2015 Doctoral Thesis

Constrained Analysis of Asymmetric dissimilarity data by using both MDS and Clustering

Graduate School of Culture and Information Science,
Doshisha University

48121001

Kensuke Tanioka

Superviser Prof. Hiroshi Yadohisa

2016, March

2015 Doctoral Thesis

Constrained Analysis of Asymmetric dissimilarity data by using both MDS and Clustering

Graduate School of Culture and Information Science,
Doshisha University

48121001

Superviser Prof. Hiroshi Yadohisa

Abstract

Asymmetric dissimilarity data exists and is obserable in a variety of fields such a marketing and psychology. Given asymmetric dissimilarity data, it is important to interpret such asymmetries between objects. A visualization method for asymmetric relations, asymmetric Multidimensional scaling (AMDS) is often used to interpret asymmetric relations between objects; however, in some situations, it becomes far too difficult to visually interpret the asymmetric relations from the results of AMDS since the number of parameters in AMDS is larger than that in ordinal MDS. To overcome this problem, we propose simultaneous methods of AMDS and clustering based on concepts of parsimonious models. Using our proposed methods, asymmetries between clusters are represented in low-dimensions rather than between objects. Results of this simultaneous approach provide us with asymmetric information with a smaller number of parameters; however various AMDS methods have been proposed. In short, various ways of representing asymmetries exist. Therefore, when one simultaneous method is proposed, AMDS model should be selected. In this study, we present from two distinct perspectives, simultaneous methods of AMDS and clustering. Furthermore, we note how these methods are characterized by various decompositions of the respective objective functions, as well as relations for feasible areas between these objective functions.

First, from Unfolding type AMDS, we propose two types of simultaneous methods, namely constrained Unfolding and the constrained slide-vector model. Unfolding type AMDS consists of two types of models, one that applies Unfolding (Coombs, 1950) to asymmetric dissimilarity data, the other method being the slide-vector model (Zieltman and Heiser, 1993), which is a parsimonious model of Unfolding. In Unfolding type AMDS, each object is represented by two sets of coordinates to describe asymmetries in low-dimensions. Then, in the simultaneous methods based on Unfolding type AMDS, each cluster is also represented by two sets of coordinates in low-dimensions. These methods can be formulated as a special case of cluster difference scaling (CDS) (Heiser, 1993). Furthermore, because the models are formulated in the same way the constrained slide-vector model can be considered a generalization of the slide-vector model, constrained Unfolding, and Unfolding from the perspective of these feasible areas.

Second, we propose two types of simultaneous methods based on the hill-climbing model (Borg and Groenen, 2005) and the radius model (Okada and Imaizumi, 1987); these methods are the constrained hill-climbing model and the constrained radius model, respectively. These methods have the same properties for the decomposition of these objective functions into symmetric and skew-symmetric parts. From the decomposition, the symmetric

part of these objective functions becomes the objective function of CDS. Therefore, these objective functions can be decomposed into CDS and skew-symmetric parts, and can be considered a simultaneous method of CDS and AMDS for skew-symmetries. Then, just as with CDS, the symmetric and skew-symmetric parts of these objective functions can be further decomposed, respectively, by using the Sokal and Michener dissimilarity (Sokal and Michener, 1958): as such the meaning of these objective functions then becomes clear. Consequently, we show that these proposed methods inherit properties of CDS and AMDS.

Contents

1 Introduction **1**

2 Notations **6**

2.1 Symbols related to inputs of AMDS 6

2.2 Symbols related to outputs of AMDS 7

2.3 Symbols related to the objective function of AMDS 10

3 Asymmetric MDS and CDS **12**

3.1 Proposerties of asymmetric dissimilarity data 12

3.2 AMDS based on Unfolding 14

3.2.1 Model and objective function of Unfolding 14

3.2.2 Model and objective function of slide-vector model 17

3.3 Asymmetric MDS based on decomposistion 19

3.3.1 Model and objective function of hill-climbing model 20

3.3.2 Model and objective function of radius model 22

3.4 Cluster difference scaling (CDS) 25

3.4.1 Model and objective function of CDS 25

3.4.2 Properties of CDS 27

4 Constrained analysis of asymmetric data based on Unfolding **29**

4.1 Background of constrained asymmetric MDS based on Unfolding 29

4.2 Constrained Unfolding based on CDS 31

4.2.1 Model and objective function of the constrained Unfolding 31

4.2.2 Properties of the constrained Unfolding 33

4.2.3 Algorithm of the constrained Unfolding 34

4.3 Constrained slide-vector model based on CDS 39

4.3.1 Model and objective function of the constrained slide-vector model . 39

4.3.2 Properties of the constrained slide-vector model 41

4.3.3 Algorithm of the constrained slide-vector model 42

4.4 Relations between constrained methods and existing methods 46

5 Constrained analysis of asymmetric data based on decomposition **50**

5.1 Background of constrained asymmetric MDS based on decomposition 50

5.2 Constrained hill-climbing model based on CDS 51

5.2.1 Model and objective function of the constrained hill-climbing model 51
5.2.2 Properties of the constrained hill-climbing model 54
5.2.3 Algorithm of the constrained hill-climbing model 58
5.3 Constrained radius model based on CDS 60
5.3.1 Model and objective function of the constrained radius model 61
5.3.2 Properties of the constrained radius model 63
5.3.3 Algorithm of the constrained radius model 66
5.4 Relations between constrained method and existing method 69

6 Simulation studies **71**
6.1 Simulation of constrained AMDS based on Unfolding 71
6.1.1 Experimental design of the simulation 72
6.1.2 Simulation results 74
6.2 Simulation of the constrained hill climbing model 76
6.2.1 Experimental design of the simulation 79
6.2.2 Simulation results 80
6.3 Simulation of the constrained radius model 80
6.3.1 Experimental design of the simulation 82
6.3.2 Simulation results 83

7 Real examples **88**
7.1 Data description 88
7.2 Results of the constrained slide-vector model 91
7.3 Results of the constrained hill-climbing model 95
7.4 Results of the constrained radius model 95

8 Conclusions **100**

Chart Contents

1.1 Example of brand switching data for 15 different cola brands (Bell and Lattin, 1998; Borg and Groenen, 2005). 1
1.2 Social exchange of resources (Foa, 1971) 2

6.1 Factors of numerical simulation for AMDS based on Unfolding 73

7.1 Brands and types of Japanese tea bottles (abbreviations shown in parentheses) 89
7.2 switching data for Japanese tea bottles 90

Figure Contents

3.1 Example of the results of Unfolding.From the example, distance from object i to object j is interpreted as closer than distance from object j to object i. 15
3.2 Example of slide-vector model. In this example, distance from object i to object j is closer to that from object j to object i. 18
3.3 Example of hill climbing model. 22
3.4 Model of raidius model 23

5.1 Example of the constrained hill-climbing model 54

6.1 Clustering results for constrained hill climbing model and the tandem clustering 75
6.2 Clustering results for Factor 1 in the simulation 75
6.3 Clustering results for Factor 2 in the simulation 76
6.4 Clustering results for Factor 3 in $k = 2$ 77
6.5 Clustering results for Factor 3 in $k = 3$ 77
6.6 Clustering results for Factor 3 in $k = 5$ 78
6.7 Clustering results for Factor 4 78
6.8 Clustering results of the constrained hill climbing model 80
6.9 Clustering results of the constrained hill climbing model for Factor 1 . . . 81
6.10 Clustering results of the constrained hill climbing model for Factor 2 . . . 81
6.11 Clustering results of the constrained hill climbing model for Factor 4 . . . 82
6.12 Clustering results of the constrained radius model 84
6.13 Clustering results of the constrained radius model for Factor 1 84
6.14 Clustering results of the constrained radius model for Factor 2 85
6.15 Clustering results of the constrained radius model for Factor 3 for $k = 2$ . . 85
6.16 Clustering results of the constrained radius model for Factor 3 for $k = 3$ . . 86
6.17 Clustering results of the constrained radius model for Factor 3 for $k = 5$ . . 86
6.18 Clustering results of the constrained radius model for Factor 4 87

7.1 Values of the objective functions for the constrained slide-vector model . . 91
7.2 Results of the constrained slide-vector model 92
7.3 Results of the Unfolding model 93
7.4 Results of the slide-vector model 94
7.5 Values of the objective functions for the constrained hill-climbing model . . 95
7.6 Result of the constrained hill-climbing model 96

7.7 Result of the hill-climbing model 97
7.8 Values of the objective function for the constrained radius model 97
7.9 Result of the constrained radius model 98
7.10 Result of the radius model 99

Chapter 1

Introduction

Asymmetric dissimilarity data exist in various fields, such as marketing and psychology. Examples of such data are provided in Table. 1.1 (Bell and Lattin, 1998; Borg and Groenen, 2005) and Table. 1.2 (Foa, 1971). For example, The row and column in Table. 1.1 represent cola brands and, in this paper, they are called "objects". The element (c, d) in Table. 1.1 indicates the frequency of switching from brand c to brand d. Therefore, these frequencies can be interpreted as the similarities between these brands. In this case, frequencies of switching from brand c to brand d are not necessarily the same as those of switching from brand d to brand c. This relation is interpreted as an asymmetric relation. For details on the analysis of such brand switching data, one can refer to Okada and Tsurumi (2012).

Table 1.1: Example of brand switching data for 15 different cola brands (Bell and Lattin, 1998; Borg and Groenen, 2005).

								To							
From	a.	b.	c.	d.	e.	f.	g.	h.	i.	j.	k.	l.	m.	n.	o.
a.	41	11	2	8	0	2	15	8	14	0	9	11	0	6	2
b.	9	341	32	3	4	8	55	78	31	1	63	16	17	14	4
c.	3	27	160	15	8	2	18	15	32	2	31	13	2	12	7
d.	7	3	17	89	2	3	16	8	4	0	3	27	1	6	3
e.	1	7	6	2	119	6	20	8	19	0	16	15	2	21	7
f.	4	4	2	1	4	73	37	8	12	3	8	33	3	36	6
g.	14	53	16	16	22	38	675	98	56	10	48	187	33	172	20
h.	5	74	14	12	7	5	108	716	123	26	92	31	11	27	18
i.	14	35	36	3	15	11	56	120	422	20	86	82	29	38	10
j.	0	5	0	1	3	3	6	30	5	12	17	6	4	14	1
k.	13	70	29	6	12	5	49	87	92	19	471	40	11	34	8
l.	8	18	9	26	19	29	204	26	91	5	29	663	24	217	51
m.	2	14	4	3	1	2	35	13	22	1	20	19	364	23	1
n.	7	10	13	7	19	34	171	30	31	10	36	230	22	440	41
o.	3	3	7	3	10	9	26	22	11	2	4	48	2	35	215

Table. 1.2 is constructed through the data obtained from psychological experiments (Foa, 1971). The objects are the "resources" for social exchange and the elements are interpreted as asymmetric similarities between the "resources".

Table 1.2: Social exchange of resources (Foa, 1971)

	Taking					
Giving	Love	Status	Information	Money	Goods	Service
Love		65	10	0	2	23
Status	62		20	10	2	5
Information	17	34		11	24	14
Money	0	16	8		60	16
Goods	6	5	21	55		14
Service	41	18	7	16	18	

This implies that asymmetric dissimilarity is the dissimilarity in which the dissimilarity from object i to object j is not necessarily the same as the dissimilarity from object j to object i. When analyzing asymmetric dissimilarity data, it is important to understand the asymmetries between objects within the asymmetric data. For an example in Table. 1.1, if the asymmetries of brand switching data are analyzed, customer loyalties and competitive relation between these brands can be interpreted. In Table. 1.1, similarity from brand $k.$ to $a.$ is larger than that from brand $a.$ to $k.$. From this fact, we can interpret that customers who were purchasing brand $a.$ tend to switch to brand $k.$ and that there is are competitive relation between brands $a.$ and $k.$. One approach for analysis is through asymmetric multidimensional scaling (AMDS). In AMDS, asymmetric dissimilarity data is the input, and the coordinates of objects in low-dimensions with the relations of the asymmetries between such objects are provided as the output. This implies that the purpose of AMDS is to visualize the asymmetric relation among objects through the estimated coordinates of objects in low-dimensions. Several types of AMDS methods have been proposed (Borg and Groenen, 2005; Chino, 2012; Saito and Yadohisa, 2005). In particular, in Chino (2012), AMDS is divided into two methods, namely descriptive AMDS and inferential AMDS (e.g. Okada, 2011; Saburi and Chino, 2008) methods; these methods satisfy the narrower definition of AMDS described by (Chino, 2012). Descriptive AMDS can be further divided into three types, i.e., the augmented distance model, non-distance model, and extended distance model, which have been described by Chino and Okada (1996) and Chino (1997). In the augmented distance model, certain types of parameters for asymmetries are added to the ordinal distance model(Chino and Okada, 1996). The augmented distance model has been adopted by Borg and Groenen (2005), Gower (1977), Krumhansl (1978), Okada and Imaizumi (1984, 1987, 1997), Saito (1991), Saito and Takeda (1990), Tobler (1976), Weeks and Bentler (1982), Yadohisa and Niki (1999), Young

(1975), and Zielman and Heiser (1993). The non-metric distance model is a model in which the distance measure is not used; however indices such as the inner product are used. Finally, the extended distance model is a model measured in Minkowski space. In this study, we focus on the augmented AMDS model.

However, there are many parameters in AMDS because the asymmetric relations between objects are described as low-dimensional data. Therefore, in some situations, it becomes difficult to interpret these asymmetric relations between the objects even if the AMDS model is applied to such asymmetric dissimilarity data. Therefore, representatives of the asymmetries between objects are required for easily interpreting the asymmetries using a small number of parameters. To overcome this problem, cluster difference scaling (CDS) (Heiser, 1993; Heiser and Groenen, 1997; Kiers, et al., 2005; Vera et al. 2008), which is the simultaneous analysis of k-means and MDS, can be used as a useful tool for detecting features of symmetric dissimilarity data through clusters. In CDS, coordinates of the clusters and not of objects are estimated; however, CDS is unable to consider and describe the asymmetries between clusters.

In this study, we propose new methods for simultaneously performing AMDS and clustering. Our proposed methods can represent the asymmetries between clusters; however, because several types of AMDS models have been proposed, there exist many representations of asymmetries between clusters. Therefore, we propose four types of methods for simultaneously performing AMDS and clustering; these four approaches can divided into two groups, namely, methods that are based on unfolding and methods that are based on the decomposition of symmetric and skew-symmetric parts. The four proposed methods adopt the approaches of CDS, primarily the estimation approach and the decomposition of the objective functions (Nocedal and Wright, 1999), to interpret the features within and between clusters.

For the first group, i.e., methods based on unfolding, each object is represented in the form of two coordinates. Concretely, constrained unfolding and constrained slide-vector models are proposed as a part of the first group, which are based on unfolding (Gower, 1977) and the slide-vector model (De Leeuw and Heiser, 1982;Zieltman and Heiser, 1993), respectively. Constrained unfolding is a method for simultaneous unfolding and clustering of asymmetric dissimilarity data. Each cluster is represented by two coordinates that represent the asymmetries, which can be used to interpret the asymmetries between clusters. However, because the number of clusters is large, it becomes difficult to understand the asymmetric relationship. Given this problem of large number of objects, we also propose a constrained slide-vector model, which is a parsimonious constrained unfolding model. The constrained slide-vector model can represent the asymmetries between clusters using a smaller number of parameters than those of the constrained unfolding model. Furthermore, in some situations, the constrained slide-vector model can be used to perform both unfolding and constrained unfolding. Therefore, the constrained slide-vector model

can be considered as a generalized unfolding model.

For the second group, i.e., methods based on the decomposition of symmetric and skew-symmetric parts, the objective functions can be decomposed into symmetric and skew-symmetric parts. Concretely, a constrained hill-climbing model and a constrained radius model are proposed as a part of the second group, which are based on the hill-climbing model (Borg and Groenen, 2005) and radius model (Okada and Imaizumi, 1987), respectively. For these models, the skew-symmetric part can be further divided into parts that are within and between the clusters using the method described by Vicari(2014). Therefore, a detailed interpretation of the objective function can be performed. Here, the constrained hill-climbing model involves simultaneous application of the hill-climbing model and clustering of the asymmetric dissimilarity data. The parameters of the hill-climbing model are represented in the same manner as those of the slide-vector model; however, the interpretation is different. Concretely, the skew-symmetric part of the asymmetric dissimilarity data is represented by one vector, called the slope-vector. In the hill-climbing model, it is easy to interpret the asymmetric part because the objective function of the hill-climbing model can be decomposed into symmetric and skew-symmetric parts. Hence, the interpretation easier in this model than in the slide-vector model due to the skew-symmetric part. Therefore, the constrained hill-climbing model can represent the asymmetries between clusters while including the advantages of the hill-climbing model. Similarly, the constrained radius model is the simultaneous implementation of the radius model and clustering for asymmetric dissimilarity data based on Tanioka and Yadohisa (2016). The parameters of the radius model are very simple, and the objective function can be decomposed into symmetric and skew-symmetric parts in the same manner as that in the hill-climbing model. In particular, data corresponding to the symmetric parts can be represented as the coordinates of the objects and the skew-symmetric parts can be represented by the radii of the objects. In the constrained radius model, the coordinates and radii of clusters, and not those of objects, are estimated to represent the asymmetries between the clusters.

The reminder of this paper is organized as follows. In Chapter 2, some notations used in this paper are defined. In Chapter 3, the definitions for AMDS are introduced, including the decomposition of asymmetric dissimilarity data as preliminaries. Furthermore, unfolding, the slide-vector model, the hill-climbing model, and the radius model are described as a part of previous studies on AMDS. Finally, CDS, which is key in this paper, is described. In Chapter 4, the models and objective functions of constrained unfolding and constrained slide-vector models are described. Next, these algorithms based on the majorizing function are described. Furthermore, we show the relationships between the constrained slide-vector model and unfolding, the constrained slide-vector model and constrained unfolding, and the constrained slide-vector model and slide-vector model. In Chapter 5, AMDS

methods, based on the decomposition of symmetric and skew-symmetric parts, are described. We also describe the models and objective functions of the constrained hill-climbing and constrained radius models. Next, as properties of these methods, we prove several ways to decompose these methods. We also show the relationships between the hill-climbing and constrained hill-climbing models, as well as the radius and constrained radius models. In Chapter 6, we provide the results of our proposed methods through numerical simulations and the effectiveness of the proposed methods. In Chapter 7, a real example of constrained unfolding, the constrained slide-vector model, the constrained hill-climbing model, and the constrained radius model are shown. In Chapter 8, conclusion and remarks are described.

Chapter 2

Notations

In this chpater, mathematical symbols are shown before describing the contents.

First, asymmetric dissimilarity data, symmetric part of asymmetric dissimilarity data and skew-symmetric dissimilarity data are defined as inputs of MDS. Second, as outputs of AMDS, coordinates of objects, clusters of objects, coordinates of clusters, slide-vectors, a slope vector, length of radii, clusters of these centroids are shown. Third, symbols related to objective function of AMDS models are defined such as distance matrieces, weights matrices and zero matrices.

2.1 Symbols related to inputs of AMDS

In this section, asymmetric dissimilarity data, symmetric dissimilarity data and symmetric parts and skew-symmetric parts of the asymmetric dissimilarity data are defined.

First, asymmetric dissimilarity data is defined.

Definition 2.1.1 *Asymmetric dissimilarity data*

Let $\mathcal{I} = \{1, 2, \cdots, n\}$ *and* $\mathcal{J} = \{1, 2, \cdots, n\}$ *be an index sets of objects, respectively, where* $n \in \mathbb{N}$ *is the number of objects. The function* Δ *is defined as follows:*

$$\Delta : \mathcal{I} \times \mathcal{J} \mapsto {\mathbb{R}_{\dagger}}^{n \times n}$$

where $\times$ *indicates Cartesian product and* $\mathbb{R}_{\dagger}^{n \times n}$ *is a set of* n *by* n *non-negative matrices but these diagonal elements are* 0*. Here, elements of* $\mathcal{I}$ *and* $\mathcal{J}$ *are called as row objects and column objects, respectively although* $\mathcal{I} = \mathcal{J}$ *is assumed in the definition of the asymmetric dissimilarity data. The range of* Δ *is described as follows:*

$$\boldsymbol{\Delta} = \Delta(\mathcal{I} \times \mathcal{J}).$$

Then, the asymmetric dissimilarity data is defined as follows:

$$\boldsymbol{\Delta} = (\delta_{ij}), \quad \delta_{ij} \in \mathbb{R}_{+} \quad (i \neq j;\ i, j = 1, 2, \cdots, n)$$

where $\neg(\forall i, j = 1, 2, \cdots, n)(\delta_{ij} = \delta_{ji})$, $(\forall i = 1, 2, \cdots, n)(\delta_{ii} = 0)$, n *is the number of objects and* $\mathbb{R}_+$ *is a set of non-negative real numbers. If* $\delta_{ij} > \delta_{st}$ *for some objects* i, j, s *and* t, *the dissimilarity from object* i *to* j *is considered as larger than the dissimilarity from object* s *to* t.

Next, symmetric dissimilarity data is defined.

Definition 2.1.2 *Symmetric dissimilarity data*

Symmetric dissimilarity data is defined as follows:

$$\boldsymbol{\Xi} = (\xi_{ij}), \quad \xi_{ij} \in \mathbb{R}_+ \quad (i \neq j; i, j = 1, 2, \cdots, n)$$

where $\boldsymbol{\Xi} = \boldsymbol{\Xi}^T$. *If* $\xi_{ij} > \xi_{st}$ *for some objects* i, j, s *and* t, *the dissimilarity between object* i *and* j *is considered as larger than the dissimilarity between object* s *and* t.

Next, symmetric part and skew-symmetric part of asymmetric dissimilarity data are defined.

Definition 2.1.3 *Symmetric part and skew-symmetric part of asymmetric dissimilarity data*

Given any asymmetric dissimilarity data $\boldsymbol{\Delta}$, *symmetric part of* $\boldsymbol{\Delta}$, $\boldsymbol{S}$ *and skew-symmetric part of* $\boldsymbol{\Delta}$, $\boldsymbol{A}$ *are defined as follows:*

$$\boldsymbol{S} = \frac{1}{2}(\boldsymbol{\Delta} + \boldsymbol{\Delta}^T) \quad \text{and} \quad \boldsymbol{A} = \frac{1}{2}(\boldsymbol{\Delta} - \boldsymbol{\Delta}^T),$$

where $\cdot^T$ *indicate transposition of matrix.*

2.2 Symbols related to outputs of AMDS

In this section, symbols used as outputs of AMDS are defined. First, coordinates of row objects and column objects are defined. Second, cluster of row objects and column objects are shown and indicator matrices of objects are defined. In addition, coordinates of cluster centorids of row objects and column objects are defined. Finally, parameters for asymmetries, cluster of centroids of clusters for column objects and indicator matrix for these centroids are shown.

First, cooridnates of objects are defined.

Definition 2.2.1 *Coordinates of row objects and column objects*

Given sets of index for objects $\mathcal{I}$ *and* $\mathcal{J}$, *coordinates of row objects* $i \in \mathcal{I}$ *are defined as follows:*

$$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)^T = (x_{it}) \quad x_{it} \in \mathbb{R} \quad (i = 1, 2, \cdots, n; t = 1, 2, \cdots, d)$$

and coordinates of column objects $j \in \mathcal{J}$ *are defined as follows:*

$$\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n)^T = (y_{jt}) \quad y_{jt} \in \mathbb{R} \quad (j = 1, 2, \cdots, n; t = 1, 2, \cdots, d)$$

where d *is the number of dimensions.*

$\boldsymbol{X}$ and $\boldsymbol{Y}$ are estimated by some kinds of AMDS to visualize the relation between objects. In many cases, d is set as 2 or 3 to visualize these relations.

Next, clusters of objects are defined.

Definition 2.2.2 *Clusters of objects*

Given a set of objects $\mathcal{I}$, clusters of these objects $C_o \subset \mathcal{I}$ $(o = 1, 2, \cdots, k)$ are defined as so that it satisfies

$$(\forall C_o, C_\ell \subset \mathcal{I})(o \neq \ell)(C_o \cap C_\ell = \phi) \tag{2.1}$$

$$\bigcup_{o=1}^{k} C_o = \mathcal{I}. \tag{2.2}$$

where k ($\leq n$) is the number of clusters.

Next, indicator matrix of objects is defined.

Definition 2.2.3 *Indicator matrix of clusters for objects*

Given clusters of objects C_o $(o = 1, 2, \cdots, k)$, indicator matrix of clusters for objects is defined as follows:

$$\boldsymbol{U} = (u_{io}) \quad u_{io} = \begin{cases} 1 & (i \in C_o) \\ 0 & (\text{others}) \end{cases} \qquad (i = 1, 2, \cdots, n; o = 1, 2, \cdots, k).$$

Next, coordinates of object clusters are defined.

Definition 2.2.4 *Coordinates of clusters for row objects and column objects*

$\mathcal{I} = \mathcal{J}$ is assumed. Coordinates of clusters for row objects are defined as follows;

$$\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \cdots, \boldsymbol{x}_k^*)^T = (x_{ot}^*) \quad x_{ot}^* \in \mathbb{R} \quad (o = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$$

and coordinates of clusters of column objects are defined as follows:

$$\boldsymbol{Y}^* = (\boldsymbol{y}_1^*, \boldsymbol{y}_2^*, \cdots, \boldsymbol{y}_k^*)^T = (y_{ot}^*) \quad y_{ot}^* \in \mathbb{R} \quad (o = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$$

where d is the number of dimensions.

$\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$ are used in simultaneous methods of clustering and AMDS based on Unfolding.

Next, three kinds of asymmetric parameters such as slide-vector (De Leeuw and Heiser, 1982; Zieltman and Heiser, 1993), slope-vector (Borg and Groenen, 2005) and radii (Okada and Imaizumi, 1987) are shown, although the ways of interpreting these parameters are not described here.

Definition 2.2.5 *Notations of slide-vectors*

Notations of slide-vectors are defined as follows:

$$\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_m)^T = (z_{st}) \quad z_{st} \in \mathbb{R} \quad z_{st} \in \mathbb{R}\ (s = 1, 2, \cdots, m; t = 1, 2, \cdots, d).$$

where m ($\leq k \leq n$) is the number of slide-vectors.

Originally, slide-vector model is defined by $m = 1$. However, in this paper, the concept of slide-vectos is extended to those of $m \geq 1$. In the slide-vector model, $\boldsymbol{Y}$ is described by using $\boldsymbol{X}$ and $\boldsymbol{Z}$ and asymmetries are described based on $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$.

Next, the slope vector is defined.

Definition 2.2.6 *Notations of slope-vector*

Notations of the slope vector is defined as follows:

$$\boldsymbol{v} = (v_t) \quad v_t \in \mathbb{R} \quad v_t \in \mathbb{R} \ (t = 1, 2, \cdots, d).$$

The slope vector is used in hill-climbing model. In hill-climbing model, the estimated slope vector represents skew-symmetric part of asymmetric dissimilarity data based on non-distance model (Chino, 2012).

Finally, radii for objects and clusters are defined.

Definition 2.2.7 *Notations of radii for these centorids*

Notations of radii for objects are defined as follows:

$$\boldsymbol{r} = (r_1.r_2, \cdots, r_n) \quad r_i \geq 0 \quad (i = 1, 2, \cdots, n),$$

and notation of radii for clusters are defined as follows:

$$\boldsymbol{r}^* = (r_1^*, r_2^*, \cdots, r_m^*) \quad r_f \geq 0 \quad (f = 1, 2, \cdots, m)$$

where m ($\leq k \leq n$) is the number of clusters.

As the same way of hill-climbing model, in the radius model, difference between radii represents skew-symmetric part of asymmetric dissimilarity data.

Next, clusters of centroids are defined. When skew-symmetric matrix of asymmetric dissimilarity data

Definition 2.2.8 *clusters of centroids for object cluster*

Given a set of clusters for objects $\mathcal{K} = \{1, , 2, \cdots, k\}$, clusters of these centorids $C_f^* \subset \mathcal{K}$ $(f = 1, 2, \cdots, m(\leq k))$ are defined as satisfying both Eq. (2.3) and Eq. (2.4).

$$(\forall C_f^*, C_q^* \subset \mathcal{K})(f \neq q)(C_f^* \cap C_q^* = \phi) \tag{2.3}$$

$$\bigcup_{f=1}^{m} C_f^* = \mathcal{K} \tag{2.4}$$

where m is the number of clusters for centroids.

These clusters of centroids are used to represent skew-symmetries by the small number of parameters in the proposed method.

Next, indicator matrix for centroids is defined.

Definition 2.2.9 *Indicator matrix of centroids*

Given clusters of centroids C_f^* $(f = 1, 2, \cdots, m)$*, indicator matrix of centroids is defined as follows:*

$$\boldsymbol{\Psi} = (\psi_{sf}) \quad \psi_{sf} = \begin{cases} 1 & (s \in C_f^*) \\ 0 & (\text{others}) \end{cases} \qquad (s = 1, 2, \cdots, k; f = 1, 2, \cdots, m).$$

2.3 Symbols related to the objective function of AMDS

In this section, distance matrices, weights matrices, zero matrices and Hadmard product are defined.

First, distance matrices are defined.

Definition 2.3.1 *Distance matrix*

Given coordinates of objects $\boldsymbol{X}$*, the distance matrix for* $\boldsymbol{X}$ *is defined as follows:*

$$\boldsymbol{D}(\boldsymbol{X}) = (d_{ij}(\boldsymbol{X})) \quad d_{ij}(\boldsymbol{X}) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \quad (i, j = 1, 2, \cdots, n).$$

where $\|\boldsymbol{x}\| = \sqrt{\sum_{t=1}^{d} x_t^2}$ *is defined as norm for arbitrary vector* $\boldsymbol{x} = (x_t) \quad x_t \in \mathbb{R}$ $(t = 1, 2, \cdots, d)$*.*

As the same way of this, given two coordinates of objects $\boldsymbol{X}$ *and* $\boldsymbol{Y}$*, distance matrix for* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *is defined as follows:*

$$\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y}) = (d_{ij}(\boldsymbol{X}, \boldsymbol{Y})) \quad d_{ij}(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{x}_i - \boldsymbol{y}_j\| \quad (i, j = 1, 2, \cdots, n).$$

These distance matrices are used when dissimilarity matrix is approximated by coordinates of objects.

Next, weight matrix for dissimilarities is defined. The weight matrix is used in the objective function of Unfolding to describe the objective function based on the unified framework as the same way of oridinal MDS.

Definition 2.3.2 *Weight matrix for dissimilarities*

Given dissimilarity matrix $\boldsymbol{\Delta} = (\delta_{ij}) \quad (i, j = 1, 2, \cdots, n)$*, weight matrix for dissimilarities is defined as follows:*

$$\boldsymbol{W}^{\dagger} = \begin{bmatrix} \boldsymbol{O}_n & \boldsymbol{W} \\ \boldsymbol{W}^T & \boldsymbol{O}_n \end{bmatrix} = (w_{i^{\dagger}j^{\dagger}}) \quad w_{i^{\dagger}j^{\dagger}} \in \{0, 1\} \ (i^{\dagger}, j^{\dagger} = 1, 2, \cdots, 2n)$$

where $\boldsymbol{W} = \mathbf{1}_n \mathbf{1}_n^T$*,* $\mathbf{1}_n = (1, 1, \cdots, 1)$ *is a vector with a length of* n *and* $\boldsymbol{O}_n = (0)$ *is* n *by* n *matrix.*

The definition of weight matrix for dissimilarities is used for AMDS based on Unfolding when the kind of AMDS is formulated based on unified framework.

Finally, Hadmard product is defined.

Definition 2.3.3 *Hadmard product*

Given n by p matrices, $\boldsymbol{B} = (b_{it})$ and $\boldsymbol{C} = (c_{it})$ $(i = 1, 2, \cdots, n; t = 1, 2, \cdots, p)$, Hadmard product is defined as follows:

$$\boldsymbol{B} \odot \boldsymbol{C} = (b_{it}c_{it}) \quad (i = 1, 2, \cdots, n; t = 1, 2, \cdots, p).$$

Chapter 3

Asymmetric MDS and CDS

In this chapter, we provide definitions and properties related to AMDS(Borg and Groenen, 2005;Chino, 2012;Saito and Yadohisa, 2005) and CDS (Heiser, 1993; Heiser and Groenen, 1997), as preliminaries for introducing our proposed methods. First, we introduce decompositions of asymmetric dissimilarity data, and show relations betweem Frobenius norm and an asymmetric dissimilarity data (e.g. Borg and Groenen, 2005). Second, we present previous studies of AMDS based on Unfolding (Gower, 1977), for example, applying Unfolding to asymmetric dissimilarity data and the slide-vector model (De Leeuw and Heiser, 1982; Zielman and Heiser, 1993). Third, we describe previous studies of AMDS based on decomposing symmetric and skew-symmetric parts, for example, the hill-climbing (Borg and Groenen, 2005) model and radius models (Okada and Imaizumi, 1987;1997). Finally, CDS (Heiser, 1993;Heiser and Groenen, 1996), which is a simultaneous analysis of MDS and clustering, is shown.

3.1 Proposerties of asymmetric dissimilarity data

In this section, we show a property for decomposition of asymmetric dissimilarity data. Furthermore, we show the relation between this decomposition and Frobenus norm. The relation is related to the decomposition of the objective function for symmetric and skew-symmetric parts.

First, we show the decomposition of asymmetric dissimilarity data; however, the proposition is shown for a square matrix because the property is satisfied for arbitrarily square matrices that include asymmetric dissimilarity data.

Proposition 3.1.1 *Decomposistion of square matrix*

Let $\boldsymbol{P} = (p_{ij}) \quad p_{ij} \in \mathbb{R}\ (i, j = 1, 2, \cdots, n)$ *be a square matrix. Then, the following decomposition exists for any* $\boldsymbol{P}$*,*

$$\boldsymbol{P} = \boldsymbol{S} + \boldsymbol{A} \tag{3.1}$$

where

$$\boldsymbol{S} = (\boldsymbol{P} + \boldsymbol{P}^T)/2, \text{ and}$$
$$\boldsymbol{A} = (\boldsymbol{P} - \boldsymbol{P}^T)/2.$$

Here, $\boldsymbol{S}$ and $\boldsymbol{A}$ are symmetric matrix and skew-symmetric matrix, respectively. where skew-symmetric matrix is $\boldsymbol{A} = (a_{ij})$ $a_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ satisfying

$$(\forall i, j = 1, 2, \cdots, n)(a_{ij} = -a_{ji})$$

Proof. *From the right side of Eq. (3.1),*

$$\boldsymbol{S} + \boldsymbol{A} = (\boldsymbol{P} + \boldsymbol{P}^T)/2 + (\boldsymbol{P} - \boldsymbol{P}^T)/2 = \boldsymbol{P}/2 + \boldsymbol{P}^T/2 + \boldsymbol{P}/2 - \boldsymbol{P}^T/2 = \boldsymbol{P}$$

and $a_{ij} = (p_{ij} - p_{ji})/2 = -(p_{ji} - p_{ij})/2 = -a_{ji}$.

In this paper, each diagonal element of $\boldsymbol{\Delta}$ is assumed to be 0. Therefore, diagonal elements of $\boldsymbol{S}$ derived from decomposition of $\boldsymbol{\Delta}$ become 0.

Next, a relation between the decomposition of symmetric and skew-symmetric matrix, and Frobenius norm is shown.

Proposition 3.1.2 *Decomposition of square matrix based on least squares*

Let $\boldsymbol{P}$ be a square matrix, $\boldsymbol{S}$, and $\boldsymbol{A}$ be symmetric and skew-symmetric parts of $\boldsymbol{P}$, respectively. In the situation, the following property is satisfied:

$$\|\boldsymbol{P}\|^2 = \|\boldsymbol{S}\|^2 + \|\boldsymbol{A}\|^2$$

where $\|\boldsymbol{B}\| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}^2}$ is Frobenius norm of arbitrary matrix $\boldsymbol{B} = (b_{ij})$ $b_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$.

Proof.

$$\|\boldsymbol{P}\|^2 = \|(\boldsymbol{P} + \boldsymbol{P}^T)/2 + (\boldsymbol{P} - \boldsymbol{P}^T)/2\|^2 = \|\boldsymbol{S} + \boldsymbol{A}\|^2$$

From the property of $\|\cdot\|^2$,

$$\begin{aligned}\|\boldsymbol{S} + \boldsymbol{A}\|^2 &= tr(\boldsymbol{S} + \boldsymbol{A})^T(\boldsymbol{S} + \boldsymbol{A}) \\ &= tr(\boldsymbol{S}^T\boldsymbol{S} + 2\boldsymbol{S}^T\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{A}) \\ &= \|\boldsymbol{S}\|^2 + 2tr(\boldsymbol{S}^T\boldsymbol{A}) + \|\boldsymbol{A}\|^2. \end{aligned} \tag{3.2}$$

Here, following equations are satisfied since $\boldsymbol{S}$ is symmetric matrix.

$$tr(\boldsymbol{S}^T\boldsymbol{A}) = tr(\boldsymbol{S}\boldsymbol{A}) = tr(\boldsymbol{A}\boldsymbol{S})$$

$$\begin{aligned}&From\ the\ property\ of\ the\ skew-symmetricmatrix \\ &= tr(\boldsymbol{S}^T\boldsymbol{A}^T) = tr(\boldsymbol{S}\boldsymbol{A}^T) = tr(-\boldsymbol{S}\boldsymbol{A}) = 0 \end{aligned} \tag{3.3}$$

From Eq (3.2) and Eq (3.3), we obtain the proposition.

Proposition 3.1.2 is used when objective functions of AMDS are decomposed into symmetric and skew-symmetric parts (e.g. Borg and Groenen, 2005; Saito and Yadohisa, 2005). Especially, it is important for interpreting the objective functions of the hill-climbing and radius models.

Next, property of skew-symmetric matrix is shown.

Proposition 3.1.3 *Property of skew-symmetric matrices*

Given arbitrarily two skew-symmetric matrices $\boldsymbol{B} = (b_{ij})$ $b_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *and* $\boldsymbol{C} = (c_{ij})$ $c_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$, $\boldsymbol{B} - \boldsymbol{C}$ *becomes skew-symmetric matrix.*

Proof. *From the definition of* $\boldsymbol{B}$ *and* $\boldsymbol{C}$, $\boldsymbol{B} = -\boldsymbol{B}^T$ *and* $\boldsymbol{C} = -\boldsymbol{C}^T$ *are satisfied. Then, following property is satisfied.*

$$\boldsymbol{B} - \boldsymbol{C} = -\boldsymbol{B}^T - (-\boldsymbol{C}^T) = -(\boldsymbol{B}^T - \boldsymbol{C}^T) = -(\boldsymbol{B} - \boldsymbol{C})^T$$

3.2 AMDS based on Unfolding

In this section, Unfolding (Coombs, 1950;Gower, 1977) and slide-vector model (De Leeuw and Heiser, 1982; Zilman and Heiser, 1993) are introduced. First, the model and objective function of Unfolding is defined as a special case of ordinary MDS (Borg and Groenen, 2005). In Unfolding, coordinates of objects are estimated based on dissimilarities between two groups, and dissimilarities within each group can not be considered. Therefore, Unfolding is considered as ordinary MDS for dissimilarities including missing corresponding to those within groups. Using the same approach, the model and objective function of the slide-vector model can be formulated on a basis of a special case of MDS.

3.2.1 Model and objective function of Unfolding

In this subsection, Unfolding was proposed for not only asymmetric dissimilarity data, however, we define them through asymmetric dissimilarity data in this paper.

