Multidimensional Scaling with Majorization

A Anish
Dept. of Mathematics and Computing
ES20BTECH11001

Thupten Dukpa
Dept. of Engineering Science
ES20BTECH11029

***Abstract*—In this paper we present the methodology of multidimensional scaling problems (MDS) solved by means of the majorization algorithm. The objective function to be minimized is known as stress and functions which majorize stress are elaborated. This strategy to solve MDS problems is called SMACOF and it is implemented in *Python*. We extend the basic SMACOF theory in terms of configuration constraints.**

***Index Terms*—SMACOF, multidimensional scaling, majorization, *Python*.**

I. Introduction

From a general point of view, multidimensional scaling (MDS) is a set of methods for discovering "hidden" structures in multidimensional data. Based on a proximity matrix derived from variables measured on objects as input entity, these distances are mapped on a lower dimensional (typically two or three dimensions) spatial representation. A classical example concerns airline distances between US cities in miles as symmetric input matrix. Applying MDS, it results in a two-dimensional graphical representation reflecting the US map. Depending on the nature of the original data various proximity/dissimilarity measures can be taken into account.

Typical application areas for MDS are, among others, social and behavioral sciences, marketing, biometrics, and ecology.

For each MDS version we provide metric and non-metric variants. Non-metric MDS will be described in a separate section since, within each majorization iteration, it includes an additional optimization step. However, for both approaches, the particular objective function (or loss function) we use in this paper is a sum of squares, commonly called stress. We use majorization to minimize stress and this MDS solving strategy is known as SMACOF (Scaling by Majorizing a Complicated Function).

Furthermore we will provide extension of the basic SMACOF approach in terms of constraints on the configuration.

II. Majorization Theory

Before describing details about MDS and SMACOF we give a brief overview on the general concept of majorization which optimizes a particular objective function; in our application referred to as stress. More details about the particular stress functions and their surrogates for various SMACOF extensions will be elaborated below.

In a strict sense, majorization is not an algorithm but rather a prescription for constructing optimization algorithms. The principle of majorization is to construct a surrogate function which majorizes a particular objective function.

From a formal point of view majorization requires the following definitions. Let us assume we have a function $f(x)$ to be minimized. Finding an analytical solution for complicated $f(x)$ can be rather cumbersome. Thus, the majorization principle suggests to find a simpler, more manageable surrogate function $g(x, y)$ which majorizes $f(x)$, i.e., for all x

$$g(x, y) \geq f(x) \tag{1}$$

where y is some fixed value called the supporting point. The surrogate function should touch the surface at y, i.e., $f(y) = g(y, y)$, which, at the minimizer x^* of $g(x, y)$ over x, leads to the inequality chain

$$f(x^*) \leq g(x^*, y) \leq g(y, y) = f(y) \tag{2}$$

called the $sandwich\ inequality$.

Majorization is an iterative procedure which consists of the following steps:

1) Choose initial starting value $y := y0$.
2) Find update $x(t)$ such that $g(x(t), y) \leq g(y, y)$.
3) Stop if $f(y) - f(x(t)) < \epsilon$, else $y := x(t)$ and proceed with step 2.

This procedure can be extended to multidimensional spaces and as long as the sandwich inequality in Equation 2 holds, it can be used to minimize the corresponding objective function. In MDS the objective function called stress is a multivariate function of the distances between objects. We will use majorization for stress minimization for various SMACOF variants as described in the following sections.

III. Basic SMACOF methodology

A. Simple SMACOF for symmetric dissimilarity matrices

MDS input data are typically a n x n matrix $\triangle$ of dissimilarities based on observed data. $\triangle$ is symmetric, non-negative, and hollow (i.e., has zero diagonal). The problem we solve is to locate i, j = 1, . . . , n points in low-dimensional Euclidean space in such a way that the distances between the points approximate the given dissimilarities δ_{ij}. Thus we want to find an n x p matrix X such that $d_{ij}(X) \approx \delta_{ij}$, where

$$d_{ij}(X) = \sqrt{\sum_{s=1}^{p}(x_{is} - x_{js})^2} = ||x_i - x_j||_2 \tag{3}$$

The index s = 1, . . . , p denotes the number of dimensions in the Euclidean space. The elements of X are called $configurations$ of the objects. Thus, each object is scaled in a p-dimensional space such that the distances between the points in the space match as well as possible the observed dissimilarities. By representing the results graphically, the configurations represent the coordinates in the configuration plot.

