A CONJUGATE GRADIENT ALGORITHM FOR THE MULTIDIMENSIONAL ANALYSIS OF PREFERENCE DATA

MING-MEI WANG
University of Iowa

PETER H. SCHÖNEMANN
Purdue University

and

JERROLD G. RUSK
University of Arizona

ABSTRACT

In continuation of earlier work on a new individual difference model for the multidimensional analysis of preference data (Schönemann and Wang, 1972), a relatively efficient algorithm for applying the model to fallible data was developed. It is based on the Method of Conjugate Gradients and thus does not require storage for second order derivatives. Several different versions of such an algorithm were compared in terms of robustness, accuracy, and speed of convergence. The results strongly suggest that the so-called "intervening conjugate gradient method" (which iterates for only two of the three sets of unknowns and solves for the third set algebraically at each stage) is the most effective method for most purposes. The algorithm was applied to a relatively large set of political choice data which had been previously analyzed by a different method. The outcome of this empirical study not only confirmed the earlier results but also led, as a consequence of the stronger metric structure of the present model, to a more detailed and informative description of the data.

INTRODUCTION

Earlier work of these authors (Schönemann, 1970a; Schönemann & Wang, 1972) has dealt with the development of a mathematically tractable model for the multidimensional analysis of preference data, either in the form of paired comparison (p.c.) probabilities or in the form of rankings. The result was a model which can be interpreted in two different ways (i) as a metricized version of the Coombs unfolding paradigm or (ii) as a multidimensional generalization of the Bradley-Terry Luce model for paired comparison data.

The basic idea under the first interpretation is that a set of stimuli (e.g., a number of political candidates) and a set of judges

This work was supported, in part, by a Small Grant of the National Institute of National Health, Department of Health, Education and Welfare, under Contract No. PRF 7183-51-1364, and in part by a David Ross Grant (PRF 2132) of the Purdue Research Foundation, both of which are gratefully acknowledged. Requests for reprints and program listings should be addressed to the senior author.

(e.g., voters) can be represented in a joint, m-dimensional (Euclidean) space. A given judge will prefer one stimulus A over annother B if A is closer to that Judge's "ideal point" in this joint space. This originally "non-metric" model was metricized by choice of a specific mathematical form of the presumed "isopreference contours." The same mathematical model can be interpreted, alternatively, as a multidimensional extension of the BTL-model where differences in scale values are replaced by differences in squared distances from the ideal point in the argument of a logistic response function which relates the observed choice probabilities to the underlying Euclidean distances. Specifically, under this second, interpretation, the model can be written

$$[1] \qquad P_{ik.j} = G(u_{ik.j})$$

where $G(u) = 1/(1+\exp(-u))$, $u_{ik.j} = d^2_{kj} - d^2_{ij}$,
and where $P_{ik.j}$ is the (observable) probability that stimulus i is chosen over stimulus k by subject j, and d_{kj}, d_{ij} are the (Euclidean) distances between the jth ideal point and stimulus i, or j, respectively. Given the $P_{ik.j}$, the objective is to solve for the vector valued coordinates of the stimuli and the ideal points.

To explain our interest in seeking practical ways for implementing such a model in the fallible case a brief statement of its major assets, as we see them, may be helpful. We think such a model merits consideration as a promising research tool because

(i) it is based on a simple, unambiguous, and easily performed experimental task — p.c. judgments — in contrast to some of the presently popular scaling models where the experimental task is often left vague or even completely unspecified or, at the other extreme, unrealistically demanding (for the sake of mathematical tractability);

(ii) the input information (probabilities) so obtained is well defined and need not and cannot be tinkered with as in some of the presently widely used "non-metric" techniques;

(iii) the model involves an intuitively plausible rationale — the Coombs unfolding paradigm, and thus should stand some chance toward verification in practice;

(iv) on the other hand, the model is not tautological or only trivially non-tautological (as are some of the "non-metric" techniques, because any set of numbers can be converted into distances, indeed Euclidean distances, by a monotonic

transformation); rather it can be falsified explicitly with the help of chi-square tests (Mosteller, 1951) against the actually observed protocol information — see (ii) above;

(v) the model is thought to be more realistic than the presently available, usually one-dimensional metric alternatives (notably the Thurstone Case V model and the BTL-model), not only because it is multidimensional, but also because

(vi) it allows for individual differences in a natural way, as they can be expected to arise in actual applications, e.g., to voting behavior or consumer behavior;

(vii) finally, in spite of its generality, the model is mathematically tractable so that its properties can be studied in some depth before resorting to iteration.

The fact that the model possesses an exact algebraic solution does not mean that there is no room for the development of efficient algorithms for use with fallible data. It became clear fairly early that the algebraic solution is not very robust in the fallible case.

We had therefore undertaken some exploratory work to study the feasibility of a simple least-squares solution based on the method of steepest descent. This algorithm performed satisfactorily for smaller data sets and it was used for the analysis of some empirical data sets which were included in Schönemann and Wang, 1972. But for larger data sets its performance was not very satisfactory. The main objective of the work to be reported here was the development of more efficient algorithms for use with fallible data. It is natural that such algorithmic work is never fully completed and that there is still room for further improvement, at least in principle. Fairly extensive empirical work (Wang, 1973), on the other hand, makes it likely that such improvements may not be easy to achieve in practice, at least not in the near future, because even highly efficient algorithms, such as the Fletcher-Powell method (which requires storage for second order derivatives), did not perform markedly better than the conjugate gradient algorithm which we finally settled on. This algorithm will be described in some detail in section 3 after a brief review of the algebraic structure of the model has been given in section 2. Finally, in section 4, the results of an empirical study, which deals with an application of the model to political preference behavior, will be presented as evidence that both the model and the algorithm perform well in situations for which they were designed.

General Formulation of the Preference Model

The model can be stated in two alternative ways — as a special case of the Coombs unfolding model or as a multidimensional generalization of the BTL model. To formulate it as a metric case of the Coombs model, one needs three basic assumptions:

(a) A set of p stimuli S_i and a set of q subject-specific (or subgroup-specific) "ideal points" P_j can be located jointly in an m-dimensional Euclidean space. Thus the squared distances d^2_{ij} between points from distinct sets are

$$[2] \qquad d^2_{ij} = (\xi_i - \eta_j)'(\xi_i - \eta_j) \quad i = 1,2,\ldots,p; \quad j = 1,2,\ldots,q \ ,$$

where $\xi_i' = (x_{i1},x_{i2},\ldots,x_{im})$ and $\eta_j' = (y_{j1},y_{j2},\ldots,y_{jm})$ are the coordinate vectors of S_i and P_j, respectively.

(b) Following the BTL model, the probability $p_{ik.j}$ that all subjects whose ideal point is at P_j will prefer stimulus S_i over stimulus S_k in a p.c. situation is assumed to be a function of p subject-specific "scale values" a_{ij} which are determined up to subject-specific multiplicative constants b_j

$$[3] \qquad p_{ik.j} = Pr(S_i > S_k|P_j) = a_{ij} \ (a_{ij} + a_{kj})^{-1} = (a^*_{ij} + a^*_{kj})^{-1} \ ,$$

where $a^*_{ij} = b_j a_{ij}$, $b_j > 0$.

(c) A function is postulated which relates the scale values a_{ij} (a^*_{ij}) to the distances d_{ij}

$$[4] \qquad a^*_{ij} = b_j a_{ij} = exp\ (-cd^2_{ij}) \ ,$$

where $c > 0$ can be chosen arbitrarily.

Alternatively, one can combine equations [3] and [4] to express a direct relation between the observables $p_{ik.j}$ and the implied Euclidean distances d_{ij}

$$[5] \qquad \begin{aligned} p_{ik.j} &= 1/[1 + e^{-c}(d^2_{kj} - d^2_{ij})] \\ &= 1/\{1 + e^{-c}[(\xi_k - \eta_j)'(\xi_k - \eta_j) - (\xi_i - \eta_j)'(\xi_i - \eta_j)]\}. \end{aligned}$$

Thus the basic model is described by equations [2]-[4] or equivalently by equations [2] and [5].

In summary, the model defines a derived scale $[B^*, R, (\xi,\eta)]$. Its derived measurement system (Suppes & Zinnes, 1963) is $B^* = [SxSxP,p]$ where $S = (S_1,S_2,\ldots,S_p)$ and $P = (P_1,P_2,\ldots,P_q)$ are the stimulus set and the ideal point set respectively, and $p_{ik.j}$ is the p.c. probability of stimulus i preferred over stimulus k by the subject (subgroup) j $(0 < p_{ik.j} < 1,\ p_{ik.j} + p_{ki.j} = 1$ and $p_{ii.j} = .5)$. The representing relation $R[p,(\xi,\eta)]$ is given by equation [5].

Algebraic Solution

In the exact case, there exists a closed algebraic solution for the model. This solution can be discussed in two stages. At the first stage, the input p.c. probabilities are mapped into between-set squared distances. From equation [5] it follows that this is accomplished by an inverse logistic transformation of the probabilities

$$[6] \qquad u_{ik.j} = L^{-1}(p_{ik.j}) = ln\ p_{ik.j} - ln\ p_{ki.j}\ ,$$

where $u_{ik.j} = c(d^2_{kj} - d^2_{ij})$, and $L(x) = 1/(1 + e^{-x})$ is the logistic function on x. Upon averaging over the rows of each matrix $\mathbf{U}_j = (u_{ik.j})$ for each j and imposing the constraints $\sum_{i=1}^{p} d_{ij}{}^{2*} = 0$, one obtains the jth column of the between set squared distance matrix up to a subgroup-specific additive constant t_j and a multiplicative constant c,

$$[7] \qquad d_{kj}{}^{2*} = (c/p) \sum_{i=1}^{p} u_{ik.j} = c(d^2_{kj} - (1/p) \sum_{i=1}^{p} d^2_{ij})$$

$$= c(d^2_{kj} + t_j)\ ,$$

where

$$t_j = (1/p) \sum_{i=1}^{p} d^2_{ij}\ ,\ k = 1,2,\ldots,p.$$

Since the multiplicative constant c corresponds to a uniform dilation of the configuration and does not affect its shape, we can assume $c = 1$ for convenience. Thus the between set squared distances are determined up to a set of column additive constants $\tau' = (t_1, t_2, \ldots, t_q)$.

