A Survey on Multidimensional Scaling

A Survey on Multidimensional Scaling

Nasir Saeed, King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah, KSA
Haewoon Nam, Hanyang University, Ansan, South Korea
Mian Imtiaz Ul Haq, University of Engineering & Technology, Peshawar, Pakistan
Dost Muhammad Saqib Bhatti, Dawood University of Engineering & Technology, Karachi, Pakistan

This survey presents multidimensional scaling (MDS) methods and their applications in real world. MDS is an exploratory and multivariate data analysis technique becoming more and more popular. MDS is one of the multivariate data analysis techniques, which tries to represent the higher dimensional data into lower space. The input data for MDS analysis is measured by the dissimilarity or similarity of the objects under observation. Once the MDS technique is applied to the measured dissimilarity or similarity, MDS results in a spatial map. In the spatial map, the dissimilar objects are far apart while objects which are similar are placed close to each other. In this survey paper, MDS is described fairly in comprehensive fashion by explaining the basic notions of classical MDS and how MDS can be helpful to analyze the multidimensional data. Later on various MDS based special models are described in a more mathematical way.

Additional Key Words and Phrases: Multidimensional Scaling, Multivariate, Similarity, Dissimilarity, Spatial Map

ACM Reference Format:
Nasir Saeed, Haewoon Nam, Mian Imtiaz Ul Haq, and Dost Muhammad Saqib Bhatti, 2016. A Survey on MDS. *ACM Comput. Surv.* 1, 1, Article 1 (January 2018), 0 pages.
DOI: 10.1145/3178155

1. INTRODUCTION

Multidimensional scaling (MDS) is one of the dimensionality reduction techniques which converts multidimensional data in to lower dimension space, while keeping the intrinsic information. The main reason for using MDS is to get a graphical display for the given data, such that it is much easier to understand. There are number of other dimensionality reduction techniques like, principal component analysis (PCA), factor analysis and isomap. But MDS is much popular among all these techniques because of its simplicity and many application areas. MDS analysis finds the spatial map for objects given that the similarity or dissimilarity information between the objects is available [Borg and Gronen 2005; Bronstein et al. 2006b; Kruskal 1956]. MDS assigns observations to specific points in a multidimensional space (2 or 3 dimensional space) in a way that the given dissimilarities or similarity information is matched closely to the distances between points [Honarkhah and Caers 2010; Cambria et al. 2014; Bai et al. 2016; Gansner et al. 2004]. MDS and factor analysis are similar techniques but MDS is preferred over factor analysis because, MDS does not depend on most common assumptions like linearity and normality [Biswas et al. 2006]. In fact, the only assumption of MDS is that, the number of dimensions should be one less than the number of points, which also means at least three variables must be entered in the model and at least two dimensions must be specified [Bronstein et al. 2006a; Bronstein et al. 2007; Young et al. 2007]. Once the proximities (similarity or dissimilarity information) are available, MDS can provide the lower dimension solution.

A rich literature is available on MDS methods [Borg et al. 2012; Hout et al. 2013; Izenman 2008; France and Carroll 2011; Kamada and Kawai 1989], resulting in several milestones for the multivariate data analysis. Some recent and classical results on MDS are discussed in [Cox and Cox 2001] and [Borg and Gronen 2005] and the references therein. The authors in [Eckart and Young 1936] and [Young and Householder 1938] are the founders of MDS. The first input metric for MDS was developed by Torgerson, considered as classical MDS (CMDS) [Torgerson 1952a]. CMDS was further

Table I. Brief history of MDS

MDS techniques	Authors	Year
Foundation for MDS	Eckart and Young	1936-1938
Classical MDS (CMDS)	Torgerson	1952
Principal coordinate analysis	Gower	1966
Non-metric MDS (NMDS)	Shepard and Kruskal	1962-1964

Table II. Development of loss function for MDS

Loss functions	Authors	Year
Sammons non-linear mapping	Sammon	1969
Unfolding model	Coombs	1964
Individual differences models	Carroll and Chang	1970
ALSCAL	Takane	1977
Maximum likelihood based MDS	Ramsay	1982
MDS with optimal scaling	Meulman	1992-1993

extended by Gower and developed the relationship between PCA and CMDS [Gower 1966]. Major difference between PCA and CMDS is that the PCA uses a correlation (or variance-covariance) matrix as an input while MDS uses a distance matrix. PCA tries to minimizes a squared loss function, which is inappropriate for non real value data. In [Collins et al. 2001] the authors proposed generalized PCA which can work with exponential, generalized linear models, and Bregman distances. PCA is used to model the linear variability in higher dimensional data. However, in many applications the higher dimensional data is nonlinear in nature. PCA gives output of optimum linear approximation for a given high-dimension data but due to the global linearity its performance is limited. Shepard and Kruskal [Kruskal 1964a] developed the non-metric MDS (NMDS) in which the similarity or dissimilarity information is monotonically related to the distances [Solaro 2011]. Brief history about the development of MDS is given in table I. NMDS opened new areas for MDS based models by specifying and optimizing the loss function [Sammon 1969]. Table I summarizes the development of loss functions for MDS. It includes the Coombs unfolding model [Coombs 1964] dealing with the data that is derived from two different set of objects and is also called individual difference model [Carroll and Chan 1970]. Coombs unfolding model is based on a geometric model for a preference and a choice. It tries to locate object and its specifics as points in a combine space, where an object will pick the specifics that is closest to its point. ALSCAL is an efficient algorithm developed by implementing the MDS model [Takane et al. 1977] and is included in statistical package for social sciences (SPSS). Maximum likelihood based MDS is proposed in [Ramsay 1982], while an optimal scaling approach for multivariate analysis is introduced in [Meulman 1992] by reformulating the CMDS.

Traditional MDS methods, consider that distances among objects are symmetric, although, this consideration is not always satisfied. For instance, [Tversky 1977] and [Tversky and Gati 1978] explained the characters of similarity among objects studied with psychological scale and concluded that cognitive similarity is mostly asymmetric. The motivation behind these asymmetric MDS methods is to remove the shortcomings of traditional MDS methods, i.e., in case where similarity or dissimilarity matrices are asymmetric in nature as they are based on supposition that similarity or dissimilarity matrices can be associated to inter-point distances in a given metric space [Torgerson 1952b], [Kruskal 1964b], [Guttman 1968a]. Many researchers have extended the traditional MDS methods by assuming that the similarity or dissimilarity among objects is not a function of only inter-point distances but is also a function of the quantities associated to these objects. For instance, the squared distances is expended by weights in weighted distance model which was first proposed by [Young 1975]. In [Saito 1983],

[Saito 1986] and [Weeks and Bentler 1982] the authors have proposed altered distance models where the distance between points is developed by a few constants associated to these points. In [Okada and Imaizumi 1984] and [Okada and Imaizumi 1987] a non-metric type of generalized altered distance model is proposed. Smallest space analysis-2 (SSA-2) is introduced in [Guttman 1968a] and [Lingoes and Guttman 1973], where column and row compatibility is applied on the data to get two solutions in metric space. Wind model is proposed in [Tobler 1975], where the asymmetries are explained by the direction of wind given to mesh point on the arrangement of objects. In [Saito 1988] and [Sato 1989] the authors proposed a model, in which the asymmetries is analyzed by utilizing the Randers metric i.e., an asymmetric metric function. [Tversky 1977] proposed the future matching model which explains the similarity or dissimilarity among two objects through a linear combination of the amount of distinctive and common characteristic of the two objects. In [Chino 1978] and [Chino 1980] the authors proposed a model using a generalization of scalar products, which fits the magnitude of cross and inner (scalar) products of solution vectors to skew-symmetric and symmetric parts of the data respectively. In [Constantine and Gower 1978] and [Gower 1977], first the asymmetric proximity matrix is split into two components, i.e., symmetric and skew-symmetric components and then deal with them separately. For symmetric component traditional MDS method is used while for skew-symmetric component canonical decomposition is used. In [Saburi and Chino 2008] the authors proposed a maximum likelihood method for asymmetric proximity matrix, which expand the work for asymmetrical data [Takane 1981].

2. APPLICATIONS OF MDS

Different Application areas of MDS include data mining, pattern recognition, information theory, psychometry, ecology and marketing [Borg and Gronen 2005]. Recently MDS have also being used for locating the wireless nodes in the network to get location aware information [Priyantha et al. 2003]. Furthermore, the MDS applications are defined below:

2.1. Scientific Visualization

In scientific visualization, MDS is used for the compression of multified visualize data. The goal of multified visualization is to portray information contained in two or more fields in the compositional manner, which facilities combination overview or visual comparison [Hansen et al. 2014; Buja et al. 2008]. In visualization, the key to achieve this goal is to exploit different visual channels available in the design space, referred to as channel fusions. The main issue with multified data is that they exceed the available number of channels. This is similar to the obstacle of multivariate data when the aim is to reduce the dimensions of the data for display, which is called low-dimension embedding through dimension reduction. MDS perform linear or non-linear projection of data into a lower dimensional space. The quality of embedding is measured by a stress metric which root mean square error of point-wise distance in data space and the respective distance in embedded space. MDS can be used on multified data embedded space. MDS can be used on multified data to set number of visual channel [Hansen et al. 2014; Lawrence et al. 2011]

2.2. Psychological Structure

One of the major application of MDS is in the field of psychology. Psychology usually brings up different questions in the mind of Psychologists. MDS is used in exploratory manner to search for the relationship between observed similarities or dissimilarities. Thus, different trends in a psychological structures can be find out using MDS methods. Fitzgerald and Hubert [Jaworska and Chupetlovska-Anastasova 2009] suggested

analyzing individual differences by using multiple proximity matrices. These matrices can be obtained from individual subjects or from subgroups. A bunch of measures can be generated if each subject's data is treated as a independent proximity matrix. Moreover, independent proximity measures can be constructed for data collection in different setting or at different times. Hence, utilization of MDS is not only for representation of interrelations between objects and determination of underlying data dimensions, but also for providing a representation of individual or group differences.

2.3. Data exploration

MDS is used to find out the structure of data, to describe the input data as a spatial map in multidimensional space. Two different situations can be detected and compared whether they are similar enough to be acted upon as the same or dissimilar enough to require different actions. It is often of great interest to have a quantitative estimate of the similarity or dissimilarity between the two items. MDS takes estimated similarity or dissimilarity information as an input. The outcome of MDS is a spatial map that gives the information, spatially, i.e., the relationships among different items, where similar items are in close proximity to each other while dissimilar items are located far apart. For example, if MDS is applied to the pairwise similarity to the set of color patches, the outcome of MDS would resemble the Netwon's famed "color wheel" [Newton 1704], with red color being near to orange, but far from green, and so on. The maps provided by MDS are valuable because they reduce complex data sets to a visual relationships in the data which are more easy to understand.

2.4. Testing structural hypothesis

One of the most fundamental issues of testing a hypothesis is that how subjective impressions of similarity come about. Why does Mike look like Julias son? How come that a Ferrari looks similar to a Porsche than to a Cadillac? Specifically, one who wants to find, if and how the facets (dimensions, factors, features, etc) by which the items are conceptually salient. In order to explain these kind of perceptions, MDS can be used with the given distance model. In such distance based models, the objects are first considered as points in a certain space based on the similarity between those objects. Then with the help of MDS data can be classified from given global similarity judgments. This is a confirmation process for the MDS, in order to verify that whether the available data supports the assumptions that can classify the original data.

2.5. Pattern Recognition

In pattern recognition, MDS is used for reducing the number of variables through feature selection (selection of subset of the original variables for classifier design) or feature extraction (determining a linear or non-linear transformation of original variable to obtain a smaller set), which can lead us to improved classifier performance and a better understanding of the data. It is applied to a class of techniques that examine a matrix of dissimilarities (the proximity matrix) or distances in order to produce a representation of the data points in a reduced dimension space, which is called representation space. Both metric and non-metric MDS are used in pattern recognition. Metric MDS considers that the data is quantitative and its procedure assumes a functional relationship between the inter-point distances and the given dissimilarities. While non-metric MDS considers that the data is qualitative and has ordinal importance and its procedure produces configuration that maintains the rank of the dissimilarities. Metric MDS was proposed into pattern recognition by Sammon in 1969 [Webb 2003]. It was inroduced to comprehend class information and to provide non-linear transformations for dimension reduction in feature extraction [Webb 2003].