Now we make the optimization problem more precise by defining $stress$ $\sigma(X)$ as

$$\sigma(X) = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij}(X))^2 \tag{4}$$

Here, W is a known n x n matrix of weights w_{ij}, also assumed to be symmetric, non-negative, and hollow. We assume, without loss of generality, that

$$\sum_{i<j} w_{ij}\delta_{ij}^2 = \frac{n(n-1)}{2} \tag{5}$$

Now from (4), stress can be decomposed as

$$\begin{aligned}\sigma(X) &= \sum_{i<j} w_{ij}\delta_{ij}^2 + \sum_{i<j} w_{ij}d_{ij}^2(X) - 2\sum_{i<j} w_{ij}\delta_{ij}d_{ij}(X)\\ &= \eta_\delta^2 + \eta^2(X) - 2\rho(X)\end{aligned}$$

From restriction (5) it follows that the first component η_δ^2 = $n(n$-1)/2. The second component $\eta^2(X)$ is a weighted sum of the squared distances $d_{ij}^2(X)$, and thus a convex quadratic. The third one, i.e., -2$\rho(X)$, is the negative of a weighted sum of the $d_{ij}(X)$, and is consequently concave.

The third component is the crucial term for majorization. Let us define the matrix A_{ij} = $(e_i - e_j)(e_i - e_j)^T$ whose elements equal 1 at a_{ii} & a_{jj}, -1 at a_{ij} & a_{ji}, and 0 elsewhere. Furthermore, we define

$$V = \sum_{i<j} w_{ij}A_{ij} \tag{6}$$

as the weighted sum of row and column centered matrices A_{ij}. Hence, we can rewrite

$$\eta^2(X) = \textbf{tr}\ (X^TVX) \tag{7}$$

For a similar representation of $\rho(X)$ we define the matrix

$$B(X) = \sum_{i<j} w_{ij}s_{ij}(X)A_{ij} \tag{8}$$

where,

$$s_{ij}(X) = \begin{cases}\delta_{ij}/d_{ij}(X) & \text{if } d_{ij}(X) > 0\\ 0 & \text{if } d_{ij}(X) > 0\end{cases}$$

Using $B(X)$ we can rewrite $\rho(X)$ as

$$\rho(X) = \textbf{tr}\ (X^TB(X)X) \tag{9}$$

The stress equation becomes

$$\sigma(X) = n(n-1)/2 + \textbf{tr}\ (X^TVX) - 2\textbf{tr}\ (X^TB(X)X) \tag{10}$$

At this point it is straightforward to find the majorizing function of $\sigma(X)$. Let us denote the supporting point by Y which, in the case of MDS, is a n x p matrix of configurations. Similar to (8) we define

$$B(Y) = \sum_{i<j} w_{ij}s_{ij}(Y)A_{ij} \tag{11}$$

$$s_{ij}(Y) = \begin{cases}\delta_{ij}/d_{ij}(Y) & \text{if } d_{ij}(Y) > 0\\ 0 & \text{if } d_{ij}(Y) > 0\end{cases}$$

The Cauchy-Schwartz inequality implies that for all pairs of configurations X and Y, we have $\rho(X) \geq$ **tr** $(X^TB(Y)Y)$.
Proof:
The Cauchy-Schwartz inequality statement:

$$\begin{aligned}(x_i - x_j).(y_i - y_j) &\leq ||x_i - x_j||_2||y_i - y_j||_2\\ ||x_i - x_j||_2 &\geq (x_i - x_j).\frac{(y_i - y_j)}{||y_i - y_j||_2}\end{aligned}$$

We know,

$$\begin{aligned}\rho(X) &= \textbf{tr}(X^TB(X)X)\\ \textbf{tr}(X^TB(X)X)) &= \sum_{i<j} w_{ij}\delta_{ij}d_{ij}(X)\\ &= \sum_{i<j} w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}d_{ij}^2(X)\\ &= \sum_{i<j} w_{ij}s_{ij}(X)(\|x_i - x_j\|)^2\\ &= \sum_{i<j} w_{ij}s_{ij}(X)(x_i - x_j).(x_i - x_j)^T\\ \sum_{i<j} w_{ij}\delta_{ij}d_{ij}(X) &\geq \sum_{i<j} w_{ij}\delta_{ij}\left((x_i - x_j).\frac{(y_i - y_j)}{||y_i - y_j||}\right)\\ &\geq \sum_{i<j} w_{ij}\frac{\delta_{ij}}{||y_i - y_j||}\left((x_i - x_j).(y_i - y_j)\right)\\ &\geq \sum_{i<j} w_{ij}\frac{\delta_{ij}}{d_{ij}(Y)}\left((x_i - x_j).(y_i - y_j)\right)\\ &\geq \sum_{i<j} w_{ij}s_{ij}(Y)\left((x_i - x_j).(y_i - y_j)\right)\\ \rho(X) &\geq \textbf{tr}(X^TB(Y)Y)\end{aligned}$$

where Y= $\begin{bmatrix} y_1^T\\ y_2^T\\ |\\ |\end{bmatrix}$.

