AN UPPER BOUND FOR SSTRESS

Jan de Leeuw
Bert Bettonvil
Department of Data Theory
Leiden University

INTRODUCTION

In a recent paper De Leeuw and Stoop (1984) proved some interesting upper bounds for Kruskal's loss function <u>stress</u>, which is used in many multidimensional scaling programs. The bounds are of the following form. Suppose $\sigma(X)$ is the stress of a configuration X, which is 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 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 <u>sstress</u>, 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. Thus the $\hat{d}_{ij}$ form a symmetric nonnegative matrix, with zero diagonal, whose off-diagonal elements are monotone with the disparities. We define

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

We use tildes above symbols to show that we are working with sstress, not with stress. The sstress of a configuration is defined as

$\tilde{\sigma}(X) = \min\{\tilde{\sigma}(X,\hat{D}) \mid \hat{D}$ 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 dependency explicitly.

Following De Leeuw and Stoop we now define

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

where $E - I$ is the matrix with all diagonal elements zero and all off-diagonal elements one. Because $\theta(E - I)$ is feasible in any nonmetric scaling problem we have the result

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

If $\tilde{\kappa}(n,p)$ is the minimum of $\tilde{\tau}(X)$ over all $n \times p$ matrices, we also have

$\min \{\tilde{\sigma}(X) \mid X\} \leq \tilde{\kappa}(n,p)$.

This is the basic upper bound result mentioned in the introduction. The rest is computation.

COMPUTATIONS

By elementary computation we find, directly from the definition,

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

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, (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 of $c_{ii}(X)$, the sum of squares of row i of X. And $b_s(X)$ is the sum of squares of column s of X. Observe that the $a_i(X)$ sum to one, and so do the $b_s(X)$.

We now have

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

In the same way

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

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

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

$$\sum_{i<j}\sum d_{ij}^4(X) \geqq n(\tfrac{1}{n}) + 2(\tfrac{1}{p}) + 1 = 2(\tfrac{p+1}{p}),$$

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

$$1 - \tilde{\tau}^2(X) \leqq n \,/\binom{n}{2}2(\tfrac{p+1}{p}) = (\tfrac{n}{n-1})(\tfrac{p}{p+1}).$$

Thus

$$\tilde{\tau}(X) \geqq \{1 - (\tfrac{n}{n-1})(\tfrac{p}{p+1})\}^{\frac{1}{2}}.$$

Now suppose we call a normalized configuration <u>regular</u> if all rows have sum of squares n^{-1} and all columns have sums of squares p^{-1}. We have proved the following result.

<u>Theorem</u>. If there exists a regular n x p configuration, then

$$\tilde{\kappa}(n,p) = \{1 - (\tfrac{n}{n-1})(\tfrac{p}{p+1})\}^{\frac{1}{2}}.$$

REGULARITY

The theorem in the previous section gives us the 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 in this sense. This follows directly from our first result.

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

<u>Result 2</u>. (n,2) is regular. This can be shown by choosing n points regularly spaced on the unit circle, and by using the summation calculus to show that the resulting configuration is regular if suitable normalized.

<u>Result 3</u>. If $(n_1,p),\dots,(n_m,p)$ are regular, then $(\sum n_j,p)$ is regular. Simply write all configurations below each other, and renormalize.

Result 4. $(n,n-1)$ is regular. Take a square orthonormal matrix whose first column has all elements equal. Delete the first column, and renormalize.

Result 5. If (n,p) is regular, then $(n,n-(p+1))$ is regular. Suppose X is the $n \times p$ regular configuration. Add a column with all elements equal, and add the $n \times (n-(p+1))$ matrix Y which makes the complete matrix square orthonormal. Then Y, suitably normalized, is also regular.

Result 6. If (n,p) is regular, and $n \leqslant p(p+3)/2$, then $(n+p+3,p+2)$ is regular. Take a $(p + 3) \times 2$ matrix X which is regular (possible by result 2). Take a $(p + 3) \times p$ matrix Y which is regular (possible by result 5). Suppose Z is the $n \times p$ regular matrix whose existence is guaranteed by hypothesis. Then form

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

A little computation shows that we can choose α and β such that the resulting configuration is regular.

It is clear that the six results mentioned above can be used to generate many regular configurations. It turns out that we can do this in a systematic way to produce 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. Suppose the theorem is true for $q < p$, where $p > 2$. Suppose in addition that p is even. Choose $1 < k \leqslant p$. 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, from result 4, that $(p + 1,p)$ is regular. The induction hypothesis also gives regularity of $(p,p - 2)$, and because $p < (p - 2)(p + 1)/2$ we find from result 6 that $(2p + 1,p)$ is regular. We now have regularity of $(p + 1,p)$, $(p + 2,p)$, ... ,$(2p + 1,p)$. Application of result 3 repeatedly shows that (n,p) for all n. Now suppose that p is odd. Again $(p + 1,p)$ is regular by result 4. Choose $1 < k \leqslant p$, 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 $(p + k,p)$. Thus $(p + 1,p)$, $(p + 2,p)$, ... , $(2p,p)$

are all regular. Again repeated application of result 3 proves regularity of (n,p) for all $n > p$, with n even. By induction the theorem follows for all p. Q.E.D.

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 explanantion for the clumping effect we have sometimes observed in real ALSCAL applications, of course these clusters may also be 'real' (of course in that case they would correspond to a good fit). 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.