A fast partitioning algorithm and its application to earthquake investigation

Rudolf Scitovski[1]
Department of Mathematics, University of Osijek
Trg Lj. Gaja 6, HR – 31 000 Osijek, Croatia
e-mail: scitowsk@mathos.hr

Sanja Scitovski
Faculty of Civil Engineering, University of Osijek
Crkvena 21, HR – 31 000 Osijek, Croatia
e-mail: sscitov@unios.hr

Abstract. In this paper a new fast partitioning algorithm able to find either a globally optimal partition or a locally optimal partition of the set $\mathcal{A} \subset \mathbb{R}^n$ close to the global one is proposed. The performance of the algorithm in terms of CPU time shows significant improvement in comparison with other incremental algorithms. Since optimal partitions with $2, 3, \ldots$ clusters are determined successively in the algorithm, it is possible to calculate corresponding clustering validity indexes for every number of clusters in a partition. In that way the algorithm also proposes the appropriate number of clusters in a partition. The algorithm is illustrated and tested on several synthetic and seismic activity data from a wider area of the Republic of Croatia in order to locate the most intense seismic activity in the observed area.

Key words: Center-based clustering; Globally optimal partition; Approximate optimization; DIRECT; Earthquake; Seismic activity.

1 Introduction

Clustering or grouping a data set into conceptually meaningful clusters is a well-studied problem in recent literature. It has practical importance in a wide variety of applications such as earthquake investigation, pattern recognition, facility location problem, text classification, machine learning, business, biology, agriculture, medicine, psychology, etc. (Adelfio et al., 2012; Colombo et al., 1997; Durak, 2011; Iyigun, 2007; Morales-Esteban et al., 2010; Pintér, 1996; Sabo et al., 2011, 2013).

Searching for an optimal partition in general is a complex global optimization problem which can have several local and global minima (Grbić et al., 2012; Evtushenko, 1985; Pardalos and Coleman, 2009; Pintér, 1996). Hence, numerous methods simplifying the problem are proposed in literature, but they may not lead to a globally optimal partition.

[1]Corresponding author: Rudolf Scitovski, e-mail: scitowsk@mathos.hr, telephone number: ++385-31-224-800, fax number: ++385-31-224-801

The second problem in cluster analysis that is often considered is determining the appropriate number of clusters in a partition. If the number of clusters is not given in advance, defining an appropriate number of clusters in a partition is a complex problem (see e.g. Gan et al. (2007); Iyigun (2007); Kogan (2007); Vendramin et al. (2009)).

In our paper, a new incremental algorithm of searching for an optimal partition is proposed. The algorithm represents a generalization of known incremental algorithms (Bagirov and Ugon, 2005; Bagirov, 2008; Bagirov et al., 2011; Likas et al., 2003), and uses the `DIRECT` algorithm for a global optimization of the Lipschitz continuous function (Gablonsky, 2001; Finkel, 2003; Jones et al., 1993) in order to find a good initial approximation for the k-means algorithm. The algorithm locates either a globally optimal partition or a locally optimal partition close to the global one.

The proposed algorithm is applied in earthquake investigation using the data freely available on the website: `http://earthquake.usgs.gov/earthquakes/eqarchives/epic/`. Only data that refer to a wider area of the Republic of Croatia have been extracted from the database. Using the aforementioned algorithm, spatial locations of seismic activity centers are detected.

The paper is organized as follows: In Section 2, some basic terms and facts about data clustering are mentioned. In Section 3, a new algorithm of searching for an optimal partition is constructed. A new algorithm is illustrated and compared with other similar algorithms on several synthetic and empirical examples. In Section 4, the mentioned algorithm is applied on the example of detecting spatial locations of seismic activity centers in a wider area of the Republic of Croatia. Conclusions and future work are discussed in Section 5.

2 Data clustering

The given data point set $\mathcal{A} = \{a_i \in \mathbb{R}^n : i = 1, \ldots, m\}$, where $n \geq 1$ represents the number of features in the data, should be partitioned into $1 \leq k \leq m$ nonempty disjoint subsets (clusters) $\pi_1, \ldots, \pi_k$. Such partition will be denoted by Π, and the set of all partitions of the set $\mathcal{A}$ consisting of k clusters will be denoted by $\mathcal{P}(\mathcal{A}; m, k)$.

Suppose also that a weight $w_i > 0$ is associated to each data point. If $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$, $\mathbb{R}_+ = [0, +\infty\rangle$ is some distance-like function (see e.g. Kogan (2007); Teboulle (2007)), which has at least positive definiteness property, then to each cluster $\pi_j \in \Pi$ we can associate its center c_j defined by

$$c_j = c(\pi_j) := \operatorname*{argmin}_{x \in \operatorname{conv}(\pi_j)} \sum_{a_i \in \pi_j} w_i d(x, a_i), \tag{1}$$

where $\operatorname{conv}(\pi_j)$ is a convex hull of the set π_j. After that, by introducing the objective function $\mathcal{F}\colon \mathcal{P}(\mathcal{A}; m, k) \to \mathbb{R}_+$ we can define the quality of a partition and search for the

globally optimal k-partition by solving the following optimization problem:

$$\underset{\Pi\in\mathcal{P}(\mathcal{A};m,k)}{\operatorname{argmin}} \mathcal{F}(\Pi), \qquad \mathcal{F}(\Pi)=\sum_{j=1}^{k}\sum_{a_i\in\pi_j} w_i d(c_j,a_i). \tag{2}$$

Conversely, for a given set of centers $c_1,\ldots,c_k\in\mathbb{R}^n$, by applying the minimal distance principle, we can define the partition $\Pi=\{\pi(c_1),\ldots,\pi(c_k)\}$ of the set $\mathcal{A}$ which consists of the clusters:

$$\pi(c_j)=\{a\in\mathcal{A}: d(c_j,a)\le d(c_s,a),\ \forall s=1,\ldots,k\}, \qquad j=1,\ldots,k,$$

where one has to take into account that every element of the set $\mathcal{A}$ occurs in one and only one cluster. Therefore, the problem of finding an optimal partition of the set $\mathcal{A}$ can be reduced to the following *global optimization problem* (GOP) (see e.g. Späth (1983); Teboulle (2007))

$$\underset{c_1,\ldots,c_k\in\mathbb{R}^n}{\operatorname{argmin}} F(c_1,\ldots,c_k), \qquad F(c_1,\ldots,c_k)=\sum_{i=1}^{m} w_i \min_{1\le s\le k} d(c_s,a_i), \tag{3}$$

The solution of (2) and (3) coincides. Namely, it is easy to verify the following equalities

$$\begin{aligned} F(c_1^\star,\ldots,c_k^\star) &= \sum_{i=1}^{m} w_i \min_{1\le s\le k} d(c_s^\star,a_i)=\sum_{j=1}^{k}\sum_{a_i\in\pi(c_j^\star)} w_i \min_{1\le s\le k} d(c_s^\star,a_i) \\ &= \sum_{j=1}^{k}\sum_{a_i\in\pi(c_j^\star)} w_i d(c_j^\star,a_i)=\mathcal{F}(\Pi^\star), \end{aligned} \tag{4}$$

where $\Pi^\star=\{\pi(c_1^\star),\ldots,\pi(c_k^\star)\}$. Thereby, the objective function F is a symmetric function which can have a large number of independent variables, it does not have to be either convex or differentiable, and generally it may have at least $k!$ local and global minima (Grbić et al., 2012). Therefore, this becomes a complex GOP.

