Journal of Classification 3:97-112 (1986)

Journal of
Classification

A Special Jackknife for Multidimensional Scaling

Jan de Leeuw
Rijksuniversiteit te Leiden

Jacqueline Meulman
Rijksuniversiteit te Leiden

Abstract: In this paper we develop a version of the Jackknife which seems especially suited for Multidimensional Scaling. It deletes one stimulus at a time, and combines the resulting solutions by a least squares matching method. The results can be used for stability analysis, and for purposes of cross validation.

Keywords: Multidimensional scaling; Jackknife; Cross validation; Stability.

1. Introduction

The results of a data analysis are not complete without some form of information about the *stability* of the solution. There are many types of stability that can be distinguished, and consequently there are many forms of stability analysis (or sensitivity analysis). A full discussion of the various forms of stability that can be studied is contained in Gifi (1981, Chapter 1). Most of these techniques have in common that they study the effect of a small perturbation of the data on the solution, or aspects of the solution. A method is stable, in a particular application, if small changes in the data produce only small changes in the solution. It is clear that statistical analysis, in the form of computation of standard errors or confidence intervals for example, is a particular kind of sensitivity analysis, which studies the *perturbations* induced by random sampling.

In most applications of Multidimensional Scaling (MDS) the stability of the solutions is not studied at all. There are exceptions, however. The method of Maximum Likelihood (ML) has been applied to MDS, for instance by Ramsay (1982). At least in some cases standard large sample theory for ML estimation can be applied to obtain information about statistical stability. It is true, however, that the optimality properties of ML

Authors' Addresses: Jan de Leeuw and Jacqueline Meulman, Department of Data Theory FSW/RUL, Rijksuniversiteit te Leiden, Middelstegracht 4, 2312 TW Leiden, Netherlands.

estimation often do not apply directly to MDS situations. This problem results from non-standard aspects of such situations, which are riddled with incidence parameters, dependencies, lack of replications, and so on (c.f. the discussion of Ramsay [1982]).

If there are independent replications in a MDS experiment, then statistical information can be obtained even without assuming parametric distributional models or specific response functions. A demonstration is given by Stoop, Heiser, and de Leeuw (1981). One can use, in this case, the standard delta method results (Rao 1973, section 6a.2). One could also apply the currently popular resampling techniques such as the Jackknife or the Bootstrap (Efron 1982). Heiser and Meulman (1983a, 1983b) and Weinberg, Carroll, and Cohen (1984) review the use of resampling in MDS.

If there are no replications, matters become more complicated. We need a parametric model to derive statistical information on stability, but very often the prior knowledge required to make the necessary assumptions with some confidence is not available. And there are other situations in which the whole idea of random sampling does not make much sense, and in which studying stability in the usual statistical way is consequently not interesting, in fact not even defined. Nevertheless it is still possible, in such cases, to perform other forms of stability analysis. The idea of a small change in the data can still be defined in various ways, and the effects of such small perturbations can still be studied. This possibility was already indicated, in the MDS context, by Kruskal and Wish (1978, pages 58-60). Our contribution makes their suggestions explicit, implements and applies them.

A first suggestion, and a rather natural one, is to make a small change in one dissimilarity in a metric MDS problem. The implicit function theorem can be used to compute the effect of such a small change on the loss function and the configuration computed by the MDS algorithm. For classical metric scaling, also known as Torgerson-Gower scaling (Torgerson, 1958; Gower (1966) or Young-Householder scaling (Young and Householder 1938) some of the necessary results have already been given by Sibson (1979). The familiar "discrete" version of the Jackknife, reviewed by Gray and Schucany (1972) and Miller (1974), corresponds with deleting one dissimilarity and treating it as missing. We then reanalyze the data with this dissimilarity omitted, and combine the $\binom{n}{2}$ different solutions obtained by omitting each dissimilarity successively. This scheme can also be applied to nonmetric MDS, and may be quite useful. For a large number of objects, however, we would have to perform many MDS analyses, and the perturbations by leaving out one dissimilarity will be very small indeed. Observe that we are using the term Jackknife in a very broad sense, for a class of techniques that studies stability by deleting part of the data, or various parts of the data in systematic ways. If one wants to restrict the use of the

Jackknife to situations with independent observations, then one should use another name for our procedure.

