

On Dimensionality of Embedding Space in Multidimensional Scaling

Julius ŽILINSKAS

*Institute of Mathematics and Informatics
Akademijos 4, LT-08663, Vilnius, Lithuania
e-mail: julius.zilinskas@mii.lt*

Received: 3 October 2007; accepted: 11 June 2008

Abstract. Multidimensional scaling is a technique for exploratory analysis of multidimensional data widely usable in different applications. By means of this technique the image points in a low-dimensional embedding space can be found whose inter-point distances fit the given dissimilarities between the considered objects. In this paper dependence of relative visualization error on the dimensionality of embedding space is investigated. Both artificial and practical data sets have been used. The images in three-dimensional embedding space normally show the structural properties of sets of considered objects with acceptable accuracy, and widening of applications of stereo screens makes three-dimensional visualization very attractive.

Keywords: multidimensional scaling, global optimization, city-block distances.

1. Introduction

Multidimensional scaling (MDS) is a technique for exploratory analysis of multidimensional data widely usable in different applications (Borg and Groenen, 2005; Cox and Cox, 2001); e.g., applications of MDS based visualization of observation points in interactive optimization systems is discussed in (Törn and Žilinskas, 1989). Let us give a short formulation of the problem. The dissimilarity between pairs of n objects is given by the matrix (δ_{ij}) , $i, j = 1, \dots, n$, and it is supposed that $\delta_{ij} = \delta_{ji}$. The points in an m -dimensional embedding space $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$, $i = 1, \dots, n$ should be found whose inter-point distances fit the given dissimilarities. Different measures of accuracy of fit can be chosen defining different images of the considered set of objects. In the case the objects are points in a high-dimensional vector space such images can be interpreted as different nonlinear projections of points in high-dimensional space to their images in an embedding space of lower dimensionality. The problem of construction of images of the considered objects is reduced to minimization of accuracy of a fit criterion, e.g., of the most frequently used least squares *STRESS* function

$$S(\mathbf{x}) = \sum_{i < j}^n w_{ij} (d(\mathbf{x}_i, \mathbf{x}_j) - \delta_{ij})^2, \quad d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^r \right)^{1/r}, \quad (1)$$

where $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$; $d(\mathbf{x}_i, \mathbf{x}_j)$ denotes the distance between the points \mathbf{x}_i and \mathbf{x}_j , and $w_{ij} > 0$, $i, j = 1, \dots, n$ denote weights. To define a particular criterion a norm in \mathbf{R}^m should be chosen implying the particular formula for calculating distances $d(\mathbf{x}_i, \mathbf{x}_j)$. The most frequently used norm is Euclidean ($r = 2$). However, MDS with other Minkowski norms in embedding space can be even more informative than MDS with Euclidean norm (Arabie, 1991). Analysis of MDS results corresponding to different norms can highlight different properties of the considered objects similarly as different orthogonal projections of a three-dimensional object complement each other.

MDS is a difficult high-dimensional ($\mathbf{x} \in \mathbf{R}^N$, $N = n \times m$) global optimization problem, especially when the number of objects is large (Bernatavičienė *et al.*, 2007). Complexity of such optimization problems is similar to the complexity of the distance geometry problems, e.g., to the problems of molecular conformation (Pardalos *et al.*, 1996). In the present paper MDS based on *STRESS* criterion with city-block distances ($r = 1$) in the embedding space is considered. Besides of multimodality the non differentiability of *STRESS* in this case cannot be ignored, *STRESS* can be non differentiable even at the minimum point. Since the objective function can be non differentiable at local minimizer, application of local descent methods with high convergence rate, e.g., of different versions of Newton method, seems questionable. Therefore to develop a local minimization algorithm we employ the piecewise quadratic structure of the objective function.

A survey of city-block MDS was presented in (Arabie, 1991). A combinatorial approach for city-block MDS was proposed in (Hubert *et al.*, 1992), where combinatorial local search is used to construct good object orders along dimensions, and least-squares are used to estimate the coordinates of image points for the objects based on the object orders. A smoothing approach for city-block MDS was proposed in (Groenen *et al.*, 1998), where smoothing excludes some local minima of *STRESS*.

For global minimization of multimodal functions randomized methods seem most promising (Zhigljavsky and Žilinskas, 2008). A heuristic algorithm based on simulated annealing for two-dimensional city-block scaling was proposed in (Brusco, 2001). A two level minimization method for the two-dimensional embedding space was proposed in (Žilinskas and Žilinskas, 2007) where a problem of combinatorial optimization is tackled by evolutionary search at the upper level, and a problem of quadratic programming is tackled at the lower level. The parallel version of the algorithm is proposed and investigated in (Žilinskas and Žilinskas, 2006). The generalized method for arbitrary dimensionality of the embedding space is developed and experimentally compared with distance smoothing approach and simulated annealing in (Žilinskas and Žilinskas, 2008a). The widening of applications of stereo screens makes such a generalization an urgent problem for stereo visualization of multidimensional data. In this paper the dependence of accuracy of fit on dimensionality of the embedding space is analyzed.

