Ann. Inst. Statist. Math.
30 (1978), Part A, 97-104

CONSTELLATION GRAPHICAL METHOD FOR REPRESENTING MULTI-DIMENSIONAL DATA

KAZUMASA WAKIMOTO AND MASAAKI TAGURI

(Received Mar. 28, 1977; revised Dec. 19, 1977)

1. Introduction

In the representation methods of multi-dimensional data on a 2-dimensional plane, various procedures have been hitherto considered. Most of these procedures are performed by transforming the given multi-dimensional data into two real values using two different real valued functions, and these are graphically represented on a perpendicularly intersecting plane. For instance, in a principal component analysis, the given data are expressed by two 1-dimensional real values, that is, the 1st principal component and the 2nd principal component, and by plotting the data on a 2-dimensional plane taking them as the ordinate and abscissa, the structure of the data is interpreted. In the Andrews' method [1], the data are transformed by a real valued function, and expressed as curves on the 2-dimensional plane.

On the contrary, in the radar chart, the Face graph [2], and the Tree graph [3], a graph is drawn directly for the values of data with very simple transformation, and by a visual treatment of the graph, it is intended to grasp the structure of data.

These methods are selected according to the object of analysis, and the field of application, and contribute to the analysis of data, but it is considered that information in the 2-dimensional plane has not been utilized sufficiently. For instance, in the graph on the 2-dimensional plane in case of the principal component analysis, it is difficult to grasp visually the actual structure of data. In the Andrews' method, a similar difficulty is accompanied. On the other hand, in the radar chart, the Face graph, and Tree graph, it is easy to grasp visually the structure of individual data, but there remains a problem in the comparison of structure among them on the whole as the number of data is increased.

In this paper, Constellation Graphical Method is proposed, in which the problematical points as mentioned above are improved in some parts. In this method, it is easy to grasp visually the actual structure of data on the whole and to grasp the characteristic for individual data simultaneously.

2. The algorithm of Constellation Graphical Method

Assume that n pairs of data are given as follows:

$$(2.1) \qquad (x_{1\alpha}, x_{2\alpha}, \cdots, x_{k\alpha}), \qquad (\alpha=1, 2, \cdots, n).$$

(1) Transform the given data (2.1) by k real valued functions $f_1, f_2, \cdots, f_k$ as follows:

$$(2.2) \qquad \xi_{j\alpha}=f_j(x_{j\alpha}), \qquad (j=1, 2, \cdots, k;\ \alpha=1, 2, \cdots, n),$$

where f_j $(j=1, 2, \cdots, k)$ is assumed to satisfy the following conditions.

$$(2.3) \qquad \begin{array}{ll} \text{(A)} & 0\leqq f_j(x_{j\alpha})\leqq\pi, \qquad (\alpha=1, 2, \cdots, n). \\ \text{(B)} & f_j \text{ is a strictly monotone function.} \end{array}$$

When our method is applied to given data, (A) must always be satisfied. Although (B) may not be satisfied necessarily, we assume (B) in order that actual meaning is clear at the time of the interpretation of results.

For example, we will show some typical transformations.

1° In case of continuous data

$$(2.4) \qquad f_j(x_{j\alpha})=\frac{x_{j\alpha}-x_{jl}}{x_{ju}-x_{jl}}\pi,$$

where

$$x_{ju}=\max_{1\leqq\alpha\leqq n} x_{j\alpha}, \quad x_{jl}=\min_{1\leqq\alpha\leqq n} x_{j\alpha}; \qquad (j=1, 2, \cdots, k).$$

2° In case of categorical data

Without loss of generality, put the set of values of $x_{j\alpha}$ to $\{1, 2, \cdots, n\}$, where n is a positive integer.

$$(2.5) \qquad f_j(x_{j\alpha})=\frac{x_{j\alpha}-1}{n-1}\pi.$$

In above transformations, f_j is linear. This reason is to make the interpretation of results easy.

