# Constrained Clustering Problems

## Vladimir Batagelj[1], Anuška Ferligoj[2]

[1] University of Ljubljana, Faculty of Mathematics and Physics, and
Institute of Mathematics, Physics and Mechanics, Dept. of TCS,
Jadranska 19, 1 000 Ljubljana, Slovenia
[2] University of Ljubljana, Faculty of Social Sciences,
P.O. Box 47, 1 109 Ljubljana, Slovenia

**Summary:** In the paper a view, based on the optimization approach, is given on different types of constrained clustering problems and methods for their solution.

**Keywords:** clustering methods, constraints, blockmodeling, optimization.

## 1.  Introduction

For constrained clustering, grouping similar units into clusters has to satisfy some additional conditions. This class of problems is relatively old. One of the most frequently treated problems in this field is regionalization: clusters of similar geographical regions have to be found, according to some chosen characteristics, where the regions included in the cluster have to be also geographically connected. A number of approaches to this problem have been taken. The majority of authors (e.g., Lebart, 1978; Lefkovitch, 1980; Ferligoj and Batagelj, 1982; Perruchet, 1983; Gordon, 1973, 1980, 1987; Legendre, 1987) solve this problem by adapting standard clustering procedures, especially agglomerative hierarchical algorithms, and local optimization clustering procedures. The geographic contiguity is a special case of *relational constraint*. Ferligoj and Batagelj (1982, 1983) first treated this clustering problem for general symmetric relations and then for nonsymmetric relations. It is possible to work also with other, non-relational conditions. Murtagh (1985) provides a review of clustering with symmetric relational constraints. A more recent survey of constrained clustering was given by Gordon (1996).

## 2.  Formalization

The (constrained) clustering problem can be posed as an optimization problem (Ferligoj and Batagelj, 1982, 1983):

Let $E$ be a finite *set of units*. Its nonempty subset $C \subseteq E$ is called a *cluster*. A set of clusters $\mathbf{C} = \{C_i\}$ forms a *clustering*.

The clustering problem $(\Phi, P, \min)$ can be expressed as: Determine the clustering $\mathbf{C}^* \in \Phi$, for which

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

where $\mathbf{\Phi}$ is a *set of feasible clusterings* and $P : \mathbf{\Phi} \to \mathbb{R}_0^+$ is a *clustering criterion function*. We denote the *set of minimal solutions* by $\mathrm{Min}(\mathbf{\Phi}, P)$.

Let us introduce some notions which we shall need in the following.

The clustering $\mathbf{C}$ is a *complete* clustering if it is a partition of the set of units $E$. We shall denote by $\mathbf{\Pi}(E)$ the set of all complete clusterings of $E$. Two among them $\mathbf{O} \equiv \{\{X\} : X \in E\}$ and $\mathbf{I} \equiv \{E\}$ deserve to be denoted by special symbols. The set of feasible clusterings $\mathbf{\Phi}$ can be decomposed into "strata" (layers)

$$\mathbf{\Phi}_k = \{\mathbf{C} \in \mathbf{\Phi} : \mathrm{card}(\mathbf{C}) = k\}$$

Let $(\mathbb{R}_0^+, \oplus, 0, \leq)$ be an ordered abelian monoid. A *simple* criterion function $P$ has the form:

$$P(\mathbf{C}) = \bigoplus_{C \in \mathbf{C}} p(C), \quad p(C) \geq 0 \quad \text{and} \quad \forall X \in E : p(\{X\}) = 0$$

For almost all criterion functions used in applications, it holds also $p(C_1 \cup C_2) \geq p(C_1) \oplus p(C_2)$. For a simple criterion function satisfying this condition, it holds for $k < \mathrm{card}\, E$: $\forall \mathbf{C} \in \mathbf{\Pi}_k \exists \mathbf{C}' \in \mathbf{\Pi}_{k+1} : P(\mathbf{C}') \leq P(\mathbf{C})$. Since $P(\mathbf{C}) \geq 0$ and $P(\mathbf{O}) = 0$, it holds that $\mathbf{O} \in \mathrm{Min}(\mathbf{\Pi}, P)$. To avoid this trivial problem we usually introduce the obvious constraint – we restrict the problem to $\mathbf{\Pi}_k$, where $k$ is a given number of clusters.