2. MDS with City-Block Distances

STRESS (1) with city-block distances can be redefined as

$$S(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left(\sum_{k=1}^m |x_{ik} - x_{jk}| - \delta_{ij} \right)^2. \quad (2)$$

Let $A(\mathbf{P})$ denote a set such that

$$A(\mathbf{P}) = \{ \mathbf{x} \mid x_{ik} \leq x_{jk} \text{ for } p_{ki} < p_{kj}, i, j = 1, \dots, n, k = 1, \dots, m \}, \quad (3)$$

where $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_m)$, $\mathbf{p}_k = (p_{k1}, p_{k2}, \dots, p_{kn})$ is a permutation of $1, \dots, n$; $k = 1, \dots, m$.

For $\mathbf{x} \in A(\mathbf{P})$, (2) can be rewritten in the following form (Žilinskas and Žilinskas, 2008a)

$$S(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left(\sum_{k=1}^m (x_{ik} - x_{jk}) z_{kij} - \delta_{ij} \right)^2,$$

where

$$z_{kij} = \begin{cases} 1, & p_{ki} > p_{kj}, \\ -1, & p_{ki} < p_{kj}, \end{cases} \quad k = 1, \dots, m, i, j = 1, \dots, n.$$

Therefore for fixed \mathbf{P} and $\mathbf{x} \in A(\mathbf{P})$, $S(\mathbf{x})$ is a quadratic function implying that the minimization problem

$$\min_{\mathbf{x} \in A(\mathbf{P})} S(\mathbf{x}) \quad (4)$$

is a quadratic programming problem. The objective function in (4) can be written in the following form:

$$\begin{aligned} S(\mathbf{x}) &= \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left(\sum_{k=1}^m (x_{ik} - x_{jk}) z_{kij} - \delta_{ij} \right)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij}^2 - 2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij} \sum_{k=1}^m (x_{ik} - x_{jk}) z_{kij} \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left(\sum_{k=1}^m (x_{ik} - x_{jk}) z_{kij} \right)^2. \end{aligned}$$

The first summand is a constant, and need not be taken into account in minimization. Let us assume $\delta_{ij} = \delta_{ji}$, $\delta_{ii} = 0$, $w_{ij} = w_{ji}$. The second summand is a linear function which

can be simplified as follows

$$\begin{aligned}
& -2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij} \sum_{k=1}^m (x_{ik} - x_{jk}) z_{kij} \\
&= -2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij} \sum_{k=1}^m x_{ik} z_{kij} + 2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij} \sum_{k=1}^m x_{jk} z_{kij} \\
&= -4 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij} \sum_{k=1}^m x_{ik} z_{kij} = -4 \sum_{k=1}^m \sum_{i=1}^n x_{ik} \sum_{j=1}^n w_{ij} \delta_{ij} z_{kij}.
\end{aligned}$$

Similarly the third summand can be written as a quadratic function

$$\begin{aligned}
& \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left(\sum_{k=1}^m (x_{ik} - x_{jk}) z_{kij} \right)^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{k=1}^m \sum_{l=1}^m (x_{ik} - x_{jk})(x_{il} - x_{jl}) z_{kij} z_{lij} \\
&= \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1}^n (x_{ik} x_{il} - x_{ik} x_{jl} - x_{jk} x_{il} + x_{jk} x_{jl}) w_{ij} z_{kij} z_{lij} \\
&= \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1}^n x_{ik} x_{il} w_{ij} z_{kij} z_{lij} + \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1}^n x_{jk} x_{jl} w_{ij} z_{kij} z_{lij} \\
&\quad - \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1}^n x_{ik} x_{jl} w_{ij} z_{kij} z_{lij} - \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1}^n x_{jk} x_{il} w_{ij} z_{kij} z_{lij} \\
&= 2 \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1}^n x_{ik} x_{il} w_{ij} z_{kij} z_{lij} - 2 \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1}^n x_{ik} x_{jl} w_{ij} z_{kij} z_{lij} \\
&= 2 \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n x_{ik} x_{il} \sum_{j=1}^n w_{ij} z_{kij} z_{lij} - 2 \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1, j \neq i}^n x_{ik} x_{jl} w_{ij} z_{kij} z_{lij} \\
&\quad - 2 \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n x_{ik} x_{il} w_{ii} z_{kii} z_{lii} \\
&= 2 \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n x_{ik} x_{il} \sum_{t=1, t \neq i}^n w_{it} z_{kit} z_{lit} - 2 \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1, j \neq i}^n x_{ik} x_{jl} w_{ij} z_{kij} z_{lij}.
\end{aligned}$$

Furthermore the linear and quadratic parts can be divided by 4 without influence to minimization.