2.1 Choice of a distance-like function

Among many well-known distance–like functions (Durak, 2011; Kogan, 2007; Teboulle, 2007) we will mention only two which will be used in numerical experiments in Section 3.2 and in application to seismic activity in Section 4. In some concrete applications, the choice of the corresponding distance–like function is very important.

The most popular distance–like function is *the Least Squares (LS) distance–like function* $d_{LS}\colon \mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}_+$, $d_{LS}(x,y)=\|x-y\|_2^2$. In this case, the cluster center is called the centroid and it can be simply obtained as a weighted arithmetic mean

$$c_j=\underset{x\in\operatorname{conv}(\pi_j)}{\operatorname{argmin}}\sum_{a_i\in\pi_j} d_{LS}(x,a_i)=\tfrac{1}{W_j}\sum_{a_i\in\pi_j} w_i a_i, \qquad W_j=\sum_{a_i\in\pi_j} w_i. \tag{5}$$

Centroid c_j has the property that the weighted sum of squares of Euclidean distances of points from the cluster π_j to its center c_j is minimal. From the physical point of view, the centroid c_j can be understood as a center of gravity of the set $\mathcal{A}$ with weights $w_i > 0$ of its points.

Mahalanobis distance–like function $d_M \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$, $d_M(x, y) = (x - y)S(x - y)^T$, ($S > 0$ symmetric positive definite matrix), takes into consideration the correlations within a data set (Durak, 2011). The matrix S is a symmetric positive definite covariance matrix. It can be easily seen that the cluster center is the same when using the LS-distance–like function. Note also that the Mahalanobis distance–like function becomes an LS-distance–like function if S is the identity matrix and both of these distance–like functions have a symmetry property, but they do not satisfy the triangle inequality.

3 Searching for a globally optimal partition

Given is a data points set $\mathcal{A} \subset [\alpha, \beta] \subset \mathbb{R}^n$, where $\alpha = (\alpha_1, \ldots, \alpha_n), \beta = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$ and $[\alpha, \beta] = \{x \in \mathbb{R}^n \colon \alpha_i \leq x_i \leq \beta_i\}$, thereby to each data point $a^i \in \mathcal{A}$ a weight $w_i > 0$ is associated. The goal is to determine a partition $\Pi^\star = \{\pi_1^\star, \ldots, \pi_k^\star\}$ with centers $c_1^\star, \ldots, c_k^\star$ as a solution of GOP (2), or equivalently (3).

Since our objective function (3) is a Lipschitz continuous function (Pintér, 1996; Sabo et al., 2013), there are numerous methods for solving this GOP (Evtushenko, 1985; Floudas and Gounaris, 2009; Neumaier, 2004; Pintér, 1996). One of the most popular algorithms for solving a GOP for the Lipschitz continuous function is the `DIRECT` (`DI`viding `RECT`angles) algorithm (Finkel, 2003; Gablonsky, 2001; Jones et al., 1993). Because of the symmetry property od the function F there are at least $k!$ solutions of this problem. That was a motive for developing a very efficient special version of the `DIRECT` algorithm for symmetric functions in Grbić et al. (2012). Complexity of this problem is specially emphasized if the number of features n or the number of data points m is large.

Instead of searching for the GOP, various simplifications are often proposed in the literature that would find a partition for which we usually do not know how close it is to the globally optimal one. The most popular algorithm of searching for a locally optimal partition is a well-known k-means algorithm (see e.g. Kogan (2007); Rizman-Žalik (2008); Späth (1983); Teboulle (2007)). If we have a good initial approximation, this algorithm can provide an acceptable solution (Volkovich et al., 2007). In case we do not have a good initial approximation, the algorithm should be restarted with various random initializations, as proposed by (Leisch, 2006).

3.1 A new algorithm

Our paper proposes a new efficient algorithm of searching for an optimal partition as a natural generalization of different incremental algorithms (Likas et al., 2003; Bagirov,

2008; Bagirov et al., 2011). For that purpose we define the sequence of objective functions

$$F_k\colon \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{k} \to \mathbb{R}_+, \qquad F_k(c_1,\ldots,c_k) = \sum_{i=1}^{m} w_i \min\{d(c_1,a^i),\ldots,d(c_k,a^i)\}. \quad (6)$$

For $k = 1$, the function F_1 attains its global minimum at the point $c_1^\star \in [\alpha,\beta]$ given by (1).

For $k > 1$, we determine an optimal k-partition with centers $c_1^\star,\ldots,c_k^\star$ by the following incremental algorithm.

Algorithm 1. (Searching for an optimal k-partition)

Step 1: Let $\hat{c}_1,\ldots,\hat{c}_{k-1}$ be the centers obtained in the previous step as an approximation of a global minimizer of the function F_{k-1} and let

$$F_{k-1}(\hat{c}_1,\ldots,\hat{c}_{k-1}) = \sum_{i=1}^{m} w_i \delta_{k-1}^i, \quad \delta_{k-1}^i = \min\{d(\hat{c}_1,a^i),\ldots,d(\hat{c}_{k-1},a^i)\}, \quad (7)$$

$$\Phi_k(c) := F_k(\hat{c}_1,\ldots,\hat{c}_{k-1},c) = \sum_{i=1}^{m} w_i \min\{\delta_{k-1}^i, d(c,a^i)\}. \quad (8)$$

Step 2: By using the `DIRECT` algorithm for global optimization determine

$$\hat{c}_k \in \operatorname*{argmin}_{c\in[\alpha,\beta]} \Phi_k(c) \quad (9)$$

Step 3: By using the k-means algorithm with initial approximations $\hat{c}_1,\ldots,\hat{c}_k$ determine new centers $c_1^\star,\ldots,c_k^\star$.

Remark 1. Note first that the objective function value decreases by increasing the cluster number (see e.g. (Späth, 1983)). Therefore,

$$F_1(c_1^\star) \geq F_2(\hat{c}_1,\hat{c}_2) \geq F_2(c_1^\star,c_2^\star) \geq \cdots \geq F_k(\hat{c}_1,\ldots,\hat{c}_k) \geq F_k(c_1^\star,\ldots,c_k^\star),$$

and, according to Bagirov and Ugon (2005), the maximum number of clusters k_{max} that makes sense to be calculated using Algorithm 1 is determined by

$$\frac{F_{k_{max}-1}^\star - F_{k_{max}}^\star}{F_1^\star} < \epsilon, \quad (10)$$

for some small $\epsilon > 0$ because the relative reduction of the objective function value for $k \geq k_{max}$ is less than ϵ.

