INFORMATICA, 1997, Vol. 8, No. 4, 465–476

CONVERGENCE OF THE PARAMETER CLUSTERING BASED ON THE SIMULATED ANNEALING

Gintautas DZEMYDA, Elvyra SENKIENĖ

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

Abstract. The problem of parameter clustering on the basis of their correlation matrix is considered. The convergence in probability of parameter clustering based on the simulated annealing is investigated theoretically.

Key words: parameter clustering, simulated annealing, combinatorial optimization.

1. Introduction. Any set of similar objects 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. The correlation matrix of parameters may be calculated during the analysis of objects composing the set. There exist groups (clusters) of parameters characterizing different properties of the object. The problem is to find these groups.

One of the major objectives of various data analysis methods is to discover relations among the parameters. The methods analysed 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].

Parameter clustering found a wide application in the optimization of multiextremal and continuous functions, too (see [3, 4, 30]).

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 $A_1, \ldots, A_p$ 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 the correlation coefficient of parameters x_i and x_j). The 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.

Here we deal with the parameter clustering [5]–[12], [31] based on the analysis of correlations and maximizing the partitioning quality

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

where F_L is the factor with a unit variance, corresponding to the group A_L; $r_{x_i F_L}$ is the 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

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

2. The problem of combinatorial optimization. The problem of parameter clustering is a combinatorial optimization problem. Combinatorial optimization problems, although largely overlooked in the early development of optimization techniques, are rather common in the operational research and management science applications. In just these domains there arose different problems in location, transport, scheduling, network design, timetabling, partitioning, path planning, assignment, cutting, and elsewhere (see, e.g., [29], [32]), with common characteristics of beeing NP-complete, that is, hard (or rather impossible at present) to solve in a reasonable time for problems in real dimensions. Recent developments like, e.g., those in communication technology, management of world scale enterprises or artificial intelligence techniques introduce new combinatorial problems and often new scales, far exceeding the dimensions considered before.

The problem of parameter clustering can be formulated as a combinatorial optimization problem. Let $X^1, \ldots, X^n$ be variables taking discrete values from 1 to p, $K = \{X = (X^1, \ldots, X^n)\colon X^i \in \{1, \ldots, p\},\ i = \overline{1, n}\}$. Let us introduce a function $f(X^1, \ldots, X^n)$ that is related with the functional I_1 in such a manner:

$$f(X^1, \ldots, X^n) = I_1, \quad \text{where} \quad x_i \in A_L \quad \text{as} \quad X^i = L.$$

It means that any point from K corresponds to the fixed parameter partition, and any partition of parameters corresponds to some point in K.

The problem of parameter clustering is proposed (see [11], [12], [19]) to be formulated as follows:

$$\max_X f(X) \tag{1}$$

subject to

$$X = (X^1, \ldots, X^n) \in K, \tag{2}$$

$$\exists i\colon X^i = 1, \ldots, \exists i\colon X^i = p. \tag{3}$$

3. Simulated annealing. The deterministic algorithms [5]–[12], [31] often find only the local maximum of I_1. All the algorithms start from some initial partition selected by some algorithm or by a certain knowledge of the problem. They are based on the analysis of parameters in consecutive order and on the search for a group of transferring the individual 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. The deterministic algorithms practically realise a single coordinate maximization of $f(X)$: consequently, the values of individual variables from $\{X^1, \ldots, X^n\}$ run from 1 to p (the values of other variables are fixed) and there is searched for the increase of the value of objective function.

In [11], [12], [19], the problem (1)–(3) was attempted to be solved using special methods for combinatorial optimization, i.e., using simulated annealing. In this case the variety of strategies for the parameter clustering was extended. The strategies of maximization of I_1 which is less sensitive on the local solutions, i.e., which allows to search for the global maximum of I_1, is proposed in [11], [12], [19].

Since [15] much work has been done on simulated annealing for discrete variables and it has been used in a wide range of contexts. A brief description of different modifications to the discrete simulated annealing algorithm can be found in [16]. Simulated annealing was used in solving partitioning problems, too [17], namely, partitioning of networks. A new applications of the simulated annealing may be found in [18] (a global discrete optimization), [19] (parameter clustering), [20] (assortment problems with cutting policies), [21] (power network design), [22] (total tardiness problem), [23] (dial-a-ride problems), [24] (shape detection), [25] (location of petrol stations).

