Algorithmica (1994) 11: 93–103

Algorithmica

Dynamic Programming and Convex Clustering[1]

Vladimir Batagelj,[2] Simona Korenjak-Černe,[2] and Sandi Klavžar[3]

Abstract. In this paper the notion of convexity of clusterings for the given ordering of units is introduced. In the case when at least one (optimal) solution of the clustering problem is convex, dynamic programming leads to a polynomial algorithm with complexity $O(kn^3)$. We prove that, for several criterion functions, convex optimal clusterings exist when dissimilarity is pyramidal for a given ordering of units.

Key Words. Clustering algorithms, Dynamic programming, Pyramidal dissimilarity, Convex clustering, Ordering constraint.

1. Introduction. This paper is inspired by the paper by Brucker [3] and, for Theorem 3, by the paper by Hwang *et al.* [9]. Its main contribution is the introduction of the notion of convex clustering for a given ordering of units. This allows us to extend and generalize the results in the papers mentioned from the case where units are represented by real numbers to arbitrary sets of units equipped by a given ordering; for example, units ordered with respect to the time scale.

Let us start with the formal setting of the clustering problem. We use the notations from [4]:

$\mathscr{E}$ is the space of units.
E is the finite set of units, $n = \text{card}(E)$.
$C \subseteq E$ is a cluster, $C \neq \emptyset$.
$\mathbf{C} = \{C_i\}$ is a clustering.
Φ is a set of feasible clusterings.
$P: \Phi \to \mathbb{R}_0^+$ is a criterion function.

Generally the clusters of a clustering $\mathbf{C}$ need not be pairwise disjoint; yet clustering theory and practice deal mainly with clusterings which are partitions of E. We denote the set of all partitions of E into k classes (clusters) by $P_k(E)$. With these notions we can express the *clustering problem* (Φ, P) as follows:

Determine the clustering $\mathbf{C}^* \in \Phi$, for which

$$P(\mathbf{C}^*) = \min_{\mathbf{C} \in \Phi} P(\mathbf{C}).$$

[1] This research was supported in part by the Research Council of Slovenia.
[2] Department of Mathematics, University of Ljubljana, Jadranska 19, 61111 Ljubljana, Slovenia.
[3] Pedagoška fakulteta Maribor, University of Maribor, Koroška c. 160, 62000 Maribor, Slovenia.

Received January 7, 1991; revised June 17, 1991. Communicated by Jean-Daniel Boissonnat.

Because the set of units E is finite, the set of feasible clusterings is also finite. Therefore, the set $\text{Min}(\Phi, P)$ of all solutions of the problem (optimal clusterings) is not empty. We denote the value of a criterion function for an optimal clustering by $\min(\Phi, P)$.

When a linear ordering $\preceq$ is given on the set E, we denote the set of feasible clusterings by $(\Phi, \preceq)$.

2. Criterion Functions. Joining individual units into a cluster C we make a certain "error," we create a certain "tension" among them—we denote this quantity by $p(C)$. The criterion function $P(\mathbf{C})$ combines these "partial/local errors" into a "global error." Usually it takes the form:

$$S: \qquad P(\mathbf{C}) = \sum_{C \in \mathbf{C}} p(C)$$

or

$$M: \qquad P(\mathbf{C}) = \max_{C \in \mathbf{C}} p(C),$$

which can be unified and generalized in the following way:

Let $(\mathbb{R}, \oplus, <)$ be an ordered abelian monoid, then

$$\oplus: \qquad P(\mathbf{C}) = \bigoplus_{C \in \mathbf{C}} p(C).$$

The "cluster-error" $p(C)$ usually has the properties:

$$p(C) \geq 0 \qquad \text{and,} \qquad \forall X \in E, \quad p(\{X\}) = 0.$$

In what follows we assume that these properties of $p(C)$ hold.