The second stage deals with a "generalized metric unfolding problem": the problem of solving for the coordinate vectors ξ_i, η_j in

$$[8] \qquad d^2_{ij} = d_{ij}^{2*} - t_j = (\xi_i - \eta_j)'(\xi_i - \eta_j), \quad i = 1,2,\ldots,p; \; j = 1,2,\ldots,q \; ,$$

given the pq d_{ij}^{2*} from the first stage.

This differs from the original "metric unfolding problem" [2] in that now only the d_{ij}^{2*}, not the d^2_{ij}, are assumed known, i.e., in the generalized problem the squared between set distances are given only up to some (unknown) column constants t_j. However, in Schönemann and Wang (1972) it was shown that Schönemann's (1970) solution of [2] is also a solution of the more general problem [8]. Therefore, the model [2]-[4] can be solved algebraically in the exact data case by first solving [6] and then [8].

Geometric Interpretation

The model can be interpreted as a metricized version of the multidimensional unfolding model. One way to see this is to set $b_j = (2\pi\sigma^2)^{m/2}$ and $c = \sigma^{-2}/2$. With these substitutions, equation [4] can be expressed as

$$[9] \qquad a_{ij} \approx \mathrm{N}_m \, (\xi_i; \, \eta_j, \, \sigma^2 I_m) \; .$$

This means that the scale value a_{ij} is proportional to the ordinate at point ξ_i (the location of stimulus S_i) of an m-variate normal distribution with mean vector $\mu = \eta_j$ (the location of the ideal point P_j) and covariance matrix $\Sigma = \sigma^2 I_m$. There are q such m-variate normal distributions, each having an identical covariance matrix Σ but a different mean vector for different person P_j.

These distributions can be regarded as defining circular "isopreference contours" (Green & Carmone, 1969, Figure 2) around each person j which decrease in intensity inversely with the dis-

tance d_{ij} between the stimulus S_i and the ideal point P_j of person j, which is at the center of this field. The subject-specific constant b_j serves as a unit which would raise or lower the isopreference contour of subject P_j but not affect the ratio of the ordinates under this contour at S_i and S_k, i.e., the p.c. probability $P_{ik.j} = (1 + a_{kj}/a_{ij})^{-1}$ would not be affected by different choice of the b_j.

In the above interpretation, the present model is a special case of the Coombs unfolding paradigm. The main difference is that the present model stipulates a specific mathematical form of the preference contours, whiles Coombs' original formulation does not. The gains from adopting such a stronger metric model are richer predictions, and the price to be paid for it is greater susceptibility to falsification.

Alternatively, one can interpret the present model as a multidimensional relative of the metric, but unidimensional BTL model. The BTL model asserts that

$$[10] \qquad P_{ik.j} = 1/[1 + e^{-(v[i] - v[k])}] = G(w_{ik.j}),$$

where $w_{ik.j} = v_i - v_k$, $G(x) = 1/(1 + e^{-x})$ is the logistic function, and v_i, v_k are the scale values. Now imagine an ideal point P_j with scale value v_j which corresponds to the maximum preference of an individual P_j. Let it be to the right of all stimuli. The term $w_{ik.j}$ in equation [10] can then be viewed as a difference in distances, i.e.,

$$[11] \qquad w_{ik.j} = v_i - v_k = (v_j - v_k) - (v_j - v_i) = d_{kj} - d_{ij} \ .$$

Thus equation [10] can be written

$$[12] \qquad P_{ik.j} = 1/[1 + e^{-(d[kj] - d[ij])}] \ .$$

Equation [12] is the same as equation [5] except that distances replace squared distances. This difference has a rather important implication. In equation [11], v_j cancels algebraically, and hence renders the ideal point irretrievable, since we are free to locate the ideal point on the v-scale anywhere to the right of the stimuli. In contrast, once we deal with differences in squared dis-

tances from the ideal point, as in equation [3], the location v_j of the ideal point P_j will affect the argument $u_{ik.j} = d^2_{kj} - d^2_{ij} = (v_j - v_k + v_j - v_i)\,(v_i - v_k) = (d_{kj} + d_{ij})\,w_{ik.j}$ and thus the observables $P_{ik.j}$. This enables us to recover the location of the ideal points, given sufficient information.

The more general multidimensional case is illustrated in Figure 1. The two stimulus points S_i and S_k are located somewhere on two concentric circles (in general, hyperspheres) with radii d_{ij} and d_{kj}, respectively, and origin at ideal point P_j. The magnitude $|d_{kj} - d_{ij}|$ of the difference in distances is simply the distance between the two concentric circles which now does depend on the common origin P_j, unless the origins are collinear with S_i and S_k. To see this, let us rotate S_i in Figure 1 to $S_i{}^*$ so that $|d_{kj} - d_{ij}|$ is simply represented by the length $|S_i{}^*, S_k|$ of the segment between $S_i{}^*$ and S_k. Similarly, S_i can be rotated to $S_i{}^{**}$ so that $|d_{kj'} - d_{ij'}|$, for an ideal point $P_{j'}$ is equivalent to the length $|S_i{}^{**}, S_k|$ of the segment between $S_i{}^{**}$ and S_k. It is obvious that $|S_i{}^*, S_k|$ is not the same as $|S_i{}^{**}, S_k|$.

In the present model, the difference in unsquared distances $w_{ik.j} = d_{kj} - d_{ij}$ is multiplied by the sum $d_{kj} + d_{ij}$ of the distances of both stimuli from the ideal point. This means, psychologically, that the preference between two stimuli S_i and S_k with fixed $d_{kj} - d_{ij}$ should become more pronounced as their joint distance $d_{kj} + d_{ij}$ from the ideal point increases.

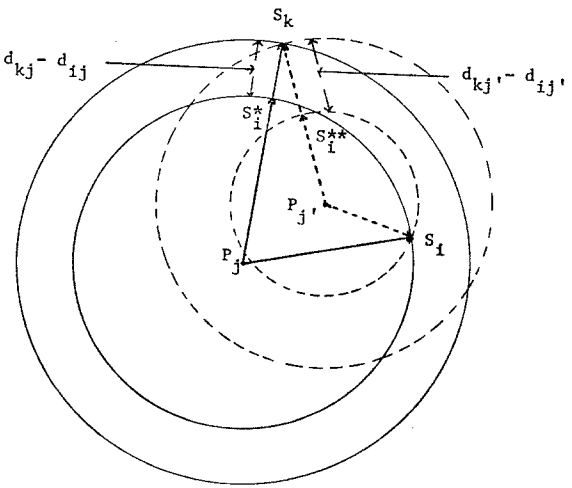

Figure 1. Geometry of the S & W model
$p_{ik.j} = G(d^2_{kj} - d^2_{ij}) = G(u_{ik.j})$.

The Fallible Case

Although an algebraic solution exists for the present model in the error-free case, a relatively robust solution (preferably with some optimal properties, e.g., least squares or maximum likelihood solution) is desirable for work with fallible data. Therefore a two-stage least squares (L.S.) solution has been developed. This solution, whose two stages correspond to those of the exact algebraic solution, will now be described.

Least Squares Fit of the BTL Model

At the first stage, the p.c. probabilities are mapped into differences between squared distances. Two problems may complicate this conversion. Firstly, there could be extreme probabilities (0 or 1) in the data. This could mean that a subject is capable of discriminating some stimuli perfectly, or it could be due to the small sample size of the estimates. Such extreme probabilities introduce technical difficulties, since $p_{ik.j} = 1$ implies an infinite mapping $d^2_{kj} - d^2_{ij} = G^{-1}(p_{ik.j}) = \infty$ when a response function G with unbounded range is used, as in the present model. Secondly, the probability estimates may contain error so that they do not conform exactly to the BTL model. Finally, there could be some p.c. probabilities which are not observed at all in the experiment. To cope with these problems, Gulliksen's (1956) L.S. solution for incomplete p.c. data was adopted which had been devised originally for Thurstone's Case V.[1]

In this procedure the L.S. estimates of the logarithmized BTL scale values $u_{ij} = \ln a_{ij}$ $(i=1,2,\ldots,p)$ for subject j are obtained so that f_j in [13] is minimized:

$$[13] \qquad f_j = \sum_{\substack{i=1 \\ (i,k)}}^{p} \sum_{\substack{k=1 \\ \varepsilon S_j}}^{n_{ij}} e^2_{ik.j} \ , \qquad \text{where}$$

$$e_{ik.j} = u_{ij} - u_{kj} - w_{ik.j} \ ,$$

$$w_{ik.j} = L^{-1}(p_{ik.j}) = \ln p_{ik.j} - \ln p_{ki.j} \ ,$$

1. A reviewer pointed out, quite correctly, that there are, of course, other and simpler ways of dealing with the missing data problem. The simplest, perhaps, would be to select a sufficiently large constant c for mapping unit probabilities into c and zero probabilities into $-c$. Whether such a course is more subjective than ours for handling missing data is not at all certain. It could be argued that perfect discriminations, if actually observed, contain more information than imperfect discriminations, and hence should not be discarded. Those who share this point of view should have no difficulty in adapting the program accordingly.

S_j = the set of all stimulus pairs (i,k) which are actually observed and not perfectly discriminated (i.e., $0 < p_{ik.j} < 1$, the pair (i,k) is treated as unobserved if the observed $p_{ik.j}$ is 0 or 1) for subject j, and n_{ij} = number of pairs involving stimulus i which are in S_j.