As the result of the transformation (2.2), the data (2.1) becomes to the following form:

$$(2.6) \qquad (\xi_{1\alpha}, \xi_{2\alpha}, \cdots, \xi_{k\alpha}), \qquad (\alpha=1, 2, \cdots, n).$$

(2) A complex number z_α is corresponded to the data (2.6) as follows:

(2.7) $$z_\alpha=\sum_{j=1}^{k} w_j \exp(i\xi_{j\alpha}), \qquad (\alpha=1, 2, \cdots, n),$$

where

$$i=\sqrt{-1}$$

and w_j is the weight to be assigned to the jth variable and the followings are assumed to be satisfied:

(2.8) $$\sum_{j=1}^{k} w_j=1 ; \quad w_j \geqq 0 , \qquad (j=1, 2, \cdots, k) .$$

Then from (2.7) and (2.8), as to the absolute value of z_α we can easily see

(2.9) $$|z_\alpha|=\left|\sum_{j=1}^{k} w_j \exp(i\xi_{j\alpha})\right| \leqq \sum_{j=1}^{k} |w_j \exp(i\xi_{j\alpha})| = \sum_{j=1}^{k} w_j=1 ,$$

where the sign of equality holds if and only if $\xi_{j\alpha}=\xi_{j'\alpha}$ for all j and j'. Moreover from (2.3), as to the argument of z_α

(2.10) $$0 \leqq \arg(z_\alpha) \leqq \pi$$

holds. Therefore from (2.9) and (2.10), z_α must exist in the upper half of the unit circle on the 1-dimensional complex plane C^1 and is just on the circumference if and only if $\xi_{1\alpha}, \cdots, \xi_{k\alpha}$ have same angles.

In this way n comlex values z_α $(\alpha=1, 2, \cdots, n)$ are plotted within the upper half of the unit circle on C^1. This pattern will be called by the name of "*Constellation Graph*" (see, Fig. 1).

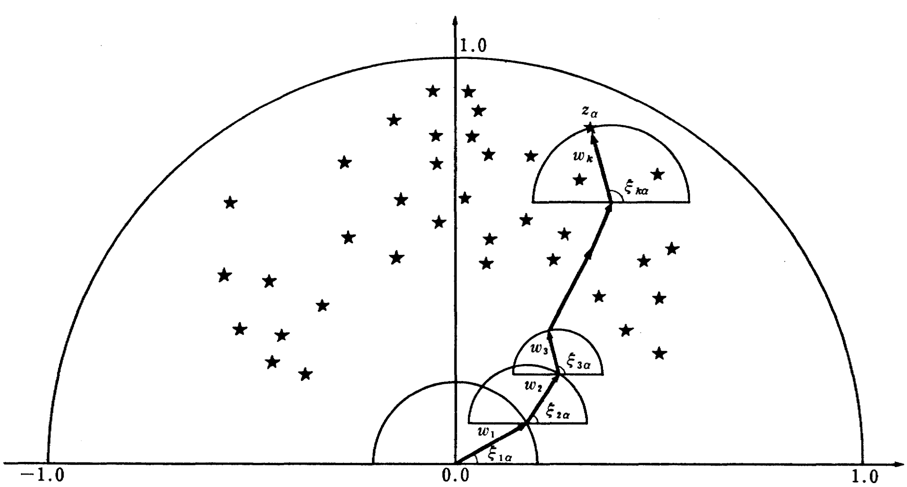

Fig. 1. Constellation Graph.

3. The features of constellation graph

In this section we will investigate the features of constellation graph, where $w_1, w_2, \cdots, w_k$ are assumed to be given. The features of constellation graph may be stated as follows:

(a) We can obtain easily the actual structure of whole data and the detailed structure for individual data simultaneously; the reason is as follows.

We rewrite (2.7) to

$$z_\alpha = \sum_{j=1}^{k} w_j(\cos \xi_{j\alpha} + i \sin \xi_{j\alpha}) , \qquad (\alpha=1, 2, \cdots, n) . \tag{3.1}$$

From (3.1), it is understood that the position of the αth data on constellation graph is the end point of k linked vectors whose arguments and lengths are $\xi_{j\alpha}$ and w_j $(j=1, 2, \cdots, k)$ respectively. We call this linked pattern to the path of the αth data. Then this gives the actual structure of individual data, that is, the path corresponds to each radar chart or each face in the face graph, or each tree in the tree graph or each curved line in the method of Andrews.

