INFORMATICA, 1997, Vol. 8, No. 2, 181–214

ON THE VISUAL ANALYSIS OF EXTREMAL PROBLEMS

Gintautas DZEMYDA

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

Abstract. The aim of investigation was to seek new ways for the analysis of extremal problems. A method of visual analysis of a set of objective function values is proposed. It allows us to find a direction where the variation of function is maximal. The method ensures a high quality of analysis when the number of used values of the objective function is small, and a possibility of identifying a specific character of the objective function. The results of analysis are used in search of a new coordinate system of the extremal problem and in a graphical representation of the observed data. The analysis will lead us to a better optimization strategy.

Key words: visual analysis, optimization, data analysis, interactive analysis.

1. Introduction. Extremal problems that arise in the design of technical systems may often be transformed into the form

$$\min_{X=(x_1,\dots,x_n)\in A^*} f(X),$$

where A^* is a bounded domain in an n-dimensional Euclidean space R^n, the objective function $f(X) : A^* \to R$ is continuous and multiextremal in the general case. In particular, A^* may be an n-dimensional rectangle $[A,B] = \{X : a_k \leqslant x_k \leqslant b_k,\ k=\overline{1,n},\ A=(a_1,\dots,a_n),\ B=(b_1,\dots,b_n)\}$.

Functions $f(X)$, occurring in practice, are often very complex, and sometimes it is difficult to solve the problem directly by classical methods. In such cases it is reasonable to analyse the problem. The analysis allows us to base optimization methods not only on functional characteristics of the objective function (linearity, convexity, etc.), but also on the information about the variation of function in various directions, relations among separate variables or their groups, the structure of a calculation process of the function value. Recent results of analysis of the set of the objective function values are presented in papers by Šaltenis and Dzemyda (1982), Šaltenis (1989), Dzemyda (1983, 1987). The analysis of extremal problems is also discussed by Dzemyda and Tiešis (1991).

In the paper by Dzemyda (1996b), a possibility of analysing visually the set of values of the objective function is determined. Here the ideas and results of Dzemyda (1996b) are extended and generalised.

The aim of analysis is to find a direction in the definition domain A^* such that

– maximizes the mean absolute difference between two values of the objective function calculated at randomly selected points in this direction (distribution of the points is uniform), or (and)

– maximizes the mean absolute difference per distance unit for the objective function values calculated at two randomly selected points in this direction (distribution of the points is uniform).

It would also be interesting to analyse cases, where the distribution of the points defined above is normal or the distance between these points is fixed. But these cases are out of the scope of this paper.

The quantities above are characteristics (not the only possible) of variation of the objective function. In the general case, the directions optimizing both characteristics of variation are not identical, and the investigator should have a possibility to choose one of them or integrate both the directions. Let us denote these directions by $\overline{Y}_1$ and $\overline{Y}_2$, respectively.

The directions defined above may be useful in developing new optimization algorithms. If we start the minimization at the randomly selected point $X_I \in A^*$, and execute a step of random length to the point $X_{II} \in A^*$, then the mean absolute change in the objective function value will be maximal, if both the points lie in the first direction defined above. The second direction defined above (and the first direction, to some extent as well) is an extension of the gradient concept: the gradient $\nabla f(X^*)$ points to the direction of the steepest slope of hyper-surface of $f(\cdot)$ at the point X^* (see Foulds, 1981).

2. Data sets. The initial information for the analysis is:

– points $X^i = (x_1^i, \ldots, x_n^i) \in A^*$, $i = \overline{1, m}$, $m \geqslant 2$, that form a discrete set $D \subset A^*$;

– values of $f(X)$ at these points.

The points $X^i = (x_1^i, \ldots, x_n^i)$, $i = \overline{1, m}$, may be selected at random from

A^* or in a deterministic manner. In the first case, these points may be considered as observations of a n-dimensional random variable $X = (x_1, \ldots, x_n)$ whose components $x_1, \ldots, x_n$ are independent and it has a uniform distribution in A^*. The methods of such an analysis are given in papers by Šaltenis and Dzemyda (1982), Dzemyda (1987), Šaltenis (1989). But sometimes it is advantageous to select points that are of some interest to the investigator (see Dzemyda, 1983).

The aim is to transform the initial information given above so that we could observe and analyse the variation of function in various directions. Two transformations η_1 and η_2 are proposed below.

Let any pair of points X^i, X^j, $i \neq j$, be taken from D with the same probability. In this case, we can observe random quantities $\eta_s = (\eta_{s1}, \ldots, \eta_{sn})$, $s = 1, 2$, and $\overline{f}$ whose values $\eta_s^{ij} = (\eta_{s1}^{ij}, \ldots, \eta_{sk}^{ij}, \ldots, \eta_{sn}^{ij})$, $s = 1, 2$, and f^{ij} are uniquely related with a randomly selected pair X^i, X^j, $i \neq j$, as follows:

$$\eta_{1k}^{ij} = (|f^i - f^j|)^\tau \cdot \frac{(x_k^i - x_k^j)}{S_{ij}},$$

$$\eta_{2k}^{ij} = \left(\frac{|f^i - f^j|}{S_{ij}}\right)^\tau \cdot \frac{(x_k^i - x_k^j)}{S_{ij}},$$

$$f^{ij} = f^i - f^j,$$

where $f^i = f(X^i)$, $S_{ij} = \sqrt{\sum\limits_{l=1}^{n}(x_l^i - x_l^j)^2}$, $\tau \in [-1, 1]$.

The transformations η_s, $s = 1, 2$, depend on τ. Therefore, in further formulae we sometimes use $\eta_s(0 < \tau < 1)$, $\eta_s(0.25)$, etc., where detailed values of τ are given in parenthesis.

Why did we introduce such transformations? Lengths of vectors $\eta_s^{ij} = (\eta_{s1}^{ij}, \ldots, \eta_{sn}^{ij})$, $s = 1, 2$, are as follows:

$$\left\|\eta_1^{ij}\right\| = (|f^i - f^j|)^\tau, \quad \left\|\eta_2^{ij}\right\| = \left(\frac{|f^i - f^j|}{S_{ij}}\right)^\tau.$$

Therefore, a longer distance of η_s^{ij} from the centre $(0, 0, \ldots, 0)$ corresponds to a greater 'variation of function', in the case of positive τ, and to a smaller 'variation of function', in the case of negative τ.

Examples of distributions of the values of $\eta_s = (\eta_{s1}, \eta_{s2})$, $s = 1, 2$, are presented graphically in Figures 1–4 for four functions dependent on two variables ($n = 2$):

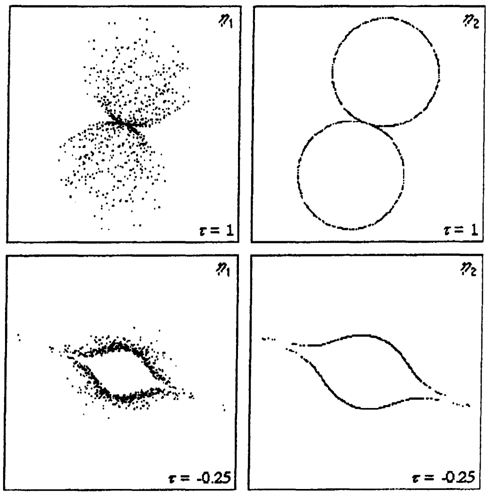

Fig. 1. Distributions of the values of η_1 and η_2 for the linear function f_1.

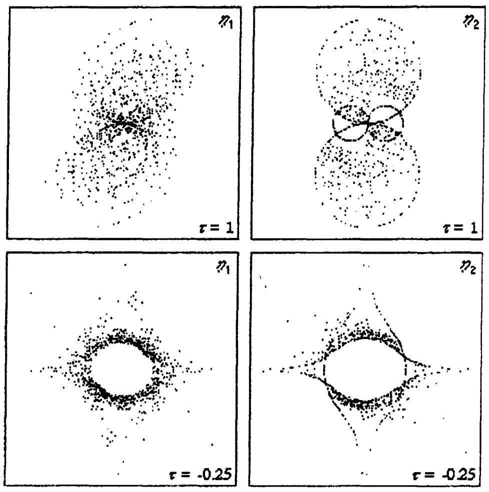

Fig. 2. Distributions of the values of η_1 and η_2 for the piecewise linear function f_2.

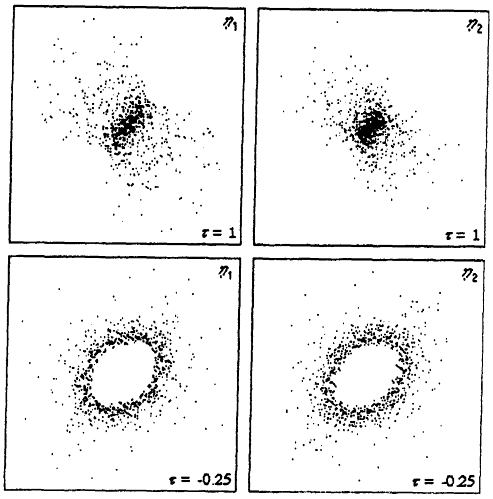

Fig. 3. Distributions of the values of η_1 and η_2 for the quadratic function f_3.

