Chapter 4

Object Sequencing and Seriation

"Things are ordered in a wonderful manner."

Joseph Conrad
Under Western Eyes (1911)

"...in consideration of the imperfection inherent in the order of the world..."

Heinrich von Kleist
The Marquise of O (1808)
Translation by Nigel Reeves

" '...so Princeton,' Dawn said, 'so *unerring*. He works so hard to be one-dimensional...' "

Philip Roth
American Pastoral (1997)

The three major sections of this chapter discuss various aspects of the task of optimally sequencing (or seriating) a set S of n objects along a continuum based on whatever proximity data may be available between the object pairs. The ultimate purpose of any sequencing technique, as is generally true for all CDA methods, is to use whatever combinatorial structure is identified (which in the present context will be object orderings) to help explain the relationships that may be present among the objects, as reflected by the proximity data. Section 4.1 is concerned with the (unconstrained) sequencing of a *single* object set, although this single set may itself be the union of two disjoint sets. In this latter case, the task of object placement is one of obtaining a joint sequencing of the two sets (i.e., we have a data analysis task usually discussed under the subject of unidimensional unfolding; e.g., see Heiser (1981)). In general, proximity data may be in the form of a one-mode matrix that may either be symmetric (which would be consistent with the discussion in the clustering framework of Chapter 3) or nonsymmetric (and more specifically, skew-symmetric), or possibly in the form of a two-mode matrix if the set S is formed by the union of two other sets. Section 4.2 discusses only two-mode proximity data, within a context where if the single object set S is formed from the union of the sets S_A and S_B, the ordering of the objects *within* S_A and/or *within* S_B is constrained in their joint placement along a continuum. In effect, we are combining (or in a slightly different interpretive sense, we are comparing) the two distinct object sets, subject to some type of constraint on the orderings of the objects within S_A and S_B. The final Section 4.3 provides an optimization problem constituting a mixture of an unconstrained sequencing task, as discussed in Section 4.1, and the clustering problem of Section 3.1. Specifically, the goal will be to construct optimal partitions of an object set S in which the classes of the partition must themselves be ordered along a continuum. Obviously, when the number of classes in the partition is set equal to the number of objects in S, this latter optimization task reduces to the sequencing of a single object set as described in Section 4.1.

In the topics outlined briefly above that are to be pursued in the sections to follow, three types of proximity data will organize the presentation. One type comes in the usual form of a one-mode $n \times n$ nonnegative dissimilarity matrix for the n objects in S denoted earlier by $\mathbf{P} = \{p_{ij}\}$, where $p_{ij} = p_{ji} \geq 0$ and $p_{ii} = 0$ for $1 \leq i, j \leq n$. A second type is an $n_A \times n_B$ two-mode nonnegative dissimilarity matrix $\mathbf{Q}$, where proximity is defined only between the objects from two distinct sets S_A and S_B containing, respectively, n_A and n_B objects and forming the rows and columns of $\mathbf{Q}$. (For convenience, the objects in S_A are again denoted as $\{r_1, \ldots, r_{n_A}\}$ and those in S_B as $\{c_1, \ldots, c_{n_B}\}$, where the letters 'r' and 'c' signify 'row' and 'column', respectively.) In the two-mode context, it will typically be convenient to construct a single mode from the objects in both S_A and S_B (i.e., $S \equiv S_A \cup S_B$), and use $\mathbf{Q}$ to generate a nonnegative symmetric proximity matrix on S that contains missing proximities for object pairs contained within S_A or within S_B. To refer to this latter matrix explicitly, we introduce the notation $\mathbf{P}^{AB} = \{p_{ij}^{AB}\}$, where $\mathbf{P}^{AB}$ has the block form

$$\begin{bmatrix} * & \mathbf{Q} \\ \mathbf{Q}' & * \end{bmatrix}$$

and the asterisk * denotes those proximities that are missing.

Finally, if the proximity information originally given between the pairs of objects in S is nonsymmetric and in the form of an $n \times n$ nonnegative matrix, say, $\mathbf{D} = \{d_{ij}\}$, this latter matrix will first be decomposed into the sum of its symmetric and skew-symmetric components,

$$\mathbf{D} = [(\mathbf{D} + \mathbf{D}')/2] + [(\mathbf{D} - \mathbf{D}')/2],$$

and each of these components will (always) be addressed separately. The matrix $(\mathbf{D} + \mathbf{D}')/2$ can be treated merely as a nonnegative symmetric dissimilarity matrix (i.e., in the notation $\mathbf{P}$), and methods appropriate for such a proximity measure can be applied. The second component of the form $(\mathbf{D} - \mathbf{D}')/2$ is an $n \times n$ matrix denoted by $\mathbf{P}^{SS} = \{p_{ij}^{SS}\}$, where $p_{ij}^{SS} = (d_{ij} - d_{ji})/2$ for $1 \leq i, j \leq n$, and the superscript 'SS' signifies 'skew-symmetric', i.e., $p_{ij}^{SS} = -p_{ji}^{SS}$ for all $1 \leq i, j \leq n$. Thus, there is an explicit directionality to the (dominance) relationship specified between any two objects depending on the order of the subscripts, i.e., p_{ji}^{SS} and p_{ij}^{SS} are equal in absolute magnitude but differ in sign. It may also be worth noting here that any skew-symmetric matrix $\{p_{ij}^{SS}\}$ can itself be interpreted as containing two separate sources of information: one is the directionality of the dominance relationship given by $\{\text{sign}(p_{ij}^{SS})\}$, where (as defined earlier) $\text{sign}(y) = 1$ if $y > 0$; $= 0$ if $y = 0$; and $= -1$ if $y < 0$; the second is in the absolute magnitude of dominance given by $\{| \, p_{ij}^{SS} \, |\}$. This latter matrix is symmetric and can be viewed as a dissimilarity matrix and analyzed as such. In fact, the first numerical examples below on object sequencing for a one-mode symmetric proximity matrix will use the paired-comparisons data from Table 1.2 in exactly this manner for constructing optimal orderings of the offenses according to their perceived seriousness.

4.1 Optimal Sequencing of a Single Object Set

The search for an optimal sequencing of a single object set S (irrespective of the type of proximity measure available and whether S itself is the union of two other sets) can be operationalized by constructing a best reordering for the rows and simultaneously the columns of an $n \times n$ proximity matrix. The row/column reordering to be identified will optimize, over all possible row/column reorderings, some specified measure of patterning for the entries of the reordered matrix. For convenience, the particular measure of pattern will usually be defined so as to be maximized (with a few exceptions, as in Section 4.1.4), and thus, the general form of the GDPP to be applied is the recursion given in (2.3). (Again, much as in the hierarchical clustering framework, the application of the max/min type of recursion in (2.5) has limited utility. When constructing an optimal ordering based on the type of measure of matrix patterning we will consider, the various positions in an ordering will have differential possible contributions to the measure of matrix pattern. As a consequence, a max/min criterion would overemphasize the placement of a very few middle objects in the ordering and effectively ignore how well the entire early and late portions of a sequence were constructed. We note, however, that not all possible measures of matrix patterning would have this property of a differential possible contribution depending on location in an ordering. Specifically, there is a particular application of a max/min (or min/max) recursion for obtaining a best sequencing through the construction of optimal paths that will be discussed later in Section 4.1.4.)

A variety of specific measures of patterning will be introduced in the next three sections within the context of the particular type of proximity measure most appropriate (i.e., symmetric ($\mathbf{P}$), skew-symmetric ($\mathbf{P}^{SS}$), or defined between two sets S_A and S_B ($\mathbf{P}^{AB}$)). However, in all three cases the same specialization of the GDPP in (2.3) will be implemented. (We should again note that a variation is introduced in Section 4.1.4 in a discussion of object sequencing based on the construction of optimal paths.)

A collection of sets $\Omega_1, \ldots, \Omega_n$ is defined (so, $K \equiv n$), where Ω_k includes all the subsets that have k members from the integer set $\{1, 2, \ldots, n\}$. The value $\mathcal{F}(A_k)$ is the optimal contribution to the total measure of matrix patterning for the objects in A_k when they occupy the first k positions in the (re)ordering. A transformation is now possible between $A_{k-1} \in \Omega_{k-1}$ and $A_k \in \Omega_k$ if $A_{k-1} \subset A_k$ (i.e., A_{k-1} and A_k differ by one integer). The contribution to the total measure of patterning generated by placing the single integer in $A_{k-1} - A_k$ at the k^{th} order position is $M(A_{k-1}, A_k)$. As always, the validity of the recursive process will require the incremental merit index $M(A_{k-1}, A_k)$ to depend only on the unordered sets A_{k-1} and A_k and the complement $S - A_k$, and specifically *not* on how A_{k-1} may have been reached beginning with Ω_1. Assuming $\mathcal{F}(A_1)$ for all $A_1 \in \Omega_1$ are available, the recursive process can be carried out from Ω_1 to Ω_n, with $\mathcal{F}(A_n)$ for the single set $A_n = \{1, 2, \ldots, n\} \in \Omega_n$ defining the optimal value for the specified measure of matrix patterning. The optimal row/column reordering is constructed, as always, by working backward through the recursion.[27]

4.1.1 Symmetric One-Mode Proximity Matrices

When the original data come in the form of a nonnegative symmetric proximity matrix $\mathbf{P}$ with no missing entries, there are (as might be expected) many different indices of patterning that could be optimized in a row/column reordered proximity matrix through the specialization of the GDPP described above. We will emphasize two general classes of such measures below.[28]

Row and/or column gradient measures. One ubiquitous concept encountered in the literature on matrix reordering is that of a symmetric proximity matrix having an anti-Robinson form (this same structure was noted briefly in the clustering context when an optimal unconstrained clustering might also be optimal when order-constrained). Specifically, suppose $\rho(\cdot)$ is some permutation of the first n integers that reorders both the rows and columns of $\mathbf{P}$ (i.e., $\mathbf{P}_\rho \equiv \{p_{\rho(i)\rho(j)}\}$). The reordered matrix $\mathbf{P}_\rho$ is said to have an anti-Robinson form if the entries within the rows and within the columns of $\mathbf{P}_\rho$ moving away from the main diagonal in any direction never decrease; or formally, two gradient conditions must be satisfied:

within rows: $p_{\rho(i)\rho(k)} \le p_{\rho(i)\rho(j)}$ for $1 \le i < k < j \le n$;
within columns: $p_{\rho(k)\rho(j)} \le p_{\rho(i)\rho(j)}$ for $1 \le i < k < j \le n$.

We might note that whenever $\mathbf{P}$ is an ultrametric, or if $\mathbf{P}$ has an exact Euclidean representation in a single dimension (i.e., $\mathbf{P} = \{|x_j - x_i|\}$ for some collection of coordinate values, $x_1, x_2, \ldots, x_n$), then $\mathbf{P}$ can be row/column reordered to display a perfect anti-Robinson pattern. Thus, the notion of an anti-Robinson form can be interpreted as generalizing either a perfect discrete classificatory structure induced by a partition hierarchy (through an ultrametric) or as the pattern expected in $\mathbf{P}$ if there exists an exact unidimensional

Euclidean representation for the objects in S. In any case, if a matrix can be row/column reordered to display an anti-Robinson form, then the objects are orderable along a continuum so that the degree of separation between objects in the ordering is reflected perfectly by the dissimilarity information in $\mathbf{P}$, i.e., for the object ordering, $O_{\rho(i)} \prec O_{\rho(k)} \prec O_{\rho(j)}$ (for $i < k < j$), $p_{\rho(i)\rho(k)} \leq p_{\rho(i)\rho(j)}$ and $p_{\rho(k)\rho(j)} \leq p_{\rho(i)\rho(j)}$ (or equivalently, because the most extreme separation in the ordering of the three distinct objects is between $O_{\rho(i)}$ and $O_{\rho(j)}$, we have $p_{\rho(i)\rho(j)} \geq \max[p_{\rho(i)\rho(k)}, p_{\rho(k)\rho(j)}]$).[29]

A natural (merit) measure for how well the particular reordered proximity matrix $\mathbf{P}_\rho$ satisfies these two gradient conditions would rely on an aggregate index of the violations/nonviolations over all distinct object triples, as given by the expression

$$\sum_{i<k<j} f(p_{\rho(i)\rho(k)}, p_{\rho(i)\rho(j)}) + \sum_{i<k<j} f(p_{\rho(k)\rho(j)}, p_{\rho(i)\rho(j)}), \tag{4.1}$$

where $f(\cdot,\cdot)$ is some function indicating how a violation/nonviolation of a particular gradient condition for an object triple within a row or within a column (and defined above the main diagonal of $\mathbf{P}_\rho$) is to be counted in the total measure of merit. The two options we concentrate on are as follows:

(1) $f(z,y) = \text{sign}(z-y) = +1$ if $z > y$; 0 if $z = y$; and -1 if $z < y$; thus, the (raw) number of satisfactions minus the number of dissatisfactions of the gradient conditions *within rows* above the main diagonal of $\mathbf{P}_\rho$ would be given by the first term in (4.1),

$$\sum_{i<k<j} f(p_{\rho(i)\rho(k)}, p_{\rho(i)\rho(j)}), \tag{4.2}$$

and the (raw) number of satisfactions minus dissatisfactions of the gradient conditions *within columns* above the main diagonal of $\mathbf{P}_\rho$ would be given by the second term in (4.1),

$$\sum_{i<k<j} f(p_{\rho(k)\rho(j)}, p_{\rho(i)\rho(j)}). \tag{4.3}$$

To refer to an application of this simple counting mechanism, the phrase *unweighted gradient measure* will be adopted.

(2) $f(z,y) = |z-y|\text{sign}(z-y)$; here, and in contrast to (1), $\text{sign}(z-y)$ is also weighted by the absolute difference between z and y, to generate a *weighted gradient measure* within rows or within columns. Thus, the weighted number of satisfactions minus the number of dissatisfactions of the gradient conditions *within rows* above the main diagonal of $\mathbf{P}_\rho$ would be given by the first term in (4.1) (labeled (4.2) above), and the weighted number of satisfactions minus the number of dissatisfactions of the gradient conditions *within columns* above the main diagonal of $\mathbf{P}_\rho$ would be given by the second term in (4.1) (labeled (4.3) above).

