(q)$ return the least value of q which maximizes and minimizes the function $\omega(q)$, respectively.

When algorithm A1 [8] considers the parameter x_s (let $x_s \in A_k$), it seeks

the greatest squared correlation coefficient $r^2_{x_s F_L}$ among $r^2_{x_s F_j}$, $j = \overline{1,p}$, $j \neq k$. If $r^2_{x_s F_L} > r^2_{x_s F_k}$, then A1 transfers x_s into the group A_L and recalculates the factors F_k and F_L.

When algorithm A2 [8] considers the parameter x_s (let $x_s \in A_k$), it seeks the group A_L, $L \neq k$, such that the value of functional I_1 increases most after transferring x_s, i.e., A2 looks for L maximizing $E = \lambda^{+s}_L + \lambda^{-s}_k - \lambda_L - \lambda_k$. If $E > 0$, then A2 transfers x_s into the group A_L and recalculates λ_k and λ_L.

REMARK 5 [8]. Algorithm A2 should yield a greater value of I_1 as compared with A1. Algorithm A1 makes a decision where to transfer the parameter x_s on the basis of correlations among this parameter and factors $F_1, \ldots, F_p$. Remark 4 implies that, in this manner, it is impossible to determine the group where the value of I_1 increases most after transferring the parameter x_s. Algorithm A2 allows us to find such a group.

Algorithm A2 is simpler than A1, because it does not calculate separate factors and the correlation between these factors and parameters. However, A2 contains the calculations of maximal eigen-values of symmetric matrices. Such calculations are very computation-intensive and make algorithm A2 significantly slower than A1.

Taking into account the result of Remark 4, one can construct an algorithm that is faster than A1, but yields similar results (values of I_1). Let us denote this algorithm by A3. When algorithm A3 [8] considers the parameter x_s (let $x_s \in A_k$), it transfers x_s into the first found group A_L such that $r^2_{x_s F_L} > r^2_{x_s F_k}$, and recalculates the factors F_k and F_L.

The algorithm proposed by Braverman and Muchnik in [5] (denote it by BM) considers all the n parameters $x_1, \ldots, x_n$ at each step and for each parameter x_s (let $x_s \in A_k$) it determines the index $L_s = \arg \max_{q=\overline{1,p}} r^2_{x_s F_q}$. Afterwards, the algorithm transfers x_s into the group A_{L_s}, $s = \overline{1,p}$, if $L_s \neq k$. Sometimes some indices may be found for the parameter x_s one of which is the same as k. There is no transfer in this case.

Various combinations of the strategies above are possible, too. They allow us to economize the computing time (as compared with A2) and to find greater values of I_1 (as compared with A1 and A3).

Algorithm B1 [8]:

1. Calculations by A1 or A3.

2. Calculations by A2.

Algorithm B2 [8]:

1. Calculations by A1 or A3.

2. One iteration of A2 (one run-through all the n parameters in consecutive order).

3. If the value of I_1 has not been changed in Step 2, terminate the partitioning process.

4. Calculations by A1 or A3.

5. If the value of I_1 has been changed in Step 4, go to Step 2; otherwise, terminate the partitioning process.

Algorithms A1, A2, A3, B1, B2, and BM need the initial partition of parameters. Any arbitrarily selected partition may be used as the initial one. Therefore, to economize calculations of partitioning algorithms, it is convenient to start from the partition obtained by some fast algorithm.

Two algorithms H1 and H2 are proposed in [8, 25] and presented below for the initial partitioning. They are grounded on finding the basic parameters (one parameter for each group), and filling up the groups by the remaining parameters. The basic parameters $x_{k_1}, \ldots, x_{k_m}$ satisfy the following system of inequalities:

$$\left|r_{x_{k_i}x_{k_j}}\right| < \min(d_{k_i}, d_{k_j}), \quad i, j = \overline{1, m},\ i \neq j,$$

where d_i is a mean correlation of the parameter x_i with other parameters [24]:

$$d_i = \frac{1}{n-1} \sum_{\substack{s=1 \\ s \neq i}}^{n} \left|r_{x_i x_s}\right| .$$

Some sets of the basic parameters are possible for the same correlation matrix. If $n > 2$ and absolute values of all the elements of the correlation matrix are not equal to 1, then the number of the basic parameters satisfies $2 \leqslant m < n$.

The number of groups can be fixed in advance or detected automatically. In the case of automatic detection of the number of groups, the search for a set of the basic parameters may be performed. The number of parameters in this set may be used as the optimal number of groups.

Algorithm H1 (partitioning into a fixed number p of groups) [25]:

1. $k_0 = 2$,
   $x_s \in A_1,\ x_k \in A_2,\ (s,k) = \underline{\arg} \min\limits_{\substack{1 \leqslant i,j \leqslant n \\ i \neq j}} \left| r_{x_i x_j} \right|$.
2. If $p = k_0$, go to Step 6.
3. $x_s \in A_{k_0+1}, \quad s = \underline{\arg} \min\limits_{i} \sum\limits_{j=1}^{k_0} \left| r_{x_i A_j} \right|$, (6)

where $r_{x_i A_j}$ is a correlation of the parameter x_i with a parameter from the group A_j, i in (6) are numbers of parameters that do not belong to the groups $A_1, \ldots, A_{k_0}$.

4. $k_0 = k_0 + 1$.
5. If $k_0 \neq p$, go to Step 3.
6. Insert parameters $x_s \notin A_1, \ldots, A_p$ into the groups according to the rule:

$$x_s \in A_L, \quad L = \underline{\arg} \max_{1 \leqslant q \leqslant p} r^2_{x_s F_q},$$

where the factor F_q is calculated on the basis of correlation matrix of the parameters from the group A_q (note that $x_s \notin A_q$).

To partition parameters into the optimal number of groups

- in Step 3, it is necessary to use i such that

$$\left| r_{x_i A_j} \right| < \min(d_i, d_{A_j}), \tag{7}$$

where d_{A_j} is a mean correlation of the parameter of group A_j with other parameters;

- Step 2 looks like this:
  $p = k_0$, go to Step 6, if in Step 3 we fail to find i satisfying (7).

Algorithm H2 differs from H1 only in the content of Step 6: insert parameters $x_s \notin A_1, \ldots, A_p$ into the groups according to the rule:

$$x_s \in A_L, \quad L = \underline{\arg} \max_{1 \leqslant q \leqslant p} \lambda_q^{+s},$$

where λ_q^{+s} is the maximal eigen-value of the correlation matrix of parameters from the group A_q^{+s} (note that $x_s \notin A_q$ during calculation of L).

The algorithm for initial clustering of parameters is also proposed in [13]. Let us denote it by C1.

The assumptions on the efficiency of H1 and H2 are analogous to those on A1 and A2 made in Remark 5. The efficiency of C1 was compared with that of A1 and A2 experimentally in further sections.

It is possible to combine a lot of algorithms for extremal parameter grouping from the algorithms above. The following combinations were investigated:

AC1 = C1+A1	AH1 = H1+A1
AC2 = C1+A2	AH2 = H1+A2
AC3 = C1+A3	AH3 = H1+A3
AC4 = C1+B1	AH4 = H1+B1
AC5 = C1+B2	AH5 = H1+B2
AC6 = C1+BM	AH6 = H2+A1
AC7 = H1+BM	AH7 = H2+A2
AC8 = H2+BM	AH8 = H2+A3
	AH9 = H2+B1
	AH10 = H2+B2

3.1. Experimental investigation of deterministic algorithms for maximizing I_1

3.1.1. Efficiency of the algorithms of initial partitioning. The first experiment was carried out on the correlation matrix of 11 frequencies influencing human mentality [13, 26]. The "ideal" partition is known:

1. $x_1, x_2, x_3 \in A_1$ are frequencies from the so-called θ-rhythm.
2. $x_4, x_5 \in A_2$ are low frequencies from the so-called α-rhythm.
3. $x_6, x_7 \in A_3$ are high frequencies from the so-called α-rhythm.
4. $x_8, x_9, x_{10}, x_{11} \in A_4$ are frequencies from the so-called β-rhythm.

The second experiment was carried out using the correlation matrix of 8 physical parameters measured on 305 schoolgirls [1, 2]: height, arm span, length of forearm, length of lower leg, weight, bitrochanteric diameter, chest girth, chest width. Investigations of these classical test data divided the parameters into two groups: $A_1 = \{x_1, \ldots, x_4\}$ and $A_2 = \{x_5, \ldots, x_8\}$: the parameters of the first group characterize shapeliness, while the parameters of the second group characterize plumpness of girls. This is an "ideal" partition.

