ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2007, Vol. 36, No. 4

REDUCING OF SEARCH SPACE OF MULTIDIMENSIONAL SCALING PROBLEMS WITH DATA EXPOSING SYMMETRIES

Julius Žilinskas
Institute of Mathematics and Informatics
Akademijos 4, LT-08663 Vilnius, Lithuania

Abstract. Multidimensional scaling addresses the problem how multidimensional data can be represented by points in a low dimensional space. The problem is reduced to global minimization of a stress function which measures a fit of dissimilarity by the distances between the respective points. Symmetries in data may exist. Performance of global optimization may be increased reducing search space so that only one of the symmetric solutions should be found. Restriction of search space is proposed and demonstrated on geometric data sets for multidimensional scaling.

1. Introduction

Multidimensional scaling (MDS) is a technique for exploratory analysis of multidimensional data widely usable in different applications [2, 4]. It is assumed that pairwise dissimilarities between n objects are given by the matrix (δ_{ij}), $i, j = 1, \ldots, n$. A set of points in an embedding space is considered as an image of the set of objects. Normally, an m-dimensional embedding metric space is used, and points $\mathbf{x}_i \in \mathbf{R}^m$, $i = 1, \ldots, n$, should be found whose inter-point distances fit the given dissimilarities.

Frequently the objects are defined by multidimensional points and the dissimilarities are defined as pairwise distances between points in the original multidimensional space. In such a case an image can be interpreted as a nonlinear projection of the high-dimensional space to the space of lower dimensionality.

The problem of constructing the image of the set of considered objects is reduced to minimization of an accuracy of fit criterion, e.g. of the frequently used least squares $STRESS$ function

$$S(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(d\left(\mathbf{x}_i, \mathbf{x}_j\right) - \delta_{ij} \right)^2, \qquad (1)$$

where $\mathbf{x} = (x_{11}, x_{21}, \ldots, x_{nm})^T$ is an $(n \cdot m)$ vector aggregating coordinates of points $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})^T$, $i = 1, \ldots, n$; $d(\mathbf{x}_i, \mathbf{x}_j)$ denotes the distance between the points $\mathbf{x}_i$ and $\mathbf{x}_j$; weights are positive: $w_{ij} > 0, i, j = 1, \ldots, n$; and $d(\mathbf{x}_i, \mathbf{x}_j)$ is defined as Minkowski distance:

$$d_r(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^{m} |x_{ik} - x_{jk}|^r \right)^{1/r}. \qquad (2)$$

Equation (2) defines Euclidean distances when $r = 2$, and city-block distances when $r = 1$. The most frequently used distances are Euclidean. However, MDS with other Minkowski distances in the embedding space can be even more informative than MDS with Euclidean distances [1].

In the present paper MDS algorithms based on the $STRESS$ criterion with city-block distances in the embedding space are considered. $STRESS$ normally has many local minima. The $STRESS$ function is invariant with respect to translation, rotation and mirroring. In the case of city-block distances $STRESS$ can be non-differentiable even at a minimum point [7]. Therefore MDS with city-block distances is a difficult high-dimensional ($\mathbf{x} \in \mathbf{R}^N$, $N = n \times m$) global optimization problem.

A survey of city-block MDS was presented in [1]. Topics include theoretical issues, algorithmic developments, and methodological discussions. A combinatorial approach for city-block MDS was proposed in [5], where combinatorial local search to construct good object orders along dimensions, and least-squares to estimate the coordinates for the objects based on the object orders are used. A heuristic algorithm based on simulated annealing for two-dimensional city-block scaling was proposed in [3]. Each coordinate axis is partitioned by uniformly spaced points, and a simulated annealing algorithm is used to search the lattice defined by these points aiming to minimize one of two types of *STRESS* either the sum of squares or the sum of corresponding absolute values. The found solution is locally improved by quadratic programming. A two-level algorithm for two-dimensional city-block scaling has been proposed in [7], where evolutionary combinatorial global search is combined with local minimization employing piecewise quadratic structure of the objective function. Generalization of the algorithm from two-dimensional scaling to general case is discussed and visualization of three-dimensional images is demonstrated in [6]. In the same paper it was shown that explicit enumeration of all feasible solutions of the upper level combinatorial problem is computationally infeasible for all but the smallest problems.