In this paper we study another data analytical version of the Jackknife to investigate the effect of somewhat larger perturbations of the data. *We emphasize, from the start, that our results are not supposed to have inferential applications.* We do not use them to assess statistical stability in some sense or another. Our purposes are purely data analytical. The starting point is that we ask ourselves what happens to our multidimensional scaling solution if we delete all information on one object from the set of objects for which a solution is sought. "Likewise stimuli could be eliminated from the data matrix, and solutions determined for the remaining stimuli using the 'Jackknife' idea of J.W. Tukey." (Kruskal and Wish 1978, page 59). Instead of analyzing the dissimilarities between n stimuli once, we analyze, in addition, n times the dissimilarities between $n-1$ stimuli, by deleting each one in turn. The stability question associated with this scheme is interesting in its own right, and the results of the $n+1$ analyses can be portrayed quite nicely in a single plot (see section 4.2). Thus our technique is well suited for data analysis. It is clear that it can be applied to both metric and nonmetric MDS. We shall indicate further that it can also be combined with cross-validation types of assessment, along the lines indicated by Stone (1974), although a great deal of further study seems required to find out what its properties are in that context. We shall investigate, tentatively, if our leave-one-out method can be used to determine a "correct" or "optimal" dimensionality. Previous procedures for choosing dimensionality are based on large scale Monte Carlo studies, and were reviewed recently by Spence (1983). Monte Carlo studies of robustness of MDS and of dimension estimation differ from our scheme, because Monte Carlo methods are used in situations in which the "correct" answer to the question that is studied is known beforehand.

2. Technique

The basic idea behind our Jackknife is simple. We perform $n+1$ multidimensional scaling analyses. The solution to the original problem is $\mathbf{X}_0$, the solutions to the n additional problems are $\mathbf{X}_i$ $(i = 1, \ldots, n)$, where $\mathbf{X}_i$ is the solution to the problem without object i. Thus $\mathbf{X}_0$ is a $n \times p$ configuration matrix, while the $\mathbf{X}_i$ only have information on the location of $n-1$ points. For convenience we also locate them in $n \times p$ matrices, where row i of $\mathbf{X}_i$ is always zero. Of course these zeroes are "structural" zeroes, which must be distinguished from other zeroes which may occur in the matrix. All $\mathbf{X}_i$ can be computed with an ordinary multidimensional scaling program, both in metric and nonmetric cases. The choice of initial configuration for the n subset analyses is rather important in this context. In our implementation we have used a *separate* rational starting

configuration for each subset problem. Of course if we intend to use this form of the Jackknife in a routine matter, then it will be convenient to build appropriate modifications into the software.

The major problem in comparing the $\mathbf{X}_i$ is, of course, that information on the location of one point is missing from each of them. In addition comparisons are complicated because of the familiar translational and rotational indeterminacy of scaling solutions. We propose to do the comparison by optimally *matching* the configurations $\mathbf{X}_i$ $(i = 1, \ldots, n)$ with respect to translation, rotation, uniform stretching or shrinking, and estimation of the location of the missing point.

Define $\mathbf{Y}_i = a_i \mathbf{X}_i \mathbf{K}_i + \mathbf{e}_i \mathbf{b}'_i + \mathbf{e} \mathbf{c}'_i$, with $\mathbf{K}_i$ an orthogonal rotation, $\mathbf{c}_i$ a translation vector, $\mathbf{b}_i$ the location of the missing point, and a_i a scalar for uniform stretching or shrinking. Vector $\mathbf{e}_i$ is the i-th unit vector, and vector $\mathbf{e}$ has all elements equal to unity. In metric scaling it is often better to set $a_i = 1$ for all i, but in nonmetric scaling we usually want to compute the optimal a_i. We perform matching by minimizing the least squares loss function.

$$\sigma(a_i, \mathbf{b}_i, \mathbf{c}_i, \mathbf{K}_i, \mathbf{Y}_0) = \sum_{i=1}^{n} tr\ (\mathbf{Y}_0 - \mathbf{Y}_i)'(\mathbf{Y}_0 - \mathbf{Y}_i) \tag{1}$$

over all unknowns. The $n \times p$ comparison matrix $\mathbf{Y}_0$ is also unknown. Minimizing (1) means that we are fitting $1/2\ n(p^2 + 5p + 2)$ parameters to $np(n-1)$ known elements of the $\mathbf{X}_i$. This parameter count applies if the a_i are unknown, if they are given there are only $1/2 n(p^2 + 5p)$ parameters. For a nontrivial solution we want the number of known elements to be larger than the number of free parameters. This gives $n > (p^2 + 7p + A) / (2p)$, with $A = 2$ if the a_i are fitted, and with $A = 0$ otherwise. For $p = 2$ it suffices that $n > 5$, for $p = 3$ we need $n > 6$. Thus, in typical multidimensional scaling situations, there is no real danger of overfitting as long as n is much larger than p.