Both the "ideal" partitions were obtained by algorithms H1 and H2 of initial partitioning. It means that these two data sets have a good degree of structure and any fast algorithm may be successfully used for the analysis of such data.

An automatic detection of the number of groups indicated the right number. This shows a good sensitivity of the algorithm for selecting the number of groups. Partitions by algorithm C1 were far from "ideal".

The third experiment was performed on the correlation matrix of 33 parameters of a tractor driver [13]. These parameters indicate various aspects of driver's working-capacity and health-state and may be grouped into six groups according their physical sense:

1. Temperature $(x_1, \ldots, x_7)$.
2. Blood pressure $(x_8, \ldots, x_{11})$.
3. Sensitivity of ear to various frequencies of sound $(x_{12}, \ldots, x_{25})$.
4. Pulse (x_{26}).
5. Reaction (x_{27}).
6. Strength $(x_{28}, \ldots, x_{33})$.

The authors in [13] suggest dividing the parameters of the last group into three subgroups: $A_6^1 = \{x_{28}, x_{29}\}$, $A_6^2 = \{x_{30}, x_{31}\}$, $A_6^3 = \{x_{32}, x_{33}\}$ (see Table 2).

An automatic detection of groups by algorithms H1 and H2 indicated that there were six groups. In the case of 7 groups, initial partitioning by H1 and H2 found the local maximum of I_1 equal to 20.44, which has not been changed using A1, A2, A3, B1, and B2. Such a number of groups was selected for comparison with the algorithm proposed in [13] which is similar to A1, maximizes I_1, is realized by the authors in [13] making some simplifications in the calculation of maximal eigen-values, and uses the initial partition obtained by C1. This algorithm found $I_1 = 18.52$ (the result is taken from [13]) only.

3.1.2. Efficiency of the deterministic algorithms maximizing I_1. Random correlation matrices (150) were generated: 50 matrices, the number of parameters being $n = 10$, 20 and 30. These matrices were analysed using the algorithms above. The number of groups was fixed beforehand: $p = 3$ as $n = 10$; $p = 4$ as $n = 20$; $p = 5$ as $n = 30$. The results of performance of the algorithms were averaged. The mean obtained value of I_1 and mean relative expenses t of computer time (as compared with the fastest algorithm, i.e., C1, as $n = 10$) for solving one problem are presented in Table 1 dependent on n (see also [8]). The maximal eigen-values of correlation matrices were calculated using the iterational gradual method [23].

The second experiment was performed on the correlation matrix of 33 pa-

Table 1. Comparison of algorithms

Algorithm	$n = 10$		$n = 20$		$n = 30$	
	t	I_1	t	I_1	t	I_1
C1	1.00	4.29	2.67	6.87	4.33	9.44
H1	1.50	4.51	5.00	7.53	10.00	10.60
H2	4.33	4.55	16.00	7.68	33.17	10.85
AC1	1.50	4.34	5.67	7.30	12.33	10.28
AC2	24.50	4.58	105.00	7.80	226.17	11.08
AC3	1.50	4.34	6.00	7.26	12.17	10.22
AC4	23.83	4.58	97.83	7.80	224.67	11.07
AC5	11.83	4.56	40.33	7.68	81.17	10.89
AC6	1.50	4.34	6.17	7.27	14.17	10.23
AC7	1.67	4.51	5.83	7.53	11.67	10.63
AC8	5.00	4.55	19.17	7.68	36.50	10.86
AH1	1.67	4.51	5.67	7.53	11.33	10.63
AH2	18.50	4.56	77.50	7.78	185.00	11.06
AH3	1.67	4.51	5.50	7.53	11.17	10.63
AH4	17.67	4.56	79.00	7.78	186.67	11.06
AH5	10.83	4.56	35.50	7.71	77.33	10.94
AH6	5.00	4.55	18.17	7.68	36.00	10.86
AH7	19.17	4.58	81.83	7.79	196.67	11.08
AH8	4.83	4.55	17.67	7.68	36.00	10.86
AH9	20.00	4.58	83.33	7.79	189.33	11.08
AH10	14.83	4.58	47.67	7.77	98.50	11.02

rameters of a tractor driver [13]. The following algorithms were applied to analyse the data:

1. The method of correlation sequences [13] (denote it by M1).
2. The first heuristic algorithm in [13] (denote it by M2).
3. The second heuristic algorithm in [13] (denote it by M3).
4. The algorithm in [5, 13] (denote it by M4) which maximizes

$$I_2 = \sum_{L=1}^{p} \sum_{x_i \in A_L} \left| r_{x_i F_L^+} \right|, \quad F_L^+ = \frac{\sum\limits_{x_i \in A_L} \delta_i^L x_i}{\sqrt{\sum\limits_{x_i, x_j \in A_L} \delta_i^L \delta_j^L r_{x_i x_j}}},$$

Table 2. Clustering results

Number of group	"Ideal"	"Ideal" [13]	M1	M2	M3	M4	M5	AH2 AH6 AH7
1	1–7	1–7	1–7	1–7	1–7	1–7	1–4,6	1–7,27
2	8–11	8–11	8,9	8,9	8,9	8–11	5,7, 10,11	8–11
3	12–25	12–25	10,11	10,11	10,11	12–17, 24,25	8,9, 26,27	12–26
4	26	26	12–21, 24	12–21, 24,25	12–21, 24,25	18–23	12–17, 24,25	28–33
5	27	27	22	22,23	22,23	28,29	18–23	
6	28–33	28,29	23	26	26,27	27, 30,31	28–29	
7		30,31	25	27	28,29	26, 32,33	30–33	
8		32,33	26	28,29	30,31			
9			27	30,31	32,33			
10			28,29	32,33				
11			30,31					
12			32,33					

where coefficients δ_i^L are equal to -1 or $+1$ and maximize $\sum_{x_i,x_j \in A_L} \delta_i^L \delta_j^L r_{x_i x_j}$.

5. The algorithm in [5, 13] (denote it by M5) which maximizes I_1 and is similar to BM.

6. All the possible combinations of H1, H2, and C1 with A1 and A2, i.e., algorithms AC1, AC2, AH1, AH2, AH6, and AH7.

It was set in [5, 13] that the functional I_1 was better to evaluate the partitioning quality in comparison with I_2.

The number of groups was selected so that the partition obtained be as similar to that discussed in Section 3.1.1 as possible. The results are given in Table 2. Algorithms that involved A1 or A2 yielded the best results in the case of partitioning into four groups. The results of AH2, AH6, and AH7 are presented in Table 2, because they found greater values of I_1 than AC1, AC2, and AH1. The results of M1–M5 are taken from [13, 34].

The difference of the results of algorithms AH2, AH6, and AH7 from the "ideal" partition is that they added

- parameter "Pulse" (x_{26}) to the group characterizing sensitivity of ear to various frequencies of sound;
- parameter "Reaction" (x_{27}) to the group characterizing temperature.

Such adding is meaningful because the pulse strongly depends on the intensity of noise, and the reaction depends on the health state a characteristic of which is temperature.

The experiments confirmed the theoretical remarks on the efficiency of algorithms maximizing I_1. The following additional conclusions may be drawn:

1. One can find greater values of I_1 (and better partitions) by using algorithms H1 and H2 in comparison with C1.

2. Algorithm A1 is better than BM.

3. If the value of I_1, corresponding to the initial partition, considerably differs from the global maximum of I_1, then it is reasonable to use algorithm A2 or other algorithms (e.g., B1 or B2) containing the elements of A2 for further maximization.

4. The values of I_1 obtained by A2 and B2 have the least dependence on the initial partition of parameters.

5. Algorithm H2 is more computation-intensive as compared with H1. Therefore, it may be used in the cases where further optimization is performed by computation-intensive algorithms, too (e.g., A2, B1, B2). In such cases, the expenses of computing time for initial partitioning will be rather low as compared with the total ones.

6. When selecting an algorithm among AH1-AH10 for solving a concrete problem, it is necessary to take into account two factors: optimization accuracy and recourses of computing time allocated for clustering. During the experiments above, greater values of I_1 are obtained by the algorithms which are more computation-intensive. This means that the fast algorithms are oriented to the problems of simple structure and the cases where a high partitioning quality is not required. Extremal situations are as follows:

- the fastest algorithm is AH3;
- the slowest algorithm is AH7;
- the greatest values of I_1 were obtained by AH7 and AH9;
- the worst results of optimization were obtained by AH1 and AH3.

