PSYCHOMETRIKA—VOL. 49, NO. 3, 391–402
SEPTEMBER 1984

UPPER BOUNDS FOR KRUSKAL'S STRESS

JAN DE LEEUW AND INEKE STOOP

LEIDEN UNIVERSITY

In this paper the relationships between the two formulas for stress proposed by Kruskal in 1964 are studied. It is shown that stress formula one has a system of nontrivial upper bounds. It seems likely that minimization of this loss function will be liable to produce solutions for which this upper bound is small. These are regularly shaped configurations. Even though stress formula two yields less equivocal results, it seems to be expected that minimization of this loss function will tend to produce configurations in which the points are clumped. These results give no clue as to which of the two loss functions is to be preferred.

Key words: nonmetric scaling, multidimensional scaling, distance geometry.

Introduction

In this paper we study some properties of STRESS, the loss function for nonmetric multidimensional scaling introduced by Kruskal (1964a, b). Actually, Kruskal introduced two different loss functions, subsequently called *stress formula one* and *stress formula two.* Both loss functions assume values between zero and one. One of the purposes of this paper is to compare the two loss functions. In this context the following quotation is of some interest.

'*Historically stress formula one was the only badness-of-fit function used for some time. Stress formula two has been in use more recently and I now tend to recommend it.*' (Kruskal and Carroll, 1969, p. 652).

A precise mathematical comparison of the two loss functions is difficult, because both are extremely complicated functions of the multidimensional scaling configurations. We shall show that it is possible to derive some interesting inequalities, and that these inequalities shed some light on the comparison. Our results are by no means definitive and complete, in fact they merely provide some indications.

Basic Notations and Definitions

X is an $n \times p$ *configuration matrix* ($n > 2$), $D(X) = \{d_{ij}(X)\}$ is the corresponding matrix of (Euclidean) distances. The matrix $\Delta = \{\delta_{ij}\}$ contains dissimilarities. Δ is a symmetric matrix of order n, with zero diagonal, whose off-diagonal elements are the integers between one and $\binom{n}{2}$, in some order. Thus the δ_{ij} are *rank-numbers* and there are no *ties.* Finally $\hat{D} = \{\hat{d}_{ij}\}$ is a matrix of *disparities,* also symmetric, nonnegative, and hollow (zero diagonal). The *raw stress* is

$$\sigma^*(X, \Delta) = \min \{\sum_{i<j} \sum (\hat{d}_{ij} - d_{ij}(X))^2 \mid \hat{D} \text{ monotonic with } \Delta\}. \tag{1}$$

By *monotonic* we mean: if $\delta_{ij} < \delta_{k\ell}$ then $\hat{d}_{ij} \leq \hat{d}_{k\ell}$.

This study has been supported by the Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek (Netherlands Organization for the Advancement of Pure Research), under grant 56-146.

Comments by Willem Heiser and Frank Critchley have been very helpful.

The second author presently is employed by the Netherlands Central Bureau of Statistics, Voorburg.

Requests for reprints should be sent to Jan De Leeuw, Department of Datatheory, FSW/RUL, Middelstegracht 4, 2312 TW Leiden, The Netherlands.

0033-3123/84/0900-5064$00.75/0

In nonmetric multidimensional scaling it will not do to minimize raw stress over configurations X. Because raw stress is not invariant under uniform stretching and shrinking of the configuration it must be *normalized.* Kruskal proposes two plausible normalization factors

$$\tau_1^*(X) = \sum_{i<j}\sum d_{ij}^2(X), \tag{2a}$$

$$\tau_2^*(X) = \sum_{i<j}\sum (d_{ij}(X) - d_{..}(X))^2, \tag{2b}$$

with $d_{..}(X)$ the average of the $d_{ij}(X)$ with $i < j$. The next step is to define the two normalized loss functions

$$\sigma_1(X, \Delta) = \{\sigma^*(X, \Delta)\}^{1/2}/\{\tau_1^*(X)\}^{1/2}, \tag{3a}$$

$$\sigma_2(X, \Delta) = \{\sigma^*(X, \Delta)\}^{1/2}/\{\tau_2^*(X)\}^{1/2}. \tag{3b}$$

Stress formula one is defined by (3a), *stress formula two* by (3b).

> '*For any given configuration, of course, stress formula two yields a substantially larger value than stress formula one, perhaps twice as large in many cases. However, in typical multidimensional scaling applications, minimizing stress formula two typically yields very similar configurations to minimizing stress formula one.*' (Kruskal and Carroll, 1969, p. 652).

In this paper we shall try to make these statements just a little bit more precise.

