J Math Chem (2012) 50:2703–2715
DOI 10.1007/s10910-012-0059-x

ORIGINAL PAPER

A benchmark calculation for the fuzzy c-means clustering algorithm: initial memberships

Alex Rodriguez · M. Santos Tomas · Jaime Rubio-Martinez

Received: 6 April 2012 / Accepted: 15 July 2012 / Published online: 3 August 2012

Abstract We report a benchmark calculation for the fuzzy c-means clustering algorithm that can be used as a reference in theoretical and practical studies related to classification methodologies. A full exploration of the hard-initialization space is done for all possible different groupings on a simple fifteen-pattern system to describe their stationary points. Numerical problems associated with the stopping criteria are discussed in relation with the calculation of some validity indexes. All necessary information to assure an easy reproduction of the obtained results is clearly reported.

Keywords Clustering · Fuzzy c-means · Benchmark

1 Introduction

Clustering is a process to partition a given set of data or objects into classes or clusters having two fundamental properties: homogeneity within the clusters, i.e. objects that belong to the same cluster have high similarity; and heterogeneity between clusters, i.e. objects that belong to different clusters are very dissimilar The goal of any clustering

A. Rodriguez
Department d'Enginyeria Química, Universitat Politècnica de Catalunya (UPC),
ETS d'Enginyeria Industrial, Av. Diagonal 647, 08028 Barcelona, Spain

M. S. Tomas
Sec. Matemàtiques i Informàtica ETSAB, Universitat Politècnica de Catalunya (UPC),
ETS d'Arquitectura, Av. Diagonal 649, 08028 Barcelona, Spain

J. Rubio-Martinez (✉)
Dept. de Química Física, Universitat de Barcelona (UB) and the Institut de Recerca en Química Teòrica i Computacional (IQTCUB), Martí i Franqués 1, 08028 Barcelona, Spain
e-mail: jaime.rubio@ub.edu

method is the classification of a data set X, composed by n pattern, $X = \{x_1, x_2, \ldots x_n\}$ where each pattern is described by p properties, into c clusters ($C_1, C_2, \ldots C_c$). The pattern matrix U of dimension c x n, represented by $U = [\mu_{ij}]$, $i = 1, 2, \ldots$ c and $j = 1, 2, \ldots$ n, describes the membership degree of data x_j to the cluster C_i. This methodology has been successfully applied in various fields and numerous methods of clustering have been proposed [1,2] that in general assign each data point of the data set to only one cluster generating a crisp or hard partition. However, fuzzy clustering, and in particular the fuzzy c-means algorithm (FCM) has clear advantages because of its concept of fuzzy membership [3]. This algorithm belongs to the unsupervised classification methods and thus, requires the number of clusters c as input. Once the number of clusters has been selected, an initial U matrix is defined and an iterative process is carried out to optimize an objective function J that includes Euclidean distance between data points and cluster centres. For the FCM algorithm, Bezdek et al. [4] proved its convergence, at least along a subsequence, to either a local minimizer or a saddle point of its objective function. Performance of iterative clustering algorithms depends highly on initial membership values that condition convergence and produce different local minima. Although some approaches have been described to overcome this problem [5–9], most references do not address this issue or simple either a user-specific or randomly selected initial points are used [10,11]. As stated in [12] "there is no simple, universally good solution to this problem". Despite of the importance of the initialisation problem, at the author's knowledge, any systematic study of the possible hard initial values for an n-pattern problem has been reported up to the data.

The fuzzy c-mean method has been widely used in the chemistry field [13–19], covering different application ranging from the analysis of molecular dynamic trajectories of proteins and polypeptides [13,14] to Metabolomics [19], but none of these works studied the problem in depth.

In this paper we report for the first time a full hard initialisation study for all possible c-clusters in a 15-pattern system, describing all different local stationary points. The obtained values are accurately described in order to make easier the results reproduction and its comparison with results derived from other approaches. Two widely used validity indexes are also computed to provide reproducible values for these important parameters. The paper is organised as follow: In the next section we review the Fuzzy c-means clustering methodology, with a special emphasis on the computational algorithm, and also on some validity indexes; next we describe how the full hard initial membership were generated and finally numerical experiments are presented and discussed.

2 Fuzzy clustering algorithm

The original FCM algorithm that recognizes spherical clouds of points in a p-dimensional space was firstly described by Bezdek [3,20–24]. This method attempts to find a representative point for each cluster which is called prototype or centroid, and the fuzzy partition matrix μ_{ij} by minimizing an objective function J. Despite numerous modifications have been proposed from the original method [25–27], it still remains widely used.

