

Features of Sequential Nonlinear Mapping

Algirdas Mykolas MONTVILAS

Institute of Mathematics and Informatics

A. Goštauto 12, 2600 Vilnius, Lithuania

e-mail: montvila@ktl.mii.lt

Received: April 2003

Abstract. In the paper the sequential nonlinear mapping has been investigated in order to reveal its features. The method was investigated by a plenty of experiments using various sorts of data. For illustrations there are presented results using “marginal” data: the first data gives the smallest mapping error, and the other data gives the largest one. The sequential nonlinear mapping has been investigated according ability to differ the data groups (clustering) when at the beginning the number of groups is taken to be less than really exists. It was showed that the sequential nonlinear mapping differs the groups of data well even though the number of them is taken to be less by one than really exists. The experiments show that the factor for correction co-ordinates on the plane for the *sequential* nonlinear mapping can be taken in the range from 0.25 to 0.75. Mapping errors depend on both the sort of initial conditions and the nature of data.

Key words: sequential nonlinear mapping, mapping error, initial conditions.

1. Introduction

The *simultaneous* nonlinear mapping of data from L -dimensional space to a lower-dimensional space was created by Sammon (1969). The inherent structure of the data is approximately preserved under the mapping. It is achieved by minimizing the error function E , which reveals the largest product of the error and partial error (Duda and Hart, 1973).

$$E = \frac{1}{\sum_{i < j}^N d_{ij}^*} \sum_{i < j}^N \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*}, \quad (1)$$

where N is a number of L -dimensional vectors being mapped, d_{ij}^* – distance between i and j vectors in L -space, d_{ij} – distance in a lower-dimensional space (two-space).

The heuristic relaxation method (Chang and Lee, 1973) runs faster and requires a less amount of memory space. However, these two methods work only having all the data, already.

For the *sequential* nonlinear mapping the triangular method was presented (Lee, *et al.*, 1977). It preserves only two distances to vectors previously mapped and, in addition,

it uses the spanning tree, so it makes the mapping dependent on the history, hence it is usable only for very special tasks.

The *sequential* nonlinear mapping (Montvilas, 1995) occurred to be very successful for sequential mapping of multidimensional data into a lower-dimensional space (frequently onto the plane). It can be applied either for the sequential clustering or for other sequential multidimensional data structure analysis or for supervising of dynamical systems, when each stable state of the system is described by parameters vector (Montvilas, 1999). It allows us to watch the dynamical system states, their change and to indicate its damage (Montvilas, 2000).

In (Montvilas, 2002) this sequential method and Sammon's simultaneous one were compared according ability to map the data onto the plane, mapping accuracy and a mapping time. It was showed that sequential nonlinear mapping has slightly bigger total mapping error but needs incomparably less calculation time, and it was recommended to use the sequential nonlinear mapping for data structure analysis, even having all the data already, especially when there is a large amount of the data.

However, this sequential method needs some investigations in order to reveal its features.

In this paper the method of *sequential* nonlinear mapping has been investigated: a) according ability to differ the data groups when at the beginning the number of groups is taken to be less than really exists; b) according mapping errors dependence on a value of \mathbf{F} ("magic factor") and, c) according mapping errors dependence on a sort of initial conditions.

2. Ability to Differ the Data Groups

The essence of any nonlinear mapping is to preserve the inherent structure of distances among the parameter's vectors being in L -dimensional space after mapping them into two-dimensional space. The *sequential* nonlinear mapping requires at the very beginning to map only M initial vectors *simultaneously* using Sammon's algorithm. After that each sequentially receiving vector has to be mapped in respect of the first M vectors. Mapping error function E_j for each receiving vector $X_j, j = M + 1, \dots, M + N$, is calculated using formula

$$E_j = \frac{1}{\sum_{i=1}^M d_{ij}^X} \sum_{i=1}^M \frac{(d_{ij}^X - d_{ij}^Y)^2}{d_{ij}^X}, \quad j=M+1, \dots, M+N; \quad (2)$$

where d_{ij}^X – distance between i and j vectors in the L -space, d_{ij}^Y – distance on the plane.

It is possible to calculate the *Sammon's mapping error* using formula (1), the *average of sequential mapping error* using formula (2) and afterwards averaging them by N and the *total mapping error* which is calculated using formula (1) after sequential mapping all the data already.

