INFORMATICA, 2003, Vol. 14, No. 1, 121–130

On Multimodality of the *SSTRESS* Criterion for Metric Multidimensional Scaling

Antanas ŽILINSKAS

Institute of Mathematics and Informatics
Akademijos 4, LT-2021 Vilnius, Lithuania &
Cardiff University, UK
e-mail: antanasz@ktl.mii.lt

Aurelija PODLIPSKYTĖ

Vytautas Magnus University &
Institute of Psychophysiology and Rehabilitation
Vydūno 4, LT-5720 Palanga, Lithuania
e-mail: aurelija@ktl.mii.lt

Abstract. Recent publications on multidimensional scaling express contradicting opinion on multimodality of *STRESS* criterion. An example has been published with rigorously provable multimodality of *STRESS*. We present an example of data and the rigorous proof of multimodality of *SSTRESS* for this data. Some comments are included on widely accepted opinion that minimization of *SSTRESS* is easier than minimization of *STRESS*.

Key words: global optimization, multidimensional scaling, symbolic computation.

1. Introduction

Multidimensional scaling (MDS) is an approach to exploratory analysis and visualization of multidimensional data (Borg and Groenen, 1997; Cox and Cox, 2001; Mathar, 1997). In the present paper metric MDS is considered. Let dissimilarities between n objects are given by a symmetric dissimilarity matrix $\delta = (\delta_{ij})$, $i, j = 1, \ldots, n$, whose diagonal elements are equal to zero. The points in m-dimensional embedding space $x_1, \ldots, x_n$ are sought, $x_i \in R^m$, whose interpoint distances $d_{ij}(X) = ||x_i - x_j||$, $X = (x_1, \ldots, x_n)$, fit the given dissimilarities. To represent a vector X by means of coordinates of points x_i in the embedding space x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$ a row form $X = (x_{11}, \ldots, x_{n1}; x_{12}, \ldots, x_{n2}; \ldots, x_{nm})$ is used. Most frequently the measure of fit is defined either by the *STRESS* criterion

$$\sigma(X) = \sum_{i<j} w_{ij} \left(\delta_{ij} - d_{ij}(X)\right)^2, \tag{1}$$

proposed by Kruskal (1964) for the case of non metric scaling, or by so called *SSTRESS* criterion (see, e.g., Cox and Cox, 2001)

$$s(X) = \sum_{i<j} w_{ij} \left(\delta_{ij}^2 - d_{ij}^2(X)\right)^2, \tag{2}$$

where w_{ij} are the weights; in the present paper weights are assumed equal to 1. A $n \times m$ dimensional vector X_{min} should be found by means of minimization of the chosen criterion. Many authors agree that these minimization problems are difficult, particularly because of multimodality of the criteria, e.g., see the theses by Groenen (1993). Various global optimization methods have been tested to attack MDS problems (Groenen, 1993; Groenen and Heiser, 1996; Groenen, Heiser and Meulman, 1999; Mathar and Zilinskas 1993; Klock and Buhman, 2000). The results of comparison of efficiency of different algorithms are presented in (Mathar, 1996, 2000). A class of test functions based on MDS problems has been proposed in (Mathar and Zilinskas, 1994) for testing global optimization algorithms. However, the multimodality of MDS problems in all the mentioned cases has been demonstrated empirically, e.g., in (Groenen, 1993) a problem is described where 1098 local minima have been indicated by the stopping of the popular in the field algorithm SMACOF-I. On the other hand, the paper by Kearsley, Tapia and Trosset (1998) has cast doubts on the real multimodality of MDS problems showing that SMACOF-I frequently stops prematurely, but the Newton's method makes further progress from different found by SMACOF-I points towards a global minimizer. To resolve the controversy to the favor of multimodality of MDS problems, the examples should be found with rigorously provable multimodality. The problems mentioned earlier are not suited because of complexity of the analysis implied by their multidimensionality. An example of multimodality of *STRESS* is proposed in (Trosset and Mathar, 1997). In the present paper a similar example has been constructed for *SSTRESS*; for that purpose we have applied the symbolic computation tool of Maple7, see reference to Symbolic Computation Group.

