4

METHODS OF CONSTRAINED CLASSIFICATION

A. D. Gordon

Department of Statistics, University of St. Andrews,
North Haugh, St. Andrews, Scotland.

SUMMARY

A review is presented of methods of classification which incorporate external constraints on the set of objects. It is convenient to consider a graph theory representation, in which each vertex represents an object, and the presence of an edge between two vertices indicates a contiguity constraint between the corresponding two objects. If the graph is such that every edge is a cut-edge, a range of methods of constrained classification is available. Several approximating algorithms are presented for situations involving more complicated contiguity constraints.

pages 161-171 in R. Tomassone (ed.)
Analyse de Données et Informatique,
I.N.R.I.A., Le Chesnay (1980).

1. INTRODUCTION

Classification methods [Cor71] are concerned with the analysis of a set of objects in order to establish whether there is any structure in the data; for example, do the objects fall into a number of distinct groups such that objects within a group are "similar" to one another and dissimilar to objects in other groups? In some instances, however, one has some external information on the objects which imposes constraints on the classification. The following two examples illustrate the types of constraints which can occur.

1. Pollen analysts take cores from lake beds and peat bogs, and extract small samples of sediment at specified intervals down the core. The sediment or peat has been accumulating over a long period of time, and will contain within it pollen grains deposited at various times in the past; it is hoped that the mixture of grains will give relevant information to aid reconstruction of the vegetation surrounding the coring site at the time that the grains were deposited. Each sample is treated chemically to separate the preserved pollen grains from the sediment matrix. A sample of the grains is then mounted in a transparent medium in a microscope slide, and each grain in the sample is assigned to its parent species. From the resulting set of pollen spectra, pollen analysts wish to trace the historical development of the vegetation "near" the site : periods of time when there was a stable surrounding vegetational community would be expected to be reflected by a group of similar *stratigraphically-neighbouring* spectra; for a fuller description of methods, assumptions and limitations, see [FIv75].

2. Soil scientists [WeB72a,b] study the properties of soil profiles at many different sites. It is convenient, for soil management purposes, to create "parcels" of land within which the variation in soil properties is small. Each parcel comprises a set of *contiguous* sites which have similar properties, and should be reasonably large and compact.

In both of these examples, there are external reasons (*i.e.* reasons not depending on the physical resemblances of the objects) for restricting the set of allowable classifications and insisting that each group must comprise *neighbouring* objects. It is convenient to represent the situation in graph-theoretic terms. Let each object (*e.g.* (1)pollen spectrum, (2)site at which soil properties have been measured) be represented by a vertex in a graph. The presence of an edge between two vertices indicates that the corresponding two objects are neighbours. For example, fig. 1a represents the graph corresponding to example 1 : it is conventional to label spectra down the core, with "younger" spectra being denoted by lower numbers. The graph corresponding to example 2 would be more complicated, and would depend on the location of the sampling sites and the area for which each site was regarded as representative : even if sites were sampled on a rectangular lattice, one would still have to decide whether a site in the centre of the region had four neighbours, or whether diagonal proximities should be allowed to increase this to eight neighbours.

In both of the above examples, the contiguity constraints were given by physical (geographical) proximities, but this is not the only type of constraint which can be envisaged, and other types of example will be given later in the text.

It is assumed throughout the paper that when the contiguity constraints are applied to the set of objects, a connected graph is obtained, *i.e.* there is a chain of edges linking every pair of vertices. If this condition does not hold, one would simply analyse separately the connected components of the graph.

The clustering problem can now be seen to be equivalent to seeking to find a cutset of the graph, *i.e.* to find a set of edges whose removal would cause the graph to fall apart into separate subgraphs. One would seek to find those particular subgraphs which contain objects which are "similar" to one another. This would give a partition of the set of objects into g, say, homogeneous groups of objects which respect the external constraints. One could find a partition into g groups for g = 2,3,...,n (where n denotes the number of objects in the data set); if these groups were hierarchically nested, the entire solution could be represented in a dendrogram, or tree diagram.