For the correct mapping we have to change the position of every mapped vector $Y_j, j = M + 1, \dots, M + N$, on the plane in such a way that the error E_j would be minimal. This is achieved by iteration procedure using recurrent algorithm. During the $r + 1$ -iteration co-ordinates of the mapped vectors Y_j will be

$$y_{jk}(r + 1) = y_{jk}(r) - \mathbf{F} \cdot \Delta_{jk}(r), \quad j = M + 1, \dots, M + N; k = 1, 2; \quad (3)$$

where

$$\Delta_{jk}(r) = \frac{\partial E_j(r)}{\partial y_{jk}(r)} \bigg/ \left| \frac{\partial^2 E_j(r)}{\partial y_{jk}^2(r)} \right|, \quad (4)$$

\mathbf{F} is the factor for correction of the coordinates (“magic factor”).

The set of the initial vectors M usually consists of the representatives of either stable state describing parameters vector of a dynamic system (Montvilas, 1999) or each cluster (Montvilas, 2002). In other words the M initial vectors represent each of M^* data groups being mapped. Of course some times may be situations when M is not equal to M^* . It is not trouble, when $M > M^*$, but it is not clear how the mapping would behave if $M < M^*$.

To determine the mapping behaviour, when $M < M^*$, a plenty of experiments has been executed using various sorts of data. For illustration, results of mapping using “marginal” data are presented: “Data 1” gives the smallest mapping error and “Data 2” gives the largest one. Let’s have 25 vectors consisting of six parameters (Table 1). Let’s call these data as “Data 1” because they give the smallest mapping error among all data used.

According the experiment these vectors belong to five classes (Table 2).

This investigation was performed at 100 iterations. In Fig. 1 the result of sequential mapping vectors into two-dimensional space at $M = 4$ is presented. The first $M = 4$ vectors mapped simultaneously are denoted by mark x with an index that means the vector’s number, and the remainder $N = 20$ vectors mapped sequentially are denoted by mark $+$ with the respective index. The representative of fifth class was not involved into M vectors, however, all vectors of fifth class were mapped into separate group with some distances to other groups.

Now let’s execute the sequential mapping using the same “Data 1” at $M = 3$.

In the Fig. 2 the result of mapping of these data at $M = 3$ is presented.

Neither fifth nor fourth class’s representatives were not involved into M vectors. This case the classes were separated as well, nevertheless, the distance between fourth and fifth classes is smaller than other distances. In Table 3 the mapping total errors are presented at $M = 5, 4$ and 3 . The mapping total errors were calculated including distances among *all* vectors like in the case of simultaneous mapping (formula(1)).

Let’s repeat the similar mapping using another kind of data (“Data 2”), which give the largest mapping error among other data used. In the Table 4 the “Data 2” consisting of 30 vectors is presented. Each vector consists of six parameters, as well.

They are distributed to five classes as well (see Table 5):

Table 1
25 vectors of the "Data 1"

Vect. No	Parameters					
	1	2	3	4	5	6
1	50.0	220.0	307.0	508.0	604.0	206.0
2	51.0	222.0	302.0	501.0	603.0	205.0
3	53.0	228.0	304.0	502.0	607.0	208.0
4	55.0	226.0	305.0	505.0	609.0	204.0
5	56.0	224.0	308.0	507.0	608.0	202.0
6	50.3	220.0	307.0	507.7	604.0	206.0
7	50.1	220.2	307.2	508.0	604.0	206.3
8	50.4	220.7	307.0	508.2	604.0	206.1
9	50.2	220.5	307.1	507.9	603.9	205.8
10	51.6	222.2	302.0	501.0	603.0	205.1
11	50.9	221.8	302.1	500.9	603.1	204.9
12	50.7	222.4	302.2	501.2	603.2	205.0
13	50.8	222.6	302.2	501.5	603.0	205.5
14	52.5	228.4	304.2	502.0	607.0	208.0
15	52.9	227.8	304.2	502.3	607.3	208.1
16	53.2	228.2	304.0	502.0	607.0	208.2
17	53.1	227.9	303.8	501.8	606.4	207.9
18	55.0	226.5	304.5	504.7	608.7	203.9
19	55.1	225.9	305.5	504.9	609.1	204.1
20	54.9	226.2	304.9	505.0	609.2	204.0
21	55.2	225.8	304.8	505.2	609.0	204.0
22	56.2	224.2	307.9	507.0	609.2	202.0
23	55.9	224.4	308.4	507.5	608.0	202.5
24	56.3	223.8	309.0	507.5	607.2	202.2
25	55.8	224.1	308.6	507.0	608.0	201.9

The result of sequential mapping of the "Data 2" at $M = 4$ is presented in Fig. 3 and at $M = 3$ in Fig. 4, respectively.