The important advantage of all incremental algorithms of searching for an optimal partition lies in the fact that we get an optimal partition for each $k \leq k_{max}$, which makes it possible to decide on the appropriate number of clusters in a partition by using various well-known indexes (see Section 4.1).

Unfortunately, it cannot be asserted that the proposed algorithm gives a globally optimal k-partition, but numerous calculations conducted in the next section show that the partition obtained by Algorithm 1 is either a globally optimal partition or a locally optimal partition close to the global one and is therefore acceptable in applied research. In the following, the partition obtained by Algorithm 1 will be simply called an *optimal partition*.

In comparison with other similar methods, the advantage of Algorithm 1 is specially manifested in significantly shorter CPU-time. As shown in several illustrative examples below, the function Φ_k can have several local and global minima and in *Step 2* we choose one of the global minimizers. This is the basic difference and advantage of the proposed algorithm compared to other incremental algorithms. Furthermore, it is known that in some cases the `DIRECT` algorithm is not efficient enough (Grbić et al., 2012; Sergeyev and Kvasov, 2011). However, that is not the case with our problem solving since the solution obtained by the `DIRECT` algorithm in *Step 2* is used only as an initial approximation in *Step 3* and it is not necessary to request a high accuracy. This efficiency of Algorithm 1 is also confirmed by demanding calculations in earthquake investigations.

Instead of solving the GOP in *Step 2*, several different approaches can be found in literature. The worst possibility (Likas et al., 2003) is instead of $\hat{c}_k$ in *Step 2*, to take the elements from set $\mathcal{A}$ successively and to execute the k-means algorithm for each of them. In this procedure the solution $(c_1^\star, \ldots, c_k^\star)$ is a k-means solution with the smallest objective function value. However, if the number of data points m is large (as, for example, in case of earthquake investigation - see Section 4), then CPU-time can be unacceptable.

The second possibility (Likas et al., 2003; Bagirov, 2008) is to find $a^j \in \mathcal{A}$ for which the difference $\Delta(a^j) := F_{k-1}(\hat{c}_1, \ldots, \hat{c}_{k-1}) - \Phi(a^j)$ will be the largest and to choose precisely that a_j for $\hat{c}_k$. A more detailed description of this method for the case of data with weights follows.

Let $I = \{1, \ldots, m\}$ be the set of all indices. First, for some $a^j \in \mathcal{A}$ we define disjoint sets of indices

$$I_1^{(j)} = \{i \in I \colon \delta_{k-1}^i \leq d(a^j, a^i)\}, \qquad I_2^{(j)} = \{i \in I \colon \delta_{k-1}^i > d(a^j, a^i)\}, \tag{11}$$

where δ_{k-1}^i is given by (7). Then

$$\begin{aligned}\Delta(a^j) &= \sum_{i=1}^{m} w_i \delta_{k-1}^i - \sum_{i \in I_1^{(j)}} w_i \delta_{k-1}^i - \sum_{i \in I_2^{(j)}} w_i d(a^j, a^i) \\ &= \sum_{i \in I_2^{(j)}} w_i \delta_{k-1}^i - \sum_{i \in I_2^{(j)}} w_i d(a^j, a^i) = \sum_{i \in I_2^{(j)}} w_i \left(\delta_{k-1}^i - d(a^j, a^i)\right).\end{aligned}$$

Since

$$\sum_{i=1}^{m} w_i \max\{0, \delta_{k-1}^i - d(a^j, a^i)\} = \sum_{i \in I_2^{(j)}} w_i \left(\delta_{k-1}^i - d(a^j, a^i)\right) = \Delta(a^j),$$

the largest difference $\Delta(a^j)$ is obtained for the $a^j \in \mathcal{A}$ for which a maximum of

$$\Delta(a^j) = \sum_{i=1}^{m} w_i \max\{0, \delta_{k-1}^i - d(a^j, a^i)\}. \tag{12}$$

is attained. This approach in literature (Likas et al., 2003; Bagirov, 2008; Bagirov et al., 2011) is called the *Global k-means Algorithm* (GKM).

The third possibility considered in (Bagirov and Ugon, 2005; Bagirov and Yearwood, 2006) is the application of the *discrete gradient method* (GRAD) for finding the minimum of the auxiliary function Φ_k. Since this is a local method, which is very sensitive to the choice of the initial approximation, these papers pay special attention to the choice of a convenient initial approximation.

The papers (Bagirov, 2008; Bagirov et al., 2011) abandon the idea of applying the discrete gradient method. Instead, they carry out a detailed analysis of sets $\left(I_2^{(j)},\, j \in I\right)$ and on the basis of that propose a good initial approximation for the k-means algorithm. This method shows some very good performances, but requires relatively large CPU-time.

3.2 Numerical experiments and illustrations

The proposed Algorithm 1 will be illustrated on several examples and compared with the GKM-algorithm and the GRAD-algorithm which incorporates GKM for finding the initial approximation.

Example 1. *In the square* $[0,1]^2 \subset \mathbb{R}^2$, $k = 7$ *points* $C_1, \ldots, C_k$, *are randomly chosen, which make vector* $c = (C_1, \ldots, C_k) \in \mathbb{R}^{7\times 2}$. *In the neighborhood of point* $C_j \in [0,1]^2$, m_j *random points are generated by using binormal random additive errors with mean vector* $0 \in \mathbb{R}^2$ *and the covariance matrix* $\sigma^2 I$, $\sigma^2 = 0.1$. *Thereby,* m_j *are random integers from* $[10, 50]$. *In this way, we obtained a data point set* $\mathcal{A}$ *with* $m = \sum_{j=1}^{7} m_j$ *random points.*

By using the LS-distance-like function and applying GKM, GRAD and Algorithm 1 we search for a reconstruction vector of centers $\hat{c} = (\hat{c}_1, \ldots, \hat{c}_k) \in \mathbb{R}^{7\times 2}$. *We define the error of reconstruction by using the Hausdorff distance*

$$\hat{d}_H := d_H(c, \hat{c}) = \max\{\max_r \min_s \|c_r - \hat{c}_s\|^2, \max_s \min_r \|c_r - \hat{c}_s\|^2\}. \tag{13}$$

Algorithm	$\hat{d}_H < .1$	$.1 \le \hat{d}_H < .2$	$.2 \le \hat{d}_H < .3$	$.3 \le \hat{d}_H$	CPU (min)
GKM	49	39	11	1	20:04
GRAD	51	37	11	1	20:42
Algorithm 1	51	35	14	-	03:41

Table 1: Frequency of errors and the corresponding CPU-time for 100 experiments

We will repeat this experiment 100 *times with different random data. The results presented in Table 1 show that all three methods reconstruct cluster centers well enough, but the* CPU*-time of Algorithm 1 is significantly shorter.*

