ISSN 1064-2269, Journal of Communications Technology and Electronics, 2016, Vol. 61, No. 6, pp. 688–694. © Pleiades Publishing, Inc., 2016.
Original Russian Text © A.A. Bedrintsev, V.V. Chepyzhov, 2015, published in Informatsionnye Protsessy, 2015, Vol. 15, No. 4, pp. 402–413.

Description of the Design Space by Extremal Ellipsoids in Data Representation Problems

A. A. Bedrintsev and V. V. Chepyzhov

Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Bol'shoi Karetnyi per. 19, str. 1, Moscow, 127051 Russia

e-mail: chep@iitp.ru, alekseybed@phystech.edu

Received November 10, 2015

Abstract—Problems of description of data sets by constructing the optimum ellipsoid are considered. The optimization problems are formulated as convex programming problems using linear matrix inequalities. The proposed methods are compared with similar methods designed earlier in accordance with two criteria: the volume of the ellipsoid and the number of points in the learning sample, which lie outside the ellipsoid.

Keywords: extremal ellipsoids, design space, linear matrix inequalities, convex programming

DOI: 10.1134/S1064226916060048

1. INTRODUCTION

Currently, the mathematical simulation and data analysis are main methods for the design of complex technical objects. The object is parametrized by multivariate vector $x \in \mathbf{R}^d$. The components of vector x include geometrical and physical characteristics, parameters of the environment, and parameters of its functioning. Some characteristic $Y = F(x)$ of the object is considered and the object is said to be optimal if its characteristic $Y(x)$ takes the maximum or the minimum value among all other admissible objects x. In other words, the problem of selection of the best technical solution can be formulated as the problem of optimization of some function $F(x)$ on set of object x.

To formulate the optimization problem, it is necessary to specify constraints on vector x. It is evident that not all d-tuples of real numbers correspond to a correct and physically sensible object. For values of some parameters, engineers can indicate intervals of admissible values known from the object domain. Thus specified domain is the parallelepiped $\Pi = \{x \in \mathbf{R}^d | l^i \le x^i \le u^i, i = 1 \dots d\}$, which may contain points far from the known vectors x describing physically correct objects. It is very likely that all these points will not correspond to correct objects.

The situation when some coordinates of vector x do not have independent meaning and sense is rather frequent. Therefore, it is difficult to specify informative constraints for them. Vector x can include detailed description of the object surface, which consists of the coordinates of the points on the grid imposed on the object surface. These descriptions are widely used in CAD systems, programs for calculation of aerodynamic characteristics and visualization.

In the process of design, the engineers conducting experiments collect data bases of digital descriptions of objects and parameters of experiments $X = \{x_i \in \mathbf{R}^d, i = 1 \dots N\}$. Vectors from set X describe real objects. It is desirable to construct description of the data domain (design space) with correct vectors on the basis of known vectors from set X. These domains should possess the following properties. First, they should contain as many as possible vectors from X. Otherwise, important information on a great number of correct objects will be lost. Second, the design space should have a small volume so as to exclude points that are distant from the points from X and, with a high probability, do not correspond to a physically sensible object.

The constructed domain can be used as constraints on values of x in optimization problems. This causes several additional requirements imposed on the domain. Convex optimization is highly developed and convex programming problems have important properties, such as existence and uniqueness of the global minimum, and the local minimum is global [1]. Efficient algorithms have been developed for solution of convex problems and their high-quality implementations in many programming languages are available. In order to make the problem convex, it is necessary to ensure that the limitations determine a convex design space. Therefore, in this study, a convex description of data set X will be constructed. The simplicity of description of the domain and easiness of generation of random points in it are also important. In the paper, we propose to look for the description of the design space in the form of an ellipsoid [2, 3].

Another application of the description of the data domain is the outlier detection problem. Outliers are the points that differ substantially from ordinary points. The description of set X of ordinary, i.e., normal points is constructed using a small-volume ellipsoid. If a new point belongs to it, it is classified as ordinary. Otherwise, the conclusion is made that this is an outlier.

