

Constrained Optimization of the Stress Function for Multidimensional Scaling

Vydunas Saltenis

Institute of Mathematics and Informatics
Akademijos 4, LT-08663 Vilnius, Lithuania
Saltenis@ktl.mii.lt

Abstract. Multidimensional Scaling (MDS) requires the multimodal Stress function optimization to estimate the model parameters, i.e. the coordinates of points in a lower-dimensional space. Therefore, finding the global optimum of the Stress function is very important for applications of MDS. The main idea of this paper is replacing the difficult multimodal problem by a simpler unimodal constrained optimization problem. A coplanarity measure of points is used as a constraint while the Stress function is minimized in the original high-dimensional space. Two coplanarity measures are proposed. A simple example presented illustrates and visualizes the optimization procedure. Experimental evaluation results with various data point sets demonstrate the potential ability to simplify MDS algorithms avoiding multimodality.

1 Introduction

Multidimensional scaling (MDS) [1, 2] is a widely used technique to visualize the dissimilarity of data points. Objects (n data points) are represented as p -dimensional vectors $Y_1, \dots, Y_n \in R^p$ so that the Euclidean distances $d_{ij}(Y)$, ($i, j = 1, \dots, n; i < j$) between the pairs of points correspond to the given dissimilarities δ_{ij} as closely as possible. Only representations onto a 2-dimensional space are used ($p=2$) as usual, since data visualization is the aim. In general, the dissimilarities δ_{ij} need not be distances between the multidimensional points.

MDS requires the multimodal Stress function optimization to estimate the model parameters, i.e. the coordinates of points (vectors Y_i) in a lower-dimensional space.

The measure of fit is usually defined by the Stress function:

$$\sigma(Y) = \sum_{\substack{i, j=1 \\ i < j}}^n w_{ij} (\delta_{ij} - d_{ij}(Y))^2 ,$$

proposed in [3]. w_{ij} are weights that may be different in various types of the Stress function. In our investigation $w_{ij} = 1$.

The aim of MDS is:

$$\min_{Y \in R^{n \times d}} \sum_{\substack{i,j=1 \\ i < j}}^n (\delta_{ij} - d_{ij}(Y))^2. \tag{1}$$

A substantial shortcoming of MDS is the existence of local minima. The examples of proved multimodality of the Stress function are constructed (for example, [4, 5]). The number of different local minima may range from a few to several thousands. MDS algorithms that minimize Stress cannot guarantee a global minimum. In general, some advice is to use multiple random starts and select the solution with the lowest Stress value (the multiple random start method). A lot of attempts have been made to improve search procedures by a proper choice of start points, however all the strategies are computationally intensive.

2 Basic Idea

Let in our case the dissimilarities δ_{ij} in (1) be the Euclidean distances $d_{ij}(X)$ between the m -dimensional points ($m > p$) with the given coordinates $X_1, \dots, X_n \in R^m$ and variable vectors $Z_1, \dots, Z_n \in R^m$ as distinct from vectors $Y_1, \dots, Y_n \in R^p$ in (1) be of the same dimensionality m . Then $n \times m$ dimensional constrained minimization problem may be formulated as:

$$\min_{Z \in R^{n \times m}} \sum_{i < j} (d_{ij}(Z) - d_{ij}(X))^2 \tag{2}$$

subject to the constraint

$$P(Z) = 0. \tag{3}$$

$P(Z)$ in (3) is some nonnegative coplanarity measure of points Z . If the points in an m -dimensional (high-dimensional) space lie on a hyperplane, then the coplanarity measure must be necessarily equal to zero.

If variable coordinates Z_i are equal to given coordinates X_i in (2), then object function value is equal to zero and coplanarity measure $P(Z) > 0$. These Z_i values are a start position to constrained optimization (2) and (3) when the influence of constraint (3) is gradually increased.

