# Visualization of High-Dimensional Data with Polar Coordinates

**Frank Rehm**
*German Aerospace Center, Germany*

**Frank Klawonn**
*University of Applied Sciences Braunschweig/Wolfenbuettel, Germany*

**Rudolf Kruse**
*University of Magdenburg, Germany*

## INTRODUCTION

Many applications in science and business such as signal analysis or costumer segmentation deal with large amounts of data which are usually high dimensional in the feature space. As a part of preprocessing and exploratory data analysis, visualization of the data helps to decide which kind of data mining method probably leads to good results or whether outliers or noisy data need to be treated before (Barnett & Lewis, 1994; Hawkins, 1980). Since the visual assessment of a feature space that has more than three dimensions is not possible, it becomes necessary to find an appropriate visualization scheme for such data sets.

Multidimensional scaling (MDS) is a family of methods that seek to present the important structure of the data in a reduced number of dimensions. Due to the approach of distance preservation that is followed by conventional MDS techniques, resource requirements regarding memory space and computation time are fairly high and prevent their application to large data sets. In this work we will present two methods that visualize high-dimensional data on the plane using a new approach. An algorithm will be presented that allows applying our method on larger data sets. We will also present some results on a benchmark data set.

## BACKGROUND

Multidimensional scaling provides low-dimensional visualization of high-dimensional feature vectors (Kruskal & Wish, 1978; Borg & Groenen, 1997). MDS is a method that estimates the coordinates of a set of objects in a feature space of specified (low) dimensionality that come from data trying to preserve the distances between pairs of objects. In the recent years much research has been done (Chalmers, 1996; Faloutsos & Lin, 1995; Morrison, Ross, & Chalmers, 2003; Williams & Munzner, 2004; Naud, 2006). Different ways of computing distances and various functions relating the distances to the actual data are commonly used. These distances are usually stored in a distance matrix. The estimation of the coordinates will be carried out under the constraint, that the error between the distance matrix of the data set and the distance matrix of the corresponding transformed data set will be minimized. Thus, different error measures to be minimized were proposed, i.e. the absolute error, the relative error or a combination of both. A commonly used error measure is the so-called Sammon's mapping (Sammon, 1969). To determine the transformed data set by means of minimizing the error a gradient descent method is used.

Many modifications of MDS are published so far, but high computational costs prevent their application to large data sets (Tenenbaum, de Silva, & Langford, 2000). Besides the quadratic need of memory, MDS, as described above is solved by an iterative method, expensive with respect to computation time, which is quadratic in the size of the data set. Furthermore, a completely new solution must be calculated, if a new object is added to the data set.

## MAIN FOCUS

With $\text{MDS}_{\text{polar}}$ and POLARMAP we present two approaches to find a two-dimensional projection of a p-dimensional data set $X$. Both methods try to find a representation in polar coordinates $Y = \{(l_1,\phi_1),\ldots,(l_n,\phi_n)\}$, where the length $l_k$ of the original vector $x_k$ is preserved and only the angle $\varphi_k$ has to be optimized. Thus, our solution is defined to be optimal if all angles between pairs of data objects in the projected data set Y coincide as good as possible with the angles in the original feature space *X*. As we will show later, it is possible to transform new data objects without extra costs.

## MDS$_{\text{polar}}$

A straight forward definition of an objective function to be minimized for this problem would be,

$$E=\sum_{k=2}^{n}\sum_{i=1}^{k-1}\left(|\phi_i-\phi_k|-\psi_{ik}\right)^2 \qquad (1)$$

where $\varphi_k$ is the angle of $y_k$, $\psi_{ik}$ is the positive angle between $x_i$ and $x_k$. The absolute value is chosen in equation (1) because the order of the minuends can have an influence on the sign of the resulting angle. The problem with this notation is that the functional *E* is not differentiable, exactly in those points we are interested in, namely, where the difference between angles $\varphi_i$ and $\varphi_k$becomes zero.

We propose an efficient method that enables us to compute an approximate solution for a minimum of the objective function (1) and related ones. In a first step we ignore the absolute value in (1) and consider

$$E=\sum_{k=2}^{n}\sum_{i=1}^{k-1}\left(\phi_i-\phi_k-\psi_{ik}\right)^2. \qquad (2)$$

When we simply minimize (2), the results will not be acceptable. Although the angle between $y_i$ and $y_k$ might perfectly match the angle $\psi_{ik}$, $\varphi_i - \varphi_k$ can either be $\psi_{ik}$ or $-\psi_{ik}$. Since we assume that $0 \leq \psi_{ik}$ holds, we always have $\left(|\phi_i-\phi_k|-\psi_{ik}\right)^2 \leq \left(\phi_i-\phi_k-\psi_{ik}\right)^2$. Therefore, finding a minimum of (2) means that this is an upper bound for the minimum of (1). Therefore, when we minimize (2) in order to actually minimize (1), we can take the freedom to choose whether we want the term $\phi_i-\phi_k$ or the term $\phi_k-\phi_i$ to appear in (2). Since

$$\begin{aligned}&\left(\phi_i-\phi_k-\psi_{ik}\right)^2\\&=\left(-\left(\phi_i-\phi_k-\psi_{ik}\right)\right)^2\\&=\left(\phi_k-\phi_i+\psi_{ik}\right)^2\end{aligned}$$

instead of exchanging the order of $\varphi_i$ and $\varphi_k$, we can choose the sign of $\Psi_{ik}$, leading to

$$E=\sum_{k=2}^{n}\sum_{i=1}^{k-1}\left(\phi_i-\phi_k-a_{ik}\psi_{ik}\right)^2 \qquad (3)$$

with $a_{ik} \in \{-1,1\}$.

In order to solve this optimization problem of equation (3) we take the partial derivatives of *E*, yielding

$$\frac{\partial E}{\phi_k}=-2\sum_{i=1}^{k-1}\left(\phi_i-\phi_k-a_{ik}\psi_{ik}\right). \qquad (4)$$

Thus, on the one hand, neglecting that we still have to choose $a_{ik}$, our solution is described by a system of linear equations which means its solution can be calculated directly without the need of any iteration procedure. On the other hand, as described above, we have to handle the problem of determining the sign of the $\Psi_{ik}$ in the form of the $a_{ik}$-values. To fulfill the necessary condition for a minimum we set equation (4) equal to zero and solve for the $\varphi_k$-values, which leads to

$$\phi_k=\frac{\sum_{i=1}^{k-1}\left(\phi_i-a_{ik}\psi_{ik}\right)}{k-1}. \qquad (5)$$

Since we only want to preserve the angles between data vectors, it is obvious that any solution will be invariant with respect to rotation of the data set. Due to the representation in polar coordinates it is necessary to apply a preprocessing step in form of a translation that makes all components of data vectors non-negative. Reasons for that and further details are given in (Rehm, Klawonn, & Kruse, 2005).

## A Greedy Algorithm for the Approximation of MDS$_{\text{polar}}$

As mentioned above, this solution describes a system of linear equations. Since the desired transformation is rotation invariant $\varphi_1$ can be set to any value, i.e. $\varphi_1 = 0$. By means of a greedy algorithm we choose $a_{ik} \in \{-1,1\}$ such that for the resulting $\varphi_k$ the error *E* of the objective function (3) is minimal. For $\varphi_2$ the exact solution can always be found, since $a_{12}$ is the only parameter to optimize. For the remaining $\varphi_k$ the greedy algorithm sets $a_{ik}$ in turn either –1 or 1, verifying the validity of the result, setting $a_{ik}$ the better value immediately and continuing with the next $a_{ik}$ until all *k*–1 values for $a_{ik}$ are set.