A Novel Geometric Approach to the Problem of Multidimensional Scaling

Gintautas Dzemyda(✉) and Martynas Sabaliauskas

Vilnius University Institute of Data Science and Digital Technologies,
Akademijos Street 4, 08412 Vilnius, Lithuania
{gintautas.dzemyda,martynas.sabaliauskas}@mii.vu.lt

Abstract. Multidimensional scaling (MDS) is one of the most popular methods for a visual representation of multidimensional data. A novel geometric interpretation of the stress function and multidimensional scaling in general (Geometric MDS) has been proposed. Following this interpretation, the step size and direction forward the minimum of the stress function are found analytically for a separate point without reference to the analytical expression of the stress function, numerical evaluation of its derivatives and the linear search. It is proved theoretically that the direction coincides with the steepest descent direction, and the analytically found step size guarantees the decrease of stress in this direction. A strategy of application of the discovered option to minimize the stress function is presented and examined. It is compared with SMACOF version of MDS. The novel geometric approach will allow developing a new class of algorithms to minimize MDS stress, including global optimization and high-performance computing.

Keywords: Multidimensional scaling · Geometric approach · Minimization · Analytical derivatives · Analytical step size · Geometric MDS

1 Introduction

Recent approaches to minimize the stress in multidimensional scaling (MDS) suggest wide possibilities for dimensionality reduction [1,2]. Recently, it finds applications of various nature: face recognition [3], analysis of regional economic development [4], image graininess characterization [5].

Suppose, we have a set $X = \{X_i = (x_{i1}, \ldots, x_{in}),\ i = 1, \ldots, m\}$ of n-dimensional data points (observations) $X_i \in \mathbb{R}^n,\ n \geqslant 3$.

Dimensionality reduction and visualization requires estimating the coordinates of new points $Y_i = (y_{i1}, \ldots, y_{id})$, $i = 1, \ldots, m$, in a lower-dimensional space ($d < n$) by holding proximities δ_{ij} between multidimensional points X_i and $X_j,\ i, j = 1, \ldots, m$, as much as possible. Proximity δ_{ij} can be measured e.g. by the distance between X_i and X_j.

Y. D. Sergeyev and D. E. Kvasov (Eds.): NUMTA 2019, LNCS 11974, pp. 354–361, 2020.
https://doi.org/10.1007/978-3-030-40616-5_30

The input data for MDS consists of the symmetric $m \times m$ matrix $\mathbf{D} = \{d_{ij}, i, j = 1, \ldots, m\}$ of proximities between pairs of points X_i and X_j. If the Minkowski distance is used as the proximity, then

$$d_{ij} = \left(\sum_{k=1}^{n} |x_{ik} - x_{jk}|^q \right)^{\frac{1}{q}}, \quad 1 \leqslant i, j \leqslant m. \tag{1}$$

If $q = 1$, then (1) defines the city-block or Manhattan distance. If $q = 2$, (1) becomes the Euclidean distance.

MDS finds the coordinates of new points Y_i representing X_i in a lower-dimensional space $\mathbb{R}^d$ by minimizing the multimodal stress function. Consider the raw stress function [6]:

$$S(Y_1, \ldots, Y_m) = \sum_{i=1}^{m} \sum_{j=i+1}^{m} (d_{ij} - d^*_{ij})^2, \tag{2}$$

where d^*_{ij} is the Euclidean distance between points Y_i and Y_j in a lower dimensional space. In (2), other proximities may be used as well. The MDS-based dimensionality reduction optimization problem may be formulated as follows:

$$\min_{Y_1, \ldots, Y_m \in \mathbb{R}^d} S(Y_1, \ldots, Y_m). \tag{3}$$

In case $1 \leqslant d < n$, the stress function has many local minima, often. The optimization problem (3) can be solved using well-known descent methods, e.g. Quasi-Newton or conjugate gradient methods [7]. However, these algorithms cannot guarantee to find a global minimum.