To express the "cluster-error" $p(C)$ we define a *dissimilarity* on the space of units:

$$d: \mathscr{E} \times \mathscr{E} \to \mathbb{R}_0^+$$

for which we require,

$$\forall X \in \mathscr{E}, \quad d(X, X) = 0 \qquad \text{and,} \qquad \forall X, Y \in \mathscr{E}, \quad d(X, Y) = d(Y, X).$$

Now we can define several "cluster-error" functions:

$$S: \qquad p(C) = \sum_{X, Y \in C} w(X) \cdot w(Y) \cdot d(X, Y),$$

$$\bar{S}: \qquad p(C) = \frac{1}{w(C)} \sum_{X, Y \in C} w(X) \cdot w(Y) \cdot d(X, Y),$$

where $w\colon \mathscr{E} \to \mathbb{R}^+$ is a *weight* on units, which is extended to clusters by

$$w(\{X\}) = w(X), \qquad X \in \mathscr{E},$$
$$w(C_1 \cup C_2) = w(C_1) + w(C_2), \qquad C_1 \cap C_2 = \varnothing.$$

Often $w(X) = 1$ holds for each $X \in \mathscr{E}$.

$$M\colon \qquad p(C) = \max_{X, Y \in C} d(X, Y).$$

Several other types of criterion functions were proposed in the literature.

We use labels in front of the names of (partial) criterion functions to denote types of criterion functions. For example,

$$SM\colon \qquad P(\mathbf{C}) = \sum_{C \in \mathbf{C}} \max_{X, Y \in C} d(X, Y).$$

Another form of "cluster-error" function, which is frequently used in practice, is based on the notion of the leader or representative of the cluster:

$$R\colon \qquad p(C) = \min_{L \in F} \sum_{X \in C} w(X) \cdot d(X, L),$$

where $F \subseteq \Psi$ is the set of *representatives* and Ψ is the space of representatives. The element $\bar{C} \in F$, which minimizes the expression on the right, is called the *representative* of cluster C. It is not always uniquely determined. In the following we assume that $\Psi \subseteq \mathscr{E}$.

Let us denote the sum from the definition of "cluster-error" function R by

$$q(C, L) = \sum_{X \in C} w(X) \cdot d(X, L),$$

then we have

$$q(C, L) \geq q(C, \bar{C}) = p(C).$$

A standard example of such a criterion function is variance or inertia: for $E \subset \mathbb{R}^m$, $F = \mathbb{R}^m$, and

$$d(X, L) = \sum_{i=1}^{m} (x_i - l_i)^2,$$

a uniquely determined representative $\bar{C}$ exists—*centre of gravity*. In this case the criterion function SR is called the *Ward criterion function.*

3. Convex Clusterings. The clustering $\mathbf{C} \in (\Phi, \preceq)$ is *convex* for the linear ordering $\preceq$ if and only if

$$\forall C_u, C_v \in \mathbf{C}, \exists Z \in E: (\forall X \in C_u, Y \in C_v: (X \preceq Z \wedge Z \prec Y) \\ \vee \forall X \in C_u, Y \in C_v: (Y \preceq Z \wedge Z \prec X)).$$

It follows immediately from the definition that each convex clustering consists of intervals $C_i = [X_p, X_q] = \{X \in E: X_p \preceq X \wedge X \preceq X_q\}$.

Let $\preceq$ be a linear ordering in the set of units E. The dissimilarity d is *compatible* with $\preceq$ if and only if

$$X \preceq Y \preceq Z \quad \Rightarrow \quad \max(d(X, Y), d(Y, Z)) \leq d(X, Z).$$

A dissimilarity d is *pyramidal* [2] over E if and only if a linear order $\preceq$ on E exists such that d is compatible with $\preceq$.

An example of the pyramidal dissimilarity for every $(E, \leq)$, $E \subseteq \mathbb{R}$, is

$$d(X, Y) = f(|X - Y|),$$

where f is an increasing function satisfying the additional condition $f(0) = 0$.