Not all clustering problems can be expressed by a simple criterion function. In some applications a *general* criterion function of the form

$$P(\mathbf{C}) = \bigoplus_{(C_1, C_2) \in \mathbf{C} \times \mathbf{C}} q(C_1, C_2), \quad q(C_1, C_2) \geq 0$$

is needed. An example of general criterion functions can be found in blockmodeling (Batagelj, Ferligoj, Doreian, 1992; Batagelj, 1997)

$$P(\mathbf{C}; \mathcal{T}) = \sum_{(C_1, C_2) \in \mathbf{C} \times \mathbf{C}} \min_{T \in \mathcal{T}} w(T) \delta(C_1, C_2; T)$$

where $\mathcal{T}$ is a set of feasible types, and $\delta$ measures the deviation of blocks, induced by a clustering, from the ideal block structure. The blockmodeling methods considering also (dis)similarities between units have still to be developed. The proposed optimization approach essentially expresses the constraints with a penalty function.

Another such example is a problem of partitioning of a generation of pupils into a given number of classes so that the classes will consist of (almost) the same number of pupils and that they will have a structure as similar as possible. An appropriate criterion function is

$$P(\mathbf{C}) = \max_{\substack{\{C_1, C_2\} \in \mathbf{C} \times \mathbf{C} \\ \mathrm{card}\, C_1 \geq \mathrm{card}\, C_2}} \min_{\substack{f : C_1 \to C_2 \\ f \text{ is surjective}}} \max_{X \in C_1} d(X, f(X))$$

where $d(X, Y)$ is a measure of dissimilarity between pupils $X$ and $Y$.

## 3. Types of Constrained Clusterings

Various types of the constraints are discussed below.

### 3.1. Relational Constraints

Generally, the set of feasible clusterings for this type of constraint can be defined as:

$$\Phi(R) = \{\mathbf{C} \in \mathbf{\Pi} : \text{each cluster } C \in \mathbf{C} \text{ is a subgraph } (C, R \cap C \times C) \text{ in the graph } (E, R) \text{ with the required type of connectedness}\}$$

We can define different types of sets of feasible clusterings for the same relation $R$ if it is nonsymmetric (Ferligoj and Batagelj, 1983). Some examples of clusterings with (nonsymmetrical) relational constraint $\Phi^i(R)$ are

| type of clusterings | type of connectedness |
|---|---|
| $\Phi^1(R)$ | weakly connected units |
| $\Phi^2(R)$ | weakly connected units that contain at most one center |
| $\Phi^3(R)$ | strongly connected units |
| $\Phi^4(R)$ | clique |
| $\Phi^5(R)$ | the existence of a trail containing all the units of the cluster |

A center of a cluster $C$ in the clustering type $\Phi^2(R)$ is the set of units $L \subseteq C$ iff the subgraph induced by $L$ is strongly connected and $R(L) \cap (C \setminus L) = 0$ where $R(L) = \{y : \exists x \in L : xRy\}$.

### 3.2. Constraining Variables

The set of feasible clusterings for this particular type of constraint is defined as follows (Ferligoj, 1986):

$$\Phi[a, b] = \{\mathbf{C} \in \mathbf{\Pi} : \text{for each cluster } C \in \mathbf{C} \text{ holds: } v_C \in [a, b]\}$$

where $v_C$ is a function of values of the constraining variable, $V$, for the units in the cluster $C$.

### 3.3. Optimizational Constraint

The set of feasible clusterings for an optimizational constraint is defined as:

$$\Phi(F) = \{\mathbf{C} \in \mathbf{\Pi} : \text{for second criterion } F \text{ the condition } F(\mathbf{C}) \leq f \text{ holds }\}$$

where $f$ is a given threshold value for the second criterion.

A combination of all three types of constraints (relational, constraining variable and optimizational) can be considered simultaneously.

### 3.4. Pre-specified Blockmodels

Given a network $\mathcal{N}$, set of types $\mathcal{T}$, and a model $\mathcal{M}$ (constraints!), determine $\mu$ (clustering) which minimizes the criterion function (for details see Batagelj, Ferligoj, Doreian, 1998).

