Behaviormetrika
1982, No. 11, 37-48

THE METHOD OF TRIADIC COMBINATIONS: A NEW TREATMENT AND ITS APPLICATION

Yoshio Takane*

In the method of triadic combinations three entities (stimuli, dissimilarities, etc.) are rank ordered in a specific order. When those entities are stimuli themselves, the subject is typically asked to choose, first, the most dominant stimulus among three stimuli presented at a time according to some prescribed criterion, and then the least dominant one among two remaining stimuli. If, on the other hand, those entities represent dissimilarities defined on a triad of stimuli, the subject is instructed to choose, among the three stimuli, the most similar stimulus pair and then the most dissimilar pair. In either case the procedure establishes an observed rank order among three entities. The rank order data collected by this method were conventionally analyzed by Thurstone's model of comparative judgments by reducing them to pair comparison data. A new approach to the method of triadic combinations is proposed in view of the fact that it is a special case of the directional ranking method (*i.e.*, rank orderings are performed in a specific direction). A maximum likelihood estimation procedure is developed and implemented in the form of a FORTRAN program. An illustrative example is given to demonstrate the feasbility of the procedure.

1. Introduction

Richardson (1938) was the first to use the method of triadic combinations. In this method three stimuli are presented at a time to the subject. He is asked first to choose a stimulus which is most dominant according to some prescribed attribute of stimuli, and then to choose the least dominant one among two remaining stimuli. The same procedure is repeatedly applied to other triads of stimuli. If the three stimuli to be compared are three dissimilarities defined on a triad of stimuli, we obtain the method of tridiac combinations as it was originally used by Richardson (1938). In this case the subject is to choose, first, the most similar stimulus pair and then the most dissimilar pair. The stimulus pair which is neither most similar nor most dissimilar is deduced from the previous two judgements. In either case rank orderings are established among three entities. In the former case the three ranked entities are stimuli themselves, while in the latter they are (dis)similarities defined on a triad of stimuli.

Since Richardson, several researchers have used the method of triadic combinations (Vlek, 1969; Levelt, van de Geer & Plomp, 1966; Roskam, 1969). However, it has never won its popularity among psychometricians as a handy data collection

Key Words and Phrases; directional ranking method, maximum likelihood estimation (MLE), multidimensional scaling (MDS)

* Department of Psychology, McGill University. The work reported in this paper was supported by grant A6394 to the author by the Natural Sciences and Engineering Research Council of Canada. Thanks are due to Justine Sergent who helped prepare Figure 2. Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Avenue Docteur Penfield, Montreal, Quebec, H3A 1B1 Canada.

method, and there seems to be a reason to it. Suppose three stimuli, S_i, S_j and S_k are presented, and S_i is chosen to be the most dominant stimulus and S_k to be the least dominant stimulus. This establishes a rank order among the three stimuli, $S_i > S_j > S_k$, where $>$ indicates that the stimulus on the left of the symbol empirically dominates the one on the right. However, due to the lack of a proper ranking model, the rank order data obtained by the method of tridiac combinations have been analyzed via Thurstone's model of comparative judgments (Thurstone, 1927) by reducing them to pair comparison information. For example, the above rank order may be decomposed to three binary relations, $S_i > S_j$, $S_j > S_k$ and $S_i > S_k$.

However, three binary relations derived from a rank order are not completely equivalent to three binary relations obtained from three independent pair comparison judgments; a rank order imposes a rather stringent constraint on the set of derived binary relations. For example, if $S_i > S_j$ and $S_j > S_k$, then it must be that $S_i > S_k$ (*i.e.*, transitivity must always hold). This implies that the last relation ($S_i > S_k$) is not statistically independent of (in fact, completely dependent on) the previous two relations (*i.e.*, $S_i > S_j$ and $S_j > S_k$). This, however, is not generally true in pair comparison judgements, in which violations of transitivity may occur quite naturally. Even if $S_i > S_j$ and $S_j > S_k$, $S_i < S_k$ can occur, and such a relation is perfectly legitimate.

In view of this fact Torgerson (1958) modified the basic constitution of the method of triadic combinations into a new method called the method of triads. This method is more conformable to the requirements of pair comparison judgments, though it is restricted to comparisons among dissimilarities rather than comparisons among stimuli themselves. In the method of triads stimuli are presented in triads as in the method of triadic combinations. The difference is in that one of the stimuli is designated as a standard stimulus, and the subject is asked to choose a stimulus which he perceives to be most similar to the standard among two comparison (=nonstandard) stimuli. Let S_i, S_j and S_k be the three stimuli presented in a particular trial, and let S_i be the standard stimulus. Then this procedure obtains a pariwise ordinal relation on $\delta(S_i, S_j)$ and $\delta(S_i, S_k)$ where δ indicates the dissimilarity between two stimuli. ($\delta(S_i, S_j) < \delta(S_i, S_k)$, if S_j is judged more similar to S_i and $\delta(S_i, S_j) > \delta(S_i, S_k)$, otherwise.) The same set of three stimuli is presented three times, one stimulus serving as a standard stimulus each time, and it is usually presented well apart in the sequence of judgment trials to ensure statistical independence of the three judgments.