Various attempts to find the global minimum are suggested. However, they are computational expensive and do not guarantee to find the global minimum, too. This lead to the conclusion that the classical approaches [8–10] to minimize the stress reached their limits in this sense. New viewpoint to the problem is necessary, including its formulation and ways of solving.

In this paper, a novel geometric interpretation of the stress function and multidimensional scaling has been proposed. It will allow developing a new class of algorithms to minimize MDS stress, including global optimization and high-performance computing. Denote this approach by Geometric MDS.

2 The Geometric Approach – Geometric MDS

A new approach, Geometric MDS, has been developed to minimize the stress function (2). Suppose, we have $m \times m$ matrix $\mathbf{D} = \{d_{ij},\ i, j = 1, \ldots, m\}$ of proximities (e.g. distances) between n-dimensional points $X_i = (x_{i1}, \ldots, x_{in}),\ i = 1, \ldots, m$. We aim to find two-dimensional points $Y_i = (y_{i1}, ..., y_{id}),\ i = 1, \ldots, m$ by solving (3).

At first, let's have some initial configuration of points $Y_1, \ldots, Y_m$. Then, let's optimize the position of the particular point Y_j when the position of remaining

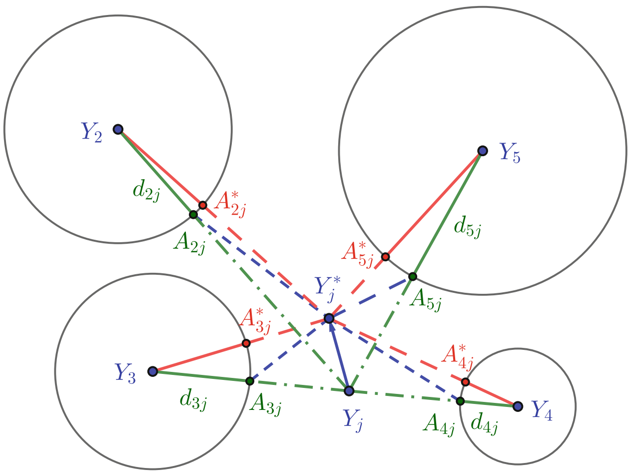

Fig. 1. An example of a single step of geometric method.

points $Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_m$ is fixed. In this case, we tend to minimize $S(\cdot)$ in (3) by minimizing the so-called local stress function $S^*(\cdot)$ depending on Y_j, only:

$$S^*(Y_j) = \sum_{\substack{i=1\\i\neq j}}^{m} \left(d_{ij} - \sqrt{\sum_{k=1}^{d} (y_{ik} - y_{jk})^2} \right)^2. \tag{4}$$

Figure 1 illustrates an example, where $m = 5$, $d = 2$. The location of points $Y_1, \ldots, Y_m$ and proximities d_{ij}, $i, j = 1, \ldots, m$ between points $X_1, \ldots, X_m$ are chosen such for better illustration of the idea. Position of point Y_1 is optimized. Y_1 is denoted by Y_j in Fig. 1 seeking for the better correspondence with notations in (4). In the centre of each circle, we have a corresponding point Y_i. Radius of the i-th circle is equal to the proximity d_{ij} between the points X_i and X_j in n-dimensional space. Point A_{ij} lies on the line between Y_i and Y_j, $i \neq j$, i.e. vectors $\overrightarrow{Y_i A_{ij}}$ and $\overrightarrow{A_{ij} Y_j}$ are collinear. Denote a new position of Y_j by Y^*_j. Let Y^*_j be chosen so that

$$\text{(a) vectors } \overrightarrow{Y_i A^*_{ij}} \text{ and } \overrightarrow{A^*_{ij} Y^*_j} \text{ are collinear, } i \neq j, \tag{5}$$

$$\text{(b) } Y^*_j = \frac{1}{m-1} \sum_{\substack{i=1\\i\neq j}}^{m} A_{ij}. \tag{6}$$

We will analyse the value of the local stress function $S^*(Y^*_j)$ and compare it with the value $S^*(Y_j)$. According to (6), Y^*_j is an average point of the points A_{ij} over $i = 1 \ldots m, i \neq j$. According to (5), when we make a step from Y_j to Y^*_j, we get new intersection points A^*_{ij} on circles that correspond to Y_j, and these points are on the line between Y_i and Y^*_j.