The mapping total errors at $M = 5, 4$ and 3 are presented in the Table 6.

Analysing the mapping results we see that at $M = 4$ the mapping procedure differs the data into the groups well. However, at $M = 3$ the first three data groups are separated well meantime the reminder data (of fourth and fifth groups) are mapped into more or less one place. The mapping total error increasing by decreasing M (Table 3 and Table 6) shows the decline of mapping quality. A great deal of experiments has been executed with various sorts of data. Always groups which representatives were involved into M vectors were mapped correctly, and the reminder data were separated into give or take close place on the screen. So, if the number of groups M^* is not known, the *sequential* nonlinear mapping is useful when the number of the initial vectors M is taken to be at least $M = M^* - 1$.

Table 2
25 vectors of "Data 1" distributed to five classes

Class	Vectors
1	1, 6, 7, 8, 9
2	2, 10, 11, 12, 13
3	3, 14, 15, 16, 17
4	4, 18, 19, 20, 21
5	5, 22, 23, 24, 25

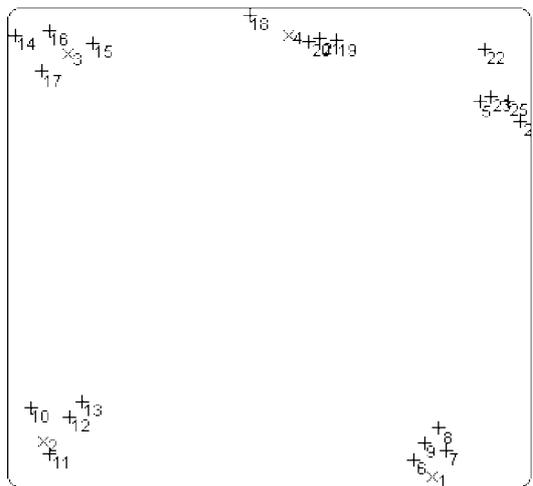


Fig. 1. Mapping of the "Data 1" at $M = 4$.

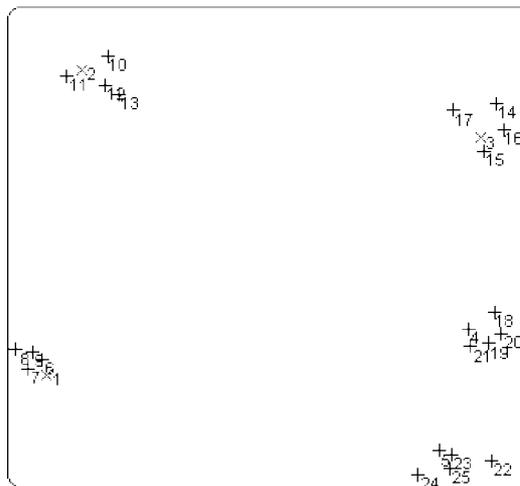


Fig. 2. Mapping of the "Data 1" at $M = 3$.

Table 3
Mapping total errors of "Data 1".

M	Total errors
5	0.005383999
4	0.007153901
3	0.012378900

Table 4
30 vectors of the "Data 2"