4. Clustering on the sphere

4.1. Formulation of the problem. Let there be given m vectors (objects, cases) $X_1 = (x_{11}, \ldots, x_{1n}), \ldots, X_m = (x_{m1}, \ldots, x_{mn}) \in S^n \subset R^n$, where S^n is a subset of an n-dimensional Euclidean space R^n containing vectors of unit length, i.e., S^n is a unit sphere [14], $\|X_s\| = 1$, $s = \overline{1,m}$. The problem is to partition these vectors into a fixed number p of non-intersecting and homogeneous, in a certain sense, groups (clusters) $A_1, \ldots, A_p$, defined on the basis of data, but not *a priori*. This problem is a special case of the classical clustering problem (e.g., see [3], [27], [33], [35]).

The homogeneity of vectors $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n) \in S^n$ may be characterized by various numerical values (see, e.g., [33]). We consider here two values:

- Euclidean distance between the vectors X and Y, i.e., length of the vector $X - Y$:

$$\rho(X,Y) = \sqrt{\sum_{j=1}^{n}(x_j - y_j)^2}, \tag{8}$$

- cosine of the angle between the vectors X and Y [14] (it is also their scalar product because X and $Y \in S^n$):

$$\cos(X,Y) = \sum_{j=1}^{n} x_j y_j. \tag{9}$$

Formulae (8) and (9) are uniquely related: $\rho^2(X,Y) = 2(1 - \cos(X,Y))$.

4.2. Functionals characterizing the partitioning quality. The formulation of the problem of cluster analysis often contains a conception of the functional (criterion) characterizing the partitioning quality. This functional is a quantitative measure allowing to compare different partitions. It is necessary to minimize (or maximize) this criterion during the search for the optimal partition of vectors into groups.

One of the most widely used criteria is the sum of interior variances of clusters [33]:

$$\mathcal{Q}_1 = \sum_{L=1}^{p} \sum_{X_i \in A_L} \rho^2(X_i, F_L),$$

where $F_L = (f_1^L, \ldots, f_n^L)$ is the weight centre of the L-th cluster:

$$f_j^L = \frac{1}{n_L} \sum_{X_i \in A_L} x_{ij}, \quad j = \overline{1,n},$$

n_L is the number of vectors in the L-th cluster A_L ($n_L \geqslant 1$). It is necessary to minimize Q_1.

REMARK 6 [6]. The vector $F_L = (f_1^L, \ldots, f_n^L)$ may be defined as follows:

$$F_L = \arg \min_{Y=(y_1,\ldots,y_n)\in R^n} \Psi_L(Y),$$

where

$$\Psi_L(Y) = \sum_{X_i \in A_L} \rho^2(X_i, Y).$$

In further formulae of Sections 4.2, 4.2.1, and 4.2.2, we will assume that $X_s = (x_{s1}, \ldots, x_{sn}) \in S^n$, $s = \overline{1,m}$, in case a special remark is not given.

Denote $\Sigma_L = \sum_{X_k \in A_L} \sum_{X_t \in A_L} \cos(X_k, X_t)$.

REMARK 7 [6]. $F_L \notin S^n$, if $n_L \geqslant 2$ and there are at least two non-coincidental vectors among $X_s \in A_L$.

In the general case, $\|F_L\| \leqslant 1$.

Taking into account that all the vectors to be clustered are located on the unit sphere ($X_s \in S^n$, $s = \overline{1,m}$), the next functional that characterizes the partitioning quality may be considered:

$$Q_2 = \sum_{L=1}^{p} \sum_{X_i \in A_L} \rho^2(X_i, F_L^*),$$

where

$$F_L^* = (f_1^{*L}, \ldots, f_n^{*L}) = \arg \min_{Y=(y_1,\ldots,y_n)\in S^n} \Psi_L(Y).$$

Two following propositions show some properties of Σ_L and $\Psi_L(\cdot)$.

PROPOSITION 7 [6]. $\Sigma_L = 0$ *and* $\|F_L\| = 0$, *if and only if*

$$\sum_{X_i \in A_L} x_{ij} = 0, \quad j = \overline{1,n}.$$

In the general case, $\Sigma_L \geqslant 0$. Let us analyse a symmetric quadratic matrix $G_L = \{\cos(X_i, X_j),\ X_i, X_j \in A_L\}$, whose dimensions are $(n_L \times n_L)$. The matrix is positive definite, if all the vectors in A_L are linearly independent, and it is positive semi-definite, if there are some linearly dependent vectors in A_L. The determinant of G_L is called Gram's determinant (see [14]). It is always non-negative. It means that the matrix G_L has no negative eigen-values.

Σ_L may be expressed like this: $\Sigma_L = Z_L G_L Z_L'$, where $Z_L = (1, \ldots, 1)$ is a vector containing n_L elements whose values are equal to 1. Z_L' is a column vector corresponding to the row vector Z_L. It follows from the formulae above that $\Sigma_L = 0$ only if G_L is a positive semi-definite matrix and the vector $Z_L = (1, \ldots, 1)$ is its eigen-vector corresponding to its least eigen-value. In this case, such an eigen-value is equal to 0.

Σ_L is strongly positive (i.e., $\Sigma_L > 0$), if $\cos(X_i, X_j) \geqslant 0,\ X_i, X_j \in A_L$.

PROPOSITION 8 [6]. *$\Psi_L(Y) = 2n_L$ for any $Y \in S^n$, if $\Sigma_L = 0$.*

PROPOSITION 9. $\sum\limits_{X_i \in A_L} \cos(X_i, Y) = 0$ *for any $Y \in S^n$, if $\Sigma_L = 0$.*

PROPOSITION 10 [6]. $f_j^{*L} = \dfrac{1}{\sqrt{\Sigma_L}} \sum\limits_{X_i \in A_L} x_{ij},\ j = \overline{1, n}$, *if $\Sigma_L > 0$.*

In case $\Sigma_L = 0$, the vector F_L^* is not defined but the values of functional Q_2 may be uniquely evaluated taking into account the results of Proposition 8: the L-th part of Q_2 will be such: $\sum\limits_{X_i \in A_L} \rho^2(X_i, F_L^*) = 2n_L$.

4.2.1. Equivalent functionals. By equivalent functionals we call functionals such that have the same partition of vectors corresponding to the global extremum of these functionals. Two groups of equivalent functionals are given in Propositions 11 and 12. '*min*' or '*max*' in front of the functional mean that it is necessary to minimize or maximize this functional during the search for the optimal partition of vectors, respectively.

PROPOSITION 11 [6]. *The following functionals are equivalent to the functional Q_1:*

$$\max Q_1^1 = \sum_{L=1}^{p} \frac{1}{n_L} \Sigma_L,$$

$$\min Q_1^2 = \sum_{L=1}^{p} \frac{1}{n_L} \sum_{X_k \in A_L} \sum_{X_t \in A_L} \rho^2(X_k, X_t),$$

$$\max Q_1^3 = \sum_{L=1}^{p} n_L \|F_L\|^2.$$

PROPOSITION 12 [6]. *The following functionals are equivalent to the functional* Q_2*:*

$$\max Q_2^1 = \sum_{L=1}^{p} \sqrt{\Sigma_L},$$

$$\max Q_2^2 = \sum_{L=1}^{p} \sum_{X_i \in A_L} \cos(X_i, F_L^*),$$

$$\max Q_2^3 = \sum_{L=1}^{p} n_L \|F_L\|.$$

We get from Propositions 11 and 12 two very simple functionals Q_1^1 and Q_2^1 which are equivalent to Q_1 and Q_2, respectively, but they are simpler than Q_1 and Q_2. New and more effective optimization strategies may be developed to optimize the equivalent functionals. These strategies should take account of the specific character of these functionals.