It is still necessary to give some consideration to the definition of the words "homogeneous" and "similar" used above. It is convenient to define a mathematical criterion assigning a value to each partition, and to seek that partition which has the optimal value of the criterion [Jar70]. For some criteria, *e.g.* the single link criterion, there are constructive algorithms which find the optimal partition [Sib73]. For other criteria, *e.g.* the total within-group sum of squares criterion, various approximating algorithms have to be used [GoH77]; there is a large number of different partitions [FoS66] and for even moderated-sized data sets it is not computationally feasible to examine all possible partitions in

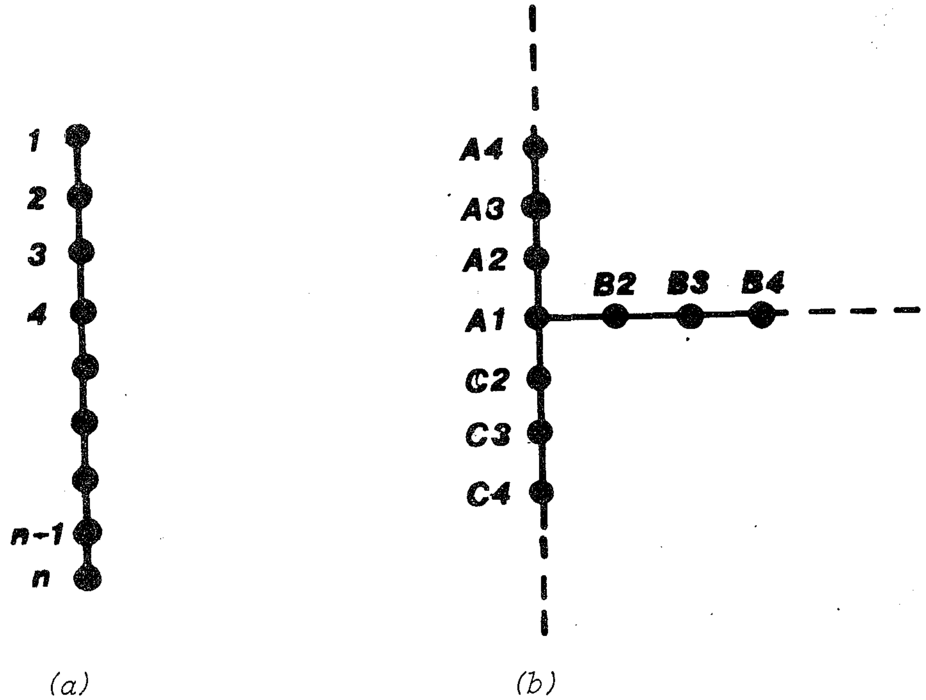

Fig. 1. Graph theory representation of objects subject to contiguity constraints: (a) stratigraphical pollen data ; (b) samples taken along three transects.

order to locate the optimal one. From a purely computational point of view, it would be advantageous to consider constrained classifications if the external constraints reduced the number of partitions which had to be examined in order to find the optimal one.

In the clustering methods to be described in this paper, one *imposes* a (constrained) partition or set of partitions on the set of objects, and there is the possibility that this has distorted the structure in the data. Two precautions, which are taken in many of the studies to be described, are :-

(i) to analyse the data using two or more different criteria of homogeneity, to see whether similar results are suggested; and

(ii) to undertake also an ordination study, in which the physical similarities between objects are represented by a low-dimensional geometrical representation of points, which can be studied in order to establish whether there is any group structure in the data.

2. *SIMPLER CONSTRAINED CLASSIFICATION PROBLEMS*

The bulk of this section will be devoted to methods of analysing pollen stratigraphical data, *i.e.* to constrained classification of data which can be represented by a linear graph (fig. 1a), but some more general problems for which the methods are relevant will also be indicated.

Let $\{x_{ik}\ (k=1,\ldots,p)\}$ represent the p measurements available on the i^{th} object $(i=1,\ldots,n)$; in example 1, x_{ik} would denote the proportion of the i^{th} spectrum which is composed of the k^{th} species of pollen. In order to investigate the relationship between the spectra, one could analyse them using a standard classification procedure without imposing any constraints based on the stratigraphical ordering of the spectra [Ada70,DaW70,MoG71,Ada74,DBS75]. While it is often of interest to note the similarity of spectra which are separated by dissimilar spectra, the manner in which the data arose suggests that, in the absence of evidence to suggest mass disturbance of the sediment, one should insist that groups be formed only from sets of neighbouring spectra.

Before describing the implementation of a linear constraint into classifications, we note one other class of methods which have been used to analyse such stratigraphical data : these obtained a measure of the difference between pairs of neighbouring readings in an attempt to detect sudden changes [Ker70,YaR72,RiY78]. This approach is likely to be considerably affected by the common practice of taking further samples close to suspected group boundaries, and while this effect can be reduced by considering a few more objects on either side [YaR72,Web73], it seems preferable to use the entire data set in the investigation of the location of the group boundaries.