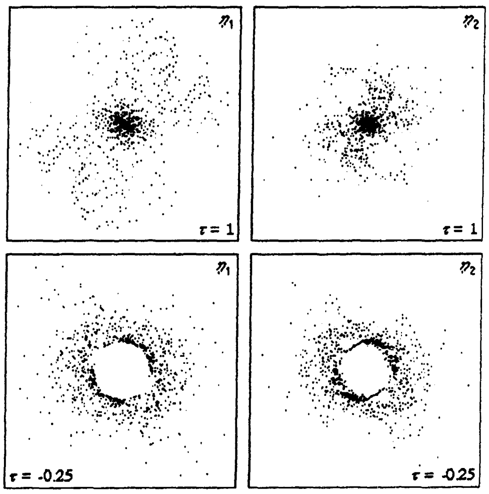

Fig. 4. Distributions of the values of η_1 and η_2 for the multiextremal function f_4.

1. Linear function: $f_1 = x_1 + 3x_2, \ -1 \leqslant x_1, x_2 \leqslant 1.$
2. Piecewise linear function: $f_2 = |x_1 - 3x_2| + x_1 + 3x_2,$
   $-1 \leqslant x_1, x_2 \leqslant 1.$
3. Quadratic function: $f_3 = (x_1 - x_2)^2 + [(x_1 + x_2)/2]^2, \ -1 \leqslant x_1, x_2 \leqslant 1.$
4. Multiextremal Branin's function (three local minima) (Dzemyda, 1985; Dixon and Szegö, 1978):
   $f_4 = (x_2 - 0.1292x_1^2 + 1.59155x_1 - 6) + 9.60211\cos(x_1) + 10,$
   $-5 \leqslant x_1 \leqslant 10, \ 0 \leqslant x_2 \leqslant 15.$

$m = 30$ points $X^i = (x_1^i, x_2^i), \ i = \overline{1,m}$, were generated randomly in the definition domain. The abscissa is meant for the values of η_{s1}, and the ordinate is meant for the values of $\eta_{s2}, \ s = 1, 2$. Surfaces of functions f_1, f_2, f_3, and f_4 are given in Fig. 5.

Data are scaled in Figures 1–4. Each figure consists of four pictures. The distance of the nearest point to the border of any picture is the same. This point has such a value of abscissa or ordinate: $\max\left|\eta_{sk}^{ij}\right|, \ i, j = \overline{1,m}, \ i \neq j,$ $k = 1, 2$. The centre of the picture corresponds to $(0, 0)$.

The distributions given in Figures 1–4 show that their analysis should lead to the detection of new peculiarities of the objective function.

2.1. Selecting a parameter τ. The analysis of distributions of the values of η_1, in the case of positive τ, allows us to search for a direction $\overline{Y}_1$ in the definition domain A^* that maximizes the mean absolute difference between two values of the objective function calculated at randomly selected points in this direction, and the analysis of distributions of the values of η_2, in the case of positive τ, allows us to search for a direction $\overline{Y}_2$ that maximizes the mean absolute difference per distance unit of the objective function values calculated at two randomly selected points in this direction.

It follows from Figures 1–4 that a lot of points concentrate in the centre of pictures in case $\tau = 1$. It means that, in case $\tau = 1$, a visual decision may often be influenced by the significantly smaller number of points located near the border of pictures. A natural problem arises: how to present the points which are located in the centre of the picture to the investigator? This may be done by varying the value of τ. The dependence of distributions of the values of η_1 and η_2 on τ is illustrated in Figures 6 and 7. We observe here that varying τ one can analyse functions better. A special attention should be paid to the analysis of cases where τ is negative.

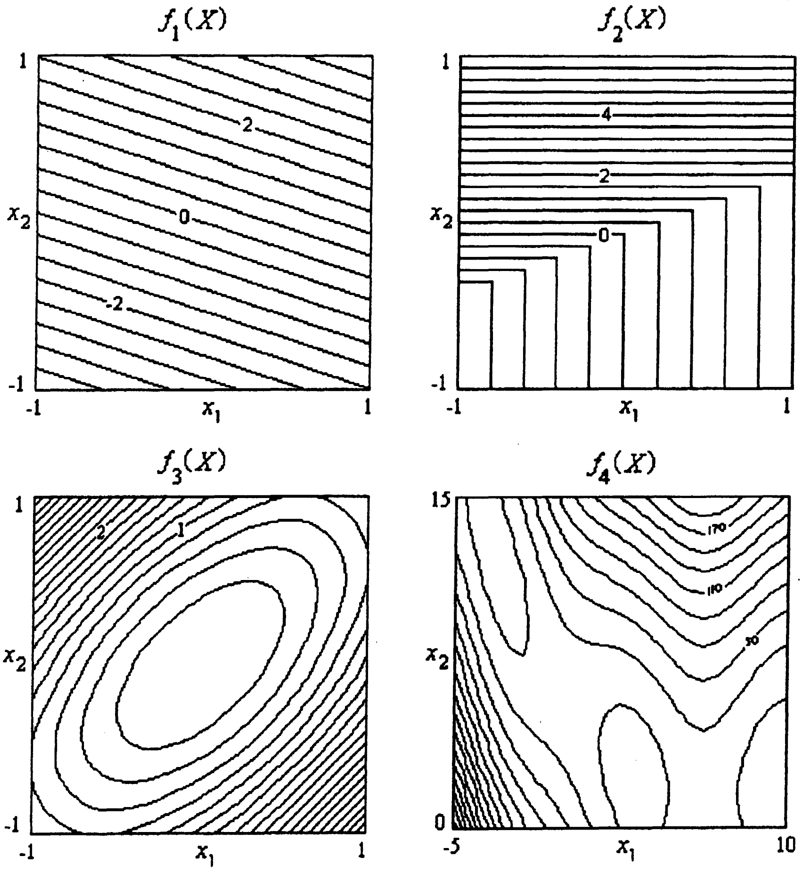

Fig. 5. Surfaces of functions f_1, f_2, f_3, and f_4.

2.2. Forming the data set D. Seeking representative data sets for the analysis of the extremal problem, the points $X^i = (x_1^i, \ldots, x_n^i)$, $i = \overline{1, m}$, should cover the definition domain A^* uniformly. Therefore, the points X^i, $i = \overline{1, m}$, may be

- selected at random in A^* (uniform distribution is essential);
- taken in a deterministic manner as
    - the nodes of the rectangular lattice (Sobolj, 1979),
    - the points of the LP-sequence (Sobolj, 1979).

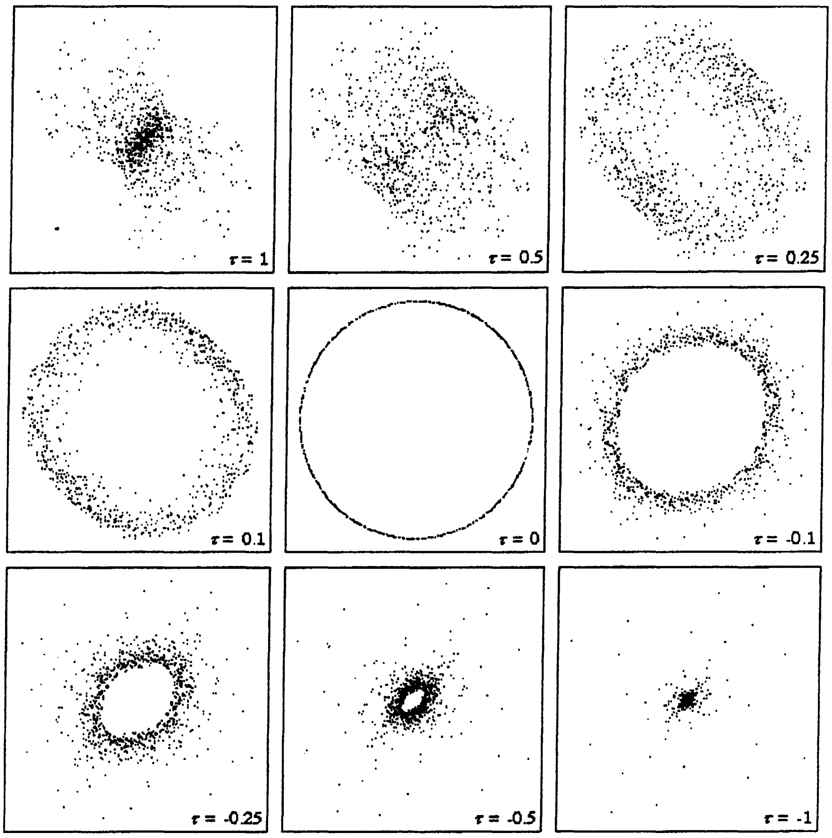

Fig. 6. Dependence of distributions of the values of η_1 on τ for the quadratic function f_3.