4.2.2. Case $\cos(X_i,X_j) \geqslant 0$, $i, j = \overline{1, m}$. The following functional, characterizing the quality of partition of the vectors $X_s = (x_{s1}, \ldots, x_{sn}) \in S^n$, $s = \overline{1, m}$, may be used in the case $\cos(X_i, X_j) \geqslant 0$, $i, j = \overline{1, m}$:

$$\max I_1^1 = \sum_{L=1}^{p} \sum_{X_i \in A_L} \cos^2(X_i, \overline{F}_L),$$

where the vector $\overline{F}_L = (\overline{f}_1^L, \ldots \overline{f}_n^L)$ is selected from S^n so that it maximizes

$$\sum_{X_i \in A_L} \cos^2(X_i, \overline{F}_L).$$

This functional has some similarities to that used in extremal parameter grouping (see, e.g., [5, 7]).

PROPOSITION 13 [6]. *The vector* $\overline{F}_L = (\overline{f}_1^L, \ldots, \overline{f}_n^L)$ *is a normalized eigen-vector corresponding to the maximal eigen-value* λ_L *of the matrix*

$$B_L = \left\{ b_{ij}^L = \sum_{X_k \in A_L} x_{ki} x_{kj},\ i, j = \overline{1, n} \right\}.$$

REMARK 8 [6]. The following functional is equivalent to the functional I_1^1:

$$\max I_1^2 = \sum_{L=1}^{p} \lambda_L.$$

It follows from Proposition 13 that the maximization of I_1^1 or I_1^2 requires to seek eigen-values and eigen-vectors of the matrices B_L, $L = \overline{1, p}$. Sometimes the investigator does not know the vectors $X_s \in S^n$, $s = \overline{1, m}$, but he knows only their interlocation defined by the matrix of cosines $G = \{\cos(X_i, X_j),\ i, j = \overline{1, m}\}$. In such a case, the interlocation of vectors from the L-th cluster is entirely defined by the matrix of cosines $G_L = \{\cos(X_i, X_j),\ X_i, X_j \in A_L\}$.

The next two propositions determine how to evaluate $\overline{F}_L = (\overline{f}_1^L, \ldots, \overline{f}_n^L)$, which is the normalized eigen-vector of the matrix B_L corresponding to its maximal eigen-value, on the basis of G_L.

PROPOSITION 14 [6]. *Matrices* B_L *of dimensions* $(n \times n)$ *and* G_L *of dimensions* $(n_L \times n_L)$ *have the same eigen-values. In addition, the matrix of greater dimensions has* $|n - n_L|$ *zero eigen-values.*

REMARK 9 [6]. Matrices B_L and G_L have the same maximal eigen-values.

PROPOSITION 15 [6]. *The vector* $\overline{F}_L = (\overline{f}_1^L, \ldots, \overline{f}_n^L)$ *may be transformed into such a form:*

$$\overline{f}_k^L = \frac{1}{\sqrt{\lambda_L}} \sum_{X_i \in A_L} x_{ik} \alpha_{Li}, \quad k = \overline{1, n},$$

where λ_L *is the maximal eigen-value of the matrix* G_L *(and* B_L*),* α_{Li} *are components of the normalized eigen-vector of the matrix* G_L *corresponding to the maximal eigen-value* λ_L *of* G_L.

The algorithms from Section 3 may be used to maximize the functionals I_1^1 and I_1^2. The data for these algorithms is the matrix of cosines between the pairs of vectors $X_s \in S^n$, $s = \overline{1, m}$.

The results of Propositions 13, 14, 15 and Remarks 8, 9 are valid for $X_s \in R^n$, $s = \overline{1,m}$, when scalar products of the vectors X_s, $s = \overline{1,m}$, and $\overline{F}_L$ are used instead of their cosines in the formulae of this section.

4.3. Transformation of the general clustering problem into the clustering on the sphere. There are some ways of transforming the initial system of vectors to be clustered $Y_s = (y_{s1}, \ldots, y_{sn}) \in R^n$, $s = \overline{1,m}$, into the system $X_s = (x_{s1}, \ldots, x_{st}) \in S^t$, $s = \overline{1,m}$.

The authors in [3] and [33] propose several ways of such transformation. In both the cases, $t = n$. In the first case, any vector Y_s is centred. Then the centred components are divided by the length of the centred vector. In the second case, the components of any vector Y_s are divided by the length of this vector. However, the gain of such transformations is doubtful.

Another way transforming the initial system of vectors is proposed in [6] and [10]. The authors in [6] and [10] propose a nonlinear transformation where an increase in the distance between two vectors of the initial system causes a nonlinear increase in the distance between the corresponding vectors in a new system. The gain of such transformation lies in this non-linearity. Let us describe such an approach more in detail.

The main idea of the transformation is to convert the initial system of vectors to be clustered $Y_s = (y_{s1}, \ldots, y_{sn}) \in R^n$, $s = \overline{1,m}$, into the system $X_s = (x_{s1}, \ldots, x_{sm}) \in S^m$, $s = \overline{1,m}$. It means that the dimension of a new space depends on the number of vectors to be clustered. Therefore, in the case of a large number of vectors, the dimension of space will also be large. But we suggest clustering the vectors $X_s \in S^m$, $s = \overline{1,m}$, on the basis of the matrix of their cosines. The strategy of clustering on the basis of the matrix of cosines of the vectors is investigated in the section above. The initial information contains $n \cdot m$ data units, and information after the transformation contains $(m-1) \cdot m/2$ data units (in fact, there are $m \cdot m$ data units, but m of them are equal to 1 and the remaining $(m-1) \cdot m$ data units make up two identical groups because of symmetry of the cosine matrix).

The aim is to find a function $K(Y_i, Y_j)$ that describes the similarity of vectors Y_i and Y_j, and bears the following properties:

1. $K(Y_i, Y_j) = K(Y_j, Y_i)$, $i, j = \overline{1,m}$.
2. $0 \leqslant K(Y_i, Y_j) \leqslant 1$, $i, j = \overline{1,m}$.
3. $K(Y_i, Y_i) = 1$, $i = \overline{1,m}$.

4. The matrix $K = \{K(Y_i, Y_j),\ i, j = \overline{1, m}\}$ is non-negative definite.

5. $K(Y_i, Y_j) < K(Y_k, Y_l)$, if the distance between Y_i and Y_j is greater than the distance between Y_k and Y_l; $K(Y_i, Y_j) = K(Y_k, Y_l)$, if the distances are equal.

Items 1–4 define a new system of vectors $X_s = (x_{s1}, \ldots, x_{sm}) \in S^m$, $s = \overline{1, m}$, which is related with the initial one as follows:

1. $\cos(X_i, X_j) = K(Y_i, Y_j)$.

2. $x_{sk} = \sqrt{\lambda_s} u_{sk},\ k = \overline{1, m},$ (10)

where λ_s is the s-th eigen-value of the matrix $K = \{K(Y_i, Y_j),\ i, j = \overline{1, m}\}$, the vector $(u_{s1}, \ldots, u_{sm})$ is a normalized eigen-vector corresponding to the eigen-value λ_s.

Equation (10) follows from the fact that any symmetric non-negative definite matrix K may be represented in such a form: $K = U\Lambda U'$, where Λ is a diagonal matrix containing the eigen-values of K; U is an orthogonal matrix, whose columns contain the normalized eigen-vectors of K, the s-th column contains the eigen-vector corresponding to the s-th eigen-value; U' is a transposition of the matrix U. Therefore,

$$K(Y_i, Y_j) = \sum_{s=1}^{m} \lambda_s u_{si} u_{sj} = \sum_{s=1}^{m} \left(\sqrt{\lambda_s} u_{si}\right) \left(\sqrt{\lambda_s} u_{sj}\right).$$

It follows from (10) that it is necessary to solve the problem of all the eigen-values and eigen-vectors of the matrix $K = \{K(Y_i, Y_j),\ i, j = \overline{1, m}\}$, if we want to determine the vectors $X_s = (x_{s1}, \ldots, x_{sm}) \in S^m$, $s = \overline{1, m}$. However, our approach is based on the clustering of the vectors $X_s \in S^m$, $s = \overline{1, m}$, on the basis of the matrix of their cosines, i.e., on the basis of the matrix K. In this case, we don't have to know the coordinates of the vectors X_s, $s = \overline{1, m}$.

The problem is to find an analytical form of the transformation $K(Y_i, Y_j)$. The idea comes from the correlation functions of random processes [16].