2.1. *HIERARCHICAL CONSTRAINED CLASSIFICATION*

The set of n objects displayed in fig. 1a can be divided into g groups of contiguous objects by removing (g-1) edges from the graph, or alternatively by placing (g-1) "markers" in some of the (n-1) gaps between pairs of neighbouring objects. A marker will be labelled "i" if it fits into the gap between objects i and (i+1), indicating the presence of a group boundary between these two objects. The number of ways in which (g-1) markers can be located, and hence the number of possible partitions of n objects into g groups of contiguous objects, is

$$(n-1)!/\{(g-1)!(n-g)!\};$$

note the great reduction in possible partitions from the unconstrained case.

Each partition will have an associated measure of homogeneity : this paper will concentrate on the within-group sum of squares criterion, *i.e.* for a specified number of groups, g, we will seek the particular partition which minimises

$$S_g \equiv \sum_{i=1}^{n} \sum_{k=1}^{p} (x_{ik} - z_{ik})^2 \qquad (1)$$

where $\{z_{ik}\ (k=1,\ldots,p)\}$ are the coordinates of the centroid of the group to which the i^{th} object belongs. No more than (gp) of the z_{ik}'s will take distinct values, corresponding to the g group centroids, although this requirement will be relaxed in subsection 2.3.

One can seek the optimal partition into g groups for a range of values of g. With the sum of squares criterion, the solutions obtained need not be hierarchically-nested [FVN71], *e.g.* the optimal three groups need not be obtainable from the optimal two groups by subdivision of one of these groups. Pollen analysts appear to prefer an hierarchical representation which allows them to investigate the relationship between various groups and subgroups. One approach might be to restrict one's attention to clustering criteria which can be guaranteed to produce an hierarchical solution, such as the constrained single link criterion [GoB72].

Alternatively, one could use a sum of squares criterion and impose an hierarchy on the data by use of a divisive [Gi170,GoB72] or agglomerative [MAP77] algorithm.

The divisive algorithm proceeds as follows. Obtain the optimal division of the data into two groups of neighbouring objects, *i.e.* find the position for a single marker which leads to minimum total within-group sum of squares. Now optimally divide one of these two groups, choosing that particular division which leads to maximum decrease in the sum of squares. This is equivalent to seeking the optimal position for a second marker, given that the first marker must remain fixed in its original position. It is clear that the three groups obtained at this stage need not be the overall optimal three groups which would have been obtained if the optimal positions had been chosen simultaneously for the two markers. The algorithm continues by successive division of existing groups, always choosing that particular division which leads to the maximum decrease possible *at that stage* in the total within-group sum of squares. The agglomerative algorithm proceeds in the converse direction, by successive amalgamations of existing groups. When the data have been divided into g groups, these groups may be regarded as an approximation to the overall optimal g groups : precautions such as those mentioned at the end of section 1 should be taken so as to satisfy oneself that this hierarchical representation does not greatly distort the structure in the data.

2.2. DYNAMIC PROGRAMMING ALGORITHM

Obtaining an hierarchical classification has seemed to be of less importance in geological and geophysical applications [HaM73], where investigators have wanted to find the overall optimal constrained sum of squares partition. Although $(n-1)!/\{(g-1)!(n-g)!\}$ is a considerable reduction on the total number of unconstrained partitions, it becomes very time-consuming to examine all possible partitions for larger data sets. Fortunately, a global search is not necessary, as the optimal solution can be built up recursively for a range of criteria of variability [Fis58,HaM73,Haw76a].

Let $s(i,j)$ denote the within-group sum of squares of objects i to j inclusive, and let $t(g,k)$ denote the total within-group sum of squares when objects 1 to k are optimally divided into g groups.We wish to find $t(g,n)$ for $2 \le g \le n$, together with the corresponding markers specifying the optimal partition. The solution is built up recursively, evaluating $\{t(g,k),\ k=g,g+1,\dots,n;\ g=1,2,\dots,n\}$ by use of the following formulae :-

$$t(1,k) = s(1,k) \qquad (1 \le k \le n)$$

$$t(g,k) = \underset{g-1 \le i \le k-1}{\text{minimum}} \{t(g-1,i) + s(i+1,k)\} \qquad (g \le k \le n;\ 2 \le g \le n).$$

The second formula considers dividing the first k objects into two classes, the first class containing (g-1) groups and the second class containing one group of objects. This is equivalent to placing the last marker at position i, and the algorithm finds the optimal value of i. The complete set of (g-1) markers is obtained using a trace-back procedure.

