Monotone Spline and Multidimensional Scaling

Z. Q. John Lu, MathSoft, Inc.
Douglas B. Clarkson, MathSoft, Inc.
Z.Q. John Lu, Data Analysis Product Division, 1700 Westlake Ave.N., Suite 500, Seattle, WA 98109, USA. jlu@statsci.com.

Keywords: monotone spline, multidimensional scaling, nonmetric scaling, cola data, graphical display.

1 Introduction

Multidimensional scaling seeks a proper geometric representation in Euclidean space of one or more matrices of dissimilarity or similarity (proximity) data. Each dissimilarity matrix represents the scores or ratings given by a subject for comparing pairs of stimuli.

In real situations the dissimilarity matrix is seldom exactly Euclidean, in the sense that it can be exactly represented as distances in Euclidean space of a given dimension, usually called the *object* or *configuration* space. Furthermore, in order to be visually explored, the object space usually has lower dimensionality, say less than 5, and so it is unlikely that high-dimensional data will be exactly represented. On the other hand, there are many situations when multivariate data expressed in higher dimensional coordinates have *lower dimension representations* such as points on a cycle, sphere, or some manifold. Chaotic time series data often have an *intrinsic* low-dimensionality. See Lu (1999) for some examples and an interesting study in the case of kernel regression. In high dimensional data analysis, an interesting question is how to approximate the data structure using some low-dimensional representations. Multidimensional scaling is one of the many useful techniques that can be used to recover hidden simplicity or order in complex situations. It is a data analytic and exploratory technique, and has been highly useful in graphical data display and visualization (Buja, Swayne, Littman, Dean 1998).

The question arises as to how to fit a distance model to dissimilarity data when the data are not exactly Euclidean. There are two main approaches in multidimensional scaling, namely "metric" and "nonmetric" scalings. In the metric approach, a parametric model is used to transform the data or distance. The actual data values are relevant and used in this approach, and the data type can be continuous (absolute), ratio, or interval (Arabie, Carroll, DeSarbo 1987). In the nonmetric approach, a nonparametric model is used to transform the data, in which only the order or rank of dissimilarity data is used (Kruskal and Wish 1978). In some situations the nonmetric approach makes more sense and it is an overall more *robust* procedure. The data types include ranking data, ordinal data as well as the data types handled by the metric approach.

The S-PLUS function, MDSCALE, is currently being developed for fitting different MDS models. It has functionalities comparable to the MDS procedure in SAS, or its precedent, the ALSCAL procedure. The well-known cola data (e.g. Clarkson et al 1999) are used to illustrate the developed procedure and its graphical display capabilities. We will focus on the implementation of *monotone spline* approach to nonmetric multidimensional scaling as developed in Ramsay (1988).

2 Monotone spline regression for transformation of data

In this section we discuss a nonparametric regression model for transformation of dissimilarity data. The most popular function is a polynomial function given as $f(x) = \sum_{i=1}^{k} a_i x^{i-1}$ at any given point x. There are at least two ways of adapting polynomials in nonparametric models. One way is via local models, see e.g. Fan and Gijbels (1996) and Lu (1996). Another approach uses splines. We will focus on polynomial splines, and especially monotone splines, which consist of monotone piecewise polynomial functions (Ramsay 1988).

Polynomial splines provide flexible transformation models using polynomials joined end-to-end in a smooth and constrained way. That is, an interval $[L, U]$ is subdivided by a mesh Δ consisting of points $L = \xi_1 < \cdots < \xi_q = U$. Within any subinterval $[\xi_j, \xi_{j+1})$ the function is a polynomial P_j of specified degree $k-1$ or order k.