Vec. No	Parameters						Vec. No	Parameters					
	1	2	3	4	5	6		1	2	3	4	5	6
1	1.0	5.0	7.0	8.0	4.0	6.0	16	8.1	2.1	4.0	2.1	7.1	8.3
2	6.0	3.0	2.0	1.0	3.0	5.0	17	9.0	2.0	4.2	2.0	7.0	8.0
3	8.0	2.0	4.0	2.0	7.0	8.0	18	8.0	3.0	4.8	2.3	7.0	8.0
4	3.0	8.0	5.0	5.0	9.0	4.0	19	5.4	2.1	4.0	7.0	7.0	2.0
5	5.0	3.0	4.0	7.0	8.0	2.0	20	7.9	1.8	3.8	1.0	6.4	7.9
6	2.5	7.5	5.1	4.8	9.0	3.0	21	7.0	1.5	3.9	1.8	6.8	8.0
7	3.0	7.8	5.5	4.9	9.1	4.1	22	6.0	4.5	2.1	0.9	3.1	6.0
8	2.9	7.9	4.9	5.0	9.2	4.0	23	6.0	3.0	2.9	1.2	3.2	5.0
9	3.1	8.1	4.8	5.2	9.0	4.0	24	6.1	3.2	2.3	1.2	3.2	4.1
10	5.2	3.3	5.0	7.2	7.5	2.0	25	6.2	3.1	1.9	1.2	3.2	4.1
11	5.1	3.2	4.8	7.1	7.8	2.1	26	6.1	2.9	1.8	1.0	3.0	4.0
12	4.8	3.1	4.2	6.8	7.7	2.2	27	1.0	5.1	7.0	7.6	4.0	6.0
13	3.1	8.7	5.3	5.2	9.3	4.1	28	1.0	5.2	8.0	8.0	4.0	6.3
14	3.3	8.3	5.5	5.5	9.8	4.3	29	1.0	5.1	7.0	7.8	4.0	6.0
15	3.2	8.4	5.4	5.1	8.2	4.4	30	1.1	5.5	6.0	9.0	4.0	6.0

Table 5
30 vectors of "Data 2" distributed to five classes

Class	Vectors
1	1, 27, 28, 29, 30
2	2, 22, 23, 24, 25, 26
3	3, 16, 17, 18, 20, 21
4	4, 6, 7, 8, 9, 13, 14, 15
5	5, 10, 11, 12, 19

3. Mapping Errors Dependence on the Value of F

F -factor in (Montvilas, 1995), or MF – "magic factor" in (Sammon, 1969) is using for correction of the vector's co-ordinates on the plane during each iteration procedure. In (Kohonen, 2001) it was proposed to take the F value between 0.3 and 0.4 and in (Groenen

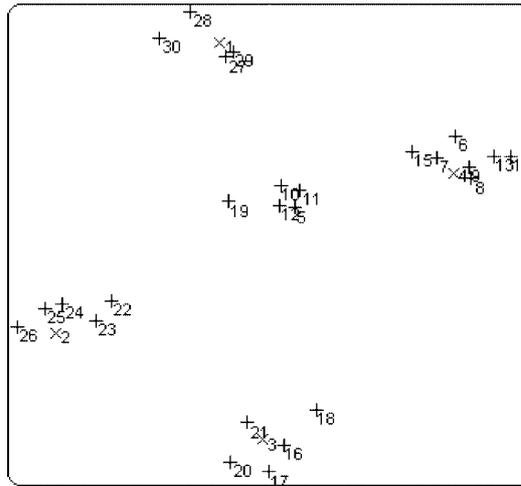


Fig. 3. Mapping of the “Data 2” at $M = 4$.

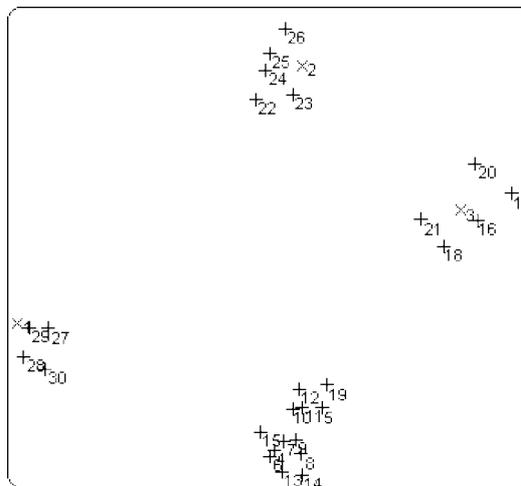
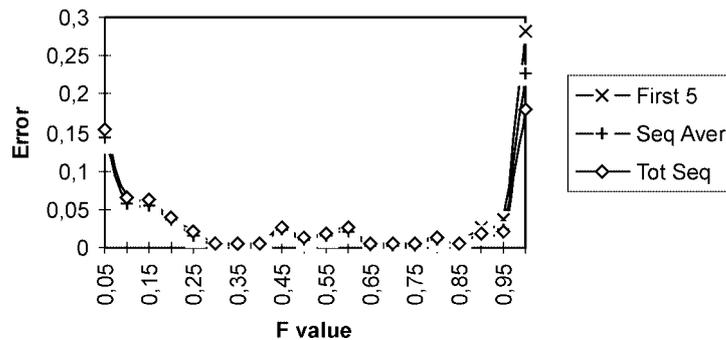


Fig. 4. Mapping of the “Data 2” at $M = 3$.