Several points can be made about this approach.
1. It can be used with measures of variability other than the sum of squares; one requires only that the variability of the entire data set is the sum of the variabilities of the constituent groups.
2. One can impose further constraints on the groups to be found; for example, one can insist that they contain at least a pre-specified number of objects [HaM73]. One could also concentrate solely on identifying "peak" groups of objects with very high readings [BeW77].
3. Sometimes, one is concerned not only with identifying groups of objects, but also with attempting to model relationships or developments within groups. For stratigraphical data, the linear ordering specifies an independent time or depth

variable, and one is led to consider piecewise regression models [MGC70,Haw76b, WaW76], or splines [Hay74] if a measure of continuity is required between curves in successive sections.

2.3. *VARIABLE BARRIERS ALGORITHM*

Vegetational conditions do not change instantaneously and, given pollen analysts' practice of taking further samples close to suspected group boundaries, it is common to have some intermediate samples between groups of homogeneous spectra corresponding to stable vegetational conditions. D.Walker [Wal66] has stressed that it is appropriate to distinguish between two different types of stratigraphical interval detectable in pollen sequences : (a) intervals of relatively stable pollen composition; (b) intervals of abrupt and/or systematic change in pollen composition interspaced between the stable intervals.

If there is a large number of transitional objects, a global measure of variability such as the sum of squares criterion will tend to place group boundaries in the middle of the second type of interval. One would like a method of analysis which distinguished between (a)the nuclei of groups, and (b)objects which are transitional between the two groups. The method to be described in this subsection [Gor73] seeks to do this; its rationale has much in common with ideas of "fuzzy" clustering [Bez74,Boc79].

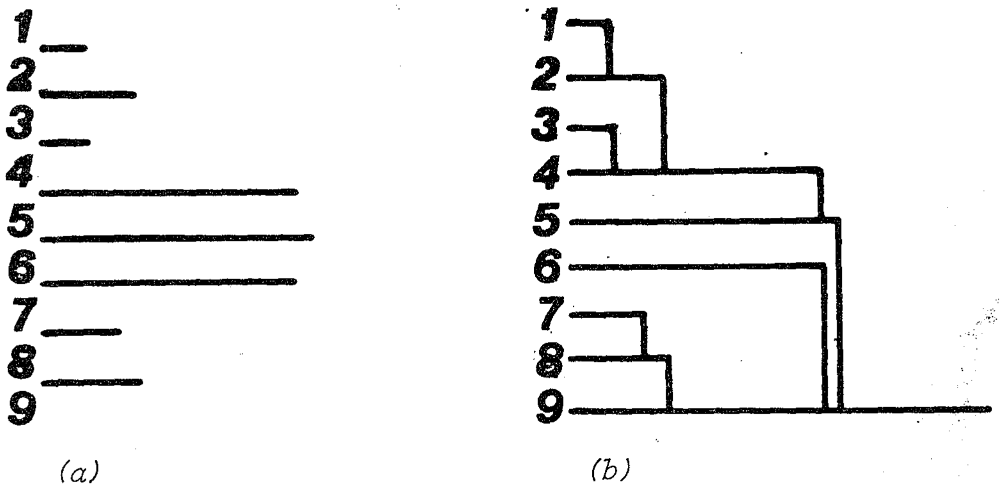

Fig. 2. Illustration of the variable barriers approach : (a) presence of barriers between neighbouring objects ; (b) representing the solution as a constrained hierarchical classification.

Consider the linearly-ordered objects (1,2,...,n). Instead of inserting markers in some of the (n-1) gaps between neighbouring objects, place a barrier of height between 0 and 1 in each of the gaps, as shown in fig. 2a. Let b_m $(1 \leq m \leq n-1)$ denote the height of the barrier placed between objects m and (m+1); by convention, $b_o = 1 = b_n$. Consider minimising the criterion (1), but with the difference that $\{z^o_{jk}$ $(k = 1,...,p)\}$ is now the *local* centroid of the j^{th} point : instead of being the average of x_{ik}'s for i belonging to the same group as j, z_{ik} is now a *weighted* average of x_{ik}'s for i lying between certain limits. The limits are found as follows : from j we can "reach" any object if this involves "jumping over" barriers of total height less than 1; the x_{ik}'s which can be "reached" from j contribute towards the evaluation of z_{jk}.