In this paper restriction of search space is proposed for MDS problems with data exposing symmetries so that only one of the symmetric solutions should be found and computational complexity of explicit enumeration would be reduced. If exchange of some objects does not change dissimilarity data, exchange of points representing these objects does not change the value of the *STRESS* function. In this case equivalent solutions and equivalent subregions of the feasible region exist. In this paper restriction of search space is proposed and demonstrated on geometric data sets for MDS.

2. MDS with city-block distances

If city-block distances $d_1(\mathbf{x}_i, \mathbf{x}_j)$ are used, *STRESS* (1) can be redefined as

$$S(\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(\sum_{k=1}^{m}|x_{ik} - x_{jk}| - \delta_{ij}\right)^2. \quad (3)$$

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

$$A(\mathbf{P}) = \{\mathbf{x} | x_{ik} \le x_{jk} \text{ for } p_{ki} < p_{kj},$$
$$i, j = 1, \ldots, n,\ k = 1, \ldots, m\},$$

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

For $\mathbf{x} \in A(\mathbf{P})$, (3) can be rewritten as quadratic function

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

where

$$z_{kij} = \left\{\begin{array}{ll} 1, & p_{ki} > p_{kj}, \\ -1, & p_{ki} < p_{kj}, \end{array}\right| k = 1, \ldots, m.$$

Since the function $S(\mathbf{x})$ is quadratic over polyhedron $\mathbf{x} \in A(\mathbf{P})$, the minimization problem

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

is a quadratic programming problem which is equivalent to

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

A standard quadratic programming method can be applied to solve this problem.

The structure of the minimization problem suggests a two level minimization algorithm: to solve a combinatorial problem at the upper level, and to solve a quadratic programming problem at the lower level:

$$\min_{\mathbf{P}} S(\mathbf{P}), \quad (4)$$
$$\text{s.t. } S(\mathbf{P}) = \min_{\mathbf{x} \in A(\mathbf{P})} S(\mathbf{x}).$$

The upper level (4) objective function is defined over the set of m-tuple of permutations of $1, ..., n$. 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 around coordinate axes and exchanging of coordinates. The feasible set can be reduced taking into account such symmetries. Refusing mirrored solutions around each coordinate axis the number of feasible solutions can be reduced to u^m, where $u = n!/2$. It can be further reduced to approximately $u^m/m!$ refusing mirrored solutions with exchanged coordinates. Therefore the number of feasible solutions is equal to u in case $m = 1$, equal to $(u^2 + u)/2$ in case $m = 2$, equal to $(u^3 + 3u^2 + 2u)/6$ in case $m = 3$, and equal to $(u^4 + 6u^3 + 11u^2 + 6u)/24$ in case $m = 4$.

Theoretically the upper level combinatorial problem can be solved using different algorithms. In this paper explicit enumeration algorithm is considered.

3. Restriction of search space for problems exposing symmetries

If exchange of some objects does not change dissimilarity data, exchange of points representing these objects does not change the value of the *STRESS* function. In this case equivalent solutions and equivalent sub-regions of the feasible region exist. We propose to restrict the search space by constraints. In continuous optimization we propose to constrain the sequence of the first coordinate values of image points of symmetric objects. In combinatorial upper level algorithm of two-level optimization it would be equivalent to allowing only some of permutations of the first coordinate $\mathbf{p}_1$.

Sets of multidimensional points corresponding to the well understood geometric objects are used to evaluate performance of MDS algorithms. Examples of such data sets are vertices of multidimensional simplices and cubes. Global optimization problems of increasing complexity correspond to increasing dimensionality of the original space (dim). Below we use the shorthands 'simplex' and 'cube' for the sets of their vertices.

