AN UPPER BOUND FOR SSTRESS

JAN DE LEEUW AND BERT BETTONVIL

DEPARTMENT OF DATA THEORY FSW
UNIVERSITY OF LEIDEN

In this note we derive an upper bound for the minimum for the multidimensional scaling loss function sstress. We conjecture that minimum sstress solution will be biased towards regular positioning of clumps of points over the surface of a sphere.

Key words: Nonmetric scaling, multidimensional scaling, loss functions, distance geometry.

Introduction

In a recent paper de Leeuw and Stoop (1984) proved some interesting upper bounds for Kruskal’s multidimensional scaling loss function stress. The bounds are of the following form. Suppose $\sigma(X)$ is the stress of a configuration $X$, which is a matrix with $n$ rows and $p$ columns. de Leeuw and Stoop define a function $\kappa(n, p)$, with the property that the minimum of $\sigma(X)$ over all configurations is always less than or equal to $\kappa(n, p)$. Thus the minimum loss in a scaling problem is always less than or equal to $\kappa(n, p)$, a number which is independent of the data. The function $\kappa(n, p)$ is not at all easy to compute if $p > 1$. de Leeuw and Stoop give some mathematical results, some numerical results, and some conjectures, which together give a fairly complete picture of the function. They also conclude, tentatively, from their results that multidimensional scaling results based on stress may have the bias of equidistributing the points over surface and/or interior of the unit sphere.

In this short note we investigate exactly the same problem for sstress, the loss function used for example in ALSCAL (Takane, Young, de Leeuw, 1977). It turns out that the theory for sstress is considerably simpler than for stress, and much more specific results can be obtained.

Preliminary Results

Our multidimensional scaling problem has $n$ points, which must be scaled in $p$ dimensions. The data are a rank ordering of the $(\frac{n(n-1)}{2})$ dissimilarities. We use $d_{ij}(X)$ for the Euclidean distance between rows $i$ and $j$ of $X$, and we use $\tilde{d}_{ij}$ for a matrix of feasible disparities (i.e., numbers monotone with the original dissimilarities). We define

$$\tilde{\sigma}(X, \tilde{D}) = \left\{ \frac{\sum_{i\neq j} (d_{ij} - \tilde{d}_{ij}(X))^2}{\sum_{i<j} \tilde{d}_{ij}^2(X)} \right\}^{1/2}.$$

We use tildes above symbols to show that we are working with sstress, not with stress.

This study has been supported by the Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (Netherlands Organization for the Advancement of Pure Research) under grant 56-97.

Requests for reprints should be sent to Jan de Leeuw, Department of Data Theory FSW/RUL, Middelsteegracht 4, 2312 TW Leiden, THE NETHERLANDS.

0033-3123/86/0300-7062$00.50/0
© The Psychometric Society
stress. The ssstress of a configuration is defined as

$$\bar{\sigma}(X) = \min \{ \bar{\sigma}(X, \bar{D}) | \bar{D} \text{ a feasible disparity matrix} \}.$$ 

Thus the ssstress depends on the ordering of the dissimilarities, but because this is fixed for the problem we do not indicate this dependence explicitly.

Following de Leeuw and Stoop we now define

$$\tilde{\sigma}(X) = \min \{ \bar{\sigma}(X, \bar{D}) | D = \theta(E - I), \theta \geq 0 \},$$

where $E - I$ is the matrix with all diagonal elements zero and all off-diagonal elements one. The matrix $\theta(E - I)$ is a feasible disparity matrix in any nonmetric scaling problem, independent of the order of the dissimilarities. Thus

$$\bar{\sigma}(X) \leq \tilde{\sigma}(X).$$

Now let $\bar{\sigma}(n, p)$ be the minimum of $\bar{\sigma}(X)$ over all $n \times p$ matrices, and let $\tilde{\sigma}(n, p)$ be the minimum of $\tilde{\sigma}(X)$ over all $n \times p$ matrices. Then

$$\bar{\sigma}(n, p) \leq \tilde{\sigma}(n, p).$$

This is the basic upper bound result mentioned in the introduction. Observe that $\bar{\sigma}(n, p)$ still depends on the order of the dissimilarities, while $\tilde{\sigma}(n, p)$ does not. The rest of this note is concerned with the properties of $\tilde{\sigma}(n, p)$. The function $\kappa(n, p)$, mentioned in the introduction, is derived by de Leeuw and Stoop in precisely the same way, starting from stress instead of ssstress.

Computations

By elementary computations we find, directly from the definition,

$$1 - \tilde{\tau}^2(X) = \frac{\left( \sum_{i<j} d_{ij}(X) \right)^2}{\left( \frac{1}{2} \sum_{i<j} d_{ij}^2(X) \right)}.$$

Without loss of generality we restrict ourselves to configurations that are normalized. By this we mean that (a) their columns sum to zero, (b) their columns are orthogonal, and (c) the sum of squares of all their elements is equal to unity. For further computation it is convenient to define $c_{ij}(X)$, which is element $(i, j)$ of $XX'$. Moreover $a_i(X)$ is short for $c_{ii}(X)$, the sum of squares of row $i$ of $X$. And $b_s(X)$ is the sum of squares of column $s$ of $X$. In the multidimensional scaling context the $a_i(X)$ are the squared distances of the $i$-th point from the origin, and the $b_s(X)$ are the eigenvalues associated with the $s$-th dimension of the configuration. Observe that the $a_i(X)$ sum to one, and so do the $b_s(X)$.

We now have

$$\sum_{i<j} d_{ij}^2(X) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2(X)$$

$$= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i(X) + a_j(X) - 2c_{ij}(X)) = n.$$

In the same way

$$\sum_{i<j} d_{ij}^2(X) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i(X) + a_j(X) - 2c_{ij}(X))^2$$

$$= n \sum_{i=1}^n a_i^2(X) + 2 \sum_{s=1}^p b_s^2(X) + 1.$$