An algebraic solution for this L.S. problem is described in Gulliksen (1956). Schönemann (1970b) gives a slight reformulation which allowed him to derive a sufficient condition for a unique solution. Under the constraints $\sum_{i=1}^{p} \hat{u}_{ij} = \text{constant } c_j$, the vector $\hat{u}_j' = (\hat{u}_{1j}, \hat{u}_{2j}, \ldots, \hat{u}_{pj})$ of the L.S. estimates of u_{ij} is given by

$$[14] \qquad \hat{u}_j = N_j^{-1}(c_j J_p + w_j) \quad \text{(assuming } N_j \text{ is nonsingular)},$$

where $N_j = (n_{ik.j})$ is a pxp matrix with elements $n_{ii.j} = n_{ij}$, the number of observed probabilities in the ith row of the p.c. matrix for the jth subject with diagonal element counted as filled, and for $k \neq i$, $n_{ik.j} = 1$ if stimulus pair (i,k) is not in S_j, 0 otherwise. The vector w_j contains the elements

$$w_{ij} = \sum_{k=1}^{n_{ij}} w_{ik.j}, \ (i.k) \ \varepsilon S_j \ .$$

In practice, it is often convenient to define the estimates $\hat{u}_{ij}$ so that they sum up to zero (i.e., $c_j = 0$). The solution [14] will then determine the L.S. estimates of u_{ij} up to an additive constant if there are more filled cells (including diagonals) than empty ones in each row of the p.c. probability matrix for subject j, i.e., a sufficient condition for a unique solution of [13] is (Schönemann, 1970b):

$$[15] \qquad n_{ii.j} > \sum_{\substack{k=1 \\ k \neq 1}}^{p} n_{ik.j} \text{ for all } i = 1,2,\ldots,p.$$

These L.S. estimates $\hat{u}_{ij}$ can be mapped into estimates of the between set squared distances up to a multiplicative constant c and a set of q subject (column) specific constants t_j by

[16] $$\hat{d}_{ij}^{2*} = c\hat{d}^2_{ij} + t_j = -\hat{u}_{ij} \ .$$

If condition [15] is not met, there may be no determinate estimates of d_{ij}^{2*} for that particular subject. In this case, the subject can be dropped from further analysis. This condition implies that the solution [14] is remarkably tolerant in its acceptance of empty cells.

Least Squares Solution for the Generalized Metric Unfolding Problem

The objective of the second phase of the solution is to find estimates of the coordinate vectors ξ_i, η_j for the stimulus points and the ideal points, given the estimates of between set squared distances up to column-specific additive constants in equation [16]. We now deal with the fallible case of the generalized metric unfolding problem because the estimates of d_{ij}^{2*} in [16] are now containing error.

Schönemann's (1970a) algebraic solution can be applied to the fallible case only if there is sufficient information (i.e., if there are enough points in both sets). In this case it has L.S. properties in the sense that it is based on a rank m L.S. approximation to $\mathbf{C}_{12}$. However, it would be preferable to have some optimal solution with more clearly defined L.S. properties. Therefore a L.S. solution was developed for the generalized metric unfolding problem. Since the relevant information are squared distances, the loss function f was defined as

[17] $$f = \sum_{i=1}^{p} \sum_{j=1}^{q} e^2_{ij} \ , \text{ where}$$

$$e_{ij} = (d_{ij}^{2*} - t_j) - (\xi_i - \eta_j)' (\xi_i - \eta_j)$$

$$= (d_{ij}^{2*} - t_j) - \sum_{r=1}^{m} (x_{ir} - y_{jr})^2 \ .$$

The first order partial derivatives of f are

[18] $$\partial f / \partial x_{ir} = -4 \sum_{j=1}^{q} e_{ij} (x_{ir} - y_{jr})$$

$$= 4 \sum_{j=1}^{q} (x_{ir} - y_{jr}) \left[\sum_{r=1}^{m} (x_{ir} - y_{jr})^2 - d_{ij}^{2*} + t_j \right],$$

[19] $$\partial f / \partial y_{jr} = 4 \sum_{i=1}^{p} e_{ij} (x_{ir} - y_{jr})$$

$$= -4 \sum_{i=1}^{p} (x_{ir} - y_{jr}) \left[\sum_{r=1}^{m} (x_{ir} - y_{jr})^2 - d_{ij}^{2*} + t_j \right],$$

[20] $$\partial f / \partial t_j = -2 \sum_{i=1}^{p} e_{ij} = 2 \sum_{i=1}^{p} \left[\sum_{r=1}^{m} (x_{ir} - y_{jr})^2 - d_{ij}^{2*} + t_j \right],$$

$$i = 1,2,\ldots,p; \; j = 1,2,\ldots,q; \; r = 1,2,\ldots,m.$$

Upon setting these derivatives to zero, one obtains the normal equations $\partial f / \partial x_{ir} = 0$, $\partial f / \partial y_{jr} = 0$, and $\partial f / \partial t_j = 0$. These equations do not seem to have an explicit algebraic solution. Therefore some iterative search method (see, for example, Wilde & Beightler, 1967) had to be used to solve the minimization problem. Such methods usually involve iteration and can be quite time-consuming. The particular choice among the many available algorithms depends upon the shape of function involved. Minimization algorithms which have been found to be relatively efficient for the present case will be considered in the next section.

Numerical Minimization Algorithm for the Second Stage

General considerations. A common feature of many minimization techniques is that they are based on the vector of first order partial derivatives (the "gradient"). This vector points in the direction of steepest slope of the surface of the function. The techniques search along this direction at each iteration until an optimal point is reached where the gradient vector vanishes. Two well-known examples are the optimal gradient method (Cauchy, 1847; which finds an optimal step size at each cycle) and the method of steepest descent ("gradient methods").

More sophisticated algorithms have been developed in the last few decades. A variant of the gradient method is the method of

resultant gradients (Finkel, 1959). This method was found by Jöreskog (1966) to be more efficient than the regular gradient method in applications to factor analysis problems. The efficiency of any particular algorithm depends, of course, on the nature of the problem dealt with. For the present problem [17], several other methods were studied. All are designed to have "quadratic convergence," i.e., they converge in n steps for n unknowns *if* the criterion function is quadratic. Specifically, we experimented with the method of deflected gradients (Fletcher & Powell, 1963), the method of conjugate gradients (Fletcher & Reeves, 1964), and a modified version of gradient method with variable step size.

Before adopting any specific algorithm it can be useful to appraise its prospective practicality in more general terms. Some methods, e.g., the method of deflected gradients, require storage space for the Hessian (the matrix of the second order derivatives). They quickly become unrealistic as the number of unknowns increases. Further, a method which involves extensive computational effort per cycle may not be faster than a simpler one in the long run. Thirdly, the particular outcome may depend on the shape of the surface. For example, the Fletcher and Powell method has been shown to be excellent for Rosenbrock's (1960) banana-shaped function with two unknowns. But it did not prove superior in speed of convergence to the other methods for the present problem. In addition, the number of unknowns (n) in the present problem is going to be large as the number of points (p,q) of both sets and the number of dimensions (m) increase [$n=(p+q)m+q$]. Thus any algorithm which requires core storage for the Hessian (a square matrix of order n) would present practical problems considering the capacity of even modern computers, no matter how superior it may be in theory.

The conjugate gradient method. This algorithm was designed by Fletcher and Reeves (1964). It does not need storage for the Hessian, but has quadratic convergence. The basic characteristic of this algorithm, which thus seems to combine the best of both possible worlds, is that it utilizes the information of the previous points to form a direction vector which guides the search to the minimum. It is a variant of the method of conjugate gradients (Hestenes & Stiefel, 1952) for solving a set of simultaneous linear equations with a symmetric positive definite matrix of coefficients.

The general principles of this method are described in Beckman (1960) and Fletcher and Reeves (1964).

(i) General description.

In the method of conjugate gradients, a set of n direction vectors $p_1\ p_2, \ldots, p_n$ are generated such that p_{i+1} is a linear combination of $-g_{i+1}$ and $p_1, p_2, \ldots, p_n$ are H-conjugate (H-orthogonal, i.e., $p_i'Hp_j = 0$, $i,j = 1,2,\ldots,n$, $i \neq j$, where H is the Hessian at the minimum). Fletcher and Reeves consider a simple form of this method where most of the coefficients in this linear combination are zero. The result is:

$$p_{i+1} = -g_{i+1} + \beta_i p_i \ ,$$

where $\beta_i = g'_{i+1}g_{i+1}/g_i'g_i$, $i = 1,2,\ldots,n$.

The process starts at an arbitrarily selected initial point x_1 (see next section) and the negative gradient vector $-g_1 = -g(x_1)$ serves as the initial direction vector p_1 $(= -g_1)$. Let p_i be the direction vector used at step i along which a new point x_{i+1} can be found so that the maximum reduction of the criterion function f is produced in the direction of p_i, and g_i be the gradient vector at the point x_i. Then, at each step i, the following computations are performed:

[21] $$x_{i+1} = x_i + \gamma_i p_i \ ,$$

where γ_i is the step size which minimizes the function $f^*(\gamma) = f(x_i + \gamma p_i)$ under choice of γ. The problem of finding such a γ is usually called the "problem of linear search." In practice, γ_i can usually not be obtained by differentiation. Rather, a quadratic interpolation technique is used to approximate the value of γ_i. This step size γ_i is then used to complete the computational cycle:

[22] $$g_{i+1} = g(x_{i+1}) \ ,$$

[23] $$\beta_i = g'_{i+1}g_{i+1}/g_i'g_i \ ,$$

[24] $$p_{i+1} = -g_{i+1} + \beta_i p_i \ ,$$

where p_{i+1} is the new direction vector along which a new point x_{i+2} is to be located at the next iteration $i+1$.

This procedure can be shown to converge to the minimum in no more than n iterations for any quadratic functions of n arguments. For higher order functions (as in our case), the convergence will be slower and its rate will depend on the closeness of the local quadratic approximation to the surface of the criterion function.

Fletcher and Reeves (1964) have modified the above basic process so that, after every $n + 1$ iterations, the process is restarted at the current x (i.e., $x_{n+2} = x_1$). They found that such a modification speeds up convergence, which therefore, was adopted for our present problem [17].

(ii) A quadratic interpolation technique for the linear search.

Fletcher and Reeves (1964) use Davidon's cubic interpolation procedure to solve the linear search problem [21]. In the present application, we preferred a simpler quadratic interpolation technique for finding the value γ_i at iteration i. A small positive value γ_o is taken as the initial step size. At each iteration i, a point $x_c = x_i + \gamma_{i-1}p_i$ along the direction of p_i is found. Its function value $f_c = f(x_c)$ is compared with $f_i = f(x_i)$. If $f(x_c)$ is less than f_i, a series of points $x_a = x_i + 2^a\ \gamma_{i-1}p_i$ $(\alpha = 1,2,\ldots)$ and their function values $f_a = f(x_a)$ are computed until some α is reached so that f_a is not less than f_{a-1}. Then, as a lower bound for the value γ_i, we use $a = 2^{a-1}\gamma_{i-1}$. If f_c is not less than f_i, we set $a = 0$.