The number of vertices of multidimensional simplex is $n = \text{dim} + 1$, and the dimensionality of the corresponding global minimization problem is $N = m \times (\text{dim}+1)$. The distances between any two vertices of a standard simplex are equal: $\delta_{ij} = 1$, $i \neq j$. 2-dimensional and 3-dimensional standard simplices are shown in Figure 1. The dissimilarity matrix of such a seven-dimensional simplex is

$$\delta_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

If i-th row and column were simultaneously changed with j-th row and column, the dissimilarity matrix would not be changed. In other words exchange of any i-th and j-th objects does not change dissimilarity data, and exchange of points representing these objects does not change the value of the *STRESS* function. It is possible to restrict the search space to find only one of the symmetric solutions by constraining the sequence of values of the first coordinate of image points. In continuous optimization the constraints would be $x_{11} \leq x_{21} \leq \ldots \leq x_{n1}$, which are equivalent to one allowed permutation of the first coordinate in the upper level combinatorial problem:

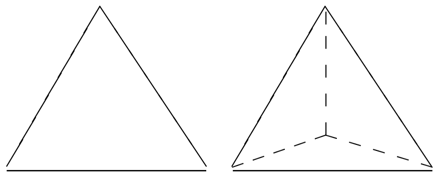

Figure 1. Standard simplices

$\mathbf{p}_1 = (1, 2, \ldots, n)$. In this case the number of feasible solutions is equal to 1 in case $m = 1$, equal to u in case $m = 2$, equal to $(u^2 + u)/2$ in case $m = 3$, and equal to $(u^3 + 3u^2 + 2u)/6$ in case $m = 4$. Computational complexity of explicit enumeration is reduced by approximately u/m times to approximately $u^{m-1}/(m-1)$.

Vertices of an unit simplex can be defined by

$$v_{ij} = \begin{cases} 1, & \text{if } i = j + 1, \\ 0, & \text{otherwise,} \end{cases} \quad \middle| \quad \begin{array}{l} i = 1, \ldots, \text{dim} + 1, \\ j = 1, \ldots, \text{dim}. \end{array}$$

2-dimensional and 3-dimensional unit simplices are shown in Figure 2. The dissimilarity between vertices is measured by city-block distance in the original vector space. For example, dissimilarity matrix of such a seven-dimensional simplex is

$$\delta_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 0 & 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 0 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 0 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 2 & 0 & 2 & 2 \\ 1 & 2 & 2 & 2 & 2 & 2 & 0 & 2 \\ 1 & 2 & 2 & 2 & 2 & 2 & 2 & 0 \end{pmatrix}.$$

If i-th ($i > 1$) row and column were simultaneously changed with j-th ($j > 1$) row and column, the dissimilarity matrix would not be changed. It is possible to restrict the search space to find only one of the symmetric solutions by constraining the sequence of values of the first coordinate of image points except one representing the vertex at the origin. In continuous optimization the constraints would be $x_{21} \leq x_{31} \leq \ldots \leq x_{n1}$, which are equivalent to n (which is further reduced to $\lceil n/2 \rceil$ by refusing mirrored solutions) allowed permutations of the first coordinate in the upper level combinatorial problem: $\mathbf{p}_1 = (l, 1, 2, \ldots, l-1, l+1, \ldots, n)$, $l \leq n/2$. In this case the number of feasible solutions is equal to $\lceil n/2 \rceil$ in case $m = 1$. It is not trivial to estimate the numbers in other cases of m, therefore experimental investigation is needed.

