

REDUCING THE SIZE OF THE NONDOMINATED SET: PRUNING BY CLUSTERING

J. N. MORSE†

Department of Business Administration, University of Delaware, Newark, DE 19711, U.S.A.

Abstract—The multicriterion simplex methods of Evans and Steuer[1] and Yu and Zeleny[2] have encouraged model builders to consider matrix criteria. When conflicting objectives are simultaneously considered, there is no such thing as an optimum solution. Rather, a preferred class of basic feasible solutions called the nondominated set results. Since this set can be extremely large, some means must be found to prune it. Steuer[3] has proposed a filtering method. Another mechanistic aid to the decision maker (DM), based on cluster analysis, is presented in this paper.

The idea is to portray the nondominated set N by a representative subset. Cluster analysis partitions N into groups of relatively homogeneous elements. In this research I added a very general evaluative criterion: minimum redundancy. Since there is a threshold of resolution beyond which the DM cannot perceive the difference between two very similar solution vectors, there is little point in making him waste time processing all of N in the search for a final solution.

Two forms of cluster analysis are tested—direct clustering and hierarchical clustering. Within the group of hierarchical methods there are eight algorithms. In the present application the two worst things that could happen are clusters that “chain” and outlying vectors (the residue set) that are obscured. Taking account of these two undesirable outcomes, three algorithms worked best on the particular data used—Ward's Method, the Group Average Method, and the Centroid Method.

The hierarchical methods are recommended over direct clustering. (However, some similarity between direct and hierarchical clustering is discovered.) Hierarchical clustering serves to minimize redundancy, and thereby reduces the chance that the selection of a final solution will stress the decision maker beyond his information endurance.

The concepts stressed in this paper are very similar to those expressed in Törn[4]. This article presents computational experience with the cluster analysis which was developed independently by Törn, whose approach and mine will be combined under an algorithmic strategy called Two-Stage Pruning (TSP). TSP first reduces the nondominated set to a representative set. This set, in turn, is interactively manipulated until a decision evolves.

1. INTRODUCTION

Matrix criteria can be implemented in the linear case by the multicriterion simplex methods of Evans and Steuer[4] and Yu and Zeleny[2]. The problem is as follows:

$$\begin{aligned} \text{Max } Cx \\ \text{s.t.} \\ Ax \leq b \\ x \geq 0 \end{aligned} \quad (1)$$

where C is a $p \times n$ matrix with rows equal to the criterial coefficients of each objective function of the problem, and A , x , and b are properly dimensioned as in a standard linear programming problem.

It is known that (1) yields a set of nondominated (also known as efficient or Pareto-optimal) solutions. This set is denoted N , or N^{ex} when the extreme points are to be considered. Of the basic solutions to (1), the elements x of N^{ex} are identified by the absence of feasible \bar{x} such that:

$$C\bar{x} \geq Cx$$

and

$$C\bar{x} \neq C\bar{x} \quad (2)$$

†Joel N. Morse received the A. B. degree in Political Economy at Williams College, and the PhD. in Business Administration at the University of Massachusetts (Amherst). Since 1976, he has been Assistant Professor of Finance at the University of Delaware. Current research interests include pruning strategies in MCDM, and capital budgeting applications of 0-1 multiobjective programming.

Under the multiobjective linear programming (MOLP) paradigm, N^{ex} can become immense. For example, Zeleny[5, p. 117 ff.] shows that a certain problem with eight decision variables, five objective functions, and eight constraints has seventy nondominated extreme points. If points along nondominated faces are considered, the size of N obviously is infinite.

To represent the process of reducing N^{ex} to one action, we will use the word pruning. The decision maker's (DM) perusal of N^{ex} can be pictured as a search on a tree; thus the etymology of pruning becomes clear. Whether pruning is heuristic or aided by algorithmic devices, it is a search process through which the DM may discover new goals, relationships, and cultural or strategic externalities. To be precise, pruning involves, by definition, minimizing redundancy among the decision vectors of N . Implementing the idea, the methods of the paper are applicable to both N^{ex} and to any other finite subset of N .†

In concrete multiple objective applications new criteria and constraints will emerge to guide the DM toward a choice. Or some of the original criteria, if found to be redundant, can be dropped, which might reduce the size of N^{ex} . However, this natural pruning is not costless. When N^{ex} is very large the DM cannot process it; time runs out and information stress ensues. All problems can be reduced to some degree by minimizing redundancy. In this paper cluster analysis is proposed for this universal aspect of pruning. Since the concepts of this approach are very similar to those developed independently by Törn[4], I will suggest a unifying description called Two-Stage Pruning (see Sec. 3 below).