Thus we minorize the convex function $\rho(X)$ with a linear function. This gives us a majorization of stress

$$\begin{aligned}\sigma(X) &= n(n-1)/2 + \textbf{tr}\ (X^TVX) - 2\textbf{tr}\ (X^TB(X)X)\\ &\leq n(n-1)/2 + \textbf{tr}\ (X^TVX) - 2\textbf{tr}\ (X^TB(Y)Y) = \tau(X, Y)\end{aligned}$$

Obviously, $\tau(X, Y)$ is a (simple) quadratic function in X which majorizes stress. Finding its minimum analytically involves

$$\frac{\partial\tau(X, Y)}{\partial X} = 2VX - 2B(Y)Y = 0 \tag{12}$$

$$VX = B(Y)Y$$

Since majorization is an iterative procedure, in step t = 0 we set $Y := X^{(0)}$ where $X^{(0)}$ is a start configuration. Within each iteration t we compute $\overline{X}^{(t)}$ which, for simple SMACOF, gives us the update $X^{(t)}$. Now the stress $\sigma(X^{(t)})$ can be calculated and we stop iterating if $\sigma(X^{(t)}) - \sigma(X^{(t-1)}) < \epsilon$ or a certain iteration limit is reached. Majorization guarantees a series of non-increasing stress values with a linear convergence rate.

B. SMACOF with restrictions on the configurations

De Leeuw and Heiser introduced a SMACOF version with restrictions on the configuration matrix X which **Borg and Groenen** call *confirmatory* MDS with *external constraints*.The basic idea behind this approach is that the researcher has some substantive underlying theory regarding a decomposition of the dissimilarities. We start with the simplest restriction in terms of a linear combination, show the majorization solution. The linear restriction in its basic form is

$$X = ZC \tag{13}$$

where Z is a known predictor matrix of dimension n x q ($q \geq p$). The predictors can be numeric in terms of external covariates. C is a q x p matrix of regression weights to be estimated.

The optimization problem is basically the same as in the former section: We minimize stress as given in (4) with respect to C. The expressions for V and B as well as $\tau(X,Y)$ can be derived analogous to simple SMACOF and, correspondingly, the Guttman transform is $V\overline{X} = B(Y)Y$. It follows that Equation 11 can be rewritten as

$$\tau(X,Y) = n(n-1)/2 + \textbf{tr}\ (X^TVX) - 2\textbf{tr}\ (X^TB(Y)Y)$$
$$\tau(X,Y) = n(n-1)/2 + \textbf{tr}\ (X^TVX) - 2\textbf{tr}\ (X^TV\overline{X})$$

After Rearraging the terms we get,

$$\tau(X,Y) = n(n-1)/2 + \textbf{tr}\left((X-\overline{X})^TV(X-\overline{X})\right) - \textbf{tr}(\overline{X}^TV\overline{X})$$

where the second term denotes the *lack of confirmation fit* and becomes zero if no restrictions are imposed. Thus, in each iteration t of the majorization algorithm we first compute the Guttman transform $\overline{X}^{(t)}$ of our current best configuration, and then solve the configuration projection problem of the form

$$min\ \textbf{tr}\left((X-\overline{X}^{(t)})^TV(X-\overline{X}^{(t)})\right) \tag{14}$$

In other words, we project $\overline{X}^{(t)}$ on the manifold of constrained configurations. With linear restrictions this projection gives us the update

$$X^{(t)} = ZC^{(t)} = Z(Z^TVZ)^{-1}Z^TV\overline{X}^{(t)} \tag{15}$$

with $\sigma(X^{(t+1)}) < \sigma(X^{(t)})$.

Basically, the smacof package allows the user to implement arbitrary configuration restrictions by specifying a corresponding update function for X as given in (15).

