



Title	A method of identifying influential data in fuzzy clustering
Author(s)	Imai, Hideyuki; Tanaka, Akira; Miyakoshi, Masaaki
Citation	IEEE TRANSACTIONS ON FUZZY SYSTEMS, 6(1), 90-101 <a href="https://doi.org/10.1109/91.660810">https://doi.org/10.1109/91.660810</a>
Issue Date	1998-02
Doc URL	<a href="http://hdl.handle.net/2115/5843">http://hdl.handle.net/2115/5843</a>
Rights	© 1998 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. " IEEE, IEEE TRANSACTIONS ON FUZZY SYSTEMS, Volume 6, Issue 1, 1998 Page(s):90 - 101
Type	article
File Information	IEEETFS6-1.pdf



[Instructions for use](#)

# A Method of Identifying Influential Data in Fuzzy Clustering

Hideyuki Imai, Akira Tanaka, and Masaaki Miyakoshi

**Abstract**—In multivariate statistical methods, it is important to identify influential observations for a reasonable interpretation of the data structure. In this paper, we propose a method for identifying influential data in the fuzzy C-means (FCM) algorithm. To investigate such data, we consider a perturbation of the data points and evaluate the effect of a perturbation. As a perturbation, we consider two cases: one is the case in which the direction of a perturbation is specified and the other is the case in which the direction of a perturbation is not specified. By computing the change in the clustering result of FCM when given data points are slightly perturbed, we can look for data points that greatly affect the result. Also, we confirm an efficacy of the proposed method by numerical examples.

**Index Terms**—Fuzzy c-means algorithm, perturbation, sensitivity analysis.

## I. INTRODUCTION

THE fuzzy C-means (FCM) is a generalization of the ISODATA, a widely used clustering method, to the fuzzy set theory. FCM was first proposed by Dunn [1] and later developed by Bezdek [2]. FCM enables us to make flexible partitions of a finite data set in a feature space. Therefore, FCM is used in many fields such as data analysis or image segmentation and can be extended to fuzzy C-variables [3], fuzzy C-elliptotype clustering [4], fuzzy C-regression model [5], and so on.

In FCM, noisy data points or outliers may greatly influence the result. Therefore, some methods to detect appropriate clusters in noisy data have been investigated [6]–[9]. In [6], a concept of noise cluster is introduced and the noise clustering algorithm is proposed. The algorithm obtained in [9] is more immune to noise by a possibilistic approach and, in [7] and [8], clustering algorithms based on the robust statistics are presented. Furthermore, it is shown that some structural noise can be captured by fuzzy relations [10]. However, it is important to identify such data points as greatly affect the clustering result. In this paper, we propose a method of identifying influential data in FCM by evaluating the effect of a perturbation of data points.

Manuscript received June 1, 1995; revised October 25, 1996. This research was supported in part by Grant-in-Aid for Scientific Research, Ministry of Education, Science, Sports, and Culture, Japan, Grants 07558149 and 0940048.

H. Imai and M. Miyakoshi are with the Division of Systems and Information Engineering, Hokkaido University, Sapporo, 060 Japan.

A. Tanaka is with Matsushita Communication Industrial Co., Ltd., Yokohama, 224 Japan.

Publisher Item Identifier S 1063-6706(98)00846-7.

## II. FUZZY C-MEANS ALGORITHM

The FCM algorithm is to solve the following problem:

$$\text{minimize } J_m(U, V) = \sum_{i=1}^N \sum_{\alpha=1}^c (u_{\alpha i})^m \text{dist}^2(\mathbf{x}_i, \mathbf{v}_\alpha) \quad (1)$$

where

- $N$  the number of data,  $I = \{1, \dots, N\}$ ;
- $c$  the fixed and known number of clusters;
- $m$  the arbitrary chosen scalar,  $m > 1$ ;
- $\mathbf{x}_1, \dots, \mathbf{x}_N \in R^k$  data points in a feature space, which is the  $k$ -dimensional real Euclidean space  $R^k$ ;
- $\mathbf{v}_1, \dots, \mathbf{v}_c \in R^k$  the unknown prototypes (cluster centers);
- $\text{dist}(\mathbf{x}_i, \mathbf{v}_\alpha)$  the distance between  $\mathbf{x}_i$  and  $\mathbf{v}_\alpha$ .

$V \equiv [\mathbf{v}_1, \dots, \mathbf{v}_c]$ : a  $k \times c$  matrix

$U \equiv (u_{\alpha i}) \in M_{fc}$ : a  $c \times N$  matrix

$M_{fc} \equiv \{U = (u_{\alpha i}) | u_{\alpha i}$ 's satisfy the following conditions 1), 2), and 3)\}

- 1)  $u_{\alpha i} \in [0, 1], \forall \alpha = 1, \dots, c, \forall i \in I$ ;
- 2)  $\sum_{\alpha=1}^c u_{\alpha i} = 1, \forall i \in I$ ;
- 3)  $0 < \sum_{i=1}^N u_{\alpha i} < N, \forall \alpha = 1, \dots, c$ .

In this paper, we use the Euclidean distance as  $\text{dist}(\mathbf{x}_i, \mathbf{v}_\alpha)$ ; that is

$$\text{dist}(\mathbf{x}_i, \mathbf{v}_\alpha) = \left[ \sum_{s=1}^k (x_{is} - v_{\alpha s})^2 \right]^{1/2}$$

where  $\mathbf{x}_i = (x_{is}), \mathbf{v}_\alpha = (v_{\alpha s})$ .

The scalar  $m$ , called the weighting exponent of FCM, determines the fuzziness of the clustering. When  $m$  is small the result is the same as hard clustering; that is,  $u_{\alpha i}$ 's tend to  $\{0, 1\}, \forall i \in I, \forall \alpha = 1, \dots, c$ , and when  $m$  is large, the result is quite fuzzy; that is,  $u_{\alpha i}$ 's tend to  $(1/c), \forall i \in I, \forall \alpha = 1, \dots, c$ .

Since it is difficult to find the solution of (1), Dunn [1] and Bezdek [2] proposed an algorithm based on the alternating least-squares method. Since the algorithm is based on a nonlinear optimization, the obtained results may fail to give the global minimum solution. The convergence properties of FCM are investigated by many researchers [11]–[15].

In what follows, we assume that the global minimum solution  $(U^*, V^*)$  is obtained by some numerical method.

### III. SENSITIVITY ANALYSIS OF FCM

#### A. Sensitivity Analysis

A descriptive statistical method such as a cluster analysis is used in the early stage of data analysis. The practical data must include observations that greatly affect the result. It is pointed out that the investigation of such observations are reported to be important for a reasonable interpretation of the data structure [16]. For example, one data point far from other data may cause inadequate results in FCM.

One way to evaluate an influence of some data points is to compute the results when these data points are removed. However, since it requires high computing costs to obtain an accurate result, some approximation for evaluation have been considered. For example, in Hayashi's methods of quantification, principal component analysis, and factor analysis, the method of searching influential data is investigated in [17]–[19]. Such a method is based on the perturbation expansion of characteristic roots and vectors of a certain symmetric matrix, such as a covariance or correlation matrix [20].

#### B. Identifying Influential Data in FCM

In this section, we consider a method for identifying influential data points in FCM.

*Definition 1:* Let  $A = [\mathbf{a}_1, \dots, \mathbf{a}_c]$  be an  $N \times c$  matrix where  $\mathbf{a}_1, \dots, \mathbf{a}_c \in R^N$ , then  $\text{vec}(A)$  is an  $N \times c$  vector defined as

$$\text{vec}(A) \equiv [\mathbf{a}'_1, \dots, \mathbf{a}'_c]'$$

*Definition 2:* An  $(N + k) \times c$  vector  $\boldsymbol{\theta}$  is defined as

$$\begin{aligned} \boldsymbol{\theta} &\equiv \{[\text{vec}(U)]', [\text{vec}(V)']'\}' \\ &= [u_{11}, \dots, u_{c1}, \dots, u_{1N}, \dots, u_{cN} \\ &\quad v_{11}, \dots, v_{1k}, \dots, v_{c1}, \dots, v_{ck}]'. \end{aligned}$$

Since  $u_{ic}, i \in I$  are not independent, we define a vector  $\bar{\boldsymbol{\theta}}$  as follows.

*Definition 3:* Let  $\bar{\boldsymbol{\theta}}$  be a  $\{(c-1) \times N + c \times k\}$  vector, made by deleting  $u_{c1}, \dots, u_{cN}$  from  $\boldsymbol{\theta}$ ; that is

$$\begin{aligned} \bar{\boldsymbol{\theta}} &\equiv [u_{11}, \dots, u_{c-1,1}, \dots, u_{1N}, \dots, u_{c-1,N} \\ &\quad v_{11}, \dots, v_{1k}, \dots, v_{c1}, \dots, v_{ck}]' \end{aligned}$$

and

$$\Theta \equiv \{\bar{\boldsymbol{\theta}} \in R^{(c-1) \times N + c \times k} | u_{\alpha i} \text{'s satisfy the following conditions 1), 2), 3), and 4)}\}$$

- 1)  $u_{\alpha i} \in [0, 1], \forall \alpha = 1, \dots, c-1, \forall i \in I;$
- 2)  $\sum_{\alpha=1}^{c-1} u_{\alpha i} \leq 1, \forall i \in I;$
- 3)  $\sum_{\alpha=1}^{c-1} u_{\alpha i} < 1, \exists i \in I;$
- 4)  $0 < \sum_{i=1}^N u_{\alpha i} < N, \forall \alpha = 1, \dots, c-1.$

Because the objective function  $J_m(\boldsymbol{\theta})$  is regarded as a function on  $\Theta$ , we denote  $\bar{J}_m(\bar{\boldsymbol{\theta}})$  defined on  $\Theta$  as follows:

$$\begin{aligned} \bar{J}_m(\bar{\boldsymbol{\theta}}) &\equiv \sum_{i=1}^N \sum_{\alpha=1}^{c-1} (u_{\alpha i})^m \sum_{s=1}^k (x_{is} - v_{\alpha s})^2 \\ &\quad + \sum_{i=1}^N \left(1 - \sum_{\alpha=1}^{c-1} u_{\alpha i}\right)^m \sum_{s=1}^k (x_{is} - v_{cs})^2, \bar{\boldsymbol{\theta}} \in \Theta, \end{aligned}$$

Here, we assume to obtain the global minimum solution  $\bar{\boldsymbol{\theta}}^* \in \Theta$ .