The results below concern the convergence of the clustering algorithms proposed and investigated experimentally in [11], [12], [19].

Let us consider the simulated annealing strategy in search of the global maximum of the combinatorial problem

$$\max_{X \in S \subset R^n} f(X),$$

where $X = (X^1, \ldots, X^n)$, $S = \{X\colon X^i \in \{A^i, \ldots, B^i\},\ i = \overline{1,n}\}$, X^i, A^i and B^i, $i = \overline{1,n}$, take integer values, $A^i \leqslant X^i \leqslant B^i$.

$A^i = 1$, $B^i = p$, $i = \overline{1,n}$, in the case of problem (1)–(3).

The performance of optimization algorithms based on simulated annealing can be generalized as follows. Let $m-1$ step be performed. The current point is $X_{m-1} = (X^1_{m-1}, \ldots, X^n_{m-1})$. The problem is to find the next current point $X_m = (X^1_m, \ldots, X^n_m)$. It may be one of the neighbours of X_{m-1}. X_{m-1} can remain as the current point after m steps, too. The selection of X_m is divided into two stages. X_m is chosen from the neighbours of X_{m-1} in the first stage. Then X_m and X_{m-1} are compared in the second stage. X_{m-1} can become X_m with some probability. The algorithms in [11], [12], [19] correspond this general scheme.

Let us denote:

$N(X_j)$ is the set of neighbours of $X_j = (X^1_j, \ldots, X^n_j)$,

$S_n = \{X_i \in S \mid f(X_j) \geqslant f(X_i) \quad \forall X_j \in N(X_i)\}$.

The search for the global maximum of $f(\cdot)$ can be performed in such a manner: the m-th step of the algorithm is as follows:

$$X^i_m = X^i_{m-1} + \xi^i, \quad m = 1, 2, \ldots,\ i = \overline{1,n}, \tag{4}$$

where ξ^i, $i = \overline{1,n}$, are integers taking values with some probabilities:

a) ξ^i, $i = \overline{1,n}$, are random numbers taking integer values in the set $\{-1, 0, 1\}$; $P\{\xi^i = 0,\ i = \overline{1,n}\} = 0$, and the probability for any other combination of ξ^i, $i = \overline{1,n}$, to appear is equal to $1/(3^n - 1)$;

b) $\xi^i \in S^i = \{A^i - X^i_{m-1}, A^i + 1 - X^i_{m-1}, \ldots, B^i - X^i_{m-1}\} \setminus \{0\}$, $i = \overline{1,n}$, with the same probability $p_i = 1/(B^i - A^i)$, i.e., $N(X_{m-1}) = S \setminus \{X_{m-1}\}$, where $N(X_{m-1})$ is the set of neighbours of X_{m-1}.

Taking into account a specific character of the functional, characterizing the partitioning quality of parameters, the authors in [11], [12], [19] suggest restricting the set of neighbours of the current point. Thus, two additional special cases of ξ^i selection for (4) are used in [11], [12], [19]:

c) $P\{\xi^k = -1\} = P\{\xi^k = 1\} = 1/2$,
$\xi^i = 0$, $i = 1, 2, \ldots, k-1, k+1, \ldots, n$, $k = 1, 2, \ldots, n$,
d) $\xi^k \in S^k = \{A^k - X^k_{m-1}, A^k + 1 - X^k_{m-1}, \ldots, B^k - X^k_{m-1}\} \backslash \{0\}$
with the same probabilities:
$p_k = 1/(B^k - A^k)$, $\xi^i = 0$,
$i = 1, 2, \ldots, k-1, k+1, \ldots, n$, $k = 1, 2, \ldots, n$.

Case c) is a restriction of case a), and case d) is that of b). The peculiarity of these two cases is that only the k-th coordinate of X_m and X_{m-1} differs, and different values of k correspond to the consequent steps. The relation of m and k may be defined in a more sophisticated way (see [12], [19] for examples of such a relation).