The method of triads, as used by Torgerson (1958) and others (*e.g.*, Krantz, 1967a, b), has some advantage over the method of triadic combinations, so far as the data arising from the latter method are reduced to pair comparison data and treated as such. While it is possible, at least in principle, to obtain statistically independent binary relations in the method of triads, it is not possible in the method of triadic combinations. Perhaps for this reason the latter method is no longer very frequently used in its original form.

A notable exception is the work by Roskam (1969); his procedure can directly deal with the kind of rank order data arising from the method of triadic combinations. Roskam invented a notion of data conditionalities (this terminology due to Takane, Young & de Leeuw, 1977), subsets of data within which they can be meaningfully compared. Furthermore, he could successfully incorporate it into a stress function, a least squares loss function typically employed in Kruskal (1964) type of nonmetric multidimensional scaling. As is well known, Kruskal's nonmetric MDS finds, based exclusively on the ordinal information about dissimilarities, a configuration of stimulus points in such a way that their mutual distances best agree with the observed ordinal relations. Roskam separately defined a stress function for each data conditionality (*i.e.*, for each triadic rank order), and combined it into a global stress function by taking a root mean square. His procedure finds a stimulus configuration which minimizes this modified stress function. Otherwise it works in a similar way to Kruskal's procedure.

In this paper we discuss a new approach to the method of triadic combinations, which, like Roskam's, can handle the triadic combination data without reducing them to pair comparison data. Unlike Roskam's, however, it is based on the maximum likelihood principle, which has some advantage over the least squares estimation in terms of its statistical inference capability.

2. Basic Formulation

In developing a maximum likelihood estimation procedure for the method of triadic combinations the most crucial thing to realize is that the ranking is performed in a specific direction (Takane & Carroll, 1981). Let us for a moment restrict our attention to the case in which stimuli, not dissimilarities, are ranked. The subject chooses the most dominant stimulus first and then the least dominant stimulus, but since choosing the least dominant stimulus among two remaining stimuli is equivalent to choosing the stimulus, which is neither most dominant nor least dominant, as the second most dominant stimulus, it amounts to rank ordering the three stimuli from the most dominant one to the least dominant one. In this case a ranking may be viewed as resulting from successive first choices.

Let stimuli S_i, S_j and S_k be rank ordered in a particular triadic comparison. Suppose the rank order obtained is $S_i > S_j > S_k$. Let μ_i, μ_j, and μ_k represent stimulus values (scale values) corresponding to these stimuli. We assume that the stimulus values are error-perturbed, and generate error-perturbed processes $y_i^{(t)}$, $y_j^{(t)}$ and $y_k^{(t)}$ (corresponding to μ_i, μ_j and μ_k, respectively) at a particular first choice which is designated by t. That is,

$$y_i^{(t)} = \mu_i + e_i^{(t)} , \tag{1}$$

(the same for j and k) where $e_i^{(t)}$ is the error random variable, which may be further assumed to follow a central normal distribution. It seems natural to assume that a stimulus is chosen on the basis of $y^{(t)}$ at the t^{th} successive first choice. The stimulus having the largest value of $y^{(t)}$ is chosen as the most dominant stimulus. Thus, S_i is judged to be the most dominant stimulus among the three stimuli, (S_i, S_j and S_k) when $y_i^{(1)}$ exceeds both $y_j^{(1)}$ and $y_k^{(1)}$. Then

$$\begin{aligned} p_{ijk}^{(1)} &\equiv Pr(S_i > S_j \text{ and } S_i > S_k) \\ &= Pr(y_i^{(1)} > y_j^{(1)} \text{ and } y_i^{(1)} > y_k^{(1)}) . \end{aligned} \tag{2}$$

