QUADRATIC ASSIGNMENT AS A GENERAL DATA ANALYSIS STRATEGY

By LAWRENCE HUBERT† and JAMES SCHULTZ
The University of Wisconsin, Madison

The quadratic assignment paradigm developed in operations research is discussed as a general approach to data analysis tasks characterized by the use of proximity matrices. Data analysis problems are first classified as being either static or non-static. The term 'static' implies the evaluation of a detailed substantive hypothesis that is posited without the aid of the actual data. Alternatively, the term 'non-static' suggests a search for a particular type of relational structure within the obtained proximity matrix and without the statement of a specific conjecture beforehand. Although the static class of problems is directly related to several inference procedures commonly used in classical statistics, the major emphasis in this paper is on applying a general computational heuristic to attack the non-static problem and in using the quadratic assignment orientation to discuss a variety of research tactics of importance in the behavioral sciences and, particularly, in psychology. An extensive set of numerical examples is given illustrating the application of the search procedure to hierarchical clustering, the identification of homogeneous object subsets, linear and circular seriation, and a discrete version of multidimensional scaling.

CONTENTS

† Requests for reprints should be sent to Lawrence Hubert, Department of Educational Psychology, The University of Wisconsin, 1025 West Johnson Street, Madison, Wisconsin 53706.

1. Introduction

One of the major practical difficulties faced by all behavioral scientists in analyzing data concerns the choice of formal techniques that are intended to be of aid in developing reasonable substantive interpretations. Much of the time the final selection of a statistical tool is guided either by tradition in the researcher's field or, at the other extreme, because one particular procedure happens to be in vogue. In either case, the chosen methodology may not be the most appropriate way to answer the specific questions posed by the behavioral scientist. The difficulty of choosing a statistical tool becomes even more acute when a research problem cannot be easily rephrased within an omnipotent general linear model, since there are very few alternative paradigms that are broad enough to formulate sufficiently powerful analyses. Consequently, because of the general inflexibility of statistical schemes that do not rely on rather strong parametric assumptions, novice researchers tend to limit the questions they ask to those that fit neatly within the analysis of variance context and its derivations, or, alternatively, embrace some other familiar strategy that may not be suitable for the particular application at hand.

The purpose of this paper is to present in some detail one specific orientation, referred to as the quadratic assignment paradigm, that can incorporate a variety of disparate data structures that are reflected in a proximity matrix defined on the objects from some set. Obviously, there will be many instances in which the tasks faced by an applied scientist in analyzing a proximity matrix do not fall into one of the categories that can be handled by the type of strategy presented in the later sections; nevertheless, the approach taken below appears flexible enough to give a general point of departure for many of the problems an individual faces in choosing an appropriate methodology, and, more importantly, is broad enough to provide a general organizing principle for an extensive theoretical analysis of structure within a proximity matrix.

A. *Overview*

The general approach discussed throughout the remainder of the paper can be formulated in a relatively easy manner. It is assumed that the collected data is on n objects that, for convenience, are denoted by $o_1, o_2, \ldots, o_n$. The term 'object' is intended to be extremely general and could refer to individuals, stimuli, societies, tests, and so on. Furthermore, it is assumed that the data on these n objects can be reduced to a single numerical value defined for each ordered object pair. For instance, if the objects are stimuli, the numerical value could be a confusability index obtained by averaging the responses from a group of experimental subjects. Alternatively, if the objects are people, the numerical values could be measures of profile similarity, where the individual profiles are based on a series of standardized psychological tests. To formalize this concept in more detail, a *data matrix* Q is defined to be an $n \times n$ matrix, where both the uth row and uth column refer to object o_u, and the entry in row u and column v is denoted by $q(o_u, o_v)$. For convenience, it is assumed that $q(o_u, o_u) \equiv 0$ for $1 \leqslant u \leqslant n$, and usually, $q(o_u, o_v) \geqslant 0$ for $1 \leqslant u, v \leqslant n$. In other words, $q(\cdot, \cdot)$ is a (non-negative) real-valued function on the Cartesian product $S \times S$, where S is the set $\{o_1, \ldots, o_n\}$; furthermore, if the two arguments in $q(\cdot, \cdot)$ are the same, the function value is defined to be zero.

In addition to the data matrix Q, a *structure matrix* C is specified that represents the type of hypothesis the researcher wishes to evaluate *against* his data or, alternatively, the type of structure he may wish to identify *in* his data. These two aims represent what will be referred to as the static and non-static data analysis problems, respectively. The rows and columns of C are labelled by the integers $1, 2, \ldots, n$, and if $N = \{1, 2, \ldots, n\}$, the entries in C are the values defined by a (non-negative) real-valued function C on $N \times N$. In particular, $C(r, s)$ is the entry in row r and column s of C and $C(r, r) \equiv 0$ for $1 \leqslant r \leqslant n$.

As an introductory example of how the data and structure matrices can be interpreted, assume that a researcher has a set of n statements reflecting various political philosophies, and he believes that these statements can be ordered along a hypothetical continuum from liberal to conservative. A subject is given each pair of statements once and is asked to rate their similarity from, say, 1 to 10, with 10 being the least similar. Consequently, the data from this single subject can be put into a data matrix Q, where $q(o_u, o_v) = q(o_v, o_u) =$ rating given to the pair $\{o_u, o_v\}$. The structure matrix C is defined to reflect the contention that the n items or objects can be ordered along a continuum, and where distances between the items are reflected perfectly in the ratings. To be more explicit, assume that the hypothesized ordering is now represented by the sequence $(o_1, \ldots, o_n)$. Then, if the proposed seriation is reasonable, the entries in Q should satisfy the relation

$$q(o_u, o_v) \leqslant q(o_u, o_{v'}) \quad \text{and} \quad q(o_u, o_v) \leqslant q(o_{u'}, o_v),$$

whenever $u \leqslant v$; $v \leqslant v'$; $u' \leqslant u$. Similarly, a structure matrix C is defined by a particular idealization of this necessary property of the measure $q(\cdot, \cdot)$; for instance, in the seriation case, it would be possible to define $C(r, s) = |r - s|$,

since $C(r, s) \leqslant C(r, s')$ and $C(r, s) \leqslant C(r', s)$, whenever $r \leqslant s$; $s \leqslant s'$; $r' \leqslant r$. The same approach taken for seriation is easily generalized to other situations by hypothesizing a specific underlying organization of the objects $o_1, \ldots, o_n$ that in turn implies some patterning of the entries in the data matrix Q. The same structure is then related to the matrix C by identifying an idealized pattern of entries that satisfies perfectly the necessary implication on the Q matrix.

Although at this point the justification for the matrix C is only definitional, its real importance follows from the use the matrix has in organizing approaches to data analysis. In this context it is convenient to define a one-to-one function ρ from the set of integers $N = \{1, 2, \ldots, n\}$ to N, i.e. a permutation of the first n integers. As a notation, let Ω be the set of all such functions and suppose $\rho_o \in \Omega$ is the identity permutation that maps each element back to itself. Furthermore, if

$$\Gamma \equiv \Gamma(\rho) = \sum_{u,v} q(o_u, o_v)\, C(\rho(u), \rho(v)),$$

then $\Gamma(\rho_o)$ is merely the sum of the products of the corresponding elements between Q and C. In general, $\Gamma(\rho)$ is the sum of the products of the corresponding elements between Q and C_ρ, where the uth row and uth column of C are now the row and column labeled $\rho(u)$ in the reordered matrix C_ρ.

There are two basic problems that a researcher can frame with the use of the index $\Gamma(\rho)$. The first will be called *static* (or confirmatory) in which a particular C matrix and permutation ρ_o are posited *a priori* and the researcher merely wishes to test whether or not this specific structure is represented in the data. In this case, the criterion $\Gamma(\rho_o)$ measures the degree to which the patterning of the entries in $C \equiv C_{\rho_o}$ corresponds to the patterning of the entries in Q. As a justification for the index $\Gamma(\rho_o)$, the general statistic $\Gamma(\rho)$ can be interpreted as an 'unnormalized correlation coefficient' since $\Gamma(\rho)$ attains a maximum value when the elements in Q and the corresponding elements in C_ρ are in a perfect monotone relationship. Specifically, if it is true that $q(o_u, o_v) \geqslant q(o_{u'}, o_{v'})$ if and only if $C(\rho(u), \rho(v)) \geqslant C(\rho(u'), \rho(v'))$, then the index $\Gamma(\rho)$ is as large as possible. It may so happen, however, that for specific Q and C matrices, this maximum cannot be attained for any ρ, which is reminiscent of a similar difficulty in attaining an upper bound encountered in assessing rank correlation using Kendall's τ when ties are present.

As an illustration of these ideas using the seriation example, if the particular ordering of the objects $(o_1, \ldots, o_n)$ is hypothesized *a priori*, then the index $\Gamma(\rho_o)$ measures the degree to which this same ordering is reflected in the data; obviously, the hypothesis is perfectly validated through the C matrix if the corresponding elements in Q and C are monotonically related, but in any event, the statistic Γ indexes the closeness of the relationship in the units of measurement implicitly defined by the functions $q(\cdot, \cdot)$ and $C(\cdot, \cdot)$.

A second *non-static* (or exploratory) data analysis problem follows naturally from the discussion of the static case. Suppose that the researcher has characterized the *form* of the structure that supposedly underlies the entries in Q by specifying the C matrix. However, no *a priori* hypothesis is available that could

be represented by an initial ordering of the rows and columns of Q and C, i.e. by an identity function ρ_o. Consequently, it now may be meaningful to find those elements in Ω that maximize the criterion $\Gamma(\rho)$. In the seriation context, for example, the researcher would be looking for that ordering ρ (or orderings) such that the matrix Q is close in patterning to the transformed matrix

$$C_\rho = \{C(\rho(u), \rho(v))\}.$$

Clearly, the degree of 'closeness' is again measured by $\Gamma(\rho)$ and indicates in an implicit sense the extent to which a common patterning exists between the corresponding elements of Q and C_ρ.

It should be obvious from the symmetry of the problem that finding a permutation ρ such that $\Gamma(\rho)$ is maximized can be redefined notationally as the search for a permutation ρ' such that $\Gamma'(\rho')$ is as large as possible, where

$$\Gamma'(\rho') = \sum_{r,s} q(o_{\rho'(r)}, o_{\rho'(s)})\, C(r, s).$$

With this interpretation we are looking for a reordering of the Q matrix that fits closely to the fixed C matrix. In any event, the C matrix numerically characterizes a necessary property of the Q matrix and, therefore, searching for a permutation ρ (or ρ') that maximizes $\Gamma(\rho)$ (or $\Gamma'(\rho')$) is an attempt to force a necessary condition to hold as close as possible, where the index of 'closeness' is given by the function $\Gamma(\rho)$ (or $\Gamma'(\rho')$). The search for these optimal permutations ρ (or ρ') is called the quadratic assignment (QA) problem, and is an extremely active area of operations research introduced by Koopmans & Beckman (1957).

As one final introductory comment on the index $\Gamma(\rho)$, it may be useful at times to have a normalized statistic $\Gamma_N(\rho)$ of the form

$$\Gamma_N(\rho) = \sum_{u,v} q(o_u, o_v)\, C(\rho(u), \rho(v)) \Big/ \left[\sum_{u,v} q(o_u, o_v)^2 \sum_{u,v} C(\rho(u), \rho(v))^2\right]^{\frac{1}{2}}.$$

Because of the Cauchy–Schwartz inequality and the positivity of $C(\cdot, \cdot)$ and $q(\cdot, \cdot)$ in most applications, $\Gamma_N(\rho)$ must lie between 0 and $+1$; consequently, $\Gamma_N(\rho)$ has some of the same characteristics of an ordinary measure of association. As an alternative index or normalization, $\Gamma(\rho)$ could be divided by the maximum possible value that $\Gamma(\rho)$ could be, i.e. when the ordered values of $q(\cdot, \cdot)$ are matched sequentially with the ordered values of $C(\cdot, \cdot)$ before the sum of the pairwise products is taken. Just as in the rank correlation context, other normalizations are possible as well, but for our purposes it is convenient to deal with the unnormalized statistic directly. A further discussion of normalization schemes that would be appropriate for the index Γ is given in Hubert & Levin (1976), and the reader interested in pursuing the topic in more depth should refer to this source.

Although many data analysis tasks can be phrased in either a static *or* a non-static manner, some of the more traditional statistical concerns that fall within this same context are *inherently* static, i.e. it is of no substantive interest

to seek permutations that maximize $\Gamma(\rho)$. The emphasis in this paper is primarily on the non-static problem, but, as a more general introduction, several of the classical statistical paradigms are discussed in the next chapter. In addition, there is a substantial literature, for instance, from the nonmetric multidimensional scaling field, that deals primarily with data analysis tasks best characterized as *non-specifiable*, since even the form of the C matrix must be derived from the data itself. Even though it may be possible to relax the requirement of defining the form of the C matrix in an *a priori* manner, and, consequently, to include within the context sketched above certain data analysis tasks that were originally non-specifiable, these extensions are not discussed within the present paper.

The overview given here is still rather sketchy but the concrete details developed later will serve to clarify this preliminary discussion. Section IIA is more specific and reviews the formulas necessary for the analysis of static problems. The later subsections are even more focused and either present applications of the earlier formulas to several representative data analysis tasks that are inherently static by emphasizing the classical two-sample problem in statistics, measures of bivariate association and several related topics, or, alternatively, discuss the use of these formulas on several problems that can be phrased in both a static or a non-static manner. The remainder of the paper reviews computational approaches to the non-static task and investigates the use of one specific heuristic algorithm on a number of data analysis tasks, for instance, seriation, circular placement, hierarchical clustering, partitioning, subset identification, and a discrete version of multidimensional scaling.

B. *Illustrative Data Set*

As a convenient source for the numerical examples to be given in the later part of the paper, Tables 1 and 2 are derived from a rather classical confusion matrix collected by Rothkopf (1957) that has been subjected to a number of reanalyses, e.g. see Shepard (1963) and Hubert (1974*a*). The original data are a set of proportions obtained from a group of experimental subjects that quantify the confusability of the 36 Morse code signals. In particular, an entry in the original asymmetric 36×36 matrix corresponds to the proportion of subjects who stated that the corresponding row and column stimuli were 'different' when the row stimulus was presented first and the column stimulus immediately thereafter. Since our paradigm assumes that the diagonal entries in a Q matrix are zero, the 'self-confusion' proportions reported by Rothkopf are omitted from both Tables 1 and 2; furthermore, only the code symbols for the 26 letters containing four or less entries are used in Table 1 and only the ten symbols for the digits are used in Table 2. Since all of the numerical discussion will be restricted to the symmetric case for convenience, each entry was obtained by first summing the two integers in Rothkopf's original matrix defined on symmetric positions with respect to the main diagonal and then subtracting this result from 200. Although no further analysis of these data is now done, Tables 1

Table 1.—Proximity Matrix between the Morse Code Symbols corresponding to the Alphabetic Characters

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	X	191	190	179	191	182	181	184	90	188	173	195	151	135	187	189	183	152	160	181	149	182	174	181	184	194
B		X	125	107	182	121	165	134	188	172	142	91	182	186	181	156	155	170	158	188	153	149	171	52	147	112
C			X	166	182	138	160	170	182	143	138	122	181	184	171	116	128	160	181	193	182	165	156	123	56	117
D				X	187	158	122	139	178	178	46	99	178	163	185	161	180	132	141	189	151	164	151	141	170	149
E					X	196	195	187	173	197	194	184	189	188	190	194	191	185	182	87	193	188	189	192	190	193
F						X	176	139	187	143	167	126	176	182	181	122	173	131	160	195	147	144	139	153	147	165
G							X	186	189	160	142	172	157	177	42	126	167	153	190	191	174	180	143	168	155	144
H								X	178	185	180	145	187	176	185	174	187	168	104	190	130	102	176	161	177	179
I									X	193	190	175	179	157	191	194	193	174	149	170	169	190	188	197	194	186
J										X	151	139	181	194	140	59	126	177	195	187	184	180	166	143	138	149
K											X	140	167	172	143	172	146	157	185	184	133	157	138	87	172	161
L												X	186	189	181	117	138	135	178	188	152	138	162	127	161	90
M													X	79	185	185	178	166	188	185	176	192	164	181	177	181
N														X	187	185	189	160	171	185	171	195	182	188	188	187
O															X	142	143	178	191	190	183	177	149	141	159	158
P																X	106	148	185	185	168	164	161	149	129	111
Q																	X	181	195	193	182	176	163	102	98	65
R																		X	166	194	137	160	77	176	183	173
S																			X	184	93	165	176	178	187	189
T																				X	186	189	192	195	194	191
U																					X	111	147	171	185	180
V																						X	163	126	152	179
W																							X	157	154	166
X																								X	108	138
Y																									X	135
Z																										X

and 2 and the correspondences between the letters or digits and the Morse code symbols given in Table 3 will be used later to provide concrete numerical illustrations of the various methodological tasks.