Next we describe the basic FCM algorithm making special emphasis in the explanation of when and how the values involved in the convergence process are calculated.

FCM algorithm

Step 1: Input the number of clusters c, the fuzzifier m and the test criterion to finish the process ε, with $\varepsilon = \varepsilon_{J_m}, \varepsilon_\mu \ or \ \varepsilon_v$

Step 2: Set the iteration number $k = 0$ and initialize the fuzzy partition matrix $\mu_{ij}^0 (i = 1, 2, \ldots c;\ j = 1, 2, \ldots n)$ satisfying:

$$\begin{cases} 0 < \sum_{j=1}^{n} \mu_{ij} < n \quad for\ i = 1, 2, \ldots c & (1a) \\ \sum_{i=1}^{c} \mu_{ij} = 1.0 \qquad for\ j = 1, 2, \ldots n & (1b) \\ \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} = n & (1c) \end{cases}$$

Step 3: Do step 4 for $k = 0$

Step 4: Compute fuzzy cluster centers v_i^k(i = 1, 2, . . .c), Euclidean distance of data to the cluster centres (d^k) and objective function (J_m^k) using:

$$v_i^k = \frac{\sum_{j=1}^{n} (\mu_{ij})^m x_j}{\sum_{j=1}^{n} (\mu_{ij})^m} \quad for\ 1 \le i \le c \tag{2}$$

$$d_{ij}^k = ||x_j - v_i|| \quad for \quad 1 \le i \le c \ , \ 1 \le j \le n \tag{3}$$

$$J_m^k = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m ||x_j - v_i||^2 \tag{4}$$

Step 5: Calculate the fuzzy partition matrix for the iteration k μ_{ij}^k(i = 1, 2, . . .c; j = 1, 2, . . .n) as:

$$\begin{cases} \mu_{ij}^k = \left[\sum_{s=1}^{c} \left(\frac{||x_j - v_i||^2}{||x_j - v_s||^2} \right)^{\frac{1}{m-1}} \right]^{-1} & if ||x_j - v_s|| > 0 \quad \forall j & (5a) \\ \mu_{ij}^k = 1 & if\ ||x_j - v_i|| = 0 & (5b) \\ \mu_{ij}^k = 0 & if \exists s \ne i ||x_j - v_s|| = 0 & (5c) \end{cases}$$

Step 6: If the selected convergence criterion ε is satisfied go to Step 7; else go to Step 4.

Step 7: Output and store information about final values for v_i, μ_{ij}, d_{ij} and J_m.

Step 8: Calculate Hessian and performance measures.

Step 9: Stop

In Step 1 convergence criterions $\varepsilon_{J_m}, \varepsilon_\mu$ *and* ε_v are defined as:

$$\max |J_m^k - J_m^{k-1}| \leq \varepsilon_{J_m} \; ; \; \max |\mu_{ij}^k - \mu_{ij}^{k-1}| \leq \varepsilon_\mu ; \max |v_j^k - v_j^{k-1}| \leq \varepsilon_v$$

Very small values of ε_{J_m} are necessary to achieve a good convergence on v_i^k and μ_{ij}^k values. This fact must be taken into account to obtain accurate calculations of performance measures.

Step 3 is separated from step 4 in the scheme of the FCM algorithm to remark the fact that we need to have all necessary values to initiate the iterative process.

In Step 8 the Hessian $H_{ij} = \frac{\partial^2 R_m(v)}{\partial v_i \partial v_j}$ of the reformulated metric $R_m(v)$ [28] is calculated using

$$R_m(v) = \sum_{j=1}^{n} \left[\sum_{i=1}^{c} ||x_j - v_i||^{2/(1-m)} \right]^{(1-m)}$$

and its eigenvalues obtained to characterise the stationary points [29,30]. Also in this step validity indices introduced by Xie and Beni [31] and Fukuyama and Sugeno [32] are calculated as:

$$V_{XB} = \frac{J_m(u, v)/n}{Sep(v)} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m ||x_j - v_i||^2}{n \; \min_{i \neq j} ||v_i - v_j||^2}$$

$$V_{FS} = J_m(u, v) - K_m(u, v) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m ||x_j - v_i||^2 - \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m ||x_j - \bar{v}||^2$$

where $\bar{v} = \frac{1}{c} \sum_{i=1}^{c} v_i$ is the mean of the cluster centroids.