2. CLUSTER ANALYSIS

The formation of m groups or clusters of similar objects, directly from the n original objects, is called cluster analysis (see Anderberg[6], Bijnen[7], Hartigan[8], Sneath and Sokal[9] and Sokal and Sneath[10]). The clusters are usually mutually exclusive, or non-overlapping. In some measurable way, the clusters are to be composed of objects that are similar, either qualitatively or quantitatively. Operationally, a cluster can be defined as whatever results from a clustering algorithm. Or it can be one of the set of groups that emerges when some objective function, such as within-groups sum of squares, is minimized (Rao[11]).

Formally speaking, a clustering is a mapping $\beta: \text{OTU} \rightarrow L$, when OTU is the set of objects to be clustered (called *operational taxonomic units*) and L is the set of clusters. Each element of L is naturally a non-empty subset of $\{\text{OTU}\}$, but the mapping β is of very general form. If β is such that each element of $\{\text{OTU}\}$ is mapped into one and only one cluster, then β is said to be a *partition*. If β forms clusters that are nested (i.e. some may be disjointed but others may be within larger clusters then it is called a *tree*.

The most appealing view of a cluster comes from Carmichael *et al.*[12]. They define *natural clusters* as areas where the OTUs are dense, surrounded by areas where they are sparse. Implementing this computationally led to a myriad of attempts as described in the literature; many are, in some way, optimization algorithms themselves.

There are three types of clustering strategies. First, the divisive approach starts with the n objects as a cluster, and successively refines this into smaller clusters. Second, the agglomerative approach starts each trial cluster as one object. Then other objects are linked to the first and larger and larger clusters are formed. Third, the iterative approach begins with an arbitrary choice of clusters and moves objects from cluster to cluster until some homogeneity criterion is met (see Anderberg[6, p. 156 ff.]).

In clustering large sets of data, optimal search techniques converge too slowly to be of practical use. Heuristic methods exist; the obvious drawback is that an optimal set of clusters is not guaranteed.

We are using cluster analysis as an aid in multiobjective programming problems.‡ There is an important difference between the treatment of outliers in most statistical applications and that of the present paper. Statisticians normally develop methods to reduce the influence of outliers because they are interested in measures of central tendency and higher moments which

†A referee has pointed out the high cost of meticulously computing N^{ex} , only to delete many of its elements. There are, as he points out, methods that *approximate* the nondominated set. If that approximation turns out to be unwieldy in size the methods of this article remain applicable.

characterize dispersion. But in pruning N^{ex} we wish to stress the outliers and the boundaries of the dense central areas.

The vectors of criterial achievement ($z = Cx$, $x \in N$) are geometric rays that are "interpreted" according to the DM's preferences. One ray will be the final solution vector. Some of the z vectors are so close to each other that the finite sensibility of the human mind cannot distinguish between them. This near-collinearity is most apparent in the dense central areas of N . It is precisely at this point that cluster analysis can be used to eliminate redundancy.

But the outlying z vectors carry much information. They are solutions to the MOLP problem that will not join any cluster. They represent unique combinations of criterial achievement, or unusual weights on those same criteria. Therefore it is important to collect and display these points systematically, even though they do not join any clusters. The RESIDUE facility of CLUSTAN (see Wishart[14]) does exactly this. The set composed of the outliers plus a representation of each cluster (this could be the centroid or any other element) will be the subject of managerial attention. Letting z' be any value vector which represents a cluster and z'' be any one that doesn't, we can form the set

$$\{g\} = G(N^{ex}) = \{z'\} \cup \{z''\} \subseteq N^{ex} \subset N. \quad (3)$$

We will refer to this set as the *generators* of N (or of N^{ex}).

Summing up, clustering involves choosing distance and joining rules so that the DM and the analyst can draw a picture of the infra-structure of N . The $\{z''\}$ must be accentuated by means of the RESIDUE facility because they carry more information (surprise value) than do the $\{z'\}$. This is analogous to simplifying and interpreting a painting by choosing the points that show what the artist wishes us to see and feel. For example, in a picture of a mountain, the point that marks the top is more crucial than one of the thousands of points that convey the details of the foothills. If the pruning process that accompanies MOLP does not expose the DM to $\{z''\}$, it fails in the same way that goal programming does. Steuer[15, p. 5] has written that GP presents the DM with "too sparse" a representation of the nondominated set.

2.1 Block clustering

Block (also known as direct) clustering was introduced in Hartigan[16] with slight amendments added in computer codes written by the author in Hartigan[8] and Biomedical Computer Programs[17]. This type of clustering is initially very appealing to the researcher in fields where taxonomic methods are relatively new. This is because block clustering does not require the specification of a distance function.

The distance function describes inter-object distances which are used to create a similarity matrix. An example is the Euclidean distance

$$d_{ij} = \left[\sum_{k=1}^n (x_{ik} - x_{jk})^2 \right]^{1/2}. \quad (4)$$

When one considers binary, continuous, and mixed data, the literature includes at least fifty different distance coefficients. Since the resulting clusters depend on the definition of distance, the analyst must be wary of spurious clusters. Theories of distance exist in disciplines such as zoology, where the use of cluster analytic methods dates back to the first brush with twentieth-century empirical research methods. It is known, for instance, that particular distance coefficients perform well with particular types of zoological and biological data.

