LEAST SQUARES METRIC, UNIDIMENSIONAL UNFOLDING

KEITH T. POOLE

GRADUATE SCHOOL OF INDUSTRIAL ADMINISTRATION
CARNEGIE-MELLON UNIVERSITY

The partial derivatives of the squared error loss function for the metric unfolding problem have a unique geometry which can be exploited to produce unfolding methods with very desirable properties. This paper details a simple unidimensional unfolding method which uses the geometry of the partial derivatives to find conditional global minima; i.e., one set of points is held fixed and the global minimum is found for the other set. The two sets are then interchanged. The procedure is very robust. It converges to a minimum very quickly from a random or non-random starting configuration and is particularly useful for the analysis of large data sets with missing entries.

Introduction

In the metric unfolding problem the data are assumed to be Euclidean distances plus some unknown observational error. Let \( x_{ik} \) be the \( i \)th individual’s estimated coordinate \((i = 1, \ldots, p)\) on the \( k \)th dimension \((k = 1, \ldots, s)\) and let \( z_{jk} \) be the estimated coordinate of the \( j \)th stimulus \((j = 1, \ldots, q)\) on the \( k \)th dimension. The corresponding \( s \) length vectors will be denoted as \( x_i \) and \( z_j \). The estimated distance between the \( i \)th individual and the \( j \)th stimulus is therefore

\[
d_{ij} = \left( \sum_{k=1}^{s} (x_{ik} - z_{jk})^2 \right)^{1/2}.
\] (1)

Let \( D^{1/2} \) be the \( p \) by \( q \) matrix of data, let \( \bar{D}^{1/2} \) be the \( p \) by \( q \) matrix of “true” but unknown distances, and let \( E \) be a \( p \) by \( q \) matrix of error. Assume that

\[
D^{1/2} = \bar{D}^{1/2} + E.
\] (2)

I will work with the standard squared error loss function

\[
\mu = tr \ E' E = \sum_{i=1}^{p} \sum_{j=1}^{q} e_{ij}^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} (d_{ij}^* - d_{ij})^2.
\] (3)

The partial derivatives of the \( z_{jk} \) and \( x_{ik} \) can be written as

\[
\frac{\partial \mu}{\partial z_{jk}} = -2 \sum_{i=1}^{p} (z_{ki} - z_{jk}),
\] (4)

\[
\frac{\partial \mu}{\partial x_{ik}} = 2 \sum_{j=1}^{q} (x_{ik} - x_{ik}),
\] (5)

where

\[
z_{ki} = x_{ik} + \frac{d_{ij}^*}{d_{ij}} (z_{jk} - x_{ik}),
\] (6)

\[
x_{ik} = z_{jk} + \frac{d_{ij}^*}{d_{ij}} (x_{ik} - z_{jk}).
\] (7)
The Geometry of the Partial Derivatives

Gleason (1967) was the first to point out one aspect of the unique geometry of these partial derivatives; namely, as (6) and (7) show, they involve a summation of equations of straight lines. For example, Figure 1 shows the geometry of (6) in the case of two dimensions.

A further unique property of the geometry of these partial derivatives is that the squared distance between the points $z_{j, i}$ (the $k$-length vector of the $z_{jkl}$) and $z_j$ is equal to the squared error between $d_{ij}^*$ and $d_{ij}$. That is,

$$e_{ij}^2 = \sum_{k=1}^{s} (z_{jkl} - z_{jk})^2 = (d_{ij}^* - d_{ij})^2$$

(see figure 1). A similar expression holds for $x_{i, j}$ and $x_i$.

In terms of this geometry, the update formulas for the steepest descent method (when the step size is fixed at one) have an interesting form. They can be written as

$$z_{jk}^{(h+1)} = \frac{\sum_{i=1}^{p} z_{jki}^{(h)}}{p}, \quad (8)$$

$$x_{ik}^{(h+1)} = \frac{\sum_{j=1}^{q} x_{ikj}^{(h)}}{q}, \quad (9)$$

where $h$ is the iteration number. The new $x_{ik}$ and $z_{jk}$ are simply the centroids of the points produced by the corresponding straight line equations. Intuitively, the process can be conceptualized as follows. Imagine that the $X$ set of points is fixed and $z_j$ is placed somewhere in the space. Think of the $d_{ij}^*$ as vectors attached to the respective $x_i$ and aim them at $z_j$ (see Figure 2 for a $p = 5, s = 2$ example). At the end of the vectors aimed at $z_j$, place points (the $z_{j,i}$ in Figure 2A). Now move $z_j$ to the centroid of these points and once again aim the $d_{ij}^*$ vectors so they are pointing at $z_j$ (the $z_{j, i}^{(2)}$ in Figure 2B).