Definition 3.2.1 *Unfolding model based on elements description*

Let $\boldsymbol{\Delta}$ *be asymmetric dissimilarity data,* $\boldsymbol{X} = (x_{it}) \quad x_{it} \in \mathbb{R}$ $(i = 1, 2, \cdots, n;\ t = 1, 2, \cdots, d)$ *and* $\boldsymbol{Y} = (y_{jt}) \quad y_{jt} \in \mathbb{R}$ $(j = 1, 2, \cdots, m; t = 1, 2, \cdots, d)$ *be coordinates of row-objects and column-objects in d dimensions, respectively. Here, the model of Unfolding is defined as following equation:*

$$\delta_{ij} = d_{ij}(\boldsymbol{X}, \boldsymbol{Y}) + \varepsilon_{ij}, \quad (i, j = 1, 2, \cdots, n)$$

where

$$d_{ij}(\boldsymbol{X}, \boldsymbol{Y}) = \Big[\sum_{t=1}^{d} (x_{it} - y_{jt})^2 \Big]^{\frac{1}{2}}, \text{ and}$$

ε_{ij} $(i, j = 1, 2, \cdots, n)$ *is a error term.*

For the determination of the number of low dimensions, two or three are set because the purpose of AMDS is visualize asymmetric relation between objects.

From the Unfolding model, we explain how to interpret the results when Unfolding is applied to asymmetric dissimilarity data. First, there exists the same object i with coordinates of a row object and a column object; i.e., in short, each object is represented as two coordinates. Therefore, when dissimilarity from object i to

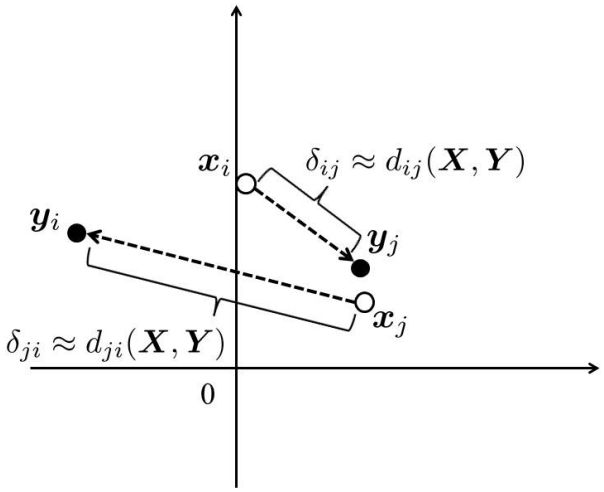

Figure 3.1: Example of the results of Unfolding.From the example, distance from object i to object j is interpreted as closer than distance from object j to object i.

object j is interpreted, we interpret the distance from the coordinate of row object i to the coordinate of column object j. From this interpretation, we apply Unfolding to asymmetric dissimilarity data to identify asymmetries between objects because there are different distances from coordinate of row object i to coordinate of column object j and from the coordinate of row object j to the coordinate of column object i.

Figure 3.1 represents example for model of Unfolding. The distance from object i to j indicates distance from $\boldsymbol{x}_i$ to $\boldsymbol{y}_j$. On the other hand, the distance from object j to i indicates distance from $\boldsymbol{x}_j$ to $\boldsymbol{y}_i$.

Next, we define the objective function of Unfolding I.

Definition 3.2.2 *Objective function of Unfolding I*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$ *and the number of low-dimensions* d*, the objective function of Unfolding I is defined as follows:*

$$L(\boldsymbol{X}, \boldsymbol{Y} |\, \boldsymbol{\Delta}) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ij} - d_{ij}(\boldsymbol{X}, \boldsymbol{Y}))^2.$$

In the objective function, $\boldsymbol{X}$ and $\boldsymbol{Y}$ such that the objective function is minimized are estimated.

Next, we define the objective function of Unfolding II; futhermore, we show the equivalence of objective functions for both Unfolding I and II.

Definition 3.2.3 *Objective function of Unfolding II*

Given $\boldsymbol{\Delta}$ and the number of dimensions d, the objective function of Unfolding II is defined as follows:

$$L(\boldsymbol{X}, \boldsymbol{Y} |\ \boldsymbol{\Delta}^\dagger) = \frac{1}{2}\|\boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}))\|^2$$

where

$$\boldsymbol{\Delta}^\dagger = \begin{bmatrix} \boldsymbol{O}_n & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}^T & \boldsymbol{O}_n \end{bmatrix},$$

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = (q_{i^\dagger t}) \quad (i^\dagger = 1, 2, \cdots, 2n; t = 1, 2, \cdots, d),$$

$$\boldsymbol{D}(\boldsymbol{Q}) = (d_{i^\dagger j^\dagger}(\boldsymbol{Q})) \quad d_{i^\dagger j^\dagger}(\boldsymbol{Q}) = \Big[\sum_{t=1}^{d}(q_{i^\dagger t} - q_{j^\dagger t})^2\Big]^{\frac{1}{2}} \ (i^\dagger, j^\dagger = 1, 2, \cdots, 2n),$$

$$\boldsymbol{W}^\dagger = \begin{bmatrix} \boldsymbol{O}_n & \boldsymbol{W} \\ \boldsymbol{W}^T & \boldsymbol{O}_n \end{bmatrix} = (w^\dagger_{i^\dagger j^\dagger}) \ (i^\dagger, j^\dagger = 1, 2, \cdots, 2n)$$

$w^\dagger_{i^\dagger j^\dagger}$ is weights for pair of $i^\dagger$ and $j^\dagger$ and $\odot$ is Hadmard product.

Next, we show the equivalence of the objective functions of Unfolding I and II.

Proposition 3.2.1 *Equivalence of objective functions of Unfolding I and II*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$, coordinates of row objects $\boldsymbol{X}$ and column objects $\boldsymbol{Y}$, and the number of low-dimensions d, the following equation is satisfied.

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(\delta_{ij} - d_{ij}(\boldsymbol{X}, \boldsymbol{Y}))^2 = \frac{1}{2}\|\boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}))\|^2 \tag{3.4}$$

Proof. *From the right hand of Eq. (3.4),*

$$\frac{1}{2}\|\boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}))\|^2 = \frac{1}{2}\sum_{i^\dagger=1}^{2n}\sum_{j^\dagger=1}^{2n} w^\dagger_{i^\dagger j^\dagger}(\delta^\dagger_{i^\dagger j^\dagger} - d_{i^\dagger j^\dagger}(\boldsymbol{Q}))^2. \tag{3.5}$$

from the definition of $\boldsymbol{\Delta}^\dagger$ *and* $\boldsymbol{W}^\dagger$*, Eq. (3.5) is described as follows:*

$$
\begin{aligned}
&\frac{1}{2}\sum_{i^\dagger=1}^{2n}\sum_{j^\dagger=1}^{2n} w^\dagger_{i^\dagger j^\dagger}(\delta^\dagger_{i^\dagger j^\dagger} - d_{i^\dagger j^\dagger}(\boldsymbol{Q}))^2 \\
=&\frac{1}{2}\sum_{i^\dagger=1}^{n}\sum_{j^\dagger=n+1}^{2n} w^\dagger_{i^\dagger j^\dagger}(\delta^\dagger_{i^\dagger j^\dagger} - d_{i^\dagger j^\dagger}(\boldsymbol{Q}))^2 + \frac{1}{2}\sum_{i^\dagger=n+1}^{2n}\sum_{j^\dagger=1}^{n} w^\dagger_{i^\dagger j^\dagger}(\delta^\dagger_{i^\dagger j^\dagger} - d_{i^\dagger j^\dagger}(\boldsymbol{Q}))^2 \\
=&\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\delta_{ij} - d_{ij}(\boldsymbol{X},\boldsymbol{Y}))^2 + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n}(\delta_{ij} - d_{ij}(\boldsymbol{X},\boldsymbol{Y}))^2 \\
=&\sum_{i=1}^{n}\sum_{j=1}^{n}(\delta_{ij} - d_{ij}(\boldsymbol{X},\boldsymbol{Y}))^2
\end{aligned}
$$

From Proposition 3.2.1, Unfolding is considered as a special case of ordinal symmetirc MDS. In short, if both dissimilarities between row-objects and between column objects are missing, the MDS model becomes Unfolding.

3.2.2 Model and objective function of slide-vector model

In this subsection, the slide-vector model (De Leeuw and Heiser, 1982; Zilman and Heiser, 1993) is described. This model is considered a parsimonious model of Unfolding; as such, the model is assumed only for situations in which a square matrix is applied, particularly for asymmetirc dissimilarity data.

We define the slide-vector model.

Definition 3.2.4 *Model of the slide-vector model*

Let $\boldsymbol{\Delta}$ *be an asymmetric dissimilarity matrix,* $\boldsymbol{X} = (x_{it}), x_{it} \in \mathbb{R}$ *and* $\boldsymbol{z} = (z_t), z_t \in \mathbb{R}$ $(t = 1, 2, \cdots, d)$ *be cooridnates of objects and slide-vector in* d *dimensions, respectively. The model of slide-vecotr model is defined as following equation:*

$$\delta_{ij} = d_{ij}(\boldsymbol{X}, \boldsymbol{X} - \mathbf{1}_n\boldsymbol{z}^T) + \varepsilon_{ij} \quad (i, j = 1, 2, \cdots, n)$$

where

$$d_{ij}(\boldsymbol{X}, \boldsymbol{X} - \mathbf{1}_n\boldsymbol{z}^T) = \Big[\sum_{t=1}^{d}(x_{it} + z_t - x_{jt})^2\Big]^{\frac{1}{2}},$$

and $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error term.*

Next, we explain how we interpret the results of the slide-vector model. In the model, when dissimilarity from object i to object j is interpreted, we interpret the distance from $\boldsymbol{x}_i$ to $\boldsymbol{x}_i - \boldsymbol{z}$, where $\boldsymbol{x}_i$ is row vector of $\boldsymbol{X}$. In short, $\boldsymbol{y}_i$ is modelized such that $\boldsymbol{y}_i = \boldsymbol{x}_i - \boldsymbol{z}$. Therefore, the slide-vector model is a special case of Unfolding. While it is difficult to interpret the distance from object i to object j in Unfolding, it is easy to interpret the distance from object i to object j in the slide-vector model

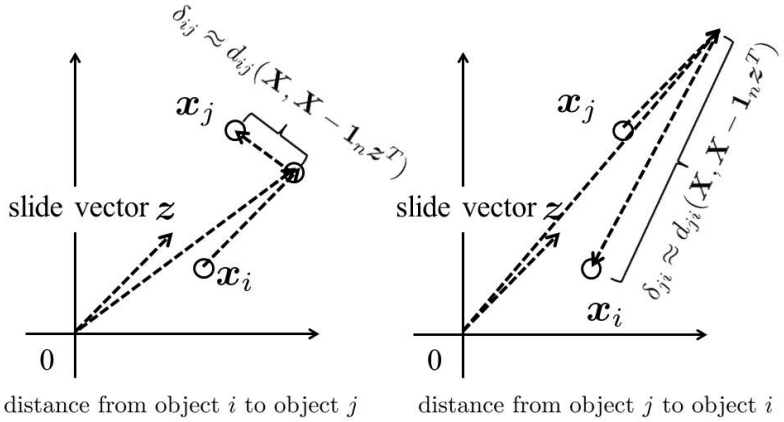

Figure 3.2: Example of slide-vector model. In this example, distance from object i to object j is closer to that from object j to object i.

because $\boldsymbol{z}$ is only considered when asymmetries are interpreted in the slide-vector model although $\boldsymbol{y}_i$ $(i = 1, 2, \cdots, n)$ have to be considered in Unfolding.

Figure 3.2 indicates example of slide-vector model. In this example, distance object i to j is closer than distance object j to i from the direction of slide- vector $\boldsymbol{z}$.

Next, we define the objective function of slide-vector model I.

Definition 3.2.5 *Objective function of slide-vector model*

Given an asymmetric dissimilarity data $\boldsymbol{\Delta}$ *and the number of low-dimensions* d*, the objective function of the slide-vector model is defined as follows:*

$$L(\boldsymbol{X}, \boldsymbol{z} |\ \boldsymbol{\Delta}) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ij} - d_{ij}(\boldsymbol{X}, \boldsymbol{X} - \mathbf{1}_n \boldsymbol{z}^T))^2.$$

In the slide-vector model, $\boldsymbol{X}$ *and* $\boldsymbol{z}$ *such as minimizing the objective function is estimated.*

To clearly show the relationships between the slide-vector model and Unfolding, we next define the objective function of slide-vector model II.

Definition 3.2.6 *Objective function of slide-vector model II*

Given $\boldsymbol{\Delta}$ *and the number of dimensions* d*, the objective function of slide-vector model II is defined as follows:*

$$L(\boldsymbol{X}, \boldsymbol{z} |\ \boldsymbol{\Delta}) = \frac{1}{2} \| \boldsymbol{W}^{\dagger} \odot (\boldsymbol{\Delta}^{\dagger} - \boldsymbol{D}(\boldsymbol{Q})) \|^2$$

where

$$\boldsymbol{\Delta}^{\dagger} = \begin{bmatrix} \boldsymbol{O}_n & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}^T & \boldsymbol{O}_n \end{bmatrix},$$

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X} - \mathbf{1}_n \boldsymbol{z}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \boldsymbol{I} & -\mathbf{1}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{z}^T \end{bmatrix} = (q_{i^{\dagger}t})$$

$$(i^{\dagger} = 1, 2, \cdots, 2n; t = 1, 2, \cdots, d)$$

$$\boldsymbol{D}(\boldsymbol{Q}) = (d_{ij}(\boldsymbol{Q})),$$

$$d_{ij}(\boldsymbol{Q}) = \Big[\sum_{t=1}^{d} (q_{it} - q_{jt})^2 \Big]^{\frac{1}{2}} \ (i, j = 1, 2, \cdots, 2n),$$

$$\boldsymbol{W}^{\dagger} = \begin{bmatrix} \boldsymbol{O}_n & \boldsymbol{W} \\ \boldsymbol{W}^T & \boldsymbol{O}_n \end{bmatrix}$$

In the slide-vector model, $\boldsymbol{X}$ and $\boldsymbol{z}$ are estimated such that the value of the objective function is minimized.

Then, equivalence of these objective functions for slide-vector model I and II is shown.

Proposition 3.2.2 *Equivalence of these objective functions of slide-vector model I and II*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$, coordinates of objects $\boldsymbol{X}$, and slide-vector $\boldsymbol{z}$, the following equation is satisfied

$$\frac{1}{2} \| \boldsymbol{W}^{\dagger} \odot (\boldsymbol{\Delta}^{\dagger} - \boldsymbol{D}(\boldsymbol{Q})) \|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ij} - d_{ij}(\boldsymbol{X}, \boldsymbol{X} - \mathbf{1}_n \boldsymbol{z}^T))^2$$

Proof. *The proof is the same way as Unfolding*

Again, from proposition 3.2.2, we interpret the objective function of the slide-vector model as a special case of Unfolding.

3.3 Asymmetric MDS based on decomposistion

In this section, we introduce the hill-climbing (Borg and Groenen, 2005) and radius model(Okada and Imaizumi, 1987). Objective functions for these models can be decomposed into symmetric and skew-symmetric parts, respectively, based on proposition 3.1.2. In particular, this means that these objective functions can be decomposed into mutually orthogonal terms. The symmetric parts are the same as that of objective functions of ordinal MDS for symmetric part of the asymmetric dissimilarity matrix; furthermore, the skew-symmetric part of the objective function of the hill-climbing model includes the same parameters of the symmetric part. In

contrast, the objective function of the radius model has no parameters included in both the symmetric and skew-symmetric parts. Therefore, when the parameters of the symmetric parts are estimated in the radius model, we do not need to consider the skew-symmetric part. In short, these decompositions reveal various important aspects of these objective functions.

Below, we show the models, objective functions and decompositions of the hill-climbing model and radius model.

3.3.1 Model and objective function of hill-climbing model

In this subsection, we show the model, objective function, and decomposition of the hill-climbing model.

First, we introduce the hill-climbing model itself.

Definition 3.3.1 *Model of the hill-climbing model*

Let $\boldsymbol{\Delta}$ *be an asymmetric dissimilarity matrix,* $\boldsymbol{X} = (x_{it})$, $x_{it} \in \mathbb{R}$ $(i = 1, 2, \cdots, n; t = 1, 2, \cdots, d)$ *and* $\boldsymbol{v} = (v_t)$ $v_t \in \mathbb{R}$ $(t = 1, 2, \cdots, d)$ *be coordinates matrix of objects and slope vector, respectively. The model of hill-climbing model is defined as following equation:*

$$\delta_{ij} = d_{ij}(\boldsymbol{X}) + \Big[\sum_{t=1}^{d}(x_{it} - x_{jt})v_t\Big]\Big(d_{ij}(\boldsymbol{X})^{-1}\Big) + \varepsilon_{ij} \quad (i, j = 1, 2, \cdots, n; i \neq j).$$

where $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error.*

In the hill-climbing model, asymmetries between objects are represented by one vector, slope vector although the interpretation of the slope vector is different from the slide-vector.

Next, we define the objective function of the hill-climbing model.

Definition 3.3.2 *Objective function of hill-climbing model*

Given asymmetric dissimialrity data $\boldsymbol{\Delta}$ *and the number of low-dimensions* d *, the objective function of the hill-climbing model is defined as follows:*

$$L(\boldsymbol{X}, \boldsymbol{v} |\, \boldsymbol{\Delta}) = \sum_{i \neq j}\Big[\delta_{ij} - \Big(d_{ij}(\boldsymbol{X}) + (\boldsymbol{x}_i - \boldsymbol{x}_j)^T\boldsymbol{v}d_{ij}(\boldsymbol{X})^{-1}\Big)\Big]^2$$

where $\boldsymbol{x}_i = (x_{it})$, $x_{it} \in \mathbb{R}(i = 1, 2, \cdots, n; t = 1, 2, \cdots, d)$ *is coordinates vector of object* i*.*

Next, we show the property of the objective function of the hill-climbing model. From the property, the relationships between the slope-vector and the skew-symmetric part become clear.

Proposition 3.3.1 *Decomposition of the objective function of the hill-climbmodel*
The objective function of the hill-climbing model can be decomposed into symmetric and skew-symmetric parts of the objective function as follows:

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{v}|\ \boldsymbol{\Delta}) &= \sum_{i \neq j} \left[\delta_{ij} - \left(d_{ij}(\boldsymbol{X}) + (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1} \right) \right]^2 \\ &= \sum_{i \neq j} \left[s_{ij} - d_{ij}(\boldsymbol{X}) \right]^2 + \sum_{i \neq j} \left[a_{ij} - (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1} \right]^2 \end{aligned}$$

where

$$\begin{aligned} &\boldsymbol{S} = (\boldsymbol{\Delta} + \boldsymbol{\Delta}^T)/2,\ \boldsymbol{S} = (s_{ij})\ (i, j = 1, 2, \cdots, n),\ \text{and} \\ &\boldsymbol{A} = (\boldsymbol{\Delta} - \boldsymbol{\Delta}^T)/2,\ \boldsymbol{A} = (a_{ij})\ (i, j = 1, 2, \cdots, n) \end{aligned}$$

are symmetric and skew-symmetric parts of asymmetric dissimilarity data, respectively.

Proof.

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{v}|\ \boldsymbol{\Delta}) =& \sum_{i \neq j} \left[\delta_{ij} - \left(d_{ij}(\boldsymbol{X}) + (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1} \right) \right]^2 \\ =& \sum_{i \neq j} \left[(\delta_{ij} + \delta_{ji})/2 + (\delta_{ij} - \delta_{ji})/2 - d_{ij}(\boldsymbol{X}) - (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1} \right]^2 \\ =& \sum_{i \neq j} \left[s_{ij} - d_{ij}(\boldsymbol{X}) \right]^2 \\ &+ \sum_{i \neq j} \left[a_{ij} - (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1} \right]^2 \\ &+ 2 \sum_{i \neq j} \left[s_{ij} - d_{ij}(\boldsymbol{X}) \right] \left[a_{ij} - (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1} \right]. \end{aligned} \tag{3.6}$$

Here, $\boldsymbol{S}$ and $\boldsymbol{D}(\boldsymbol{X})$ are symmetric matrices, respectively and $\boldsymbol{A}$ and $\boldsymbol{A}^\dagger = ((\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1})\ (i, j = 1, 2, \cdots, n)$ are skew-symmetric matrices, respectively, since $(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1} = -(\boldsymbol{x}_j - \boldsymbol{x}_i)^T \boldsymbol{v} d_{ij}(\boldsymbol{X})^{-1}$ is satisfied. $\boldsymbol{S} - \boldsymbol{D}$ is symmetric matrix and $\boldsymbol{A} - \boldsymbol{A}^\dagger$ is skew-symmetric matrix from proposition 3.1.3. Therefore, from proposition 3.1.2, Eq (3.6) becomes 0 and the proposition is proved.

From proposition 3.3.1, the distance between estimated coordinates of objects indicate corresponding dissimilarities for symmetric part of the asymmetric dissimilarity data. In constrast, the inner product between the slope-vector and difference vectors between objects represent the skew-symmetries part of asymmetric dissimilarity data. Figure 3.3 shows an example of the results of hill-climbing model. In the example, these distances on the estimated coordinates between objects correspond to symmetric part of the asymmetric dissimilarity data. On the other hand, the value of innner product between $\boldsymbol{x}_j - \boldsymbol{x}_o$ and $\boldsymbol{v}$ tends to be higher than that

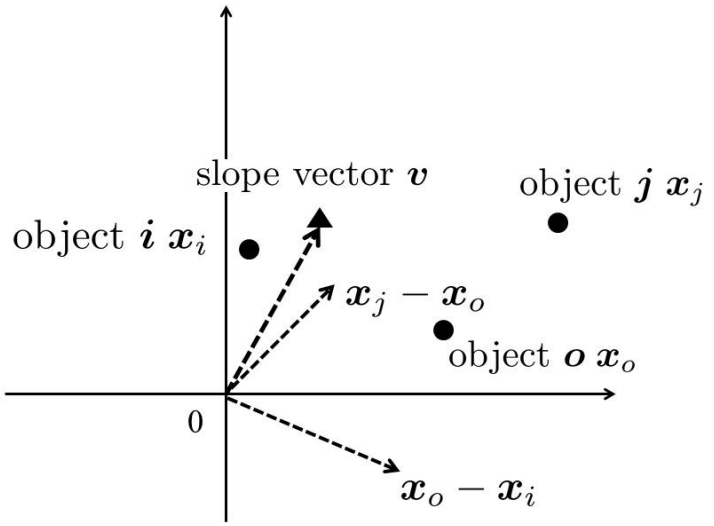

Figure 3.3: Example of hill climbing model.

between $\boldsymbol{x}_o - \boldsymbol{x}_i$ and $\boldsymbol{v}$ in the example. Then, the skew-symmetry between object j and o is interpreted as higher than that between object j and o. In addition, skew-symmetries from object o to object j is larger than that from object j to object o.

3.3.2 Model and objective function of radius model

In this subsection, we introduce the model, objective function, and property of the radius model.

First, we define the radius model.

Definition 3.3.3 *Model of the radius model*

Let $\boldsymbol{\Delta}$ *be asymmetric dissimilarity matrix,* $\boldsymbol{X} = (x_{ij}), x_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *and* $\boldsymbol{r} = (r_i)$ $r_i > 0$ $(i = 1, 2, \cdots, n)$ *be coordinate matrix of objects and the length of the radii for objects, respectively. The model of the radius model is defined as following equation:*

$$\delta_{ij} = d_{ij}(\boldsymbol{X}) - r_i + r_j + \varepsilon_{ij} \ (i, j = 1, 2, \cdots, n) \tag{3.7}$$

where $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error term. The model of the radius model based on matrix representation is defined as follows:*

$$\boldsymbol{\Delta} = \boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T + \boldsymbol{E}$$

where

$$\boldsymbol{D}(\boldsymbol{X}) = (d_{ij}(\boldsymbol{X})) \quad d_{ij}(\boldsymbol{X}) = \Big[\sum_{t=1}^{d} (x_{it} - x_{jt})^2 \Big]^{\frac{1}{2}} \ (i, j = 1, 2, \cdots, n)$$

is Euclidean distance matrix, $\mathbf{1}_n$ *is a vector whose elemtns are all* 1 *with the length of* n*, and* $\boldsymbol{E} = (\varepsilon_{ij})$ $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error matrix.*

From Eq. (3.7), differences between objects are described as distance between estimated coordinates of objects. For the interpretation of asymmetries, see Figure 3.4. If length of radius for object j is large and length of radius for object i is small, these relations are interpreted as large asymmetries. In the situation, distance from object j to object i is closer than that from object i to j.

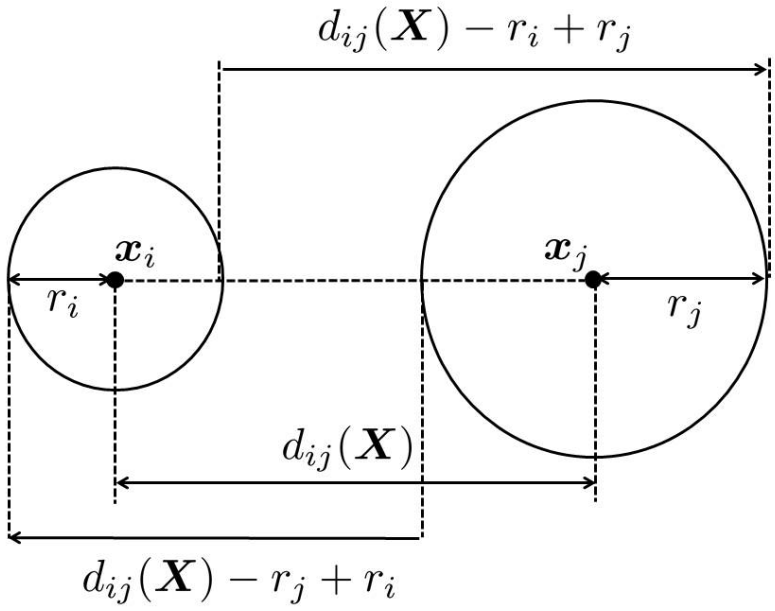

Figure 3.4: Model of raidius model

Next, objective function of radius model is defined.

Definition 3.3.4 *Objective function of the radius model*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$*, and the number of low-dimensions* d*, the objective function of the radius model is defined as follows:*

$$L(\boldsymbol{X}, \boldsymbol{r} |\ \boldsymbol{\Delta}) = \left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T) \right\|^2.$$

In the radius model, $\boldsymbol{X}$ *and* $\boldsymbol{r}$ *such that the value of the objective function is minimized are estimated.*

Next, the decomposition of the objective function for the radius model is shown.

.

Proposition 3.3.2 *Decomposition of the objective function of the radius model*

The objective function of the radius model can be decomposed into symmetric part and skew-symmetric part of the objective function, respectively, as follows:

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{r} |\ \boldsymbol{\Delta}) &= \left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T) \right\|^2 \\ &= \left\| \boldsymbol{S} - \boldsymbol{D}(\boldsymbol{X}) \right\|^2 + \left\| \boldsymbol{A} - (\boldsymbol{r}\mathbf{1}_n^T - \mathbf{1}_n \boldsymbol{r}^T) \right\|^2 \end{aligned} \tag{3.8}$$

where $\boldsymbol{S}$ and $\boldsymbol{A}$ are symmetric part and skew-symmetric part of $\boldsymbol{\Delta}$, respectively.

Proof. *The objective function of the radius model can be decomposed as follows:*

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{r} |\ \boldsymbol{\Delta}) =& \left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T) \right\|^2 \\ =& \left\| \boldsymbol{S} + \boldsymbol{A} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T) \right\|^2 \\ =& \left\| \boldsymbol{S} - \boldsymbol{D}(\boldsymbol{X}) + \boldsymbol{A} - (\boldsymbol{r}\mathbf{1}_n^T - \mathbf{1}_n \boldsymbol{r}^T) \right\|^2 \end{aligned}$$

$\boldsymbol{S} - \boldsymbol{D}(\boldsymbol{X})$ is symmetric matrix since both $\boldsymbol{S}$ and $\boldsymbol{D}(\boldsymbol{X})$ are symmetric matrix. In addition, $\boldsymbol{A} - (\boldsymbol{r}\mathbf{1}_n^T - \mathbf{1}_n \boldsymbol{r}^T)$ becomes skew-symmetric matrix from proposition 3.1.3 since $\boldsymbol{A}$ is skew-symmetric matrix and

$$-(\boldsymbol{r}\mathbf{1}_n^T - \mathbf{1}_n \boldsymbol{r}^T)^T = -\mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}_n \mathbf{1}_n^T = \boldsymbol{r}_n \mathbf{1}_n^T - \mathbf{1}_n \boldsymbol{r}^T.$$

Therefore, this proposition is proved from proposition 3.1.2.

From Eq. (3.8), the difference between radius lengths between object i and object j indicate skew-symmetry between object i and object j. Furthermore, parameter $\boldsymbol{X}$ is included only in the symmetric part of Eq. (3.8), while parameter $\boldsymbol{r}$ is included only in the skew-symmetric part of Eq. (3.8). Therefore, estimating $\boldsymbol{X}$ in the radius model is performed in the same manner as that of ordinal MDS. Furthermore, when $\boldsymbol{r}$ is estimated, the estimation depends only on the skew-symmetric part of Eq. (3.8).

In the objective function of the radius model, the feasible area of r_i $(i = 1, 2, \cdots, n)$ is greater than zero; however, we do not need to estimate r_i above zero because the objective function of the radius model includes indefiniteness. Therefore, the optimization problem of r_i becomes a non-constrained optimization problem with the indefiniteness of the objective function of the radius model shown.

Proposition 3.3.3 *Indefiniteness of the objective function of radius model*

For the objective function of the radius model, following property is satisfied:

$(\forall c \in \mathbb{R})$

$$\left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T) \right\|^2 = \left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n (\boldsymbol{r} + c\mathbf{1}_n)^T + (\boldsymbol{r} + c\mathbf{1}_n)\mathbf{1}_n^T) \right\|^2$$

Proof. *For any $c \in \mathbb{R}$,*

$$\begin{aligned} & \left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n (\boldsymbol{r} + c\mathbf{1}_n)^T + (\boldsymbol{r} + c\mathbf{1}_n)\mathbf{1}_n^T) \right\|^2 \\ =& \left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T) + c\mathbf{1}_n\mathbf{1}_n^T - c\mathbf{1}_n\mathbf{1}_n^T \right\|^2 \\ =& \left\| \boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T) \right\|^2. \end{aligned}$$

From proposition 3.3.3, the optimization problem becomes a non-constrained problem. In short, even if there exists an i such that estimated $r_i < 0$, the value of the objective function does not change for transformation $\boldsymbol{r}^{\dagger} = \boldsymbol{r} - c\boldsymbol{1}$, where $c = \min_i r_i$.

3.4 Cluster difference scaling (CDS)

In this section, we introduce cluster difference scaling (CDS) (Heiser, 1993; Heiser and Groenen, 1997). CDS is a simutaneous method for MDS and clustering of symmetric dissimilarity data. As a result of CDS, clustering results and coordinates of cluster centorids are estimated. Therefore, it is useful when the number of objects is large. Here, the objective function of CDS can be decomposed into four parts, i.e., *Among cluster error sum of squares (SSQ)*, *Within-clusters Error SSQ*, *Lack of spatial fit* and *Lack of homogeneity*. From this decomposition, the needs for simultaneous analysis becomes clear.

Among cluster error sum of squares (SSQ) indicates variations of dissimilarities for between clusters, *Within-clusters Error SSQ* indicates variations of dissimilarities between objects within clusters, *Lack of spatial fit* represent MDS for dissimilarities between clusters, and *Lack of homogeneity* is variation of self dissimilarities for clusters. For the details of interpretaion of these terms, see Heiser and Groenen (1997).

3.4.1 Model and objective function of CDS

In this subsection, we define the CDS model. Next, we introduce the objective function of both CDS and constrained MDS; the equivalence of these objective functions is also shown.

Definition 3.4.1 *Model of the cluster difference scaling*

Let $\boldsymbol{\Xi} = (\xi_{ij})$ $\xi_{ij} \in \mathbb{R}_+$ $(i, j = 1, 2, \cdots, n; i \neq j)$ *be symmetric dissimilarity matrix, and* C_ℓ $(\ell = 1, 2, \cdots, k)$ *be cluster of objects, where* k *is the number of clusters. The model of the cluster difference scaling is defined as following equation:*

$$(\forall i, j = 1, 2, \cdots, n)(\exists! C_o;\ i \in C_o)(\exists! C_\ell;\ j \in C_\ell)(\xi_{ij} = d_{o\ell}(\boldsymbol{X}^*) + \varepsilon_{ij})$$

where $\boldsymbol{X}^* = (x^*_{ot})$ $x^*_{ot} \in \mathbb{R}$ $(o = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$ *is a coordinates matrix of cluster centroids,* $d_{o\ell}(\boldsymbol{X}^*)$ *is an Euclidean distance between cluster* C_o *and* C_ℓ *on* $\boldsymbol{X}^*$*, and* $\varepsilon_{ij} \in \mathbb{R}$ *is error. .*

Next, we define the objective function of CDS.

Definition 3.4.2 *Objective function of the cluster difference scaling*

Given symmetric dissimilarity data $\boldsymbol{\Xi}$, the number of clusters k and the number of low-dimensions d, the objective function of the cluster difference scaling is defined as follows:

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{U} |\ \boldsymbol{\Xi}) &= \sum_{\ell=1}^{k} \sum_{o=1}^{k} \sum_{i \in C_o} \sum_{j \in C_\ell} (\xi_{ij} - d_{o\ell}(\boldsymbol{X}^*))^2 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\ell=1}^{k} \sum_{o=1}^{k} u_{i\ell} u_{jo} (\xi_{ij} - d_{o\ell}(\boldsymbol{X}^*))^2 \end{aligned}$$

where $\boldsymbol{U} = (u_{i\ell})$ $u_{i\ell} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; \ell = 1, 2, \cdots, k)$ is an indicator matrix. Here, $\boldsymbol{X}^*$ and $\boldsymbol{U}$ are estimated such that the value of the objective function is minimized.

Next, we define the constrained MDS model.

Definition 3.4.3 *Model of the constrained MDS*

Let $\boldsymbol{\Xi}$ be a symmetric dissimilarity matrix, and $\boldsymbol{U} = (u_{i\ell})$ $u_{i\ell} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; \ell = 1, 2, \cdots, k)$ be indicator matrix. The model of constrained MDS is defined as following equation:

$$\boldsymbol{\Xi} = \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) + \boldsymbol{E}.$$

where

$$\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) = (d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)) \quad d_{ij}(\boldsymbol{U}\boldsymbol{X}^*) = \left\| \sum_{o=1}^{k} u_{io} \boldsymbol{x}_o - \sum_{\ell=1}^{k} u_{j\ell} \boldsymbol{x}_\ell \right\|,$$

and $\boldsymbol{E} = (\varepsilon_{ij})$ $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ is error matrix. Here, $\boldsymbol{x}_o$ is the oth row-vector of $\boldsymbol{X}^*$.

Next, we define the objective function of the constrained MDS.

Definition 3.4.4 *Objective function of the constrained MDS*

Given symmetric dissimilarity matrix $\boldsymbol{\Xi}$, the number of clusters k and the number of low-dimensions d, the objective function of the constrained MDS is defined as follows:

$$L(\boldsymbol{X}, \boldsymbol{U} |\ \boldsymbol{\Delta}) = \left\| \boldsymbol{\Xi} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) \right\|^2.$$

In the objective function, $\boldsymbol{U}$ and $\boldsymbol{X}^*$ are estimated such that a value of the objective function is minimized.

Next, we show the equivalence of the objective functions of CDS and constrained MDS.

Proposition 3.4.1 *Equivalence of CDS and constrained MDS*

Given symmetric dissimilarity matrix $\boldsymbol{\Xi}$, Indicator matrix $\boldsymbol{U}$, coordinate matrix of cluster centorids $\boldsymbol{X}^*$, the number of clusters k and the number of low-dimensions d, the following property is satisfied;

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\ell=1}^{k}\sum_{o=1}^{k} u_{i\ell}u_{jo}(\xi_{ij} - d_{\ell o}(\boldsymbol{X}^*))^2 = \left\|\boldsymbol{\Delta} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\right\|^2. \tag{3.9}$$

Proof. *From the right term of Eq (3.9),*

$$\begin{aligned}\left\|\boldsymbol{\Xi} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\right\|^2 =& \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\xi_{ij} - d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)\right)^2 \\ =& \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\xi_{ij} - \left\|\sum_{\ell=1}^{k} u_{i\ell}\boldsymbol{x}^*_\ell - \sum_{o=1}^{k} u_{jo}\boldsymbol{x}^*_o\right\|\right)^2, \end{aligned} \tag{3.10}$$

where $\boldsymbol{x}^*_\ell$ $(\ell = 1, 2, \cdots, k)$ is row-vector of $\boldsymbol{X}^*$.

Here, for all i, ℓ^* exists such that $u_{i\ell^*} = 1$ and $u_{i\ell} = 0$ $(\ell^* \neq \ell)$ from the definition of indicator matrix. In the same way as i, for all j, o^* exists such that $u_{jo^*} = 1$ and $u_{jo} = 0$ $(o^* \neq o)$. Therefore, Eq (3.10) for i, j is described as follows:

$$\begin{aligned}(\xi_{ij} - \|\boldsymbol{x}^*_{\ell^*} - \boldsymbol{x}^*_{o^*}\|)^2 =& u_{i\ell^*}u_{jo^*}(\xi_{ij} - \|\boldsymbol{x}^*_{\ell^*} - \boldsymbol{x}^*_{o^*}\|)^2 \\ =& \sum_{\ell=1}^{k}\sum_{o=1}^{k} u_{i\ell}u_{jo}(\xi_{ij} - \|\boldsymbol{x}^*_\ell - \boldsymbol{x}^*_o\|)^2.\end{aligned}$$

Then, we have this proposition.

From Eq.(3.9), we note that CDS is equivalent to constrained MDS in terms of coordinates of matrix with indicator matrix.

3.4.2 Properties of CDS

In this subsection, we show how to interpret the features of the objective function of CDS by the decomposition of the objective function for CDS. When the objective function of CDS can be decomposed into four parts, we use the Sokal-Michener dissimilarity (Sokal and Michener, 1958), where Sokal-Michener dissimilarity indicates aggregated dissimilarities within and between clusters. From the decomposition, we also interpret the effects of clustering and MDS results.

Next, we show the decomposition of the objective function of CDS.