2. Two Examples Concerning *STRESS*

The first example due to de Leeuw (1988) is defined by the dissimilarity matrix $\delta_{ij} = 1$, $i \neq j$. In the three dimensional embedding space the global minimizer comprises the vertices of a tetrahedron and has zero *STRESS*. In the two dimensional embedding space the global minimizer is believed to comprise the vertices of a square with the side length equal to $\frac{2+\sqrt{2}}{4}$. It was claimed that $X^* = (0, 2a, a, a, 0, 0, a\sqrt{3}, a\sqrt{3}/3)$ is a non global minimizer, where $a = (3+\sqrt{3})/8$. The configuration of points corresponding to X^* comprises the vertices of an equilateral triangle plus its centroid. However, it has been proved in (Trosset and Mathar, 1997) that X^* is a saddle point.

The example by Trosset and Mathar (1997) with at least two different local minimizers of *STRESS* is defined by the dissimilarity matrix $\delta_{12} = \delta_{14} = \delta_{23} = \delta_{34} = 1$, $\delta_{13} = \delta_{24} = \sqrt{2}$. A global minimizer (with the zero *STRESS*) obviously corresponds to the vertices of the unit square. It is proved in (Trosset and Mathar, 1997) that $X_2 = (0, b, 0, b, 0, 0, a, a)$ is a non global minimizer, where $b = (3+\sqrt{3})/6$, $a = b\sqrt{2}$.

3. An Example of Multimodality of *SSTRESS*

In the case of a smooth objective function the rigorous prove that a point is a local minimizer normally is based on the analysis of the gradient and of the matrix of second derivatives. It is easy to show that gardient and Hessian of $s(X)$ are defined by the following formulae

$$\nabla s(X) = \left(\frac{\partial s(X)}{\partial x_{11}}, \ldots, \frac{\partial s(X)}{\partial x_{n1}}, \frac{\partial s(X)}{\partial x_{12}}, \ldots, \frac{\partial s(X)}{\partial x_{n2}}, \ldots, \frac{\partial s(X)}{\partial x_{nm}}\right),$$

$$\frac{\partial s(X)}{\partial x_{ij}} = 4\sum_{k=1}^{n} w_{ik}\left(d_{ik}^2(X) - \delta_{ik}^2\right)(x_{ij} - x_{kj}),$$

$$H(X) = \begin{pmatrix} H_{11}(X) & \ldots & H_{1m}(X) \\ \ldots & \ldots & \ldots \\ H_{m1}(X) & \ldots & H_{mm}(X) \end{pmatrix}, \quad H_{ij}(X) = \left(\frac{\partial^2 s(X)}{\partial x_{ki}\partial x_{lj}}\right),$$

$$\frac{\partial^2 s(X)}{\partial x_{ki}^2} = \sum_{j=1}^{n} w_{kj}\left((d_{kj}^2(X) - \delta_{kj})^2 + 2(x_{ki} - x_{ji})^2\right),$$

$$\frac{\partial^2 s(X)}{\partial x_{ki}\partial x_{li}} = -4w_{kl}\left((d_{kl}^2(X) - \delta_{kl}^2) + 2(x_{ki} - x_{li})^2\right),$$

$$\frac{\partial^2 s(X)}{\partial x_{ki}\partial x_{lj}} = -8w_{kl}(x_{ki} - x_{li})(x_{kj} - x_{lj}), \quad i, j = 1, \ldots, m, \quad k, l = 1, \ldots, n.$$

However, the criteria (1) and (2) are invariant with respect to translations and rotations of the point configuration $x_1, x_2, \ldots, x_n$ in the embedding space involving the degeneracy of the matrices of second derivatives. It means that the local and global minimizers are not isolated, but minima are attained on some hypersurfaces. To exclude such an invariance some restrictions should be introduced. Configurations centered at the origin of the embedding space are considered by means of the majorization approach, see, e.g., (Borg and Groenen, 1997; Cox and Cox, 2001). In such a case only the invariance with respect to translations is excluded, but not the invariance with respect to rotations. To exclude the invariances of both kinds Mathar and Zilinskas (1993, 1994) have proposed to solve a constrained optimization problem whose feasible optimization region consists from centered configurations with orthogonal components, i.e.,

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

However, analysis of optimality conditions for restricted problems is more complicated than for unconstrained problems. Therefore, we will use here another approach normally used in distance geometry. Several coordinates of X are fixed equal to zero; the number of fixed coordinates is equal to the number of degrees of freedom in the original formulation of unrestricted optimization problem, see, e.g., (Kearsley, Tapia and Trosset, 1998).