## 4. Solving Constrained Clustering Problems

With few exceptions the clustering problem is too hard to be exactly solved efficiently. Therefore, approximative/heuristic methods have to be used. Among these, agglomerative (hierarchical) and local optimization (relocation) methods are the most popular.

### 4.1. Hierarchical Algorithm

The set of feasible clusterings $\mathbf{\Phi}$ determines the *feasibility predicate* $\Phi(\mathbf{C}) \equiv \mathbf{C} \in \mathbf{\Phi}$ defined on $\mathcal{P}(\mathcal{P}(E) \setminus \{\emptyset\})$; and conversely $\mathbf{\Phi} \equiv \{\mathbf{C} \in \mathcal{P}(\mathcal{P}(E) \setminus \{\emptyset\}) : \Phi(\mathbf{C})\}$.

In the set of all clusterings $\mathbf{\Phi}$ the relation of *clustering inclusion* $\sqsubseteq$ can be introduced by

$$\mathbf{C}_1 \sqsubseteq \mathbf{C}_2 \equiv \forall C_1 \in \mathbf{C}_1, C_2 \in \mathbf{C}_2 : C_1 \cap C_2 \in \{\emptyset, C_1\}$$

we say also that the clustering $\mathbf{C}_1$ is a *refinement* of the clustering $\mathbf{C}_2$.

It is well known that $(\mathbf{\Pi}, \sqsubseteq)$ is a partially ordered set (even more, semimodular lattice). Because any subset of partially ordered set is also partially ordered, we have: Let $\mathbf{\Phi} \subseteq \mathbf{\Pi}$ then $(\mathbf{\Phi}, \sqsubseteq)$ is a partially ordered set.

The clustering inclusion determines two related relations (on $\mathbf{\Phi}$):

$\mathbf{C}_1 \sqsubset \mathbf{C}_2 \equiv \mathbf{C}_1 \sqsubseteq \mathbf{C}_2 \wedge \mathbf{C}_1 \neq \mathbf{C}_2$    and

$\mathbf{C}_1 \sqsubset\!\!\!\cdot\, \mathbf{C}_2 \equiv \mathbf{C}_1 \sqsubset \mathbf{C}_2 \wedge \neg \exists \mathbf{C} \in \mathbf{\Phi} : (\mathbf{C}_1 \sqsubset \mathbf{C} \wedge \mathbf{C} \sqsubset \mathbf{C}_2)$.

#### 4.1.1. Conditions on the Structure of the Set of Feasible Clusterings.

In the following we shall assume that the set of feasible clusterings $\mathbf{\Phi} \subseteq \mathbf{\Pi}$ satisfies the following conditions (Batagelj, 1984):

**F1.** The feasibility predicate $\Phi$ is *local* – it has the form $\Phi(\mathbf{C}) = \bigwedge_{C \in \mathbf{C}} \varphi(C)$ where $\varphi(C)$ is a predicate defined on $\mathcal{P}(E) \setminus \{\emptyset\}$ (clusters).

The intuitive meaning of $\varphi(C)$ is: $\varphi(C) \equiv$ the cluster $C$ is "good". Therefore the locality condition can be read: a "good" clustering $\mathbf{C} \in \mathbf{\Phi}$ consists of "good" clusters.

**F2.** $\mathbf{O} \in \mathbf{\Phi}$

**F3.** The predicate $\Phi$ has the property of *binary heredity* with respect to the *fusibility* predicate $\psi(C_1, C_2)$, i.e.,

$$C_1, C_2 \neq \emptyset \wedge C_1 \cap C_2 = \emptyset \wedge \varphi(C_1) \wedge \varphi(C_2) \wedge \psi(C_1, C_2) \Rightarrow \varphi(C_1 \cup C_2)$$

This condition means: in a "good" clustering, a fusion of two "related" clusters produces a "good" clustering.