LEMMA 1. *Suppose that the dissimilarity d is pyramidal for $(E, \preceq)$. Then, for each $C \subseteq E$, the equality*

$$\max_{X, Y \in C} d(X, Y) = d(\min(C), \max(C)),$$

where $\min(C)$ *and* $\max(C)$ *are the least and the greatest element of C, holds.*

PROOF. Suppose $d(X_1, Y_1) = \max d(X, Y)$ over all $X, Y \in C$, and let $X_2 = \min(C)$, $Y_2 = \max(C)$. Then $X_2 \preceq Y_1 \preceq Y_2$ and, by pyramidality of d, $d(X_2, Y_2) \geq d(X_2, Y_1)$. Again by pyramidality of d and maximality of $d(X_1, Y_1)$, the equality $d(X_2, Y_1) = d(X_1, Y_1)$ follows. Therefore, $d(X_1, Y_1) = d(X_2, Y_2)$, proving the lemma. □

The role of the pyramidality can be seen from the following theorem:

THEOREM 1. *Let d be pyramidal for $(E, \preceq)$ and let the operation $\oplus$ be compatible with the relation $\leq$, i.e.,*

$$a \leq b \quad \Rightarrow \quad a \oplus c \leq b \oplus c.$$

Then the problem $((P_k, \preceq), \oplus M)$ has a convex solution.

PROOF. Suppose that $\mathbf{C} \in \text{Min}((P_k, \preceq), \oplus M)$ is not convex. We can transform this clustering to a convex optimal clustering. To show this let X be the smallest unit with respect to $\preceq$ for which a cluster $C_u \in \mathbf{C}$ exists such that $X \notin C_u$ and

$$\min(C_u) \prec X \prec \max(C_u).$$

Because $\mathbf{C}$ is not convex such a unit always exists. Note that $X = \min(C_v)$. Define a new clustering

$$\mathbf{C}' = (\mathbf{C} \setminus \{C_u, C_v\}) \cup \{C'_u, C'_v\},$$

where the clusters C'_u and C'_v are determined as follows:

(a) $\max(C_u) \prec \max(C_v)$; in this case we define

$$C'_u = [\min(C_u), X], \qquad C'_v = (C_u \cup C_v) \setminus C'_u.$$

Because $\min(C_u) \prec X = \min(C_v) \prec \max(C_u) \prec \max(C_v)$ we have, using Lemma 1,

$$p(C'_u) = d(\min(C_u), X) \leq d(\min(C_u), \max(C_u)) = p(C_u),$$

$$p(C'_v) = d(\min(C'_v), \max(C'_v)) \leq d(\min(C_v), \max(C_v)) = p(C_v).$$

(b) $\max(C_u) \succeq \max(C_v)$; now we set

$$C'_u = \{\min(C_u)\}, \qquad C'_v = (C_u \cup C_v) \setminus C'_u,$$

and we obtain, again using Lemma 1,

$$p(C'_u) = 0 \leq p(C_v),$$

$$p(C'_v) = d(\min(C'_v), \max(C'_v)) \leq d(\min(C_u), \max(C_u)) = p(C_u).$$

In both cases, because $\oplus$ is compatible with $\leq$, we have

$$p(C'_u) \oplus p(C'_v) \leq p(C_u) \oplus p(C_v).$$

Therefore,

$$\begin{aligned} P(\mathbf{C}') &= p(C'_u) \oplus p(C'_v) \oplus \bigoplus_{C \in \mathbf{C}' \setminus \{C'_u, C'_v\}} p(C) \\ &\leq p(C_u) \oplus p(C_v) \oplus \bigoplus_{C \in \mathbf{C} \setminus \{C_u, C_v\}} p(C) = \bigoplus_{C \in \mathbf{C}} p(C) = P(\mathbf{C}) \end{aligned}$$

is an optimal clustering, which has a longer initial interval covered by convex clusters. Because the set E is finite we obtain a convex optimal clustering in a finite number of steps by repeatedly applying this procedure. □

