PSYCHOMETRIKA—VOL. 51, NO. 1, 149–153
MARCH 1986
NOTES AND COMMENTS

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 $\binom{n}{2}$ dissimilarities. We use $d_{ij}(X)$ for the Euclidean distance between rows i and j of X, and we use $\hat{d}_{ij}$ for a matrix of feasible disparities (i.e., numbers monotone with the original dissimilarities). We define

$$\tilde{\sigma}(X, \hat{D}) = \left\{ \frac{\sum\sum_{i<j} (\hat{d}_{ij} - d_{ij}^2(X))^2}{\sum\sum_{i<j} d_{ij}^4(X)} \right\}^{1/2}.$$

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

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, Middelstegracht 4, 2312 TW Leiden, THE NETHERLANDS.

0033-3123/86/0300-7062$00.50/0

stress. The sstress of a configuration is defined as

$$\tilde{\sigma}(X) = \min \{\tilde{\sigma}(X, \hat{D}) \mid \hat{D} \text{ a feasible disparity matrix}\}.$$

Thus the sstress 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

$$\tilde{\tau}(X) = \min \{\tilde{\sigma}(X, \hat{D}) \mid \hat{D} = \theta(E - I), \theta \geq 0\},$$

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

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

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

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

This is the basic upper bound result mentioned in the introduction. Observe that $\tilde{\sigma}(n, p)$ still depends on the order of the dissimilarities, while $\tilde{\tau}(n, p)$ does not. The rest of this note is concerned with the properties of $\tilde{\tau}(n, p)$. 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 sstress.

Computations

By elementary computations we find, directly from the definition,

$$1 - \tilde{\tau}^2(X) = \frac{(\sum\sum_{i<j} d_{ij}(X))^2}{\binom{n}{2} \sum\sum_{i<j} d_{ij}^4(X)}.$$

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

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

In the same way

$$\sum\sum_{i<j} d_{ij}^4(X) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i(X) + a_j(X) - 2c_{ij}(X))^2$$
$$= n \sum_{i=1}^{n} a_i^2(X) + 2 \sum_{s=1}^{p} b_s^2(X) + 1.$$

By Cauchy-Schwartz, applied to the first two terms,

$$\sum_{i<j}\sum d_{ij}^4(X) \geq n\left(\frac{1}{n}\right) + 2\left(\frac{1}{p}\right) + 1 = 2\left(\frac{p+1}{p}\right),$$

with equality if and only if all $a_i(X)$ are equal to n^{-1} and all $b_s(X)$ are equal to p^{-1}. Combining our computations so far gives

$$1 - \tilde{\tau}^2(X) \leq n^2 \bigg/ \left\{\binom{n}{2}^2\left(\frac{p+1}{p}\right)\right\} = \left(\frac{n}{n-1}\right)\left(\frac{p}{p+1}\right).$$

Thus

$$\tilde{\tau}(X) \geq \left\{1 - \left(\frac{n}{n-1}\right)\left(\frac{p}{p+1}\right)\right\}^{1/2}.$$

Now suppose we call a normalized configuration *regular* if all rows have sum of squares n^{-1} and all columns have sums of squares p^{-1}. If there exists a regular configuration with n rows and p columns ($n > p$) then the pair (n, p) is also called regular. We have proved the following result

Theorem. If (n, p) is regular then

$$\tilde{\tau}(n, p) = \left\{1 - \left(\frac{n}{n-1}\right)\left(\frac{p}{p+1}\right)\right\}^{1/2}.$$

Observe that, as intuition suggests, $\tilde{\tau}(n, p)$ is strictly increasing in n and strictly decreasing in p. Also observe that $\tilde{\tau}(n, n-1) = 0$, reflecting the fact that $(E - I)$ is also a squared Euclidean distance matrix in $p = n - 1$ dimensions.

Regularity

The theorem in the previous section gives an upper bound for all pairs (n, p) for which regular configurations exist. But not all pairs of natural numbers (n, p), with $n > p$, are regular. This follows directly from our first result in this section.

Result 1. $(n, 1)$ is regular if and only if n is even. In this case regularity means existence of nonzero real numbers, of equal modulus, which add up to zero. These numbers can only exist if half of them are negative, and the other half are positive.

Result 2. $(n, 2)$ is regular. This can be seen by choosing n points regularly spaced on the circle with center in the origin and radius $n^{-1/2}$. The summation calculus can be used to show that the resulting configuration is indeed regular.

Result 3. If $(n_1, p), \ldots, (n_m, p)$ are regular, then $(\sum n_j, p)$ is regular. Suppose the regular configurations are X_j. Let $n_+ = \sum n_j$. By writing the m configurations $(n_j/n_+)^{1/2}X_j$ on top of each other we create a regular configuration of dimension $n_+ \times p$.