The next stage is a search between $x_a = x_i + ap_i$ and x_a (when $a = 0$, we have $\alpha = 0$) in the direction of p_i to find an upper bound b for the value of γ_i. To obtain b, we examine the middling points $x_u = 2^{-u}[(2^u - 1)x_{a-1} + x_a]$ and the function values $f_u = f(x_u)$ $(u=1,2,\ldots)$ until some integer u is reached for which $f_u \leq f_{a-1}$. This means that x_{u-1} is a limit point where further decrease of the criterion function along the direction of p_i cannot be obtained. The upper bound of γ_i is then set to $b = 2^{-u+1}\,[(2^{u-1} - 1).2^{a-1}\gamma_{i-1} + 2^a\ \gamma_{i-1}] = 2^{a-u+1}(2^{u-2} + \tfrac{1}{2})\ \gamma_{i-1}$. In the case $a = 0$, $b = 2^{-u+1}\gamma_{i-1}$ is taken as the upper bound of γ_i. To avoid unnecessary waste of time, this intensive search for γ_i is limited to $u \leq 4$. Experience showed that the improvement in the approximation to γ_i by more intensive search for a very precise range of γ_i is not worth the extra time spent in the search.

Having obtained a range for γ_i $(a \leq \gamma_i \leq b)$, the best value of γ_i is approximated by quadratic interpolation of the three points $x_a = x_i + ap_i$ $(x_a = x_i$ if $a = 0)$, $x_u = \tfrac{1}{2}$ $(x_a + x_b) = x_i + (a + b)$ $p_i/2 = x_i + 2^{a-u}(2^{v-1} + \tfrac{1}{2})\ \gamma_{i-1}p_i$ (in the case $a = 0$, $b = 2^{-u+1}\ \gamma_{i-1}$,

we have $x_u = x_i + 2^{-u}\,\gamma_{i-1}p_i)$, and $x_b = x_{u-1} = x_i + bp_i$ as a function of γ, i.e., $f^*(\gamma) = f(x_i + \gamma p_i)$ where the corresponding values of γ assumed by the three points are a, $a + b/2$, and b, respectively.

Thus, the value $\gamma_{min} = \gamma_i$ which gives the minimum of f^* under choice of γ can be approximated by

$$[25] \qquad \gamma_i = [(f_1^* - f_3)h/2(f_1^* + f_3^* - 2f_2^*)] + h_o \ ,$$

where $f_1^* = f(x_a) = f^*(a)$, $f_2^* = f(x_u) = f^* \quad (a+b)/2$, $f_3^* = f(x_{u-1}) = f^*(b)$ and $h = (b-a)/2$, $h_o = (a+b)/2$ [i.e., $f_1^* = f^*(h_o - h)$, $f_2^* = f^*(h_o)$ and $f_3^* = f^*(h_o + h)$ in terms of $f^*(\gamma) = f(x_i + \gamma p_i)$].

The Fletcher and Reeves method with such a quadratic interpolation procedure [instead of Davidon's method in the linear search (eq. [21])] will be called "conjugate gradient method" (CJGMC) in the sequel.

(iii) "Intervening conjugate gradient method" (CJGMI).

Another algorithm was tried which iterates only for X and Y. At every point of the linear search, estimates of the additive constants t_j for the given values of X and Y are obtained algebraically from equation [23], which can be solved exactly, given X, Y, and $d_{ij}{}^{2*}$:

$$[26] \qquad \hat{t}_j = 1/p \sum_{i=1}^{p} [d_{ij}{}^{2*} - \sum_{r=1}^{m} (x_{ir} - y_{jr})^2] \ , \ j = 1,2,\ldots,q$$

These values of $\hat{t}_j$ are then entered into the computations [21]-[24] as required by the conjugate gradient method at all the points studied in the linear search process. Experience with this method showed that the extra efforts spent in obtaining L.S. estimates of t_j at every intermediate point of the linear search process paid off in terms of overall speed of convergence.

This revised form of the conjugate gradient method will be called "intervening conjugate gradient method" (CJGMI) in the sequel.

(iv) Summary of the intervening conjugate gradient method for the generalized metric unfolding problem.

The intervening conjugate gradient method has been applied

to the generalized metric unfolding problem [8] with satisfactory results. It requires only slightly more storage than the gradient methods (GRDMC and GRDMI, analogously defined) n more locations are needed for saving the direction vector of the previous iteration. It does not require much more computational work and is not difficult to program. In practical applications, it performed much better than the gradient method, especially when the number of unknowns was relatively large. In particular, it considerably improved the rate of convergence in application to the metric unfolding problem.

A subroutine CJGMI has been written for solving the L.S. problem [17] by the intervening conjugate gradient method. Its major computational steps are given in Table 1.

Table 1. Flow Chart for CJGMI

Iterative process:

DO [5] to [6.4] ($L = 1,2,\ldots N$, number of major cycles each consisting of $n+1$ inner iterations).

DO [5] to [6.2] ($i = 1,2,\ldots,n+1$, n is the number of unknowns) (Step [5] provides a practical way for obtaining the bounds a, b by the procedure described in (ii) of *The Conjugate Gradient Method.*

[5] set $a = 0$, $b = 0$ and $f_1{}^* = f_i$.

[5.1] compute $x_a = x_i + c_o p_i$ and $f_a = f(x_a, \tau_a)$ where τ_a is obtained from eg. [26], given $x_a = (X_a, Y_a)$.

[5.2] if $f_a < f_1{}^*$, set $a = c_o$, $c_o = 2c_o$, $f_1{}^* = f_a$ and GO TO [5.1].

[5.3] set $b = c_o$, $f_3{}^* = f_a$, $u = 0$.

[5.4] compute $x_u = x_i + c_1 p_i$ and $f_u = f(x_u, \tau_u)$, set $u = u+1$, where $c_1 = (a+b)/2$.

[5.5] if $f_u > f_1{}^*$, and $u < 4$, set $b = c_1$, $f_3{}^* = f_u$ and GO TO [5.4]. (Note: [5.4]-[5.5] is not repeated more than 4 times for practicality.)

[5.6] set $f_2{}^* = f_u$.

[5.7] compute γ_i from eq. [25] given $f_1{}^*$, $f_2{}^*$, $f_3{}^*$ and $h = (b-a)/2$, $h_o = (a+b)/2$ ($a \leq \gamma_i \leq b$), and set $c_o = \gamma_i$ to be used in [5].

[5.8] obtain a new point $x_{i+1} = x_i + \gamma_i p_i$, its corresponding gradient vector $g_{i+1} = g(x_{i+1}, \tau_{i+1})$ and its function value $f_{i+1} = f(x_{i+1}, \tau_{i+1})$ where τ_{i+1} is the L.S. estimate of τ given x_{i+1}.

[5.9] construct a new direction vector $p_{i+1} = -g_{i+1} + \beta_i p_i$ from eqs. [23]-[24].

Test of convergence:

[6] check stopping criteria.

[6.1] if stopping criteria (e.g., $g_{i+1} \approx \Phi$ or $g_i' g_i \leq \varepsilon$ (a very small positive value such as 10^{-20})) are met, GO TO [7].

[6.2] if number of inner iterations $n+1$ is not completed, GO TO [5].

[6.3] if number of major cycles N is completed, GO TO [7].

[6.4] replace $x_1 = x_{n+2}$ and $p_1 = -g_{n+2}$, $c_o = \gamma_{n+1}$, then GO TO [5].

Starting configuration and stopping criteria. Two minor technical problems with iterative algorithms are the choice of a starting point and the problem of setting some criteria to terminate the iterative process. These two problems will now be discussed in detail:

(i) The choice of a starting point.

In principle, any starting point is equally satisfactory for quadratic functions when the iterative algorithm has quadratic convergence. For functions of higher order, one would wish to start with a location from which the minimization process will lead to the minimum as quickly as possible. Sometimes it may be possible to find a good starting point by some prior theoretical analysis of the criterion function, or sometimes from one's past experience with the particular problem. But since we usually lack such prior knowledge of an approximate solution, it is helpful to have a general way of generating a starting point to bring about reasonably fast convergence.

In the present case, we decided to use an Eckart-Young decomposition of the quasi-scalar product matrix $\mathbf{C}_{12}$ for this purpose. Thus, if

$$\mathbf{C}_{12} = \mathbf{V}\mathbf{D}_m\mathbf{W}' = \mathbf{V}\mathbf{D}_m^{1/2}\mathbf{D}_m^{1/2}\mathbf{W}' = \mathbf{G}_1\mathbf{H}_1' \ ,$$

(where $\mathbf{D}_m$ contains the positive square roots of the latent roots of $\mathbf{C}'_{12}\mathbf{C}_{12}$, and $\mathbf{G}_1 = \mathbf{V}\mathbf{D}_m^{1/2}$, $\mathbf{H}_1 = \mathbf{W}\mathbf{D}_m^{1/2}$) the iterative process starts with $\mathbf{X}_1 = \mathbf{G}_1$ and $\mathbf{Y}_1 = \mathbf{H}_1$. This particular choice is partially based on its practical convenience but also on some theoretical insights gleaned from the algebraic solution. Our preference for this starting point is largely, but not exclusively, empirically motivated. In Schönemann (1970a) it was shown that $\mathbf{G}$ and $\mathbf{H}$ relate to the solution matrices $\mathbf{X}$ and $\mathbf{Y}$ through a joint non-singular matrix $\mathbf{T}$. Extensive experience with the algebraic solution showed that the product $\mathbf{M} = \mathbf{T}\mathbf{T}'$ is practically always near diagonal, so that $\mathbf{T}$ can be chosen near diagonal for this choice of $\mathbf{G}_1$, $\mathbf{H}_1$. Subsequent experience with the L.S. solution confirmed that this particular starting point compares favorably with all others which were tried.

If the original modified gradient method or conjugate gradient method is used, the initial values τ_1 of the column additive constants τ can be obtained from eq. [26], given $\mathbf{X}_1$ and $\mathbf{Y}_1$.