To carry out the GDPP based on the measure in (4.1), an explicit form must be given for the incremental contribution, $M(A_{k-1}, A_k)$, to the total merit measure of patterning generated by placing the single integer in $A_k - A_{k-1}$ at the k^{th} order position. We observe first that the index in (4.1) for the given matrix $\mathbf{P}$ in its original (identity) order can be rewritten as

$$\sum_{k=1}^{n} I_{row}(k) + \sum_{k=1}^{n} I_{col}(k),$$

where

$$I_{row}(k) \equiv \sum_{i=1}^{k-1} \sum_{j=k+1}^{n} f(p_{ik}, p_{ij})$$

and

$$I_{col}(k) \equiv \sum_{i=1}^{k-1} \sum_{j=k+1}^{n} f(p_{kj}, p_{ij}).$$

Neither $I_{row}(k)$ nor $I_{col}(k)$ depends on the ordering of the objects placed either before or after the index k. Thus, generalizing such a decomposition for any ordering $\rho(\cdot)$ of the rows and columns of $\mathbf{P}$, the merit increment for placing an integer, say, k' ($\equiv \rho(k)$) (i.e., $\{k'\} = A_k - A_{k-1}$) at the k^{th} order position can be defined as

$$\sum_{k=1}^{n} I_{row}(\rho(k)) + \sum_{k=1}^{n} I_{col}(\rho(k)),$$

where

$$I_{row}(\rho(k)) = \sum_{i' \in A_{k-1}} \sum_{j' \in S - A_k} f(p_{i'k'}, p_{i'j'}),$$

$$I_{col}(\rho(k)) = \sum_{i' \in A_{k-1}} \sum_{j' \in S - A_k} f(p_{k'j'}, p_{i'j'}),$$

and $A_{k-1} = \{\rho(1), \ldots, \rho(k-1)\}$, $S - A_k = \{\rho(k+1), \ldots, \rho(n)\}$. Thus, letting $\mathcal{F}(A_1) = 0$ for all $A_1 \in \Omega_1$ and using one of the two possible specifications for $f(\cdot, \cdot)$ suggested above (i.e., $f(z, y) = \text{sign}(z - y)$ or $f(z, y) = |z - y|\text{sign}(z - y)$), the recursion in (2.3) can be carried out to identify an optimal row/column reordering of the given proximity matrix $\mathbf{P}$ to maximize either the unweighted or the weighted gradient measure over all row/column reorderings of $\mathbf{P}$.

In using (4.1), *both* the row and column gradient measures in (4.2) and (4.3), respectively, are considered as a sum in optimizing a total merit measure of patterning, where each is defined above the main diagonal of the symmetric reordered matrix $\mathbf{P}_\rho$. It might be noted that either (4.2) or (4.3) could be used by itself and a best reordering of $\mathbf{P}$ could be constructed that would maximize one or the other through the separate consideration of $\sum_{k=1}^{n} I_{row}(k)$ or $\sum_{k=1}^{n} I_{col}(k)$.[30] Because $\mathbf{P}$ is symmetric, an optimal row/column reordering based solely on the row gradient measure will also identify an optimal row/column reordering based only on the column gradient measure. Specifically, a complete reversal of the row/column reordering that is optimal for the row gradient measure will be an optimal row/column reordering for the column gradient measure (and conversely).[31]

Several numerical applications are given at the end of this section to illustrate the optimal sequencing of a single object set based on optimizing the merit measure in (4.1), as well as just one of its constituent terms in (4.2) or (4.3). For these illustrations, in addition to giving the optimal merit values achieved, convenient descriptive indices are provided for how well the gradient conditions are satisfied. These descriptive ratios are defined by the optimal index values divided by the sum of the contributions for the nonviolations and violations, where the denominators can be interpreted as the maximum the index could be for the given proximities and supposing the gradient conditions were perfectly satisfied. To be explicit, the merit measure in (4.1), for any permutation $\rho(\cdot)$, can be written as the difference between two nonnegative terms, with the first corresponding to the nonviolations of the gradient conditions and the second to the violations. Thus, when the *sum* of these two nonnegative terms is used to divide the difference, the ratio obtained (which is bounded by $+1$ and -1) reflects the (possibly weighted) ratio of the observed index to the maximum that is possible if the gradient conditions were satisfied perfectly. Formally, the descriptive ratio could be given by

$$\frac{\sum_{i<k<j} f(p_{\rho(i)\rho(k)}, p_{\rho(i)\rho(j)}) + \sum_{i<k<j} f(p_{\rho(k)\rho(j)}, p_{\rho(i)\rho(j)})}{\sum_{i<k<j} |f(p_{\rho(i)\rho(k)}, p_{\rho(i)\rho(j)})| + \sum_{i<k<j} |f(p_{\rho(k)\rho(j)}, p_{\rho(i)\rho(j)})|}. \tag{4.4}$$

Analogously, when only the gradient conditions, say, within rows are considered by the use of (4.2), a descriptive ratio could be given by

$$\frac{\sum_{i<k<j} f(p_{\rho(i)\rho(k)}, p_{\rho(i)\rho(j)})}{\sum_{i<k<j} |f(p_{\rho(i)\rho(k)}, p_{\rho(i)\rho(j)})|}. \tag{4.5}$$

As noted, in the numerical applications at the end of this current section, but also in some later generalizations as well, descriptive ratios for an optimal solution will be routinely reported in addition to giving the optimal values achieved for their numerators.

Measures of matrix patterning based on coordinate representations. There are several measures of matrix patterning that can be derived indirectly from the auxiliary problem of attempting to fit a given proximity matrix $\mathbf{P}$ by some type of unidimensional scaling representation. Because detailed discussions of this task are available in the literature (e.g., see Defays (1978); Hubert and Arabie (1986); Groenen (1993); and Hubert, Arabie, and Meulman (1997a)), we merely report the necessary results here and give the derived measure of matrix pattern (used below in the numerical illustrations). Explicitly, suppose we wish to find a set of n ordered coordinate values, $x_1 \leq \cdots \leq x_n$ (such that $\sum_k x_k = 0$), and a permutation $\rho(\cdot)$ to minimize the least-squares criterion

$$\sum_{i<j} (p_{\rho(i)\rho(j)} - \mid x_j - x_i \mid)^2.$$

After some algebraic reduction, this latter least-squares criterion can be rewritten as

$$\sum_{i<j} p_{ij}^2 + n \sum_k \left[x_k - \left(\frac{1}{n}\right) G(\rho(k)) \right]^2 - \left(\frac{1}{n}\right) \sum_k [G(\rho(k))]^2,$$

where

$$G(\rho(k)) = \sum_{i=1}^{k-1} p_{\rho(k)\rho(i)} - \sum_{i=k+1}^{n} p_{\rho(k)\rho(i)}.$$

If the measure

$$\sum_{k=1}^{n} [G(\rho(k))]^2 \tag{4.6}$$

is maximized over all row/column reorderings of $\mathbf{P}$ and the optimal permutation is denoted by $\rho^*(\cdot)$, then $G(\rho^*(1)) \leq \cdots \leq G(\rho^*(n))$, and the optimal coordinates can be retrieved as $x_k = (1/n)G(\rho^*(k))$ for $1 \leq k \leq n$. The minimum value for the least-squares criterion is

$$\sum_{i<j} p_{ij}^2 - \left(\frac{1}{n}\right) \sum_k [G(\rho^*(k))]^2.$$

To execute the GDPP recursion using (4.6), the merit increment for placing the integer, say, k' ($\equiv \rho(k)$) (i.e., $\{k'\} = A_k - A_{k-1}$), in the k^{th} order position can be written as $[G(\rho(k))]^2$, where

$$G(\rho(k)) = \sum_{i' \in A_{k-1}} p_{k'i'} - \sum_{j' \in S - A_k} p_{k'j'},$$

with $A_{k-1} = \{\rho(1), \ldots, \rho(k-1)\}$, $S - A_k = \{\rho(k+1), \ldots, \rho(n)\}$, and $\mathcal{F}(A_1)$ for $A_1 = \{k'\} \in \Omega_1$ defined by

$$\left(\sum_{j' \in S - \{k'\}} p_{k'j'} \right)^2 .$$

The recursion in (2.3) can then be carried out to identify an optimal row/column reordering $\rho^*(\cdot)$ of $\mathbf{P}$ based on the measure in (4.6); also, as noted above, the optimal coordinates x_k ($1 \leq k \leq n$) are $(1/n)G(\rho^*(k))$, where the integer $\rho^*(k)$ is placed at the k^{th} order position in an optimal row/column reordering.

To give a more intuitive sense of the measure being optimized in (4.6), we first note that for an optimal row/column reordering $\rho^*(\cdot)$, the variance of the coordinate estimates is maximized (using the fact that the sum of the optimal coordinates, $\sum_k (1/n)G(\rho^*(k))$, is zero). Thus, the pattern of entries sought in an optimally reordered matrix is one in which, within rows, the sum of entries from the left to the main diagonal, versus the sum of entries to the right away from the main diagonal, are maximally separated. The difference between these two sums is $G(\rho(k))$, which when squared and aggregated over all rows is the measure in (4.6) that is maximized to identify the optimal permutation $\rho^*(\cdot)$.

As a slight variation on coordinate representation, suppose an equally spaced representation for $\mathbf{P}$ is to be identified by minimizing the least-squares loss function (over all permutations, $\rho(\cdot)$)

$$\sum_{i<j}(p_{\rho(i)\rho(j)} - \alpha \mid j-i \mid)^2,$$

where α is a multiplicative parameter to be estimated. The measure in (4.6) would be replaced by

$$\sum_k kG(\rho(k)).$$

Thus, the merit increment for placing an integer k' ($\equiv \rho(k)$) (i.e., $\{k'\} = A_k - A_{k-1}$) in the k^{th} order position is $kG(\rho(k))$, and $\mathcal{F}(A_1)$ for $A_1 = \{k'\} \in \Omega_1$ defined by

$$-\sum_{j' \in S - \{k'\}} p_{k'j'}.$$

If we wish, the least-squares estimate for the multiplicative constant α can be obtained as

$$\hat{\alpha} = \sum_{i<j} p_{\rho^*(i)\rho^*(j)} \mid j-i \mid /[n^2(n+1)(n-1)/12],$$

where $\rho^*(\cdot)$ is an optimal row/column reordering of $\mathbf{P}$.

Numerical illustrations. To give several examples of sequencing an object set along a continuum based on a symmetric proximity matrix and the various measures of matrix pattern just described, we first derive two separate symmetric proximity matrices from the data given in Table 1.2 on the rated seriousness of thirteen offenses. In particular, for both the 'before' and 'after' proximity matrices, the entry defined for each pair of offenses is the absolute value of the difference in the proportions of rating one offense more serious than the other. (For example, because the proportion judging a bootlegger more serious than a bankrobber is .29 before the movie was shown, a symmetric proximity of .42 = $|.29 - .71|$ is given for the pair (bootlegger, bankrobber) in the corresponding proximity matrix.) The two proximity matrices so constructed are given in the upper- and lower-triangular portions of Table 4.1 (where, for graphical convenience, the rows and columns have been already reordered according to the optimal sequencings to be described immediately below).

We first give the optimal orderings associated with an explicit coordinate representation and the measure of matrix pattern given in (4.6).

Optimal orderings based on coordinate estimation, obtained using the program DPSE1U:

Before viewing the movie:

order:	9	7	10	4	2	11	3	5	13	6	12	8	1
coordinates:	−.82	−.78	−.33	−.26	−.23	−.17	−.02	.27	.29	.32	.50	.59	.64

Table 4.1: *Symmetric proximity matrices constructed for thirteen offenses using the absolute values of the skew-symmetric proximities generated from the entries in Table 1.2. The above-diagonal entries are before showing the film* Street of Chance*; those below the diagonal are after viewing the motion picture. The entities followed by an asterisk were negative before taking the absolute value.*

offense	9	7	10	4	2	11	3	5	13	6	12	8	1
9:tramp	x	.16	.82	.94	.90	.94	.96	.96	1.0	.98	.98	.97	1.0
7:beggar	.28	x	.72	.98	.86	.92	.96	.96	.98	1.0	1.0	.92	.98
10:speeder	.74	.58	x	.26	.18	.16	.22	.76	.84	.82	.80	.84	.88
4:drunkard	.92	.88	.34	x	.04*	.24	.50	.62	.74	.90	.82	.84	.90
11:petty thief	.90	.88	.28	.06	x	.02*	.42	.52	.62	.84	.80	.84	.86
3:pickpocket	.96	.94	.36	.40	.24	x	.16	.56	.56	.48	.96	.84	.96
2:gambler	.92	.90	.46	.40	.28	.02	x	.34	.36	.50	.74	.72	.84
6:bootlegger	.96	.96	.80	.74	.52	.40	.38	x	.10	.02*	.28	.40	.46
13:smuggler	.96	.96	.78	.68	.52	.40	.38	.02	x	.00	.46	.38	.58
5:quack doctor	.98	.96	.78	.68	.62	.44	.28	.00	.08*	x	.36	.58	.42
12:kidnapper	.98	.96	.84	.84	.94	.68	.46	.40	.28	.30	x	.28	.46*
8:gangster	.98	.98	.88	.78	.78	.74	.64	.46	.32	.36	.24	x	.00
1:bankrobber	1.0	.96	.88	.90	.94	.86	.58	.40	.46	.34	.24*	.00	x

The optimal value for the index in (4.6) is 471.16, with a residual sum-of-squares for the original least-squares task of 3.307 (we might also note that the correlation between the original proximities and the reconstructed absolute values of the coordinate differences is .864. However, this correlation must be considered a conservative association measure, because the optimization process itself did not include the estimation of an additive constant for the proximities, and thus, the correlation was not explicitly optimized).

After viewing the movie:

order:	9	7	10	4	11	3	2	6	13	5	12	8	1
coordinates:	$-.81$	$-.75$	$-.39$	$-.26$	$-.21$	$-.05$	.02	.27	.27	.29	.48	.55	.58

Here, the optimal index in (4.6) is 436.71 with a residual sum-of-squares of 2.302 for the original least-squares task (the conservative association measure of .903 was observed for the correlation between the original proximities and the reconstructed absolute values of the coordinate differences).

Both the before and after orderings of the offenses are obviously arranged from least to worst in seriousness, with a clear difference in the position of 'gambler' (which is #2) from the 5^{th} least serious to the 7^{th}, reflecting a change in subjective assessment after viewing the movie. The order of the offenses numbered 5 (quack doctor), 6 (bootlegger), and 13 (smuggler) also vary among themselves, but as is apparent in the estimated coordinates both before and after, these three offenses are very close to one another and any interchange from before to after is most likely attributable to minor fluctuations in the data. The same object orders are also optimal for an equally spaced set of coordinates. For both the before and after proximity matrices, the multiplicative constant α was estimated as .12, with, respectively, residual sums of squares of 4.735 and 3.624 and correlations of .759 and .801 obtained between the absolute values of the equally spaced coordinate differences and the proximities.