Usually, anomalies are very rare and the learning samples for outlier detection problems contain a small number of negative examples. In a similar problem of novelty detection, learning samples contain only ordinary points [4]. Thus, it is necessary to construct a geometrical body, containing a great number of points from the learning sample in order to minimize the probability of false alarm errors. However, this body should have a small volume since, otherwise, the outliers will be often recognized as ordinary points. The anomaly and novelty detection problems are widely used in statistics, model-constructing problems for the credit scoring, and computer-aided detection of fraudulent activities [5, 6].

Study [7] presents one more application of minimum-volume ellipsoids. It has been proved that the center of the minimum-volume ellipsoid containing more than a half of the points of set X is a stable affine-invariant location estimation. Papers [8, 9] consider other properties of this estimation, namely, rate of convergence, consistency and continuity with respect to the distribution used for generation of set X. The main drawback of this estimation is a high computational complexity of exact calculation. In [10], an approximate method for approximation of the location estimation using the minimum ellipsoid is described. It is reasonably complex in small-dimension spaces.

This paper is organized as follows. Section 2 contains formal formulation of the problem and the proof of existence of an exact solution with description of a naive but resource-capacious algorithm. Section 3 presents a survey of two known methods for obtaining an approximate solution to the problem. In Section 4, we propose two generalizations of known methods. It will be shown that some properties of the known methods are preserved during the generalization. This allows one to efficiently use new methods in practical simulation. Section 5 contains the results of numerical experiments on artificial data sets with different statistical characteristics. The final Section 6 summarizes the results of the study.

2. FORMULATION OF THE PROBLEM

Let us have sample $X = \{x_i \in \mathbf{R}^d, i = 1...N\}$ of vectors from a d-dimensional space. Let E be an ellipsoid that we should construct for description of set X. The ellipsoid can be specified using its center $a \in \mathbf{R}^d$, a $d \times d$ symmetrical positively defined matrix $P = P^T > 0$, and a squared effective radius $R \geq 0$ in the form of the following inequality:

$$E = \{x \in \mathbf{R}^d | (x-a)^T P^{-1} (x-a) \leq R\}. \quad (1)$$

Let us denote the volume of ellipsoid E by Vol(E). Its value is calculated by the formula

$$\mathrm{Vol}(E) = w_d \sqrt{\det P} R^{d/2}, \quad (2)$$

where $w_d = \pi^{d/2}/\Gamma(d/2 + 1)$ is the volume of the a unit ball in space $\mathbf{R}^d$.

Let us introduce a discrete function of the number of points from set X that do not belong to ellipsoid E:

$$K(E) = \#\{x \in X | x \notin E\}. \quad (3)$$

Let us pose the following bi-objective optimization problem of the search for a small-volume ellipsoid containing many points from the specified set X:

$$\min_E (\mathrm{Vol}(E), K(E)). \quad (4)$$

It is impossible to decrease to zero the ellipsoid volume, without leaving beyond its bounds virtually all points from learning set X. Therefore, a solution to the problem will be a set of Pareto optimal ellipsoids. This is the set of ellipsoids that are not dominated by any admissble and not Pareto optimal ellipsoid. It is said that ellipsoid E_1 dominates ellipsoid E_2 if and only if $K(E_1) \leq K(E_2)$ and $\mathrm{Vol}(E_1) \leq \mathrm{Vol}(E_2)$ and if at least one of these inequalities is strictly met. The Pareto frontier of the problem is formed by pairs of volume (2) and number (3) of outliers among points of the learning data set calculated for Pareto optimal ellipsoids. A particular ellipsoid is selected by the expert from the Pareto optimal set, depending on the demands of a particular data analysis problem.

The following theorem establishes the fundamental solvability of problem (4).

Theorem 1. *The Pareto frontier and the set of Pareto optimal ellipsoids for problem (4) exist and can be determined.*

Proof. Looking through all subsets $G \subseteq X$, solving problems

$$\begin{aligned} &\min_E (\mathrm{Vol}(E)) \\ &s.t.\ G \subset E \end{aligned} \quad (5)$$

and selecting a set of nondominated ellipsoids among solutions of (5), it is possible, evidently, to solve problem (4). Note that the constraint $G \subset E$ and objective function Vol(E) can be expressed in the convex form [1]. Hence, each problem (5) has a unique solution if the convex hull of set G has positive volume. To find the Pareto frontier, it is necessary to select a minimum-volume ellipsoid containing exactly j points from X for each value $j = 1...N$, i.e., the minimum of the bounded from below set (since the volume is not negative).