Table 2.—Proximity Matrix between the Morse Code Symbols corresponding to the Digits

	1	2	3	4	5	6	7	8	9	0
1	X	75	169	187	176	177	159	126	86	95
2		X	82	154	185	172	151	150	145	163
3			X	125	147	133	166	167	183	181
4				X	89	132	153	174	185	186
5					X	141	164	181	190	190
6						X	70	156	184	164
7							X	70	138	170
8								X	83	122
9									X	41
0										X

Table 3.—Morse Code Symbols corresponding to the Twenty-six Letters and Ten Digits

Letter	Symbol	Letter	Symbol
A	·–	S	···
B	–···	T	–
C	–·–·	U	··–
D	–··	V	···–
E	·	W	·––
F	··–·	X	–··–
G	––·	Y	–·––
H	····	Z	––··
I	··	1	·––––
J	·–––	2	··–––
K	–·–	3	···––
L	·–··	4	····–
M	––	5	·····
N	–·	6	–····
O	–––	7	––···
P	·––·	8	–––··
Q	––·–	9	––––·
R	·–·	0	–––––

II. General Formulas and Procedures for the Static Case

Within the past few years numerous data analytic techniques have been developed in the behavioral and social sciences that generally can be subsumed under the names of clustering and multidimensional scaling. Most of these procedures, however, have been solely descriptive and little work within a traditional statistical framework has been undertaken. Ideally, the literature developed in the future will include hypothesis testing schemes appropriate

for these newer data analytic alternatives and the type of structure they impose in defining the relationships within a set of objects, e.g. for structures that are commonly identified by underlying partitions, seriations, subsets, placements in m-dimensional space and alike. Nevertheless, until tailor-made statistical routines are available, some recent work by Graves & Whinston (1970*a*, *b*) and Mantel (1967) discussed in this section can be used in an attempt to test the reasonableness of certain of these assumed formal structures, as long as they are hypothesized *a priori*.

The sections below review the results given by Graves & Whinston and then connect their formulas to some of the relevant literature both in psychometrics and statistics. In all cases, it is shown how the mean and variance of a particular statistic can be derived or rederived under a suitable randomization assumption as a special case of the Graves & Whinston expressions. Although some rather common statistical procedures that are inherently static are reviewed in Section B, the discussion in Section C deals with problems that could be viewed as being static or non-static and presents statistics that measure in some way the departure of an obtained set of data from an ideal relational structure conjectured without the aid of the data matrix Q.

A. *Formulas*

Given a data matrix Q and a structure matrix C, the static problem is one of comparing Q and C and assessing whether the pattern represented by C is also present in Q. In particular, the general approach to the static problem will be as follows: if the structure defined by the matrix C is not reflected in Q, then the value of $\Gamma(\rho_o)$ should not be unusually large compared to the distribution we would expect if all labelings of the rows and corresponding columns of C were equally likely. Specifically, $\Gamma(\rho)$ is evaluated for each function $\rho \in \Omega$, and the frequency table constructed for all $n!$ (possibly nondistinct) values of Γ, generating what is typically called a permutation distribution. The statistic $\Gamma(\rho_o)$ is then compared to this distribution and if $\Gamma(\rho_o)$ is at a suitable extreme percentage point, the hypothesis of an equally likely *a priori* labeling is rejected in favor of the structure defined by C. Usually, the actual permutation distribution is too computationally laborious to obtain explicitly each time a new data matrix is obtained; however, as will be shown below, the mean and variance parameters needed for an approximate test can be obtained directly by formula.

As a rather trivial and artificial example of how a permutation distribution can be constructed, suppose the two matrices Q and C are as follows:

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 3 & 2 \\ 1 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix} \end{array}, \qquad C = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array}.$$

Then, for the $3! = 6$ possible reorderings of the C matrix, the following values

for Γ are obtained:

Permutation	Γ
123	5
132	3
213	3
231	3
312	4
321	2

Obviously, the frequency distribution for this simple example is as follows:

Γ value	Frequency
2	1
3	3
4	1
5	1
	6

It should be emphasized at the outset that the use of a randomization distribution for evaluating a relabeling ρ is restricted to these permutations that are hypothesized without the aid of the actual data defined by $q(\cdot, \cdot)$. In other words, it makes no sense to evaluate a function ρ through an equally likely labeling assumption if by construction that function already must generate a value at an extreme percentage point of the reference distribution. Clearly, such a practice is akin to using a standard t test on two independent groups already selected for their extreme differences. Although this admonition is obvious when stated explicitly, it is not always heeded in the literature, e.g. see Fillenbaum & Rapoport's (1971) use of Johnson's (1968*a*) cluster statistic.

To make the concepts already introduced notationally more precise, suppose Ω is the set of all one-to-one mappings of the set of integers $N = \{1, 2, \ldots, n\}$ into N, and let $\rho(\cdot)$ be an arbitrary element in Ω. For convenience, it is assumed that $n \geqslant 4$ and the two matrices $Q = \{q(o_u, o_v)\}$ and $C = \{C(r, s)\}$ are both of order $n \times n$, where $q(o_u, o_u) = C(r, r) \equiv 0$ for $1 \leqslant u, r \leqslant n$. Associated with each $\rho \in \Omega$, define

$$\Gamma(\rho) = \sum_{u,v} q(o_u, o_v)\, C(\rho(u), \rho(v)),$$

and assume that a uniform probability distribution is imposed over the elements in Ω. Then $E_\rho(\Gamma(\rho))$ and $\mathrm{var}_\rho(\Gamma(\rho))$ may be expressed by the formulas given in equations (1) and (2), respectively (see Graves & Whinston, 1970*a*, *b*, and Mantel, 1967, for independent derivations):

$$E_\rho(\Gamma) \equiv E_\rho(\Gamma(\rho)) = [1/(n(n-1))]\left[\sum_{u,v} q(o_u, o_v)\right]\left[\sum_{r,s} C(r, s)\right]. \tag{1}$$

$$\begin{aligned}\mathrm{var}_\rho(\Gamma) \equiv \mathrm{var}_\rho(\Gamma(\rho)) = &-[1/(n(n-1))]^2 B_1 + [1/(n(n-1))]\,[B_2 + B_3] \\ &+ [1/(n(n-1)(n-2))]\,[B_4 + 2B_5 + B_6] \\ &+ [1/(n(n-1)(n-2)(n-3))]\, B_7,\end{aligned} \tag{2}$$

where

$$B_1 = \left(\sum_{u,v} q(o_u, o_v)\right)^2 \left(\sum_{r,s} C(r,s)\right)^2,$$

$$B_2 = \left(\sum_{u,v} q(o_u, o_v)^2\right) \left(\sum_{r,s} C(r,s)^2\right),$$

$$B_3 = \left\{\sum_u \left[\sum_v q(o_u, o_v)\, q(o_v, o_u)\right]\right\} \left\{\sum_r \left[\sum_s C(r,s)\, C(s,r)\right]\right\},$$

$$B_4 = \left\{\sum_u \left[\left(\sum_v q(o_u, o_v)\right)^2 - \sum_v q(o_u, o_v)^2\right]\right\} \left\{\sum_r \left[\left(\sum_s C(r,s)\right)^2 - \sum_s C(r,s)^2\right]\right\},$$

$$B_5 = \left\{\sum_u \left[\left(\sum_v q(o_u, o_v) \sum_v q(o_v, o_u)\right) - \sum_v q(o_u, o_v)\, q(o_v, o_u)\right]\right\}$$
$$\times \left\{\sum_r \left[\left(\sum_s C(r,s) \sum_s C(s,r)\right) - \sum_s C(r,s)\, C(s,r)\right]\right\},$$

$$B_6 = \left\{\sum_v \left[\left(\sum_u q(o_u, o_v)\right)^2 - \sum_u q(o_u, o_v)^2\right]\right\} \left\{\sum_s \left[\left(\sum_r C(r,s)\right)^2 - \sum_r C(r,s)^2\right]\right\},$$

$$B_7 = \left[\left(\sum_{u,v} q(o_u, o_v)\right)^2 - \sum_v \left(\sum_u q(o_u, o_v)\right)^2 - 2 \sum_u \left(\sum_v q(o_u, o_v)\right) \left(\sum_v q(o_v, o_u)\right)\right.$$
$$\left. - \sum_u \left(\sum_v q(o_u, o_v)\right)^2 + \sum_u \left(\sum_v q(o_u, o_v)\, q(o_v, o_u)\right) + \sum_{u,v} q(o_u, o_v)^2\right]$$
$$\times \left[\left(\sum_{r,s} C(r,s)\right)^2 - \sum_s \left(\sum_r C(r,s)\right)^2 - 2 \sum_r \left(\sum_s C(r,s)\right) \left(\sum_s C(s,r)\right)\right.$$
$$\left. - \sum_r \left(\sum_s C(r,s)\right)^2 + \sum_r \left(\sum_s C(r,s)\, C(s,r)\right) + \sum_{r,s} C(r,s)^2\right].$$

Two special cases of these expressions deserve particular mention for later use. First of all, if the matrices Q and C are symmetric, then the variance formula in (2) simplifies somewhat, since in this case $B_2 = B_3$, and $B_4 = B_5 = B_6$. Specifically,

$$\text{var}_\rho(\Gamma(\rho)) = -[1/(n(n-1))]^2 D_1 + [2/(n(n-1))] D_2$$
$$+ [4/(n(n-1)(n-2))] D_3 + [1/(n(n-1)(n-2)(n-3))] D_4, \quad (3)$$

where

$$D_1 = \left(\sum_{u,v} q(o_u, o_v)\right)^2 \left(\sum_{r,s} C(r,s)\right)^2,$$

$$D_2 = \left(\sum_{u,v} q(o_u, o_v)^2\right) \left(\sum_{r,s} C(r,s)^2\right),$$

$$D_3 = \left\{\sum_u \left[\left(\sum_v q(o_u, o_v)\right)^2 - \sum_v q(o_u, o_v)^2\right]\right\}\left\{\sum_r \left[\left(\sum_s C(r,s)\right)^2 - \sum_s C(r,s)^2\right]\right\},$$

$$D_4 = \left[\left(\sum_{u,v} q(o_u, o_v)\right)^2 - 4\sum_u \left(\sum_v q(o_u, o_v)\right)^2 + 2\sum_{u,v} q(o_u, o_v)^2\right]$$

$$\times \left[\left(\sum_{r,s} C(r,s)\right)^2 - 4\sum_r \left(\sum_s C(r,s)\right)^2 + 2\sum_{r,s} C(r,s)^2\right].$$

Secondly, even though it is generally assumed as a convention that the functions $q(\cdot, \cdot)$ and $C(\cdot, \cdot)$ are non-negative, the formulas given in (1) and (2) do hold more generally as well. One particular extension is worth noting in which the matrices Q and C are skew-symmetric, i.e. $C(r,s) = -C(s,r)$ and $q(o_u, o_v) = -q(o_v, o_u)$, $1 \leqslant r, s \leqslant n$, $1 \leqslant u, v \leqslant n$, since then $B_1 = 0$, $B_7 = 0$, $B_2 = B_3$, $B_4 = B_5 = B_6$ and $E_\rho(\Gamma(\rho)) = 0$. Consequently, the variance expression is now

$$\text{var}_\rho(\Gamma(\rho)) = [2/(n(n-1))]\,[D_2 + (2/(n-2))\,D_3]. \tag{4}$$

Clearly, other special constraints on the Q or C matrices could reduce the variance formulas in a different manner, and, in fact, many of these cases will be presented in detail later in conjunction with specific data analysis problems.

Given the mean and variance parameters of the permutation distribution for $\Gamma(\rho)$, a difficulty still remains in actually testing the equally likely labeling hypothesis. The most optimistic approach is to assume that the index $\Gamma(\rho)$ is normally distributed, in which case the statistic

$$\Gamma_N(\rho) = [\Gamma(\rho_o) - E_\rho(\Gamma(\rho))]/\sqrt{[\text{var}_\rho(\Gamma(\rho))]}$$

would be compared to the well-known percentage points for a standard normal with mean zero and variance one. In many instances, the permutation distribution for $\Gamma(\rho)$ appears to be fairly well approximated by a normal even for moderate values of n, and, in fact, various empirical demonstrations of the approximation appear in several related papers, e.g. Schultz & Hubert (1976), Hubert & Schultz (1975*a*, *b*). Nevertheless, it is also possible to obtain conservative significance levels, i.e. the 'true' significance level can be no larger, by employing Cantelli's inequality. Specifically, since for most applications that are of interest, only large positive values of $\Gamma_N(\rho_o)$ lead to rejection of the hypothesis of equally likely labelings, it is always true that

$$P(\Gamma(\rho_o) - E_\rho(\Gamma(\rho)) \geqslant k\sqrt{[\text{var}_\rho(\Gamma(\rho))]}) \leqslant 1/(1+k^2).$$

Thus, if $\Gamma(\rho_o)$ were three standard deviations away from the mean, then the 'true' significance level generated by this statistic would be no larger than $1/(1+3^2) = 0{\cdot}10$ irrespective of the form of permutation distribution for $\Gamma(\rho)$. Clearly, the significance level generated by the normal approximation would always be much smaller, but it is reassuring to know that the Cantelli inequality provides a conservative way to approach the problem. If a two-tailed test is desired, the well-known Chebyschev inequality could be used:

$$P(|\Gamma(\rho_o) - E_\rho(\Gamma(\rho))| \geqslant k\sqrt{[\text{var}_\rho(\Gamma(\rho))]}) \leqslant 1/k^2.$$

B. *Inherently Static Data Analysis Problems*

As mentioned above, one of the more interesting aspects of the general formulas given in (1), (2), (3) and (4) is the close connection these expressions have with several classical paradigms considered in statistics, for instance, assessing bivariate association and the two-independent or two-dependent sample problems. In all cases, a suitable redefinition of the Q and C matrices places these paradigms within the standard permutational approach discussed in, say, Bradley (1968). Several of these connections will be pointed out in this section and the appropriate specialization of the mean and variance formulas indicated. Since many different statistical procedures have an associated permutational theory, examples other than those given below could be developed in a similar manner. Many of the more complicated illustrations, however, are best handled through a case specific approach, and, consequently, the presentation will be confined to several of the simpler possibilities and used only to illustrate the unifying principle that the QA point of view can provide even for data analysis tasks that are inherently static.

1. *Bivariate Association*

One of the most convenient ways to approach the problem of indexing the association between two sets of observations is through Daniels' concept of a generalized correlation coefficient as discussed in Kendall (1970). We assume that n pairs of observations have been collected on a pair of random variables X and Y. For convenience, these pairs will be denoted by $(x_1, y_1), \ldots, (x_n, y_n)$, and for each pair of observations on X, say x_u and x_v, a variable a_{uv} is defined subject to the constraints that $a_{uv} = -a_{vu}$, and $a_{uu} = 0$ for $1 \leqslant u, v \leqslant n$. Similarly, for each pair of observations on Y, say y_u and y_v, a variable b_{uv} is defined subject to the same sort of restrictions. Finally, a generalized correlation coefficient using these derived scores is calculated as follows:

$$\nabla = \sum_{u,v} a_{uv} b_{uv} \Big/ \left(\sum_{u,v} a_{uv}^2 \sum_{u,v} b_{uv}^2 \right)^{\frac{1}{2}}.$$

By making appropriate definitions for the variables a_{uv} and b_{uv}, the standard Pearson product-moment correlation, Spearman's correlation and Kendall's τ can all be obtained as special cases. Specifically, if $a_{uv} = (x_u - x_v)$ and $b_{uv} = (y_u - y_v)$, Pearson's index is obtained; if

$$a_{uv} = \operatorname{sign}(x_u - x_v), \quad b_{uv} = \operatorname{sign}(y_u - y_v),$$

Kendall's τ is defined, and finally, when

$$a_{uv} = \operatorname{rank} x_u - \operatorname{rank} x_v, \quad b_{uv} = \operatorname{rank} y_u - \operatorname{rank} y_v,$$

where the ranks are within the xs or within the ys separately, Spearman's index is obtained.

The importance of the generalized correlation coefficient for us should be obvious. If we view the a_{uv}s as defining the data matrix Q and the b_{uv}s as defining

a structure matrix C, the formula given in (4) for skew-symmetric matrices immediately gives the permutational variance for ∇ and, consequently, for each of the three special measures of association. More importantly, our normalized measure Γ_N is a generalization of ∇ and since the denominators of Γ_N and of ∇ are both constant over all permutations, the parameters of the permutation distribution for the numerator can be achieved directly. Alternative forms for the variance expression can be obtained by simple algebra from (4) and are given in standard sources such as Kendall (1970). It should be noted that these specializations include the rather cumbersome variance terms that result when some of the original observations are tied.

2. *Two-independent samples*

Suppose we have two groups of independent observations where the observations in group 1 are denoted by $x_1, \ldots, x_{n_1}$, and those in group 2 by $x_{n_1+1}, \ldots, x_n$. If $Q = \{q(o_u, o_v)\} = \{x_u\}$ for $u \neq v$, and

$$C(r,s) = \begin{cases} 1/(n_1-1) & \text{if } 1 \leqslant r \neq s \leqslant n_1, \\ 0 & \text{else,} \end{cases}$$

then $\Gamma(\rho_o)$ is just the sum of the observations in the first group ($n_1 \geqslant n_2$).