2.6. Ecology

The MDS is used for the ordination in the ecology, which summarizes the patterns and main trends of multidimensional data spaces in to concise low-dimensional ordination space. The type used for the ordination is non-metric MDS [Falk et al. 2006]. According to [Gauch 1982], "ordination primarily endeavors to represent sample and species relationships as faithfully as possible in a low-dimensional space". The ordination is necessary as it is impossible to visualize multiple dimensions simultaneously and most importantly a single multivariate analysis saves huge time as compare to separate uni-variate analysis of each piece [Gauch 1982]. There are many techniques used for ordination including NMDS, i.e. polar ordination (PO) and Principal coordinates analysis (PCoA). PO organizes samples between endpoints or poles according to distance matrix. PCoA is similar to PO, it shows the distance between samples [Bray and Curtis 1957]. PO and PCoA has number of flaws, MDS rectifies those flaws by increasing the rank order correlation. This is the main reason why NMDS is widely used in ecology.

2.7. Sports Visualization

MDS is also being used in sports to visualize the competing teams performance [Raalte et al. 1990; Carter et al. 2001]. In [Machado and Lopes 2017], authors have studied the behavior of competing teams in soccer league using MDS technique. The dissimilarities between the soccer teams are measured using the results of each round and that information is imposed on MDS algorithm to visualize the performance of teams. The dissimilarities between teams are shown using three different approaches. In first, one dissimilarity matrix and one MDS per map are generated. In second approach, whole data is combined into dissimilarity matrix, which leads to a single global MDS chart. In third approach, the result of each round is used to generate time series for all teams [Machado and Lopes 2017]. In [Carter et al. 2001], using multidimensional scaling authors have shown the analysis of sports photographs. To explore the properties of still photographs, observers judged the similarity under two conditions, image focus and movement focus. The image focus has shown how similar are images and movement focus has shown how similar is the type of movement.

2.8. Earthquake

The seismic data is difficult to analyze and classical mathematical tools impose strong limitations in unveiling the hidden relationships between earthquakes [Machado and Lopes 2013]. MDS is one of the approach which is useful to get the information regarding earthquakes. MDS maps are proven as an intuitive and useful visual representation of the complex relationships that are present among seismic events, which may not be perceived on traditional geographic maps [Lopes et al. 2013; 2014]. MDS produces spatial or geometric representation of complex objects, such that those objects that are perceived to be similar in some sense are placed on the MDS maps forming a cluster [Lopes et al. 2013]. The analysis of earthquake using MDS is studied in [Lopes et al. 2013] and [Lopes et al. 2014] using the data over three million seismic occurrences from the period of January 1 1904 to March 14 2012. In which, space-time and space-frequency correlation indices are proposed to quantify the similarities among events. According to [Lopes et al. 2014], MDS has the advantage of avoiding sensitivity to the non-uniform spatial distribution of seismic data, resulting from poorly instrumented areas, and is well suited for accessing dynamics of complex system.

2.9. Temperature Analysis

The climate change and global warming have major impact on economic, social and health aspects of the human life. The surface temperature time-series characterize earth as a slow dynamic spatio-temporal system [Lopes and Machado 2014]. The complex correlations between global temperature time-series using MDS has been studied in [Lopes and Machado 2014]. In which, MDS provides a graphical representation of the pattern of climatic similarities between regions around the globe. The similarities are evaluated through mathematical indices that correlate the monthly average temperatures observed in meteorological stations over a period of time.[Lopes and Machado 2014].

2.10. Forest Fire Analysis

Forest fires caused by natural factors, human negligence or human intent consume every year vast areas of vegetation [Machado and Lopes 2014]. Forest fires has direct impact upon economy due to the destruction of property and infrastructures. It raises the carbon dioxide emissions to the atmospheres, affects the water cycle, contributes to soil erosion and has economic implications associated with climate change. The authors in [Machado and Lopes 2014] has shown that the public domain forest fires information of events of Portugal from 1980 to 2012. In which, MDS approach has been used as a visualizing tool to generate maps where objects that are similar to each other are placed on the map forming clusters.

2.11. Virus Diseases Analysis

Viruses are infectious agents that replicate inside organisms and reveal a plethora of distinct characteristics. Viral infections spread in many ways, but often have devastating consequences and represent a huge danger for public health [Lopes and Machado 2016]. To design statistical and computational techniques capable of handling the available data and highlighting the most important features, [Lopes and Machado 2016] has used MDS. In which, proposed methodology has represent a solid mathematical tool to tackle a large number of virus and additional information about these infectious agents.

2.12. Localization

Recently MDS is used for localization purpose in wireless sensor networks. In [Shang et al. 2003] the authors addressed a sensor network localization problem based on MDS, where binary information is used as a proximity information. The main idea in [Shang et al. 2003] is to construct a global configuration using CMDS. The MDS based localization algorithms in [Shang et al. 2003] and [Vivekanandan and Wong 2006] are centralized having higher computational complexity [Saeed and Nam 2015]. A semi-centralized MDS is used in [Ji and Zha 2004], [Moore et al. 2004], [Yu and Wang 2008], [Shon et al. 2010], [Saeed and Nam 2016] and [Stojkoska et al. 2008] to compute local coordinates of nodes and these local coordinates are refined to find the final position of nodes. In [Patwari et al. 2003], [Li and Zhang 2007] and [Yin et al. 2012] the authors proposed manifold learning to estimate the sensor nodes location in wireless sensor networks. In [Macagnano and de Abreu 2013] the authors proposed Nystrom approximation for the proximity information matrix in MDS to reduce its size for better localization accuracy in sensor networks. Distributed MDS based localization algorithm is proposed in [Wei et al. 2015] with noisy range measurements, where they assume that the distances are corrupted with independent Gaussian random noise. MDS with different refinement schemes to get better localization accuracy for the sensor nodes location in WSNs have also been proposed in literature [Latsoudas and Sidiropoulos

2005], [Shi et al. 2007], [De Marziani et al. 2009], [Zhu and Ding 2012], [Saeed and Nam 2014].

There are various accurate localization algorithm based on MDS are proposed for 2-dimensional WSNs but in real world applications 3-dimensional localization is needed for better estimation and accuracy. In 2-dimensional WSNs with the help of three anchors (nodes with already known position), location of all other nodes in the network can be computed while in 3-dimensional networks atleast 4 anchor nodes are required. All the localization techniques that are developed for 2-dimensional WSNs are violated in 3-dimensional networks [Alam and Haas 2006]. 3-dimensional localization technique can not be directly extended from the 2-dimensional solution by just increasing one parameter. There are several problems which can be solved using 2-dimensional localization but are much more complex when modeled as 3-dimensional. Localization of WSNs in 3-dimensional is an interesting and challenging task. In [Stojkoska 2014], [Chaurasiya et al. 2014], [jun Peng and wei Li 2014] and [Stojkoska 2013] the authors proposed 3-dimensional localization for WSNs based on MDS.

The analogy between object distances and node distances in a network is used for the purpose of wireless sesnsro network (WSN) localization. MDS algorithm can use inter node distances in order to produce 2 or 3 dimensional representation, which corresponds to the real nodes deployment. Since nodes are capable to measure the inter node distances with respect to their neighboring nodes, the only problem remains obtaining the non neighboring inter node distances. In MDS-MAP, these distances are approximated with the distances calculated by Floyd Warshall shortest path algorithm [Floyd 1962].

Distances between every node in the network are collected at the central station (sink). The remaining (non neighboring) distances would be calculated by the sink. Thus, MDS can be classified as centralized, range-based localization algorithm. The well known MDS-MAP for 2-D network consists of the following steps:

1. Compute the shortest path distances between every node in the network (using either Dijkstra or Floyd algorithm). These shortest path distances work as an input data for MDS.

2. CMDS is applied to the shortest path distance matrix. The first 2 largest eigenvalues and eigenvectors forms the relative location of every node in the network.

3. Finally the relative locations are transformed to the absolute global map using the anchor nodes. This transformation includes optimal rotation, translation and reflection. This type of transformation is also called rigid or Euclidean transformation. Singular value decomposition (SVD) is one of the most stable transformation[Eggert et al. 1997].

3. MDS ALGORITHM

MDS is the approach that maps the original high dimensional data (m dimensions) in to a lower dimensional data (d dimensions). It addresses the problem of constructing a configuration between the n points from $k \times k$ matrix $\boldsymbol{D}$, which is called distance affinity matrix if it is symmetric, i.e., $d_{ii} = 0$, and $d_{ij} > 0,\ i \neq j$. MDS finds n data points $y_1, ...y_n$ from the distance matrix $\boldsymbol{D}$ in a d dimension space, such that if $\hat{d}_{ij}$ is the Euclidean distance between y_i and y_j, then $\hat{\boldsymbol{D}}$ is similar to $\boldsymbol{D}$. In [Ghodsi 2006; Cox and Trevor F 2008], MDS considered is

$$\min_{Y} \sum_{i=1}^{k} \sum_{i=1}^{k} (d_{ij}^{X} - d_{j}^{Y})^2 \tag{1}$$

where $d_{ij}^X = ||x_i - x_j||^2$ and $d_{ij}^Y = ||y_i - y_j||^2$. The distance matrix $\boldsymbol{D}^X$ is converted to a kernel matrix of inner product $\boldsymbol{X}^T\boldsymbol{X}$ by

$$\boldsymbol{X}^T\boldsymbol{X} = -\frac{1}{2}\boldsymbol{H}\boldsymbol{D}^X\boldsymbol{H} \tag{2}$$

where $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{t}\boldsymbol{e}\boldsymbol{e}^T$ and $\boldsymbol{e}$ is a column vector of 1's. The above equation can be written as

$$\min_Y \sum_{i=1}^{k}\sum_{i=1}^{k} \left(x_i^T x_j - y_i^T y_i\right)^2 \tag{3}$$

The solution is $\boldsymbol{Y} = \Lambda^{1/2}\boldsymbol{V}^T$ where $\boldsymbol{V}$ is the eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$ and Λ is the d eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$.

3.1. Input Data for MDS

In MDS based applications, the input data plays an important about the final spatial map in multidimensional space. The closeness of points data is generally termed as proximity and it is the measurement to which pair of objects are same or different [Okada and Imaizumi 1997]. The proximity measure for MDS based application is very diverse in nature. One of the major difference between the proximities is the way they are collected in direct or indirect fashion. The proximities are said to be direct if they are based on qualitative judgment or quantitative appraisals that directly shows the state of the point that how similar or dissimilar the point is in a set of points. These judgments can be transformed to preference rankings when points are needed to be sorted in terms of perception, preference or opinion [Bove 2006]. Indirect proximities are measured from certain type of information available, most often this information is in the form of variables in a matrix [Davidson and Skay 1991]. Confusion data is one of the examples of indirect proximity that is derived from perceptual mistakes [Gilmore et al. 1979]. A data matrix may consist of quantitative, qualitative or mixed information, therefore, it is a crucial task to set up the indirect proximities by choosing the most suitable measure to handle the variables based on their metric or non-metric characteristics.