Proposition 1. *The gradient of local stress function $S^*(\cdot)$ is as follows:*

$$\nabla S^*|_{Y_j} = \left(2\sum_{\substack{i=1\\i\neq j}}^{m} \frac{d_{ij} - \sqrt{\sum_{l=1}^{d}(y_{il}-y_{jl})^2}}{\sqrt{\sum_{l=1}^{d}(y_{il}-y_{jl})^2}}(y_{ik}-y_{jk}),\ k=1,\ldots,d\right).$$

The proof follows from (4) by differentiating $S^*(\cdot)$.

Proposition 2. *The step direction from Y_j to Y_j^* corresponds to the anti-gradient of the function $S^*(\cdot)$ at the point Y_j:*

$$Y_j^* = Y_j - \frac{1}{2(m-1)}\nabla S^*|_{Y_j}. \tag{7}$$

Proof

$$Y_j^* - Y_j = \left(\frac{1}{m-1}\sum_{\substack{i=1\\i\neq j}}^{m}\left(\frac{d_{ij}(y_{jk}-y_{ik})}{\sqrt{\sum_{l=1}^{d}(y_{il}-y_{jl})^2}} + y_{ik} - y_{jk}\right),\ k=1,\ldots,d\right)$$

$$= \left(-\frac{1}{2(m-1)}2\sum_{\substack{i=1\\i\neq j}}^{m}\frac{d_{ij} - \sqrt{\sum_{l=1}^{d}(y_{il}-y_{jl})^2}}{\sqrt{\sum_{l=1}^{d}(y_{il}-y_{jl})^2}}(y_{ik}-y_{jk}),\ k=1,\ldots,d\right)$$

$$= -\frac{\nabla S^*|_{Y_j}}{2(m-1)}. \quad \square$$

Proposition 3. *Size of a step from Y_j to Y_j^* is equal to*

$$\frac{||\nabla S^*|_{Y_j}||}{2(m-1)} = \frac{1}{m-1}\sqrt{\sum_{k=1}^{d}\left(\sum_{\substack{i=1\\i\neq j}}^{m}\frac{d_{ij} - \sqrt{\sum_{l=1}^{d}(y_{il}-y_{jl})^2}}{\sqrt{\sum_{l=1}^{d}(y_{il}-y_{jl})^2}}(y_{ik}-y_{jk})\right)^2}.$$

Proposition 4. *Let Y_j does not match to any local extreme point of the function $S^*(\cdot)$. If Y_j^* is chosen by (6), then a single step from Y_j to Y_j^* reduces a local stress $S^*(\cdot)$:*

$$S^*(Y_j^*) < S^*(Y_j).$$

Proof. Let's have following functions:

$$S^*(Y_j) = \sum_{\substack{i=1\\i\neq j}}^{m} d^2(A_{ij}, Y_j), \tag{8}$$

$$S_A^*(Y_j^*) = \sum_{\substack{i=1\\i\neq j}}^{m} d^2(A_{ij}, Y_j^*),\ S^*(Y_j^*) = \sum_{\substack{i=1\\i\neq j}}^{m} d^2(A_{ij}^*, Y_j^*). \tag{9}$$

where $d(\cdot\ ,\cdot)$ is the Euclidean distance between two points.