From (3), $\mathbf{x} \in A(\mathbf{P})$ means that

$$x_{ik} \leq x_{jk} \quad \text{for } p_{ki} < p_{kj}, \quad i, j = 1, \dots, n, \quad k = 1, \dots, m,$$

which is equivalent to constraints

$$x_{\{j|p_{kj}=i\},k} \leq x_{\{j|p_{kj}=i+1\},k}, \quad k = 1, \dots, m, \quad i = 1, \dots, n - 1,$$

and can be redefined as

$$x_{\{j|p_{kj}=i+1\},k} - x_{\{j|p_{kj}=i\},k} \geq 0, \quad k = 1, \dots, m, \quad i = 1, \dots, n - 1.$$

STRESS function and constraints are invariant with respect to translation (addition of constant values to x_{ik} , $i = 1, \dots, n$). We exclude this unfavorable property (w.r.t. optimization) by centering the solution by means of the following constraints

$$\sum_{i=1}^n x_{ik} = 0, \quad k = 1, \dots, m.$$

Summarizing the above algebra the problem of minimization of $S(\mathbf{x})$ over $A(\mathbf{P})$ is reduced to the quadratic programming problem

$$\begin{aligned} \min & \left[- \sum_{k=1}^m \sum_{i=1}^n x_{ik} \sum_{j=1}^n w_{ij} z_{kij} \right. \\ & + \frac{1}{2} \left(\sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n x_{ik} x_{il} \sum_{t=1, t \neq i}^n w_{it} z_{kit} z_{lit} \right. \\ & \quad \left. \left. - \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^n \sum_{j=1, j \neq i}^n x_{ik} x_{jl} w_{ij} z_{kij} z_{lij} \right) \right] \\ \text{s.t.} & \sum_{i=1}^n x_{ik} = 0, \quad k = 1, \dots, m, \\ & x_{\{j|p_{kj}=i+1\},k} - x_{\{j|p_{kj}=i\},k} \geq 0, \quad k = 1, \dots, m, \quad i = 1, \dots, n - 1, \end{aligned}$$

which can be written in the matrix form (Žilinskas and Žilinskas, 2008a) as presented below:

$$\min \left[- \mathbf{d}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{D} \mathbf{x} \right] \tag{5}$$

$$\text{s.t.} \quad \mathbf{A}_0 \mathbf{x} = 0, \tag{6}$$

$$\mathbf{A}_k \mathbf{x} \geq 0, \quad k = 1, \dots, m, \tag{7}$$

where

$$d_{kn-n+i} = \sum_{j=1}^n w_{ij} \delta_{ij} z_{kij} \quad \left| \begin{array}{l} k = 1, \dots, m, \\ i = 1, \dots, n, \end{array} \right. \tag{8}$$

$$D_{kn-n+i,ln-n+j} = \begin{cases} \sum_{t=1, t \neq i}^n w_{it} z_{kit} z_{lit}, & i = j, \\ -w_{ij} z_{kij} z_{lij}, & i \neq j, \end{cases} \left| \begin{array}{l} k, l = 1, \dots, m, \\ i, j = 1, \dots, n, \end{array} \right. \quad (9)$$

$$A_{0kj} = \begin{cases} 1, & j = kn - n + 1, \dots, kn, \\ 0, & \text{otherwise,} \end{cases} \left| \begin{array}{l} k = 1, \dots, m, \\ j = 1, \dots, mn, \end{array} \right. \quad (10)$$

$$A_{kij} = \begin{cases} 1, & p_{k,j-kn+n} = i + 1, \\ -1, & p_{k,j-kn+n} = i, \\ 0, & \text{otherwise,} \end{cases} \left| \begin{array}{l} k = 1, \dots, m, \\ i = 1, \dots, n - 1, \\ j = 1, \dots, mn. \end{array} \right. \quad (11)$$

Polyhedron $\mathbf{x} \in A(\mathbf{P})$ is defined by linear inequality constrains (7). Equality constrains (6) ensure centering of the solution with respect to each coordinate.

For the problem (5) a standard quadratic programming method can be applied. A solution of a quadratic programming problem is not necessarily a local minimizer of the initial problem of minimization of *STRESS* (1). This is because *STRESS* is minimized with respect to \mathbf{P} as well. If a solution of a quadratic programming problem is on the border of polyhedron $A(\mathbf{P})$, a local minimizer possibly is located in a neighboring polyhedron. Therefore a search can be continued by solving a quadratic programming problem over the polyhedron on the opposite side of the active inequality constrains. The permutations in \mathbf{P} should be updated to define the neighboring polyhedron. If i, \dots, j inequality constrains $\mathbf{A}^k \mathbf{x} \geq 0$ are active, $i \leq p_{kt} \leq j + 1$ should be updated to $i + j + 1 - p_{kt}$. Solution of quadratic programming problems is repeated while better values are found, and some inequality constrains are active.