Proposition 3.4.2 *Decomposition of the objective function of CDS*

Given the objective function of CDS, the objective function can be decomposed as follows:

$$
\begin{aligned}
L(\boldsymbol{X}, \boldsymbol{U} |\, \boldsymbol{\Xi}) =& \|\boldsymbol{\Xi} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\|^2 \\
=& \sum_{o \neq \ell} \|\mathrm{diag}(\boldsymbol{u}_{(o)})[\boldsymbol{\Xi} - \boldsymbol{P}_U \boldsymbol{\Xi} \boldsymbol{P}_U]\mathrm{diag}(\boldsymbol{u}_{(\ell)})\|^2 && (3.11) \\
&+ \sum_{o=1}^{k} \|\mathrm{diag}(\boldsymbol{u}_{(o)})[\boldsymbol{\Xi} - \boldsymbol{P}_U \boldsymbol{\Xi} \boldsymbol{P}_U]\mathrm{diag}(\boldsymbol{u}_{(o)})\|^2 && (3.12) \\
&+ \sum_{o \neq \ell} \|\mathrm{diag}(\boldsymbol{u}_{(o)})[\boldsymbol{P}_U \boldsymbol{\Xi} \boldsymbol{P}_U - \boldsymbol{P}_U \boldsymbol{\Xi} \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\boldsymbol{P}_U]\mathrm{diag}(\boldsymbol{u}_{(\ell)})\|^2 && (3.13) \\
&+ \sum_{o=1}^{k} \|\mathrm{diag}(\boldsymbol{u}_{(o)})[\boldsymbol{P}_U \boldsymbol{\Xi} \boldsymbol{P}_U - \boldsymbol{P}_U \boldsymbol{\Xi} \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\boldsymbol{P}_U]\mathrm{diag}(\boldsymbol{u}_{(o)})\|^2 && (3.14)
\end{aligned}
$$

where $(\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T\boldsymbol{\Xi}\boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}$ *is Sokal-Michener dissimilarity matrix,* $\boldsymbol{P}_U = \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T$ *is projection matrix and* $\boldsymbol{u}_{(o)}$ $(o = 1, 2, \cdots, k)$ *is oth column vector of* $\boldsymbol{U}$*.*

Proof. *The objective function of CDS can be decomposed as fllows:*

$$
\begin{aligned}
\|\boldsymbol{\Xi} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\|^2 &= \|\boldsymbol{\Xi} - \boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U + \boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\|^2 \\
&= \|\boldsymbol{\Xi} - \boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U\|^2 && (3.15) \\
&\quad + \|\boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\|^2 && (3.16) \\
&\quad + tr(\boldsymbol{\Xi} - \boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U)^T(\boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)). && (3.17)
\end{aligned}
$$

Eq.(3.17) is transformed as follows:

$$
\begin{aligned}
&tr(\boldsymbol{\Xi}^T\boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U) - tr(\boldsymbol{\Xi}^T\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)) \\
&\quad - tr(\boldsymbol{P}_U^T\boldsymbol{\Xi}^T\boldsymbol{P}_U^T\boldsymbol{P}_U\boldsymbol{\Xi}\boldsymbol{P}_U) + tr(\boldsymbol{P}_U^T\boldsymbol{\Xi}^T\boldsymbol{P}_U^T\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)) && (3.18)
\end{aligned}
$$

Since $\boldsymbol{P}_U = \boldsymbol{P}_U^T$ *and* $\boldsymbol{P}_U\boldsymbol{P}_U = \boldsymbol{P}_U$*, the sum of the first term and the third term in Eq.(3.18) becomes* 0*. In addition, since* $\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) = \boldsymbol{U}\boldsymbol{D}(\boldsymbol{X}^*)\boldsymbol{U}^T$*, sum of the second term and forth term becomes* 0*. Therefore, Eq.(3.18) becomes* 0*.*

Eq. (3.15) can be decomposed into Eq. (3.11) and Eq. (3.12), and Eq. (3.16) can be decomposed into Eq. (3.13) and Eq. (3.14).

Therefore, this proposition is proved.

Next we show how to interpret for Eq (3.11), Eq (3.12), Eq (3.13), and Eq (3.14). Here Eq (3.11), Eq (3.12), Eq (3.13), and Eq (3.14) are interpreted as *Among-cluster Error Sum of Squares (SSQ)*, *Within-clusters Error SSQ*, *Lack of spatial fit*, and *Lack of homogeneity*, respectively. From the terms of *Among-cluster Error Sum of Squares (SSQ)* and *Lack of spatial fit*, CDS can be interpreted as simultaneous analysis involving non-hierarchical clustering applied to dissimilarity data and MDS for these centorids. Furthermore, CDS considers the effects of self dissimilarities for clustering from *Within-clusters Error SSQ* and *Lack of homogeneity*.

Chapter 4

Constrained analysis of asymmetric data based on Unfolding

4.1 Background of constrained asymmetric MDS based on Unfolding

As noted in the introduction, asymmetric dissimilarity data exists and is observed in different research areas, including marketing and psychology. It is defined as dissimilarity from i to j that is not necessarily the same as that from j to i. If asymmetries are considered as informative, it is important to these interpret the asymmetries accordingly. As one method for achieving this, AMDS is useful. In particular AMDS is a visualization method for interpreting asymmetries that exist in asymmetric dissimilarity data. To date, nimerous AMDS models have been proposed (Borg and Groenen, 2005; Chino, 2012;Saito and Yadohisa, 2005;Zielman and Heiser, 1996).

In Saito and Yadohisa (2005), AMDS models are classified with Unfolding-type models introduced as one group of methods. Unfolding was originally proposed for two-way two-mode dissimilarity data by Coombs (1964). In the results of Unfolding, coordinates of row objects and column objects are simutaneously estimated although relationships within coordinates of row objects and column objects are no-meaning, respectively. Gower (1977) and Constantine and Gower (1978) suggested the application of Unfolding to asymmetric dissimilarity data. In their work, each object has two coordinates; when dissimilarity from object i to object j is interpreted, it correspond to the distance from the estimated coordinates of row object i to that of column object j. In other words, the distance from estimated coordinates of row object i to that of column object j is not necessarily the same as the distance from estimated coordinates of row object j to that of column object i. Therefore, asymmetries can be described by applying Unfolding to asymemtric dissimilarity data.

However, the number of estimated parameters is large for Unfolding. In response to this, the slide-vector model was proposed, as a special case of Unfolding based on parsimonious notions (De Leeuw and Heiser, 1982; Zielman and Heiser, 1993). In the model, one vector called slide-vector is introduced and coordinates of column objects are described by combinations of coordinates of row objects and this slide-vector.

Also noted in the introduction, improved information technology has provided us with substantial amounts of large and complex data. Therefore, asymmetric dissimilarity data has also become large and complex. For example, brand swithing data, which is analyzed to reveal competitive relations for customer loyalty, become large and complex when data is calculated from purchase behavior of e-commerce web sites. It therefore becomes difficult to interpret asymmetries between objects because the number of objects is so large. To overcome this problem of scale, constrained Unfolding and the constrained slide-vector model, which involve the simultaneous analysis of clustering and AMDS, have been proposed on a basis of CDS (Heiser, 1993). Simultaneous analysis of Unfolding and CDS for two-way two-mode dissimilarity data was proposed by Vera et al. (2013) and called cluster difference Unfolding (CDU). This method corresponds to two-mode clustering of row objects and column objects. However, when simultaneous analysis of Unfolding and clustering is applied to asymmetric dissimilarity data, clustering results for both row and column objects should be the same because it is easier and more natural to interpret the features. Therefore, we proposed constrained Unfolding subject to indicator matrices of row and column objects being the same. Furthermore, there are certainly cases in which the number of clusters is large; in such situation, even if constrained Unfolding based on CDS is applied to the asymmetric dissimilarity data, it becomes difficult to interpret the asymmetries between clusters. Therefore, we propose a method that simultaneously involved the slide-vector model and clustering; this method is called the constrained slide-vector model and adopts not only one slide-vector but also several numbers of slide-vectors to more flexibly represent given asymmetries. Finally, we show that our constrained slide-vector model is considered a generalization of Unfolding and constrained Unfolding.

In the remaining sections of this chapter, the model, objective function and algorithm of the constrained Unfolding and constrained slide-vector model are shown in section 3.2 and section 3.3, respectively. Then, in section 3.4, we show relations between constrained Unfolding and constrained slide-vector model, Unfolding and constrained slide-vector models, and the slide-vector and constrained slide-vector models.

4.2 Constrained Unfolding based on CDS

This section comprise three parts, i.e., a description of the model and objective function of constrained Unfolding, the property, and the algorithm.

4.2.1 Model and objective function of the constrained Unfolding

In this subsection, we show the model and objective function of constrained Unfolding. Furthermore, we introduce two types of descriptions for objective functions of constrained Unfolding as well as the equivalence based on Heiser and Groenen (1997).

Definition 4.2.1 *Model of the constrained Unfolding*

Let $\boldsymbol{\Delta} = (\delta_{ij})$ $\delta_{ij} \in \mathbb{R}_+$ $(i, j = 1, 2, \cdots, n)$ *be an asymmetric dissimilarity matrix,* $\boldsymbol{X}^* = (x^*_{ot})$ $x^*_{ot} \in \mathbb{R}$ $(o = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$ *and* $\boldsymbol{Y}^* = (y^*_{\ell t})$ $y^*_{\ell t} \in \mathbb{R}$ $(\ell = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$ *be cooridinates of clusters for row-objects and column-clusters in* d *dimensions, respectively, where* k *is the number of clusters for objects . Here, the constrained Unfolding model is defined as follows:*

$$(\forall i, j = 1, 2, \cdots, n)(\exists! C_o; i \in C_o)(\exists! C_\ell; j \in C_\ell)(\delta_{ij} = d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*) + \varepsilon_{ij})$$

where

$$d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*) = \|\boldsymbol{x}^*_o - \boldsymbol{y}^*_\ell\| = \Big[\sum_{t=1}^{d}(x^*_{ot} - y^*_{\ell t})^2\Big]^{\frac{1}{2}},$$

C_o *and* C_ℓ *are clusters of objects, respectively, and* $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error.*

For the constraind Unfolding model, dissimilarity from objects belonging to cluster o to objects belonging to cluster ℓ is represented by the distance from $\boldsymbol{x}^*_o$ to $\boldsymbol{y}^*_\ell$.

Next, we define the first type of objective function of constrained Unfolding, i.e., constrained Unfolding I.

Definition 4.2.2 *Objective function of the constrained Unfolding I*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$*, the number of clusters* k*, and the number of low-dimensions* d*, the objective function of the constrained Unfolding is defined as follows:*

$$L(\boldsymbol{X}^*, \boldsymbol{Y}^*, \boldsymbol{U} |\ \boldsymbol{\Delta}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k}\sum_{\ell=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2$$

where $\boldsymbol{U} = (u_{io})$ $u_{io} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; o = 1, 2, \cdots, k)$ *is indicator matrix of objects. In the constrained Unfolding model,* $\boldsymbol{X}^*$*,* $\boldsymbol{Y}^*$ *and* $\boldsymbol{U}$ *are estimated such that values of the objective finction is minimized.*

Here, we define the objective function of constrained Unfolding II; we also prove the equivalence of the objective functions of these constrained Unfolding models.

Definition 4.2.3 *The objective function of constrained Unfolding II*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$*, the number of clusters of objects* k *and the number of low-dimensions* d*, the objective function of constrained Unfolding II is defined as follows:*

$$L(\boldsymbol{X}^*, \boldsymbol{Y}^*, \boldsymbol{U} |\ \boldsymbol{\Delta}) = \frac{1}{2}\left\| \boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\dagger)) \right\|^2$$

where $\boldsymbol{W}^\dagger = (w^\dagger_{i^\dagger j^\dagger})$ $w^\dagger_{i^\dagger j^\dagger} \in \{0, 1\}$ $(i^\dagger, j^\dagger = 1, 2, \cdots, 2n)$ *are weight matrix for dissimilarities*

$$\boldsymbol{\Delta}^\dagger = \begin{bmatrix} \boldsymbol{O}_n & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}^T & \boldsymbol{O}_n \end{bmatrix}, \text{ and}$$

$$\boldsymbol{Q}^\dagger = \begin{bmatrix} \boldsymbol{U}\boldsymbol{X}^* \\ \boldsymbol{U}\boldsymbol{Y}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{O}_{n,k} \\ \boldsymbol{O}_{n,k} & \boldsymbol{U} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}^* \\ \boldsymbol{Y}^* \end{bmatrix}.$$

Here, $\boldsymbol{O}_{n,k} = (0)$ *is* n *by* k *matrix. In the constrained Unfolding II,* $\boldsymbol{U}$*,* $\boldsymbol{X}^*$ *and* $\boldsymbol{Y}^*$ *are estimated such that the value of the objective function is minimized.*

Next, we show the equivalence of the objective functions of constrained Unfolding I and II.

Proposition 4.2.1 *Equivalence of constraind Unfolding I and II*

Given asymmetric dissimilarity matrix $\boldsymbol{\Delta}$*, indicator matrix of objects* $\boldsymbol{U}$*, coordinates of cluster centroids of row-objects* $\boldsymbol{X}^*$ *and of column-objects* $\boldsymbol{Y}^*$*, respectively, and the number of low-dimensions* d*, the following property is satisfied;*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k}\sum_{\ell=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 = \frac{1}{2}\left\| \boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\dagger)) \right\|^2. \tag{4.1}$$

Proof. *From the right term of Eq (4.1),*

$$\frac{1}{2}\left\| \boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\dagger)) \right\|^2 = \frac{1}{2}\sum_{i^\dagger=1}^{2n}\sum_{j^\dagger=1}^{2n} w^\dagger_{i^\dagger j^\dagger}(\delta^\dagger_{i^\dagger j^\dagger} - d_{i^\dagger j^\dagger}(\boldsymbol{Q}^\dagger))^2. \tag{4.2}$$

From the definition of $\boldsymbol{W}^\dagger$ *and* $\boldsymbol{\Delta}^\dagger$*, Eq (4.2) is described as follows:*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(\delta_{ij} - d_{ij}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*))^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\delta_{ij} - \left\| \sum_{o=1}^{k} u_{io}\boldsymbol{x}^*_o - \sum_{\ell=1}^{k^*} u_{j\ell}\boldsymbol{y}^*_\ell \right\|\right)^2 \tag{4.3}$$

where $\boldsymbol{x}^*_o$ *and* $\boldsymbol{y}^*_\ell$ *are row vectors of* $\boldsymbol{X}^*$ *and* $\boldsymbol{Y}^*$*, respectively.*

Here, for all i and j, o^* and ℓ^* exists such that $u_{io^*} = 1$ and $u_{io} = 0(o \neq o^*)$, and $u_{j\ell^*} = 1$ and $u_{j\ell} = 0(\ell \neq \ell^*)$,respectively from the definition of indicator matrices. Therefore, i, j part of Eq (4.3) is described as follows:

$$\begin{aligned}\Big(\delta_{ij} - \Big\| \sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{y}_\ell^* \Big\|\Big)^2 =& \Big(\delta_{ij} - \|u_{io^*}\boldsymbol{x}_{o^*}^* - u_{j\ell^*}\boldsymbol{y}_{\ell^*}^*\|\Big)^2 \\ =& \Big(\delta_{ij} - \|\boldsymbol{x}_{o^*}^* - \boldsymbol{y}_{\ell^*}^*\|\Big)^2 \\ =& \sum_{o=1}^{k}\sum_{\ell=1}^{k} u_{io}u_{j\ell}\Big(\delta_{ij} - \|\boldsymbol{x}_o^* - \boldsymbol{y}_\ell^*\|\Big)^2\end{aligned}$$

Therefore, Eq (4.1) is proved.

From proposition 4.2.1, constrained Unfolding is considered as special case of ordinal MDS. The advantage of describing constrained Unfolding by the objective function II is that it is easy to derive the majorizing function (De Leeuw, 1994; Heiser, 1995; Lange, Hunter, and Yang, 2000;Kiers, 2002; Hunter and Lange, 2004) using the same framework of that of ordinal MDS.

4.2.2 Properties of the constrained Unfolding

In this subsection, we show the decomposition of constrained Unfolding; however, there is no homogeneity term and *Within-cluster Error SSQ* in the objective function of constrained Unfolding because the distance between the same object does not become zero. Therefore, the objective function of constrained Unfolding can be decomposed into two mutually orthogonal terms.

Proposition 4.2.2 *Decomposition of the constrained unfolding*
Given the objective function of constrained Unfoding, the objective function can be decomposed as follows:

$$\begin{aligned}L(\boldsymbol{X}^*, \boldsymbol{Y}^*, \boldsymbol{U} |\ \boldsymbol{\Delta}) =& \|\boldsymbol{\Delta} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*)\|^2 \\ =& \|\boldsymbol{\Delta} - \boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U\|^2 \qquad (4.4) \\ &+ \|\boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*)\|^2 \qquad (4.5)\end{aligned}$$

where $(\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T\boldsymbol{\Delta}\boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}$ is Sokal-Michener dissimilarity matrix (Sokal and Michener, 1958) and $\boldsymbol{P}_U = \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T$ is a projection matrix.

Proof. *The objective function of constrained Unfolding can be decomposed as follows:*

$$\begin{aligned}&\|\boldsymbol{\Delta} - \boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U + \boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*)\|^2 \\ =&\|\boldsymbol{\Delta} - \boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U\|^2 + \|\boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*)\|^2 \\ &+ tr(\boldsymbol{\Delta} - \boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U)^T(\boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*)) \qquad (4.6)\end{aligned}$$

The third term of Eq (4.6) becomes as follows:

$$tr(\boldsymbol{\Delta}^T\boldsymbol{P}_U\boldsymbol{\Delta}\boldsymbol{P}_U) - tr(\boldsymbol{\Delta}^T\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*)) \\ - tr(\boldsymbol{P}_U^T\boldsymbol{\Delta}^T\boldsymbol{P}_U^T\boldsymbol{\Delta}) + tr(\boldsymbol{P}_U^T\boldsymbol{\Delta}^T\boldsymbol{P}_U^T\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*)) \tag{4.7}$$

Since $\boldsymbol{P}_U = \boldsymbol{P}_U^T$ *and* $\boldsymbol{P}_U\boldsymbol{P}_U = \boldsymbol{P}_U$*, the sum of the first term and the third term in Eq.(4.7) becomes* 0*. In addition, the sum of the second term and forth term in Eq. (4.7) becomes* 0 *from* $\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}\boldsymbol{Y}^*) = \boldsymbol{U}\boldsymbol{D}(\boldsymbol{X}^*, \boldsymbol{Y}^*)\boldsymbol{U}^T$*. Therefore, this proposition is proved.*

Eq. (4.4) and Eq. (4.5) represent the *Among-clusters Error SSQ* and the *Lack of spatial fit*, respectively. The difference between the decomposition of CDS and constrained Unfolding is *Within-clusters error SSQ* and *Lack of homogeneity*. In short, the decomposition of constrained unfolding does not include the *Within-clusters error SSQ* and *Lack of homogeneity*. Furthermore, from the dcomposition, we consider the objective function of constrained Unfolding as simultaneous analysis of clustering of asymmetric dissimilarity data and Unfolding for the Sokal and Michener dissimilarities.

4.2.3 Algorithm of the constrained Unfolding

In this subsection, we show the algorithm of constrained Unfolding. These parameters are estimated on the basis of ALS (Young et al., 1980). The flow of the proposed algorithm is described as follows.

Algorithm of the constrained Unfolding

Step 0 Set k and d, and initial values of $\boldsymbol{X}^*$, $\boldsymbol{Y}^*$ and $\boldsymbol{U}$

Step 1 Update $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$, given $\boldsymbol{U}$

Step 2 Update $\boldsymbol{U}$, given $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$

Step 3 If stop condition is satisfied, stop the algorithm, else return to the Step 1

To update $\boldsymbol{X}^*$ and $\boldsymbol{Y}$, we adopt the majorization algorithm (Borg and Groenen, 2005). Next, we derive the majorizing function of the constrained Unfolding model.

Proposition 4.2.3 *Majorizing function of the constrained Unfolding*

Given objective function of the constrained unfolding, the majorizing function of the constrained Unfolding is given as follows:

$$\frac{1}{2}\left\|\boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\dagger))\right\|^2 \\ \leq \frac{1}{2}\eta_\delta^2 + \frac{1}{2}tr\boldsymbol{Q}^T\boldsymbol{\Phi}^T\boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{Q} - tr\boldsymbol{Q}^T\boldsymbol{\Phi}^T\boldsymbol{B}(\boldsymbol{\Phi}\boldsymbol{H})\boldsymbol{\Phi}\boldsymbol{H} = L_M(\boldsymbol{Q}, \boldsymbol{H}, \boldsymbol{\Phi}|\ \boldsymbol{\Delta}^\dagger) \tag{4.8}$$

where

$$\eta_\delta^2 = \sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger \delta_{i^*j^*}^{\dagger 2}$$

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{O}_{n,k} \\ \boldsymbol{O}_{n,k} & \boldsymbol{U} \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{X}^* \\ \boldsymbol{Y}^* \end{bmatrix},$$

$$\boldsymbol{V} = \sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger (\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})^T,$$

$$\boldsymbol{e}_{i^*} = (e_{i^*s^*}) \quad e_{i^*s^*} = \begin{cases} 1 & (i^* = s^*) \\ 0 & (i^* \neq s^*) \end{cases}, \quad (i^*, s^* = 1, 2, \cdots, 2n),$$

$$\boldsymbol{B}(\boldsymbol{\Phi H}) = (b_{i^*j^*})\ (i^*, j^* = 1, 2, \cdots, 2n)$$

$$b_{i^*j^*} = \begin{cases} -\frac{w_{i^*j^*}^\dagger \delta_{i^*j^*}^\dagger}{d_{i^*j^*}(\boldsymbol{\Phi H})} & (if\ i^* \neq j^*\ and\ d_{i^*j^*}(\boldsymbol{\Phi H}) \neq 0) \\ 0 & (if\ i^* \neq j^*\ and\ d_{i^*j^*}(\boldsymbol{\Phi H}) = 0) \end{cases}$$

$$b_{i^*i^*} = -\sum_{(j^*=1)\wedge(i^*\neq j^*)}^{n} b_{i^*j^*} \quad and$$

$$\boldsymbol{H} = (h_{s^\dagger t}) \quad h_{s^\dagger t} \in \mathbb{R}\ (s^\dagger = 1, 2, \cdots, 2k; t = 1, 2, \cdots, d).$$

Here, $L_M(\boldsymbol{Q}, \boldsymbol{H}, \boldsymbol{\Phi} |\ \boldsymbol{\Delta}^\dagger)$ is called as majorizing function.

Proof. *The left term of inequation (4.8) can be described as follows:*

$$\begin{aligned} &\frac{1}{2}\left\| \boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\dagger)) \right\|^2 \\ =&\frac{1}{2}\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger (\delta_{i^*j^*}^\dagger - d_{i^*j^*}(\boldsymbol{\Phi Q}))^2 \\ =&\frac{1}{2}\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger \delta_{i^*j^*}^{\dagger 2} + \frac{1}{2}\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger d_{i^*j^*}(\boldsymbol{\Phi Q})^2 - \sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger \delta_{i^*j^*}^\dagger d_{i^*j^*}(\boldsymbol{\Phi Q}) \end{aligned} \tag{4.9}$$

The second term of Eq. (4.9) is described as follows:

$$\begin{aligned} \frac{1}{2}\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger d_{i^*j^*}(\boldsymbol{\Phi Q})^2 &= \frac{1}{2}\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger \sum_{t=1}^{d} \boldsymbol{q}_{(t)}^T \boldsymbol{\Phi}^T (\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})^T \boldsymbol{\Phi q}_{(t)} \\ &= \frac{1}{2}\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger \mathrm{tr} \boldsymbol{Q}^T \boldsymbol{\Phi}^T (\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})^T \boldsymbol{\Phi Q} \\ &= \frac{1}{2}\mathrm{tr} \boldsymbol{Q}^T \boldsymbol{\Phi}^T \Big[\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^\dagger (\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})^T \Big] \boldsymbol{\Phi Q} \\ &= \frac{1}{2} tr \boldsymbol{Q}^T \boldsymbol{\Phi}^T \boldsymbol{V} \boldsymbol{\Phi Q} \end{aligned} \tag{4.10}$$

where $\boldsymbol{q}_{(t)}$ $(t = 1, 2, \cdots, d)$ *is column vector of* $\boldsymbol{Q}$.

For the third term of Eq. (4.9), we derive the inequality for $-d_{i^*j^*}(\boldsymbol{\Phi Q})$ *by using Cauchy-Schwarz inequality;*

$$\begin{aligned}
&\sum_{t=1}^{d}(\boldsymbol{\phi}_{i^*}^T\boldsymbol{q}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{q}_{(t)})(\boldsymbol{\phi}_{i^*}^T\boldsymbol{h}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{h}_{(t)}) \\
&\qquad\qquad \leq \Big[\sum_{t=1}^{d}(\boldsymbol{\phi}_{i^*}^T\boldsymbol{q}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{q}_{(t)})^2\Big]^{\frac{1}{2}}\Big[\sum_{t=1}^{d}(\boldsymbol{\phi}_{i^*}^T\boldsymbol{h}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{h}_{(t)})^2\Big]^{\frac{1}{2}} \\
&\qquad\qquad = d_{i^*j^*}(\boldsymbol{\Phi Q})d_{i^*j^*}(\boldsymbol{\Phi H}) \\
&\iff -d_{i^*j^*}(\boldsymbol{\Phi Q}) \leq -\frac{\sum_{t=1}^{d}(\boldsymbol{\phi}_{i^*}^T\boldsymbol{q}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{q}_{(t)})(\boldsymbol{\phi}_{i^*}^T\boldsymbol{h}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{h}_{(t)})}{d_{i^*j^*}(\boldsymbol{\Phi H})}, \qquad (4.11)
\end{aligned}$$

where $\boldsymbol{\phi}_{i^*}$ $(i^* = 1, 2, \cdots, 2n)$ *is row vecotr of* $\boldsymbol{\Phi}$ *and* $\boldsymbol{h}_{(t)}$ *is* t*th column vector of* $\boldsymbol{H}$ *. If* $\boldsymbol{Q} = \boldsymbol{H}$*, inequality (4.11) satisfies the equality from the property of Cauchy-Schwarz inequality. Therefore, inequality of the third term of Eq. (4.9) is derived as follows:*

$$\begin{aligned}
&-\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^{\dagger}\delta_{i^*j^*}^{\dagger}d_{i^*j^*}(\boldsymbol{\Phi Q}) \\
&\leq -\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^{\dagger}\delta_{i^*j^*}^{\dagger}d_{i^*j^*}(\boldsymbol{\Phi H})^{-1}\sum_{t=1}^{d}(\boldsymbol{\phi}_{i^*}^T\boldsymbol{q}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{q}_{(t)})(\boldsymbol{\phi}_{i^*}^T\boldsymbol{h}_{(t)} - \boldsymbol{\phi}_{j^*}^T\boldsymbol{h}_{(t)}) \\
&= -\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^{\dagger}\delta_{i^*j^*}^{\dagger}d_{i^*j^*}(\boldsymbol{\Phi H})^{-1}\sum_{t=1}^{d}\boldsymbol{q}_{(t)}^T\boldsymbol{\Phi}^T(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})^T\boldsymbol{\Phi h}_{(t)} \\
&= -\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^{\dagger}\delta_{i^*j^*}^{\dagger}d_{i^*j^*}(\boldsymbol{\Phi H})^{-1}tr\boldsymbol{Q}^T\boldsymbol{\Phi}^T(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})^T\boldsymbol{\Phi H} \\
&= -tr\boldsymbol{Q}^T\boldsymbol{\Phi}^T\Big[\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w_{i^*j^*}^{\dagger}\delta_{i^*j^*}^{\dagger}d_{i^*j^*}(\boldsymbol{\Phi H})^{-1}(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})(\boldsymbol{e}_{i^*} - \boldsymbol{e}_{j^*})^T\Big]\boldsymbol{\Phi H} \\
&= -tr\boldsymbol{Q}^T\boldsymbol{\Phi}^T\boldsymbol{B}(\boldsymbol{\Phi H})\boldsymbol{\Phi H} \qquad (4.12)
\end{aligned}$$

From (4.11) and (4.12), the majorizing function of the constrained Unfolding can be derived.

Next, updating formulas for $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$ are shown based on the majorizing function.

Proposition 4.2.4 *Updating formula of* $\boldsymbol{Q}$

Given $\boldsymbol{\Phi}$ *and* $\boldsymbol{H}$ *, updating formula of* $\boldsymbol{Q}$ *minimizing Eq. (4.8) is derived as follows:*

$$\boldsymbol{Q} = [\boldsymbol{\Phi}^T\boldsymbol{V}\boldsymbol{\Phi}]^{+}\boldsymbol{\Psi}^T\boldsymbol{B}(\Phi\boldsymbol{H})\boldsymbol{\Phi H} \qquad (4.13)$$

where $[\boldsymbol{\Phi}^T\boldsymbol{V}\boldsymbol{\Phi}]^{+}$ *is the Moore-Penrose inverse of* $\boldsymbol{\Phi}^T\boldsymbol{V}\boldsymbol{\Phi}$.

Proof. *The partial differential of Eq. (4.8) for* $\boldsymbol{Q}$ *is given as follows:*

$$\begin{aligned}\frac{\partial L_M(\boldsymbol{Q},\boldsymbol{H},\boldsymbol{\Phi}|\ \boldsymbol{\Delta}^{\dagger})}{\partial \boldsymbol{Q}} &= 2\boldsymbol{\Phi}^T\boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{Q} - 2\boldsymbol{\Phi}^T\boldsymbol{B}(\boldsymbol{\Phi}\boldsymbol{H})\boldsymbol{\Phi}\boldsymbol{H} = \boldsymbol{O}\\ \Longleftrightarrow \boldsymbol{\Phi}^T\boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{Q} &= \boldsymbol{\Phi}^T\boldsymbol{B}(\boldsymbol{\Phi}\boldsymbol{H})\boldsymbol{\Phi}\boldsymbol{H}\\ \Longleftrightarrow \boldsymbol{Q} &= [\boldsymbol{\Phi}^T\boldsymbol{V}\boldsymbol{\Phi}]^{+}\boldsymbol{\Phi}^T\boldsymbol{B}(\boldsymbol{\Phi}\boldsymbol{H})\boldsymbol{\Phi}\boldsymbol{H}\end{aligned}$$

Then, algorithm of estimating $\boldsymbol{Q}$ is shown based on the majorizing function.

Algorithm of estimating $\boldsymbol{Q}$

Step 0 Set iteration number $\beta \leftarrow 1$, initial value of ${}^{(\beta)}\boldsymbol{H}$, and threshold value $\varepsilon > 0$.

Step 1 Update ${}^{(\beta)}\boldsymbol{Q}$ based on Eq. (4.13)

Step 2 ${}^{(\beta)}\boldsymbol{Q}$ is substituted into ${}^{(\beta+1)}\boldsymbol{H}$ and $\beta \leftarrow \beta + 1$

Step 3 if $|{}^{(\beta)}L_M(\boldsymbol{Q},\boldsymbol{H},\boldsymbol{\Phi}|\ \boldsymbol{\Delta}^{\dagger}) - {}^{(\beta-1)}L_M(\boldsymbol{Q},\boldsymbol{H},\boldsymbol{\Phi}|\ \boldsymbol{\Delta}^{\dagger})| < \varepsilon$ is satisfied, stop, else back to step 1

where ${}^{(\beta)}\boldsymbol{H}$, ${}^{(\beta)}\boldsymbol{Q}$ and ${}^{(\beta)}L_M(\boldsymbol{Q},\boldsymbol{H},\boldsymbol{\Phi}|\ \boldsymbol{\Delta}^{\dagger})$ are $\boldsymbol{H}$, $\boldsymbol{Q}$ and a value of the objective function corresponding to β th iteration, respectively.

In this algorithm, $\boldsymbol{U}$ is estimated by each row vector of $\boldsymbol{U}$ although value of the objective function depends on the order of estimating row vector of $\boldsymbol{U}$.

Next, the way of updating ith row vector of $\boldsymbol{U}$ is shown.

Proposition 4.2.5 *Updating* $\boldsymbol{U}$

Given $\boldsymbol{X}^*$, $\boldsymbol{Y}^*$ *and* $\boldsymbol{U}$ *without* i^**th row vector of* $\boldsymbol{U}$*, if updating rule of* i*th row vector of* $\boldsymbol{U}$ *is used, values of the objective function of the constrained Unfolding does not increase.*

$$u_{i^*o} = \begin{cases} 1 & \Big((\forall \ell = 1,2,\cdots,k)(\gamma_{i^*o}(\boldsymbol{Q}) \leq \gamma_{i^*\ell}(\boldsymbol{Q}))\Big) \\ 0 & (others) \end{cases} \tag{4.14}$$

for $o = 1,2,\cdots,k$ *where*

$$\begin{aligned}\gamma_{i^*o}(\boldsymbol{Q}) = &\sum_{j\neq i^*}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{i^*j} - d_{o\ell}(\boldsymbol{X}^*,\boldsymbol{Y}^*))^2 + \sum_{j\neq i^*}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{ji^*} - d_{o\ell}(\boldsymbol{X}^*,\boldsymbol{Y}^*))^2\\ &+ d_{oo}(\boldsymbol{X}^*,\boldsymbol{Y}^*).\end{aligned}$$

and o $(o = 1,2,\cdots,k)$.

Proof. *We assume that object i^* satisfies $\gamma_{io}(\boldsymbol{Q}) \leq \gamma_{i\ell}(\boldsymbol{Q})$ for all $\ell = 1, 2, \cdots, k$. In the situation, the objective function of the constrained Unfolding can be decomposed as follows:*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k}\sum_{\ell=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2$$

$$= \sum_{i\neq i^*}\sum_{o=1}^{k} u_{io}\sum_{j=1}^{n}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 \tag{4.15}$$

$$+ \sum_{o=1}^{k} u_{i^*o}\sum_{j=1}^{n}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{i^*j} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 \tag{4.16}$$

Eq. (4.15) is described as follows:

$$\sum_{i\neq i^*}\sum_{o=1}^{k} u_{io}\sum_{j=1}^{n}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2$$

$$= \sum_{i\neq i^*}\sum_{o=1}^{k} u_{io}\sum_{j\neq i^*}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 \tag{4.17}$$

$$+ \sum_{i\neq i^*}\sum_{o=1}^{k} u_{io}\sum_{\ell=1}^{k} u_{i^*\ell}(\delta_{ii^*} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 \tag{4.18}$$

Eq. (4.18) can be described as follows:

$$\sum_{i\neq i^\dagger}\sum_{o=1}^{k} u_{io}\sum_{\ell=1}^{k} u_{i^\dagger\ell}(\delta_{ii^\dagger} - d_{ii^\dagger}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2$$

$$= \sum_{\ell=1}^{k} u_{i^\dagger\ell}\sum_{i\neq i^\dagger}\sum_{o=1}^{k} u_{io}(\delta_{ii^\dagger} - d_{ii^\dagger}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 \tag{4.19}$$

Eq. (4.16) can be described as follows:

$$\sum_{o=1}^{k} u_{i^*o}\sum_{j=1}^{n}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{i^*j} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2$$

$$= \sum_{o=1}^{k} u_{i^*o}\sum_{j\neq i^*}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{i^*j} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 \tag{4.20}$$

$$+ \sum_{o=1}^{k} u_{i^*o}\sum_{\ell=1}^{k} u_{i^*\ell}(\delta_{i^*i^*} - d_{o\ell}(\boldsymbol{X}^*, \boldsymbol{Y}^*))^2 \tag{4.21}$$

In short, the objective function of constrained Unfolding can be decomposed into Eq. (4.17), Eq.(4.19), Eq.(4.20) and Eq. (4.21). The sum of Eq.(4.19), Eq.(4.20), and Eq.(4.21) is minimized for arbitrarily o ($o = 1, 2, \cdots, k$) from the rule of Eq. (4.14). In addition, Eq.(4.17) is not affected by update becasue Eq. (4.17) does not include i^*. Therefore, the proposition is proved.

Algorithm of updating $\boldsymbol{U}$

Step 1 $i \leftarrow 1$

Step 2 Calcurate $\gamma_{io}(\boldsymbol{Q})$ for $o = 1, 2, \cdots, k$

Step 3 Update ith row vector of $\boldsymbol{U}$ based on thr rule of Eq. (4.14) among $\gamma_{io}(\boldsymbol{Q})$ $(o = 1, 2, \cdots, k)$. If the number of o such that $u_{io} = 1$ is over 1, one of o is selected randomly.

Step 4 If $i = n$ stop, otherwise, $i \leftarrow i + 1$ back to Step 2

However, the rule given by Eq. (4.14) is affected by the order of i.

4.3 Constrained slide-vector model based on CDS

This section comprises three parts, i.e., a description of the model and objective function for the constrained slide-vector model, the property, and the algorithm.

4.3.1 Model and objective function of the constrained slide-vector model

In this subsection, we show the model and objective function of the constrained slide-vector model. Furthermore, we introduce two types of descriptions for the objective function of the constrained slide-vector model; we then show their equivalence based on Heiser and Groenen (1997).

Definition 4.3.1 *Model of the constrained slide-vector model*
Let $\boldsymbol{\Delta} = (\delta_{ij})$ $\delta_{ij} \in \mathbb{R}_+$ $(i, j = 1, 2, \cdots, n)$ *be an asymmetric dissimilarity matrix,* $\boldsymbol{X}^* = (x^*_{ot})$ $x^*_{ot} \in \mathbb{R}$ $(o = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$ *and* $\boldsymbol{Z} = (z_{st})$ $z_{st} \in \mathbb{R}$ $(s = 1, 2, \cdots, m; t = 1, 2, \cdots, d; m \leq k)$ *be coordinates of clusters for objects and coordinates of slide-vectors, respectively, where* k *and* m *are the number of clusters for objects and slide-vectors, respectively. Here, the constrained slide-vector model is defined as follows:*

$$(\forall i, j = 1, 2, \cdots, n)(\exists! C_o; i \in C_o)(\exists! C_\ell; j \in C_\ell)(\exists! s = 1, 2, \cdots, m)$$
$$(\delta_{ij} = d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}) + \varepsilon_{ij})$$

where

$$d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}) = \|\boldsymbol{x}_o - (\boldsymbol{x}_\ell - \boldsymbol{z}_s)\| = \Big[\sum_{t=1}^{d} \Big(x^*_{ot} - (x^*_{\ell t} - z_{st})\Big)^2\Big]^{\frac{1}{2}},$$

C_o *is cluster* o *of objects,* $\boldsymbol{U} = (u_{i\ell})$, $u_{i\ell} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; \ell = 1, 2, \cdots, k)$ *is indicator matrix and* $\boldsymbol{\Psi} = (\psi_{\ell s})$, $\psi_{\ell s} \in \{0, 1\}$, $(\ell = 1, 2, \cdots, k; s = 1, 2, \cdots, m)$ *is indicator matrix for centroids, respectively, and* $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error.*

From definition 4.3.1, dissimilarity from object i to object j is interpreted by the distance from $\boldsymbol{x}_o$ to $\boldsymbol{x}_\ell - \boldsymbol{z}_s$. In the model, there are two types of indicator matrices; one indicator matrix $\boldsymbol{U}$ is introduced to classify objects, while the other $\boldsymbol{\Psi}$ is introduced to allocate slide-vectors to clusters. The constrained slide-vector model includes several slide-vectors, while the original slide-vector model includes only one slide-vector.

Next, we define the objective function of the constrained slide-vector model I.

Definition 4.3.2 *Objective function of constrained slide-vector model I*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$, the number of clusters of objects and slide-vectors, k and m, respectively, and the number of dimensions d, the objective function of constrained slide-vectors model is defined as follows:

$$L(\boldsymbol{X}^*, \boldsymbol{Z}, \boldsymbol{U}, \boldsymbol{\Psi} |\ \boldsymbol{\Delta}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k}\sum_{\ell=1}^{k}\sum_{s=1}^{m} u_{io}u_{j\ell}\psi_{\ell s}(\delta_{ij} - d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2$$

In the constrained slide-vector model, $\boldsymbol{X}^*$, $\boldsymbol{Z}$, $\boldsymbol{U}$ and $\boldsymbol{\Psi}$ are estimated such that the value of the objective function is minimized.

Next, we define the objective function of another constrained slide-vector model (i.e., slide-vector moel II), and prove the equivalence of the objective functions of these constrained slide-vector models.