It is possible that, for some point (V, K) on the Pareto frontier, there are two (or more) different optimal ellipsoids $E_1 \neq E_2$ for which $\mathrm{Vol}(E_1) = \mathrm{Vol}(E_2)$ and $K(E_1) = K(E_2)$. Since they do not dominate each other they both are included into the set of Pareto optimal ellipsoids of problem (4).

3. KNOWN APPROXIMATE METHODS

The method used in the proof of Theorem 1 is unrealizable in practice. It requires solution of about 2^N (number of subsets $G \subseteq X$) problems (5). In this paper, methods for approximate solution of problem (4) are described.

We can replace the ellipsoid volume by any monotonically increasing function $\phi(\mathrm{Vol}(E))$. The following optimization problem is equivalent to problem (4):

$$\min_{E}(\phi(\mathrm{Vol}(E)), K(E)).$$

The authors of the papers that will be mentioned below in this section do not solve problem (4). They replace discrete objective function $K(E)$ by other continuous convex function, e.g., by a sum of quantities ξ_i, where nonnegative quantity $\xi_i \geq 0$ is the measure of remoteness of point x_i from the ellipsoid. This measure of remoteness can be selected by different methods, depending on the problem conditions. If $x_i \in E$, then $\xi_i = 0$. By denoting the vector of all ξ_i by $\xi = \{\xi_i\}_{i=1}^{N}$, we obtain the following formulation instead of problem (4):

$$\min_{E, \xi}\left(\phi(\mathrm{Vol}(E)), \sum_{i=1}^{N}\xi_i\right). \quad (6)$$

The scalarization method is a generally accepted method for solving multicriterion optimization problems [1]. For any nonnegative constant $C \geq 0$, we consider the following problem with one objective function:

$$\min_{E, \xi}\phi\mathrm{Vol}(E) + \sum_{i=1}^{N}\xi_i.$$

For convex problems (if all objective functions are convex and the constraints specify a convex set in space ξ and ellipsoid parameters P, a, R), by changing the scalarization parameter C in the range $[0; \infty]$, it is possible to obtain all Pareto optimal ellipsoids [1]. For nonconvex problems, scalarization gives a subset of the Pareto frontier of problem (6).

For a fixed matrix P of the ellipsoid or a fixed value of determinant $\det P$, the volume is a function of the effective radius of the ellipsoid. Therefore, we take $\phi(\mathrm{Vol}(E)) = R$.

In [11], the description of the domain is constructed as a sphere with radius $\sqrt{R}$. The points x_i from the learning set X lie in a sphere with radius $\sqrt{R + \xi_i}$ for some $\xi_i \geq 0$. Thus, the matrix of ellipsoid P is the identity matrix and the scalarized form of the problem of description of the data set has the following form:

$$\begin{gathered}\min_{R, a, \xi} R + C\sum_{i=1}^{N}\xi_i \\ s.t. (x_i - a)^T(x_i - a) \leq R + \xi_i, \quad i = 1\ldots N, \\ R \geq 0, \quad \xi_i \geq 0, \quad i = 1\ldots N.\end{gathered} \quad (7)$$

The authors of paper [12] propose a method for description of set X in which the data set is described by an ellipsoid whose matrix is equal to the covariance matrix of sample X. Such a selection of this matrix is related to the aim of taking into account correlation between different characteristics of simulated objects. In [12], the center of the ellipsoid is fixed at the arithmetic mean of points X:

$$\mu = \frac{1}{N}\sum_{i=1}^{N}x_i.$$

Matrix P is calculated by the formula

$$P = \mathrm{Cov}(X, X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T.$$

As a result, paper [12] proposes to solve the following optimization problem:

$$\begin{gathered}\min_{R, \xi} R + C\sum_{i=1}^{N}\xi_i \\ s.t. (x_i - \mu)^T\mathrm{Cov}^{-1}(X, X)(x_i - \mu) \leq R + \xi_i, \quad i = 1\ldots N, \\ R \geq 0, \quad \xi_i \geq 0, \quad i = 1\ldots N.\end{gathered} \quad (8)$$

Both problems (7) and (8) can be written in the convex form (see Section 4.2) and, hence, using the scalarization method, it is possible to find their Pareto frontier.