The probability of transition to the point X_m is defined by the formula:

$$P\{X_m\} = \begin{cases} 1, & \text{as } f(X_m) > f(X_{m-1}), \\ \exp\{[f(X_m) - f(X_{m-1})]/T_m\}, & \text{as } f(X_m) \leqslant f(X_{m-1}), \end{cases} \tag{5}$$

i.e., $P\{X_m\} = 1$ as $f(X_m) > f(X_{m-1})$; in the other case a random number $\eta \in [0, 1]$ is generated: the point X_m will be initial for a new step ($(m+1)$-st) of the algorithm and in formula (4) it will replace X_{m-1} if

$$\eta < \exp\{[f(X_m) - f(X_{m-1})]/T_m\},$$

and the point X_{m-1} remains as the initial one for a new step, otherwise.

$$T_m = c/\ln(1 + m_0 + m), \tag{6}$$

or

$$T_m = c/\ln[\ln(1 + m_0 + m)], \tag{7}$$

$m = 1, 2, \ldots$ is the number of a step, c is a positive constant, m_0 is some constant from $[1, \infty)$.

The proof of convergence of the algorithm in probability to the global maximum of $f(\cdot)$ is based on the results presented in [14].

Theorem 1. *If*
1. $T_m \leqslant T_{m-1}$,
2. $\lim\limits_{m \to \infty} T_m = 0$, *where* $T_m = c/\ln(1 + m_0 + m)$,
3. $c \geqslant rL$, *where*

$$L = \max_{X_i \in S} \max_{X_j \in N(X_i)} |f(X_i) - f(X_j)|,$$

$$r = \max_{X_i \in S \backslash S_n} \max_{X_j \in S} \Big(\sum_{k=1}^{n} (X_i^k - X_j^k)^2 \Big)^{1/2},$$

then the annealing algorithm (4)–(6) *converges in probability to the global maximum of* $f(X)$, *i.e.,* $\lim_{m\to\infty} P\{|X_m - q| < \varepsilon\} = 1$, *where* q *is in the set of all the points that are global maxima of* $f(X)$.

The proof of Theorem 1 follows from [14].

The transition probability $P\{X_m\}$ with an unknown parameter c (see (5)–(7)) may be modified into the form with an unknown parameter $\delta \in (0, 1]$. If we use some initial probability $P\{X_1\} = \delta$ as $m = 1$ and if T_m has the form (6), then the constant c can be expressed:

$$c = [f(X_1) - f(X_0)] \ln(2 + m_0)/\ln \delta, \tag{8}$$

where X_0 and X_1 are such that $f(X_1) < f(X_0)$. Then (5) will have such a form (for $m = 2, 3, \ldots$):

$$P\{X_m\} = \begin{cases} 1, & \text{as } f(X_m) > f(X_{m-1}) \\ (1 + m_0 + m)^{\frac{[f(X_m) - f(X_{m-1})]}{[f(X_1) - f(X_0)]} \frac{\ln \delta}{\ln(2+m_0)}}, & \text{as } f(X_m) \leqslant f(X_{m-1}) \end{cases} \tag{9}$$

If T_m has the form (7), then the transition probability may be transformed as follows:

$$P\{X_m\} = \begin{cases} 1, & \text{as } f(X_m) > f(X_{m-1}) \\ [\ln(1 + m_0 + m)]^{\frac{[f(X_m) - f(X_{m-1})]}{[f(X_1) - f(X_0)]} \frac{\ln \delta}{\ln[\ln(2+m_0)]}}, & \text{as } f(X_m) \leqslant f(X_{m-1}) \end{cases} \tag{10}$$

PROPOSITION 1. If

1. $T_m = c/\ln[\ln(1 + m_0 + m)]$,
2. $c \geqslant rL$,

then

$$\sum_{k=k_0}^{\infty} \exp\left\{ - \frac{rL}{T_{kr-1}} \right\} = \infty, \quad k_0 < \infty. \tag{11}$$

Proof.

$$\begin{aligned} \sum_{k=k_0}^{\infty} \exp\left\{ - \tfrac{rL}{T_{kr-1}} \right\} &= \sum_{k=k_0}^{\infty} \exp\left\{ - \tfrac{rL}{c} \ln\left[\ln(kr + m_0) \right] \right\} \\ &= \sum_{k=k_0}^{\infty} \left[\ln(kr + m_0) \right]^{-\frac{rL}{c}} = \sum_{k=k_0}^{\infty} \left[\ln\left\{ r(k + m_0/r) \right\} \right]^{-\frac{rL}{c}} \\ &= \sum_{k=k_0}^{\infty} \left[\ln r + \ln(k + m_0/r) \right]^{-\frac{rL}{c}} = \infty. \end{aligned}$$

The proposition is proved.