3 Generation of initial membership

Although it has been demonstrated that the iterative process used to solve the FCM equations always converges to a stationary point, it seems clear that there is not a general initialization method that can be used for all theoretical or real problems. However, different methods have been suggested for initializing cluster centres, basically devoted to the K-means algorithm [8,9,33–35]. Also, a comparison of four different initialization methods for this algorithm has been reported [36] suggesting the random initialization as one of the best methods.

Thus, to try to overcome this problem, an exhaustive hard-generation of initial points has been done. However, if we define a hard c-partition space for data X as the matrix set

$$M_c = \left\{ U \; | \mu_{ij} \in \{0, 1\} \; , \; \sum_{i=1}^{c} \mu_{ij} = 1, \;\; 0 < \sum_{j=1}^{n} \mu_{ij} < n \right\}$$

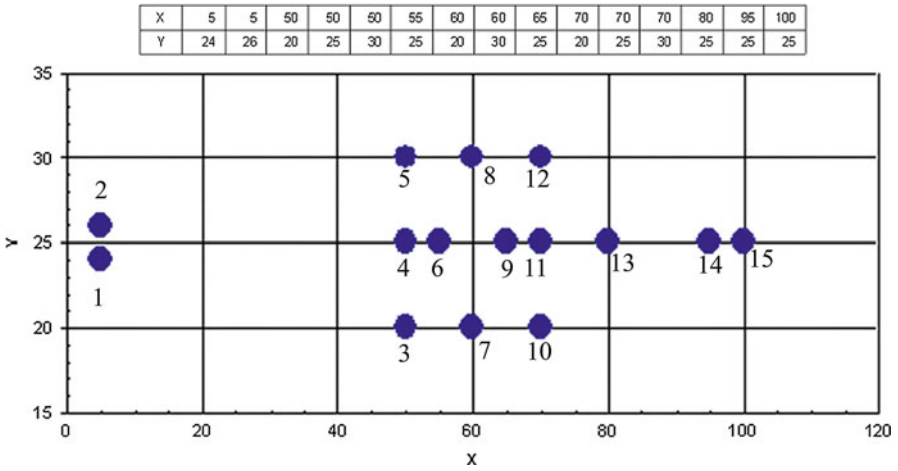

Fig. 1 Graphical representation of the fifteen data points used to perform the fuzzy c-means calculations

then, any matrix $U \in M_c$ is a hard c-partition with cardinality

$$\eta_{M_c} = \frac{1}{c!}\left[\sum_{i=1}^{c}\binom{c}{i}(-1)^{c-i}i^n\right]$$

It is easy to see that for a given c-partition this number becomes unaffordable when the number of data points increase. Thus, to retain the cardinality of the problem in a computational feasible dimension, it is necessary to study a reasonable small system. A dimension of fifteen was considered adequate to maintain equilibrium between the number of data points and the cardinality. In Fig. 1 it is shown a graphical representation of the selected points including its numerical values for reproducibility purposes. Cardinality for all possible clusters that can be generated from fifteen data points is showed in Table 1. Although it includes big numbers these magnitudes are still computationally affordable..

Different computational approximations were tested to generate all hard initialization possibilities for a given c-partition, but finally approximation showed in Fig. 2 was used. Concretely, an example of fourth clusters (N_clus = 4) and fifteen data points (N_data = 15) is showed. This approximation, although time consuming, is affordable from a computer memory resources point of view and the program works storing all necessary data in memory. Thus, once the number of clusters (N_clus) for a given number of data points (N_data) has been selected, the program proceeds as follows: Firstly, all possible non-redundant distributions of the data between the clusters are generated, annotating those clusters with the same occupation in order to generate only non-redundant combinations (Fig. 2a). Next, all possible combinations of N_data over all cluster occupations appearing in any of the distributions are generated and stored in main memory (Fig. 2b). Finally, N_clus nested do are done (Fig. 2c). Inside each do, a test to discard non-unique hard occupations is performed and do limits are modified to take into account possible equal cluster occupations. In Fig. 2d a combination of data

Table 1 Cardinality values for all possible clusters that can be generated from fifteen data points

Number of clusters	Cardinality
2	16,383
3	2,375,101
4	42,355,950
5	210,766,920
6	420,693,273
7	408,741,333
8	216,627,840
9	67,128,490
10	12,662,650
11	1,479,478
12	106,470
13	4,550
14	105