Formally, let y_{ji} denote the "weight" which the i^{th} object contributes to the evaluation of the j^{th} object's local centroid,

i.e. $$z_{jk} = \sum_{i=1}^{n} (y_{ji}.x_{ik}) / \sum_{i=1}^{n} y_{ji} \qquad (1 \leq j \leq n \; ; \; 1 \leq k \leq p)$$

where $$y_{ji} = \begin{cases} \max\left(0, 1 - \sum_{m=j}^{i-1} b_m\right) & \text{if } i > j \\ 1 & \text{if } i = j \\ \max\left(0, 1 - \sum_{m=i}^{j-1} b_m\right) & \text{if } i < j . \end{cases}$$

The aim is then to minimise

$$S_g(\underline{b}) \equiv \sum_{i=1}^{n} \sum_{k=1}^{p} \{x_{ik} - z_{ik}(\underline{b})\}^2 \qquad (2)$$

i.e., to find the particular values of the barrier heights $\underline{b} \equiv (b_1, b_2, \ldots, b_{n-1})$ which lead to the minimisation of (2), subject to the constraints

$$0 \leq b_m \leq 1 \quad (1 \leq m \leq n-1), \text{ and } \sum_{m=1}^{n-1} b_m = g-1 \qquad (3)$$

If the second constraint in (3) were not imposed, a solution with S = 0 would be obtained from $\underline{b} = \underline{1}$.

It can be seen that restricting the barriers to be all of height either 0 or 1 would reduce the problem to the one considered in the previous two subsections : the z_{ik}'s would then be unweighted averages of all objects lying between adjacent barriers of height 1 (or markers), and, from (3), there would be (g-1) of these markers, specifying the g groups. Minimising the sum of squares under the more flexible conditions described in this subsection will ensure that dissimilar objects will be separated by high barriers, but instead of barriers of height 1 abruptly marking group boundaries, it is possible for there to be a configuration of lower barriers. Transitional objects should appear between adjacent higher barriers, *e.g.* if the configuration shown in fig. 2a gives the optimal solution, one could identify objects 5 and 6 as transitional between the homogeneous groups (1 to 4) and (7 to 9).

An alternative interpretation of the above procedure is to regard it as defining a measure of dissimilarity $d^*_{ij} = (1-y_{ij})$ between each pair of objects. This "dissimilarity" depends not only on the physical difference between objects i and j but also on their relationships with neighbouring objects. Analysis of the matrix of dissimilarities (d^*_{ij}) using the single link method will yield precisely the same results as the consideration of barrier heights given above: this is because b_m, the height of the barrier between objects m and (m+1), is the "branch" of the dendrogram associated with object m ($1 \leq m \leq n-1$; recall also that $b_n = 1$). The single link dendrogram can be constructed from the barriers given in fig. 2a as follows : move each barrier up "half a notch" so that the barrier of height b_m is opposite object m ($1 \leq m \leq n$). From the end of the first (n-1) barriers, drop a perpendicular until it hits some barrier below it; this yields the single link dendrogram, as is illustrated in fig. 2b. Transitional objects will now be represented as late additions to groups visualised as growing in an agglomerative manner.

The problem of obtaining the optimal $\underline{b}$ has not been solved analytically; an approximating function-minimisation algorithm has been used. The results do not appear to be critically dependent on the value of (g-1), the sum of the barrier heights; broadly similar results (with a different scale) were obtained over a range of values of g. Examples of the use of the algorithm are given in [Gor73, GoB74,BiM78].

2.4. *GENERALISATIONS*

The methods described in this section were developed for the analysis of data on which a linear ordering had been imposed, as represented schematically by the linear graph shown in fig. 1a. They can, however, be used to analyse data subject to more complex contiguity constraints. The necessary condition for applicability is that the corresponding graph representation shall be such that *every* edge is a cut-edge, *i.e.* removing *any* edge will section the graph into two connected subgraphs. Another example of such a data set is represented by the graph in fig. 1b: this depicts a set of three transects (A, B and C) radiating from the centre of a bog; at regular intervals along each transect surface pollen samples were taken, these corresponding to the objects to be analysed. Constrained classifications of these data have been carried out,using an approximating divisive algorithm [CaG78] and a dynamic programming algorithm [Cam78].