Figure 1 illustrates a case, where position of point Y_j is optimized to Y_j^*. It is enough to show that

$$S^*(Y_j^*) < S_A^*(Y_j^*) < S^*(Y_j).$$

Firstly, we show that $S_A^*(Y_j^*) < S^*(Y_j)$. Define $A_{ij} = (a_{ij1}, \dots, a_{ijd})$. From (8), it follows that the gradient of $S^*(Y_j)$ is equal to

$$\nabla S^*(Y_j) = \Big(\sum_{\substack{i=1 \\ i \neq j}}^{m} 2(a_{ijk} - y_{jk}),\ k = 1, \dots, d \Big).$$

At the local minimum Y_j of function $S^*(Y_j)$, the condition $\nabla S^*(Y_j) = (0, \dots, 0)$ is valid, and then we have a unique solution of Y_j:

$$(m-1)y_{jk} - \sum_{\substack{i=1 \\ i \neq j}}^{m} a_{ijk} = 0,\ k = 1, \dots, d \implies (m-1)Y_j - \sum_{\substack{i=1 \\ i \neq j}}^{m} A_{ij} = 0.$$

We see that the solution is defined as Y_j^*, which is given in (6). Such Y_j^* corresponds to minimized local stress $S_A^*(Y_j^*)$. Therefore, $S_A^*(Y_j^*) < S^*(Y_j)$.

For the proof that $S^*(Y_j^*) < S_A^*(Y_j^*)$, it is enough to show that

$$d(Y_j^*, A_{ij}^*) < d(Y_j^*, A_{ij}), \quad i = 1, \dots, m, \quad i \neq j.$$

Using the triangle inequality, we have a valid condition

$$d(Y_i^*, Y_j^*) = d(Y_i^*, A_{ij}^*) + d(A_{ij}^*, Y_j^*) < d(Y_i^*, A_{ij}) + d(A_{ij}, Y_j^*).$$

Since the radius of the i-th circle satisfies condition $d(Y_i^*, A_{ij}^*) = d(Y_i^*, A_{ij})$, then $d(Y_j^*, A_{ij}^*) < d(Y_j^*, A_{ij})$. □

Proposition 5. *The value of the local stress function $S^*(\cdot)$ (4) will converge to a local minimum when repeating steps (7) and $Y_j := Y_j^*$.*

Proposition 6. *Let Y_j does not match to any local extreme point of the function $S^*(\cdot)$. Movement of any projected point by the geometric method reduces the stress (2) of MDS: if Y_j^* is chosen by (6), then the stress function $S(\cdot)$, defined by (2), decreases:*

$$S(Y_1, \dots, Y_{j-1}, Y_j^*, Y_{j+1}, \dots, Y_m) < S(Y_1, \dots, Y_{j-1}, Y_j, Y_{j+1}, \dots, Y_m).$$

Proof. Before the step from Y_j to Y_j^*, we have following stress function

$$S(Y_1, \dots, Y_{j-1}, Y_j, Y_{j+1}, \dots, Y_m) = S^*(Y_j) + \sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{\substack{k=i+1 \\ k \neq j}}^{m} (d_{ik} - d_{ik}^*)^2.$$

Since $S^*(Y_j^*) < S^*(Y_j)$ and $\sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{\substack{k=i+1 \\ k \neq j}}^{m} (d_{ik} - d_{ik}^*)^2$ remain constant after the step, the stress function $S(\cdot)$ is reduced after the step. □

3 Multimodality of the Local Stress Function of Geometric MDS

Proposition 7. *Function* $f(\delta) = S^* \left(Y_j - \delta \frac{\nabla S^*|_{Y_j}}{||\nabla S^*|_{Y_j}||} \right)$ *is not unimodal, where* δ *is a step size.*

Proof. Consider a dataset X of six five-dimensional points and their Euclidean distances as proximities:

$X_1 = (3.142, 2.718, 1.618, 1.202, 0.2078)$, $X_2 = (16.462, 2.718, 1.618, 1.202, 0.2078)$, $X_3 = (3.142, 7.648, 1.618, 1.202, 0.2078)$, $X_4 = (3.142, 2.718, 4.818, 1.202, 0.2078)$, $X_5 = (3.142, 2.718, 1.618, 4.952, 0.2078)$, $X_6 = (3.142, 2.718, 1.618, 1.202, 4.0278)$.