We give below a brief summary of the optimal orders achieved for the other measures of matrix pattern, i.e., for the weighted or unweighted gradients within rows (or columns) alone or within both rows and columns. In each case the optimal index value is given and separated into the constituent positive and negative contributions, depending on the nonviolations and violations for the achieved optimal order. Also, as given by the terms in (4.4) and (4.5) and defined by a ratio, a descriptive index is provided in each case for how well the gradient condition is satisfied.

Optimal orderings based on other measures of matrix pattern, obtained using DPSE1U:

Before viewing the movie:

Criterion	Index	Order
within row (or column) unweighted gradient	236 = (256−20) (ratio = .855)	9 7 4 10 2 11 3 13 6 5 12 8 1
within row and column unweighted gradient	431 = (489−58) (ratio = .788)	9 7 4 10 2 11 3 5 13 6 12 8 1
within row (or column) weighted gradient	95.48 = (97.56−2.08) (ratio = .958)	9 7 10 4 2 11 3 13 5 6 12 8 1
within row and column weighted gradient	161.45 = (165.66−4.21) (ratio = .950)	9 7 10 4 2 11 3 5 13 6 12 8 1

After viewing the movie:

Criterion	Index	Order
within row (or column) unweighted gradient	263 = (272−9) (ratio = .936)	9 7 10 4 11 3 2 6 5 13 12 8 1
within row and column unweighted gradient	491 = (515−24) (ratio = .911)	9 7 10 4 11 3 2 13 6 5 12 8 1
within row (or column) weighted gradient	104.44 = (105.00−.56) (ratio = .989)	9 7 10 4 11 3 2 6 13 5 12 8 1
within row and column weighted gradient	165.08 = (166.56−1.48) (ratio = .982)	9 7 10 4 11 3 2 6 13 5 12 8 1

These latter results are very consistent with those generated from the coordinate representations — the offense ‘gambler’ is 5^{th} least serious in all orders for the before matrix and 7^{th} for the after matrix; there is also some of the same unsystematic interchange among the ‘close’ offenses 5, 6, and 13, and two instances of an interchange of the offenses 4 (drunkard) and 10 (speeder) when using the unweighted gradient measure for the before matrix. We might also observe that for the optimal indices and their descriptive ratios calculated using (4.4) and (4.5), the after matrix is apparently structured slightly better than the before matrix (this difference was also reflected in the lower residual sums of squares and higher correlations provided in the context of coordinate estimation).

4.1.2 Skew-Symmetric One-Mode Proximity Matrices

As noted earlier, any skew-symmetric matrix $\mathbf{P}^{SS} = \{p_{ij}^{SS}\}$ contains two distinct types of information about the relationship between any pair of objects. First, for each object pair, $O_i, O_j \in S$, $\mid p_{ij}^{SS} \mid$ indicates a (symmetric) degree of dissimilarity between O_i and O_j, whereas $\text{sign}(p_{ij}^{SS})$ indicates the directionality of dominance. Thus, if p_{ij}^{SS} is positive, O_i can be interpreted as dominating O_j with the magnitude of p_{ij}^{SS} reflecting the degree of dominance. In the presence of a skew-symmetric matrix $\mathbf{P}^{SS}$, a natural extension of the notion of an anti-Robinson form appropriate for a symmetric proximity matrix $\mathbf{P}$ would be to a row/column reordering of $\mathbf{P}^{SS}$, say, by a permutation $\rho(\cdot)$, in which (a) the degree of dominance is perfectly depicted by having an anti-Robinson form in the reordered matrix $\{\mid p^{SS}_{\rho(i)\rho(j)} \mid\}$, and (b) using the *same* row/column reordering, the direction of dominance is perfectly depicted in the reordered matrix $\{\text{sign}(p_{\rho(i)\rho(j)})\}$ in that all above-diagonal entries are nonnegative and all below-diagonal entries are (therefore) nonpositive.

The various classes of measures appropriate for indexing matrix patterning and for obtaining an optimal reordering of $\mathbf{P}^{SS}$ could rely on either $\{\mid p^{SS}_{ij} \mid\}$ or $\{\text{sign}(p^{SS}_{ij})\}$ alone or, instead, attempt to use both types of information jointly by considering the matrix $\mathbf{P}^{SS}$ as is. (Thus, there is the possibility of assessing the extent to which the same [or similar] optimal reorderings arise from the use of the separate information sources.) If $\{\mid p^{SS}_{ij} \mid\}$ were considered by itself, the measures described in the previous section for symmetric proximities would obviously be appropriate; in fact, the numerical examples given in the last section for sequencing a symmetric one-mode proximity matrix relied on dissimilarity matrices of this latter form constructed from the nonsymmetric data of Table 1.2. The more distinctive class of alternatives to be considered below is best viewed as emphasizing $\{\text{sign}(p^{SS}_{ij})\}$ and forcing the above-diagonal entries in an optimally reordered matrix to be as consistently positive as possible.

More pointedly, we will *not* discuss at this juncture the optimization through row/column reorderings of weighted or unweighted gradient measures for the matrix $\mathbf{P}^{SS}$ that would attempt to force an anti-Robinson form for the above-diagonal entries in the reordered matrix and which would parallel those developed for symmetric matrices. The authors' experience has found that these *may* work well, but anomalies can arise frequently. For example, a row/column reordering of $\mathbf{P}^{SS}$ may exist in which the reordered matrix $\{\mid p^{SS}_{\rho(i)\rho(j)} \mid\}$ is very close to being anti-Robinson in form, and for this same reordering the above-diagonal entries in $\{p^{SS}_{\rho(i)\rho(j)}\}$ tend to be positive, but an optimal reordering of the matrix $\mathbf{P}^{SS}$ based on the above-diagonal gradient conditions analogous to those used for symmetric matrices finds an even better reordering, where the above-diagonal entries may no longer be predominantly positive. Although formally better according to the particular gradient index chosen, the latter reordering is also hard to explain in any clear substantive fashion. Therefore, the use of gradient conditions for the matrix $\mathbf{P}^{SS}$, as is, may be best avoided when we might want to identify an anti-Robinson form for the mostly positive above-diagonal entries in a reordered matrix. We note, in a discussion at the end of this section, certain kinds of gradient conditions we might want to identify in a reordered matrix $\{p^{SS}_{\rho(i)\rho(j)}\}$, but in an unfolding context where the original nonsymmetric matrix (before generating the skew-symmetric proximity matrix $\mathbf{P}^{SS}$) is used. Here, both subjects and objects are assumed orderable along the same continuum and subjects generally make their preference judgments as a function of their own (estimated) distances to the objects.

Given $\{p_{ij}^{SS}\}$ (or possibly just $\{\text{sign}(p_{ij}^{SS})\}$), an obvious class of measures of a matrix pattern would be the sum of the above-diagonal entries

$$\sum_{i<j} p_{ij}^{SS} \qquad \left(\text{or} \sum_{i<j} \text{sign}\left(p_{ij}^{SS}\right)\right).$$

In fact, because $p_{ij}^{SS} = \text{sign}(p_{ij}^{SS}) \mid p_{ij}^{SS} \mid$, the index $\sum_{i<j} p_{ij}^{SS}$ can be interpreted merely as a weighted version of the one based just on $\text{sign}(p_{ij}^{SS})$. We will assume in our discussion below that $\sum_{i<j} p_{ij}^{SS}$ is being considered, but the obvious replacement of p_{ij}^{SS} by $\text{sign}(p_{ij}^{SS})$ could incorporate the use of $\sum_{i<j} \text{sign}(p_{ij}^{SS})$ directly.

To carry out the GDPP based on $\sum_{i<j} p_{ij}^{SS}$, the incremental contribution to the total merit measure of patterning generated by placing the single integer, say, k' ($\equiv \rho(k)$) in $A_k - A_{k-1}$, at the k^{th} order position can be defined as

$$J(\rho(k)) \equiv \sum_{j' \in A_{k-1}} p_{j'k'}^{SS}. \tag{4.7}$$

Letting $\mathcal{F}(A_1) = 0$ for $A_1 \in \Omega_1$, the recursion in (2.3) can be executed to identify an optimal row/column reordering of $\mathbf{P}^{SS}$ that will maximize the sum of the above-diagonal entries over all row/column reorderings of $\mathbf{P}^{SS}$.

The optimization task of reordering the rows/columns of a matrix to maximize the sum of above-diagonal entries and its solution by the type of DP strategy sketched above was first introduced by Lawler (1964) to identify minimum feedback arc sets in a directed graph. The optimization task itself, however, has several other distinct substantive incarnations, e.g., in maximum likelihood paired comparison ranking (Flueck and Korsh (1974)) or in triangulating an input-output matrix (Korte and Oberhofer (1971)). For an extensive review of the variety of possible applications up to the middle 1970's, the reader is referred to Hubert (1976). For more up-to-date surveys and current work, see Charon, Hudry, and Woirgard (1996), Barthélemy et al. (1995), and Charon et al. (1997).

Measures of matrix pattern for a skew-symmetric matrix $\mathbf{P}^{SS}$ that might be derivable indirectly from an attempt to generate some type of coordinate representation have a rather different status than they had for a symmetric matrix $\mathbf{P}$. In the skew-symmetric framework, closed-form least-squares solutions are possible, thus eliminating the need for any GDPP optimization strategy. For example, suppose we wish to find a set of n coordinate values $x_1, \ldots, x_n$ such that the least-squares criterion,

$$\sum_{i<j} (p_{ij}^{SS} - (x_j - x_i))^2,$$

is minimized. An optimal set of coordinates can be obtained analytically (e.g., see Hubert and Arabie (1986)) by letting x_j be the average proximity within column j, i.e., $x_j = (1/n)\sum_i p_{ij}^{SS}$, with a minimum least-squares loss value of $\sum_{i<j}(p_{ij}^{SS})^2 - n\sum_j(\sum_i p_{ij}^{SS}/n)^2$. Thus, an optimal row/column reordering of $\mathbf{P}^{SS}$ can be obtained merely by using the order of the optimal coordinates from smallest (most negative) to largest (most positive). Similarly, if we consider an equally spaced coordinate representation obtained by minimizing the least-squares loss function

$$\sum_{i<j}(p_{ij}^{SS} - \alpha(x_j - x_i))^2,$$

where $x_1, \ldots, x_n$ are the integers $1, \ldots, n$ in some order and α is some multiplicative constant to be estimated, the optimal row/column reordering of $\mathbf{P}^{SS}$ induced by the integer coordinates would again be generated by the ordering of $(1/n)\sum_i p_{ij}^{SS}$ for $1 \leq j \leq n$. In the numerical examples given below, these latter analytic solutions are illustrated in addition to those measures described earlier defined by an above-diagonal sum. Because no analytic solution is possible when using these latter measures, a GDPP application is required for their optimization.

Numerical illustration. To provide a few examples of sequencing an object set along a continuum based on a skew-symmetric proximity matrix, we reconsider the data of Table 1.2 and first form two skew-symmetric matrices (one associated with before showing the motion picture and one after) based on the signed differences between the proportions of rating one offense as more serious than the other. The absolute values of these differences were given in Table 4.1 and used to illustrate the sequencing of object sets based on symmetric proximity information. We merely indicate by asterisks beside the relevant entries in Table 4.1 which differences were negative before the absolute values were taken; thus, with this annotation, both the before and after skew-symmetric matrices can be considered displayed in Table 4.1 as well (although using the object reorderings obtained from coordinate representations derived from their absolute values).

The optimal orderings based on maximizing the sum of above-diagonal entries for the two skew-symmetric proximity matrices are as follows (where the offense of being a gambler, #2, is bracketed for emphasis):

before: 9 7 10 [2] 4 11 3 6 5 13 1 12 8;

after: 9 7 10 4 11 3 [2] 6 5 13 1 12 8.

The before order given above produces a reordered skew-symmetric matrix in which all the entries above the main diagonal are nonnegative except the (2,11) pair, which has a value of $-.02$; the after order produces a reordered skew-symmetric matrix where all entries above the main diagonal are nonnegative without exception. In comparison with the orderings based on the symmetric proximities of Table 4.1, there are a few minor local interchanges among (2,4), (5,6,13), and (1,8,12) that force a little more nonnegativity above the main diagonals (in comparison with the orderings used in presenting Table 4.1), but the results are otherwise very consistent. Based on the optimal orderings for the skew-symmetric information, there is an obvious change in the position of 'gambler' (i.e., offense #2) from the 4^{th} least serious before the showing of the movie to the 7^{th} thereafter. We also note that these same two before and after optimal reorderings (in the present case) are also optimal if only the sign information from the original skew-symmetric proximity matrices is considered.

For completeness we give below the object sequences obtained from closed-form coordinate representations based on the before and after skew-symmetric proximity matrices along with some associated descriptive information. Again, there is an expected consistency with respect to the orderings obtained by maximizing the sum of above-diagonal entries with some minor local interchange differences among (2,4), (5,6,13), and (1,8,12).

Optimal orderings based on coordinate estimation for the skew-symmetric proximity matrices, obtained using DPSE1U; the residual sum-of-squares and the correlations between the skew-symmetric proximities and the signed coordinate differences are also provided:

before:	9	7	10	4	2	11	3	5	13	6	12	1	8;
coordinates:	-.82	-.78	-.33	-.26	-.23	-.18	-.02	.27	.29	.32	.57	.57	.59;

before: equally spaced coordinates ($\hat{\alpha} = .12$): 5.182; correlation = .739;
estimated coordinates: 3.457; correlation = .852;

after:	9	7	10	4	11	3	2	6	5	13	12	1	8;
coordinates:	-.81	-.75	-.39	-.26	-.21	-.05	.02	.27	.28	.29	.51	.55	.55.

after: equally spaced coordinates ($\hat{\alpha} = .12$): 3.853; correlation = .788;
estimated coordinates: 2.369; correlation = .896.

The Greenberg Form:

As noted earlier in this section, there is at least one specific context in which it is of interest to identify whether the matrix $\mathbf{P}^{SS}$, as is, may be row/column reorderable to satisfy a particular set of gradient conditions within its rows and columns. Explicitly, suppose the set S of n objects is assumed orderable along a continuum and, in addition, a group of judges can be placed along the same continuum so that when asked which of two presented objects they would prefer, the response is made according to the closer distance between the judge's location and the location of the two objects presented. If the skew-symmetric matrix $\mathbf{P}^{SS} = \{p_{ij}^{SS}\}$ is constructed by letting p_{ij}^{SS} = the proportion of judges preferring object O_i to O_j − the proportion of judges preferring object O_j to

O_i, then when $\mathbf{P}^{SS}$ is row/column reordered according to the placement of the objects along the continuum, say, as $\{p^{SS}_{\rho(i)\rho(j)}\}$, the specific set of gradient conditions we give below will be satisfied (these were first observed and discussed by Greenberg (1965) and will be referred to as defining a Greenberg form for the matrix $\{p^{SS}_{\rho(i)\rho(j)}\}$):

within rows: $p^{SS}_{\rho(i)\rho(k)} \leq p^{SS}_{\rho(i)\rho(j)}$ for $1 \leq i < k < j \leq n$;
within columns: $p^{SS}_{\rho(k)\rho(j)} \geq p^{SS}_{\rho(i)\rho(j)}$ for $1 \leq i < k < j \leq n$.