Let us start with the example by Trosset and Mathar (1997) but concerning *SSTRESS* criterion. Four point configuration is considered in two dimensional embedding space; the data of the example is presented in previous section. Since the embedding space is two dimensional a configuration is invariant with respect of translations along two coordinates and with respect to rotation around the origin. Therefore three degrees of freedom should be excluded. Following the approach used in distance geometry we assume that the first point coincides with the origin, and the second point is on the horizontal axes:

$$x_{11} = x_{12} = x_{22} = 0. \tag{3}$$

A four point configuration in the two dimensional embedding space is defined by a five dimensional vector Z consisting of free coordinates of X which are denoted as follows $z_1 = x_{21}$, $z_2 = x_{31}$, $z_3 = x_{32}$, $z_4 = x_{41}$, $z_5 = x_{42}$. The *SSTRESS* criterion as a function of Z is denoted by $s(Z)$. The analysis of the stationary points of $s(Z)$ involves its gradients and matrices of second derivatives $\nabla s(Z)$ and $H(Z)$ which may be easily obtained from the formulae of $\nabla s(X)$ and $H(X)$ presented above.

We analyze stationary points of $s(Z)$ following the approach proposed by Trosset and Mathar (1997). It is assumed there that at least some of local minimizers correspond to the rectangular configurations of points in the two dimensional embedding space. This is equivalent to the assumption that either $z_1 = z_2 = u$, $z_3 = z_5 = v$, $z_4 = 0$ or $z_1 = z_4 = u$, $z_3 = z_5 = v$, $z_2 = 0$, enabling to reduce the minimization of $s(Z)$ to the minimization of *SSTRESS* with respect to only two variables (u, v). The first case version of $s(Z)$ may be expressed as

$$s(Z) = S_1(u, v) = 2(u^2 - 1)^2 + 2(u^2 + v^2 - 2)^2 + 2(v^2 - 1)^2,$$

and the second case version of $s(Z)$ may be expressed as

$$s(Z) = S_2(u, v) = 2(u^2 - 1)^2 + 2(u^2 + v^2 - 1)^2 + (v^2 - 2)^2.$$

We start with analytical solution of the systems of equations $\nabla S_1(u, v) = 0$ and $\nabla S_2(u, v) = 0$ defining the stationary points of *SSTRESS* for the set of rectangular configurations

$$\begin{aligned} &4u^3 - 6u + 2uv^2 = 0, \\ &4v^3 - 6v + 2u^2v = 0, \end{aligned}$$

and

$$\begin{aligned} &4u^3 - 4u + 2uv^2 = 0, \\ &4v^3 - 6v + 2u^2v = 0. \end{aligned}$$

The first system of equations has the following solutions

$$u_1 = 1,\ \ v_1 = 1,\ \ u_2 = 0,\ \ v_2 = 0,\ \ u_3 = 0,\ \ v_3 = \sqrt{\frac{3}{2}},\ \ u_4 = \sqrt{\frac{3}{2}},\ \ v_4 = 0,$$

and the second system of equations has the following solutions

$$u_2 = 0,\ \ v_2 = 0,\ \ u_5 = 1,\ \ v_5 = 0,\ \ u_3 = 0,\ \ v_3 = \sqrt{\frac{3}{2}},\ \ u_6 = \sqrt{\frac{1}{3}},\ \ v_6 = \sqrt{\frac{2}{3}}.$$

Six stationary points found in the two dimensional space (u, v) are presented in the Table 1. However, these points have been found assuming the configuration rectangular. A point Z corresponding to a two dimensional stationary point need not be a stationary point of $s(Z)$ considered as a function in the five dimensional space. Therefore the obtained solutions should be tested by means of analysis of gradients and Hessians of $s(Z)$ at the corresponding five dimensional points.