The contiguity constraints can be imposed for reasons other than objects being physical neighbours. For example, one might wish to classify objects for which a family tree was known, taking into account the family structure. A slightly different problem was considered by T.Calinski and J.Harabasz [CaH74] : in seeking the optimal sum of squares partition of n objects into g groups in the *unconstrained* classification problem, they aimed to reduce the number of partitions to be examined by considering only those which could be obtained by deleting (g-1) edges of the minimum spanning tree.

3. MORE COMPLICATED CONSTRAINED CLASSIFICATION PROBLEMS

This section considers problems which can not be represented by graphs with the property that every edge is a cut-edge. A fair amount of related work has been done in different disciplines, some of which might be relevant to specific constrained classification problems, and it is convenient to begin by summarising some of this work.

Engineers and computer scientists have wanted to arrange various types of linked components into a smaller number of modules (of limited size) such that there are few inter-module connections. This has led to their seeking efficient ways of partitioning graphs [Ker71,Luk75,ChB76] and trees [Luk74] so as to ensure that (i) the sum of vertex weights in any subgraph is less than a pre-specified value, (ii)the sum of the values of the edges joining the different subgraphs is a minimum.

Geographers and social scientists have wanted to split a country up into regions for administrative or electoral purposes [Mil67,HOR72,Tay73,Ope77] : one aim has been to obtain compact regions containing comparable populations. Although they seek to optimise different criteria, they are still concerned with efficient ways of obtaining optimal partitions into groups of contiguous objects (enumeration districts), and some of their algorithms can be applied to the general constrained classification problem.

One approach to problems of classification in which the constraints are specified by geographical location has been to ignore the constraints in the initial analysis [GoY73,Bir76,NaG77]. Another, rather dubious, approach has been to build the geographical coordinates or distances apart, suitably weighted, into the pairwise measure of dissimilarity, which is then analysed by a standard unconstrained method of classification [WeB72b,DaH74].

For many general graphs, it would be extremely time-consuming (and is often not feasible with present computing facilities) to examine all possible divisions into g subgraphs in order to find the corresponding partition with optimal value of the specified criterion of homogeneity. This operation is more straightforward for some criteria, such as the single link criterion for which there is a constructive algorithm which will obtain optimal partitions; for other criteria, however, one at present has to be content with approximating algorithms. Two such approximating algorithms leading to constrained classifications are described below.

1. Agglomerative algorithm [WeB72b] : This resembles agglomerative algorithms in unconstrained classification in that one starts with n single-member groups and successively amalgamates a pair of groups which give minimum increase in the criterion of variability (*e.g.* total within-group sum of squares). However, there is the added condition that two groups may only be amalgamated if they are contiguous. This procedure yields an hierarchical classification : when there are g groups, these can be regarded as an approximation to the optimal g groups.

2. Iterative relocation algorithm [Ope77] : One attempts to improve an initial partition into g groups of contiguous objects by moving objects from one group to another; an object can only be moved from group G_1 to group G_2 if the object is contiguous to G_2, and if moving the object would not result in G_1's splitting into several unconnected subgraphs. Relocation is continued until the criterion of variability cannot be reduced further by any movement of objects.

This review is completed with an account of some work by L.P.Lefkovitch [Lef78], originally introduced to consider the unconstrained classification problem. A set of n objects has 2^n-1 non-empty subsets : these can be stored in the $n \times (2^n-1)$ matrix $A \equiv (a_{ij})$,

$$\text{where } a_{ij} = \begin{cases} 1 \text{ if object } i \text{ is present in the } j^{th} \text{ subset } (1 \le i \le n;\ 1 \le j \le 2^n-1) \\ 0 \text{ otherwise.} \end{cases}$$

A partition (or *covering*, into overlapping subsets) is specified by selecting some of these subsets :

$$\text{let } x_j = \begin{cases} 1 \text{ if the } j^{th} \text{ subset belongs to a partition (covering)} \\ 0 \text{ otherwise ;} \end{cases}$$

and let c_j denote the cost of including the j^{th} subset in the partition (covering). One then seeks to optimise some function of $\underline{x}$ and $\underline{c}$ subject to

$$A\underline{x} = \underline{1} \quad \text{for a partition; or} \quad A\underline{x} \ge \underline{1} \text{ for a covering.}$$

Of course, it is not usually feasible to examine all 2^n-1 possible subsets, and Lefkovitch restricts his attention to a smaller number which satisfy certain extra conditions, *e.g.* if three or more objects belong to a group, it seems reasonable that any object belonging to their convex hull in attribute space should also belong to the group [Dag66]. As Lefkovitch notes [Lef78], this idea can be extended to constrained classification problems : a subset will be considered as a possible group if its constituent members not only (1) satisfy some conditions on their dissimilarities, but also (2) satisfy other conditions based on external criteria, *e.g.* geographical distance apart, time ordering; an example where both of these latter constraints are present is the data set analysed by P.M.North [Nor77].