In words, the reordered matrix $\mathbf{P}^{SS}_{\rho} = \{p^{SS}_{\rho(i)\rho(j)}\}$ has a Greenberg form if the entries within a row moving to the right from the main diagonal never decrease, and if the entries within a column moving up from the main diagonal never increase. (In comparison to the gradient conditions for an anti-Robinson form for a symmetric matrix $\mathbf{P}$, the obvious difference here is in the reversal of the within-column inequality when identifying a Greenberg form for a skew-symmetric matrix $\mathbf{P}^{SS}$.)

The natural merit measures of how well a particular reordered skew-symmetric proximity matrix satisfies the Greenberg gradient conditions, which could be used in the GDPP to order the objects along the continuum, would be the analogue of (4.1),

$$\sum_{i<k<j} f(p^{SS}_{\rho(i)\rho(k)}, p^{SS}_{\rho(i)\rho(j)}) + \sum_{i<k<j} f(p^{SS}_{\rho(i)\rho(j)}, p^{SS}_{\rho(k)\rho(j)}),$$

where, in contrast to (4.1), the second within-column comparison term, $\sum_{i<k<j} f(p^{SS}_{\rho(i)\rho(j)}, p^{SS}_{\rho(k)\rho(j)})$, has an interchange of its two components.[32]

Numerical illustrations. To give an example of the use of the Greenberg gradient conditions in optimally ordering a set of objects along a continuum, we use a data set reported in Orth (1989) on the preferences for five German political parties collected in 1980. The five parties are ordered below according to their position on a political left-right dimension:

DKP: Deutsche Kommunistische Partei (German Communist Party),
SPD: Sozialdemokratische Partei Deutschlands (Social Democratic Party),
FDP: Freie Demokratische Partei (Free Democratic Party),
CDU/CSU: Christliche Demokratische Union/Christliche Soziale Union (Christian Democratic Union/Christian Social Union),
NPD: Nationaldemokratische Partei Deutschlands (National Democratic Party).

Based on the complete preference rankings for 1316 German voters, Table 4.2 provides a 5×5 skew-symmetric matrix for the five political parties derived from the choice proportions.

The optimal orderings for the matrix of Table 4.2 for both the weighted and unweighted gradient measures are given below, along with their indices of merit

Table 4.2: *A skew-symmetric proximity matrix among five German political parties, obtained from the complete preference rankings for 1316 German voters.*

party	DKP	SPD	FDP	CDU/CSU	NPD
DKP	x	-.982	-.968	-.862	-.015
SPD	.982	x	.181	.131	.933
FDP	.968	-.181	x	.114	.964
CDU/CSU	.862	-.131	-.114	x	.945
NPD	.015	-.933	-.964	-.945	x

and descriptive ratios defined by the direct extension of equation (4.4) to the Greenberg gradient conditions. In addition, the indices of merit are given as if the parties were merely ordered, as discussed by Orth (1989), according to their position on the political left-right dimension.

unweighted gradient:

optimal order: DKP $\rightarrow$ SPD $\rightarrow$ CDU/CSU $\rightarrow$ FDP $\rightarrow$ NPD
index = 16 = (18−2)
ratio = .800 = (18−2)/(18+2)

left-right order: DKP $\rightarrow$ SPD $\rightarrow$ FDP $\rightarrow$ CDU/CSU $\rightarrow$ NPD
index = 14 = (17−3)
ratio = .700 = (17−3)/(17+3)

weighted gradient:

optimal order: DKP $\rightarrow$ FDP $\rightarrow$ SPD $\rightarrow$ CDU/CSU $\rightarrow$ NPD
index = 11.707 = (11.771−.064)
ratio = .989 = (11.771−.064)/(11.771+.064)

left-right order: DKP $\rightarrow$ SPD $\rightarrow$ FDP $\rightarrow$ CDU/CSU $\rightarrow$ NPD
index = 11.683 = (11.769−.086)
ratio = .987 = (11.769−.086)/(11.769+.086)

Obviously, for either the weighted or unweighted gradient measures, the left-right order is *not* optimal and better orderings can be identified. This observation casts doubt on the reasonableness of the initial assumption that voters and political parties are jointly orderable along a common left-right continuum and that voters would prefer those parties that are closer to their own locations. The two extreme parties, DKP and NPD, are placed appropriately at the ends of the continuum in the optimal orderings given above, but there is difficulty with the three intermediates, SPD, FDP, and CDU/CSU. Orth (1989) addresses this point explicitly and notes that because coalition governments involving SPD, FDP, and CDU/CSU have been the norm for the last several decades, consistent preferences based on a clear-cut left-right continuum are probably inherently difficult to observe for these intermediate parties.

Besides attempting to obtain an ordering of the parties along a continuum that would induce an approximate Greenberg form within the rows and columns of the reordered matrix, it is also possible to reorder the matrix of Table 4.2 using the earlier criterion of maximizing the sum of above-diagonal entries. Doing so is an attempt to define a dominance ordering among the political parties. An optimal reordering of Table 4.2 maximizing the above-diagonal sum is SPD $\rightarrow$ FDP $\rightarrow$ CDU/CSU $\rightarrow$ NPD $\rightarrow$ DKP (the maximum sum is .610, and all above-diagonal entries are nonnegative).[33] This specific ordering can be interpreted as a consensus (or societal) ordering for the political parties over the set of judges.

We might note in closing this illustration that there is an extensive literature on the use of binary choice proportions in obtaining a transitive societal ordering for a set of objects (that form the items of choice) by maximizing the above-diagonal sum in a reordered proximity matrix. The reader is referred to Bowman and Colantoni (1973; 1974) or the review by Hubert (1976) for many more details.

4.1.3 Two-Mode Proximity Matrices

As noted in the introduction to Chapter 4, if the available proximity data are between two distinct sets, S_A and S_B, and in the form of an $n_A \times n_B$ nonnegative dissimilarity matrix $\mathbf{Q}$, a symmetric proximity matrix $\mathbf{P}^{AB}$ can be constructed for the single object set $S = S_A \cup S_B$ ($n = n_A + n_B$) that will have missing entries for object pairs within S_A and within S_B. An optimal joint sequencing of $S_A \cup S_B$ along a continuum can then be attempted using some measure of matrix patterning defined for $\mathbf{P}^{AB}$, with the obvious candidates being the same as those considered in Section 4.1.1. Either the weighted or the unweighted gradient measures of (4.1) based on the function $f(\cdot,\cdot)$ extend directly to a use with $\mathbf{P}^{AB}$ merely by defining $f(z,y)$ to be identically 0 if either z or y refers to a pairwise proximity within S_A or within S_B (and thus, either z or y refers to a missing proximity); applications of these measures will be given in the numerical examples. In contrast, however, the obvious analogues for the measures of matrix patterning based on coordinate representation fail to provide a way of constructing merit increments that are independent of the order of the objects previously placed (as was briefly noted in the introduction to this monograph), and thus cannot be directly implemented within a GDPP framework. We will return to this topic at the end of this section and give a more detailed explanation of why this difficulty arises.

Analogous to the case of complete symmetric proximity matrices $\mathbf{P}$, the use of the weighted or unweighted gradient measure to reorder $\mathbf{P}^{AB}$ optimally can be interpreted as an attempt to find a row/column reordering as close as possible to an anti-Robinson form in its nonmissing entries. In turn, any reordering of $\mathbf{P}^{AB}$ must also induce separate row and column reorderings for the original $n_A \times n_B$ proximity matrix $\mathbf{Q}$. If $\mathbf{P}^{AB}$ can be placed in a perfect anti-Robinson form (as, for example, when a one-dimensional Euclidean representation exists for the objects in $S = S_A \cup S_B$, or a perfect ultrametric defines the nonmissing entries in $\mathbf{P}^{AB}$), the reordered matrix $\mathbf{Q}$ would also display a perfect order pattern for its entries within each row and within each column. Explicitly, the entries within each row (or column) would be nonincreasing to a minimum and nondecreasing thereafter. Such a pattern can be called a Q-form (within rows and within columns) in the reordered matrix $\mathbf{Q}$ (as named by Kendall (1971a; 1971b)), and has a very long history in the literature of unidimensional unfolding (e.g., see Coombs (1964), Chapter 4; Hubert and Arabie (1995a)).

In the application of the weighted or unweighted gradient conditions for reordering a symmetric matrix $\mathbf{P}$, one possible variation discussed was to allow the gradient measures (defined above the main diagonal) to be optimized only within rows (or equivalently, only within columns). For two-mode data and the derived matrix $\mathbf{P}^{AB}$, a different restriction on gradient comparisons may also be useful. Specifically, suppose that the gradient conditions are used to define an optimal reordering of $\mathbf{P}^{AB}$ within both the rows and columns of $\mathbf{P}^{AB}$ but only for those in which the common object on which a gradient comparison is made is a member of S_B (containing those objects forming the columns of $\mathbf{Q}$). This approach can be carried out directly in the use of the gradient measures in (4.1) merely by defining $f(z, y)$ to be identically zero whenever the common object associated with the two proximities z and y is not a member of S_B. If this restricted measure is adopted as the means to reorder $\mathbf{P}^{AB}$ optimally, and if there are no violations of these gradient conditions in the optimally reordered matrix $\mathbf{P}^{AB}$, then the induced reordering of the $n_A \times n_B$ matrix $\mathbf{Q}$ would have a perfect Q-form within each column (but not necessarily within each row). More generally, the attempt to reorder $\mathbf{P}^{AB}$ optimally based on this restricted measure can be interpreted as a mechanism for reordering the rows of $\mathbf{Q}$ to approximate a Q-form within columns (alone). Analogously, if the function $f(z, y)$ is defined to be identically zero when the common object associated with the proximities z and y is not a member of S_A, then an optimal reordering of $\mathbf{P}^{AB}$ can be interpreted as a mechanism for reordering the columns of $\mathbf{Q}$ to approximate a Q-form within rows (alone).

In addition to these possible variations on how the weighted or unweighted gradient conditions can be evaluated in a reordered matrix, there is one additional alternative for the choice of an index for matrix patterning (although for brevity, we will not explicitly illustrate it in our numerical examples below). Instead of maximizing the weighted or unweighted gradient conditions as they have been defined, we can also allow the minimization of just the violations (either weighted or unweighted). For some data sets, slightly different optimal joint orderings of S_A and S_B might arise when emphasizing the latter measure because it relies only on the reduction of gradient violations and not on the maximization of a difference between the gradient nonviolations and violations.[34]

Numerical illustrations. As a numerical example of optimally reordering a two-mode proximity matrix $\mathbf{Q}$, we will use the data of Table 1.3 on the dissimilarities between the goldfish retinal receptors (the eleven rows of $\mathbf{Q}$) and the specific wavelengths of light (the nine columns of $\mathbf{Q}$). The joint optimal sequencings of the rows/columns of the derived matrix $\mathbf{P}^{AB}$ are given below for the weighted and unweighted gradient measures, along with the relevant descriptive information (the column objects are shown by an underline in each joint ordering).

Optimal orderings for the combined row/column object set based on the gradient measures, obtained using DPSE2U:

within row and column unweighted gradient:
joint order: $\underline{3}$ 10 11 5 7 $\underline{8}$ $\underline{2}$ 1 4 6 2 $\underline{7}$ $\underline{1}$ $\underline{5}$ $\underline{6}$ 8 9 3 $\underline{4}$ $\underline{9}$
row order: 10 11 5 7 1 4 6 2 8 9 3
column order: $\underline{3}$ $\underline{8}$ $\underline{2}$ $\underline{7}$ $\underline{1}$ $\underline{5}$ $\underline{6}$ $\underline{4}$ $\underline{9}$
index = 464 = (517−53); ratio = .814 = (517−53)/(517+53)

within row and column weighted gradient:
joint order: $\underline{3}$ $\underline{8}$ 10 11 $\underline{2}$ 5 7 1 $\underline{7}$ $\underline{1}$ 4 6 2 $\underline{5}$ $\underline{6}$ $\underline{9}$ 9 $\underline{4}$ 8 3
row order: 10 11 5 7 1 4 6 2 9 8 3
column order: $\underline{3}$ $\underline{8}$ $\underline{2}$ $\underline{7}$ $\underline{1}$ $\underline{5}$ $\underline{6}$ $\underline{9}$ $\underline{4}$
index = 31025 = (32019−994); ratio = .940 = (32019−994)/(32019+994)

Although there is some variation in how the row and column objects are combined for the weighted and unweighted gradient measures in the joint orders given above, the separate row and column reorderings are very consistent, with one adjacent receptor interchange for 8 and 9 and one adjacent wavelength interchange for $\underline{4}$ and $\underline{9}$. The column ordering of the stimuli for the unweighted gradient measure (i.e., $\underline{3}$ $\underline{8}$ $\underline{2}$ $\underline{7}$ $\underline{1}$ $\underline{5}$ $\underline{6}$ $\underline{4}$ $\underline{9}$) is exactly consistent with decreasing wavelength (for the weighted gradient measure, the orders of the two lowest wavelength stimuli of $\underline{4}$ and $\underline{9}$ are reversed but otherwise the order is the same as for the unweighted gradient measure).

As noted in our earlier discussion, we can allow some variation in how the gradient measures are obtained, particularly when allowing the comparisons to be defined only within the row or only within the column objects of the original two-mode matrix $\mathbf{Q}$. For example, if gradient comparisons are limited to within the rows (the receptors) of $\mathbf{Q}$ to emphasize an approximate Q-form over columns (the light stimuli), the optimal orderings of the eight wavelength stimuli for both the unweighted and weighted gradient measures turn out to be identical to those obtained for the unweighted gradient measure within both rows and columns (i.e., $\underline{3}$ $\underline{8}$ $\underline{2}$ $\underline{7}$ $\underline{1}$ $\underline{5}$ $\underline{6}$ $\underline{4}$ $\underline{9}$) but for obviously different optimal index values (i.e., within-row unweighted gradient: index = 264 = (289−25); within-row weighted gradient: index = 19293 = (19926−633)).