The proximity information computed through direct or indirect method is organized into proximity matrices. This proximity matrix represents the basic input data for any MDS based application. One of the major obstacle in MDS based applications is to collect proximity information, which is an expensive task. The cheapest way to overcome this obstacle is that, the input data can be replaced by some assumptions. There are basic two major assumptions for MDS based applications. First, the dissimilarity ρ_{ij} between points i and j is always symmetric, because of this ρ_{ji} is not needed. The proximity information of a point to itself, ρ_{ii}, is zero [Gower 1966]. The dissimilarity or similarity between points i and j should satisfy the conditions of non-negativity ($\rho_{ij} \geq 0$), identity $\rho_{ii} = 0$ and symmetry $\rho_{ij} = \rho_{ji}$. In [Gower and Legendre 1986] and [Caillez and Kuntz 1996] the authors disagree with the above conditions, even though $\rho_{ij} = 0$ hold for two points i and j, there may be $\rho_{ik} \neq \rho_{jk}$ for third point k. This problem may not arise if dissimilarity or similarity information satisfies the definiteness property or triangle inequality i.e., both points i and j coincides for $\rho_{ij} = 0$ or $\rho_{ij} \leq \rho_{ik} + \rho_{jk}$ [Caillez and Kuntz 1996].

Generally, dichotomous (binary) variables are used for the similarities. Similarity concept is based on co-presence where a phenomena is present on points i and j. The opposite concept is of co-absence where a phenomena is absent on points i and j [Tan et al. 2005]. Qualitative variables are modeled as dichotomous variables where as quantitative variables can be modeled using well known Minkowski family of distances

[Klock and Buhmann 2000]. Consider m quantitative variables $\boldsymbol{Y}_l$ with observations y_{il}, then the dissimilarity is given as

$$\rho_{ij} = \left(\sum_{l=1}^{m} |y_{il} - y_{jl}|^r \right)^{\frac{1}{r}}, \tag{4}$$

where $r = 1$ for Manhatton or City-block [Groenen et al. 1997] , $r = 2$ for Euclidean and $r = \infty$ for Lagrange and Chebyshev models. Table 1 gives an overview of popular proximity measures for dissimilarity ρ_{ij} and similarity σ_{ij} between any two points i and j. Note that 1 to 4 models in Table III are the Minkowski distances for four different cases, which combine dimensional differences directly. The proximity models from 4 to 8 in Table 1 are to control the dispersion of variables [Gower 1966]. Absolute differences are corrected in the Canberra distance model along each dimension for the size of the coordinates along the axis. In addition, if y_{il} have negative values, then ρ_{ij} reaches an asymptote of infinity. Therefore, Canberra distance model is good to use when y_{il} have positive values. The Bray-Curtis distance model [Gerry and Keough 2004] is mostly used in the field of ecology, it tries to correct the sum of absolute differences. Bray-Curtis distance model also works well for positive values of y_{il}. The Chord distance [Ault 2007] also requires positive y_{il}. The angular separation computes the angle between y_i and y_j from the origin and its index lies between -1 and 1. The values of correlation coefficient also lies between -1 to 1 and it works well if their are large number of dimensions m.

MDS can handle mixed type of data which involves both quantitative and qualitative variables using general coefficient of similarity (GCS) [Gower 1971]. The GCS computes give wights to the similarity values and average it. The general form of similarity between points i and j is given by

$$\sigma_{ij} = \frac{\sum_{l=1}^{m} w_{ijl}\sigma_{ijl}}{\sum_{l=1}^{m} w_{ij}}, \tag{5}$$

where σ_{ijl} is the similarity values and w_{ijl} is the non negative weights between points i and j respectively. Weighting is used to handle the missing values, a weight $w_{ijl} = 0$ when the observation between points i and j is not available. Finally, GCS give much weights to the qualitative variables, when the variables are heterogeneous, such as Gini's or Shannon's heterogeneity [Solaro 2010].

3.2. Loss Functions for MDS

MDS models need to map each dissimilarity exactly into its Euclidean distance, which may not be practical due to many factors like noise added to the measurement, sampling effect and unreliability [Borg and Gronen 2005]. In order to get as close result as possible to ensure the equality requirement in $f(\rho_{ij}) = d_{ij}(\boldsymbol{X})$ as mentioned in section 3.3, most of the computer programs used to solve this problem usually starting with some initial configuration and tries to minimize the difference between $f(\rho_{ij})$ and $d_{ij}(\boldsymbol{X})$ iteratively. To ensure the equality sign holds, a statistical concept of error is used in literature also called squared error, defined as

$$e_{ij}^2 = \Big(f(\rho_{ij}) - d_{ij}(\boldsymbol{X}) \Big)^2. \tag{6}$$

Table III. Proximities

S. No	Model	Formula
1	Euclidean	$\rho_{ij} = \left(\sum_{l=1}^{m} \lvert y_{il} - y_{jl}\rvert^2\right)^{\frac{1}{2}}$
2	City block and Manhatton	$\rho_{ij} = \left(\sum_{l=1}^{m} \lvert y_{il} - y_{jl}\rvert\right)$
3	Langrange and Chebyshev	$\max_{l=1}^{m} \lvert y_{il} - y_{jl}\rvert$
3	Minkwoski	$\rho_{ij} = \left(\sum_{l=1}^{m} \lvert y_{il} - y_{jl}\rvert^r\right)^{\frac{1}{r}}$ with $r \geq 1$
4	Canberra	$\rho_{ij} = \sum_{l=1}^{m} \frac{\lvert y_{il} - y_{jl}\rvert}{\lvert y_{il} + y_{jl}\rvert}$
5	Bray-Curtis	$\rho_{ij} = \frac{\sum_{l=1}^{m} \lvert y_{il} - y_{jl}\rvert}{\sum_{l=1}^{m} (y_{il} + y_{jl})}$
6	Chord	$\rho_{ij} = \left(\sum_{l=1}^{m} \lvert y_{il}^{\frac{1}{2}} - y_{jl}^{\frac{1}{2}}\rvert^2\right)^{\frac{1}{2}}$
7	Angular Separation	$\sigma_{ij} = \frac{\sum_{l=1}^{m} y_{il} y_{jl}}{(\sum_{l=1}^{m} y_{il}^2)^{\frac{1}{2}} (\sum_{l=1}^{m} y_{jl}^2)^{\frac{1}{2}}}$
8	Correlation Coefficient	$\sigma_{ij} = \frac{\sum_{l=1}^{m} (y_{il} - \bar{y_i})(y_{jl} - \bar{y_j})}{(\sum_{l=1}^{m} (y_{il} - \bar{y_i})^2)^{\frac{1}{2}} (\sum_{l=1}^{m} (y_{jl} - \bar{y_j})^2)^{\frac{1}{2}}}$

For all pairs of points (i, j) combines the total error among the configurations yielding raw Stress, represented as

$$\xi_r = \sum_{i=1}^{n} \sum_{j<i}^{n} \Big(f(\rho_{ij}) - d_{ij}(\boldsymbol{X}) \Big)^2. \tag{7}$$

The main problem with the raw Stress in (7) is that, it is not invariant under coordinate scaling transformation. To overcome this problem of invariability some forms of normalization are introduced. The most suitable choice for normalization is using the disparities $\hat{d}_{ij}$. Normalization using the disparities develop the famous Kruskal's Stress-1 [Kruskal 1964a], given as

$$\xi_1 = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j<i}^{n} \Big(\hat{d}_{ij} - d_{ij}(\boldsymbol{X}) \Big)^2}{\sum_{i=1}^{n} \sum_{j<i}^{n} d_{ij}^2(\boldsymbol{X})}}. \tag{8}$$

Kruskal's Stress-2 formula is derived by inducing the mean of the distances $\bar{d}$ as well, for $1 \leq i < j \leq n$, given as

$$\xi_2 = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j<i}^{n} \Big(\hat{d}_{ij} - d_{ij}(\boldsymbol{X}) \Big)^2}{\sum_{i=1}^{n} \sum_{j<i}^{n} \big(d_{ij}(\boldsymbol{X}) - \bar{d} \big)^2}}. \tag{9}$$

Minimizing the Stress function always requires to find the optimal coordinates for $\boldsymbol{X}$ in a given dimension m. Therefore, in [Borg and Gronen 2005] the authors proposed normalized Stress based on the sum of squares of the dissimilarities ρ_{ij} which are not accounted for the distances, defined as

$$\xi_n = \frac{\xi_r}{\eta_\rho^2} = \frac{\sum_{i=1}^{n} \sum_{j<i}^{n} \Big(f(\rho_{ij}) - d_{ij}(\boldsymbol{X}) \Big)^2}{\sum_{i=1}^{n} \sum_{j<i}^{n} w_{ij} \rho_{ij}^2}. \tag{10}$$

Clearly, if $\sum_{i=1}^{n}\sum_{j<i}^{n} w_{ij}\rho_{ij}^2 = 1$, then $\xi_n = \xi_r$, where w_{ij} is the associated weight for each dissimilarity. In [Leeuw et al. 1977] it is shown that how the normalized Stress ξ_n is related to the Tucker's coefficient. Consider that $\bar{\boldsymbol{X}}$ is a local minimum for the raw Stress ξ_r, which implies that $c\bar{\boldsymbol{Y}} = \bar{\boldsymbol{X}}$ must be a local minima with $c > 0$. To find the optimal value of c the homogeneity property of Euclidean distance is considered i.e., $d_{ij}c\bar{\boldsymbol{Y}} = cd_{ij}\bar{\boldsymbol{Y}}$. Then $\xi_r(c\bar{\boldsymbol{Y}})$ can be written as

$$\begin{aligned}\xi_r(c\bar{\boldsymbol{Y}}) &= \sum_{i<j} w_{ij}\Big((\rho_{ij}) - d_{ij}(c\bar{\boldsymbol{Y}})\Big)^2 \\ &= \sum_{i<j} w_{ij}\rho_{ij}^2 + c^2\sum_{i<j} w_{ij}d_{ij}^2(\bar{\boldsymbol{Y}}) - 2c\sum_{i<j} w_{ij}\rho_{ij}d_{ij}(\bar{\boldsymbol{Y}}) \\ &= \eta_\rho^2 + c^2\eta_d^2(\bar{\boldsymbol{Y}}) - 2c\zeta(\bar{\boldsymbol{Y}}).\end{aligned} \tag{11}$$

The minimum of (11) for optimum c is obtained by setting the first derivative of $\xi_r(c\bar{\boldsymbol{Y}}) = 0$, $2c\eta_d^2(\bar{\boldsymbol{Y}}) - 2\zeta(\bar{\boldsymbol{Y}}) = 0$, which gives the optimum value of c as, $c^\star = \frac{\zeta(\bar{\boldsymbol{Y}})}{\eta_d^2}$ [Mathar 1990]. Inserting the optimum value of c in $\xi_r(c\bar{\boldsymbol{Y}})$ yields

$$\xi_r(c^\star\bar{\boldsymbol{Y}}) = \eta_\rho^2 - \left(\frac{\zeta(\bar{\boldsymbol{Y}})}{\eta_d(\bar{\boldsymbol{Y}})}\right)^2. \tag{12}$$

Dividing both sides by η_ρ^2 gives

$$\xi_n(c^\star\bar{\boldsymbol{Y}}) = 1 - \left(\frac{\zeta(\bar{\boldsymbol{Y}})}{\eta_\rho\eta_d(\bar{\boldsymbol{Y}})}\right)^2, \tag{13}$$

where the last term of (13) gives the value of the Tucker's congruence coefficient, which lies in the range of -1 and 1 because of the Cauchy-Schwarz inequality. Thus, using the normalized Stress provides a clear interpretation of the loss function which does not depends on proximities scale.