From the n end points and the paths on the constellation graph, it is possible easily to grasp the structure of whole data and obtain the detailed structure for individual data simultaneously.

(b) The argument and absolute value of z_α have the information of the weighted mean and standard deviation of k variables respectively in some sense.

(c) Many kinds of information can be written on a same constellation graph; For example, when a complex number z_α is plotted on constellation graph, we can incorporate much information by changing the shape or color of a end point of k linked vectors.

4. Examples

Example 1. Representation of educational data

Suppose that in a class of a middle school, the records of five subjects $(x_{1\alpha}, x_{2\alpha}, x_{3\alpha}, x_{4\alpha}, x_{5\alpha})$, $(\alpha=1, 2, \cdots, 41)$ are given (see Table 1). Here, $x_{j\alpha}$ satisfies $0.0 \leqq x_{j\alpha} \leqq 10.0$, $(j=1, 2, 3, 4, 5;\ \alpha=1, 2, \cdots, 41)$, and five subjects are as follows: Japanese, Social Studies, Mathematics, Science and English.

Then we obtain the constellation graph such as Fig. 2 by using the following transformation:

$$\xi_{j\alpha} = f_j(x_{j\alpha}) = \frac{x_{j\alpha} - 0.0}{10.0 - 0.0}\pi , \qquad (j=1, 2, 3, 4, 5;\ \alpha=1, 2, \cdots, 41) , \tag{4.1}$$

$$z_\alpha=\frac{1}{5}\sum_{j=1}^{5}\exp(i\xi_{j\alpha}), \qquad (\alpha=1, 2, \cdots, 41). \tag{4.2}$$

The reason why a monotone increasing function was used as the transformation formula is as follows. From (4.1), the pupils whose records of each subject are all 10, 5, 0 are respectively plotted at the positions (−1.0, 0.0), (0.0, 1.0), (1.0, 0.0) on the constellation graph. From this fact it is understood that the records of each subject of the pupils are better, as the argument of the points plotted on the constellation graph becomes larger. Therefore, if such a transformation is used, the results of average record will be found at once by considering the argument of plotted point, which is extremely convenient in the interpretation of the results on the constellation graph.

The reason why an equal weight such as $w_1=w_2=w_3=w_4=w_5=1/5$ was used in (4.2) is as follows. If there is any particularly important subject among the 5 subjects, the value of weight to be given to the subject must be large, but the five subjects taken up in the example are same important for the pupils of middle school. Therefore, it is considered that putting an equal weight is most natural.

Table 1. Records of five subjects

Boy pupil						Girl pupil					
No.	Jap.	Soc.	Math.	Sci.	Engl.	No.	Jap.	Soc.	Math.	Sci.	Engl.
1	3	6	4	3	4	22	10	9	8	9	10
2	8	7	7	6	2	23	5	5	6	7	5
3	4	5	2	5	5	24	2	3	2	3	2
4	5	2	3	5	3	25	6	9	5	8	6
5	8	7	2	3	8	26	9	9	8	9	9
6	4	2	7	2	2	27	7	7	3	5	7
7	8	10	3	8	9	28	5	6	3	4	5
8	8	8	8	7	10	29	7	5	3	2	6
9	7	8	7	9	9	30	9	10	9	9	10
10	3	2	2	2	4	31	6	6	6	7	6
11	7	8	9	8	8	32	3	3	1	1	2
12	6	8	2	8	7	33	4	4	3	8	9
13	5	6	8	9	6	34	4	3	6	8	7
14	9	8	7	10	9	35	7	6	6	5	6
15	2	1	3	2	0	36	2	1	5	1	2
16	9	9	8	8	9	37	8	6	7	5	6
17	3	5	8	7	5	38	3	6	0	2	0
18	9	8	9	8	9	39	9	10	8	9	8
19	6	3	1	6	5	40	6	5	3	6	8
20	10	5	7	7	4	41	8	9	10	10	8
21	3	3	6	7	2						