The number of vertices of a multidimensional cube is $n = 2^{\text{dim}}$, and the dimensionality of the corresponding global minimization problem is

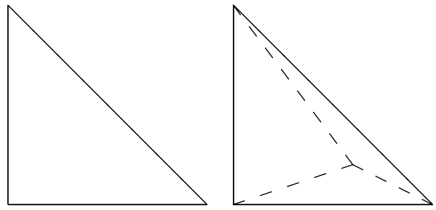

Figure 2. Unit simplices

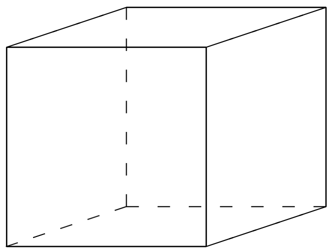

Figure 3. Cube

$N = m \times 2^{\text{dim}}$. The coordinates of i-th vertex of a dim-dimensional cube are equal either to 0 or to 1, and they are defined by a binary code of $i = 0, ..., n-1$. The dissimilarity matrix of a three-dimensional cube is

$$\delta_{ij} = \begin{pmatrix} 0 & 1 & 1 & 2 & 1 & 2 & 2 & 3 \\ 1 & 0 & 2 & 1 & 2 & 1 & 3 & 2 \\ 1 & 2 & 0 & 1 & 2 & 3 & 1 & 2 \\ 2 & 1 & 1 & 0 & 3 & 2 & 2 & 1 \\ 1 & 2 & 2 & 3 & 0 & 1 & 1 & 2 \\ 2 & 1 & 3 & 2 & 1 & 0 & 2 & 1 \\ 2 & 3 & 1 & 2 & 1 & 2 & 0 & 1 \\ 3 & 2 & 2 & 1 & 2 & 1 & 1 & 0 \end{pmatrix}.$$

It is difficult to define changes of objects which does not change dissimilarity matrix, but the data are symmetric as can be seen in Figure 3 where a (three-dimensional) cube is shown. It is possible to restrict the search space so that at least the vertex at the origin would be represented by the leftmost point in the image. In continuous optimization the constraints would be $x_{11} \le x_{i1}, i = 2, \ldots n$, which are equivalent to allowed permutations of the first coordinate with $p_{11} = 1$.

4. Experimental investigation

The efficiency of the two-level algorithm with explicit enumeration of combinatorial problem (with and without coping with symmetries of data) and standard quadratic programming method has been evaluated experimentally. The accuracy of scaling evaluated as minimum of $S(\mathbf{x})$ depends on n and $\delta_{ij}, i, j = 1, \ldots, n$. Therefore such a criterion is

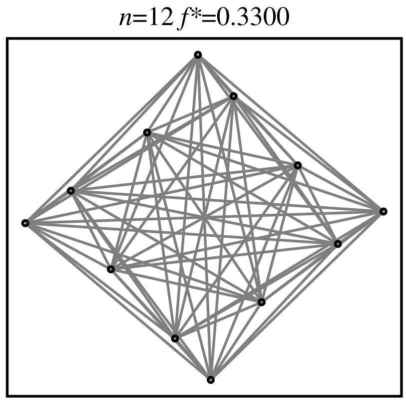

Figure 4: Image of a standard eleven-dimensional simplex ($n = 12$)

not very convenient to compare accuracies of scaling for different sets of objects. To reduce such an undesirable impact, a relative scaling error

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

is used in the experimental investigation below. Performance of deterministic global optimization algorithms is measured using the optimization time t, s, and the smallest relative scaling error f^*. We also present the number of the lower level quadratic programming problems solved (NQP).

Performance of the two-level MDS algorithm on standard simplices is shown in Table 1. Numbers of quadratic programming problems solved coincide with theoretical estimates derived in previous section. As it can be seen from the table, problems of $m = 1$ of standard simplices can be solved by solving one quadratic programming problem when the search space is restricted. The number of quadratic programming problems solved to find the global solution in m-dimensional scaling with restricted search space is equal to the number of quadratic programming problems solved in $(m-1)$-dimensional scaling of the original problem. Although the solution time in such a case is increased because of larger quadratic problems, the dimensionality of the global optimization problems solved in acceptable time is the same as for $(m-1)$-dimensional scaling for the original problem. In two-dimensional scaling $n = 12$ problem has been solved instead of $n = 8$ and in three-dimensional scaling $n = 8$ instead of $n = 6$ has been solved. The two-dimensional image of a standard eleven-dimensional simplex is shown in Figure 4. The dimensionality of the largest global optimization problems solved is $N = 24$.