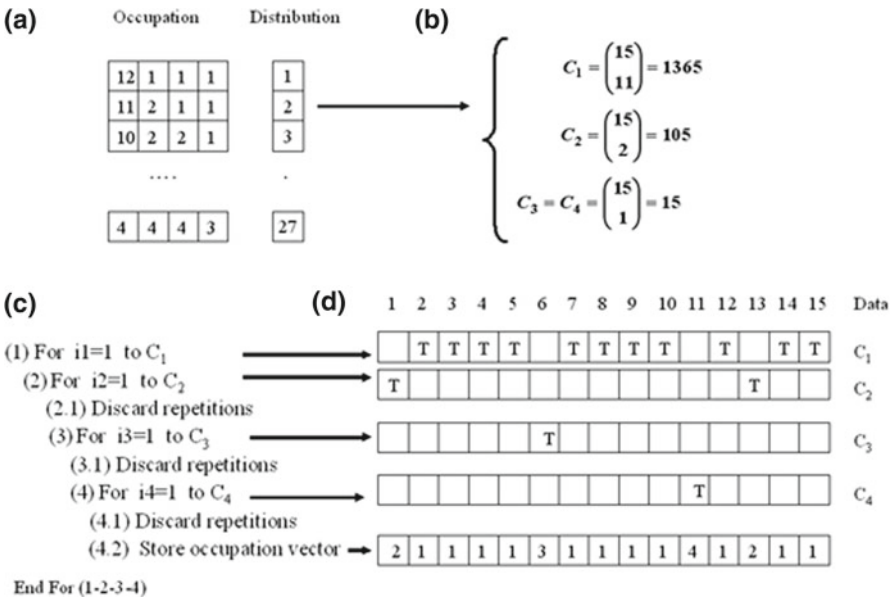

Fig. 2 A schematic description of the process used to generate a full hard-distribution for a case with four clusters and fifteen data points. **a** Generation of the 27 possible non-equivalent distributions. **b** Generation of all possible combinations for each cluster occupation. **c** Four nested do to compare occupation of each combination for each cluster. **d** An example of one combination for each cluster that generated a proper partitioning

that generated a proper cluster distribution is showed. At the end of this procedure, the occupation vector for each possible value of the hard N_data$_{N_clus}$ partition is stored in a file that can be used as input in the fuzzy c-means computer program.

4 Numerical experiments and discussion

As the main goal of this article is to generate a reference calculation for the FCM methodology, our first objective was to asses that our reported values were really stables. Thus, we tested convergence of the validity indices of Xie and Beni (V_{FS}) [31] and Fukuyama and Sugeno (V_{FS}) [32] at the end of the iterative process of the FCM algorithm as a function of different convergence criterions. In Tables 2 and 3 we show the V_{XB} and V_{FS} values obtained stopping the FCM iterative process with different convergence criterions (ε_{J_m}, ε_{μ}) and using different precision ε. As it can be seen, both indices clearly present very different behaviour. Thus, V_{XB} index reaches a stable value at high precision values independently of the criterion used to stop the iterative process. However, the V_{FS} index only reaches convergence in three decimal numbers at very small values of ε reflecting the more complex structure of this validity index. As expected, when using ε_{J_m} smaller values are needed to obtain a desired precision than when using the ε_{μ} criterion. We can conclude that an important dispersion is presented in the obtained results, mainly for big convergence criterion values and that this variability is index dependent. Thus, to compare validity indices for different clusters, in order to locate the best number of clusters, the use of low values of ε seems

Table 2 Values of Xie and Beni validity index (V_{XB}) for our data using c = 2, m = 2.0, 16, 383 full hard initial trials and different convergence criterion defined as (1) $\varepsilon = \max|U_{ij}(n) - U_{ij}(n-1)|$ or (2) $\varepsilon = \max|-J_m(U,v)(n) - J_m(U,v)(n-1)|$

Convergence criterion (ε)	Maximum (1) V_{XB}	Minimum (1) V_{XB}	Maximum (2) V_{XB}	Minimum (2) V_{XB}
10^{-2}	0.05527	0.05517	0.05521	0.05519
10^{-3}	0.05520	0.05519	0.05520	0.05520
10^{-4}	0.05520	0.05520	0.05520	0.05520

Five decimal numbers are showed in all cases for comparison

Table 3 Values of Fukuyama and Sukeno validity index (V_{FS}) for our data using c = 2, m = 2.0, 16,383 full hard initial trials and different convergence criterion defined as (1) $\varepsilon = \max|U_{ij}(n) - U_{ij}(n-1)|$ or (2) $\varepsilon = \max|-J_m(U,v)(n) - J_m(U,v)(n-1)|$