PROPOSITION 16 [10]. *Let there be given m vectors $Y_s = (y_{s1}, \ldots, y_{sn}) \in R^n$, $s = \overline{1, m}$. The matrix $K = \{K(Y_i, Y_j),\ i, j = \overline{1, m}\}$ is non-negative definite, if*

$$K(Y_i, Y_j) = e^{-\alpha \rho^2(Y_i, Y_j)}, \tag{11}$$

where $\alpha > 0$, $\rho(Y_i, Y_j)$ is a generalized Euclidean distance between the

vectors Y_i *and* Y_j [4]:

$$\rho(Y_i, Y_j) = \sqrt{\sum_{k,l=1}^{n} r^*_{kl}(y_{ik} - y_{jk})(y_{il} - y_{jl})}, \tag{12}$$

where $R^* = \{r^*_{kl},\ k, l = \overline{1,n}\}$ *is a non-negative definite symmetric matrix.*

Formula (12) is a generalization of distance functions wide used in the cluster analysis. For example, if R^* is a diagonal matrix, then $\rho(\cdot)$ is the weighted Euclidean distance [18]; if R^* is a diagonal matrix and $r^*_{ll} = 1,\ l = \overline{1,n}$, then $\rho(\cdot)$ is the Euclidean distance. Formula (12) envelops the Mahalanobis distance [18, 19, 33].

The function $K(Y_i, Y_j)$ defined by (11) is a monotone and positive function of the distance $\rho(Y_i, Y_j)$. A greater distance between the vectors Y_i and Y_j produces a lower value of the function $K(Y_i, Y_j)$. Therefore, the interlocation of the vectors $Y_s,\ s = \overline{1,m}$, is entirely defined by the matrix $K = \{K(Y_i, Y_j),\ i, j = \overline{1,m}\}$, which is, as shown above, a matrix of cosines of the vectors $X_s = (x_{s1}, \ldots, x_{sm}) \in S^m,\ s = \overline{1,m}$.

The form of $K(Y_i, Y_j)$ defined by (11) is not the only one possible. Using the criterion from [15] one can find more transformations $K(Y_i, Y_j)$.

4.3.1. Clustering algorithm. Any algorithm maximizing I_1 (see Section 3) may be used to analyse the matrix $K = \{K(Y_i, Y_j),\ i, j = \overline{1,m}\}$.

The well-known K-means clustering algorithm (see its description, e.g., in [2] or [4]) may also be transformed so that to cluster the vectors $X_s = (x_{s1}, \ldots, x_{sm}) \in S^m,\ s = \overline{1,m}$, according to the matrix of their cosines. It is easy to make a transformation and we don't discuss it here. However, this transformation was used in the experimental investigation the results of which are presented in the sequel. The K-means algorithm minimizes the functional Q_1.

The clustering algorithm proposed in [6, 10] has such a general structure: determination and analysis of the matrix $K = \{K(Y_i, Y_j),\ i, j = \overline{1,m}\}$. Let us detail it:

1. Compute the elements of matrix $\rho = \{\rho(Y_i, Y_j),\ i, j = \overline{1,m}\}$, where the function $\rho(\cdot)$ is defined by (12).

2. Find the maximal element $\rho_{\max}$ of the matrix ρ.

3. If the number of clusters is unknown and $\rho_{\max} < \delta_2$, set the number of clusters p equal to 1. Here the positive parameter δ_2 defines the threshold when the process of partitioning into different clusters starts.

4. Compute the elements of matrix $K = \{K(Y_i, Y_j),\ i,j = \overline{1,m}\}$, where the function $K(\cdot)$ is defined by (11) and the value of parameter α is selected so that the minimal element of the matrix K not exceed some positive value δ_1 $(0 < \delta_1 < 1)$, i.e., $\delta_1 = e^{-\alpha\rho^2_{\max}}$ and $\alpha = -\frac{\ln \delta_1}{\rho^2_{\max}}$.

5. Determine the number of clusters p (if it is unknown and not defined in Step 3) and partition initially the vectors $Y_1, \ldots, Y_m$ by using algorithms H1 or H2 (see Section 3).

6. Partition the vectors $Y_1, \ldots, Y_m$ into p clusters corresponding to the maximum of the functional I_1^1 (or I_1^2) by using algorithms A1, A2, A3, B1, B2, etc. (see Section 3).

Let us denote the clustering algorithm above by HA1, if it contains H1 and A1, and by HA2 if it contains H2 and A2. These versions of the algorithm are examined experimentally in the next sections. Algorithm HA1 is much faster than HA2 because HA1 contains A1, and HA2 contains A2. The computational efficiency of algorithms A1 and A2 is shown in Section 3.1.2: A2 yields better partitions as compared with A1 but requires more computer time. It means that, in the general case, HA2 will yield better results as compared with HA1.

We can vary the values of δ_1 and δ_2. The parameter δ_1 influences the form of the function $K(\cdot)$, and the clustering results depend on the value of δ_1. The results of clustering also significantly depend on the matrix $R^* = \{r^*_{kl},\ k,l = \overline{1,n}\}$ in the formula describing the distance $\rho(\cdot)$. The simplest (and, maybe, most often used) form of $\rho(\cdot)$ is the Euclidean distance. However, sometimes it is possible to estimate the degrees of importance of various components $y_1, \ldots, y_n$, e.g., from the point-of-view of measurement units of these components or the accuracy of their measurement. The generalized Euclidean distance may also be used (see, e.g., [4]) for the description of distance.

5. The problem of parameter clustering as a problem of clustering of vectors of S^n. The physical meaning of the matrix $\widetilde{X} = \{x_{kl},\ k = \overline{1,m},\ l = \overline{1,n}\}$ is the matrix of some experimental data or observations. The rows of the matrix $\widetilde{X}$ correspond to different observed objects (cases, vectors) $X_1, \ldots, X_m \in R^n$. The columns of the matrix $\widetilde{X}$ correspond to the parameters (variables)

$x_1, \ldots, x_n$ characterizing the state of objects. The previous section was devoted to the clustering of vectors. This section deals with the clustering of parameters according to their correlation matrix $R = \{r_{ij},\ i,j = \overline{1,n}\}$: r_{ij} is a correlation coefficient of parameters x_i and x_j.

A specific character of the problem of parameter clustering lies in the fact that the parameters x_i and x_j are related more strongly if the value of the correlation coefficient $|r_{ij}|$ is greater, and less strongly if the value of $|r_{ij}|$ is lower (see [5]). The minimal relationship between the parameters is equal to 0. The maximal relationship between the parameters is equal to 1 or -1.

The goal of clustering is to partition the parameters $x_1, \ldots, x_n$ into a fixed number p of non-intersecting and homogeneous, in a certain sense, groups $A_1, \ldots, A_p$ by the correlation matrix R, characterizing the relations among the parameters (see Sections 2 and 3). The relations among the parameters from the same group must be close, and the relations among the parameters from different groups must by weak. The values of elements of the matrix R may be both positive and negative.

The maximization of the functional I_1 by using the algorithms discussed in Sections 2 and 3 is rather computation-intensive. Various attempts to simplify the functional, characterizing the partitioning quality, by involving the absolute or squared values of correlation coefficients have been made. For example, the algorithm proposed in [5] maximizes the following heuristic functional:

$$F = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{n_k - 1}\sum_{\substack{x_i, x_j \in A_k \\ i \neq j}} \alpha_{ij},$$

where n_k is the number of parameters in the group A_k; α_{ij} is the absolute or squared value of the correlation coefficient r_{ij}. The package IMSL [43] also gives opportinity to analyse the correlation matrix: subroutines CLINK and DCLINK perform a hierarchical cluster analysis of the given distance matrix. The authors of CLINK and DCLINK suggest converting correlations to distances by taking the reciprocal of the absolute value of the correlation coefficient. Clustering algorithms, realized in CLINK and DCLINK, do not "care" whether the distances that are "fed" to them are actual real distances.

In [6] the application of matrices $|R| = \{|r_{ij}|,\ i,j = \overline{1,n}\}$ and $R^2 = \{r_{ij}^2,\ i,j = \overline{1,n}\}$ in clustering of parameters is investigated both theoretically and experimentally.

To apply functionals that describe the quality of vector clustering (see Section 4.2.1) to the clustering of parameters, it is necessary to determine a system of vectors $Y_1, \ldots, Y_n \in S^n$ corresponding to the system of parameters $x_1, \ldots, x_n$ in such a manner: $\cos(Y_i, Y_j) = |r_{ij}|$ or $\cos(Y_i, Y_j) = r_{ij}^2$.

Then the clustering of vectors $Y_1, \ldots, Y_n \in S^n$ should be performed.