Result 4. $(n, n-1)$ is regular. Take a square orthonormal matrix, that is, an X with $X'X = XX' = I$, of order n whose first column has all elements equal to $n^{-1/2}$. Delete the first column and multiply by $(n-1)^{-1/2}$. The resulting configuration is regular.

Result 5. If (n, p) is regular, and $n > p + 1$, then $(n, n - (p + 1))$ is regular. Suppose X is regular, and $n \times p$. Multiply by $p^{1/2}$. Add a column with all elements equal to $n^{-1/2}$ and add the $n \times (n - (p + 1))$ matrix Y which makes the complete $n \times n$ matrix square orthonormal. Then $(n - (p + 1))^{-1/2}Y$ is regular.

Result 6. If (n, p) is regular, and $n \leq \frac{1}{2}p(p + 3)$, then $(n + p + 3, p + 2)$ is regular. This result make look a bit contrived, but we need it in the proof of the theorem below. Suppose X is a $(p + 3) \times 2$ matrix which is regular. Such an X exists by Result 2. Suppose Y is a $(p + 3) \times p$ matrix which is regular. Such a Y always exists because of results 5 and

2. Suppose Z is a regular $n \times p$ matrix, which exists by hypothesis. Then form

$$\begin{vmatrix} \alpha X & \beta Y \\ 0 & \gamma Z \end{vmatrix}$$

with

$$\alpha = (2/(p+2))^{1/2}, \qquad \gamma = (n/(n+p+3))^{1/2},$$

and

$$\beta = ((p^2 + 3p - 2n))/\{(p+2)(n+p+3)\})^{1/2}.$$

The resulting configuration is regular. Of course we can only choose β in this way if $p^2 + 3p - 2n \geq 0$, i.e. if $n \leq \frac{1}{2}p(p+3)$.

It is clear that these six results can be used to generate many regular configurations. It turns out that we can do this in a very systematic way, producing the following interesting theorem.

Theorem. If n is even and/or p is even, and $p < n$, then (n, p) is regular.

Proof. Results 1 and 2 show that the theorem is true for $p = 1$ and $p = 2$. We now use strong induction on p, distinguishing the cases in which p is odd and even. Suppose the theorem is true for $q < p$, where $p > 2$. This is the induction hypothesis.

First consider the case with p odd. Then $(p + 1, p)$ is regular by Result 4. Choose $1 < k \leq p$, with k odd. By the induction hypothesis $(p + k, k - 1)$ is regular, because $k - 1$ is even, and Result 5 then gives regularity of $(p + k, p)$. This shows that $(p + 1, p)$, $(p + 3, p), \ldots, (2p, p)$ are all regular. Now repeated application of Result 3 proves regularity of (n, p) for all $n > p$, with n even.

Now suppose p is even. Choose $1 < k \leq p$. Then $(p + k, k - 1)$ is regular by the induction hypothesis, because $p + k$ is even if k is even and $k - 1$ is even if k is odd. Result 5 shows that $(p + k, p)$ is regular. We know again from Result 4 that $(p + 1, p)$ is regular. From the induction hypothesis $(p, p - 2)$ is regular. Because $p < (p - 2)(p + 1)/2$ we find from Result 6 that $(2p + 1, p)$ is regular. Thus $(p + 1, p), (p + 2, p), \ldots, (2p + 1, p)$ are all regular. Again Result 3 can be applied repeatedly to prove regularity of (n, p) for all n with $n > p$.

The theorem now follows by strong induction. Q.E.D. □

Observe that the theorem only provides a sufficient condition for regularity. It is still possible that pairs (n, p), with neither n nor p even, are regular, although some such pairs are excluded by our earlier results. Thus $(n, 1)$ with n odd is not regular by Result 1. By combining Results 5 and 1 we also see that $(n, n - 2)$ is not regular if n is odd. But these are about the only results we have in the direction of showing that our conditions are also necessary.

Conclusion

In keeping with the results and interpretations of de Leeuw and Stoop we conclude that, especially in the case of poor fit, multidimensional scaling solutions based on sstress may be biased towards distributing clusters of points regularly over the surface of a sphere. There may be many local minima of the same type in such bad-fitting cases, because it is largely arbitrary how many clumps there are, and which points are assigned to which clumps. Again we emphasize that this may be a possible explanation for the clumping effect we have sometimes observed in real ALSCAL applications. Of course the clusters one finds in these cases may also be "real," although one expects real clusters to correspond with a good fit. And, of course, we have only *proved* the upper bound. We conjecture a bias towards regularity.

References

de Leeuw, J., & Stoop, I. (1984). Upper bounds for Kruskal's stress. *Psychometrika, 49*, 391–402.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42*, 7–68.

Manuscript received 1/14/85
Final version received 8/21/85