



# Recent advances in data dimensionality reduction using multidimensional scaling

Gintautas Dzemyda

Vilnius University Institute of Data Science and Digital Technologies, Akademijos str. 4, LT-08412 Vilnius, Lithuania

ICCCC2020, May 11-15, 2020, Romania, Oradia

# Introduction (1)

Human participation plays an essential role in most decisions when analyzing data. The huge storage capacity and computational power of computers cannot replace the human flexibility, perceptual abilities, creativity, and general knowledge.

A proper interaction between human and computer is essential. Moreover, such an interaction is one of the areas in computer science that has evolved a lot in recent years.

Real data in technologies and sciences are often high-dimensional. So it is very difficult to understand these data and extract patterns.



# Introduction (2)

One way of such an understanding is to make a visual insight into the data set. Here, a hopeful view may be put on the visualization of multidimensional data.

The goal of visualization methods is to represent the multidimensional data in a low-dimensional space so that certain properties (e.g. clusters, outliers) of the structure of the data set were preserved as faithfully as possible.

The dimensionality reduction or visualization methods are recent techniques to discover knowledge hidden in multidimensional data sets.



# Visualization

The human being can comprehend visual information more quickly than textual one.



### Visualization problem

The goal of the projection (visualization) methods is to represent the input data items in a lower-dimensional space so that certain properties of the structure of the data set were preserved as faithfully as possible.

	A	В	C	D	E	F	G	H		J									
	1,11091	7,04704	2,68999	3,51433	6,24851	-7,70285	3,98986	-15,0384	5,9192	5,87187	5	•		-					-
	-5,15074	1,14707	1,49299	2,55285	-2,27442	-14,6506	7,86902	-7,36946	9,63875	7,11384									
3	1,91532	-4,58991	-1,90508	4,07452	-11,9971	-7,92136	9,69161	0,222057	10,8834	5,87123									
4	0,020048	0,630172	1,08796	2,26984	-4,79242	-8,15183	4,29621	-0,46226	7,42289	6,15558									
5	-0,89494	8,11653	-6,95531	6,61701	-8,21592	-11,2322	7,56654	-8,13229	10,6122	5,83966									
6	-0,10323	-10,4471	-3,17563	5,31002	0,14175	-9,40686	9,52215	1,64615	-1,88292	6,64164		o		_					-
7	1,72396	0,97497	5,85485	5,75813	1,33302	-4,35716	3,95199	-1,21654	4,97955	4,7983									
в	4,52856	11,1303	-0,52264	3,58449	-5,35348	-9,28854	9,6102	-3,04153	5,72042	5,28749									
ĺ	-0,83436	1,05809	7,03569	2,13677	-4,65407	-8,54866	6,28365	-2,71864	17,3228	5,42663		1							
0	2,58882	8,18563	4,12082	5,86436	-3,93221	-0,99236	6,98705	-2,16161	3,14953	5,16401		1						/	
1	-1,00736	0,836051	-0,27273	5,757	-8,00482	-2,68832	4,08388	2,19967	4,11293	6,58779	3	•		_		-	-	6	
12	-2,78432	-0,36631	-5,02407	2,12856	-5,31326	-8,53505	4,3084	-4,39105	5,59385	6,73681					. I.		/		
3	6,34916	4,67636	3,14195	2,93509	-4,95993	-11,4985	4,62661	-7,7016	7,09844	7,54618									
4	2,35358	9,6925	0,786942	5,20636	-1,26773	-11,0521	5,5869	-0,09746	10,3834	6,49727									
5	-3,13531	-3,0484	3,19325	3,82446	1,29874	-6,93801	4,42466	-4,00488	15,2039	4,52903						6			
16	0,192948	9,22675	1,49804	3,56301	-3,13276	-7,37238	11,2558	-1,01955	6,3389	5,51749	2		-	E CE		-	-		-
17	-6,14694	11,5993	5,03532	3,68279	2,66215	-12,0582	8,4349	-3,82024	-4,39669	8,02664				77					
8	1,66801	-1,53136	2,95147	0,441002	-6,85443	-6,11669	7,83389	-8,50664	12,1942	6,17958									
9	-2,42808	-2,35968	-0,9866	2,95973	-0,71374	-6,25115	7,01922	-4,05128	0,133749	7,84149									
0	-3,23398	0,739948	1,61084	3,85228	-9,73604	-3,64236	11,4746	-11,9365	9,5091	5,65706				<b>.</b>					
1	-0,02485	3,72569	-5,69587	1,96643	-8,95218	0,753423	8,5288	-6,39854	1,8114	7,20848									
2	-1,7316	0,14579	-8,76263	0,305276	-0,84276	-13,4019	7,17862	-2,53552	1,32054	5,16366		] 🚅				-	-		
23	-0,26349	2,65934	-0,38402	4,20786	-9,42951	-3,06817	6,16654	-7,08867	-3,23498	6,06282								$\mathbf{i}$	$\sim$ 1
24	0,935001	7,62704	-5,89728	0,043108	-8,97537	-17,8392	5,97077	-13,8924	6,5347	5,41031					1		÷ .	)	_\
25	-1,13148	-3,6351	3,4953	4,03109	11,0707	2,57909	6,92097	-4,71983	5,97176	5,41212		1							
26	-0,42632	10,1015	-4,25269	2,9947	-8,99443	-5,9449	3,81592	-0,83499	17,2262	3,8634					1	_		1	
7	1,67647	4,57368	2,22492	2,66839	-7,65192	-8,25385	6,32712	-2,59647	7,04368	6,75276		9		_	$\sim$		1	/	/
8	0,42202	5,57122	-3,12425	4,19927	-16,3641	-0,344	9,59166	-7,06135	11,1317	1,69255			6	78	100	_	-	_	19