Example 2. *As an illustration, we observe a chosen experiment from Example 1.*

(a) GKM (b) GRAD (c) Algorithm 1

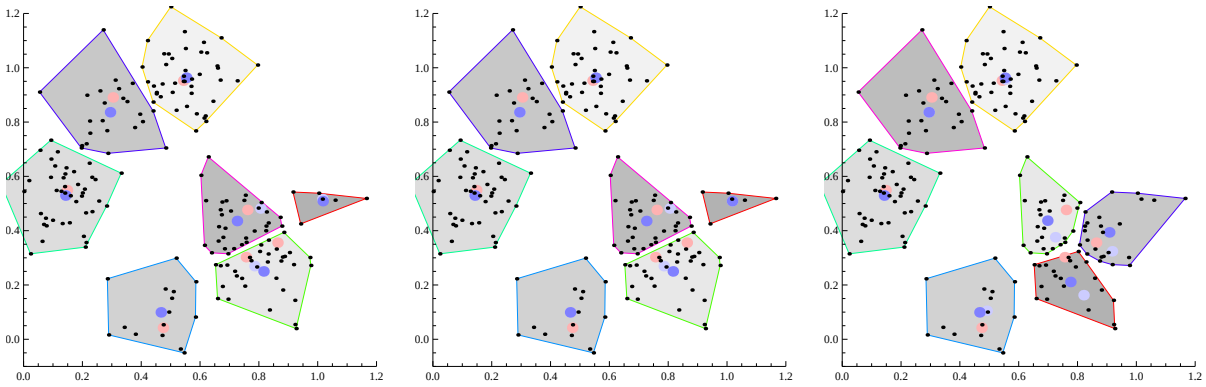


Figure 1: Reconstruction of the points $c_1, \ldots, c_7$

Fig. 1 shows original points $C_1, \ldots, C_k$ (red points), data (black points), initial approximation $\hat{c}_1, \ldots, \hat{c}_k$ (light blue points) and reconstructed centers $c_1^\star, \ldots, c_k^\star$ (blue points) with clusters for every method. Table 2 shows Hausdorff distances of points $\hat{c}_1, \ldots, \hat{c}_k$ and points $c_1^\star, \ldots, c_k^\star$ to the original points $C_1, \ldots, C_k$ with corresponding objective function values for every method. Centers $c_1^\star, \ldots, c_k^\star$ and corresponding objective function values do not differ substantially with all three algorithms, but the CPU-time for Algorithm 1 is significantly shorter.

Method	$d_H(c, \hat{c})$	$F(\hat{c})$	$d_H(c, c^\star)$	$F(c^\star)$	CPU-time (sec)
GKM	0.221342	3.2576	.216829	3.12839	8.52
GRAD	0.216829	3.25725	.216829	3.12839	8.78
Algorithm 1	0.153964	3.28666	.093465	3.10509	1.57

Table 2: Details of a chosen experiment from Example 1

Example 3. *We will compare Algorithm 1 with GKM and GRAD on the data set* $\mathcal{A}_3$ ($|\mathcal{A}_3| = 3184$) *with weights* $w_i = M_i \geq 3$ *from Section 4.2 by using the Mahalanobis distance-like function. We chose* $\hat{c}_1 = (18, 43)$ *as the initial center. Table 3 shows the flow of the iterative process for all three algorithms. All algorithms will be terminated for* $k_{max} = 14$ *according to* (10) *because* $\frac{F_{14}^\star - F_{13}^\star}{F_1^\star} < .005$. *As one can see, centers* $c_1^\star, \ldots, c_{13}^\star$

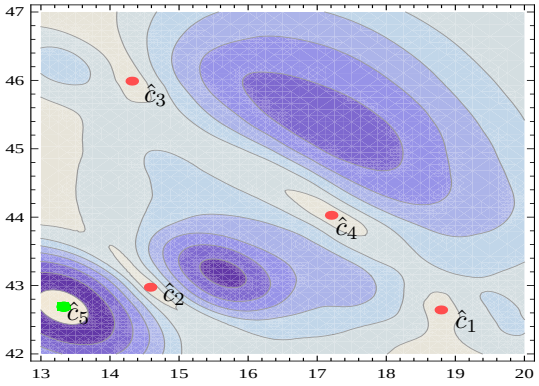
k	GKM $F(\hat{c})$	GKM CPU	GKM $F(c^\star)$	GKM CPU	GRAD $F(\hat{c})$	GRAD CPU	GRAD $F(c^\star)$	GRAD CPU	Algorithm 1 $F(\hat{c})$	Algorithm 1 CPU	Algorithm 1 $F(c^\star)$	Algorithm 1 CPU
2	14576.7	4:12	14158.7	5:14	14455.2	4:12	14158.7	4:14	14455.3	0:04	14158.7	0:06
3	8995.7	5:55	5846.0	7:14	8995.2	5:56	5846.0	5:57	8995.4	0:06	5846.0	0:07
4	4861.0	7:34	4510.1	9:16	4858.7	7:35	4510.1	7:42	4858.7	0:09	4510.1	0:16
5	3768.1	9:14	3223.7	11:20	3767.5	9:15	3223.7	9:16	3767.6	0:11	3223.7	0:13
6	2687.3	11:12	2376.7	13:31	2685.7	11:12	2376.7	11:16	2686.1	0:12	2376.7	0:16
7	2137.6	12:48	1996.0	15:33	2137.3	12:49	1996.0	12:51	2137.3	0:15	1996.0	0:17
8	1804.2	14:44	1769.4	17:21	1803.7	14:45	1769.4	14:47	1803.6	0:19	1769.4	0:22
9	1619.7	15:56	1581.5	19:22	1619.4	15:57	1581.5	15:59	1619.4	0:19	1581.5	0:22
10	1454.9	17:58	1408.3	21:17	1454.6	17:59	1408.3	18:01	1463.6	0:18	1408.3	0:20
11	1294.0	19:57	1199.6	23:57	1293.8	19:58	1199.6	20:04	1303.8	0:21	1271.8	0:24
12	1101.4	21:49	1071.9	25:21	1097.4	21:51	1071.9	21:53	1161.7	0:25	1068.1	0:30
13	981.3	23:22	941.0	27:24	978.9	23:22	941.0	23:26	966.3	0:27	940.9	0:30
		2:44:41		3:16:50		2:44:43		2:45:26		0:03:06		0:03:46

Table 3: Comparison of algorithms of searching for $k = 2, \ldots, 13$ spatial locations of seismic activity centers

and corresponding objective function values for different algorithms differ a bit only in the last few iterations, but the CPU-time for Algorithm 1 is significantly shorter.