Define

$$y_{ijk}^{(1)}=\begin{pmatrix} y_i^{(1)} \\ y_j^{(1)} \\ y_k^{(1)} \end{pmatrix}, \quad \Sigma_{ijk}^{(1)}=V(y_{ijk}^{(1)}), \quad \mu_{ijk}^{(1)}=\begin{pmatrix} \mu_i \\ \mu_j \\ \mu_k \end{pmatrix},$$

and

$$A^{(1)}=\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

We may then rewrite (2) as

$$\begin{aligned} p_{ijk}^{(1)} &= Pr(A^{(1)}y_{ijk}^{(1)}>0) \\ &= \int_{S^{(1)}} \phi(z_{ijk}^{(1)})\, dz_{ijk}^{(1)}, \end{aligned} \tag{3}$$

where $z_{ijk}^{(1)}=A^{(1)}y_{ijk}^{(1)} \sim N(A^{(1)}\mu_{ijk}^{(1)}, A^{(1)}\Sigma_{ijk}^{(1)}A^{(1)\prime})$ and $S^{(1)}$ is a multidimensional region such that $z_{ijk}^{(1)}>0$ (0 is the zero vector). If the variance-covariance matrix $\Sigma_{ijk}^{(1)}$ of $y_{ijk}^{(1)}$ can be expresssed in the form of

$$\Sigma_{ijk}^{(1)}=(\sigma^2/2)\, I_3+1_3a_3'+a_31_3' \tag{4}$$

for any arbitrary three-component vector a_3 and the vector of ones, 1_3, the variance-covariance matrix $A^{(1)}\Sigma_{ijk}^{(1)}A^{(1)\prime}$ of $z_{ijk}^{(1)}$ will have σ^2 in the diagonals and $\sigma^2/2$ in the off-diagonals. In this case the bivariate normal integral in (3) can be aproximated by the bivariate logistic distribution (Bock, 1975), which is given by

$$\begin{aligned} p_{ijk}^{(1)} &= [1+\exp\{s(\mu_j-\mu_i)\}+\exp\{s(\mu_k-\mu_i)\}]^{-1} \\ &= \frac{\exp(s\mu_i)}{\exp(s\mu_i)+\exp(s\mu_j)+\exp(s\mu_k)}, \end{aligned} \tag{5}$$

where s is a dispersion parameter and is approximately equal to $\pi/\sqrt{3}\,\sigma$. (Note that this s does not have to be explicitly estimated. Since μ's are only determined up to a multiplicative constant plus an additive constant, we may let the scale factor of μ's take over the size of s.) The logistic distribution is much easier to evaluate than the normal integral, and is more convenient to use for computational purposes.

The probability $p_{jk}^{(2)}$ that S_j is judged to be more dominant than S_k after S_i is eliminated from the comparison set is similarly obtained.
We have

$$\begin{aligned} p_{jk}^{(2)} &= Pr(A^{(2)}y_{jk}^{(2)}>0) \\ &= \int_{S^{(2)}} \phi(z_{jk}^{(2)})\, dz_{jk}^{(2)} \end{aligned} \tag{6}$$

where $z_{jk}^{(2)} \sim N(A^{(2)}\mu_{jk}^{(2)}, A^{(2)}\Sigma_{jk}^{(2)}A^{(2)\prime})$,

$$y_{jk}^{(2)}=\begin{pmatrix} y_j^{(2)} \\ y_k^{(2)} \end{pmatrix}, \quad \Sigma_{jk}^{(2)}=V(y_{jk}^{(2)}), \quad \mu_{jk}^{(2)}=\begin{pmatrix} \mu_j \\ \mu_k \end{pmatrix},$$

$$A^{(2)}=(1\ \ -1),$$

and $S^{(2)}$ is a region such that $z_{jk}^{(2)}>0$. Since in the present case $z_{jk}^{(2)}=A^{(2)}y_{jk}^{(2)}=$

$y_j^{(2)}-y_k^{(2)}$ and $A^{(2)}\mu_{jk}^{(2)}=\mu_j-\mu_k$, (6) can alternatively be written as

$$p_{jk}^{(2)}=\int_{-\infty}^{b_{jk}}\phi(z)\,dz \tag{7}$$

where $b_{jk}=(\mu_j-\mu_k)/\sigma$ $(\sigma^2=A^{(2)}\Sigma_{jk}^{(2)}A^{(2)\prime})$ and z is univariate standard normal. It is well known that the normal integral in (7) is closely approximated by the following logistic distribution:

$$\begin{aligned} p_{jk}^{(2)} &= [1+\exp\{s(\mu_k-\mu_j)\}]^{-1} \\ &= \frac{\exp(s\mu_j)}{\exp(s\mu_j)+\exp(s\mu_k)}\,. \end{aligned} \tag{8}$$

Finally, we may identically set $p_k^{(3)}=1$, since, after the two most dominant stimuli are chosen, S_k is the only one stimulus left, which is bound to be chosen as the least dominant stimulus.