The optimal coordinates Z^{opt} of the problem (2), (3) are m -dimensional and the distances between them $d_{ij}(Z^{opt})$ are the same as that between the p -dimensional optimal coordinates Y^{opt} obtained from (1):

$$d_{ij}(Z^{opt}) = d_{ij}(Y^{opt}).$$

3 Coplanarity Measures

3.1 Coplanarity Measure Based on the Volumes of Tetrahedra

One of the possible coplanarity measures is based on the volumes V_{ijkl} of tetrahedra whose four vertices are multidimensional points Z_1, \dots, Z_n . We use the sum of squared volumes of all possible tetrahedra as coplanarity measure:

$$P(Z) = \sum_{i=1}^{n-3} \sum_{j=i+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^n V_{ijkl}^2,$$

where the volume V is given by the Cayley-Menger determinant [6]:

$$V_{ijkl}^2 = \frac{1}{28} \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{ij}^2 & d_{ik}^2 & d_{il}^2 \\ 1 & d_{ji}^2 & 0 & d_{jk}^2 & d_{jl}^2 \\ 1 & d_{ki}^2 & d_{kj}^2 & 0 & d_{kl}^2 \\ 1 & d_{li}^2 & d_{lj}^2 & d_{lk}^2 & 0 \end{vmatrix}.$$

For the simplicity, the notation d_{ij} is used instead of $d_{ij}(Z)$.

This coplanarity measure was used in our experimental evaluation.

3.2 Coplanarity Measure Based on Point-Plane Distances

Another possible coplanarity measure is consequent upon the coplanarity definition [7, 8]. The points Z_1, \dots, Z_n can be tested for coplanarity by finding the point-plane distances of the points $Z_i, i = 4, \dots, n$ from the plane determined by Z_1, Z_2, Z_3 and checking if all of them are zero. If so, all the points are coplanar.

The point-plane distance from the plane determined by three points Z_1, Z_2, Z_3 may be computed [9] as follows:

$$D_i = \hat{n} \cdot (Z_k - Z_i),$$

where Z_k is any of the three points Z_1, Z_2, Z_3 and \hat{n} is the unit normal

$$\hat{n} = \frac{(Z_2 - Z_1) \times (Z_3 - Z_1)}{|(Z_2 - Z_1) \times (Z_3 - Z_1)|}.$$

Then one of possible coplanarity measures may be introduced:

$$P(Z) = \sum_{i=4}^n D_i^2.$$

The measure depends on the selection of the three points Z_1, Z_2, Z_3 .

4 Constrained Optimization

The optimization problem (2), (3) was solved by the penalty function method [10]. A constrained optimization problem is transformed into a sequence of unconstrained optimization problems by modifying the objective function. In our case, we use such a sequence of unconstrained problems:

$$\min_{Z \in R^{n \times m}} \left(\sum_{i < j} (d_{ij}(Z) - d_{ij}(X))^2 + r_k P^2(Z) \right), \quad (k = 1, 2, \dots),$$

where k is the number of sequence, r_k is a positive penalty parameter. The problem is solved with a sequence of parameters r_k tending to ∞ :

$$r_{k+1} = \Delta r \cdot r_k . \tag{4}$$

The modified problem can be solved by the methods of unconstrained local optimization: Quasi-Newton and Conjugate gradient methods, in our case.

The Torgerson scaling technique [1, 11] may be used for recovery of coordinates of dimensionality 2 from the optimal distances $d_{ij}(Z^{opt})$. The method yields an analytical solution, requiring no iterations.

5 Simple Illustrative Example

In order to visualize and better understand the new approach and the optimization procedure, a simple unidimensional MDS illustrative example was constructed. It uses only three data points, two optimization coordinates, and ($p=1$).

Let the initial distances between three points be: $\delta_{12} = \delta_{13} = 5$; $\delta_{23} = 6$. Only two of the distances d_{13} and d_{23} will be optimized. The distance d_{12} will be fixed: $d_{12} = \delta_{12} = 5$. Then the Stress function in our case is:

$$\sigma(d_{13}, d_{23}) = (d_{13} - 5)^2 + (d_{23} - 6)^2 .$$

In our example the coplanarity measure based on the volumes of tetrahedra reduces to linearity measure L , which is based on the area of a triangle with side lengths d_{12} , d_{13} , d_{23} :

$$L(d_{13}, d_{23}) = (5 + d_{13} + d_{23})(5 + d_{13} - d_{23})(5 - d_{13} + d_{23})(-5 + d_{13} + d_{23}) .$$

L is proportional to the square of the triangle area, calculated by Heron’s formula. There are three local optima of the constrained optimization problem:

1. $d_{13} = 3$; $d_{23} = 8$, with a minimal value of the Stress function $\sigma(3,8) = 8$. The constrained local optimum is marked by point A in Fig. 1.