**F4.** The predicate $\psi$ is *compatible* with clustering inclusion $\sqsubseteq$, i.e.,

$$\forall \mathbf{C}_1, \mathbf{C}_2 \in \mathbf{\Phi} : (\mathbf{C}_1 \sqsubseteq \mathbf{C}_2 \wedge \mathbf{C}_1 \setminus \mathbf{C}_2 = \{C_1, C_2\} \Rightarrow \psi(C_1, C_2) \vee \psi(C_2, C_1))$$

**F5.** The *"interpolation"* property holds in $\mathbf{\Phi}$, i.e., $\quad \forall \mathbf{C}_1, \mathbf{C}_2 \in \mathbf{\Phi}$ :

$$(\mathbf{C}_1 \sqsubset \mathbf{C}_2 \wedge \mathrm{card}(\mathbf{C}_1) > \mathrm{card}(\mathbf{C}_2) + 1 \Rightarrow \exists \mathbf{C} \in \mathbf{\Phi} : (\mathbf{C}_1 \sqsubset \mathbf{C} \wedge \mathbf{C} \sqsubset \mathbf{C}_2))$$

It is easy to verify that the sets of feasible clusterings $\mathbf{\Phi}^i(R)$, $i = 1, 2, 4, 5$ from Ferligoj and Batagelj (1983) satisfy the conditions F1 – F5. But, in the case of $\mathbf{\Phi}^3(R)$, the property F5 fails (in general). The counterexample is given in Figure 5 in Ferligoj and Batagelj (1983), for which we have $\{\{1, 2\}, \{3, 4\}, \{5, 6\}\} \sqsubset \{\{1, 2, 3, 4, 5, 6, \}\}$ .

**4.1.2. Criterion Function and Agglomerative Clustering.** A *dissimilarity between clusters* is a function $d : (C_1, C_2) \to \mathbb{R}_0^+$ which is symmetric, i.e., $d(C_1, C_2) = d(C_2, C_1)$. Let $(\mathbb{R}_0^+, \oplus, 0, \leq)$ be an ordered abelian monoid. Then the criterion function $P$ is *compatible* with dissimilarity $d$ over $\mathbf{\Phi}$ iff:

$$\forall C \subseteq E : (\varphi(C) \wedge \mathrm{card}(C) > 1 \Rightarrow p(C) = \min_{(C_1, C_2) \in \Psi(C)} (p(C_1) \oplus p(C_2) \oplus d(C_1, C_2))$$

Now we can state our main theorem:

**Theorem 1** *Let $P$ be a simple criterion function compatible with $d$ over $\mathbf{\Phi}$, $\oplus$ distributes over $\min$, and F1 – F5 hold, then*

$$P(\mathbf{C}_k^*) = \min_{\mathbf{C} \in \mathbf{\Phi}_k} P(\mathbf{C}) = \min_{\substack{C_1, C_2 \in \mathbf{C} \in \mathbf{\Phi}_{k+1} \\ \psi(C_1, C_2)}} (P(\mathbf{C}) \oplus d(C_1, C_2))$$

From this theorem the following "greedy" approximation can be seen:

$$P(\mathbf{C}_k^*) \approx P(\mathbf{C}_{k+1}^*) \oplus \min_{\substack{C_1, C_2 \in \mathbf{C}_{k+1}^* \\ \psi(C_1, C_2)}} d(C_1, C_2)$$

which is the basis for the following agglomerative (binary) procedure for solving the clustering problem :

```
1.      k := n(= card E); C(k) := O;
2.      while ∃Cᵢ, Cⱼ ∈ C(k): (i ≠ j ∧ ψ(Cᵢ, Cⱼ)) repeat
2.1.        (Cₚ, C_q) := argmin{d(Cᵢ, Cⱼ): i ≠ j ∧ ψ(Cᵢ, Cⱼ)};
2.2.        C := Cₚ ∪ C_q; k := k − 1;
2.3.        C(k) := C(k + 1) \ {Cₚ, C_q} ∪ {C};
2.4.        determine d(C, Cₛ) for all Cₛ ∈ C(k)
3.      m := k
```

Note that this procedure is not an exact procedure for solving the clustering problem. But, because of the nature of clustering problem (Garey and Johnson, 1979; Shamos, 1976; Brucker, 1978), it seems that we are forced to search for and to use such procedures. This procedure has also some nice properties:

**Theorem 2** *All clusterings* $\mathbf{C}(k), k = n, n - 1, \ldots, m$ *obtained by the described procedure are feasible. It holds* $\mathbf{C}(k) \in \mathbf{\Phi}_k, \ k = n, n - 1, \ldots, m$ *and* $\mathbf{C}(m) \in \mathrm{Max} \, \mathbf{\Phi}$.