Definition 4.3.3 *The objective function of constrained slide-vector model II*

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$, the number of clusters of objects k and slide-vectors m, respectively, and the number of low-dimensions d, the objective function of constrained slide-vector model II is defined as follows:

$$L(\boldsymbol{X}^*, \boldsymbol{Z}, \boldsymbol{U}, \boldsymbol{\Psi} |\ \boldsymbol{\Delta}) = \frac{1}{2}\left\| \boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\ddagger)) \right\|^2$$

where $\boldsymbol{W}^\dagger = (w^\dagger_{i^\dagger j^\dagger})$ $w^\dagger_{i^\dagger j^\dagger} \in \{0, 1\}$ $(i, j = 1, 2, \cdots, 2n)$ are weights for dissimilarities,

$$\boldsymbol{\Delta}^\dagger = \begin{bmatrix} \boldsymbol{O}_n & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}^T & \boldsymbol{O}_n \end{bmatrix}, \text{ and}$$

$$\boldsymbol{Q}^\ddagger = \begin{bmatrix} \boldsymbol{U}\boldsymbol{X}^* \\ \boldsymbol{U}\boldsymbol{X}^* - \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{Z} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{O}_{n,m} \\ \boldsymbol{U} & -\boldsymbol{U}\boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}^* \\ \boldsymbol{Z} \end{bmatrix}.$$

Here, $\boldsymbol{O}_{n,m} = (0)$ is n by m matrix. In the constrained slide-vector model, $\boldsymbol{X}^*$, $\boldsymbol{Z}$, $\boldsymbol{U}$ and $\boldsymbol{\Psi}$ are estimated such that the value of the objective function is minimized.

Next, we show the equivalence of the objective functions of the constrained slide-vector model I and II.

Proposition 4.3.1 *Equivalence of constrained slide-vector model I and II*

Given asymmetric dissimilarity matrix $\boldsymbol{\Delta}$*, indicator matrix of objects* $\boldsymbol{U}$ *and slide-vecotrs* $\boldsymbol{\Psi}$*, respectively, coordinates of cluster centroids* $\boldsymbol{X}^*$ *and slide-vectors* $\boldsymbol{Z}$*, and the number of low-dimensions* d*, the following property is satisfied;*

$$\begin{aligned}&\frac{1}{2}\left\|\boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\ddagger))\right\|^2 \\ &\qquad = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k}\sum_{\ell=1}^{k}\sum_{s=1}^{m} u_{io}u_{j\ell}\psi_{\ell s}(\delta_{ij} - d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2 \end{aligned} \tag{4.22}$$

Proof. *From the left term of Eq (4.22),*

$$\frac{1}{2}\left\|\boldsymbol{W}^\dagger \odot (\boldsymbol{\Delta}^\dagger - \boldsymbol{D}(\boldsymbol{Q}^\ddagger))\right\|^2 = \frac{1}{2}\sum_{i^*=1}^{2n}\sum_{j^*=1}^{2n} w^\dagger_{i^*j^*}(\delta^\dagger_{i^*j^*} - d_{i^*j^*}(\boldsymbol{Q}^\ddagger))^2. \tag{4.23}$$

From the definition of $\boldsymbol{W}^\dagger$ *and* $\boldsymbol{\Delta}^\dagger$*, Eq (4.23) is described as follows:*

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n}(\delta_{ij} - d_{ij}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z})))^2 \\ =&\sum_{i=1}^{n}\sum_{j=1}^{n}\Big(\delta_{ij} - \Big\|\sum_{o=1}^{k} u_{io}\boldsymbol{x}^*_o - \Big(\sum_{\ell=1}^{k} u_{j\ell}\Big(\boldsymbol{x}^*_\ell - \sum_{s=1}^{m}\psi_{\ell s}\boldsymbol{z}_s\Big)\Big)\Big\|\Big)^2 \end{aligned} \tag{4.24}$$

Here, for all i *and* j*, there exists* o^**,* ℓ^**, uniquely, such that* $u_{io^*} = 1$ *and* $u_{io} = 0$ $(o \neq o^*)$ *and* $u_{j\ell^*} = 1$ *and* $u_{j\ell} = 0$ $(\ell \neq \ell^*)$*, respectively, from the definition of indicator matrix. As the same way, for all* i*,* j *and* ℓ*, there exists* s^**, uniquely, such that* $\psi_{\ell s^*} = 1$ *and* $\psi_{\ell s^*} = 1$*. Therefore,* i *and* j *part of Eq. (4.23) is descirbed as follows:*

$$\begin{aligned}&\Big(\delta_{ij} - \Big\|\sum_{o=1}^{k} u_{io}\boldsymbol{x}^*_o - \Big(\sum_{\ell=1}^{k}\Big(u_{j\ell}\boldsymbol{x}^*_\ell - \sum_{s=1}^{m}\psi_{\ell s}\boldsymbol{z}_s\Big)\Big)\Big\|\Big)^2 \\ =&(\delta_{ij} - \|u_{io^*}\boldsymbol{x}^*_{o^*} - (u_{j\ell^*}\boldsymbol{x}^*_{\ell^*} - \psi_{\ell s^*}\boldsymbol{z}_{s^*})\|)^2 \\ =&(\delta_{ij} - \|\boldsymbol{x}^*_{o^*} - (\boldsymbol{x}^*_{\ell^*} - \boldsymbol{z}_{s^*})\|)^2 \\ =&\sum_{o=1}^{k}\sum_{\ell=1}^{k}\sum_{s=1}^{m} u_{io}u_{j\ell}\psi_{\ell s}(\delta_{ij} - \|\boldsymbol{x}^*_{o^*} - (\boldsymbol{x}^*_{\ell^*} - \boldsymbol{z}_{s^*})\|)^2. \end{aligned}$$

Therefore, the proposition is proved.

From Eq. (4.22), the objective function of the constrained slide-vector model is considered as special case of ordinal MDS, just as with constrained Unfolding.

4.3.2 Properties of the constrained slide-vector model

In this subsection, we show the decomposition of the objective function of the constrained slide-vector model. For the decomposition, just as with constrained

Unfolding, the objective function of the constrained slide-vector model can be decomposed into two parts, i.e., *Among cluster error SSQ* and *Lack of spatial fit*. The objective function of constrained slide-vector model does not include terms for *Within cluster error SSQ* and *homogeneity*.

Proposition 4.3.2 *Decomposition of the objective function of the constrained slide-vector model*

Given the objective function of the constrained slide-vector model, the objective function can be decomposed as follows:

$$\begin{aligned} L(\boldsymbol{X}^*, \boldsymbol{Z}, \boldsymbol{U}, \boldsymbol{\Psi} |\ \boldsymbol{\Delta}) =& \|\boldsymbol{\Delta} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))\|^2 \\ =& \|\boldsymbol{\Delta} - \boldsymbol{P}_U \boldsymbol{\Delta} \boldsymbol{P}_U\|^2 \qquad (4.25) \\ &+ \|\boldsymbol{P}_U \boldsymbol{\Delta} \boldsymbol{P}_U - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*, \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))\|^2 \qquad (4.26) \end{aligned}$$

where $(\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T\boldsymbol{\Delta}\boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}$ *are Sokal-Michener dissimilarity (Sokal and Michener, 1958) and* $\boldsymbol{P}_U = \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T$ *is projection matrix.*

Proof. *This proof can be conducted in the same way as Proposition 4.2.2.*

From the decomposition, we consider objective function of the constrained slide-vector model as simultaneous analysis of clustering for asymmetric dissimilarity data and slide-vector model for Sokal-Michener dissimilarities.

4.3.3 Algorithm of the constrained slide-vector model

In this subsection, we show the algorithm of the constrained slide-vector model. The parameters are estimated on a basis of ALS (Young et al., 1980). The flow of our proposed algorithm is described as follows.

Algorithm of the constrained slide-vector model

Step 0 Set k, m and d, and initial values of $\boldsymbol{X}^*$, $\boldsymbol{Z}$, $\boldsymbol{U}$, and $\boldsymbol{\Psi}$

Step 1 Update $\boldsymbol{X}^*$ and $\boldsymbol{Z}$, given $\boldsymbol{U}$ and $\boldsymbol{\Psi}$

Step 2 Update $\boldsymbol{U}$, given $\boldsymbol{\Psi}$, $\boldsymbol{X}^*$ and $\boldsymbol{Z}$

Step 3 Update $\boldsymbol{\Psi}$, given $\boldsymbol{X}^*$, $\boldsymbol{Z}$ and $\boldsymbol{U}$

Step 4 If stop condition is satisfied, stop the algorithm, else return to the Step 1

To update $\boldsymbol{X}^*$ and $\boldsymbol{Z}$, we adopt a majorization algorithm. thus deriving the majorizing function of the constrained slide-vector model.

Proposition 4.3.3 *Majorizing function of the constrained slide-vector model*

Given objective function of the constrained slide-vector model, the majorizing function of the constrained slide-vector model is given as follows:

$$\frac{1}{2}\left\|\boldsymbol{W}^{\dagger}\odot(\boldsymbol{\Delta}^{\dagger}-\boldsymbol{D}(\boldsymbol{Q}^{\ddagger}))\right\|^{2}$$
$$\leq\frac{1}{2}\eta_{\delta}^{2}+\frac{1}{2}tr\boldsymbol{Q}^{*T}\boldsymbol{\Phi}^{\dagger T}\boldsymbol{V}\boldsymbol{\Phi}^{\dagger}\boldsymbol{Q}^{*}-tr\boldsymbol{Q}^{*T}\boldsymbol{\Phi}^{\dagger T}\boldsymbol{B}(\boldsymbol{\Phi}^{\dagger}\boldsymbol{H})\boldsymbol{\Phi}^{\dagger}\boldsymbol{H}=L_{M}^{*}(\boldsymbol{Q}^{*},\boldsymbol{H},\boldsymbol{\Phi}^{\dagger}) \quad (4.27)$$

where

$$\eta_{\delta}^{2}=\sum_{i^{*}=1}^{2n}\sum_{j^{*}=1}^{2n}w_{i^{*}j^{*}}^{\dagger}\delta_{i^{*}j^{*}}^{\dagger 2}$$
$$\boldsymbol{\Phi}^{\dagger}=\begin{bmatrix}\boldsymbol{U} & \boldsymbol{O}_{n,m}\\ \boldsymbol{U} & -\boldsymbol{U}\boldsymbol{\Psi}\end{bmatrix},\quad \boldsymbol{Q}^{*}=\begin{bmatrix}\boldsymbol{X}^{*}\\ \boldsymbol{Z}\end{bmatrix},$$
$$\boldsymbol{V}=\sum_{i^{*}=1}^{2n}\sum_{j^{*}=1}^{2n}w_{i^{*}j^{*}}^{\dagger}(\boldsymbol{e}_{i^{*}}-\boldsymbol{e}_{j^{*}})(\boldsymbol{e}_{i^{*}}-\boldsymbol{e}_{j^{*}})^{T},$$
$$\boldsymbol{e}_{i^{*}}=(e_{i^{*}s^{*}})\quad e_{i^{*}s^{*}}=\begin{cases}1 & (i^{*}=s^{*})\\ 0 & (i^{*}\neq s^{*})\end{cases},\quad (i^{*},s^{*}=1,2,\cdots,2n),$$
$$\boldsymbol{B}(\boldsymbol{\Phi}^{\dagger}\boldsymbol{H})=(b_{i^{*}j^{*}})\ (i^{*},j^{*}=1,2,\cdots,2n)$$
$$b_{i^{*}j^{*}}=\begin{cases}-\frac{w_{i^{*}j^{*}}^{\dagger}\delta_{i^{*}j^{*}}^{\dagger}}{d_{i^{*}j^{*}}(\boldsymbol{\Phi}^{\dagger}\boldsymbol{H})} & (if\ i^{*}\neq j^{*}\ and\ d_{i^{*}j^{*}}(\boldsymbol{\Phi}^{\dagger}\boldsymbol{H})\neq 0)\\ 0 & (if\ i^{*}\neq j^{*}\ and\ d_{i^{*}j^{*}}(\boldsymbol{\Phi}^{\dagger}\boldsymbol{H})=0)\end{cases}$$
$$b_{i^{*}i^{*}}=-\sum_{(j^{*}=1)\wedge(i^{*}\neq j^{*})}^{2n}b_{i^{*}j^{*}}\quad and$$
$$\boldsymbol{H}=(h_{s^{\dagger}t})\quad h_{s^{\dagger}t}\in\mathbb{R}\ (s^{\dagger}=1,2,\cdots,(k+m);t=1,2,\cdots,d).$$

Proof. *This proposition is proved in the same way as Proposition 4.2.3.*

Next, we show updating formula for $\boldsymbol{X}^{*}$ and $\boldsymbol{Z}$ based on the majorizing function.

Proposition 4.3.4 *Updating formula of* $\boldsymbol{Q}^{\ddagger}$

Given $\boldsymbol{\Phi}^{*}$ *and* $\boldsymbol{H}$ *, updating formula of* $\boldsymbol{Q}^{\ddagger}$ *minimizing Eq. (4.27) is derived as follows:*

$$\boldsymbol{Q}^{\ddagger}=[\boldsymbol{\Phi}^{\dagger T}\boldsymbol{V}\boldsymbol{\Phi}^{\dagger}]^{+}\boldsymbol{\Phi}^{\dagger T}\boldsymbol{B}(\boldsymbol{\Phi}^{\dagger}\boldsymbol{H})\boldsymbol{\Phi}^{\dagger}\boldsymbol{H} \quad (4.28)$$

where $[\boldsymbol{\Phi}^{\dagger T}\boldsymbol{V}\boldsymbol{\Phi}^{\dagger}]^{+}$ *is the Moore-Penrose inverse of* $\boldsymbol{\Phi}^{\dagger T}\boldsymbol{V}\boldsymbol{\Phi}^{\dagger}$*.*

Proof. *This proposition is proved as the same way of Proposition 4.2.4.*

Then, the algorithm for estimating $\boldsymbol{Q}^{\ddagger}$ is shown as follows.

Algorithm of updating $\boldsymbol{Q}^{*}$

Step 0 Set iteration number $\beta\leftarrow 1$, initial value of ${}^{(\beta)}\boldsymbol{H}$, and threshold value $\varepsilon>0$.

Step 1 Update ${}^{(\beta)}\boldsymbol{Q}^*$ based on Eq. (4.13)

Step 2 ${}^{(\beta)}\boldsymbol{Q}^*$ is substituted into ${}^{(\beta+1)}\boldsymbol{H}$ and and $\beta \leftarrow \beta + 1$

Step 3 if $|{}^{(\beta)}L_M^*(\boldsymbol{Q}^*, \boldsymbol{H}, \boldsymbol{\Phi}^\dagger|\ \boldsymbol{\Delta}^\dagger) - {}^{(\beta-1)}L_M^*(\boldsymbol{Q}^*, \boldsymbol{H}, \boldsymbol{\Phi}^\dagger|\ \boldsymbol{\Delta}^\dagger)| < \varepsilon$ is satisfied, stop, else back to step 1

where ${}^{(\beta)}\boldsymbol{H}$, ${}^{(\beta)}\boldsymbol{Q}^*$ and ${}^{(\beta)}L_M(\boldsymbol{Q}^*, \boldsymbol{H}, \boldsymbol{\Phi}|\ \boldsymbol{\Delta}^\dagger)$ are $\boldsymbol{H}$, $\boldsymbol{Q}^*$ and a value of the objective function corresponding to β th iteration, respectively.

In this algorithm, $\boldsymbol{U}$ is estimated by each row vector of $\boldsymbol{U}$ although value of the objective function depends on the order of estimating row vector of $\boldsymbol{U}$.

Next, the way of updating ith row vector of $\boldsymbol{U}$ is shown.

Proposition 4.3.5 *Updating formula of* $\boldsymbol{U}$

Given $\boldsymbol{X}^*$, $\boldsymbol{Z}$, $\boldsymbol{\Psi}$ *and* $\boldsymbol{U}$ *without* i^**th row vector of* $\boldsymbol{U}$*, if updating rule of* i^**th row vector of* $\boldsymbol{U}$ *is used, values of the objective function of the constrained slide-vector model does not increase.*

$$u_{i^*o} = \begin{cases} 1 & \Big((\forall \ell = 1, 2, \cdots, k)(\gamma_{i^*o}(\boldsymbol{Q}^*, \boldsymbol{\Psi}) \leq \gamma_{i\ell}(\boldsymbol{Q}^*, \boldsymbol{\Psi}))\Big) \\ 0 & (others) \end{cases} \tag{4.29}$$

where

$$\begin{aligned} \gamma_{io}(\boldsymbol{Q}, \boldsymbol{\Psi}) = & \sum_{j \neq i} \sum_{\ell=1}^{k} \sum_{s=1}^{m} u_{j\ell} \psi_{\ell s} (\delta_{ij} - d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2 \\ & + \sum_{j \neq i} \sum_{\ell=1}^{k} \sum_{s=1}^{m} u_{j\ell} \psi_{\ell s} (\delta_{ji} - d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2 + d_{oo}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}). \end{aligned}$$

and o^* *is the index of cluster for objects.*

Proof. *This proof can be conducted as the same way of Proposition 4.2.5.*

Next, we show how to update $\boldsymbol{U}$.

Algorithm of updating $\boldsymbol{U}$

Step 1 $i \leftarrow 1$

Step 2 Calcurate $\gamma_{io}(\boldsymbol{Q}, \boldsymbol{\Psi})$ for $o = 1, 2, \cdots, k$

Step 3 Update ith row vector of $\boldsymbol{U}$ based on the rule of Eq.(4.29) among $\gamma_{io}(\boldsymbol{Q}, \boldsymbol{\Psi})$ $(o = 1, 2, \cdots, k)$. If the number of o such that $u_{io} = 1$ is over 1, one of o is selected randomly.

Step 4 If $i = n$ stop, otherwise, $i \leftarrow i + 1$ and back to Step 2

Note that the rule of Eq. (4.29) is affected by the order of i.

Next, we show the update rule for $\boldsymbol{\Psi}$. In this algorithm, $\boldsymbol{\Psi}$ is estimated by each row vector of $\boldsymbol{\Psi}$ although values of the objective function depend on the order of estimating row vector of $\boldsymbol{\Psi}$.

Proposition 4.3.6 *Updating formula of* $\boldsymbol{\Psi}$

Given $\boldsymbol{X}^*$, $\boldsymbol{Z}$, $\boldsymbol{U}$ *and* $\boldsymbol{\Psi}$ *without* ℓ^**th vector of* $\boldsymbol{\Psi}$*, if following updating rule of* ℓ^**th row vector of* $\boldsymbol{\Psi}$ *is used, values of the objective function of the constrained slide-vector model does not increase.*

$$\psi_{\ell^* s} = \begin{cases} 1 & \Big((\forall s^* = 1, 2, \cdots, m)(\kappa_{\ell s}(\boldsymbol{Q}, \boldsymbol{U}) \leq \kappa_{\ell s^*}(\boldsymbol{Q}, \boldsymbol{U}))\Big) \\ 0 & (others) \end{cases} \tag{4.30}$$

where

$$\kappa_{\ell s}(\boldsymbol{Q}, \boldsymbol{U}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2$$

and s^* *is the index of cluster of centrodis.*

Proof. *In the situation, the objective function of the constrained slide-vector model can be decomposed as follows:*

$$\begin{aligned} &\sum_{\ell=1}^{k}\sum_{s=1}^{m}\psi_{\ell s}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2 \\ =&\sum_{\ell\neq\ell^\dagger}^{k}\sum_{s=1}^{m}\psi_{\ell s}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - d_{o\ell s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2 \qquad (4.31) \\ &+\sum_{s=1}^{m}\psi_{\ell^\dagger s}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k} u_{io}u_{j\ell^\dagger}(\delta_{ij} - d_{o\ell^\dagger s}(\boldsymbol{X}^*, \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))^2 \qquad (4.32) \end{aligned}$$

In short, the objective function of constrained slide-vector model can be decomposed into Eq. (4.31) and Eq. (4.32). From the rule of Eq. (4.30), Eq. (4.32) is minimized, and Eq. (4.31) is not affected by updating ℓ^**. Therefore, the proposition is satisfied.*

Algorithm of updating $\boldsymbol{\Psi}$

Step 1 $\ell \leftarrow 1$

Step 2 Calcurate $\kappa_{\ell s}(\boldsymbol{Q}, \boldsymbol{U})$ for all $s = 1, 2, \cdots, m$

Step 3 Update ℓth row vector of $\boldsymbol{\Psi}$ based on the rule of Eq. (4.30) among $\kappa_{\ell s}(\boldsymbol{Q}, \boldsymbol{U})$ $(s = 1, 2, \cdots, m)$. If the number of s such that $\psi_{\ell s} = 1$ is over 1, one of s is selected randomly.

Step 4 If $\ell = k$ stop, otherwise, $\ell \leftarrow \ell + 1$ and back to Step 2

4.4 Relations between constrained methods and existing methods

The constrained slide-vector model is considered to be a generalization of the slide-vector model and Unfolding for asymmetric dissimilarity data. Furthermore, in some situations, the objective functions of the constrained slide-vector model and constrained Unfolding become equivalent. In this section, we show the following three relations: (1) the relation between constrained Unfolding and the constrained slide-vector model; (2) the relation between Unfoldng and the constrained slide-vecotr model; and (3) the relation between the constrained slide-vector model and the slide-vector model.

First, we show the relation between constrained Unfoldng and the constrained slide-vecotr model.

Theorem 4.4.1 *Relation between constrained Unfolding and slide-vector model*
If $m = k$ for the objective function of constrained slide-vector model, these objective functions of constrained Unfolding and of constrained slide-vetor model are equivalent for the asymmetric dissimilarity data.

Proof. *From the proposition 4.2.1 and the proposition 4.3.1, the objective function of definition 4.2.3 and definition 4.3.3 will be compared. The differences between definition 4.2.3 and definition 4.3.3 is depending on models of coordinates matrices.*
Therefore, the following property will be shown.

$$\boldsymbol{Y} \in \mathcal{CU} = \left\{ \boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Y}^* \middle| \ \boldsymbol{U} \in \mathcal{S}_I(n,k),\ \boldsymbol{Y}^* \in \mathbb{R}^{k\times d} \right\} \tag{4.33}$$

$$\Longleftrightarrow \boldsymbol{Y} \in \mathcal{CS} = \left\{ \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}) \middle| \ \boldsymbol{U} \in \mathcal{S}_I(n,k),\ \boldsymbol{\Psi} \in \mathcal{S}_I(k,m),\ \boldsymbol{X}^* \in \mathbb{R}^{k\times d},\ \boldsymbol{Z} \in \mathbb{R}^{m\times d} \right\} \tag{4.34}$$

where

$$\mathcal{S}_I(n,k) = \left\{ \boldsymbol{U} \middle| \ \boldsymbol{U} = (u_{io}),\ \sum_{o=1}^{k} u_{io} = 1 (i = 1, 2, \cdots, n),\ 0 < \sum_{i=1}^{n} u_{io} < n (1 < o \leq k) \right\}.$$

First Eq. (4.33) is assumed and Eq.(4.34) will be proved.
We assumed $\boldsymbol{Y} \in \mathcal{CU}$ and $m = k$. From Eq. (4.33),

$$(\forall \boldsymbol{Y} \in \mathcal{CU})(\exists \boldsymbol{U} \in \mathcal{S}_I(n,k))(\exists \boldsymbol{Y}^* \in \mathbb{R}^{k\times d})(\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Y}^*).$$

Here, if $k = m$, $\boldsymbol{\Psi} \in \mathcal{S}_I(k,m)$ exists such that $\boldsymbol{\Psi} = \boldsymbol{I}$. Therefore, the following is satrisfied:

$$(\forall \boldsymbol{Y}^* \in \mathbb{R}^{k\times d})(\forall \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\exists \boldsymbol{\Psi} \in \mathcal{S}_I(k,k))(\exists \boldsymbol{Z} \in \mathbb{R}^{k\times d})(\boldsymbol{Y}^* = \boldsymbol{X}^* - \boldsymbol{I}\boldsymbol{Z} = \boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z})$$

because $\boldsymbol{Y}^*, \boldsymbol{X}^*$ and $\boldsymbol{Z} \in \mathbb{R}^{k\times d}$. Then $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Y}^* = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}) \in \mathcal{CS}$.
Next, Eq.(4.34) is assumed and Eq. (4.33) will be proved.

We assumed $\boldsymbol{Y} \in \mathcal{CS}$. *From Eq. (4.34)*

$$(\forall \boldsymbol{Y} \in \mathcal{CS})(\exists \boldsymbol{U} \in \mathcal{S}_I(n,k))(\exists \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\exists \boldsymbol{\Psi} \in \mathcal{S}_I(k,m))(\exists \boldsymbol{Z} \in \mathbb{R}^{m\times d}) \\ (\boldsymbol{Y} = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))$$

Here,

$$(\forall \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}) \in \mathcal{CS})(\exists \boldsymbol{Y}^* \in \mathbb{R}^{k\times m})(\boldsymbol{U}(\boldsymbol{X} - \boldsymbol{\Psi}\boldsymbol{Z}) = \boldsymbol{U}\boldsymbol{Y}^*).$$

Therefore, $\boldsymbol{U}\boldsymbol{Y}^* \in \mathcal{CU}$ *is satisfied.*

Next, relation between Unfolding and the constrained slide-vector model is shown.

Theorem 4.4.2 *Relation between Unfolding and the constrained slide-vector model*

If $m = k = n$ *for the objective function of the constrained slide-vector model, these objective functions of Unfolding and the constrained slide-vector model are equivalent for the asymmetric dissimilarity data.*

Proof. *In the same way as proof for theorem 4.4.1, the objective function of Unfolding and the constrained slide-vector model can be described based on the same style of the objective functions. Therefore, the following properties will be shown.*

$$\Big[\boldsymbol{X}^\dagger \in \mathcal{U}_X = \{\boldsymbol{X} = \boldsymbol{X}^\dagger |\ \boldsymbol{X} \in \mathbb{R}^{n\times d}\} \\ \iff \boldsymbol{X}^\dagger \in \mathcal{CS}_X = \{\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^* |\ \boldsymbol{X}^* \in \mathbb{R}^{k\times d}, \boldsymbol{U} \in \mathcal{S}_I(n,k)\}\Big] \tag{4.35}$$

$$\wedge \Big[\boldsymbol{Y}^\dagger \in \mathcal{U}_Y = \{\boldsymbol{Y}^\dagger = \boldsymbol{Y} |\ \boldsymbol{Y} \in \mathbb{R}^{n\times d}\} \\ \iff \boldsymbol{Y}^\dagger \in \mathcal{CS}_Y = \{\boldsymbol{Y}^\dagger = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}) |\ \boldsymbol{U} \in \mathcal{S}_I(n,k), \\ \boldsymbol{\Psi} \in \mathcal{S}_I(k,m),\ \boldsymbol{X}^* \in \mathbb{R}^{k\times d},\ \boldsymbol{Z} \in \mathbb{R}^{m\times d}\}\Big] \tag{4.36}$$

First, Eq. (4.35) will be shown.

We assumed $\boldsymbol{X}^\dagger \in \mathcal{U}_X$. *If* $n = k$, $\boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k)$ *such that* $\boldsymbol{I} = \boldsymbol{U}$. *Therefore,*

$$(\forall \boldsymbol{X}^\dagger \in \mathcal{U}_X = \mathbb{R}^{n\times d})(\exists \boldsymbol{U} \in \mathcal{S}_I(n,k))(\exists \boldsymbol{X}^* \in \mathbb{R}^{n\times d})(\boldsymbol{X}^\dagger = \boldsymbol{I}\boldsymbol{X}^* = \boldsymbol{U}\boldsymbol{X}^*).$$

and $\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^* \in \mathcal{CS}_X$.

We assumed $\boldsymbol{X}^\dagger \in \mathcal{CS}_X$. *The following property is satisfied:*

$$(\forall \boldsymbol{X}^\dagger \in \mathcal{CS}_X)(\exists \boldsymbol{U} \in \mathcal{S}_I(n,k))(\exists \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^*), \text{ and} \\ (\forall \boldsymbol{U}\boldsymbol{X}^* \in \mathcal{CS}_X)(\exists \boldsymbol{X} \in \mathbb{R}^{n\times d})(\boldsymbol{U}\boldsymbol{X}^* = \boldsymbol{X}).$$

Therefore, $\boldsymbol{U}\boldsymbol{X}^* \in \mathcal{U}_X$.

Next, Eq. (4.36) will be proved.

We assumed $\boldsymbol{Y}^\dagger \in \mathcal{U}_Y$. *If* $n = k = m$, *there exists that* $\boldsymbol{U} \in \mathcal{S}_I(n,k)$, *and* $\boldsymbol{\Psi} \in \mathcal{S}_I(k,m)$ *such that* $\boldsymbol{I} = \boldsymbol{U}$ *and* $\boldsymbol{I} = \boldsymbol{\Psi}$, *respectively. Therefore, following property is satified.*

$$
\begin{aligned}
&(\forall \boldsymbol{Y}^\dagger \in \mathcal{U}_Y)(\exists \boldsymbol{U} \in \mathcal{S}_I(n,k))(\exists \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\exists \boldsymbol{\Psi} \in \mathcal{S}_I(k,m))(\exists \boldsymbol{Z} \in \mathbb{R}^{m\times d}) \\
&(\boldsymbol{Y}^\dagger = \boldsymbol{X}^* - \boldsymbol{Z} = \boldsymbol{I}\boldsymbol{X}^* - \boldsymbol{I}\boldsymbol{I}\boldsymbol{Z} = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z})).
\end{aligned}
$$

Then, $\boldsymbol{X}^\dagger \in \mathcal{CS}_Y$.

We assumed $\boldsymbol{X}^\dagger \in \mathcal{CS}_Y$. *The following property is satisfied from the definition of* $\mathcal{CS}_Y$*:*

$$
\begin{aligned}
&(\forall \boldsymbol{Y}^\dagger \in \mathcal{CS})(\exists \boldsymbol{U} \in \mathcal{S}_I(n,k))(\exists \boldsymbol{\Psi} \in \mathcal{S}_I(k,m))(\exists \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\exists \boldsymbol{Z} \in \mathbb{R}^{m\times d}) \\
&(\boldsymbol{Y}^\dagger = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z})) \quad \text{and} \\
&(\forall \boldsymbol{U} \in \mathcal{S}_I(n,k))(\forall \boldsymbol{\Psi} \in \mathcal{S}_I(k,m))(\forall \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\forall \boldsymbol{Z} \in \mathbb{R}^{m\times d})(\exists \boldsymbol{Y} \in \mathbb{R}^{n\times d}) \\
&(\boldsymbol{Y} = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z})).
\end{aligned}
$$

Therefore $\boldsymbol{Y} \in \mathcal{U}_Y$, *and the theorem are proved.*

Finally, we show the relation between the constrained slide-vector model and the slide-vector model.

Theorem 4.4.3 *Relation between slide-vector model and constrained slide-vector model*

If $m = 1$ *and* $n = k$, *objective functions of constrained slide-vector model and slide-vector model are equivalent.*

Proof. *The difference between definition 3.2.6 and definition 4.3.3 is depending on models of coordinates matrix. Therefore the following properties will be proved.*

$$
\begin{aligned}
&\Big[\boldsymbol{X}^\dagger \in \mathcal{CS}_X = \{\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^* |\ \boldsymbol{U} \in \mathcal{S}_I(n,k),\ \boldsymbol{X}^* \in \mathbb{R}^{k\times d}\} \\
\Longleftrightarrow &\boldsymbol{X}^\dagger \in \mathcal{S}_X = (\boldsymbol{X}^\dagger = \boldsymbol{X} |\ \boldsymbol{X} \in \mathbb{R}^{n\times d})\Big] \wedge \qquad (4.37) \\
&\Big[\boldsymbol{Y}^\dagger \in \mathcal{CS}_Y = \{\boldsymbol{Y}^\dagger = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}) |\ \boldsymbol{U} \in \mathcal{S}_I(n,k),\ \boldsymbol{X}^* \in \mathbb{R}^{k\times d}, \boldsymbol{\Psi} \in \mathcal{S}_I(k,m), \\
&\qquad\qquad \boldsymbol{Z} \in \mathbb{R}^{m\times d}\} \\
\Longleftrightarrow &\boldsymbol{Y}^\dagger \in \mathcal{S}_Y = \{\boldsymbol{Y}^\dagger = \boldsymbol{X} - \mathbf{1}\boldsymbol{z}^T |\ \boldsymbol{X} \in \mathbb{R}^{n\times d}, \boldsymbol{z} \in \mathbb{R}^d\}\Big] \qquad (4.38)
\end{aligned}
$$

Eq. (4.37) can be proved as the same way of proof for Eq. (4.35) if $n = k$. *Next, Eq. (4.38) will be proved.*

We assumed $\boldsymbol{Y}^\dagger \in \mathcal{CS}_Y$, $n = k$, *and* $m = 1$.

$$
\begin{aligned}
&(\forall \boldsymbol{Y}^\dagger \in \mathcal{CS}_Y)(\exists \boldsymbol{U} \in \mathcal{S}_I(n,k))(\exists \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\exists \boldsymbol{\Psi} \in \mathcal{S}_I(k,m))(\exists \boldsymbol{Z} \in \mathbb{R}^{m\times d}) \\
&(\boldsymbol{Y}^\dagger = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z})) \quad \textit{and} \\
&(\forall \boldsymbol{U} \in \mathcal{S}_I(n,k))(\forall \boldsymbol{\Psi} \in \mathcal{S}_I(k,m))(\forall \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\forall \boldsymbol{Z} \in \mathbb{R}^{m\times d})(\exists \boldsymbol{Y} \in \mathbb{R}^{n\times d}) \\
&(\boldsymbol{Y} = \boldsymbol{U}(\boldsymbol{X}^* - \boldsymbol{\Psi}\boldsymbol{Z}))
\end{aligned}
$$

Then, $\boldsymbol{Y}^{\dagger} = \boldsymbol{Y} \in \mathcal{S}_Y$.

We assumed $\boldsymbol{Y}^{\dagger} \in \mathcal{S}_Y$.

Here, if $n = k$*, there exists* $\boldsymbol{U} \in \mathcal{S}_i(n, k)$ *such that* $\boldsymbol{U} = \boldsymbol{I}$*. In addition to that, if* $m = 1$*,* $\boldsymbol{\Psi} \in \mathcal{S}_I(k, m)$ *must be* $\boldsymbol{\Psi} = \mathbf{1}_k$ *from the defitnion of indicator matrix. Therefore, if* $n = k$ *and* $m = 1$*,*

$$(\forall \boldsymbol{Y} \in \mathbb{R}^{n \times d})(\exists \boldsymbol{Z} \in \mathbb{R}^{m \times d})(\exists \boldsymbol{X}^* \in \mathbb{R}^{n \times d})(\exists \boldsymbol{U} \in \mathcal{S}_I(n, k))(\exists \boldsymbol{\Psi} \in \mathcal{S}_I(k, m))$$
$$(\boldsymbol{Y} = \boldsymbol{X}^* - \mathbf{1}\boldsymbol{z}^T = \boldsymbol{I}(\boldsymbol{X}^* - \mathbf{1}\boldsymbol{Z}))$$

Then $\boldsymbol{Y}^{\dagger} \in \mathcal{CS}_Y$.

From theorem 4.4.1, theorem 4.4.2, and theorem 4.4.3, the constrained slide-vector model can be considered a generalization of these Unfolding type methods for asymmetric dissimilarity data; however, the constrained slide-vector model does not include CDU (Vera et al., 2013) for asymmetric dissimilarity data.

Chapter 5

Constrained analysis of asymmetric data based on decomposition

5.1 Background of constrained asymmetric MDS based on decomposition

AMDS is a method for estimating coordinates of objects in low-dimensions from asymmetric dissimilarity data, where asymmetric dissimilarity is defined as dissimilarity from object i to object j not necessarily being the same dissimilarity from object j to object i. In AMDS, asymmetries between objects are represented in estimated low-dimensions (Borg and Groenen, 2005; Chino, 2012; Saito and Yadohisa, 2005). Asymmetric dissimilarity data exists in a wide variety of areas; therefore, it is important to interpret asymmetries between objects using AMDS.

Chino (2012) reviewed AMDS in a narrower sense and classified descriptive AMDS models into three methods, in particular, augmented models, non-metric distance models and extend distance models (Chino and Okada, 1996; Chino, 1997). In short, various AMDS methods have been proposed to describe asymmetries in low-dimensions. Among them, objective functions of the hill-climbing model (Borg and Groenen, 2005) and radius model (Okada and Imaizumi, 1987) can be decomposed into symmetric and skew-symmetric parts. Symmetric parts of these objective functions are equivalent to ordinal MDS for the symmetric part of asymmetric dissimilarity data. In contrast, the skew-symmetric parts of these objective functions are modelized in different manner for skew-symmetric part of asymmetric dissimilarity data, respectively. Therefore, these methods can be considered as extended models of ordinal MDS. The interpretation of the asymmetries from described by these methods is rather straighforward.

In the hill-climbing model, a vector called the slope-vector is used to describe asymmetries. Therefore, the representation here is similar to that of the slide-vector

model (Zieltman and Heiser, 1993); however, the interpretation is different. In hill climbing model, skew-symmetry is represented by notion of non-distance model although, in the slide-vector model, that is represented by notion of augmented distance model. If inner product between the slope vector and the difference vector between two objects tends to be relatively large , the asymmetric relation between these two objects is considered to be relatively large. On the other hand, if the inner product tends to be relatively small, the asymmetric relation between these two objects is considered to be a relatively small. In the radius model, coordinates and radii are used to represent the symmetric part and skew-symmetric parts. if one object has a relatively large radius length and the other object has a relatively small radius length, the relation is considered to be a large asymmetry in this model.

Although these methods are useful for asymmetric dissimilarity data, it becomes increasingly difficult to interpret the results of these methods as the asymmetric dissimilarity data grows. To overcome this problem for the interpretation, we propose two types of simultaneous methods for AMDS and clustering, i.e., the constrained hill-climbing model and the constrained radius model; these methods are proposed on a basis of cluster difference scaling (CDS) (Heiser, 1993; Heiser and Groenen, 1997). In the constrained hill-climbing and constrained radius models, relations for asymmetries between clusters are represented. These objective functions can be decomposed into symmetric and skew-symmetric parts, the symmetric part being equivalent to the objective fucntion of CDS for the symmetric part of the asymmetric dissimilarity data. Here, the symmetric and skew-symmetric parts can be further decomposed using the Sokal-Michener dissimilarities (Sokal and Michener, 1958) of the symmetric and skew-symmetric parts, respectively. Therefore, the interpretations of these objective functions become relatively straightforward. For the constrained radius model, the radii are also classified on the basis of the notion of parsimonious models (Tanioka and Yadohisa, 2016).

For the remainder of this chapter, the model, objective function and algorithm of the constrained hill-climbing model and constrained radius model are shown in section 4.2 and section 4.3, respectively. Finally, in section 4.4, we show the relations between the hill-climbing and constrained hill-climbing models, and the radius and constrained radius models.

5.2 Constrained hill-climbing model based on CDS

This section comprise three parts, i.e., a description of the model and objective function for the constrained hill-climbing model, the property and the algorithm.

5.2.1 Model and objective function of the constrained hill-climbing model

In this section, we show the model and objective function of the constrained hill-climbing model. Furthermore, we introduce two types of descriptions for the

objective functions of the constrained hill-climbing model; furthermore, we show the equivalence based on Heiser and Groenen (1997).