Taking into account the structure of the minimization problem (4) we apply a two-level minimization algorithm:

$$\min_{\mathbf{P}} S(\mathbf{P}), \quad (12)$$

$$\text{s.t. } S(\mathbf{P}) = \min_{\mathbf{x} \in A(\mathbf{P})} S(\mathbf{x}), \quad (13)$$

where the upper level problem is a combinatorial problem defined over the set \mathbf{P} of m -tuple of permutations of $1, \dots, n$ (one permutation per each coordinate of embedding space), and the lower level problem is a quadratic programming problem with a positively definite quadratic objective function and linear constraints setting the sequences of values of coordinates of originals defined by permutations \mathbf{P} . The problem at lower level is solved using quadratic programming algorithm. Globalism of search is ensured by the upper level algorithms.

The number of feasible solutions of the upper level combinatorial problem is $(n!)^m$. A solution of MDS with city-block distances is invariant with respect to mirroring when changing direction of coordinate axes or exchanging of coordinates. The number of feasible solutions can be reduced to approximately $(n!/2)^m/m!$ refusing mirrored solutions. It can be further reduced when the data exposes symmetries (Žilinskas, 2007).

Guaranteed methods can be applied for the upper level combinatorial problem when the number of objects is small. Branch and bound algorithm has been developed in (Žilinskas and Žilinskas, 2008b). Genetic algorithms seem prospective to solve the upper level

problem when the number of objects or dimensionality of embedding space is larger. The idea of genetic algorithm is to maintain a population of best (with respect to *STRESS* value) solutions whose crossover can generate better solutions. The permutations in \mathbf{P} are considered as chromosomes representing an individual. The initial population of individuals is generated randomly, and it is improved performing the local search method described above. A mutation is performed with a predefined probability by exchanging two random objects of randomly chosen individual. The population evolves generating offsprings from two randomly chosen individuals of the current population with the chromosomes $\tilde{\mathbf{P}}$ and $\hat{\mathbf{P}}$, where the first corresponds to the better fitted parent. The fitness of an individual is defined by the optimal value of the corresponding lower level problem. The chromosome of the offspring is defined by the following formula

$$\mathbf{p}_k = (\hat{p}_{k1}, \dots, \hat{p}_{k\xi_1}, \tilde{p}_{k1}, \dots, \tilde{p}_{k(\xi_2-\xi_1)}, \hat{p}_{k\xi_2}, \dots, \hat{p}_{kn}), \quad k = 1, \dots, m,$$

where ξ_1, ξ_2 are two integer random numbers from the set $\{1, \dots, n\}$; and \tilde{p}_{ki} are numbers from the set $\{1, \dots, n\}$ not included into the set $\{\hat{p}_{k1}, \dots, \hat{p}_{k\xi_1}, \hat{p}_{k\xi_2}, \dots, \hat{p}_{kn}\}$, and ordered in the same way as they are ordered in $\tilde{p}_{k1}, \dots, \tilde{p}_{kn}$. The offspring is improved performing the local search method described above. An elitist selection is applied: if the offspring is better fitted than the worst individual of the current population, then the offspring replaces the latter. The minimization continues generating new offsprings and terminates after the predetermined computing time t_c . The structure of the hybrid algorithm for multidimensional scaling is presented in Algorithm 1.

Algorithm 1: Hybrid algorithm for multidimensional scaling

Input: $N_{init}; pp; t_c; n; m; \delta_{ij}; w_{i,j}, i, j = 1, \dots, n$

Output: S^*

- 1: Generate N_{init} uniformly distributed random vectors \mathbf{x} of dimension $n \cdot m$.
 - 2: Perform search for local minima starting from the best pp generated vectors.
 - 3: Form the initial population from the found local minimizers.
 - 4: **while** t_c time has not passed **do**
 - 5: Randomly with uniform distribution select two parents from a current population.
 - 6: Produce an offspring by means of crossover and local minimization.
 - 7: **if** the offspring is more fitted than the worst individual of the current population **then**
 - 8: the offspring replaces the latter.
 - 9: **end if**
 - 10: **end while**
-

3. Experimental Investigation of the Dependence of Accuracy of Fit on Dimensionality of the Embedding Space

The visualization accuracy cannot be predicted theoretically because of the complexity of the underlying global optimization problem. We investigate the error of fit (express-

ing reciprocal value of accuracy) experimentally. Several sets of multidimensional points corresponding to well understood geometric objects are needed for the experimental investigation. We want to choose difficult test problems, i.e., difficult to visualize geometric objects. The data with desired properties correspond to the multidimensional objects equally extending in all dimensions of the original space, e.g., sets of vertices of multidimensional cubes and simplices. Dissimilarity between vertices is measured by the distance in the original vector space defined by its metric, city-block metrics in our case.