# Examples of visualization (1)



Recent advances in data dimensionality reduction using multidimensional scaling

# Examples of visualization (2)



Gintautas Dzemyda

Recent advances in data dimensionality reduction using multidimensional scaling

### Example of multidimensional data (breast cancer data)

	$x_1$	<i>x</i> <sub>2</sub>	<i>x</i> <sub>3</sub>	<i>x</i> <sub>4</sub>				<i>x</i> <sub>8</sub>	<i>x</i> 9	
<i>X</i> <sub>1</sub>	5	1	1	1	2	1	3	1	1	b
<i>X</i> <sub>2</sub>	5	4	4	5	7	10	3	2	1	b
<i>X</i> <sub>3</sub>	3	1	1	1	2	2	3	1	1	b
$X_4$	6	8	8	1	3	4	3	7	1	b
$X_5$	4	1	1	3	2	1	3	1	1	b
$X_6$	1	1	1	1	2	10	3	1	1	b
X7	2	1	2	1	2	1	3	1	1	b
X <sub>8</sub>	2	1	1	1	2	1	1	1	5	b
$X_9$	4	2	1	1	2	1	2	1	1	b
X <sub>460</sub>	8	10	10	8	7	10	9	7	1	m
X <sub>461</sub>	5	3	3	3	2	3	4	4	1	m
X462	8	7	5	10	7	9	5	5	4	m
X463	7	4	6	4	6	1	4	3	1	m
X464	10	7	7	6	4	10	4	1	2	m
X465	7	3	2	10	5	10	5	4	4	m
X466	10	5	5	3	6	7	7	10	1	m
X467	5	2	3	4	2	7	3	6	1	m

University of Wisconsin, Clinical Sciences Center

- $x_1$  clump thickness,
- $x_2$  uniformity of cell size,
- $x_3$  uniformity of cell shape,
- $x_4$  marginal adhesion,
- $x_5$  single epithelial cell size,
- $x_6$  bare nuclei,
- $x_7$  bland chromatin,
- $x_8$  normal nucleoli,
- $x_9 \text{mitoses},$
- C class (benign, malignant) (Nepiktybinis. Piktybinis)

# Visualization for Early Diagnosis

• Breast cancer data analysis



### Introduction

Visualization of multidimensional data is a complicated problem followed by extensive researches because it allows to the investigator

- to observe data clusters
- to estimate the inter-nearness between the multidimensional points
- to make proper decisions

Let us have *m* multidimensional (*n*-dimensional) vectors

$$X_1, X_2..., X_m \in \mathbb{R}^n$$
  $X_i = (x_{i1}, x_{i2}..., x_{in}), i = 1, ..., m$ 

The problem is to get a projection of this set of vectors on the visually perceived low dimensional space  $\mathbb{R}^2$  or  $\mathbb{R}^3$ . Denote projections on the plane by  $Y_i = (y_{i1}, y_{i2}), i = 1, ..., m$ .