In closing this section, we will illustrate the difficulties encountered in a two-mode context when we attempt to derive a measure of matrix pattern based on a unidimensional coordinate representation that could then be applied in the GDPP framework (which would be intended to parallel the measure in (4.6) for a one-mode symmetric proximity matrix $\mathbf{P}$). The two-mode coordinate representation task can be phrased as follows: given the $n \times n$ matrix $\mathbf{P}^{AB}$ =

$\{p_{ij}^{AB}\}$ (where for computational convenience the asterisk entries denoting the missing values can be considered replaced by zeros) and an $n \times n$ indicator matrix $\mathbf{W} = \{w_{ij}\}$, where $w_{ij} = 0$ for $1 \leq i, j \leq n_A$; $n_A + 1 \leq i, j \leq n$, and $= 1$ otherwise, find a set of n ordered coordinate values, $x_1 \leq \cdots \leq x_n$ (such that $\sum_k x_k = 0$), minimizing the least-squares criterion

$$\sum_{i<j} w_{\rho(i)\rho(j)} \left(p_{\rho(i)\rho(j)}^{AB} - \mid x_j - x_i \mid \right)^2 .$$

(Obviously, the purpose of the indicator value $w_{\rho(i)\rho(j)}$ is to choose only squared discrepancies that involve the nonmissing proximities in $\mathbf{P}^{AB}$.) Relying on the extensive work of Heiser (1981), Chapter 6, a solution to the (necessary) stationary equations derived for the least-squares loss function would produce a set of coordinates $x_1 \leq \cdots \leq x_n$ (where $\sum_k x_k = 0$) and an associated permutation $\rho(\cdot)$ of the n objects in S $(= S_A \cup S_B)$, such that if we let

$$K(\rho(k)) = \sum_{i=1}^{k-1} w_{\rho(k)\rho(i)} p_{\rho(k)\rho(i)} - \sum_{i=k+1}^{n} w_{\rho(k)\rho(i)} p_{\rho(k)\rho(i)},$$

and if $\rho(k)$ is a row object in S_A, then

$$x_k = \frac{1}{n_B} \left[K(\rho(k)) - \frac{1}{n_A + n_B} \sum_{\rho(j) \in S_A} K(\rho(k)) \right];$$

also, if $\rho(k)$ is a column object in S_B, then

$$x_k = \frac{1}{n_A} \left[K(\rho(k)) - \frac{1}{n_A + n_B} \sum_{\rho(j) \in S_B} K(\rho(k)) \right].$$

Now, the least-squares loss function for any permutation $\rho(\cdot)$ can be rewritten as

$$\sum_{i<j} (p_{ij}^{AB})^2 \; + n_B \sum_{\rho(k) \in S_A} \left[x_k - \frac{1}{n_B} K(\rho(k)) + \frac{1}{n_A + n_B} \sum_{\rho(j) \in S_A} K(\rho(j)) \right]^2$$
$$+ \, n_A \sum_{\rho(k) \in S_B} \left[x_k - \frac{1}{n_A} K(\rho(k)) + \frac{1}{n_A + n_B} \sum_{\rho(j) \in S_B} K(\rho(j)) \right]^2$$
$$-2 \left(\sum_{\rho(k) \in S_A} x_k - \sum_{\rho(k) \in S_A} \left[\frac{1}{n_B} K(\rho(k)) - \frac{1}{n_A + n_B} \sum_{\rho(j) \in S_A} K(\rho(j)) \right] \right)$$
$$\times \left(\sum_{\rho(k) \in S_B} x_k - \sum_{\rho(k) \in S_B} \left[\frac{1}{n_A} K(\rho(k)) - \frac{1}{n_A + n_B} \sum_{\rho(j) \in S_B} K(\rho(j)) \right] \right)$$

$$- \left[\frac{1}{n_B} \sum_{\rho(k) \in S_A} (K(\rho(k)))^2 + \frac{1}{n_A} \sum_{\rho(k) \in S_B} (K(\rho(k)))^2 \right.$$
$$\left. + \frac{1}{n_A n_B} \sum_{\rho(k) \in S_A} K(\rho(k)) \sum_{\rho(k) \in S_B} K(\rho(k)) \right].$$

Thus, if $\rho(\cdot)$ is derived from a stationary solution, all the nonconstant terms are zero except the last three in the equation above, i.e.,

$$\frac{1}{n_B} \sum_{\rho(k) \in S_A} (K(\rho(k)))^2 + \frac{1}{n_A} \sum_{\rho(k) \in S_B} (K(\rho(k)))^2$$
$$+ \frac{1}{n_A n_B} \sum_{\rho(k) \in S_A} K(\rho(k)) \sum_{\rho(k) \in S_B} K(\rho(k)). \tag{4.8}$$

Analogous to results from using a symmetric proximity matrix with (4.6), if one could show that maximizing (4.8) leads to a stationary solution and if (4.8) could produce additive merit increments (again, analogous to the use of (4.6)), a GDPP recursive solution could be generated. Unfortunately, it does not appear possible to develop such additive merit increments based on (4.8) because of the presence of the last term

$$\frac{1}{n_A n_B} \sum_{\rho(k) \in S_A} K(\rho(k)) \sum_{\rho(k) \in S_B} K(\rho(k)),$$

which requires knowledge of how an entity in A_{k-1} was reached (i.e., a knowledge of the values of $K(\rho(k))$ for those objects $\rho(k) \in A_{k-1}$), and the latter depends on the order of object placement in the construction of A_{k-1}. If only the first two terms in (4.8) were present, a GDPP recursive process could be carried out, but with the inclusion of the third, it appears that an obvious GDPP specialization is not possible.

4.1.4 Object Sequencing for Symmetric One-Mode Proximity Matrices Based on the Construction of Optimal Paths

The type of GDPP recursive process for the optimal sequencing of an object set introduced in Section 4.1 (and implemented in the preceding sections, 4.1.1 through 4.1.3) is based on a definition for the sets $\Omega_k, 1 \le k \le n$, characterized by all subsets containing k of the subscripts on the objects in S. There is, however, at least one other possibility for how these basic sets might be redefined and how a recursive process might then be carried out, which would allow a different measure of matrix patterning to be used in optimally reordering a symmetric one-mode proximity matrix $\mathbf{P}$. The specific variation discussed here involves the construction of a sequencing of the objects in S by identifying an optimal path between the objects (that includes each object exactly once) based on some function of the proximities between adjacently-placed objects (for example, we might minimize or maximize either the sum of such adjacent proximities or their maximum or minimum). The matrix pattern being assessed in the reordered matrix $\mathbf{P}_\rho$ would involve the magnitudes of those proximities immediately adjacent to the principal diagonal, i.e., $p_{\rho(i)\rho(i+1)}$ for $1 \leq i \leq n-1$. For example, in one manifestation of the various optimization options possible, we may wish to minimize the sum

$$\sum_{i=1}^{n-1} p_{\rho(i)\rho(i+1)}$$

over all possible reorderings of $\mathbf{P}$.

To be explicit in the required tailoring of the GDPP (and for the moment emphasizing the minimization of the sum of adjacent object proximities in constructing a path among the objects in S), a collection of sets $\Omega_1, \ldots, \Omega_n$ is defined (thus again, $K \equiv n$) so that each entity in $\Omega_k, 1 \leq k \leq n$, is now an ordered pair (A_k, j_k). Here, A_k is a k-element subset of the n subscripts on the objects in S, and j_k is one subscript in A_k (to be interpreted as the subscript for the last-placed object in a sequencing of the objects contained within A_k). The function value $\mathcal{F}((A_k, j_k))$ is the optimal contribution to the total measure of matrix patterning for the objects in A_k when they are placed in the first k positions in the (re)ordering, and when the object with subscript j_k occupies the k^{th}. A transformation is possible between $(A_{k-1}, j_{k-1}) \in \Omega_{k-1}$ and $(A_k, j_k) \in \Omega_k$ if $A_{k-1} \subset A_k$ and $A_k - A_{k-1} = \{j_k\}$ (i.e., A_{k-1} and A_k differ by the one integer j_k). The cost increment $C((A_{k-1}, j_{k-1}), (A_k, j_k))$ is simply $p_{(j_{k-1})j_k}$ for the contribution to the total measure of patterning generated by placing the object with the single integer subscript in $A_k - A_{k-1}$ at the k^{th} order position (i.e., the proximity between the adjacently-placed objects with subscripts j_{k-1} and j_k).

As usual, the validity of the recursive process requires the incremental cost index, $C((A_{k-1}, j_{k-1}), (A_k, j_k)) = p_{(j_{k-1})j_k}$, to depend only on $(A_{k-1}, j_{k-1}) \in \Omega_{k-1}$ and $(A_k, j_k) \in \Omega_k$, but in contrast to the GDPP specializations of Sections 4.1.1 to 4.1.3, we now know which was the last-placed subscript j_{k-1} in A_{k-1}; thus, cost increments can be defined using those objects with subscripts j_{k-1} and j_k as the last-placed indices in A_{k-1} and A_k, respectively. The values of $\mathcal{F}((A_1, j_1))$ can be assumed zero for all $(A_1, j_1) \in \Omega_1$, and the recursive process can be carried out from Ω_1 to Ω_n. The value defined by the minimum of $\mathcal{F}((A_n, j_n))$ over all j_n, $1 \leq j_n \leq n$, for $(A_n, j_n) \in \Omega_n$ and $A_n = \{1, 2, \ldots, n\}$ provides the optimal (minimal) value for the sum of adjacent proximities over all paths among the n objects in S. As always, an optimal row/column reordering of $\mathbf{P}$ attaining this minimal value can be identified by working backward through the recursion.

There are several variations on the optimal construction of a path that can be directly implemented through the type of GDPP specialization just described.

One is an interpretation through the use of merit increments (rather than cost increments), $M((A_{k-1}, j_{k-1}), (A_k, j_k)) = p_{(j_{k-1})j_k}$, and adopts a maximization optimization criterion. This change is immediate and offers no difficulty because the GDPP recursion given in (2.3) can be applied. Similarly, a min/max or max/min criterion can also be used merely by selecting the general recursive structure of (2.5) or (2.6), which would identify optimal paths among the objects in S that either minimize the maximum adjacent proximity or maximize the minimum adjacent proximity.[35] In contrast to the GDPP recursion of Sections 4.1.1 through 4.1.3, there are now greater storage requirements because in the construction of optimal paths, the sets $\Omega_k, 1 \leq k \leq n$, are defined by pairs $(A_k, j_k) \in \Omega_k$. In any case, this general type of recursive process defined on sets having this latter form was first described independently by Bellman (1962) and Held and Karp (1962) for what is called the traveling salesman problem. We will return to this specific topic briefly at the end of this section.

The symmetric proximity matrix $\mathbf{P}$ used in the construction of an optimal path between the objects in S has been considered arbitrary up to this point. There are, however, several substantive applications for this optimization task already suggested in the literature, which depend on specific definitions of how $\mathbf{P}$ may be constructed from other data available on the objects in S. To be explicit, and to mention briefly a few of these applications (for a more detailed review, see Hubert and Baker (1978)), suppose an $n \times p$ data matrix $\mathbf{X} = \{x_{ij}\}$ is given, where the rows of $\mathbf{X}$ refer to the objects in S, and the columns of $\mathbf{X}$ refer to p attributes measured on each of the n objects. Depending on how $\mathbf{P}$ is constructed, several data analysis applications can be given as specific exemplars:

Profile smoothing. Given an $n \times p$ data matrix $\mathbf{X}$, one visual means for displaying the information it contains is to first place the n objects along a horizontal axis and then graph the p profiles for each attribute over the n objects. Depending on the object order used along the horizontal axis (as discussed by Hartigan (1975), pp. 28-34, Späth (1980), Chapter 5, and Wegman (1990)), there may be a way of reducing the complexity of the graphical representation by minimizing the number of instances in which the profiles cross. If a proximity matrix $\mathbf{P} = \{p_{ij}\}$ is defined as

$$p_{ij} = \sum_{k<k'} g(x_{ik}, x_{ik'}, x_{jk}, x_{jk'}),$$

where

$$g(x_{ik}, x_{ik'}, x_{jk}, x_{jk'}) = \begin{cases} 1 & \text{if } (x_{ik} - x_{ik'})(x_{jk} - x_{jk'}) < 0; \\ 0 & \text{otherwise,} \end{cases}$$

then the ordering of the objects in S, minimizing the sum of proximities between adjacent objects along the path, also minimizes the number of instances in which the profiles cross.

Data array reordering. To help interpret the patterning of data present in $\mathbf{X}$, it may be of value to reorder the rows (and possibly the columns as well) of **X** so that the numerically larger elements of the array are placed as close as possible to each other. As discussed by McCormick, Schweitzer, and White (1972) (and clarified by Lenstra (1974); see Arabie and Hubert (1990) for a review), one possibility would be to define a proximity matrix $\mathbf{P} = \{p_{ij}\}$ among the n rows of **X** (now with a similarity interpretation) through a simple cross-product measure over the p column attributes, i.e.,

$$p_{ij} = \sum_k x_{ik} x_{jk}.$$

The optimal path sought among the n objects would maximize the sum of proximities between adjacent objects, and the object order thus obtained could be used to reorder the rows of **X**.

As a related suggestion for possibly reordering the rows of **X** to help interpret the pattern of information present, Kendall (1971a; 1971b) observed that if the rows of **X** could be reordered to display a perfect Q-form within columns (see Section 4.1.3), and if $\mathbf{P} = \{p_{ij}\}$ is defined as

$$p_{ij} = \sum_k \max(x_{ik}, x_{jk}),$$

then **P** can be reordered to display a perfect anti-Robinson form. Using this same reordering on the rows of **X**, a perfect Q-form within columns would be displayed. Thus, one possible strategy for attempting to find an approximate Q-form for **X** would be to identify the minimum length path using **P** and use the object order so identified to reorder the rows of **X**. (Such a method obviously depends on the result that if any proximity matrix **P** can be reordered to display a perfect anti-Robinson pattern, then the minimum-length path can be used to identify such an ordering.)

Numerical illustrations. As examples of constructing optimal paths based on a symmetric proximity matrix **P**, we again consider the before and after submatrices of Table 4.1 on the rated seriousness of thirteen offenses. The two optimal reorderings minimizing the sum of proximities between adjacently-placed objects are given graphically in Figure 4.1, where the respective optimal lengths are 2.30 and 2.34 for the before and after data. The two orders are very consistent with the results given earlier (e.g., with the explicit coordinate representation using the measure of matrix pattern in (4.6)); again, there are some differences among offenses 5, 6, and 13, which are very close to one another.