IV. Code Results

We firstly tested MDS with majorization on the Swiss Roll dataset and this we what we got:

Fig. 1. Swiss roll Dataset

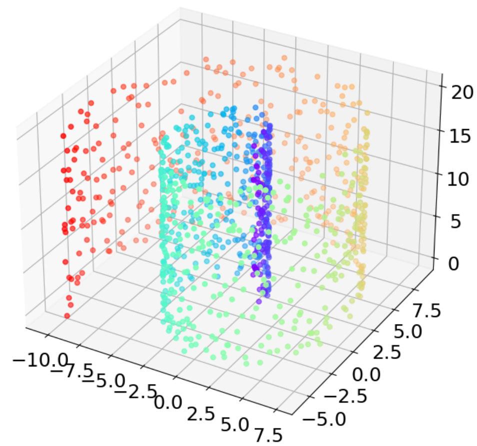
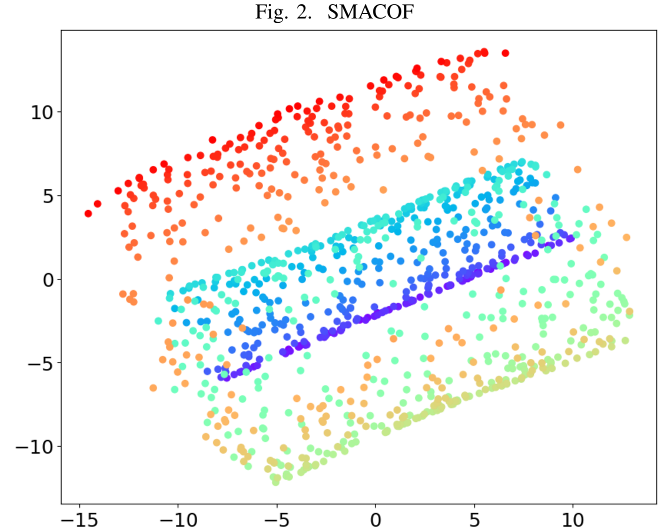


Fig. 2. SMACOF



Then we tested it out on the Arrow datset:

Fig. 3. Arrow Dataset

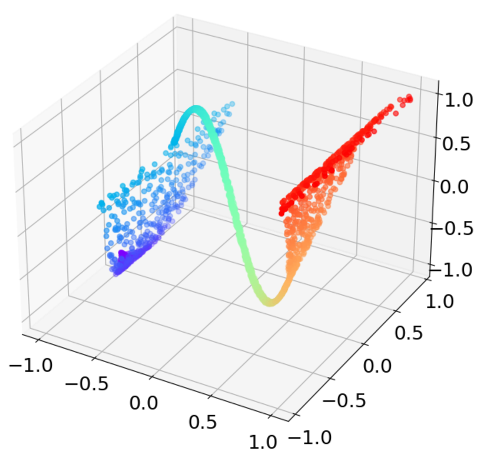
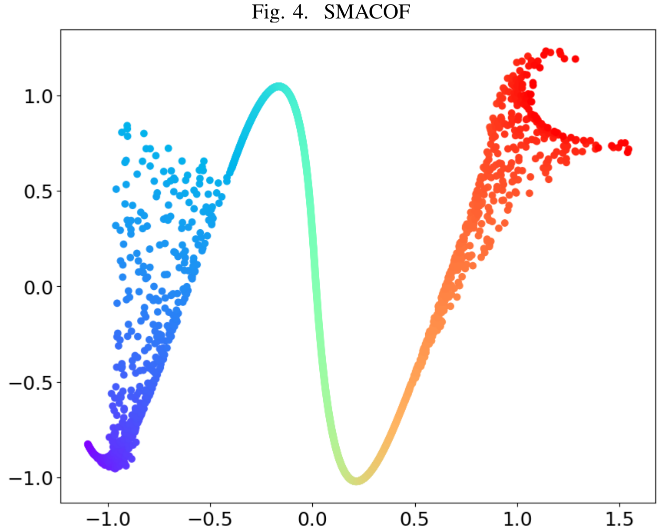


Fig. 4. SMACOF



Acknowledgment

We extend our heartfelt appreciation to Professor C.S Sastry for not only guiding us in choosing this compelling topic but also for imparting his extensive knowledge and teaching us the fundamental concepts that have enriched our understanding. His unwavering support has been pivotal in shaping of this project.

References

[1] Jan de Leeuw and Patrick Mair, ”Multidimensional Scaling Using Majorization: SMACOF in R”,August 2009.

[2] Patrick J. F. Groenen, Jan de Leeuw and Patrick Mair, ”Multidimensional Scaling in R: SMACOF”