*Definition 4:* Let  $H(\bar{\boldsymbol{\theta}})$  be a  $\{(c-1) \times N + c \times k\} \times \{(c-1) \times N + c \times k\}$  matrix defined as follows:

$$H(\bar{\boldsymbol{\theta}}) \equiv \begin{bmatrix} \frac{\partial^2 \bar{J}_m}{\partial u_{11} \partial u_{11}} & \cdots & \frac{\partial^2 \bar{J}_m}{\partial u_{11} \partial v_{ck}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \bar{J}_m}{\partial v_{ck} \partial u_{11}} & \cdots & \frac{\partial^2 \bar{J}_m}{\partial v_{ck} \partial v_{ck}} \end{bmatrix}$$

that is,  $H(\bar{\boldsymbol{\theta}})$  is the Hessian matrix of  $\bar{J}_m(\bar{\boldsymbol{\theta}})$ . The elements of  $H(\bar{\boldsymbol{\theta}})$  are explicitly found in the Appendix.

Since  $\bar{\boldsymbol{\theta}}^*$  is the global minimum solution of  $\bar{J}_m(\bar{\boldsymbol{\theta}})$

$$\frac{\partial \bar{J}_m}{\partial \bar{\boldsymbol{\theta}}}(\bar{\boldsymbol{\theta}}^*) = \mathbf{0},$$

holds and  $H(\bar{\boldsymbol{\theta}}^*)$  is a nonnegative definite matrix.

Now we evaluate the effect of perturbation. Let a nonvoid subset  $S \subset I$  be a set of indexes of the data points, which we wish to perturb. By renumbering data points, we can assume that  $S = \{1, \dots, l\}$ . Then, the perturbed data points  $\mathbf{y}_i, i \in I$  is represented by

$$\mathbf{y}_i = \begin{cases} \mathbf{x}_i + \varepsilon \mathbf{d}_i + O(\varepsilon^2), & i = 1, \dots, l \\ \mathbf{x}_i, & i = l+1, \dots, N \end{cases}$$

where the unit vector  $\mathbf{d}_i, i = 1, \dots, l$  is the direction of the perturbation of  $\mathbf{x}_i$  and  $O(\varepsilon^2)$  denotes the vector whose elements are all  $O(\varepsilon^2)$ ; that is, bounded for small  $\varepsilon^2$ . As the direction  $\mathbf{d}_i$ , we consider the following two cases.

- 1) All  $\mathbf{d}_i, i = 1, \dots, l$  are specified, which means that an evaluation of the effect of the particular direction of perturbation is needed. For example, we wish to evaluate the effect when some components of data points are perturbed.
- 2) All  $\mathbf{d}_i, i = 1, \dots, l$  are not specified. In this case, we wish to obtain the directions of perturbation  $\mathbf{d}_i, i = 1, \dots, l$ , which make the effect of perturbation maximum.

For the case 1), we obtain Theorem 2 and Corollary 1 and for the case 2), we obtain Corollaries 2 and 3.

*Definition 5:* Let  $\bar{J}_m(\bar{\boldsymbol{\theta}}, \varepsilon)$ , for  $\bar{\boldsymbol{\theta}} \in \Theta, \varepsilon \in R$  denote the objective function for perturbed data points  $\mathbf{y}_i, i \in I$ .

For the perturbed data points  $\mathbf{y}_i, i \in I$ , the problem to be solved is

$$\text{minimize}_{\bar{\boldsymbol{\theta}} \in \Theta} \bar{J}_m(\bar{\boldsymbol{\theta}}, \varepsilon). \quad (2)$$

Notice that the objective function without perturbation is  $\bar{J}_m(\bar{\boldsymbol{\theta}})$ . To clarify that  $\bar{\boldsymbol{\theta}}^*$  is the solution of (2) in the case of

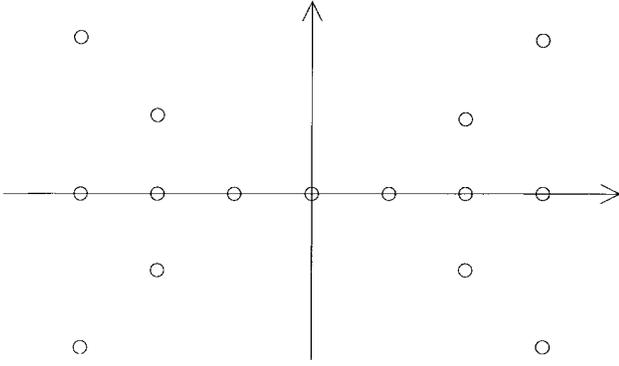
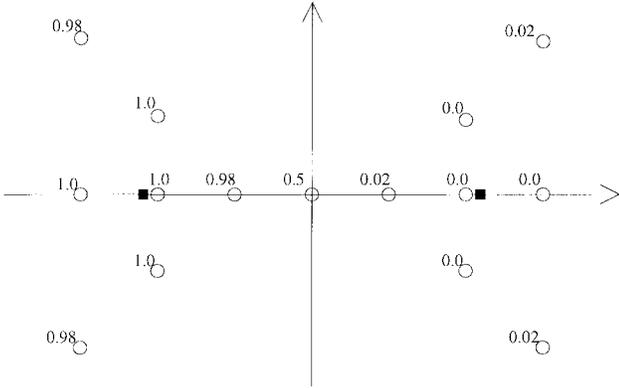


Fig. 1. Butterfly data.

Fig. 2. The result of FCM for butterfly data ( $m = 1.5$ ).

$\varepsilon = 0$ , we let  $\bar{\theta}_0^*$  denote the solution of (2) instead of  $\bar{\theta}^*$  and  $u_{0,\alpha i}^*$ ,  $v_{0,\alpha i}^*$  denote the elements of  $\bar{\theta}_0^*$ . Our aim is to evaluate the difference between  $\bar{\theta}_0^*$  and the solution of (2). About the solution, the following theorem holds.

*Theorem 1:* If either of the following two conditions

- 1)  $m \geq 2$ ;
- 2)  $u_{0,\alpha i}^* \in (0, 1)$ ,  $\forall i \in I$ ,  $\forall \alpha = 1, \dots, c-1$

is satisfied and if  $H(\bar{\theta}_0^*)$  is nonsingular, then

$$\frac{\partial \bar{J}_m}{\partial \bar{\theta}}(\bar{\theta}, \varepsilon) = 0$$

can be analytically solved as  $\bar{\theta} = \bar{\theta}(\varepsilon)$  in the neighborhood of  $(\bar{\theta}, \varepsilon) = (\bar{\theta}^*, 0)$  and  $\bar{\theta}^*(0) = \bar{\theta}_0^*$ . Moreover,

$$\bar{\theta}^*(\varepsilon) - \bar{\theta}_0^* = H^{-1}(\bar{\theta}_0^*)L(\bar{\theta}_0^*)\text{vec}(D) \cdot \varepsilon + O(\varepsilon^2) \quad (3)$$

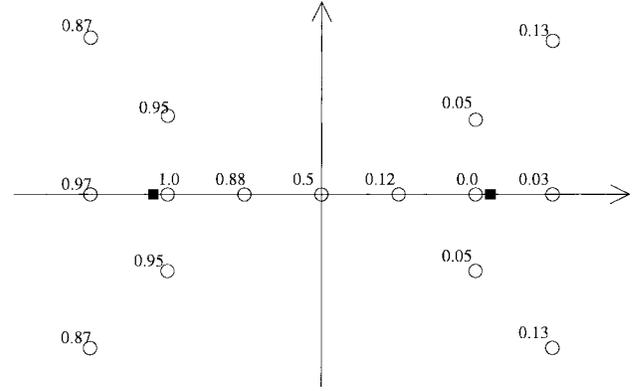
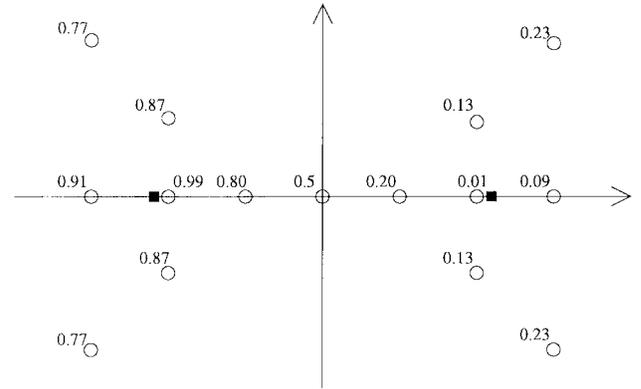
where  $L(\bar{\theta})$  denotes a  $[(c-1) \times N + c \times k] \times (l \times k)$  matrix defined in the Appendix.