In the administrative and policy sciences, only marketing research has passed the beginning of the learning curve with cluster analysis. The application at hand, pruning the nondominated sets in MOLP problems, is the first attempt of its kind. Steuer[3, p. 11] uses a filtering technique that is equivalent to forming clusters from given cluster centroids. But he does not call the procedure cluster analysis, nor is the implied algorithm one that really finds clusters.

Since clustering in this context is new, there is no applicable theory of distance. The question remains: "Of all the possible distance measures, which best describes the DM's

(psychological) reactions as he compares elements of N^{ex} ?" Zeleny[5] addresses this issue. His distance functions are part of a rather mechanistic approach requiring the use of an *ideal point*. This point is that intersection of rays (usually not feasible) where each objective function finds its maximum in z space (criterion space).

In this Sec. I will assume that the DM wishes to do some searching before the learning from the display of an ideal point is initiated. Owing to a lack of such an *anchor* and to interdependencies of the structure of the feasible set, block clustering is very appealing because it is not necessary to specify a distance function. Seventy block clustering runs were completed using the data associated with a solved problem of Zeleny[5, pp. 117 ff.].[†] The objects clustered are the seventy nondominated solutions (in criterion space). The criteria and constraints are arbitrary; thus the extreme points themselves have no intuitive meaning. The clustering behavior of various algorithmic strategies is what interests us.

Block clustering transforms the original multidimensional data (in this case nondominated solution vectors) by classifying each observation as follows. For a z vector to be pruned, each criterial dimension is partitioned into intervals. The actual components of z become 1, 2, 3, etc. as they happen to fall in the first, second or third divisions of the range spanned by the numbers attained on each objective function. For example, if the nondominated solutions for the i th objective ranged from one to four, with three intervals specified, a 1.6 would be recoded as a one.

This "classified" data set is manipulated (for details see Hartigan[8, 16]) into *data blocks*. These are obtained by switching rows and columns until similar values are positioned near each other. Then at the horizontal margin one can perceive clusters of objectives, and along the vertical margin appear clusters of N^{ex} points. Since only clusters with more than one z vector achieve the goal of reducing the redundancy within the nondominated set, we change various technical parameters of the computer code until we achieve a manageable number of multiple element clusters. From each one we arbitrarily choose a z vector to represent its cluster and thus obtain $G(N^{ex})$, the generators of the nondominated set.

Table 1 summarizes the block clustering of the solutions (N^{ex}) of the Zeleny[3, pp. 117 ff.] problem.

The data blocks which consist of only one solution are often discarded or joined to other clusters; however they do contain important information. As stated previously, these outliers

Table 1. Zeleny data-block clustering

# of intervals within each obj.	# of blocks not single	Objective functions that cluster
1, 1, 1, 1, 1	1	2nd, 4th, 3rd
1, 1, 6, 6, 1	11	2nd, 5th, 4th
2, 2, 2, 2, 2	10	3rd, 2nd, 1st
2, 2, 4, 4, 2	22	
3, 3, 4, 4, 3	36	3rd, 5th, 2nd
3, 3, 3, 3, 3	23	5th, 2nd, 3rd
4, 4, 3, 3, 4	30	3rd, 5th, 1st
4, 4, 4, 4, 4	32	3rd, 5th, 1st
5, 5, 5, 4, 5	42	3rd, 5th, 1st
5, 5, 5, 5, 5	39	3rd, 5th, 1st
6, 6, 6, 6, 6	46	3rd, 5th, 1st
7, 1, 1, 1, 1	7	2nd, 5th, 4th
7, 7, 7, 7, 7	52	3rd, 5th, 1st
8, 8, 8, 8, 8	55	3rd, 5th, 1st
9, 9, 9, 9, 9	50	3rd, 2nd, 1st
10, 10, 10, 10, 10	55	3rd, 2nd, 1st
11, 11, 11, 11, 11	56	3rd, 2nd, 1st
12, 12, 12, 12, 12	51	3rd, 2nd, 1st

[†]The reader may obtain the data and the computer analysis by corresponding with the author. Also available is similar experience with a problem of Steuer[18].

are associated with unusual weights on objectives. Since they represent areas of value space that the DM should consider, some portion of the single element data blocks should be included in $G(N^{\text{ex}})$. An appealing sampling procedure is suggested in Törn[4].

Block clustering is a two-way splitting technique. That means that clustering occurs simultaneously for the rows and for the columns. Revealed clusters in the columns show goal collinearity, or correlation. Since this may indicate the need for model respecification, it is necessary for the analyst to check for this phenomenon. The methods of Morse[19] help do that job in a formal manner. This non-orthogonality of objective functions leads to unwanted amplification and distortion of goals (see, for example, Morse[19] or Einhorn[20]). But since block clustering reveals this pathology in a very unstable way, we are led to consider more classical numerical approaches to clustering.