# Dimensionality reduction methods

There exist a lot of methods that can be used for reducing the dimensionality of data, and, particularly, for visualizing the *n*-dimensional vectors.

- Traditional methods
  - Multidimensional scaling (SMACOF, Relative MDS, Diagonal Majorization Algorithm (DMA)...)
  - Sammon's projection
  - Principal components
  - Direct methods (Chernoff faces, Andrews curves, star)
  - Others
- Neural networks
  - Self-organizing map (SOM)
  - Feed-forward networks
- Combinations of traditional methods and neural networks
- Manifold learning methods (locally linear embedding (LLE), Laplacian Eigenmaps (LE), lsomap...)

11 / 35

# Abstract (1)

- Multidimensional scaling (MDS) is one of the most popular methods for a visual representation of multidimensional data.
- Recently, it finds wide applications of various nature: face recognition, analysis of regional economic development, image graininess characterization.
- Classical approaches to minimize the stress in multidimensional scaling (MDS) reached their limits. New viewpoint to the problem is necessary, including its formulation and ways of solving.
- A novel geometric interpretation of the stress function and multidimensional scaling in general (Geometric MDS) has been proposed.
- A strategy of application of the discovered option to minimize the stress function is presented and examined.
- The novel geometric approach will allow developing a new class of algorithms to minimize MDS stress, including global optimization and high-performance computing.

Gintautas Dzemyda

Recent advances in data dimensionality reduction using multidimensional scaling



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,
- the direction coincides with the steepest descent direction, and the analytically found step size guaranties almost the optimal step in this direction.

# Introduction (1)

- Suppose, we have a set X = {X<sub>i</sub> = (x<sub>i1</sub>,...,x<sub>in</sub>), i = 1,...,m} of n-dimensional data points (observations) X<sub>i</sub> ∈ ℝ<sup>n</sup>, n ≥ 3.
- Dimensionality reduction and visualization requires estimating the coordinates of new points Y<sub>i</sub> = (y<sub>i1</sub>,..., y<sub>id</sub>), i = 1,..., m, in a lower-dimensional space (d < n) by holding proximities δ<sub>ij</sub> between multidimensional points X<sub>i</sub> and X<sub>j</sub>, i, j = 1,..., m, as much as possible.
- Proximity  $\delta_{ij}$  can be measured e.g. by the distance between  $X_i$  and  $X_j$ .
- The input data for MDS consists of the symmetric m × m matrix
   D = {δ<sub>ij</sub>, i, j = 1,..., m} of proximities between pairs of points X<sub>i</sub> and X<sub>j</sub>. If the Minkowski distance is used as the proximity, then

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

• When q = 2, the distance becomes the Euclidean distance.

# Introduction (2)

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 raw stress function:

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

where  $d_{ij}^*$  is the Euclidean distance between points  $Y_i$  and  $Y_j$  in a lower-dimensional space. The MDS-based dimensionality reduction optimization problem may be formulated as follows:

$$\min_{Y_1,\ldots,Y_m\in\mathbb{R}^d} S(Y_1,\ldots,Y_m).$$
(2)