The probability p_{ijk} of rank order, $S_i>S_j>S_k$, can now be stated, under the independence assumption on $p_{ijk}^{(1)}$ and $p_{jk}^{(2)}$, as

$$p_{ijk}\equiv Pr(S_i>S_j>S_k)=p_{ijk}^{(1)}\cdot p_{jk}^{(2)}\,. \tag{9}$$

The independence of $p_{ijk}^{(1)}$ and $p_{jk}^{(2)}$ can be obtained, even if $y_{ijk}^{(1)}$ and $y_{jk}^{(2)}$ are not completely independent. It suffices that the covariance between them $\Sigma^{(1,2)}$ has the following structure:

$$\Sigma^{(1,2)}=\mathrm{Cov}\,(y_{ijk}^{(1)},\,y_{jk}^{(2)})=1_3b_2'+c_31_2'\,, \tag{10}$$

where b and c are arbitrary vectors and 1 is the vector of ones. Subscripts on the vectors indicate their dimensions.

There are six possible rank orders among three stimuli. For example, for S_i, S_j and S_k we have (1) $S_i>S_j>S_k$, (2) $S_i>S_k>S_j$, (3) $S_j>S_i>S_k$, (4) $S_j>S_k>S_i$, (5) $S_k>S_i>S_j$ and (6) $S_k>S_j>S_i$. Only one of these occurs in a particular ranking. Let $p_{(ijk)}$ denote the probability that one of them occurs in a particular trial. Define

$$Z_{ijk}=\begin{cases}1, \text{ if } S_i>S_j>S_k \text{ is observed}\\ 0, \text{ otherwise.}\end{cases}$$

Then

$$p_{(ijk)}=\prod_{ijk}(p_{ijk})^{Z_{ijk}} \tag{11}$$

where the product is taken over all permutations of i, j and k (i.e., ijk, ikj, jik, jki, kij and kji). Suppose $N_{(ijk)}$ repeated trials are made, involving the same three stimuli and a particular rank order, $S_i>S_j>S_k$, is observed Z_{ijk}^* times. The probability $p_{(ijk)}^*$ that possible rank orders are observed certain fixed numbers of times is given by

$$p_{(ijk)}^*=\prod_{ijk}(p_{ijk})^{Z_{ijk}^*}\,, \tag{12}$$

where the product is again taken over all possible permutations of i, j and k. We also have $N_{(ijk)}=\sum_{ijk}Z_{ijk}^*$. The likelihood of the total set of observed rank orders is in turn given by the product of $p_{(ijk)}^*$ over triads of stimuli, again assuming the statistical independence of $p_{(ijk)}^*$ and $p_{(stu)}^*$. The independence can be obtained under a similar condition to (10).

When the ranking is performed from the least dominant stimulus to the most dominant (rather than in the other direction as has been assumed throughout), only a minor modification is necessary in the above formulation. Only s in (5) and (8) need be replaced by $-s(s>0)$, provided that certain symbols are systematically redefined. The $p_{ijk}^{(1)}$ in (2) now indicates $Pr(S_i<S_j$ and $S_i<S_k)$ and $A^{(1)}$ has to be replaced by $-A^{(1)}$. Similarly, $p_{jk}^{(2)}$ indicates $Pr(S_j<S_k)$ and $A^{(2)}$ should be replaced by $-A^{(2)}$. Also, p_{ijk} in (9) now indicates $Pr(S_i<S_j<S_k)$.

3. The Case of Dissimilarity Rankings

The method of triadic combinations has most often been used as a method to collect dissimilarity data. In this case three stimuli are presented to the subject as before, but instead of obtaining a rank order among them we obtain a rank order among dissimilarities between them. As noted earlier, this is done by asking the subject to choose the most similar stimulus pair first and then the most dissimilar pair. It is obvious that this procedure obtains a rank order among three dissimilarities defined on three stimuli from the smallest to the largest. Again some modification is necessary in the above formulation in order to deal with this situation.

When the data are ranked dissimilarities, we would usually like to not only scale them, but also represent them by some distance model. Let us assume that the distance model is euclidean; i.e.,

$$d_{ij} = \{\sum_{a=1}^{A} (x_{ia}-x_{ja})^2\}^{1/2}, \tag{13}$$

where d_{ij} is the euclidean distance between stimuli i and j, x_{ia} is the coordinate of stimulus i on dimension a, and A is the dimensionality of the space. This d_{ij} is analogous to μ's in the preceding discussion. We again assume that d_{ij} is error-perturbed at each particular first choice, and that the most similar stimulus pair is chosen on the basis of a particular value of error-perturbed d_{ij} realized at that occasion. The rest of the procedure then goes much the same way as in the previous development. Notice that the likelihood in this case is maximized over the stimulus coordinates x_{ia} rather than d_{ij}. Note also that this case can be construed as a special case of the previous formulation in which μ's are constrained to have a special structure implied by (13).