In the results below a relative error at a solution

$$f(\mathbf{x}) = \sqrt{S(\mathbf{x}) / \sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij}^2}$$

is used for comparison instead of the value of *STRESS* function to reduce the impact of the number of objects; this criterion is also less sensitive to the metric in the original space.

Sets of vertices of multidimensional simplices (Žilinskas, 2008) have been used as data sets in this paper. The number of vertices of d -dimensional simplex is $n = d + 1$, and the dimensionality of the global minimization problem is $N = m \times (d + 1)$. The distances between any two vertices of the standard simplex are equal: $\delta_{ij} = 1, i \neq j$. Vertices of the unit simplex can be defined by

$$v_{ij} = \begin{cases} 1, & \text{if } i = j + 1, \\ 0, & \text{otherwise,} \end{cases} \quad \left| \begin{array}{l} i = 1, \dots, d + 1, \\ j = 1, \dots, d. \end{array} \right.$$

Below we use shorthand ‘simplex’ for sets of its vertices.

The hybrid global optimization algorithm composed of genetic search and quadratic programming for MDS with city-block distances has been used to visualize simplices of different dimensionality in embedding space of different dimensionality. Fig. 1 shows

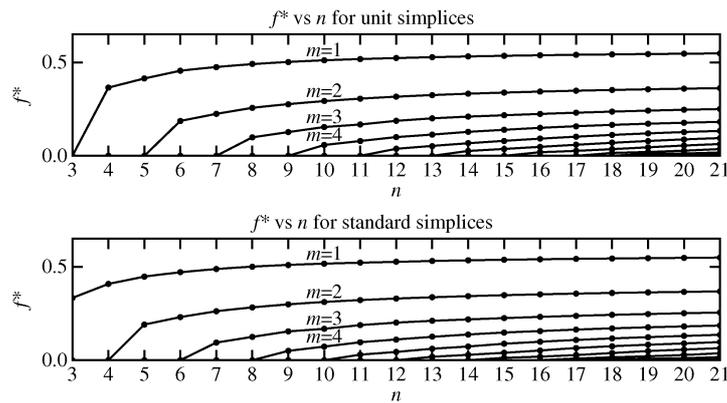


Fig. 1. Dependence of relative visualization error on the number of vertices n of multidimensional simplices and the dimensionality of embedding space m .

how values of relative error at the best found solutions depend on the number of vertices n of multidimensional simplices. Different graphs represent different dimensionalities of embedding space m . Black circles represent the results when the hybrid global optimization algorithm finds the same best function value in at least 10 runs out of 100. Grey circles represent less reliable results. Relative visualization error grows with dimensionality of simplices as expected, however the rate of growth is fast decreasing. There is a big decrease of visualization error between unidimensional and two-dimensional scaling, as well as between two-dimensional and three-dimensional scaling.

Fig. 2 shows how values of relative error at the best found solutions depend on the dimensionality of embedding space m . Different graphs represent simplices of different dimensionality. The gray circles should be again considered with care, since the same best function value has been found in less than 10 runs out of 100. As expected, relative visualization error reduces when the dimensionality of embedding space is increased, as it becomes more easy to fit image to the data. Quite a large decrease of visualization error can be reached by changing the two-dimensional embedding space to the three-dimensional one. Although visualization error can be further decreased using embedding space of larger dimensionality, practical use of the dimensionality larger than three is questionable.

A frequently used test problem for MDS algorithms is based on experimental testing of several soft drinks (Green *et al.*, 1989). 38 students have tested ten different brands of soft drinks. Each pair was judged on its dissimilarity on a 9 point scale (1 – very similar, 9 – completely different). The scaled down accumulated dissimilarities have been used as a practical data set in our experiments. This problem is referred as ‘cola’ problem in the results below, $n = 10$ in this problem. As the data for this problem is dissimilarity matrix, actual dimensionality of data is not known.

Problems of analysis of pharmacological binding affinity data (Žilinskas, 2006) have been used as other practical data sets. ‘ruusk1’ represents binding affinity data (Ruuskanen *et al.*, 2005) analyzed as properties of three human and five zebrafish α_2 -adrenoceptor