Definition 5.2.1 *Model of constrained hill-climbing model*

Let $\boldsymbol{\Delta}$ *be an asymmetric dissimilarity matrix,* $\boldsymbol{X}^* = (x^*_{ot})$ $x^*_{ot} \in \mathbb{R}$ $(o = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$ *and* $\boldsymbol{v} = (v_t)$ $v_t \in \mathbb{R}$ $(t = 1, 2, \cdots, d)$ *be coordinates of clusters and a slope vector in* d *dimensions, respectively, where* k *is the number of clusters for symmetric part. Here, the model of constrained hill-climbing model is defined as follows:*

$$
\begin{aligned}
&(\forall i, j = 1, 2, \cdots, n)(\exists! C_o; i \in C_o)(\exists! C_\ell; j \in C_\ell) \\
&\delta_{ij} = \begin{cases} d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}^*_o - \boldsymbol{x}^*_\ell)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1} + \varepsilon_{ij} & (o \neq \ell) \\ 0 & (o = \ell) \end{cases}
\end{aligned}
$$

where

$$
d_{o\ell}(\boldsymbol{X}^*) = \|\boldsymbol{x}^*_o - \boldsymbol{x}^*_\ell\| = \Big[\sum_{t=1}^{d} (x^*_{ot} - x^*_{\ell t})^2 \Big]^{\frac{1}{2}}
$$

C_o *is cluster* o *of objects, and* $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error.*

Next, we define the objective function of the constrained hill-climbing model I.

Definition 5.2.2 *Objective function of constrained hill-climbing model I*

Given an asymmetric dissimilarity data $\boldsymbol{\Delta}$*, the number of clusters for objects* k*, and the number of low-dimensions* d*, the objective function of the constrained hill-climbing model is defined as follows:*

$$
L(\boldsymbol{X}^*, \boldsymbol{v}, \boldsymbol{U} | \, \boldsymbol{\Delta}) = \sum_{(i,j) \in \Omega} \sum_{o=1}^{k} \sum_{\ell=1}^{k} u_{io} u_{j\ell} \Big(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}^*_o - \boldsymbol{x}^*_\ell)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}) \Big)^2
$$

where

$$
\Omega = \{(i, j) | \, \sum_{o=1}^{k} \sum_{\ell=1}^{k} u_{io} u_{j\ell} d_{o\ell}(\boldsymbol{X}^*) \neq 0 \ (i, j = 1, 2, \cdots, n)\}
$$

and $\boldsymbol{U} = (u_{io}), u_{io} \in \{0, 1\}$, $(i = 1, 2, \cdots, n; o = 1, 2, \cdots, k)$ *is indicator matrix. In the constrained hill-climbing model,* $\boldsymbol{X}^*$*,* $\boldsymbol{v}$ *and* $\boldsymbol{U}$ *are estimated such that the value of the objective function is minimized.*

Next, we define another objective function of the constrained hill-climbing model II, and show the equivalence of these objective functions.

Definition 5.2.3 *Objective function of constrained hill-climbing model II*

Given an asymmetric dissimilarity data $\boldsymbol{\Delta}$*, the number of clusters* k *and the number of low-dimensions* d*, the objective function of the constrained hill-climbing model II is defined as follows:*

$$L(\boldsymbol{X}^*, \boldsymbol{v}, \boldsymbol{U} |\ \boldsymbol{\Delta}) = \sum_{(i,j)\in\Omega^*} \Big(\delta_{ij} - \Big[d_{ij}(\boldsymbol{U}\boldsymbol{X}^*) + (\sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)^{-1}\Big]\Big)^2$$

where

$$d_{ij}(\boldsymbol{U}\boldsymbol{X}^*) = \Big\| \sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_o^* \Big\| \ (i, j = 1, 2, \cdots, n),$$

and

$$\Omega^* = \{(i,j) |\ d_{ij}(\boldsymbol{U}\boldsymbol{X}) \neq 0 \ (i, j = 1, 2, \cdots, n)\}$$

Next, we prove the equivalence of these objective functions.

Proposition 5.2.1 *Equivalence of constrained hill-climb model*

Given an asymmetric dissimilarity matrix $\boldsymbol{\Delta}$*, an indicator matrix of objects* $\boldsymbol{U}$*, coordinates of cluster centorids* $\boldsymbol{X}^*$*, a slope-vector* $\boldsymbol{v}$*, and the number of low-dimensions* d*, the following property is satisfied;*

$$\begin{aligned}
&\sum_{(i,j)\in\Omega} \sum_{o=1}^{k} \sum_{\ell=1}^{k} u_{io}u_{j\ell} \Big(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o - \boldsymbol{x}_\ell)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1})\Big)^2 \\
&= \sum_{(i,j)\in\Omega^*} \Big(\delta_{ij} - \Big[d_{ij}(\boldsymbol{U}\boldsymbol{X}^*) + (\sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)^{-1}\Big]\Big) \qquad (5.1)
\end{aligned}$$

Proof. *From the right term of Eq. (5.1),*

$$\begin{aligned}
&\sum_{(i,j)\in\Omega^*} \Big(\delta_{ij} - \Big[d_{ij}(\boldsymbol{U}\boldsymbol{X}^*) + (\sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)^{-1}\Big]\Big)^2 \\
&= \sum_{(i,j)\in\Omega^*} \Big(\delta_{ij} - \Big[\Big\| \sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_\ell^* \Big\| + \\
&\Big(\sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T \boldsymbol{v} \Big\| \sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_\ell^* \Big\|^{-1}\Big)^2. \qquad (5.2)
\end{aligned}$$

Here, for all i *and* j*, there exists* o^**,* ℓ^**, uniquely, such that* $u_{io^*} = 1$ *and* $u_{io} = 0$ $(o \neq o^*)$ *and* $u_{j\ell^*} = 1$ *and* $u_{j\ell} = 0$ $(\ell \neq \ell^*)$*, respectively, from the definition of indicator matrix. Therefore,* i *and* j *part of Eq. (5.2) is as follows;*

$$\begin{aligned}
&\Big(\delta_{ij} - \Big[\|\boldsymbol{x}_{o^*}^* - \boldsymbol{x}_{\ell^*}^*\| + (\boldsymbol{x}_{o^*}^* - \boldsymbol{x}_{\ell^*}^*)^T \boldsymbol{v} \|\boldsymbol{x}_{o^*}^* - \boldsymbol{x}_{\ell^*}^*\|^{-1}\Big]\Big)^2 \\
&= \sum_{o=1}^{k} \sum_{\ell=1}^{k} u_{io}u_{j\ell} \Big(\delta_{ij} - \Big[\|\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*\| + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} \|\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*\|^{-1}\Big]\Big)^2
\end{aligned}$$

Therefore, the proposition is proved.

5.2.2 Properties of the constrained hill-climbing model

In this subsection, we introduce three types of decompositions of the constrained hill-climbing model. First, we show the decomposition of the objective function of the constrained hill-climbing model into the symmetric and skew-symmetric parts of the objective function. From the decomposition, the interpretation of the slope vector becomes clear as the same way of the hill climbing model.

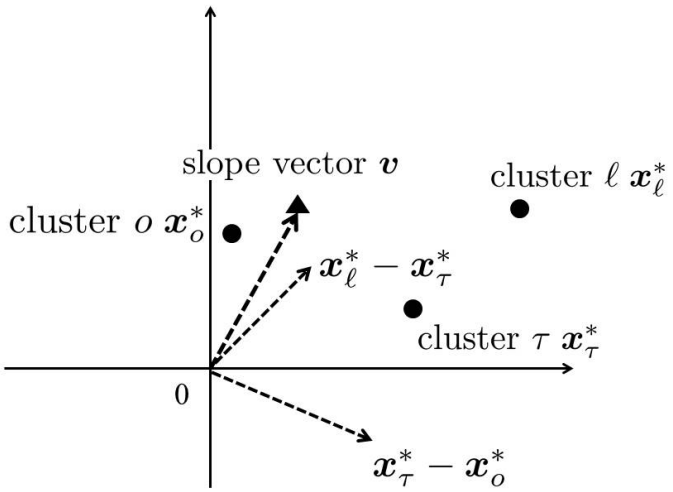

Figure 5.1: Example of the constrained hill-climbing model

See Figure 5.1. In Figure 5.1, the asymmetric relation between coordinates of cluster τ, $\boldsymbol{x}^*_\tau$ and cluster o $\boldsymbol{x}^*_o$ is smaller than tha between coordinates of cluster τ, $\boldsymbol{x}^*_\tau$ and cluster ℓ, $\boldsymbol{x}^*_\ell$ because inner product between slope vector $\boldsymbol{v}$ and $\boldsymbol{x}^*_{tau} - \boldsymbol{x}^*_o$ is smaller than that between $\boldsymbol{v}$ and $\boldsymbol{x}^*_\ell - \boldsymbol{x}^*_\tau$. In addition, in the example, distance from $\boldsymbol{x}^*_\tau$ to $\boldsymbol{x}^*_\ell$ is closer than that from $\boldsymbol{x}^*_\ell$ to $\boldsymbol{x}^*_\tau$ from the direction of $\boldsymbol{v}$.

Second, the symmetric part of the objective function can be decomposed into four parts in the same manner as that of CDS. Third, the skew-symmetric part of the objective function can also be decomposed into three parts using the Sokal and Michener dissimilarities.

Proposition 5.2.2 *Decomposition of the objective function of the constrained hill-climbing model*

The objective function of the constrained hill-climbing model can be decomposed

into symmetric part and skew-symmetric part, respectively, as follows:

$$\sum_{(i,j)\in\Omega^*}\Big(\delta_{ij}-\Big[d_{ij}(\boldsymbol{UX}^*)+\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{UX}^*)^{-1}\Big]\Big)^2$$

$$=\sum_{(i,j)\in\Omega^*}\Big(s_{ij}-d_{ij}(\boldsymbol{UX}^*)\Big)^2 \tag{5.3}$$

$$+\sum_{(i,j)\in\Omega^*}\Big(a_{ij}-\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{z}d_{ij}(\boldsymbol{UX}^*)^{-1}\Big)^2 \tag{5.4}$$

where

$$\boldsymbol{S}=\frac{1}{2}(\boldsymbol{\Delta}+\boldsymbol{\Delta}^T)=(s_{ij})\ (i,j=1,2,\cdots,n)\ \textit{and}$$
$$\boldsymbol{A}=\frac{1}{2}(\boldsymbol{\Delta}-\boldsymbol{\Delta}^T)=(a_{ij})\ (i,j=1,2,\cdots,n)$$

are symmetric and skew-symmetric parts, respectively.

Proof.

$$\sum_{(i,j)\in\Omega^*}\Big(\delta_{ij}-\Big[d_{ij}(\boldsymbol{UX}^*)+\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{UX}^*)^{-1}\Big]\Big)^2$$

$$=\sum_{(i,j)\in\Omega^*}\Big(s_{ij}-d_{ij}(\boldsymbol{UX}^*)+a_{ij}-\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{UX}^*)^{-1}\Big)^2$$

$$=\sum_{(i,j)\in\Omega^*}\Big(s_{ij}-d_{ij}(\boldsymbol{UX}^*)\Big)^2 \tag{5.5}$$

$$+\sum_{(i,j)\in\Omega^*}\Big(a_{ij}-\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{UX}^*)^{-1}\Big)^2 \tag{5.6}$$

$$+2\sum_{(i,j)\in\Omega^*}\Big(s_{ij}-d_{ij}(\boldsymbol{UX}^*)\Big)\Big(a_{ij}-\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{UX}^*)^{-1}\Big) \tag{5.7}$$

In Eq. (5.7), for arbitrarily $(i,j)\in\Omega$, $s_{ij}-d_{ij}(\boldsymbol{UX}^*)=s_{ji}-d_{ji}(\boldsymbol{UX}^*)$ since both s_{ij} and $d_{ij}(\boldsymbol{UX}^*)$ are symmetric. In addition, for arbitrarily $(i,j)\in\Omega$

$$\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{UX}^*)^{-1}=-\Big(\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*-\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*\Big)^T\boldsymbol{v}d_{ji}(\boldsymbol{UX}^*)^{-1}.$$

and these parts become skew-symmetric. Therefore, the following term becomes skew-symmetric dissimilarities from Proposition 3.1.3

$$\Big(a_{ij}-\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{UX}^*)^{-1}\Big)$$

and Eq. (5.7) becomes 0 *from Proposition 3.1.2.*

Moreover, the symmetric and skew-symmetric parts are further decomposed, respectively. Again, the decomposition of the symmetric part for the constrained hill-climbing model is equivalent to that of CDS. Both decompositions of the objective functions of the constrained hill-climbing model for the symmetric and skew-symmetric parts are obtained using the Sokal and Michener dissimilarities.

Proposition 5.2.3 *Decomposition of symmetric part of the constraied hill-climbing model*

Given the symmetric part of the objective function of the constrained hill-climbing model, the objective function can be decomposed as follows:

$$
\begin{aligned}
L(\boldsymbol{X}^*, \boldsymbol{U} |\, \boldsymbol{S}) &= \sum_{(i,j)\in\Omega^*} (s_{ij} - d_{ij}(\boldsymbol{U}\boldsymbol{X}^*))^2 \\
&= \sum_{(i,j)\in\Omega^*} \sum_{o\neq\ell} u_{io}u_{j\ell}(s_{ij} - \tilde{s}_{o\ell})^2 && (5.8)\\
&+ \sum_{(i,j)\in\Omega^*} \sum_{o=1}^{k} u_{io}u_{jo}(s_{ij} - \tilde{s}_{oo})^2 && (5.9)\\
&+ \sum_{o\neq\ell} \tilde{w}_{o\ell}(\tilde{s}_{o\ell} - d_{o\ell}(\boldsymbol{X}^*))^2 && (5.10)\\
&+ \sum_{o=1}^{k} \tilde{w}_{oo}\tilde{s}_{oo}^2 && (5.11)
\end{aligned}
$$

where

$$
\tilde{s}_{o\ell} = \sum_{i=1}^{n}\sum_{j=1}^{n} u_{io}u_{j\ell}\frac{s_{ij}}{\tilde{w}_{o\ell}} \quad \text{and} \quad \tilde{w}_{o\ell} = \sum_{i=1}^{n}\sum_{j=1}^{n} u_{io}u_{j\ell}\ (o, \ell = 1, 2, \cdots, k),
$$

are Sokal and Michener dissimilaries and weights among clusters, respectively.

Proof. *See the Heiser and Groenen (1997).*

For the symmetric part of the constrained hill-climbing model's objective function, Eq. (5.8), Eq. (5.9), Eq. (5.10) and Eq. (5.11) are called the *Among-cluster Error Sum of Squares (SSQ)*, *Within-cluster Error SSQ*, *Lack of Spatial fit* and *Lack of homogeneity*, respectively.

Next, we show the decomposition of the skew-symmetric part of the constrained hill-climbing model.

Proposition 5.2.4 *Decomposition of skew-symmetric part of the constrained hill-climbing model*

Given the skew-symmetric part of the objective function of the constrained hill-climbing model, the objective function can be decomposed as follows:

$$\begin{aligned}
&\sum_{(i,j)\in\Omega^*}\Big(a_{ij}-\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)^{-1}\Big)^2\\
=&\sum_{o\neq\ell}\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}(a_{ij}-\tilde{a}_{o\ell})^2 &(5.12)\\
&+\sum_{o=1}^{k}\sum_{(i,j)\in\Omega^*}u_{io}u_{jo}a_{ij}^2 &(5.13)\\
&+\sum_{o\neq\ell}\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}(\tilde{a}_{o\ell}-(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X})^{-1})^2 &(5.14)
\end{aligned}$$

where

$$\tilde{a}_{o\ell}=\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}\frac{a_{ij}}{\tilde{w}_{o\ell}},\quad \textit{and}\quad \tilde{w}_{o\ell}=\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}\ (o,\ell=1,2,\cdots,k),$$

are Sokal and Michener dissimilarities for skew-symmetries and weights for clusters, respectively.

Proof.

$$\begin{aligned}
&\sum_{(i,j)\in\Omega^*}\Big(a_{ij}-\Big(\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^*-\sum_{\ell=1}^{k}u_{j\ell}\boldsymbol{x}_\ell^*\Big)^T\boldsymbol{v}d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)^{-1}\Big)^2\\
=&\sum_{(i,j)\in\Omega^*}\sum_{o=1}^{k}\sum_{\ell=1}^{k}u_{io}u_{j\ell}\Big(a_{ij}-(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\Big)^2\\
=&\sum_{(i,j)\in\Omega^*}\sum_{o=1}^{k}\sum_{\ell=1}^{k}u_{io}u_{j\ell}\Big(a_{ij}-\tilde{a}_{o\ell}+\tilde{a}_{o\ell}-(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\Big)^2\\
=&\sum_{(i,j)\in\Omega^*}\sum_{o=1}^{k}\sum_{\ell=1}^{k}u_{io}u_{j\ell}\Big(a_{ij}-\tilde{a}_{o\ell}\Big)^2 &(5.15)\\
&+\sum_{(i,j)\in\Omega^*}\sum_{o=1}^{k}\sum_{\ell=1}^{k}u_{io}u_{j\ell}\Big(\tilde{a}_{o\ell}-(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\Big)^2 &(5.16)\\
&+2\sum_{(i,j)\in\Omega^*}\sum_{o=1}^{k}\sum_{\ell=1}^{k}u_{io}u_{j\ell}\Big(a_{ij}-\tilde{a}_{o\ell}\Big)\Big(\tilde{a}_{o\ell}-(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\Big) &(5.17)
\end{aligned}$$

Eq. (5.17) is described as follows:

$$
\begin{aligned}
&2\sum_{(i,j)\in\Omega^*}\sum_{o=1}^{k}\sum_{\ell=1}^{k}u_{io}u_{j\ell}\Big(a_{ij}-\tilde{a}_{o\ell}\Big)\Big(\tilde{a}_{o\ell}-(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\Big)\\
=&2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\tilde{a}_{o\ell}\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}a_{ij}-2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\tilde{a}_{o\ell}^2\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}\\
&-2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}a_{ij}(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\\
&+2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\tilde{a}_{o\ell}\sum_{(i,j)\in\Omega^*}u_{io}u_{j\ell}(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\\
=&2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\tilde{a}_{o\ell}^2\tilde{w}_{o\ell}-2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\tilde{a}_{o\ell}^2\tilde{w}_{o\ell}-2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\tilde{a}_{o\ell}\tilde{w}_{o\ell}(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}\\
&+2\sum_{o=1}^{k}\sum_{\ell=1}^{k}\tilde{a}_{o\ell}\tilde{w}_{o\ell}(\boldsymbol{x}_o^*-\boldsymbol{x}_\ell^*)^T\boldsymbol{v}d_{o\ell}(\boldsymbol{X}^*)^{-1}=0
\end{aligned}
$$

Eq. (5.15) can be decomposed into Eq. (5.12) and Eq. (5.13). In addition, Eq. (5.16) can be described as Eq. (5.14) by using $(\boldsymbol{x}_o^*-\boldsymbol{x}_o^*)=\mathbf{0}$ *since*

$$
\begin{aligned}
\tilde{a}_{oo}=&\sum_{i=1}^{n}\sum_{j=1}^{n}u_{io}u_{jo}\frac{a_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n}u_{io}u_{jo}}=n_o^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}u_{io}u_{jo}a_{ij}\\
=&n_o^{-2}\sum_{i<j}u_{io}u_{jo}(a_{ij}+a_{ji})=n_o^{-2}\sum_{i<j}u_{io}u_{jo}(a_{ij}-a_{ij})=0
\end{aligned}
\tag{5.18}
$$

where n_o *is the number of object belonging to cluster* C_o*.*

For the skew-symmetric part of the constrained hill-climbing model's objective function, Eq. (5.12), Eq. (5.13) and Eq. (5.14) are called *Among-cluster Error Sum of Squares (SSQ)*, *Within-cluster Error SSQ*, *Lack of Spatial fit* and *Lack of homogeneity*, respectively.

From proposition 5.2.3 and proposition 5.2.4, the objective function of the constrained hil-climbing model can be decomposed into seven terms. Therefore, *Within-cluster Error SSQ*, *Among-cluster Error SSQ* and *Lack of homogeneity* for the symmetric and skew-symmetric parts indicate terms related to clustering for asymmetric dissimilarity data; the clustering can consider both the symmetric and skew-symmetric parts of the asymmetric dissimilarity data. In addition, *Lack of spatial fit* for the symmetric and skew-symmetric parts represent MDS for Sokal and Michener dissimilarities corresponding to symmetric and skew-symmetric parts, respectively.

5.2.3 Algorithm of the constrained hill-climbing model

In this subsection, we show the algorithm of the constrained hill-climbing model. The parameters are estimated on the basis of ALS (Young et al., 1980). The flow

of our proposed algorithm is described as follows.

Algorithm of the constrained hill-climbing model

Step 0 Set k and d, and initial values of $\boldsymbol{X}^*$, $\boldsymbol{v}$ and $\boldsymbol{U}$

Step 1 Update $\boldsymbol{X}^*$ and $\boldsymbol{v}$, given $\boldsymbol{U}$

Step 2 Update $\boldsymbol{U}$, given $\boldsymbol{X}^*$ and $\boldsymbol{v}$

Step 3 If stop condition is satisfied , stop the algorithm, else return to the Step 1

Here, $\boldsymbol{X}^*$ and $\boldsymbol{v}$ are estimated by numerical solutions such as the Broyden-Fletcher-Goldfarb-Shanno method (BFGS) (Nocedal and Wright, 1999).

In this algorithm, $\boldsymbol{U}$ is estimated by each row vector of $\boldsymbol{U}$, although value of the objective function depends on the order of estimating row vector of $\boldsymbol{U}$.

Next, we show how to ith row vector of update $\boldsymbol{U}$.

Proposition 5.2.5 *Updating* $\boldsymbol{U}$

Given $\boldsymbol{X}^*$, $\boldsymbol{v}$ *and* $\boldsymbol{U}$ *without* i*th row vector of* $\boldsymbol{U}$, *if updating rule of* i*th row vector of* $\boldsymbol{U}$ *is used, the value of the objective function of the constrained slide-vector model does not increase.*

$$u_{io} = \begin{cases} 1 & \Big((\forall o^* = 1, 2, \cdots, k)(\gamma_{io}(\boldsymbol{X}^*, \boldsymbol{v}) \leq \gamma_{io^*}(\boldsymbol{X}^*, \boldsymbol{v}))\Big) \\ 0 & (others) \end{cases} \tag{5.19}$$

$(o = 1, 2, \cdots, k)$ *where*

$$\begin{aligned} \gamma_{io}(\boldsymbol{X}^*, \boldsymbol{v}) = &\sum_{j \neq i} \sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \\ &+ \sum_{j \neq i} \sum_{\ell=1}^{k} u_{j\ell}(\delta_{ji} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \end{aligned}$$

and o^* *is index set of cluster for objects.*

Proof. *We assumed that object* $i^\dagger$ *is satisfying* $\gamma_{i^\dagger o}(\boldsymbol{X}^*, \boldsymbol{z}) \leq \gamma_{i^\dagger o^*}(\boldsymbol{X}^*, \boldsymbol{v})$ *for all* $o^* = 1, 2, \cdots, k$.

In the situation, the objective function of the constrained hill-climbing model can be decomposed as follows:

$$\begin{aligned} &\sum_{i=1}^{n} \sum_{o=1}^{k} u_{io} \sum_{j \neq i} \sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \\ =&\sum_{i \neq i^\dagger}^{n} \sum_{o=1}^{k} u_{io} \sum_{j \neq i} \sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \qquad (5.20) \\ &+ \sum_{o=1}^{k} u_{i^\dagger o} \sum_{j \neq i^\dagger} \sum_{\ell=1}^{k} u_{j\ell}(\delta_{i^\dagger j} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \qquad (5.21) \end{aligned}$$

Eq. (5.20) can be decomposed as follows:

$$\sum_{i\neq i^\dagger}^{n}\sum_{o=1}^{k} u_{io}\sum_{j\neq i}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2$$

$$= \sum_{i\neq i^\dagger}^{n}\sum_{o=1}^{k} u_{io}\sum_{(j\neq i)\wedge(j\neq i^\dagger)}\sum_{\ell=1}^{k} u_{j\ell}(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \quad (5.22)$$

$$+ \sum_{i\neq i^\dagger}^{n}\sum_{o=1}^{k} u_{io}\sum_{\ell=1}^{k} u_{i^\dagger\ell}(\delta_{ii^\dagger} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \quad (5.23)$$

Eq. (5.23) can be described as follows:

$$\sum_{i\neq i^\dagger}^{n}\sum_{o=1}^{k} u_{io}\sum_{\ell=1}^{k} u_{i^\dagger\ell}(\delta_{ii^\dagger} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2$$

$$= \sum_{\ell=1}^{k} u_{i^\dagger\ell}\sum_{i\neq i^\dagger}^{n}\sum_{o=1}^{k} u_{io}(\delta_{ii^\dagger} - (d_{o\ell}(\boldsymbol{X}^*) + (\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*)^T \boldsymbol{v} d_{o\ell}(\boldsymbol{X}^*)^{-1}))^2 \quad (5.24)$$

Therefore, the objective function of constrained hill-climbing model is decomposed into Eq. (5.20), Eq. (5.22), and Eq. (5.24). Eq. (5.20) and Eq. (5.24) is minizmied for arbitrarily o from the rule Eq. (5.19). In addition, Eq. (5.22) is not affected by the updating rule because Eq. (5.22) does not include part of $i^\dagger$. Therefore, the proposition is satisfied.

Then, updating algorithm of $\boldsymbol{U}$ is shown.

Algorithm of updating $\boldsymbol{U}$

Step 1 $i \leftarrow 1$

Step 2 Calcurate $\gamma_{io}(\boldsymbol{X}^*, \boldsymbol{v})$ for $o = 1, 2, \cdots, k$

Step 3 Update ith row vector of $\boldsymbol{U}$ based on the rule of Eq.(5.19) among $\gamma_{io}(\boldsymbol{X}^*, \boldsymbol{v})$ ($o = 1, 2, \cdots, k$). If the number of o such that $u_{io} = 1$ is over 0, one of o is selected randomly.

Step 4 If $i = n$ stop, otherwise, $i \leftarrow i + 1$ and back to Step 2

5.3 Constrained radius model based on CDS

This section comprise three parts, i.e., a description of the model and objective function for the constrained radius model, the property, and the algorithm.

5.3.1 Model and objective function of the constrained radius model

In this subsection, we show the model and objective function of the constrained radius model. Furthermore, we introduce two types of descriptions for the objective function of the constrained radius model, and show their equivalence based on Heiser and Groenen (1997).

Definition 5.3.1 *Model of constrained radius model*

Let $\boldsymbol{\Delta} = (\delta_{ij})$ $\delta_{ij} \in \mathbb{R}_+$ $(i, j = 1, 2, \cdots, n)$ *be an asymmetric dissimilarity matrix,* $\boldsymbol{X}^* = (x^*_{o\ell})$ $x^*_{ot} \in \mathbb{R}$ $(o = 1, 2, \cdots, k; t = 1, 2, \cdots, d)$ *and* $\boldsymbol{r}^*_f = (r^*_f)$, $r^*_f > 0$ $(f = 1, 2, \cdots, m; m \leq k)$ *be coordinate matrix of object clusters in* d *dimensions and the length of radii for clusters, respectively, where* k *and* m *are the number of clusters for objects and centroids, respectively. Here, the model of the constrained radius model is defined as following equation:*

$$(\forall i, j = 1, 2, \cdots, n)(\exists! C_o; i \in C_o)(\exists! C_\ell; j \in C_\ell)(\exists! C^*_f; o \in C^*_f)(\exists C^*_q; \ell \in C^*_q)$$
$$(\delta_{ij} = d_{o\ell}(\boldsymbol{X}^*) - r^*_f + r^*_q + \varepsilon_{ij})$$

where

$$d_{o\ell}(\boldsymbol{X}^*) = \|\boldsymbol{x}^*_o - \boldsymbol{x}^*_\ell\| = \Big[\sum_{t=1}^{d}(x^*_{ot} - x^*_{\ell t})^2\Big]^{\frac{1}{2}},$$

C_o *and* C^*_f *are clusters for objects and centroids, respectively, and* $\varepsilon_{ij} \in \mathbb{R}$ $(i, j = 1, 2, \cdots, n)$ *is error.*

From the constrained radius model, coordinates of clusters and radii of clusters are estimated. Furthermore, radius length depend on $m(\leq k)$ from the parsimonious notion.

Next, we define the objective function of the constrained radius model I.

Definition 5.3.2 *Objective function of the constrained radius model I*

Given an asymmetric dissimilarity matrix $\boldsymbol{\Delta}$*, the number of clusters for objects* k *and for centroids* m*, and the number of low-dimensions* d*, the objective function of the constrained radius model I is defined as follows;*

$$L(\boldsymbol{X}^*, \boldsymbol{r}^*, \boldsymbol{U}, \boldsymbol{\Psi} | \boldsymbol{\Delta}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k}\sum_{\ell=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) - \sum_{f=1}^{m}\psi_{of}r^*_f + \sum_{q=1}^{m}\psi_{\ell q}r^*_q))^2$$

subject to $r^*_f > 0$ $(f = 1, 2, \cdots, m)$*, where* $\boldsymbol{U} = (u_{io})$ $u_{io} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; o = 1, 2, \cdots, k)$ *and* $\boldsymbol{\Psi} = (\psi_{of})$ $\psi_{of} \in \{0, 1\}$ $(o = 1, 2, \cdots, k; f = 1, 2, \cdots, m)$*. In the constrained radius model,* $\boldsymbol{X}^*$*,* $\boldsymbol{r}^*$*,* $\boldsymbol{U}$ *and* $\boldsymbol{\Psi}$ *are estimated such that the value of the objective functin is minimized.*

Next, we define the objective function of another constrained radius model (i.e., the constrained radius model II), and show the equivalence of these objective function.

Definition 5.3.3 *Objective function of the constrained radius model II*

Given an asymmetric dissimilarity data $\boldsymbol{\Delta}$*, the number of object clusters* k *and centroids clusters* m*, respectively, and the number of low-dimensions* d*, the objective function of the constrained radius model II is defined as follows;*

$$L(\boldsymbol{X}^*,\boldsymbol{r}^*,\boldsymbol{U},\boldsymbol{\Psi}|\ \boldsymbol{\Delta}) = \left\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\right\|^2$$

subject to $r_f^* > 0\ (f = 1, 2, \cdots, m)$ *where*

$$\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) = (d_{ij}(\boldsymbol{U}\boldsymbol{X}^*)),\ d_{ij}(\boldsymbol{U}\boldsymbol{X}^*) = \left\| \sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{j\ell}\boldsymbol{x}_\ell^* \right\|\ (i, j = 1, 2, \cdots, n),$$

and $\mathbf{1}_n = (1)$ *is vector with length of* n*. In the constraind radius model,* $\boldsymbol{X}^*$*,* $\boldsymbol{r}^*$*,* $\boldsymbol{U}$ *and* $\boldsymbol{\Psi}$ *are estimated such that the value of the objective function is minimized.*

Proposition 5.3.1 *Equivalence of constrained radius model I and II*

Given an asymmetric dissimilarity matrix $\boldsymbol{\Delta}$*, indicator matrices for objects* $\boldsymbol{U}$ *and for centroids* $\boldsymbol{\Psi}$*, respectively, coordinates of cluster centroids* $\boldsymbol{X}^*$ *and radii* $\boldsymbol{r}^*$*, and the number of low-dimensions* d*, the following property is satisfied;*

$$\begin{aligned}
&\left\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\right\|^2 \\
=&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{o=1}^{k}\sum_{\ell=1}^{k} u_{io}u_{j\ell}(\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) - \sum_{f=1}^{m}\psi_{of}r_f^* + \sum_{q=1}^{m}\psi_{\ell q}r_q^*))^2
\end{aligned} \tag{5.25}$$

Proof. *From the left term of Eq. (5.25),*

$$\begin{aligned}
&\left\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\right\|^2 \\
=&\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\delta_{ij} - \left(d_{ij}(\boldsymbol{U}\boldsymbol{X}^*) - \sum_{o=1}^{k}\sum_{f=1}^{m}u_{io}\psi_{of}r_f^* + \sum_{o=1}^{k}\sum_{q=1}^{m}u_{j\ell}\psi_{\ell q}r_q^*\right)\right)^2 \\
=&\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\delta_{ij} - \left(\|\sum_{o=1}^{k}u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k}u_{i\ell}\boldsymbol{x}_\ell^*\| - \sum_{o=1}^{k}u_{io}\sum_{f=1}^{m}\psi_{of}r_f^* + \sum_{o=1}^{k}u_{j\ell}\sum_{q=1}^{m}\psi_{\ell q}r_q^*\right)\right)^2
\end{aligned} \tag{5.26}$$

Here, for all i *and* j*, there exists* o^* *and* ℓ^**, uniquely, such that* $u_{io^*} = 1$ *and* $u_{io} = 0\ (o \neq o^*)$*, and* $u_{j\ell^*} = 1$ *and* $u_{j\ell} = 0\ (\ell \neq \ell^*)$*, respectively, from the definition*

of indicator matrix. Therefore i and j part of Eq. (5.26) is described as follows:

$$
\begin{aligned}
&\Big(\delta_{ij} - \Big(\|\sum_{o=1}^{k} u_{io}\boldsymbol{x}_o^* - \sum_{\ell=1}^{k} u_{i\ell}\boldsymbol{x}_\ell^*\| - \sum_{o=1}^{k} u_{io}\sum_{f=1}^{m}\psi_{of}r_f^* + \sum_{o=1}^{k} u_{j\ell}\sum_{q=1}^{m}\psi_{\ell q}r_q^*\Big)\Big)^2 \\
=&\Big(\delta_{ij} - \Big(\|u_{io^*}\boldsymbol{x}_{o^*}^* - u_{j\ell^*}\boldsymbol{x}_{\ell^*}^*\| - u_{io^*}\sum_{f=1}^{m}\psi_{o^*f}r_f^* + u_{j\ell^*}\sum_{q=1}^{m}\psi_{\ell^* q}r_q^*\Big)\Big)^2 \\
=&\Big(\delta_{ij} - \Big(\|\boldsymbol{x}_{o^*}^* - \boldsymbol{x}_{\ell^*}^*\| - \sum_{f=1}^{m}\psi_{o^*f}r_f^* + \sum_{q=1}^{m}\psi_{\ell^* q}r_q^*\Big)\Big)^2 \\
=&\sum_{o=1}^{k}\sum_{\ell=1}^{k} u_{io}u_{j\ell}\Big(\delta_{ij} - \Big(\|\boldsymbol{x}_o^* - \boldsymbol{x}_\ell^*\| - \sum_{f=1}^{m}\psi_{of}r_f^* + \sum_{q=1}^{m}\psi_{\ell q}r_q^*\Big)\Big)^2
\end{aligned}
\tag{5.27}
$$

From proposition 5.3.1, the constrained radius model can be considered a special case of the radius model. In short, if $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{X}^*$ and $\boldsymbol{r} = \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*$, the radius model becomes the constrained radius model.

5.3.2 Properties of the constrained radius model

In this subsection, we introduce three types of decompositions of the constrained radius model. First, we show the decomposition of the objective function of the constrained radius model into symmetric and skew-symmetric parts. From this decomposition, the interpretation of radius lengths becomes clear. Therefore, if the difference of radius lengths between clusters is large, the asymmetric relation between the two clusters is also interpreted as large. Furthermore, the symmetric part is equivalent to the objective function of CDS for the symmetric part of asymmetric dissimilarity data. For the estimation of these parameters, the manner which we estimate $\boldsymbol{X}$ is the same as that of CDS because $\boldsymbol{X}$ is included only in the symmetric part of the objective function for the constrained radius model. Similarly, the manner of which we estimate $\boldsymbol{r}^*$ and $\boldsymbol{\Psi}$ depend only on the skew-symmetric part. Second, the symmetric part of the objective function can be decomposed into four parts in the same manner as that of CDS. Third, the skew-symmetric part of the objective function can also be decomposed into three parts using the Sokal and Michener dissimilarities.

In the remainder of this subsection, we show the indefiniteness of the objective function for the constrained radius model. From the property, manner in which we estimate $\boldsymbol{r}^*$ becomes a non-constrained optimization problem.

Proposition 5.3.2 *Decomposition of the objective function of the constrained radius model*

Given the objective function of the constrained radius model, the objective function

can be decomposed as follows:

$$\begin{aligned}
&\left\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\right\|^2 \\
=&\left\|\boldsymbol{S} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\right\|^2 \qquad (5.28)\\
&+ \left\|\boldsymbol{A} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)\right\|^2 \qquad (5.29)
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{S} =&(\boldsymbol{\Delta} + \boldsymbol{\Delta}^T)/2,\ \boldsymbol{S} = (s_{ij})\ (i,j = 1,2,\cdots,n)\ \textit{and}\\
\boldsymbol{A} =&(\boldsymbol{\Delta} - \boldsymbol{\Delta}^T)/2,\ \boldsymbol{A} = (a_{ij})\ (i,j = 1,2,\cdots,n).
\end{aligned}$$

Proof.

$$\begin{aligned}
&\left\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\right\|^2 \\
=&\left\|(\boldsymbol{\Delta} + \boldsymbol{\Delta}^T)/2 + (\boldsymbol{\Delta} - \boldsymbol{\Delta}^T)/2 - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\right\|^2 \\
=&\left\|\boldsymbol{S} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\right\|^2 \\
&+ \left\|\boldsymbol{A} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)\right\|^2 \\
&+ 2tr\Big(\boldsymbol{S} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\Big)^T\Big(\boldsymbol{A} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T\Big) \qquad (5.30)
\end{aligned}$$

In Eq. (5.30), $\boldsymbol{S} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)$ *is symmetric matrix since both* $\boldsymbol{S}$ *and* $\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)$ *are symmetric matrices. In addition,* $(\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)$ *is skew-symmetric matrix since*

$$(\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T) = -(\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)^T$$

Therefore, $\boldsymbol{A} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)$ *is skew-symmetric matrix from Proposition 3.1.3.*

Then, Eq. (5.30) becomes 0 *from Proposition 3.1.2.*

In the constrained radius model, Eq. (5.28) and Eq. (5.29) are called as symmetric and skew-symmetric part of the constrained radius model.

There exists a further decomposition of the objective function corresponding to Eq. (5.28) and Eq. (5.29), respectively. The decomposition of Eq. (5.28) is the same for proposition 5.2.3; therefore, we show the decomposition of the objective function corresponding to Eq. (5.29).

Proposition 5.3.3 *Decomposition of skew-symetric part of the constrained radius model*

Given the skew-symmetric part of the objective fucntion of the objective function of the constrained radius model, the objective function can be decomposed as follows:

$$\begin{aligned}
&\|\boldsymbol{A} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)\|^2 \\
=&\sum_{o\neq\ell}\|\mathrm{diag}(\boldsymbol{u}_{(o)})[\boldsymbol{A} - \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}_U]\mathrm{diag}(\boldsymbol{u}_{(\ell)})\|^2 &(5.31)\\
&+\sum_{o=1}^{k}\|\mathrm{diag}(\boldsymbol{u}_{(o)})\boldsymbol{A}\mathrm{diag}(\boldsymbol{u}_{(o)})\|^2 &(5.32)\\
&+\|\boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}_U - \boldsymbol{P}_U(\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)\boldsymbol{P}_U\|^2 &(5.33)
\end{aligned}$$

where $\boldsymbol{P}_U = \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T$ is projection matrix.

Proof. *From the same way of proposition Proposition 4.2.2, we have*

$$\begin{aligned}
&\|\boldsymbol{A} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)\|^2 \\
=&\|\boldsymbol{A} - \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P} + \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)\|^2 \\
=&\|\boldsymbol{A} - \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}\|^2 &(5.34)\\
&+\|\boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P} - (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T)\|^2 &(5.35)
\end{aligned}$$

Eq. (5.34) can be further decomposed as follows

$$\begin{aligned}
\|\boldsymbol{A} - \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}_U\|^2 =& \sum_{o\neq\ell}\|\mathrm{diag}(\boldsymbol{u}_{(o)})[\boldsymbol{A} - \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}_U]\mathrm{diag}(\boldsymbol{u}_{(\ell)})\|^2 \\
&+\sum_{o=1}^{k}\|\mathrm{diag}(\boldsymbol{u}_{(o)})[\boldsymbol{A} - \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}_U]\mathrm{diag}(\boldsymbol{u}_{(o)})\|^2
\end{aligned}$$

From Eq. (5.18),

$$\mathrm{diag}(\boldsymbol{u}_{(o)})\boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}_U\mathrm{diag}(\boldsymbol{u}_{(o)}) = \boldsymbol{O}_n$$

for any $o = 1, 2, \cdots, k$, this proposition is proved.