When dissimilarities are rank ordered, it may also be useful to incorporate the log normal assumption (multiplicative error model) on e in (1) as well as the normal assumption (additive error model). In this case d_{ij} and $y_{ij}^{(m)}$ have to be replaced by $\ln d_{ij}$ and $\ln y_{ij}^{(m)}$,) respectively. Defining $\delta_{ij}=\delta_i(S_i, S_j)$, we have

$$\begin{aligned} p_{ijk}^{(1)} &= Pr(\delta_{ij}<\delta_{jk} \text{ and } \delta_{ij}<\delta_{ik}) \\ &= Pr(y_{ij}<y_{ik} \text{ and } y_{ij}<y_{ik}) \\ &= [1+\exp\{s(\ln d_{jk}-\ln d_{ij})\} + \exp\{s(\ln d_{ik}-\ln d_{ij})\}]^{-1} \\ &= d_{ij}^s/(d_{ij}^s+d_{jk}^s+d_{ik}^s), \end{aligned} \tag{14}$$

and

$$\begin{aligned} p_{ijk}^{(2)} &= Pr(\delta_{jk}<\delta_{ik}) = Pr(y_{jk}<y_{ik}) \\ &= [1+\exp\{s(\ln d_{ik}-\ln d_{jk})\}]^{-1} \\ &= d_{jk}^{s}/(d_{jk}^{s}+d_{ik}^{s}), \end{aligned} \tag{15}$$

where $y_{ij}^{(t)}=d_{ij}+e_{ij}^{(t)}$ and $s<0$.

We may compare the goodness of fit of the two error models on an empirical basis and choose the one which fits to the data better.

4. Numerical Method and Derivatives

The log likelihood can be maximized by various numerical methods. The method we prefer to use is Fisher's scoring algorithm, which has proven to be very efficient in similar situations (Takane, 1978, 1981, 1982; Takane & Carroll, 1981). It is an iterative procedure, in which current parameter estimates are updated by

$$\theta^{(q+1)} = \theta^{(q)} + I(\theta^{(q)})^{-1} u(\theta^{(q)}), \tag{16}$$

where $u(\theta)=(\partial \ln L/\partial\theta)$ is the score vector (L is the likelihood), $I(\theta)=E[(\partial \ln L/\partial\theta)(\partial \ln L/\partial\theta)']$ is Fisher's information matrix and the parenthesized superscripts indicate iteration numbers. When the regular inverse of the information matrix does not exist due to singularity of the matrix, it can be replaced by the Moore-Penrose inverse (Ramsay, 1978). In the present case the singularity of the information matrix occurs due to nonuniqueness of parameters.

In order to apply the above algorithm we have to have explicit expressions for $u(\theta)$ and $I(\theta)$. Let $L=\Pi p_{(ijk)}{}^{*}$ be the likelihood. We then have $\ln L=\Sigma \ln p_{(ijk)}{}^{*}$, where the summation extends over the triads of stimuli actually compared. We have

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu_m} &= \Sigma\left(\frac{\partial \ln p_{(ijk)}{}^{*}}{\partial \mu_m}\right) \\ &= \Sigma \sum_{ijk} Z_{ijk}{}^{*} \frac{\partial \ln p_{ijk}}{\partial \mu_m} \end{aligned} \tag{17}$$

(the second summation denoted by ijk is over all permutations of i, j and k), where

$$\frac{\partial \ln p_{ijk}}{\partial \mu_m} = \delta_{im}\frac{\partial \ln p_{ijk}}{\partial \mu_m} + \delta_{jm}\frac{\partial \ln p_{ijk}}{\partial \mu_m} + \delta_{km}\frac{\partial \ln p_{ijk}}{\partial \mu_m} \tag{18}$$

($\delta\cdots$ is a Kronecker delta). Note that only one of the three terms on the right hand side of the above equation is nonzero, since i, j and k are all distinct. We have $\ln p_{ijk}=\ln p_{ijk}{}^{(1)}+\ln p_{jk}{}^{(2)}$, so that $\partial \ln p_{ijk}/\partial\mu_m=\partial \ln p_{ijk}{}^{(1)}/\partial\mu_m+\partial \ln p_{jk}{}^{(2)}/\partial\mu_m$, and thus

$$\frac{\partial \ln p_{ijk}{}^{(1)}}{\partial \mu_m} = \delta_{im}s - c_1\left(\frac{\partial \exp(s\mu_i)}{\partial \mu_m} + \frac{\partial \exp(s\mu_j)}{\partial \mu_m} + \frac{\partial \exp(s\mu_k)}{\partial \mu_m}\right) \tag{19}$$

where

$$c_1 = [\exp(s\mu_i)+\exp(s\mu_j)+\exp(s\mu_k)]^{-1},$$

and

$$\frac{\partial \exp(s\mu_i)}{\partial \mu_m} = \delta_{im}\, s\exp(s\mu_i).$$

Similarly,

$$\frac{\partial \ln p_{jk}^{(2)}}{\partial \mu_m} = \delta_{jm} s - c_2\left(\frac{\partial \exp(s\mu_j)}{\partial \mu_m} + \frac{\partial \exp(s\mu_k)}{\partial \mu_m}\right), \tag{20}$$

where $c_2 = [\exp(s\mu_j) + \exp(s\mu_k)]^{-1}$.