An agglomerative procedure is said to be *compatible* with $\mathbf{\Phi}$ iff: every clustering obtained by the procedure is feasible, and every feasible clustering can be obtained by the procedure if we can at each step fuse any pair of "related" clusters. For our procedure it can be shown:

**Theorem 3** *If* $\mathbf{\Phi}$ *satisfies the conditions F1 – F5, then the described procedure is compatible with* $\mathbf{\Phi}$.

### 4.2. The Relocation Algorithm

The basic scheme for an adapted relocation algorithm is: Suppose that a reflexive and symmetric *neighborhood* relation $N \subseteq \mathbf{\Phi} \times \mathbf{\Phi}$ is given between feasible clusterings. Usually, for clustering problems, $N$ is determined by the following two transformations: *moving* a unit $X$ from cluster $C_p$ to cluster $C_q$ (*transition*); and *interchanging* units $X$ and $Y$ from different clusters $C_p$ and $C_q$ (*transposition*).

1.  determine the initial feasible clustering $\mathbf{C}$;
2.  **while** there exist $\mathbf{C}' \in N(\mathbf{C})$ such that $P(\mathbf{C}') \leq P(\mathbf{C})$ **repeat**:
2.1.  move to $\mathbf{C}'$: $\mathbf{C} := \mathbf{C}'$ .

The following two features are crucial in developing of an algorithm of this type: a method for randomly generating initial feasible clusterings; and an efficient procedure to scan the neighbourhood $N(\mathbf{C})$.

If the constraints are not too stringent, the relocation method can be applied directly on $\mathbf{\Phi}$; otherwise, we can transform (penalty function method) the problem to an equivalent nonconstrained problem $(\mathbf{\Pi}_k, Q, \mathrm{min})$ with $Q(\mathbf{C}) = P(\mathbf{C}) + \alpha K(\mathbf{C})$ where $\alpha > 0$ is a large constant and

$$K(\mathbf{C}) = \begin{cases} 0 & \Phi(\mathbf{C}) \\ > 0 & \text{otherwise} \end{cases}$$

There exist several improvements of the basic relocation algorithm: simulated annealing, tabu search, ... (Aarts and Lenstra, 1997).

### 4.3.  Dynamic Programming

Suppose that $\mathrm{Min}(\mathbf{\Phi}_k, P) \neq \emptyset$. Denoting $P^*(E, k) = P(\mathbf{C}_k^*(E))$ we can derive the generalized Jensen equality (Batagelj, Korenjak and Klavžar, 1994):

$$P^*(E, k) = \begin{cases} p(E) & \{E\} \in \mathbf{\Phi}_1 \\ \min_{\substack{\emptyset \subset C \subset E \\ \exists \mathbf{C} \in \mathbf{\Phi}_{k-1}(E \setminus C): \mathbf{C} \cup \{C\} \in \mathbf{\Phi}_k(E)}} (P^*(E \setminus C, k-1) \oplus p(C)) & k > 1 \end{cases}$$

This is a *dynamic programming* (Bellman) equation which, for some special constrained problems, allows us to solve the clustering problem by the adapted Fisher's algorithm. For an application of dynamic programming to hierarchies see Lebbe and Vignes (1996).

### 4.4.  Multicriteria Clustering and Constraints

In a *multicriteria clustering problem* $(\mathbf{\Phi}, P_1, P_2, \dots, P_k, \min)$ we have several criterion functions $P_t, t = 1, \dots, k$ over the same set of feasible clusterings $\mathbf{\Phi}$, and our aim is to determine the clustering $\mathbf{C} \in \mathbf{\Phi}$ in such a way that $P_t(\mathbf{C}) \to \min, t = 1, \dots, k$.

In general, solutions minimal for distinct criteria will differ from each other. This creates the problem how to find the 'best' solution so as to satisfy as many of the criteria as possible. In this context, it is useful to define the set of *Pareto efficient* clusterings: a clustering is Pareto efficient if it cannot be improved on any criterion without sacrificing on some other criterion. A multicriteria clustering problem can be approached in different ways (Ferligoj and Batagelj, 1992). It can be solved also by using constrained clustering algorithms where a selected criterion is considered as the clustering criterion and all other criteria determine the (optimizational) constraints. And conversely: a constrained clustering problem can be transformed to a multiciteria clustering problem by expressing the deviations from constraints by penalty functions.