There exist a simple relation between Kruskal's Stress-1 ξ_1 and the normalized Stress ξ_n. Consider the local minimum $\bar{\boldsymbol{X}}$ after minimizing the normalized Stress ξ_n, which implies that $\eta_d(\bar{\boldsymbol{X}}) = \zeta\bar{\boldsymbol{X}}$ and

$$\xi_n(\bar{\boldsymbol{X}}) = 1 - \frac{\eta_d^2(\bar{\boldsymbol{X}})}{\eta_\rho^2}, \tag{14}$$

where as the Stress-1 can be written as

$$\begin{aligned}\xi_1(\bar{\boldsymbol{X}}) &= \frac{\eta_\rho^2 - \eta_d^2(\bar{\boldsymbol{X}})}{\eta_d^2(\bar{\boldsymbol{X}})} \\ &= \frac{\eta_\rho^2}{\eta_d^2(\bar{\boldsymbol{X}})} - 1.\end{aligned} \tag{15}$$

From (14) $\frac{\eta_d^2(\bar{\boldsymbol{X}})}{\eta_\rho^2 - \eta_d^2(\bar{\boldsymbol{X}})} = \frac{1}{1-\xi_n(\bar{\boldsymbol{X}})}$, yielding

$$\xi_1(\bar{\boldsymbol{X}}) = \frac{\xi_n(\bar{\boldsymbol{X}})}{1 - \xi_n(\bar{\boldsymbol{X}})}. \tag{16}$$

Inducing the scaling factor c, the Stress-1 becomes

$$\xi_1(c\bar{\boldsymbol{X}}) = \frac{\eta_\rho^2 + c^2\eta_d^2(\bar{\boldsymbol{X}}) - 2c\zeta(\bar{\boldsymbol{X}})}{c^2\eta_d^2(\bar{\boldsymbol{X}})}. \tag{17}$$

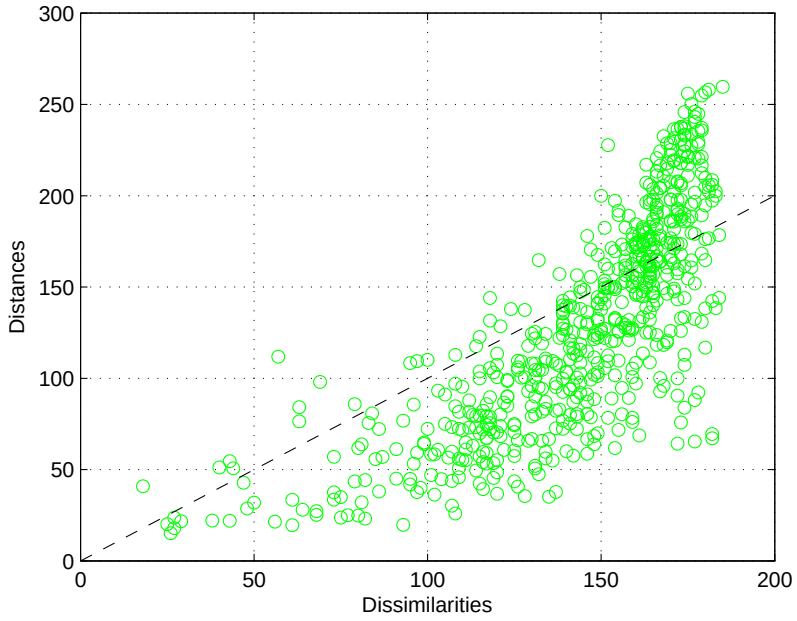

Fig. 1. Shephard Scatter Plot

In order to find the optimum value of scaling factor c, differentiating $\xi_1(c\bar{\boldsymbol{X}})$ with respect to c, that is,

$$\begin{aligned}\frac{\partial \xi_1(c\bar{\boldsymbol{X}})}{\partial c} &= \frac{2c^2(c-1)\eta_d^4(\bar{\boldsymbol{X}}) - 2c\eta_d^2(\bar{\boldsymbol{X}})(\eta_\rho^2 + (c^2-2c)\eta_d^2(\bar{\boldsymbol{X}}))}{c^4\eta_d^2(\bar{\boldsymbol{X}})}, \\ &= \frac{2c\eta_d^2(\bar{\boldsymbol{X}}) - 2\eta_\rho^2}{c^3},\end{aligned} \tag{18}$$

which is equal to zero for $c^\star = \frac{\eta_\rho^2}{\eta_d^2(\bar{\boldsymbol{X}})}$, inserting the optimum value of c in $\xi_1(c\bar{\boldsymbol{X}})$ yields

$$\xi_1(c^\star\bar{\boldsymbol{X}}) = \frac{\eta_\rho^2/\eta_d^2(\bar{\boldsymbol{X}}) - 1}{\eta_\rho^2/\eta_d^2(\bar{\boldsymbol{X}})} = \xi_n(\bar{\boldsymbol{X}}). \tag{19}$$

Thus, it is proved in (19), that normalized Stress and Stress-1 are same at local minimum. In [Leeuw et al. 1977] the authors introduce a methodology to minimize the Stress function, which shows that beforin starting the minimization process the proximities are normalized. This type of normalization is also known as scaling by majorizing a complicated function (SMACOF). The majorizing operation is efficient from a computational point of view and is being used in many MDS models [Cox and Cox 2001].

Loss functions basically tells us the mismatch between the dissimilarities and the corresponding distances, while the Stress function shows the badness-of-fit between them. The scatter diagram also called Shepard diagram [Leeuw 2005] plots the dissimilarities and distances as shown in Fig. 1. A regression line also exist which shows how the dissimilarities and their distances are related.

3.3. MDS Output and Further Processing

Once MDS produces the coordinates for points in a m-dimensional space, the output configuration is not explicit in terms of reflection, dilation and rotation. For the configuration of points $\boldsymbol{X}$ in m-dimensions, with $\boldsymbol{x}_i$ representing the i-th row, then $\hat{\boldsymbol{x}}_i = s\boldsymbol{T}\boldsymbol{x}_i + \boldsymbol{\tau}$, where s is the dilation factor ($s < 1$ for shrinking and $s > 1$ for stretching), $\boldsymbol{T}$ is the orthogonal matrix representing rotation or reflection and $\boldsymbol{\tau}$ is the translation. The non-unique solution of MDS creates an ambiguity about comparing the two configurations taking into account the above transformations. Solution to this problem lies in Procrustes analysis [Borg and Gronen 2005]. Consider that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the two set of coordinates for n number of points in m and $\bar{m}$ different dimensions respectively, where $\bar{m} \leq m$. Moreover, when $\bar{m} \leq m$, then $m - \bar{m}$ columns of zeros are added to $\boldsymbol{Y}$ to convert it into $n \times m$ matrix. Then the matching between the two configurations considers all the possible reflections, rotations, dilation and translations of $\boldsymbol{X}$. The miss-match between the two configurations can be formulated as a minimization problem given as

$$\min_{s,\boldsymbol{\tau},\boldsymbol{T}} R^2 = \min_{s,\boldsymbol{\tau},\boldsymbol{T}} \sum_{i=1}^{n} (\boldsymbol{y}_i - s\boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{\tau})^T (\boldsymbol{y}_i - s\boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{\tau}), \tag{20}$$

where R^2 is the summed and squared intra-point distances. All the mathematical derivation to find the optimal values for s,$\boldsymbol{\tau}$ and $\boldsymbol{T}$ can be found in [Cox and Cox 2001] and [Mardia 1979].

4. VARIOUS MDS TECHNIQUES

Until now number of different MDS techniques have been proposed. Most of them tries to represent the coordinates for the observed dissimilarities in m-dimensional space. The dissimilarities are mapped in such a way that it tries to match the Euclidean distances. Thus, the dissimilarity ρ_{ij} between points i and j is mapped into its Euclidean distance d_{ij} with the minimum information loss. Let $\boldsymbol{X}$ is an $n \times m$ configuration of points, where m defines the dimensional space and n is the number of points. The the dissimilarities are related to the Euclidean distance by function $f : \rho_{ij} \rightarrow d_{ij}(\boldsymbol{X})$, where $d_{ij}(\boldsymbol{X})$ implies that the distance d_{ij} depends on the unknown coordinates of $\boldsymbol{X}$. Different models of MDS can be defined based on this mapping of dissimilarities to the distances. Hence every MDS model starts with the equation $f(\rho_{ij}) = d_{ij}(\boldsymbol{X})$. The equality sign however has only theoretical value in real world applications it does not hold exactly. Therefore from a practical point of view usually a weaker statement should be defined with approximation relation $f(\rho_{ij}) \approx d_{ij}(\boldsymbol{X})$, where as the exact equality holds when the dissimilarity is in the form of disparity. Dissimilarities ρ_{ij} are converted into disparities by applying function f to it i.e., $\hat{d}_{ij} = f(\rho_{ij})$. Some of the commonly used functions for defining the disparities from the dissimilarities are, linear function $\hat{d}_{ij} = b\rho_{ij}$ also called ratio MDS, identity function $\hat{d}_{ij} = \rho_{ij}$ i.e., absolute MDS and finally $\hat{d}_{ij} = g + h\rho_{ij}$, which corresponds to interval MDS, parameters g and h are selected such that it guarantees the relation to hold as far as possible. Thus, any kind of function can be considered for the mapping which is continuous, parametric and monotone in nature. According to proximities, MDS have 2 different types, metric MDS (MMDS) and NMDS. In MMDS, proximity is related to distance $\rho_{ij} = \hat{d}_{ij}$, whose cost function is [Zhang 2007]

$$X = \sum_{i \neq j} \left(\hat{d}_{ij} - d_{ij}\right)^2. \tag{21}$$

Each proximity value ρ_{ij} should correspond exactly to the distance between points i and j in the m-dimensional MDS Space. The NMDS model presents only ordinal properties of the data, e.g. if $\rho_{12} = 6$ and $\rho_{34} = 3$, an ordinal model reads this as $\rho_{12} > \rho_{34}$ [Zhang 2007]. In other words, MMDS considers that the the proximity information, e.g., display of numeric data, like the distances measured between objects. Thus, the MMDS tries to preserve the distances as close as possible to the intervals and ratios between the proximities. While NMDS only consider proximty information order. In NMDS the order of the distances gives the information regarding the order of the proximity information. There are four types of models comes under MMDS, classical MDS (CMDS), replicated MDS (RMDS), Generalized MDS (GMDS), weighted MDS (WMDS) and Isomap.

4.1. CMDS

CMDS is similar to generating of a map. By starting with the distances between different objects and then try to arrange points on a piece of paper so that the distances between the points in the map representing each of those objects are related to the true distances between the objects [Borg and Gronen 2005; Amar et al. 2010]. Stating this more mathematically CMDS endeavors to find an isometry between points distributed in a higher dimensional space and in a low dimensional space. In other words, if there are n m-dimensional points, $\boldsymbol{X}$, then the the dissimilarities between the pairs of points is, p_{ij}. The CMDS tries to create n projections, x, of the high dimensionality points in a d-dimensional linear space by trying to arrange the projections so that the Euclidean distances between pairs of them, d_{ij}, resemble the dissimilarities between the high dimensional points. In short CMDS tries to minimize:

$$\chi = \sum_{i \neq j} (p_{ij} - d_{ij})^2 \tag{22}$$

where p_{ij} is the dissimilarity between point X_i and point X_j, and d_{ij} is the distance between the projection of X_i, x_i, and the projection of X_j, x_j. The classical MDS finds the location of points in matrix form $\boldsymbol{X}$ by taking the eigenvalue decomposition of the double centered matrix $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{X}'$. Double centered matrix $\boldsymbol{B}$ is constructed from the proximity matrix $\boldsymbol{P}$ by multiplying it with the centering matrix $\boldsymbol{J} = \boldsymbol{I} - n^{-1}\boldsymbol{1}\boldsymbol{1}'$. Following are the major steps for CMDS:

— Compute the squared proximity matrix $\boldsymbol{P} = [p^2]$.
— Double center the proximity information i.e., $\boldsymbol{B} = \frac{1}{2}\boldsymbol{J}\boldsymbol{P}\boldsymbol{J}$ using the centering operator $\boldsymbol{J} = \boldsymbol{I} - n^{-1}\boldsymbol{1}\boldsymbol{1}'$, where n tells us about the total number of objects. Double centering is the method of subtracting the column and row means of a matrix from its elements and adding the grand mean.
— Next step is to extract m eigenvectors $e_1, ..., e_m$ and the corresponding eigenvalues $\lambda_1, ..., \lambda_m$.
— Finally the coordinates for the n objects in m-dimensional space are derived from $X = \boldsymbol{E}_m \boldsymbol{\Lambda}_m^{\frac{1}{2}}$, where $\boldsymbol{E}_m$ are the m eigenvectors and $\boldsymbol{\Lambda}_m$ are the m eigenvalues, respectively.