Derivatives can be obtained in an analogous manner when dissimilarities are rank ordered rather than stimuli themsevles. In this case μ is replaced by $\boldsymbol{d}$ which is further related to x by (13). Since x is the parameter to be estimated we need a derivative of ln L with respect to x. We have

$$\frac{\partial \ln L}{\partial x_{ma}} = \frac{\partial \ln L}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial x_{ma}}, \tag{21}$$

where $\partial \ln L/\partial d_{ij}$ is analogous to $\partial \ln L/\partial \mu_m$, and

$$\frac{\partial d_{ij}}{\partial x_{ma}} = (\delta_{im} - \delta_{jm}) \frac{(x_{ia} - x_{ja})}{d_{ij}}. \tag{22}$$

When the log normal assumption is made, we obtain

$$\frac{\partial \ln p_{ijk}^{(1)}}{\partial d_{mn}} = \delta_{(ij)(mn)} s/d_{mn} - c_1\left(\frac{\partial d_{ij}^s}{\partial d_{mn}} + \frac{\partial d_{ij}^s}{\partial d_{mn}} + \frac{\partial d_{ik}^s}{\partial d_{mn}}\right), \tag{23}$$

where $c_1 = [\ln(d_{ij}^s + d_{jk}^s + d_{ik}^s)]^{-1}$ and $\partial d_{ij}^s/\partial d_{mn} = \delta_{(ij)(mn)}\, s d_{ij}^{s-1}$, and

$$\frac{\partial \ln p_{jk}^{(2)}}{\partial d_{mn}} = \delta_{(jk)(mn)} s/d_{mn} - c_2\left(\frac{\partial d_{jk}^s}{\partial d_{mn}} + \frac{\partial d_{ik}^s}{\partial d_{mn}}\right), \tag{24}$$

where $c_2 = [\ln(d_{jk}^s + d_{ik}^s)]^{-1}$.

Equations (23) and (24) follow from $\partial \ln f(x)/\partial x = (1/f(x))(\partial f(x)/\partial x)$.

It is not very easy to obtain an explicit expression of the information matrix directly. However, it can be fairly easily obtained by noting the equivalence between the Gauss-Newton method for the weighted least squares and the scoring algorithm for maximum likelihood estimation when the assumed population distribution is one of the exponential family of distributions (Jennrich & Moore, 1975). In the present case we have the multinomial distribution (we see the kernel of multinomial distribution in (12)), which is a special case of the exponential family of distributions. The equivalent least squares criterion in this case is written as

$$\tau = \Sigma \sum_{ijk} \frac{N_{(ijk)}}{p_{ijk}} (Z_{ijk}/N_{(ijk)} - p_{ijk})^2, \tag{25}$$

where the first summation is over the triads of stimuli actually compared and the second summation over permutations of i, j and k. It is well known that the Hessian used in the Guass-Newton method to minimize τ is given by

$$h_{mm'} = 2\Sigma \sum_{ijk} \left(\frac{\partial p_{ijk}}{\partial \theta_m}\right) \frac{N_{(ijk)}}{p_{ijk}} \left(\frac{\partial p_{ijk}}{\partial \theta_{m'}}\right), \tag{26}$$

where $h_{mm'}$ is the (m, m') element of the Hessian matrix $H(\theta)$, θ_m the m^{th} parameter and the two summations extend over the same range as in (25). When the population distribution is one of the exponential family of distributions as in the present case, the information matrix derived from the likelihood function is known to be proportional to $H(\theta)$ [*i.e.*, $I(\theta)=1/2\ H(\theta)$].

5. Example of Application

In this section we report some empirical results obtained by the procedure described in the previous sections. The data to be analyzed are dissimilarity data collected by the method of triadic combinations. Dissimilarity judgments were obtained between nine colors originally employed by Torgerson (1958) in his study of classical multidimensional scaling. He used the method of triads to obtain dissimilarity judgments, employed Thurstone's model of comparative judgments to scale them, and applied Young and Householder's (1938) method to the scaled dissimilarities to find a spatial representation. In contrast we used the method of triadic combinations and the analysis procedure specifically designed to deal with the kind of rank order data obtained by this method.