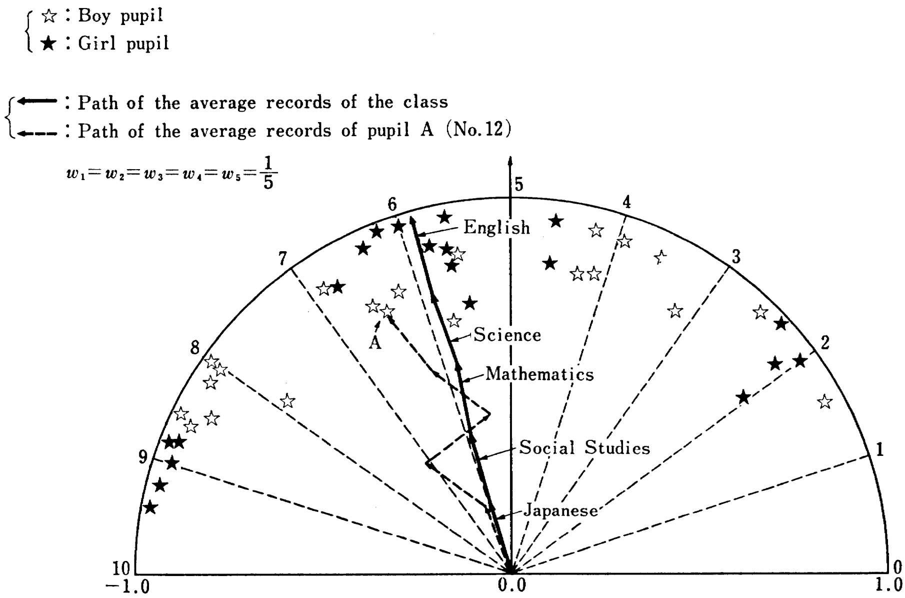

Fig. 2. The Constellation Graph in Example 1.

From Fig. 2, the following facts are clearly observed.

1° From the comparison between the path of the pupil A (No. 12) and the path corresponding to five average values of each subject for the class, we can easily see the following points:

(i) The records of five subjects of the pupil A are superior than the average values of the class except for mathematics.

(ii) The record of mathematics of the pupil A is extremely inferior.

2° Judging from the positions of whole points, the total record of the pupil A belongs to the upper of middle group.

3° The girl pupils are separated into three groups, because the marks ★ are positioned collectively at the arguments nearly of $\pi/5$, $3\pi/5$, $9\pi/10$.

Example 2. Analysis of environmental pollution data

Suppose that the data regarding SO_2 (ppm), NO_x (ppm), and floating dust (mg/m^3) in the 16 observation points of a certain city are given (see, Table 2). These data are the daily average values of a certain day in the observation points. Using the transformation method of (2.4) and varying the value of weights, two constellation graphs as shown in Fig. 3 and Fig. 4 are obtained for the weights $w_1=w_2=w_3=1/3$ and $w_1=0.002$, $w_2=0.378$, $w_3=0.620$ respectively. In these figures, the observation points near by a highway are indicated by the mark • and the other observation points by the mark ∘. And the pollution matters corresponding to w_1, w_2, w_3 are SO_2, floating dust, and NO_x respectively.

From Fig. 3 and Fig. 4, we find that the separation between the

Table 2. The data of Environmental Pollution

No.	SO_2 (ppm)	D-D (mg/m^3)	NO_x (ppm)	No.	SO_2 (ppm)	D-D (mg/m^3)	NO_x (ppm)
1	0.025	0.113	0.21	9	0.008	0.082	0.04
2	0.015	0.073	0.11	10	0.008	0.045	0.14
3	0.011	0.072	0.09	11	0.007	0.089	0.05
4	0.014	0.070	0.08	12	0.009	0.074	0.04
5	0.016	0.075	0.08	13	0.000	0.078	0.11
6	0.013	0.047	0.05	14	0.017	0.078	0.00
7	0.007	0.063	0.10	15	0.015	0.064	0.09
8	0.007	0.039	0.05	16	0.025	0.074	0.14