3. Covariance matrices. The formulae above implies that if set D contains m points X^i, $i = \overline{1,m}$, then random quantities $\overline{f}$ and $\eta_s = (\eta_{s1}, \dots, \eta_{sn})$, $s = 1, 2$, can acquire $m(m-1)$ different values. The mean values of $\overline{f}$ and η_{sk}, $s = 1, 2$, $k = \overline{1,n}$, are equal to 0, the variance of $\overline{f}$ is

$$D\overline{f} = \frac{2}{m(m-1)} \sum_{i=1}^{m} \sum_{j=i+1}^{m} (f^i - f^j)^2 = \frac{2}{(m-1)} \sum_{i=1}^{m} \left(f^i - \frac{1}{m} \sum_{j=1}^{m} f^j \right)^2,$$

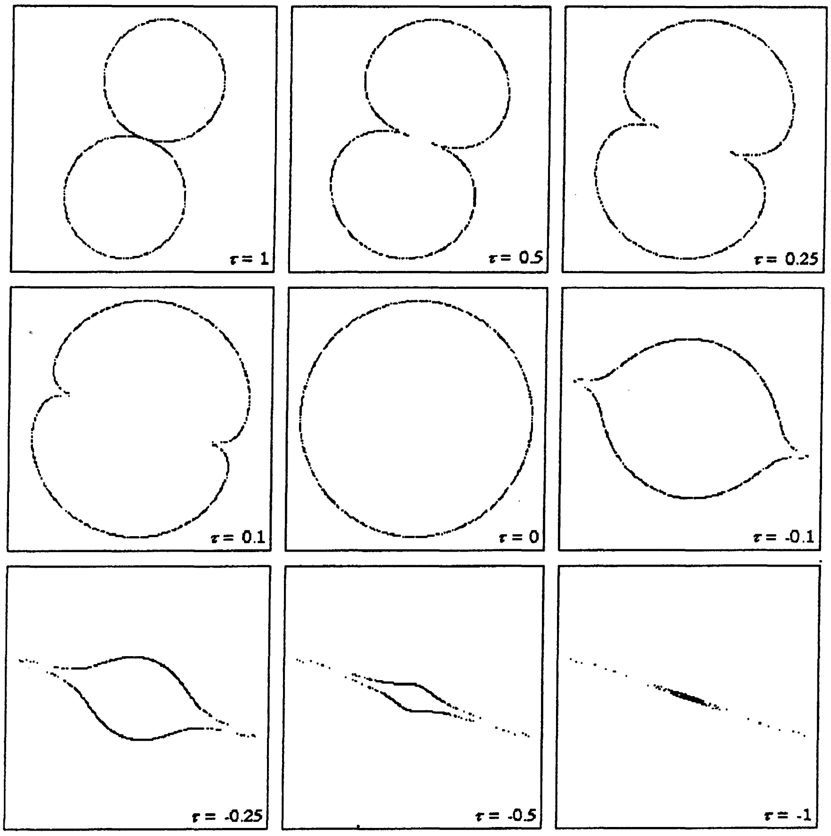

Fig. 7. Dependence of distributions of the values of η_2 on τ for the linear function f_1.

and covariance matrices $K_s = \{K_{\eta_{sk}\eta_{sl}},\ k,l = \overline{1,n}\}$ of $\eta_s = (\eta_{s1}, \dots, \eta_{sn})$, $s = 1,2$, are as follows:

$$K_{\eta_{1k}\eta_{1l}} = \frac{2}{m(m-1)} \sum_{i=1}^{m} \sum_{j=i+1}^{m} \left[(f^i - f^j)^{2\tau} \cdot \frac{(x_k^i - x_k^j)(x_l^i - x_l^j)}{S_{ij}^2} \right],$$

$$K_{\eta_{2k}\eta_{2l}} = \frac{2}{m(m-1)} \sum_{i=1}^{m} \sum_{j=i+1}^{m} \left[(f^i - f^j)^{2\tau} \cdot \frac{(x_k^i - x_k^j)(x_l^i - x_l^j)}{S_{ij}^{2(\tau+1)}} \right].$$

REMARK. The variances of random quantities $\eta_{11}, \dots, \eta_{1n}$, in case $\tau = 1$, are related with the estimate D^*f of the variance of $f(X)$, calculated on the basis of all m points X^i, $i = \overline{1,m}$, from D, as follows:

$$\sum_{k=1}^{n} K_{\eta_{1k}\eta_{1k}} = D\overline{f} = \frac{2}{(m-1)} \sum_{i=1}^{m} \left(f^i - \sum_{j=1}^{m} f^j \right)^2 = 2D^* f.$$

The covariance matrix K_1^* of the n-dimensional random quantity $\eta_1(1)\sqrt{2}$ was used by Šaltenis and Dzemyda (1982), Dzemyda (1987), Dzemyda (1983) for the analysis of extremal problems and for the construction of optimization algorithms. A multiplier $1/\sqrt{2}$ was introduced to simplify the further formulae only. It does not have any influence on eigen-vectors of the covariance matrix and on the proportionality among diagonal elements and among eigen-values of the covariance matrix. Namely, this proportionality was used in making a decision on the problem structure. So the covariance matrix K_1 of $\eta_1(1)$ may be used for the analysis like in papers by Šaltenis and Dzemyda (1982), Dzemyda (1983, 1987) as well. The covariance matrix K_1 was analysed by Šaltenis and Dzemyda (1982), Dzemyda (1983) using the factor analysis, and by Dzemyda (1985, 1987) using extremal parameter grouping (Dzemyda, 1988, 1996a). As a result of analysis, the efficiency of local optimization was increased and the extremal problem was simplified, in the first case, and the efficiency of global optimization (LP-search) was increased, in the second case. In the paper by Šaltenis and Dzemyda (1982), the dependence of optimization errors, occurring as a result of fixing separate variables, on the values of diagonal elements and eigen-values of K_1^* are investigated.

Sometimes m is sufficiently great and the values of function $f(X)$ are not calculated before the analysis starts. In such a case, sometimes it is impossible to calculate the covariance matrices $K_s = \{K_{\eta_{sk}\eta_{sl}}, k, l = \overline{1,n}\}$, $s = 1, 2$, analytically. In this case, the covariance matrices may be estimated using m_1 pairs of points. $m_1 < m(m-1)$ because the use of $m_1 \geqslant m(m-1)$ is senseless. Let us denote these pairs by $X^{1t} = (x_1^{1t}, \dots, x_n^{1t})$, $X^{2t} = (x_1^{2t}, \dots, x_n^{2t})$, where t is the number of pair, $t < m(m-1)$. In this case, $\eta_s = (\eta_{s1}, \dots, \eta_{sn})$, $s = 1, 2$, will take the following values: $\eta_s^t = (\eta_{s1}^t, \dots, \eta_{sn}^t)$, $t = \overline{1, m_1}$, $s = 1, 2$. Taking into account that the mean values of η_{sk}, $s = 1, 2$, $k = \overline{1,n}$, are equal to 0, not shifted estimates of covariance matrices of $\eta_s = (\eta_{s1}, \dots, \eta_{sn})$, $s = 1, 2$, are as follows:

$$\overline{K}_{\eta_{1k}\eta_{1l}} = \frac{1}{m_1} \sum_{t=1}^{m_1} \left[\left(f^{1t} - f^{2t} \right)^{2\tau} \cdot \frac{(x_k^{1t} - x_k^{2t})(x_l^{1t} - x_l^{2t})}{S_t^2} \right],$$

$$\overline{K}_{\eta_{2k}\eta_{2l}} = \frac{1}{m_1}\sum_{t=1}^{m_1}\left[\left(f^{1t}-f^{2t}\right)^{2\tau}\cdot\frac{(x_k^{1t}-x_k^{2t})(x_l^{1t}-x_l^{2t})}{S_t^{2(\tau+1)}}\right],$$

where $S_t = \sqrt{\sum_{l=1}^{n}(x_l^{1t}-x_l^{2t})^2}$ and $f^{ut} = f(X^{ut}),\ u = 1,2$.

4. Analysis. Three possibilities of the directional analysis of covariance matrices are discussed in this paper:

1. The factor analysis of covariance matrices of $\eta_s = (\eta_{s1},\dots,\eta_{sn})$, $s = 1,2$.
2. The interactive visual analysis of observations of $\eta_s = (\eta_{s1},\dots,\eta_{sn})$, $s = 1,2$.
3. Combination of the visual and the factor analysis.

The stress is put on the interactive visual analysis.

4.1. Application of factor analysis. The factor analysis allows us to find a linear combination $\vartheta_1^s = a_{11}^s\eta_{s1} + a_{12}^s\eta_{s2} + \dots + a_{1n}^s\eta_{sn}$ of $\eta_{s1},\dots,\eta_{sn}$, $s = 1,2$, having the maximal variance. The solution $a_1^s = (a_{11}^s,\dots,a_{1n}^s)$ is a normalized eigen-vector (i.e., a vector of unit length) corresponding to the maximal eigen-value of K_s. Therefore, applying factor analysis for the case $\tau > 0$, the criterion of 'variation of function' may be the maximal eigen-value of the covariance matrix. The eigen-vector corresponding to the eigen-value represents the desired direction. The factor analysis also allows us to find a second combination $\vartheta_2^s = a_{21}^s\eta_{s1} + a_{22}^s\eta_{s2} + \dots + a_{2n}^s\eta_{sn}$, where $a_2^s = (a_{21}^s,\dots,a_{2n}^s)$ is a normalized eigen-vector, corresponding to the second in size eigen-value of the covariance matrix. The variance of ϑ_2^s is maximal for any direction perpendicular to a_1^s. Thus, it is possible to find all n directions $a_k^s = (a_{k1}^s,\dots,a_{kn}^s)$, $k = \overline{1,n}$, that create a new coordinate system. The factor analysis may also be used in the case $\tau < 0$, but the first found direction $a_1^s = (a_{11}^s,\dots,a_{1n}^s)$ will show where 'variation of function' is the least.