Six complete replications were made on 84 triads of stimuli using a single subject (male adult, normal vision). Several authors (Messick, 1956; Nakatani, 1972; Saito, 1977; Shepard, 1958, 1962; Takane, 1978) have reported stimulus configurations on the same set of stimuli. However, in all these cases replications were taken over different subjects. This is the first time the data replicated within a single subject are analyzed.

One of the major advantages of the current procedure is that it allows various model comparisons through a statistic called AIC (Akaike, 1974). The AIC of model π is defined by

$$\mathrm{AIC}(\pi) = -2\ln L + 2\,n_\pi \tag{27}$$

where L is the maximum likelihood and n_π is the effective number of parameters in model π. The model with the smallest value of AIC is considered the best fitting model. (Relatively nontechnical discussion on, and the use of, AIC in maximum likelihood MDS may be found in Takane (1981)). One of the most fundamental model comparisons in the present case is the choice of an appropriate error model. Another is the test of dimensionality. A third is the test of specific hypotheses about a stimulus configuration. By use of the AIC statistic all these could be done in a relatively straightforward manner.

Major analysis results are summarized in Table 1. Three figures are reported in each cell of the table. The top one is minus twice the log likelihood, the middle is the effective number of parameters (n_π) in the model and the bottom is the value of the AIC statistic.

The data were analyzed under several assumptions; under two error models (normal and log normal) and in three different dimensionalities (1, 2 & 3). We see that in all dimensionalities the AIC values are smaller in the normal error model than in the log normal error model, indicating that the former is the better fitting model. This finding is consistent with our previous findings (Takane, 1978; Takane & Carroll, 1981) that the normal error model generally shows a better fit, when the data are collected by the methods which involve direct comparison of dissimilarities (*e.g.*,

Table 1
Summary of MAXSCAL-4 Analyses of the Color Data obtained by the Method of Triadic Combinations

Dimensionality		Normal error model (Additive error model)	Log normal error model (Multiplicative error model)
3	−2(LL)	498.8	515.2
	n_π	21	21
	AIC	540.8	557.2
2	−2(LL)	521.5	528.5
	n_π	15	15
	AIC	551.5	555.5
1	−2(LL)	792.7	895.4
	n_π	8	8
	AIC	808.7	911.4

pair comparisons and directional rankings). Assuming that the normal error model is more appropriate, the next question to be posed is the appropriate number of dimensions in the representation space. We see that the three dimensional solution has the minimum AIC value, indicating that the dimensionality is at least three. (A four dimensional solution was not attempted.)

The set of stimuli employed has a spatial representation, like the one depicted in Figure 1, in terms of the Munsell Value (brightness) and Chroma (saturation) dimensions.

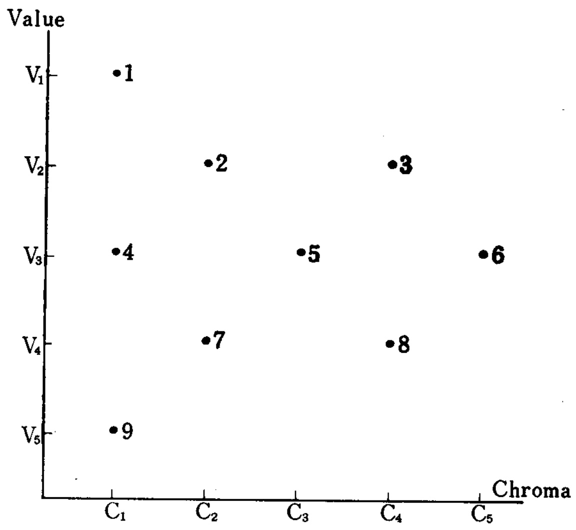

Fig. 1 The Munsell Configuration of the Nine Colors

The Munsell dimensions are presumed to represent psychological dimensions of colors. The above results cast some doubt about the credibility of the Munsell system as a psychological model. Takane (1978) also obtained some negative results against the system from reanalyses of Saito's (1974) data. His data were collected on the same set of stimuli by the method of tetrads. However, even in that case the dimensionality was found to be two, and the derived stimulus configuration was only topologically, though significantly, distorted in the two dimensional space, whereas here the configuration seems to require at least three dimensions. The three dimensional solution which happens to be the minimum AIC solution is presented in Figure 2.