4. THE PROPOSED NEW APPROACH

The ellipsoid constructed by means of problem (8), has a fixed center. If it is allowed to vary, as in problem (7), we evidently obtain ellipsoids that are not worse after solution of the following problem:

$$\begin{gathered}\min_{R, a, \xi} R + C\sum_{i=1}^{N}\xi_i \\ s.t. (x_i - a)^T\mathrm{Cov}^{-1}(X, X)(x_i - a) \leq R + \xi_i, \quad i = 1\ldots N, \\ R \geq 0, \quad \xi_i \geq 0, \quad i = 1\ldots N.\end{gathered} \quad (9)$$

Problems (7), (8), and (9) described above have a common drawback: the matrix of the ellipsoid is fixed. We propose to include the matrix of the ellipsoid into a list of optimization variables. Then, we obtain the following optimization problem:

$$\min_{P, a, R, \xi} R + C\sum_{i=1}^{N} \xi_i$$
$$s.t.\ (x_i - a)^T P^{-1}(x_i - a) \le R + \xi_i, \quad i = 1 \dots N,$$
$$P = P^T > 0, \quad R \ge 0, \quad \xi_i \ge 0, \quad i = 1 \dots N.$$

The parameters of ellipsoid (1) include both matrix P and the squared effective radius R. By multiplying elements of P by some number $z > 1$, it is possible to decrease R and ξ_i without violating the constraint by dividing them by z and to decrease the $(R + C\sum_{i=1}^{N} \xi_i)$ to a value arbitrarily close to zero. To avoid this, let us require that an ellipsoid with the unit effective radius would be not too large. Let us add one inequality for the ellipsoid matrix determinant into the list of constraints of the problem. In addition, let us replace C by other dimensionless parameter $C = 1/N\nu$, the meaning of which will be explained later. Thus, we obtain the following problem:

$$\begin{gathered}\min_{P, a, R, \xi} R + \frac{1}{N\nu}\sum_{i=1}^{N} \xi_i \\ s.t.\ (x_i - a)^T P^{-1}(x_i - a) \le R + \xi_i, \quad i = 1 \dots N, \\ P = P^T > 0, \quad \det P \le 1, \quad R \ge 0, \quad \xi_i \ge 0, \quad i = 1 \dots N.\end{gathered} \tag{10}$$

4.1. Interpretation of the Scalarization Parameter

The following theorem explains the meaning of the replacement $C = 1/N\nu$.

Theorem 2. *The fraction of points of set X that do not belong to ellipsoid (1) constructed by solving problem (10) does not exceed* ν.

Proof. The Lagrangian for problem (10) can be written as

$$\begin{aligned}L(P, a, R, \xi, \alpha, \beta, \gamma, \delta) &= R + \frac{1}{N\nu}\sum_{i=1}^{N} \xi_i \\ &+ \sum_{i=1}^{N} \alpha_i((x_i - a)^T P^{-1}(x_i - a) - R\xi_i) \\ &- \sum_{i=1}^{N} \beta_i \xi_i - \gamma R + q(P, \delta),\end{aligned}$$

where $\alpha = \{\alpha_i \ge 0\}_{i=1}^{N}$, $\beta = \{\beta_i \ge 0\}_{i=1}^{N}$, and $\gamma \ge 0$ are the Lagrange multipliers, $q(P, \delta)$ is the function of the ellipsoid matrix and the Lagrange multipliers corresponding to the remaining constraints. At the optimal point, $\nabla L = 0$. We obtain the following system of equations:

$$\frac{\partial L}{\partial R} = 1 - \sum_{i=1}^{N} \alpha_i - \gamma = 0, \tag{11}$$

$$\frac{\partial L}{\partial \xi_i} = \frac{1}{N\nu} - \alpha_i - \beta_i = 0. \tag{12}$$

From the Karush–Kuhn–Tucker (KKT) conditions, we also have equations of complementary slackness:

$$\beta_i \xi_i = 0. \tag{13}$$

Point $x_i \notin E$ if and only if when $\xi_i > 0$, which, together with (13), means that $\beta_i = 0$. From (12), we obtain $\alpha_i = 1/N\nu$ for outliers (points out of the ellipsoid). Since $\gamma \ge 0$, then, taking into account Eq. (11), we conclude that $\sum_{i=1}^{N} \alpha_i \le 1$. Using nonnegativeness of all α_i, we state that the number of α_i that are equal to $1/N\nu$ does not exceed $N\nu$. Theorem is proved.