Basic Results

We can formulate our main theorem by introducing some additional definitions. In the first place we need the function

$$\tau(X) = \{\tau_2^*(X)\}^{1/2}/\{\tau_1^*(X)\}^{1/2}. \tag{4}$$

As $\tau(X)$ is the ratio of the standard deviation and the root-mean-square of the distances, it is a coefficient of variation, although not *the* coefficient of variation as it is usually defined, viz. the ratio of the standard deviation and the mean. Also

$$\kappa(n, p) = \min\{\tau(X) \mid X \text{ all } n \times p \text{ matrices}\}, \tag{5}$$

and

$$\sigma_1(\Delta, n, p) = \min\{\sigma_1(X, \Delta) \mid X \text{ all } n \times p \text{ matrices}\}, \tag{6a}$$

$$\sigma_2(\Delta, n, p) = \min\{\sigma_2(X, \Delta) \mid X \text{ all } n \times p \text{ matrices}\}, \tag{6b}$$

and finally

$$\mu_1(n, p) = \max\{\sigma_1(\Delta, n, p) \mid \Delta\}, \tag{7a}$$

$$\mu_2(n, p) = \max\{\sigma_2(\Delta, n, p) \mid \Delta\}. \tag{7b}$$

Many inequalities can be derived from these definitions. The more important ones are collected in the following theorem.

Theorem 1

a: $\sigma_1(\Delta, n, p) \leq \mu_1(n, p) \leq \kappa(n, p)$,

b: $\sigma_2(\Delta, n, p) \leq \mu_2(n, p) \leq 1$,

c: $\sigma_1(X, \Delta)/\sigma_2(X, \Delta) \geq \kappa(n, p)$.

Proof. Because the disparity matrix with all $\hat{d}_{ij}$ equal to $d_{..}(X)$ is clearly monotonic

with Δ it follows that $\sigma^*(X, \underline{\Delta}) \leq \tau_2^*(X)$, i.e. that $\sigma_2(X, \Delta) \leq 1$. This immediately proves part (b) of the theorem.

On the other hand it follows from

$$\sigma_1(X, \Delta) = \tau(X)\sigma_2(X, \Delta) \tag{8}$$

that

$$\sigma_1(X, \Delta) \leq \tau(X), \tag{9}$$

and thus that

$$\sigma_1(\Delta, n, p) \leq \kappa(n, p). \tag{10}$$

This proves part (a) of the theorem.

Part (c) follows directly from (8). □

With regard to part (c) of the theorem it should be noted that the configuration X is the same in both σ_1 and σ_2. Numerical minimization of σ_1 and σ_2 on the other hand does not necessarily yield the same configuration X in both cases, although '*... minimizing stress formula two typically yields very similar configurations to minimizing stress formula one.*', according to the conclusion of the Kruskal and Carroll (1969) quotation presented above, with respect to typical multidimensional scaling applications.

Theorem 1 is quite trivial, but does have some interesting applications. If $\kappa(n, p)$, for instance, can be computed or approximated, we know if and when stress formula two exceeds stress formula one. Part (a) is also useful, because Monte Carlo studies of non-metric multidimensional scaling (reviewed below) have indicated that $\sigma_1(\Delta, n, p) \leq .60$. Moreover Lingoes and Roskam (1973) have computed $\mu_1(4, 1)$, and we can try to relate that result to our work.

In order to apply parts (a) and (c) however, we have to compute $\kappa(n, p)$ in a number of situations. Part (b) of the theorem is not informative, it is only given here for completeness. It does show the fact that we have been unable to derive an upper bound for the minimum of stress formula two. This is interesting as it follows from (8) that we can find configurations with a small value of $\sigma_1(X, \Delta)$ by making $\tau(X)$ small or by making $\sigma_2(X, \Delta)$ small (provided that making one of them small does not have the reverse effect on the other) or both. This makes it worthwhile to try to find out whether programs minimizing $\sigma_1(X, \Delta)$ are liable to produce configurations with small $\tau(X)$, and consequently to find out for what kind of configurations $\tau(X)$ is minimized.

On the other hand, (8) implies that we can find configurations with a small value of $\sigma_2(X, \Delta)$ by making $\sigma_1(X, \Delta)$ small or making $\tau(X)$ large (provided that small values of the former do not correspond to small values of the latter, or large values of the latter to large values of the former) or both. In this case it would be interesting to find out if programs that minimize $\sigma_2(X, \Delta)$ are prone to producing configurations with large $\tau(X)$, and, if they are, to find out what kind of configurations they are liable to produce. As we only have been able to derive an upper bound for the minimum of stress formula one, the next sections will be entirely devoted to this normalization. In the final section of this paper however, some marginal results for stress formula two will be presented.

Results in One Dimension

As the function $\tau(X)$ is a coefficient of variation, configurations which give small values of τ will tend to have distances which do not vary a great deal. This occurs typically if the configurations are *regularly shaped.* This idea can be used as a guideline when looking for configurations that minimize τ, or at least give a small value of τ.