Let the values of Y $(d = 2)$ be such:

$Y_1 = (18.723, -1.880)$, $Y_2 = (19.025, 6.247)$, $Y_3 = (12.147, 11.208)$, $Y_4 = (11.338, 2.585)$, $Y_5 = (3.909, 3.546)$, $Y_6 = (10.560, -4.654)$. Consider point Y_1 for its moving to a new position Y_1^* according to the anti-gradient direction by (7). See Fig. 2 for details. The local stress function reaches its two different local minima depending on the step δ. □

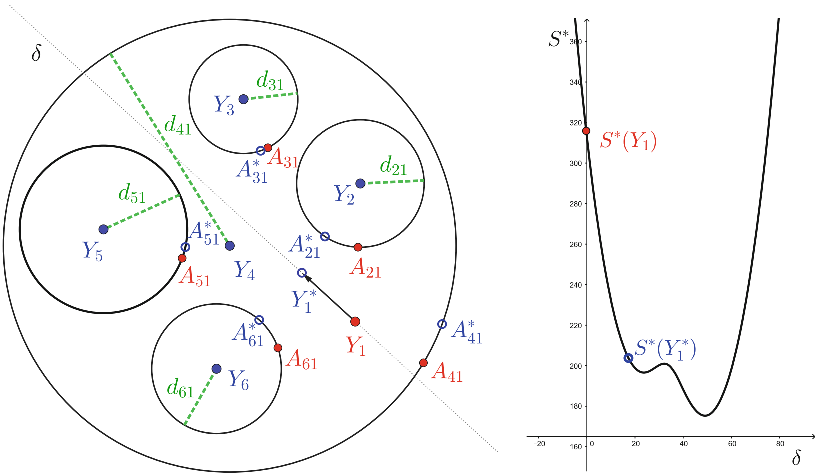

Fig. 2. Example of the anti-gradient search

4 Experiments with Geometric MDS

Simple realizations of Geometric MDS are based on fixing some initial positions of points $Y_i = (y_{i1}, \dots, y_{id})$, $i = 1, \dots, m$ (at random, using principal component analysis, etc.), and further changing the positions of Y_j (once by (7) or multistep descent by several steps using (7)) in consecutive order from $j = 1$ to $j = m$

many times till some stop condition is met: e.g. number of runs from $j = 1$ to $j = m$ reaches some limit or the decrease of stress function $S(\cdot)$ becomes less than some small constant after two consecutive runs.

In the experiments, minimization of stress $S(\cdot)$ was performed by consecutive one-step (not multistep) changing of positions of points $Y_1, \ldots, Y_m$ many times. 1000 random sets X of 30 points ($m = 30$) were generated inside the 4-dimensional unit hypercube ($n = 4$) and represented in $d = 2$ and $d = 3$ spaces. For comparison, the same data sets were analysed by multidimensional scaling based on stress $S(\cdot)$ minimization using majorization (SMACOF) that is realized in R [11,12]. Both Geometric MDS and SMACOF used the same initial values of points $Y_1, \ldots, Y_m$ obtained by Torgerson Scaling [13] realized in R [14].

When $d = 2$, Geometric MDS and SMACOF gave the same results (stress values) in 997 cases, however the average value of $S(\cdot)$ is obtained a bit better by Geometric MDS and equals 13.7570 as compared with 13.7613 by SMACOF. When $d = 3$, Geometric MDS gave the same results in 922 cases. Average values of $S(\cdot)$ are almost the same: 2.9789 (Geometric MDS) and 2.9787 (SMACOF). These preliminary results are very promising, because the evaluated efficiency of the Geometric MDS and the SMACOF is the same, however Geometric MDS is much easier realizable and interpreted.

5 Conclusions

A novel geometric interpretation of the stress function and multidimensional scaling in general (Geometric MDS) has been proposed. Following this interpretation, the step size and direction forward the minimum of the stress function are found analytically for a separate point in a projected space without reference to the analytical expression of the stress function, numerical evaluation of its derivatives and the linear search. It is proved theoretically that the direction coincides with the steepest descent direction, and the analytically found step size guarantees the decrease of stress in this direction.