Although our discussion of constructing optimal paths based on **P** has been phrased as obtaining a sequence of adjacent objects that includes each object in S exactly once and some function of the $n - 1$ proximities between adjacently placed objects, the more traditional discussion of optimal path construction in the literature (e.g., see Lawler et al. (1985) for an extensive review in book form) is concerned with the generation of optimal circular paths in which each object is also included exactly once but the path is closed and now includes n proximities between the adjacently-placed objects. This topic is typically discussed

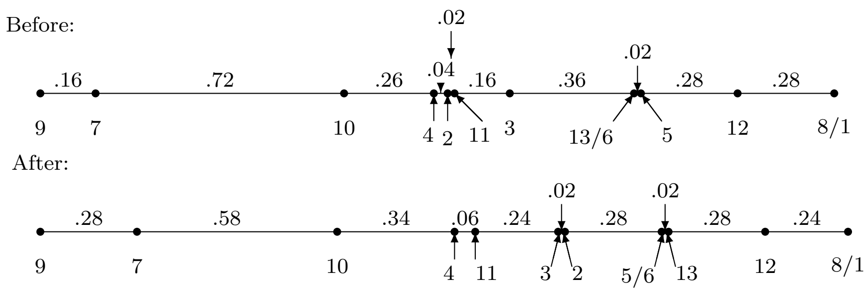

Figure 4.1: *The optimal orders for the before and after matrices of Table 4.1 minimizing the sum of proximities between adjacently-placed objects.*

under the label of the Traveling Salesman Problem, where interpretatively, a salesperson must visit each of n cities once and only once, return to the city of origin, and minimize the length of the tour. The type of GDPP recursion used for the construction of optimal linear paths can be modified easily for the construction of optimal circular paths:

Choose object O_1 as an (arbitrary) origin and force the construction of the optimal linear paths to include O_1 as the initial object by defining $\mathcal{F}((A_1, j_1)) = 0$ for $j_1 = 1$ and $A_1 = \{1\}$, and otherwise equal to a very large positive or negative value (depending on whether the task is a minimization or a maximization, respectively). The function values $\mathcal{F}((A_n, j_n))$ for all $j_n, 1 \leq j_n \leq n$, for $(A_n, j_n) \in \Omega_n$, and for $A_n = \{1, 2, \ldots, n\}$ can then be used to obtain the optimal circular paths depending on the chosen optimization criteria as follows:

minimum path length: $\min[\mathcal{F}((A_n, j_n)) + p_{j_n 1}]$;

maximum path length: $\max[\mathcal{F}((A_n, j_n)) + p_{j_n 1}]$;

minimax path length: $\min[\max(\mathcal{F}((A_n, j_n)), p_{j_n 1})]$;

maximin path length: $\max[\min(\mathcal{F}((A_n, j_n)), p_{j_n 1})]$.

Numerical illustrations (continued). As an example of how such an optimization might be carried out, Figure 4.2 represents an optimal circular path for the proximity matrix from Shepard, Kilpatric, and Cunningham (1975) on the ten digits of Table 1.1, minimizing the sum of the ten pairwise input proximities between the adjacently-placed objects (the minimal value is 3.411). Interpretatively, the multiples of 2 (2,4,8) and of 3 (3,6,9) appear at adjacent locations, with the odd numbers that are not multiples of 3 (5,7) and the identities (0,1) placed between these two groups and arranged to be as consistent as possible with digit magnitude, e.g., 5 and 6, and 7 and 8 are adjacent; and 0 and 1 are placed close to 2 and 3.

In closing this section, we make three final observations about the type of GDPP recursive process discussed above. First, although it has been explicitly assumed that the proximity matrix $\mathbf{P}$ is symmetric, if an $n \times n$ nonsymmetric

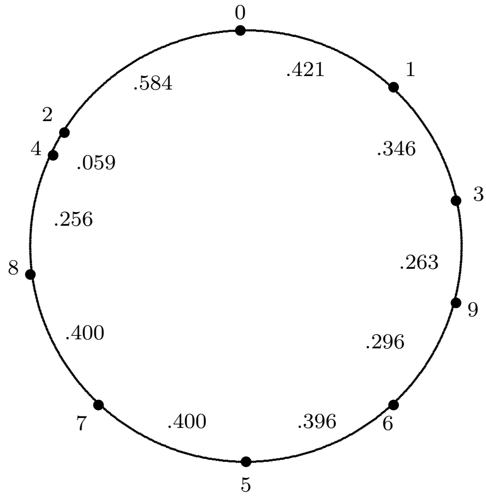

Figure 4.2: *An optimal circular ordering of the ten digits that minimizes the sum of proximities from Table 1.1 between adjacently-placed objects.*

matrix were to be used, then the same recursive process could be implemented to construct optimal *directed* linear or circular paths among the n objects in S, where each link in the path has an implied direction, i.e., p_{ij} (p_{ji}) is the proximity from object O_i *to* O_j (or from object O_j *to* O_i).[36] Second, it would be possible to redefine further the basic sets $\Omega_k, 1 \leq k \leq n$, to include additional specific information about the placement of certain objects other than just the last, e.g., we might define Ω_k to be (A_k, j_k, l_k), where j_k is the last-placed object in the set A_k and l_k denotes the second to the last. Although possible, this latter type of extension would obviously require much greater storage space; therefore, we will not explicitly pursue any of these generalizations here. Finally, although we have suggested the use of a GDPP recursive solution for the task of locating optimal paths (given the obvious emphases of this monograph), there are alternative optimization approaches that would be much better for dealing with large object sets and which could also provide optimal solutions (much as for the LA task introduced in Section 2.1). A detailed and comprehensive survey of these optimization options for the construction of optimal paths is available in the previously cited volume edited by Lawler et al. (1985).

4.2 Sequencing an Object Set Subject to Precedence Constraints

In the presentation of object sequencing in the various sections of 4.1, the search for optimality using the GDPP was carried out without any constraints on the object orderings. In all the cases discussed, however, it is straightforward to include precedence constraints that must be satisfied by the object orderings through the inclusion of an $n \times n$ indicator matrix $\mathbf{V} = \{v_{ij}\}$. The latter is defined over object pairs formed from S (where S may be $S_A \cup S_B$ for a two-mode proximity matrix), by letting $v_{ij} = 1$ if O_j must precede O_i in an (optimal) ordering, and 0 otherwise. Thus, by the device of incorporating $\mathbf{V}$ and not allowing transitions to occur between entities in Ω_{k-1} and Ω_k whenever the precedence conditions are not met, an optimal sequencing of S (or a joint sequencing of S_A and S_B) is sought that must satisfy all the constraints formalized by $\mathbf{V}$.[37]

Although precedence constraints can be incorporated directly into the use of the GDPP by the mechanism of including a precedence matrix $\mathbf{V}$, it is also possible, at least for a two-mode proximity matrix and in the joint sequencing of S_A and S_B, to use precedence constraints given by linear orderings (for example) of the row and/or column objects to reduce the storage requirements needed by an application of the GDPP. In particular, if the row and/or column objects in S_A and S_B are subject to linear order constraints, the sets $\Omega_1, \ldots, \Omega_K$ over which the recursive process is carried out can be redefined, thereby allowing larger object sets to be jointly sequenced optimally.

To be explicit, if there is an assumed linear ordering of the column objects in $S_B = \{c_1, \ldots, c_{n_B}\}$ (but none is assumed for the row objects in $S_A = \{r_1, \ldots, r_{n_A}\}$) that without loss of generality can be taken in the subscript order $c_1 \prec c_2 \prec \cdots \prec c_{n_B}$, then Ω_k for $1 \leq k \leq n$ may be defined so that each entity in Ω_k is of the form (B_k, j_k), where j_k is an integer from 0 to k indicating that the first j_k column objects have been placed, and B_k is a $k - j_k$ element subset of the row object subscripts $\{1, \ldots, n_A\}$ (which is the empty subset $\emptyset$ when $k - j_k = 0$). The function value $\mathcal{F}((B_k, j_k))$ denotes the optimal contribution to the total measure of matrix patterning when the first k positions in the reordering are occupied by the first j_k column objects and the $k - j_k$ row objects in B_k. Thus, because Ω_1 contains the members $(\emptyset, 1), (\{1\}, 0), \ldots, (\{n_A\}, 0)$, and assuming, say, the use of the weighted or unweighted gradient measures of matrix pattern of Section 4.1.3, $\mathcal{F}((B_1, j_1))$ can be assumed zero for all $(B_1, j_1) \in \Omega_1$, and the recursive process carried out from Ω_1 to Ω_n. The possible transformations between $(B_{k-1}, j_{k-1}) \in \Omega_{k-1}$ and $(B_k, j_k) \in \Omega_k$ are

(1) $j_{k-1} = j_k$ and $B_{k-1} \subset B_k$, where $B_{k-1} - B_k$ contains a single row object placed at the k^{th} order position;

(2) $j_{k-1} + 1 = j_k$ and $B_{k-1} = B_k$, where the single column object with subscript j_k is placed at the k^{th} order position.

The optimal value is achieved for the one entity $(\{1, \ldots, n_A\}, n_B) \in \Omega_n$, i.e., $\mathcal{F}((\{1, \ldots, n_A\}, n_B))$; as usual, working backward through the recursion identifies an optimal ordering of the $n = n_A + n_B$ objects in $S = S_A \cup S_B$, where the column objects in S_B appear in the subscript order $c_1 \prec c_2 \prec \cdots \prec c_{n_B}$. (Obviously, order constraints on the row objects alone [and not on the column objects] could be handled merely by reversing the roles of the row and column objects in the original $n_A \times n_B$ proximity matrix $\mathbf{Q}$.)

If linear orderings can be assumed for *both* the row and column objects in S_A and S_B (that again, without loss of generality, can be taken in the subscript orders $c_1 \prec c_2 \prec \cdots \prec c_{n_B}$ and $r_1 \prec r_2 \prec \cdots \prec r_{n_A}$, respectively), an even more substantial reduction in storage requirements can be achieved. Here, the sets Ω_k for $1 \leq k \leq n$ would be defined so that each entity in Ω_k is of the form (i_k, j_k), where both i_k and j_k are integers within the range 0 to k, subject to $i_k + j_k = k$, which indicates that the first i_k row objects and the first j_k column objects have been placed. The function value $\mathcal{F}((i_k, j_k))$ denotes the optimal contribution to the total measure of matrix patterning when the first k positions in the reordering are occupied by the first i_k and the first j_k row and column objects. The possible transformations between $(i_{k-1}, j_{k-1}) \in \Omega_{k-1}$ and $(i_k, j_k) \in \Omega_k$ are

(1) $j_k = j_{k-1}$ and $i_k = i_{k-1} + 1$;
(2) $j_k = j_{k-1} + 1$ and $i_k = i_{k-1}$.

Beginning with Ω_1, which contains the entities (i_1, j_1), where $i_1 + j_1 = 1$ (i.e., the two pairs $(0,1)$ and $(1,0)$) and $\mathcal{F}((i_1, j_1)) = 0$ (for, say, the weighted or unweighted gradient measures of matrix pattern of Section 4.1.3), the recursive process proceeds from Ω_1 to Ω_n. The latter contains the single entity (n_A, n_B), and $\mathcal{F}((n_A, n_B))$ defines the optimal value for the chosen measure of matrix pattern. The optimal ordering is again identified by working backward through the recursion from Ω_n to Ω_1.

The type of DP recursive process just described for finding a joint sequencing of the sets S_A and S_B when both are subject to linear order constraints was first described in Delcoigne and Hansen (1975) for a particular measure of matrix patterning discussed by Gordon (1973). This measure of matrix patterning (to be referred to as the DHG measure) for a joint sequencing of row and column objects is the sum of the proximities between the row objects and the adjacently-placed column objects plus the proximities between the column objects and the adjacently-placed row objects. So, assuming proximity has a dissimilarity interpretation, an optimal joint sequencing would minimize the DHG measure. (As one technicality, if there is only a single adjacent column object for a particular row object in the joint sequence, then the proximity to this single adjacent column object is doubled; similarly, if there is only a single adjacent row object for a particular column object in the joint sequence, then the proximity to this single adjacent row object is doubled). The same form of the recursive process just described can be carried out for the DHG measure by defining $\mathcal{F}((i_1, j_1))$ for $(i_1, j_1) \in \Omega_1$ to be $2q_{11}$ and the cost increment in moving from $(i_{k-1}, j_{k-1}) \in \Omega_{k-1}$ to $(i_k, j_k) \in \Omega_k$ to be either

(1) $q_{i_k j_k} + q_{i_k (j_k+1)}$ when $j_k = j_{k-1}$ and $i_k = i_{k-1} + 1$ (if $j_k = 0$, the increment is $2q_{i_k(j_k+1)}$; if $j_k = n_B$, the increment is $2q_{i_k j_k}$);

(2) $q_{i_k j_k} + q_{(i_k+1) j_k}$ when $j_k = j_{k-1} + 1$ and $i_{k-1} = i_k$ (if $i_k = 0$, the increment is $2q_{(i_k+1) j_k}$; if $i_k = n_A$, the increment is $2q_{i_k j_k}$).[38]

Numerical illustrations. To provide an example of imposing row and/or column constraints, we will again use the 11×9 data matrix of Table 1.3 on the dissimilarities between the eleven goldfish retinal receptors and the nine specific wavelengths of light. In Section 4.1 the optimal orderings were provided for the combined row/column object set based on both the unweighted and weighted gradient measures. The row orders for the two gradient measures differed by an adjacent receptor interchange for 8 and 9; the separate column orders differed by the adjacent wavelength interchange for $\underline{4}$ and $\underline{9}$ with the unweighted gradient measure giving an ordering that was completely consistent with decreasing wavelengths. To illustrate the effect of imposing row and/or column constraints, we use the weighted gradient measure in the comparisons below but impose for the column constrained joint sequencing that column order obtained for the unweighted gradient measure (i.e., $\underline{3} \rightarrow \underline{8} \rightarrow \underline{2} \rightarrow \underline{7} \rightarrow \underline{1} \rightarrow \underline{5} \rightarrow \underline{6} \rightarrow \underline{4} \rightarrow \underline{9}$). When the additional row constraints are to be imposed, we use the row order obtained also from the unweighted gradient measure (i.e., $10 \rightarrow 11 \rightarrow 5 \rightarrow 7 \rightarrow 1 \rightarrow 4 \rightarrow 6 \rightarrow 2 \rightarrow 8 \rightarrow 9 \rightarrow 3$). To allow comparison, the unrestricted joint sequencing of S_A and S_B from Section 4.1 is also reproduced below based on the weighted gradient measure; also, in all cases the descriptive ratios obtained from (4.4) are provided.