Observe that

$$\sigma_1(\Delta, n, p) \leq \kappa(n, p) \leq \tau(X), \tag{11}$$

which means that $\tau(X)$ is an upper bound for $\kappa(n, p)$ for all X. It turns out that for $p = 1$, however, we can compute $\kappa(n, p)$ exactly.

Theorem 2

$$\kappa(n, p) \leq \kappa(n, 1) = \left(\frac{n-2}{3n}\right)^{1/2} \leq 1/\sqrt{3} = .5773502692\ldots$$

Proof. According to the definitions $\kappa(n, p) \leq \kappa(n, 1)$. We now compute $\kappa(n, 1)$. Without loss of generality we assume that $\sum x_i^2 = 1$ and $\sum x_i = 0$. Thus

$$\kappa(n, 1) = \min\{\tau(X) \mid X \text{ all } n \times 1 \text{ matrices standardized as above}\}. \tag{12}$$

and consequently minimizing $\tau(X)$ in one dimension amounts to the same thing as maximizing $\sum_i \sum_j |x_i - x_j|$ over all n-vectors that are centered and that satisfy $\sum x_i^2 = 1$. This term can be reformulated as

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j| = 4\sum_{i=1}^{n} x_i\{(r_i - \tfrac{1}{2}(n+1)\}, \tag{13}$$

where the r_i are the rank numbers of the x_i and $\frac{1}{2}(n+1)$ is the average of the rank numbers. This is maximized by Cauchy-Schwartz, if X is the unit length vector proportional to the centered rank numbers, i.e., if the n points are equally spaced on the line. By consequence the optimal X cannot contain ties, as in that case some of the r_i also would be equal, and the vector r would have a smaller length. Substitution of n equally spaced points gives the formula

$$\kappa(n, 1) = (1/\sqrt{3})(1 - 2/n)^{1/2}, \tag{14}$$

which clearly shows that $\kappa(n, 1)$ increases to $1/\sqrt{3}$ as $n \to \infty$. Computational details are in De Leeuw (1973a). □

Theorem 2 implies that $\sigma_1(\Delta, n, p) \leq 1/\sqrt{3}$, which explains the upper bounds found in the Monte Carlo experiments of Klahr (1969), Stenson and Knoll (1969), Wagenaar and Padmos (1971), and Spence and Ogilvie (1973). From Lingoes and Roskam (1973, table 18) it is known that $\mu_1(4, 1) = .1691$. Although this is true only if we assume that their algorithm has found the global minimum in all cases, strictly speaking we only know that $\mu_1(4, 1) \leq .1691$. A comparison with our bound $\mu_1(4, 1) \leq \kappa(4, 1) = 1/\sqrt{6} = .4082482905$. shows our bound to be a long way off here, but this may be due to the very small value of n. No information is available about $\mu_1(n, 1)$ for $n > 4$, but we expect it to get closer to $\kappa(n, 1)$ as n increases.

Monte Carlo studies may provide some evidence for the truth of this conjecture too. Results such as Klahr's (1969) give information of the form $\mu_1(n, 1) \geq \mu_1^*(n, 1)$, where μ_1^* is the so-called *maximin* stress, i.e. the largest value of $\sigma_1(\Delta, n, p)$ encountered in the study. Thus

$$\mu_1^*(n, 1) \leq \mu_1(n, 1) \leq \kappa(n, 1). \tag{15}$$

It should be emphasized however, that (15) is based on the assumption that a global minimum is found in all cases, and that consequently (15) is invalid if local minima are present. Here the discussion in Arabie (1973, 1978), Clark (1976), Spence (1974) and Spence and Young (1978) is relevant. The same major restriction holds with regard to the results of Stenson and Knoll (1969).

An Excursion

In the previous section is shown that random ranking studies can be used to compute lower bounds for $\mu_1(n, 1)$ or, more generally, for $\mu_1(n, p)$. Some of these studies are marred by an unfortunate choice of the initial configuration. Klahr's (1969) results for instance are based on the L-shaped initial configuration from the older version of MDSCAL. Other studies do not have the range of n and p we require. Therefore we have replicated parts of the random ranking studies, using the recently developed nonmetric multidimensional scaling program SMACOF (Stoop and De Leeuw, 1982). This program starts with the p-dimensional Young-Householder solution, iterates to convergence in p dimensions, rotates to principal components, uses the first $p - 1$ components as a starting point for the iterations in $p - 1$ dimensions, and so on. In this study the number of dimensions is $p = 5, 4, 3, 2, 1$ and the number of points is $n = 10, 20, 30, 40, 50$. The results for each combination of n and p are given in table 1, formatted in the same way as

TABLE 1

Monte Carlo Results

Final Values of Stress Formula One[a]