2. $d_{13} = 2$; $d_{23} = 3$, with a minimal value of the Stress function $\sigma(2,3) = 18$. The constrained local optimum is marked by point *B*.
3. $d_{13} = 8$; $d_{23} = 3$, with a minimal value of the Stress function $\sigma(8,3) = 18$. The constrained local optimum is marked by point *C*.

The global constrained optimum is the first one.

At the beginning of optimization (point *O* in Fig.1) the Stress function is equal to zero, the constraint value is equal to 2304. At each step of constrained optimization, when increasing the penalty parameter r_k , the value of constraint decreases and in the last step (point *A* in Fig. 1) it achieves zero value. At the same time the Stress value increases and, in the last step, achieves the global optimum value 8.

A contour plot diagram demonstrates that, with slightly different data, the global optimum point may be different and, consequently, the result of constrained optimization also changes.

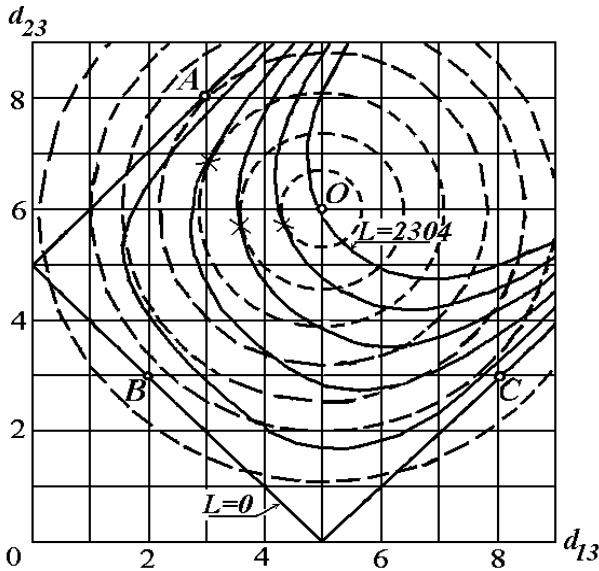


Fig. 1. The contour plots of the Stress function $\sigma(d_{13}, d_{23})$ (dotted contour lines) and of constraint function $L(d_{13}, d_{23})$ (solid contour lines). The points of constrained local minima are marked as *A*, *B*, *C*. The start point is denoted as *O*, and transitional points of constrained optimization are marked as *X*.

6 Experimental Evaluation

Two types of data sets were used in the experimental investigation: regular and irregular. The points of regular data sets were the vertices of a multidimensional cube of various dimensionality and the irregular ones were obtained randomly.

All the results of computational experiments with the proposed algorithm were compared with the results of global optimization obtained using usual MDS random multistart optimization of the Stress function. The number of multistart local optimizations was equal to 200-500.

Coplanarity measure based on the volumes of tetrahedra was used.

Table 1 presents the results of investigation with the regular data points of various dimensionality. The average local optimization error, the number of local minima and the probability to find a global minimum by multistart random local optimization were evaluated from the results of multistart local optimization.

The stopping rule of the constrained optimization was: $P(Z) < 10^{-7}$. The precision of local optimization was 10^{-8} .

Two local optimization methods were compared. We can see that the interval of successful values of Δr is greater for Quasi-Newton method. This method was used in the investigations of Table 2-3.