In situations where the dissimilarity p_{ij} is not Euclidean distance, CMDS can still be applied, such that matrix $\boldsymbol{B}$ is positive semi-definite or the dissimilarity matrix $\boldsymbol{P}$ has Euclidean properties. If matrix $\boldsymbol{B}$ is not positive semi-definite then there can be large number of negative eigenvalues. Under these circumstances dissimilarity matrix $\boldsymbol{P}$ is considered Euclidean by transformations [Gower 1966], while very small eigenvalues are ignored [Cox and Cox 2001].

4.2. GMDS

GMDS is the extension of CMDS in which the dissimilarity matrix is non Euclidean [Bronstein et al. 2006b]. CMDS evaluates all of the permutations between two different surfaces, which is an expensive task, while GMDS can work with both, partial as well as full surface matching. In other words, GMDS is a technique which computes the map that best preserves the inter-geodesic distances while embedding one surface into another. There are number of other choices than Euclidean distances e.g., hyperbolic [Walter and Ritter 2002] and spherical [Elad and Kimmel 2002] spaces, which can be more advantageous for different surfaces in GMDS. Consider that, if $\boldsymbol{G_1}$ and $\boldsymbol{G_2}$ represent the inter-geodesic distances matrix of C_1 and C_2 configurations, respectively, then GMDS attempts to find the permutation matrix $\boldsymbol{W}$ minimizing $\parallel \boldsymbol{W}\boldsymbol{G}_1 - \boldsymbol{G}_2\boldsymbol{W} \parallel_2^2$, i.e.,

$$\min_{\boldsymbol{W}} \parallel \boldsymbol{W}\boldsymbol{G}_1 - \boldsymbol{G}_2\boldsymbol{W} \parallel_2^2, \tag{23}$$

where $\parallel . \parallel_2^2$ represents the discretization of the L_2 norm of a mapping between C_1 and C_2 and $\boldsymbol{W}_{ij} = \{0, 1\}$. In continuous setting,

$$||F||_2^2 = \int \int_{S_1,S_2} F^2(x, y)da(x_1)da(x_2) \tag{24}$$

and for the discrete setting,

$$||F||_2^2 \approx trace(\mathbf{F}^T\mathbf{A}_2\mathbf{F}\mathbf{A}_1) \tag{25}$$

The L_2 in above equation can be written as

$$\parallel \boldsymbol{W}\boldsymbol{G}_1 - \boldsymbol{G}_2\boldsymbol{W} \parallel_2^2 = trace\left((\boldsymbol{W}\boldsymbol{G}_1 - \boldsymbol{G}_2\boldsymbol{W})^T\boldsymbol{G}_2(\boldsymbol{W}\boldsymbol{G}_1 - \boldsymbol{G}_2\boldsymbol{W})\boldsymbol{G}_1)\right) \tag{26}$$

which can be written as

$$-2\ trace\left(\boldsymbol{W}^T\boldsymbol{G}_2\boldsymbol{W}\boldsymbol{G}_1\right). \tag{27}$$

Problem in (27) appears to be non-deterministic polynomial-time (NP) hard problem which ignores the smooth relation between the two configurations. Several methods have been proposed to overcome this problem [Bronstein et al. 2006b]. Recently spectral GMDS (S-MDS) is proposed in [Aflalo et al. 2013] to further reduce the complexity of GMDS and to overcome the non-convex nature of GMDS.

4.3. WMDS

In WMDS an extra parameter is also computed for the proximities to fit the points and their corresponding dissimilarities in a better fashion. The WMDS also know as INDSCAL was first by proposed by Carroll and Chang [Carroll and Chan 1970], which removes the rotational in-variance existing in the CMDS, thus providing the user with dimensions that are potentially psychologically meaningful [Greenacre 2005]. Once these weights are estimated, the rest of the procedure is similar to CMDS. The loss function for WMDS is defined as

$$\chi = \sum_{j=i+1}^{n} w_{ij}\left(p_{ij} - d_{ij}(\boldsymbol{X})\right)^2, \tag{28}$$

where w_{ij} are the associated weights which quantify the accuracy of the dissimilarity p_{ij}. if there is no dissimilarity information available between points i and j then $w_{ij} = 0$. Weights w_{ij} tells us about the accuracy of proximities, such that more accurate measurements are given higher weight-age in the overall loss function. For a given measurement noise model, w_{ij} is modeled based on the variance [Costa et al.

2006]. In [Leeuw and Heiser 1980] the authors used the idea of weighting, some of them are S-Stress, STRAIN and Sammon's mapping.

Consider the S-Stress or STRAIN defined for ALSCAL [Young et al. 1978], the error function for S-Stress is defined as

$$\xi_s = \sum_{j=i+1}^{n} \Big(p_{ij} + d_{ij}(\boldsymbol{X})\Big)^2 \Big(p_{ij} - d_{ij}(\boldsymbol{X})\Big)^2, \tag{29}$$

where the S-Stress depends on the two terms, $\Big(p_{ij} - d_{ij}(\boldsymbol{X})\Big)^2$ and the weighting term $\Big(p_{ij} + d_{ij}(\boldsymbol{X})\Big)^2$, both of terms are dependent on $d_{ij}(\boldsymbol{X})$. Assuming that the residual error is considerably small, then the weighing factor $\Big(p_{ij} + d_{ij}(\boldsymbol{X})\Big)^2$ can be approximated such that

$$\Big(p_{ij} + d_{ij}(\boldsymbol{X})\Big)^2 \approx 4p_{ij}^2. \tag{30}$$

Therefore, the S-Stress can be minimized by setting the weight $w_{ij} = 4p_{ij}^2$. In [McGee 1966] the authors proposed elastic distances for the MDS, which provides equal importance to the fitting of both the small and large dissimilarities. The error function for such weighting is

$$\xi_e = \sum_{j=i+1}^{n} \left(1 - \frac{d_{ij}(\boldsymbol{X})}{p_{ij}}\right)^2 = \sum_{i<j} p_{ij}^{-2} \Big(p_{ij} - d_{ij}(\boldsymbol{X})\Big)^2 \tag{31}$$

Sammon mapping [Sammon 1969] is popular in pattern recognition, it also considers WMDS [Lerner et al. 1998], where the weight $w_{ij} = p_{ij}^{-1}$. The loss function for Sammon mapping is defined as

$$\xi_{sam} = \sum_{j=i+1}^{n} \left(1 - \frac{d_{ij}(\boldsymbol{X})}{p_{ij}}\right)^2 = \sum_{i<j} p_{ij}^{-1} \Big(p_{ij} - d_{ij}(\boldsymbol{X})\Big)^2, \tag{32}$$

which is similar to the raw Stress. The objective of Sammon mapping is somehow similar to elastic scaling. In [Jourdan and Melangon 2004] the authors proposed MULTISCALE loss function which can be written as

$$\xi_m = \sum_{j=i+1}^{n} \log^2 \left(\frac{d_{ij}(\boldsymbol{X})}{p_{ij}}\right)^2, \tag{33}$$

which tries to minimize the relative error logarithmically. In most of the above techniques weights w_{ij} are specified on the basis of some formal considerations. One way to choose w_{ij} is to equalize it with reliability of the proximity information, which means that more reliable proximities are given more weight while unreliable proximities have less weight.

4.4. Isomap

Isomap is also a dimensionality reduction technique which maps the high dimensional structures into low dimensional space [Tenenbaum et al. 2000]. Isomap is the extension of MDS such that it considers the practical geodesic distances instead of the Euclidean distances between each pair of the points in the configuration. In Isomap, first

of all the geodesic distances are computed between each pair of points to construct a graph G. In graph, there is a link between point i and j if the geodesic distance g_{ij} is smaller than the threshold ϵ, $g_{ij} < \epsilon$, and the value of this edge is equal to g_{ij}. Once the value for each vertex is computed the shortest path distances to every other point in the configuration is computed using Floyd Warshall algorithm. Once the distances for each pair of points are available, a low dimensional embedding of the points is needed to be find out. The possible loss function for Isomap is defined as

$$\xi_i = \sum_{i \neq j} (g_{ij} - d_{ij})^2. \tag{34}$$

The geodesic distance with respect data set D, a distance $d(u, v)$ and neighborhood k as [Bengio et al. 2003]

$$\bar{D}(a, b) = \min_p \sum_i d(p_i, p_{i+1}) \tag{35}$$

where p is the sequence of length $l \geq 2$ with $p_1 = a$, $p_l = b$, $p_i \in D \forall i \in (2, ..., l - 1)$ and (p_i, p_{i+1}) are k nearest neighbors. The length l is free in minimization. The Isomap algorithm gets the normalized matrix M from which the embedding is derived by transforming the raw pairwise distance matrices. The matrix $M_{ij} = \bar{D}^2(x_i, x_j)$ of squared geodesics distances with respect to the data D is computed first, then distance-to-dot product transformation is applied to this matrix as

$$\bar{M_{ij}} = -\frac{1}{2} \left(M_{ij} - \frac{1}{n} S_i - \frac{1}{n} S_j + \frac{1}{n^2} \sum_k S_k \right) \tag{36}$$

where $S_i = \sum_j M_{ij}$.

4.5. NMDS

Non-metric MDS (NMDS) and CMDS, tries to compute the coordinates of the objects in m-dimensional space, such that the proximities matches the inter-point distances. [Zhang and Cheng 2010; Nhat et al. 2008; Xin et al. 2009]. The foundation of NMDS is motivated by two drawbacks of CMDS [Guttman 1968b]:

— Define a function explicitly such that the given dissimilarities are transformed into distances.
— Object configuration is restricted to be determined in Euclidean geometry.

Basically NMDS demands a less constrained relationship between the proximities and the distances. Consider a function f which is monotonically increasing, such that $d_{ij} = f(\delta_{ij})$, generates a set of distances d_{ij} from the given proximities δ_{ij}. The function f works in such a fashion that $\delta_{ij} < \delta_{rs}$ for $f(\delta_{ij}) < f(\delta_{rs})$. The input data to the NMDS is matrix of dissimilarities

$$D = |\delta_{ij}| \tag{37}$$

where d_{ij} is the dissimilarity matrix of i and j.

NMDS depends on the rank order of the proximities rather than its true values, therefore it is ordinal in nature and also called ordinal MDS.

Hence, the distance of the final configuration should be in same rank order as the original data. Therefore, the purpose of NMDS is to find the configuration of points whose distances reflect as near as possible the rank order of the data.