Let the aim of analysis be to make a new coordinate system $Y' = UX'$:

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots \\ u_{n1} & u_{n2} & \dots & u_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

for the extremal problem, where $U = \{u_{ij},\ i,j = \overline{1,n}\}$ is a square matrix, Y' and X' are column vectors obtained from row vectors $Y = (y_1,\dots,y_n)$

and $X = (x_1, \dots, x_n)$. As a result of the factor analysis, the rows of matrix U would be normalized eigen-vectors $a_i^s = (a_{i1}^s, \dots, a_{in}^s)$, $i = \overline{1,n}$ of K_s. The characteristics of variables x_i, $i = \overline{1,n}$, are respective diagonal elements $K_{\eta_{sk}\eta_{sk}}$, $k = \overline{1,n}$, of K_s. The characteristics of variables y_i, $i = \overline{1,n}$, are respective eigen-values λ_i^s, $i = \overline{1,n}$, of K_s. If new coordinates are introduced according to the decreasing value of λ_i^s then the following inequalities are valid:

$$\lambda_1^s = \max_{k=\overline{1,n}} \lambda_k^s \geqslant \max_{k=\overline{1,n}} K_{\eta_{sk}\eta_{sk}},$$

$$\lambda_n^s = \min_{k=\overline{1,n}} \lambda_k^s \leqslant \min_{k=\overline{1,n}} K_{\eta_{sk}\eta_{sk}},$$

$$\sum_{k=1}^{n} \lambda_k^s = \sum_{k=1}^{n} K_{\eta_{sk}\eta_{sk}}.$$

$K_{\eta_{sk}\eta_{sk}}$, $k = \overline{1,n}$, and λ_i^s, $i = \overline{1,n}$, are related with the optimization errors occurring as a result of fixing separate variables (see Šaltenis and Dzemyda (1982) on experimental investigation of the covariance matrix K_1^* of the n-dimensional random quantity $\eta_1(1)/\sqrt{2}$).

The quality of analysis depends, to a great extent, on the number m of calculated values of $f(X)$. For instance, let us analyse $m = 50$ values of the function $f_5(X) = x_1 + 2x_2 + 3x_3 + 4x_4$ calculated at randomly selected and uniformly distributed points in $[A, B]$ where $A = (-1,-1,-1,-1)$ and $B = (1,1,1,1)$. These points form the discrete set D. The ideal direction when analysing the covariance matrix K_2 of $\eta_2(1)$ is $(1,2,3,4)$, which, as a result of normalization (division by $\sqrt{1+2^2+3^2+4^2}$), is as follows: $(0.1826, 0.3651, 0.5477, 0.7303)$. However, the obtained solutions are very sensitive on the set D. For example, the factor analysis of K_2 (in case $\tau = 1$) calculated for four different sets D gave solutions which are presented in the first four rows of Table 1. Δ is the distance between the point from Table 1 and the ideal solution. The last two rows show the averaged and normalized solutions after the analysis of K_2 obtained using 100 and 1000 randomly filled sets D, respectively. A similar sensitivity of the result to m is observed in the analysis of any other covariance matrix, too. This indicates that the analysis of covariance matrices gives poor results in the case of small m. However, even for small m such an analysis allows to find a new coordinate system, where (see Šaltenis and Dzemyda, 1982)

• the first variable is most essential (its fixing causes the greatest error as compared with any other variable from both systems),

• variables with larger order numbers are less significant (their fixing causes smaller errors), and the variables with the largest order numbers are often less significant than any variable from the old system,

• coordinate descent gives better results.

Table 1. Factor analysis of matrix K_2

K_2	
X	Δ
(0.1544, 0.3952, 0.6265, 0.6538)	0.117
(0.1770, 0.3311, 0.4199, 0.8263)	0.164
(0.1412, 0.4368, 0.5333, 0.7106)	0.086
(0.1806, 0.2602, 0.6578, 0.6833)	0.159
(0.1702, 0.3776, 0.5466, 0.7278)	0.053
(0.1726, 0.3622, 0.5553, 0.7285)	0.045

These advantages of the new coordinate system require to look for new and more effective analysis methods.

4.2. Interactive visual analysis. The main idea of such an analysis is to present the sets of observations of $\eta_s = (\eta_{s1}, \ldots, \eta_{sn})$, $s = 1, 2$, to the investigator graphically. The investigator makes a decision on the best direction. Therefore, the main direction (not the system of perpendicular directions) may be found in this manner only.

Sometimes it may be useful to analyse the distributions of η_1 and η_2 at the same time because the directions $\overline{Y}_1$ and $\overline{Y}_2$ are frequently often similar (see Section 7 and Fig. 8). In this case, the investigator's decision is influenced by a specific character of distributions of the values of both random quantities η_1 and η_2. Our experience shows that it would be better to present to the investigator the distributions of $\eta_1(\tau < 0)$ and $\eta_2(\tau < 0)$ rotated by the 90° angle, if he prefers to analyse the distributions of η_1 and η_2 for both positive and negative values of τ at the same time.

The algorithm of visual analysis:

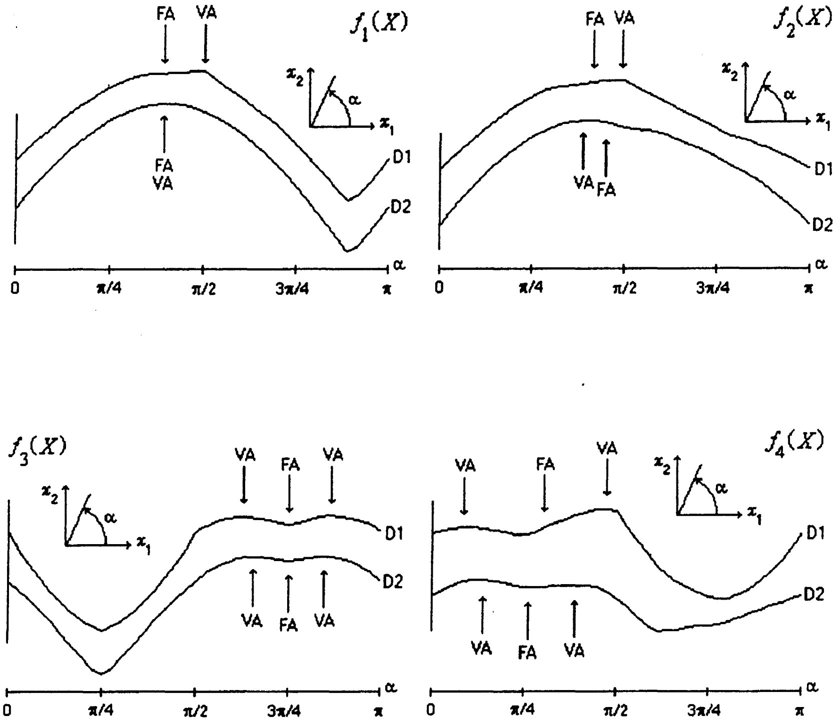

Fig. 8. Dependence of D1 and D2 on the angle α.

1. The pictures of distributions of $(\eta_{sk}^{ij}, \eta_{sl}^{ij})$, $i, j = \overline{1, m}$, $i \neq j$, are presented graphically to the investigator for any pair (k, l), $k, l = \overline{1, n}$, $k < l$, of variables in consecutive order.

2. The investigator analyses the distributions visually and shows the best, to his mind, direction $a_{kl}x_k - a_{lk}x_l = 0$ for any pair (k, l), $k, l = \overline{1, n}$, $k < l$, of variables. The decision depends on the goal of analysis and may be made on the basis of distributions of separate random quantities or groups of distributions (e.g., quantities η_s, $s = \overline{1, 2}$, for various values of τ).

3. The integral direction is determined by $n(n-1)/2$ subdirections

$$a_{kl}x_k - a_{lk}x_l = 0, \quad k, l = \overline{1, n}, \ k < l.$$

This approach has been investigated an compared with the factor analysis in the sections below.

4.3. Combination of the visual and the factor analysis. Various combinations of the visual and the factor analysis are possible. For example:

1. Factor analysis precedes the visual analysis: determination of direction using the factor analysis; an interactive visual analysis taking into account the results of the factor analysis.

2. Visual analysis precedes the factor analysis: an interactive visual analysis taking into account the results of the factor analysis; creation of a new coordinate system by finding other $n-1$ directions using the factor analysis.

3. Combination of the previous two strategies: determination of direction using the factor analysis; an interactive visual analysis taking into account the results of the factor analysis; creation of a new coordinate system by finding other $n-1$ directions using the factor analysis.

In the first case, the results of the factor analysis serve for the visual analysis as initial data.

The aim of analysis in the second case is to make a new coordinate system $Y' = UX'$ taking into account the results of visual analysis. As a result of visual analysis, let the direction $a = (\overline{a}_1, \ldots, \overline{a}_n)$ be found. $||a|| = 1$. Let us modify the covariance matrix K_s by adding a new matrix $K_a = \{\lambda \overline{a}_i \overline{a}_j,\ i, j = \overline{1, n}\}$, where $\lambda > 0$. A further examination of the matrix $K_a + K_s$ is performed using the factor analysis (see Section 4.1).

The matrix K_a has only one non-zero eigen-value λ. The vector a is a normalized eigen-vector corresponding to λ. The matrix $K_a + K_s$ is non-negative definite because $\lambda > 0$, and in the general case, the matrix K_s is non-negative definite. If the value of λ is chosen sufficiently high then

- the values of elements of the matrix K_s are insignificant in the search for the maximal eigen-value and the corresponding eigen-vector of the joint matrix $K_a + K_s$ because the absolute values of elements of the matrix $K_a + K_s$ become significantly greater than the absolute values of elements of the matrix K_s,
- the normalized eigen-vectors corresponding to the maximal eigen-value of both matrices K_a and $K_a + K_s$ are similar,
- in the case of identical normalized eigen-vectors corresponding to the maximal eigen-value of both matrices K_a and K_a+K_s, all the remaining eigen-values and respective eigen-vectors of the matrix $K_a + K_s$ are determined by the matrix K_s only.