Performance of the two-level MDS algorithm on unit simplices is shown in Table 2. For the origi-

Table 1. Experimental results on standard simplices

	$m = 1$		$m = 2$		$m = 3$	
n	t, s (NQP)	f^*	t, s (NQP)	f^*	t, s (NQP)	f^*
3	0.00 (3)	0.3333	0.00 (6)	0.00	0.00 (10)	0.00
4	0.00 (12)	0.4082	0.00 (78)	0.00	0.01 (364)	0.00
5	0.00 (60)	0.4472	0.03 (1830)	0.1907	1.79 (37820)	0.00
6	0.00 (360)	0.4714	1.71 (64980)	0.2309	589.72 (7840920)	0.00
7	0.03 (2520)	0.4879	118.59 (3176460)	0.2621		
8	0.21 (20160)	0.5000	10229.00 (203222880)	0.2825		
9	2.24 (181440)	0.5092				
10	26.63 (1814400)	0.5164				
11	351.09 (19958400)	0.5222				
12	4702.00 (239500800)	0.5270				
			Restricted search space			
3	0.00 (1)	0.3333	0.00 (3)	0.00	0.00 (6)	0.00
4	0.00 (1)	0.4082	0.00 (12)	0.00	0.00 (78)	0.00
5	0.00 (1)	0.4472	0.00 (60)	0.1907	0.09 (1830)	0.00
6	0.00 (1)	0.4714	0.01 (360)	0.2309	5.01 (64980)	0.00
7	0.00 (1)	0.4879	0.10 (2520)	0.2621	379.88 (3176460)	0.0945
8	0.00 (1)	0.5000	1.01 (20160)	0.2825	31681.00 (203222880)	0.1250
9	0.00 (1)	0.5092	11.89 (181440)	0.2991		
10	0.00 (1)	0.5164	153.88 (1814400)	0.3115		
11	0.00 (1)	0.5222	2121.56 (19958400)	0.3217		
12	0.00 (1)	0.5270	31170.00 (239500800)	0.3300		

nal problem the number of quadratic programming problems solved is the same as for the case of standard simplices. As it can be seen from the table, problems of $m = 1$ of unit simplices can be solved by solving $\lceil n/2 \rceil$ quadratic programming problems when the search space is restricted. Although the number of quadratic programming problems solved to find the global solution in m-dimensional scaling with the restricted search space is larger than the number of quadratic programming problems solved in $(m-1)$-dimensional scaling of the original problem, the dimensionality of the global optimization problems solved in acceptable time is the same as for $(m-1)$-dimensional scaling for the original problem. In two-dimensional scaling $n = 12$ problem has been solved instead of $n = 8$ and in three-dimensional scaling $n = 8$ instead of $n = 6$ has been solved. The two-dimensional image of standard eleven-dimensional simplex is shown in Figure 5. The dimensionality of the largest global optimization problems solved is again $N = 24$.

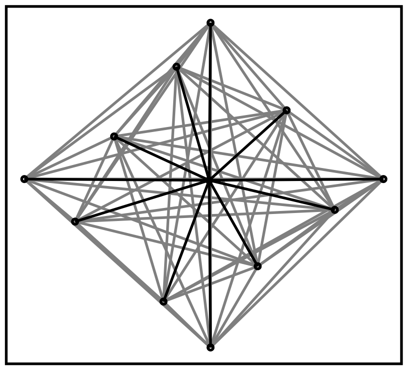

Figure 5: Image of unit eleven-dimensional simplex ($n = 12$)

Performance of the two-level MDS algorithm on two- and three-dimensional cubes is shown in Table 3. A small increase of performance is seen. A larger increase would be expected if all symmetries of such problems were considered.