THEOREM 2. *The problem $((P_k, \preceq), \oplus S)$ does not always have a convex solution.*

PROOF. Consider the following (counter)example:

$$E = \{X, Y_1, Y_2, \ldots, Y_n, Z\},$$

where $X \prec Y_1 \prec Y_2 \prec \cdots \prec Y_n \prec Z$ and

$$d(X, Y_i) = d(Z, Y_i) = a, \quad a > 0, \qquad d(Y_i, Y_j) = 0, \qquad d(X, Z) = b > a.$$

Evidently the dissimilarity d is pyramidal for $(E, \preceq)$. For any clustering in two clusters we have either

(a) units X and Z belong to the same cluster

$$\mathbf{C}_a = \{C_{a1}, C_{a2}\}, \qquad C_{a1} = \{X, Z\} \cup C, \qquad C_{a2} = E \setminus C_{a1},$$

or

(b) units X and Z belong to different clusters

$$\mathbf{C}_b = \{C_{b1}, C_{b2}\}, \qquad C_{b1} = \{X\} \cup C, \qquad C_{b2} = E \setminus C_{b1}.$$

In both cases $X, Z \notin C$. Suppose that the set C consists of k units. Then

$$p(C_{a1}) = 2b + 4ka, \qquad p(C_{b1}) = 2ka,$$

$$p(C_{a2}) = 0, \qquad p(C_{b2}) = 2(n - k)a.$$

Therefore, for the criterion function of the form SS,

$$P(\mathbf{C}_a) = 2b + 4ka \quad \text{and} \quad P(\mathbf{C}_b) = 2na$$

and, for the criterion function of the form MS,

$$P(\mathbf{C}_a) = 2b + 4ka \quad \text{and} \quad P(\mathbf{C}_b) = 2a \max(k, n - k) \geq 2a\lceil n/2 \rceil.$$

In both cases we have

$$\mathbf{C}^* = \{\{X, Z\}, \{Y_1, Y_2, \ldots, Y_n\}\}$$

provided that n is big enough ($n > 2b/a$). The clustering $\mathbf{C}^*$ is not convex. □

LEMMA 2 [1]. *In a minimal clustering for the problem (P_k, SR) each unit is assigned to a nearest representative. Precisely: if $\mathbf{C}^*$ is a minimal clustering, then*

$$\forall C_u \in \mathbf{C}^*, \quad \forall X \in C_u, \quad \forall C_v \in \mathbf{C}^* \setminus \{C_u\}, \qquad d(X, \bar{C}_u) \leq d(X, \bar{C}_v).$$

PROOF. Let $\mathbf{C}$ be any neighboring clustering to $\mathbf{C}^*$ with respect to transitions

$$\mathbf{C} = (\mathbf{C}^* \setminus \{C_u, C_v\}) \cup \{C_u \setminus \{X\}, C_v \cup \{X\}\}.$$

Since $\mathbf{C}^*$ is minimal we have

$$\begin{aligned} 0 &\leq P(\mathbf{C}) - P(\mathbf{C}^*) \\ &= q(C_u \setminus \{X\}, \overline{C_u \setminus \{X\}}) + q(C_v \cup \{X\}, \overline{C_v \cup \{X\}}) - q(C_u, \bar{C}_u) - q(C_v, \bar{C}_v) \\ &\leq q(C_u \setminus \{X\}, \bar{C}_u) + q(C_v \cup \{X\}, \bar{C}_v) - q(C_u, \bar{C}_u) - q(C_v, \bar{C}_v) \\ &= w(X)(d(X, \bar{C}_v) - d(X, \bar{C}_u)) \end{aligned}$$

and therefore, because $w(X) > 0$, finally

$$d(X, \bar{C}_u) \leq d(X, \bar{C}_v).$$ □

In the following we denote by Γ the set of units and representatives of clusters from feasible clusterings over these units.