The last proposition requires additional proof. Let

- μ_i be the i-th eigen-value of the matrix $K_a + K_s$: $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$;
- z^i be a normalized eigen-vector corresponding to μ_i.

Then

$$\mu_i = \max_{z=(z_1,\ldots,z_n)\in Z_i} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l \left(\lambda \bar{a}_k \bar{a}_l + K_{\eta_{sk}\eta_{sl}}\right)$$

$$= \max_{z=(z_1,\ldots,z_n)\in Z_i} \left\{\lambda\left(\sum_{k=1}^{n} z_k \bar{a}_k\right)^2 + \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l K_{\eta_{sk}\eta_{sl}}\right\},$$

where Z_i, $i = \overline{2,n}$, are sets of vectors of unit length in R^n perpendicular to z^k, $k = \overline{1,i-1}$, Z_1 is a set of all the vectors of unit length in R^n.

If the vectors z^1 and a are identical, then μ_i may be expressed as follows:

$$\mu_1 = \lambda + \max_{z=(z_1,\ldots,z_n)\in Z_1} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l K_{\eta_{sk}\eta_{sl}},$$

$$\mu_i = \max_{z=(z_1,\ldots,z_n)\in Z_i} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l K_{\eta_{sk}\eta_{sl}}, \quad i > 1,$$

i.e., vectors z^i, $i = \overline{2,n}$ are determined analysing the matrix K_s.

All normalized eigen-vectors of both matrices K_s and K_a+K_s are identical, if $a = (\bar{a}_1, \ldots, \bar{a}_n)$ and $a_1^s = (a_{11}^s, \ldots, a_{1n}^s)$ are identical. Slight differences between a and a_1^s cause small changes in the coordinate system defined by analysing K_s. The introduction of a new coordinate system is illustrated by the analysis of the following matrix:

$$K_s = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}.$$

New coordinates are shown in Tables 2–4. The results of factor analysis (without visual analysis) are given in Table 2. In Tables 3 and 4, we illustrate a combination of the visual and the factor analysis for two different vectors a.

Table 2. Results of the factor analysis

	x_1	x_2	x_3
y_1	.397	.521	.756
y_2	.233	.739	$-.632$
y_3	.888	$-.427$	$-.172$

Table 3. Results of a combined analysis. $a = (.415, .461, .784)$

	x_1	x_2	x_3
y_1	.415	.461	.784
y_2	.251	.770	$-.586$
y_3	.874	$-.440$	$-.204$

Table 4. Results of a combined analysis. $a = (.577, .577, .577)$

	x_1	x_2	x_3
y_1	.577	.577	.577
y_2	$-.211$	$-.577$	.789
y_3	.788	$-.578$	$-.212$

5. Problems of finding the integral direction. The direction in R^n is entirely defined if the coordinates of start and end points of the direction vector are known. The start point in our case is $X^0 = (0, \ldots, 0)$. The end point $X^* = (x_1^*, \ldots, x_n^*)$ must be determined from the system of $n(n-1)/2$ equations:

$$a_{kl}x_k = a_{lk}x_l, \quad k, l = \overline{1, n}, \ k < l, \tag{1}$$

where a_{kl} is the coefficient at x_k when x_k is in the same equation as x_l.

There are two equivalent solutions:

$X^* = (x_1^*, \ldots, x_n^*)$ and $(-1) \cdot X^* = (-x_1^*, \ldots, -x_n^*)$.

Let the distance between X^0 and X^* be equal to 1, i.e., $\sum\limits_{l=1}^{n} x_l^{*2} = 1$.

The problem is as follows:

$$a_{kl}x_k - a_{lk}x_l = 0, \quad k,l = \overline{1,n}, \; k < l, \tag{2}$$

$$\sum_{l=1}^{n} x_l^2 = 1. \tag{3}$$

The number of equations in the system of linear equations (2) starting from $n > 2$ is greater than the number of variables. However, it is necessary to find a solution. In this case, problem (2)–(3) may be formulated and solved as an optimization one using the least squares approach:

$$\min_{X \in R^n} \varphi(X) = \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} (a_{kl}x_k - a_{lk}x_l)^2 \tag{4}$$

subject to (3).

The Lagrangian (Reklaitis *et al.*, 1983) of the problem is

$$L(X,\nu) = \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} (a_{kl}x_k - a_{lk}x_l)^2 - \nu\left(\sum_{l=1}^{n} x_l^2 - 1\right).$$

The goal of analysis of $L(X,\nu)$ is to find $\nu \neq 0$ and X, where

- $\partial L/\partial x_k = 0, \quad k = \overline{1,n},$ (5)
- $\partial L/\partial \nu = 0,$ (6)
- the Hesse matrix $H(X,\nu) = \{h_{kl} = \partial^2 L/\partial x_k \partial x_l, \; k,l = \overline{1,n}\}$ is positive definite.

The obtained X value is the point of local minimum for $\varphi(\cdot)$ in this case.

Any stationary point of $L(X,\nu)$ satisfies system (5)–(6) of $n+1$ equations with $n+1$ unknowns.

The following $n-1$ parts of $\varphi(X)$ depend on x_k: $(a_{1k}x_1 - a_{k1}x_k)^2$, $(a_{2k}x_2 - a_{k2}x_k)^2, \ldots, (a_{k-1\,k}x_{k-1} - a_{k\,k-1}x_k)^2, (a_{k\,k+1}x_k - a_{k+1\,k}x_{k+1})^2, \ldots,$ $(a_{kn}x_k - a_{nk}x_n)^2$. Thus, equations (5) will be as follows:

$$\begin{aligned} &-2\sum_{l=1}^{k-1} (a_{lk}x_l - a_{kl}x_k)\, a_{kl} \\ &+2\sum_{l=k+1}^{n} (a_{kl}x_k - a_{lk}x_l)\, a_{kl} - 2\nu x_k = 0, \quad k = \overline{1,n}, \end{aligned}$$

and system (5)–(6) will consist of such equations:

$$\sum_{\substack{l=1\\ l\neq k}}^{n} (a_{kl}x_k - a_{lk}x_l)\,a_{kl} - \nu x_k = 0, \quad k = \overline{1,n}, \tag{7}$$

$$\sum_{l=1}^{n} x_l^2 - 1 = 0. \tag{8}$$

Let us analyse the matrix $H(X,\nu)$ and determine ν when the matrix is positive definite. $H(X,\nu)$ is positive definite, if and only if its eigen-values are positive. Therefore, a further investigation is oriented to the analysis of eigen-values of $H(X,\nu)$.

The elements of $H(X,\nu)$ are independent of X:

$$\partial^2 L/\partial x_k^2 = 2\sum_{\substack{l=1\\ k\neq l}}^{n} a_{kl}^2 - 2\nu, \tag{9}$$

$$\partial^2 L/\partial x_k \partial x_l = -2a_{lk}a_{kl}. \tag{10}$$

Let

– $B^* = \Big\{ b_{kk} = \sum_{\substack{i=1\\ i\neq k}}^{n} a_{ki}^2,\ k = \overline{1,n};$

$b_{kl} = -a_{lk}a_{kl},\ k,l = \overline{1,n},\ k \neq l \Big\}, \quad (11)$

– $\lambda_1, \ldots, \lambda_n$ be eigen-values of the matrix B^*,

– $\lambda_{\min}$ be the least eigen-value of B^*.

PROPOSITION 1. *$H(X,\nu)$ is positive definite, if and only if $\nu < \lambda_{\min}$.*

Proof. The Hesse matrix H of $L(X,\nu)$ is defined by (9) and (10). This matrix is a sum of two matrices: $H = 2(B^* + C)$, where B^* is defined by (11), and C is a diagonal matrix, all n diagonal elements of which are equal to $-\nu$, i.e.,

$$C = \left\{ c_{kk} = -\nu,\ k = \overline{1,n};\quad c_{kl} = 0,\ k,l = \overline{1,n},\ k \neq l \right\}.$$

Let

– μ_i be the i-th eigen-value of the matrix H: $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$;

– z^i be a normalized eigen-vector corresponding to μ_i;

- Z_1 be a set of vectors of unit length in R^n;
- Z_i, $i = \overline{2,n}$, be sets of vectors of unit length in R^n perpendicular to z^k, $k = \overline{1,i-1}$.

Then

$$
\begin{aligned}
\mu_i &= \max_{z=(z_1,\dots,z_n)\in Z_i} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l h_{kl} \\
&= 2 \max_{z=(z_1,\dots,z_n)\in Z_i} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l (b_{kl} + c_{kl}) \\
&= 2 \max_{z=(z_1,\dots,z_n)\in Z_i} \left\{ \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l b_{kl} + \sum_{k=1}^{n} z_{k}^{2} c_{kk} \right\} \\
&= 2 \max_{z=(z_1,\dots,z_n)\in Z_i} \left\{ \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l b_{kl} - \nu \right\} \\
&= 2 \max_{z=(z_1,\dots,z_n)\in Z_i} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l b_{kl} - 2\nu. \qquad (12)
\end{aligned}
$$

It follows from (12):

$$\arg \max_{z=(z_1,\dots,z_n)\in Z_i} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l h_{kl} = \arg \max_{z=(z_1,\dots,z_n)\in Z_i} \sum_{k=1}^{n}\sum_{l=1}^{n} z_k z_l b_{kl},$$

i.e., matrices H and B^* have the same eigen-vectors. Hence

$$\mu_i = 2\lambda_i - 2\nu. \qquad (13)$$

From (13), it follows that H is positive definite, if and only if $\lambda_i - \nu > 0$, $i = \overline{1,n}$, i.e., if $\nu < \lambda_{\min}$.