The classification of the stationary points Z_i is important to draw conclusions on multimodality of $s(Z)$ in the five dimensional space, defined as *SSTRESS* minimization space by means of fixing some variables (3). Analytical investigation of a gradient and a Hessian of $s(Z)$ is difficult because of complexity of corresponding formulae. The analysis of necessary minimum conditions at the candidate points may be aided by the symbolic computation tool Maple, see reference to (Symbolic Computation Group). Let us specify the goal of the analysis. The two dimensional vectors (u_i, v_i) in Table 1 and (3) define a four point configurations in the two dimensional embedding space. On the other hand, the (u_i, v_i) corresponds to the five dimensional point Z_i, and the latter is a candidate to be a stationary point. The testing of the optimality conditions at the point Z_i includes testing of equality $\nabla s(Z_i) = 0$ and positve definiteness of $H(Z_i)$. Of course, gradient and Hessian of $s(Z)$ may be obtained using general formulae presented above for $s(X)$. However, such analytical manipulations are rather cumbersome because $s(Z)$ does not posses symmetry specific for $s(X)$, which is lost by fixing specific values of some variables (3). The formula of $s(Z)$ has been programmed in Maple7. Using the symbolic differentiation subroutine **diff** the formulae of the gradient and of the Hessian have been obtained. The calculation of values of the gradient norm has shown that all points Z_i are stationary points except the point Z_5. To test definiteness the eigenvalues of $H(Z_i)$ have been computed. The results confirmed the obvious fact that points Z_1 and Z_2 are global minimizer and local maximizer correspondingly. The points Z_3 and Z_4 are saddle points. Because of the similarity of our case with the case in (Trosset and Mathar, 1997) the point Z_6 might be expected to be a local minimizer. However, it is a saddle

Table 1

Stationary points in the two dimensional space (u, v)

Point number	1	2	3	4	5	6
u	1	0	0	$\sqrt{3/2}$	1	$1/\sqrt{3}$
v	1	0	$\sqrt{3/2}$	0	0	$2/\sqrt{3}$
s	0	12	3	3	4	2 2/3

point. The Hessian and the vector of its eigenvalues Λ_6 are presented below

$$H(Z_6) = \begin{pmatrix} c & -d & d & c & 0 \\ -d & c & -d & 0 & 0 \\ d & -d & e & 0 & a \\ c & 0 & 0 & c & d \\ 0 & 0 & a & d & e \end{pmatrix}, \tag{4}$$

$$\Lambda_6 = (-3.484,\, 0.802,\, 5.119,\, 18.694,\, 24.203),$$

where $c = 2\frac{2}{3}$, $d = 5\frac{1}{3}$, $e = 18\frac{2}{3}$, and eigenvalues are presented with the truncated precision 10^{-3}, although they were calculated by the procedure **Eigenvals** of Maple7 with the precision 10^{-10}.

We must conclude that for the data of (Trosset and Mathar, 1997) $s(Z)$ is unimodal over the region corresponding to configurations comprising vertices of rectangles. In this special case *STRESS* and *SSTRESS* behave differently: the latter is unimodal, but the former has two different local minima. To test $s(Z)$ for the multimodality in a practically interesting region, $s(Z)$ has been minimized by means of the damped Newton's method 100 times from random starting points uniformly distributed in the hypercube $-2 \leqslant z_i \leqslant 2$, $i = 1, \ldots, 5$. A MATLAB code has been used with the formulae of first and second derivatives imported from Maple7. In all cases the descent was convergent to the point Z_1. Therefore, the hypothesis of unimodality of $s(Z)$ for the data of (Trosset and Mathar, 1997) can not be rejected.