### The geometric approach (1)

- Without loosing the generality and for better perception of the approach, let's consider case, where the dimensionality of projected space d = 2.
- Suppose, we have m × m matrix D = {d<sub>ij</sub>, i, j = 1,...,m} of Euclidean distances between *n*-dimensional points X<sub>i</sub> = (x<sub>i1</sub>,...,x<sub>in</sub>), i = 1,...,m. We aim to find two-dimensional points Y<sub>i</sub> = (y<sub>i1</sub>, y<sub>i2</sub>), i = 1,...,m by solving the problem (2).
  At first, let's have some initial configuration of points Y<sub>1</sub>,...,Y<sub>m</sub>.
- Then, let's optimize the position of the particular points  $Y_1, \ldots, Y_m$ . Then, let's optimize the position of the particular point  $Y_j$  when the position of remaining 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 (2) by minimizing the so-called local stress function depending on  $Y_j$ , only:

$$S^{*}(Y_{j}) = \sum_{i=1, i \neq j}^{m} \left( d_{ij} - \sqrt{\sum_{k=1}^{d} (y_{ik} - y_{jk})^{2}} \right)^{2}.$$
 (3)

Gintautas Dzemyda

# The geometric approach (2)



Figure 1: An example of a single iteration of geometric method.

# The geometric approach (3)

- In the centre of each circle, we have a corresponding point Y<sub>i</sub>. Radius of the *i*-th circle is equal to the distance d<sub>ij</sub> between the points X<sub>i</sub> and X<sub>j</sub> in n-dimensional space. Point A<sub>ij</sub> lies on the line between Y<sub>i</sub> and Y<sub>j</sub>, i ≠ j.
- Let Y<sub>j</sub><sup>\*</sup> be chosen as an average point of the points A<sub>ij</sub> over i = 1...m, i ≠ j:

$$Y_j^* = \frac{1}{m-1} \sum_{i=1, i \neq j}^m A_{ij}.$$
 (4)

- When we make a step from Y<sub>j</sub> to Y<sup>\*</sup><sub>j</sub>, we get new intersection points A<sup>\*</sup><sub>ij</sub> on circles that correspond to Y<sub>j</sub>, and these points are on the line between Y<sub>i</sub> and Y<sup>\*</sup><sub>i</sub>.
- We will analyse the value of the local stress function  $S^*(Y_j^*)$  and compare it with the value  $S^*(Y_j)$ .

Gintautas Dzemyda

#### Proposition 1

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

$$\nabla S^*|_{\mathbf{Y}_j} = \left(2\sum_{\substack{i=1\\i\neq j}}^m \frac{d_{ij} - \sqrt{\sum_{k=1}^d (y_{ik} - y_{jk})^2}}{\sqrt{\sum_{k=1}^d (y_{ik} - y_{jk})^2}} (y_{ik} - y_{jk}), \ k = 1, \dots, d\right).$$

#### Proposition 2

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

$$Y_j^* = Y_j - \frac{1}{2(m-1)} \nabla S^*|_{Y_j}.$$
 (5)

#### **Proposition 3**

Size of a step from  $Y_j$  to  $Y_i^*$  is equal to

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

#### Proposition 4

Let  $Y_j$  does not match to any local extreme point of the function  $S^*(\cdot)$ . If  $Y_j^*$  is chosen by (4):  $Y_j^* = \frac{1}{m-1} \sum_{i=1, i \neq j}^m A_{ij}$ , then a single step from  $Y_j$  to  $Y_i^*$  reduces a local stress  $S^*(\cdot)$ :

$$S^{*}(Y_{j}^{*}) < S^{*}(Y_{j}).$$

#### Proposition 5

The value of the local stress function (3)

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

will converge to a local minimum when repeating steps

$$Y_j^* = Y_j - \frac{1}{2(m-1)} \nabla S^*|_{Y_j}$$
 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  $S(\cdot)$  of MDS, i. e. if  $Y_j^*$  is chosen by (4):  $Y_j^* = \frac{1}{m-1} \sum_{i=1, i \neq j}^m A_{ij}$ , then the stress function

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

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

Proposition 7

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

### Proof of the proposition 7



Figure 2: Example of the anti-gradient search

23 / 35

# Multiextremal local stress



Figure 3: Example of the multiextremal local stress

### Simple realization of Geometric MDS

- Simple realizations of Geometric MDS are based on fixing some initial positions of points  $Y_i = (y_{i1}, \ldots, y_{id})$ ,  $i = 1, \ldots, m$  (at random, using principal component analysis, etc.), and further changing the positions of  $Y_j$  once or several times by (5) 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 for minimization of function S(·), the random selection of a set Y of d-dimensional points was used and further minimization of S(·) seeking for its local minimum was performed by consecutive one-step changing of positions of points Y<sub>1</sub>,..., Y<sub>m</sub> many times.
- This multistart procedure was repeated several times expecting to find a better local minimum or even the global minimum.