As an illustration, let us look closely at $k = 5$. *The known centers* $\hat{c}_1, \ldots, \hat{c}_4$ *in Fig. 2a are shown on* `ContourPlot` *of the function* $\Phi_5(c) = F(\hat{c}_1, \ldots, \hat{c}_4, c)$ *with the red dots. By solving the GOP in Step 2 we get the point* $\hat{c}_5$ *(green point in Fig. 2a). After that, by using the k-means algorithm (Step 3), we get the optimal solution* $c^\star = (c_1^\star, \ldots, c_5^\star)$ *(green points in Fig. 2b).*

(a) Initial approximations



(b) k-means optimized centers

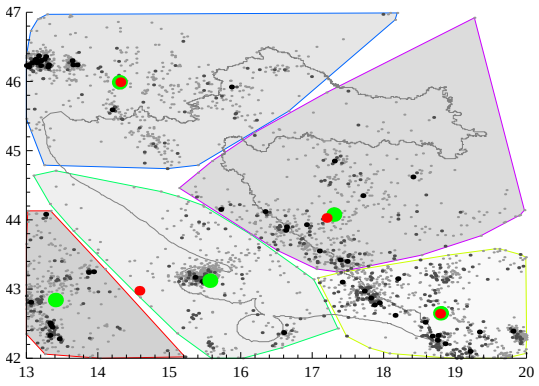


Figure 2: Searching for the global minimizer of the function Φ_5

4 Detection of spatial locations of seismic activity centers

Earthquakes usually come without any warning, they can destroy entire cities in just a few seconds and kill or severely injure a large number of people and cause enormous property damage. Therefore, a lot of attention is paid to the study of earthquakes.

There are several publicly available databases of earthquakes that occurred in the past around the world. For example, the web site
`http://earthquake.usgs.gov/earthquakes/eqarchives/epic/`
contains data on earthquakes around the world since 1973. The following data are given for each earthquake that occurred in that period of time:

$$\textit{Year / Month / Day / Orig.Time/Latitude}(\varphi_i)\textit{/Longitude}(\lambda_i)\textit{/ Depth / Magnitude } (w_i)$$

Based upon this data, the set

$$\mathcal{A}_M = \{a_i = (\lambda_i, \varphi_i) \in \mathbb{R}^2 \colon L_\lambda \leq \lambda_i \leq U_\lambda,\ L_\varphi \leq \varphi_i \leq U_\varphi,\ M_i \geq M\},$$

which contains earthquake locations determined by longitude $\lambda_i \in [L_\lambda, U_\lambda]$, latitude $\varphi_i \in [L_\varphi, U_\varphi]$ and magnitude $M_i \geq M > 0$ is constructed. The set $\mathcal{A}_M$ of the data location with the weights $w_i = M_i$ should be partitioned into $1 \leq k \leq m$ clusters. Thereby, an appropriate number of clusters should be taken into consideration. Centers of these clusters will represent centers of seismic activity in the neighborhood of which stronger earthquakes most frequently occurred in this period in the area under consideration.

If the rectangle $[L_\lambda, U_\lambda] \times [L_\varphi, U_\varphi]$ is relatively small (such that relative distances in this rectangle do not significantly differ from relative distances in the corresponding rectangle in the Gauss-Krüger coordinate system), then searching for the optimal partition can be carried out directly with the data from the set $\mathcal{A}_M$. Else, it would be necessary to transform the data set in the Gauss-Krüger coordinate system.

An important application of cluster analysis in earthquake investigation is forecasting the location, time and magnitude of an earthquake in future. Knowing the centers of seismic activity is important when making a decision on choosing the location for building major construction facilities (Modirzadeh et al., 2012). The forecast of the occurrence and the maximum magnitude of a possible earthquake is one of the most difficult problems in seismic hazard assessment. Naturally, forecasts based only on a posteriori criteria are not representative enough, but they can indicate seismic hazard. (Cho et al., 2010; Colombo et al., 1997; Holliday et al., 2006; Morales-Esteban et al., 2010).

4.1 Determining the appropriate number of clusters in the partition

Automatically determining the number of clusters has been one of the most difficult problems in data clustering processes. In some cases, the number of clusters in a partition

is determined by the nature of the problem itself. If the number of clusters in a partition is not given in advance (as is the case with the problem under consideration), then it is natural to search for an optimal partition which consists of clusters that are as compact and well-separated as possible. Thereby, choosing a large number of clusters does not necessarily imply better classifications. There are many different approaches to solving this problem in the literature (see e.g. Gan et al. (2007); Iyigun (2007); Kogan (2007); Vendramin et al. (2009)).

The observed data on earthquake activity are naturally grouped as extended and not spherical units. For that purpose we will use a Mahalanobis distance-like function (Durak, 2011) and the following validity indexes for determining the appropriate number of clusters in a partition: Davies-Bouldin (Davies and Bouldin, 1979), Silhouette With Criterion and Simplify Silhouette Width Criterion (Kaufman and Rousseeuw, 2005). These indexes are proposed as most appropriate in similar situations (Adelfio et al., 2012; Morales-Esteban et al., 2010; Vendramin et al., 2009).

Davies - Bouldin Index (DB) for the optimal partition with k clusters is defined as

$$DB(k) = \frac{1}{k}\sum_{j=1}^{k} \max_{s\neq j} \frac{V(\pi_j^\star)+V(\pi_s^\star)}{d(c_j^\star,c_s^\star)}, \tag{14}$$

where $V(\pi_j^\star)$ is a variance of the cluster $\pi_j^\star$ given by

$$V(\pi_j^\star) = \frac{1}{W_j}\sum_{a_s\in\pi_j^\star} w_s d(c_j^\star, a_s), \qquad W_j = \sum_{a_s\in\pi_j^\star} w_s.$$

More compact and better separated clusters in an optimal partition will result in a lower DB index.

Silhouette Width Criterion (SVC) is very popular in cluster analysis and applications. For the optimal partition with k clusters $\pi_1^\star,\ldots,\pi_k^\star$ the SVC is defined as follows: For each $a_i \in \mathcal{A}\cap\pi_r^\star$ we calculate the numbers

$$\alpha_{ir} = \frac{1}{\Omega_r}\sum_{a_s\in\pi_r^\star} w_s d(a_i,a_s), \qquad \beta_{ir} = \min_{q\neq r}\frac{1}{\Omega_q}\sum_{a_s\in\pi_q^\star} w_s d(a_i,a_s), \qquad \Omega_p = \sum_{a_s\in\pi_p^\star} w_s, \tag{15}$$

and the corresponding index is then defined as

$$SWC(k) = \frac{1}{m}\sum_{i=1}^{m}\frac{\beta_{ir}-\alpha_{ir}}{\max\{\alpha_{ir},\beta_{ir}\}}. \tag{16}$$

More compact and better separated clusters in an optimal partition will result in a greater SWC number.