	# of Dimensions				
	1	2	3	4	5
10 points					
Mean	.3770	.1847	.1008	.0552	.0287
S.D.	.0316	.0228	.0149	.0120	.0100
Min.	.3016	.1357	.0629	.0189	.0137
Max.	.4443	.2322	.1447	.1045	.0659
20 points					
Mean	.4675	.2822	.1945	.1454	.1132
S.D.	.0110	.0055	.0034	.0019	.0014
Min.	.4397	.2585	.1710	.1269	.1002
Max.	.4928	.3089	.2100	.1602	.1294
30 points					
Mean	.5198	.3417	.2467	.1902	.1540
S,D.	.0059	.0053	.0048	.0045	.0043
Min.	.5079	.3282	.2342	.1779	.1446
Max.	.5334	.3541	.2626	.2013	.1644
40 points					
Mean	.5352	.3619	.2683	.2109	.1745
S.D.	.0041	.0040	.0035	.0033	.0029
Min.	.5237	.3512	.2577	.2028	.1673
Max.	.5432	.3812	.2801	.2204	.1813
50 points					
Mean	.5278	.3563	.2688	.2180	.1851
S.D.	.0040	.0038	.0043	.0034	.0031
Min.	.5192	.3494	.2585	.2098	.1768
Max.	.5387	.3701	.2811	.2306	.1933

[a] Mean, standard deviation, minimum and maximum of stress formula one ; for each number of points based on 100 samples of random rankings.

TABLE 2

Theoretical and Numerical Bounds when p=1

n	Upper Bound on stress one[a] $\kappa(n,1)$	Lower Bound on $\kappa(n,1)$[b] $\mu_1^*(n,1)$
10	.516	.444
20	.548	.493
30	.558	.533
40	.563	.543
50	.566	.539

a Points equally spaced on a circle
b Monte Carlo results

table 1 in Levine (1978). Each analysis is based on 100 randomly chosen samples. Note that this is a very small fraction indeed of the $\{\frac{1}{2}n(n-1)\}!/n!$ possible rankings of distances between n points. Thus the average value of the stress and its standard deviation very probably will be more reliable than the extreme values in table 1.

Comparison with table 1 of Spence and Young (1978) shows that our results are in close agreement with earlier random ranking studies. Comparison with Levine (1978) shows that stress formula two has a much larger range and variance than stress formula one. Our results also confirm another conclusion of Levine:

> *For two-dimensional solutions, the average stress values for formula two were approximately two and one-half times the corresponding average stress values of formula one. For three dimensions, the average stress values were three and one-half times the corresponding average stress values for formula one when there is no inherent data structure.'* (Levine, 1978, p. 313).

Table 2 compares $\mu_1^*(n, 1)$ with $\kappa(n, 1)$. The results are indicating that the bounds get tighter with increasing n, indeed get much more precise than suggested by the Lingoes-Roskam result for $\mu_1(4, 1)$.

Average, maximum and minimum from table 1 are also presented graphically in figure 1.

Results in Two Dimensions

The results in one dimension give an upper bound $\sigma_1(\Delta, n, p) \leq 1/\sqrt{3}$, and the indication that minimizing stress formula one will tend to produce *equal-space-prone* solutions in one dimension. We shall now try to improve that upper bound for $\sigma_1(\Delta, n, p)$ by taking the dimensionality into account, starting with $p = 2$.

Theorem 3

$$\kappa(n, 2) \leq \kappa^*(n, 2) = \underline{\left\{1 - \frac{2\,(\text{ctg}\ \Pi/2n)}{n(n-1)}\right\}^{1/2}} \leq \{(\Pi^2 - 8)/\Pi^2\}^{1/2} = .435236..$$

Proof. The upper bound for $\kappa(n, 2)$, $\kappa^*(n, 2)$, is the value of $\tau(X)$ if X consists of n points equally spaced on a circle. Details of the computation are in De Leeuw (1973a). The result is also obtainable from Fejes Tóth (1956). The upper bound on the right is obtained by letting $n \to \infty$, using $\text{ctg}\ (x/n) = (n/x) + O(n^{-2})$. Again more details are in De Leeuw (1973a). □