Suppose that the DM deemed it practical to consider about ten clusters, plus a small sampling of outliers. Note in Table 1 that we must drop the interval specification to two in order to achieve that goal. In other words, block clustering degrades the data, in this case, to a "high-low" categorization. This might result in a loss of information. If that loss is warranted in order to conform to practical size constraints, then the data has simply been recoded into ordinality-scaled data. For example, the set of N^{ex} values (7, 9.41, 3, 27) could be coded into three intervals on the range 3–27. The new data would be (1, 1, 1, 3). But there is a cluster analysis literature on distance functions for non-ratio-scaled data such as this. The claim that block clustering does not require the user to define a distance function is misleading. Hartigan's methodology[16] does implicitly define one when it chooses the level of aggregation (number of intervals) for each objective. As a matter of face, Hartigan's *thresholds* explicitly use distance concepts. Perhaps this is why he found "disappointingly few" differences between block clustering and average distance linkage (to be explained in Sec. 2.2) when both were applied to state voting data.

There is one more way in which direct clustering behaves like the hierarchical distance methods which are discussed in the next section. Hartigan[16, p. 126] explains that direct clustering "splits up" the data matrix in such a way that the downside change in error sum-of-squares (SSQ) is maximized at each stage. Ward's method (described below) is one of the best hierarchical methods, and it fuses clusters at each stage using essentially the same minimizing objective function!

2.2 Hierarchical clustering.

As the attractiveness of block clustering diminishes, the use of the *hierarchical* methods is motivated. This type of clustering system does not "split" data blocks as does direct clustering. Instead it proceeds only on *cases* to join OTUs into successively larger clusters.

Eight different hierarchical algorithms were applied to the Zeleny N^{ex} points. In the following paragraphs, brief summaries of these methods are given. The original references can be found in the bibliographies of Anderberg[6], Hartigan[8] and Wishart[14].

First, there are three *linkage* methods. In *single linkage* algorithms the similarity between any two clusters is equal to the highest single similarity coefficient between two cases, each of which resides in a different cluster. In other words, when searching for two intermediate clusters to fuse, the nearest neighbor rule prevails. The single linkage method finds long serpentine clusters very well. But on larger data sets this can lead to *chaining*. Chaining occurs when most of the OTUs join one cluster, and the rest form a meaningless residue.

The *complete linkage* method uses the smallest similarity (or highest dissimilarity) to amalgamate two points or two clusters into a new one. In other words a farthest neighbor rule is employed. This strategy produces spherical clusters. But these may be spurious clusters, since group structure is not considered. Outliers may be too influential in forcing clustering.

The third method is the *average linkage* method, which considers the average of all the similarity coefficients for all pairs of individuals across clusters. Thus group structure is recognized. This algorithm, which has been introduced independently by several authors, is a well-behaved search strategy which creates spherical clusters.

The fourth hierarchical option employed was *centroid sorting* (also known in the literature as the *weighted group* method of Sokal and Michener). Clusters are joined by finding the center of

gravity, or the mean of the clusters. Chaining occurs, but less perniciously than in the single linkage method.

The fifth approach is the *median* method. Valid mainly with distance type coefficients, this method computes a distance $s(r, p + q)$ where r is an arbitrary cluster and $p + q$ is a cluster fused during the last stage of the process. The distance $s(r, p + q)$ is measured from the centroid of r and the midpoint of the line which joins the centroids of p and q . Again, chaining is likely for large populations.

The sixth type is known as *Ward's method*, although Orloci and Wishart have also written about it. First, the distance of each OTU to the centroid of its parent cluster is recorded. For each cluster these distances are added up and called the error sum of squares. The next two clusters to be joined are those for which the increase in this error sum of squares is minimized. Recall that block clustering reversed this; a split was performed which maximized the decrease in the error sum of squares. Ward's method computes minimum-variance spherical clusters.

The *Lance-Williams flexible beta method* offers the chance to dilate or contract space by varying only one parameter of a joining measure. In the expression

$$s(r, p + q) = \{[s(r, p) + s(r, q)(1 - \beta)]/2s(p, q)\}\beta \quad (5)$$

it is β that acts like the dilation and contraction operators of Zeleny[3, p. 172] or Zadeh[21]. With $\beta = -0.25$, the flexible beta method is reported to act like Ward's method.

The last major variant of hierarchical clustering is *McQuitty's similarity analysis*. Two clusters. Two clusters are joined to form one new one by minimizing the coefficient

$$s(r, p + q) = [s(r, p) + s(r + q)]/2. \quad (6)$$

Notice that this is equivalent to the Lance-Williams method with $\beta = 0$. Again, chaining occurs with large populations.