We discuss some normalization conventions. In the first place the data matrices $\mathbf{X}_i$ are supposed to be centered columnwise. If we solve for the a_i we suppose in addition that $tr\ \mathbf{X}'_i \mathbf{X}_i = 1$. This condition merely implies a particular scaling of the data, it has no influence on the solution. For identification purposes we also require that the columns of $\mathbf{Y}_0$ sum to zero. Again this constraint causes no loss of generality. If we solve for the optimal a_i we need an additional constraint to prevent degeneracy. We use $\Sigma\ a_i^2 = 1$. We minimize (1) in three steps.

In the first step we minimize over $\mathbf{b}_i$ and $\mathbf{c}_i$, for fixed a_i, $\mathbf{K}_i$, and $\mathbf{Y}_0$. The optimal $\mathbf{c}_i$ and $\mathbf{b}_i$ are

$$\hat{\mathbf{c}}_i = -\mathbf{y}_i^0 / (n-1) \quad , \tag{2a}$$

$$\hat{\mathbf{b}}_i = n\mathbf{y}_i^0 / (n-1) \quad . \tag{2b}$$

Here $\mathbf{y}_i^0$ is row i of $\mathbf{y}_0$. If we substitute (2) in (1) we find the "partial minimum"

$$\sigma(a_i, *, *, \mathbf{K}_i, \mathbf{Y}_0) = \sum_{i=1}^{n} tr(\mathbf{Y}_0 - a_i\mathbf{X}_i\mathbf{K}_i)'(\mathbf{Y}_0 - a_i\mathbf{X}_i\mathbf{K}_i)$$

$$- \frac{n}{n-1} tr\ \mathbf{Y}'_0\mathbf{Y}_0 \quad . \tag{3}$$

Arguments over which we have minimized are replaced by an asterisk. In the second step we minimize (3) over $\mathbf{Y}_0$, for fixed a_i and $\mathbf{K}_i$. The solution is

$$\hat{\mathbf{Y}}_0 = \frac{n-1}{n(n-2)} \sum_{i=1}^{n} a_i\ \mathbf{X}_i\ \mathbf{K}_i \quad . \tag{4}$$

The new partial minimum is

$$\sigma(a_i, *, *, \mathbf{K}_i, *) = \sum_{i=1}^{n} a_i^2\ tr\ (\mathbf{X}'_i\mathbf{X}_i)$$

$$- \frac{n-1}{n(n-2)} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j\ tr(\mathbf{K}'_i\mathbf{X}'_i\mathbf{X}_j\mathbf{K}_j) \quad . \tag{5}$$

If the a_i are fixed, as in metric scaling, then all that remains to be done is maximization of $\Sigma\Sigma\ tr(\mathbf{K}'_i\mathbf{X}'_i\mathbf{X}_j\mathbf{K}_j)$ over the $\mathbf{K}_i$. If the a_i are free, then we must maximize $\Sigma\Sigma\ a_i a_j\ tr(\mathbf{K}'_i\mathbf{X}'_i\mathbf{X}_j\mathbf{K}_j)$ over the $\mathbf{K}_i$ and over the a_i. This is exactly the problem considered by Gower (1975) and Ten Berge (1977a, 1977b), who discuss simple alternating least squares algorithms to solve this problem and find the optimal $\mathbf{K}_i$ (and a_i). Together with (2) and (4) this procedure defines $\mathbf{Y}_i$, the optimally matched and completed configuration. Finally we rotate $\mathbf{X}_0$ in such a way that $\mathbf{X}_0$ and $\mathbf{Y}_0$ are matched optimally, which again facilitates comparisons between our different solutions.

3. How to Use the Results

After matching we have $n + 2$ solutions that can be compared: we have our original $\mathbf{X}_0$, we have the n matched solutions $\mathbf{Y}_i$ $(i = 1, \ldots, n)$, and their average $\mathbf{Y}_0$. Our first assessment of stability is obtained by comparing the $\mathbf{Y}_i$. This step can be done graphically, by plotting all $\mathbf{Y}_i$ as the endpoints of stars with centers given by $\mathbf{Y}_0$. This approach will be illustrated in section 4. In addition to the graphical comparisons, we can compute dispersions of the rows of $\mathbf{Y}_i$ around their centroids $\mathbf{Y}_0$, or around the original solution $\mathbf{X}_0$. In the illustrations in section 4 we will use the following measure:

$$\text{STAB} : 1 - \frac{\sum_{i=1}^{n} || \mathbf{Y}_i - \mathbf{Y}_0 ||^2}{\sum_{i=1}^{n} || \mathbf{Y}_i ||^2}$$

This stability measure can be interpreted as the ratio of Between to Total variance (cf. Heiser and Meulman 1983b).