### 4.5.  Other

Some other optimizational approaches for solving constrained clustering problems can be found in Klauer (1994) and Hansen, Jaumard, and Sanlaville (1994).

## 5.  Conclusion

In the paper we presented an overview of constrained clustering problems viewed through 'optimizational' glasses. For details the reader should consult the references. Some related papers are available at

**`http://vlado.fmf.uni-lj.si/pub/cluster/`**

# References

Aarts, E. & Lenstra, J.K. (Eds.)(1997). *Local Search in Combinatorial Optimization.* Wiley, Chichester.

Batagelj, V. (1984). Agglomerative methods in clustering with constraints. *Preprint Series Dept. Math. Univ. Ljubljana*, 22(1984)102, 5-19.

Batagelj, V. (1985). Algorithmic aspects of clustering problem. (invited talk) *Proceedings of the 7 th International Symposium Computer at the University*, Cavtat, 27-29.5.1985, SRCE, Zagreb.

Batagelj, V. (1997). Notes on Blockmodeling. *Social Networks* **19**, 143-155.

Batagelj, V., Doreian, P. & Ferligoj, A. (1992). An Optimizational Approach to Regular Equivalence. *Social Networks* **14**, 121-135.

Batagelj, V., Ferligoj, A. & Doreian, P. (1992). Direct and Indirect Methods for Structural Equivalence. *Social Networks* **14**, 63-90.

Batagelj, V., Ferligoj, A. & Doreian, P. (1998). Fitting Pre-specified Blockmodels. C. Hayashi, & (Eds.): *Data Science, Classification, and Related Methods / IFCS'96*, Springer, Tokyo, 199-206.

Batagelj, V., Korenjak-Černe, S. & Klavžar, S. (1994). Dynamic Programming and Convex Clustering. *Algorithmica* **11**, 93-103.

Brucker, P. (1978). On the complexity of clustering problems, in: R.Henn, B. Korte, W. Oettli (Eds.): *Optimization and Operations Research.* Lecture Notes in Economics and Mathematical Systems, Vol. 157, Springer-Verlag, Berlin.

Ferligoj, A. (1986). Clustering with constraining variable. *Journal of Mathematical Sociology*, **12**, 299-313.

Ferligoj, A. & Batagelj, V. (1982). Clustering with relational constraint. *Psychometrika*, **47**, 4, 413-426.

Ferligoj, A. & Batagelj, V. (1983). Some types of clustering with relational constraints. *Psychometrika*, **48**, 4, 541-552.

Ferligoj, A. & Batagelj, V. (1992). Direct Multicriteria Clustering Algorithms. *Journal of Classification* **9**, 1, 43-61.

Garey, M.R. & Johnson, D.S. (1979). *Computer and intractability.* Freeman, San Francisco.

Gordon, A.D. (1996). A survey of constrained classification. *Computational Statistics & Data Analysis* **21**, 17-29.

Hansen, P., Jaumard, B. & Sanlaville, E. (1994). A Review of Mathematical Programming Approaches, in: P. Bertrand, B. Burtschy (Eds.): *New Approaches in Classification and Data Analysis / IFCS'93*, Springer-Verlag, Berlin, 228-240.

Klauer, K.C. (1994). Representing proximities by network models, in: P. Bertrand, B. Burtschy (Eds.): *New Approaches in Classification and Data Analysis / IFCS'93*, Springer-Verlag, Berlin, 493-501.

Lebbe, J. & Vignes, R. (1996). Optimal Hierarchical Clustering with Order Constraint, in: E. Diday, Y. Lechevallier, O. Opitz (Eds.): *Ordinal and Symbolic Data Analysis / OSDA'95*, Springer-Verlag, Berlin, 265-276.

Murtagh, F. (1985). A survey of algorithms for contiguity-constrained clustering and related problems. *The Computer Journal* **28**, 82-88.

Shamos, M.I. (1976). Geometry and statistics: Problems at the interface, in: J. Traub (Ed.): *Algorithms and Complexity (New directions and recent results).* Academic Press, New York, 251-288.