Theorem 2. *If*

1. $T_m \leqslant T_{m-1}$, (12)
2. $\lim_{m\to\infty} T_m = 0$, (13)
   where $T_m = c/\ln\left[\ln(1 + m_0 + m)\right]$,
3. $c \geqslant rL$, (14)

then the annealing algorithm (4), (5), (7) *converges in probability to the global maximum of* $f(X)$, *i.e.,* $\lim_{m\to\infty} P\{|X_m - q| < \varepsilon\} = 1$, *where* q *is in the set of all the points that are global maxima of* $f(X)$.

If conditions (11)–(14) are satisfied, then theorems, analogous to 4.1, 5.1, and 5.2 in [14], may be formulated and proved for our case, i.e., algorithm (4), (5), (7) converges to the global maximum.

The following theorem, analogous to Hajek's theorem [27], [28], and propositions deal with the convergence of the annealing algorithm, too.

Theorem 3. *Let*

1. $\lim_{m\to\infty} T_m = 0$,
2. $T_m \leqslant T_{m-1}, \quad m = 1, 2, \ldots$.

The annealing algorithm (4), (5) *converges in probability to the global maximum of* $f(X)$ *if and only if*

$$\sum_{m=1}^{\infty} \exp\left\{-\frac{D}{T_m}\right\} = \infty, \tag{15}$$

where

- $D = \max_{X_i \in S \setminus S_{\max}} \{d(X_i),\ X_i \notin S_{\max},\ X_i$ *is the local maximum*$\}$;
- S_{max} *is the set of all the points that are global maxima of* $f(X)$;
- $d(X_i)$ *is some function which value depends on* X_i, *and* $d(X_i) = \infty$ *if* $X_i \in S_{\max}$ *(for more details see* [27], [28]*).*

PROPOSITION 2. Let

1. $T_m = c/\ln(1 + m_0 + m), \quad m = 1, 2, \ldots$,
2. c be some positive constant,
3. $1 \leqslant m_0 < \infty$.

The annealing algorithm (4), (5) converges in probability to the global maximum of $f(X)$ if and only if $c \geqslant D$.

Proof. From $c \geqslant D$ we obtain

$$\exp\left\{-\frac{D}{T_m}\right\} = \exp\left\{-\frac{D\ln(1+m_0+m)}{c}\right\} \geqslant \exp\left\{-\ln(1+m_0+m)\right\} = 1/(1+m_0+m).$$

By $\sum_{m=1}^{\infty} 1/(1+m_0+m) = \infty$ we obtain (15) of Theorem 3.
The proposition is proved.

PROPOSITION 3. Let
1. $T_m = c/\ln\left[\ln(1+m_0+m)\right], \quad m = 1, 2, \ldots,$
2. c be some positive constant,
3. $1 \leqslant m_0 < \infty$.

The annealing algorithm (4), (5) converges in probability to the global maximum of $f(X)$ if and only if $c \geqslant D$.

Proof. From $c \geqslant D$ we obtain

$$\exp\left\{-\frac{D}{T_m}\right\} = \exp\left\{-\frac{D\ln[\ln(1+m_0+m)]}{c}\right\} \geqslant \exp\left\{-\ln[\ln(1+m_0+m)]\right\} = 1/(1+m_0+m).$$

By $\sum_{m=1}^{\infty} 1/(1+m_0+m) = \infty$ we obtain (15) of Theorem 3.
The proposition is proved.

4. Conclusions. In this paper, the convergence in probability of algorithms of parameter clustering proposed in [11], [12], [19] has been investigated. However, the results of this paper do not give the answer about the rate of convergence. The rate depends on the value of parameter c (or δ) and the form of T_m. The optimal value of δ is determined in [19] experimentally, only. The influence of the form of T_m on the rate of convergence is investigated in [19] experimentally, too.

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. Dzemyda, G. (1997). *Knowledge Discovery Seeking a Higher Optimization Efficiency. Research Report Presented for Habilitation.* Mokslo Aidai, Vilnius. ISBN 9986–479–28–2.
4. 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. pp. 499–502.