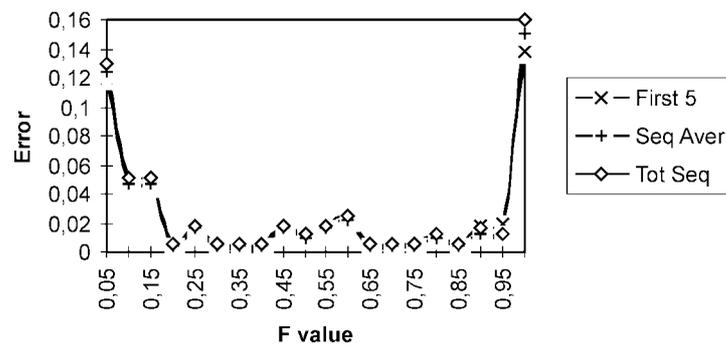
Table 6
Mapping total errors of “Data 2”

M	Total errors
5	0.02750241
4	0.03240145
3	0.07661533

"Data 1", 50 iterations

Fig. 5. Dependence the mapping errors on F for "Data 1" at 50 iterations.

"Data 1", 200 iterations

Fig. 6. Dependence the mapping errors on F for "Data 1" at 200 iterations.

and Heiser, 1996) between 0.25 and 0.45. However, investigations carried out show that for the sequential nonlinear mapping the range of F could be taken more wide. A lot of experiments show that mapping error's curve by changing F in the range from 0.05 to 1.0 has rather wide part with minimum mapping error, and only at the ends of the range the mapping errors grow up. The range of F with minimum mapping error and minimum error value depend on the nature of data and the number of iterations. For illustration that, the mapping errors both the first $M = 5$ vectors and sequential average and total sequential for "Data 1" and "Data 2" at 50 and 200 iterations are presented in Fig. 5 – Fig. 8, respectively. The errors were calculated at several sort of initial conditions and then averaged.

A great deal of experiments have been executed using various sorts of data and experiment's conditions. They showed that the factor F for correction of the co-ordinates on the plane for the *sequential* nonlinear mapping could be taken in the range from 0.25 to 0.75.

"Data 2", 50 iterations

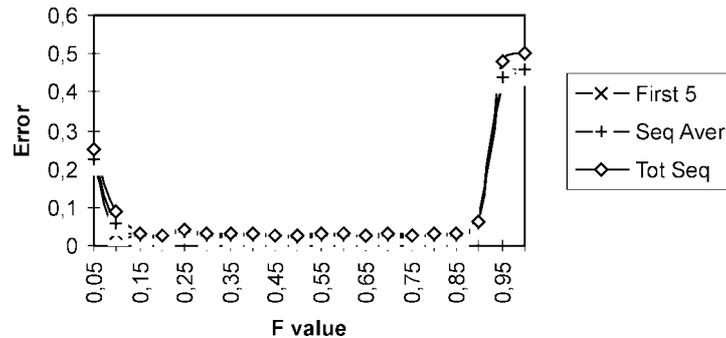


Fig. 7. Dependence the mapping errors on **F** for "Data 2" at 50 iterations.

"Data 2", 200 iterations

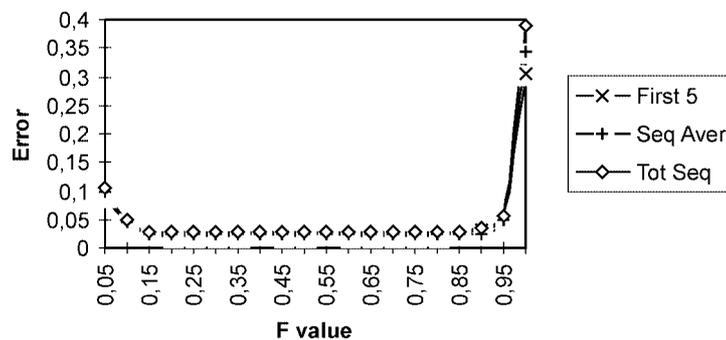


Fig. 8. Dependence the mapping errors on **F** for "Data 2" at 200 iterations.