Finally, we note that the Eckart-Young roots of $\mathbf{C}_{12}$ provide a basis for selecting m. In the fallible case, the number of dominant

roots m_o (the remaining roots should be much smaller and near zero) is an estimate of the true dimensionality m. In the exact case, this number m_o is equal to the dimensionality of the underlying common space (if a subspace case is involved, m_o will be the dimensionality of the joint subspace).

(ii) Stopping criteria.

Although the formal requirement for a point x_i to be at the minimum is that the gradient g_i vanishes, one does not expect in practice that all elements in g_i vanish completely because of accumulated rounding errors. Hence, one usually decides on a small positive value epsilon (ε) as the acceptable size of the elements in g_i for x_i to be considered a minimum. The particular size of ε depends on the practical situation. In some cases a very precise solution may be crucial, and one will adopt a very stringent acceptance criterion (e.g., a very small ε). In other cases, a somewhat less stringent criterion may be more appropriate. Frequently, a lenient criterion (e.g., $\varepsilon \leq .01$) is sufficient to yield a satisfactory approximation to the solution, e.g., when the model is used in some exploratory study where an extremely high accuracy is not warranted. In general the user is advised not to insist on a too stringent value for ε.

Other convergence criteria, such as the length of g, the reduction of the criterion function f, the amount of changes of the unknowns etc., have been employed in various iterative minimization algorithms. However, it is not rare to find that a "nonsignificant reduction" of the criterion function f at a particular iteration is achieved far from the actual minimum so that solution can still be improved and the size of g remains significantly large. Similarly, the length of g and certain other combined indices of the total changes in the unknowns at an iteration may be too vague as criteria for confident judgment of the convergence.

We adopted the maximum gradient element as a stopping criterion. Upon exit, these routines print the gradients as well as the gradual reductions of f in the searching paths for inspection by the user. This information enables the user to decide how satisfactory a solution is for his purposes or whether to continue the iterations in an effort to improve the solution.

Finally we note that in the method of conjugate gradients some constant β_i, which is the ratio of the squared length of two gradient vectors (eq. [23]), must be calculated at each iteration. Hence care must be taken that the denominator $g_i'g_i$ does not ap-

proach zero to cause arithmetic errors (overflows), and at each iteration the length of the current gradient g_i should be checked. If it is near zero (e.g., less than 10^{-20} since the numerator is also negligible in the vicinity of a minimum), the process will terminate and the current point is used as an estimate of the minimum.

Empirical comparison of different algorithms. In addition to the methods described above (CJGMC and CJGMI), a subroutine FMFPC (which is based on the method of deflected gradients, Fletcher and Powell, 1963) was adapted for the generalized metric unfolding problem. Two more subroutines, GRDMC and CJGMC, based on the original modified gradient method and conjugate gradient method respectively, were also written. We thus have experimented with five subroutines (GRDMC, GRDMI, CJGMC, CJGMI and FMFPC) and applied them to several constructed data sets. The comparisons among them provide a basis for our choice among these algorithms. In Table 2, these five routines are compared in terms of their convergence speed for two problems (a) and (b) (described in Table 2). The problems chosen were of different size in order to study the relation of the efficiency to the size of a problem.

The starting point was generated in all cases from the Eckart-Young decomposition of $\mathbf{C}_{12}$ as discussed in (i) of the previous section. The convergence criterion was $\varepsilon = .0005$ for the first four routines. The subroutine FMFPC has a built-in criterion (the length $|g|$ of g, $|g| \leq 10^{-8}$). The following four measures are presented for comparison: (i) the processing time t used for the total iterative process (from entering to the exit of the iterations); (ii) the criterion function $\hat{f}_{min}$ upon exit of the process; (iii) the total number of iterations n_t (for CJGMC and CJGMI, $n_t = n\text{x}N$); and (iv) the maximum absolute magnitude g_{max} of all the elements in the gradient vector upon exit. Note that n_t is not a very good index of efficiency, because different amounts of computations are involved per iteration for different routines. Hence, a meaningful comparison should be based on t.

For the smaller problem (a), CJGMI and GRDMI are fastest in convergence, FMFPC is second. Both GRDMC and CJGMC failed to converge within a minute. For the bigger problem (b), GRDMC and CJGMC again failed to converge. GRDMI did not converge within the specified number of iterations (5000) and took longer than for CJGMI and FMFPC to converge. FMFPC was

Table 2. Comparison of GRDMC, GRDMI, CJGMC, CJGMI, and FMFPC for the Generalized Metric Unfolding Problem

Problem (a): An example of fallible data from Schönemann (1970a, Table 3, p. 363f). A set of column constants $\tau_q' = (-5,-10,-15,-4,-8)$ is introduced to yield $\Delta_{12}^{(2)*} = \Delta_{12}^{(2)} + j_p\tau_q'$ to be used in the present problem. $p=8$, $q=5$, $m=2$.

Problem (b): An exact case ($f_{min} = 0$) with $\tau_q' = \Phi'$ constructed for testing the subroutines. $p=11$, $q=6$, $m=3$. A solution should reproduce the between and within set distances exactly and the vector of column constants τ should be null.

t: Time in seconds spent in the iterative process alone.

$\hat{f}_{min}$: Criterion function given the output configuration $\hat{X}$, $\hat{Y}$, and $\hat{\tau}$.

n_t: Total number of iterations (in CJGMC and CJGMI, $n_t = n \times N$).

g_{max}: Maximum (in magnitude) of the elements in the gradient vector g given $\hat{X}$, $\hat{Y}$ and $\hat{\tau}$.

* in entry g_{max}: fail to converge by the criterion $\varepsilon = .0005$ (with the exception of FMFPC where the criterion is $|g| \leq 10^{-8}$).

	routine	t	$\hat{f}_{min}$	n_t	g_{max}
	GRDMC	87.969	66.3375	3000	1.3483*
problem (a)	GRDMI	12.136	63.2568	362	.0005
$p=8,q=5,m=2$	CJGMC	77.793	63.2568	960	.0261*
fallible	CJGMI	11.431	63.2568	92	.0004
	FMFPC	15.176	63.2568	88	.0018
	GRDMC	298.819	2.11051	5000	.1635*
problem (b)	GRDMI	349.212	1.37964×10^{-2}	5000	.0452*
$p=11,q=6,m=3$	CJGMC	280.426	4.23394×10^{-2}	1740	1.6045*
exact	CJGMI	239.085	1.21012×10^{-7}	907	.0004
	FMFPC	153.105	9.15695×10^{-11}	259	.0019

faster than CJGMI for this particular problem. Nevertheless, we are content with the performance of CJGMI because the excessive core storage requirement makes FMFPC an impractical routine for bigger problems.

All these routines can be adapted to provide L.S. solutions for the metric unfolding problem which differs from [17] only in that all $t_i = 0$. We therefore have three routines GRDM, CJGM and FMFPM (revised from FMFPC) for this problem. Table 3 gives a comparison of their efficiency. For problems (a) and (b), CJGM did best. For problem (b), FMFPM is more efficient than GRDM, but the convergence rates in both do not differ significantly for the smaller problem (a).

We therefore concluded that for the metric unfolding problem the intervening conjugate gradient is to be preferred. The column constants introduce complications in the generalized prob-

Table 3. Comparison of GRDM, CJGM, and FMFPM for the Metric Unfolding Problem

Problem (a): A fallible case from Schönemann (a metric unfolding problem corresponding to problem (a) in Table 2). $p=8$, $q=5$, $m=2$.

Problem (b): An exact case (problem (b) in Table 2 treated as a metric unfolding problem). $p=11$, $q=6$, $m=3$.

t, $\hat{f}_{min}$, n_t, g_{max}: as explained in Table 2.

**: $\hat{f}_{min} = 1.63156\times10^{-9}$ reached at $n_t = 143$ or $t \approx 60$.

	routines	t	$\hat{f}_{min}$	n_t	g_{max}
problem (a)	GRDM	9.235	82.2515	323	.0005
$p=8,q=5,m=2$	CJGM	6.068	82.2515	76	.0004
fallible	FMFPM	9.775	82.2515	74	.0001
problem (b)	GRDM	203.676	1.23115×10^{-7}	3501	.0004
$p=11,q=6,m=3$	CJGM	36.645	5.13554×10^{-9}	233	.0005
exact	FMFPM	67.900	1.40235×10^{-23}**	154	.0001

lem [17]. In this case, the superiority of CJGMI is less pronounced, especially for the bigger problem, but it still is the most satisfactory algorithm at the present stage of our knowledge. The occasionally somewhat better convergence of FMFPC does not convince us of its absolute superiority, since the storage problem often becomes unmanageable when the number of unknowns is large, as it is likely to be in the applications to the present model. The simpler intervening modified gradient method was found to perform satisfactorily for most problems of moderate size.

An iterative algorithm is usually more time-consuming than an algebraic solution where it is applicable. However, this fact does not vitiate the value of iterative L.S. algorithms. The L.S. solution for the present problem [17] has its own merits. (i) It is robust; (ii) it does not impose as severe restrictions on the number of points in each set for a solution to be possible; (for example, a matrix of between set distances of two sets each having three points in a two-dimensional Euclidean space ($p=3$, $q=3$, $m=2$) can be solved by the present L.S. routines, but it should be pointed out that the solution might not be determined in this case); (iii) its solution possesses a clearly defined L.S. property with the loss function f.

A final word on the problem of local minima. So far as we presently know local minima do not seem to arise as long as there are sufficiently many points (e.g., enough for a determinate algebraic solution). Of course, any local minimum problem which

arises in the metric case may also arise in the nonmetric case. There might even be more local minima in the latter due to the greater degree of indeterminacy. The user is warned that the present L.S. algorithms, like any others, do not preclude the possibility of converging to a local minimum.