All eight of the hierarchical clustering algorithms were used on the Zeleny N^{ex} points.[†] As Table 2 indicates, five of these methods caused chaining problems. About 60 of the 70 nondominated points tended to join one cluster, with the remainder forming clusters each with only one, two, or three elements. This outcome was useless managerially, since it does not reduce the redundancy within the N^{ex} set. The DM would still have to process all the points in the "chained" cluster heuristically. The human brain may be poorly suited to this task.

Three of the methods performed well on the Zeleny test data. They were Ward's method, the group average method, and the centroid method. as Appendix I shows, the three methods produce clusters with fairly similar composition. There is too much data to subject to a

Table 2. Hierarchical clustering methods that caused chaining

-
1. The single linkage method. NOTE: This method was implemented by both BMDP02M (Biomedical Computer Programs /17/) and CLUSTAN 1C (Wishart /14/).
 2. The complete linkage method (i.e., the farthest neighbor rule).
 3. Gowers median method.
 4. The Lance-Williams flexible beta method.
 5. McQuitty's similarity analysis.
-

[†]Sample print-outs for problems of both Zeleny and Steuer are available from the author. Of particular interest are features of CLUSTAN which help the analyst understand the structure of N^{ex} , such as the order of agglomeration, a dendrogram table, and k -linkage lists.

contingency analysis or chi-square tests. But since each method was tested with between ten and five clusters "requested," the behavior of the three clustering strategies is not extremely difficult to follow by inspection.

Only one method guarantees minimum-variance spherical clusters. This feature makes Ward's method the most attractive of the three non-chaining algorithms that appear in Table 2. It is important that the DM be exposed to a wide sample of the efficient region. Since we propose that $G(N^{ex})$ be composed of one representative of each cluster plus some outliers, unequal size clusters lead to a superficial review of regions that are dense in nondominated points. Too many such points would be represented by only one generator $g \in G(N^{ex})$. An even more subtle coverage concept due to Steuer and Harrison[22] is described in Sec. 3. For the present discussion, Ward's method is favored because of its tendency to yield clusters of fairly equal size and shape.

There are commonly used methods to help absorb the masses of information provided by a good computer code for clustering. For example, in Fig. 1 the positions of the N^{ex} points are plotted and numbered. Since the portrayal occurs in the plane, only objective function No. 1 vs objective function No. 2 can be visualized. Naturally, in a real business decision all pairs of objective functions could be so plotted.

Figure 2 reduces some of the visual clutter of Fig. 1 by drawing circles around the points in each cluster, and eliminating the actual case, numbers. Since the computer, in this particular case, was coded to produce seven clusters, there are seven circles. Except for cluster No. 6, the clusters are reasonably tight, spherical and disjointed as seen along the first two criterial dimensions.

The most informative graphical aid is the *dendrogram*, as seen in Fig. 3. The dendrogram draws the joining of two OTUs and indicates, on its vertical axis, the level of similarity at which this occurs. This is similar to, but much clearer than, the tree joining diagrams which are provided by Hartigan's block clustering. Figure 3 shows both the clusters that are formed at varying levels of amalgamation and the order of that joining.

For the DM who prefers to think algebraically rather than graphically, some interesting data manipulations are available via the *minimum spanning tree*. This has appeared in graph theory and in operations research (for example, see Kruskal[23]).

Consider this structure as it appears in what is called the traveling salesman problem. A group of cities is to be connected by a system of roadways. Various pairs of cities will be linked, and over each link there will be a known fixed cost per mile. The problem is to find a route through the maze of cities, such that each one is visited and the cost is minimized.

The solution to the problem, in graph theory terms is the *minimum spanning tree* (MST). The DM can develop clusters directly from an examination of the MST. By deleting links whose length is greater than g , maximal single-linkage type clusters of diameter g are discovered. Table 3 gives the minimal spanning tree for the Zeleny problem.

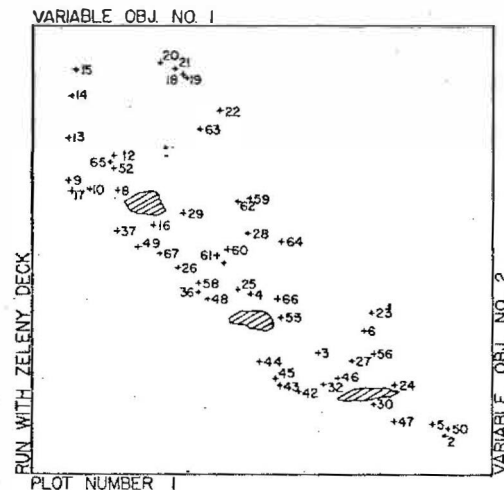


Fig. 1. Scatter diagram of Zeleny problem.