To provide a convenient way of defining a spline

This work is funded by the NIMH SBIR Grant 2R44MH57559-02.

function, it is useful to incorporate the interior knot specification or mesh Δ and the continuity conditions ν_j into a *knot sequence* $t = \{t_1, \cdots, t_{n+k}\}$ where n is the number of free parameters that specify the spline function having the specified continuity characteristics. The knot sequence has the properties:

1. $t_1 \leq \cdots \leq t_{n+k}$.
2. For all i there is some j such that $t_i = \xi_j$.
3. The continuity characteristics are determined by:
   (a) $t_1 = \cdots = t_k = L$ and $U = t_{n+1} = \cdots = t_{n+k}$;
   (b) $t_i < t_{i+k}$ for all i;
   (c) if $t_i = \xi_j$ and $t_{i-1} < \xi_j$, then
   $$t_i = \cdots = t_{i+k-\nu_j-1}.$$

In the most common situation where maximal continuity $k-1$ holds at each boundary point in the interior of $[L, U]$, there is one knot at each interior boundary, the related sequence is referred to as a *simple* knot sequence. The number of free parameters n is equal to the number of knots interior to $[L, U]$ plus the order k.

It is convenient in the application of splines to have a suitable set of *basis splines*, $M_i(\cdot|k,t), i = 1, \cdots, n$, such that any piecewise polynomial or spline f of order k associated with knot sequence t can be represented as the linear combination $f = \sum_{i=1}^{n} a_i M_i$.

A set of basis splines particularly appealing to statisticians is the M-spline family in which $M_i, i = 1, \cdots, n$, is defined such that it is positive in (t_i, t_{i+k}), zero elsewhere, and has the normalization $\int_L^U M_i(x)dx = 1$ (Curry and Schoenberg 1966). A computationally convenient specification for $t_i \leq x < t_{i+1}$ is the recursion

$$M_i(x|1,t) = \begin{cases} \frac{1}{t_{i+1}-t_i}, & \text{if } t_i \leq x \leq t_{i+1} \text{ and,} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

$$M_i(x|k,t) = \begin{cases} [(k-1)(t_{i+k}-t_i)]^{-1} \\ \{k[(x-t_i)M_{i-1}(x|k-1,t) \\ +(t_{i+k}-x)M_i(x|k-1,t)]\}, \\ \text{if } t_i < x < t_{i+k} \text{ and ,} \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

for $k > 1$, where $M_0(\cdot|k-1,t), M_n(\cdot|k-1,t)$ are defined as zeros. (Note slight modifications from Ramsay (1988).)

Each M_i has the properties of a probability density function and is nonzero only over the interval $[t_i, t_{i+k}]$. This *localization* property of the M-spline basis shares some features with *wavelets* basis functions (Bruce and Gao 1996). This property has advantages in computational efficiency and numerical stability. For example, a change in coefficient a_i will only affect f within the interval $[t_i, t_{i+k}]$, achieving a desirable local sensitivity to coefficient values. Moreover, in order for $f = \sum_{i=1}^{n} a_i M_i$ to be a density function, we only need to require the coefficients to satisfy $a_i \geq 0$ for all $i \geq 0$ and $\sum_{i=1}^{n} a_i = 1$.

The B-spline basis $B_i, i = 1, \cdots, n$ is closely related to the M-spline basis, in that

$$B_i = (t_{i+k} - t_i)M_i/k. \tag{3}$$

Note that the B-splines are also nonnegative and $B_i(x) > 0$ only for $t_i < x < t_{i+k}$. The B-splines have the normalization property $\sum_{i=1}^{n} B_i(x) = 1$ for all x, which is quite useful in some applications.

Monotone splines Our focus is on monotone splines, and first we define a set of basis consisting of monotone splines. We derive the integrated splines I_i, or I-splines, from the M-spline basis, as follows

$$I_i(x|k,t) = \int_L^x M_i(u|k,t)du,$$

for $i = 1, \cdots, n$. Since M_i is nonnegative, I_i is a nondecreasing function, and so any function of the form $f = \sum_{i=1}^{n} a_i I_i$ where $a_i \geq 0$ is always nondecreasing. Any monotone function f or $-f$ may be approximated this way if the I-spline basis is flexible enough. Note that I_i is a piecewise polynomial of degree k since M_i is a piecewise polynomial of degree $k-1$. Further, $I_i(x) = 0$ for $x \leq t_i$ and 1 if $x > t_{i+k}$.