Many problems in data analysis can be formulated in terms of the classification of a set of objects where there are external constraints on the class of permissible solutions. Methods of constrained classification merit further investigation.

REFERENCES

Ada70 Adam,D.P. "Some palynological applications of multivariate statistics", Ph.D. thesis, University of Arizona, 1970.

Ada74 Adam,D.P. "Palynological applications of principal component and cluster analyses", J. Res. U.S. Geol. Survey 2, 1974, 727-741.

Bez74 Bezdek,J.C. "Numerical taxonomy with fuzzy sets", J. Math. Biology 1, 1974, 57-71.

BeW77 Bement,T.R. & Waterman,M.S. "Locating maximum variance segments in sequential data", Math. Geology 9, 1977, 55-61.
BiM78 Birks,H.H. & Mathewes,R.W. "Studies in the vegetational history of Scotland V", New Phytologist 80, 1978, 455-484.
Bir76 Birks,H.J.B. "The distribution of European pteridophytes : a numerical analysis", New Phytologist 77, 1976, 257-287.
Boc79 Bock,H.-H. "Fuzzy clustering procedures", in this volume.
CaG78 Caseldine,C.J. & Gordon,A.D. "Numerical analysis of surface pollen spectra from Bankhead Moss, Fife", New Phytologist 80, 1978, 435-453.
CaH74 Calinski,T. & Harabasz,J. "A dendritic method for cluster analysis", Comm. Statist. 3, 1974, 1-27.
Cam78 Campbell,M.M. "Classification of some pollen data subject to contiguity constraints", Honours dissertation, University of St. Andrews, 1978.
ChB76 Christofides,N. & Brooker,P. "The optimal partitioning of graphs", SIAM J. App. Math. 30, 1976, 55-69.
Cor71 Cormack,R.M. "A review of classification", J. R. Statist. Soc. A 134, 1971, 321-367.
Dag66 Dagnelie,P. "A propos des différentes méthodes de classification numérique", Revue de Statistique Appliquée 14, 1966, 55-75.
DaH74 Dahmström,P. & Hagnell,M. "The formation of strata using cluster analysis", Research report 1974:4, Department of Statistics, University of Lund, 1974.
DaW70 Dale,M.B. & Walker,D. "Information analysis of pollen diagrams I", Pollen et Spores 12, 1970, 21-37.
DBS75 Davis,R.B., Broadstreet,T.E., Stuckenrath,R. & Borns,H.W. "Vegetation and associated environments during the past 14,000 years near Moulton pond, Maine", Quaternary Research 5, 1975, 435-465.
Fis58 Fisher,W.D. "On grouping for maximum homogeneity", J. Amer. Statist. Assoc. 53, 1958, 789-798.
FIv75 Faegri,K. & Iversen,J. Textbook of pollen analysis, 3rd edition, Blackwell, Oxford etc., 1975.
FoS66 Fortier,J.J. & Solomon,H. "Clustering procedures", in : Krishnaiah,P.R. (editor), Proc. Symp. Multivariate Analysis, Academic Press, New York, 1966, 493-506.
FVN71 Fisher,L. & Van Ness,J.W. "Admissible clustering procedures", Biometrika 58, 1971, 91-104.
Gil70 Gill,D. "Application of a statistical zonation method to reservoir evaluation and digitized-log analysis", Am. Assoc. Petroleum Geologists Bull. 54, 1970, 719-729.
GoB72 Gordon,A.D. & Birks,H.J.B. "Numerical methods in Quaternary palaeoecology I Zonation of pollen diagrams", New Phytologist 71, 1972, 961-979.
GoB74 Gordon,A.D. & Birks,H.J.B. "Numerical methods in Quaternary palaeoecology II Comparison of pollen diagrams", New Phytologist 73, 1974, 221-249.
GoH77 Gordon,A.D. & Henderson,J.T. "An algorithm for Euclidean sum of squares classification", Biometrics 33, 1977, 355-362.
Gor73 Gordon,A.D. "Classification in the presence of constraints", Biometrics 29, 1973, 821-827.
GoY73 Golder,P.A. & Yeomans,K.A. "The use of cluster analysis for stratification", Applied Statistics 22, 1973, 213-219.
HaM73 Hawkins,D.M. & Merriam,D.F. "Optimal zonation of digitized sequential data", Math. Geology 5, 1973, 389-395.
Haw76a Hawkins,D.M. "Fortran IV program to segment multivariate sequences of data", Computers & Geosciences 1, 1976, 339-351.
Haw76b Hawkins,D.M. "Point estimation of the parameters of piecewise regression models", Applied Statistics 25, 1976, 51-57.
Hay74 Hayes,J.G. "Numerical methods for curve and surface fitting", Bull. Inst. Math. Applics. 10, 1974, 144-152.
HOR72 Helbig,R.E., Orr,P.K. & Roediger, R.R. "Political redistricting by computer", Comm. ACM 15, 1972, 735-741.