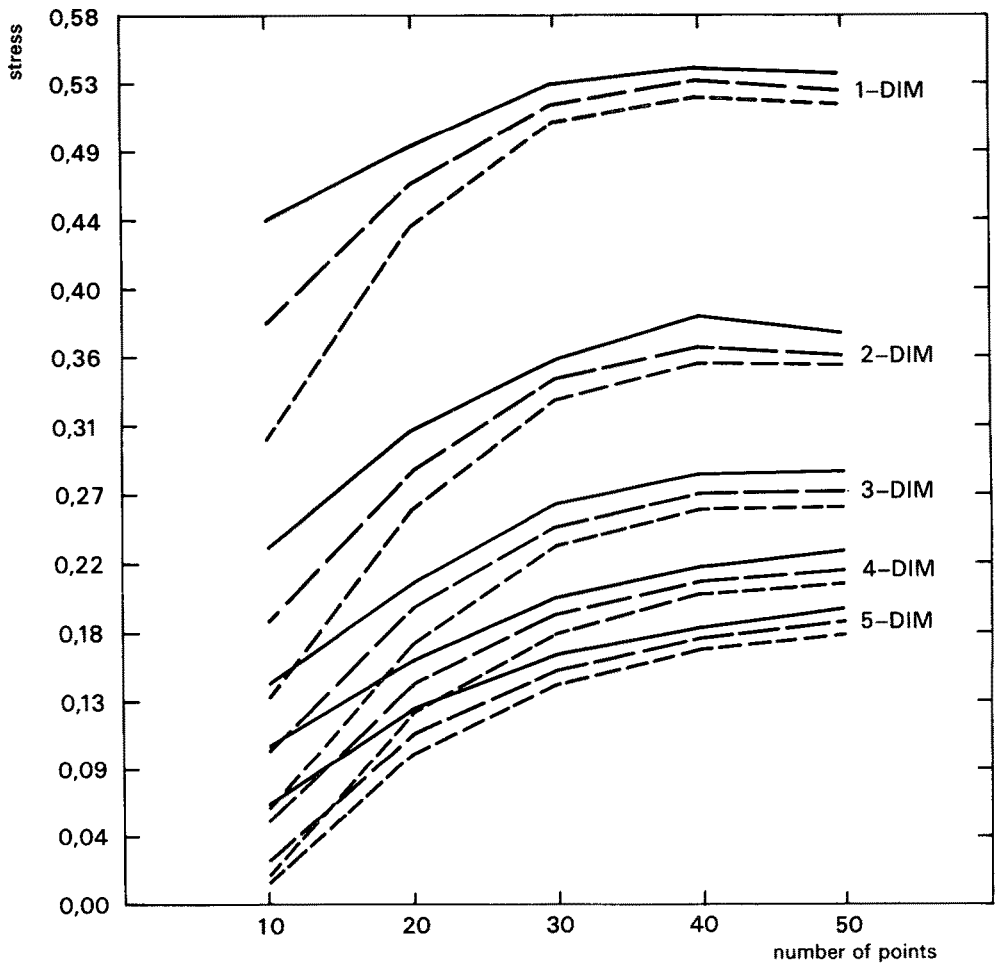

FIGURE 1
Monte Carlo results, final values of stress formula one. Maximum: ———; mean: – – –; minimum: - - - -.

The upper bound $\kappa^*(n, 2)$ in Theorem 3 is the minimum of $\tau(X)$ where all n points x_i are restricted to have length one (to be on a circle with unit radius). The formula for $\kappa^*(n, 2)$ is given by Fejes Tóth (1956). We have not proved, however, that $\kappa(n, 2) = \kappa^*(n, 2)$. In fact, numerical minimization of $\tau(X)$ indicates that $\kappa(n, 2) < \kappa^*(n, 2)$ for $n > 7$, and $\kappa(n, 2) = \kappa^*(n, 2)$ for $n \leq 7$. In this sense the results for $p = 2$ are far less satisfactory than those for $p = 1$.

We already have indicated that another way to find upper bounds for $\kappa(n, 2)$, and consequently for $\sigma_1(\Delta, n, p)$, is minimizing $\tau(X)$ numerically over the $n \times 2$ matrices X. As it happens, this can be done by means of any ordinary metric multidimensional scaling program minimizing stress formula one, if all dissimilarities are set equal. The upper bound found in this way is called $\kappa^{**}(n, 2)$. In table 3 some values of $\kappa^*(n, 2)$ and $\kappa^{**}(n, 2)$ are presented. It turns out that for $n > 7$ the configurations minimizing $\tau(X)$ are constructed by distributing n points equally spaced over two or more concentric circles. Another indication that minimizing stress formula one may tend to produce solutions liable to have distances that do not vary a great deal, especially of course if the fit is rather bad. More extensive information on $\kappa^{**}(n, 2)$ can be obtained from figure 2, from the line labeled *2*. This line interconnects the values of stress formula one for different numbers of

TABLE 3

Theoretical and Numerical Bounds when p=2

n	$\kappa^{*}(n,2)^{a}$	$\kappa^{**}(n,2)^{b}$
4	.169	.169
5	.230	.230
6	.267	.267
7	.293	.293
8	.312	.308
9	.326	.321
10	.338	.332
15	.371	.362
25	.397	.386
50	.417	.403

a Points equally spaced on a circle
b Numerical minimization of $\tau(X)$ by means of metric MDS

points, two dimensions and all dissimilarities equal. The line labeled *1* gives $\kappa^{**}(n, 1)$, and thus $\kappa(n, 1)$ for each value of n.

Figure 2 is sufficiently regular to assume that local minima are rare, and thus that it may actually represent $\kappa(n, p)$. The minor irregularities around the points $n = 2p$ will be discussed below.