Table IV. Comparison of MDS Techniques

MDS Techniques	**Linear/ Non-Linear**	**Closed form/ No-Closed form**	**Complexity**	**Stress**	**Application**
PCA	Linear [Chengan Guo and Zeng 2013]	Closed form	$O(N^2)$ [R. Leszek and Zurada 2010]	High	(1) Scientific visualization [Cheng 2006], (2) Pattern Recognition [Thomas Villmann and Seiffert 2008], (3) Localization (4) Ecology [Janzekovic and Novak 2012], (5) Psychological Structure [Schinka and Velicer 2013], (6) Data Exploration [Zuur et al. 2009] and Testing Structural Hypothesis [Yamamoto et al. 2014]
CMDS	Linear [Werner Dubitzky and Berrar 2007]	Closed form	$O(N^3)$ [Pawliczek and Dzwinel 2013; Tzeng et al.]	High	(1) Scientific Visualization [Ghodsi 2006], (2) Pattern recognition [Hancock and Vento 2003], (3) Localization, (4) Psychological structures [Jaworska and Chupetlovska-Anastasova 2009], (5) Data Exploration [Filzmoser and Hron] and (6) Ecology [Greenacre and Primicerio 2013]
GMDS	Non-Linear	No-closed form	$O(N^2logN$ [Sahillioglu and Yemez 2011]	Low	(1) Scientific Visualization [Lezoray and Grady 2012], (2) Pattern recognition [Bronstein et al. 2006b], (3) Localization
NMDS	Non-linear	No-closed form	$O(N\sqrt{N})$ [Tzeng et al.]	Higher	(1) Psychological structures [Jaworska and Chupetlovska-Anastasova 2009], (2) Data Exploration [Naud and Duch 2000], (3) Testing Structural Hypothesis [Kruskal 1964a] (4) Ecology [Greenacre and Primicerio 2013]
Isomap	Non-linear	No-closed form	$O(N^2logN+ N^2D)$ [Burges 2005]	Low	(1) Scientific Visualization [Ghodsi 2006], (2) Pattern recognition [Yang 2002] (3) Data Exploration [Lim et al. 2003]

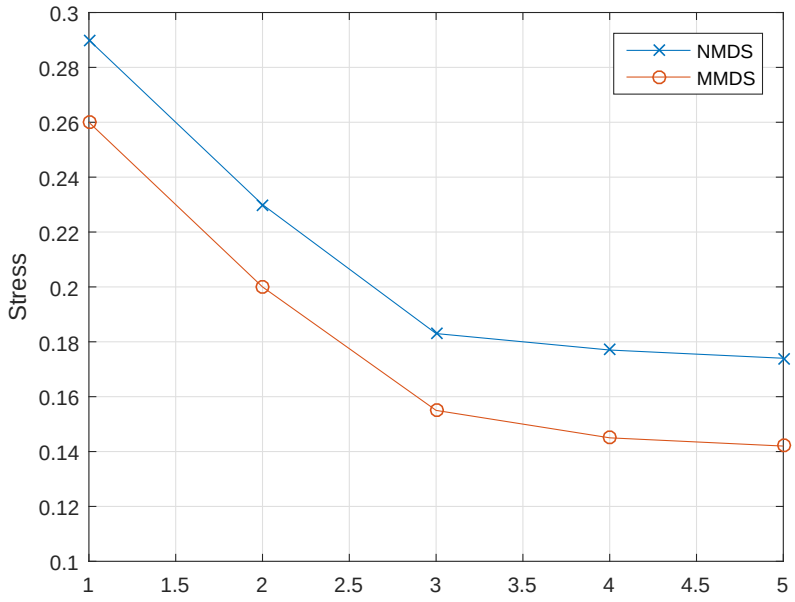

Fig. 2. Stress of MMDS, NMDS and Isomap

5. COMPARISON OF VARIOUS MDS TECHNIQUES

In this section, different MDS techniques are compared. It is mentioned earlier that GMDS is simply an extension of CMDS, in which target space is an arbitrary smooth non-Euclidean space. GMDS is used in the situation when the dissimilarities are distances on a surface and the target space is another surface. It allows to find the minimum distortion embedding of one surface into another surface[Bronstein et al. 2006b]. Isomap is also an extension of CMDS, but it uses geodesic distances as an input matrix to MDS [Zhang 2007]. The WMDS generalizes the distance model so that numerous matrices could be assumed to be distinguished from each other in a systematically non-linear or monotonic ways. The WMDS incorporates a model to account for individual difference in the fundamental perceptual or cognitive process that generate the responses [Young 2013]. The weight in WMDS is given on the basis of distance measurement, i.e., if estimated distance have high error, then less weight is given and vice versa [Zhang 2007]. We have compared the MDS techniques according to their stress function using linear data, the performance of metric and non-metric is shown in Fig. 2. Total 30 nodes are considered, in which 27 nodes are randomly distributed in the area of $10\text{m}\times10\text{m}$, while 3 nodes are considered as anchor nodes with known position. Fig. 2 clearly shows that the performance of MMDS is better than NMDS technique. In Fig. 2 MMDS refers to CMDS and GMDS. It is shown that the stress function of MMDS is less than NMDS. As in MMDS, actual distances are used while in NMDS the rank is used as input data. In Isomap and WMDS, stress will be the same as of MMDS due to the fact that both techniques use euclidean distances as an input matrix.

The comparison of all techniques using non-linear toy data is shown in Fig.3. It shows the effect of different types of MDS on the selected non-linear data. As seen in the figure, GMDS and Isomap preserve the given (non-linear) data in the output mostly due to the fact that those techniques use geodesic distances. However, the performance

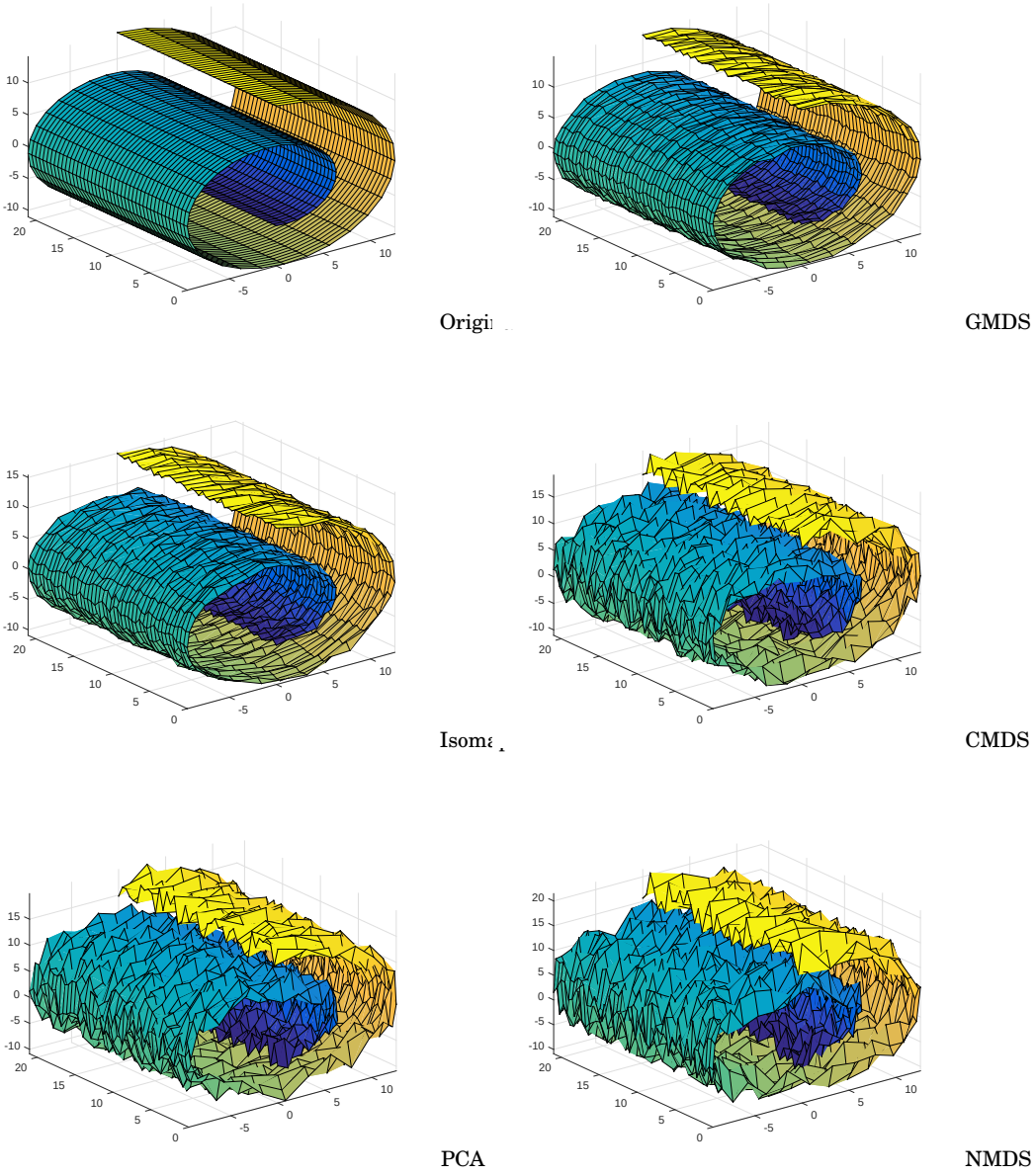

Fig. 3. Comparison of Different MDS techniques with non-linear data

of GMDS is better than Isomap though both of them are using geodesic distances. The performance of CMDS and PCA is also shown, which clearly indicates that both are worse than GMDS and Isomap due to the fact that CMDS and PCA are designed to work better when the given data is embedded linearly in the observation space.

Furthermore, different models of MDS are compared in Table IV-E in terms of the linearity/non-linearity, closed form solution, complexity, stress and corresponding ap-

plications. In which, N is the number of points and D is the number of dimensions. It is clear from Table. IV-E that PCA and CMDS are linear techniques, which have closed form expressions, while GMDS, NMDS and Isomap are non linear techniques and their closed form expressions do not exist. The complexity of each MDS technique is also shown along with their corresponding stress. As shown in the table, many applications use PCA and CMDS, while rest of MDS techniques are used by limited number of applications.

6. CONCLUSIONS

In this survey paper we have provided an overview of well known multidimensional scaling techniques, namely classical MDS and non-metric MDS and their real world applications. We have briefly discussed several other techniques based on CMDS, that is, the weighted MDS, generalized MDS, and Isomap. We have also discussed in detail the input data for each technique, which is a major part for any all the MDS methods. Furthermore the loss functions for each MDS technique is devised and all the MDS methods are compared.

REFERENCES

Y. Aflalo, A. Dubrovina, and R. Kimmel. 2013. Spectral Generalized Multi-Dimensional Scaling. *CoRR* 1311 (Nov 2013).

S M N Alam and Z J Haas. 2006. *Topology Control and Network Lifetime in Three-Dimensional Wireless Sensor Networks*. Technical Report cs.NI/0609047. CERN.

A. Amar, Yiyin Wang, and G. Leus. 2010. Extending the Classical Multidimensional Scaling Algorithm Given Partial Pairwise Distance Measurements. *IEEE Signal Processing Letters* 17, 5 (May 2010), 473–476.

Jerald S. Ault. 2007. *Biology and Management of the World Tarpon and Bonefish Fisheries*. CRC Press.

Song Bai, Xiang Bai, Longin Jan Latecki, and Qi Tian. 2016. (2016).

Yoshua Bengio, Jean-Francois Paiement, and Pascal Vincent. 2003. Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. In *In Advances in Neural Information Processing Systems*. MIT Press, 177–184.

Pratik Biswas, Tzu-Chen Lian, Ta-Chung Wang, and Yinyu Ye. 2006. Semidefinite Programming Based Algorithms for Sensor Network Localization. *ACM Trans. Sen. Netw.* 2, 2 (May 2006), 188–220.

Ingwer Borg, Patrick JF Groenen, and Patrick Mair. 2012. *Applied multidimensional scaling*. Springer Science & Business Media.

I. Borg and P. Gronen. 2005. *Modern Multidimensional Scaling: theory and applications* (2nd ed.). Springer.

Giuseppe Bove. 2006. Approaches to Asymmetric Multidimensional Scaling with External Information. In *Data Analysis, Classification and the Forward Search*. Springer Berlin Heidelberg, 69–76.

J. Roger Bray and J. T. Curtis. 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* (1957).

Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. 2006a. Efficient computation of isometry-invariant distances between surfaces. *SIAM Journal on Scientific Computing* 28 (2006), 2–6.

Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. 2006b. Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences of the United States of America* 103, 5 (2006), 1168–1172.

Er M. Bronstein, Michael M. Bronstein, Alfred M. Bruckstein, and Ron Kimmel. 2007. Analysis of two-dimensional non-rigid shapes. *IJCV* (2007).