The proposition is proved.

$\lambda_{\min}$ may be successfully found because there are effective algorithms to search for it (see, e.g., Krylov *et al.*, 1976).

The next problem is to find a solution of the system of nonlinear equations (7)–(8). We suggest formulating an optimization least squares problem:

$$
\begin{aligned}
\min_{\substack{x_i\in[-1,1],\ i=\overline{1,n} \\ X\neq(0,0,\dots,0) \\ \nu<\lambda_{min}}} \gamma(x_1,\dots,x_n,\nu) &= \sum_{k=1}^{n} \left(\sum_{\substack{l=1 \\ l\neq k}}^{n} (a_{kl}x_k - a_{lk}x_l)\, a_{kl} - \nu x_k \right)^2 \\
&\quad + \left(\sum_{l=1}^{n} x_l^2 - 1 \right)^2. \qquad (14)
\end{aligned}
$$

Problem (14) may be simplified by analytical search for the optimal value of ν, when the values of other variables are fixed. From

$$\partial\gamma/\partial\nu = -2\sum_{k=1}^{n}\Bigg(\sum_{\substack{l=1\\ l\neq k}}^{n}(a_{kl}x_k - a_{lk}x_l)\,a_{kl} - \nu x_k\Bigg)x_k = 0$$

it follows

$$\nu = \frac{\sum\limits_{k=1}^{n} x_k \sum\limits_{\substack{l=1\\ l\neq k}}^{n}(a_{kl}x_k - a_{lk}x_l)\,a_{kl}}{\sum\limits_{l=1}^{n} x_l^2}. \tag{15}$$

From

$$\partial^2\gamma/\partial\nu^2 = \sum_{k=1}^{n} x_k^2 > 0$$

it follows that ν defined by (15) corresponds to the minimum of γ , when the values of $x_1,\ldots,x_n$ are fixed and $x_k \in [-1,1]$, $k=\overline{1,n}$, $X \neq (0,0,\ldots,0)$. Problem (14) may be transformed as follows:

$$\min_{\substack{x_k\in[-1,1],\, k=\overline{1,n}\\ X\neq(0,0,\ldots,0)}} \gamma^*(x_1,\ldots,x_n) = \sum_{k=1}^{n}\Bigg(\sum_{\substack{l=1\\ l\neq k}}^{n}(a_{kl}x_k - a_{lk}x_l)\,a_{kl} - \nu x_k\Bigg)^2 + \Bigg(\sum_{l=1}^{n} x_l^2 - 1\Bigg)^2 + \beta(\lambda_{\min}-\nu)^2, \tag{16}$$

where $\beta(\lambda_{\min}-\nu)^2$ is a penalty function,

$$\beta = \begin{cases} 0 & \text{if } \nu < \lambda_{\min}, \\ 1 & \text{if } \nu \geqslant \lambda_{\min}. \end{cases}$$

The starting point for local search may be $\overline{X} = (1\sqrt{n},\ldots,1/\sqrt{n})$. The value of γ^* is not defined at the point $X = (0,\ldots,0)$. So, it would be better to start optimization from the point where $\gamma^* < 1$. Then, one of two equivalent solutions will be found depending on the starting point. The variable metric algorithm (Tiešis, 1975; Dzemyda, 1985) gave good results in minimizing the function $\gamma^*(x_1,\ldots,x_n)$.

The investigations above showed that the analysis of problem (3)–(4) and application of the Lagrange method gave a possibility of finding efficient ways for solving the system of equations (2)–(3). This approach may be used to solve other similar optimization problems having restriction (3).

6. Experimental illustration of search for the integral direction. The first experiment was executed analysing the function $f_5(X) = x_1 + 2x_2 + 3x_3 + 4x_4$ calculated at $m = 50$ randomly selected and uniformly distributed points in D, where $A = (-1, -1, -1, -1)$ and $B = (1, 1, 1, 1)$. Distributions of the values of $\eta_1(1)$, $\eta_1(-0.25)$, $\eta_2(1)$, and $\eta_2(-0.25)$ were analysed visually. The matrix $\widehat{A} = \{a_{kl},\ k, l = \overline{1, n}\}$ obtained as a result of the visual analysis is given in Table 5. The aim is to find a solution of problem (16). The starting point $\overline{X} = (\overline{x}_1, \ldots, \overline{x}_4)$ and the corresponding value of $\gamma^*(\overline{X})$ are given in the first column of Table 6. The results of optimization ($\gamma^*(X^*)$ and $X^* = (x_1^*, \ldots, x_4^*)$) are given in the second column of Table 6. The respective values of $\varphi(X)$ (see (4)) at the points $\overline{X} = (\overline{x}_1, \ldots, \overline{x}_4)$ and $X^* = (x_1^*, \ldots, x_4^*)$ are given in Table 6, too. The visual analysis was facilitated by the known analytical expression of the function. This resulted in integer numbers in Table 5. If we have an unknown analytical expression, then the numbers in Table 5 would differ to some unessential extent only because this data set bears some features that predestine the decision. The function with the known structure is also a good test for the optimization algorithm from the section above.

Table 5. The matrix $\widehat{A}$ obtained during the visual analysis

k/l	1	2	3	4
1	–	2	3	4
2	1	–	3	4
3	1	2	–	4
4	1	2	3	–

The second experiment was performed on the function $f_6(X)$, $X = (x_1, \ldots, x_7)$, which is an objective function in the problem of computer-aided synthesis of the external circuit of the tunable subnanosecond pulse TRAPATT-generator (Dzemyda *et al.*, 1984, 1990; Dzemyda, 1993, 1995). It is often used as a test for comparison of the efficiency of optimization methods (Dzemyda *et al.*, 1990; Dzemyda, 1993, 1995). The number of variables is equal to 7. $x_k \in [0, 1]$, $k = \overline{1, 7}$. The function $f_6(x_1, \ldots, x_7)$ has two minima near zero.

Table 6. The results of optimization

$\gamma^*(X)$	15	$2.301 \cdot 10^{-11}$
$\varphi(X)$	5	$1.400 \cdot 10^{-13}$
x_1	0.5	0.1826
x_2	0.5	0.3651
x_3	0.5	0.5477
x_4	0.5	0.7303

The surface of the function $f_6(x_1, \ldots, x_7)$ is given in the paper by Dzemyda (1995) for the case, where the values of $x_1, \ldots, x_5$ are set to be equal to 0.5, and those of x_6 and x_7 are varied in [0,1].

The number m of points X^i, $i = \overline{1,m}$, was selected equal to 30, distributions of values of η_1 and η_2 were analysed visually in the case $\tau = -1, -0.25, 1$. The points X^i, $i = \overline{1,m}$, were generated randomly. The matrix $\widehat{A} = \{a_{kl},\ k,l = \overline{1,n}\}$ obtained as a result of visual analysis is given in Table 7. The starting point $\overline{X} = (\overline{x}_1, \ldots, \overline{x}_7)$ and the respective value of $\gamma^*(\overline{X})$ are given in the first column of Table 8. The results of optimization ($\gamma^*(X^*)$ and $X^* = (x_1^*, \ldots, x_7^*)$) are given in the second column of Table 8. The respective values of $\varphi(X)$ (see (4)) at the points $\overline{X} = (\overline{x}_1, \ldots, \overline{x}_7)$ and $X^* = (x_1^*, \ldots, x_7^*)$ are given in Table 8, too.

Table 7. The matrix $\widehat{A}$ obtained during the visual analysis

k	1	2	3	4	5	6	7
1	–	1.000	1.000	1.000	1.000	−1.000	−1.000
2	0.915	–	−1.000	1.000	−0.890	1.000	1.000
3	0.221	0.246	–	−1.000	0.459	−1.000	0.057
4	0.353	0.175	0.996	–	−1.000	1.000	0.822
5	0.674	1.000	1.000	0.825	–	−1.000	1.000
6	0.218	0.245	0.914	0.777	0.339	–	1.000
7	0.514	0.385	1.000	1.000	0.787	1.000	–

Table 8. The results of optimization

$\gamma^*(X)$	11.45	$3.106 \cdot 10^{-14}$
$\varphi(X)$	3.6783	0.5676
x_1	0.3780	−0.0451
x_2	0.3780	0.1128
x_3	0.3780	−0.6074
x_4	0.3780	0.4309
x_5	0.3780	−0.1693
x_6	0.3780	0.5860
x_7	0.3780	0.2422

7. Experimental investigation of functions dependent on two variables (n=2). Experiments were performed on functions f_1, f_2, f_3, and f_4. The mean absolute difference (D1) and the mean absolute difference per distance unit (D2) for these functions calculated at randomly selected pairs of points in the direction whose orientation defines the angle α are presented in Fig. 8. Each curve in Fig. 8 is drawn on the basis of the values of D1 or D2 at $\alpha = t \cdot \pi/36$, $t = \overline{0,35}$. Each value of D1 and D2 is determined on the basis of 500000 randomly selected pairs of (x_1, x_2).

Distributions of the values of $\eta_1(1)$ and $\eta_2(1)$ were investigated by the factor analysis (FA) and the visual analysis (VA). The final results of factor analysis have been obtained after averaging the results of analysis of 100 different data sets containing 300 randomly selected points in A^* $(m = 300)$. This was done with a view to find a sufficiently precise direction that tends to be indicated by the factor analysis. Pointers show the directions found as a result of the analysis.