4. Mapping Error Dependence on Initial Conditions

Mapping error depends on the nature of data, but the error especially depends on initial conditions because any nonlinear mapping algorithm often finds the local maximum of a functional that characterises the mapping quality which is not global (Dzemyda, 1997) because of bad initial conditions. There are many ways to choose initial conditions using certain knowledge of the data. The matter is that most often we have only the data without any knowledge about it. Formerly, a random distribution of initial points on the plane for nonlinear mapping was used. It was noticed by many researches that there is no way to choose optimal case of distribution which would give the smallest mapping error. Along it was noticed that the mapping error is comparable small if initial points are distributed along more or less shifted diagonal. So it was expedient to use it in the experiments.

The experiments have been carried out using several kinds of slightly shifted descending or ascending diagonals:

- D – descending diagonal,
- DS – descending shifted diagonal,
- DMS – descending more shifted diagonal,
- DOMS – descending one more shifted diagonal,
- A – ascending diagonal,
- AS – ascending shifted diagonal,
- AMS – ascending more shifted diagonal,
- AOMS – ascending one more shifted diagonal.

The mapping error of the first $M = 5$ (number of data groups) vectors, the average of sequential mapping errors, the sequential mapping total error along with Sammon’s mapping error were calculated. In Fig. 9 the results are presented for “Data 1” and in Fig. 10 for “Data 2”, respectively. 500 iterations were used because Sammon’s method does need much more iterations than sequential does.

The experiments show that for the “Data 1” every mapping error is of a little value

"Data 1", 500 iterations

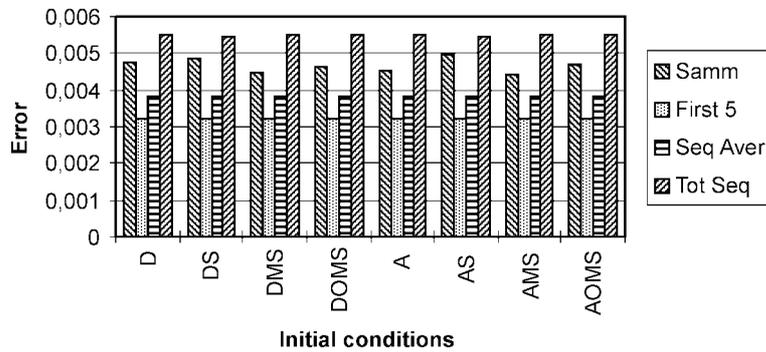


Fig. 9. Magnitudes of the errors at various initial conditions for “Data 1”.

"Data 2", 500 iterations

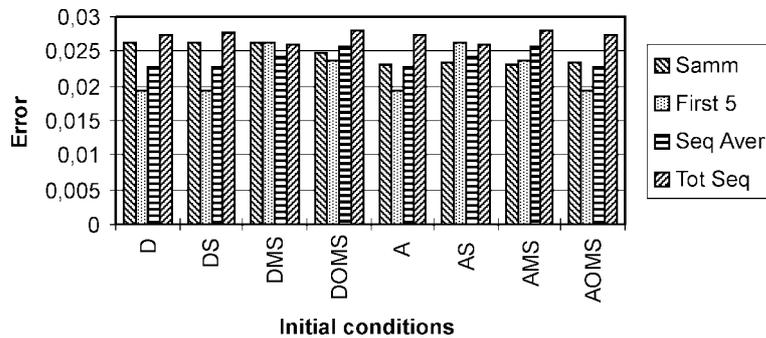


Fig. 10. Magnitudes of the errors at various initial conditions for “Data 2”.

and they change only a little by changing the initial conditions. For the “Data 2”, on the contrary, the errors are rather bigger than that of the “Data 1”, and their values change itself by changing the initial conditions in a more extent range. It means that for the “Data 1” this sort of initial conditions is almost optimal and the mapping quality functional is near to the global maximum.