The system of vectors $Y_1, \ldots, Y_n \in S^n$ exists, if the matrix of their scalar products is non-negative definite.

PROPOSITION 17 [6]. *The matrix R^2 is non-negative definite.*

REMARK 10. The non-negative definiteness of the matrix $|R|$ does not follow from the non-negative definiteness of the matrix R: the matrix $|R|$ may not be non-negative definite.

This fact is proved in [20] by examples. It is interesting that 148 matrices among 150 random ones used for the experiments in Section 5.1 remained non-negative definite after transforming into $|R|$.

5.1. Application of the K-means algorithm to parameter clustering on the basis of correliations. In this section, we show a possibility to maximize I_1 via minimization of the functional $\mathcal{Q}_1$. This functional defines the partitioning quality in the K-means clustering algorithm.

The problem is to analyse a system Y of vectors $Y_1, \ldots, Y_n \in S^n$, corresponding to the system of parameters $x_1, \ldots, x_n$, on the basis of the matrix of cosines between pairs of vectors from the system Y via minimization of the functional $\mathcal{Q}_1$ by using the K-means algorithm.

The experiments were carried out on the modification of the K-means algorithm presented in [33] (subroutine KMEANS). It is based on the analysis of vectors in consecutive order and on the search for a cluster of transferring a separate vector with a view to decrease the $\mathcal{Q}_1$ value. When KMEANS considers the vector Y_s (let $Y_s \in A_k$), it seeks the cluster A_L, where the value of functional $\mathcal{Q}_1$ decreases most after transferring Y_s.

The same 150 random correlation matrices R as in Section 3.1.2 were analysed. The results of performance of the algorithms were averaged. The mean obtained value of functional I_1 and the mean relative expenses t of computer time for solving one problem are presented in Table 3 dependent on n. The values of I_1 are computed on the basis of the matrix R.

KMEANS analysed matrices $|R|$ and R^2. The matrix R^2 is non-negative

Table 3. Clustering results

Algorithm	$n = 10$		$n = 20$		$n = 30$	
	t	I_1	t	I_1	t	I_1
R						
C1	1.00	4.29	2.67	6.87	4.33	9.44
H1	1.50	4.51	5.00	7.53	10.00	10.60
H2	4.33	4.55	16.00	7.68	33.17	10.85
$R + \|R\|$						
C1+KMEANS	1.67	4.57	5.33	7.77	11.33	11.03
H1+KMEANS	2.00	4.57	7.33	7.79	15.83	11.05
H2+KMEANS	4.67	4.58	17.83	7.79	37.33	11.08
$R + \|R\| + R$						
H2+KMEANS+A2	16.17	4.59	57.83	7.79	130.83	11.09
$R + R^2$						
C1+KMEANS	1.67	4.56	5.50	7.78	11.50	11.05
H1+KMEANS	2.00	4.56	7.50	7.78	16.50	11.04
H2+KMEANS	4.83	4.57	17.83	7.78	37.83	11.06

definite (see Proposition 17) and, in this case, the system of vectors $Y_1,\ldots,Y_n \in S^n$ exists. It follows from Remark 10 that the matrix $|R|$ may not be non-negative definite. The system of vectors $Y_1,\ldots,Y_n \in S^n$ does not exist in such a case, and KMEANS analyses a matrix of fictitious cosines.

The greatest values of functional I_1 were obtained by H2+KMEANS during the analysis of the matrix $|R|$. A further analysis of the obtained results using algorithm A2 on the basis of the matrix R improved them. However, the computation expenditure grew significantly. The following question arises: how often the partition of parameters, obtained by KMEANS, is a local maximum of I_1? To answer the question, the case C1+KMEANS ($n = 30$) was analysed more in detail. The analysis showed that all the partitions among 50 were locally optimal in the sense of Definition 1, and 31 partition was locally optimal in the

sense of Remark 3, when KMEANS used the matrix $|R|$. These numbers were equal to 49 and 14, respectively, in the case of matrix R^2.

The following conclusions may be drawn on the application of the K-means clustering algorithm in maximizing the functional I_1:

1. The initial partition has a weak influence on the final result.

2. The expenditure of computing time is significantly lower as compared with algorithms A2, B1, and B2 yielding similar values of I_1.

3. The expenditure of computing time of the K-means algorithm is similar to that of A1 and A3.

4. There is no necessity to seek a local maximum of I_1 in the sense of Definition 1 because,

- in most cases, the partition obtained by the K-means algorithm corresponds to the local maximum;
- it is impossible to answer uniquely, which functional (I_1 or Q_1) is better for clustering of parameters and better describes the partitioning quality.

These conclusions indicate that the proposed application of the K-means algorithm for analysis of the matrices $|R|$ and R^2 is an effective strategy for the clustering of parameters.

6. Experimental investigation of clustering algorithms. The aim of experimental investigation is to estimate the efficiency of the algorithm proposed in Section 4.3.1 (versions HA1 and HA2) in comparison with other well-known clustering algorithms.

The values of functionals (criteria) I_1^1, Q_1, and

$$Q_3 = \sum_{L=1}^{p} \sum_{X_i, X_j \in A_L} \rho^2(X_i, X_j),$$

where $\rho^2(X_i, X_j)$ is the Euclidean distance between the vectors X_i and $X_j \in R^n$, were evaluated for any obtained partition of the vectors $X_s = (x_{s1}, \ldots, x_{sn}) \in R^n$, $s = \overline{1, m}$, in order to compare them. Thus, the efficiency of application of I_1^1 as a criterion describing the partitioning quality was shown.

The criteria Q_1 and Q_3 are wide used in the comparison of cluster procedures (see [18]). Lower values of Q_1 and Q_3 correspond to better partitions. In algorithms HA1 and HA2, greater values of I_1^1 correspond to better partitions. Naturally, not only these three criteria are possible (see, e.g., Section 4.2 or

[29]). But a review and comparison of all possible criteria is out of the scope of this work.

The following algorithms were investigated experimentally:

HA1, HA2 – two versions of the algorithm proposed in Section 4.3.1;
MDC – the algorithm based on the method of dynamic condensations [21];
KC1, KC2 – two versions of the K-means algorithm [2, 4];
MC – the centroid hierarchical cluster analysis [2];
MW – Ward's minimum variance hierarchical cluster analysis [2];
MCC – the average linkage hierarchical cluster analysis [2].

The method of dynamic condensations [21] is a generalization of the K-means algorithm. The results of MDC are taken from [21].

MC, MW, and MCC are hierarchical clustering algorithms. Extensive investigations in the field of cluster analysis showed that MW and MCC are the best algorithms of hierarchical clustering (see, e.g., [22]). The experiments were carried out on the realizations of MC, MW, and MCC included into the system of statistical analysis SAS [2, 37]. The Euclidean distance was used in the algorithms to define a neighbourhood of vectors.

There is a lot of realizations of the K-means algorithm. Their differences are presented, for example, in [4]. KC1 is taken from [4], and KC2 is taken from [2].

The partitions obtained by algorithms MC, MW, MCC, KC1, and KC2 have such main peculiarities:

1. Any vector from the partition obtained by KC1 and KC2 is located closer to the centre of its own cluster than to the centres of other clusters.

2. Algorithms MC, MW, MCC, KC1, and KC2 find compact groups of vectors: any pair of groups may be divided by a hyperplane.

The peculiarities of HA1 and HA2:

1. The Euclidean distance is used as the function $\rho(\cdot)$.

2. $\delta_1 = 0.01$.

3. $\delta_2 = 0$, i.e., the number of clusters is assumed to be larger than 1.

4. Algorithm HA1 contains H1 and A1; algorithm HA2 contains H2 and A2.

In [6, 10], the experiments were carried out on various test data sets located on a plane $x_1 0 x_2$ and one real data set (Fisher's data on irises [2, 17]). We present here the results of analysis of three data sets, only.

Fisher's data on irises [2, 17] are wide used to demonstrate the discriminant and cluster analysis. 150 irises of three different sorts (50 irises of setosa, versicolor and virginica) were selected, and four parameters (in millimetres) of any iris were measured: length of sepal, width of sepal, length of petal, width of petal. It is useful to analyse these data because the "ideal" partition into clusters is known, and it is possible to estimate the efficiency of cluster-procedures by comparing the obtained and the "ideal" partition. An automatic detection of the number of clusters by HA1 and HA2 indicated that $p = 3$. The results of partitioning into three clusters are presented in Table 4. The column '%' means the percent of misclassified irises. The values of functionals Q_1, Q_3, and I_1^1 indicate that the second partition is the worst. The values of Q_1 and I_1^1 indicate that the first partition is the best. The last partition is the best from the standpoint of Q_3.