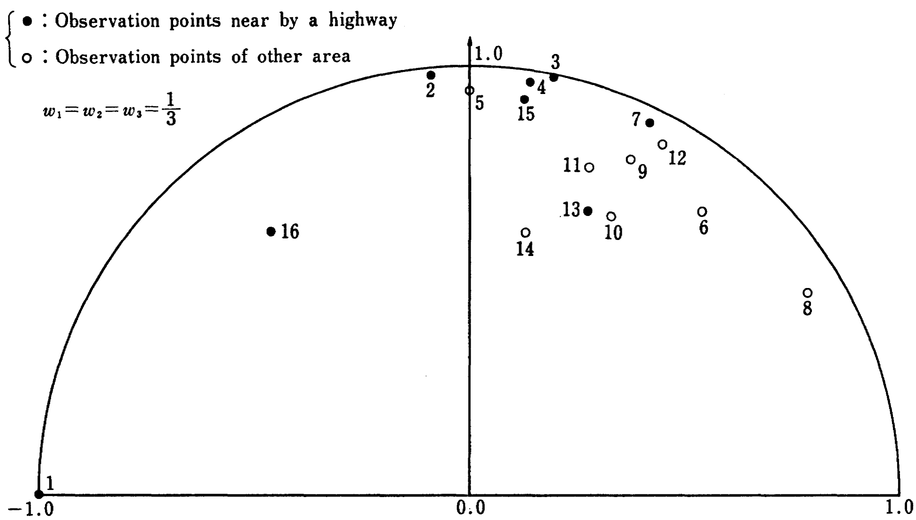

Fig. 3. The Constellation Graph in Table 2.

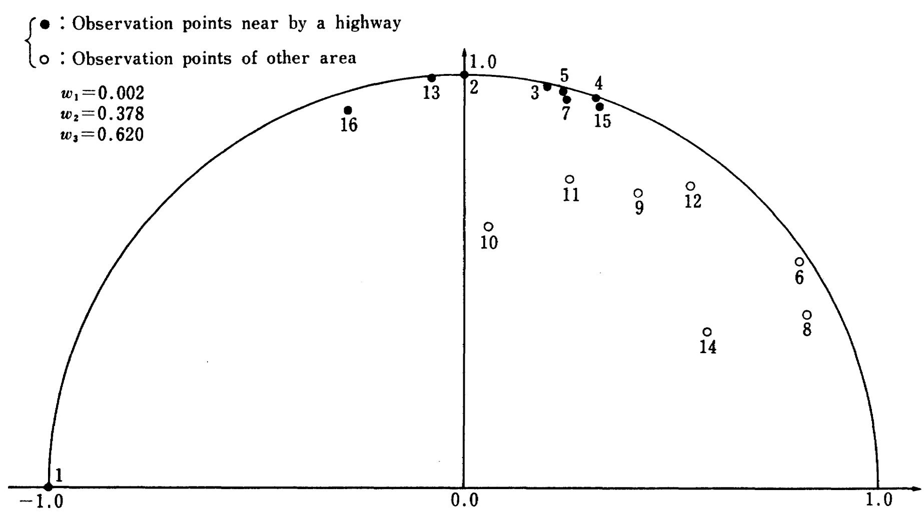

Fig. 4. The Constellation Graph in Table 2.

set of mark • and the set of mark ○ in Fig. 4 is more clear than it in Fig. 3. In the Fig. 4, the weight for NO_x is maximum and that for SO_2 is very small. This is probably explained by the fact that NO_x exhausted from motor cars running on the highway gives a large influence on the separation between the observation points near by a highway and the other observation points.

Remark. From this example, we can propose that the constellation graph is useful for discriminant or cluster analysis.

Acknowledgement

The authors are grateful to the referee for his helpful comments.

OKAYAMA UNIVERSITY
CHIBA UNIVERSITY

REFERENCES

[1] Andrews, D. F. (1972). Plots of high-dimensional data, *Biometrics*, **28**, 125-136.
[2] Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically, *J. Amer. Statist. Ass.*, **68**, 361-368.
[3] Wakimoto, K. (1977). Tree graph method for visual representation of multidimensional data, *J. Japan Statist. Soc.*, **7**, 27-34.