*Proof:* If either condition 1) or 2) is satisfied, then  $H(\bar{\theta}_0^*)$  exists and is nonsingular by the assumption. Then, by the implicit function theorem [21], we obtain  $\bar{\theta} = \bar{\theta}(\varepsilon)$  and

$$\bar{\theta}^*(\varepsilon) - \bar{\theta}_0^* = H^{-1}(\bar{\theta}_0^*) \frac{\partial^2 \bar{J}_m}{\partial \bar{\theta} \partial \varepsilon}(\bar{\theta}_0^*, 0) \varepsilon + O(\varepsilon^2).$$

The elements of  $(\partial^2 \bar{J}_m / \partial \bar{\theta} \partial \varepsilon)(\bar{\theta}_0^*, 0)$  are explicitly found in the Appendix. By the definition of  $L(\bar{\theta}_0^*)$  in the Appendix

$$\frac{\partial^2 \bar{J}_m}{\partial \bar{\theta} \partial \varepsilon}(\bar{\theta}_0^*, 0) = L(\bar{\theta}_0^*)\text{vec}(D)$$

Fig. 3. The result of FCM for butterfly data ( $m = 2.0$ ).Fig. 4. The result of FCM for butterfly data ( $m = 2.5$ ).

since all elements of  $(\partial^2 \bar{J}_m / \partial \bar{\theta} \partial \varepsilon)(\bar{\theta}_0^*, 0)$  are linear combinations of  $d_i$ ,  $i = 1, \dots, l$ . ■

In (3) of Theorem 1, the influence of perturbation on  $u_{ic}$ ,  $i \in I$  is not evaluated. Noticing that

$$\sum_{\alpha=1}^c u_{0,\alpha i} = \sum_{\alpha=1}^c u_{\alpha i}(\varepsilon) = 1, \quad \forall i \in I$$

then we obtain

$$u_{ci}^*(\varepsilon) - u_{0,\alpha i}^* = \sum_{\alpha=1}^{c-1} [u_{\alpha i}^*(\varepsilon) - u_{0,\alpha i}^*] \varepsilon + O(\varepsilon^2), \quad \forall i \in I.$$

From Theorem 1 and above, we get the following corollary.

*Corollary 1:* Let

$$H^{-1}(\bar{\theta}_0^*)L(\bar{\theta}_0^*) = \left[ \underbrace{c-1}_{E'_1} \quad \dots \quad \underbrace{c-1}_{E'_N} \quad \underbrace{k}_{F'_1} \quad \dots \quad \underbrace{k}_{F'_c} \right]'$$

then

$$\bar{\theta}^*(\varepsilon) - \bar{\theta}_0^* = P \text{vec}(D) \cdot \varepsilon + O(\varepsilon^2) \quad (4)$$

where

$$P = [E'_1 \quad -E'_1 \mathbf{1} \quad \dots \quad E'_N \quad -E'_N \mathbf{1} \quad F'_1 \quad \dots \quad F'_c]'$$

and  $O(\varepsilon^2)$  denotes an  $(N+c) \times k$  vector whose elements are all  $O(\varepsilon^2)$  and  $\mathbf{1}$  denotes a  $(c-1)$  vector whose elements are all one.

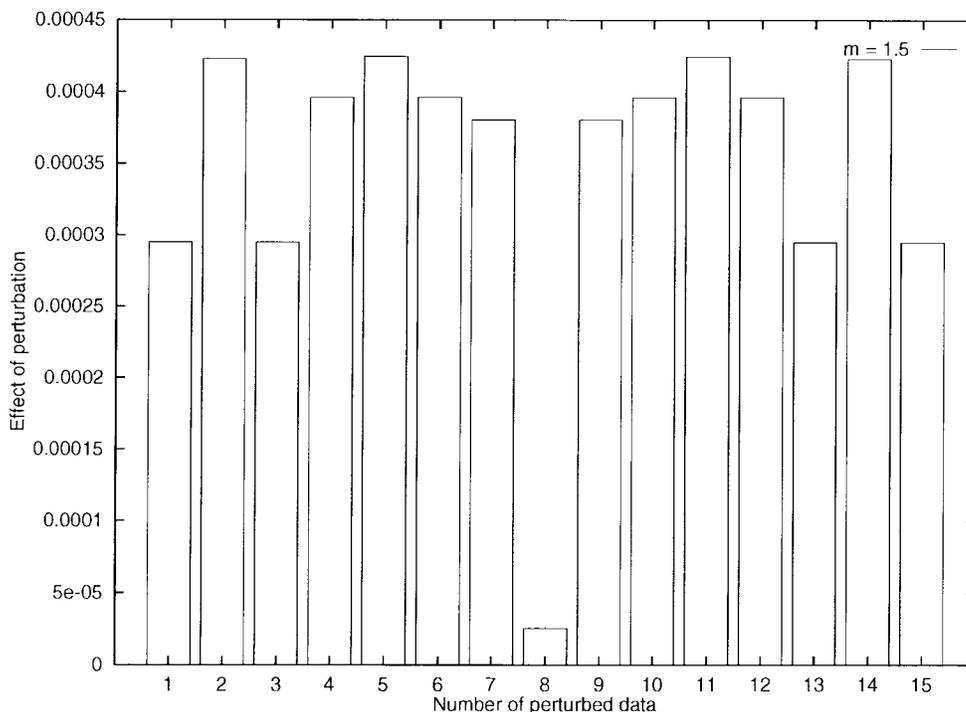


Fig. 5. The effect of perturbation for butterfly data ( $m = 1.5$ ).

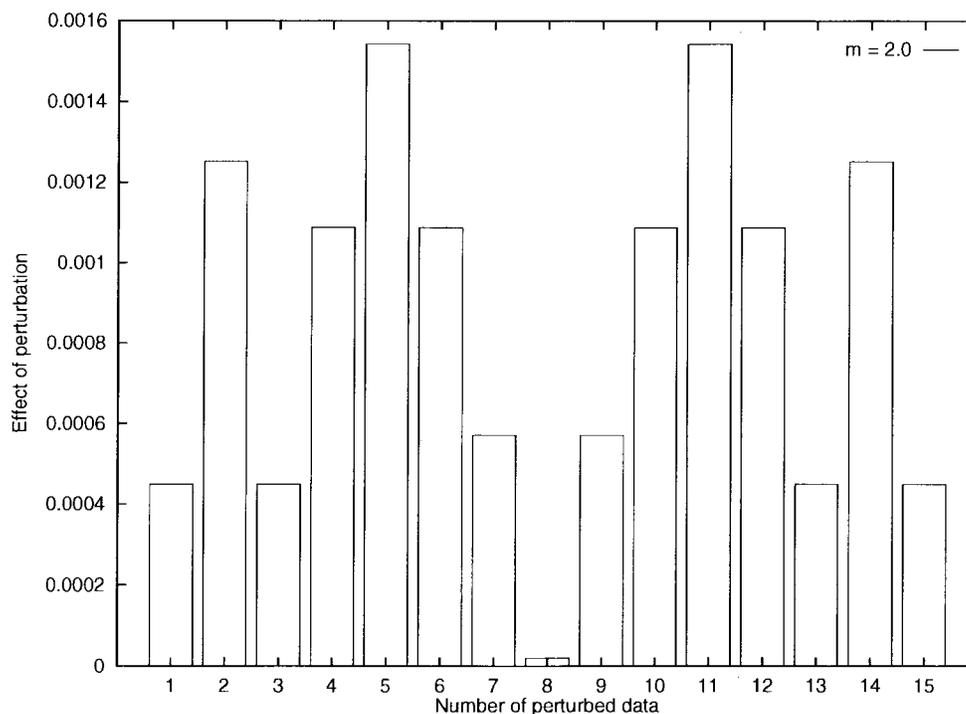


Fig. 6. The effect of perturbation for butterfly data ( $m = 2.0$ ).

By using (4), we can compute the change in the clustering result of FCM when perturbations  $\mathbf{d}_1, \dots, \mathbf{d}_l$  are added to the data points  $\mathbf{x}_1, \dots, \mathbf{x}_l$ .

Now, we consider the case that the directions of the perturbation  $\mathbf{d}_i, i = 1, \dots, l$  are not specified. As a measure of influence,  $[\boldsymbol{\theta}^*(\varepsilon) - \boldsymbol{\theta}_0^*]^T W [\boldsymbol{\theta}^*(\varepsilon) - \boldsymbol{\theta}_0^*]$  is used where  $W$  is a given nonnegative definite matrix. By choosing a suitable matrix  $W$  as a weight, we can consider various measures

of influence. For example, to consider the changes both in prototypes and in membership grade, the identity matrix should be used as  $W$ . To consider the only changes in prototypes,  $\begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & I_{k \times c} \end{bmatrix}$  should be used where  $I_{k \times c}$  is the  $(k \times c) \times (k \times c)$  identity matrix.

When we evaluate the influence by the weighted pseudonorm of vector mentioned above, then we get the following corollaries.