We will reserve the term *order* k to an M-spline of degree $k-1$ or an I-spline of degree k. For *simple* knot sequences, which have only one knot at each interior boundary, n is equal to order k plus the number of interior knots. For such simple knot sequence, the I-spline I_i can be expressed in the convenient form as a simple function of an M-spline basis of order $k+1$, i.e. $M_i(\cdot|k+1,t), i = 1, \ldots, n+1$. Let t be the simple knot sequence of order $k+1$, then for all x such that $t_j \leq x < t_{j+1}$,

$$I_i(x|k,t) = \begin{cases} 0, \ i > j; \\ \sum_{m=i}^{j}(t_{m+k+2} - t_{m+1}) \\ M_{m+1}(x|k+1,t)/(k+1), \\ j-k+1 \leq i \leq j; \\ 1, \ i < j-k+1; \end{cases} \tag{4}$$

where $i = 1, \cdots, n$. Note that $t_{n+1} < t_{n+2}$, since t is a sequence of order $k+1$. Also, notice that the terms

inside the summation are exactly B-splines of order $k+1$ from (3). (Again note the subtle differences from Ramsay (1988)'s formulas.)

Figure 1 shows the six I-splines and M-splines of order 3, and the associated B-splines of order 4 on the interval $[0,1]$, where the interior knots are chosen at $0.3, 0.5, 0.6$. Note that the M-splines are piecewise quadratic and are the first derivatives of the corresponding I-splines. The B-splines are piecewise cubic and have the characteristic of summing to one at each location.

A monotone spline transformation is a linear combination of I-splines with nonnegative coefficients. Since the I-splines are nondecreasing piecewise polynomials, the spline transformation is also a nondecreasing piecewise polynomial. The idea is that with enough knots, this type of transformation can approximate any monotone function at desired accuracy. For a given bivariate data set, the *monotone spline regression* is the nonnegative least squares fit of such a linear model. This approach to data modeling is certainly feasible in practice as demonstrated from the extensive study of local polynomial regression (cf. Fan and Gijbel 1996, Lu 1999).

While local polynomial regression requires a decision on a bandwidth or a size of the neighborhood to be used at each fitting, the polynomial spline regression requires the choice of knots. It turns out that the location of knots does not have significant impact on the fit, and it is the number of knots that is much more crucial. In practice, however, there is often enough flexibility in a curve defined by a single interior knot, and it is not usually necessary to use a large numbers of knots. As a rule of thumb, we suggest using two interior knots placed at roughly equally spaced locations.

At one extreme, when the data are ordered in nondecreasing order, one may place knots at each distinct data point, and this type of spline models is also very useful. In particular, when the linear spline ($k=1$) is used, this model corresponds exactly to the isotonic regression as discussed by Barlow et al (1972) and Kruskal (1964). In another important situation when the basis functions consist of some particular class of expansion series, not necessarily restricted to the B-spline basis, the resulting spline model is called a *smoothing spline*, though this name is often reserved for situations when a penalty term is used to make the estimator smoother(cf. Ramsay and Silverman 1997).

When the data consist of multi-way observations or measurements by multiple subjects, there is a class of *curves* that need to be fitted, each of which corresponds to a specific subject. Such curves, called *functional data*, occur in many areas, including meteorology, psychology, and biomedicine. We refer to Ramsay and Silverman (1997) for detailed discussions on this promising area. Next, we present a nontrivial application to individual difference models in multidimensional scaling.

3 Nonmetric multidimensional scaling

To fix ideas, consider a data example analyzed in Clarkson, Gonzalez, and Lu (1999) among others. The data set consists of pairwise comparison of 10 cola brands by 10 subjects. (Thus, there is a total of $45 \times 10 = 450$ data points.) In this example, the main interest is to discern the relative difference among the different cola brands as perceived by the testers, and it is thus natural to assume a *nonmetric* multidimensional scaling model. To fix notation, we assume there are multiple subjects or populations, each one producing a set of dissimilarity data, and a separate transformation is likely required for each subject. Thus, there can be as many spline transformations as the number of subjects.