The standard randomization or permutation distribution for the sum of the observations in group 1 is usually obtained by calculating this sum for all $\binom{n}{n_1}$ equally likely ways of picking n_1 observations from the pool of n and then tabling these results. However, since for each subset of n_1 observations there are $n_1!(n-n_1)!$ permutations of the total set of n observations that would lead to the same observations in group 1, the actual permutation distribution is the 'same' except that there are now $n_1!(n-n_1)!$ repeats for each subset of n_1 objects in the complete permutation distribution. In any event, the mean and variance of the obtained distributions do not change, and specializing formulas (1) and (2) gives

$$E_\rho(\Gamma) = (n_1/n) \sum_u x_u$$

and

$$\operatorname{var}_\rho(\Gamma) = (n_1(n-n_1)/(n(n-1))) \left(\sum_u x_u^2 - \left[\left(\sum_u x_u \right)^2 \Big/ n \right] \right).$$

If the actual observations $x_1, \ldots, x_n$ are replaced by given numbers $a_1, \ldots, a_n$, the general *linear rank statistic* is obtained and, obviously, the same mean and variance formulas given above apply with a_u replacing x_u. In particular, by choosing the numbers $a_1, \ldots, a_n$ appropriately, a wide variety of two-sample location and two-sample scale tests can be obtained. Several ways of choosing these constants are mentioned below along with the standard names attached to the specific tests. For other alternatives that fit within this same framework

and a further discussion of linear rank statistics, the reader is referred to Gibbons (1971):

Suppose a_u is the rank, say k, of x_u within the set $\{x_1, \ldots, x_n\}$.

Location

(a) Wilcoxon: $a_u = k$. The more common Mann–Whitney U test is an equivalent procedure, since the U statistic plus $n_1(n_1+1)/2$ is equal to the Wilcoxon statistic.

(b) Terry–Hoeffding (expected normal scores two-sample test): a_u is the expected value of the kth-order statistic from a standard normal distribution.

(c) Van der Waerden (inverse normal scores two-sample test): a_u is $\phi^{-1}[k/(n+1)]$, where $\phi^{-1}[k/(n+1)]$ is the ordinate value on the standard normal distribution that leaves $k/(n+1)$ probability in the lower tail.

(d) Bell–Doksum (random normal scores two-sample test): a_u is the value of the kth-order statistic when a sample of n observations is chosen from a standard normal distribution. Notice, the values of a_u change for each application.

Other examples include the percentile modified rank tests and the two-sample median tests discussed in Gibbons (1971).

Scale

(a) Mood: $a_u = \{k-[(n+1)/2]\}^2$.

(b) Freund–Ansari–Bradley–David–Barton: $a_u = |k-[(n+1)/2]|$.

Other definitions for a_u define the Siegel–Tukey test, Klotz (inverse) normal scores test, and scale analogues of the Terry–Hoeffding, Bell–Doksum and percentile modified rank tests for location. These alternatives and several others are discussed more fully in Gibbons (1971).

3. *Two-dependent samples*

The data for a two-dependent sample problem is similar in form to that collected in measuring bivariate association. In particular, suppose

$$(x_1, y_1), \ldots, (x_n, y_n)$$

are n pairs of observations and the absolute values of the differences, $d_u = |x_u - y_u|$, are obtained. One of the statistics that may be used to assess the differences between the two samples equals the sum, say T, of those d_us such that $x_u > y_u$. The randomization distribution for T is obtained by assuming that all 2^n assignments of signs to the n values of d_u are equally likely *a priori*; consequently, the 2^n possibly non-distinct values of T can be tabled to provide an exact null distribution.

The easiest way to obtain the mean and variance of T under the 2^n equally likely assignments of sign is to proceed in stages. First of all, the mean and variance of T can be obtained by assuming n_1 signs are positive and using the results for the two-independent sample case. More precisely, letting $d_u = x_u$ in

formulas (1) and (2), where the two groups are of size n_1 and $n-n_1$, we obtain

$$E(T|n_1 \text{ positives}) = (n_1/n)\sum_{u=1}^{n} d_u,$$

$$\text{var}(T|n_1 \text{ positives}) = (n_1(n-n_1))/(n(n-1))\left[\sum_{u=1}^{n} d_u^2 - \left(\left(\sum_{u=1}^{n} d_u\right)^2/n\right)\right].$$

Using the standard conditional formulas for two random variables X and Y, $E(E(Y|X)) = E(Y)$ and $\text{var}(Y) = E(V(Y|X)) + \text{var}(E(Y|X))$, we get

$$E(T) = \sum_{u=1}^{n} d_u/2,$$

$$\text{var}(T) = \sum_{u=1}^{n} d_u^2/4.$$

Instead of using the original absolute differences d_u, other functions of the d_us could be used as well. For example, if the rank r_u of d_u replaces d_u, Wilcoxon's test for two-dependent samples is obtained. Alternatively, if we replace d_u by the expected value of the r_uth smallest observation in a sample of size n from a standard absolute normal variate, that is, a distribution equal to the upper half of the standard normal in which all probabilities are doubled, a normal scores test presented by Klotz is obtained (see Gibbons, 1971).

4. *k-Independent Samples*

The extension of the two-independent sample paradigm to k samples requires an alternative formula used by Gini for the sum of squares for a set of numbers, say, $y_1, \ldots, y_n$:

$$\sum_{u=1}^{n}(y_u - \bar{y}_.)^2 = (1/2n)\sum_{u,v}(y_u - y_v)^2.$$

Now, suppose that we have k independent groups with $n_1, \ldots, n_k$ observations in each group. If we denote the observations in group j by $x_{j1}, x_{j2}, \ldots, x_{jn_j}$, then the standard within sum of squares term can be defined by

$$W = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ji} - \bar{x}_{j.})^2.$$

For convenience, we are interested in the distribution of W over all

$$(n_1 + \ldots + n_k)!/(n_1! \ldots n_k!)$$

equally likely ways of choosing the observations for the k groups from the total sample of size $n_1 + \ldots + n_k = n$. Alternatively, since the total sum of squares is constant for all possible ways of allocating the observations to the k groups, the between sum of squares could be used as the quantity for which the permutation distribution is defined.

Within our paradigm, the mean and variance parameters can be obtained for W by defining specific Q and C matrices. Suppose $x_1, \ldots, x_n$ denote the n observations in all k groups and define

$$Q = \{q(o_u, o_v)\} = \{(x_u - x_v)^2\},$$

$$C = \{C(r,s)\} = \begin{cases} 1/2n_1 & \text{if } 1 \leqslant r \neq s \leqslant n_1, \\ 1/2n_t & \text{if } 2 \leqslant t \leqslant k, \text{ and } n_1 + \ldots + n_{t-1} < r \neq s \leqslant n_1 + \ldots + n_t, \\ 0 & \text{else.} \end{cases}$$

In this case, the Q and C matrices are symmetric and formulas (1) and (3) define the mean and variance parameters for the permutation distribution for the W criterion.

5. *Barton & David's Intersection of Two Graphs*

A somewhat more nonstandard statistical problem that can be developed as a special case of the Graves & Whinston orientation requires the concept of an undirected graph. In particular, a graph $G(n, q)$ is defined by a set of n nodes and a set of q edges joining certain pairs of nodes. Initially, G will be viewed as codifying a symmetric relation on the nodes and no orientation or direction will be given to the edges. Consequently, two undirected graphs on the same set of nodes $G_1(n, q_1)$ and $G_2(n, q_2)$ characterize two possibly different symmetric relations on the same set of objects $\{o_1, \ldots, o_n\}$ and our problem is to develop some way of measuring the correspondence between G_1 and G_2.

Barton & David (1966) in developing a statistical approach to the study of contagious diseases define a measure of correspondence between G_1 and G_2 by the number of object pairs joined by edges in *both* G_1 and G_2. Under the hypothesis that the node labels in G_2 are fixed but the $n!$ possible labelings of the nodes in G_1 are equally likely, Barton & David obtain the mean and variance of this statistic. By suitable definitions of Q and C in the Graves & Whinston formulas, the same first two moments can be found. Specifically, let

$$C(r,s) = \begin{cases} 1 & \text{if node } r \text{ is joined to node } s \text{ in } G_2 \text{ and } 1 \leqslant r \neq s \leqslant n, \\ 0 & \text{else,} \end{cases}$$

$$q(u,v) \equiv q(o_u, o_v) = \begin{cases} 1 & \text{if node } u \text{ is joined to node } v \text{ in } G_1 \text{ and } 1 \leqslant u \neq v \leqslant n, \\ 0 & \text{else.} \end{cases}$$

The general statistic $\Gamma(\rho)$ in this case is *twice* the number of distinct unordered object pairs that are joined in G_1 and G_2, where the nodes in G_2 have been relabeled by ρ. Using the following notation, the mean and variance over all $n!$ equally likely labelings can be given by simplifying the expressions given in equations (1) and (3).

Let

$$e_1 = \sum_{u,v} q(u,v) = 2q_1 \quad (q_1 \text{ is the number of edges in } G_1),$$

$$e_2 = \sum_{r,s} C(r,s) = 2q_2 \quad (q_2 \text{ is the number of edges in } G_2),$$

$$f_1 = \sum_u \left(\sum_v q(u,v)\right)^2 = \sum_u d_{1u}^2 \quad (d_{1u} \text{ is the number of nodes joined to } o_u \text{ in } G_1, \text{ i.e. the degree of } o_u \text{ in } G_1),$$

$$f_2 = \sum_r \left(\sum_s C(r,s)\right)^2 = \sum_r d_{2r}^2 \quad (d_{2r} \text{ is the degree of } o_r \text{ in } G_2).$$

Then,

$$E_\rho(\Gamma) = (4/(n(n-1)))\, q_1 q_2,$$

$$\begin{aligned}\mathrm{var}_\rho(\Gamma) = {} & (8q_1q_2/(n(n-1)))\{1-[2q_1q_2/(n(n-1))]\} \\ & + (4/(n(n-1)(n-2)))\left(\sum_u d_{1r}^2 - 2q_1\right)\left(\sum_r d_{2r}^2 - 2q_2\right) \\ & + (16/(n(n-1)(n-2)(n-3))) \\ & \times \left[q_1(q_1-1) - \sum_u d_{1u}^2\right]\left[q_2(q_2-1) - \sum_r d_{2r}^2\right].\end{aligned}$$

As mentioned previously, these two formulas are equivalent to those obtained by Barton & David.

For the case of directed graphs, similar formulas may be obtained. In this instance, $C(r,s) = 1$ if and only if a directed edge exists from node r to node s, and $q(u,v) = 1$ if and only if a directed edge exists from node u to node v. The statistic Γ is then the number of common directed edges in G_1 and G_2, where the nodes in G_2 have been relabeled by the function ρ. Thus, e_1 and e_2 are the number of directed edges in G_1 and G_2, respectively; and in addition, the terms d_{1u} and d_{2r} defining f_1 and f_2 are now the 'out' degrees of the nodes u and r in G_1 and G_2, respectively. Several further terms are also needed.

Let

$$g_1 = \sum_u \left(\sum_v q(u,v)\, q(v,u)\right),$$

$$g_2 = \sum_r \left(\sum_s C(r,s)\, C(s,r)\right).$$

(g_1 and g_2 are twice the number of unordered object pairs in which edges exist in both directions for G_1 and G_2, respectively.)

$$h_1 = \sum_u \left(\sum_v q(u,v)\right)\left(\sum_v q(v,u)\right),$$

$$h_2 = \sum_r \left(\sum_s C(r,s)\right)\left(\sum_s C(s,r)\right).$$

(h_1 and h_2 are the sums over all nodes in G_1 and G_2, respectively, of the products of 'in' and 'out' degrees.)

$$c_1 = \sum_v \left(\sum_u q(u,v) \right)^2,$$

$$c_2 = \sum_s \left(\sum_r C(r,s) \right)^2.$$

(c_1 and c_2 are analogues of f_1 and f_2, but the terms that are squared are now 'in' degrees for the nodes v and s, respectively.)

Finally,

$$E_\rho(\Gamma) = [1/(n(n-1))]\, e_1 e_2,$$

and

$$\begin{aligned} \mathrm{var}_\rho(\Gamma) = &-\{[1/(n(n-1))]\, e_1 e_2\}^2 + [1/(n(n-1))]\,(e_1 e_2 + g_1 g_2) \\ &+ [1/(n(n-1)(n-2))]\,((f_1 - e_1)(f_2 - e_2) + 2(h_1 - g_1)(h_2 - g_2) \\ &+ (c_1 - e_1)(c_2 - e_2)) + [1/(n(n-1)(n-2)(n-3))]\,(e_1(e_1 - 1) \\ &+ g_1 - (c_1 + 2h_1 + f_1))\,(e_2(e_2 - 1) + g_2 - (c_2 + 2h_2 + f_2)). \end{aligned}$$

In addition to Barton & David's measure of the correspondence between two graphs, other statistics have been proposed that can be handled in a somewhat similar manner, for instance, the symmetric set difference index used by Zahn (1964), Restle (1959) and others. Since nothing conceptually new is added, however, the details of these generalizations are left to the reader. In a very loose sense, however, the problem of finding means and variances under the common randomization distribution has been 'solved' for many of the standard statistics that measure the correspondence between two graphs G_1 and G_2. A large number of such statistics are based upon some function of four quantities: the number of node pairs related in G_1 but not in G_2, the number of pairs related in G_2 but not in G_1, the pairs that are related in both graphs and, finally, the pairs that are unrelated in both. Given a fixed number of edges in G_1 and G_2, any one of these quantities determines the remainder. This statement extends as well to graphs that represent partitions or equivalence relations and the measures that have been proposed in this special case, e.g. Rand (1971). This includes the sorting experimental paradigms discussed by Arabie & Boorman (1973) and others in which the correspondence is measured between an obtained 'sort' or partition generated by an experimental subject and some theoretical partition hypothesized by the researcher. In fact, within the sorting context, Johnson (1968*b*) discusses essentially the same measure used by Barton & David but obtains the mean given above from a different conceptual point of view. In a similar manner a direct equivalence can be identified for the measures used in assessing clustering in free recall (see Hubert & Levin, 1976).

6. *Run Statistics*

The standard statistical concepts regarding runs of like and unlike objects can be developed within the QA framework with immediate extensions to runs of more than two types of objects (see Bradley, 1968). For instance, suppose we have r_1 objects of type A and $r_2 = n - r_1$ objects of type B, and the n objects are placed along a one-dimensional continuum at random. Using the following definitions of Q and C the criterion Γ is twice the number of runs minus 1:

$$q(u, v) \equiv q(o_u, o_v) = \begin{cases} 1 & \text{if } 1 \leqslant u \leqslant r_1 \text{ and } r_1 + 1 \leqslant v \leqslant n \\ & \text{or } r_1 + 1 \leqslant u \leqslant n \text{ and } 1 \leqslant v \leqslant r_1, \\ 0 & \text{else.} \end{cases}$$

$$C(r, s) = \begin{cases} 1 & \text{if } r = s + 1 \text{ or } r = s - 1, \\ 0 & \text{else.} \end{cases}$$

Consequently, the well-known mean and variance of the number of runs under random allocation are special cases of the expressions in (1) and (2), respectively. In a similar way the number of runs around a circle can be studied by augmenting $C(r, s)$ to be 1 when $r = 1$ and $s = n$, and when $s = 1$ and $r = n$. The criterion Γ is then twice the number of circular runs (see David & Barton, 1962).

Several other statistical problems that are inherently static can be handled in a similar way and lead to fairly concise mean and variance formulas. For instance, one general application that is immediate deals with the general problem of testing the similarity of structure underlying two proximity matrices. In other words, suppose both $q(o_u, o_v)$ and $C(r, s)$ are different proximity measures obtained in some way on the same set of objects. Specifically, the objects may be tests and the 'correlations' obtained on the same sample of subjects at two time points or from two-independent samples; or possibly $q(o_u, o_v)$ may be a standard proximity matrix, but $C(r, s)$ obtained, for example, by reconstructing a correlation matrix from a hypothesized target pattern within a factor analytic framework a hypothesized ultrametric defined by some hierarchical clustering of another proximity matrix (Johnson, 1967) the reconstructed Euclidean distances from a multidimensional scaling an additive tree metric (Dobson, 1974), and so on.

In summary, the main point of this section is to show the relative generality of the QA approach even for a class of problems that forms the basis for much of a beginning applied statistics course in the experimental behavioral sciences. Once the appropriate definitions of Q and C are obtained, there is no need to revert continually to first principles in generating the permutation mean and variance parameters. All that is required is a direct algebraic simplification of the formulas given in (1) and (2).

C. *Approximate Statistical Tests for* a priori *Hypothesis of Structure*

By changing the definition of the elements in the matrices Q and C, a number of other special cases of the Graves & Whinston formulas can be developed for

problems that could be phrased in either a static or a non-static manner. In most instances, some specific function ρ_o and a particular form of the C matrix will be the characterization of a posited structure that supposedly underlies the data contained in the matrix Q. The elements in Q are typically proximity measures between object pairs from a set of elements $S = \{o_1, o_2, \ldots, o_n\}$, and thus, the more extreme the value of $\Gamma(\rho_o)$, the more the *a priori* structure is mirrored in the data. Since the mean and variance of a randomization distribution for $\Gamma(\rho)$ are available, the value of $\Gamma(\rho_o)$ may be compared to these parameters for possible rejection of the hypothesis that ρ_o does not capture a substantial part of the structure that underlies the proximity entries in Q.