Convergence criterion (ε)	Maximum (1) V_{FS}	Minimum (1) V_{FS}	Maximum (2) V_{FS}	Minimum (2) V_{FS}
10^{-2}	−9,194.04171	−9,245.77509	−9,226.00548	−9,236.23064
10^{-3}	−9,229.18765	−9,236.01727	−9,230.62347	−9,233.68238
10^{-4}	−9,232.36778	−9,232.98419	−9,232.05999	−9,232.98419
10^{-5}	−9,232.68238	−9,232.73777	−9,232.51010	−9,232.79124
10^{-6}	−9,232.71375	−9,232.71882	−9,232.65184	−9,232.73777
10^{-7}	−9,232.71689	−9,232.71736	−9,232.69664	−9,232.73777
10^{-8}	−9,232.71720	−9,232.71727	−9,232.71070	−9,232.72294
10^{-9}	−9,232.71724	−9,232.71724	−9,232.71517	−9,232.71882

Five decimal numbers are showed in all cases for comparison

critical, independently of the specific criterion used. Moreover, the use of a criterion based on membership values seems to be the best choice (similar results were obtained using ε_v as convergence criterion. Data not show). To minimise possible errors derived from the selection of the convergence criterion, all reported calculations were obtained using a value of $\varepsilon_\mu = 10^{-8}$. A value of m = 2.0 and a maximum of 500 iterations for the convergence process was uses anywhere. As stationary points with negative eigenvalues of $R_m(v)$ do not represent logic partitions, its validity indices were not calculated. On the other hand, two solutions were considered different if they differs in more than ε_v.

Calculations for two clusters involve 16,383 possible values for a full hard initialization of memberships (see Table 1). In this simple case, only two stationary points were located by the FCM algorithm (see Table 4). The first one, with all the eigenvalues of $R_m(v)$ positives, contains points 1 and 2 in the first cluster and all the other in the second cluster. The second one, with one negative eigenvalue of $R_m(v)$, has an equal occupation of $\mu_{ij} = 0.5$ for all data; that is, the two cluster centres are located at the same point (59.00000, 25.00000). It is worth to note that we use a value of $\mu_{ij,\,hard} = 0.5$ to transform a fuzzy partition into a hard partition. Thus, if there is not an occupation greater than this value for a point j, we consider that this point belong to a fuzzy distribution along the most representative clusters.

Three stationary points were located for the three clusters case: one with zero negative eigenvalues of $R_m(v)$ and two with one negative eigenvalue. The last two stationary points have two equal cluster centres and one different. The first one, with all the eigenvalues of $R_m(v)$ positives, contains points 1 and 2 in the first cluster, points 13, 14 and 15 belong to the second cluster and all the remaining points are included into the third cluster.

Two stationary points with zero negative eigenvalues of $R_m(v)$ were located for the four clusters case whose hard occupations are 1–2, 3–8, 9–13, 14–15 and 1, 2, 3–12, 13–15 for the first and second stationary points, respectively. For this case, there are some stationary points with two degenerate negative eigenvalues. When this occurs, three clusters centres have the same value. When the two negative eigenvalues are not degenerated, two pair of equal clusters centres or two equal clusters centres exists. One negative eigenvalue indicates that two equal clusters centres exist.

For the five clusters case, three stationary points with zero negative eigenvalues of $R_m(v)$ were located. The first and third stationary points, in increasing evaluation function values, have as a hard occupation of 1–2, 3–6, 7–9, 10–13, 14–15 and 1, 2, 3–8, 9–13, 14–15. However, for the second stationary point, a hard occupation is not possible using the abovementioned $\mu_{ij,hard}$ hard limit value. Thus, points 7 and 8 have an occupation of 0.38838 and 0.48284 for the second and third clusters being the other points distributed over the different clusters as 1–2, 3–5, 9–12, 13, 14–15.