From Fig. 8 it follows that both D1 and D2 have one maximum for f_1 and f_2, and two maxima for f_3 and f_4. The visual analysis, using different data sets containing 30 randomly selected points in A^* $(m = 30)$, indicated all of them by different respondents: decisions made by the respondents are located around the pointers denoted as VA. But the direction $\alpha = \pi/2$ was indicated in the analysis of distributions of the values of $\eta_2(1)$ for f_2, too. In this case, the investigator tries to recognize and take into account some regularity of the distribution.

The factor analysis yielded a direction that is an attempt to integrate both maxima of D1 and D2 for functions f_3 and f_4. In most cases of the analysis of distributions of the values of $\eta_1(1)$ and $\eta_2(1)$, the factor analysis indicated directions different from those corresponding to maxima of D1 and D2. Indeed, the factor analysis showed the global orientation of distributions of the values of $\eta_s(1)$ on the plane. The visual analysis of distributions of the values of $\eta_1(-0.25)$ and $\eta_2(-0.25)$ gave similar results to that of the factor analysis. It means that the visual analysis of distributions of the values of $\eta_1(\tau < 0)$ and $\eta_2(\tau < 0)$ also allows us to estimate the global orientation of distributions of the values of $\eta_1(\tau > 0)$ and $\eta_2(\tau > 0)$, respectively.

The investigations showed that in most cases it was better to use $0 < \tau < 1$ and $-1 < \tau < 0$ instead of $\tau = 1$ and $\tau = -1$, respectively. However, τ has to be sufficiently far from 0, e.g., $|\tau| = 0.5, 0.25, 0.1$.

8. Dependence of the distribution of values of η_1 and η_2 on the set of values of X. The experiments were carried out on the problem of computer-aided synthesis of the external circuit of the tunable subnanosecond pulse TRAPATT-generator (Dzemyda *et al.*, 1984, 1990; Dzemyda, 1993, 1995). The sets of values of η_1 were investigated. The number m of points $X^i = (x_1^i, \dots, x_n^i)$, $i = \overline{1,m}$, was selected equal to 70. The values of $x_2, \dots, x_5$ and x_7 were fixed and set to be equal to 0, and the values of x_1 and x_6 were generated randomly in $[0, 1]$ for the first experiment. In Fig. 9, six sets of all the possible values of $\eta_1(1)$ and $\eta_1(-0.1)$ for six different sets of points X^i, $i = \overline{1,m}$ are presented graphically. We observe a consequent evolution from the pictures of the first row to the pictures of the last one. Indeed, the pictures occurred in random order, because the sets of points X^i, $i = \overline{1,m}$, were generated at random.

The second experiment illustrates a situation where the values of $x_3, \dots, x_7$ are fixed and equal to 0, and the values of x_1 and x_2 are varied in $[0, 1]$. In Fig. 10, eight sets of all the possible values of $\eta_1(1)$ for eight different sets of points X^i, $i = \overline{1,m}$ is presented. Here we also have a possibility to construct a sequence of pictures with a consecutively evolving view.

In Figures 9 and 10, the results for η_1 are presented, but the same evolution of pictures is observed for η_2, too.

The experimental results suggest an idea, that any set of m points $X^i = (x_1^i, \dots, x_n^i)$, $i = \overline{1,m}$, may be related with the picture of the consecutively evolving sequence. Thus we can introduce a visual measure of similarity of sets. We have a possibility to compare the pictures interactively,

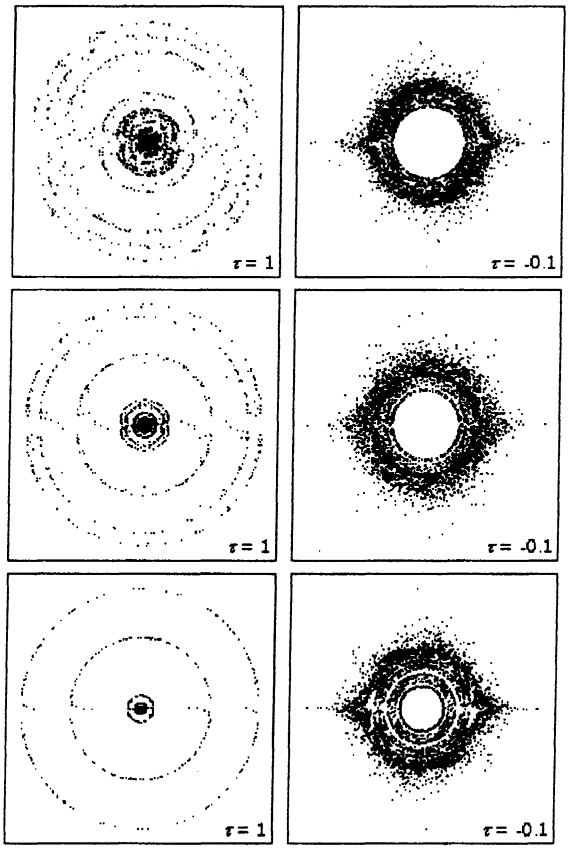

Fig. 9. $\eta_1(1)$ and $\eta_1(-0.1)$ for six different sets of points X^i, $i = \overline{1,m}$ (to be continued).

i.e., we can answer "which picture of the two follows next?". The pictures may be ranked using specialized decision support methods, e.g., the paired comparisons method, introduced by Saaty (1980; 1982) and included into the decision support computer system (Dzemyda and Šaltenis, 1994). It means that we can rank and compare the sets of m points X^i, $i = \overline{1,m}$. A visual human criterion may be used here. Further investigations require to answer many interesting questions. For instance,

- what is a formal (not visual-human) similarity criterion of the sets,
- how to compare the sets visually when $n > 2$; whether any two-dimensional projections may be used?

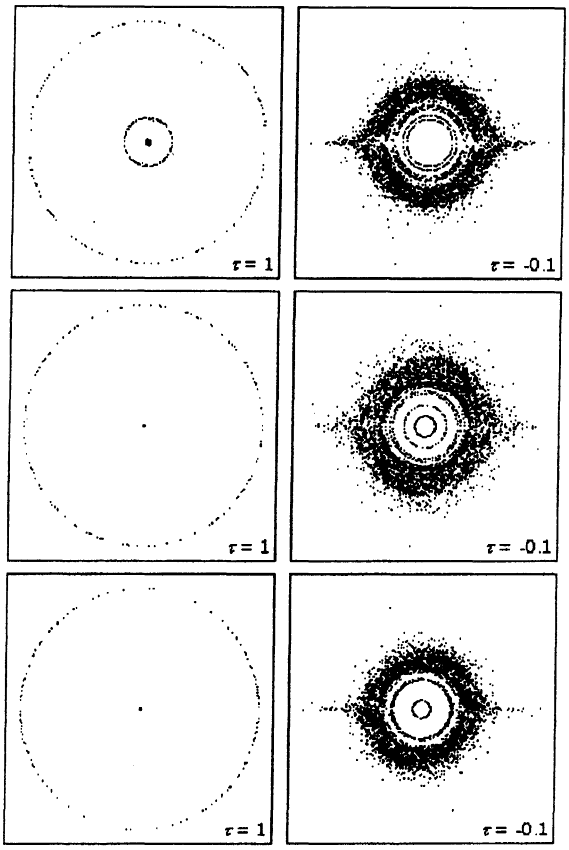

Fig. 9 (Continuation). $\eta_1(1)$ and $\eta_1(-0.1)$ for six different sets of points X^i, $i = \overline{1, m}$.

We observe a tendency of dependence of the pictures of the sequence on variance $D^* f$, calculated on the basis of m points X^i, $i = \overline{1, m}$. The pictures from Fig. 10 are put in an increasing order of $D^* f$, and only one pair of pictures violates the monotony of $D^* f$ in Fig. 9.

We also see that such an evolution of views may be well detected in the case of functions whose variance $D^* f$ varies in a wide range for different sets of points X^i, $i = \overline{1, m}$. Such functions often appear in practice. The problem of computer-aided synthesis of the external circuit of the TRAPATT-generator possesses such a property, too. Function $f_6(X)$ is positive for any

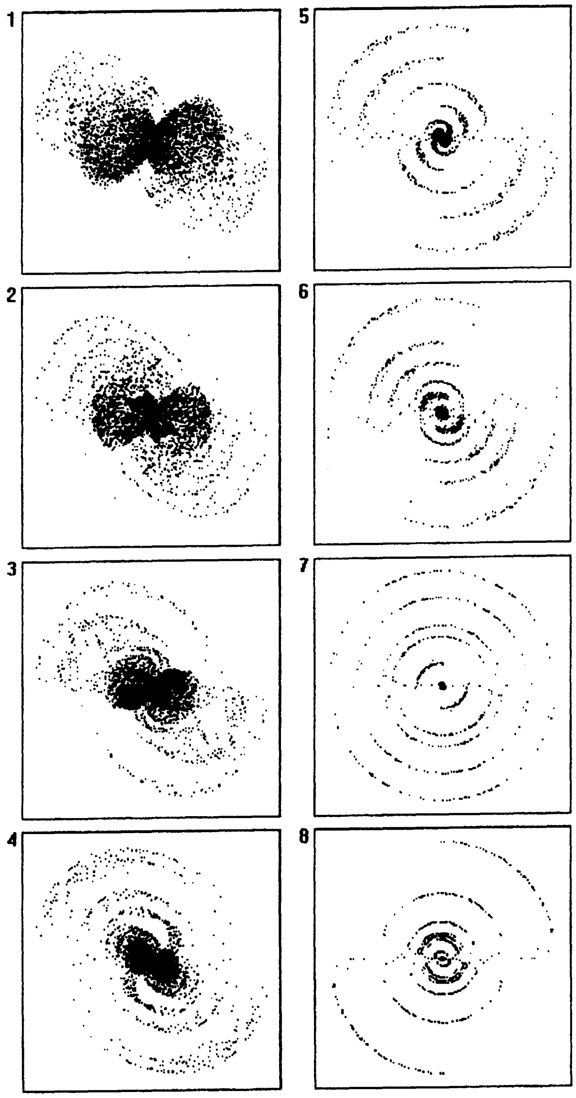

Fig. 10. $\eta_1(1)$ **for eight different sets of points** X^i, $i = \overline{1, m}$.