In the sections to follow, $q(\cdot, \cdot)$ is to be interpreted, unless otherwise noted, as a general proximity measure defined on $S \times S$ in which large values assigned by $q(\cdot, \cdot)$ correspond to dissimilar object pairs. Also, any reference to the index $\Gamma(\rho)$ will be in relation to a specific function defined by some $C(\cdot, \cdot)$, and the context will make the particular usage clear.

1. *Johnson's Cluster Statistic*

In an unpublished paper, Johnson (1968*a*) developed a statistic for assessing the homogeneity of a single subset of objects that recently has found some acceptance in the substantive literature, for example, in Fillenbaum & Rapoport (1971). Johnson's argument is as follows: using the convention stated above, suppose S is a set of n objects $\{o_1, o_2, \ldots, o_n\}$ and Q is a (symmetric) matrix of proximity values between pairs of elements in S, where small proximities denote similar objects. To each subset or cluster S_1 of S containing k elements, where $k \geqslant 2$, a statistic λ is defined that measures the homogeneity of the subset by taking the average of the $k(k-1)/2$ within cluster proximities minus the average of the $k(n-k)$ proximities between the objects in S_1 and $S - S_1$. Under the assumption that the subset S_1 is drawn at random from all possible k element subsets of S, Johnson obtains the mean and variance of λ. Consequently, when compared to these parameters, large (negative) values of λ imply rejection of the hypothesis that the specific subset S_1 is formed randomly.

The formulas derived by Johnson can also be obtained as special cases of those given in (1) and (3). Here, Q is symmetric, and

$$C(r,s) = \begin{cases} 0 & r = s, \text{ or } k+1 \leqslant r, s \leqslant n, \\ 1/(k(k-1)) & 1 \leqslant r \neq s \leqslant k, \\ -1/(2k(n-k)) & \text{else.} \end{cases}$$

With these preconditions, $\Gamma(\rho)$ is equal to Johnson's λ for the subset S_1 defined by those objects $o_{i_1}, \ldots, o_{i_k}$ such that $\rho(i_1) = 1, \ldots, \rho(i_k) = k$. Moreover, the mean and variance of $\Gamma(\rho)$ under all $n!$ equally likely functions in Ω are the same expressions derived by Johnson under the hypothesis that all $\binom{n}{k}$ subsets are

equally likely *a priori*. In particular,

$$\sum_{r,s} C(r,s) = 0,$$

$$\sum_{r,s} C(r,s)^2 = [1/(2k)]\{[2/(k-1)]+[1/(n-k)]\},$$

$$\sum_{r}\left[\left(\sum_{s} C(r,s)\right)^2 - \sum_{s} C(r,s)^2\right] = [1/(4k)]\{[(n-2)/(n-k)]-[4/(k-1)]\},$$

$$\left(\sum_{r,s} C(r,s)\right)^2 - 4\sum_{r}\left(\sum_{s} C(r,s)\right)^2 + 2\sum_{r,s} C(r,s)^2 = [-1/k] \times\{[(k-3)/(k-1)]+[(k-1)/(n-k)]\}.$$

Thus, if for later notational purposes we let

$$A_1 = \left(\sum_{u,v} q(u,v)\right)^2,$$

$$A_2 = \sum_{u}\left(\sum_{v} q(u,v)\right)^2,$$

$$A_3 = \sum_{u,v} q(u,v)^2,$$

where $q(u,v) \equiv q(o_u, o_v)$, then

$$E_\rho(\Gamma) = 0,$$

$$\begin{aligned}\mathrm{var}_\rho(\Gamma) = {} & [1/(kn(n-1))]\,[(2/(k-1))+(1/(n-k))]\,A_3+[1/(kn(n-1)(n-2))] \\ & \times[((n-2)/(n-k))-(4/(k-1))]\,(A_2-A_3) \\ & -[1/(kn(n-1)(n-2)(n-3))] \\ & \times[((k-3)/(k-1))+((k-1)/(n-k))]\,(A_1-4A_2+2A_3).\end{aligned}$$

As mentioned previously, these last two formulas are equivalent to those derived by Johnson.

For an asymmetric proximity matrix Q and Γ defined by the average of the $k(k-1)$ proximity values within S_1 minus the average of the $2k(n-k)$ proximity values defined between the elements in S and $S-S_1$, the expectation is still 0 but the variance changes slightly to take the asymmetry of Q into account, i.e. let

$$A_4 = \sum_{u}\left(\sum_{v} q(u,v)\,q(v,u)\right),$$

$$A_5 = \sum_{u}\left[\left(\sum_{v} q(u,v)\right)\left(\sum_{v} q(v,u)\right)\right],$$

$$A_6 = \sum_{v}\left(\sum_{u} q(u,v)\right)^2,$$

then

$$\begin{aligned}\mathrm{var}_\rho(\Gamma) = {} & [1/(2kn(n-1))]\,[(2/(k-1))+(1/(n-k))]\,(A_3+A_4)\\ & +[1/(4kn(n-1)(n-2))]\,[((n-2)/(n-k))-(4/(k-1))]\\ & \times(A_2+2A_5+A_6-2(A_3+A_4))\\ & -[1/(kn(n-1)(n-2)(n-3))]\,[((k-3)/(k-1))+((k-1)/(n-k))]\\ & \times(A_1+A_3+A_4-(A_2+2A_5+A_6)).\end{aligned}$$

Within a slightly different context, an alternative cluster statistic has been studied by Bloemena (1964) that is defined by the sum of the $k(k-1)$ within cluster proximities only. Since some fairly nice results on asymptotic normality have been obtained by Bloemena, the first two moments of the statistic will be given below. These expressions correspond to his and the asymptotic normality theorem he presents supports Johnson's conjecture that for k close to $n/2$, λ should be approximately normal.

Suppose Q is symmetric as before, but

$$C(r,s) = \begin{cases} 1 & \text{if } 1 \leqslant r \neq s \leqslant k, \\ 0 & \text{else.} \end{cases}$$

Then, $\Gamma(\rho)$ denotes twice the sum of the within cluster proximities for distinct object pairs,

$$\sum_{r,s} C(r,s) = k(k-1) = \sum_{r,s} C(r,s)^2, \qquad \sum_r \left(\sum_s C(r,s)\right)^2 = k(k-1)^2,$$

and

$$\begin{aligned}E_\rho(\Gamma) = {} & [(k(k-1))/(n(n-1))] \sum_{u,v} q(u,v),\\ \mathrm{var}_\rho(\Gamma) = {} & [(2k(k-1)(n-k))/(n(n-1)(n-2)(n-3))]\\ & \times [((3(n+k-1)-2nk)/(n(n-1)))\,A_1+2(k-2)\,A_2\\ & +(n-k-1)\,A_3].\end{aligned}$$

Obviously, the asymmetric Q case can be handled in a similar way.

Bloemena also presents two rather cumbersome expressions for the third and fourth moments of his statistic when Q is symmetric and, furthermore, gives a rather cumbersome general expression for all the moments of our criterion Γ when both Q and C are symmetric and a particular regularity condition holds. Since this expression will not be used explicitly, the reader is merely refered to Bloemena (1964) for details.

2. *Seriation Statistics*

One of the most well-studied methodological problems in psychology that was used as an illustration in the introduction involves the seriation of objects along a one-dimensional continuum. Given a set of symmetric proximity measures Q,

the goal of the seriation strategy is some ordering of the objects, i.e. some permutations ρ' of the indices $1, \ldots, n$, such that the induced sequence $(o_{\rho'(1)}, \ldots, o_{\rho'(n)})$ is adequate according to a specific criterion. For instance, as one measure of goodness-of-fit, Szczotka (1972) suggests the following index:

$$\Gamma(\rho) = \sum_{u,v} q(u,v)\,|\,\rho(u) - \rho(v)\,|.$$

Within the Graves & Whinston context, if $C(r,s) = |r-s|$, then the expectation and variance of $\Gamma(\rho)$ can be obtained as special cases of the formulas in (1) and (3), assuming the $n!$ possible orderings are equally likely *a priori*. Specifically,

$$\sum_{r,s} C(r,s) = n(n^2-1)/3,$$

$$\sum_{r}\left(\sum_{s} C(r,s)\right)^2 = n(n^2-1)(7n^2-8)/60,$$

$$\sum_{r,s} C(r,s)^2 = n^2(n^2-1)/6,$$

and

$$E_\rho(\Gamma) = [(n+1)/3]\sum_{u,v} q(u,v),$$

$$\mathrm{var}_\rho(\Gamma) = [(n+1)/45]\,[-A_1 + (n-4)\,A_2 + 4(n-1)\,A_3].$$

where the quantities A_1, A_2 and A_3 have been given previously.

A second criterion for the seriation context that may be developed in a similar manner is defined by the length of the Hamiltonian path traced by an ordering of the nodes, where a Hamiltonian path is identified by a series of contiguous edges passing through all nodes once and only once. Consequently, if the suggestion given by Wilkinson (1971) and discussed more extensively in Hubert (1974*a*, *b*) is followed, the sum of the proximities between adjacent objects in the seriation is found and a small total length implies a potentially adequate seriation.

All that is required in this case to specialize formulas (1) and (3) is the definition of $C(r,s) = 1$ if $|r-s| = 1$, and 0 otherwise. The general criterion $\Gamma(\rho)$ is then twice the length of the Hamiltonian path defined by the permutation ρ. Following the standard argument,

$$\sum_{r,s} C(r,s) = 2(n-1) = \sum_{r,s} C(r,s)^2,$$

$$\sum_{r}\left(\sum_{s} C(r,s)\right)^2 = 2(2n-3).$$

Thus,

$$E_\rho(\Gamma) = (2/n)\sum_{u,v} q(u,v),$$

$$\mathrm{var}_\rho\ (\Gamma) = (4/(n(n-1)))\,(A_1 - 2A_2) - (4/n^2)\,A_1 + (4/n)\,A_3.$$

Asymmetric generalizations of the Szczotka and Wilkinson criteria are immediate and include the possibility of defining the criterion in one 'direction' only, i.e. letting $C(r,s) = s-r$, if $s-r \geqslant 0$, and 0 otherwise; or defining $C(r,s) = 1$ if $s-r = 1$, and 0 otherwise. As an interesting alternative generalization, suppose $q(u,v)$ equals zero or one depending upon a subject preferring u over v or v over u, respectively, and let $C(r,s) = 1$ if $s-r > 0$, and 0 otherwise. The criterion Γ is the number of 'inversions' defined by the Hamiltonian path identified by ρ and is equivalent to a measure used for paired comparisons by Remage & Thompson (1966) and others in defining nearest adjoining linear orderings. A more thorough discussion of the seriation problem using the QA formulas is given in Hubert & Schultz (1975*b*); similarly, the maximum likelihood paired comparison problem is developed in more detail in Hubert & Schultz (1975*a*).

3. *Placement along a Circle*

A problem related to the seriation of objects along a continuum involves the ordering of objects around a circle. Circular placements have been discussed by Guttman (1954), Hubert (1974*c*) and others, and complement rather naturally the last section on seriation. For example, a Hamiltonian cycle can be defined by some circular ordering of the nodes $\{o_1, \ldots, o_n\}$ and the sum of the n proximities between adjacent objects is the length of the cycle. Again, small total length indicates a potentially adequate ordering. To be precise, let $C(r,s) = 1$ if $|r-s| = 1$ or $n-1$, and 0 otherwise. Consequently, the criterion $\Gamma(\rho)$ is twice the length of the Hamiltonian cycle identified by the permutation ρ. Under the assumption that all permutations are equally likely, which is equivalent in terms of the mean and variance to assuming that the $(n-1)!$ circular arrangements are equally likely, we have

$$\sum_{r,s} C(r,s) = 2n = \sum_{r,s} C(r,s)^2,$$

$$\sum_{r} \left(\sum_{s} C(r,s) \right)^2 = 4n,$$

and thus,

$$E_\rho(\Gamma) = [2/(n-1)] \sum_{u,v} q(u,v),$$

$$\text{var}_\rho(\Gamma) = [4/(n-1)] \{ [(A_1 - 2(n-1)A_2)/((n-1)(n-2))] + A_3 \}.$$

As a special case of these two formulas that really belongs to our previous discussion of inherently static problems, suppose $x_1, \ldots, x_n$ are n observations and $q(u,v) = x_u x_v$ for $u \neq v$. Then $\Gamma(\rho)$ is twice the common cyclical serial correlation statistic of lag 1 discussed by Wald & Wolfowitz (1943) and others; and thus, the expressions given above provide the mean and variance formulas. Lag k statistics require $C(r,s) = 1$ if $|r-s| = k$ or $n-k$, and 0 otherwise. In a similar way, serial correlation statistics of lag k can be developed from the Hamiltonian chain material of the previous section.

A second criterion similar to that suggested by Szczotka (1972) for the seriation problem can also be defined, and its mean and variance derived under all $(n-1)!$ equally likely circular arrangements of the n objects. In particular, if n is even, then suppose

$$C(r,s) = \begin{cases} |r-s| & \text{if } |r-s| \leqslant n/2, \\ n-|r-s| & \text{if } |r-s| > n/2, \end{cases}$$

and if n is odd, define

$$C(r,s) = \begin{cases} |r-s| & \text{if } |r-s| \leqslant (n-1)/2, \\ n-|r-s| & \text{if } |r-s| > (n-1)/2. \end{cases}$$

Then, for n even,

$$\sum_{r,s} C(r,s) = n^3/4,$$

$$\sum_{r}\left(\sum_{s} C(r,s)\right)^2 = n^5/16,$$

$$\sum_{r,s} C(r,s)^2 = n^2(n^2+2)/12,$$

and

$$E_\rho(\Gamma) = (n^2/(4(n-1)))\sum_{u,v} q(u,v),$$

$$\mathrm{var}_\rho(\Gamma) = (n(n^2-2n+4))/(12(n-1)(n-3))\,[(A_1/(2(n-1))) + ((n-2)/2)\,A_3 - A_2].$$

For n odd,

$$E_\rho(\Gamma) = ((n+1)/4)\sum_{u,v} q(u,v),$$

$$\mathrm{var}_\rho(\Gamma) = ((n+1)/(24(n-2)))\,(A_1 - 2(n-1)\,A_2 + (n-1)(n-2)\,A_3),$$

using

$$\sum_{r,s} C(r,s) = n(n^2-1)/4,$$

$$\sum_{r}\left(\sum_{s} C(r,s)\right)^2 = n(n^2-1)^2/16,$$

$$\sum_{r,s} C(r,s)^2 = n^2(n^2-1)/12.$$

Again, asymmetric generalizations are possible as well, including criteria defined in one 'direction' only.

4. *Partition Statistics*

As another very common hypothesized structure among a set of objects, consider a specific partition of the set $\{o_1, \ldots, o_n\}$ into K classes of size $n_1, \ldots, n_k$, where $\sum_{i=1}^{K} n_i = n$. The adequacy of a particular allocation of objects to the K classes defined by a permutation ρ can be measured by the sum of the K

average within-class proximity measures, and implemented by the statistic $\Gamma(\rho)$ using

$$C(r,s) = \begin{cases} 1/(n_i(n_i-1)) & \text{if } i \geqslant 2,\ n_i \neq 1, \text{ and } n_1+\ldots+n_{i-1} < r \neq s \leqslant n_1+\ldots+n_i; \\ 1/(n_1(n_1-1)) & \text{if } 1 \leqslant r \neq s \leqslant n_1 \text{ and } n_1 \neq 1, \\ 0 & \text{else.} \end{cases}$$

The mean and variance of $\Gamma(\rho)$ over all equally likely functions ρ are the same as those obtained by picking a specific partition at random from all partitions containing $n_1, \ldots, n_k$ classes. For symmetric Q, we find

$$\sum_{r,s} C(r,s) = K,$$

$$\sum_{r,s} C(r,s)^2 = \sum_{k=1}^{K} 1/(n_k(n_k-1)),$$

$$\sum_{r}\left(\sum_{s} C(r,s)\right)^2 = \sum_{k=1}^{K} 1/n_k.$$

Consequently,

$$E_\rho(\Gamma) = (K/(n(n-1)))\sum_{u,v} q(u,v),$$

$$\begin{aligned} \mathrm{var}_\rho(\Gamma) = &-(K/(n(n-1)))^2 A_1 + \left(\sum_{k=1}^{K} 1/(n_k(n_k-1))\right)(2/(n-1))A_3 \\ &+ \left(\sum_{k=1}^{K}(n_k-2)/(n_k(n_k-1))\right)(4/(n(n-1)(n-2)))(A_2-A_3) \\ &+ \left(K^2 - 2\sum_{k=1}^{K}(2n_k-3)/(n_k(n_k-1))\right)(1/(n(n-1)(n-2)(n-3))) \\ &\times (A_1-4A_2+2A_3). \end{aligned}$$

A formula appropriate for an asymmetric matrix Q can be derived easily as well. For an explicit discussion of this criterion, where the proximity measure is actually a squared distance, consult Rao (1971) and Jensen (1969).