The six clusters case has the biggest cardinal value and sixteen stationary point were located, ten of then having zero negative eigenvalues of $R_m(v)$. The evaluation functions of the two best stationary points have very similar values (138.37079 and 138.73261) differing in the distributions of points between the clusters. Thus, for the first one, point seven has fuzzy occupation distributed in clusters two, three and four (with 0.24617, 0.26005 and 0.40620 μ_{ij} values) with the other points having a hard occupation of 1–2, 3–6, 8, 9–12, 13, 14–15. For the second one, we can assign a hard

Table 4 Representative data of the obtained results using a full hard-initialisation procedure for two to six clusters

Number of clusters	Stationary points (1)	Objective function	Number of evaluations (2)	Smallest eigenvalue (3)	Negative eigenvalues (4)	V_{XB}	V_{FS}
2	2/1	2,995.459278	16,370	3.43054	0	0.05520	−9,232.71722
		5,031.000000	13	−17.02415	1		
3	3/1	858.696145	2,373,500	2.46746	0	0.05218	−8,108.88400
		1,580.570219	1,299	−13.68222	1		
		2,556.696520	302	−7.95180	1		
4	8/2	307.957701	42,180,936	3.21391	0	0.07630	−9,012.45779
		832.904872	34,922	1.24481	0	14.37092	−10,885.26464
		501.754495	56,687	−8.84470	2		
		662.354793	79,489	−6.37942	1		
		833.493697	3,705	−1.13395	2		
		1,073.132322	175	−6.24503	2 (deg)		
		1,497.729639	30	−12.55117	2		
		1,904.782205	6	−4.67244	2 (deg)		
5	20/3	201.19437	190,888,997	1.04937	0	0.18242	−8,998.15388
		205.77839	13,497,009	0.99827		0.10067	−9,224.08259
		302.61311	4,502,501	1.98363		5.05390	−10,690.47324
6	60/10	138.37079	55,167,339	0.53102	0	0.10828	−9,308.42976
		138.73261	17,121,904	0.05103		0.15097	−9,289.51701
		151.50788	53,301,409	0.33634		0.25817	−9,031.76720
		153.26422	105,825,466	0.50825		0.32558	−8,934.67725
		153.31950	30,402,476	0.28445		0.21656	−8,957.59314
		153.41123	40,187,011	0.92197		0.14494	−8,990.91970

Table 4 continued

Number of clusters	Stationary points (1)	Objective function	Number of evaluations (2)	Smallest eigenvalue (3)	Negative eigenvalues (4)	V_{XB}	V_{FS}
		179.52214	2,369,230	0.82422		0.40781	−9,545.98140
		189.81637	94,151	0.94709		0.49634	−10,113.48344
		197.85323	17,999,769	1.03607		3.30040	−9,734.45059
		202.08816	2,730,206	1.03513		3.37219	−9,710.59544

Only stationary points with no negative eigenvalues are showed for five and six clusters and validity indexes are only showed for these types of stationary points. (1) Total number of stationary points and number of stationary points with no negative eigenvalues. (2) Number of evaluations standing for the number of times a particular solution appears over the total number of initial points. (3) Smallest eigenvalue of $R_m(\nu)$. (4) Number of negative eigenvalues of $R_m(\nu)$for each stationary point

occupation as 1–2, 3–5, 6–8, 9–12, 13, 14–15. Another interesting stationary point corresponds to the most populated distribution. This distribution has points seven and eight with fuzzy occupation (0.41282 and 0.32125) over clusters three and four and the other points with a hard occupation of 1–2, 3–5, 6, 9, 10–13, 14–15.

As it can be seen in Table 4, as the number of cluster increases, the number of located stationary points increases. This is also true for the number of stationary points having all its $R_m(\nu)$ eigenvalues positives. On the other hand, up to five clusters the number of evaluations that fail into the smallest evaluation function is clearly bigger than the other possibilities. Thus, it is expected that any reasonable initialization method would conduct easily to the right solution.

However, when the number of clusters is set to 6, the number of evaluations that fail into the smallest evaluation function value is not the most important. There are other stationary points with higher number of occurrences. As a practical test we performed a calculation with 5×10^3 hard-random initializations to compare with our full results. The two smallest stationary points appeared 536 and 168 times respectively and the most populated 1795 times. Another remarkable point was that only ten stationary points were located, but corresponding exactly to the ten having zero negative eigenvalues of $R_m(\nu)$. On the other hand, a calculation with 10^5hard-random initializations adds only one stationary point with a negative eigenvalue of $R_m(\nu)$ to the ten located in the smallest test.

For cases with a number of clusters greater than six, the number of stationary points found is really important. For this reason, only the one with the smallest evaluation function value for each number of clusters is showed in Table 5. In these cases, even the best partition need the use of a fuzzy occupation.

Results from the present particular system indicate that every time a stationary point has negative eigenvalues for $R_m(\nu)$, cluster centre values are not unique. Thus, this fact can also be used to asses the validity of the stationary point located.