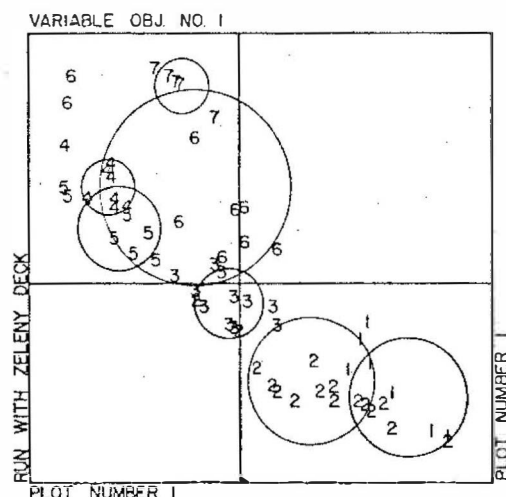


Fig. 2. Cluster diagram of Zeleny problem.

Conceptually similar to the minimum spanning tree is the k -linkage list. This shows the k nearest neighbors for each OTU. It is rather like Steuer's and Harrison's "filter" method[22], which focuses the DM's attention on a particular set of N^{ex} points. These authors suggest exhibiting a subset of N^{ex} and asking, "Which point do you like best?" Then the k nearest neighbors to that point are printed.

This procedure is appealing but non-optimal. Certain attractive areas of the nondominated set may be prematurely discarded in this way. The DM should be encouraged to examine a neighborhood around a *tentatively* desirable point and compare it to several other neighborhoods.

2.3 Block clustering vs hierarchical clustering: A summary

Block clustering has two main advantages, computational speed and direct interpretation of the data. With modern computers the first is only relevant on extremely large problems. Hartigan[8, p. 267] states that two-way splitting (another name for the type of direct clustering used in this chapter) draws "beautiful (data) pictures." But he cautions that the data must be rescaled so that an error of one unit along one variable dimension will be equal to a one-unit error along another dimension. The rescaling creates problems when the data are of several sorts such as binary, nominal, interval or ratio. Hierarchical clustering also has difficulty overcoming mixed data bases. Shift phenomena are particularly well captured by block clustering. In Hartigan[16] presidential election data is analyzed. In South Carolina, Louisiana,

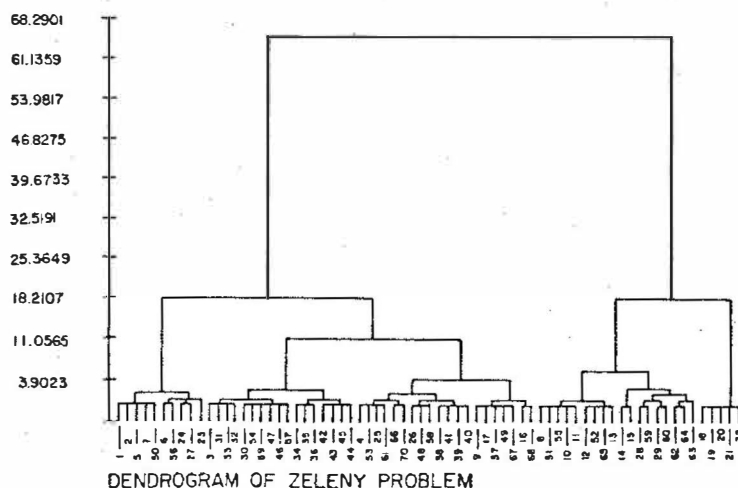


Fig. 3. Dendrogram of Zeleny problem.

Table 3. The minimum spanning tree of the Zeleny problem *

Edge No.	First Vertex	Second Vertex	Edge Length	Edge No.	First Vertex	Second Vertex	Edge Length
1	1	7	0.005	36	36	38	0.111
2	2	1	0.001	37	37	49	0.042
3	3	53	0.093	38	38	41	0.000
4	4	53	0.032	39	39	41	0.000
5	5	7	0.007	40	40	39	0.000
6	6	56	0.099	41	41	48	0.044
7	7	50	0.007	42	42	43	0.008
8	8	55	0.031	43	43	45	0.001
9	9	55	0.092	44	44	35	0.082
10	10	51	0.062	45	45	44	0.012
				46	46	69	0.017
11	11	51	0.047	46	46	69	0.017
12	12	65	0.021	47	47	30	0.024
13	13	65	0.046	48	48	58	0.020
14	14	12	0.212	49	49	67	0.052
15	15	14	0.147	50	50	24	0.113
16	16	70	0.167	51	51	8	0.016
17	17	37	0.071	52	52	11	0.061
18	18	19	0.011	53	53	66	0.084
19	19	21	0.021	54	54	69	0.003
20	20	18	0.016	55	55	67	0.179
21	21	63	0.511	56	56	53	0.336
22	22	19	0.059	57	57	54	0.040
23	23	6	0.140	58	58	25	0.089
24	24	56	0.178	59	59	28	0.180
25	25	4	0.053	60	60	29	0.074
26	26	58	0.033	61	61	4	0.064
27	27	24	0.051	62	62	59	0.414
28	28	60	0.174	63	63	62	0.395
29	29	52	0.130	64	64	62	0.075
30	30	54	0.003	65	65	52	0.016
31	31	42	0.081	66	66	70	0.086
32	32	31	0.074	67	67	58	0.110
33	33	31	0.006	68	68	16	0.016
34	34	36	0.052	69	69	33	0.051
35	35	34	0.002				

* Edge length is Euclidean distance.

and Mississippi the percentage of votes cast for Republicans went up greatly from 1960 to 1964, and then down again in 1968. This shift over time would not be tracked well by hierarchical methods because states and years would not be simultaneously clustered.