Although the analysis of the example by Mathar and Trosset (1997) has not supported the hypothesis on multimodality of *SSTRESS*, it has induced an idea of the following example:

$$\delta_{12} = \delta_{14} = \delta_{23} = \delta_{34} = 1, \quad \delta_{13} = \delta_{24} = 1.2. \tag{5}$$

We have not carried out in this case an analytical investigation of stationary points of $s(Z)$ corresponding to rectangular configurations, since we had no intention to compare the behaviour of *SSTRESS* and *STRESS* for the data set (5). Instead, a version of the damped Newton's method mentioned above has been used to minimize $s(Z)$ with the data (5) from the starting points Z_1 and Z_6. The approximations of two different local minimizers have been found. The latter have been used to define the ranges of the variables for the Maple7 procedure **fsolve** used to solve the system of equations $\nabla s(Z) = 0$ with precision 10^{-10}. The solutions are presented below with the truncated precision 10^{-3}:

$$s(Z_7) = 0.209, \quad Z_7 = (0.902, 0.902, 0.902, 0, 0.902),$$
$$s(Z_8) = 1.382, \quad Z_8 = (0.721, 0, 0.980, 0.721, 0.980).$$

Both points are local minimizers, as follows from the eigenvalues of the matrices of the second derivatives presented below:

$$H(Z_7) = \begin{pmatrix} 12.267 & 0.747 & 0 & -7.253 & 6.507 \\ 0.747 & 12.267 & 6.507 & -5.760 & 0 \\ 0 & 6.507 & 12.267 & 0 & 0.747 \\ -7.253 & -5.760 & 0 & 12.267 & -6.507 \\ 6.507 & 0 & 0.747 & -6.507 & 12.267 \end{pmatrix},$$

$$\Lambda_7 = (1.759, 5.529, 8.611, 18.231, 27.203),$$

$$H(Z_8) = \begin{pmatrix} 6.400 & -6.080 & 5.652 & 1.920 & 0 \\ -6.080 & 6.400 & -5.652 & -2.240 & 0 \\ 5.652 & -5.652 & 13.440 & 0 & 1.920 \\ 1.920 & -2.240 & 0 & 6.400 & 5.652 \\ 0 & 0 & 1.920 & 5.652 & 13.440 \end{pmatrix},$$

$$\Lambda_8 = (0.294, 1.440, 6.727, 15.683, 21.936).$$

The configurations corresponding to the global optimum point Z_7 and to the local optimum point Z_8 are shown in Fig. 1 by $+$ and $\times$ correspondingly.

The results of analysis of two examples above confirm the empirical evidence that minimization of *SSTRESS* is normally less complicated with respect to multimodality than minimization of STRESS. The number of local minimizers of *SSTRESS* increases when increases the difference in structure of dissimilarities data and a structure of distances in embedding space. The hypersurface of *SSTRESS* may have many saddle points. Therefore a version of Newton method (Trosset and Mathar, 1997) or a version of conjugate gradient method (Zilinskas, 1996) seem most appropriate for local minimization of *SSTRESS*.

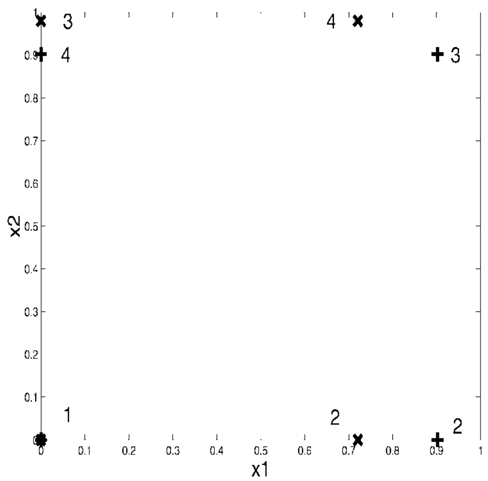

Fig. 1. Optimal configuration is shown by $+$, and suboptimal configuration is shown by $\times$.

4. Conclusions

There exist rather simple examples of the rigorously provable multimodality of *SSTRESS*. Although hypersurface of *SSTRESS* normally is less picky than hypersurface of *STRESS*, minimization of *SSTRESS* may be difficult for the algorithms not using information on second derivatives. The difficulties may be caused by the saddle points, particularly corresponding to the coalescence of x_i and x_j, $i \neq j$ not possible for *STRESS*. The results of the present paper support the opinion that the development of efficient algorithms for the metric MDS may be prospective in combining global search strategies with local descent guided by the quadratic model of an objective function.