In Table 4, a numerical example of fallible data is given to illustrate the present L.S. solution for problem [17]. The between set squared distance matrix $\Delta_{12}^{(2)*}$ (determined up to a set of column additive constants τ) is generated from the data $\Delta_{12}^{(2)}$ con-

Table 4. Numerical Example of a L.S. Solution of the Generalized Metric Unfolding Problem in the Fallible Case ($p = 8, q = 5, m = 2$)

$$\Delta_{12}^{(2)} = \begin{bmatrix} 16 & 9 & 9 & 16 & 36 \\ 9 & 81 & 81 & 4 & 16 \\ 9 & 25 & 36 & 16 & 4 \\ 4 & 16 & 16 & 9 & 9 \\ 16 & 4 & 9 & 25 & 25 \\ 16 & 1 & 4 & 25 & 25 \\ 9 & 81 & 81 & 9 & 4 \\ 16 & 36 & 49 & 16 & 4 \end{bmatrix} \qquad \Delta_{12}^{(2)*} = \begin{bmatrix} 11 & -1 & -6 & 12 & 28 \\ 4 & 71 & 66 & 0 & 8 \\ 4 & 15 & 21 & 12 & -4 \\ -1 & 6 & 1 & 5 & 1 \\ 11 & -6 & -6 & 21 & 17 \\ 11 & -9 & -11 & 21 & 17 \\ 4 & 71 & 66 & 5 & -4 \\ 11 & 26 & 34 & 12 & -4 \end{bmatrix}$$

$\tau_1' = (-5 \; -10 \; -15 \; -4 \; -8)$ $\qquad \Delta_{12}^{(2)*} = \Delta_{12}^{(2)} + \mathbf{J}_p \tau_q'$

$$\mathbf{X}_1' = \begin{bmatrix} -2.5215 & 3.8289 & .1241 & -.8115 & -2.6228 & -2.9181 & 4.0461 & .8748 \\ 1.9157 & 1.5896 & -1.4811 & .0847 & -.1736 & -.1312 & -.0832 & -1.7209 \end{bmatrix}$$

$$\mathbf{Y}_1' = \begin{bmatrix} 2.1191 & -4.0557 & -3.9715 & 2.7849 & 3.1233 \\ .8768 & -.9724 & .7766 & 1.7718 & -2.4438 \end{bmatrix}$$

$\hat{f}_{min} = 63.2568$

L. S. estimates of column constants and coordinate vectors (with origin at the joint centroid):

$\hat{\tau}' = (-5.8834 \quad -12.9899 \quad -15.3569 \quad -5.1509 \quad -8.7033)$

$$\hat{\mathbf{X}}' = \begin{bmatrix} -2.1182 & 4.3229 & .4479 & -.4438 & -2.3477 & -2.5423 & 5.5038 & 1.2073 \\ 2.5233 & 2.2949 & -2.2034 & -.1915 & -.9967 & -.3929 & -.1597 & -2.8492 \end{bmatrix}$$

$$\hat{\mathbf{Y}}' = \begin{bmatrix} 1.4481 & -4.5633 & -4.5417 & 2.0711 & 2.5559 \\ .9283 & -.1593 & 1.0129 & 1.5146 & -1.3214 \end{bmatrix}$$

(Table continued on next page.)

Reproduced within and between set distances:

$$\hat{\Delta}_{11} = \begin{bmatrix} 0. & & & & & & & \\ 6.4452 & 0. & & & & & & \\ 5.3783 & 5.9372 & 0. & & & & & \\ 3.1897 & 5.3763 & 2.2006 & 0. & \text{symmetric} & & & \\ 3.5275 & 7.4385 & 3.0448 & 2.0671 & 0. & & & \\ 2.9469 & 7.3727 & 3.4956 & 2.1081 & .6345 & 0. & & \\ 7.1449 & 2.4613 & 4.5417 & 4.9477 & 6.9024 & 7.0500 & 0. & \\ 6.3185 & 6.0141 & .9970 & 3.1288 & 4.0087 & 4.4826 & 4.2544 & 0. \end{bmatrix}$$

$$\hat{\Delta}_{22} = \begin{bmatrix} 0. & & & & \\ 6.1089 & 0. & & \text{symmetric} & \\ 5.9904 & 1.1724 & 0. & & \\ .8555 & 6.8423 & 6.6318 & 0. & \\ 2.5077 & 7.2134 & 7.4716 & 2.8772 & 0. \end{bmatrix}$$

$$\hat{\Delta}_{12} = \begin{bmatrix} 3.9067 & 3.6297 & 2.8557 & 4.3090 & 6.0522 \\ 3.1831 & 9.2189 & 8.9569 & 2.3832 & 4.0249 \\ 3.2875 & 5.4120 & 5.9364 & 4.0569 & 2.2851 \\ 2.1985 & 4.1196 & 4.2712 & 3.0391 & 3.2055 \\ 4.2560 & 2.3686 & 2.9753 & 5.0826 & 4.9143 \\ 4.2035 & 2.0344 & 2.4442 & 4.9922 & 5.1821 \\ 3.2437 & 9.0671 & 9.1212 & 2.9532 & 2.2680 \\ 3.7852 & 6.3667 & 6.9259 & 4.4485 & 2.0378 \end{bmatrix}$$

Reproduced between set squared distances up to column constants:

$$\hat{\Delta}_{12}^{(2)*} = \begin{bmatrix} 9.4289 & .1850 & -7.2022 & 13.4166 & 27.9256 \\ 4.2990 & 71.9979 & 64.8687 & .5288 & 7.4967 \\ 4.9744 & 16.2998 & 19.8836 & 11.3075 & -3.4817 \\ -1.0001 & 3.9809 & 2.8865 & 4.0850 & 1.5716 \\ 12.2801 & -7.3797 & -6.5043 & 20.6816 & 15.4470 \\ 11.8356 & -8.8512 & -9.3830 & 19.7715 & 18.1508 \\ 4.6879 & 69.2218 & 67.8396 & 3.5707 & -3.5595 \\ 8.4943 & 27.5456 & 32.6110 & 14.6384 & -4.5505 \end{bmatrix}$$

structed by Schönemann (1970a, Table 3, p. 363f). The matrix $\Delta_{12}^{(2)*} = \Delta_{12}^{(2)} + \mathbf{J}_p\tau_q'$ (see Table 4) is used as input to the above five subroutines for problem [17]. All of the routines yield identical solutions, so that the reproduced between and within set distances are the same in all cases. This means that the L.S. estimates of X, Y obtained from these routines are related to one another by a rigid motion. The specific contents of Table 4 are taken from

the output of CJGMI. The agreement among these results suggests that there was no local minimum in this case. To confirm this, several other, arbitrary, starting configurations were tried which all led to the same solution.

Empirical Verification of the S&W Model: An Analysis of the 1968 Presidential Election

A fairly substantial body of voting data were collected, nationwide, by the Survey Research Center, Institute of Social Research, The University of Michigan, at the time of the 1968 Presidential election. These data are based on a large and carefully chosen sample, and they include much stratifying information about the political and socioeconomic background of the interviewees. Thus they seem ideally suited for an analysis by our model. Moreover, previous research by Weisberg and Rusk (1970) has already established a rather thorough understanding of the substantive content of these data by means of other methods of analysis, which can be used to check and corroborate the results of our own analysis.

Description of the Study

The data were taken from interviews of 1673 respondents in the 1968 election study of the University of Michigan's Survey Research Center. The respondents were asked to rate twelve candidates on a 0-to-100 ("feeling thermometer") scale (see Weisberg and Rusk, 1970). For a detailed description of the data collecting procedures the reader is referred to "The SRC 1968 American National Election Study" (Inter-Consortium for Political Research Edition, 1971, SRC 45523, Ann Arbor, Michigan). It was assumed that an individual's preference order of the candidates corresponds to the order of the scores given to the candidates. Weisberg and Rusk computed Pearson correlations from these preferential ratings and analyzed them as similarity measures with Kruskal's (1964a,b) nonmetric scaling method.

Since the data were incomplete for some respondents, only 1182 subjects could be in this reanalysis. The subjects were classified into 22 relatively homogeneous subgroups in terms of their race, party identification, geographical region, and education. Since there were not enough Negroes in this sample, they were divided into only two groups (by region). The whites were broken

down into 20 groups according to the above stratifiers. Since there were only four people in the subgroup SRSH (whites, strong Republican, south, higher education), this subgroup had to be omitted from the reanalysis. The 12 candidates and the 21 subgroups included in this study are given in Table 5.

Table 5. 1968 Election Study: The 12 Candidates and 21 Political Subgroups

Candidates:

1. G. Wallace (W)	2. H. Humphrey (H)	3. R. Nixon (N)
4. E. McCarthy (Mc)	5. R. Reagan (Rg)	6. N. Rockefeller (Rk)
7. L. Johnson (J)	8. G. Romney (Rm)	9. R. Kennedy (K)
10. E. Muskie (M)	11. S. Agnew (A)	12. C. LeMay (L)

Subgroups:

Descriptions (Race, Party, Region, and Education)		Sample size N_j
1. Negro, South	(NS)	88
2. Negro, Non-south	(NN)	77
3. White, S. Dem., South, High	(SDSH)	17
4. White, S. Dem., South, Low	(SDSL)	43
5. White, W. Dem., South, High	(WDSH)	27
6. White, W. Dem., South, Low	(WDSL)	79
7. White, S. Dem., Non-south, High	(SDNH)	21
8. White, S. Dem., Non-south, Low	(SDNL)	85
9. White, W. Dem., Non-south, High	(WDNH)	65
10. White, W. Dem., Non-south, Low	(WDNL)	180
11. White, Indept., South, High	(ISH)	8
12. White, Indept., South, Low	(ISL)	27
13. White, Indept., Non-south, High	(INH)	25
14. White, Indept., Non-south, Low	(INL)	46
15. White, S. Rep., South, Low	(SRSL)	13
16. White, S. Rep., Non-south, High	(SRNH)	40
17. White, S. Rep., Non-south, Low	(SRNL)	60
18. White, W. Rep., South, High	(WRSH)	34
19. White, W. Rep., South, Low	(WRSL)	36
20. White, W. Rep., Non-south, High	(WRNH)	90
21. White, W. Rep., Non-south, Low	(WRNL)	117

Abbreviations: S—strong, W—weak, Dem.—democrat, Rep.—republican and Indept.—independent.