We also assume the *individual difference* model for subject differences. Thus, the overall model is defined as follows: There is a *monotone* function f_k such that

$$f_k(y_{kij}) = d_{kij} + \varepsilon_{kij} \tag{5}$$

where $i, j = 1, \cdots, n$ for each subject $k = 1, \cdots, m$. Since separate transformations may be employed for each different subject, f_k may be different across k. Usually d_{kij} is assumed to be the weighted distance

$$d_{kij} = \sqrt{\sum_{\ell=1}^{\tau} w_{k\ell}(x_{i\ell} - x_{j\ell})^2}. \tag{6}$$

The goal is to estimate the *configuration* matrix $X = (x_{ij})$ in the object space and the *weight* matrix $W = (w_{k\ell})$ in the subject space. The transformed data $f_k(y)$ can be interpreted as the "unobserved" perceptive distance by the kth subject. The transformed data are also called the *disparities*. Define $\hat{y}_{kij} = \hat{f}_k(y_{kij})$.

If there is no smoothness assumption on monotone f_k, the resulting model fitting is based on isotonic regression (Kruskal 1964, Barlow et al 1972). The isotonic regression is usually too rough and it is desirable to impose certain smoothness, leading to the use of monotone spline transformation as discussed in Section 2. In particular, the linear spline transformation with knots at every distinct data points corresponds to the isotonic transformation.