As another application of the general partitioning problem, suppose the measure $q(u,v) = 1$ if an undirected edge exists between u and v now considered as nodes in a graph, and 0 otherwise. The general graph partition context as discussed by Donath (1968) and Donath & Hoffman (1973), among many others, can be specified in terms of the measure $C(r,s)$ as follows.

Let

$$C(r,s) = \begin{cases} 1 & \text{for } i \geqslant 2,\ n_1+\ldots+n_{i-1} < r \neq s \leqslant n_1+\ldots+n_i, \text{ or } 1 \leqslant r \neq s \leqslant n_1, \\ 0 & \text{else,} \end{cases}$$

then the criterion $\Gamma(\rho)$ is $n(n-1)$ minus twice the number of edges spanning disjoint subsets in the partition defined by ρ. In this case, the appropriate

moments under the hypothesis that a partition of the specified form is chosen at random can be obtained in the usual manner, including an obvious asymmetric generalization. Similarly, the expressions for a closely related problem in which the edges spanning two subsets are weighted and the criterion defined by summing these weights may be approached in the same way.

5. *Cuts*

A *cut* (see Ford & Fulkerson, 1962) in an undirected graph G is defined as the set of edges that connect two complementary subsets of nodes S_1 and $S-S_1$. The weight of a cut is the sum of the (positive) weights attached to the edges in the cut and measures the degree to which the subsets are 'related' to one another. If the weights are all 1, then the sum is merely the number of edges in the cut. Suppose the number of nodes in S_1 is fixed at n_1, and hence, the number of nodes in $S-S_1$ is $n-n_1=n_2$. The randomization mean and variance of the weight of a cut defined by node sets of these fixed sizes are obtained by letting $q(u,v)$ be the (positive) weight attached to the edge between u and v if $u\neq v$, and 0 if $u=v$ or if no edge exists between u and v. Also,

$$C(r,s)=\begin{cases}1 & \text{if } 1\leqslant r\leqslant n_1 \text{ and } n_1+1\leqslant s\leqslant n,\\ 0 & \text{else.}\end{cases}$$

Again, asymmetric generalizations are possible. A similar problem has been discussed by Bloemena (1964) who gives the first two moments in the symmetric case. Also, when n_1 and n_2 are not 'too extreme' in size, he justifies an asymptotic approach to normality for a criterion. For a further discussion of the concept of a cut and its use in psychological data analysis, the reader is referred to Hubert (in press).

D. *Summary*

It should be apparent that the quadratic assignment (QA) framework is capable of providing a unifying approach for many of the data analysis paradigms that in the past have been discussed in comparative isolation. The basic 'trick' of all such applications is in the definition of a 'goodness-of-fit' index that may be rephrased in terms of a data matrix Q and a structure matrix C. Once this identification is completed, the randomization mean and variance formulas for Γ may be used directly or, possibly, subjected to some algebraic simplification to generate a more convenient set of expressions.

As one simple numerical illustration of the results obtainable by using the formulas given above, let P be the set of eight letters that correspond to the Morse code signals containing three symbols, i.e.

$$P=\{\mathrm{D},\mathrm{G},\mathrm{K},\mathrm{O},\mathrm{R},\mathrm{S},\mathrm{U},\mathrm{W}\}.$$

Conjecture: the subset P is homogeneous with respect to Johnson's cluster statistic.

Using the appropriate formulas given earlier, the following values are obtained:

$$\Gamma(\rho_o) = -23{\cdot}13,$$
$$E(\Gamma) = 0,$$
$$\sqrt{\text{var}\,(\Gamma)} = 6{\cdot}88,$$
$$[\Gamma(\rho_o) - E(\Gamma)]/\sqrt{\text{var}\,(\Gamma)} = -3{\cdot}36.$$

Thus, using Cantelli's inequality, the probability of obtaining a value equal to or less than $-23{\cdot}13$ when a subset of size 8 is chosen randomly from the 26 objects is less than or equal to 0·08; consequently, the conjecture is given statistical support. Clearly, a host of other such hypotheses could be evaluated in exactly the same way.

III. Computational Procedures for the Non-static Task

Given a structure matrix C and a data matrix Q, the non-static data analysis problem consists of finding those permutations ρ that maximize the criterion Γ. As mentioned previously, this same general framework has been discussed extensively in the operations research literature under the label of quadratic assignment or QA and, consequently, the intent of this first section will be to give a fairly extensive survey of alternative computational schemes. Since QA techniques attempt to attack the optimization of a general criterion Γ, any efficiencies are lost that could be gained from the special constraints imposed by particular forms for the Q and C matrices. In some instances these gains may be substantial if the specific matrices are employed, and thus, other computational alternatives may be more appropriate. The clear advantage of the QA paradigm is in the generality of the strategies that can be developed and the ability to deal with a variety of problems with one single algorithm.

As an attempt to clarify the general non-static problem even further, it is convenient to introduce an alternative point of view that is related to a class of data analysis strategies in psychology referred to as nonmetric multidimensional scaling (NMDS). Given a set of proximity values between n objects and a set of 'distances' between n_o fixed positions in some appropriately dimensioned space, the non-static QA task can be interpreted as an attempt to allocate the n objects optimally to the n_o positions. The criterion of goodness-of-fit used to evaluate any specific allocation, Γ, is similar in form to a measure developed by Guttman and others in their approaches to nonmetric multidimensional scaling (see Shepard, 1962; Kruskal, 1964*a*, *b*; Guttman, 1968). For instance, Guttman's index may be formulated in terms of a sum, defined over all object pairs, of proximity values times fixed distances values with appropriate normalization analogous to our previous index Γ_N.

As a way of introducing an extremely simplified Guttman concept of the NMDS problem, suppose as before that S is a set of n objects $\{o_1, o_2, \ldots, o_n\}$ and $q(\cdot, \cdot)$ is the usual symmetric, positive and real-valued proximity function on

$S \times S$. The main objective of NMDS is to place these n objects in an m-dimensional space by maximizing the following function, where $m < n$ and all the summations are over u and v:

$$\mu = [\sum \hat{d}(o_u, o_v)\, d(o_u, o_v)]/[\sum d(o_u, o_v)^2 \sum d(o_u, o_v)^2]^{\frac{1}{2}}. \quad (5)$$

Here, $d(o_u, o_v)$ is some symmetric 'distance' between the objects o_u and o_v in the m-dimensional space, usually assumed to be Euclidean, and $\hat{d}(o_u, o_v) = d(o_k, o_{k'})$ for some object pair $\{o_k, o_{k'}\}$. The pair $\{o_k, o_{k'}\}$ is chosen in such a way that the rank of $d(o_k, o_{k'})$ within the set $\{d(o_r, o_s) \mid 1 \leqslant r \neq s \leqslant n\}$ is consistent with the rank of $q(o_u, o_v)$ within the set $\{q(o_r, o_s) \mid 1 \leqslant r \neq s \leqslant n\}$. Numerous problems and technical difficulties arise when ties exist in the values assigned by $q(\cdot, \cdot)$ or $d(\cdot, \cdot)$, but, for our purposes, these details can be ignored; the function $\hat{d}(\cdot, \cdot)$ will be viewed merely as some type of monotone transformation of the original proximity values.

Given this background, the non-static problem of QA can be developed in a special case by redefining the quantities used in obtaining a goodness-of-fit criterion. First of all, the elements $\hat{d}(o_u, o_v)$ will be replaced by the original proximity values $q(o_u, o_v)$, and secondly, it is assumed that the n objects are to be placed into n_o *fixed* positions specified *a priori* in the m-dimensional space, where $n_o \geqslant n$. For now, n_o will be assumed actually equal to n.

With these preconditions, the coefficient μ in (5) becomes

$$\mu' = (\text{constant}) \cdot \sum q(o_u, o_v)\, d(o_u, o_v).$$

If the n positions are labeled by $p_1, \ldots, p_n$, then maximizing μ' is equivalent to finding a permutation $\rho(\cdot)$ defined on the integer set $\{1, 2, \ldots, n\}$ such that

$$\mu'' = \sum q(o_u, o_v)\, d(p_{\rho(u)}, p_{\rho(v)})$$

is as large as possible, where $d(p_k, p_{k'})$ denotes the distance between positions p_k and $p_{k'}$. Obviously, constructing an optimal ρ is essentially the non-static problem attacked by *QA*, where $\mu'' \equiv \Gamma$ and $d(p_{\rho(u)}, p_{\rho(v)}) \equiv C(\rho(u), \rho(v))$. Conversely, finding a ρ that maximizes Γ can be interpreted as allocating the n objects in S to n fixed positions in which the 'distance' between the positions is given by $C(\cdot, \cdot)$.

A. *Review*

As a convenient categorization of approaches to the QA problem, the survey below is subdivided under three major headings: quadratic assignment methods that guarantee optimal solutions, those that provide suboptimal results and are iterative, and, finally, strategies that produce suboptimal results by construction. Under each of these major subdivisions, the contributions from a number of researchers are reviewed in a general and nontechnical manner. Empirical comparisons using several of the presented suggestions are given in Hillier & Connors (1966), Nugent *et al.* (1968) and Ritzman (1972). As one final introductory comment, most of the published discussions of QA involve finding

minima and not maxima. Since the necessary translation to the 'maxima' terminology is carried out automatically in the remainder of the paper, the reader is advised at the outset of the conflicting terms he will encounter in the primary literature cited on QA. Also, for convenience, the structure and data matrices will be assumed to be symmetric from now on.

1. *Strategies Generating Optimal Solutions*

If Ω denotes the set of all permutations of the integers in the set $\{1, 2, \ldots, n\}$, maximizing $\Gamma(\rho)$ requires finding at least one element $\rho \in \Omega$ such that no other member of Ω produces a larger value of Γ. Clearly, Ω contains $n!$ possible members and even for a moderate value of n, an exhaustive search of all elements in Ω is computationally impossible. Nevertheless, procedures have been developed that *implicitly* enumerate all alternatives to generate an optimal permutation. These techniques are generally referred to as 'branch and bound' algorithms, and have been developed by Gilmore (1962), Land (1963), Lawler (1963), Gavett & Plyter (1966) and Pierce & Crowston (1971) for the QA problem. An excellent expository introduction to branch and bound schemes in general is given by Agin (1966).

Gilmore's approach, called a single-assignment algorithm by Pierce & Crowston (1971), depends upon the existence of some cut-off value for Γ that is either known to be obtainable by some permutation or merely hypothesized to be obtainable. Partial permutations that specify the placement for only a subset of the first n integers $\{1, 2, \ldots, n\}$ are generated in some systematic fashion, and for each such partial permutation, an upper bound on Γ is calculated for all possible completions, i.e. extensions of the partial permutation to the whole set $\{1, 2, \ldots, n\}$. If this bound is not greater than the cut-off value, all completions of that partial permutation are then eliminated from the implicit enumeration. Obviously, the efficiency of this procedure depends upon the adequacy of the bound on Γ used for the completions of any specific partial permutation.

Gilmore provides two possible bounds and suggests that when $n \leqslant 15$, his procedures will be computationally feasible although still rather expensive. Explicitly, suppose $\alpha(\cdot)$ denotes a partial permutation of the integers $\{1, 2, \ldots, n\}$ and let ρ_α be some completion of α. The first bound $W_{\rho\alpha}$ depends upon the following general relation.

Let

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}$$

be two arbitrary vectors consisting of non-negative elements. If x^* and y^* are two vectors defined by reordering the elements in x and y in some way, the inner product of x^* and y^* attains the maximum if the entries in x^* and y^* are both ordered from smallest to largest. As a convenience, any such maximum inner product will be denoted by $T(x, y)$. The relationship presented above essentially

justified our use of Γ as a measure of correspondence between the numerical values assigned by the functions $q(\cdot,\cdot)$ and $C(\cdot,\cdot)$. A maximum value of Γ is achieved when a monotone relationship exists between the corresponding values assigned by the functions and only then.

Suppose D is the set of all integers for which the partial permutation α is defined, and let R be the set of all integer values taken by α on D. Four different vectors have to be defined.

Let $u \in D$, and suppose

y_u is a vector of all elements $q(u,v)$ in some order, where $v \notin D$,

w_u is a vector of all elements $C(\alpha(u),s)$ in some order, where $s \notin R$.

Furthermore, suppose

z is a vector of all elements $q(u,v)$ in some order where $u \neq v$ and $u, v \notin D$;

x is a vector of all elements $C(r,s)$ in some order, where $r \neq s$ and $r, s \notin R$.

Then,

$$W_{\rho\alpha} = \sum_{u,v \in D} q(u,v)\, C(\alpha(u),\alpha(v)) + 2\sum_{u \in D} T(y_u, w_u) + T(z,x).$$

The three terms in the expression for $W_{\rho\alpha}$ correspond, respectively, to contributions from pairs of objects that are both placed, pairs of objects in which only one object in the pair is placed and, finally, pairs in which both objects are unplaced. Obviously if D is the empty set, then the term $T(z,x)$ itself provides a bound for all permutations ρ; in this case, z is the vector of all off-diagonal elements of $\{q(u,v)\}$ in some order and x is the vector of all off-diagonal elements of $\{C(r,s)\}$ in some order.

A second bound that is theoretically better is also given by Gilmore but requires more computational effort. Before the details are given, however, the general idea may be clarified by describing the results for any complete permutation ρ, i.e. where D is the empty set. Specifically, let c_u be the uth row of the matrix $\{q(u,v)\}$ and d_r be the rth row of $\{C(r,s)\}$. Furthermore, suppose $A = \{T(c_u, d_r)\}$ and solve the ordinary assignment problem for A using one of the many available efficient algorithms (e.g. Munkes, 1957); in particular, find n elements in A having the property that no two are in the same row or column, and the sum of the n chosen entries is a maximum. This maximum sum is then an upper bound for the value of Γ obtainable with any permutation $\rho \in \Omega$.

When D is not empty and contains m members, the more general bound $W_{\rho\alpha}'$ can be given as follows:

$$W_{\rho\alpha}' = \sum_{u,v \in D} q(u,v)\, C(\alpha(u),\alpha(v)) + f_{\rho\alpha},$$

where $f_{\rho\alpha}$ is computed by solving the ordinary assignment problem for an $n-m$ by $n-m$ matrix $T_{\rho\alpha}$; explicitly, $f_{\rho\alpha}$ is the maximum sum of $n-m$ elements in $T_{\rho\alpha}$ no two of which are in the same row or column. As a notation, $T_{\rho\alpha} = \{t_{ur}\}$,

and for $u \notin D$, $r \notin R$,

$$t_{ur} = T(a_u, b_r) + 2 \sum_{v \in D} q(u,v)\, C(\alpha(v), r), \tag{6}$$

where a_u is the vector of elements $q(u, v)$ for $u \neq v$, $v \notin D$, in some order; and b_r is the vector of elements $C(r, s)$ for $r \neq s$, $s \notin R$, in some order. Clearly, the first term defining t_{ur} uses the unplaced elements and is a bound upon the remaining placement of elements once o_u is assigned to r; the second term is the additional contribution to the sum if o_u is assigned to r and depends upon the previously placed elements. Besides forming the basis of Gilmore's optimal algorithms, these two bounds $W_{\rho\alpha}$ and $W_{\rho\alpha}'$ form the crux of several suboptimal procedures discussed in a later section as well.

The branch and bound procedure suggested by Lawler (1963) is in all important respects equivalent to Gilmore's approach using the bound $W_{\rho\alpha}'$, although Lawler developed it independently. In addition, Lawler (1963), Breuer (1966*a*, *b*) and others discuss an explicit equivalence between QA and a specific integer linear programming problem. Although this correspondence is conceptually interesting, the use of general integer programming techniques appears to be much more computationally prohibitive than trying to develop specific branch and bound procedures that take advantage of the special nature of the QA formulation.

The branch and bound procedure suggested by Gavett & Plyter (1966) and Land (1963) is somewhat different in approach. Conceptually, the Gavett & Plyter approach, called a pair assignment algorithm by Pierce & Crowston (1971), is more involved than Gilmore's and does not seem to provide any computational advantage. Nevertheless, since a number of papers follow the Gavett & Plyter paradigm, a brief discussion of the strategy is included.

Wimmert (1958), Conway & Maxwell (1961), Land (1963), Gavett & Plyter (1966) and others reformulate the quadratic assignment problem as an ordinary assignment task with additional constraints on a solution. Specifically, an $N \times N$ matrix G is defined, where $N = n(n-1)/2$, and the elements in G are of the form $\{q(u, v)\, C(r, s)\}$ for $1 \leqslant u \neq v$, $r \neq s \leqslant n$. In other words, the matrix contains the costs of assigning every possible pair of objects to every possible pair of locations. Within the set of all $N!$ possible solutions of the assignment problem using G, there are $n!$ solutions of the original QA problem. Consequently, a solution is sought that is acceptable as well as best in terms of the sum of the defining $n(n-1)/2$ elements in G, i.e. one that maximizes Γ. Explicit constraints on the solution to make it acceptable can be defined in a number of ways. For instance, Lawler (1963) uses the notion of the Kronecker second power of a permutation matrix, whereas Land (1963) develops the constraints more explicitly in terms of what entries in G are precluded from an acceptable solution if some other entries are already present. For a more thorough discussion, the excellent overview of branch and bound procedures for the QA problem given in Pierce & Crowston (1971) should be consulted, which, incidentally, includes a third branch and bound procedure called a pair-exclusion algorithm.