LEMMA 3. *Let d be pyramidal for $(\Gamma, \preceq)$. Then, for every $C \in \mathbf{C} \in \mathrm{Min}((P_k, \preceq), SR)$, there is a representative $\bar{C}$ such that $\min(C) \preceq \bar{C} \preceq \max(C)$.*

PROOF. Suppose that, for a representative $\bar{C}$, $\bar{C} \prec \min(C)$. Since, for every $X \in C$, $\min(C) \preceq X \preceq \max(C)$ hold, we have, by pyramidality, $d(X, \min(C)) \leq d(X, \bar{C})$. It follows that $\min(C)$ is a representative as well. □

LEMMA 4. *Let d be pyramidal for $(\Gamma, \preceq)$. Let $C_u, C_v \in \mathbf{C} \in \mathrm{Min}((P_k, \preceq), SR)$ and let $X \in C_v$. If $X \preceq \bar{C}_u \preceq \bar{C}_v$, then $d(X, \bar{C}_u) = d(X, \bar{C}_v)$.*

PROOF. By Lemma 2 we have $d(X, \bar{C}_v) \leq d(X, \bar{C}_u)$. On the other hand, it follows by pyramidality that $d(X, \bar{C}_u) \leq d(X, \bar{C}_v)$. □

THEOREM 3. *Let d be pyramidal for $(\Gamma, \preceq)$. Then the problem $((P_k, \preceq), SR)$ has a convex solution.*

PROOF. Suppose that $\mathbf{C} \in \mathrm{Min}((P_k, \preceq), SR)$ is not convex. We are going to transform this clustering to a convex optimal clustering.

Let X be the smallest unit with respect to $\preceq$ for which a cluster $C_u \in \mathbf{C}$ exists such that $X \notin C_u$ and

$$\min(C_u) \prec X \prec \max(C_u).$$

Because $\mathbf{C}$ is not convex such a unit always exists. Assume $X \in C_v$.

Note first that $X = \min(C_v)$. Hence $X \preceq \bar{C}_v$. Also by assumption,

$$\min(C_u) \prec X \prec \max(C_u).$$

Therefore, according to Lemma 3, there are five cases to be considered (with respect to the position of X and $\bar{C}_v$):

Case 1. $\min(C_u) \prec X \preceq \bar{C}_v \preceq \bar{C}_u \preceq \max(C_u)$.
Case 2. $\min(C_u) \prec X \preceq \bar{C}_u \preceq \bar{C}_v \preceq \max(C_u)$.
Case 3. $\min(C_u) \prec X \preceq \bar{C}_u \preceq \max(C_u) \preceq \bar{C}_v$.
Case 4. $\min(C_u) \preceq \bar{C}_u \preceq X \preceq \max(C_u) \preceq \bar{C}_v$.
Case 5. $\min(C_u) \preceq \bar{C}_u \preceq X \preceq \bar{C}_v \preceq \max(C_u)$.

In all the cases we can define a new clustering

$$\mathbf{C}' = (\mathbf{C} \setminus \{C_u, C_v\}) \cup \{C'_u, C'_v\}$$

such that we obtain a longer initial interval covered by convex clusters.

Case 1. Define the clusters C'_u and C'_v as follows:

$$C'_u = C_u \setminus [\min(C_u), X),$$

$$C'_v = C_v \cup [\min(C_u), X).$$

Here $[\min(C_u), X)$ denotes the interval between $\min(C_u)$ and X including $\min(C_u)$ and excluding X. Due to the choice of X, all the units between $\min(C_u)$ and X belong to C_u. Let Y be such a unit. Then, by Lemma 4, $d(Y, \bar{C}_u) = d(Y, \bar{C}_v)$. It follows that the new clustering is also optimal.