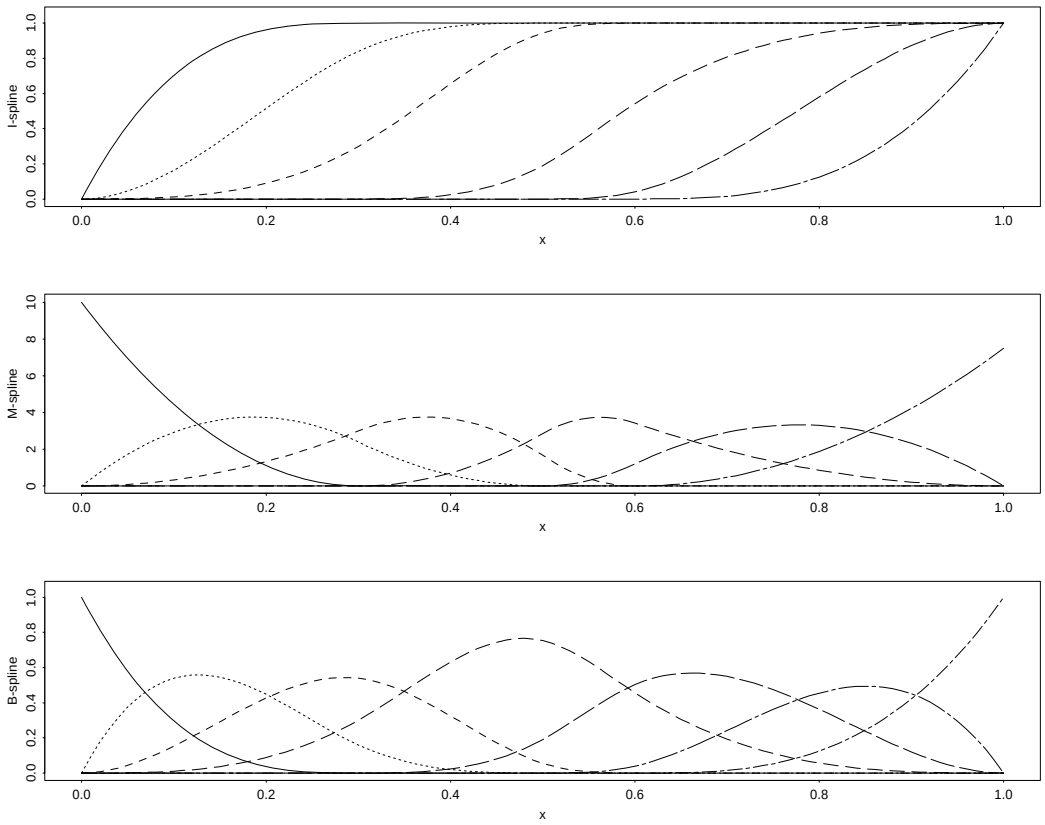

Figure 1: Illustration of definitions of different splines
The six I-splines and M-splines of order 3, and the associated B-splines of order 4. The interior knots are at $0.3, 0.5, 0.6$.

Following Kruskal's formulation, our objective is to minimize the fit measure, or "stress" function, defined as

$$S = \sum_{k=1}^{m} \delta_k \sum_{i,j} (\hat{y}_{kij} - d_{kij})^2, \tag{7}$$

where δ_k is a weighting function often given by

$$\delta_k = 1/\sum_{ij} \hat{y}_{kij}^2. \tag{8}$$

(or sometimes given by $1/\sum_{ij}(\hat{y}_{kij} - \bar{y}_k)^2$ where $\bar{y}_k = \frac{1}{n(n-1)} \sum_{ij} y_{kij}$.) The parameters to be fitted include elements in X and W as well as coefficients in the spline transformation. We use a *sequential* optimization procedure in which iteratively S is minimized over X and W, given the disparities $\hat{y}_{kij}$; while given X and W, we fit $\hat{y}_{kij}$ by using a nonnegative least square procedure, and then reweighting. Within the spline fitting step, since S is a sum of m terms, each of which involves coefficients for a separate spline, the spline fit can be implemented separately for each subject.

In the case of isotonic transformation, there is more direct algorithms such as Kruskal (1964)'s up and down algorithm or equivalently the "pooled adjacent violators (PAV)" algorithm of Barlow et al (1972) for finding the *optimal* monotone transformation.

Cola example. We fit the nometric individual difference diagonal model to this data set using dimension $p = 3$, spline transformation of degree 2 and 2 interior knots. The transformed data, or disparities versus the original data are plotted in Figure 2, where different symbols represent different subjects ("1", subject one; "2", subject two, and so on, with "0" representing subject ten). To evaluate the fitness of the selected model, we also plot the residuals (disparities-distance) for each subject, and this is given in Figure 3. It is seen that the spline transformation model is fairly adequate in this data set.