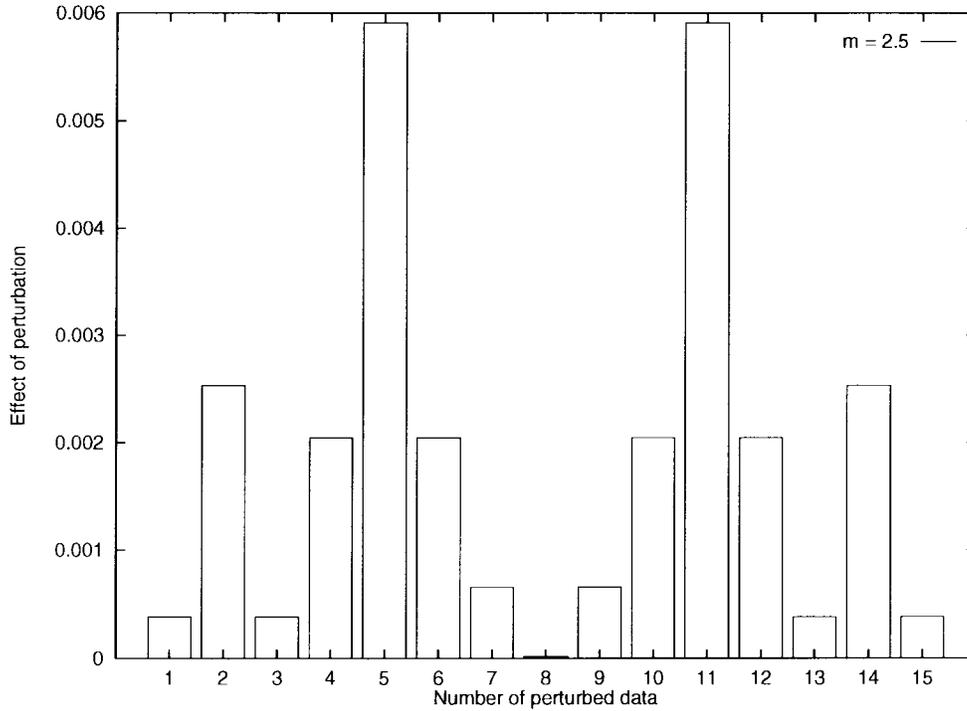


Fig. 7. The effect of perturbation for butterfly data ( $m = 2.5$ ).

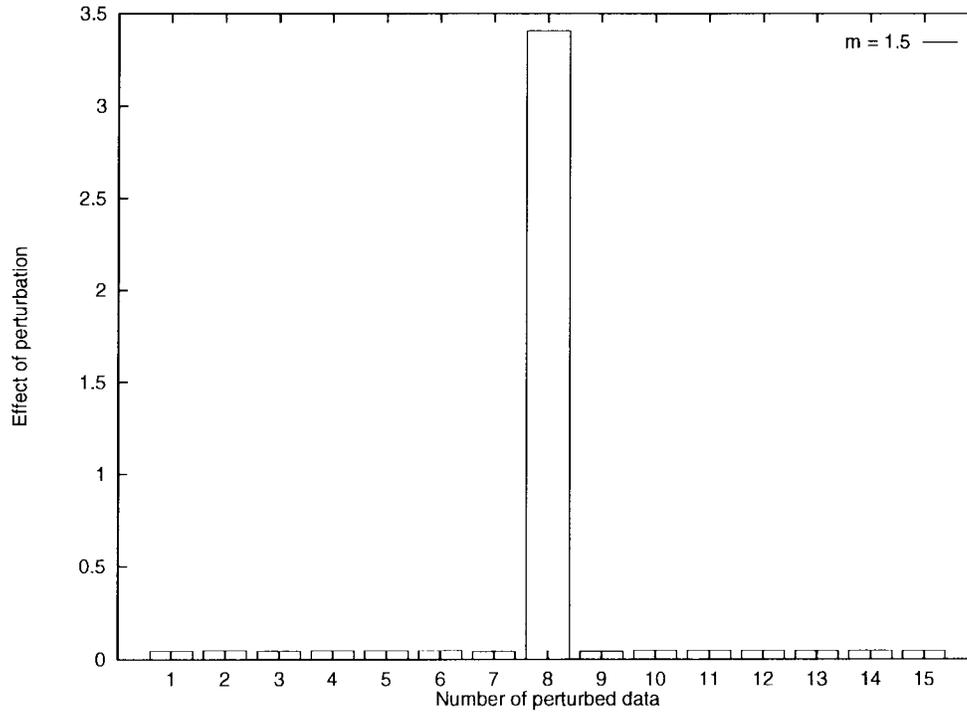


Fig. 8. The effect of perturbation for butterfly data ( $m = 1.5$ ).

*Corollary 2:* The coefficient of  $\varepsilon^2$  of the pseudonorm  $[\theta^*(\varepsilon) - \theta_0^*]^T W [\theta^*(\varepsilon) - \theta_0^*]$  is

$$[\mathbf{d}'_1 \ \cdots \ \mathbf{d}'_l] P' W P \begin{bmatrix} \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_l \end{bmatrix}, \quad \mathbf{d}'_i \mathbf{d}_i = 1, \forall i \in S.$$

*Corollary 3:* If  $S = \{i_1\}$  and we let  $\lambda_1$  be the largest eigenvalue of  $P' W P$  and  $\mathbf{l}_1$  be the corresponding eigenvector

to  $\lambda_1$ , then the coefficient of  $\varepsilon^2$  of  $[\theta^*(\varepsilon) - \theta_0^*]^T W [\theta^*(\varepsilon) - \theta_0^*]$  is maximized by  $\mathbf{d}_1 = \mathbf{l}_1$  and the maximum value is equal to  $\lambda_1$ .

Now, we evaluated the effect of the perturbation of each data point of Butterfly. In this example, we perturb one data point at a time for all 15 data points.

By using Corollaries 2 and 3, we can compute the maximal change in the clustering result of FCM when the data points are perturbed.

TABLE I  
THE BUTTERFLY DATA

Data	$x_i$	Membership grade					
		$m = 1.5$		$m = 2.0$		$m = 2.5$	
		$u_{1i}$	$u_{2i}$	$u_{1i}$	$u_{2i}$	$u_{1i}$	$u_{2i}$
1	(-3.0, 2.0)	0.977	0.023	0.866	0.134	0.773	0.227
2	(-3.0, 0.0)	0.999	0.001	0.973	0.027	0.912	0.089
3	(-3.0, -2.0)	0.977	0.023	0.866	0.134	0.773	0.227
4	(-2.0, 1.0)	0.997	0.003	0.947	0.053	0.872	0.128
5	(-2.0, 0.0)	1.000	0.000	0.999	0.001	0.992	0.008
6	(-2.0, -1.0)	0.997	0.003	0.947	0.053	0.872	0.128
7	(-1.0, 0.0)	0.982	0.018	0.883	0.117	0.798	0.202
8	(0.0, 0.0)	0.500	0.500	0.500	0.500	0.500	0.500
9	(1.0, 0.0)	0.018	0.982	0.117	0.883	0.202	0.798
10	(2.0, 1.0)	0.003	0.996	0.053	0.947	0.128	0.872
11	(2.0, 0.0)	0.000	1.000	0.001	0.999	0.008	0.992
12	(2.0, -1.0)	0.003	0.997	0.053	0.947	0.128	0.872
13	(3.0, 2.0)	0.023	0.977	0.134	0.866	0.227	0.773
14	(3.0, 0.0)	0.001	0.999	0.027	0.973	0.088	0.912
15	(3.0, -2.0)	0.023	0.977	0.134	0.866	0.227	0.773

TABLE II  
PROTOTYPES OF CLUSTERS

Cluster number	Prototypes		
	$m = 1.5$	$m = 2.0$	$m = 2.5$
	1	(-2.166, 0)	(-2.145, 0)
2	( 2.166, 0)	( 2.145, 0)	( 2.121, 0)

TABLE III  
EVALUATION OF A PERTURBATION WITH A DIRECTION (0, 1)

Data	Effect of Perturbation		
	$m = 1.5$	$m = 2.0$	$m = 2.5$
1	2.951e-04	4.491e-04	3.8271e-04
2	4.229e-04	1.253e-03	2.5313e-03
3	2.951e-04	4.491e-04	3.8271e-04
4	3.962e-04	1.088e-03	2.0439e-03
5	4.246e-04	1.543e-03	5.9102e-03
6	3.962e-04	1.088e-03	2.0439e-03
7	3.805e-04	5.713e-04	6.5676e-04
8	2.535e-05	1.922e-05	1.7425e-05
9	3.805e-04	5.713e-04	6.5676e-04
10	3.962e-04	1.088e-03	2.0439e-03
11	4.246e-04	1.543e-03	5.9102e-03
12	3.962e-04	1.088e-03	2.0439e-03
13	2.951e-04	4.491e-04	3.8271e-04
14	4.229e-04	1.253e-03	2.5313e-03
15	2.951e-04	4.491e-04	3.8271e-04

TABLE IV  
MAXIMUM EFFECT OF PERTURBATION AND ITS DIRECTION

Data	Maximum effect and its direction			
	$m = 1.5$		$m = 2.0$	
	1	0.045	(0.999, 0.034)	0.030
2	0.048	(1.000, 0.000)	0.041	(1.000, 0.000)
3	0.045	(0.999, -0.034)	0.030	(0.924, -0.382)
4	0.048	(1.000, -0.000)	0.042	(0.979, 0.204)
5	0.049	(1.000, 0.000)	0.050	(1.000, 0.000)
6	0.048	(1.000, 0.000)	0.042	(0.979, -0.204)
7	0.042	(1.000, 0.000)	0.118	(1.000, 0.000)
8	3.407	(1.000, 0.000)	0.654	(1.000, 0.000)
9	0.042	(1.000, 0.000)	0.118	(1.000, 0.000)
10	0.048	(1.000, 0.000)	0.042	(0.979, -0.204)
11	0.049	(1.000, 0.000)	0.050	(1.000, 0.000)
12	0.048	(1.000, -0.000)	0.042	(0.979, 0.204)
13	0.045	(0.999, -0.034)	0.030	(0.924, -0.382)
14	0.048	(1.000, 0.000)	0.041	(1.000, 0.000)
15	0.045	(0.999, 0.034)	0.030	(0.924, 0.382)

Data	Maximum effect and its direction	
	$m = 2.5$	
1	0.027	(0.853, -0.522)
2	0.050	(1.000, 0.000)
3	0.027	(0.853, 0.522)
4	0.053	(0.776, -0.631)
5	0.088	(1.000, 0.000)
6	0.053	(0.776, 0.631)
7	0.142	(1.000, 0.000)
8	0.282	(1.000, 0.000)
9	0.142	(1.000, 0.000)
10	0.053	(0.776, 0.631)
11	0.088	(1.000, 0.000)
12	0.053	(0.776, -0.631)
13	0.027	(0.853, 0.522)
14	0.050	(0.000, 1.000)
15	0.027	(0.853, -0.522)

In Corollary 3, it is shown that the maximal value of the coefficient is the largest eigenvalue of  $PWP$ . Nevertheless, if we wish to perturb multiple data points, maximizing the coefficient is not reduced to the ordinary eigen problem as in Corollary 3. To maximize the quadratic form in Corollary 2, we use the numerical optimization stated in [22].