Numerous experiments, have been executed using various sorts of data, show that mapping errors depend on both the sort of initial conditions and the nature of data. Notice that comparative values of four kinds of errors mentioned above remain give or take the same.

5. Conclusions

The method of the sequential nonlinear mapping has been investigated according ability to differ the data groups when the number of groups M is taken less than really exists M^* , according mapping errors dependence on a value of the factor \mathbf{F} for correction of co-ordinates on the plane and according mapping errors dependence on a sort of initial conditions.

It was showed that the sequential nonlinear mapping differs the groups of data when the number of initial vectors M , mapped simultaneously, is taken to be at least $M = M^* - 1$. The factor \mathbf{F} for correction co-ordinates on the plane for the sequential nonlinear mapping can be taken in the range from 0.25 to 0.75. Mapping errors depend on both the sort of initial conditions and the nature of data.

Acknowledgement

The author is very grateful to unknown referee for his helpful comments that greatly helped to improve the original version of this manuscript.

References

- Chang C.L., and R.C.T. Lee (1973). A heuristic relaxation method for nonlinear mapping in cluster analysis. In *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3. pp. 197–200.
- Duda R.O., and P.E. Hart (1973). *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, London, Sydney, Toronto.
- Dzemyda G. (1997). Clustering of parameters on the basis of correlations: a comparative review of deterministic approaches. *Informatica*, **8**(1), 83–118.
- Groenen P.J.F., and W.J. Heiser (1996). Tunnelling method for global optimisation in multidimensional scaling. *Psychometrika*, **61**, 529–550.
- Kohonen T. (2001). *Self-Organizing Maps*, 3rd ed. Springer Series in Information Sciences, Vol. 30. Springer-Verlag.
- Lee R.C.T., J.R. Slangle and H. Blum (1977). A triangulation method for the sequential mapping of points from N -space to two-space. In *IEEE Trans. on Computers*, Vol. c-26(3). pp. 288–292.
- Montvilas A.M. (1995). On sequential nonlinear mapping for data structure analysis. *Informatica*, **6**(2), 225–232.

- Montvilas A.M. (1999). Issues for design of information system for supervision and control of dynamic systems. *Informatica*, **10**(3), 289–296.
- Montvilas A.M. (2000). Processing of information for supervision and control of technological processes. In *Proceedings of the IFAC Workshop*, Vienna, Pergamon press. pp. 39–43.
- Montvilas A.M. (2002). Sequential nonlinear mapping versus simultaneous one. *Informatica*, **13**(3), 333–343.
- Sammon J.W. (1969). A nonlinear mapping for data structure analysis. In *IEEE Trans. on Computers*, Vol. c-18(5), pp. 401–409.

A.M. Montvilas is a doctor of technical sciences, a senior researcher of the Recognition Processes Department at the Institute of Mathematics and Informatics and an associate professor of the Radioelectronics Department at the Vilnius Gediminas Technical University. His research interests include processing of random processes, detection of dynamic system states and their changes, data visualisation, discrete signal processing, sequential nonlinear mapping.

Nuoseklus netiesinio atvaizdavimo savybės

Algirdas Mykolas MONTVILAS

Straipsnyje pateikti nuoseklus netiesinio atvaizdavimo metodo (Montvilas, 1995) duomenų struktūros analizei bei vizualizavimui tyrimų rezultatai, norint atskleisti jo savybes. Metodas buvo tiriamas pagal gebėjimą skirstyti duomenis į grupes (klasterizavimą), kai duomenų grupių kiekis pradžioje imamas mažesnis, negu yra iš tikrųjų, pagal atvaizdavimo paklaidos priklausomybę nuo koordinacių plokštumoje koregavimo koeficiento dydžio bei pagal atvaizdavimo paklaidos priklausomybę nuo pradinių sąlygų ir duomenų tipo.

Parodyta, kad nuoseklus netiesinio atvaizdavimo metodas gerai skirsto duomenis į grupes, net kai pradinis duomenų grupių kiekis imamas vienetu mažesnis už tikrą; koordinacių plokštumoje koregavimo koeficientą nuosekliam atvaizdavimui galima imti ribose nuo 0,25 iki 0,75; atvaizdavimo paklaida priklauso nuo pradinių sąlygų neįžymaus pakeitimo bei nuo duomenų tipo.