Graves & Whinston (1970*a*, *b*) discuss an alternative strategy that can be made 'almost optimal' in the sense that it eliminates from consideration those sets of completions of partial permutations that *probably* do not contain optimal permutations. In particular, for the criterion Γ they derive the mean, $\mu_{\rho\alpha}$, and variance, $\sigma_{\rho\alpha}{}^2$, over all equally likely completions ρ_α of some partial permutation α; if $\mu_{\rho\alpha}+k\sigma_{\rho\alpha}$ is not greater than the known obtainable value of Γ, then the completions of α are eliminated from further consideration. Since the constant k is chosen by the researcher, some *crude* control over the probability of 'missing' an optimal permutation can be obtained using the standard Chebychev inequality theory.

The mean value $\mu_{\rho\alpha}$ is directly interpretable within the notation introduced in defining $W_{\rho\alpha}{}'$. Specifically,

$$\mu_{\rho\alpha} = \sum_{u,v\in D} q(u,v)\, C(\alpha(u),\alpha(v)) + (1/(n-m)) \times \sum_{u\notin D, r\notin R}\left[2\sum_{v\in D} q(u,v)\, C(\alpha(v),r)\right] + (1/((n-m)(n-m-1))) \times \left(\sum_{u,v\notin D} q(u,v)\right)\left(\sum_{r,s\notin R} C(r,s)\right).$$

The use of $\mu_{\rho\alpha}$ leads to several suboptimal constructive methods discussed in a later section, which provide known assignments that give obtainable values of Γ before the probabilistic branch and bound procedure is implemented.

2. *Strategies that are Suboptimal and Iterative*

Numerous approaches have been developed for iteratively improving upon an initial permutation of objects until no further significant change is possible. None of these iterative procedures can guarantee an optimal solution, but they can be used for problems that are too large to attack either with the branch and bound or integer programming techniques and, rather importantly for the applied researcher, are conceptually easy to understand. Several strategies that fall within the general iterative context are discussed below and, in fact, the scheme that is used for our numerical examples later falls within this same category.

One of the more elegant approaches to QA developed by Steinberg (1961) depends upon a reduction to an ordinary assignment problem similar to the Gilmore (1962) and Lawler (1963) branch and bound strategies. Within our framework, the basic proximity function $q(\cdot,\cdot)$ must be initially dichotomized at some value d:

$$q_d(u,v) = \begin{cases} q(u,v) & \text{if } q(u,v)\leqslant d, \\ 0 & \text{if } q(u,v) > d. \end{cases}$$

All the internally stable subsets $\{U_1,\ldots,U_r\}$ that have the following properties are then found by some algorithm, e.g. through a simple modification of those routines for finding maximal cliques referenced in Hubert (1974*c*):

(a) for any U_i, $1\leqslant i\leqslant r$, if $o_k, o_{k'}\in U_i$, then $q_d(k,k')=0$,

(*b*) for any U_i, the set U_i is not a subset of U_j, $1\leqslant i\neq j\leqslant r$.

The basic approach used by Steinberg is to start with the placement made at some step k and select a set U_i. The elements in $S-U_i$ are assumed to be fixed in their positions, i.e. a partial permutation α is defined with $S-U_i=D$ as the domain of α with, say, m elements. A complete placement at step $k+1$ is then obtained by reallocating the objects in U_i by solving the ordinary assignment problem using t_{ur} in equation (6), where $T(a_u, b_r)$ in the expression is now 0 due to the definition of $q_{\bar{d}}(\cdot,\cdot)$. This step $k+1$ permutation and a new maximal set U_i' are then used to reinitialize the process; termination occurs when no change in the current permutation results through any choice of a set in $\{U_1, \ldots, U_r\}$. A fairly clear presentation of the Steinberg approach is available in Rutman (1964), along with several suggested modifications. The most important new feature is an evaluation of all pairwise interchanges of objects at several points during the operation of the algorithm in addition to the simple recycling through the sets U_i, $1 \leqslant i \leqslant r$.

Armour & Buffa (1963) suggest an iterative procedure in which the pairwise exchange concept is the basic motivation for the whole method and, in effect, is the algorithm to be used in the later chapters. Given some initial placement, the effect of all pairwise interchanges of elements on the function Γ is evaluated, and that single interchange is made that gives the largest increase in Γ. This process is repeated until none of the interchanges provide an increase in Γ. For the more specific context of seriating objects along a continuum, Szczotka (1972) labels such a final placement as quasi-optimal. An alternative procedure introduced by Nugent *et al.* (1968) is based upon the Armour & Buffa strategy but chooses the single interchange to perform according to a probability distribution over all pairwise interchanges that induce increases in Γ.

An excellent general discussion of the local (or quasi-) optimal properties satisfied by search procedures involving the interchange of objects is available in Nicholson (1971); and for the specific case of the QA problem, in the article by Garside & Nicholson (1968). In this latter reference, a 'good' initial permutation is selected by construction (see the details in Garside & Nicholson, 1968) and k-level cyclic interchanges are applied to improve the solution. In particular, k-level cyclic interchange requires the selection of any k objects $o_{i_1}, \ldots, o_{i_k}$ and placing o_{i_t} into the position occupied previously by $o_{i_{t+1}}$, $1 \leqslant t \leqslant k-1$, and placing o_{i_k} into the position occupied previously by o_{i_1}. If no k-level cyclic interchange can increase the value of Γ, the obtained permutation is said to be optimal with respect to a k-level cyclic interchange. For $k=2$, Garside & Nicholson suggest finding that object whose contribution to Γ is smallest and tentatively exchanging this element with all others. That interchange is made for which the increase in Γ is the greatest, or if no increase is possible, the object with the second smallest contribution is chosen for trial interchange, and so forth. When no interchange can be made, a level 2 optimal permutation has been found. The level k optimal permutations are obtained in a similar fashion but the object selected for testing must be exchanged cyclically with the position of $k-1$ other elements. Obviously, many classes or types of interchange strategies could be used and the reader is referred to Nicholson (1971) for a number of alternative proposals.

The general notion of using pairwise interchanges has been suggested by many authors for the QA problem or for several special cases, for instance, see Glaser (1959) and Pomentale (1965, 1967). Glaser orders the objects according to the magnitude of the proximity or connections to other objects and attempts to interchange the highly related objects first to improve the permutation. For a more general context, the problems of local optimum points are discussed in Reiter & Sherman (1965) and Pegels (1966) who investigate the use of an induced probability distribution over local optima by using random starting points and some heuristic improvement algorithm.

A final set of iterative procedures for the QA problem includes 'relaxation' methods and depends upon the physical analogue of force vectors operating upon the objects placed in some particular manner. Objects are then repositioned sequentially to reduce the 'tension' on each object. The reader is referred to Hanan and Kurtzberg (1972*a*, *b*) for a more detailed discussion of the general idea behind relaxation strategies as well as for the specification of several detailed algorithms that implement these concepts.

3. *Strategies that are Suboptimal and Constructive*

In addition to a discussion of a branch and bound strategy for QA, Gilmore (1962) presents two suboptimal alternatives based upon a reduction to an ordinary assignment problem. In both instances, the final permutation is built up in stages and, consequently, these two schemes can be characterized as suboptimal and constructive. More precisely, a sequence of partial permutations $\alpha_o, \ldots, \alpha_n$ is formed, where α_k has k elements in its domain. Given α_k, the next partial permutation α_{k+1} is formed from α_k by picking a single element from the matrix $\{t_{ur}\}$ of order $n-k \times n-k$ given in equation (6). For example, if $t_{u_o r_o}$ is selected, then o_{u_o} is placed in position r_o, where o_{u_o} is some element not in the domain of α_k and r_o is some element not in the range of α_k. Thus, since α_{k+1} defines the same mapping as α_k except that o_{u_o} is now mapped to r_o, Gilmore's suboptimal alternatives can be defined by two different procedures for selecting a single element from $\{t_{ur}\}$.

(a) Find the maximum of each row and column and take the minimum of the maxima.

(b) Solve the ordinary assignment problem and select the smallest of the $n-k$ entries appearing in the solution.

Besides the two possibilities for choosing an element of $\{t_{ur}\}$ proposed by Gilmore, a third criterion has been suggested by Hillier and Connors (1966).

(c) Within each row and column, the arithmetic difference between the largest and the second largest entry is obtained; these differences indicate a minimum 'penalty' for not choosing the largest cell in a particular row or column. That specific row or column with the largest difference is found and the largest cell within the row or column is chosen for the assignment.

Graves & Whinston (1970*a*) follow a similar approach in selecting a known feasible solution prior to the application of their probabilistic implicit enumeration, although matrices other than $\{t_{ur}\}$ are used. In particular, using the formula

for the mean completion $\mu_{\rho\alpha}$ of a partial permutation α given in (7), three different approaches are suggested. Suppose $\bar{\alpha}_{k-1}$ is the mean completion of α_{k-1} and $\bar{\alpha}_k(u,r)$ is the mean completion of α_k, where α_k is defined by mapping o_u to r, i.e. if D is the domain of α_{k-1}, then $u \notin D$, and if R is the range of α_{k-1}, then $r \notin R$. If $u^* \notin D$, $r^* \notin R$ are the indices defining α_k, the Graves & Whinston approaches can be characterized as follows.

(a) Define α_k from α_{k-1} by finding u^*, r^* such that

$$\bar{\alpha}_k(u^*, r^*) = \max\{\bar{\alpha}_k(u,r) \,|\, u \notin D, r \notin R\}.$$

(b) Let u^* be the index not in D that maximizes $\sum_{r \notin R}(\bar{\alpha}_k(u,r) - \bar{\alpha}_{k-1})^2$ and let r^* be that index not in R that maximizes $\bar{\alpha}_k(u^*, r)$.

(c) Define u^* to be the index not in D that minimizes $\sum_{v \notin D} q(u,v)$, and let r^* be the index not in R that maximizes $\bar{\alpha}_k(u^*, r)$.

One constructive technique suggested by Edwards, Gillett & Hale (1970) as a way of providing an initial starting point for an iterative routine is very similar in form to the hierarchical clustering strategies discussed by Johnson (1967). In fact, their method can be viewed loosely as an attempt to cluster objects and locations simultaneously, and in the process to allocate the objects to the locations. The Edwards *et al.* procedure begins with all object pairs ranked from most similar to least similar and the fixed locations from closest to furthest apart. Objects are positioned by scanning in sequence the object pairs from most similar to least similar and using the following decision rule.

(i) If an object pair is reached that contains one object o_u that is assigned and the other object o_v unassigned, the location pairs are scanned from closest to most distant until a location in a pair is occupied by o_u and the other location is empty. Object o_v is then allocated to the empty position.

(ii) If neither object o_u nor o_v in the pair is assigned, the next most similar object pair is found that contains o_u or o_v plus some third object o_k. The same procedure is followed in the location pair list giving empty locations u', v', k'. If the two object pairs are $\{o_u, o_v\}$ and $\{o_v, o_k\}$, and the two locations are (u', v') and (v', k') then o_u, o_v and o_k are assigned to u', v' and k', respectively.

If only two objects remained to be assigned, they are placed in the two remaining locations in the order that maximizes Γ. Otherwise, the termination will be after step (i).

Several other procedures very similar in intent to the Edwards *et al.* suggestion are given in Kurtzberg (1965) and Hanan & Kurtzberg (1972*a*, *b*); like clustering procedures, however, these schemes are capable of almost infinite variation and variety with little hope of obtaining reasonable criteria for a final choice. In any event, these constructive procedures provide fairly fast strategies for defining an initial starting point for the iterative procedures and probably mitigate the problem of finding local optima inherent in any of the search-oriented methods.

Although the discussion in the previous subsections has been concerned with a particular simplified form of the QA problem, there are many variations of the basic format that have been presented by the authors already cited. Since these

digressions may be valuable for projected users of the QA paradigm in psychology, several brief comments on these variations are in order.

(i) The restriction to symmetric proximity and 'distance' functions is unnecessary. In fact, much of the basic literature on QA does not depend upon the symmetry restriction, e.g. Gilmore (1962) and Pierce & Crowston (1971); moreover, all the Graves & Whinston (1970*a*, *b*) formulas extend to the asymmetric case.

(ii) 'Cost' functions more general than Γ have been discussed explicitly, e.g. Steinberg (1961), Gilmore (1962), Hillier & Connors (1966) and Pierce & Crowston (1971), including the use of a constant penalty cost for assigning a particular object o_u to a position p_r. Consequently, in this generalization, the final solution can be forced either toward or away from a specific allocation. Other generalizations are given in Greenberg (1969).

(iii) For the city block metric, several additional routines have been developed that depend upon the specialized nature of the distance function, e.g. Hillier & Connors (1966), Breuer (1966*a*, *b*), Nugent *et al.* (1968) and Ritzman (1972).

(iv) Within the branch and bound alternatives suggested by Lawler (1963) and Gilmore (1962) it is possible to terminate the search for an optimal solution once some permutation is identified that is within a specified degree of accuracy to an upper bound.

B. *The Steepest Ascent Heuristic*

Although branch and bound and related linear integer programming techniques that generate optimal solutions to the QA problem may be appropriate for some small problems when the computational burden is not too great, it is clear that a generally applicable routine must be based upon a method with reasonable cost requirements and great flexibility. For this reason, the sections that follow rely upon an iterative improvement scheme that guarantees only locally optimal solutions, i.e. solutions that cannot be bettered by an application of a defined class of operations. In particular, once an initial permutation, say ρ_1, is chosen either through external construction or randomly, a 'steepest ascent' procedure will be used to modify ρ_1 in the following way (see Francis & White, 1974, for a historical introduction to this algorithm).

Let $\rho_i(u : v)$ denote a permutation that is identical to a permutation ρ_i except now u and v are interchanged, $u \neq v$. If $\Gamma(\rho_i) < \Gamma(\rho_i(u : v))$ for some pair (u, v), define $\rho_{i+1} = \rho_i(u^* : v^*)$, where $\Gamma(\rho_i(u^* : v^*)) = \max_{(u,v)} \Gamma(\rho_i(u : v))$, and tied maxima are broken in some arbitrary manner. If no interchange can increase Γ, the final solution has been obtained.

Many variations on this basic interchange strategy could be defined, but for purposes of providing a 'clean' computational baseline, it is reasonable to consider only the basic interchange alternative, without arbitrary or *ad hoc* modifications. However, it is the intent of the authors to compare systematically the effect of various 'improvements' in the basic computational scheme and to report on these studies at a later date.

Ending up at a local minimum for the procedure suggested above is still a difficulty, but there is an alternative point of view that provides a reassuring perspective on this problem. In optimization applications encountered in operations research, the failure of a strategy to find global optima is usually of secondary importance if a computationally feasible scheme can still improve upon a solution obtained without the aid of the heuristic. Similarly, in many behavioral science applications the emphasis should be on a procedure's ability to detect or identify an underlying structure in the presence of a *reasonable* amount of noise. Consequently, the criterion that should be used in evaluating the search strategy is in how well it identifies a reasonably prominent structure and not on how many times it finds the optimal solution for any data set whatsoever. In fact, local optima that are fairly close to a global optimum may be very important substantively in defining alternative representations of a data set; conceivably, local optima may be more important than the *single* 'best' solution.

IV. Application of the General Steepest Ascent Heuristic to Selected Non-static Problems

A. *Hierarchical Clustering*

In recent years one of the most active areas of methodological research in the behavioral sciences has dealt with the problem of hierarchical clustering (see Sneath & Sokal, 1973). Given a set of objects $S = \{o_1, o_2, \ldots, o_n\}$, the aim of any hierarchical clustering strategy is to find a 'good' sequence of partitions of S, say $l_0, \ldots, l_m$, with the following properties.

(i) l_0 is the disjoint partition in which each object defines a separate class; l_m is the conjoint partition containing all objects within one single class.

(ii) l_k is a refinement of l_{k+1}, i.e. if two objects are within a single subset in l_k, then they are also within a single subset in l_{k+1}; and furthermore, l_{k+1} contains fewer subsets than l_k.

Given n objects in S, obviously m must be less than or equal to $n-1$.

As an example, if S contains six objects, then the following four partitions form a hierarchy:

$$l_0 = \{\{o_1\}, \{o_2\}, \{o_3\}, (o_4\}, \{o_5\}, \{o_6\}\},$$
$$l_1 = \{\{o_1, o_3, o_6\}, \{o_2\}, \{o_4, o_5\}\},$$
$$l_2 = \{\{o_1, o_3, o_6\}, \{o_2, o_4, o_5\}\},$$
$$l_3 = \{\{o_1, o_2, o_3, o_4, o_5, o_6\}\}.$$

Any partition hierarchy corresponds to a class of functions on $S \times S$ containing *ultrametrics* (Johnson, 1967), where an ultrametric is defined by a non-negative, symmetric and real-valued function $d(\cdot, \cdot)$ on $S \times S$ with the following additional properties:

(i) $d(o_i, o_j) = 0$ if and only if $o_i = o_j$,

(ii) $d(o_i, o_j) \leqslant \max\{d(o_i, o_k), (o_k, o_j)\}$ for all $o_i, o_j, o_k \in S$.