Jar70 Jardine,N. "Algorithms, methods and models in the simplification of complex data", Computer J. 13, 1970, 116-117.

Ker70 Kershaw,A.P. "A pollen diagram from Lake Euramoo, North East Queensland, Australia", New Phytologist 69, 1970, 785-805.

Ker71 Kernighan,B.W. "Optimal sequential partition of graphs", J. Assoc. Comp. Mach. 18, 1971, 34-40.

Lef78 Lefkovitch,L.P. "Cluster generation and grouping using mathematical programming", Math. Biosciences 41, 1978, 91-110.

Luk74 Lukes,J.A. "Efficient algorithm for the partitioning of trees", IBM Jour. Res. Develop. 18, 1974, 217-224.

Luk75 Lukes,J.A. "Combinatorial solution to the partitioning of general graphs", IBM Jour. Res. Develop. 19, 1975, 170-180.

MAP77 Mehringer,P.J., Arno,S.F. & Petersen,K.L. "Postglacial history of Lost Trail Pass Bog, Bitterfoot Mountains, Montana", Arctic and Alpine Research 9, 1977, 345-368.

MGC70 McGee,V.E. & Carleton,W.T. "Piecewise regression", J. Amer. Statist. Assoc. 65, 1970, 1109-1124.

Mil67 Mills,G. "The determination of local government electoral boundaries", Op. Res. Quarterly 18, 1967, 243-255.

MoG71 Mosimann,J.E. & Greenstreet,R.L. "Representation-insensitive methods for paleoecological pollen studies", in : Patil,G.P., Pielou,E.C. & Waters,W.E. (editors), Statistical Ecology, volume I Spatial Patterns and Statistical Distributions, Pennsylvania State University Press, 1971, 23-58.

NaG77 Narendra,P.M. & Goldberg,M. "A nonparametric clustering scheme for LANDSAT", Pattern Recognition 9, 1977, 207-215.

Nor77 North,P.M. "A novel clustering method for estimating numbers of bird territories", Applied Statistics 26, 1977, 149-155.

Ope77 Openshaw,S. "A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling", Trans. Inst. Brit. Geog. 2(NS), 1977, 459-472.

RiY78 Ritchie,J.C. & Yarranton,G.A. "The late Quaternary history of the Boreal forest of central Canada, based on standard pollen stratigraphy and principal components analysis", J. Ecology 66, 1978, 199-212.

Sib73 Sibson,R. "SLINK : an optimally efficient algorithm for the single-link cluster method", Computer J. 16, 1973, 30-34.

Tay73 Taylor,P.J. "Some implications of the spatial organisation of elections", Trans. Inst. Brit. Geog. 60, 1973, 121-136.

Wal66 Walker,D. "The Late Quaternary history of the Cumberland lowland", Phil. Trans. Roy. Soc. Lond. B 251, 1966, 1-210.

WaW78 Walker,D. & Wilson,S.R. "A statistical alternative to the zoning of pollen diagrams", J. Biogeography 5, 1978, 1-21.

WeB72a Webster,R. & Burrough,P.A. "Computer-based soil mapping of small areas from sample data I Multivariate classification and ordination", J. Soil Sci. 23, 1972, 210-221.

WeB72b Webster,R. & Burrough,P.A. "Computer-based soil mapping of small areas from sample data II Classification smoothing", J. Soil Sci. 23, 1972, 222-234.

Web73 Webster,R. "Automatic soil-boundary location from transect data", Math. Geology 5, 1973, 27-37.

YaR72 Yarranton,G.A. & Ritchie,J.C. "Sequential correlations as an aid in placing pollen zone boundaries", Pollen et Spores 14, 1972, 213-223.