X and its values varies in a wide range. Therefore, it is possible to analyse the transformation $\lg[f_6(X)]$. In this case, the evolution of views is not so impressive, however the decision on the new coordinates is more exact.

9. Application of new coordinates in the optimization. Fig. 11 illustrates the introduction of a new coordinate system and its application in the optimization. The extremal problem, containing Branin's function $f_4(x_1, x_2)$, has been analysed. The problem was simplified by introducing a new coordinate system $Y = (y_1, y_2)$ and by fixing the variable y_2. The new problem has only one variable y_1. This variable is entirely determined by the point (x_1^*, x_2^*) and the angle α (for α see Fig. 11). Let us denote the objective function of the simplified problem by $\widehat{f}_4(y_1)$. 100000 different points (x_1^*, x_2^*) were generated at random in the definition domain A^*. The values of f_4 were calculated at five different points y_1^i, $i = \overline{1,5}$, for each (x_1^*, x_2^*): $\widehat{f}_4(y_1^1) = f_4(x_1^*, x_2^*)$ and the remaining four points y_1^i, $i = \overline{2,5}$, were selected at random in the definition domain. Thus, some steps of random search for the minimum of the function, dependent on a single variable, have been made. The results were averaged by all the 100000 searches. The averaged minimal value $f_4^{\min}$ is presented in Fig. 11 dependent on α. The pointers repeat the results of visual and factor analysis from Fig. 8. The results of Fig. 11 prove the possibility of applying the analysis presented in this paper.

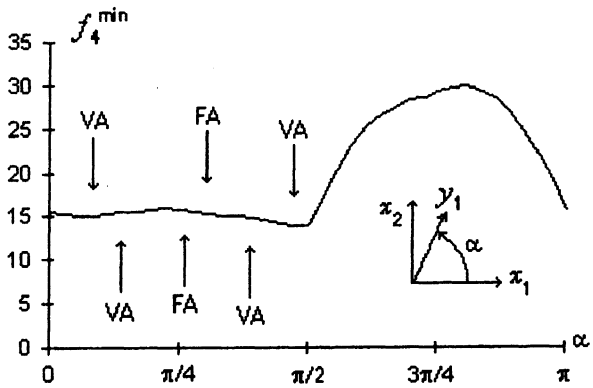

Fig. 11. Averaged results of minimization.

10. Conclusions. The results presented in this paper make the basis for new directions in the analysis of extremal problems. The method of visual analysis of the set of objective function values has been proposed. It ensures a high quality of analysis when the used number of function values is small, a minor dependence of results on the analysed set of function values, and a possibility of identifying a specific character of the function. The results of analysis may be used in creating a new coordinate system of the extremal problem and in the graphical representation of the observed data.

The efficiency of analysis depends on the software abilities. The WINDOWS-oriented application with a mouse managed choice of the best direction has been made. It simplifies the visual analysis of data sets.

When the investigator makes a decision on the new coordinates, he can better evaluate or predict the evolution of a picture and make some approximations. Fig. 9 and 10 illustrate such an evolution. The investigator often has an opportunity to observe some different data sets for the same function. This also increases the quality of final analysis. The factor analysis gives a new direction taking into account only one concrete data set. In this case, any new point may essentially influence the new direction.

Let a set of points X^i, $i = \overline{1,m}$, be generated at random. The variance $D^* f$ depends on this set. The search for new coordinates (both visually and using the factor analysis) will be more effective in the case of slight dependence of $D^* f$ on different sets of points X^i, $i = \overline{1,m}$. Therefore, such a dependence may be reduced by using greater m. A sufficient value of m depends on the complexity of the function $f(X)$. For example, $m = 30$ is sufficient for the functions f_1, f_2, f_3, and f_4, but insufficient for f_6. Also, the reduction of variance may sometimes be achieved by using, e.g., logarithmic transformations of the objective function.

The factor analysis may be applied in search of directions just like the visual analysis. However, the factor analysis uses another criterion of 'changes of function', and its results are often different as compared with the visual analysis. In a global sense however, we can observe similarities of results obtained by both these approaches (see Fig. 8).

Further investigations may be directed to the extension of areas of application of the proposed analysis in the optimization.

Acknowledgment. The author wishes to thank his colleague Vytautas Tiešis for many detailed and helpful comments.

REFERENCES

Dixon, L.C.W., and G.P. Szegö (1978). The global optimization problem: an introduction. In L.C.W.Dixon and G.P.Szegö (Eds.), *Towards Global Optimization 2.* North-Holland, Amsterdam. pp. 1–15.

Dzemyda, G. (1983). Investigation of the method of structural analysis of extremal problems and its application in local optimization. In A. Žilinskas and G. Dzemyda (Eds.), *Teorija Optimaljnych Reshenij,* Vol. 9. Inst. Math. Cybern., Vilnius. pp. 21–38 (in Russian).

Dzemyda, G., *et al.* (1984). Solution of the problems of optimal design and selection of model parameters' values using the package of applied programs MINIMUM. In A. Žilinskas (Ed.), *Teorija Optimaljnych Reshenij,* Vol. 10. Inst. Math. Cybern., Vilnius. pp. 77–98 (in Russian).

Dzemyda, G. (Ed.) (1985). *The Package of Applied Programs for Dialogue Solving of Multiextremal Problems MINIMUM: A Description of Using.* The State Fund of Algorithms and Programs (reg.No50860000112), Inst.Math.Cybern., Vilnius. 41 pp. (in Russian).

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,* Vol. 2. Pergamon Press, Oxford (Wheaton, Exeter)–New York–Toronto. pp. 499–502.

Dzemyda, G. (1988). The algorithms of extremal parameter grouping. In A. Sydow, S.G. Tzafestas and R. Vichnevetsky (Eds.), *Mathematical Research,* Band 46. Akademie-Verlag, Berlin. pp. 133–136.

Dzemyda, G. (1993). Investigation of the multiextremal problem. In *16th IFIP Conference on System Modelling and Optimization, Abstracts,* Vol. 2. INRIA, Compiegne (France). pp. 893–895.

Dzemyda, G. (1995). Multiextremal problem of computer-aided design. *Informatica,* **6**(3), 249–263.

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

Dzemyda, G. (1996b). Visual analysis of a set of function values. In *13-th International Conference on Pattern Recognition, Pattern Recognition and Signal Analysis,* Vol. 2, Track B. IEEE Computer Society Press, Los Alamitos, California. pp. 700–704.

Dzemyda, G., and V. Šaltenis (1994). Multiple criteria decision support system: methods, user's interface and applications. *Informatica,* **5**(1–2), 31–42.

Dzemyda, G., and V. Tiešis (1991). On the use of coordinated calculations in the solution of extremal problems. *Informatica,* **2**(2), 171–194.

Dzemyda, G., J. Valevičienė and L. Lukšan (1990). Investigation of the TRAPATT-generator design problem. In A. Žilinskas (Ed.), *Teorija Optimaljnych Reshenij,* Vol. 14. Inst. Math. Cybern., Vilnius. pp. 9–21 (in Russian).

Foulds, L.R. (1981). *Optimization Techniques. An Introduction.* Springer–Verlag, New York.

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

Reklaitis, G.V., A. Ravidran and K.M.Ragsdell (1983). *Engineering Optimization. Methods and Applications.* John Wiley and Sons, New York.

Saaty, T.L. (1980). *The Analytic Hierarchy Process.* McGraw-Hill, New York.

Saaty, T.L. (1982). *Decision Making for Leaders.* Lifetime Learning Publications, Belmont.

Sobolj, I.M. (1979). On systematic search in a hypercube. *SIAM J. Numer.Anal.,* **16**, 790–793.

Šaltenis, V. (1989). *Structure Analysis of Optimization Problems.* Mokslas, Vilnius (in Russian).

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).

Tiešis, V. (1975). The variable metric method for local minimization of multivariate functions subject to simple bounds. In R. Chomskis (Ed.), *Vyczislitelnaja Technika, Proceedings of the Conference.* Kaunas Polytechnic Institute, Kaunas. pp. 111–114 (in Russian).

Received March 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.

OPTIMIZAVIMO UŽDAVINIŲ VIZUALI ANALIZĖ

Gintautas DZEMYDA

Tyrimo tikslas yra naujų optimizavimo uždavinio analizės būdų paieška. Pasiūlytas tikslo funkcijos reikšmių aibės vizualios analizės metodas. Jis įgalina rasti kryptį, kur funkcijos kaita yra didžiausia. Pasiūlytas metodas užtikrina gerą analizės kokybę, kuomet turime nedaug tikslo funkcijos stebėjimų, ir galimybę identifikuoti tikslo funkcijos specifiką. Analizės rezultatai yra naudojami kuriant naują koordinačių sistemą.