For cross validation purposes we can investigate if the "predicted" position of object i, which is row i of $\mathbf{Y}_i$, corresponds with the "actual" position of object i, which is row i of $\mathbf{X}_0$. It follows from (2a) and (2b) that row i of $\mathbf{Y}_i$ is equal to row i of $\mathbf{Y}_0$. Thus each row of $\mathbf{Y}_0$ comes from a different Multidimensional Scaling analysis, although all rows are only known after completing the matching analysis. Similarity between $\mathbf{Y}_0$ and $\mathbf{X}_0$, established either numerically or graphically, can consequently also be used for cross-validation. It is possible that, in some sense, the difference between these two configurations can be used to correct for bias (as in the ordinary Jackknife). We do not have any theory yet which would indicate how such a combination can be made. To measure cross-validity we will use

$$\text{CROSS} : 1 - \frac{n \; || \mathbf{X}_0 - \mathbf{Y}_0 ||^2}{\sum_{i=1}^{n} || \mathbf{Y}_i ||^2}$$

By normalizing STAB and CROSS in this manner, solutions are comparable across methods, and the dispersion around the original solution $\mathbf{X}_0$ (DISP) can easily be derived. Since

$$\frac{1}{n}\sum_{i=1}^{n} || \mathbf{Y}_i - \mathbf{X}_0 ||^2 = \frac{1}{n}\sum_{i=1}^{n} || \mathbf{Y}_i - \mathbf{Y}_0{}^2 || + || \mathbf{X}_O - \mathbf{Y}_0 ||^2 \ ,$$

it follows that DISP = 2 − (STAB + CROSS).

4. Illustrations

In the following sections our version of the Jackknife will be illustrated by various analyses of two different sets of data. For the first set of data the technique is used to determine the dimensionality. Analyses of the second set of data address three separate issues: the additive constant problem, the choice of dimensionality, and the comparison of three major MDS loss functions. Throughout both sections 6 different approaches to MDS have been applied, which are summarized in Table 1. To avoid ambiguous terminology we will denote them with acronyms in terms of the loss function that is minimized by them.

Table 1 uses the following notation:

δ_{ij}	dissimilarity between object i and object j
$\mathbf{X}$	configuration in p-dimensional space
d_{ij}	distance between object i and object j in configuration $\mathbf{X}$
$\mathbf{B}$	− 1/2 double centered matrix of squared dissimilarities
α	additive constant
$\hat{d}$	optimally transformed dissimilarities by a monotone function

- LSSP is least squares on the *scalar products*, it is a *metric* method and also known as classical, or Torgerson-Gower MDS;
- LSD is least squares on the *distances*, it minimizes Kruskal's (1964) STRESS for the *metric* case;
- LSDA is like LSD, with in addition an additive constant estimated;
- LSDN is least squares on the *distances, nonmetric*, and also known as Kruskal-Shepard MDS (Shepard 1962, Kruskal 1964).
- LSDR is a special version of LSDN. It denotes that the monotone regression is performed across the separate rows of the data matrix.
- LSSD is least squares on the *squared* distances. The metric version has been applied (see below);
- LSLD is least squares on the *logarithms* of the distances (see below).

LSD, LSDA, LSDN, and LSDR have been minimized by means of the SMACOF algorithm (see e.g., de Leeuw and Heiser 1980). For LSSD the ALSCAL program has been used (see e.g., Takane, Young and de Leeuw

TABLE 1.

MDS Methods that have been applied.

1. LSSP $\sigma^2(X) = \text{tr}\,(B\text{-}XX')^2$

2. LSD $\sigma^2(X) = \sum_i \sum_j (\delta_{ij} - d_{ij}(X))^2$

3. LSDA $\sigma^2(X;\alpha) = \sum_i \sum_j ((\delta_{ij}+\alpha) - d_{ij}(X))^2$

4. LSDN $\sigma^2(X;\hat{d}) = \sum_i \sum_j (\hat{d}_{ij} - d_{ij}(X))^2$

5. LSSD $\sigma^2(X) = \sum_i \sum_j (\delta^2_{ij} - d^2_{ij}(X))^2$

6. LSLD $\sigma^2(X) = \sum_i \sum_j (\log \delta_{ij} - \log d_{ij}(X))^2$

1977), and for LSLD MULTISCALE has been applied (see e.g., Ramsay 1977). LSSP is performed by the computation of the standard initial configuration in the SMACOF program.