It is possible to prove similar statements for problems (7), (8), and (9).

Three following notes show how Theorem 2 can be applied in practice.

Note 1. It follows from Theorem 2 that the number of points of set X that lie out of the ellipsoid does not exceed $N\nu$, which allows one to set the range of possible values of parameter ν. In practice, too small or too large values of ν out of the range $[1/N, 1]$ are not used. For small values of ν, all points belong to the ellipsoid and the change of the scalarization parameter below $1/N$ does not change number (3) of outliers among the learning set; it remains $K = 0$.

For $\nu > 1$, all points may lie out of the ellipsoid. For large ν, the coefficient $C = 1/N\nu$ for the measure of remoteness of the point from the ellipsoid becomes small and the optimal ellipsoid is degenerated into the point with $R = 0$. In fact, Eq. (12) proves that $0 \le \alpha_i \le 1/N\nu$. If $\nu > 1$, then $\sum_{i=1}^{N} \alpha_i < 1$, and, hence (see (11)), $\gamma > 0$. The KKT condition $\gamma R = 0$ proves that $R = 0$, i.e., the ellipsoid is actually degenerated into the point.

Note 2. By solving problem (10), it is possible to approximate part of the Pareto frontier of problem (4). Theorem 2 allows one to select the value of the scalarization parameter if the engineer wants to obtain an ellipsoid with a specified number of outliers (3).

Note 3. It is possible to diagnose changes in the typical behavior of the simulated system. In the novelty detection problem, the analysts operate with an available data sample and train the classifier. When the system is deployed, normal behavior of the system may change and the parameters that were earlier anomalous may become normal in future and vice versa. This can mean that the resource of the mechanism is exhausted, or, alternatively, the evolution of the system may be typical (e.g., the climate change) and it is necessary to construct the ellipsoid again on the basis of new data. Theorem 2 allows one to automatically detect the need in retraining of the model. The system can track the intensity of detection of outliers. The ellipsoid should be considered out-of-date if the fraction of outliers exceeds the value of ν used during construction of the ellipsoid.

4.2. Convex Formulation of the Optimization Problem

Problem (10) can be reformulated in the convex form. Let $Q = P^{-1/2}$ and $b = Qa$. Then problem (10) is equivalent to the following problem:

$$\min_{Q,\, b,\, R,\, \xi} R + \frac{1}{N\nu}\sum_{i=1}^{N} \xi_i$$
$$s.t. (Qx_i - b)^T I (Qx_i - b) \le R + \xi_i, \quad i = 1\ldots N,$$
$$Q = Q^T > 0, \quad \det Q \ge 1, \quad R \ge 0, \quad \xi_i \ge 0, \quad i = 1\ldots N.$$

It is known that the function $f(Q) = -\ln\det Q$ is convex on a set of positively detfined symmetric matrices [1]. Using the Schur's lemma, the quadratic constraint can be rewritten in the form of a linear matrix inequality (LMI). The set of solutions of the LMI is convex [13]. Hence, problem (10) can be reduced to the convex programming problem

$$\min_{Q,\, b,\, R,\, \xi} R + \frac{1}{N\nu}\sum_{i=1}^{N} \xi_i$$
$$s.t. \begin{pmatrix} R + \xi_i & (Qx_i - b)^T \\ (Qx_i - b) & I \end{pmatrix}^T > 0, \quad i = 1\ldots N, \tag{14}$$
$$Q = Q^T > 0, \quad -\ln\det Q \le 0, \quad R \ge 0, \quad \xi_i \ge 0, \quad i = 1\ldots N.$$

Similarly, problems (7), (8), and (9) can be formulated in the convex form.