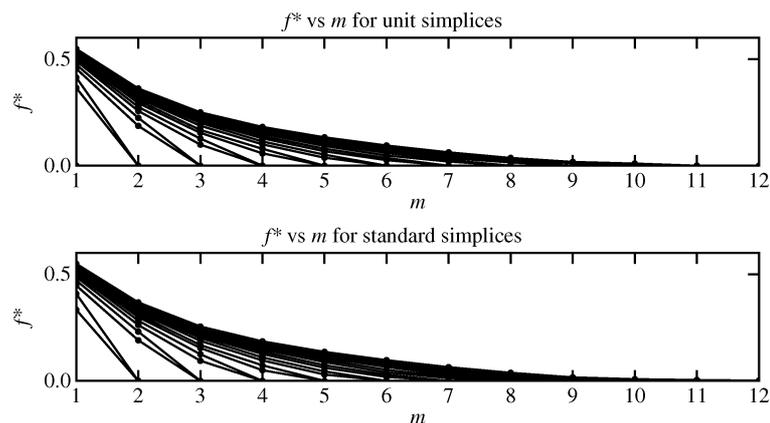


Fig. 2. Dependence of relative visualization error of multidimensional simplices on the dimensionality of embedding space m .

proteins, $n = 8$, each protein is represented by 20-dimensional vector; ‘ruusk2’ represents binding affinity data (Ruuskanen *et al.*, 2005) analyzed as properties of 20 ligands known to bind to the human α_2 -adrenoceptors, $n = 20$, each ligand is represented by 8-dimensional vector; ‘uhlen’ represents binding affinity data (Uhlén *et al.*, 1998) analyzed as properties of human, rat, guinea pig and pig α_2 -adrenoceptor proteins, $n = 12$, each protein is represented by 5-dimensional vector; ‘hwa’ represents binding affinity data (Hwa *et al.*, 1995) analyzed as properties of wild type and mutant proteins, $n = 12$, each protein is represented by 6-dimensional vector. The binding affinity data is represented through a matrix, one dimension formed by the different ligands tested in a series of experiments while the other dimension represents the different proteins. Dissimilarities of proteins are computed as distances between vectors of the \log_{10} -transformed binding affinities representing properties of the proteins. Dissimilarities of ligands are computed as distances between vectors of the \log_{10} -transformed binding affinities representing ligands.

Fig. 3 shows how values of relative error at the best found solutions of practical MDS problems depend on the dimensionality of embedding space m . Different graphs represent different problems which are specified at vertical axis. As expected, relative visualization error reduces when the dimensionality of embedding space is increased, as it becomes more easy to fit image to the data. However for these problems the relative visualization error decreases more rapidly than for high-dimensional simplices, even for the problem ‘ruusk1’, which is composed by 20-dimensional vectors. Visualization errors are largest for ‘cola’ problem; this is probably because dissimilarity data in this problem are collected by psychological testing, in contrast data of other problems are distances between multidimensional points. Visualization in three-dimensional embedding space seems most appropriate for the considered practical data sets.

The images in three-dimensional embedding space normally show the structural properties of sets of considered objects better than in two-dimensional space. The widening

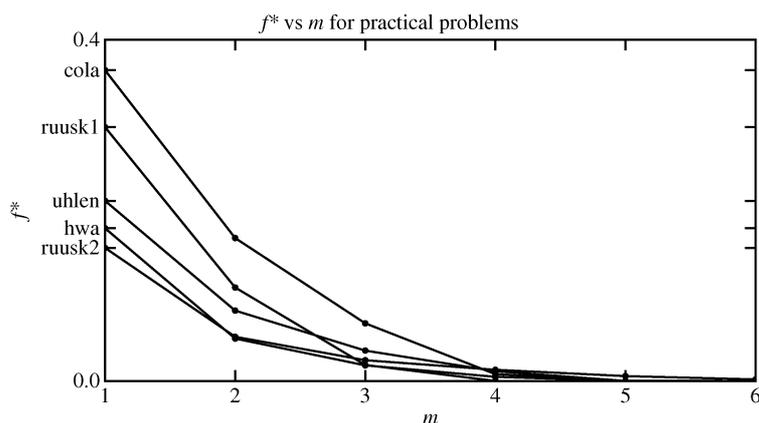


Fig. 3. Dependence of relative visualization error of practical data sets on the dimensionality of embedding space m .

of applications of stereo screens makes three-dimensional visualization very attractive, however it is difficult to demonstrate the advantages of stereo images in the paper. The classical methods of two-dimensional visualization by means of orthogonal and isometric projections are much weaker than dynamic visualization on stereo screens. Nevertheless some properties of the data can be grasped from such projections. For example, projections of three-dimensional image of the problem ‘ruusk2’ are shown in Fig. 4. Four projections are shown: the left upper picture shows the orthogonal projection on the xz-plane, the right upper picture shows the orthogonal projection on the yz-plane, the left lower picture shows the orthogonal projection on the xy-plane, and the right lower picture shows isometric projection. In the orthogonal projections the x-axis directs left, the z-axis directs up, and the y-axis directs right in the right upper picture and down in the left lower picture. In the isometric projection the x-axis directs down left, the y-axis directs down right, and the z-axis directs up. The ligands are numbered according to (Ruuskanen *et al.*, 2005). Agonists are indicated by the symbol ‘+’. Antagonists are indicated by the symbol ‘x’.