From a managerial point of view, block clustering, as mentioned earlier, cannot pre-specify the desired number of clusters. An allied problem is the interpretation of a case cluster. Since these form separately along various subsets of the objectives, the DM cannot tell how the cases would cluster if all objectives (or criteria) were considered. This is an implicit 0 – 1 weighting of criteria; formation of the case clusters in this way does not interface well with the DM's need to learn his preference structure.

In the section on hierarchical clustering we saw how important the RESIDUE facility of CLUSTAN was. In block clustering this feature could be copied by taking the small data blocks and listing them as additions to the list of generators $geG(N^{ex})$. But sometimes to keep the size of $G(N^{ex})$ manageable, we need to assign outlying z vectors to existing clusters. This can be done with the RELOCATE option of CLUSTAN, but cannot be done in direct clustering.

The conclusion here is that hierarchical clustering methods show more promise than block clustering for the purpose of pruning the nondominated set N^{ex} . However, the choice of clustering strategy is not as risky as it is in a static taxonomic study. In the dynamic decision context we are never quite certain of the effects of our strategies. As the problem evolves, new information overwhelms the effect of less than optimal clustering in the previous stage.

3. TWO STAGE PRUNING

Next we discuss pruning by clustering in the broader context of Multiple Criteria Decision Making. The process of *decision making* could be divided into two stages. In the first stage, most of the work is done by a computer-based model. In the second stage, much more of the

task is completed by the decision maker. Since both stages drive toward the adoption of one of the nondominated solutions, the name Two Stage Pruning (TSP) is appropriate.

Cluster analysis has been suggested as a mechanistic decision aid. A new evaluative criterion, minimizing redundancy, has been used to discover the generators of N^{ex} , much as an artist creates a picture by accentuating a limited set of points which suggest the whole visual mass. As a matter of fact, the proper choice of points can be more informative than the original flood of sensory data. In the MCDM case, the generators $G(N^{ex})$ focus the DM on the bunching of solutions, the dispersions of implied preference weights, and the trade-offs among objectives.

Two other papers in this issue address themselves to pruning. Törn samples points from the feasible region. Then he uses optimization techniques to move toward an efficient frontier. If there are any points to be presented to the DM, he proposes cluster analysis to obtain *solution types*. These techniques of Törn are first stage strategies in a TSP paradigm. The second stage (recall that this part involves much greater interaction with the DM) invokes a confidence concept. The DM, with the aid of a decision analyst (DA), tries different efficient regions of the feasible space; he adds new evaluative constraints as desired. The problem is considered solved when the DM has confidence that he is near an optimum.

The paper of Steuer and Harris can also be viewed as TSP. In stage one the efficient points are found. *Forward filtering* is used to temporarily discard redundant solutions. When a particular point causes a positive response from the DM, nearby nondominated points are displayed. This opportunity for detailed assessment of a subset of N^{ex} is called *reverse filtering*. A further decision aid is intra-set point generation. Since attention to the extreme points of N may fail to represent its infinity adequately, we may wish to consider non-extreme points on efficient faces. For example, imagine a non-equilateral triangle in criterion space. a very long edge might have a higher probability of containing the best choice than do the shorter edges of the corner points.

This is an important addition to the body of first stage tools. Steuer and Harrison can generate the intra-set points by drawing weight vectors from the uniform and the Weibull distributions. Embedded in this is another check for efficiency, since not all convex combinations of efficient points are themselves efficient. Given an expanded candidate list, the DM can filter forward and backward to the final solution.

The two papers just cited rely almost exclusively on the heuristic processes of the DM for the second stage of pruning. It is desirable to continue computer-based aid in stage two also. This is a lengthy subject, but tentative ideas can be found in Morse[24].

4. SUMMARY AND CONCLUSIONS

A draw back of the very promising MCDM approach is the sheer size of the nondominated set. For this reason a decision support system to pre-process N^{ex} has been outlined. The redundancy inherent in that set is reduced by using a statistical technique called cluster analysis.

Numerical experiments were conducted with block clustering and hierarchical clustering. The best dispersion of generators $G(N^{ex})$ was obtained for this limited data base by means of Ward's method (one of the hierarchical strategies).