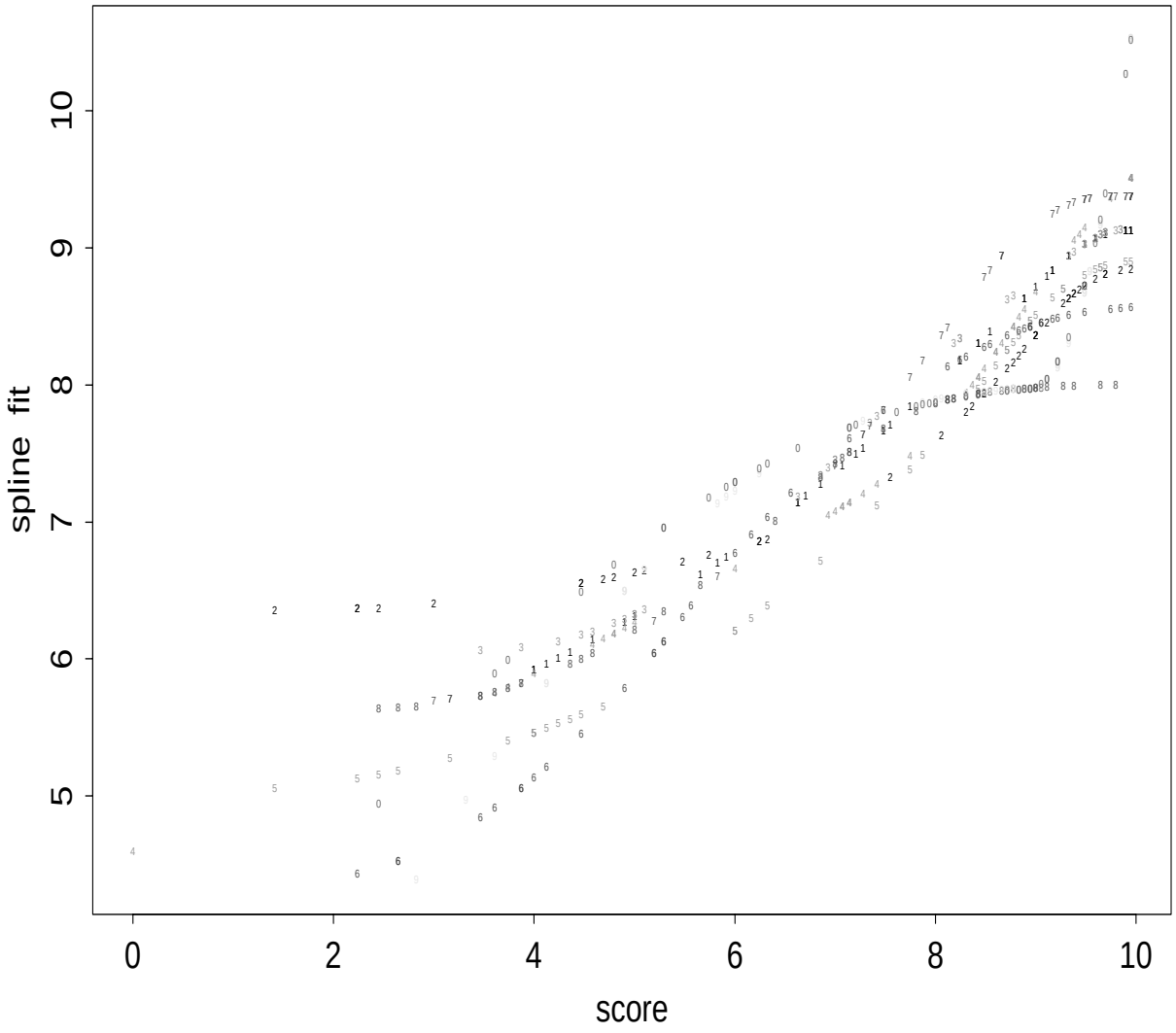

Figure 2: Monotone spline transformations of the cola data
Separate transformations applied to each subject. The I-spline basis of degree 2 is used, placed at two equally spaced interior knots. ("1" for subject one; "2" for subject two, and so on, with "0" for subject ten.)

References

Arabie, P., J.D. Carroll, W.S. DeSarbo (1987). *Three-Way Scaling and Clustering.* Sage Publications, Newbury Park.

Barlow, R.E., D.J. Bartholomew, J.M. Bremner, H.D. Brunk (1972). *Statistical Inference under Order Restrictions: The Theory and Applications of Isotonic Regression.* John Wiley, New York.

Bruce, A. and H.Y. Gao (1996). *Applied Wavelet Analysis with S-PLUS.* Springer and MathSoft Inc.

Buja, A., D.F. Swaye, M. Littman, N. Dean (1998). XGvis: Interactive Data Visualization with Multidimensional Scaling. *J. of Computational and Graphical Statistics*, to appear.