Binary tree clustering and principal component analysis have been used for analysis of the problem in (Ruuskanen *et al.*, 2005). In the binary tree all agonist ligands cluster

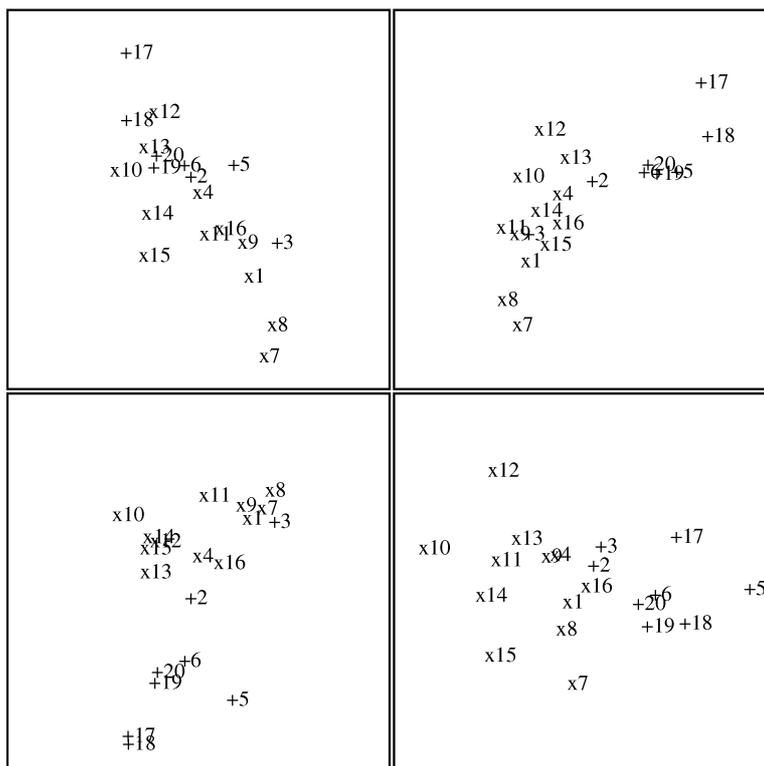


Fig. 4. Projections of three-dimensional image of the properties of 20 ligands binding human and zebrafish α_2 -adrenoceptors; $n = 20$, $f^* = 0.0243$.

together except agonist 3, which clusters together with antagonist ligands. In the image of principal components all agonist cluster together, but antagonist 16 is near this cluster. Similarly in the image of two-dimensional scaling with Euclidean distances in (Žilinskas, 2006) antagonist 16 makes separation of agonist and antagonist clusters difficult. However, in the image of two-dimensional scaling with city-block distances in (Žilinskas, 2006) it is possible to draw a line which would separate agonists and antagonists.

The orthogonal projection shown in the left lower picture in Fig. 4 is quite similar to mirrored two-dimensional image shown in (Žilinskas, 2006). The agonists 2 and 3 are near the cluster of antagonists in this projection, but it is possible to draw a line separating agonists and antagonists. The clusters of agonists and antagonists are separable more clearly in the three-dimensional image. Although it is difficult to demonstrate this on the paper, but more clear separability can be seen in the isometric projection in the right lower picture of Fig. 4. For better understanding of three-dimensional images acquired using multidimensional scaling we have created rotating pictures and placed them on Internet (<http://www.mii.lt/enoc/>). The rotating images are stored in PDF format therefore no specific software or hardware is required to analyze them. It is easier to see volume in the rotating pictures than in the projections.

4. Conclusions

The images acquired using multidimensional scaling expose the structural properties of sets of considered objects better in three-dimensional embedding space than in two-dimensional space. The decrease of visualization error between two-dimensional and three-dimensional scaling is similar to the decrease between unidimensional and two-dimensional scaling. The widening of applications of stereo screens makes three-dimensional scaling very attractive.

Acknowledgments

The research is partially supported by the Agency for International Science and Technology Development Programmes in Lithuania through COST programme and Lithuanian State Science and Studies Foundation within the project B-03/2007 “Global optimization of complex systems using high performance computing and GRID technologies”.

References

- Arabie, P. (1991). Was Euclid an unnecessarily sophisticated psychologist? *Psychometrika*, **56**(4), 567–587.
- Bernatavičienė, J., G. Dzemyda and V. Marcinkevičius (2007). Conditions for optimal efficiency of relative MDS. *Informatica*, **18**(2), 187–202.
- Borg, I., and P. Groenen (2005). *Modern Multidimensional Scaling*, 2nd Edition. Springer, New York.
- Brusco, M.J. (2001). A simulated annealing heuristics for unidimensional and multidimensional (city-block) scaling of symmetric proximity matrices. *Journal of Classification*, **18**(1), 3–33.