IV. NUMERICAL EXAMPLE

In this section, we confirm the efficacy of the method proposed in this paper by numerical examples. As a weight  $W$  stated in the previous section, we use the  $(N \times c + k \times c) \times (N \times c + k \times c)$  identity matrix; that is, the change both in prototypes and membership grade is considered.

*Example 1:* The data set used in this example consists of the 15 points in  $R^2$  listed in Table I and illustrated graphically in Fig. 1, which is called Butterfly [2], [23]. Table II shows the results of FCM with weighting exponents  $m = 1.5$ ,  $m = 2.0$ , and  $m = 2.5$  where the cluster number  $c = 2$ . These are illustrated in Fig. 2 ( $m = 1.5$ ), in Fig. 3 ( $m = 2.0$ ), and in Fig. 4 ( $m = 2.5$ ). In these figures, the numbers near the data points are the grades in cluster 1 (rounded off of the third decimal place) and the filled rectangles are prototypes of each cluster, which are listed in Table II. This result is the

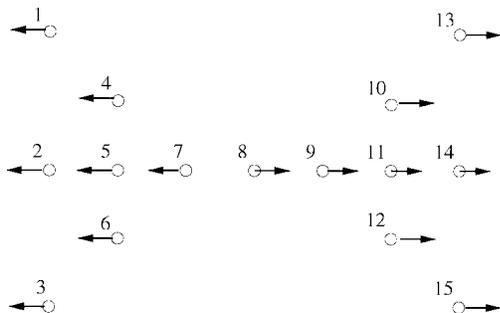


Fig. 9. Directions that make effect of perturbation maximum ( $m = 1.5$ ).

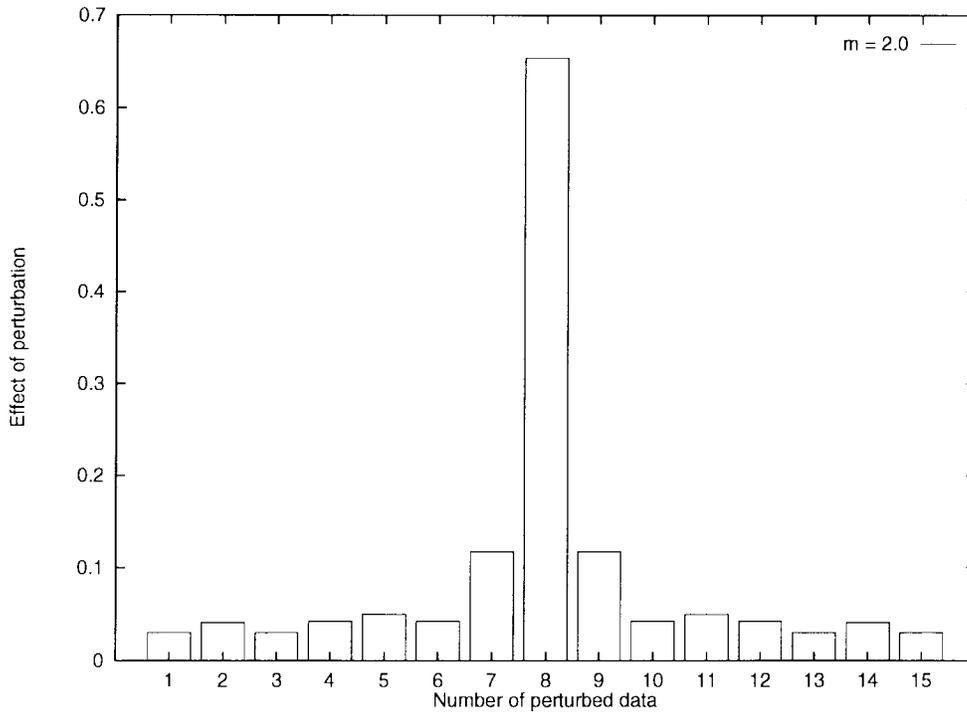


Fig. 10. The effect of perturbation for butterfly data ( $m = 2.0$ ).

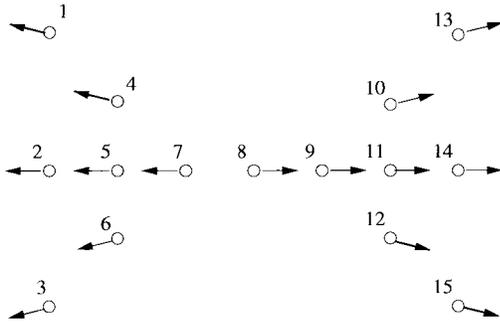


Fig. 11. Directions that make effect of perturbation maximum ( $m = 2.0$ ).

$(U, V)$ , which minimizes the objective function  $J_m$  among 100 experiments with a random initial partition matrix  $U^{(0)}$ .

At first, we consider the case where each direction of perturbations  $\mathbf{d}_i$ ,  $i = 1, \dots, 15$  is specified. Table III shows the coefficient  $\varepsilon^2$  of the norm  $[\boldsymbol{\theta}^*(\varepsilon) - \boldsymbol{\theta}_0^*]'[\boldsymbol{\theta}^*(\varepsilon) - \boldsymbol{\theta}_0^*]$  when the  $i$ th data point is perturbed with  $\mathbf{d}_i = (0, 1)$ . These are illustrated in Fig. 5 ( $m = 1.5$ ), in Fig. 6 ( $m = 2.0$ ), and in Fig. 7 ( $m = 2.5$ ).

As can be seen from the tables and these figures, the effects of perturbation are not so large for all 15 data points for the weighting exponents  $m = 1.5$ ,  $m = 2.0$ , and  $m = 2.5$ . Especially, the effect of data eight is significantly small.

Second, we consider the case in which the direction of a perturbation is not specified. By Corollary 3, we can find the direction of a perturbation which maximizes the coefficient of  $\varepsilon^2$  of the norm  $[\boldsymbol{\theta}^*(\varepsilon) - \boldsymbol{\theta}_0^*]'[\boldsymbol{\theta}^*(\varepsilon) - \boldsymbol{\theta}_0^*]$  when one data point is perturbed. Table IV shows the maximum coefficients and these vectors when  $i$ th data point is perturbed. The maximum coefficient and the directions are illustrated in Figs. 8 and 9

TABLE V  
NUTRIENTS IN MEAT, FISH, AND FOWL

Food	Ener- gy	Pro- tein	Fat	Cal- cium	Iron
BB (Beef, braised)	11	29	28	1	26
HR (Hamburger)	8	30	17	1	27
BR (Beef, roast)	13	21	39	1	20
BS (Beef, steak)	12	27	32	1	26
BC (Beef, canned)	6	31	10	2	37
CB (Chicken, broiled)	4	29	3	1	14
CC (Chicken, canned)	5	36	7	2	15
BH (Beef heart)	5	37	5	2	59
LL (Lamb leg, roast)	8	29	20	1	26
LS (Lamb shoulder, roast)	9	26	25	1	25
HS (Smoked ham)	11	29	28	1	25
PR (Pork roast)	11	27	29	1	25
PS (Pork simmered)	11	27	30	1	25
BT (Beef tongue)	6	26	14	4	25
VC (Veal cutlet)	6	33	9	1	27
FB (Bluefish, baked)	4	31	4	3	6
AR (Clams, raw)	2	16	1	10	60
AC (Clams, canned)	1	10	1	9	54
TC (Crabmeat, canned)	3	20	2	5	8
HF (Haddock, fried)	4	23	5	2	5
MB (Mackerel, broiled)	6	27	13	1	10
MC (Mackerel, canned)	5	23	9	20	18
PF (Perch, fried)	6	23	11	2	13
SC (Salmon, canned)	4	24	5	20	7
DC (Sardines, canned)	6	31	9	46	25
UC (Tuna, canned)	5	36	7	1	12
RC (Shrimp, canned)	3	33	1	12	26

( $m = 1.5$ ), Figs. 10 and 11 ( $m = 2.0$ ), and Figs. 12 and 13 ( $m = 2.5$ ), respectively.