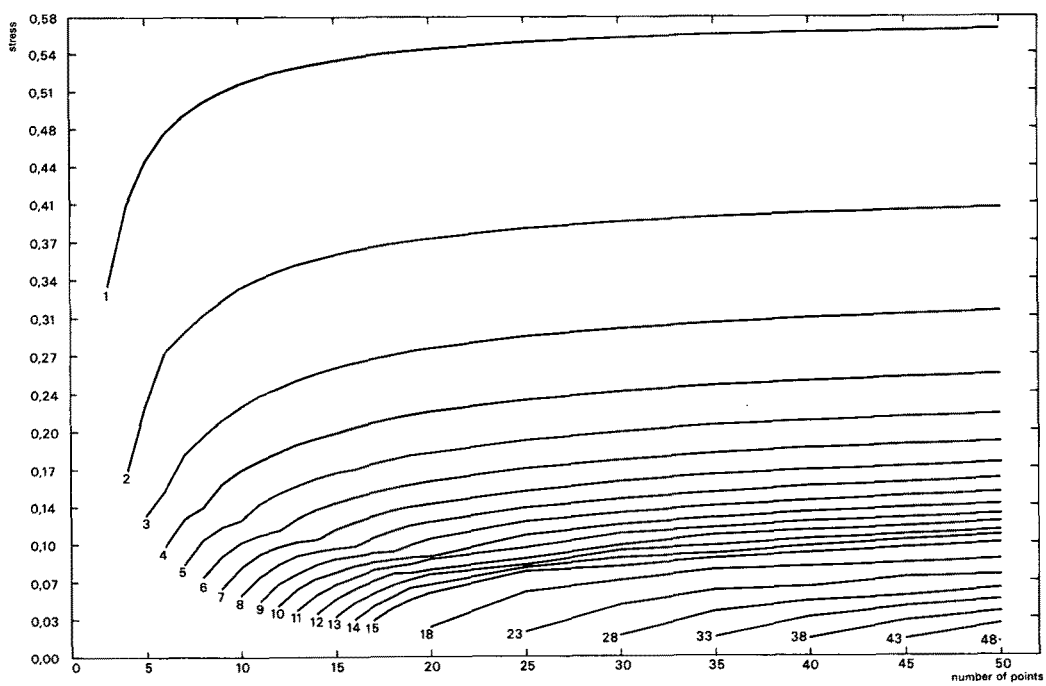

FIGURE 2

$\kappa^{**}(n, p)$ for equal dissimilarities. Values based on the same number of dimensions are interconnected; the labels of the lines refer to the dimensionality.

The information concerning $\mu_1(n, 2)$ has to be derived from Monte Carlo studies once more. We do not discuss this in detail here, as the reader can identify any desired quantity in the figure. The results are suggesting that for large n $\kappa^*(n, 2)$ and $\kappa^{**}(n, 2)$ do not differ a great deal. Because

$$\mu_1^*(n, 2) \leq \mu_1(n, 2) \leq \kappa(n, 2) \leq \kappa^{**}(n, 2) \leq \kappa^*(n, 2), \tag{16}$$

this implies that there can not be much difference between $\mu_1(n, 2)$ and $\kappa(n, 2)$.

A final conjecture is that $\kappa(n, 2)$ increases to .424548, as $n \to \infty$. This value is obtained by computing the probability limit of $\tau(X)$ if points x_i are sampled randomly from the interior of a circle, and if the sample size tends to infinity. The necessary computations, and a weak law of large numbers that can be applied here, are in De Leuuw (1973b). It is also shown there that sampling from a circular binormal distribution gives $\kappa(n, 2) \leq .463251$, and sampling from a circle gives $\kappa(n, 2) \leq .435236$, which conforms with the result of theorem 3.

Results in Three and More Dimensions

In this section we briefly summarize the results we have in more than two dimensions. They are even more fragmentary, because even $\kappa^*(n, 3)$, the minimum of $\tau(X)$ for n points on the unit sphere, can not be computed in closed form any more. It is nevertheless of some interest to show that most of the methods we have used in the previous section also work to some extent if $p \geq 3$.

In the first place $\tau(X)$ can be minimized numerically. This gives $\kappa^{**}(n, p)$, plotted for various values of n and p in figure 2. As there are hardly any irregularities apart from the systematic indentations around the points $n = 2p$, the obtained values of the stress seem to indicate global minima. A possible tentative explanation of the small *irregularities*, the indentations, may be found in De Leeuw (1973a). Here $\tau(X)$ is computed for the vertices of the classical regular polytopes in $p = 3$ and $p = 4$ dimensions. He also computes $\tau(X)$ for the vertices of the p-dimensional octahedrons. This gives an upper bound for $\kappa(2p, p)$. And finally computing $\tau(X)$ for the vertices of the unit hypercube in p dimensions gives an upper bound for $\kappa(2^p, p)$. It might very well be true that the octahedron causes the dips in figure 2 at $n = 2p$. In general it seems to be true that other small dips can be found at other very regular figures. Thus for $p = 3$ and $n = 12$ another small dip is found (the isocahedron).