Unrestricted joint sequencing of S_A and S_B, obtained using DPSE2U (taken from Section 4.1):

$\underline{3}\ \underline{8}\ 10\ 11\ \underline{2}\ 5\ 7\ 1\ \underline{7}\ \underline{1}\ 4\ 6\ 2\ \underline{5}\ \underline{6}\ \underline{9}\ 9\ \underline{4}\ 8\ 3$

weighted gradient measure $= 31025 = (32019 - 994)$
ratio $= .940 = (32019 - 994)/(32019 + 994)$

Column order restricted joint sequencing of S_A and S_B, obtained using DPSE2R:

$\underline{3}\ \underline{8}\ 10\ 11\ \underline{2}\ 5\ 7\ 1\ \underline{7}\ \underline{1}\ 4\ 6\ 2\ \underline{5}\ \underline{6}\ [9\ \underline{4}\ \underline{9}]\ 8\ 3$

weighted gradient measure $= 30985 = (31998 - 1013)$
ratio $= .939 = (31998 - 1013)/(31998 + 1013)$

Row and column order restricted joint sequencing of S_A and S_B, obtained using DPSE2R:

$\underline{3}\ \underline{8}\ 10\ 11\ \underline{2}\ 5\ 7\ 1\ \underline{7}\ \underline{1}\ 4\ 6\ 2\ \underline{5}\ \underline{6}\ [\underline{4}\ \underline{9}\ 8\ 9]\ 3$

weighted gradient measure $= 30867 = (31933 - 1066)$
ratio $= .935 = (31933 - 1066)/(31933 + 1066)$

DHG index of matrix patterning:

$\underline{3}\ 10\ \underline{8}\ 11\ 5\ 7\ \underline{2}\ 1\ 4\ \underline{7}\ 6\ \underline{1}\ 2\ \underline{5}\ \underline{6}\ 8\ \underline{4}\ 9\ 3\ \underline{8}$

DHG index $= 2317.0$

In comparison to the unrestricted joint sequencing of S_A and S_B, the imposition of row and/or column constraints does have an effect, albeit somewhat small, in the optimal weighted gradient measures. Also, the comparison of the joint sequencings for the DHG and weighted gradient measures does suggest that the specific type of intermixing of the row and column objects will be influenced heavily by the choice of index.

Although we will not pursue the topic in any further detail here, it may be of some interest to note an extensively developed area in the literature that addresses the task of comparing two linearly-ordered object sets (in contrast to combining them and finding an optimal joint sequencing), where the preferred strategy of comparison is again through a DP recursive process. The volume edited by Sankoff and Kruskal (1983) provides a variety of applications to the comparison of genetic sequences, time-warping problems in the processing of speech, and string-correction methods in computer science, among others. In our notation, the comparison process can be characterized as follows: we are given two sets $S_A = \{r_1, \ldots, r_{n_A}\}$ and $S_B = \{c_1, \ldots, c_{n_B}\}$, where it is assumed that the linear orderings within S_A and S_B are $r_1 \prec r_2 \prec \cdots \prec r_{n_A}$ and $c_1 \prec c_2 \prec \cdots \prec c_{n_B}$. Each of these sequences can be augmented by the inclusion of null elements, say, $\oslash$, to produce two new sequences, $r_1^{'} \prec r_2^{'} \prec \cdots \prec r_{n_A+n_B}^{'}$ and $c_1^{'} \prec c_2^{'} \prec \cdots \prec c_{n_A+n_B}^{'}$, where each $r_i^{'}$ is either a null element $\oslash$ or r_j for some integer j, $1 \le j \le n_A$, and each $c_i^{'}$ is either a null element $\oslash$ or c_j for some integer j, $1 \le j \le n_B$, and the nonnull entities in either sequence are in their given linear ordering. The measure of comparison between the two sequences is based on the one-to-one correspondence between $r_i^{'}$ and $c_i^{'}$, i.e.,

$$\sum_{i=1}^{n_A+n_B} u(r_i^{'}, c_i^{'}), \tag{4.9}$$

where the costs $u(r_i^{'}, c_i^{'})$ are assumed to be given and depend on whether $r_i^{'}$ and $c_i^{'}$ are both nonnull entities, or whether either $r_i^{'}$ or $c_i^{'}$ or both are null entities (it can be supposed that $u(r_i^{'}, c_i^{'}) = 0$ whenever $r_i^{'}$ and $c_i^{'}$ are both null entities because they could merely be deleted from their respective sequences and, thus, would play no role in assessing the dissimilarity of the two sequences). Computationally, we wish to minimize the index in (4.9) and find an optimal matching between the two sequences (as augmented by the possible inclusion of null entities in each); the optimal value is assumed to provide a reasonable measure of dissimilarity between the two original (linearly-ordered) sequences.

An application of the general form of the GDPP to minimize the index in (4.9) is very direct. The sets $\Omega_1, \ldots, \Omega_n$ (where $n = n_A + n_B$) can be defined so that Ω_k contains the ordered pairs (i_k, j_k), where i_k and j_k are integers ($0 \le i_k \le k$; $0 \le j_k \le k$) and refer to the indices in the ordered sets S_A and S_B matched up to this point, i.e., i_k row objects and $k - i_k$ null entities have been matched with j_k column objects and $k - j_k$ null entries. The set Ω_1 contains four members, $(0,0), (0,1), (1,0)$, and $(1,1)$, that correspond to the matching that occurs in the first position, i.e., this matching can be, respectively, two null entities, a null entity and c_1, r_1 and a null entity, or r_1 and c_1. The values $\mathcal{F}((i_1, j_1))$ for $(i_1, j_1) \in \Omega_1$ are assumed given as $u(\oslash, \oslash)$, $u(\oslash, c_1)$, $u(r_1, \oslash)$, and $u(r_1, c_1)$ for these four pairs. A transformation of an entity in $(i_{k-1}, j_{k-1}) \in \Omega_{k-1}$ to an entity $(i_k, j_k) \in \Omega_k$ is possible if

(a) $i_{k-1} = i_k, j_{k-1} + 1 = j_k$ (with a cost increment of $u(\oslash, c_{j_k})$);
(b) $i_{k-1} + 1 = i_k, j_{k-1} = j_k$ (with a cost increment of $u(r_{i_k}, \oslash)$);
(c) $i_{k-1} + 1 = i_k, j_{k-1} + 1 = j_k$ (with a cost increment of $u(r_{i_k}, c_{j_k})$);
(d) $i_{k-1} = i_k, j_{k-1} = j_k$ (with a cost increment of $u(\oslash, \oslash) \equiv 0$).

Using the minimization form of the GDPP in (2.4), the recursion proceeds from Ω_1 to Ω_n, where Ω_n contains the single entity (n_A, n_B). The optimal value for a matching of the two ordered sequences is defined by $\mathcal{F}((n_A, n_B))$, with the actual solution (or matching) obtained by working backward through the recursive process. As noted, the very extensive literature on this comparison task can be accessed through the volume edited by Sankoff and Kruskal (1983).

4.3 Construction of Optimal Ordered Partitions

In Section 3.1, which dealt with the partitioning of an object set, the classes of an optimal partition, based on some measure of subset heterogeneity, had no particular order imposed on them. In contrast, the emphasis in Chapter 4, up to this point, has been on the optimal reordering of an object set, but this task has effectively been carried out by placing one object at a time. As a possible (and obvious) conjunction of these two tasks of partitioning and sequencing, this section discusses the problem of constructing an optimal (ordered) partition of an object set in which the classes of the partition must be sequenced along the continuum. We will only consider one-mode proximity data that come in the usual form of a symmetric matrix $\mathbf{P}$ or a skew-symmetric matrix $\mathbf{P}^{SS}$, and the relevant measures of matrix patterning from Sections 4.1.1 and 4.1.2 will be generalized. There are, as might be expected, a host of variations and extensions that could be pursued, which would parallel many topics discussed earlier in this monograph. Some of these possibilities will at least be noted at the end of this section.[39]

The basic task of constructing an ordered partition of an object set $S = \{O_1, \ldots, O_n\}$ into M ordered classes, $S_1 \prec S_2 \prec \cdots \prec S_M$, using some (merit) measure of matrix patterning and a proximity matrix $\mathbf{P}$ or $\mathbf{P}^{SS}$, can be approached through the general type of GDPP recursive process applied for the partitioning task of Section 3.1 but with appropriate variation in defining the merit increments. Explicitly, the sets $\Omega_1, \ldots, \Omega_K$ (where $K \equiv M$) will each contain all $2^n - 1$ nonempty subsets of the n object subscripts; $\mathcal{F}(A_k)$ for $A_k \in \Omega_k$ is the optimal value for placing k classes in the first k positions, and the subset A_k is the union of these k classes. A transformation from $A_{k-1} \in \Omega_{k-1}$ to

$A_k \in \Omega_k$ is possible if $A_{k-1} \subset A_k$; the merit increment $M(A_{k-1}, A_k)$ is based on placing the class $A_{k-1} - A_k$ at the k^{th} position (which will depend on A_{k-1}, A_k, and $S - A_k$). Beginning with $\mathcal{F}(A_1)$ for all $A_1 \in \Omega_1$ (i.e., the merit of placing the class A_1 at the first position), the recursion proceeds from Ω_1 to Ω_K, with $\mathcal{F}(A_K)$ for $A_K = S \in \Omega_K$ defining the optimal merit value for an ordered partition into K ($\equiv M$) classes (which can then be identified as usual by working backward through the recursion). Also, optimal ordered partitions that contain from 2 to $M - 1$ classes are identified by $\mathcal{F}(A_2), \ldots, \mathcal{F}(A_{M-1})$ when $S = A_2 = \cdots = A_{M-1}$. It is necessary to specify in a particular context the merit increment, $M(A_{k-1}, A_k)$; this is discussed below for various measures of matrix patterning generalized from Sections 4.1.1 and 4.1.2.

For a Symmetric Proximity Matrix P:

To generalize the weighted or unweighted gradient measure given in (4.1), the merit increment for placing the *class* $A_k - A_{k-1}$ at the k^{th} order position is $I_{row}(A_k - A_{k-1}) + I_{col}(A_k - A_{k-1})$, where

$$I_{row}(A_k - A_{k-1}) = \sum_{i' \in A_{k-1}} \sum_{k' \in A_k - A_{k-1}} \sum_{j' \in S - A_k} f(p_{i'k'},\ p_{i'j'})$$

and (4.10)

$$I_{col}(A_k - A_{k-1}) = \sum_{i' \in A_{k-1}} \sum_{k' \in A_k - A_{k-1}} \sum_{j' \in S - A_k} f(p_{k'j'},\ p_{i'j'}).$$

To initialize the recursion, we let $\mathcal{F}(A_1) = 0$ for all $A_1 \in \Omega_1$.

A merit measure based on a coordinate representation for each of the M ordered classes, $S_1 \prec S_2 \prec \cdots \prec S_M$, that generalizes (4.6) can also be developed directly. Here, M coordinates, $x_1 \leq \cdots \leq x_M$, are to be identified so that the residual sum-of-squares,

$$\sum_{k \leq k'} \sum_{i_k \in S_k,\ j_{k'} \in S_{k'}} (p_{i_k j_{k'}} - \mid x_{k'} - x_k \mid)^2,$$

is minimized (the notation $p_{i_k j_{k'}}$ indicates those proximities in $\mathbf{P}$ defined between objects with subscripts $i_k \in S_k$ and $j_{k'} \in S_{k'}$). A direct extension of the argument that led to optimal coordinate representation for single objects would require the maximization of

$$\sum_{k=1}^{M} \left(\frac{1}{n_k}\right) (G(A_k - A_{k-1}))^2, \qquad (4.11)$$

where $G(A_k - A_{k-1}) =$

$$\sum_{k' \in A_k - A_{k-1}} \sum_{i' \in A_{k-1}} p_{k'i'} \; - \sum_{k' \in A_k - A_{k-1}} \sum_{i' \in S - A_k} p_{k'i'},$$

and n_k denotes the number of objects in $A_k - A_{k-1}$. The merit increment for placing the subset $A_k - A_{k-1}$ at the k^{th} order position would be $(1/n_k)(G(A_k - A_{k-1}))^2$, with the recursion initialized by

$$\mathcal{F}(A_1) = \left(\frac{1}{n_1}\right)\left(\sum_{k' \in A_1}\sum_{i' \in S - A_1} p_{k'i'}\right)^2$$

for all $A_1 \in \Omega_1$. If an optimal ordered partition that maximizes (4.11) is denoted by $S_1^* \prec \cdots \prec S_M^*$, the optimal coordinates for each of the M classes can be given as

$$x_k^* = \left(\frac{1}{nn_k}\right) G(S_k^*),$$

where $x_1^* \leq \cdots \leq x_M^*$ and $\sum_k n_k x_k^* = 0$. The residual sum-of-squares has the form

$$\sum_{i<j} p_{ij}^2 - \left(\frac{1}{n}\right)\sum_k \left(\frac{1}{n_k}\right)(G(S_k^*))^2.$$

For a Skew-Symmetric Proximity Matrix $\mathbf{P}^{SS}$:

In extending the measure in (4.7) for skew-symmetric matrices that led to maximizing the above-diagonal entries in $\mathbf{P}_\rho^{SS}$, the merit increment is now defined as

$$J(A_k - A_{k-1}) = \sum_{i' \in A_{k-1}}\sum_{k' \in A_k - A_{k-1}} p_{i'k'},$$

to indicate the contribution from the class $A_k - A_{k-1}$ when placed at the k^{th} order position. Again, the recursive process is initiated by letting $\mathcal{F}(A_1) = 0$ for all $A_1 \in \Omega_1$.