Table 1. Results of investigation with the regular data points on the vertices of a multidimensional cube of various dimensionality

| | | |
|---|--------------|------------|
| Dimensionality | 3 | 4 |
| Number of points | 8 | 16 |
| Optimal Stress value | 2,854261 | 23,089651 |
| Average local optimization error in % | 2,42 | 0,69 |
| Number of local minima | 6 | 7 |
| Probability of finding a global minimum by multistart random local optimization | 0.68 | 0.74 |
| Values of Δr from (4) (Quasi-Newton method) | 1.1–10000000 | 1.1–700000 |
| Values of η_1 from (4) (Quasi-Newton method) | 1 | 1 |
| Values of Δr from (4) (Conjugate gradient method) | 1.1 – 100 | 1.1 – 80 |
| Values of η_1 from (4) (Conjugate gradient method) | 1 | 1 |

Table 2. Results of investigation with the random data points of dimensionality $m=4$ (number of points $n=16$)

| | | |
|---|---------|---------|
| Optimal Stress value | 2,16464 | 1.77626 |
| Average local optimization error in % | 25.47 | 26.02 |
| Number of local minima | 10 | 8 |
| Probability of finding a global minimum by multistart random local optimization | 0.25 | 0.39 |
| Δr value | 100 | 100 |
| η_1 value | 0.01 | 0.01 |

Tables 2-3 present two examples of numerous investigations with the random data points. Table 2 presents the results with the random data points of dimensionality $m=4$ (number of points $n=16$). Table 3 presents the results with the random data points of dimensionality $m=6$ (number of points $n=20$).

In all the experiments (not only presented in Tables 1-3) the proposed constrained optimization achieved the global minimum.

Table 3. Results of investigation with the random data points of dimensionality $m=6$ (number of points $n=20$)

| | | |
|---|---------|----------|
| Optimal Stress value | 8.16616 | 10.32240 |
| Average local optimization error in % | 22.47 | 9.29 |
| Number of local minima | 38 | 37 |
| Probability of finding a global minimum by multistart random local optimization | 0.37 | 0.38 |
| Δr value | 100 | 100 |
| r_1 value | 0.01 | 0.01 |

The proposed optimization procedure is more computation exhaustive. The number of variables is larger in comparison with the optimization by the usual approach. For example, the execution time is 5 times greater in comparison to single start of the usual approach for the data set of Table 2.

6 Conclusions

The new approach replaces the difficult multimodal optimization problem by a simpler optimization problem that uses the constrained local optimization procedure. It minimizes the Stress function in the original high-dimensional space subjected to zero planarity constraint.

This approach eliminates the problem of the initial choice of variables and the difficulties caused by the Stress function multimodality.

However, the optimization procedure is more computation exhaustive. The number of variables is larger in comparison with the optimization by the usual approach; the constrained optimization requires some steps.

We did not test any evaluations of the computational efficiency of the new approach for various data, neither did we consider possible performance improvement observations in the paper. These issues remain as a possible trend of further research.

Acknowledgements

The research was partially supported by the Lithuanian State Science and Studies Foundation, Grant No. C 03013.

References

1. Borg, L., Groenen, P.: *Modern Multidimensional Scaling: Theory and Applications*, Springer (1997)
2. Cox, T., Cox, M.: *Multidimensional Scaling*, Chapman and Hall (2001)
3. Kruskal, J.: Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika*, Vol.29 (1964) 115-129
4. Trosset, M., Mathar R.: On existence of nonglobal minimizers of the STRESS Criterion for Metric Multidimensional Scaling. *Proceedings of the Statistical Computing Section, American Statistical Association, Alexandria, VA, (1997) 158-162*
5. Zilinskas, A., Podlipskyte, A.: On multimodality of the SSTRESS criterion for metric multidimensional scaling, *Informatica*, Vol. 14, No. 1, (2003) 121-130
6. Sommerville, D. M. Y.: *An Introduction to the Geometry of n Dimensions*. New York: Dover, (1958)
7. Abbott, P. (Ed.). In and Out: Coplanarity. *Mathematica J.* 9 (2004) 300-302
8. Weisstein, E. W.: Coplanar. *MathWorld - A Wolfram Web Resource*. <http://mathworld.wolfram.com/Coplanar.html>
9. Weisstein, E. W.: Point-Plane Distance. *MathWorld - A Wolfram Web Resource*. <http://mathworld.wolfram.com/Point-PlaneDistance.html>
10. Bertsekas, D. P.: *Nonlinear programming*. Athena Scientific (1999)
11. Torgerson, W. S.: *Theory and methods of scaling*. New York: Wiley (1958)