4.1 Illustration 1: Color Data

The data for illustration 1 have been taken from Torgerson (1958), and concern dissimilarities between 9 Munsell colors all of the same red hue but differing from each other in brightness and saturation. The judgments were obtained from 38 subjects. These data have been used in two different versions: an asymmetric data matrix, and a symmetric one with an additive constant eliminated (Torgerson 1958, Table 2 on page 285, and Table 7 on page 287, respectively). The symmetric data have been analyzed by LSSP, LSD, and LSDN, while LSDR has been applied to the asymmetric table.

The Jackknife has been performed for this example to investigate whether Torgerson's conclusions about the dimensionality of the data could be sustained. Table 2 summarizes results for STAB and CROSS for the different methods with varying dimensionality. Within the same dimensionality results for STAB and CROSS are almost perfectly consistent, while across dimensions the performances of the various methods slightly differ.

One-dimensional scaling problems are qualitatively different from higher-dimensional ones, because the one-dimensional problems are basically combinatorial optimization problems (Hubert and Arabie 1986).

TABLE 2.

Jackknife Results for the Color Data.

METHOD	STAB			CROSS		
	1-dim	2-dim	3-dim	1-dim	2-dim	3-dim
LSSP	.8596	.9994	.9951	.9801	1.000	.9994
LSD	.8423	.9994	.9974	.9678	1.000	.9999
LSDN	.8600	.9988	.9979	.9829	.9998	.9997
LSDR	.7721	.9991	.9960	.9400	.9999	.9992

Accordingly it does not come as a surprise to us that the one-dimensional solutions are clearly discredited by STAB and CROSS results. The values for STAB and CROSS in two and three dimensions look very much alike, both being very high. If we had normalized STAB and CROSS differently, however, the discrepancies would turn out larger. Since in addition the results are consistent across MDS-approaches, we conclude on the basis of the third digit of STAB and CROSS values that, apart from the global argument of parsimony, the two-dimensional solutions are preferred. STAB and CROSS confirm Torgerson's conclusion, which he had drawn using a completely different criterion while performing LSSP. One could have suspected that values of STAB are related to the values of STRESS for the various approaches. Since the metric methods LSSP and LSD are more *restricted*, and accordingly have higher stress, than the nonmetric methods LSDN and LSDR, the former could be expected to be more stable. For none of the dimensionalities this speculation is sustained.

4.2 Illustration 2: Nations Data

The second illustration concerns data from Kruskal and Wish (1978, page 31). The degree of overall similarity between 12 nations was rated by 18 students on a scale from 1 (very different) to 9 (very similar) and was averaged to obtain a mean similarity matrix. Since MDS methods approximate dissimilarities rather than similarities, the data were transformed by subtracting the mean rating from 9, the largest value of the scale. This transformation resulted in a dissimilarity matrix labeled $\mathbf{D}^+$. This procedure immediately suggests the first application of the Jackknife, which is to investigate whether we are dealing with an additive constant problem contained in the transformed data. Various metric analyses of $\mathbf{D}^+$ have been performed.

TABLE 3.

Jackknife Results for the Nations Data: Additive Constant Problem.

METHOD	STAB	CROSS
	2-dim	2-dim
LSSP ($\mathbf{D}^+$)	.9814	.9993
LSSP ($\mathbf{D}^-$)	.9812	.9993
LSD ($\mathbf{D}^+$)	.9633	.9926
LSD ($\mathbf{D}^-$)	.9721	.9959
LSDA ($\mathbf{D}^+$)	.9728	.9933

In addition they have been applied to $\mathbf{D}^-$, a dissimilarity matrix with an additive constant eliminated. This estimation was done by applying the Messick and Abelson (1956) procedure, as described in Torgerson (1958, pages 273-276). To complete the results $\mathbf{D}^+$ has been analyzed by minimizing LSDA, i.e., by iterative estimation of the additive constant.

Bearing in mind the remarks about noticeable differences stated above, it is clear from Table 3 that LSD suffers most from the additive constant problem. Both STAB and CROSS increase when the additive constant is eliminated, either by analyzing $\mathbf{D}^-$ or by analyzing $\mathbf{D}^+$ and estimating the additive constant iteratively. LSSP, which involves the double-centering of the squared dissimilarity matrix, is not much affected by the additive constant.

The second application of the Jackknife for the nations data again concerns the dimensionality. Kruskal and Wish remark that the data might be four or even five dimensional (Kruskal and Wish 1978, page 56). To see if this speculation is correct, LSSP and LSD have been applied to $\mathbf{D}^-$, and LSDA and LSDN to $\mathbf{D}^+$, with varying dimensionality. Results are given in Table 4.