Fit of the Model

Each respondent's relative scores for the 12 candidates were converted into rank orders of preferences. These rank orders were then transformed into p.c. probabilities for each subgroup which served as input data. The BTL scale values of the 12 candidates for each subgroup are presented in Table 6. The reader is remind-

Table 6. 1968 Election Study: BTL Scale Values of the 12 Candidates for Each of the 21 Subgroups (Scaled to Multiply to Unity for Each Subgroup)

	BTL scale values for the candidates											
Subgroups	*W*	*H*	*N*	*Mc*	*Rg*	*Rk*	*J*	*Rm*	*K*	*M*	*A*	*L*
NS	.11	9.00	.95	.75	.28	.66	9.10	.51	21.70	1.52	.35	.14
NN	.03	12.09	.95	1.27	.29	1.78	7.54	.58	16.02	2.15	.26	.11
SDSH	.49	3.82	1.00	1.17	.42	.94	1.48	.45	1.89	3.43	.62	.43
SDSL	.86	2.64	1.09	.58	.46	.70	2.54	.53	1.89	1.70	.74	.67
WDSH	.72	1.08	2.82	1.01	.84	1.21	1.27	.56	1.37	1.52	.69	.44
WDSL	1.24	1.20	2.30	.76	.68	.66	1.12	.64	1.45	.93	1.03	.86
SDNH	.09	4.72	1.11	1.67	.40	1.07	2.64	1.23	5.92	4.44	.38	.09
SDNL	.26	3.80	.95	.86	.43	.81	3.12	.64	5.99	2.38	.49	.25
WDNH	.12	2.99	1.46	1.68	.42	1.61	1.54	.92	5.13	2.46	.49	.19
WDNL	.37	1.99	1.57	.98	.59	.82	1.58	.68	3.69	1.71	.70	.40
ISH	.43	1.24	4.07	.76	.89	.97	.88	.60	2.88	1.10	1.34	.31
ISL	.68	.90	3.40	.87	.83	.86	.95	.51	2.53	.87	1.03	.71
INH	.43	1.77	2.54	1.49	.66	.84	1.01	.77	1.69	1.80	.88	.30
INL	.37	1.48	2.30	1.13	.74	.95	1.12	.71	2.66	1.82	.83	.31
SRSL	.11	.66	20.37	.55	1.43	.82	.86	.78	1.93	.77	3.20	.34
SRNH	.16	.55	14.29	1.05	1.98	1.44	.54	1.29	1.26	.98	1.19	.26
SRNL	.28	.62	8.49	1.02	1.39	1.02	.61	.95	1.28	.87	1.78	.41
WRSH	.76	.45	7.53	.64	1.78	.99	.82	.65	.86	.82	1.15	.80
WRSL	.85	.56	5.23	1.07	1.03	1.10	.78	.62	1.55	.55	1.23	.66
WRNH	.28	.89	4.78	1.56	1.12	1.46	.68	.75	1.38	1.34	1.01	.34
WRNL	.33	.86	5.84	1.06	1.08	1.06	.76	.73	1.75	.99	1.17	.43

The approximate chi-square tests do not reject the fit of the BTL model at the p.c. probability level for any of the 21 subgroups.

ed that these scale values are scaled to multiply to unity for each subgroup and thus are not comparable across subgroups. The fits of the BTL model to the p.c. probabilities at the first stage were satisfactory.

At the next step the BTL scale values were mapped into between set squared distances (up to column-specific constants). The L.S. solution of the generalized metric unfolding problem was then applied. The five largest Eckart-Young roots of the matrix of quasi-scalar products were 13.34, 6.82, 3.92, 2.11 and 1.73. A two-dimensional analysis was found to be unsatisfactory (the chi-square test rejected a fit with upper tail probability $\pi = .00$). The output coordinates X and Y for the 12 candidates and the 21 subgroups' ideal points, respectively, are given in Table 7. The overall

Table 7. Coordinates for Candidates and Subgroups Ideal Points in Three Dimensions

	X for stimulus points (candidates) dimensions				Y for ideal points (subgroups) dimensions		
	1	2	3		1	2	3
W	–1.17	–1.40	–.31	NS	.66	.21	–.39
H	1.41	–.31	.24	NN	.74	.51	–.42
N	–1.21	.15	–.09	SDSH	.21	–.53	.29
Mc	.07	.53	1.42	SDSL	.14	–.59	.20
Rg	–.93	.56	1.19	WDSH	–.06	–.22	.10
Rk	–.03	.63	1.39	WDSL	–.11	–.43	.12
J	1.02	–.89	–.80	SDNH	.48	.22	–.08
Rm	–.12	.71	1.47	SDNL	.37	–.16	–.02
K	1.16	–.46	–.54	WDNH	.32	.20	–.10
M	.53	.29	1.28	WDNL	.15	–.12	–.03
A	–.96	–.89	–1.05	ISH	–.05	.22	–.26
L	–.96	–1.34	–.91	ISL	–.11	–.06	–.12
				INH	.03	–.07	.03
				INL	.06	.01	–.05
				SRSL	–.25	1.14	–.92
				SRNH	–.27	.85	–.50
				SRNL	–.26	.53	–.39
				WRSH	–.34	.09	–.15
				WRSL	–.25	.03	–.14
				WRNH	–.13	.28	–.15
				WRNL	–.15	.32	–.27
latent roots of X'_oX_o				latent roots of Y'_oY_o			
17.454	9.818	.872		5.015	1.844	.097	

Compound chi-squares test for overall fit of the model:

Chi-square statistic = 1139.847
d.f. = 1155

Upper tail probability [$\Pr(\chi^2_{1155} \geq 1139.847)$] = .619

fit of the model to these data is now satisfactory as judged from the compound chi-square test ($\pi = .619$).

Interpretation of the Results

Subspace checks indicated that the ideal points lie in a two-dimensional subspace of the three-dimensional common space (the third latent root of $Y_o'Y_o$ was nearly zero, where Y_o is Y translated to its own centroid). Under such circumstances, the two-dimensional joint subspace is interpretable as long as it is kept orthogonal to the extraneous dimension for the candidates (Schönemann and Wang, 1972). Hence, X and Y were rotated so that their dimensions coincide with the principal axes of Y. The resulting coordinates are given in Table 8. The candidates (represented by

Table 8. Subspace Rotation: Coordinates at the Principal Axes Position for the Ideal Points

candidates	X_p for stimulus points dimensions			subgroup	Y_p for ideal points dimensions		
	1	2	3		1	2	3
W	.99	−1.38	−.97	NS	−.10	.63	−.14
H	.78	1.24	.14	NN	−.40	.75	−.00
N	−.20	−1.25	.05	SDSH	.80	.03	.04
Mc	.49	−.09	1.56	SDSL	.79	−.04	−.08
Rg	.18	−1.05	1.35	WDSH	.40	−.18	.03
Rk	.37	−.18	1.58	WDSL	.57	−.26	−.07
J	.66	.89	−1.06	SDNH	.02	.42	.13
Rm	.32	−.26	1.69	SDNL	.35	.26	−.02
K	.46	1.06	−.61	WDNH	.00	.26	.09
M	.69	.34	1.31	WDNL	.28	.05	−.01
A	.21	−1.04	−1.31	ISH	−.16	−.08	−.04
L	.66	−1.11	−1.44	ISL	.14	−.19	−.07
				INH	.25	−.07	.05
				INL	.15	−.02	.03
				SRSL	−1.31	−.09	−.10
				SRNH	−.85	−.19	.09
				SRNL	−.52	−.23	.01
				WRSH	−.04	−.40	−.02
				WRSL	.02	−.32	−.04
				WRNH	−.16	−.16	.09
				WRNL	−.26	−.17	.01

x) and the subgroups' ideal points (represented by .) are plotted in Figure 2 with respect to the two principal axes of Y. These two axes are then rotated and translated for identification of the two underlying dimensions.

The two dimensions could be identified as a Republican-

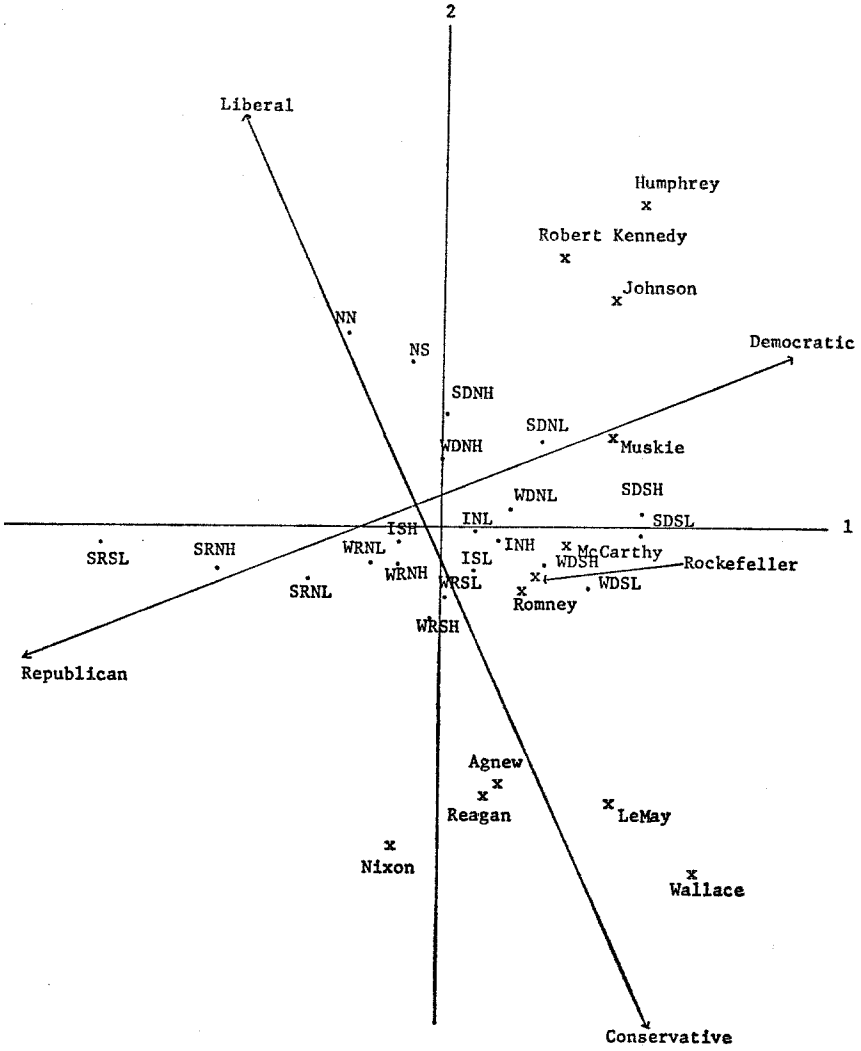

Figure 2. Election study: Candidates and subgroup's ideal points in two-dimensional joint subspace.