4.3. Refinement of the Solution

Problems (9) and (14) are the approximations of problem (4). Ellipsoid E on which the corresponding minimum value is attained contains some subset $U(E) \subseteq X$ of vectors from the learning sample. Points from $U(E)$ form a set of points at which ellipsoid E is Pareto optimal for problem (6).

A classical problem of construction of the minimum-volume ellipsoid containing specified points (Lowner ellipsoid) is known. Let us fix in the equation of ellipsoid (1) the squared effective radius $R = 1$ and replace variables: $Q = P^{-1/2}$ and $b = P^{-1/2}a$. Then the parameters of the minimum-volume ellipsoid containing points from $U(E)$ can be determined by solving the following problem [13]:

$$\min_{Q,\, b} -\ln\det Q$$
$$s.t. \begin{pmatrix} 1 & (Qx_i - b)^T \\ (Qx_i - b) & I \end{pmatrix}^T \succcurlyeq 0, \quad x_i \in U(E). \tag{15}$$

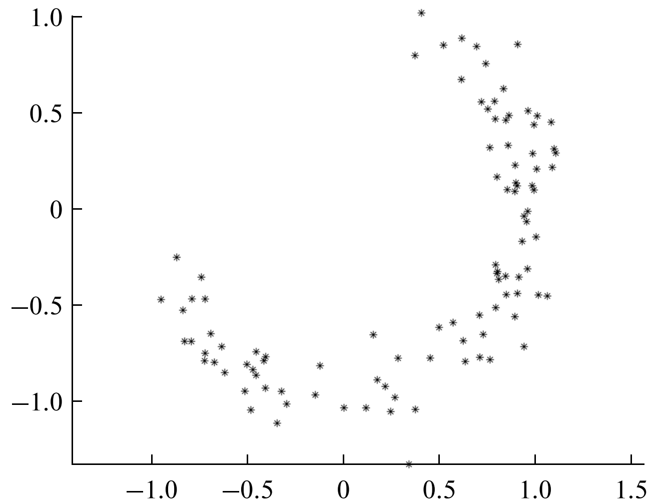

Fig. 1. Set of 100 points **banana** in $\mathbf{R}^2$, which form a nonconvex domain.

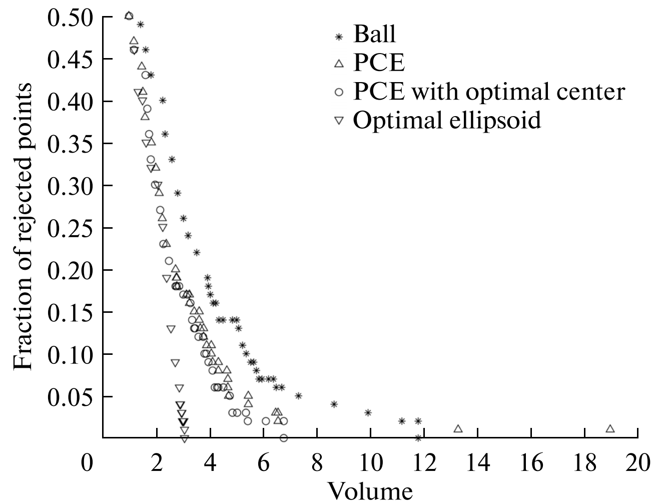

Fig. 2. Approximate Pareto frontier for the **box_10_100** data set.

Let us have ellipsoid E that is optimal for problem (6). Let us find points from X, that belong to ellipsoid E: $U(E) = \{x \in X | x \in E\}$. Using procedure (15), we construct minimum-volume ellipsoid E' containing points from $U(E)$. Then $\mathrm{Vol}(E') \le \mathrm{Vol}(E)$ and $\mathrm{K}(E') \le \mathrm{K}(E)$. In practice, the first inequality usually strictly holds. In the general case, ellipsoid E' does not belong to ellipsoid E and, in addition to points from $U(E)$, may contain additional points from X. Thus, application of procedure (15) allows one to obtain ellipsoid E' dominating ellipsoid E in the sense of the main problem (4).