For the skew-symmetric part of the objective function, from proposition 5.3.3, Eq. (5.31), Eq. (5.32) and Eq. (5.33) are called *Among-cluster Error SSQ* , *Within-cluster Error SSQ* and *Lack of Spatial fit*, respectively.

Next, we show the indefiniteness of the objective function of the constrained radius model. From the property, the manner in which we estimate $\boldsymbol{r}^*$ becomes a non-constrained optimization problem.

Proposition 5.3.4 *Indefiniteness of the objective function of constrained radius model*

For the objective function of the constrained radius model, following property is satisfied:

$$\begin{aligned}
&(\forall c \in \mathbb{R}) \\
&\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\|^2 \\
=&\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n(\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + c\mathbf{1}_n^T) + (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^* + c\mathbf{1}_n)\mathbf{1}_n^T)\|^2 &(5.36)
\end{aligned}$$

Proof. *For any $c \in \mathbb{R}$, Eq. (5.36) is described as follows:*

$$\begin{aligned}
&\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n(\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + c\mathbf{1}_n^T) + (\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^* + c\mathbf{1}_n)\mathbf{1}_n^T)\|^2 \\
=&\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T - c\mathbf{1}_n\mathbf{1}_n^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T + c\mathbf{1}_n\mathbf{1}_n^T)\|^2 \\
&\|\boldsymbol{\Delta} - (\boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*) - \mathbf{1}_n\boldsymbol{r}^{*T}\boldsymbol{\Psi}^T\boldsymbol{U}^T + \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*\mathbf{1}_n^T)\|^2
\end{aligned}$$

5.3.3 Algorithm of the constrained radius model

In this subsection, the algorithm of the constrained radius model is shown. Here, the parameters are estimated on the basis of ALS (Young et al., 1980). The flow of our proposed algorithm is described as follows.

Algorithm of the constrained radius model

Step 0 Set k, m and d and initial values of $\boldsymbol{X}^*$, $\boldsymbol{r}^*$, $\boldsymbol{U}$ and $\boldsymbol{\Psi}$

Step 1 Update $\boldsymbol{X}^*$ based on Eq. (5.28), given $\boldsymbol{U}$, $\boldsymbol{\Psi}$ and $\boldsymbol{r}^*$

Step 2 Update $\boldsymbol{r}$ based on Eq. (5.29), given $\boldsymbol{X}^*$, $\boldsymbol{U}$ and $\boldsymbol{\Psi}$

Step 3 Update $\boldsymbol{\Psi}$ based on Eq. (5.29), given $\boldsymbol{X}^*$, $\boldsymbol{U}$ and $\boldsymbol{r}^*$

Step 4 Update $\boldsymbol{U}$, given $\boldsymbol{X}^*$, $\boldsymbol{r}^*$ and $\boldsymbol{\Psi}$

Step 5 If stop condition is satisfied, stop the algorithm, else return to the Step 1

From the decomposition of the constrained radius model, when $\boldsymbol{X}^*$ is estimated, the estimation only depends on Eq. (5.28). As with estimating $\boldsymbol{X}^*$, when $\boldsymbol{r}^*$ is estimated, the estimation only depends on Eq. (5.29); however, estimating $\boldsymbol{U}$ depends on the objective function of the constrained radius model because both Eq. (5.28) and Eq. (5.29) include $\boldsymbol{U}$.

To update $\boldsymbol{X}^*$, we adopt a majorizing algorithm which we derive from the constrained hill-climbing model. The majorizing function is derived depending only on Eq. (5.28). In short, the majorizing function is equivalent to that of CDS for the symmetric part of the asymmetric dissimilarity data.

Proposition 5.3.5 *Majorizing function of the constrained radius model*

Given objective function of the constrained slide-vector model, the majorizing function of the constrained slide-vector model is derived as follows:

$$\begin{aligned}
&\left\|\boldsymbol{S} - \boldsymbol{D}(\boldsymbol{U}\boldsymbol{X}^*)\right\|^2 \\
&\leq \eta_\delta^2 + tr\boldsymbol{X}^{*T}\boldsymbol{U}^T\boldsymbol{V}\boldsymbol{U}\boldsymbol{X}^* - 2tr\boldsymbol{X}^{*T}\boldsymbol{U}^T\boldsymbol{B}(\boldsymbol{U}\boldsymbol{H})\boldsymbol{U}\boldsymbol{H} = L_M(\boldsymbol{X}^*, \boldsymbol{H}, \boldsymbol{U}|\boldsymbol{S}) \quad (5.37)
\end{aligned}$$

where

$$\eta_\delta^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^2$$

$$\boldsymbol{V} = \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T,$$

$$\boldsymbol{e}_i = (e_{iq}) \quad e_{iq} = \begin{cases} 1 & (i = q) \\ 0 & (i \neq q) \end{cases}, \quad (i, q = 1, 2, \cdots, n),$$

$$\boldsymbol{B}(\boldsymbol{UH}) = (b_{ij})\ (i, j = 1, 2, \cdots, n)$$

$$b_{ij} = \begin{cases} -\frac{s_{ij}}{d_{ij}(\boldsymbol{UH})} & (if\ i \neq j\ and\ d_{ij}(\boldsymbol{UH}) \neq 0) \\ 0 & (if\ i \neq j\ and\ d_{ij}(\boldsymbol{UH}) = 0) \end{cases}$$

$$b_{ii} = -\sum_{(j=1)\wedge(i\neq j)}^{n} b_{ij} \quad and$$

$$\boldsymbol{H} = (h_{ot}) \quad h_{ot} \in \mathbb{R}\ (o = 1, 2, \cdots, k; t = 1, 2, \cdots, d).$$

Proof. *This proorf can be conducted as the same way of Proposition 4.2.3.*

Next, updating formula of $\boldsymbol{X}^*$ is shown based on the majorizing function.

Proposition 5.3.6 *Updating formula of* $\boldsymbol{X}^*$

Given $\boldsymbol{U}$ *and* $\boldsymbol{H}$ *, updating formula of* $\boldsymbol{X}^*$ *minimizing Eq. (5.37) is derived as follows:*

$$\boldsymbol{X}^* = [\boldsymbol{U}^T\boldsymbol{V}\boldsymbol{U}]^+\boldsymbol{U}^T\boldsymbol{B}(\boldsymbol{UH})\boldsymbol{UH} \tag{5.38}$$

where $[\boldsymbol{U}^T\boldsymbol{V}\boldsymbol{U}]^+$ *is the Moore-Penrose inverse of* $\boldsymbol{U}^T\boldsymbol{V}\boldsymbol{U}$*.*

Proof.

Then, we show the algorithm for estimating $\boldsymbol{X}^*$.

Algorithm of estimating $\boldsymbol{X}^*$

Step 0 Set iteration number $\beta \leftarrow 1$, initial value of ${}^{(\beta)}\boldsymbol{H}$, and threhold value $\varepsilon > 0$

Step 1 Update ${}^{(\beta)}\boldsymbol{X}^*$ based on Eq.(5.38)

Step 2 ${}^{(\beta)}\boldsymbol{X}^*$ is substituted into ${}^{(\beta+1)}\boldsymbol{H}$ and $\beta \leftarrow \beta + 1$

Step 3 $|{}^{(\beta)}L_M(\boldsymbol{X}^*, \boldsymbol{H}, \boldsymbol{U}|\boldsymbol{S}) - {}^{(\beta+1)}L_M(\boldsymbol{X}^*, \boldsymbol{H}, \boldsymbol{U}|\boldsymbol{S})| < \varepsilon$ is satisfied, stop, $i \leftarrow i+1$ and else back to Step 1.

where ${}^{(\beta)}\boldsymbol{H}$, ${}^{(\beta)}\boldsymbol{Q}$ and ${}^{(\beta)}L_M(\boldsymbol{X}^*, \boldsymbol{H}, \boldsymbol{U}|\boldsymbol{S})$ are $\boldsymbol{H}$, $\boldsymbol{Q}$ and a value of the objective function corresponding to βth iteration, respectively.

To estimate $\boldsymbol{r}^*$, we use numerical solutions such as the BFGS method.

Next, we show how we update ith row vector of $\boldsymbol{U}$.

Proposition 5.3.7 *Updating* $\boldsymbol{U}$

Given $\boldsymbol{X}^*$, $\boldsymbol{r}^*$ *and* $\boldsymbol{U}$ *without* i*th row vector of* $\boldsymbol{U}$*, if updating rule of* i*th row vector of* $\boldsymbol{U}$ *is used, values of the objective function of the constrained slide-vector model does not increase.*

$$u_{io} = \begin{cases} 1 & \Big((\forall o^* = 1, 2, \cdots, k)(\gamma_{io}(\boldsymbol{X}^*, \boldsymbol{r}^*, \boldsymbol{\Psi}) \leq \gamma_{io^*}(\boldsymbol{X}^*, \boldsymbol{r}^*, \boldsymbol{\Psi})\Big) \\ 0 & (others) \end{cases} \tag{5.39}$$

where

$$\begin{aligned} \gamma_{io}(\boldsymbol{X}^*, \boldsymbol{r}^*, \boldsymbol{\Psi}) = &\sum_{j \neq i} \sum_{\ell=1}^{k} w_{ij} u_{j\ell} (\delta_{ij} - (d_{o\ell}(\boldsymbol{X}^*) - \sum_{f=1}^{m} \psi_{of} r_f + \sum_{q=1}^{m} \psi_{\ell q} r_q))^2 \\ &+ \sum_{j \neq i} \sum_{\ell=1}^{k} w_{ij} u_{j\ell} (\delta_{ji} - (d_{o\ell}(\boldsymbol{X}^*) - \sum_{f=1}^{m} \psi_{of} r_f + \sum_{q=1}^{m} \psi_{\ell q} r_q))^2 \end{aligned}$$

and o^* *is index set of cluster for objects.*

Proof. *The proof can be conducted as the same way of Proposition 5.2.5.*

The updating algorithm of $\boldsymbol{U}$ is shown.

Algorithm of updation $\boldsymbol{U}$

Step 1 $i \leftarrow 1$

Step 2 Calcurate $\gamma_{io}(\boldsymbol{X}^*, \boldsymbol{r}^*, \boldsymbol{\Psi})$ for $o = 1, 2, \cdots, k$

Step 3 Update ith row vector of $\boldsymbol{U}$ based on the rule of Eq.(5.39) among $\gamma_{io}(\boldsymbol{X}^*, \boldsymbol{r}^*, \boldsymbol{\Psi})$ $(o = 1, 2, \cdots, k)$

Step 4 If $i = n$ stop, otherwise, $i \leftarrow i + 1$ back to Step 2

Next we show how we update ℓth row vector of $\boldsymbol{\Psi}$.

Proposition 5.3.8 *Updating* $\boldsymbol{\Psi}$

Given $\boldsymbol{U}$, $\boldsymbol{r}$ *and* $\boldsymbol{\Psi}$ *without* ℓ*th row vector of* $\boldsymbol{\Psi}$*, if updating rule of* $\boldsymbol{\Psi}$ *is used, values of the objective function of the constrained radius model does not increase.*

$$\psi_{of} = \begin{cases} 1 & \Big((\forall f^* = 1, 2, \cdots, m)(\kappa_{of}(\boldsymbol{U}, \boldsymbol{r}^*) \leq \kappa_{of^*}(\boldsymbol{U}, \boldsymbol{r}^*)\Big) \\ 0 & (others) \end{cases} \tag{5.40}$$

where

$$\begin{aligned} \kappa_{of}(\boldsymbol{U}, \boldsymbol{r}^*) = &\sum_{q \neq f} \sum_{\ell=1}^{k} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} u_{io} u_{j\ell} \psi_{\ell q} (a_{ij} - (\sum_{f=1}^{m} \psi_{of} r_f^* - \sum_{q=1}^{m} \psi_{\ell q} r_q^*))^2 \\ &+ \sum_{q \neq f} \sum_{\ell=1}^{k} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} u_{io} u_{j\ell} \psi_{\ell q} (a_{ji} - (\sum_{f=1}^{m} \psi_{of} r_f^* - \sum_{q=1}^{m} \psi_{\ell q} r_q^*))^2 \end{aligned}$$

and ℓ^* *is index set of cluster for centroids.*

Proof. *This proposition is proved as the same way of updating U.*

The updating algorithm of $\boldsymbol{\Psi}$ is shown.

Algorithm of updation $\boldsymbol{\Psi}$

Step 1 $\ell \leftarrow 1$

Step 2 Calcurate $\kappa_{of}(\boldsymbol{U}, \boldsymbol{r}^*)$ for $f = 1, 2, \cdots, m$

Step 3 Update ℓth row vector of $\boldsymbol{\Psi}$ based on the rule of Eq. (5.40) among $\kappa_{of}(\boldsymbol{U}, \boldsymbol{r}^*)$ $(o = 1, 2, \cdots, k)$. If the number of f such that $\psi_{of} = 1$ is over 1, one of f is selected randomly.

Step 4 If $\ell = k$ stop, otherwise, $\ell \leftarrow \ell + 1$ and back to Step 2

5.4 Relations between constrained method and existing method

In this section, the relation between constrained hill-climbing model and hill-climbing model, and relation between constrained radius model and radius model are shown.

Proposition 5.4.1 *Relation between constrained hill-climbing model and hill-climbing model*

If $n = k$ for the objective function of constrained hill-climbing model, objective function for constrained hill-climbing model and hill-climbing model are equivalent.

Proof. *From proposition 5.2.1, the difference between the slide-vector model and constrained slide-vector model depend only on models of coordinates. Therefore, in this proposition, we will prove following;*

$$\begin{aligned} &\boldsymbol{X}^\dagger \in \mathcal{H} = \{\boldsymbol{X}^\dagger = \boldsymbol{X} |\ \boldsymbol{X} \in \mathbb{R}^{n \times d}\} \\ \Longleftrightarrow\ &\boldsymbol{X}^\dagger \in \mathcal{CH} = \{\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^* |\ \boldsymbol{X}^* \in \mathbb{R}^{k \times d},\ \boldsymbol{U} \in \mathcal{S}_I(n, k)\} \end{aligned}$$

where

$$\mathcal{S}_I(n, k) = \{\boldsymbol{U} |\ \boldsymbol{U} = (u_{io}),\ \sum_{o=1}^{k} u_{io} = 1\ (i = 1, 2, \cdots, n),$$

$$0 < \sum_{i=1}^{n} u_{io} < n\ (o = 1, 2, \cdots, k)\}$$

$\Longrightarrow$ *We assumed $\boldsymbol{X}^\dagger \in \mathcal{H}$. If $n = k$, there exists $\boldsymbol{U} \in \mathcal{S}_\mathcal{I}(n, n)$ such that $\boldsymbol{I} = \boldsymbol{U}$ from the definition of indicator matrix. Therefore,*

$$(\forall \boldsymbol{X}^\dagger \in \mathcal{H})(\exists \boldsymbol{U} \in \mathcal{S}_\mathcal{I}(n, n))(\exists \boldsymbol{X}^* \in \mathbb{R}^{n \times n})(\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^*)$$

and $\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^* \in \mathcal{CH}$.

$\Longleftarrow$ *We assumed* $\boldsymbol{X}^\dagger \in \mathcal{CH}$*. From definition of* $\mathcal{CH}$*, we have*

$$(\forall \boldsymbol{X}^\dagger \in \mathcal{CH})(\exists \boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k))(\exists \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^*).$$

Then, following property is satisfied;

$$(\forall \boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k))(\forall \boldsymbol{X}^* \in \mathbb{R}^{k\times d})(\exists \boldsymbol{X} \in \mathbb{R}^{n\times d})(\boldsymbol{U}\boldsymbol{X}^* = \boldsymbol{X} \in \mathcal{H}).$$

Therefore, we have the proposition.

Proposition 5.4.2 *Relation between constrained radius model and radius model*

If $n = k = m$ *for the objective function of constrained radius model, objective function for constrained radius model and radius model are equivalent.*

Proof. *From proposition 5.3.1, the difference between the radius and constrained radius models is depend on models of coordinates and parameters of skew-symmetries. Therefore, we will prove following;*

$$\begin{aligned}
&\Big(\boldsymbol{X}^\dagger \in \mathcal{H} = \{\boldsymbol{X}^\dagger = \boldsymbol{X} |\ \boldsymbol{X} \in \mathbb{R}^{n\times d}\} \\
&\Longleftrightarrow \boldsymbol{X}^\dagger \in \mathcal{CH} = \{\boldsymbol{X}^\dagger = \boldsymbol{U}\boldsymbol{X}^* |\ \boldsymbol{X}^* \in \mathbb{R}^{k\times d},\ \boldsymbol{U} \in \mathcal{S}_I(n,k)\}\Big) \qquad (5.41)\\
&\wedge\Big(\boldsymbol{r}^\dagger \in \mathcal{R} = \{\boldsymbol{r}^\dagger = \boldsymbol{r} |\ \boldsymbol{r} \in \mathbb{R}^n\} \\
&\Longleftrightarrow \boldsymbol{r}^\dagger \in \mathcal{CR} = \{\boldsymbol{r}^\dagger = \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^* |\ \boldsymbol{r}^* \in \mathbb{R}^m,\ \boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k),\ \boldsymbol{\Psi} \in \mathcal{S}_{\mathcal{I}}(k,m)\} \qquad (5.42)
\end{aligned}$$

For Eq. (5.41), there is equivalent to proposition 5.4.1.

Next, Eq. (5.42) will be proved.

$\Longrightarrow$ *We assumed* $\boldsymbol{r}^\dagger \in \mathcal{R}$ *and* $n = k = m$*. From the definition of* $\mathcal{S}_{\mathcal{I}}$*,* $\boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k)$ *and* $\boldsymbol{\Psi} \in \mathcal{S}_{\mathcal{I}}(k,m)$ *exist such that* $\boldsymbol{I} = \boldsymbol{U}$ *and* $\boldsymbol{I} = \boldsymbol{\Psi}$*, respectively. Therefore, following property is satisfied;*

$$(\forall \boldsymbol{r}^\dagger \in \mathcal{R})(\exists \boldsymbol{I} = \boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k))(\exists \boldsymbol{I} = \boldsymbol{\Psi} \in \mathcal{S}_{\mathcal{I}}(k,m))(\exists \boldsymbol{r}^* \in \mathbb{R}^m)(\boldsymbol{r}^\dagger = \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*),$$

and $\boldsymbol{r}^\dagger \in \mathcal{CR}$*.*

$\Longleftarrow$ *We assumed* $\boldsymbol{r}^\dagger \in \mathcal{CR}$*. From the definition of* $\mathcal{CR}$*, we have*

$$(\forall \boldsymbol{r}^\dagger \in \mathcal{CR})(\exists \boldsymbol{r}^* \in \mathbb{R}^m)(\exists \boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k))(\exists \boldsymbol{\Psi} \in \mathcal{S}_{\mathcal{I}}(k,m))(\boldsymbol{r}^\dagger = \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^*).$$

Therefore, following property is satisfied;

$$(\forall \boldsymbol{r}^* \in \mathbb{R}^m)(\forall \boldsymbol{U} \in \mathcal{S}_{\mathcal{I}}(n,k))(\forall \boldsymbol{\Psi} \in \mathcal{S}_{\mathcal{I}}(k,m))(\exists \boldsymbol{r} \in \mathbb{R}^n)(\boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^* = \boldsymbol{r})$$

and we have $\boldsymbol{r}^\dagger = \boldsymbol{U}\boldsymbol{\Psi}\boldsymbol{r}^* = \boldsymbol{r} \in \mathcal{R}$*.*

Then, this proposition is satisfied.

From proposition 5.4.1 and proposition 5.4.2, the constrained hill-climbing model and the constrained radius model are considered as generalization of hill-climbing model and radius model, respectively.

Chapter 6

Simulation studies

In this section, the clustering results of the proposed methods, namely, constrained Unfolding, constrained slide-vector model, constrained hill-climbing model, and constrained radius model are shown through numerical simulations. For these methods, the asymmetric dissimilarity data are assumed to have different structures because these are different kinds of models. Therefore, three kinds of numerical simulations are conducted each for the constrained slide-vector model, the constrained hill-climbing model, and the constrained radius model.

6.1 Simulation of constrained AMDS based on Unfolding

In this section, the experimental design of the simulation performed to verify the clustering results of constrained unfolding and constrained slide-vector model is shown. Next, the clustering results are compared based on the adjusted rand index (ARI) values (Hubert and Arabie, 1985).

Here, ARI is defined as follows.

Definition 6.1.1 *Adjusted Rand Index*

Given two clustering structure $\mathcal{C} = \{C_1,\ C_2, \cdots, C_k\}$ *and* $\mathcal{V} = \{V_1,\ V_2, \cdots\ V_k^*\}$ *of objects* $I = \{1, 2, \cdots, n\}$ *such that*

$$I = \bigcup_{o=1}^{k} C_o, \quad C_o \cap C_\ell = \phi\ (o \neq \ell), \quad I = \bigcup_{o=1}^{k^*} V_o \quad \text{and} \quad V_o \cap V_\ell = \phi\ (o \neq \ell),$$

ARI is defined as follows:

$$\frac{\sum_{o=1}^{k}\sum_{\ell=1}^{k^*}\binom{n_{o\ell}}{2} - \Big[\sum_{o=1}^{k}\binom{n_{o\cdot}}{2}\sum_{\ell=1}^{k^*}\binom{n_{\cdot\ell}}{2}\Big]/\binom{n}{2}}{\frac{1}{2}\Big[\sum_{o=1}^{k}\binom{n_{o\cdot}}{2} + \sum_{\ell=1}^{k^*}\binom{n_{\cdot\ell}}{2}\Big] - \Big[\sum_{o=1}^{k}\binom{n_{o\cdot}}{2}\sum_{\ell=1}^{k^*}\binom{n_{\cdot\ell}}{2}\Big]/\binom{n}{2}}$$

where $n_{o\ell}$ *is the number of objects that are belonging to cluster* C_o *and* V_ℓ, $n_{o\cdot}$ *is the number of objects that are belonging to cluster* C_o *and* $n_{\cdot\ell}$ *is the number of objects that are belonging to cluster* V_ℓ.

6.1.1 Experimental design of the simulation

In the simulation, artificial asymmetric dissimilarity data with true clustering structure is generated and then the method is applied to the asymmetric dissimilarity data. Then, the ARI between the true clustering structure and the clustering result obtained is calculated. In the simulation, the artificial asymmetric dissimilarity data with true clustering structure

$\boldsymbol{\Delta}^{\ddagger}$ is generated as follows:

$$\boldsymbol{\Delta}^{\ddagger} = \boldsymbol{D}(\boldsymbol{X}^{\dagger}, \boldsymbol{X}^{\dagger} - \boldsymbol{U}^{\dagger}\boldsymbol{\Psi}^{\dagger}\boldsymbol{Z}^{\dagger})$$

where $\boldsymbol{X}^{\dagger} = (x_{it}^{\dagger})$ $x_{it}^{\dagger} \in \mathbb{R}$ $(i = 1, 2, \cdots, 150; t = 1, 2)$ denotes the coordinates of the objects with the true clustering structure, $\boldsymbol{U}^{\dagger} = (u_{io})$ $u_{io} \in \{0, 1\}$ $(i = 1, 2, \cdots, 150; o = 1, 2, \cdots, k)$ is the indicator matrix for objects corresponding to the true clustering structure, $\boldsymbol{\Psi}^{\dagger} = (\psi_{of})$ $\psi_{of} \in \{0, 1\}$ $(o = 1, 2, \cdots, k; f = 1, 2, \cdots, m)$ is the indicator matrix for centroids with the true clustering structure and $\boldsymbol{Z}^{\dagger} = (z_{ft})$ $z_{ft} \in \mathbb{R}^{m \times d}$ denotes the slide-vectors with the true clustering structure.

The method of generating these parameters is described below. In the simulation, the number of objects is set as 150 and the number of dimension is set as 2. $\boldsymbol{X}^{\dagger}$ is generated as follows:

$$\begin{aligned} &{}^{(o)}x_{i1}^{\ddagger} \sim N(\mu_{o1}, 0.5) \ (i = 1, 2, \cdots, n_o; \ o = 1, 2, \cdots, k) \\ &\mu_{o1} = 0, \quad \mu_{\ell 1} = \mu_{(\ell-1)1} + 1 \ (\ell = 2, 3, \cdots, k) \end{aligned}$$

where ${}^{(o)}x_{i1}^{\ddagger}$ denotes the cooridnates of object i belonging to cluster o in 1st dimension, n_o is the number of objects belonging to cluster o, N is normal distribution and μ_o is the mean of cluster o. The way of generating the way of the cooridnates of the objects in the second dimension depends on the estimated coordinates of the objects in the first dimension.

$$\begin{aligned} &{}^{(o)}x_{i2}^{\ddagger} \sim N(\mu_{o2}, 0.5) \ (i = 1, 2, \cdots, n_o; \ o = 1, 2, \cdots, k) \\ &\mu_{o2} \sim U(x_{min}, \frac{2}{3}(x_{max} - x_{min})) \ (o = 1, 2, \cdots, k) \end{aligned}$$

where x_{min} and x_{max} denote the minimum and maximum values of the coordinates of the objects in the 1st dimension, respectively and U denotes uniform distribution. $\boldsymbol{U}^{\ddagger}$ is generated based on Factor 1 and $\boldsymbol{\Psi}^{\ddagger}$ is generated at random subject to whether the constraint of indicator matrix is satisfied. Finally, the slide vectors are generated as follows:

$$\boldsymbol{Z}^{\dagger} = (z_{ft}) \ z_{ft} \sim U(-1, 1) \ (f = 1, 2, \cdots, m; t = 1, 2)$$

To add the true clustering structure to $\boldsymbol{X}^{\dagger}$, $\boldsymbol{U}^{\dagger}$, $\boldsymbol{\Psi}^{\dagger}$ and $\boldsymbol{Z}^{\dagger}$, five factors in this simulation are set as shown Table. 6.1. The simulation design is based on the studies of Milligan (1985) and Milligan and Cooper (1988).

Table 6.1: Factors of numerical simulation for AMDS based on Unfolding

	name of these factors	the number of levels
Factor 1	cluster sizes	3
Factor 2	the number of clusters	3
Factor 3	the true number of slide-vectors	2 or 3 or 5
Factor 4	noise of dissimilarities	2
Factor 5	Methods	3

Next, we describe these five factors.

Factor 1: cluster sizes

Here, cluster size is defined as the number of objects belonging to the same cluster. In this simulation, three levels are set based on Milligan and Cooper (1988) as follows:

level 1: The cluster sizes are equal

level 2: One cluster has half the number of objects, and the other clusters have the same number of objects.

level 3: One cluster has 20% of the number of objects, and the other clusters have the same number of objects.

From these levels, the true clustering structure includes three combinations of cluster sizes. In level 2, if the number of clusters is small, each cluster will be almost of the same size. On the other hand, if the number of clusters is large, the cluster sizes are different from each other.

Factor 2: number of clusters

There are three levels in Factor 2 where the number of clusters are 2, 3, or 5.

Factor 3: number of slide-vectors

In this factor, the number of levels depends on the number of clusters. When the number of clusters is 2, 3 and 5, the number of levels will be 2, 3 and 5, respectively.

Factor 4: noise of dissimilarities

In the factor, there are two levels as follows:

level 1: There is no noise

level 2: Here, 30% of the dissimilarities, i.e. 6750 dissimilarities, out of $\boldsymbol{\Delta}^{\ddagger}$ are replaced with $\delta_{ij}^{\ddagger} \sim U(0, 15)$

The purpose of the Factor 4 is to reveal the effect of noise in the proposed methods.

Factor 5: Methods

In this factor, there are three levels, namely, such as constrained Unfolding, constrained hill-climbing model and tandem clustering for Unfolding and constrained two-mode clustering. For two mode clustering, refer to Van Mechelen, et al. (2004).

Here, the tandem clustering method is described, which consists of the following two steps.

Tandem clustering

Step 1: Unfolding is applied to asymmetric dissimilarity data $\boldsymbol{\Delta}^{\ddagger}$ and distance matrix $\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$ is obtained such that the values of the objective function are minimized.

Step 2: Constrained clustering is applied to $\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$ estimated in Step 1 as follows:

$$\|\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y}) - \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^T\|^2$$

where $\boldsymbol{U} = (u_{io})$ $u_{io} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; o = 1, 2, \cdots, k)$ denotes the indicator matrix and $\boldsymbol{\Gamma} = (\gamma_{o\ell})$ $(o, \ell =, 1, 2, \cdots, k)$ denotes the centorid matrix and $\boldsymbol{U}$ and $\boldsymbol{\Gamma}$ are estimated.

Finally, number of combinations is 180 = 3 levels (Factor 1) × 2 levels (Factor 4) × 3 levels (Factor 5) × (2 levels + 3 levels + 5 levels (Factor 2 × Factor 3)). For each element of these combinations, asymmetric dissimilarity data is generated 50 times and the number of initial values is 10.

6.1.2 Simulation results

In this subsection, the results corresponding to the ARI are shown. In particular, the boxplot for the proposed method and tandem clustering are shown to verify the effect of constrained unfolding and constrained slide-vector model. Remember constrained unfolding is equal to the constrained hill-climbing model with $m = 1$. Next, the effect of each factor is shown.

Figure 6.1 shows the results of the constrained slide-vector model and tandem clustering for all situations. The median of ARI corresponding to the constrained slide-vector model is higher than that for tandem clustering. In addition, the range of the ARI corresponding to the constrained slide-vector model is lower than that of tandem clustering. From this, we can confirm that the clustering results of the constrained slide-vector model are superior to those of tandem clustering. Next, the effects of the cluster size are shown in Figure 6.2. For all the levels, the results of the constrained slide-vector model are better than those of tandem clustering. For the number of clusters, see Figure 6.3. For both the constrained slide-vector model and tandem clustering, the clustering results tend degrade if the number of clusters is large. The tendency is the same as that observed by Milligan and Cooper (1988) and Steinley (2004). Next, the effects of the number of slide vectors on the clustering results are shown. First, the effects of the number of slide-vectors, when the number of clusters is 2 , are shown in Figure 6.4. From Figure 6.4, it can be observed that if the number of slide-vectors is large, the clustering results degrade. The reason

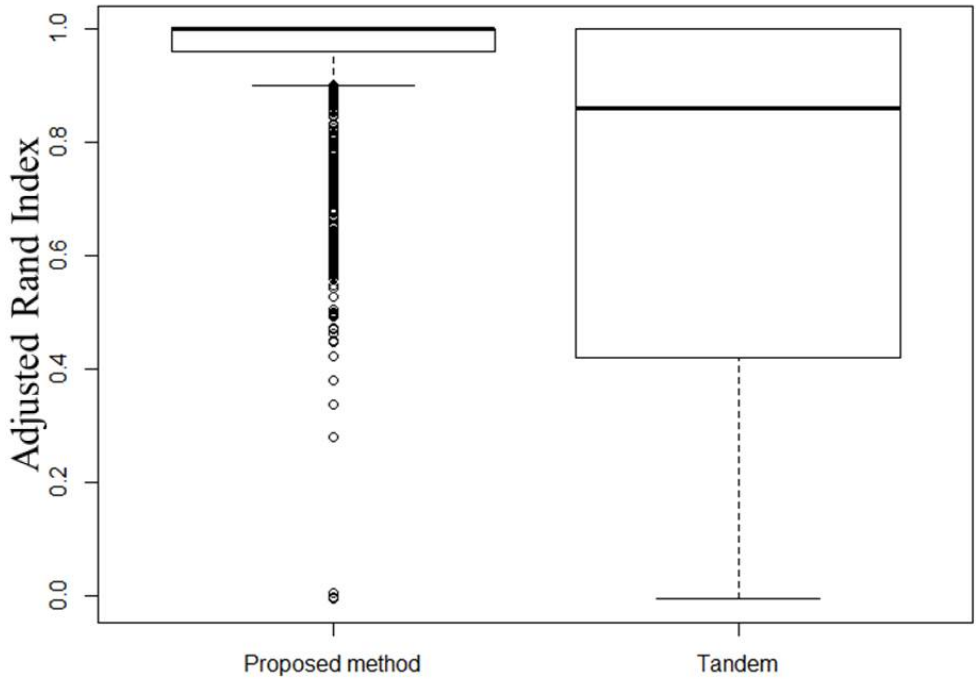

Figure 6.1: Clustering results for constrained hill climbing model and the tandem clustering

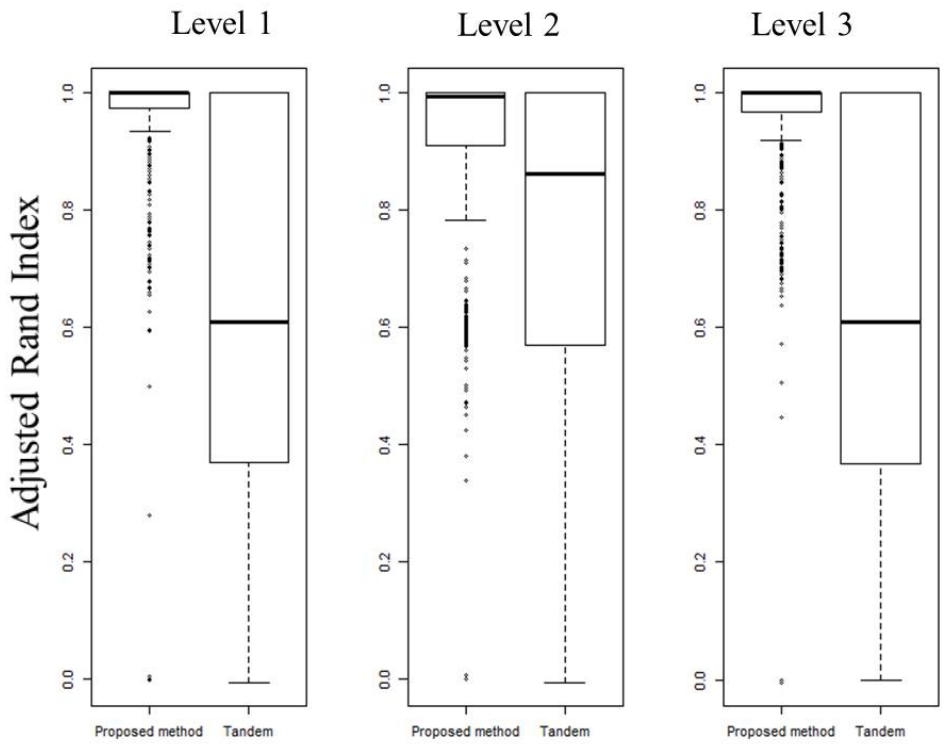

Figure 6.2: Clustering results for Factor 1 in the simulation

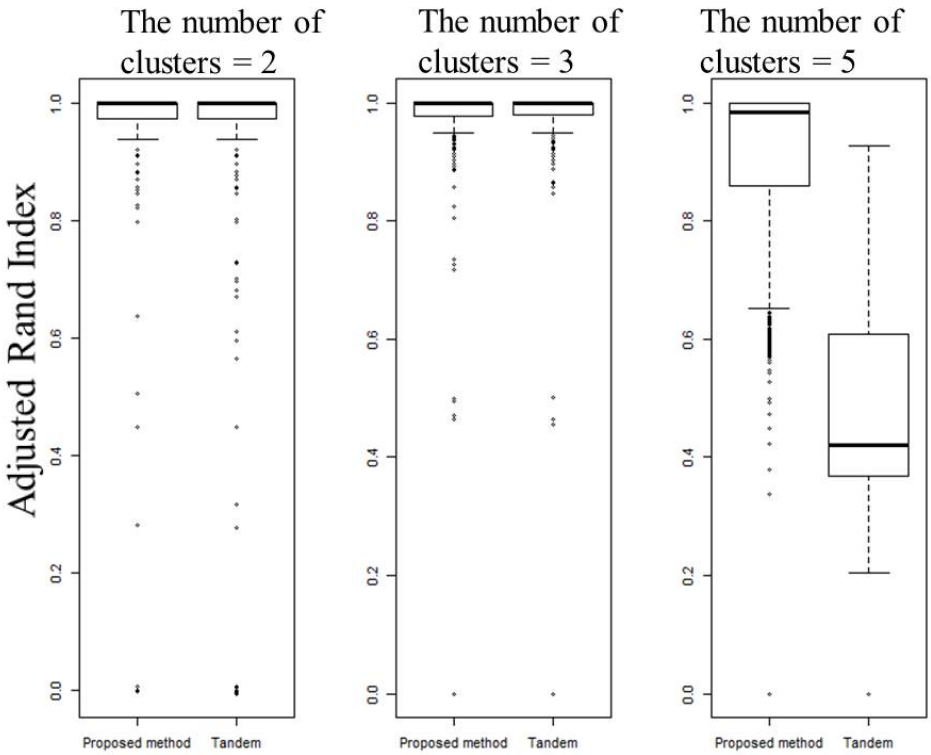

Figure 6.3: Clustering results for Factor 2 in the simulation

is that if the number of slide-vectors is large for a true structure, the structure of asymmetric dissimilarity data will no longer be simple and these clustering results become unstable. However, when the number of clusters are and , these clustering results are not affected by the number of true slide vectors, as shown in Figure 6.5 and Figure 6.6, respectively. This is because the differences in the true asymmetric dissimilarities between clusters increase for the data generated in this simulation if the number of clusters is large.

Finally, the effects on the clustering results due to noise are shown in Figure 6.7. From Figure 6.7, it can be seen that the clustering results of the constrained slide-vector model are robust to noise dissimilarities. On the other hand, the results of tandem clustering are affected by the noise dissimilarities. This is because in the constrained slide-vector model, the coordinates of clusters are estimated based on Sokal and Michener dissimilarities. That is, estimated Sokal and Michener dissimilarities tend to be close to the true Sokal and Michener dissimilarities even if some of these asymmetric dissimilarities include noise.