5. Conclusions

$N = 24$ problems of global optimization have been solved exactly using two level algorithm for MDS with city-block distances composed of explicit enumeration of the upper level combinatorial problem and standard quadratic programming for the lower level with the search space restricted considering symmetries of geometric data sets. Without restriction the largest problems solved exactly with such an algorithm were with $N = 18$.

Table 2. Experimental results on unit simplices

	$m=1$		$m=2$		$m=3$	
n	t, s (NQP)	f^*	t, s (NQP)	f^*	t, s (NQP)	f^*
3	0.00 (3)	0.00	0.00 (6)	0.00	0.00 (10)	0.00
4	0.00 (12)	0.3651	0.00 (78)	0.00	0.01 (364)	0.00
5	0.00 (60)	0.4140	0.04 (1830)	0.00	2.02 (37820)	0.00
6	0.01 (360)	0.4554	2.05 (64980)	0.1869	661.11 (7840920)	0.00
7	0.02 (2520)	0.4745	137.12 (3176460)	0.2247		
8	0.23 (20160)	0.4917	11662.00 (203222880)	0.2569		
9	2.51 (181440)	0.5018				
10	29.78 (1814400)	0.5113				
11	378.45 (19958400)	0.5176				
12	5265.00 (239500800)	0.5236				
			Restricted search space			
3	0.00 (2)	0.00	0.00 (4)	0.00	0.00 (7)	0.00
4	0.00 (2)	0.3651	0.00 (18)	0.00	0.01 (99)	0.00
5	0.00 (3)	0.4140	0.01 (108)	0.00	0.14 (2574)	0.00
6	0.00 (3)	0.4554	0.02 (720)	0.1869	8.49 (101160)	0.00
7	0.00 (4)	0.4745	0.25 (5760)	0.2247	695.19 (5446080)	0.00
8	0.00 (4)	0.4917	2.90 (50400)	0.2569	66686.00 (381049200)	0.0992
9	0.00 (5)	0.5018	37.16 (504000)	0.2759		
10	0.00 (5)	0.5113	560.84 (5443200)	0.2936		
11	0.00 (6)	0.5176	7813.00 (65318400)	0.3058		
12	0.00 (6)	0.5236	122360.00 (838252800)	0.3167		

Table 3. Experimental results on hyper-cubes

	$m=1$		$m=2$		$m=3$	
n	t, s (NQP)	f^*	t, s (NQP)	f^*	t, s (NQP)	f^*
4	0.00 (12)	0.4082	0.00 (78)	0.00	0.02 (364)	0.00
8	0.24 (20160)	0.4787	12572.00 (203222880)	0.2245		
			Restricted search space			
4	0.00 (6)	0.4082	0.00 (57)	0.00	0.01 (308)	0.00
8	0.06 (5040)	0.4787	5483.00 (88908120)	0.2245		

Acknowledgement

The research is supported by the NATO Reintegration grant CBP.EAP.RIG.981300 and Lithuanian State Science and Studies Foundation.

References

[1] Arabie P (1991) Was Euclid an unnecessarily sophisticated psychologist? Psychometrika 56: 567–587

[2] Borg I, Groenen P (2005) Modern Multidimensional Scaling. 2nd ed. Springer, New York

[3] Brusco MJ (2001) A simulated annealing heuristics for unidimensional and multidimensional (city block) scaling of symmetric proximity matrices. Journal of Classification 18: 3–33

[4] Cox T, Cox M (2001) Multidimensional Scaling. Chapman and Hall/CRC, Boca Raton

[5] Hubert L, Arabie P, Hesson-Mcinnis M (1992) Multidimensional scaling in the city-block metric: a combinatorial approach. Journal of Classification 9: 211–236

[6] Žilinskas A, Žilinskas J (2006) On visualization of multidimensional data using three-dimensional embedding space. Technological and Economic Development of Economy 12: 353-359

[7] Žilinskas A, Žilinskas J (2007) Two level minimization in multidimensional scaling. Journal of Global Optimization 38: 581–596

Received September 2007.

DOI: 10.5755/j01.itc.36.4.11896