The most important conclusion that can be drawn from Table 4 is that the three-dimensional solution has to be preferred, according to both STAB and CROSS for all methods. Furthermore, LSSP is most stable in three dimensions, compared with the other methods, but this statement is not true for the four-dimensional solution, where LSD performs best. These findings are consistent with the results in Table 2 for the color data in two

TABLE 4.

Jackknife Results for the Nations Data:
Dimensionality Problem.

METHOD	STAB			CROSS		
	2-dim	3dim	4dim	2-dim	3dim	4-dim
LSSP ($\mathbf{D}^{-}$)	.9812	.9891	.9630	.9993	.9997	.9974
LSD ($\mathbf{D}^{-}$)	.9721	.9880	.9809	.9959	.9996	.9995
LSDA ($\mathbf{D}^{+}$)	.9728	.9808	.9640	.9933	.9980	.9970
LSDN ($\mathbf{D}^{+}$)	.9564	.9585	.9533	.9892	.9905	.9870

TABLE 5.

Jackknife Results for the Nations Data:
three major loss functions.

METHOD	STAB	CROSS
	3-dim	3-dim
LSSD	.9887	.9998
LSD	.9880	.9996
LSLD	.9697	.9956

and three dimensions, and might very well be explained by the way in which the dissimilarities are approximated. By LSSP they are approximated *from below*, while in the other methods the approximation is simultaneously from below and from above.

Our final application concerns the comparison of LSD with its most important rivals, which are LSSD and LSLD. Our hypothesis is that LSSD, because of its relation with LSSP (cf. de Leeuw and Heiser 1982), will give similar results, thus LSSD will be more stable than LSD. This hypothesis is confirmed by both STAB and CROSS. According to both criteria LSLD gives less stable results compared with the other methods.

To complete our second illustration we present parts of the Jackknife results graphically. We have chosen the solutions for LSSP, LSSD, LSD and LSLD (Figures 1 through 4, respectively), which can be thus compared.

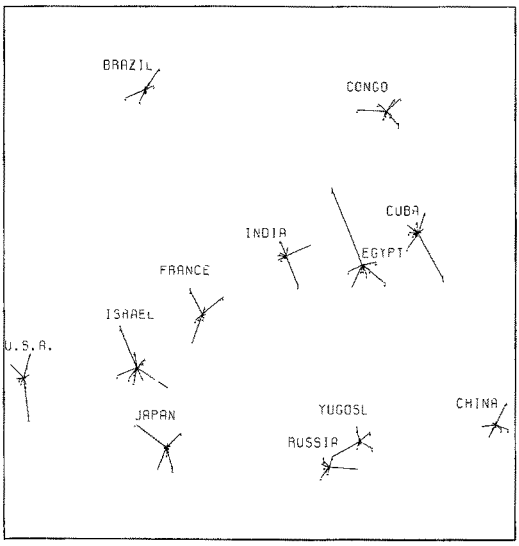

Figure 1. Stability results for LSSP: first 2 dimensions of the three-dimensional solution for the nations data.

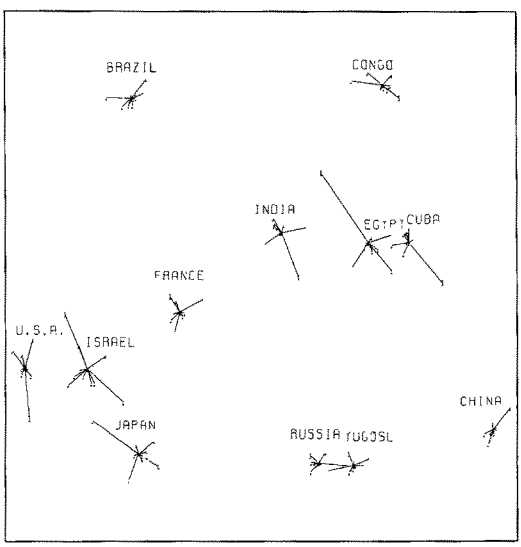

Figure 2. Stability results for LSSD: first 2 dimensions of the three-dimensional solution for the nations data.