Cases 2–4. In all these cases define the clusters C'_u and C'_v as follows:

$$C'_u = C_u \cup \{X\},$$

$$C'_v = C_v \setminus \{X\}.$$

Now we have

$$\begin{aligned}
p(C'_u) + p(C'_v) &= q(C'_u, \overline{C'_u}) + q(C'_v, \overline{C'_v}) \\
&\le q(C'_u, \bar{C}_u) + q(C'_v, \bar{C}_v) \\
&= q(C_u, \bar{C}_u) + w(X) \cdot d(X, \bar{C}_u) + q(C_v, \bar{C}_v) - w(X) \cdot d(X, \bar{C}_v) \\
&= p(C_u) + p(C_v) + w(X) \cdot (d(X, \bar{C}_u) - d(X, \bar{C}_v)).
\end{aligned}$$

We claim that in all the cases we have $d(X, \bar{C}_u) = d(X, \bar{C}_v)$. Indeed, for Cases 2 and 3 the fact follows from Lemma 4. For Case 4 we have the following sequence of inequalities, which hold by the pyramidality and by Lemma 2:

$$d(X, \bar{C}_u) \leq d(\max(C_u), \bar{C}_u) \leq d(\max(C_u), \bar{C}_v) \leq d(X, \bar{C}_v) \leq d(X, \bar{C}_u).$$

It follows that $d(X, \bar{C}_u) = d(X, \bar{C}_v)$.

Case 5. In this last case define the clusters C'_u and C'_v as follows:

$$C'_u = C_u \setminus \{\max(C_u)\},$$

$$C'_v = C_v \cup \{\max(C_u)\}.$$

By pyramidality, $d(\max(C_u), \bar{C}_v) \leq d(\max(C_u), \bar{C}_u)$ and, on the other hand, by Lemma 2, $d(\max(C_u), \bar{C}_v) \geq d(\max(C_u), \bar{C}_u)$. It follows that we again have an optimal clustering. Yet it is not necessary that we obtained a longer initial interval covered by convex clusters. However, we can repeat this procedure until we (in finite number of steps) reach the unit X. Note that it is possible in this procedure that at some step we no longer have Case 5. However, then one of the cases 1–4 occurs. In any case, after a finite number of steps we obtain a longer initial interval covered by convex clusters. □

4. Dynamic Programming and Clustering. Combinatorial optimization problems can often be successfully solved by the branch and bound method or by dynamic programming. However, it seems that, in general, because of the combinatorial explosion, they are not appropriate for solving the clustering problem [7], [10], [11]; yet, for some special problems, dynamic programming leads to efficient algorithms.

Let us consider the problem $(P_k(E), \oplus)$ with a solution

$$\mathbf{C}^*_k(E) = \{C^*_1, C^*_2, \ldots, C^*_k\}.$$

Then

$$\mathbf{C}^*_{k-1}(E \setminus C^*_k) = \{C^*_1, C^*_2, \ldots, C^*_{k-1}\}$$

is a solution of the problem $(P_{k-1}(E \setminus C^*_k), \oplus)$ and

$$P(\mathbf{C}^*_k(E)) = P(\mathbf{C}^*_{k-1}(E \setminus C^*_k)) \oplus p(C^*_k).$$

Denoting $P^*(E, k) = P(\mathbf{C}^*_k(E))$ we get the Jensen equality [10]:

$$P^*(E, k) = \begin{cases} p(E), & k = 1, \\ \min\limits_{\varnothing \subset C \subset E} (P^*(E \setminus C, k-1) \oplus p(C)), & k > 1. \end{cases}$$