6.2 Simulation of the constrained hill climbing model

In this section, the clustering results of the constrained hill-climbing model are shown. The experimental design is described as first and these clustering results are

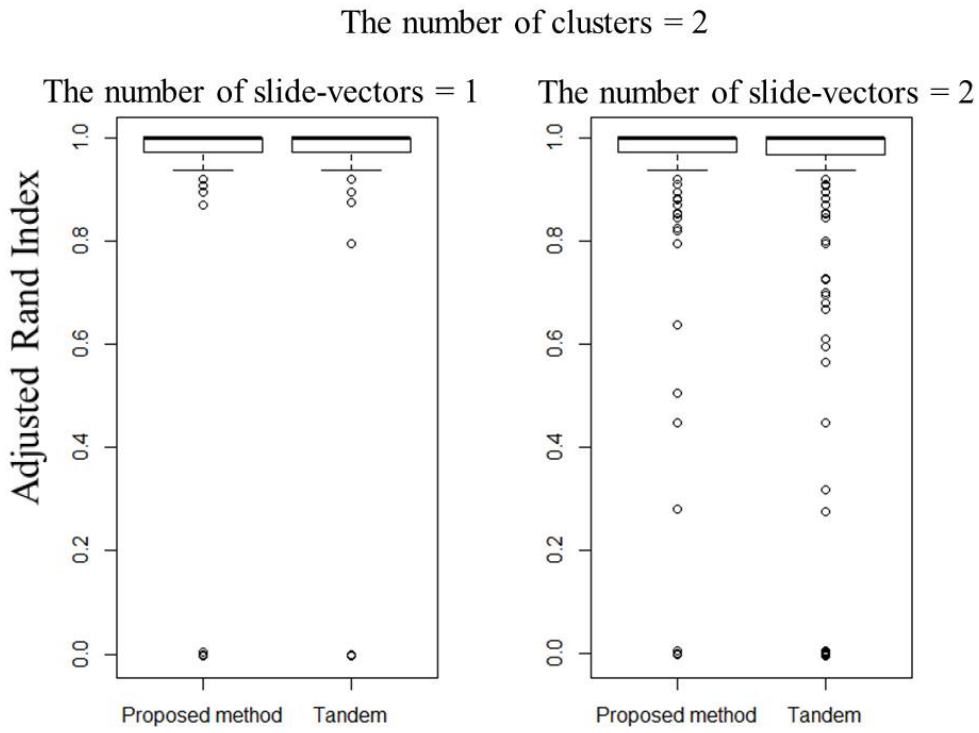

Figure 6.4: Clustering results for Factor 3 in $k = 2$

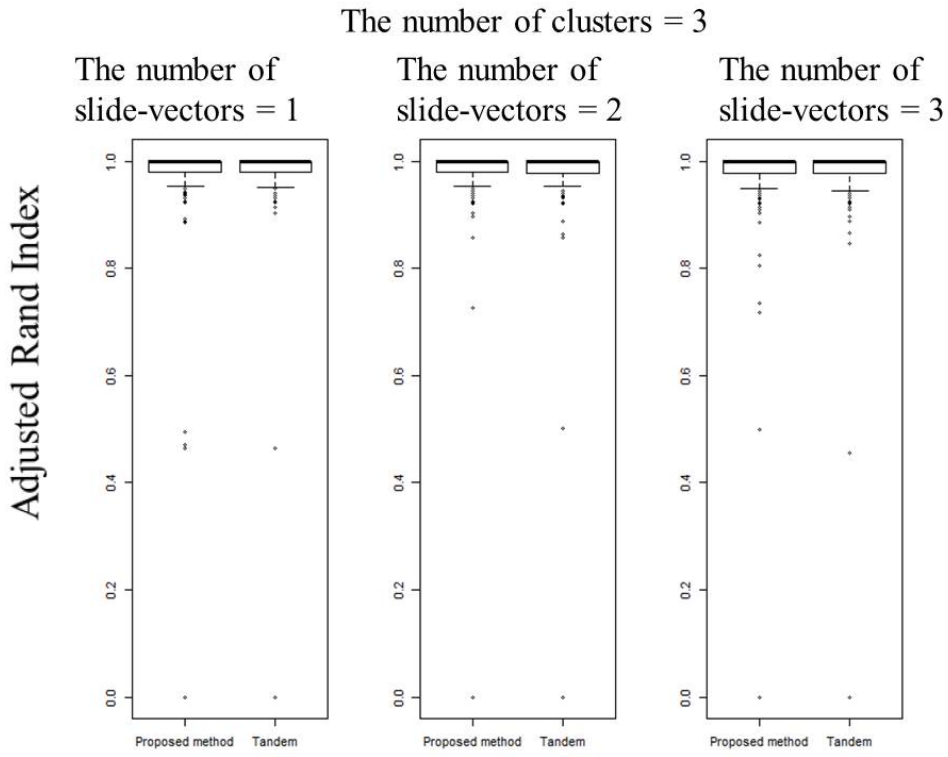

Figure 6.5: Clustering results for Factor 3 in $k = 3$

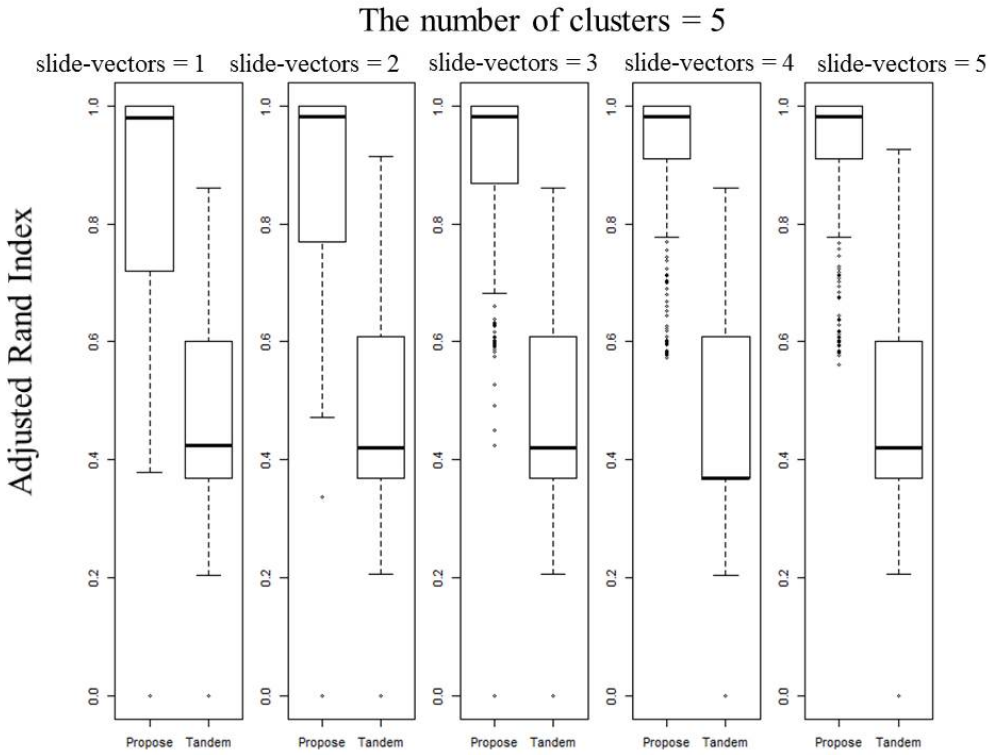

Figure 6.6: Clustering results for Factor 3 in $k = 5$

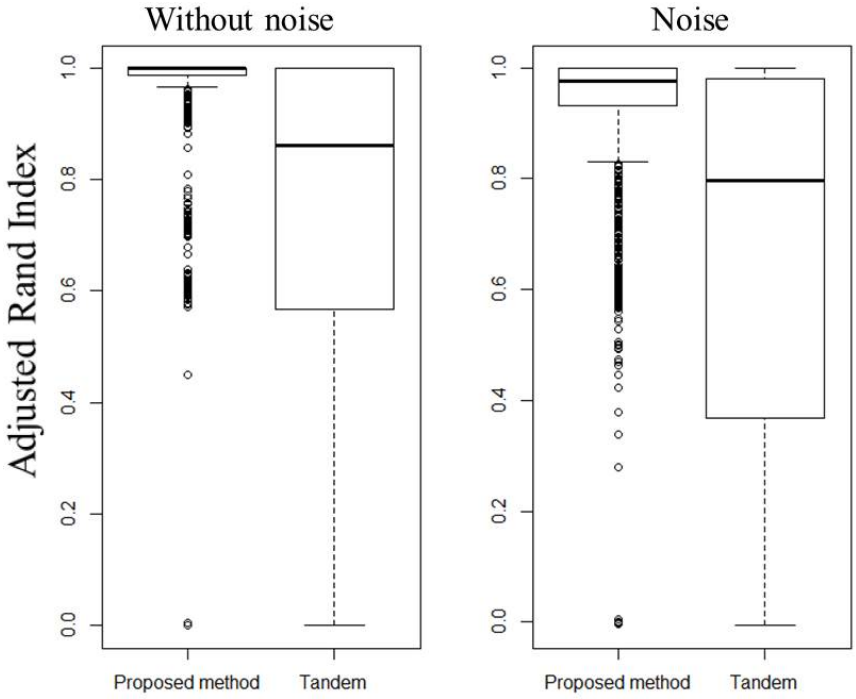

Figure 6.7: Clustering results for Factor 4

shown corresponding to the ARI.

6.2.1 Experimental design of the simulation

Similar to Section 6.1, artificial asymmetric dissimilarity data with a true clustering structure is generated and then the constrained hill-climbing model is applied to the asymmetric dissimilarity data to verify the effects of the constrained hill-climbing model. In the simulation, ARI is adopted to verify the clustering results.

The number of objects is set as 90 and the number of dimensions is set as 2. The artificial asymmetric dissimilarity data $\boldsymbol{\Delta}^{\ddagger} = (\delta_{ij}^{\ddagger})$ $(i, j = 1, 2, \cdots, 90)$ is generated as follows:

$$\delta_{ij}^{\ddagger} = d_{ij}(\boldsymbol{X}^{\dagger}) + (\boldsymbol{x}_i^{\dagger} - \boldsymbol{x}_j^{\dagger})^T \boldsymbol{v} d_{ij}(\boldsymbol{X}^{\dagger})^{-1}$$

where $\boldsymbol{X}^{\dagger} = (\boldsymbol{x}_1^{\dagger}, \boldsymbol{x}_2^{\dagger}, \cdots, \boldsymbol{x}_{90}^{\dagger})^T = (x_{it}^{\dagger})$ $(i = 1, 2, \cdots, 90; t = 1, 2)$ denote the coordinates of objects with the true clustering structure and $\boldsymbol{v} = (v_t)$ $v_t \in \mathbb{R}$ $(t = 1, 2)$ denotes the true slope vector. The method for generating $\boldsymbol{X}^{\dagger}$ is the same as that described in Section 6.1. We generate $\boldsymbol{v}$ is as follows:

$$v_t \sim N(0, 1) \quad (t = 1, 2).$$

In addition, Factor 1, Factor 2, Factor 4, and Factor 5, shown in Table 3, are adopted in this simulation. While Factor 1, Factor 2, and Factor 4 are the same as those considered in the simulation of the constrained slide-vector model, the content of Factor 5 is different because the constrained hill-climbing model is different from the constrained slide-vector model.

Tandem clustering for the constrained hill-climbing model

Step 1: hill-climbing model is applied to the asymmetric dissimilarity data $\delta_{ij}^{\ddagger}$ $(i, j = 1, 2, \cdots, n)$ and get $d_{ij}(\boldsymbol{X}) + (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v}$ such that values of the objective function is minimized

Step 2: Constrained clustering is applied to $d_{ij}(\boldsymbol{X}) + (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v}$ estimated in Step 1 as follows:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(d_{ij}(\boldsymbol{X}) + (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{v} - \sum_{o=1}^{k} \sum_{\ell=1}^{k} u_{io} u_{j\ell} \gamma_{o\ell} \right)^2$$

where $\boldsymbol{U} = (u_{io})$ $u_{io} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; o = 1, 2, \cdots, k)$ denotes the indicator matrix of objects and $\boldsymbol{\Gamma} = (\gamma_{o\ell})$ $\gamma_{o\ell} \in \mathbb{R}$ $(o, \ell = 1, 2, \cdots, k)$ denotes the centroid matrix.

Finally, the number of combinations is 36 = 3levels (Factor 1) × 3 levels (Factor 3) × 2 levels (Factor 4) × 2 levels (Factor 5). For each element of these combinations, asymmetric dissimilarity data is generated 50 times and the number of initial values is 5.

6.2.2 Simulation results

In this subsection, clustering results of the constrained hill climbing model are compared to those of tandem clustering based on ARI. Figure 6.8 shows the clustering results of the constrained hill-climbing model and tandem clustering for all situations. The median and hinge values are almost the same.

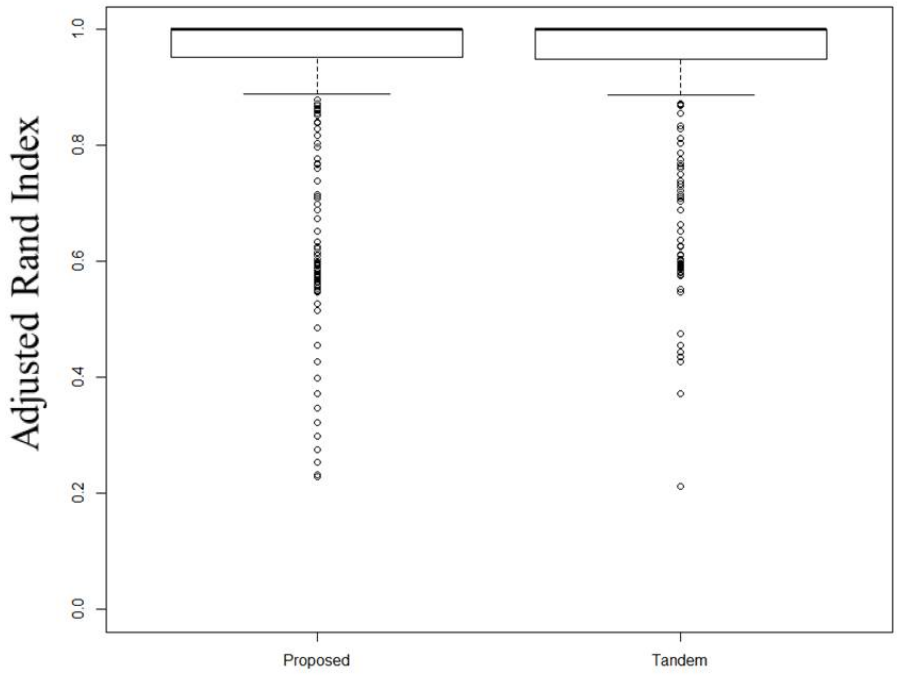

Figure 6.8: Clustering results of the constrained hill climbing model

Next, the effects of cluster sizes are verified as shown in Figure 6.9. For level 1, the results of the constrained hill-climbing model are stable, although the results of the tandem clustering are not stable. However, in levels 2 and 3, the clustering results of the constrained hill-climbing model are not stable.

Next, the effects of the number of clusters are shown in Figure 6.10. From Figure 6.10, the clustering results of tandem clustering degrade if the number of clusters is large, while the clustering results of the constrained hill-climbing model are stable.

Finally, the effects on clustering results due to noise are shown in Figure 6.11. In conditions when there is noise, the clustering results of tandem clustering are stable, whereas those of the constrained hill climbing model are not stable.

6.3 Simulation of the constrained radius model

In this section, experimental design and the results of the constrained radius model are shown similar to Section 6.1 and Section 6.2.

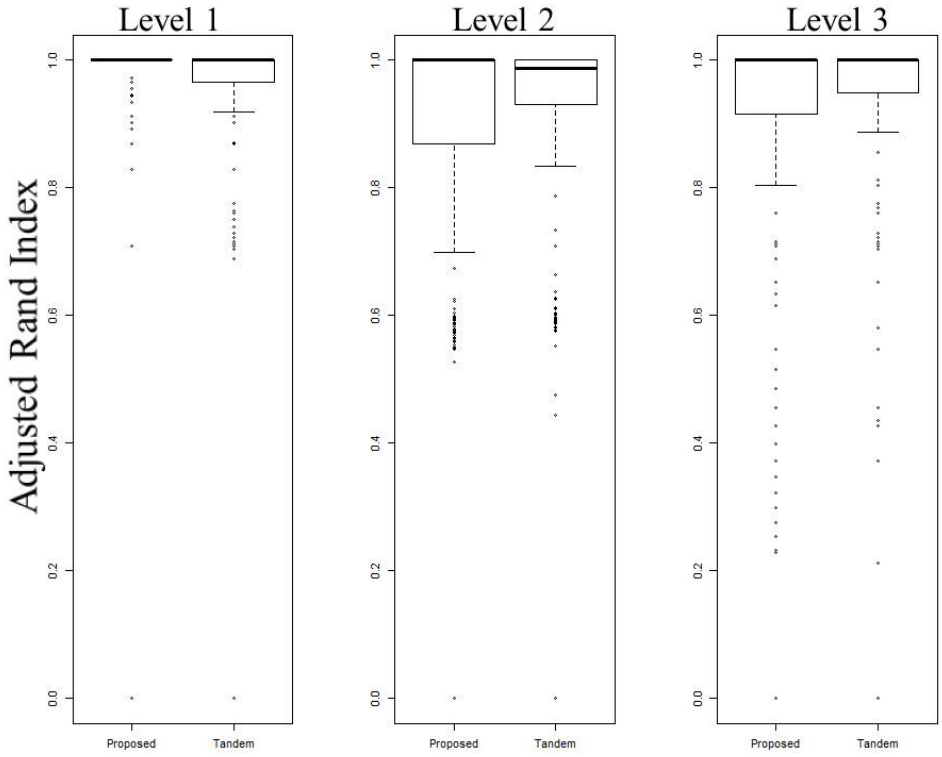

Figure 6.9: Clustering results of the constrained hill climbing model for Factor 1

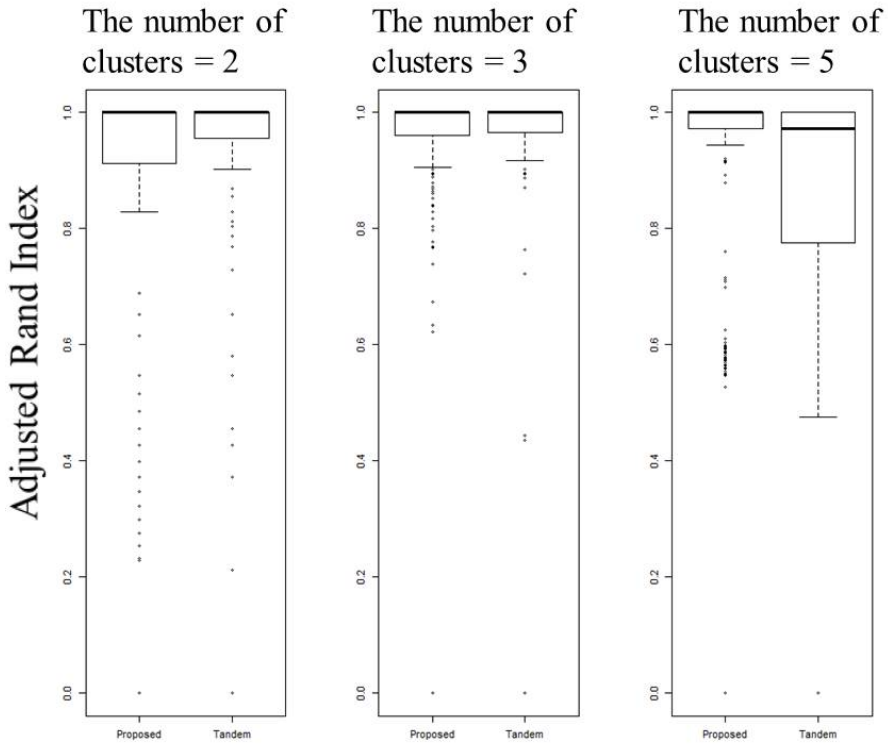

Figure 6.10: Clustering results of the constrained hill climbing model for Factor 2

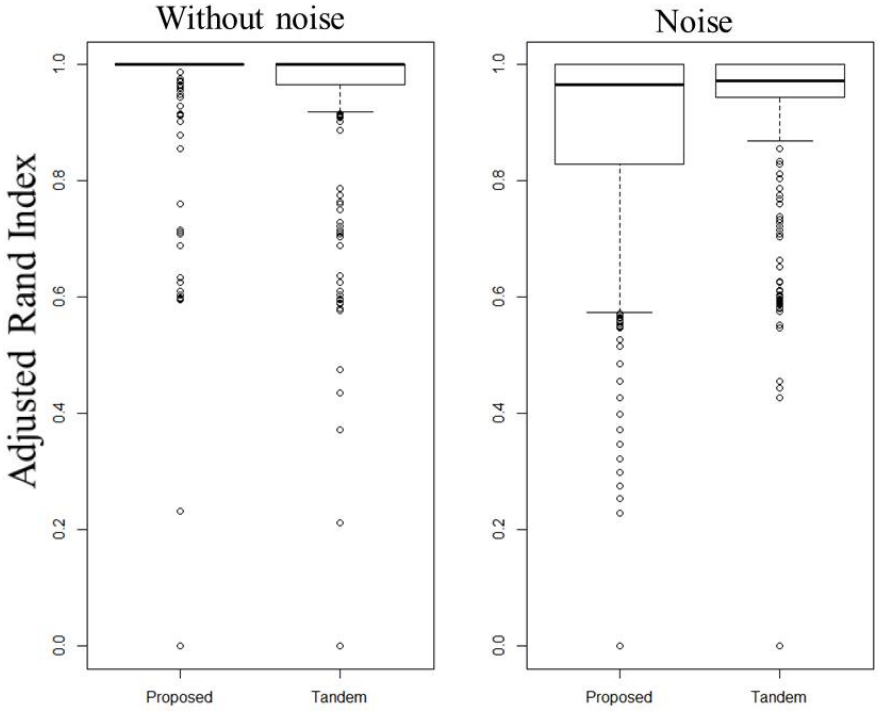

Figure 6.11: Clustering results of the constrained hill climbing model for Factor 4

6.3.1 Experimental design of the simulation

In this simulation, artificial asymmetric dissimilarity data with the true structure is generated and then the constrained radius model is applied to the asymmetric dissimilarity data.

In the simulation, the artificial asymmetric dissimilarity data with the true structure $\boldsymbol{\Delta}^{\ddagger}$ is generated as follows:

$$\boldsymbol{\Delta}^{\ddagger} = \boldsymbol{D}(\boldsymbol{X}^{\dagger}) - \mathbf{1}_n \boldsymbol{r}^{\dagger T} \boldsymbol{\Psi}^{\dagger T} \boldsymbol{U}^{\dagger T} + \boldsymbol{U}^{\dagger} \boldsymbol{\Psi}^{*} \boldsymbol{r}^{\dagger} \mathbf{1}_n^T$$

where $\boldsymbol{X}^{\dagger} = (x_{it}^{\dagger})$ $x_{it}^{\dagger} \in \mathbb{R}$ $(i = 1, 2, \cdots, 90; t = 1, 2)$ denotes the coordinates of objects with the true clustering structure, $\boldsymbol{r}^{\dagger} = (r_f^{\dagger})$ $r_f^{\dagger} \leq 0$ $(f = 1, 2, \cdots, m)$, denotes radii for the cluster centorids, $\boldsymbol{U}^{\dagger} = (u_{io}^{\dagger})$ $u_{io}^{\dagger} \in \{0, 1\}$ $(i = 1, 2, \cdots, 90; o = 1, 2, \cdots, k)$ denotes the indicator matrix for objects, $\boldsymbol{\Psi}^{\dagger} = (\psi_{\ell f}^{\dagger})$ $\psi_{\ell f}^{\dagger} \in \{0, 1\}$ $(\ell = 1, 2, \cdots, k; f = 1, 2, \cdots, m)$ denotes the indicator matrix for the centroids and $\mathbf{1}_n = (1)$ denotes a vector of n length.

The method of generating $\boldsymbol{X}^{\dagger}$, $\boldsymbol{U}^{\dagger}$ and $\boldsymbol{\Psi}^{\dagger}$ is the same as that shown in section 6.1. Here, $\boldsymbol{r}^{\dagger}$ is generated as follows:

$$\boldsymbol{r}^{\dagger} = (r_f) \quad r_f \sim U(0.5, 1.5) \quad (f = 1, 2, \cdots, m).$$

In addition, all factors, i.e., Factor 1, Factor 2, Factor 3, Factor 4 and Factor 5 shown in Table 6.1 are adopted in this simulation. However, tandem clustering in

the constrained radius model is different from the tandem clustering in both the constrained slide-vector model and the constrained hill-climbing mode. We now describe tandem clustering for the constrained radius model.

Tandem clustering for the constrained radius model

Step 1: Radius model is applied to the asymmetric dissimilarity data $\boldsymbol{\Delta}^{\ddagger}$ and $\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T$ is obtained such that the values of the objective function are minimized.

Step 2: Constrained clustering is applied to $\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T$ estimated in Step 2 as follows,

$$\|\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}_n \boldsymbol{r}^T + \boldsymbol{r}\mathbf{1}_n^T - \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^T\|^2$$

where $\boldsymbol{U} = (u_{io})$ $u_{io} \in \{0, 1\}$ $(i = 1, 2, \cdots, n; o = 1, 2, \cdots, k)$ denotes the indicator matrix and $\boldsymbol{\Gamma} = (\gamma_{o\ell})$ $\gamma_{o\ell} \in \mathbb{R}$ $(o, \ell = 1, 2, \cdots, k)$ denotes the centroid matrix.

Finally, the number of the combinations is 180 = 3 levels (Factor 1) × 2 levels (Factor 4) × 3 levels (Factor 5) × (2 levels + 3 levels + 5 levels (Factor 2 × Factor 3)). For each element of these combinations, the asymmetric dissimilarity data is generated 50 times and the number of initial values is 5.

6.3.2 Simulation results

In this subsection, the clustering results for the constrained radius model corresponding to ARI are shown similar to Section 6.1.2 and 6.2.2.

Figure 6.12 shows the clustering results for the constrained radius model and the tandem clustering for all situations. From Figure 6.12, it can be seen that the clustering results of the constrained radius model are superior to those of tandem clustering in terms of both median and range.

Figure 6.13 shows the effects of cluster sizes on the clustering results. For any levels, the results of the constrained radius model are superior to those of tandem clustering; however, the results of the constrained radius model are affected in levels 2 and 3.

Figure 6.14 shows the effects of the number of clusters on the clustering results. it can be seen from the figure that the results of the constrained radius model are superior to those of tandem clustering. When the number of clusters is large, the results of tandem clustering become worse and the tendency is the same as that described in Milligan and Cooper (1988).

The effects of the number of slide vectors on the clustering results are shown in Table 6.15, Table 6.16, and Table 6.17. The clustering results of the constrained radius model are very good. In addition, the clustering results for both are not affected by the number of radii.

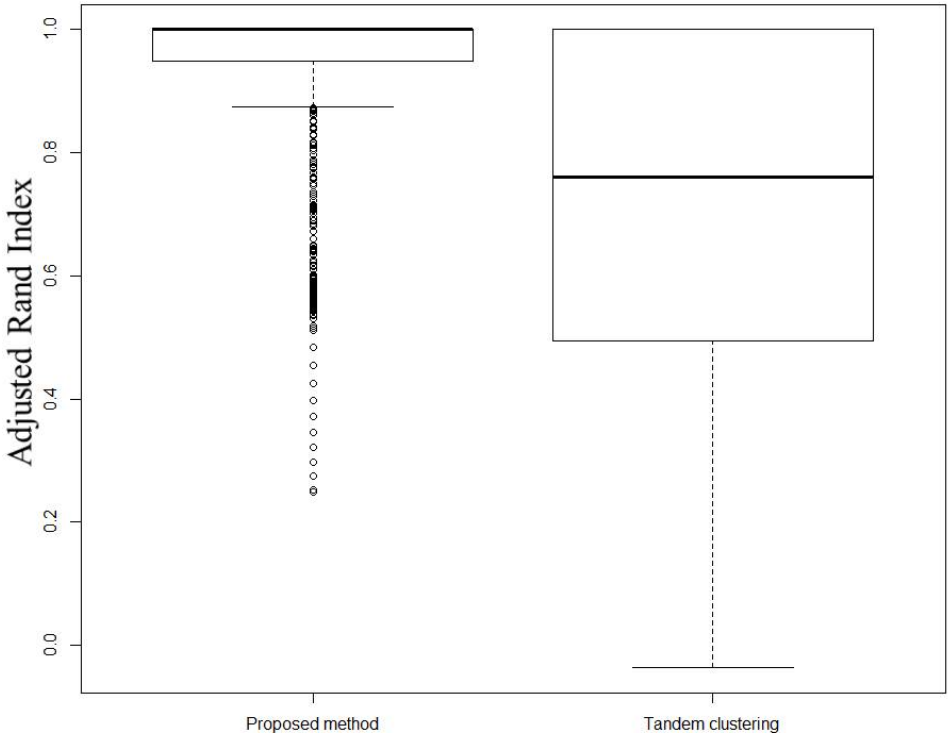

Figure 6.12: Clustering results of the constrained radius model

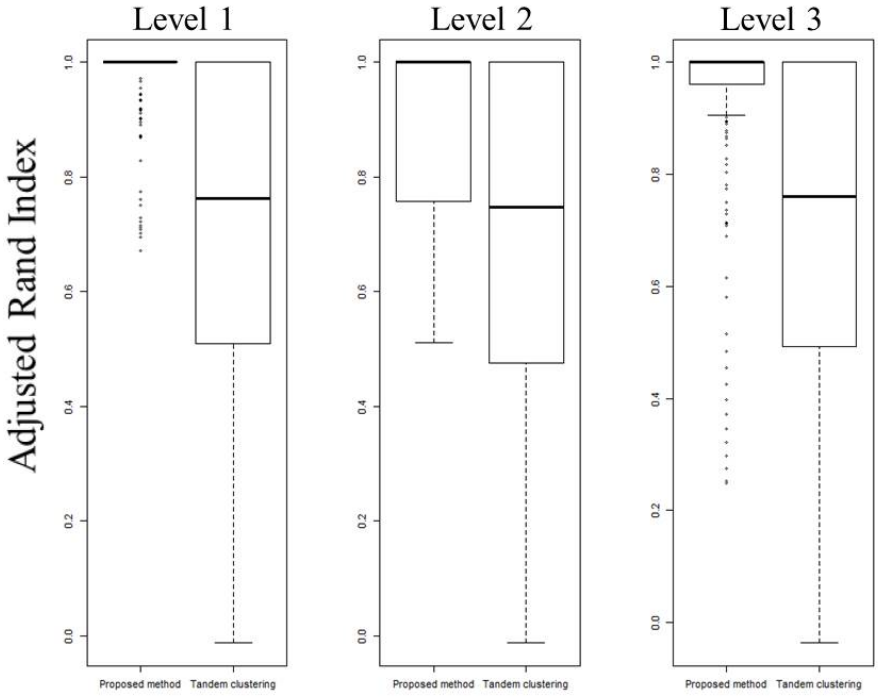

Figure 6.13: Clustering results of the constrained radius model for Factor 1

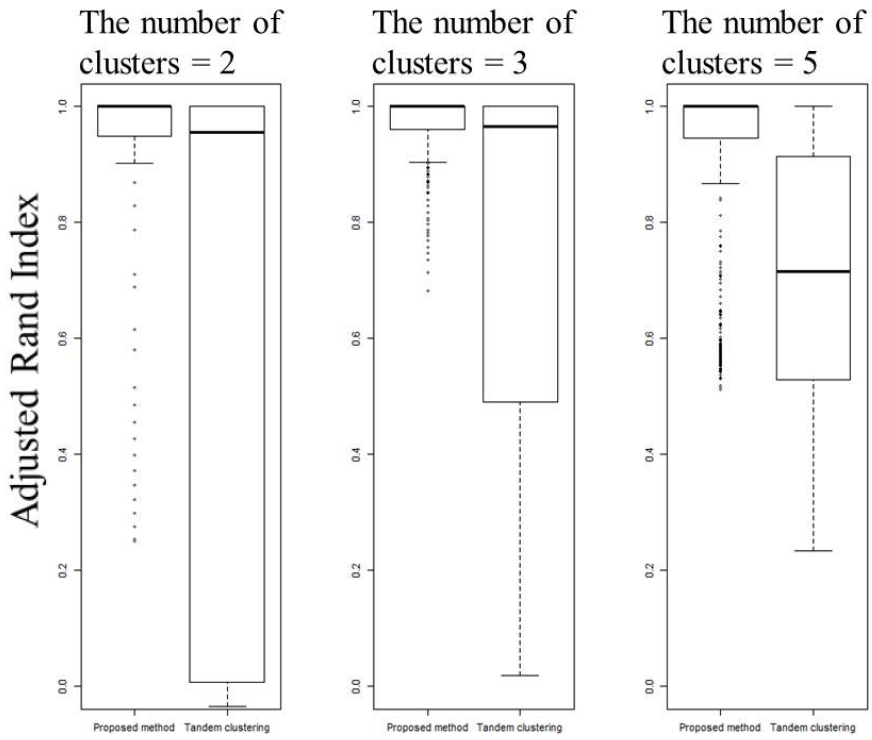

Figure 6.14: Clustering results of the constrained radius model for Factor 2

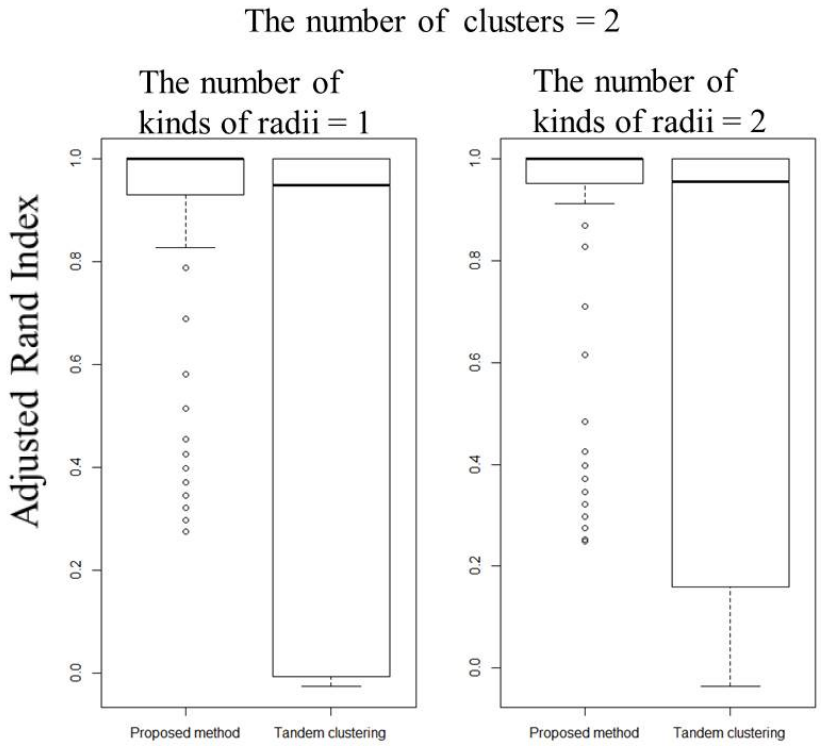

Figure 6.15: Clustering results of the constrained radius model for Factor 3 for $k = 2$

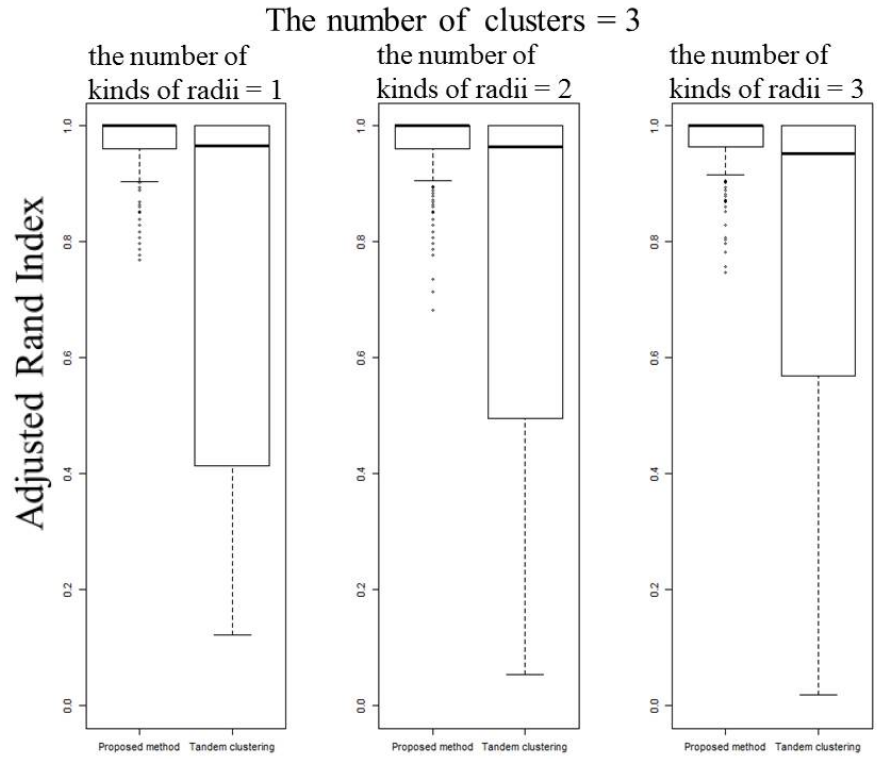

Figure 6.16: Clustering results of the constrained radius model for Factor 3 for $k = 3$

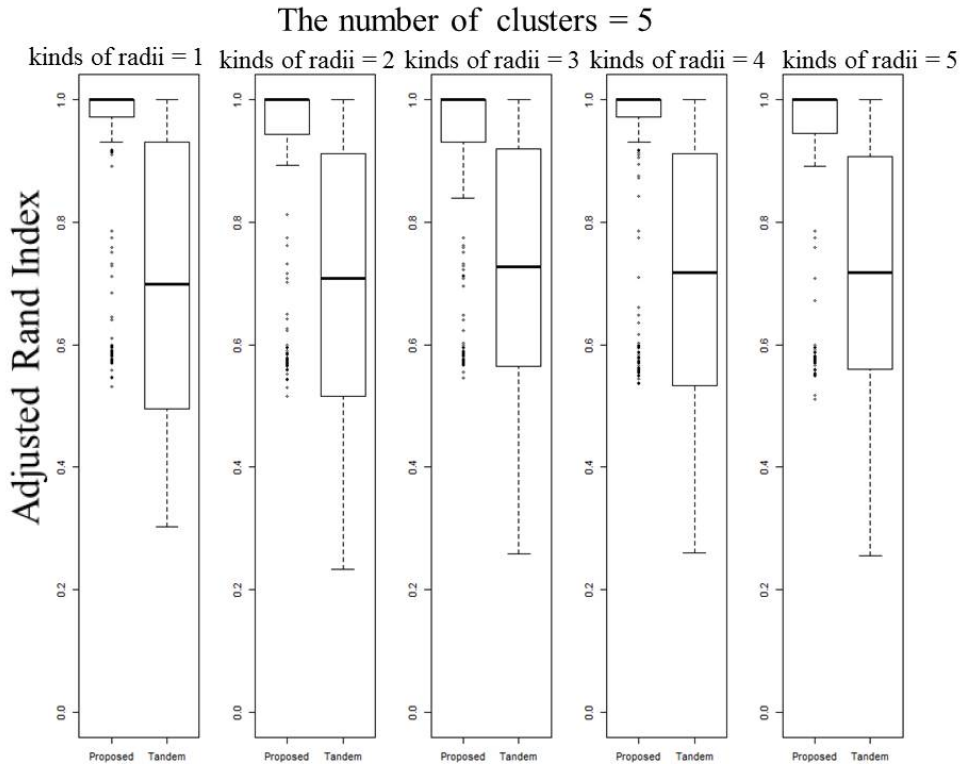

Figure 6.17: Clustering results of the constrained radius model for Factor 3 for $k = 5$

Finally, effects of noise for these clustering results are shown in Figure 6.18. From Figure 6.18, it can be seen that the results of the constrained radius model are good and are not affected by noise. Hence, the constrained radius model is robust to noise.