Linkages with the similar research of Törn[4] and Steuer and Harris[22] were possible by exploiting Two Stage Pruning. The passage of control from machine to human is thus described and facilitated.

REFERENCES

1. J. P. Evans and R. E. Steuer, A revised simplex method for linear multiple objective programs. *Math. Programming* 5 54-72 (1973).
2. P. L. Yu and M. Zeleny, The set of all nondominated solutions in linear cases and a multicriteria simplex method. *J. Math. Analysis and Applications* 49, 430-468 (1975).
3. R. E. Steuer, A five phase procedure for implementing a vector maximum algorithm for multiple objective linear programming problems. *Working Paper*, University of Kentucky (1976).
4. A. A. Törn, A sampling-search-clustering approach for exploring the feasible/efficient solutions of MCDM problems. *Comp. Op. Res.* this issue.
5. M. Zeleny, *Linear Multiobjective Programming*. New York (1974).
6. M. R. Anderberg, *Cluster Analysis for Application*. New York (1973).
7. E. J. Bijnen, *Cluster Analysis: Survey and Evaluation of Techniques*. Holland (1973).

8. J. A. Hartigan, *Clustering Algorithms*. New York (1975).
9. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy; The Principles and Practice of Numerical Classification*. San Francisco (1973).
10. R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy*. San Francisco (1963).
11. Mr. Rao, Cluster analysis and mathematical programming. *J. Am. Stat. Association* 66(335), 622-626 (1941).
12. J. Carmichael, J. George and R. Julius, Finding natural clusters. *Systematic Zoology* 17, 144-150 (1968).
13. P. Hanson, Bicriterion cluster analysis as an exploratory tool. *Proc. Conf. Multiple Criteria Problem Solving*, Buffalo (1977).
14. D. Wishart, CLUSTAN 1C London (1975).
15. R. E. Steuer, Interval criterion weights programming. *Working Paper*, University of Kentucky (1975).
16. J. A. Hartigan, Direct clustering of a data matrix. *JASA* 67, 123-129 (1972).
17. Biomedical Computer Programs, BMD P Series Los Angeles (1975).
18. R. E. Steuer, Linear multiple objective programming: Theory and computational experience. Ph. D. Thesis, University of North Carolina (1973).
19. J. N. Morse, A theory of naive weights, in: S. Zionts (ed.) *Multiple Criteria Problem Solving*. 384-401 (1978).
20. H. J. Einhorn, Unit weighting schemes for decision making. *Organizational Behavior and Human Performance* 13, 171-192 (1968).
22. R. E. Steuer and F. W. Harris, Intra-set point generation and filtering in decision and criterion space. *Comp. Op. Res.* 7(1-2), 41-53 (1980).
23. J. B. Kruskal, On the shortest spanning subtree of a graph, and a traveling salesman problem. *Proc. Am. Math. Soc.* Vol. 7 48-50 (1956).
24. J. N. Morse, A ranking algorithm for pruning, Working Paper, University of Delaware, Newark, Delaware (1980).

APPENDIX I

A comparison of three hierarchical clustering algorithms on the Zeleny data

Note: The CLUSTAN code was set to yield ten clusters. It can be seen that the seventy (numbered consecutively) nondominated extreme points of the Zeleny solutions join to form ten fairly equal clusters under Ward's method, and ten unequal clusters under the group average method and the centroid method.

	Ward's method	Group average method	Centroid method
(with 10 clusters specified)			
No. 1	1	1 50	1
	2	2 56	2
	5	5	5
	7	6	6
	50	7	7
		24	24
		27	27
			50
			56
No. 2	3 46	3 44	3 42
	30 47	30 45	4 43
	31 54	31 46	25 44
	32 57	32 47	26 45
	33 69	33 54	30 46
		34 57	31 47
		35 59	32 48
		36	33 53
		42	34 54
		43	35 57
			38 61
			39 66
			40 69
			41 70
No. 3	4 48	4 48	8 51
	25 53	25 53	9 52
	26 58	26 58	10 55
	38 61	38 61	11 60
	39 66	39 66	12 65
	40 70	40 70	13
	41	41	29

	Ward's method		Group average method		Centroid method
No. 4	6		8	55	14
	23		9	60	15
	24		10	65	
	27		11		
	56		12		
			13		
			29		
			51		
			52		
No. 5	8	51	14		16
	10	52	15		17
			1155		37
	12	65			49
	13				68
No. 6	9	68	16		18
	16		17		19
	17		37		20
	37		49		21
	49		67		22
	67		68		
No. 7	14		18		23
	15		19		
			20		
			21		
			22		
No. 8	18		23		28
	19		28		59
	20		59		
	21				
	22				
No. 9	28	63	62		62
	29	64	64		64
	59				
	60				
	62				
No. 10	34		63		63
	35				
	36				
	42				
	43				
	43				
	44				
	45				