Democrat and a Liberal-Conservative dimension. The ordering of the candidates' projections on the Republican-Democrat dimension is Nixon, Reagan, Agnew, (Romney, LeMay), (Rockefeller, Wallace), McCarthy, Muskie, Kennedy, Johnson, Humphrey. This ordering corresponds closely to their party identifications. The ordering of the projections of the ideal points on this dimension reflects the strength of party affiliation. On the second dimension, it shows (Humphrey, Johnson, Kennedy) on one end, (LeMay, Wallace) on the other extreme, and (Muskie, McCarthy,

Rockefeller, Romney, Nixon, Reagan, Agnew) in the middle. The projections of the candidates and the subgroups on this second dimension suggest that this dimension expresses a Liberal-Conservative stand on domestic rather than foreign policies.

Weisberg and Rusk (1970) found that a third dimension in the nonmetric analysis of the candidates space had little explanatory power. Consequently, they reported also a two-dimensional space for the candidates which is very similar to our present results. The candidate clusters [(H,J,K), (Mc,Rm,Rk), (N,Rg,A), and (L,W)] showed up in both analyses. A major difference is that in the present case we are able to embed the subgroups and the candidates in a joint space. This provides a more complete picture for identifying the underlying dimensions. In addition, the stronger metric model makes more specific predictions about the probabilities.

Prediction of Choice Probabilities

P.c. probabilities can be predicted from the present model [5] for the total population. Within group p.c. probabilities $\hat{p}_{ik.j}$ can be computed from eq. [5] with $\hat{d}^2_{kj}$ and $\hat{d}^2_{ij}$ based on the fitted X and Y. similarly, any choice probabilities within a subset of candidates can be computed as long as the model is interpreted within the framework of the Luce choice axiom. For example, the predicted choice probability for assigning first choice to S_i, given the three stimuli S_i, S_k, S_l, $\hat{p}_i(i,k,l|P_j)$ for a subgroup with ideal point P_j is given by

$$p_i(i,k,l|P_j) = \hat{a}_{ij}/(\hat{a}_{ij}+\hat{a}_{kj}+\hat{a}_{lj}),\ (S_i,S_k,S_l)\ \varepsilon S,$$

where

$$\hat{a}_{ij} = exp\ (-\hat{d}^2_{ij})\ .$$

It is of interest to compare such predicted p.c. and choice probabilities $\hat{p}_{ik}$, $\hat{p}_i(i,k,l)$ among the three presidential candidates in 1968 from the results of the present analysis with the corresponding estimates p^*_{ik}, $p_i^*(i,k,l)$ which can be obtained directly from the interview records. The overall predicted probabilities were computed from the formula of conditional probabilities:

$$\hat{p}_{ik} = \sum_j \hat{w}_j \hat{p}_{ik.j} \qquad ,j = 1,2,\ldots,21\ ,$$

and

$$\hat{p}_i(i,k,l) = \sum_j \hat{w}_j \hat{p}_i(i,k,l|P_j)\ ,$$

where $\hat{w}_j$ is the estimated ("marginal") proportion of each subgroup in the population which can be estimated from the sample in this study. Since the probability of the strata was proportional to the 1960 population of each stratum, it can be assumed to be a reasonably representative sample for the purpose of computing the estimates of $\hat{w}_j$.

Table 9 gives the overall probabilities as predicted from the

Table 9: Predicted Versus Estimates Pair Comparison and Choice Probabilities for the Three Presidential Contenders. (The estimates of the p.c. and choice probabilities from the actual votes recorded for the sample of this study are presented below in paretheses.)

P.c. probabilities ($\hat{p}_{ik}$ vs. p_{ik}*)

	W	*H*	*N*
W	—	.2311	.1422
		(.2092)	(.1855)
H	.7689	—	.4211
	(.7908)		(.4626)
N	.8578	.5789	—
	(.8145)	(.5374)	

Choice probabilities [$\hat{p}_i(i,k,l)$ vs. $p_i^*(i,k,l)$]

W	*H*	*N*
.0797	.3891	.5311
(.1091)	(.4122)	(.4788)

fitted S&W model as well as the corresponding estimates obtained from the actual votes recorded in the data. The predicted values for preferring Nixon over the other two candidates are slightly higher than the actual estimates from the sample. One possible explanation of this might be a "bandwagon effect" in favor of president-elect Nixon, since the scores on the "feeling thermometer" were collected after the election. Consequently, the predicted choice probabilities in the same table show a similar bias—slightly higher predicted value for Nixon, lower for Humphrey and Wallace. Of course, other explanations for these small but apparently systematic discrepancies are possible. For example, the presence of a vice-presidential running mate on the ticket might have influenced the actual voting outcome. In any case, the fact that the actual discrepancies, though apparently not random, are quite small leads us to conclude that the overall fit of the model in all its aspects is quite encouraging.

Discussion

It is contended that the results of this empirical study lend support to the expectation that the model can become a useful and practical research tool in areas where the underlying Coombs' unfolding paradigm is intuitively plausible.

In the present instance of political choice behavior this premise is satisfied. It was found that the analysis of the same data by two rather different techniques (the correlational technique followed by a nonmetric scaling analysis employed by Weisberg and Rusk versus the probabilistic treatment within the present stronger metric model) led to practically identical results for the description of the candidate space. Therefore both strategies appear to support each other where they cover the same ground. The present model provides additional information about the subgroups and, moreover, enabled us to derive probability predictions which could be verified independently. There can be little doubt that it proved to be an adequate model for describing these 1968 election data. In addition to yielding richer predictions and a more complete description of the underlying multidimensional choice space, the present model has the further advantage of being explicitly falsifiable, in part and in total, in terms of well understood statistical tests.

Summary and Conclusions

As stated in the introduction the present work was primarily concerned with the fallible case of the model. In particular, an attempt was made to develop a more robust and reasonably fast converging least squares solution for the generalized metric unfolding problem.

The results reported in the section entitled "The Fallible Case" strongly suggest that, given the storage limitations imposed by a large number of unknowns (which rule out use of second order derivatives in actual applications), the "intervening conjugate gradient method" (CJGMI) is probably the most useful minimization technique for the present purpose. This method approaches in overall convergence the Fletcher and Powell method, which requires storage for second order derivatives. There still is room for for further improvement on this technical issue. The larger data iterative process is still rather slow and thus might have to be

terminated before the gradients vanish, even for the Fletcher and Powell method. However, the present results suggest that any further progress on this technical issue is unlikely to result from experimenting with other minimization techniques. Rather, it is feared, that only a drastic change in the definition of the loss function, or perhaps, further work on a better starting configuration can further speed up convergence. Both these possibilities are likely to be difficult technically and of course are not assured of any success.

Finally, the empirical example in the preceding section lends support to the expectation that the present model has a good chance of becoming a useful research tool in areas (e.g., consumer behavior and political choice behavior) where the basic choice paradigm of the unfolding model is likely to hold, at least approximately. The results of this particular empirical study closely match those obtained by Weisberg and Rusk in their original, technically quite different, analysis of the same data. It was also seen that an analysis by the present, metric, model yields more detailed predictions, which can be checked independently. Hence, the stronger assumptions of the present model, if met, are rewarded with more detailed and informative results.

References

Beckman, F. S. The solution of linear equations by the conjugate gradient method. In A. Ralston, and H. S. Wilf (Eds.), *Mathematical methods for digital computers.* New York: Wiley, 1960, 62-72.

Cauchy, A. Methode generale pour la resolution des systemes d'equations simultanees. *Compt. rend. Acad. Sci.* Paris, *25,* 1847, 536-538. (Cited by D. J. Wilde and C. S. Beightler).

Finkel, R. W. The method of resultant descents for the minimization of an arbitrary function. Paper 71, Preprints of papers presented at *14th National Meeting of Association of Computing Machinery,* 1959.

Fletcher, R. and Powell, M. J. D. A rapidly convergent descent method for minimization. *Computer Journal,* 1963, *6,* 163-168.

Fletcher, R. and Reeves, C. M. Function minimization by conjugate gradients. *Computer Journal,* 1964, *7,* 149-153.

Green, P. E. and Carmone, F. J. Multidimensional scaling: An introduction and comparison of nonmetric unfolding techniques. *Journal of Marketing Research,* 1969, *VI,* 330-341.

Gulliksen, H. A least squares solution for paired comparisons with incomplete data. *Psychometrika,* 1956, *21,* 125-134.

Hestenes, M. R. and Stiefel, E. Methods of conjugate gradients for solving linear systems. *Journal of Research of National Bureau of Standards,* 1952, *49,* 409-436.

Jöreskog, K. G. Testing a simple structure hypothesis in factor analysis. *Psychometrika,* 1966, *31,* 165-178.

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

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, *29*, 115-129. (b)

Mosteller, F. Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, 1951, *16*, 207-218.

Schönemann, P. H. On metric multidimensional unfolding. *Psychometrika*, 1970, *35*, 349-366. (a)

Schönemann, P. H. A note on Gulliksen's (1956) least squares solution for incomplete data. *British Journal of Mathematical and Statistical Psychology*, 1970, *23*, 69-71. (b)

Schönemann, P. H. and Carroll, R. M. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 1970, *35*, 245-255.

Schönemann, P. H. and Wang, M. M. An individual difference model for the multidimensional analysis of preference data. *Psychometrika*, 1972, *37*, 275-309.

Suppes, P. and Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. I. New York: Wiley, 1963, 1-76.

Wang, M. M. The multidimensional analysis of preference data. Unpublished doctoral dissertation, Purdue University, 1973.

Weisberg, H. F. and Rusk, J. G. Dimensions of candidate evaluation. *The American Political Science Review*, 1970, *64*, 1167-1185.

Wilde, D. J. and Beightler, C. S. *Foundations of optimization*. Englewood Cliffs, N. J.: Prentice-Hall, 1967.