References

Borg, I., and P. Groenen (1997). *Modern Multidimensional Scaling: Theory and Applications*. Springer.

Cox, T., and M. Cox (2001). *Multidimensional Scaling*. Chapman and Hall.

Groenen, P. (1993). *Majorization Approach to Multidimensional Scaling: Some Problems and Extensions*. DSWO Press, Leiden.

De Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification*, **5**, 163–180.

Groenen, P., and W. Heiser (1996). The tunneling method for global optimization in multidimensional scaling. *Psychometrica*, **61**, 529–550.

Groenen, P., W. Heiser and J. Meulman (1999). Global optimization in least-squares multidimensional scaling by distance smoothing. *Journal of Classification*, **16**, 225–254.

Kearsley, A., R. Tapia and M. Trosset (1998). The solution of the metric *STRESS* and *SSTRESS* problems in multidimensional scaling using newton's method. *Computational Statistics*, **13**, 369–396.

Klock, H., J. Buhmann (2000). Data visualization by multidimensional scaling: deterministic annealing approach. *Pattern Recognition*, **33**, 651–669.

Kruskal, J. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrica*, **29**, 115–129.

Mathar, R. (1996). A hybrid global optimization algorithm for multidimensional scaling. In R. Klar and O. Opitz (Eds.), *Classification and Knowledge Organization*. Springer.

Mathar, R. (1997). *Multidimensionale Skalierung*. Teubner, Stuttgart.

Mathar, R. (2000). Global optimization method for multidimensional scaling applied to mobile communications. In W. Gaul and O. Opitz (Eds.), *Data Analysis: Scientific Modelling and Practical Applications*. Springer. pp. 459–475.

Mathar, R., and A. Zilinskas (1993). On global optimization in two dimensional scaling. *Acta Applicandae Mathematica*, **33**, 109–118.

Mathar, R., and A. Zilinskas (1994). A class of test functions for global optimization. *Journal of Global Optimization*, **5**, 195–199.

Symbolic Computation Group. `http://scg.uwaterloo.ca`.

Trosset, M., and R. Mathar (1997). On existence of nonglobal minimizers of the stress criterion for metric multidimensional scaling. *Proceedings of the Statistical Computing Section*, American Statistical Association, Alexandria, VA, 158–162.

Zilinskas, A. (1996). A quadratically converging algorithm of multidimensional scaling. *Informatica*, **7**(2), 268–274.

A. Žilinskas is a senior researcher of the Institute of Mathematics and Informatics and a professor and head of Applied Informatics Department of Vytautas Magnus University. He received doctor degree from Kaunas University of Technology in 1973 and degree of doctor habilitus from St. Petersburg University in 1985. His research interests are global optimization, optimal design and visualization of multidimensional data. A. Žilinskas is a member of IEEE, Computer Society, American Mathematical Society, IFIP working group WG7.6, Lithuanian Computer Union. He is a member of editorial boards of Journal of Global Optimization, Control and Cybernetics, and Informatica. He was visiting professor/researcher at Dortmund, Aachen Technical, Copenhagen and London Universities. Recently he worked at Cardiff university.

A. Podlipskytė is a doctor degree student at Vytautas Magnus University and a researcher at the Institute of Psychophysiology and Rehabilitation in Palanga. She graduated Kaunas University of Technology in 1994. Her research interests are in visualization and analysis of biomedical data.

Daugiadimensinių skalių SSTRESS kriterijaus daugiaekstremalumas

Antanas ŽILINSKAS, Aurelija PODLIPSKYTĖ

Pastarųjų metų publikacijose išryškėjo prieštara dėl daugiamačių skalių sudaryme vartojamų kriterijų STRESS bei SSTRESS daugiaekstremalumo. Neseniai buvo paskelbtas pavyzdys, kuriam STRESS daugiaekstremalumas griežtai įrodytas. Šiame straipsnyje sukonstruotas panašus SSTRESS daugiaekstremalumo pavyzdys. Diskutuojama plačiai paplitusi nuomonė, kad SSTRESS minimizavimo uždavinys paprastesnis už STRESS minimizavimo uždavinį.