The definition of a merit measure based on a coordinate representation for a skew-symmetric matrix as discussed in Section 4.1.2 is very direct when each object is treated separately because a closed-form solution was possible for the optimal coordinates. The situation changes, however, when attempting to obtain an optimally ordered partition based on $\mathbf{P}^{SS}$ through a coordinate representation. Explicitly, suppose we want to find a set of M coordinates, $x_1, \ldots, x_M$, and a set of M classes, $S_1, \ldots, S_M$ (not necessarily ordered), such that

$$\left(\frac{1}{2}\right)\sum_{k,k'}\sum_{i_k \in S_k,\ j_{k'} \in S_{k'}} (p_{i_k j_{k'}}^{SS} - (x_k - x_{k'}))^2 \tag{4.12}$$

is minimized, where again $p_{i_k j_{k'}}^{SS}$ indicates a proximity between objects with subscripts $i_k \in S_k$ to $j_{k'} \in S_{k'}$. If the classes $S_1, \ldots, S_M$ with respective sizes $n_1, \ldots, n_M$ were given, the M coordinates, say, $x_1, \ldots, x_M$, could be obtained by

$$x_k = \left(\frac{1}{nn_k}\right)\sum_{k' \in S_k}\sum_{i' \in S - S_k} p_{k'i'}^{SS}.$$

There is, however, a preliminary need to choose $S_1, \ldots, S_M$ optimally if the loss function in (4.12) is to be minimized. This optimal choice can be carried out by defining an increment of merit in the GDPP for placing the subset $A_k - A_{k-1}$ at the k^{th} order position, as $(1/n_k)(L(A_k - A_{k-1}))^2$, where

$$L(A_k - A_{k-1}) = \sum_{k' \in A_k - A_{k-1}} \sum_{i' \in S - (A_k - A_{k-1})} p_{k'i'}^{SS},$$

and initializing the recursion by

$$\mathcal{F}(A_1) = \left(\frac{1}{n_1}\right) \left(\sum_{k' \in A_1} \sum_{i' \in S - A_1} p_{k'i'}^{SS} \right)^2 .$$

The optimal classes thus identified, say, $S_1^*, \ldots, S_M^*$, lead directly to the M optimal coordinates $x_1^*, \ldots, x_M^*$, where

$$x_k^* = (1/nn_k) \sum_{k' \in S_k^*} \sum_{i' \in S - S_k^*} p_{k'i'}^{SS},$$

and a residual sum-of-squares of

$$\sum_{i<j} (p_{ij}^{SS})^2 - n \sum_{k=1}^{n} n_k (x_k^*)^2.$$

Numerical illustrations.[40] For constructing optimal ordered partitions, we will again consider the data of Table 4.1 on the rated seriousness of thirteen offenses both before and after viewing a film, and also both its skew-symmetric and symmetric forms (where the latter are based on taking absolute values of the skew-symmetric proximities). For brevity, we will present for the various measures of matrix patterning the original ordering using single objects (i.e., partitions with thirteen ordered classes) and an optimally ordered partition with five classes. In all cases, there was a very precipitous change in the merit measures when moving from five to four classes, and, therefore, the choice of presenting only those optimal ordered partitions with five classes is not arbitrary. As shown in the results summarized below, the five ordered (according to increasing severity) classes of offenses consistently include 'gambler' (#2) in the second class before viewing the film and in the third class after (we might also comment that in most of the analyses reported, the 5-class and the 13-class ordered partitions are completely consistent in the sense that the classes in the former are defined by consecutively-placed single objects in the latter; the exceptions are for the gradient measures and a few adjacently-located objects). In general, the ordered partitions into thirteen and five classes are very similar within the before and within the after conditions, and are consistent over the various measures of matrix patterning for either the symmetric or skew-symmetric proximity matrices (although for completeness we give the exhaustive listing of results for all the options below).

Optimal ordered partitions into thirteen and five classes, obtained from the program DPOP1U:

Before viewing:

symmetric proximity; coordinate representation

```
object order      9       7    10      4       2     11      3     5   13   6   12   8    1
coordinates     −.82    −.78 −.33   −.26    −.23 −.17    −.02   .27 .29 .32 .50 .59 .64
(5 classes):  {     −.80   }     {          −.25           }   {−.02}  {   .29   }   {  .58  }
residual sum-of-squares: (13 classes) 3.307; (5 classes) 3.635
```

symmetric proximity; unweighted gradient

```
object order 9 7 4 10 2 11 3 5 13 6 12 8 1
(5 classes)    {     } {      } { } {13 6 8} {12 1}
index of gradient comparisons: (13 classes) 431; (5 classes) 289
```

symmetric proximity; weighted gradient

```
object order 9 7 10 4 2 11 3 5 13 6 12 8 1
(5 classes)    {          } {      } {  } {6 8} {12 1}
index of gradient comparisons: (13 classes) 161.45; (5 classes) 100.08
```

skew-symmetric proximity; above-diagonal sum

```
object order 9 7 10 2 4 11 3 6 5 13 1 12 8
(5 classes)    { } {      } { } {      } {      }
above-diagonal sum: (13 classes) 49.93; (5 classes) 48.27
```

skew-symmetric proximity; coordinate representation

```
object order      9       7    10      4      2     11      3     5   13   6    1   12   8
coordinates     −.82    −.78 −.33   −.26   −.23 −.18    −.02   .27 .29 .32 .57 .57 .59
(5 classes)    {    −.80    }     {         −.25           }   {−.02}  {  .29   }   {  .58   }
residual sum-of-squares: (13 classes) 3.457; (5 classes) 3.635
```

After viewing:

symmetric proximity; coordinate representation

```
object order      9       7    10      4      11      3       2     6   13   5   12   8    1
coordinates     −.81    −.75 −.39   −.26   −.21 −.05    −.02   .27 .27 .29 .48 .55 .58
(5 classes)    {    −.78    }     {         −.29           }   {−.02}  {  .28   }   {  .54   }
residual sum-of-squares: (13 classes) 2.302; (5 classes) 2.674
```

symmetric proximity; unweighted gradient

```
object order 9 7 10  4  11 3 2 13 6 5 12 8 1
(5 classes)    { } {         } { } {      } {      }
index of gradient comparisons: (13 classes) 491; (5 classes) 331
```

symmetric proximity; weighted gradient

```
object order  9  7 10   4   11  3  2 6 13  5  12   8   1
(5 classes)    {9 7 4} {10 11 3} {      } {5 8} {12 1}
index of gradient comparisons: (13 classes) 165.08; (5 classes) 101.84
```

skew-symmetric proximity; above-diagonal sum

```
object order 9 7 10 4 11 3 2 5 6 13 1 12 8
(5 classes)    { } {      } { } {      } {      }
above-diagonal sum: (13 classes) 47.42; (5 classes) 45.86
```

skew-symmetric proximity; coordinate representation

```
object order      9       7    10     4     11       3      2   6    5   13   12   1    8
coordinates     −.81    −.75 −.39 −.26 −.21     −.05    .02 .27 .28 .29 .51 .55 .55
(5 classes)    {    −.78    }     {   −.29    }   {    −.02   }   {  .28   }   {  .54   }
residual sum-of-squares: (13 classes) 2.369; (5 classes) 2.674
```

The task of identifying optimal ordered partitions has been limited to a discussion of one-mode matrices, but there are numerous variations that could be pursued, related to topics already raised in earlier sections. For example, extensions would be possible to the use of order (or precedence) restrictions on a one-mode matrix (as in Section 3.1.1), or to two-mode proximity matrices that may be row and/or column order (or precedence) restricted (as in Section 3.1.2). Similarly, the classes of the ordered partitions as M varies from 1 to n might be restricted to be hierarchical (as in Section 3.2) and with or without a consecutive order restriction on which objects can form the classes of each partition. Relatedly, as in the discussion of object sequencing in Sections 4.1 and 4.2, gradient comparisons might be restricted to be only within rows or within columns of a one-mode proximity matrix or, for a two-mode matrix, only within the rows or columns. Alternative gradient measures could also be adopted, e.g., minimizing only the gradient violations, attempting to use the Greenberg form, or concentrating on an equally spaced coordinate representation. For a further discussion of the task of constructing optimal ordered partitions, the reader is referred to Hubert, Arabie, and Meulman (1997b).

Endnotes

[27]Because the general recursive process just described requires the storage of intermediate results for all possible subsets of an object set with n members, the two programs to be used in the next three sections, DPSE1U and DPSE2U (where the suffix '1U' refers to '1-mode unrestricted' and '2U' to '2-mode unrestricted'), are effectively limited to object set sizes in their low 20's given the typical RAM configurations currently available, although no formal upper limits are built into either program.

[28]Although we choose not to do so, the discussion of the GDPP applications in this chapter could employ the terminology of posets (see Chapter 2, endnote 6), as well as several more restrictive concepts usually introduced in that framework, e.g., lattices, maximal chains, and the like. To be a little more specific (but still only in a very schematic form based on the notation in Chapter 2, endnote 6), the set Ω would contain all partitions of S, and the relation, $\preceq$, would be defined by partition refinement $A \preceq A'$ (in other words, A is a "refinement" of A') if all the classes in A are in A' or can be formed from subdividing those present in A'. The pair $(\Omega, \preceq)$ is a poset; oreover, it is a lattice in that any two elements A and A' in Ω have a greatest-lower-bound, denoted $A \wedge A'$ (and read as A "meet" A' or as the "meet" of A and A'), and a least-upper-bound, denoted $A \vee A'$ (and read as A "join" A' or as the "join" of A and A'). Formally, $A \wedge A'$ is the unique element of Ω such that $A \wedge A' \preceq A$, $A \wedge A' \preceq A'$, and for any other member A'' of Ω, if $A'' \preceq A$ and $A'' \preceq A'$, then $A'' \preceq A \wedge A'$. Constructively, $A \wedge A'$ is generated from all pairwise intersections of the classes in A and A'. The partition $A \vee A'$ is the unique element of Ω such that $A \preceq A \vee A'$, $A' \preceq A \vee A'$, and for any other member A'' of Ω, if $A \preceq A''$ and $A' \preceq A''$, then $A \vee A' \preceq A''$. Constructively, $A \vee A'$ is generated from the meet of all partitions in Ω that are greater than or equal to both A and A' according to the order relation $\preceq$.

The set Ω is a lattice under the join and meet operations, and any nonempty subset of Ω, say, Ω', that is also a lattice under the same operations is called a sublattice of Ω. If, in addition, A and A' in Ω' imply $A \preceq A'$ or $A' \preceq A$, then Ω' is said to be totally (or simply) ordered and is called a chain. Consequently, the later discussions in Section 3.2 of constructing a hierarchical sequence of partitions could be phrased, if we wished, as finding totally ordered sublattices or chains. Also, in our later use of the term "full partition hierarchy," where successive partitions are constructed by uniting only a single pair of classes from the one given previously, we could refer to a maximal chain, i.e., a totally ordered sublattice of Ω in which each element except the first covers its predecessor.

[29]There is now a rather extensive literature on graphically representing a matrix having either a Robinson or an anti-Robinson form (see Chapter 3, endnote 13, for the distinction between these two terms). In this monograph our emphasis is solely on the main combinatorial optimization tasks, and in this chapter specifically on identifying optimal object orders for a proximity matrix; therefore, we will not go further into the subsidiary issues of graphically representing an (anti-) Robinson matrix that is being used as an approximation. The reader interested in pursuing some of the relevant literature might begin with Diday (1986) and the introduction to graphically representing an (anti-)Robinson matrix by pyramids, and then continue on with the review by Durand and Fichet (1988), who point out the necessity of strengthening the basic (anti-)Robinson condition to one that is strongly-(anti-)Robinson if a consistent graphical (pyramidal) representation is to be possible — otherwise, unresolvable graphical anomalies can arise. Finally, there are two comprehensive review papers on fitting a given proximity matrix (through least-squares) by a sum of matrices each having the (anti-)Robinson form (Hubert and Arabie (1994)) or the strongly-(anti-)Robinson variation (Hubert, Arabie, and Meulman (1998)). The latter discusses in detail, and with all the appropriate historical background, the need to strengthen the basic (anti-)Robinson condition to one that is strongly-(anti-)Robinson if any type of consistent graphical representation is to be achieved (and then, we might add, extends the whole graphical representation to the use of circular orders rather than the linear orders that underlie matrices having an anti-Robinson form).

[30]The program DPSE1U allows such a choice.

[31]Although we will not pursue the topic in any detail here, there are several nice theoretical relationships between these choices of a row and/or column gradient measure and the (approximate) construction of representations for the objects in S as intervals along a continuum based on 0/1 dichotomizations of the optimally reordered proximity matrix (see Roberts (1978), Chapters 3–4; Mirkin (1979), Chapter 1). In particular, the use of only one of the row or column gradient measures is relevant to the construction of (general) interval graph representations (e.g., see Mirkin (1979), Chapter 1), and there is also the use of both simultaneously in the construction of *proper* interval graphs, i.e., no interval is properly contained within another (e.g., see Roberts (1978), Chapters 3–4).

[32]Both weighted and unweighted gradient measures having this latter structure are options in the program DPSE1U and will be illustrated below.

[33]There is no approximate anti-Robinson form for the latter, however, because voters apparently do not have the same location along the continuum. If

they did, and all voters evaluated the parties in the same manner, one might expect such an anti-Robinson pattern in addition to all nonnegative above-diagonal entries.

[34]Although not an option in DPSE2U, another obvious variation would be to maximize the gradient nonviolations alone.

[35]The program DPSEPH, where the suffix ‘PH’ refers to ‘path’, that implements all these optimization variations just described has an effective limit of object set sizes of about 20, given typical current (as of 1999) RAM configurations, although as usual no formal upper-limit is built into the program.

[36]The program DPSEPH accommodates such a nonsymmetric proximity matrix in the construction of optimal linear or circular paths.

[37]The inclusion of precedence constraints through an indicator matrix is an option in each program mentioned thus far — DPSE1U, DPSE2U, and DPSEPH.

[38]The latter DHG measure is an option in the program DPSE2R, used in the numerical illustrations to follow (the suffix ‘2R’ again denotes ‘2-mode restricted’). DPSE2R parallels DPSE2U in all the various options of the latter (including the possible restriction of comparisons to the rows or to the columns of either the original proximity matrix $\mathbf{Q}$ or the derived $n \times n$ matrix $\mathbf{P}^{AB}$) but allows the imposition of either column, or row and column, order constraints that can be provided by the user. It includes (as does DPSE2U) the maximization of the weighted or unweighted gradient measures of matrix patterning (and, as in DPSE2U, only the minimization of the weighted or unweighted discrepancies). As noted, when both row and column order constraints are imposed, the DHG measure just described is an additional option as well. Although again no formal limits are present in DPSE2R, practical RAM configurations will allow row object sizes of about 20 if only the (reasonably-sized) column object set is subject to an order constraint; much larger row and column object sizes (e.g., n_A's and n_B's in the hundreds) are possible when both the row and column objects are subject to order constraints.

[39]Although we will phrase our discussion as one of constructing optimal ordered partitions (or linearly ordered partitions), these are the same entities that are referred to by other names, e.g., as linear quasi-orders (Mirkin (1979), pp. 95–96), and much more commonly as weak orders (Krantz et al. (1971), pp. 14–17).

[40]The numerical illustrations rely on a program DPOP1U (where ‘OP’ refers to ‘ordered partition’ and ‘1U’ to ‘1-mode unrestricted’) that implements the options just described for constructing optimal ordered partitions for either symmetric or skew-symmetric matrices. The program has a built-in limit to object set sizes of 30, but given typical RAM configurations, the effective limit may actually be about 20.