Andreas Buja, Deborah F Swayne, Michael L Littman, Nathaniel Dean, Heike Hofmann, and Lisha Chen. 2008. Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics* 17, 2 (2008), 444–472.

Christopher J. C. Burges. 2005. Geometric Methods for Feature Extraction and Dimensional Reduction. *Data Mining and Knowledge Discovery Handbook* (2005).

Francis Caillez and Pascale Kuntz. 1996. A contribution to the study of the metric and Euclidean structures of dissimilarities. *Psychometrika* 61, 2 (1996), 241–253.

E. Cambria, Yangqiu Song, Haixun Wang, and N. Howard. 2014. Semantic Multidimensional Scaling for Open-Domain Sentiment Analysis. *IEEE Intelligent Systems* 29, 2 (Mar 2014), 44–51.

J. D. Carroll and J. J. Chan. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika* 35, 3 (1970), 283–319.

M. L. Carter, S. S. Valenti, and R. F. Goldberg. 2001. Perception of sports photographs: a multidimensional scaling analysis. *National center for biotechnology information* 92, 3 (2001), 643–652.

Vijay K. Chaurasiya, Neeraj Jain, and G.C. Nandi. 2014. A novel distance estimation approach for 3D localization in wireless sensor network using multi dimensional scaling. *Information Fusion* 15, 0 (2014), 5 – 18.

Q. Cheng. 2006. Spatial and Spatially Weighted Principal Component Analysis for Images Processing. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*. 972–975. DOI:http://dx.doi.org/10.1109/IGARSS.2006.250

Zeng-Guang Hou Chengan Guo and Zhigang Zeng. 2013. *Advances in Neural Networks- International Symposium on Neural Networks*. Springer.

Naohito Chino. 1978. A graphical technique for representing the asymmetric relationships between N objects. *Behaviormetrika* 5, 5 (1978), 23–44.

N Chino. 1980. A unified geometrical interpretation of the MDS techniques for the analysis of asymmetry and related techniques. In *Paper presented at the symposium on" Asymmetric multidimensional scaling" at the Spring Meeting of the Psychometric Society, Iowa, USA*.

Michael Collins, Sanjoy Dasgupta, and Robert E. Schapire. 2001. A generalization of principal component analysis to the exponential family. In *Advances in Neural Information Processing Systems*.

AG Constantine and JC Gower. 1978. Graphical representation of asymmetric matrices. *Applied Statistics* (1978), 297–304.

C. H. Coombs. 1964. *A theory of data*. New York: Jhon Wiley and Sons, lnc.

Jose A. Costa, Neal Patwari, and Alfred O. Hero, III. 2006. Distributed Weighted-multidimensional Scaling for Node Localization in Sensor Networks. *ACM Trans. Sen. Netw.* 2, 1 (Feb 2006), 39–64.

Michael AA Cox and Cox Trevor F. 2008. *Multidimensional scaling*. Springer. 315–347 pages.

T. Cox and M. Cox. 2001. *Multidimensional Scaling* (2nd ed.). C. Hall and B. Raton.

M. L. Davidson and C. L. Skay. 1991. Multidimensional Scaling and Factor Models of Test and Item Responses. *Psychological Bulletin* 110, 3 (1991), 551–556.

C. De Marziani, J. Urena, A. Hernandez, J.J. Garcia, A. Jimenez, M.C. Perez, F. Alvarez, and J.M. Villadangos. 2009. Relative localization and mapping combining multidimensional scaling and Levenberg-Marquardt optimization. In *IEEE International Symposium on Intelligent Signal Processing*. 43–47.

C. Eckart and G. Young. 1936. THE APPROXIMATION OF ONE MATRIX BY ANOTHER OF LOWER RANK. *Psycometrika* 1, 3 (Sep 1936), 211–218.

D. W. Eggert, A. Lorusso, and R. B. Fisher. 1997. Estimating 3-D Rigid Body Transformations: A Comparison of Four Major Algorithms. *Mach. Vision Appl.* 9, 5-6 (Mar 1997), 272–290.

A. Elad and R. Kimmel. 2002. *Geometric Methods in Bio-Medical Image Processing*. Springer-Verlag Berlin Heidelberg.

Donald A Falk, Margaret A Palmer, Joy B Zedler, and others. 2006. *Foundations of restoration ecology*. Island Press Washington, DC.

Peter Filzmoser and Karel Hron. Outlier Detection for Compositional Data Using Robust Methods. *Mathematical Geosciences* 40, 3 (????), 223–248.

Robert W. Floyd. 1962. Algorithm 97: Shortest Path. *Commun. ACM* 5 (June 1962), 345.

Stephen L France and J Douglas Carroll. 2011. Two-way multidimensional scaling: A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41, 5 (2011), 644–661.

Emden R Gansner, Yehuda Koren, and Stephen North. 2004. Graph drawing by stress majorization. In *International Symposium on Graph Drawing*. Springer, 239–250.

Hugh G. Gauch. 1982. *Multivariate Analysis in Community Ecology*. Cambridge University Press.

P. Q. Gerry and M. J. Keough. 2004. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press.

Ali Ghodsi. 2006. Dimensionality reduction a short tutorial. *Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada* 37 (2006), 38.

G.C. Gilmore, H. Hersh, A. Caramazza, and J. Griffin. 1979. Multidimensional letter similarity derived from recognition errors. *Perception and Psychophysics* 25, 5 (1979), 425–431.

J.C. Gower and P. Legendre. 1986. Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* 3, 1 (1986), 5–48.

J. C. Gower. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 3 (Dec 1966), 325–338.

J. C. Gower. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27, 4 (1971).

John C Gower. 1977. The analysis of asymmetry and orthogonality. *Recent developments in statistics* (1977), 109–123.

M. Greenacre. 2005. Weighted Metric Multidimensional Scaling. In *New Developments in Classification and Data Analysis*. Springer Berlin Heidelberg, 141–149.

Michael Greenacre and Raul Primicerio. 2013. *Multivariate Analysis of Ecological Data*. Rubes Editorial.

P.J.F. Groenen, W.J. Heiser, and J.J. Meulman. 1997. Global Optimization in Least Squares Multidimensional Scaling by Distance. *Journal of Classification* 16 (1997), 225–254.

Louis Guttman. 1968a. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika* 33, 4 (1968), 469–506.

L. Guttman. 1968b. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika* 33, 4 (1968), 469–506.

Edwin Hancock and Mario Vento. 2003. Graph Based Representation in Pattern Recognition. In *4th IAPR International Workshop, Proceedings*.

Charles D Hansen, Min Chen, Christopher R Johnson, Arie E Kaufman, and Hans Hagen. 2014. *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*. Springer.

Mehrdad Honarkhah and Jef Caers. 2010. Stochastic Simulation of Patterns Using Distance-Based Pattern Modeling. *Mathematical Geosciences* 42, 5 (2010), 487–517.

Michael C. Hout, Megan H. Papesh, and Stephen D. Goldinger. 2013. (2013).

Alan Julian Izenman. 2008. *Modern multivariate statistical techniques*. Vol. 1. Springer.

Franc Janzekovic and Tone Novak. 2012. PCA A Powerful Method for Analyze Ecological Niches. *Principal Component Analysis - Multidisciplinary Applications* (2012).

Natalia Jaworska and Angelina Chupetlovska-Anastasova. 2009. A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in quantitative methods for psychology* 5, 1 (2009), 1–10.

X. Ji and H. Zha. 2004. Sensor positioning in wireless ad-hoc sensor networks with multidimensional scaling. In *Proc. IEEE Int. Conf. on Computer Commun. (INFOCOM)*, Vol. 4. 2652–2661.

F. Jourdan and G. Melangon. 2004. Multiscale hybrid MDS. In *Proceedings of international Conference on Information Visualisation*. 388–393.

Li jun Peng and Wen wei Li. 2014. The improvement of 3D wireless sensor network nodes localization. In *The 26th Chinese Conference on Control and Decision*. 4873–4878.

Tomihisa Kamada and Satoru Kawai. 1989. An algorithm for drawing general undirected graphs. *Information processing letters* 31, 1 (1989), 7–15.

Hansjrg Klock and Joachim M. Buhmann. 2000. Data visualization by multidimensional scaling: a deterministic annealing approach. *Pattern Recognition* 33, 4 (2000), 651–669.

Joseph B. Kruskal. 1956. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. In *Proc. American Mathematical Society*. 48–50.

J. B. Kruskal. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1 (1964), 1–27.

Joseph B Kruskal. 1964b. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1 (1964), 1–27.

G. Latsoudas and N.D. Sidiropoulos. 2005. A two-stage fastmap-MDS approach for node localization in sensor networks. In *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*. 64–67.

J. Lawrence, S. Arietta, M. Kazhdan, D. Lepage, and C. O'Hagan. 2011. A User-Assisted Approach to Visualizing Multidimensional Images. 17, 10 (Oct 2011), 1487–1498. DOI:http://dx.doi.org/10.1109/TVCG.2010.229

J. D. Leeuw. 2005. *Shepard Diagram*. John Wiley and Sons, Ltd.

Jan De Leeuw, In J. R. Barra, F. Brodeau, G. Romier, and B. Van Cutsem. 1977. Applications of convex analysis to multidimensional scaling. In *Recent Developments in Statistics*. 133–146.

J. D. Leeuw and W. Heiser. 1980. *Multidimensional scaling with restrictions on the configuration*. Amsterdam: North-Holland.

Boaz Lerner, Hugo Guterman, Mayer Aladjem, Its'hak Dinsteint, and Yitzhak Romem. 1998. On pattern classification with Sammon's nonlinear mapping an experimental study. *Pattern Recognition* 31, 4 (1998), 371 – 381.

Olivier Lezoray and Leo Grady. 2012. *Image Processing and Analysis with Graph Theory and Pratice*. Taylor and Francis group.

S. Li and D. Zhang. 2007. A Novel Manifold Learning Algorithm for Localization Estimation in Wireless Sensor Networks. *IEICE TRANSACTIONS on Communications* E90-B, 12 (Dec 2007), 659–668.

Ik Soo Lim, P. de Heras Ciechomski, S. Sarni, and D. Thalmann. 2003. Planar arrangement of high-dimensional biomedical data sets by isomap coordinates. In *Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium*. 50–55. DOI:http://dx.doi.org/10.1109/CBMS.2003.1212766

James C Lingoes and Louis Guttman. 1973. *The Guttman-Lingoes nonmetric program series*. Mathesis Press.

A. M. Lopes and J. A. T. Machado. 2014. Analysis of temperature time-series: Embedding dynamics into the MDS method. *Communications in Nonlinear Science and Numerical Simulation* 19, 4 (2014), 851–871.

A. M. Lopes and J. A. T. Machado. 2016. Multidimensional scaling analysis of virus diseases. *Computer Methods and Programs in Biomedicine* 131 (2016), 97–110.

A. M. Lopes, J. A. T. Machado, C. M. A. Pinto, and A. M. S. F. Galhano. 2013. Fractional dynamics and MDS visualization of earthquake phenomena. *Computer and mathematics with applications* 66, 5 (2013), 647–658.

A. M. Lopes, J. A. T. Machado, C. M. A. Pinto, and A. M. S. F. Galhano. 2014. Multidimensional scaling visualization of earthquake phenomena. *Journal of Seismology* 18, 1 (2014), 163–179.

D. Macagnano and G.T.F. de Abreu. 2013. Algebraic Approach for Robust Localization with Heterogeneous Information. 12, 10 (Oct 2013), 5334–5345.

J. A. T. Machado and A. M. Lopes. 2013. Analysis and Visualization of Seismic Data Using Mutual Information. *Entropy* 15, 9 (2013), 3892–3909.

J. A. T. Machado and A. M. Lopes. 2014. Analysis of Forest Fires by means of Pseudo Phase Plane and Multidimensional Scaling Methods. *Mathematical Problems in Engineering* (2014).