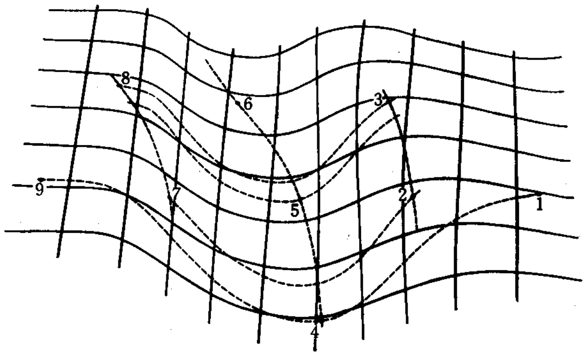

Fig. 2 Derived three-dimensional stimulus configuration

The configuration is not only distorted relative to the Munsell configuration, but also curved in an interesting way in the three dimensional space. It looks like a valley in the middle of mountains or ridge if seen upside down. The discrepancy between Takane's (1978) and the current results may be due to the difference in the data collection method. On the other hand, it may be due to the fact that the present data were obtained from a single subject, while Saito's data were collected over 60 different subjects. If the nature of individual differences is such that they differ with regard to the way the configuration is curved along the third dimension, the data aggregated over different subjects will not yield a clear third dimension. Torgerson's (1958) data analyzed by Takane's (1978) procedure also reveals some mild degree of curvedness along the third dimension, though this is not as distinct as in the present case. The curved configuration in the present case itself might be due to the noneuclidean nature of the perceptual space of colors, or to some peculiarity in judgmental processes involved in direct comparisons of stimuli or dissimilarities (Takane, 1980). A further investigation is necessary on this point.

6. Summary

In this paper we proposed a new approach to the method of triadic combinations. In this approach the method of triadic combinations is veiwed as a special instance of the method of directional rank orders in which rankings are supposedly obtained in a specific order. A model of psychological processes is postulated which transform scale values of stimuli into observed rank orders. Through this model the likelihood of the observed rank orders is related to the scale values of stimuli. A procedure to estimate the scale values of stimuli by the maximum likelihood principle was developed. Example data were analyzed to illustrate practical uses of the procedure. The data analyzed were dissimilarity data between nine colors previously employed by Torgerson

(1958) and by several others. Some interesting new evidence has been found about the perceptual structure of the colors.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Bock, R.D. (1975). *Multivariate statistical methods in behavioral research.* New York: McGraw Hill.

Jennrich, R.I. and Moore, R.H. (1975). Maximum likelihood estimation by means of nonlinear least squares. *Proceedings of the American Statistical Association, Statistical Computing Section*, 57–65.

Krantz, D.H. (1967a). Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology*, **4**, 226–245.

Krantz, D.H. (1967b). Small-step and large-step color differences for monochromatic stimuli of constant brightness. *Journal of Optical Society of America*, **57**, 1304–1316.

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

Levelt, W.J.M., van de Geer, J.P. and Plomp, R. (1966). Triadic comparisons of musical intervals. *British Journal of Mathematical and Statistical Psychology*, **19**, 163–179.

Messick, S.J. (1956). An empirical evaluation of multidimensional successive categories. *Psychometrika*, **21**, 367–375.

Nakatani, L.H. (1972). Confusion-choice model for multidimensional psychophyscis. *Journal of Mathematical Psychology*, **9**, 104–127.

Ramsay, J.O. (1978). Confidence regions for multidimensional scaling analysis. *Psychometrika*, **43**, 145–160.

Richardson, M.W. (1938). Multidimensional psychophysics. *Psychological Bulletin*, **35**, 659–660.

Roskam, E.E. (1970). The method of triads for multidimensional scaling. *Nederlands Tijdschrift v. Psych.*, **25**, 404–417.

Saito, T. (1974). Multidimensional Thurstonian scaling and its applications. *Nippon Univac Soken Kiyo*, **4**, 87–112.

Saito, T. (1977). Multidimensional Thurstonian scaling with an application to color metrics. *Japanese Psychological Research*, **19**, 78–89.

Shepard, R.N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, **55**, 509–523.

Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function, I & II. *Psychometrika*, **27**, 125–140 and 219–246.

Takane, Y. (1978). A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent – theory and evaluations. *Japanese Psychological Research*, **20**, 7–17 and 105–114.

Takane, Y. (1980). Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgement. *Japanese Psychological Research*, **22**, 9–28.

Takane, Y. (1981). Multidimensional successive categories scaling: a maximum likelihood method. *Psychometrika*, **46**, 188–196.

Takane, Y. (1982). Maximum likelihood additivity analysis. *Psychometrica*, **47**, (in press).

Takane, Y. and Carroll, J.D. (1981). Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika*, **46**, (in press).

Takane, Y., Young, F.W. and de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7–67.

Thurstone, L.L. (1927). A law of comparative judgments. *Psychological Review*, **34**, 273–286.

Torgerson, W.S. (1958). *Theory and methods of scaling.* New York: Wiley.

Vlek, C. (1969). Metric and nonmetric approaches to a simple psychophysical continuum. Reports from the Psychological Laboratories, The University of Stockholm.

Young, G. and Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, **3**, 19–22.

(Received July, 1981)