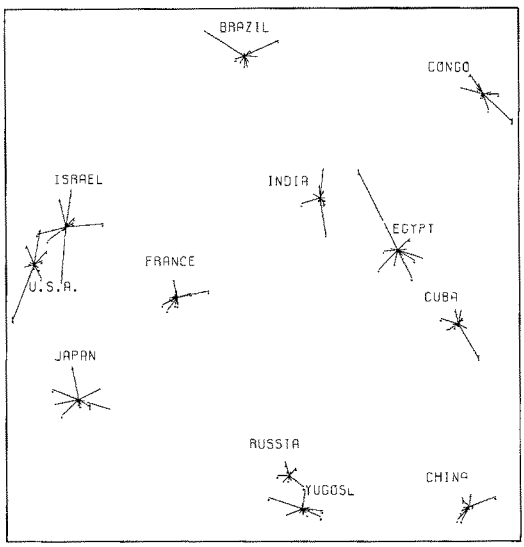

Figure 3. Stability results for LSD: first 2 dimensions of the three-dimensional solution for the nations data.

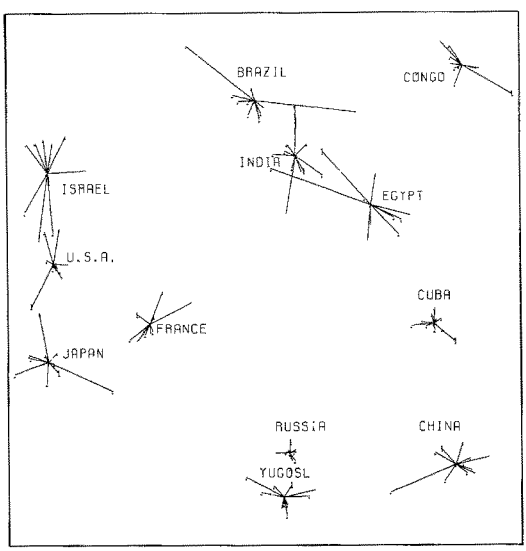

Figure 4. Stability results for LSLD: first 2 dimensions of the three-dimensional solution for the nations data.

By looking at the separate stars in a configuration, specific stability information can be deduced for each object. Because of the large overall stability, the endpoints of the stars have not all been labeled, since this would yield very smudgy plots. When performing an actual data analysis, however, it might be very worthwhile to enlarge the scale of the figure and look for special patterns. To give some idea, we have labeled the endpoint of each star indicating the position of the object when object 1 (Brazil) was left out of the analysis. We see, e.g., that Egypt and Cuba move away from each other; the same is true for the USA and Israel, while Russia, Yugoslavia and Cuba retain their interrelations. This pattern is consistent across the four methods.

It is beyond the scope of this paper to scrutinize all aspects of the various solutions. We want to draw attention, however, to the close resemblance between the solutions for LSSP and LSSD. It seems that apart from details, like the position of Russia and Yugoslavia, the ALSCAL solutions hardly move away from their initial configuration, which is obtained by the Torgerson scaling method. Apart from the fact that LSD is more stable than LSLD, the SMACOF and MULTISCALE configurations don't show major differences. They are more similar to each other than to the ALSCAL solution.

We do not present graphical results concerning the cross validation aspects of our Jackknife. As can be deduced from Tables 3-5 the "actual" position and the "predicted" position in the three dimensional solutions are almost identical. This state of affairs, of course, credits the analyses we have performed. Multidimensional Scaling again proves to be a very robust technique, even when large perturbations of the data, by deleting all information about one object, are applied.

4.3 Some Technical Remarks

To conclude this section, we want to report a general experience in our stability study, that has implications for other studies as well, i.e., the important role of the starting configuration for each subset problem. When the (stable) LSSP solution is used, as is done in our study and is common practice in most MDS programs, incomplete convergence, either due to a too lenient convergence criterion or to imprecision, may lead to overoptimistic stability results. By imprecision we mean that a FORTRAN program may stop according to the convergence criterion, while the critical difference between the two values of stress involved is relatively too large compared to preceding differences. This suspicion has been checked by using analogous programs written in APL. If random initial configurations are used, we expect that incomplete convergence will lead to less stable solutions.

5. Discussion

In this paper a special Jackknife is presented that gives information about stability and cross-validation aspects of MDS solutions. This data analytical Jackknife is especially suitable when there are no replications.

In the applications presented, the evaluative measures for stability and cross-validation were highly consistent and accordingly specific decisions were easily made. If they had been contradictory, we would have given prevalence to the cross-validation index. This is justified by previous experience with resampling (e.g., Meulman 1982), where very stable solutions could be due to degeneracies and accordingly were hardly informative from a data analytic point of view.

Our procedure gives very useful results, that are however obtained at the cost of much additional computation. Each analysis has to be performed $n + 1$ times, where n is the number of objects in the data matrix. This might be prohibitive when the available CPU-time is limited. Moreover, a lot of separate manipulations are needed as long as the procedures like the Jackknife are not built into standard MDS programs. Nevertheless these efforts are warranted, since we obtain stability results in cases where a short cut to inference is just not feasible.