Table 4. Clustering the data on irises

Algorithms	%	Q_1	Q_3	I_1^1
HA1, HA2, KC2	10.67	7884.68	826774	137.13
MDC	9.33	11789.06	1248794	132.64
MC	9.33	7944.16	851110	137.05
MW	10.67	7929.27	852210	137.07
MCC	10.00	7913.92	939198	137.09
KC1	11.33	7885.03	816886	137.12

The data set given in Fig. 1 has four visually evident clusters: $A_1 = \{X_1 - X_9\}$, $A_2 = \{X_{10} - X_{18}\}$, $A_3 = \{X_{19} - X_{27}\}$, $A_4 = \{X_{28} - X_{36}\}$. The experiments showed that this partition corresponds to the local maximum of I_1^1. The results of clustering are presented in Table 5.

The data set given in Fig. 2 has two visually evident clusters which cannot be separated by a line: $A_1 = \{X_1 - X_{27}\}$, $A_1 = \{X_{28} - X_{36}\}$. The experiments showed that this partition corresponds to the local maximum of I_1^1. It means that the algorithm proposed in Section 4.3.1 may yield a partition into clusters that,

in the general case, are not convex and cannot be separated by a hyperplane. Such a peculiarity of the algorithm arises from the non-linearity of the function $K(\cdot)$ which describes the similarity of vectors.

Table 5. Results of clustering the data set from Fig. 1

Algorithms	A_1	A_2	A_3	A_4	Q_1	Q_3	I_1^1
HA1, HA2, KC2	1–9	10–18	19–27	28–36	0.480	8.64	29.09
MC	1–10, 30	11–15, 17, 18	16, 20–27	19, 28, 29, 31–36	0.552	10.44	28.45
MW	1–4, 6–9, 30	5, 10–18, 21	19, 28, 29, 31–36	20, 22–27	0.552	10.44	28.45
MCC	1–4, 6–9, 30	5, 10–15, 17, 18	16, 20–27	19, 28, 29, 31–36	0.551	9.92	28.47
KC1	1, 5–10, 14–18	2–4, 11–13	19, 23–28, 32–36	20–22, 29–31	1.445	29.54	22.46

The results of investigation indicate that the proposed clustering algorithm may successfully compete with other well-known algorithms. Its efficiency depends on the following aspects:

- the function $K(\cdot)$ and its parameters,
- the function $\rho(\cdot)$ (when $K(\cdot) = K(\rho(\cdot))$,
- the algorithm of initial partitioning,
- the algorithm of maximization of a functional characterizing the partitioning quality.

One can to construct various versions of the clustering algorithm by varying the components above. Therefore, the problem remains to find the optimal configuration of the algorithm for specific classes of problems or for the general case.

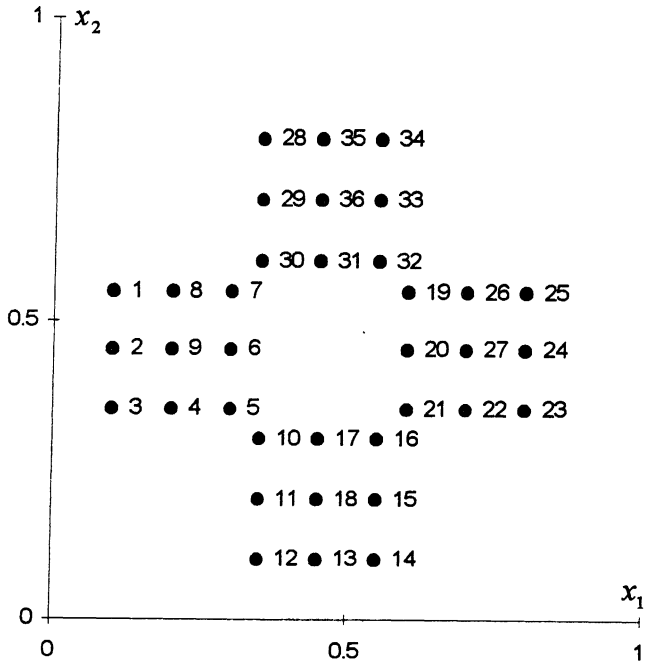

Fig. 1. Data set having four evident clusters.

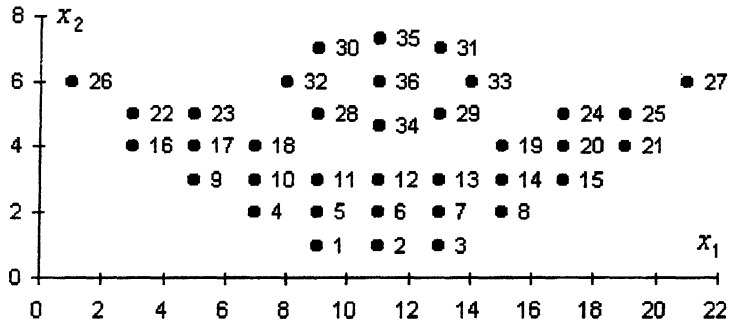

Fig. 2. Data set having two evident clusters.

The algorithms of initial partitioning H1 and H2, which were included into HA1 and HA2, may suggest the optimal, in their convenience, number of clusters. However, the problem of reliability of the determined number of clusters in a data set remains (see [28, 30, 32] for more details on the investigations regarding the optimal number of clusters).

The quality of partitioning may be increased by using some initial knowledge about the number of groups and the known vectors from different groups. The scheme of algorithms of initial partitioning H1 and H2 allows us to take account of such information.

The experiments showed that the functional I_1^1 may be successfully used to determine the partitioning quality of data sets. The values of I_1^1 have a small variance in the neighbourhood of sufficiently good solutions (see, e.g., data on irises, where five really different solutions have very similar values of I_1^1).

A new way to economize the computing time and to get very good solutions is the application of the **K**-means-type algorithm in the proposed clustering algorithm for the analysis of matrix K instead of A1 or A2. The results of Section 5.1 allow us to assume that, in this case, the obtained solutions will be similar to those obtained by HA2, and the computational expenditure will be similar to that of HA1.

7. Conclusions. Theoretical and experimental investigations of various deterministic strategies for solving the problem of extremal parameter grouping indicate that the problem of selection of an algorithm is a multiple criteria one: it is necessary to take into consideration two contradictory factors: optimization accuracy and recourses of computing time allocated for the grouping. During the experiments above, greater values of I_1 are obtained by the algorithms which are more computation-intensive. The possibility of finding the optimal number of clusters is considered. In this case, one may search for some set of the basic parameters: the number of parameters in it may be used as the optimal number of groups.

The transformation of a general clustering problem into the clustering on the sphere allows us to change the interlocation of points nonlinearly resulting in a better separation of clusters. The clustering algorithm proposed in [6, 10] may successfully compete with other well-known algorithms. It may find clusters that, in the general case, are not convex and cannot be separated by a hyperplane. Such a peculiarity of the algorithm comes from a non-linearity of

the function which characterizes the similarity of vectors. The experiments of test data sets on a wide range showed that such transformation leads to better partitions of vectors into clusters.

The relation, discovered in [6, 10], between the clustering of parameters on the basis of their correlation matrix and the clustering of vectors of the unit sphere enables us to use simpler functionals characterizing the partitioning quality. Modifications of considerably faster classical algorithms (e.g., K-means) may be used to optimize these functionals.

REFERENCES

[1] Harman, H.H. (1976). *Modern Factor Analysis,* 3 rd ed. University of Chicago Press, Chicago.

[2] *SAS User's Guide: Statistic,* 1982 ed. (1982). SAS Institute Inc, Cary.

[3] Anderberg, M.R. (1973). *Cluster Analysis for Applications.* Academic Press, New York.

[4] Hartigan, J. (1975). *Clustering Algorithms.* Wiley – Interscience, New York.

[5] Braverman, E.M., and I.B. Muchnik (1983). *The Structural Methods for Empirical Data Processing.* Nauka, Moscow (in Russian).

[6] Dzemyda, G. (1990). Grouping on the sphere. In A. Žilinskas (Ed.), *Teorija Optimaljnych Reshenij,* Vol. 14. Inst. Math. Cybern., Vilnius. pp. 22–40 (in Russian).