In the second place, if $p \geq 3$, we can study $\kappa^*(n, p)$, the minimum of $\tau(X)$ for n points on the surface of a unit hypersphere in p dimensions. Recently this quantity has been studied a great deal. We give some of the key references. The field is reviewed and placed in context by Alexander and Stolarsky (1973). Sharp inequalities bounding $\kappa^*(n, p)$ are derived for $p = 3$ by Alexander (1970), and for general p by Stolarsky (1972, 1973). Berman and Hanes (1977) discuss an algorithm to compute $\kappa^*(n, p)$ numerically, and use it to determine tentative values of $\kappa^*(n, 3)$ for $n = 5, 6, 7, 8, 9, 10$.

Although these mathematical results are interesting indeed, they add very little practically useful information to figure 2. Information that would have practical significance, like the limit of $\kappa(n, p)$ for fixed p if $n \to \infty$, is difficult to obtain from this figure. Here the probabilistic methods of De Leeuw (1973b) come in handy. His results lead to the conjecture that $\kappa(\infty, 3) = .33333..$, the value obtained by sampling from the surface of a unit hypersphere. This is also indicated by the results of Alexander and Stolarsky on $\kappa^*(n, 3)$. Other upper bounds (and conjectured values) for $\kappa(\infty, p)$ can be found in De Leeuw (1973b). There the unit hypercube and the p-dimensional octahedron illustrate what happens if both n and p tend to infinity simultaneously.

Conclusion

We have shown that the minimum value of stress formula one never exceeds $1/\sqrt{3}$, and we have given many results improving on this upper bound by taking into account the values of n and p. It is also argued that our developments lead to the conjecture that minimizing stress formula one will lead to configurations with a liability to equal spacing or regularity, certainly if the fit is bad. If $p = 2$, this may lead to points regularly spaced on a circle, or a small number of concentric circles. If the configuration is essentially one-dimensional, on the other hand, this liability will tend to bend this configuration into a hoop. In this way the order of the points is recovered along the hoop, whereas the distances will vary less than those of the corresponding one-dimensional configuration.

This by no means implies that stress formula two is to be preferred to stress formula one. Not being able to find a nontrivial system of upper bounds for stress formula two does not prove that no such system does exist. The liability of stress formula one to produce some kind of configurations does not prove that stress formula two is not prone to deliver a (different) special kind of configurations. It is not difficult indeed to make part (b) of theorem 1 somewhat less trivial. Suppose δ_{12} is the largest dissimilarity. In the one-dimensional configuration X ($x_1 = 0$, $x_2 = \cdots = x_n = 1$) it is equally true that d_{12} is the largest distance, which implies that $\hat{d}_{12} = d_{12}$. In the worst possible case all other distances are averaged to form the remaining disparities, all equal to this average. Making the necessary substitutions proves that

$$\mu_2(n, p) \leq \left\{\frac{n}{n+1}\right\}^{1/2}. \tag{17}$$

This result, which could possibly be extended by taking p into account, can be used in the same way as above to argue that stress formula two will tend towards configurations consisting of clumps, i.e. towards configurations with a large value of $\tau(X)$, as follows from (8). Also, the variance is largest relative to the mean when the distribution is bimodal, which in multidimensional scaling means that the points are in two clumps. This, however, should be investigated much more closely.

Our computation of $\kappa(n, p)$ also gives information, by Theorem 1c, on the size of the ratio of stress formula two to stress formula one which is somewhat more precise than previously available information. It is still in the form of an equality however. More important is the fact that the configuration is the same in both stress formula one and stress formula two, while we have argued that each stress formula is prone to produce different kinds of configurations. This tendency will be most extreme if the fit is bad. On the other hand, if the fit is not bad, or more generally '... *in typical multidimensional scaling situations*...' (referring once more to the quotation of Kruskal and Carroll (1969)), both stress formula yield very similar configurations. In this case the idiosyncratic tendencies of each stress formula will be less outspoken, whereas the ratio of stress formula one to stress formula two given here, is of much more practical significance if the fit is bad.

Kruskal's stress measures badness-of-fit by computing the sum of squared differences between dissimilarities and distances. The most important alternative loss function probably is *sstress*, proposed by Takane, Young and De Leeuw (1977), which is defined entirely analogously, except for the fact that it incorporates the squared distances in the definition of raw stress. Thus there is also sstress formula one and sstress formula two, and—not surprisingly—the entire development in this paper can be repeated for sstress. As it turns out, $\lambda(n, p)$, the analogue of $\kappa(n, p)$ for sstress, even is fairly easy to compute

(De Leeuw and Heiser, 1982). As a matter of fact

$$\lambda(n, p) = \left\{1 - \frac{n}{n-1}\frac{p}{p+1}\right\}^{1/2}. \tag{18}$$

This value is attained for clumps each consisting of the same number of points, and distributed regularly on the surface of a unit hypersphere. A precise proof will be published elsewhere.