Each partition hierarchy $l_0, \ldots, l_m$ defines a representative ultrametric of the form $d'(o_i, o_j) = \min\{k \mid o_i \text{ and } o_j \text{ belong to the same subset within the partition } l_k\}$, and, in particular, a class of ultrametrics by using any monotone increasing transformation of the $d'(\cdot, \cdot)$ function that maps 0 to 0 ('zero invariant'). Conversely, a partition can be defined by subsets in which objects o_i and o_j are within a single subset if and only if $d'(o_i, o_j) \leqslant h$ for some constant h, or for that matter, if and only if $d(\cdot, \cdot) \leqslant h$ for any constant h and any $d(\cdot, \cdot)$ within the appropriate equivalence class. As h is varied from 0 to ∞, a complete partition hierarchy is constructed.

Given a set of objects $S = \{o_1, \ldots, o_n\}$, a particular ultrametric $d(\cdot, \cdot)$ on $S \times S$, and a symmetric measure of proximity $q(\cdot, \cdot)$ on $S \times S$, our problem is this: find that relabeling of the objects in S that maximizes Γ, where $C(\cdot, \cdot) \equiv d(\cdot, \cdot)$. In other words, we are looking for a partition hierarchy of a particular form, where the form is specified by $d(\cdot, \cdot)$, that best 'fits' the proximity data defined by the function $q(\cdot, \cdot)$. Obviously, this formulation is still extremely broad and requires further specification; for instance, the given ultrametric $d(\cdot, \cdot)$ has been selected from a class of ultrametrics that generate the same partition hierarchy, and depending upon the choice, the best relabeling may vary. More specifically, by choosing certain ultrametrics from the class corresponding to a particular partition hierarchy, the final best relabeling can be made to reflect certain parts of the hierarchy more than others.

Clearly, the choice of an ultrametric for representing a partition hierarchy needs study and some type of sensitivity analysis familiar in linear programming should be developed. Our purpose here is only to demonstrate the feasibility of obtaining reasonable results through the 'steepest ascent' algorithm for maximizing Γ. Consequently, it is easy to see that the problem considered here reduces to one of obtaining a relabeling of the objects in S that maximizes the standard Pearson product-moment correlation between the *given* proximity and *given* ultrametric values.

To illustrate the type of analysis that can be handled within the framework sketched above, the Rothkopf data of Table 1 will be used along with two different target ultrametrics, where in each case the representative ultrametric $d'(\cdot, \cdot)$ is used.

(a) Chained hierarchy: at partition level k, there are $n-k$ subsets; one subset contains $k+1$ objects and the remaining $n-k-1$ subsets contain a single object, i.e. $m = n-1 = 25$.

(b) Single partition: when $m = 2$ the steepest ascent procedure essentially fits a single partition. In our case, the non-trivial level 1 partition will be defined according to the number of possible Morse code signals containing a specific number of components, i.e. we have four subsets at level 2 where the subsets contain 2, 4, 8 and 12 objects corresponding to 1, 2, 3 or 4 dots or dashes in the signal, respectively.

The single partition example is particularly instructive since maximizing Γ in this case is equivalent to minimizing the sum of all within subset proximity

values, i.e. an attempt is made to allocate the objects optimally to the subsets of fixed sizes. In other words, for the specific example introduced in (b) it is of interest to see if the computational heuristic can recapture the decomposition defined by the number of symbols in a Morse code signal or, alternatively, find partitions of exactly the same type that are even more homogeneous with respect to the index Γ.

It is generally assumed that by initializing a suboptimal search at a reasonably good solution, suboptimal strategies can better avoid entrapment in local optima. To study this empirical generalization with our algorithm, 50 random relabelings were selected as initial solutions for each of the two target ultrametrics. The folklore hypothesis regarding initial solutions was not justified (in fact, the converse may be more reasonable) as may be seen in the two tables given below; the categorizations were constructed by dichotomizing the initial and final Γ values at their respective medians.

Chained hierarchy:

		Final Γ	
		Above median	Below median
Initial Γ	Above median	9	16
	Below median	15	10

Single partition hierarchy:

		Final Γ	
		Above median	Below median
Initial Γ	Above median	11	14
	Below median	12	13

The few improved Γ values actually tied at the medians were assumed to be in the categories labeled 'below median'. In short, these results and those given later suggest that Shepard's (1974) pessimism regarding initial solution procedures may be justified.

In terms of z-score units, the improved Γ values did not vary widely for the first target ultrametric and ranged from 7·13 to 7·20; for the single partition example the range was somewhat broader and extended from 6·37 to 9·83. In the case of the chained ultrametric, 15 different local optima were observed that provided almost identical hierarchies; the major differences were in how the four element symbols were sequentially added. The best chained hierarchy obtained may be represented in terms of an ordering of the 26 symbols, i.e. the level k partition is defined by uniting the first $k+1$ elements in the following order:

(–·––; –·–·; ––·–; ––··; –···–; –····; ·–··; ·––·;

·–––; –·–; –··; ···–·; ––·; –––; ·––; ·–·; ···–;

···–; ····; ···; ––; –·; ·–; ··; –; ·)

A fairly clear gradient from four elements in a symbol to a single element exists in the chained clustering with some evidence that initial position is also important.

Using the second target ultrametric the best partition found may be given as follows:

$$\{\{-;\ \cdot\},\ \{--;\ -\cdot;\ \cdot-;\ \cdot\cdot\},\ \{\cdot--;\ \cdot\cdot\cdot;\ -\cdot\cdot;\ \cdot\cdot-;\ \cdot\cdot\cdot\cdot;$$
$$\cdot\cdot-\cdot;\ \cdot\cdot\cdot-;\ --\cdot\cdot;\},\ \{\cdot-\cdot;\ ---;\ --\cdot;\ -\cdot-;\ -\cdot\cdot-;$$
$$-\cdot-\cdot;\ \cdot-\cdot\cdot;\ --\cdot-;\ -\cdot--;\ -\cdot\cdot\cdot;\ \cdot--\cdot;\ \cdot---\}\}$$

Thirty different local optima were obtained using 50 random starting configurations, with only two starts out of the 50 producing the largest Γ. Almost all of these results, however, were similar to the type of partition given above, i.e. a general subdivision using the number of components with some minor intermixing of the three and four element symbols. A similar intermixing of three and four element symbols appears in the later analyses as well. It is interesting to note that the 'natural' partition into number of symbol components constructed a z-value of 9·12; more significantly, 21 out of the 50 starting configurations give better improved Γ indices than this 'natural 'specification even though the partition itself was also a local optimum and obtained twice in the sample of 50.

The use of partition hierarchies is well known in psychological research but this same general steepest ascent procedure can be used as well for some recent generalizations suggested in the clustering literature that allow overlap between the subsets at a particular level (see Jardine & Sibson, 1971; Hubert, 1974*c*). Instead of a hierarchy of partitions, a hierarchy of subsets could be defined in which the subsets at any particular level are formed by uniting certain of the subsets at a lower level. The distance function that corresponds to $d'(\cdot,\cdot)$ in the partitioning context would again assign an index to each object pair in relation to the minimum level at which the pair first appears within the same subset. In a somewhat similar sense the literature on the partitioning of graphs falls under this same rubic (see Gorinshteyn, 1969).

B. *Subset Identification*

Instead of seeking single partitions or partition hierarchies, the related problem of identifying homogeneous subsets of objects has also been discussed in the psychological literature (for instance, see Fillenbaum & Rapoport, 1971). As a convenience in our maximization problem, Johnson's (1968*a*) definition is modified slightly by the average of the proximities from the objects within the subset to the objects outside the subset minus the average of the within subset proximities. More precisely, if

$$C(r,s) = \begin{cases} 0 & r=s, \text{ or } k+1 \leqslant r,\ s \leqslant n, \\ -1/(k(k-1)) & 1 \leqslant r \neq s \leqslant k, \\ 1/(2k(n-k)) & \text{else.} \end{cases}$$

then $\Gamma(\rho)$ is Johnson's index calculated for a subset of k objects $\{o_{i_1}, \ldots, o_{i_k}\}$, where $\rho(i_1) = 1, \ldots, \rho(i_k) = k$.

As discussed previously, many other definitions of subset homogeneity are possible but the same general procedures will be followed. It is interesting to note that even the hierarchical partitioning discussion can include a specific search for a homogeneous subset, where the index is just the sum of within group proximity values and the only non-trivial partition in the hierarchy is defined by a simple subset containing k elements with the remainder of the objects in separate classes. In fact, the nesting of subsets could be approached in a manner similar to the chained example of hierarchical clustering. Because certain of the Morse code signals have the same number of components, k was set at 4, 8 and 12.

Table 4.—*Observed Local Optima for the Subset Specification*

k	Subset	Times observed	Z
4	{C, Q, Y, Z}	10	5·07
	{B, D, K, X}	7	5·04
	{J, P, Q, Z}	8	4·90
	{B, L, X, Z}	3	4·49
	{G, J, O, P}	5	4·42
	{H, S, U, V}	6	4·26
	{A, I, M, N}	4	3·92
	{D, K, R, W}	1	3·58
	{G, O, R, W}	1	3·44
	{C, G, O, Y}	2	3·22
	{A, E, I, T}	1	2·85
	{E, M, N, S}	2	2·47
8	{B, C, L, P, Q, X, Y, Z}	35	7·32
	{B, D, H, K, L, U, V, X}	9	5·25
	{B, D, F, K, L, R, W, X}	2	4·79
	{D, G, J, K, O, P, R, W}	1	3·88
	{D, G, K, M, N, O, R, W}	1	3·32
	{A, I, M, N, R, S, U, W}	1	3·23
	{A, G, I, M, N, O, R, W}	1	2·67
12	{B, C, D, F, J, K, L, P, Q, X, Y, Z}	45	7·66
	{B, D, F, H, K, L, R, S, T, U, V, W}	5	5·39

For each subset size, 50 random permutations were selected as starting configurations, but, again, there was no apparent relationship between the size of the initial and final improved Γ values. Table 4 presents the local optima found for each of the three values of k and also the number of times in the sample of 50 it was actually observed. It should be remembered, however, that these subsets define local optima only and other subsets not given in the table could be more homogeneous. For example, for $k = 8$ one typical starting subset that produced the largest observed value also gave three other subsets in the process that were more homogeneous with respect to Γ than any of the other local optima:

Sequence	Subset	Γ	Z-value
0 (start)	{G, N, O, P, R, S, T, Y}	−4·28	−0·62
1	{G, N, O, P, R, S, W, Y}	5·47	0·80
2	{G, J, N, O, P, R, W, Y}	15·71	2·28
3	{C, G, J, O, P, R, W, Y}	27·38	3·98
4	{C, G, J, O, P, Q, W, Y}	34·19	4·97
5	{C, G, J, O, P, Q, Y, Z}	43·98	6·39
6	{C, J, O, P, Q, X, Y, Z}	44·51	6·47
7	{C, J, L, P, Q, X, Y, Z}	49·46	7·19
8	{B, C, L, P, Q, X, Y, Z}	50·38	7·32

The best values for Γ for each of the three values of k are given below and all appear to be subsets that generally contain symbols with four or more elements:

k

4 {– · – ·; – – · –; – · – –; – – · ·}

8 {– · · ·; – · – ·; · – · ·; · – – ·; – – · –; – · · –; – · – –; – – · ·}

12 {– · · ·; – · – ·; – · ·; · · – ·; · – – –; · – ·; · – · ·; · – – ·; – – · –; – · · –; – · – –; – – · ·}

C. *Linear and Circular Seriation*

Besides encompassing hierarchical clustering and its subtle variations, QA can also be used to order objects along a circle or line. Again, several approaches could be followed (e.g. see Gelfand, 1971) but for convenience the discussion here relies upon a measure discussed by Szczotka (1972). In particular, if we wish to locate a group of n objects that appear to be orderable along a circle or a one-dimensional line, then the structure matrix takes one of the following forms.

Linear seriation:

$$C(r,s) = |r-s|.$$

Circular seriation:

If n is even,

$$C(r,s) = \begin{cases} |r-s| & \text{if } |r-s| \leqslant n/2, \\ n-|r-s & \text{if } |r-s| > n/2. \end{cases}$$

If n is odd,

$$C(r,s) = \begin{cases} |r-s| & \text{if } |r-s| \leqslant (n-1)/2, \\ n-|r-s| & \text{if } |r-s| > (n-1)/2. \end{cases}$$

As an illustration, the proximity data given in Table 2 on the ten digits were used to obtain the local optima given in Tables 5 and 6 for 50 random starting configurations. The best linear and circular seriation were given by the following sequences, where the end points for the circular ordering are assumed to be joined:

Linear:

(·····; ····–; ···––; –····; ––···; ··–––;
–––··; ·––––; ––––·; –––––)

Circular:

(–····; ·····; ····–; ···––; ··–––; ·––––;
–––––; ––––·; –––··; ––···)

Table 5.—Observed Local Optima for the Linear Seriation Specification

Linear seriation	Times observed	Z
(5, 4, 3, 6, 7, 2, 8, 1, 9, 0)	48	4·96
(3, 2, 1, 0, 9, 8, 7, 6, 4, 5)	1	3·93
(5, 4, 0, 9, 1, 8, 7, 6, 2, 3)	1	2·60

Table 6.—Observed Local Optima for the Circular Seriation Specification

Circular seriation	Times observed	Z
(6, 5, 4, 3, 2, 1, 0, 9, 8, 7)	46	4·67
(8, 2, 7, 3, 6, 4, 5, 0, 9, 1)	4	3·39

Shepard's (1963) number of dots and dashes interpretation is fairly clear in the linear case, whereas the circular model discussed by Hubert (1974*a*) is evident in the circular case.

D. *Discrete Placement in Space*

As a final illustration of the use of QA, the objects in S can also be allocated to a p-dimensional space characterized by some specific distance function. As before, several approaches could be taken using different distance functions and/or different values of p, but, for convenience, p will be assumed to equal two and the distance function Euclidean. As a first step, the two-dimensional plane is 'gridded' by equally spaced calibrations along both the X and Y axes; in particular, for analyzing the Rothkopf data, six intervals were defined at the integers $0, 1, \ldots, 5$, generating a checkerboard pattern of 36 vertices, i.e. $\{(x, y) \mid 0 \leqslant x, y \leqslant 5\}$. Consequently, the structure matrix C is defined to be 36×36, where if $r = (x_r, y_r)$, $s = (x_s, y_s)$,

$$C(r, s) = ((x_r - x_s)^2 + (y_r - y_s)^2)^{\frac{1}{2}}.$$

Since the original Q matrix is of size 26×26, an augmented Q matrix of size 36×36 must be defined by adding 'dummy objects' in the following way:

$$q'(o_u, o_v) = \begin{cases} q(o_u, o_v) & \text{if } 1 \leqslant u, v \leqslant 26, \\ 0 & \text{if } 27 \leqslant u, v \leqslant 36. \end{cases}$$

In this manner, the final object placement in the two-dimensional space allows the presence of 'holes', and clearly, the 'fineness' of the gridding is up to the researcher.

In our example, the two 36×36 matrices defined above were used in conjunction with the steepest ascent routine, and 50 random starting permutations or configurations were chosen to initiate the search. Again, there was no noticeable correspondence between the initial values of Γ and the improved indices. All of the local maxima were unique for each of the 50 runs and, using z-score units, generated final Γ values that ranged from 5·31 to 5·72. Moreover, all of the configurations produced patternings similar to that given in Table 7, which illustrates the placement corresponding to the best attained value of Γ.

Table 7.—A Discrete Placement using the Steepest Ascent Heuristic—Corresponds to the Largest Final Γ Index

··	···	··–	····	··–·	···–
·–	·–·		–··	–·–	–···
–·					–··–
––				·–··	–·–·
–	·––			––··	–·––
·	––·	–––	·–––	·––·	––·–

As can be seen from Table 7, there appears to be a general left to right gradient that is characterized by the number of elements in a symbol; also, a top to bottom gradient is present that loosely corresponds to the number of dots versus dashes. This representation is obviously cruder than that given by NMDS (see Shepard, 1963), but the same type of patterning is still evident. To some extent the 'obvious' inconsistencies in the patterning can be explained on an *ad hoc* basis by a more detailed examination of the actual proximity matrix. For instance, the two symbols – and · are very similar to each other but almost form an isolated pair in relationship to the other 24 objects. The symbols –·· and –·– are positioned in the manner shown since each is more similar to a four-element symbol than to any of the other three-element symbols, even though they are very 'close' to each other. The symbol ···· is essentially an isolate and is not highly related to any other object; the weak correspondences that can be identified for this symbol are to objects that are adjacent or very near in the obtained configuration.