This is a *dynamic programming* (Bellman) equation which, at least in theory, allows us to solve the clustering problem by the following algorithm:

```
for all C: ∅ ⊂ C ⊆ E do begin P*[C, 1] := p(C); opt[C, 1] := C end;
for s := 2 to k do begin
  for all D: { D ⊂ E ∧ card(D) ≥ s,   s < k,
             { D = E,                 s = k,     do begin
    pbest := ∞;
    for all C: C ⊂ D ∧ card(D\C) ≥ s − 1 ∧ max(D) ∈ C do begin
      pt := P*[D\C, s − 1] ⊕ p(C);
      if pt < pbest then begin pbest := pt; cbest := C end
    end;
    P*[D, s] := pbest; opt[D, s] := cbest
  end
end;
C := E; C* := ∅;
for s := k downto 1 do begin
  D := opt[C, s]; C* := C* ∪ {D}; C := C\D
end;
```

In the algorithm max(D) is defined with respect to a given numeration of the set of units E.

In general this algorithm has exponential complexity. Therefore it can be used only for small problems (less than 20 units).

In the case when at least one solution of the problem $((P_k, \preceq), \oplus)$ is convex, dynamic programming leads to a polynomial algorithm which is essentially a generalization of Fisher's algorithm [5], [8], [6, p. 63]:

```
for i := 1 to n do begin P*[i, 1] := p([1, i]); opt[i, 1] = 1 end;
for s := 2 to k do begin
  if s = k then m := n else m := s;
  for i := m to n do begin
    pbest := ∞;
    for j := s to i do begin
      pt := P*[j − 1, s − 1] ⊕ p([j, i]);
      if pt < pbest then begin pbest := pt; cbest := j end
    end;
    P*[i, s] := pbest; opt[i, s] := cbest
  end
end;
r := n; C* := ∅;
for s := k downto 1 do begin
  l := opt[r, s]; C* := C* ∪ {[l, r]}; r := l − 1
end;
```

The reduction of complexity is achieved by limiting the search for an optimal clustering to the intervals with respect to a given ordering $\preceq$. It is easy to see that the complexity of the generalized Fisher algorithm is $O(kn^3)$.

5. Conclusion. The dynamic programming approach is not applicable for solving the general clustering problem; but it leads to a polynomial algorithm for the problems with at least one convex optimal solution. In this paper we proved that such a solution often exists when the underlying dissimilarity is pyramidal for a given ordering of units.

Acknowledgment. We are grateful to a referee for pointing out a mistake in the first version of this paper and for several useful remarks.

References

[1] V. Batagelj, Optimal clusterings, Paper presented at the Fourth European Meeting of the Psychometric Society and the Classification Societies, Cambridge, 2–5 July 1985, Unpublished manuscript.

[2] P. Bertrand and E. Diday, A visual representation of the compatibility between an order and a dissimilarity index: the pyramids, *Comput. Statist. Quart.*, **2** (1985), 31–42.

[3] P. Brucker, On the complexity of clustering problems, in *Optimization and Operations Research* (R. Henn, B. Korte and W. Oettli, eds.), Lecture Notes in Economics and Mathematical Systems, Vol. 175, Springer-Verlag, Berlin, 1978, pp. 45–54.

[4] A Ferligoj and V. Batagelj, Clustering with relational constraint, *Psychometrika*, **47** (1982), 413–426.

[5] W. D. Fisher, On grouping for maximum homogeneity, *J. Amer. Statist. Assoc.*, **53** (1958), 789–798.

[6] A. D. Gordon, *Classification*, Chapman & Hall, London, 1981.

[7] D. J. Hand, Branch and bound in statistical data analysis, Manuscript, 1981.

[8] J. A. Hartigan, *Clustering Algorithms*, Wiley, New York, 1975.

[9] F. K. Hwang, J. Sun, and E. Y. Yao, Optimal set partitioning, *SIAM J. Algebraic Discrete Methods*, **6** (1985), 163–170.

[10] R. E. Jensen, A dynamic programming algorithm for cluster analysis, *Oper. Res.*, **17** (1969), 1034–1057.

[11] M. R. Rao, Cluster analysis and mathematical programming, *J. Amer. Statist. Assoc.*, **66** (1971), 622–626.