Gintautas Dzemyda

# Experiments (1)

- 20 random sets Y of 30 points (m = 30) were generated inside the 4-dimensional unit hypercube (n = 4) and represented in d = 2, 3, 4 spaces using random start position for optimization.
- Values in Table 1 show the minima. 11 descents among 20 ones reached the global minimum, when d = 2. When d = 3, all descents reached the best minimum.

Table 1: Stress value of 20 experiments for the same data set	Table	1:	Stress	value o	f 20	experiments	for	the	same	data	set
---	-------	----	--------	---------	------	-------------	-----	-----	------	------	-----

4 - 2	19.146	19.146	19.146	19.146	19.146	19.146	19.146	19.146	19.146	19.146
u _ 2	19.146	19.154	19.214	19.214	20.584	20.584	20.584	20.792	21.453	22.479
4-2	2.779	2.779	2.779	2.779	2.779	2.779	2.779	2.779	2.779	2.779
u _ 5	2.779	2.779	2.779	2.779	2.779	2.779	2.779	2.779	2.779	2.779
d _ 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
u = 4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

# Experiments (2)

- For comparison, the same 1000 data sets of 30 random 4-dimensional points (m = 30, n = 4) were generated and analysed by Geometric MDS and multidimensional scaling based on stress  $S(\cdot)$  minimization using majorization (SMACOF) that is realized in R.
- Both Geometric MDS and SMACOF used the same initial values of points Y<sub>1</sub>,..., Y<sub>m</sub> obtained by Torgerson Scaling.
- When d = 2, Geometric MDS and SMACOF gave the same results 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).

# GeoGebra example (1)



Figure 4: Minimization of local stress function M.Sabaliauskas, G.Dzemyda, "Visual analysis of multidimensional scaling using GeoGebra"

Gintautas Dzemyda

# GeoGebra example (2)



Figure 5: Minimization of local stress function M.Sabaliauskas, G.Dzemyda, "Visual analysis of multidimensional scaling using GeoGebra"

Gintautas Dzemyda

Recent advances in data dimensionality reduction using multidimensional scaling

# GeoGebra example (3)



Figure 6: Minimization of local stress function M.Sabaliauskas, G.Dzemyda, "Visual analysis of multidimensional scaling using GeoGebra"

Gintautas Dzemyda

30 / 35

# GeoGebra example (4)



Figure 7: Minimization of local stress function M.Sabaliauskas, G.Dzemyda, "Visual analysis of multidimensional scaling using GeoGebra"

Gintautas Dzemyda

# Conclusions (1)

- A novel geometric interpretation of the stress function and multidimensional scaling in general (Geometric MDS) has been proposed.
- According to the experiments, the realization of Geometric MDS gives very similar results as SMACOF. The results are a bit better often.
- 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.
- More sophisticated realizations of ideas presented in this paper should be developed.

# Conclusions (2)

- The reasons that a good performance of the proposed algorithm can be expected as compared with other (e.g. majorization) algorithms:
  - In Geometric MDS, 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 almost the optimal step in this direction.
  - Obspite the fact that the Geometric MDS uses the simplest (raw) stress function, there is no need for its normalization depending on the number *m* of data points, the scale of features x<sub>ik</sub> and proximities δ<sub>ij</sub>.



Lithuanian Computer Society Vilnius University Institute of Mathematics and Informatics Lithuanian Academy of Sciences



https://www.mii.lt/damss/

### DAMSS-2020 12th International Workshop **Data Analysis Methods for Software Systems**

Druskininkai, Lithuania, Hotel "Europa Royale" 25.11-28.11 2020









# Thank you for your attention!

This research has received funding from the Research Council of Lithuania (LMTLT), agreement No S-MIP-20-19.