Table 5 Representative data of the obtained results using a full hard-initialisation procedure for seven to fourteen clusters

Number of clusters	Stationary points (1)	Objective function	Number of evaluations (2)	Smallest eigenvalue (3)	V_{XB}	V_{FS}
7	176/31	97.94157	43,470,701	1.61093	0.08080	−9,338.55509
8	466/68	71.29311	36,186,836	0.37056	0.07810	−9,361.57394
9	937/105	49.89783	3,309,444	0.45042	0.05276	−9,483.72244
10	1119/122	37.13382	981,138	0.71280	0.05738	−9,482.79322
11	819/113	25.15625	40,712	0.73422	0.06702	−10,046.92061
12	350/74	14.39228	3,994	1.44314	0.03836	−10,090.23500
13	70/31	7.926580	755	1.52584	0.02115	−10,115.68385
14	10/7	1.992189	27	2.00000	0.00531	−10,228.97848

Only the best stationary point with no negative eigenvalues for each distribution is showed. (1) Total number of stationary points and number of stationary points with no negative eigenvalues. (2) Number of evaluations standing for the number of times a particular solution appears over the total number of initial points. (3) Smallest eigenvalue of $R_m(\nu)$

Another point to remark is that for distributions up to five clusters, all initialization points converged in the predefined number of iterations to some stationary point. However, from six to ten clusters systems many initial points fail to converge. As an example, for the six cluster case, 88,515,155 initial points do not converged.

As it is expected, evaluation function always decreases with the number of clusters. However, validity indexes have very different behaviour. Thus, XB validity index has a minimum value at c = 3, increases up to c = 5, has another minimum at c = 9, increases slightly its value and finally decreases until the end. On the other hand, FS validity index has a minimum value at c = 2, another minimum at c = 4 and decreases more ore less continuously to the end.

It is important to note that validity indexes values were only calculated for the solution with the best evaluation function. An important point but out of the scope of this paper, is to analyse the behaviour of different validity indexes when taking the best validity index from all the good solutions.

Complete calculation for c = 6 took 26 days in our Intel ® $Xeon^{TM}$ 2.8GHz personal computer using the Intel ® 10.1.015 FORTRAN compiler. The computer time needed to carry out this calculation is a clear indication of the difficulty to do this full calculation for bigger systems. However, even for this simple system, the complexity appears when the number of clusters is big and thus, the present results can be used as a test of effectiveness for other faster initialization methods.

Full numerical results for all the studied cases are available upon request to the authors.

Acknowledgments We acknowledge financial support from the Spanish Ministry of Science and Technology projects CTQ2011-29285-C02-02 and TIN2009-07235. Also to the Generalitat de Catalunya, project 2009SGR1308.

References

1. M.R. Anderberg, *Cluster Analysis for Application* (Academic Press, NewYork, 1973)
2. P.A. Devijver, J. Kittler, *Pattern Recognition: A Statistical Approach* (Prentice-Hall, London, 1982)
3. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, New York, 1981)
4. J.C. Bezdek, R.J. Hathaway, M.J. Sabin, et al. Convergence theory for fuzzy c-means: counterexamples and repairs. IEEE Trans. Syst., Man Cybernet. **SMC17**, 873–877 (1987)
5. S.J. Redmond, C. Heneghan, A method for initialising the K-means clustering algorithm using kd-trees. Pattern Recognit. Lett. **28**, 965–973 (2007)
6. S.S. Khan, A. Ahmad, Cluster center initialization algorithm for K-means clustering. Pattern Recognit. Lett. **25**, 1293–1302 (2004)
7. P.S. Bradley, U.M. Fayyad, in *Refining Initial Points for K-Means Clustering. Proceedings of the Fifteenth International Conference on Machine Learning*. pp. 91–99 (1998)
8. F. Cao, J. Liang, G. Jiang, An initialization method for the K-means algorithm using neighbourhood model. Comput. Math. Appl. **58**, 474–483 (2009)
9. F. Cao, J. Liang, L. Bai, A new initialization method for categorical data clustering. Expert Syst. Appl. **36**, 10223–10228 (2009)
10. E. Rasmussen, in *Clustering Algorithms, Information Retrieval Data Structures and Algorithms*, ed. by Frakes (Prentice Hall, New Jersey, 1992), pp. 419–442
11. L. Kaufman, P. Rousseeuw, *Finding Groups in Data* (Wiley, New York, 1989)
12. R.O. Duda, P.E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973)
13. M. Feher, J.M. Schmidt, Fuzzy clustering as a means of selecting representative conformers and molecular alignments. J. Chem. Inf. Comput. Sci. **43**(3), 810–818 (2003)