The discovered option to minimize the stress function was examined on the simple realization of the Geometric MDS. According to the experiments, the realization of Geometric MDS gives very similar results as SMACOF [11]. The results are a bit better often.

In fact, the proposed algorithm is some version of the coordinate-wise descent using d-coordinate blocks. For the objective functions with curved valleys, the convergence of those algorithms normally is slow. However, the geometric approach guarantees the decrease of stress in every step, where the direction and size of the step is determined analytically. In the realisation of Geometric MDS, one step of descent is done only for a separate block taking into account that the most decrease in stress is in the first steps, usually. Despite the fact that the Geometric MDS uses the simplest stress function, there is no need for its normalization depending on the number m of data points and the scale of proximities d_{ij}. These are the reasons that a good performance of the proposed algorithm can be expected as compared with other (e.g. majorization) algorithms. Moreover, more sophisticated realizations of ideas presented in this paper should be developed.

Acknowledgements. This research is funded by Vilnius University, grant No. MSF-LMT-4. The authors are grateful to the reviewers for their comments that made the results of this paper more valuable.

References

1. Dzemyda, G., Kurasova, O., Žilinskas, J.: Multidimensional Data Visualization: Methods and Applications. Springer Optimization and its Applications, vol. 75. Springer, New York (2013). https://doi.org/10.1007/978-1-4419-0236-8
2. Borg, I., Groenen, P.J.F., Mair, P.: Applied Multidimensional Scaling and Unfolding, 2nd edn. Springer, Heidelberg (2018)
3. Li, F., Jiang, M.: Low-resolution face recognition and feature selection based on multidimensional scaling joint L-2, L-1-norm regularisation. IET Biom. **8**(3), 198–205 (2019)
4. Dzemyda, G., Kurasova, O., Medvedev, V., Dzemydaitė, G.: Visualization of data: methods, software, and applications. In: Singh, V.K., Gao, D., Fischer, A. (eds.) Advances in Mathematical Methods and High Performance Computing. AMM, vol. 41, pp. 295–307. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-02487-1_18
5. Perales, E., Burgos, F.J., Vilaseca, M., Viqueira, V., Martinez-Verdu, F.M.: Graininess characterization by multidimensional scaling. J. Mod. Opt. **66**(9), 929–938 (2019)
6. Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika **29**(1), 1–27 (1964)
7. Žilinskas, A.: A quadratically converging algorithm of multidimensional scaling. Informatica **7**(2), 268–274 (1996)
8. Orts Gomez, F.J., Ortega Lopez, G., Filatovas, E., Kurasova, O., Garzon, G.E.M.: Hyperspectral image classification using Isomap with SMACOF. Informatica **30**(2), 349–365 (2019)
9. Groenen, P., Mathar, R., Trejos, J.: Global optimization methods for multidimensional scaling applied to mobile communication. In: Gaul, W., Opitz, O., Schander, M. (eds.) Data Analysis: Scientific Modeling and Practical Applications, pp. 459–475. Springer, Heidelberg (2000). https://doi.org/10.1007/978-3-642-58250-9_37
10. Orts, F., et al.: Improving the energy efficiency of SMACOF for multidimensional scaling on modern architectures. J. Supercomput. **75**(3), 1038–1050 (2018)
11. De Leeuw, J., Mair, P.: Multidimensional scaling using majorization: SMACOF in R. J. Stat. Softw. **31**(3), 1–30 (2009)
12. Symmetric Smacof. https://www.rdocumentation.org/packages/smacof/versions/2.0-0/topics/smacofSym
13. Borg, I., Groenen, P.J.F.: Modern Multidimensional Scaling, 2nd edn. Springer, New York (2005). https://doi.org/10.1007/0-387-28981-X
14. Torgerson Scaling. https://www.rdocumentation.org/packages/smacof/versions/2.0-0/topics/torgerson