Simplify Silhouette Width Criterion (SSC) uses the distances of elements $a_i \in \mathcal{A}\cap\pi_r^\star$ to cluster centers $c_1^\star,\ldots,c_k^\star$ instead of the average value from (15)

$$\alpha_{ir} = d(a_i,c_r^\star), \quad \beta_{ir} = \min_{q\neq r} d(a_i,c_q^\star), \qquad SSC(k) = \frac{1}{m}\sum_{i=1}^{m}\frac{\beta_{ir}-\alpha_{ir}}{\max\{\alpha_{ir},\beta_{ir}\}}. \tag{17}$$

4.2 An application to seismic activity in a wider area of the Republic of Croatia

We will apply Algorithm 1 to determining locations of the most intense seismic activity in a wider area of the Republic of Croatia. In this case the data set

$$\mathcal{A}_3 = \{a_i = (\lambda_i, \varphi_i) \in \mathbb{R}^2 \colon 13 \leq \lambda_i \leq 20,\ 42 \leq \varphi_i \leq 47,\ w_i = M_i \geq 3\}, \tag{18}$$

consists of locations in this area that have been affected by the earthquake of magnitude at least 3 since 1973. One can find 10 018 data in the database for this area, whereby $m = 3184$ of them refer to earthquakes of magnitude greater than or equal to 3. Locations of these earthquakes are denoted in Fig. 3a, where high magnitude earthquakes are marked by bigger black dots.

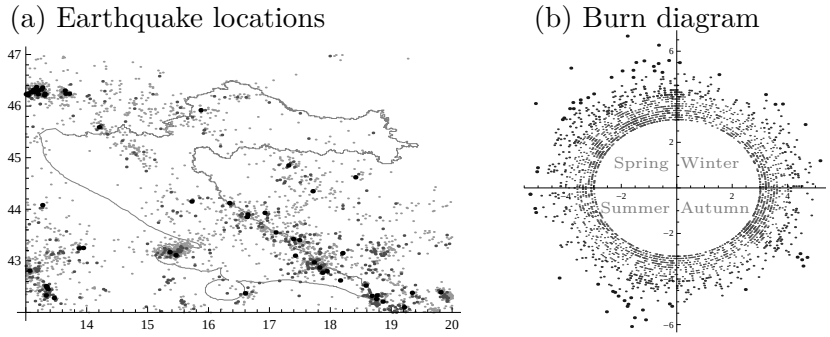


Figure 3: Locations in a wider area of the Republic of Croatia affected by the earthquake of magnitude at least 3 since 1973 and the corresponding moments in a year

Fig. 3b shows time distribution of seismic moments throughout the year for this data set using the so-called Burn diagram (see Parajka et al. (2010)), i.e. the set

$$\mathcal{B} = \{w_i(\cos t_i, \sin t_i) \in \mathbb{R}^2 \colon t_i = 2\pi T_i (\mathrm{mod}\ 2\pi) \in [0, 2\pi],\ i = 1, \ldots, m\},$$

where $T_i \in [0, 39]$ are time moments inside 39 successive years since 1973. This diagram also confirms that seismic moments can be considered as stationary Poisson processes with a fixed occurrence rate over time (Cho et al., 2010; Stipčević et al., 2011).

The data point set $\mathcal{A}_3$ given by (18) will be partitioned into several clusters by using the Mahalanobis distance-like function $d_M \colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$ (Durak, 2011)

$$d_M(x, y) = (x - y)\Sigma^{-1}(x - y)^T, \tag{19}$$

$$\Sigma = \frac{1}{W} \begin{bmatrix} \sum\limits_{i=1}^{m} w_i(\lambda_i - \bar{\lambda})^2 & \sum\limits_{i=1}^{m} w_i(\lambda_i - \bar{\lambda})(\varphi_i - \bar{\varphi}) \\ \sum\limits_{i=1}^{m} w_i(\lambda_i - \bar{\lambda})(\varphi_i - \bar{\varphi}) & \sum\limits_{i=1}^{m} w_i(\varphi_i - \bar{\varphi})^2 \end{bmatrix} = \begin{bmatrix} 4.6646 & -1.3706 \\ -1.3706 & 1.8571 \end{bmatrix},$$

where Σ is a covariance matrix and $\bar{\lambda} = \frac{1}{W}\sum_{i=1}^{m} \lambda_i = 16.2268$, $\bar{\varphi} = \frac{1}{W}\sum_{i=1}^{m} \varphi_i = 43.7542$, and $W = \sum_{i=1}^{m} w_i = 11418.4$. By using Algorithm 1 we calculate the weighted Mahalanobis optimal partition (WM-optimal partition) for $k = 2, \ldots, 14$ (see Example 3).

In each iteration of Algorithm 1, according to Subsection 4.1, the corresponding Davies-Bouldin (DB) index, Silhouette Width Criterion (SWC) and Simplify Silhouette Width Criterion (SSC) are calculated after Step 3. Fig. 4 shows graphs of these indexes for a WM-optimal partition obtained by Algorithm 1. Clustering validity indexes show that the partition with $k = 3$ and $k = 13$ clusters achieves relatively more compact and well-separated clusters.

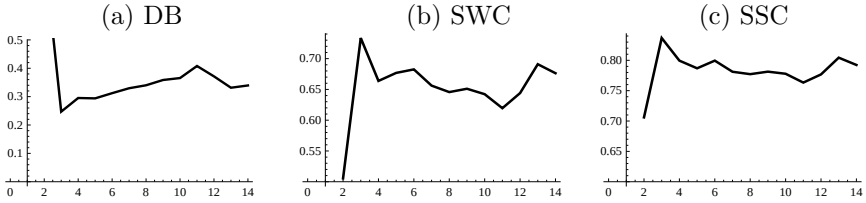


Figure 4: Clustering validity indexes for a WM-optimal partition

A WM-optimal partition with $k = 3$ clusters shows global distribution of seismic activities in a wider area of the Republic of Croatia. Centers of their clusters are situated in the neighborhood of the following locations (see also Fig. 5a):

Mostar (Bosnia and Herzegovina)	$(\lambda = 17.80, \varphi = 43.34)$	$c_1^\star = (18.23, 43.12)$
San Benedetto del Tronto (Italy)	$(\lambda = 13.87, \varphi = 42.96)$	$c_2^\star = (14.60, 42.99)$
Ljubljana (Slovenia)	$(\lambda = 14.50, \varphi = 46.05)$	$c_3^\star = (14.53, 45.90)$

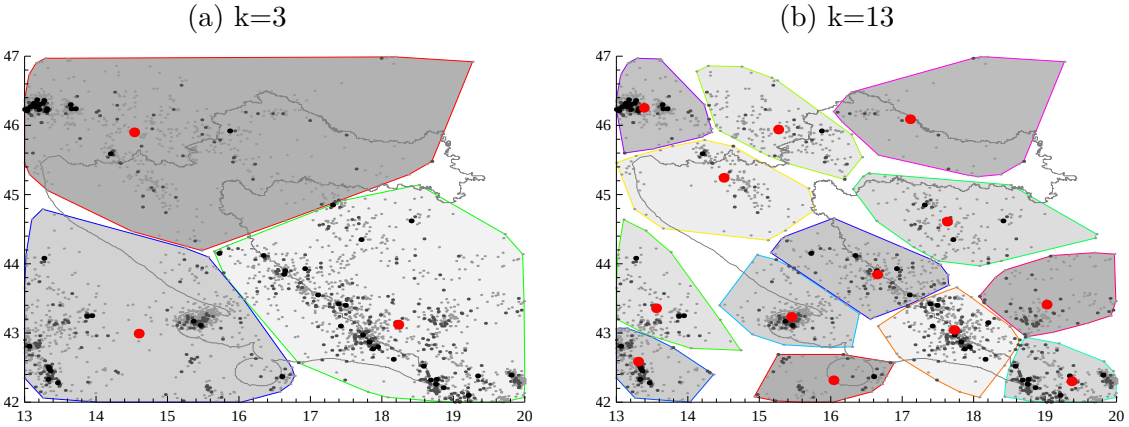


Figure 5: WM-optimal partitions

A WM-optimal partition with $k = 13$ clusters points out at 13 locations in which the most intense seismic activity in the observed area can be expected (see Fig. 5b). Centers of these clusters and locations with a geographical position closest to these clusters