References

DE LEEUW, J., and HEISER, W.J. (1980), "Multidimensional Scaling with Restrictions on the Configuration," in *Multivariate Analysis, Volume 5*, Ed., P.R. Krishnaiah, Amsterdam: North Holland, 501-522.

DE LEEUW, J., and HEISER, W.J. (1982), "Theory of Multidimensional Scaling," in *Handbook of Statistics, Volume 2*, Eds., P.R. Krishnaiah and L. Kanal, Amsterdam: North Holland, 285-316.

EFRON, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS NSF Regional Conference Series in Applied Mathematics, No. 38, Philadelphia: Society for Industrial and Applied Mathematics.

GIFI, A. (1981), *Nonlinear Multivariate Analysis*, Leiden: Department of Data Theory.

GOWER, J.C. (1966), "Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis," *Biometrika, 53*, 325-338.

GOWER, J.C. (1975), "Generalized Procrustes Analyses," *Psychometrika, 40*, 33-51.

GRAY, H.L., and SCHUCANY, W.R. (1972), *The Generalized Jackknife*, New York: Marcel Dekker.

HEISER, W.J., and MEULMAN, J. (1983a), "Analyzing Rectangular Tables by Joint and Constrained Multidimensional Scaling," *Journal of Econometrics, 22*, 139-167.

HEISER, W.J., and MEULMAN, J. (1983b), "Constrained Multidimensional Scaling, Including Confirmation," *Applied Psychological Measurement, 7*, 381-404.

HUBERT, L.J., and ARABIE, P. (1986), "Unidimensional Scaling and Combinatorial Optimization," in *Multidimensional Data Analysis*, Eds., J. de Leeuw, W.J. Heiser, J. Meulman, and F. Critchley, Leiden: DSWO Press.

KRUSKAL, J.B. (1964), "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika, 29*, 1-28.

KRUSKAL, J.B., and WISH, M. (1978), *Multidimensional Scaling*, Beverly Hills and London: Sage Publications.

MESSICK, S.J., and ABELSON, R.P. (1956), "The Additive Constant Problem in Multidimensional Scaling," *Psychometrika, 21*, 1-17.

MEULMAN, J. (1982), *Homogeneity Analysis of Incomplete Data*, Leiden: DSWO Press.

MILLER, R.G. (1974), "The Jackknife – A Review," *Biometrika, 61*, 1-17.

RAMSAY, J.O. (1977), "Maximum Likelihood Estimation in Multidimensional Scaling," *Psychometrika, 42*, 241-266.

RAMSAY, J.O. (1982), "Some Statistical Approaches to Multidimensional Scaling Data," *Journal of the Royal Statistical Association, Series A, 145*, 285-312.

RAO, C.R. (1973), *Linear Statistical Inference and Its Applications*, New York: Wiley.

SHEPARD, R.N. (1962), "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function I," *Psychometrika, 27*, 125-140.

SIBSON, R. (1979), "Studies in the Robustness of Multidimensional Scaling: Perturbational Analysis of Classical Scaling," *Journal of the Royal Statistical Society, Series B, 41*, 217-229.

SPENCE, I. (1983), "Monte Carlo Simulation Studies," *Applied Psychological Measurement, 7*, 405-425.

STONE, M. (1974), "Cross-validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society, Series B, 36*, 111-147.

STOOP, I., HEISER, W.J., and DE LEEUW, J. (1981), *How to Use SMACOF-IA*, Leiden: Department of Data Theory.

TAKANE, Y., YOUNG, F.W., and DE LEEUW J. (1977), "Nonmetric Individual Differences Scaling: An Alternating Least Squares Method with Optimal Scaling Features," *Psychometrika, 42*, 1-67.

TEN BERGE, J.M.F. (1977a), *Optimizing Factorial Invariance*, Doctoral Dissertation, University of Groningen.

TEN BERGE, J.M.F. (1977b), "Orthogonal Procrustes Rotation for Two or More Matrices," *Psychometrika, 42*, 267-276.

TORGERSON, W.S. (1958), *Theory and Methods of Scaling*, New York: Wiley.

WEINBERG, S.L., CARROLL, J.D., and COHEN, H.S. (1984), "Confidence Regions for INDSCAL Using the Jackknife and the Bootstrap Techniques," *Psychometrika, 49*, 475-491.

YOUNG, G., and HOUSEHOLDER, A.S. (1938), "Discussion of a Set of Points in Terms of Their Mutual Distances," *Psychometrika, 3*, 19-22.