From these table and figures, in the case of  $m = 1.5$ , we find that the perturbation of the data point  $(0, 0)$  has a large effect on the result when the direction of the perturbation is almost  $(1, 0)$ . However, in the case of  $m = 2.0$  and  $m = 2.5$ , the effect of the perturbation of the data point  $(0, 0)$  is not so

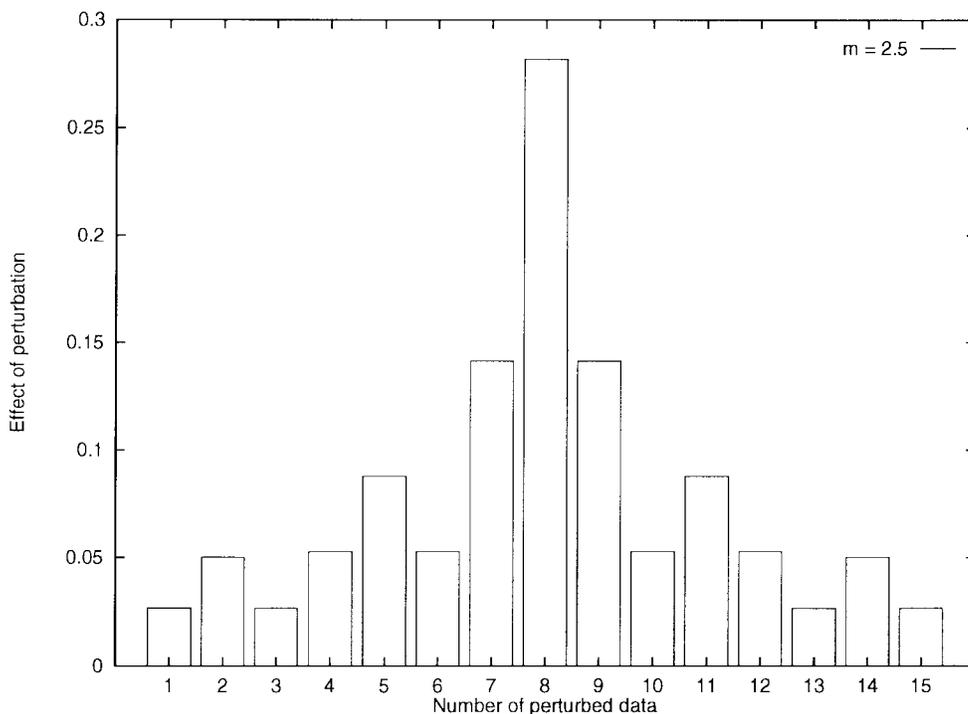


Fig. 12. The effect of perturbation for butterfly data ( $m = 2.5$ ).

TABLE VI  
RESULT OF FCM WITH  $c = 2$  AND  $m = 1.5$

Food	Membership grade		Food	Membership grade	
	$u_{1i}$	$u_{2i}$		$u_{1i}$	$u_{2i}$
BB	0.9840	0.0160	VC	0.5770	0.4230
HR	0.9857	0.0143	FB	0.0168	0.9832
BR	0.8604	0.1396	AR	0.6899	0.3101
BS	0.9668	0.0332	AC	0.6662	0.3338
BC	0.9006	0.0994	TC	0.0199	0.9801
CB	0.0091	0.9909	HF	0.0265	0.9735
CC	0.0382	0.9618	MB	0.0529	0.9471
BH	0.7548	0.2452	MC	0.0882	0.9118
LL	0.9941	0.0059	PF	0.0347	0.9653
LS	0.9895	0.0105	SC	0.0411	0.9589
HS	0.9789	0.0211	DC	0.3786	0.6214
PR	0.9767	0.0233	UC	0.0383	0.9617
PS	0.9724	0.0276	RC	0.1161	0.8839
BT	0.8189	0.1811			

TABLE VII  
PROTOTYPES OF CLUSTERS

Cluster number	Prototypes				
	Energy	Protein	Fat	Calcium	Iron
1	8.487	27.04	20.73	2.952	30.44
2	4.534	27.56	6.355	7.630	14.96

large as one of  $m = 1.5$ . This makes sense that the sensitivity generally decreases as  $m \rightarrow \infty$ .

*Example 2:* Data used in this example consists of the 27 points in  $R^5$ , listed in Table V. This data is investigated for

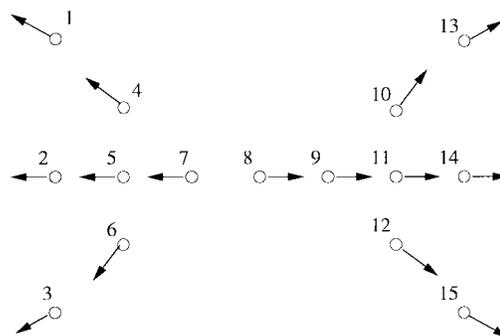


Fig. 13. Directions that make effect of perturbation maximum ( $m = 2.5$ ).

$k$ -means algorithm in [22]. Food energy, protein, calcium, and iron show percentages of content to their recommended daily dietary allowance; fat shows the weight (grams) of content of each food (rounded off of the first decimal place) [22]. For example, braised beef (BB) delivers about 11% of the daily allowance of calories and 28 grams of fat.

Tables VI and VII show membership grade and prototypes of each cluster by FCM with cluster size  $c = 2$  and weighting exponent  $m = 1.5$ . The membership grade in cluster 1 is illustrated in Fig. 14.

In the same manner of Example 1, we consider the case that one data point is perturbed at a time for all 27 data points and a direction of perturbation is not specified. Therefore, we find the direction of perturbation which maximizes the coefficient of  $\epsilon^2$  of the norm  $[\theta^*(\epsilon) - \theta_0^*]^T [\theta^*(\epsilon) - \theta_0^*]$  by Corollary 3. The second column of Table VIII shows the maximum effect of perturbation and the third column shows direction maximizing the coefficient.

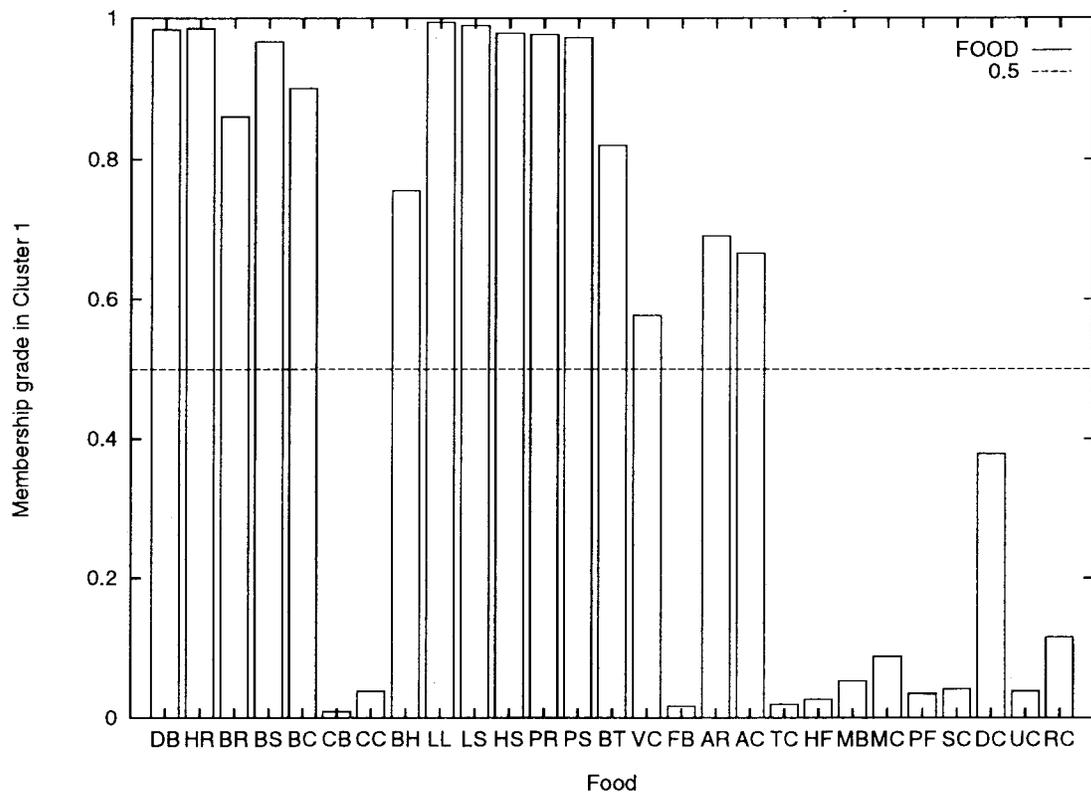


Fig. 14. Membership grade in cluster 1.

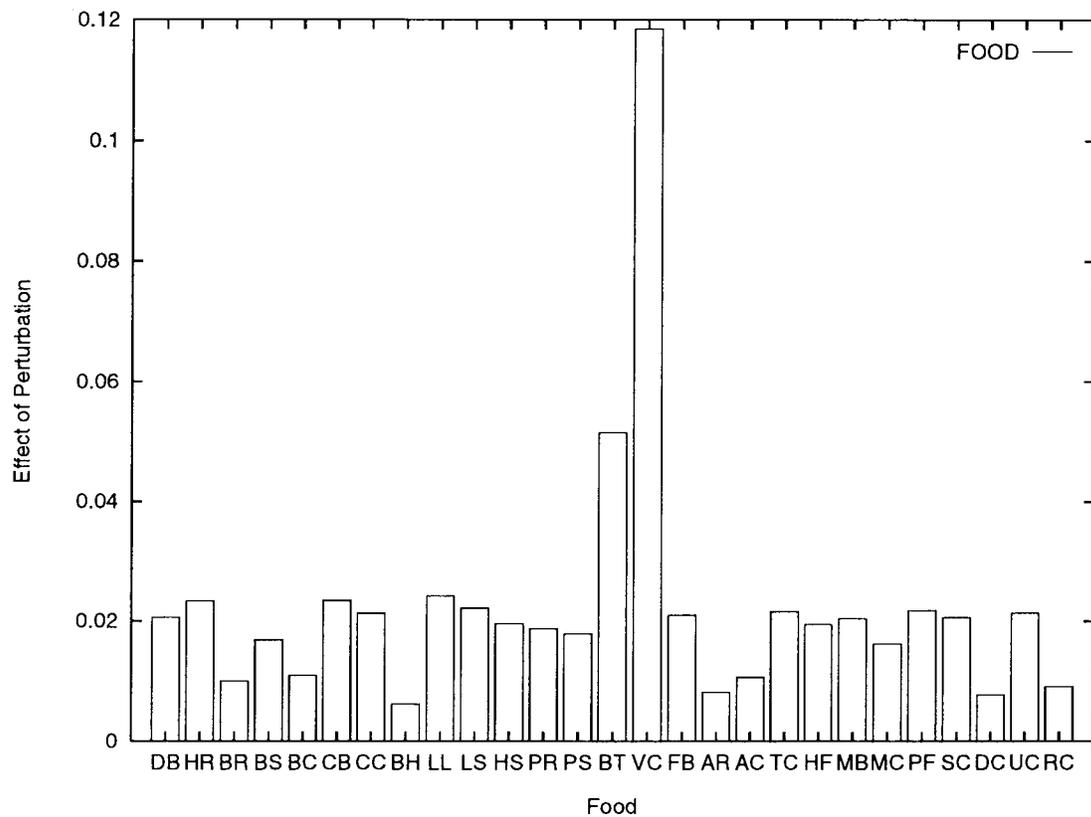


Fig. 15. Effect of perturbation.