are shown in Table 4. The number of all earthquakes N_{all} and the number N_5 of corresponding stronger earthquakes ($M_i \geq 5$) for each cluster are also shown in the table previously mentioned. The most intense seismic activity appears in clusters with centers near Metković $(17.64, 43.05)$, Podgorica $(19.26, 42.44)$, and Udine $(13.23, 46.07)$. If the databases with longer time period were used, it could be expected that the results are more realistic.

k	WM-centers	N_5/N_{all}	Location	Position
1	(17.73, 43.04)	11/400	Metković (HR)	(17.64, 43.05)
2	(14.50, 45.24)	2/152	Rijeka (HR)	(14.44, 45.32)
3	(15.26, 45.94)	2/200	Novo mesto (SLO)	(15.17, 45.80)
4	(13.56, 43.36)	3/165	Recanati (IT)	(13.54, 43.40)
5	(17.63, 44.61)	4/175	Banja Luka (BiH)	(17.18, 44.76)
6	(19.37, 42.30)	11/462	Podgorica (Montenegro)	(19.26, 42.44)
7	(15.45, 43.23)	3/450	Šibenik (HR)	(15.89, 43.73)
8	(13.30, 42.59)	7/258	L'Aquila (IT)	(13.39, 42.34)
9	(16.65, 43.85)	6/272	Sinj (HR)	(16.63, 43.70)
10	(13.38, 46.25)	20/329	Udine (IT)	(13.23, 46.07)
11	(17.11, 46.09)	2/60	Koprivnica (HR)	(16.83, 46.16)
12	(19.02, 43.41)	1/159	Goražde (BiH)	(18.98, 43.66)
13	(16.04, 42.32)	1/102	Rodi Garganico (IT)	(15.88, 41.92)

Table 4: Cluster centers of a WM-optimal partition with 13 clusters

It is also interesting to analyze the geometric position of WM-optimal cluster centers $c_1^\star, \ldots, c_{13}^\star$. Let $\mathcal{K}(c_0, \rho) = \{x \in \mathbb{R}^2 \colon d_M(c_0, x) = \rho\}$ be a Mahalanobis circle (M-circle) with radius ρ and centre c_0 and let $\{x_i \in \mathbb{R}^2 \colon i = 1, \ldots, s\}$ be a given set of points in the plane. The optimal M-circle can be determined (Nievergelt, 2002) by solving the global optimization problem

$$\operatorname*{argmin}_{c_0,\rho} G(c_0, \rho), \qquad G(c_0, \rho) = \sum_{i=1}^{s} |d_M(c_0, x_i) - \rho|.$$

Hence, for points $c_5^\star, c_7^\star, c_9^\star, c_{13}^\star$ we get an M-circle $\mathcal{K}_1((18.53, 42.65), 2.11)$, and for points

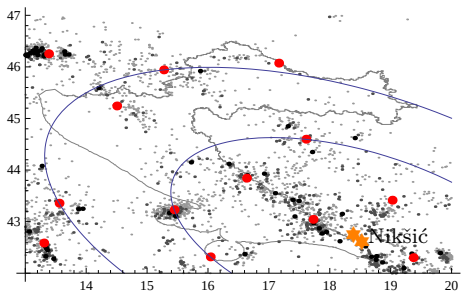


Figure 6: Geometric position of WM-optimal cluster centers

$c_2^\star, c_3^\star, c_4^\star, c_8^\star, c_{11}^\star$ an M-circle $\mathcal{K}_2((18.39, 42.79), 5.51)$ (see Fig. 6). It is interesting to notice that the centers of M-circles $\mathcal{K}_1$, $\mathcal{K}_2$ almost coincide (d_M the distance of their centers is 0.1) near Nikšić (Montenegro) with geographical position $(18.94, 42.77)$, and in the observed period (1973-2012) in the close vicinity of that place (d_M away from Nikšić for 0.25 at most) there have been more than 80 earthquakes of magnitude greater than 4, out of which 17 earthquakes had a magnitude greater than 5.

It is possible to carry out an additional necessary analysis for each cluster, that is, we can observe occurrence of earthquakes in some cluster in the whole period of time or analyze locations and intensity of these processes just in some selected time interval, etc. Similarly, in (Borghi et al., 2009), GPS monitoring and earthquake prediction for the cluster with the center near Udine (Italy) is investigated.

Identified centers of the most intense seismic activity in the observed area confirm the results obtained in other ways (Ivančić et al., 2006; Stipčević et al., 2011), and the proposed method can be used in the work of other researchers as in (Adelfio et al., 2012; Colombo et al., 1997; Morales-Esteban et al., 2010).

5 Conclusion

The problem of determining an optimal partition of the set $\mathcal{A} \subset \mathbb{R}^n$ is a complex global optimization problem. Therefore, in recent literature special emphasis is put on construction of an efficient algorithm of searching for a partition which is as close to the optimal one as possible. The algorithm proposed in this paper is a generalization of already known incremental algorithms of searching for an optimal partition able to find either a globally optimal partition or a locally optimal partition close to the global one. Thereby, in each iteration a well-known `DIRECT` algorithm for global optimization and the k-means algorithm are combined. We should point out high efficiency of the proposed algorithm which requires significantly shorter CPU-time than other incremental algorithms. Since optimal partitions with $2, 3, \ldots$ clusters are determined successively, another advantage of the proposed algorithm is the possibility of proposing an appropriate number of clusters in a partition by calculating corresponding clustering validity indexes.

It has been shown that a complex problem of detecting seismic activity centers can be efficiently solved by using the proposed algorithm. The results obtained in such way can be useful for further analysis and prediction of seismic activity (Morales-Esteban et al., 2010). In addition to that, the proposed algorithm could be used very efficiently in other complex applications as well.

Acknowledgements
The authors would like to thank Prof. Kristian Sabo and Dr. Ivan Vazler (University of Osijek, Croatia) for their useful comments and remarks. We are also thankful to anonymous referees and journal editors for their careful reading of the paper and insightful

comments that helped us improve the paper.

This work was supported by the Ministry of Science, Education and Sports, Republic of Croatia, through research grant 235-2352818-1034.