[7] Braverman, E.M. (1970). Methods of extremal grouping of parameters and the problem of apportionment of essential factors. *Avtomatika i Telemechanika,* **1**, 123–132 (in Russian).

[8] Dzemyda, G. (1987). On the extremal parameter grouping. In A. Žilinskas (Ed.), *Teorija Optimaljnych Reshenij,* Vol. 12. Inst. Math. Cybern., Vilnius. pp. 28–42 (in Russian).

[9] Dzemyda, G. (1988). The algorithms of extremal parameter grouping. In A. Sydow, S.G. Tzafestas and R. Vichnevetsky (Eds.), *Mathematical Research,* Band 46, *Systems Analysis and Simulation 1988,* Vol. 1, *Theory and Foundations, Proceedings of the International Symposium Held in Berlin (GDR).* Akademie-Verlag, Berlin. pp. 133–136.

[10] Dzemyda, G., and J. Valevičienė (1988). The extremal parameter grouping in cluster analysis. In A. Žilinskas (Ed.), *Teorija Optimaljnych Reshenij,* Vol. 13. Inst. Math. Cybern., Vilnius. pp. 36–53 (in Russian).

[11] Dzemyda, G., and E. Senkienė (1992). Simulated annealing for parameter grouping. In *Trans. of the Eleventh Prague Conf. on Information Theory, Statistical*

Decision Functions, Random Processes. Academia, Prague. pp. 373–383.

[12] Dzemyda, G., E. Senkienė and J. Valevičienė (1990). On the problem of parameter grouping. In H.-M.Voigt, H.Mühlenbein and H.-P.Schwefel (Eds.), *Evolution and Optimization'89. Selected Papers on Evolution Theory, Combinatorial Optimization, and Related Topics.* Akademie-Verlag, Berlin. pp. 216–224.

[13] Lumelsky, V.Y. (1970). Grouping of parameters on the basis of communication matrix. *Avtomatika i Telemechanika,* **1**, 133–143 (in Russian).

[14] Shilov, G.E. (1969). *Mathematical Analysis (Finite-Dimensional Linear Spaces).* Nauka, Moscow.

[15] Aizerman, M.A., E.M. Braverman and L.I. Rozanoer (1970). *The Method of Potential Functions in the Training of Machines,* Nauka, Moscow (in Russian).

[16] Gixman, I.I., and A.V. Skoroxod (1971). *Theory of Random Processes,* Vol. 1. Nauka, Moscow (in Russian).

[17] Fisher, R.A. (1936). The use of multiple measurements in tachonomic problems. *Annals of Eugenic,* **7**, 643–663.

[18] Aivazian, S.A., Z.I. Bezhaeva and O.V. Staroverov (1974). *Classification of the Multidimensional Observations.* Statistika, Moscow (in Russian).

[19] Bolshev, L.N. (1969). Cluster analysis. *Bull. Int. Stat. Inst.,* **43**, 441–425.

[20] Horn, P., and C. Johnson (1989). *Matrix Analysis.* Mir, Moscow (in Russian).

[21] Diday, E. (1985). *Data Analysis Methods: Approach Based on the Dynamical Condensations.* Finansy i Statistika, Moscow (in Russian).

[22] Milligan, G.W. (1980). An examination of the effect of six types of error perturbation of fifteen clustering algorithms. *Psychometrika,* **45**(3), 325–343.

[23] Krylov, V.I., V.V. Bobkov and P.I. Monastyrnyj (1976). *Computing Methods.* Vol. 1. Nauka, Moscow (in Russian).

[24] Dubrovskij, S.A. (1982). *Applied Multidimensional Statistical Analysis.* Finansy i Statistika, Moscow (in Russian).

[25] Dzemyda, G. (1987). LP-search with extremal problem structure analysis. In N.K. Sinha and L.A. Telksnys (Eds.), *Proceedings of the 2nd IFAC Symposium, IFAC Proceedings Series,* 1987, Number 2. Pergamon Press, Oxford (Wheaton, Exeter) - New York - Toronto. pp. 499–502.

[26] Nebylicyn, V.D. (1966). *The Main Properties of Human Mentality.* Prosveszenije, Moscow (in Russian).

[27] Mezzich, J.E., and H. Solomon (1980). *Taxonomy and Bexavioral Science, Comparative Performance of Grouping Methods.* Academic Press, London-New York.

[28] Dubes, R.C. (1987). How many clusters are best? - An experiment. *Pattern Recognition,* **20**(6), 643–663.

[29] Jambu, M. (1988). *Hierarchical Cluster Analysis and Associations.* Finansy i Statistika, Moscow (in Russian).

[30] Baker, F.B., and L.J. Hubert (1975). Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association,* **70**, 31–38.

[31] Dzemyda, G. (1996). Clustering of parameters on the basis of correlations via simulated annealing. *Control and Cybernetics,* **25**(1), R.V.V.Vidal and Z.Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization,* 55–74.

[32] Milligan, G.W., and M.C. Cooper (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika,* **50**(2), 159–179.

[33] Späth, H. (1980). *Cluster Analysis Algorithms for Data Reduction and Classification of Objects.* Ellis Horwood, Chichester.

[34] Braverman, E.M., A.A. Dorofejuk, V.Y. Lumelsky and I.B. Muchnik (1970). Diagonalization of the communication matrix and apportionment of hidden factors. In M.A. Aizerman (Ed.), *Problemy Rasshyrenija Vozmozhnostej Avtomatov,* Vol. 1. Institute of Control Problems, Moscow. pp. 42–79 (in Russian).

[35] Owsinski, J.W. (Ed.) (1986). *Control and Cybernetics, Special Issue on Optimization Approaches in Clustering,* **15**(2).

[36] Reyment, R A., and K.G. Joreskog (1993). *Applied Factor Analysis in the Natural Sciences.* Cambridge University Press, Cambridge.

[37] SAS/STAT (1990). *SAS/STAT User's Guide Vols. I&II – Version 6 Edition.* SAS Institute, Cary.

[38] Boisvert, R.F., S.E. Howe, D.K. Kahaner and J.L. Springmann (1990). *The Guide to Available Mathematical Software.* National Institute of Standards and Technology Internal Report NISTIR 90-4237 (also PB90-216508/AS, National Technical Information Service, Springfield, VA 22161), March 1990.

[39] SPPS, Inc. (1997). *Systat 6.0 for Windows: Statistics.* Prentice-Hall.

[40] Wilkinson, L., G. Blank and C. Gruber (1996). *Desktop Data Analysis with SYSTAT.* Prentice-Hall.

[41] StatSoft, Inc. (1995). *STATISTICA for Windows [Computer Program Manual].* StatSoft, Inc., Tulsa, OK.

[42] *NAG Fortran Library Mark 16 Manual,* Vol. 1–12 (1994). Numerical Algorithms Group, Oxford.

[43] *IMSL STAT/LIBRARY User's Manual,* Vol. 1–3 (1987). IMSL Inc., Houston, TX (QA76.73 in NIST Gaithersburg Research Information Center, Admin E-120).

[44] Jöreskog, K.G., J.E. Klovan and R.A. Reyment (1976). *Geological Factor Analysis.* Elsevier, Amsterdam.

Received November 1996

G. Dzemyda received his Ph.D. degree from the Kaunas Polytechnic Institute, Kaunas, Lithuania, in 1984. He is a senior researcher at the Optimization Department of the Institute of Mathematics and Informatics, and an Associate Professor at the Vilnius Pedagogical University. His research interests include knowledge discovery seeking a better optimization efficiency and interaction of optimization and data analysis.

PARAMETRŲ KLASTERIZAVIMAS KORELIACIJOS PAGRINDU: LYGINAMOJI DETERMINISTINIŲ STRATEGIJŲ APŽVALGA

Gintautas DZEMYDA

Darbe nagrinėjamos deterministinės parametrų klasterizavimo strategijos, kai parametrų sąveiką nusako tiktai jų koreliacinė matrica. Ypatingas dėmesys skiriamas ekstremalinio parametrų grupavimo metodo realizacijoms. Parodyta optimalaus klasterių skaičiaus paieškos galimybė. Nustatytasis ryšys tarp parametrų klasterizavimo ir daugiamačių vektorių klasterizavimo uždavinių yra naudojamas klasterizavimo (ir parametrų, ir daugiamačių vektorių) efektyvumo gerinimui.