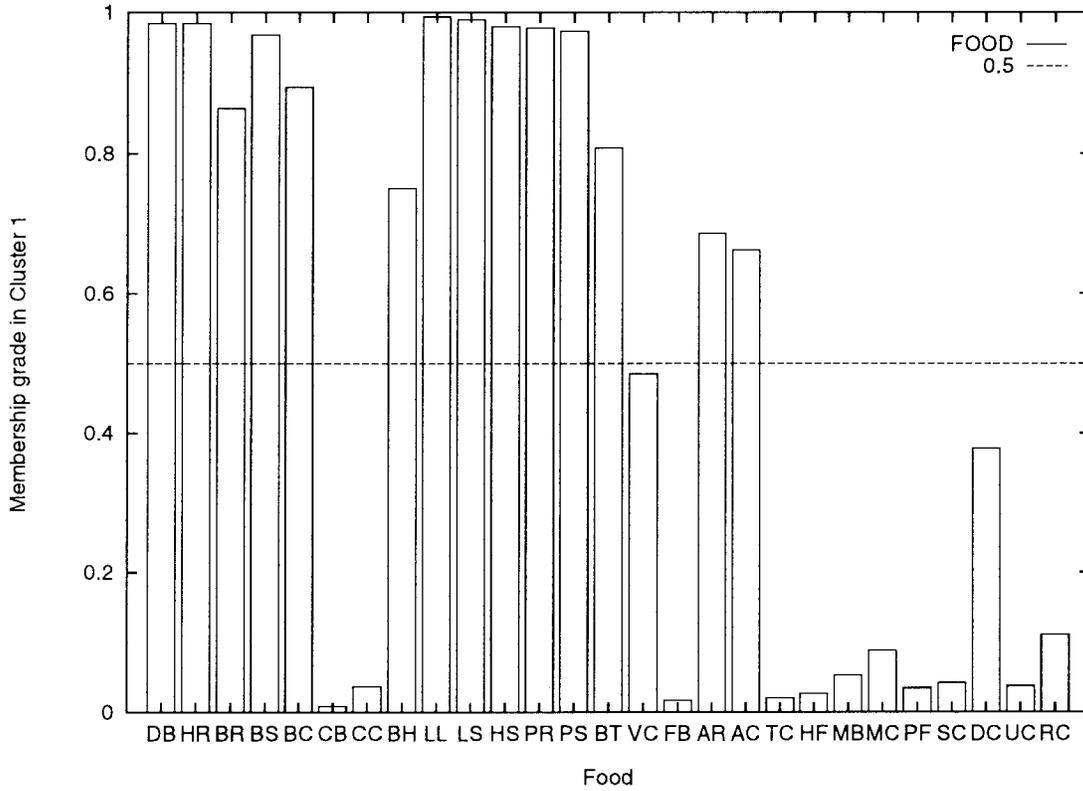


Fig. 16. Membership grade in cluster 1 (percentage of iron of VC is changed).

These results are illustrated in Fig. 15. From these results, we find that the effect of veal cutlet (VC) is quite large and the percentage of iron seems to cause the large effect. To confirm the effect of veal cutlet, we now change the percentage of iron of veal cutlet from 27 to 28 and apply FCM algorithm. The membership grade and prototypes are found in Tables IX and X, respectively, and the membership grade in cluster 1 is illustrated in Fig. 16. From Table IX, we find that the membership grade in the cluster 1 for veal cutlet changes to 0.4848 and the membership grade in the cluster 2 is larger than one in the cluster 1. The results is counter to one for original (unchanged) data; so we confirm that the percentage of iron of veal cutlet greatly affects the results.

## V. CONCLUSIONS

In this paper, we propose a method of identifying influential observations in FCM. As the direction of a perturbation, we consider both specified and unspecified cases and confirm the effectiveness of the method. In a numerical example, we deal with the influence of a single individual. We must develop an efficient algorithm for finding the maximum of the quadratic form stated in Corollary 2.

In examples, we evaluated the effects of perturbation by both changes in membership grades and changes in prototypes. From the practical point of view, various measures of influence may be considered, for example, only changes in prototypes. Moreover, we will investigate the issue of fuzziness in fuzzy clustering from the viewpoint of the sensitivity analysis.

## APPENDIX

### A. Calculation of Hessian of $\bar{J}_m(\bar{\theta})$

Let  $H(\bar{\theta})$  be the Hessian matrix of  $\bar{J}_m(\bar{\theta})$ ; then we get

$$H(\bar{\theta}) = \begin{bmatrix} A_1 & \mathbf{O} & B_{11} & \cdots & B_{1c} \\ & \ddots & & & \\ \mathbf{O} & & A_N & B_{N1} & \cdots & B_{Nc} \\ B_{11} & \cdots & B_{1N} & D_1 & & \mathbf{O} \\ & \cdots & & & \ddots & \\ B_{c1} & \cdots & B_{cN} & \mathbf{O} & & D_c \end{bmatrix}$$

where

$$A_i = \left[ \frac{\partial^2 \bar{J}_m}{\partial u_{\beta i} \partial u_{\gamma i}} \right], \quad B_{i\alpha} = \left[ \frac{\partial^2 \bar{J}_m}{\partial u_{\beta i} \partial v_{\alpha t}} \right]$$

$$D_\alpha = \left[ \frac{\partial^2 \bar{J}_m}{\partial v_{\alpha s} \partial v_{\alpha t}} \right].$$

The elements of  $A_i$ ,  $B_{i\alpha}$ ,  $D_\alpha$  are as follows:

$$\frac{\partial^2 \bar{J}_m}{\partial u_{\beta i} \partial u_{\gamma i}} = \begin{cases} m(m-1) \{ (u_{\beta i})^{m-2} \sum_s^k (x_{is} - v_{\beta s})^2 \\ + (u_{\gamma i})^{m-2} \sum_s^k (x_{is} - v_{\gamma s})^2 \}, & \beta = \gamma \\ m(m-1) (u_{\beta i})^{m-2} \sum_s^k (x_{is} - v_{\beta s})^2, & \beta \neq \gamma \end{cases}$$

TABLE VIII  
MAXIMUM EFFECT OF PERTURBATION AND ITS DIRECTION

Food	Energy	Protein	Fat	Cal.	Iron
BB	0.021	(0.2275	0.1064	-0.0153	0.1757 0.9516 )
HR	0.023	(0.2240	0.1314	0.0969	0.1742 0.9449 )
BR	0.010	(0.2545	0.0790	0.1228	0.9551 -0.0421 )
BS	0.017	(0.2263	0.1022	-0.0179	0.1763 0.9523 )
BC	0.011	(0.0179	-0.1736	0.1927	0.9525 0.1586 )
CB	0.024	(-0.0216	-0.0894	-0.0621	0.2707 0.9562 )
CC	0.021	(-0.0213	-0.0263	0.2450	0.9671 0.0593 )
BH	0.006	(0.1710	0.1771	0.1355	0.9511 -0.1285 )
LL	0.024	(0.2316	0.0966	-0.0200	0.1765 0.9516 )
LS	0.022	(0.2443	0.0837	-0.0210	0.1521 0.9538 )
HS	0.020	(0.2310	0.1001	-0.0137	0.1734 0.9520 )
PR	0.019	(0.2360	0.0935	-0.0193	0.1785 0.9504 )
PS	0.018	(0.2422	0.0884	-0.0180	0.1760 0.9499 )
BT	0.052	(0.1578	0.2074	-0.0278	0.1350 0.9556 )
VC	0.118	(0.1721	0.2012	0.0404	0.1170 0.9563 )
FB	0.021	(-0.0219	-0.0773	0.2326	0.1262 0.9610 )
AR	0.008	(0.2230	-0.0253	0.1057	0.9526 -0.1764 )
AC	0.011	(0.2200	-0.0468	0.0596	0.9472 -0.2206 )
TC	0.022	(-0.0313	-0.0631	0.1995	0.2516 0.9444 )
HF	0.020	(-0.0187	-0.0752	0.2230	0.1601 0.9585 )
MB	0.021	(0.0320	0.0714	0.1132	0.9758 0.1700 )
MC	0.016	(-0.0203	0.2213	0.1326	0.9605 0.1022 )
PF	0.022	(0.0011	0.0100	0.2361	0.9368 -0.2579 )
SC	0.021	(-0.0189	-0.1854	0.0801	0.1895 0.9607 )
DC	0.008	(0.1811	0.1956	0.0957	0.9589 0.0182 )
UC	0.022	(-0.0103	-0.0308	0.2463	0.9678 0.0415 )
RC	0.009	(-0.1363	0.1134	0.1785	0.9625 -0.1018 )

$$\frac{\partial^2 \bar{J}_m}{\partial u_{\beta i} \partial v_{\alpha t}} = \begin{cases} -2m(u_{\beta i})^{m-1}(x_{it} - v_{\beta t}), & \alpha \neq c, \alpha = \beta \\ 0, & \alpha \neq c, \alpha \neq \beta \\ 2m(u_{c i})^{m-1}(x_{it} - v_{c t}), & \alpha = c \end{cases}$$

$$\frac{\partial^2 \bar{J}_m}{\partial v_{\alpha s} \partial v_{\alpha t}} = \begin{cases} 2 \sum_{i=1}^N (u_{\alpha i})^m, & s = t \\ 2 \sum_{i=1}^N (u_{\alpha i})^m, & s \neq t \end{cases}$$

where  $u_{c i} = 1 - \sum_{\alpha=1}^{c-1} u_{\alpha i}$ .