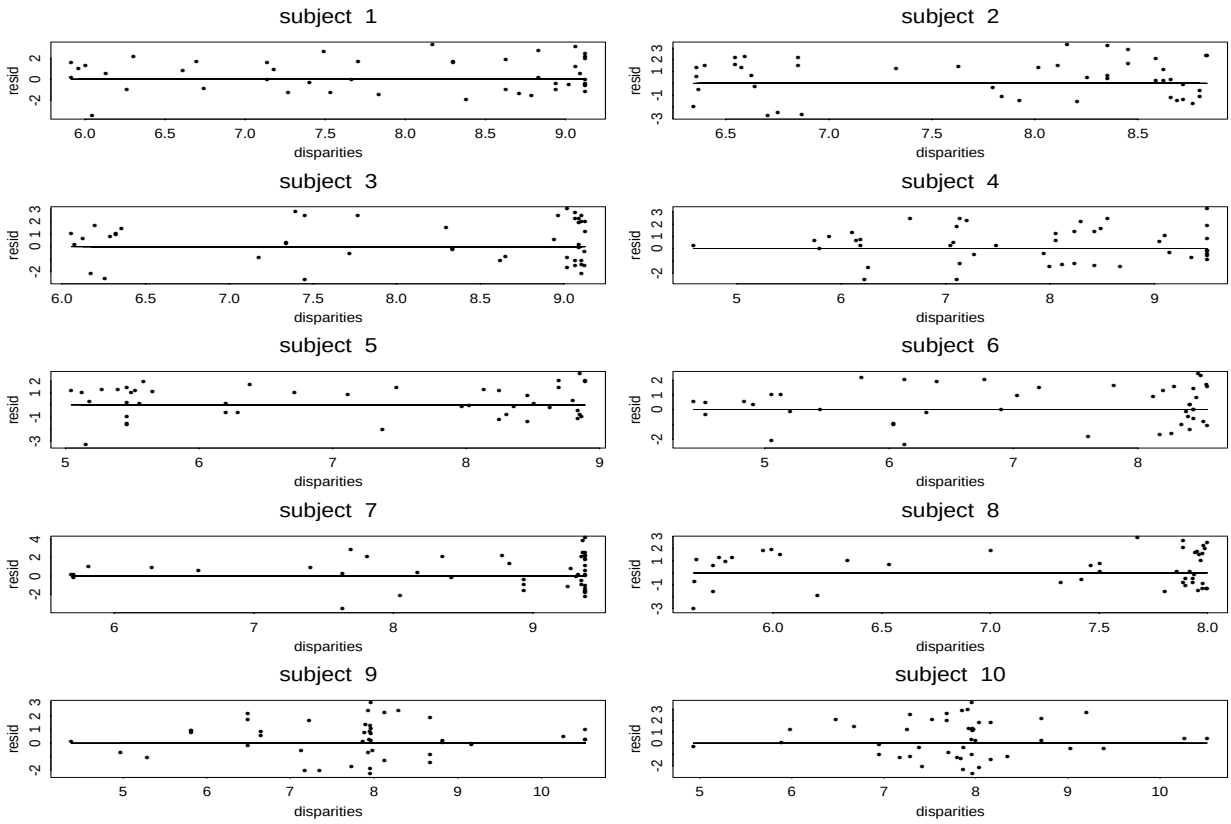

Figure 3: Residual plot after fitting the nonmetric diagonal MDS model to the cola data. It is emphasized that a proper way to see the goodness of fit is to plot separately for each subject, as shown here.

Clarkson, D.B., R. Gonzalez, Z.Q. Lu (1999). Random effects multidimensional scaling. *Proceedings of the 31st Symposium on the Interface: Computing Science and Statistics*, to appear.

Fan, J. and I. Gijbel (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Kruskal, J.B.(1964). Nonmetric multidimensional scaling. *Psychometrika*, **29**, 1-27, 115-129.

Kruskal, J.B. and M. Wish (1978). *Multidimensional Scaling.* Sage Publications, Newbury Park.

Lu, Z. Q. (1999). Nonparametric regression with singular design. *J. of Multivariate Anal.*, **70**, 177-201.

Ramsay,J.O. (1988). Monotone Regression Splines in Action. *Statistical Science*, Vol.3, No.4, 425-441.

Ramsay, J.O. and B.W. Silverman (1997). *Functional Data Analysis.* Springer, New York.