- Cox, T., and M. Cox (2001). *Multidimensional Scaling*. Chapman and Hall/CRC, Boca Raton.
- Green, P., F. Carmone and S. Smith (1989). *Multidimensional Scaling: Concepts and Applications*. Allyn and Bacon, Boston.
- Groenen, P.J.F., W.J. Heiser and J.J. Meulman (1998). City-block scaling: smoothing strategies for avoiding local minima. In: I. Balderjahn, R. Mathar and M. Schader (Eds.), *Classification, Data Analysis, and Data Highways*. Springer. pp. 46–53.
- Hubert, L., P. Arabie and M. Hesson-Mcinnis (1992). Multidimensional scaling in the city-block metric: a combinatorial approach. *Journal of Classification*, **9**(2), 211–236.
- Hwa, J., R.M. Graham and D.M. Perez (1995). Identification of critical determinants of α_1 -adrenergic receptor subtype selective agonist binding. *Journal of Biological Chemistry*, **270**(39), 23189–23195.
- Pardalos, P., D. Shalloway and G. Hue (Eds.) (1996). *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*. American Mathematical Society.
- Ruuskanen, J.O., J. Laurila, H. Xhaard, V.-V. Rantanen, K. Vuoriluoto, S. Wurster, A. Marjamäki, M. Vainio, M.S. Johnson and M. Scheinin (2005). Conserved structural, pharmacological and functional properties among the three human and five zebrafish α_2 -adrenoceptors. *British Journal of Pharmacology*, **144**(2), 165–177.
- Törn, A., and A. Žilinskas (1989). Global Optimization. *Lecture Notes in Computer Science*, **350**, 1–250.
- Uhlén, S., M. Dambrova, J. Näsman, H.B. Schiöth, Y. Gu, A. Wikberg-Matsson and J.E.S. Wikberg (1998). [3 H]RS79948-197 binding to human, rat, guinea pig and pig α_{2A} -, α_{2B} - and α_{2C} -adrenoceptors. Comparison with MK912, RX821002, rauwolscine and yohimbine. *European Journal of Pharmacology*, **343**(1), 93–101.
- Zhigljavsky, A., and A. Žilinskas (2008). *Stochastic Global Optimization*. Springer, NY.
- Žilinskas, A., and J. Žilinskas (2006). Parallel hybrid algorithm for global optimization of problems occurring in MDS-based visualization. *Computers & Mathematics with Applications*, **52**(1–2), 211–224.
- Žilinskas, A., and J. Žilinskas (2007). Two level minimization in multidimensional scaling. *Journal of Global Optimization*, **38**(4), 581–596.
- Žilinskas, A., and J. Žilinskas (2008a). A hybrid method for multidimensional scaling using city-block distances. *Mathematical Methods of Operations Research*, **68**(3). doi:10.1007/s00186-008-0238-5.
- Žilinskas, A., and J. Žilinskas (2008b). Branch and bound algorithm for multidimensional scaling with city-block metric. *Journal of Global Optimization*. doi:10.1007/s10898-008-9306-x.
- Žilinskas, J. (2006). Multidimensional scaling in protein and pharmacological sciences. In: I.D.L. Bogle and J. Žilinskas (Eds.), *Computer Aided Methods in Optimal Design and Operations. Series on Computers and Operations Research*, World Scientific, Vol. 7. pp. 139–148.
- Žilinskas, J. (2007). Reducing of search space of multidimensional scaling problems with data exposing symmetries. *Information Technology and Control*, **36**(4), 377–382.
- Žilinskas, J. (2008). Branch and bound with simplicial partitions for global optimization. *Mathematical Modelling and Analysis*, **13**(1), 145–159.

J. Žilinskas is a principal researcher in Systems Analysis Department at the Institute of Mathematics and Informatics, Lithuania. His research interests include global optimization, parallel computing, data analysis and visualization. He is a member of editorial board of *Optimization Letters*.

Apie vaizdo erdvės matmenų skaičių daugiamatėse skalėse

Julius ŽILINSKAS

Daugiamatės skalės – tai tiriamasis daugiamatė duomenų analizės metodas, taikomas daugelyje sričių. Šiuo metodu ieškoma mažo matmenų skaičiaus vaizdo erdvės taškų, tarp kurių atstumai atitiktų duotus objektų skirtingumus. Šiame straipsnyje tiriama, kaip santykinė vizualizavimo paklaida priklauso nuo vaizdo erdvės matmenų skaičiaus. Dirbtinės geometrinės ir taikomosios duomenų aibės yra tiriamos. Trimatėse skalėse duomenų aibės objektų struktūrinių savybių atvaizdavimas yra priimtinas, dėl to stereo ekranų ir kitų erdvinio vizualizavimo priemonių populiarėjimas daro trimatės skales ypač patraukliomis.