14. H.L. Gordon, R.L. Somorjai, Fuzzy cluster analysis of molecular dynamics trajectories. Proteins **14**(2), 249–264 (1992)
15. T.N. Doman, J.M. Cibulskis, M.J. Cibulskis, P.D. McCray, D.P. Spangler, Algorithm5: A technique for fuzzy similarity clustering of chemical inventories. J. Chem. Inf. Comput. Sci. **36**, 1195–1204 (1996)
16. D.N. Rassokhin, V.S. Lobanov, D.K. Agrafiotis, Nonlinear mapping of massive data sets by fuzzy clustering and neural networks. J. Comput. Chem. **22**(4), 373–386 (2001)
17. TH. Lin, GM. Wang, YH. Hsu , Classification of some active HIV-1 protease inhibitors and their inactive analogues using some uncorrelated three-dimensional molecular descriptors and a fuzzy c-means algorithm. J. Chem. Inf. Comput. Sci. **42**(6), 1490–1504 (2002)
18. A. Banerjee, M. Misra, D. Pai, O.LY. Shih, R. Woodley, XJ. Lu, A.R. Srinivasan, W.K. Olson, R.N. Dave, C.A. Venanzi, Feature extraction using molecular planes for fuzzy relational clustering of a flexible dopamine reuptake inhibitor. J. Chem. Inf. Model. **47**, 2216–2227 (2007)
19. X. Li, X. Lu, J. Tian, P. Gao, H. Kong, G. Xu, Application of fuzzy c-means clustering in data analysis of metabolomics. Anal. Chem. **81**(11), 4468–4475 (2009)
20. J.C. Bezdek, Fuzzy Mathematics in Pattern Classification Ph.D. Dissertation, IEEE Trans. Cornell University, Ithaca, 1973
21. J.C. Bezdek, Cluster validity with fuzzy sets. J. Cybernet. **3**, 58–73 (1974)
22. J.C. Bezdek, Numerical taxonomy with fuzzy sets. J. Math. Biol. **1**, 57–71 (1974)
23. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, NewYork, 1981)
24. J.C. Bezdek, *Pattern Recognition in Handbook of Fuzzy Computation* (IOP Publishing Ltd., Boston, 1998)
25. R.N. Dave, K. Bhaswan, Adaptive fuzzy c-shells clustering and detection of ellipses. IEEE Trans. Neural Netw. **3**(5), 643–662 (1992)
26. R. Krishnapuram, O. Nasraoui, J. Keller, The fuzzy c spherical shells algorithm: a new approach, IEEE Trans. Neural Netw. **3**(5), 663–671 (1992)
27. Y. Man, I. Gath, Detection and separation of ring-shaped clusters using fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. **16**(8), 55–861 (1994)
28. R.J. Hathaway, J.C. Bezdek, Optimization of clustering criteria by reformulation, IEEE Trans. Fuzzy Syst. **3**(2), 241–245 (1995)
29. W. Wei, J.M. Mendel, Optimality test for the fuzzy c-means algorithm. Pattern Recognit. **27**(11), 1567–1573 (1994)
30. J. Yu, M.S. Yang, Optimality test for generalized FCM and its application to parameter selection, IEEE Trans. Fuzzy Syst. **13**(1), 164–176 (2005)
31. X.L. Xie, G. Beni, A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. **13**, 841–847 (1991)
32. Y. Fukuyama, M. Sugeno, *A new method of choosing the number of clusters for the fuzzy c-means method, Proc. Fifth Fuzzy Systems Symp.*, pp. 247–250 (1989)
33. F. Cao, J. Liang, G. Jiang, An initialization method for the K-Means algorithm usingneighborhood model. Proc. Comput. Math. Appl. **58**, 474–483 (2009)
34. S.J. Redmond, C. Heneghan, A method for initialising the K-means clustering algorithm using kd-trees. Pattern Recognit. Lett. **28**, 965–973 (2007)
35. S.S. Khan, A. Ahmad, Cluster center initialization algorithm for K-means clustering. Pattern Recognit. Lett. **25**, 1293–1302 (2004)
36. J.M. Peña, J.A. Lozano, P. Larrañaga, An empirical comparison of four initialization methods for the K-means algorithm. Pattern Recognit. Lett. **25**, 1027–1040 (1999)