By (18) we have another illustration of the general fact that the mathematics connected with sstress is simpler than that connected with stress. On the other hand, the conditions of equality in (18), especially the *degeneration* into clumps, may be less attractive. A liability of sstress formula one to produce configurations for which $\theta(X)$, the analogue of $\tau(X)$ for sstress, is minimized, relative to $\sigma\sigma_1(X, \Delta)$ (the analogue of $\sigma_1(X, \Delta)$), may have a more serious effect than the particular liability of stress formula one, because of the clumps associated with the minimum of $\theta(X)$. From a practical point of view sstress (or ALSCAL) solutions thus may tend to be more crude and less detailed than stress (or KYST, MDSCAL, MINISSA, SMACOF) solutions.

REFERENCES

Alexander, R. (1970). On the sum of distances between n points on a sphere. *Acta Math. Acad. Sci. Hungar., 23*, 443–448.

Alexander, R. & Stolarsky, K. B. (1973). Extremal problems in distance geometry related to energy integrals. *Trans. Amer. Math. Soc., 193*, 1–31.

Arabie, P. (1973). Concerning Monte Carlo evaluations of nonmetric multidimensional scaling algorithms. *Psychometrika, 38*, 607–608.

Arabie, P. (1978). The difference between 'several' and 'single': a reply to Spence and Young. *Psychometrika, 43*, 119.

Berman, J. & Hanes, K. (1977). Optimizing the arrangement of points on the unit sphere. *Math. of Computation, 31*, 1006–1008.

Clark, A. K. (1976). Re-evaluation of Monte Carlo studies in nonmetric multidimensional scaling. *Psychometrika, 41*, 401–404.

De Leeuw, J. (1973a). Bounds for SFORM1. Part I. Department of Data Theory, University of Leiden, Research Bulletin RB-001-73.

De Leeuw, J. (1973b). Bounds for SFORM1. Part II. Department of Data Theory, University of Leiden, Research Bulletin RB-004-73.

De Leeuw, J. & Heiser, W. (1982). Theory of multidimensional scaling. In P. R. Krishnaiah & L. Kanal (eds): *Handbook of Statistics, volume II*. Amsterdam, North Holland Publishing Co., 285–316.

Fejes Tóth, L. (1956). On the sum of distances determined by a point set. *Acta. Math. Acad. Sci. Hungar., 7*, 397–401.

Klahr, D. A. (1969). Monte Carlo investigation of the statistical significance of Kruskal's scaling procedure. *Psychometrika, 34*, 319–330.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness-of-fit to a non-metric hypothesis. *Psychometrika, 29*, 1–27.

Kruskal, J. B. (1964b). Non-metric multidimensional scaling: a numerical method. *Psychometrika, 29*, 115–129.

Kruskal, J. B. & Carroll, J. D. (1969). Geometric models and badness-of-fit functions. In P. R. Krishnaiah (ed): *Multivariate Analysis II*. New York, Academic Press.

Levine, D. M. (1978). A Monte Carlo study of Kruskal's variance based measure on stress. *Psychometrika, 43*, 307–315.

Lingoes, J. C. & Roskam, E. E. (1973). A mathematical and empirical analysis of two multidimensional scaling algorithms. Psychometric Monograph Supplement, *38* (1, Pt. 2).

Spence, I. (1974). On random rankings studies in nonmetric scaling. *Psychometrika, 39*, 267–268.

Spence, I. & Ogilvie, J. C. (1973). A table of expected stress values for random rankings in nonmetric multidimensional scaling. *Multivariate Behavioural Research, 8*, 511–517.

Spence, I. & Young, F. W. (1978). Monte Carlo studies in nonmetric scaling. *Psychometrika, 43*, 115–117.

Stenson, H. H. & Knoll, R. L. (1969). Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychol. Bulletin, 72*, 122–126.

Stolarsky, K. B. (1972). Sums of distances between points on a sphere. *Proc. Amer. Math. Soc., 35*, 547–549.
Stolarsky, K. B. (1973). Sums of distances between points on a sphere. *Proc. Amer. Math. Soc., 41*, 575–582.
Stoop, I. & De Leeuw, J. (1982). How to use SMACOF-1B. Department of Data Theory, University of Leiden.
Takane, Y., Young, F. W. & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika, 42*, 7–68.
Wagenaar, W. A. & Padmos, P. (1971). Quantitative interpretation of stress in Kruskal's multidimensional scaling technique. British J. Math. Statist. Psychol., *24*, 101–110.

Manuscript received 1/25/83
Final version received 5/15/84