J. A. T. Machado and A. M. Lopes. 2017. Multidimensional scaling analysis of soccer dynamics. *Applied Mathematical modeling* 45 (2017), 642–652.

K. V. Mardia. 1979. *Multivariate Analysis*. London: Academic Press.

R. Mathar. 1990. Multidimensional scaling with constraints on the configuration. *Journal of Multivariate Analysis* 33, 2 (1990), 151–156.

V. E. McGee. 1966. THE MULTIDIMENSIONAL ANALYSIS OF ELASTIC DISTANCES. *Brit. J. Math. Statist. Psych.* 19, 2 (1966), 181–196.

J. J. Meulman. 1992. The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika* 57, 4 (1992), 539–565.

David Moore, John Leonard, Daniela Rus, and Seth Teller. 2004. Robust Distributed Network Localization with Noisy Range Measurements. In *Proceedings of the International Conference on Embedded Networked Sensor Systems*. 50–61.

Antoine Naud and Wlodzislaw Duch. 2000. Interactive Data Exploration Using Mds Mapping. In *5th Conf. on Neural Networks and Soft Computing*. 255–260.

I. Newton. 1704. *Opticks: a Treatise of the Reflexions, Refractions, Inflexions and Colours of Light. Also Two Treatises of the Species and Magnitude of Curvilinear Figures*. London: Smith and Walford.

Vo Dinh Minh Nhat, Nhat Vo, S. Challa, and Sungyoung Lee. 2008. Nonmetric MDS for sensor localization. In *3rd International Symposium on Wireless Pervasive Computing, (ISWPC)*. 396–400.

A Okada and T Imaizumi. 1984. Geometric models for asymmetric similarity data. *Behaviormetrika* 21 (1984), 81–96.

Akinori Okada and Tadashi Imaizumi. 1987. Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika* 14, 21 (1987), 81–96.

Akinori Okada and Tadashi Imaizumi. 1997. Asymmetric multidimensional scaling of two-mode three-way proximities. *Journal of Classification* 14, 2 (1997), 195–224.

N. Patwari, A. Hero, III, A. .O. Perkins, N. Correal, and R. J. Odea. 2003. Relative location estimation in wireless sensor networks. 51 (Aug. 2003), 2137–2148.

Piotr Pawliczek and Witold Dzwinel. 2013. Interactive Data Mining by Using Multidimensional Scaling. *Procedia Computer Science* 18 (2013), 40 – 49. DOI:http://dx.doi.org/10.1016/j.procs.2013.05.167

N. Priyantha, H. Balakrishnan, E. Demaine, and S. Teller. 2003. Anchor Free distributed localization in sensor networks. In *Technical report MIT-LCS-TR-892.MIT, Cambridge, MA, USA*.

Ryszard Tadeusiewicz Lotfi A. Zadeh R. Leszek, Rafal Scherer and Jacek M. Zurada. 2010. *Artificial Intelligence and Soft Computing ICAISC*. Cambridge University Press.

J. L. V. Raalte, B. W. Brewer, D. E. Linder, and N. Delange. 1990. Perceptions of Sport-Oriented Professionals: A Multidimensional Scaling Analysis. *Human Kinectic Journals* 4, 3 (1990), 228–234.

J. O. Ramsay. 1982. Some Statistical Approaches to Multidimensional Scaling Data. *Journal of the Royal Statistical Society. Series A* 145, 3 (1982), 285–312.

Shingo Saburi and Naohito Chino. 2008. A maximum likelihood method for an asymmetric MDS model. *Computational Statistics & Data Analysis* 52, 10 (2008), 4673–4684.

N. Saeed and Haewoon Nam. 2014. MDS-LM for Wireless Sensor Networks Localization. In *IEEE 79th Vehicular Technology Conference*. 1–6.

N. Saeed and H. Nam. 2015. Robust Multidimensional Scaling for Cognitive Radio Network Localization. *IEEE Transactions on Vehicular Technology* 64, 9 (Sept 2015), 4056–4062.

N. Saeed and H. Nam. 2016. Cluster Based Multidimensional Scaling for Irregular Cognitive Radio Networks Localization. *IEEE Transactions on Signal Processing* 64, 10 (Feb 2016), 2649–2659.

Y. Sahillioglu and Y. Yemez. 2011. Coarse-to-fine combinatorial matching for dense isometric shape correspondence. *Computer Graphics Forum* 30, 5 (2011), 1461–1470.

T Saito. 1983. A method of multidimensional scaling to obtain a sphere configuration. *Hokkaido Behavioral Science Report, Series-M* 4 (1983).

Takayuki Saito. 1986. Multidimensional scaling to explore complex aspects in dissimilarity judgment. *Behaviormetrika* 13, 20 (1986), 35–62.

Takayuki Saito. 1988. Cluster analysis based on nonparametric tests. *Classification and related methods of data analysis.(Bock, HH ed., North Holland)* (1988), 257–266.

J.W. Sammon. 1969. A Nonlinear Mapping for Data Structure Analysis. 18, 5 (May 1969), 401–409.

Y Sato. 1989. On distance representation using asymmetric metric function. In *17th annual meeting of the Behaviormetric Society of Japan, Okayama, Japan (in Japanese)*.

John A. Schinka and Wayne F. Velicer. 2013. *Handbook of Psychology*. Vol. 2. John Willey & Sons, Ltd.

Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz. 2003. Localization from mere connectivity. In *Proc. ACM Mobihoc, Annapolis, MD*. 201–212.

Q.Q. Shi, H. Huo, T. Fang, and R.D. Li. 2007. Using steepest descent method to refine the node positions in wireless sensor networks. In *IET Conference on Wireless, Mobile and Sensor Networks*. 1055–1058.

Minhan Shon, Wook Choi, and Hyunseung Choo. 2010. A Cluster-Based MDS Scheme for Range-Free Localization in Wireless Sensor Networks. In *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. 42–47.

N. Solaro. 2010. Sensitivity Analysis and Robust Approach in Multidimensional Scaling: An Evaluation of Customer Satisfaction. *Quality Technology and Quantitative Management* 7, 2 (2010).

N. Solaro. 2011. *Multidimensional Scaling*. John Wiley & Sons, Ltd. 357–390 pages.

Biljana Stojkoska, Danco Davcev, and Andrea Kulakov. 2008. Cluster-based MDS Algorithm for Nodes Localization in Wireless Sensor Networks with Irregular Topologies. In *Proceedings of the 5th International Conference on Soft Computing As Transdisciplinary Science and Technology*. 384–389.

B. R. Stojkoska. 2013. Improved MDS-MAP Algorithm for Nodes Localization in 3D Wireless Sensor Networks. In *Proceedings of the Annual South-East European Doctoral Student Conference*. 340–348.

B. R. Stojkoska. 2014. Nodes Localization in 3D Wireless Sensor Networks Based on Multidimensional Scaling Algorithm. *International Scholarly Research Notices* 2014, 2014 (Jun 2014), 1–10.

Yoshio Takane. 1981. Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika* 46, 1 (1981), 9–28.

Y. Takane, F. W. Young, and J. D. Leeuw. 1977. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* 42, 1 (1977), 7–67.

P. Tan, M. Steinbach, and V. Kumar. 2005. *Introduction to Data Mining*. Addison-Wesley.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. 2000. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290, 5500 (2000), 2319–2323.

Ersebet Merenyi Thomas Villmann and Udo Seiffert. 2008. Machine Learning Approaches and Pattern Recognition for Spectral Data. In *European Symposium on Artificial Neural Networks- Advances in Computational Intelligence and Learning Bruges (ESANN 2008)*.

Waldo Tobler. 1975. Spatial interaction patterns. (1975).

W. S. Torgerson. 1952a. Multidimensional scaling: I. Theory and method. *Psycometrika* 17, 4 (Dec 1952), 401–419.

Warren S Torgerson. 1952b. Multidimensional scaling: I. Theory and method. *Psychometrika* 17, 4 (1952), 401–419.

Amos Tversky. 1977. Features of similarity. *Psychological review* 84, 4 (1977), 327.

Amos Tversky and Itamar Gati. 1978. Studies of similarity. *Cognition and categorization* 1, 1978 (1978), 79–98.

Jengnan Tzeng, Henry HS Lu, and Wen-Hsiung Li. Multidimensional scaling for large genomic data sets. (????).

V. Vivekanandan and V.W.S. Wong. 2006. Ordinal MDS-Based Localization for Wireless Sensor Networks. In *IEEE 64th Vehicular Technology Conference*. 1–5.

Jörg A. Walter and Helge Ritter. 2002. On Interactive Visualization of High-dimensional Data Using the Hyperbolic Plane. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 123–132.

Andrew R Webb. 2003. *Statistical pattern recognition*. John Wiley & Sons.

David G Weeks and PM Bentler. 1982. Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika* 47, 2 (1982), 201–208.

Mingzhu Wei, R. Aragues, C. Sagues, and G.C. Calafiore. 2015. Noisy Range Network Localization Based on Distributed Multidimensional Scaling. 15, 3 (Mar 2015), 1872–1883.

Martin Granzow Werner Dubitzky and Daniel P. Berrar. 2007. *Fundamentals of Data Mining in Genomics and Proteomics*. Springer.

Wei Xin, Wang Quandi, Yang Chenghe, and Yang Yongming. 2009. A three-dimensional node localization algorithm of high-precision in wireless sensor network. In *IEEE International Conference on Industrial Technology, (ICIT)*. 1–5.

Hiroyuki Yamamoto, Tamaki Fujimori, Hajime Sato, Gen Ishikawa, Kenjiro Kami, and Yoshiaki Ohashi. 2014. Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis. *BMC Bioinformatics* 15, 1 (2014), 1–9. DOI:http://dx.doi.org/10.1186/1471-2105-15-51

Ming-Hsuan Yang. 2002. Extended Isomap for classification. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, Vol. 3. 615–618 vol.3. DOI:http://dx.doi.org/10.1109/ICPR.2002.1048014

Y.H. Yin, J.Y. Xie, Li Da Xu, and Hao Chen. 2012. Imaginal Thinking-Based Human-Machine Design Methodology for the Configuration of Reconfigurable Machine Tools. 8, 3 (Aug 2012), 659–668.

ForrestW. Young, Yoshio Takane, and Rostyslaw Lewyckyj. 1978. ALSCAL: A nonmetric multidimensional scaling program with several individual-differences options. *Behavior Research Methods and Instrumentation* 10, 3 (1978), 451–453.

Forrest W Young. 1975. An asymmetric Euclidean model for multi-process asymmetric data. In *Paper presented at US-Japan Seminar on MDS, San Diego*.

Forrest W. Young. 2013. *Multidimensional Scaling History, THeory and Applications*. LEA.

G. Young and A. S. Householder. 1938. Discussion of a set of points in terms of their mutual distances. *Psycometrika* 3, 1 (Mar 1938), 19–22.

S.D. Young, B.D. Adelstein, and S.R. Ellis. 2007. Calculus of Nonrigid Surfaces for Geometry and Texture Manipulation. *Visualization and Computer Graphics, IEEE Transactions on* 13, 5 (Sept 2007), 902–913.

Gwo-Jong Yu and Shao-Chun Wang. 2008. A Hierarchical MDS-Based Localization Algorithm for Wireless Sensor Networks. In *International Conference on Advanced Information Networking and Applications*. 748–754.

Gang Zhang and Liang-Lun Cheng. 2010. Semi-supervised Classification with Metric Learning. In *Second WRI Global Congress on Intelligent Systems (GCIS)*, Vol. 3. 123–126.

Jin Zhang. 2007. *Visualization for information Retrieval*. Cambridge University Press.

S. Zhu and Z. Ding. 2012. Bridging gap between multi-dimensional scalingbased and optimum network localisation via efficient refinement. *IET Signal Processing* 6, 2 (April 2012), 132–142.

Alain F. Zuur, Elena N. Leno, and Chris S. Elphick. 2009. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* (2009).

Received April 2016; revised October 2016; accepted January 2018