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. Mitra, D., F. Romeo and A. Sangiovanni-Vincentelli (1986). Convergence and finite-time behavior of simulated annealing. *Adv.Appl.Prob.*, **18**(3), 747–771.
15. Kirkpatrick, S., C.D. Gelatt and M.P. Vecchi (1983). Optimization by simulated annealing. *Science*, **220**(4598), 671–680.
16. Eglese, R.W. (1990). Simulated annealing: a tool for operational research. *European Journal of Operational Research*, **46**, 271–281.
17. Trzebiatowski, G.W. (1985). Thermodynamic simulation procedure for partitioning problems. In H.-M. Voigt (Ed.), *Informatis: Informationen-Reporte*, Vol. 1, No. 12. Institut für Informatik und Rechentechnik, Berlin. pp. 86–96.
18. Ingber, L. (1996). Adaptive simulated annealing (ASA): lessons learned. *Control and Cybernetics*, **25**(1), R.V.V. Vidal and Z. Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 33–54.
19. 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.

20. Borges, P.C., and J.S. Ferreira (1996). Assortment problems with cutting policies. *Control and Cybernetics*, **25**(1), R.V.V. Vidal and Z. Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 75–95.
21. Ravn, H.F., J.M. Rygaard and B. Wibbels (1996). Power network design. *Control and Cybernetics*, **25**(1), R.V.V. Vidal and Z. Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 97–120.
22. Antony, S.R., and C. Koulamas (1996). Simulated annealing applied to the total tardiness problem. *Control and Cybernetics*, **25**(1), R.V.V. Vidal and Z. Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 121–130.
23. Hart, S.M. (1996). The modelling and solution of a class of dial-a-ride problems using simulated annealing. *Control and Cybernetics*, **25**(1), R.V.V. Vidal and Z. Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 131–157.
24. Di Ianni, M., R.Diekmann, R.Lüling, J.Schulze and S.Tschöke (1996). Simulated annealing and genetic algorithms for shape detection. *Control and Cybernetics*, **25**(1), R.V.V.Vidal and Z.Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 159–175.
25. Czyżak, P., and A. Jaszkiewicz (1996). A multiobjective metaheuristic approach to the location of petrol stations by the capital budgetting model. *Control and Cybernetics*, **25**(1), R.V.V. Vidal and Z. Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 177–187.
26. Nebilicin, V.D. (1966). *The Main Properties of Human Mentality*. Prosvesczenije, Moscow (in Russian).
27. Hajek, B. (1988). Cooling schedules for optimal annealing. *Mathematics of Operations Research*, **113**, 311–329.
28. Van Laarhoven, P.J.M. (1988). *Theoretical and Computational Aspects of Simulated Annealing*. Amsterdam.
29. Pirlot, M., and R.V.V. Vidal (1996). Simulated annealing: a tutorial. *Control and Cybernetics*, **25** (1), R.V.V.Vidal and Z. Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 9–31.
30. Šaltenis, V., and G.Dzemyda (1982). The structure analysis of extremal problems using some approximation of characteristics. In A. Žilinskas (Ed.), *Teorija Optimaljnych Reshenij*, Vol. 8. Inst. Math. Cybern., Vilnius. pp. 124–138 (in Russian).
31. Dzemyda, G. (1997). Clustering of parameters on the basis of correlations: A comparative review of deterministic approaches. *Informatica*, **8**(1), 83–118.
32. Vidal, R.V.V. (Ed.) (1993). *Applied Simulated Annealing*. Springer Verlag.

Received October 1997

G. Dzemyda received his Ph.D. and Habil. Dr. degrees from the Kaunas Polytechnic Institute, Kaunas, Lithuania, in 1984 and 1997, respectively. He is a senior researcher at the Optimization Department of the Institute of Mathematics, 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.

E. Senkienė graduated from Vilnius University, Lithuania, in 1966. She received a Ph.D. degree of mathematical sciences from Vilnius University, Vilnius, Lithuania, in 1974. She is a senior researcher at the Department of Optimization, Institute of Mathematics and Informatics, Vilnius, Lithuania. Her research interests include stochastic processes and optimization problems.

PARAMETRŲ GRUPAVIMO, NAUDOJANT MODELIUOJAMĄ ATKAITINIMĄ, KONVERGAVIMAS

Gintautas DZEMYDA, Elvyra SENKIENĖ

Straipsnyje nagrinėjamas parametrų grupavimo remiantis jų koreliacine matrica uždavinys. Apžvelgtos parametrų grupavimo taikymo sritys ir galimos šio uždavinio sprendimo strategijos. Teoriškai ištirtas modeliuojamu atkaitinimu grindžiamų grupavimo strategijų konvergavimas.