TABLE IX  
RESULT OF FCM WITH  $c = 2$  AND  $m = 1.5$   
(PERCENTAGE OF IRON OF VC IS CHANGED)

Food	Membership grade		Food	Membership grade	
	$u_{1i}$	$u_{2i}$		$u_{1i}$	$u_{2i}$
BB	0.9847	0.0153	VC	0.4848	0.5152
HR	0.9844	0.0156	FB	0.0171	0.9829
BR	0.8633	0.1367	AR	0.6860	0.3140
BS	0.9681	0.0319	AC	0.6623	0.3377
BC	0.8941	0.1059	TC	0.0204	0.9797
CB	0.0088	0.9912	HF	0.0271	0.9729
CC	0.0366	0.9634	MB	0.0529	0.9471
BH	0.7501	0.2499	MC	0.0885	0.9115
LL	0.9939	0.0061	PF	0.0346	0.9654
LS	0.9901	0.0099	SC	0.0421	0.9579
HS	0.9798	0.0202	DC	0.3783	0.6217
PR	0.9777	0.0223	UC	0.0374	0.9626
PS	0.9735	0.0265	RC	0.1113	0.8887
BT	0.8078	0.1922			

TABLE X  
PROTOTYPES OF CLUSTERS (PERCENTAGE OF IRON OF VC IS CHANGED)

Cluster number	Prototypes				
	Energy	Protein	Fat	Calcium	Iron
1	8.523	26.99	20.88	2.959	30.40
2	4.543	27.60	6.369	7.574	15.08

### B. Calculation of $\partial^2 \bar{J}_m / \partial \bar{\theta} \partial \varepsilon$

Let

$$\frac{\partial^2 \bar{J}_m}{\partial \bar{\theta} \partial \varepsilon} (\bar{\theta}_0^*, 0) = \left[ \underbrace{g_1^{(1)'}}_{c-1} \quad \dots \quad \underbrace{g_N^{(1)'}}_{c-1} \quad \underbrace{g_1^{(2)'}}_k \quad \dots \quad \underbrace{g_c^{(2)'}}_k \right]'$$

where

$$g_i^{(1)} = -2m \begin{bmatrix} (u_{1i})^{m-1} \sum_j (x_{ij} - v_{1j}) d_{ij} \\ \vdots \\ (u_{c-1,i})^{m-1} \sum_j (x_{ij} - v_{c-1,j}) d_{ij} \\ (u_{ci})^{m-1} \sum_j (x_{ij} - v_{cj}) d_{ij} \\ \vdots \\ (u_{ci})^{m-1} \sum_j (x_{ij} - v_{cj}) d_{ij} \end{bmatrix},$$

$$i = 1, \dots, l$$

$$g_i^{(1)} = \mathbf{0}, \quad i = l+1, \dots, N$$

$$g_\alpha^{(2)} = \begin{bmatrix} -2 \sum_{i=1}^l (u_{\alpha i})^m d_{i1} \\ \vdots \\ -2 \sum_{i=1}^l (u_{\alpha i})^m d_{ik} \end{bmatrix}$$

and  $u_{c i} = 1 - \sum_{\alpha=1}^{c-1} u_{\alpha i}$ .

### C. Definition of $L(\bar{\theta})$

We obtain  $L(\bar{\theta}_0^*)$  as follows:

$$L(\bar{\theta}) = \begin{bmatrix} L_1^{(1)} & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & L_i^{(1)} \\ L_{11}^{(2)} & \cdots & L_{1l}^{(2)} \\ & \ddots & \\ L_{c1}^{(2)} & \cdots & L_{cl}^{(2)} \end{bmatrix}$$

where

$$L_i^{(1)} = -2m \begin{bmatrix} u_{1i}^{m-1}(x_{i1} - v_{11}) & \cdots & u_{1i}^{m-1}(x_{ik} - v_{1k}) \\ \vdots & \ddots & \vdots \\ u_{c-1,i}^{m-1}(x_{ik} - v_{c-1,k}) & \cdots & u_{c-1,i}^{m-1}(x_{ik} - v_{c-1,k}) \end{bmatrix} \\ + 2m \begin{bmatrix} u_{ci}^{m-1}(x_{i1} - v_{c1}) & \cdots & u_{ci}^{m-1}(x_{ik} - v_{ck}) \\ \vdots & \ddots & \vdots \\ u_{ci}^{m-1}(x_{i1} - v_{c1}) & \cdots & u_{ci}^{m-1}(x_{ik} - v_{ck}) \end{bmatrix},$$

$$L_{\alpha i}^{(2)} = -2(u_{\alpha i})^m I_k$$

and  $I_k$  is the  $k \times k$  identity matrix.

### ACKNOWLEDGMENT

The authors would like to thank the referees for valuable comments and suggestions.

### REFERENCES

- [1] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1974.
- [2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*. New York: Plenum, 1981.
- [3] J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson, "Detection and characterization of cluster substructure 1—Linear structure: Fuzzy c-lines," *SIAM J. Appl. Math.*, vol. 40, no. 2, pp. 339–357, 1981.
- [4] R. N. Dave, "An adaptive fuzzy c-elliptotype clustering algorithm," in *Proc. NAFIPS'90: Quarter Century Fuzziness*, Toronto, Canada, June 1990, pp. 9–12.
- [5] R. J. Hathaway and J. C. Bezdek, "Switching regression models and fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 195–204, Aug. 1993.
- [6] R. N. Dave, "Characterization and detection of noise in clustering," *Pattern Recogn. Lett.*, vol. 12, pp. 657–664, Nov. 1991.
- [7] H. Frigui and R. Krishnapuram, "A robust clustering algorithm based on the M-estimator," in *Proc. Neural, Parallel, Scientific Computat.*, 1995, vol. V-1, pp. 163–166, 1995.
- [8] K. Jajuga, " $l_1$ -norm based fuzzy clustering," *Fuzzy Sets Syst.*, vol. 39, pp. 43–50, 1991.
- [9] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 98–110, May 1993.
- [10] W. Pedrycz, "Fuzzy multimodels," *IEEE Trans. Fuzzy Syst.*, vol. 4, pp. 139–148, May 1996.
- [11] J. C. Bezdek, "A convergence theorem for the fuzzy isodata clustering algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, pp. 1–8, Jan. 1980.
- [12] J. C. Bezdek, R. J. Hathaway, M. J. Sabin, and W. T. Tucker, "Convergence theory for fuzzy C-means: Counterexample and repairs," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-17, pp. 873–877, Sept. 1987.
- [13] R. J. Hathaway and J. C. Bezdek, "Local convergence of the fuzzy c-means algorithm," *Pattern Recogn.*, vol. 19, no. 6, pp. 477–480, 1986.
- [14] S. Z. Selim and M. A. Ismail, "K-means-type algorithms: A generalized convergence theorem and characterization of local optimality," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, pp. 81–87, Jan. 1984.
- [15] ———, "On the local optimality of the fuzzy ISODATA clustering algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 2, pp. 284–288, Mar. 1986.
- [16] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics*. New York: Wiley, 1980.
- [17] Y. Tanaka, "Sensitivity analysis in Hayashi's third method of quantification," *Behaviormetrika*, vol. 16, pp. 31–44, 1984.
- [18] ———, "Influence functions related to eigenvalue problems which appear in multivariate analysis," *Communicat. Statistics*, vol. A18, pp. 3991–4010, 1989.
- [19] ———, "Recent advance in sensitivity analysis in multivariate statistical methods," *J. Japanese Soc. Computat. Statistics*, vol. 7, no. 1, pp. 1–25, 1994.
- [20] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. New York: Springer-Verlag, 1995.
- [21] J. Dieudonné, *Foundations of Modern Analysis*. New York: Academic, 1960.
- [22] J. A. Hartigan, *Clustering Algorithms*. New York: Wiley, 1975.
- [23] E. Ruspini, "Numerical methods for fuzzy clustering," *Inform. Sci.*, vol. 6, pp. 273–284, 1972.



**Hideyuki Imai** received the M.E. degree in information engineering from Hokkaido University, Sapporo, Japan, in 1987.

He is currently a member of the Faculty of Engineering at Hokkaido University. His research interests include statistical inference.

Mr. Imai is a member of the Japan Statistical Society.



**Akira Tanaka** received the B.E and M.E. degrees in information engineering from Hokkaido University, Sapporo, Japan, in 1994 and 1996, respectively.

He is currently with Matsushita Communication Industrial Co., LTD, Yokohama, Japan. His research interests include digital image processing.

Mr. Tanaka is a member of the IEICE of Japan.



**Masaaki Miyakoshi** received the D.E. degree from Hokkaido University, Sapporo, Japan, in 1985.

He is currently a member of the Faculty of Engineering at Hokkaido University. His research interests include fuzzy theory.

Dr. Miyakoshi is a member of the IEICE of Japan.