The use of dummy objects is not limited to the problem of object placement in space and is appropriate for use with other structure matrices as well; for instance, in hierarchical clustering or seriation. In particular, when searching for a labeling of the objects in S that fits a particular ultrametric well, the inclusion of dummy objects allows some implicit relaxing of the form of the target ultrametric. In a similar way, the use of dummy objects in the seriation task would allow the possibility of spacing the objects along a continuum by steps other than unity. For convenience, the inclusion of dummy objects in Q will be referred

to as the 'dummy object option' and can be used as a basic principle in the non-static hypothesis testing framework as well. Here, Ω would be the set of all one–one mappings of $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n_o\}$ when $n \leqslant n_o$. The same type of mean and variance formulas given in (1) and (2) still hold except that now the initial constant multipliers are reciprocal functions of n_o rather than of n.

V. Discussion

In conclusion, the use of QA as an approach to data analysis has a number of salient features that should be reiterated. First of all, the data analysis problems we have discussed can be classified as either static or non-static. The term 'static' implies that some sort of a posited relational structure represented by a matrix C and a particular relabeling ρ_o is to be tested against the data matrix Q. On the other hand, a non-static problem implies a search for some relabeling ρ of the rows and columns of C to fit the data matrix Q, i.e. conditional on the *form* of the structure matrix C, we attempt to locate a reasonable relabeling ρ by using the data matrix Q. Some problems, however, can be treated either as static or non-static and the search for a 'good' relabeling can be made or, alternatively, the test of a particular labeling stated *a priori*. The 'steepest ascent' heuristic appears flexible enough to serve as a general strategy for locating a reasonable ρ in the class of structures representable by a structure matrix C. Moreover, this particular computational approach is extremely promising as a general basis for an interactive orientation to many of the data analysis tasks encountered in the behavioral sciences concerned with proximity matrices and identifying a structural organization within a set of objects.

One potentially useful generalization of the Graves & Whinston expressions (see Graves & Whinston, 1970*a*, *b*) allows for an initial fixing of certain elements of the map ρ; in other words, the mean and variance of the criterion Γ over all equally likely completions of a partial permutation can be obtained. The necessary expressions may be developed formally as follows.

Suppose Ω_α is the set of all one–one mappings of the set of integers $M = \{1, 2, \ldots, n\}$ into the set $N = \{1, 2, \ldots, n\}$, where k fixed elements in M are mapped into k fixed elements in N. Again, $4 \leqslant n$, and suppose D is the set of fixed elements in M, R is the set of fixed elements in N, and $\alpha(\cdot)$ specifies the one–one assignment between D and R. Then, under the assumption that all elements ρ_α in Ω_α are equally likely, the criterion

$$\Gamma(\rho_\alpha) = \sum_{u,v} q_{uv} C_{\rho\alpha(u)\rho\alpha(v)}$$

has expectations and variance given in (8) and (9) below. The conditions on Q and C are the same as before, and for notational convenience, $q(o_u, o_v) \equiv q_{uv}$ and $C(r, s) \equiv C_{rs}$.

Let A be an $n-k \times n-k$ matrix, where

$$a_{ur} = \sum_{v \in D} [q_{vu} C_{\rho\alpha(v)r} + q_{uv} C_{r\rho\alpha(v)}].$$

Then,

$$E_{\rho_\alpha}(\Gamma(\rho_\alpha)) = \sum_{u,v\in D} q_{uv}\, C_{\rho_\alpha(u)\rho_\alpha(v)} + (1/(n-k)) \sum_{u\notin D, r\notin R} a_{ur} + (1/((n-k)(n-k-1)))\left(\sum_{u,v\notin D} q_{uv}\right)\left(\sum_{r,s\notin R} C_{rs}\right). \quad (8)$$

Furthermore, define

$$F_1 = \left(\sum_{u\notin D, r\notin R} a_{ur}\right)^2,$$

$$F_2 = \sum_{u\notin D}\left(\sum_{r\notin R} a_{ur}\right)^2,$$

$$F_3 = \sum_{r\notin R}\left(\sum_{u\notin D} a_{ur}\right)^2,$$

$$F_4 = \sum_{u\notin D, r\notin R} a_{ur}{}^2,$$

$B_1', B_2', \ldots, B_7'$ are the same terms as $B_1, B_2, \ldots, B_7$ defined previously except now the summations are over u and $v\notin D$ and over r and $s\notin R$;

$$G_1 = \left(\sum_{u\notin D, r\notin R} a_{ur}\right)\left(\sum_{u,v\notin D} q_{uv}\right)\left(\sum_{r,s\notin R} C_{rs}\right),$$

$$G_2 = \sum_{u\notin D, r\notin R} a_{ur}\left(\left(\sum_{v\notin D} q_{uv}\right)\left(\sum_{s\notin R} C_{rs}\right) + \left(\sum_{v\notin D} q_{vu}\right)\left(\sum_{s\notin R} C_{sr}\right)\right),$$

$$G_3 = \sum_{u\notin D, r\notin R} a_{ur}\left(\left(\sum_{u',v\notin D} q_{u'v} - \sum_{v\notin D} q_{uv} - \sum_{v\notin D} q_{vu}\right)\left(\sum_{r',s\notin R} C_{r's} - \sum_{s\notin R} C_{rs} - \sum_{s\notin R} C_{sr}\right)\right).$$

Then,

$$\begin{aligned}\mathrm{var}_{\rho_\alpha}(\Gamma(\rho_\alpha)) = {} & [1/((n-k)^2(n-k-1))](F_1 - G_1) \\ & + [-1/((n-k)(n-k-1))](F_2 + F_3 - G_2 - B_2' - B_3') \\ & + [1/(n-k-1)]F_4 \\ & + [1/((n-k)(n-k-1)(n-k-2))](G_3 + B_4' + 2B_5' + B_6') \\ & + [-1/((n-k)^2(n-k-1)^2)]B_1' \\ & + [1/((n-k)(n-k-1)(n-k-2)(n-k-3))]B_7'. \end{aligned} \quad (9)$$

Again, there are numerous data analytic problems that correspond to special cases of the 'ρ_α' framework, and in fact, similar formulas hold when

$$N = \{1, 2, \ldots, n_o\},$$

where $n_o \geqslant n$. Several rather natural possibilities are indicated below with the algebraic details of reducing the formulas in (8) and (9) left to the reader. In

particular, within the Hamiltonian chain context introduced earlier, the formulas in (8) and (9) allow the fixing of specific end points of a seriation; secondly, the change in Johnson's λ statistic can be assessed that results from adding a specific number of additional objects to a fixed subset; and finally, the distribution of the weights of cuts of a specified size can be evaluated, where the cut separates two known objects. Within a somewhat broader context, the set of 'fixed' elements could correspond to one particular type of object and the variable or unplaced elements could correspond to a distinct second set. For instance, the fixed elements may be stimuli and the variable objects may be people. In this context the placements would correspond to the location of 'ideal points' within the general framework proposed by Coombs (1964) and his co-workers.

The research of the first author was supported in part by a Postdoctoral Research Training Fellowship from the Social Science Research Council (USA) and by the National Science Foundation through NSFSOC75–07860. Equal authorship is implied.

References

AGIN, M. (1966). Optimum seeking with branch and bound. *Manag. Sci.* **13**, 176–185.

ARABIE, P. & BOORMAN, S. A. (1973). Multidimensional scaling of measures between partitions. *J. math. Psychol.* **10**, 148–203.

ARMOUR, G. C. & BUFFA, E. S. (1963). A heuristic algorithm and simulation approach to relative location of facilities. *Manag. Sci.* **9**, 294–309.

BARTON, D. E. & DAVID, F. N. (1966). The random intersection of two graphs. In F. N. David (ed.), *Research Papers in Statistics.* New York: Wiley.

BLOEMENA, A. R. (1964). *Sampling from a Graph.* Amsterdam: Mathematisch Centrum.

BRADLEY, J. V. (1968). *Distribution-free Statistical Tests.* Englewood Cliffs, N.J.: Prentice-Hall.

BREUER, M. A. (1966*a*). The application of integer programming in design automation. *Design Automation Workshop* **3**, 1–19.

BREUER, M. A. (1966*b*). The formulation of some allocation and connection problems as integer programs. *Naval Res. Logist. Q.* **13**, 83–95.

CONWAY, R. W. & MAXWELL, W. L. (1961). A note on the assignment of facility location. *J. Indust. Engin.* **12**, 34–36.

COOMBS, C. H. (1964). *A Theory of Data.* New York: Wiley.

DAVID, F. N. & BARTON, D. E. (1962). *Combinatorial Chance.* New York: Hafner.

DOBSON, A. J. (1974). Unrooted trees for numerical taxonomy. *J. appl. Prob.* **11**, 32–42.

DONATH, W. E. (1968). Statistical properties of the placement of a graph. *SIAM J. appl. Math.* **16**, 439–457.

DONATH, W. E. & HOFFMAN, A. J. (1973). Lower bounds for the partitioning of graphs. *IBM J. Res. Dev.* **17**, 420–425.

EDWARDS, H., GILLETT, B. & HALE, M. (1970). Modular allocation technique (MAT). *Manag. Sci.* **17**, 161–169.

FILLENBAUM, S. & RAPOPORT, A. (1971). *Structures in the Subjective Lexicon.* New York: Academic Press.

FORD, L. R. & FULKERSON, D. R. (1962). *Flows in Networks.* Princeton, N.J.: Princeton University Press.

FRANCIS, R. L. & WHITE, J. A. (1974). *Facility Layout and Location: An Analytical Approach.* Englewood Cliffs, N.J.: Prentice-Hall.

GARSIDE, R. G. & NICHOLSON, T. A. J. (1968). Permutation procedures for the backboard-wiring problem. *Proc. IEE* **115**, 27–30.

GAVETT, J. W. & PLYTER, N. V. (1966). The optimal assignment of facilities to locations by branch and bound. *Operat. Res.* **14**, 210–232.

GELFAND, A. E. (1971). Rapid seriation methods with archaeological applications. In F. R. Hodson *et al.* (eds), *Mathematics in the Archaeological and Historical Sciences*. Edinburgh: Edinburgh University Press.

GIBBONS, J. D. (1971). *Nonparametric Statistical Inference*. New York: McGraw-Hill.

GILMORE, P. C. (1962). Optimal and suboptimal algorithms for the quadratic assignment problem. *SIAM J. appl. Math.* **10**, 305–313.

GLASER, R. H. A. (1959). A quasi-simplex method for designing suboptimum packages of electronic building blocks (Burroughs 220). In *Proc. 1959 Computer Applications Symp.* Armour Research Foundation, Illinois Institute of Technology, Chicago.

GORINSHTEYN, L. L. (1969). The partitioning of graphs. *Engin. Cyber.* **1**, 76–82.

GRAVES, G. W. & WHINSTON, A. B. (1970*a*). An algorithm for the quadratic assignment problem. *Manag. Sci.* **17**, 453–471.

GRAVES, G. W. & WHINSTON, A. B. (1970*b*). Derivation of the mean and variance for the quadratic assignment problem. Unpublished report.

GREENBERG, H. (1969). A quadratic assignment problem without column constraints. *Naval Res. Logist. Q.* **16**, 417–421.

GUTTMAN, L. (1954). A new approach to factor analysis: The Radex. In P. Lazarsfeld (ed.), *Mathematical Thinking in the Social Sciences*. New York: Free Press.

GUTTMAN, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika* **33**, 469–506.

HANAN, M. & KURTZBERG, J. M. (1972*a*). A review of the placement and quadratic assignment problems. *SIAM Rev.* **14**, 324–342.

HANAN, M. & KURTZBERG, J. M. (1972*b*). Placement techniques. In M. A. Breuer (ed.), *Design Automation of Digital Systems*, vol. 1. Englewood Cliffs, N.J.: Prentice-Hall.

HILLIER, F. S. & CONNORS, M. M. (1966). Quadratic assignment problem algorithms and the location of indivisible facilities. *Manag. Sci.* **13**, 42–57.

HUBERT, L. (1974*a*). Some applications of graph theory and related nonmetric techniques to problems of approximate seriation: The case of symmetric proximity measures. *Br. J. math. statist. Psychol.* **27**, 133–153.

HUBERT, L. (1974*b*). Problems of seriation using a subject by item response matrix. *Psychol. Bull.* **81**, 976–983.

HUBERT, L. (1974*c*). Some applications of graph theory to clustering. *Psychometrika* **39**, 283–309.

HUBERT, L. Data analysis implications of some concepts related to the cuts of a graph. *J. math. Psychol.* (in press).

HUBERT, L. & LEVIN, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Psychol. Bull.* **83** (in press).

HUBERT, L. & SCHULTZ, J. (1975*a*). Maximum-likelihood paired-comparison ranking and quadratic assignment. *Biometrika* **62**, 665–659.

HUBERT, L. & SCHULTZ, J. (1975*b*). Seriation and quadratic assignment. Unpublished manuscript.

JARDINE, N. & SIBSON, R. (1971). *Mathematical Taxonomy*. New York: Wiley.

JENSEN, R. E. (1969). A dynamic programming algorithm for cluster analysis. *Operat. Res.* **12**, 1034–1057.

JOHNSON, S. C. (1967). Hierarchical clustering schemes. *Psychometrika* **32**, 241–254.

JOHNSON, S. C. (1968*a*). A simple cluster statistic. Unpublished manuscript.

JOHNSON, S. C. (1968*b*). Metric clustering. Unpublished manuscript.

KENDALL, M. G. (1970). *Rank Correlation Methods*, 4th ed. New York: Hafner.

KOOPMANS, T. C. & BECKMANN, M. (1957). Assignment problems and the location of economic activities. *Econometrica* **25**, 53–76.

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

KRUSKAL, J. B. (1964*b*). Multidimensional scaling: A numerical method. *Psychometrika* **29**, 115–129.

KURTZBERG, J. M. (1965). Algorithms for backplane formation. In *Microelectronics and Large Systems*. Washington D.C.: Spartan Books.

LAND, A. H. (1963). A problem of assignment with inter-related costs. *Operat. Res. Q.* **14**, 185–199.

LAWLER, E. L. (1963). A quadratic assignment problem. *Manag. Sci.* **9**, 586–599.

MANTEL, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.

MUNKES, J. (1957). Algorithms for the assignment and transportation problem. *SIAM J. appl. Math.* **5**, 32–38.

NICHOLSON, T. A. J. (1971). A method for optimizing permutation problems and its industrial applications. In P. J. A. Welsh (ed.), *Combinatorial Mathematics and its Applications*. New York: Academic Press.

NUGENT, C. E., VOLLMANN, T. E. & RUML, J. (1968). An experimental comparison of techniques for the assignment of facilities to locations. *Operat. Res.* **16**, 150–173.

PEGELS, C. C. (1966). Plant layout and discrete optimizing. *Int. J. Production Res.* **5**, 81–92.

PIERCE, J. F. & CROWSTON, W. B. (1971). Tree-search algorithms for quadratic assignment problems. *Naval Res. Logist. Q.* **18**, 1–36.

POMENTALE, T. (1965). An algorithm for minimizing backboard wiring functions. *Commun. ACM* **8**, 699–703.

POMENTALE, T. (1967). On minimization of backboard wiring functions. *SIAM Rev.* **9**, 564–568.

RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Am. statist. Ass.* **66**, 846–850.

RAO, M. R. (1971). Cluster analysis and mathematical programming. *J. Am. statist. Ass.* **66**, 622–626.

REITER, S. & SHERMAN, G. (1965). Discrete optimizing. *SIAM J. appl. Math.* **13**, 864–889.

REMAGE, R. & THOMPSON, W. A. (1966). Maximum-likelihood paired comparison rankings. *Biometrika* **53**, 143–149.

RESTLE, F. (1959). A metric and an ordering on sets. *Psychometrika* **24**, 207–219.

RITZMAN, L. (1972). The efficiency of computer algorithms for plant layout. *Manag. Sci.* **18**, 240–248.

ROTHKOPF, E. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *J. exp. Psychol.* **53**, 94–101.

RUTMAN, R. A. (1964). An algorithm for placement of interconnected elements based upon minimum wire length. In *AFIPS Proc., 1964 Spring Joint Computer Conf.*

SCHULTZ, J. & HUBERT, L. (1976). A nonparametric test for the correspondence between two proximity matrices. *J. educ. Statist.* **1**, 59–67.

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

SHEPARD, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Hum. Factors* **5**, 19–34.

SHEPARD, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika* **39**, 373–421.

SNEATH, P. H. A. & SOKAL, R. R. (1973). *Numerical Taxonomy*. San Francisco: Freeman.

STEINBERG, L. (1961). The backboard wiring problem: A placement algorithm. *SIAM Rev.* **3**, 37–50.

SZCZOTKA, F. (1972). On a method of ordering and clustering of objects. *Zastosowania Math.* **13**, 23–33.

WALD, A. & WOLFOWITZ, A. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *Ann. math. Statist.* **14**, 387–388.

WILKINSON, E. M. (1971). Archaeological seriation and the travelling salesman problem. In F. R. Hodson *et al.* (eds), *Mathematics in the Archaeological and Historical Sciences*. Edinburgh: Edinburgh University Press.

WIMMERT, R. J. (1958). A mathematical method of equipment location. *J. indust. Engin.* **9**, 498–505.

ZAHN, C. T. (1964). Approximating symmetric relations by equivalence relations. *SIAM J. appl. Math.* **12**, 840–846.