5. RESULTS OF THE EXPERIMENTS

The methods based on solution of optimization problems (9) and (14), after the use of procedure (15) are compared with the known methods (7) and (8). Below, we will refer to method (7) as the *Ball*; method (8) as the *Principal Component Ellipsoid* or, in abbreviated form, *PCE*; our method (9) together with subsequent application of procedure (15) as the *PCE with optimal center*; and method (14) and (15) as the *Optimal ellipsoid*.

The convex programming problems were solved using the CVX package for MATLAB [14].

The methods were compared using the following artificial data sets:

1. Set **box_10_100** consists of 100 points uniformly distributed in cube $[0, 1]^{10}$.

2. In order to generate set **normal_COV_6_100** containing 100 points, a multivariate normal distribution in space $\mathbf{R}^6$ with fixed non-unitary covariance matrix $\Sigma \ne I$ was used.

3. Two-dimensional data set **banana** consists of 100 points taken from a nonconvex domain. It is shown in Fig. 1.

Plots 2, 3, and 4 show approximations of the Pareto frontier for problem (4) obtained by the considered methods for data sets **box_10_100**, **normal_COV_6_100**, and **banana**, respectively. The fraction of $\overline{K(E)}/N$ points that do not belong to the ellipsoid is placed along the ordinate axis. The d-dimension volume $\mathrm{Vol}(E)$ of the ellipsoid is placed along the abscissa axis. For each method and data set, 40 different values of ν in the range (0, 0.5] were used.

By analyzing the experimental results, we come to the following conclusions:

1. In spaces with low dimensions d, when the size of the sample N is sufficiently large, all methods give close approximations of the Pareto frontier.

2. The *Ball* method gives points that are much more distant from the Pareto frontier than other methods if the variance of the data along one direction substantially differs from the variance along other directions.

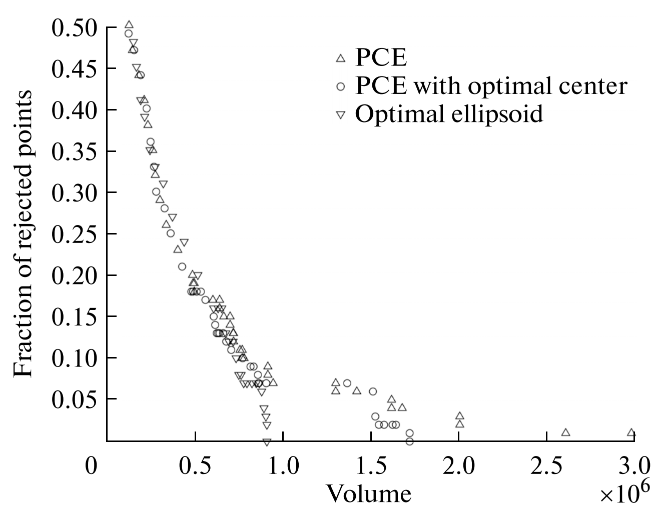

Fig. 3. Approximate Pareto frontier for the **normal_COV_6_100** data set.

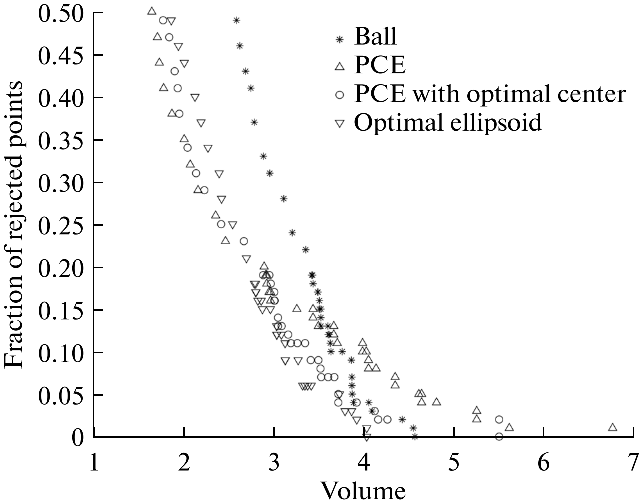

Fig. 4. Approximate Pareto frontier for the **banana** data set.