References

Adelfio, G., M.Chiodi, A.D'Alessandro, D.Luzio, G.D'Anna, G.Mangano, 2012. Simultaneous seismic wave clustering and registration. Computers & Geosciences 44, 60–69.

Bagirov, A.M., 2008. Modified global k-means algorithm for minimum sum-of-squares clustering problems. Pattern Recognition 41, 3192–3199.

Bagirov, A.M., Ugon, J., 2005. An algorithm for minimizing clustering functions. Optimization 54, 351–368.

Bagirov, A.M., Ugon, J., Webb, D., 2011. Fast modified global k-means algorithm for incremental cluster construction. Pattern Recognition 44, 866–876.

Bagirov, A.M., Yearwood, J., 2006. A new nonsmooth optimization algorithm for minimum sum-of-squares clustering problems. European Journal of Operational Research 170, 578–596.

Borghi, A., Aoudia, A., Riva, R.E., Barzaghi, R., 2009. GPS monitoring and earthquake prediction: A success story towards a useful integration. Tectonophysics 465, 177–189.

Cho, N.F., Tiampo, K.F., Mckinnon, S.D., Vallejos, J.A., Klein, W., Dominguez, R., 2010. A simple metric to quantify seismicity clustering. Nonlin. Processes Geophys. 17, 293–302.

Colombo, D., Gitis, V., de Franco, R., 1997. Application of pattern recognition techniques to long-term earthquake prediction in central Costa Rica. Engineering Geology 48, 7–18.

Davies, D., Bouldin, D., 1979. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 2, 224–227.

Durak, B., 2011. A Classification Algorithm Using Mahalanobis Distances Clustering of Data with Applications on Biomedical Data Set. Ph.D. thesis. The Graduate School of Natural and Applied Sciences of Middle East Technical University.

Evtushenko, Y.G., 1985. Numerical Optimization Techniques (Translations Series in Mathematics and Engineering). Springer-Verlag, Berlin.

Finkel, D.E., 2003. DIRECT Optimization Algorithm User Guide. Center for Research in Scientific Computation. North Carolina State University. `http://www4.ncsu.edu/definkel/research/index.html`.

Floudas, C.A., Gounaris, C.E., 2009. A review of recent advances in global optimization. Journal of Global Optimization 45, 3–38.

Gablonsky, J.M., 2001. DIRECT Version 2.0. Technical Report. Center for Research in Scientific Computation. North Carolina State University.

Gan, G., Ma, C., Wu, J., 2007. Data Clustering: Theory, Algorithms, and Applications. SIAM, Philadelphia.

Grbić, R., Nyarko, E.K., Scitovski, R., 2012. A modification of the DIRECT method for Lipschitz global optimization for a symmetric function. Journal of Global Optimization Published online: 23 December 2012.

Holliday, J.R., Rundle, J.B., Tiampo, K.F., Turcotte, D.L., 2006. Using earthquake intensities to forecast earthquake occurrence times. Nonlinear Processes in Geophysics 13, 585–593.

Ivančić, I., Herak, D., Markušić, S., Sović, I., Herak, M., 2006. Seismicity of Croatia in the period 2002–2005. Geofizika 23, 87–103.

Iyigun, C., 2007. Probabilistic Distance Clustering. Ph.D. thesis. Graduate School – New Brunswick, Rutgers.

Jones, D.R., Perttunen, C.D., Stuckman, B.E., 1993. Lipschitzian optimization without the Lipschitz constant. Journal of Optimization Theory and Applications 79, 157–181.

Kaufman, L., Rousseeuw, P.J., 2005. Finding groups in data: An introduction to cluster analysis. Jonh Wiley & Sons, Hoboken.

Kogan, J., 2007. Introduction to Clustering Large and High-dimensional Data. Cambridge University Press.

Leisch, F., 2006. A toolbox for k-centroids cluster analysis. Computational Statistics & Data Analysis 51, 526–544.

Likas, A., Vlassis, N., Verbeek, J.J., 2003. The global k-means clustering algorithm. Pattern Recognition 36, 451–461.

Modirzadeh, M., Tesfamariam, S., Milani, A.S., 2012. Performance based earthquake evaluation of reinforced concrete buildings using design of experiments. Expert Systems with Applications 39, 2919–2926.

Morales-Esteban, A., Martínez-Álvarez, F., Troncoso, A., Justo, J., Rubio-Escudero, C., 2010. Pattern recognition to forecast seismic time series. Expert Systems with Applications 37, 8333–8342.

Neumaier, A., 2004. Complete search in continuous global optimization and constraint satisfaction. Acta Numerica 13, 271–369.

Nievergelt, Y., 2002. A finite algorithm to fit geometrically all midrange lines, circles, planes, spheres, hyperplanes, and hyperspheres. Numerische Mathematic 91, 257–303.

Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Dumitrescu, S.C.A., Gaume, E., Hlavčová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J., Blöschl, G., 2010. Seasonal characteristics of flood regimes across the Alpine–Carpathian range. Journal of Hydrology 394, 78–89.

Pardalos, P.M., Coleman, T.F. (Eds.), 2009. Lectures on Global Optimization. volume 55 of *Fields Institute Communications Series*, AMS.

Pintér, J.D., 1996. Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications). Kluwer Academic Publishers, Dordrecht.

Rizman-Žalik, K., 2008. An efficient k'-means clustering algorithm. Pattern Recognition Letters 29, 1385–1391.

Sabo, K., Scitovski, R., Vazler, I., 2013. One-dimensional center-based l_1-clustering method. Optimization Letters 7, 5–22.

Sabo, K., Scitovski, R., Vazler, I., Zekić-Sušac, M., 2011. Mathematical models of natural gas consumption. Energy Conversion and Management 52, 1721–1727.

Sergeyev, Y.D., Kvasov, D.E., 2011. Lipschitz global optimization, in: Cochran, J. (Ed.), Wiley Encyclopedia of Operations Research and Management Science. Wiley, New York. volume 4, pp. 2812–2828.

Späth, H., 1983. Cluster-Formation und Analyse. R. Oldenburg Verlag, München.

Stipčević, J., Tkalčić, H., Herak, M., Markušić, S., Herak, D., 2011. Crustal and uppermost mantle structure beneath the External Dinarides, Croatia, determined from teleseismic receiver functions. Geophys. J. Int. 185, 1103–1119.

Teboulle, M., 2007. A unified continuous optimization framework for center-based clustering methods. Journal of Machine Learning Research 8, 65–102.

Vazler, I., Sabo, K., Scitovski, R., 2012. Weighted median of the data in solving least absolute deviations problems. Communications in Statistics - Theory and Methods 41:8, 1455–1465.

Vendramin, L., Campello, R.J.G.B., Hruschka, E.R., 2009. On the comparison of relative clustering validity criteria, in: Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 - May 2, 2009, Sparks, Nevada, USA, SIAM. pp. 733–744.

Volkovich, V., Kogan, J., Nicholas, C., 2007. Building initial partitions through sampling techniques. European Journal of Operational Research 183, 1097–1105.