3. Sometimes our method *Optimal ellipsoid* gives points that are dominated by ellipsoids constructed by other methods. This is possible, since formulation (6) is an approximation of problem (4) and gives an approximate solution. The ellipsoids constructed using one approximation can dominate other approximate solutions.

4. If ν increases, i.e., the number of outliers in the learning set is large, all methods (except for *Ball*) give close Pareto frontiers in most experiments.

5. As compared to other methods, the *Optimal ellipsoid* method gives substantially better results in high-dimension spaces and when set X consists of a not very large number of points. These situations arise in practice. The computer experiment can be very long-term and, at the beginning of the study, the data base of the experiments contains a small number of multidimensional vectors.

6. Our approach *Optimal ellipsoid* also substantially surpasses the known methods when the fraction of points lying out of the ellipsoid (ν) is sufficiently small. This part of the Pareto frontier is most interesting in practice.

6. CONCLUSIONS

We have considered the data representation with the use of extremal ellipsoids. Main problem (4) with continuous and discrete objective functions has been solved approximately by replacing initial criteria with other convex functions. The scalarization method has been used to solve the multicriterion problem.

In future, alternative approximations of problem (4) should be studied. Minimization of the sums of distances in the Mahalanobis metric from the boundary of the ellipsoid to the point instead of minimization of their squares may increase stability of location estimation and give more exact results for the number of outliers $K(E)$, since large distances from the point to the ellipsoid are penalized weaker. More exact but nonconvex approximations of this problem are also possible. However, there are no universal fast solution methods for them.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation, project no. 14-50-00150.

REFERENCES

1. S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge Univ. Press, Cambridge, 2004).
2. A. A. Bedrintsev, V. V. Chepyzhov, and S. S. Chernova, "Extreme ellipsoids as approximations of design space in data predictive metamodelling," Artificial intelligence and decision making, No. 2, 35–44 (2015).
3. A. A. Bedrintsev and V. V. Chepyzhov, "Two-objective problem of optimal ellipsoids constructing for data representation," in *Information Technologies and Systems – 2014, (Proc. 38th Conf.–School, Nizhny Novgorod, Sept. 1–5, 2014)* (ITAS RAS, Moscow, 2014) [in Russian].
4. D. T. J. Tax and R. P. W. Duin, "Support vector data description," Mach. Learning **54**, 45–66 (2004).
5. Karanjit Singh and Shuchita Upadhyaya, "Outlier detection: applications and techniques," Int. J. Comput. Sci. Issues **9**, 307–323 (2012).
6. Rana Poonam, Pahuja Deepika, and Gautam Ritu, "A critical review on outlier detection techniques" Int. J. Sci. Res. **3**, 2394–2403 (2014).
7. P. Rousseeuw, "Multivariate estimation with high breakdown point," in *Mathematical Statistics and Applications*, Ed. by W. Grossmann, G. Pflug, I.Vincze, and W.Wertz (Reidel Publishing Co., Dordrecht, 1985), Vol. B, p. 283–297.
8. D. Laurie, "The asymptotics of Rousseeuw's minimum volume ellipsoid estimators," Ann. Stat., **20**, 1828–1843 (1992).
9. Xuming He and Gang Want, "Cross-checking using the minimum volume ellipsoid estimator," Stat. Sinica **6**, 367–375 (1996).
10. S. Val Aelst and P. Rousseeuw, "Minimum volume ellipsoid," WIREs Comput. Stat. **1**, 71–82 (2009).
11. Wei-Cheng Chang, Ching-Pei Lee and Chih-Jen Lin, *A Revisit to Support Vector Data Description. Tech. Rep.* (Nat. Taiwan Univ. Science and Technology, Taiwan, 2013).
12. Yang Zhang, Nirvana Meratnia, and P. J. M. Havinga, "Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine," Ad Hoc Networks **11**, 1063–1074 (2012).
13. S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, "Linear matrix inequalities in system and control theory," in *SIAM Studies in Applied Mathematics* (SIAM, Philadelphia, 1994), Vol. 15.
14. CVX: Matlab Software for Disciplined Convex Programming. http://cvxr.com/cvx/.

Translated by N. Pakhomova