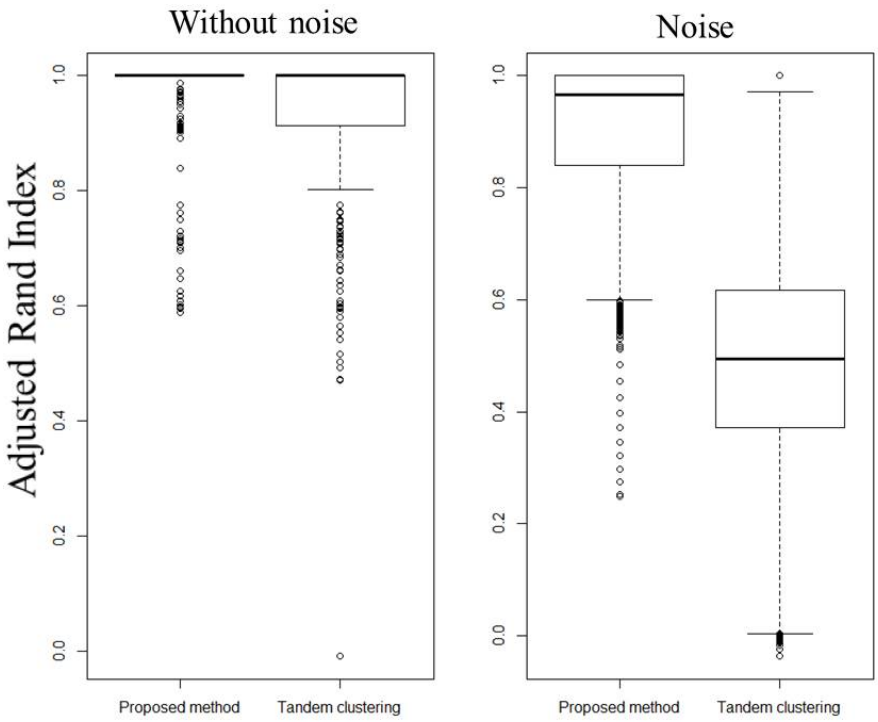

Figure 6.18: Clustering results of the constrained radius model for Factor 4

Chapter 7

Real examples

In this section, the results for real data when using the constrained slide-vector model, the constrained hill-climbing model, and the constrained radius model are shown.

7.1 Data description

Here, we evaluate the methods for a real application, namely, the switching data of Japanese tea bottles. The data were generated from the scan panel data collected in Tokyo in July 2012; this data was provided by MACROMILL, Inc. The 25 brands of Japanese tea bottles are listed in Table 7.1, and the switching data are presented in Table 7.2. Each entry in Table 7.2 τ_{ij} $(i, j = 1, 2, \cdots, 25)$ corresponds to the frequency of switching from brand i to another brand j. The switching data are considered to be the similarities. However, AMDS requires asymmetric dissimilarity data and not similarity data. Therefore, brand switching data are converted into asymmetric dissimilarity data by using the gravity model (Tobler and Wineburg, 1971) as follows:

$$\delta_{ij} = \left(\frac{\tau_{i\cdot}\tau_{\cdot j}}{\tau_{ij} + 0.1} \right) \ (i, j = 1, 2, \cdots, 25)$$

where $\tau_{i\cdot} = \sum_{j=1}^{25} \tau_{ij}$ $(i = 1, 2, \cdots, 25)$ and $\tau_{\cdot j} = \sum_{i=1}^{25} \tau_{ij}$ $(j = 1, 2, \cdots, 25)$.

In earlier studies, for the application of AMDS to brand switching data, the gravity model was used (Borg and Groenen, 2005; Heiser and Groenen, 1997; Zielman and Heiser, 1993).

The purpose of this application is to detect the clustering structure and interpret the asymmetric relation between clusters. In addition, we compare the results for the methods visually.

Table 7.1: Brands and types of Japanese tea bottles (abbreviations shown in parentheses)

Abbreviation	Tea bottle brand
1. IG1	Itoen Green tea1
2. 7&i	7&i Green tea
3. IG2	Itoen Green tea2
4. IG3	Itoen Green tea3
5. IG4	Itoen Green tea4
6. IRG	Itoen Roasted Green tea
7. ABa	Asahi Barley tea
8. SaG	Sangaria Green tea
9. KaG	Kao Green tea
10. SG1	Suntory Green tea1
11. SG2	Suntory Green tea2
12. SRG	Suntory Roasted Green tea
13. SG3	Suntory Green tea3
14. DG	Daido Green tea
15. KG	Kirin Green tea
16. SO1	Suntory Oolong tea1
17. PO	Pokka Oolong tea
18. SO2	Suntory Oolong tea2
19. ABr	Asahi Brended tea
20. CBr1	Coca cola brended tea1
21. CBr2	Coca cola brended tea2
22. SBa	Suntory Barley tea
23. KaBr	Kao Brended tea
24. CBr3	Coca cola Brended tea3
25. KBr	Kirin Brended tea

Table 7.2: switching data for Japanese tea bottles

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.IG1	283	0	0	0	18	0	0	2	2	3	1	2	40
2.7&i	4	56	2	5	0	3	0	0	0	0	0	1	3
3.IG2	5	0	43	0	4	0	0	0	0	1	3	2	4
4.IG3	41	0	3	177	9	9	0	2	0	0	1	4	13
5.IG4	0	0	0	0	18	0	0	0	2	2	3	4	9
6.IRG	10	0	1	0	2	23	0	0	1	0	1	1	3
7.ABa	6	1	4	6	1	2	19	1	1	2	2	0	1
8.SaG	0	0	0	0	0	0	0	24	0	0	0	0	0
9.KaG	0	0	0	0	0	0	0	0	58	0	1	1	3
10.SG1	0	0	0	0	0	0	0	1	0	0	0	0	0
11.SG2	0	0	0	0	0	0	0	1	0	6	14	0	0
12.SRG	0	0	0	0	0	0	0	0	0	4	1	13	0
13.SG3	0	0	0	0	0	0	0	3	0	6	8	5	246
14.DG	0	0	0	0	0	0	0	0	0	0	0	0	0
15.KG	0	0	0	0	0	0	0	0	0	0	0	0	0
16.SO1	0	0	0	0	0	0	0	1	0	0	2	1	0
17.PO	0	0	0	0	0	0	0	0	0	0	0	0	0
18.SO2	0	0	0	0	0	0	0	0	0	0	0	0	0
19.ABr	15	4	10	11	5	2	24	0	0	2	2	1	11
20.CBr1	0	0	0	0	0	0	0	0	0	0	0	0	0
21.CBr2	0	0	0	0	0	0	0	3	0	0	0	0	0
22.SBa	0	0	0	0	0	0	0	0	0	0	2	1	0
23.KaBr	0	0	0	0	0	0	0	0	0	0	0	1	2
24.CBr3	0	0	0	0	0	0	0	5	0	0	0	0	0
25.KBr	0	0	0	0	0	0	0	0	0	0	0	0	0

	14	15	16	17	18	19	20	21	22	23	24	25
1.IG1	2	17	4	0	7	0	0	5	4	0	18	2
2.7&i	0	2	0	1	0	0	0	1	0	0	1	0
3.IG2	0	4	1	3	2	0	0	5	2	2	5	3
4.IG3	2	6	0	0	3	0	0	2	1	0	10	0
5.IG4	0	9	0	0	0	0	1	2	0	0	8	0
6.IRG	1	0	2	0	0	0	1	1	0	1	3	0
7.ABa	0	1	1	0	0	0	0	1	1	2	3	3
8.SaG	2	2	0	2	0	0	0	0	0	0	0	0
9.KaG	0	0	3	1	0	0	0	1	0	6	1	1
10.SG1	1	5	0	1	2	0	0	1	0	0	5	5
11.SG2	0	8	0	0	2	0	0	1	0	0	2	0
12.SRG	0	6	0	0	1	0	1	5	0	0	8	0
13.SG3	3	21	7	4	18	0	2	2	1	0	26	1
14.DG	32	1	0	0	0	0	0	0	0	0	0	0
15.KG	0	192	0	0	12	0	0	0	0	0	0	13
16.SO1	0	3	156	1	1	0	0	1	6	0	7	1
17.PO	0	3	0	53	0	0	0	0	0	0	0	0
18.SO2	0	0	0	0	167	0	0	0	0	0	0	0
19.ABr	0	5	0	0	6	137	2	5	0	2	16	2
20.CBr1	0	3	0	0	2	0	22	0	0	0	0	1
21.CBr2	0	13	0	2	5	0	4	114	0	0	0	7
22.SBa	0	1	0	0	0	0	0	2	115	0	2	0
23.KaBr	0	1	2	1	0	0	0	3	2	28	2	2
24.CBr3	2	23	0	0	15	0	7	22	0	0	201	5
25.SKBr	0	0	0	0	6	0	0	0	0	0	0	5

7.2 Results of the constrained slide-vector model

Before applying the constrained slide-vector model to asymmetric dissimilarity data, the number of clusters has to be determined. To determine the number of clusters, first, we set candidates for the number of clusters such as 2, 3, 4 and 5. Then, the constrained slide vector model with $m = 1$ is applied to the brand switching data for the case when the number of clusters range from 2 to 5. The number of initial values are set as 100 in the application and the results such that the values of the objective function are minimized are adopted for all value of number of clusters. Figure 7.1 shows the values of the objective function for the constrained slide vector model.

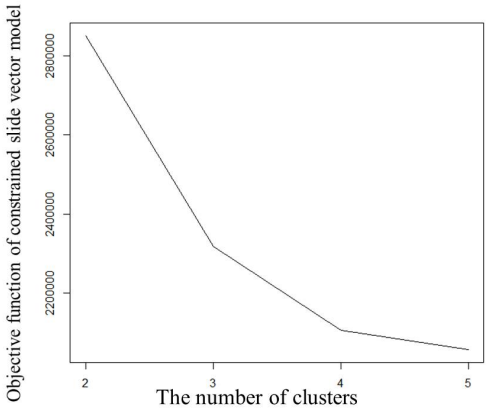

Figure 7.1: Values of the objective functions for the constrained slide-vector model

From Figure 7.1, it can be seen that the number of clusters is set as 4 because the difference between the values for $k = 4$ and $k = 5$ is smaller than that between other values.

The results of the constrained slide-vector model are shown in Figure 7.2. The number of brands belonging to cluster 1 is the largest. From the direction of the estimated slide vector, asymmetries from cluster 3 to the other clusters are smaller than the other relations. In fact, the sum of rows corresponding to the elements of cluster 3 in Table 7.2 tends to be larger than that of the others.

Similarly, the results of slide-vector model and Unfolding are shown in Figure 7.2 and Figure 7.3, respectively. For the constrained slide-vector model, the interpretaion is simple. On the other hand, the results of Unfolding and slide-vector model become difficult to interpret because the number of parameters are very large and these configurations become complex. In the results of the slide-vector model, although the representation of the slide vector is simple, many brands are located around $(0, 0)$ in clumps and the interpretation becomes difficult from Figure 7.4.

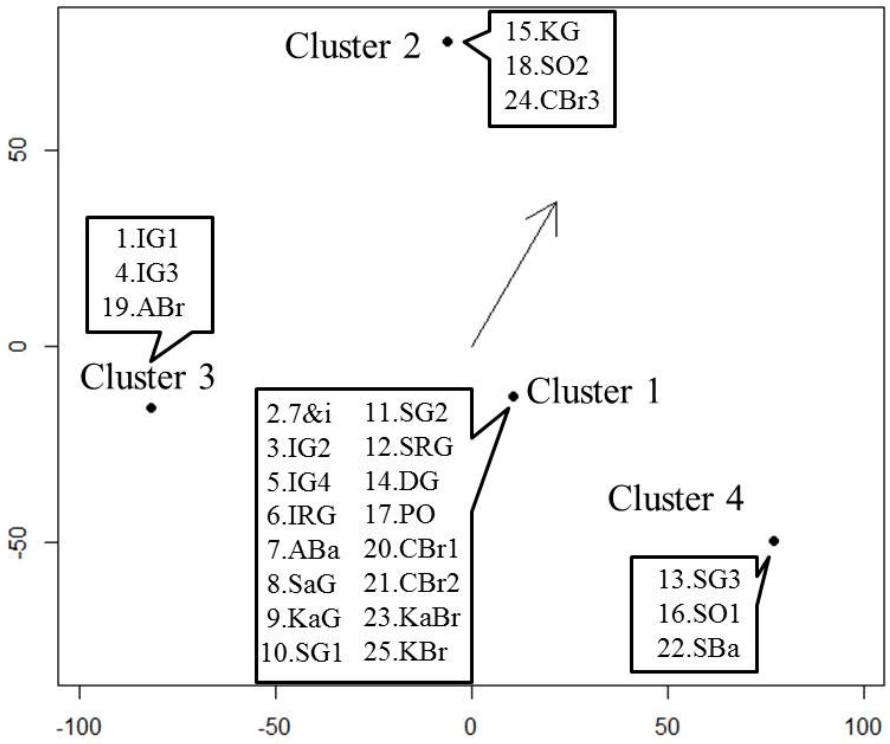

Figure 7.2: Results of the constrained slide-vector model

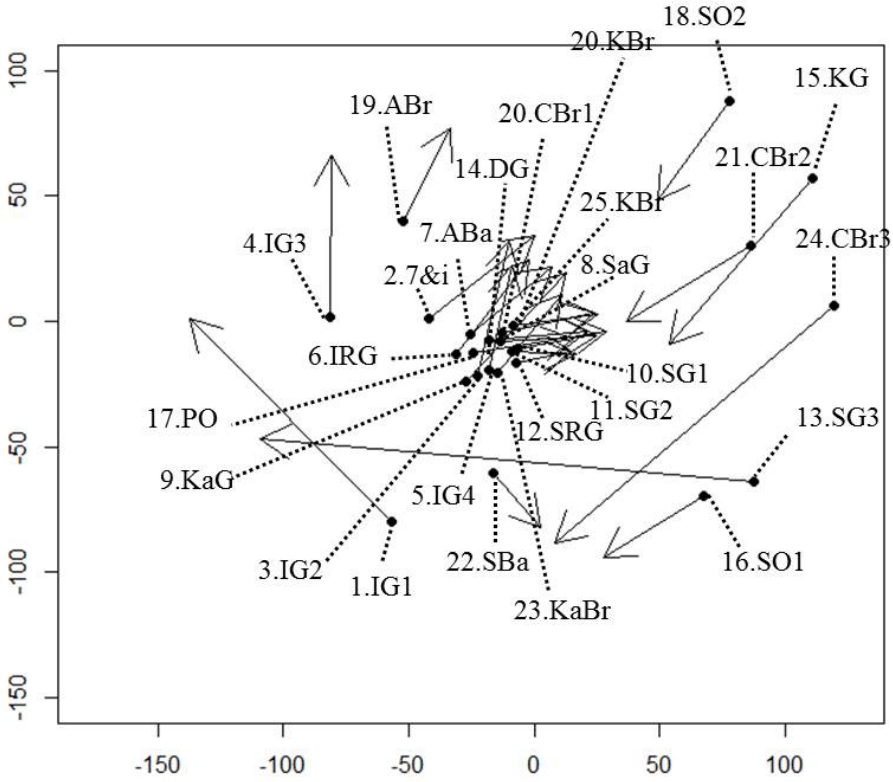

Figure 7.3: Results of the Unfolding model

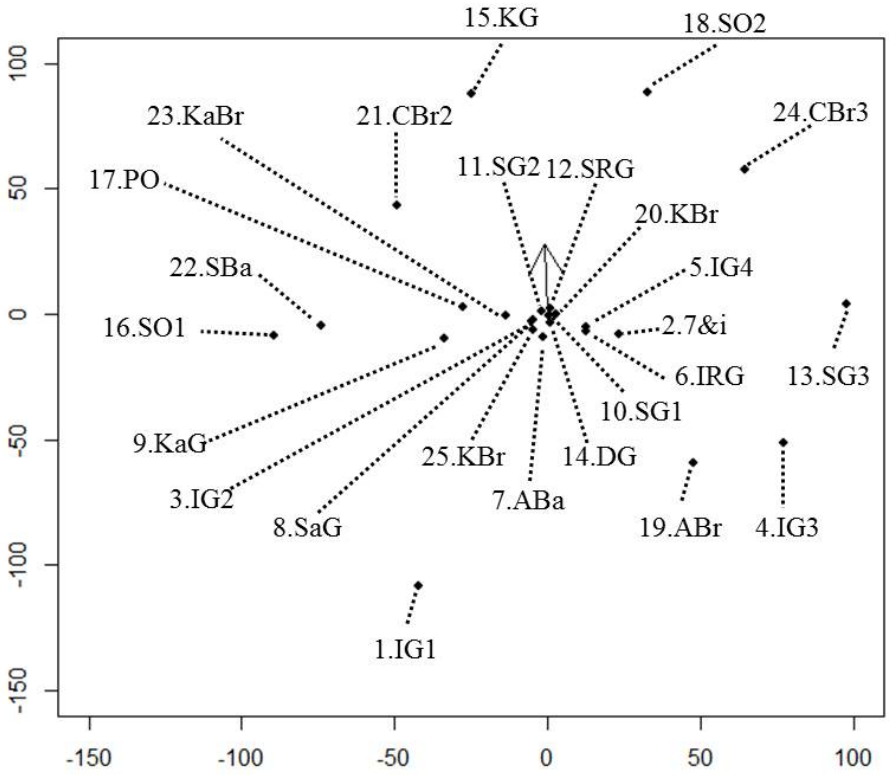

Figure 7.4: Results of the slide-vector model

7.3 Results of the constrained hill-climbing model

In this section, the results of hill-climbing model are shown. First, to determine the number of clusters, the values of the objective function are calculated in a manner similar to that shown in section 7.2 and shown in Figure 7.5. From Figure 7.5, the number of clusters is set as 4 for the same reason described in Section 7.2.

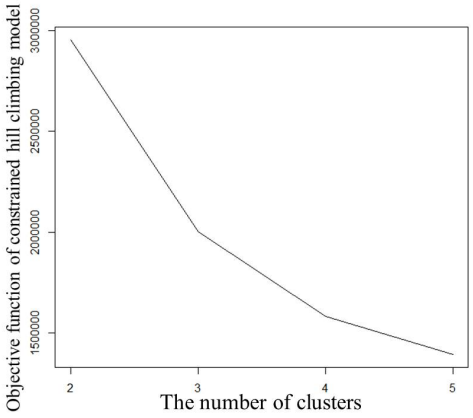

Figure 7.5: Values of the objective functions for the constrained hill-climbing model

Figure 7.6 shows the result of the constrained hill-climbing model. These cluster sizes are almost the same way of the result of the constrained slide-vector model. In terms of the slope vector, the asymmetric relation between cluster 2 and cluster 3 is large because the value of inner product between the difference vector and the slope vector is larger than those of the others. On the other hand, the asymmetric relation between cluster 1 and cluster 4 is small because the value of inner product between the difference vector and the slope vector tends to be smaller than those of the others. The elements of cluster 3, shown in Figure 7.6 are the same as those of cluster 3, shown in Figure 7.2. Therefore, the interpretation of the asymmetries between clusters, shown in Figure 7.6, can be performed as shown in Figure 7.2.

Figure 7.7 shows the results of the hill-climbing model. The interpretation becomes easier than that shown in Figure 7.4 because the coordinates of the brands in Figure 7.7 are dispersed. However, the interpretation is not too easy because the number of brands is large.

7.4 Results of the constrained radius model

In this section, the results of the constrained radius model are shown. The number of clusters will be determined similar to the method described in Section 7.2 and 7.3, Figure 7.8 shows the results of the objective function for the constrained radius model for each value of the number of clusters. Similar to Section 7.2 and 7.3, the number of clusters are set as 4.

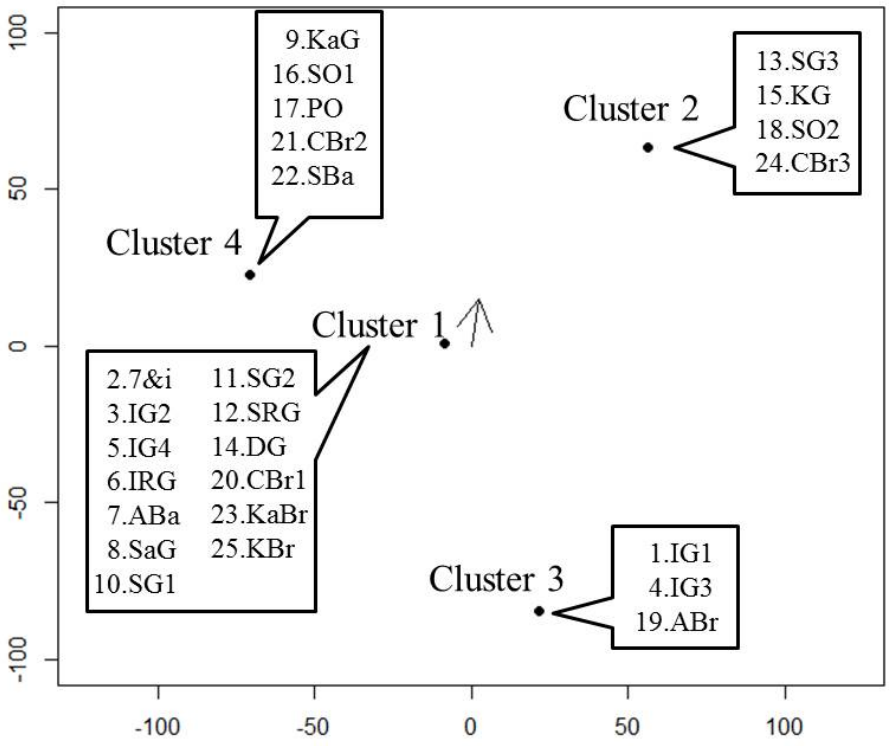

Figure 7.6: Result of the constrained hill-climbing model

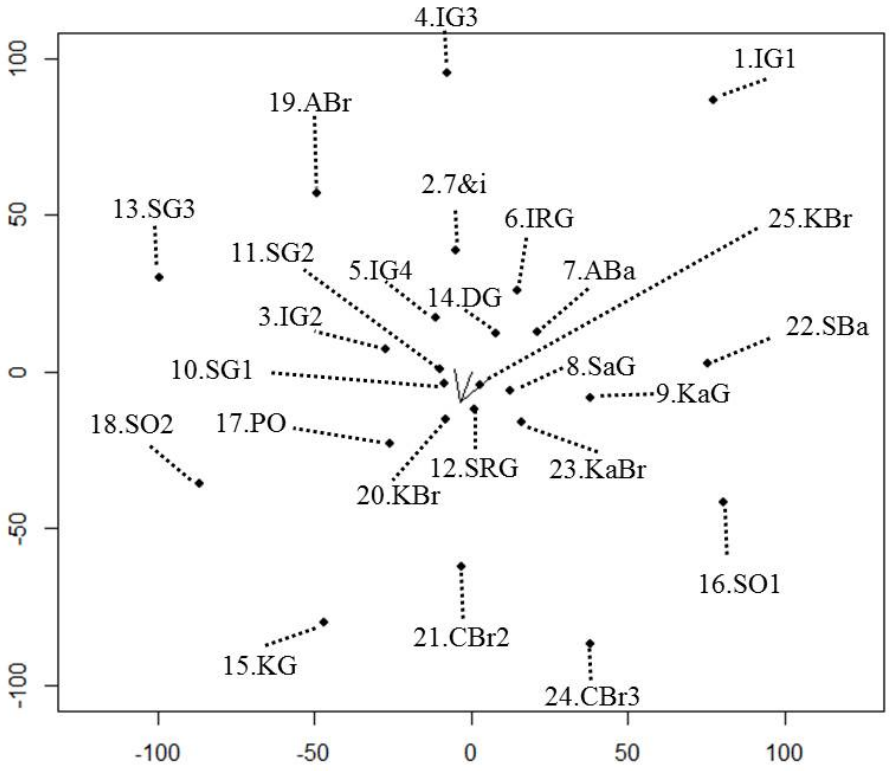

Figure 7.7: Result of the hill-climbing model

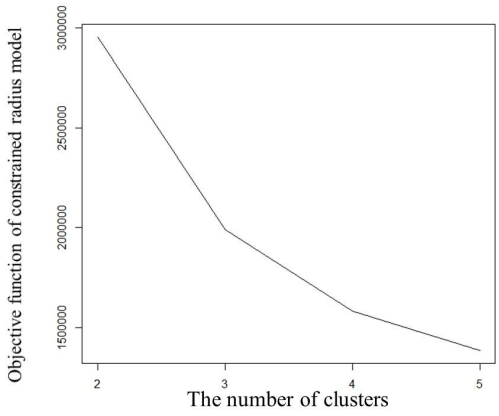

Figure 7.8: Values of the objective function for the constrained radius model

Figure 7.9 shows the results of the constrained radius model. The estimated cluster sizes tend to be almost the same as those of the constrained slide-vector model and the constrained hill-climbing model. In particular, the brands of cluster 1, shown in Figure 7.9 are exactly the same as those of cluster 2, shown in Figure 7.6. In addition, for the interpretation of the asymmetries between clusters from Figure 7.9 becomes change those in both Figure 7.2 and Figure 7.6. In the results of the constrained radius model, the distances from cluster 2, cluster 3, and cluster 4 to cluster 1 tend to be small; whereas the distance from cluster 1 to the other clusters tend to not be small. This is because the number of parameters in the constrained radius model is large and the model can represent the asymmetries between clusters in detail.

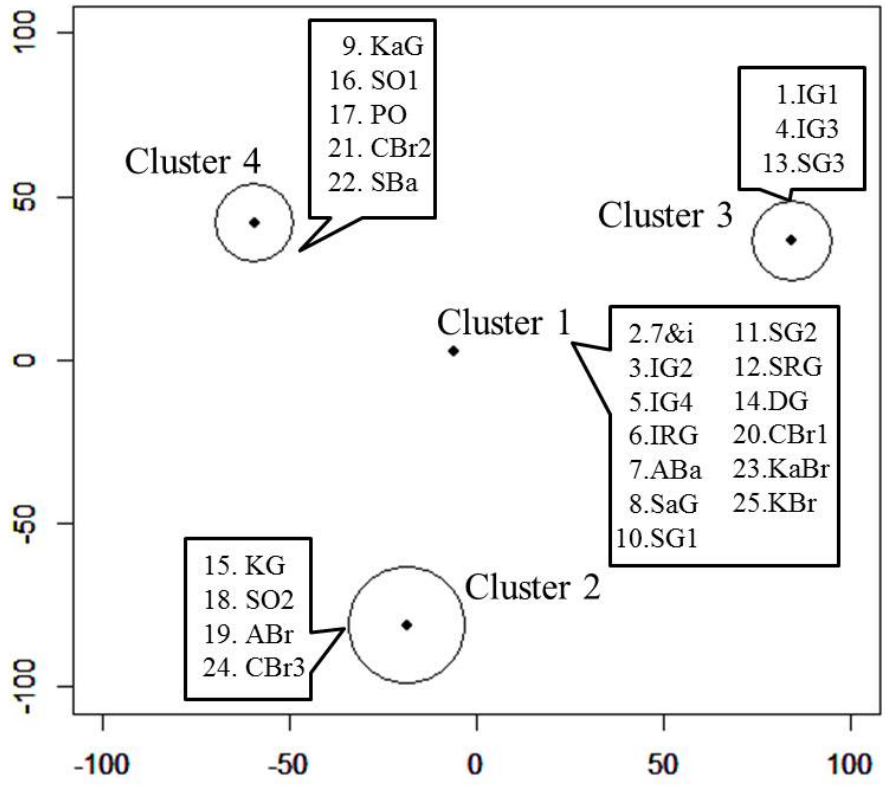

Figure 7.9: Result of the constrained radius model

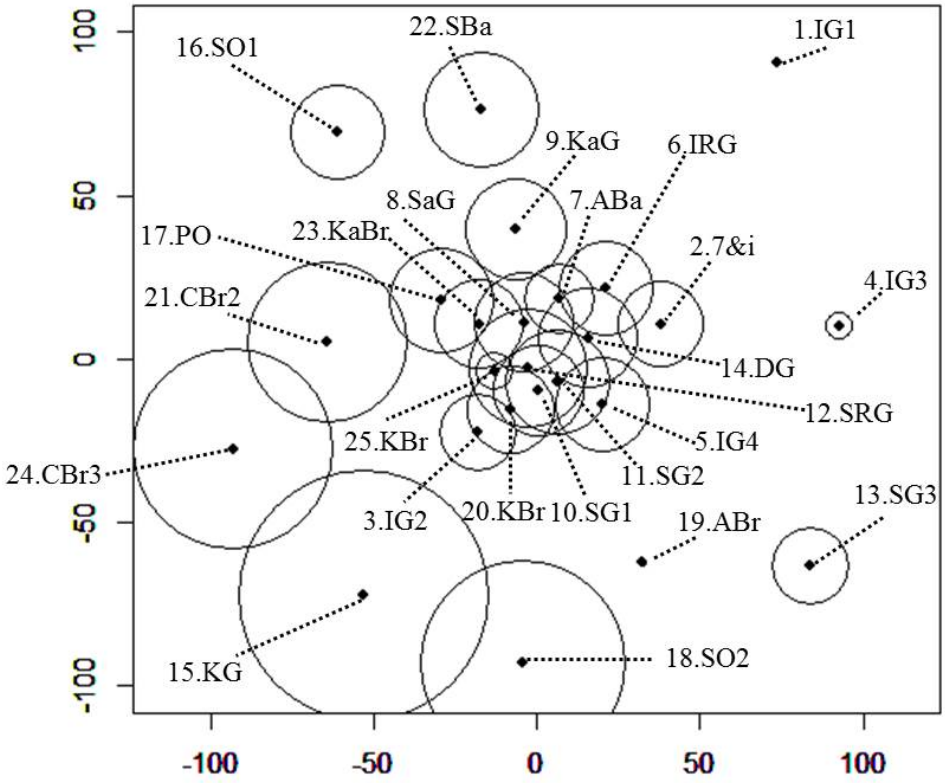

Figure 7.10: Result of the radius model

Chapter 8

Conclusions

In this study, simultaneous methods for AMDS and clustering were proposed from two distinct perspectives, one being AMDS based on Unfolding and the other being AMDS based on the decomposition into symmetric and skew-symmetric parts. From the formulation, the simultaneous methods based on Unfolding can be considered a special case of CDS. Further, these objective functions can be decomposed into two parts, i.e., clustering and MDS terms for these centroids. Then, these methods can be considered a simultaneous method of clustering and a special case of MDS for cluster centroids. Conversely, the simultaneous method based on the decomposition can also be considered a simultaneous method of CDS and the skew-symmetric clustering model. When interpretaion of ordinal AMDS is difficult these simultaneous methods become useful because asymmetries between clusters are represented based on concepts of parsimonious models. In addition, from the simulation results, the effectiveness of these constrained methods is revealed. However, there are several areas of future work to consider.

First, the way by which weights for pairs of objects are identified should be improved for the above methods. The advantage of dimensional reduction clustering (e.g., De Soete and Carroll, 1994; Vichi and Kiers, 2001) is that clustering results are good even if the data include noise. In short, for asymmetric dissimilarity data, if some of the asymmetric dissimilarities do not have a clear clustering structure, clustering results of dimensional reduction clustering should be fine; however, in our proposed methods, the estimation does not work well because the number of parameters is very high. To overcome this problem, the number of parameters should be reduced. For example, weights between clusters are added to these models as the solution.

Second, the number of clusters and the number of low-dimensions should be determined. As a meaning of determining these hyper parameters, cross validation is useful. For example, Yamamoto and Hwang (2014), whose work focused on dimensional reduction clustering, adopted cross validation based on Wang (2010). For determining tuning parameters without considering the number of clusters, cross validation is useful (e.g., Sun et al., 2012). In addition, the cross validation can be

applied to determine the weights for terms of these objective functions such that the clustering result is stable.

Third, although our proposed methods can represent asymmetries between clusters in low-dimensions, it cannot represent asymmetries within clusters because the objective functions of the simultaneous methods consist of terms for clustering based on asymmetric dissimilarity data, and not on low-dimensions, and terms for the MDS of cluster centroids. Therefore, when these coordinates are estimated, the information within clusters is already lost in the type of our proposed objective functions. To overcome this problem, terms for clustering and MDS related to asymmetries within each cluster should be added to these objective functions as penalty terms.

Fourth, given asymmetric dissimilarity data and external information for the same objects as multivariate data, it is difficult to interpret relations between these asymmetries and variables of the multivariate data by using the results of our proposed simultaneous method. When brand switching data is analyzed, related multivariate data is given in many cases, and it is important to detect the relations between the asymmetries and features of brands. Kiers et al. (2005) proposed simultaneous classification and MDS with external information, although the method proposed by Kiers et al. (2005) does not consider the asymmetries. Conversely, Vicari (2014) proposed a clustering model for skew-symmetric data that included external information; however, these methods are not based on the MDS model. Therefore, new simultaneous methods for clustering and AMDS that consider external information must be proposed.

Fifth, various AMDS models have been proposed, and it is difficult to select the appropriate AMDS model among AMDS models even though these AMDS models are characterized based on such common features as Unfolding type and decomposition. Therefore, we need to identify criteria for selecting such AMDS models. However, there is no criterion for selecting AMDS models and the simultaneous method for clustering and AMDS also inherit this problems. A Monte Carlo simulation should also be conducted in various situations to verify clustering results based on the Adjusted Rand Index (Hubert and Arabie, 1985). In short, although it is difficult to select one model from the various AMDS models for the meaning of fitting between models and data, it is possible to select models of these simultaneous methods for clustering and AMDS through the clustering results because these clustering results have the same structure regardless of AMDS models.

Finally, clustering reuslts of the proposed methods should be evaluated through the real data, although we confirmed that the clustering results is superior to the results of tandem clustering via Monte Carlo Simulation. It is difficult to evaluate the performance of the clustering results via real data because it is difficult to determine the clustering structure of real data. Therefore, it needs to compare the clustering results of the proposed methods to that of tandem clustering for various real data to validate the properties of the proposed methods.

Aknowledgement

I am grateful to Prof. Hiroshi Yadohisa for the dedicated comments and suggestions since my undergraduate days. He always support both my research and life. I could not keep research on multivariate data analysis without his support. Valuable comments from Prof. Kohkichi Kawasaki, Prof. Kohei Adachi, Prof. Tamaki Yano and Prof. Mingzhe Jin.

Thanks are due to all members of Yadohisa's laboraotory members who participated discussions and MACROMILL, INC for providing the scan panel data.

Finally I would like to express my appreciation to my family.

Reference

Bell, D. R. and Lattin, J.M. (1998) Shopping behavior and consumer performance for store price format: Why 'large basket' shoppers prefer EDLP, *Marketing Science*, **17**, 66-88.

Borg, I. and Groenen, P.J.F. (2005) *Modern multidimensional scaling: Theory and applications*. 2nd Ed. New York: Springer.

Chino, N. (1997) *Hitaisho Tajigen Syakudo Koseiho* [Asymmetric Multidimensional Scaling]. Kyoto: Gendai Sugaku (in Japanese).

Chino, N. (2012) A brief survey of asymmetric MDS and some open problems, *Behaviormetrika*, **39**, 127-165.

Chino, N., and Okada, A. (1996) Hitaisho tajigen syakudo Koseiho to sono syuhen [Asymmetric multidimensional scaling and related topics], *Kodokeiryogaku*, **23**, 130-152 (in Japanese).

Constantine, A. G., and Gower, J. C. (1978). Graphical representation of asymmetric matrices. *Applied Statistics*, **27**, 297-304.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological review*, **57**, 145-158.

Coombs, C. H. (1964). *A theory of data*. New York: Wily.

De Leeuw, J. (1994). Theory of multidimensional scaling. In P. R. Krishnaiah and L. N. Kanal (Eds.), *Handbook of Statistics* (Volume 2, pp.285-316). Amsterdam, The Netherlands: North-Holland.

De Leeuw, J. and Heiser, W. J. (1982). Block relaxation algorithms in statistics, In H. -H. Bock, W. Lenski and M. M. Richter (Eds.), *Information systems and data analysis* (pp.308-324). Berilin: Springer.

De Soete, G., and Carroll, J. D. (1994). K-means clustering in a low-dimensional Euclidean space. *In New approaches in classification and data analysis* (pp. 212-219). Springer Berlin Heidelberg.

Foa, U. G. (1971). Interpersonal and economic resources, *Science*, **171**, 345-351.

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. *Recent developments in statistics*, (pp.109-123). Amsterdam: North Holland.

Heiser, W. J. (1993). Clustering in low-dimensional space. In O. Optizm B. Lausen, and R.Klar (Eds.), *Information and classification: Concepts, methods, and applications*, (pp. 162-173). Berlin: Springer.

Heiser, W. J. (1995). Convergent computation by iterative majorization:Theory and applications in multidimensional data analysis. In W. J. Krzanowski (Eds.), *Recent advance in descriptive multivariate analysis*, (pp. 157-189). Oxford: Oxford University Press.

Heiser, W. J., and Groenen, P. J. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, **62**, 529-550.

Hubert, L., and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, **2**, 193-218.

Hunter, D. R. and Lange, K. (2004). A autorial on MM algorithms. *The American Statistician*, **39**, 30-37.

Kiers, H. A. L. (2002). Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems. *Computational Statistics & Data Analysis*, **41**, 157-170.

Kiers, H. A. L., Vicari, D., and Vichi, M. (2005). Simultaneous classification and multidimensional scaling with external information. *Psychometrika*, **70**, 433-460.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, **85**, 445-4

Lange, K., Hunter, D. R., and Yang, I. (2000). Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, **9**, 1-20.

Milligan, G.W. (1985). An algorithm for generating artificial test clusters. *Psychometrika*, **50**, 123-127.

Milligan, G.W. and Cooper, M.C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification*, **5**, 181-204.

Nocedal, J. and Wright, S. (1999). *Numerical optimizations*, Springer series in operation research, New York:Springer.

Okada, K. (2011). Beizu suitei ni yoru hitaisyo MDS. [Asymmetric MDS by Bayesian estimation]. *Proceedings of the 39th annual meeting of the Behaviormetric Society of Japan*. Okayama Science University (in print).

Okada, A., and Imaizumi, T. (1984). Geometric models for asymmetric similarity. *Research Reports of School of Social Relations*, Rikkyo (St. Paul's) University.

Okada, A., and Imaizumi,T. (1987). Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika*, **21**, 81-96.

Okada, A., and Imaizumi, T. (1997). Asymmetric multidimensional scaling of two-mode three-way proximities. *Journal of Classification*, **14**, 195-224.

Okada, A., and Tsurumi, H. (2012). Asymmetric multidimensional scaling of brand switching among margarine brands . *Behaviormetrika*, **39**, 111-126.

Saburi, S. and Chino, N. (2008) A maximum likelihood method for an asymmetric MDS model, *Computational Statistics & Data Analysis*, **52**, 4673-4684.

Saito, T. (1991). Analysis of asymmetric proximity matrix by a model of distance and additive terms. *Behaviormetrika*, **29**, 45-60.

Saito, T., and Takeda, S. I. (1990). Multidimensional scaling of asymmetric proximity: model and method. *Behaviormetrika*, **28**, 49-80.

Saito, T. and Yadohisa, H. (2005) *Data analysis of Asymmetric Structures: Advanced Approaches in Computational Statistics*: New York : Marcel Dekker.

Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin*, **38**, 1409-1438.

Steinley, D. (2004). Standardizing variables in k-means clustering. In Banks D., House L, McMorris F.R, Arabie P, Gaul W. (eds) *Classification, clustering, and data mining application.* , (pp53-60). Berlin: Springer.

Sun, W., Wang, J., and Fang, Y. (2012). Regularized k-means clustering of high-dimensional data and its asymptotic consistency. *Electronic Journal of Statistics*, **6**, 148-167.

Tanioka, K. and Yadohisa, H. (2016). Discriminant coordinates for asymmetric dissimilarity data based on radius model . *to appear in Behaviormetrika*.

Tobler, W. (1976-77). Spatial interaction patterns. *Journal of Environmental Systems*, **6**, 271-301.

Tobler, W. and Wineburg (1971). A cappadocian speculation. *Nature*, **231**, 39-41.

Van Mechelen, I., Bock, H.H., and De Boechk, P. (2004). Two-mode clustering methods: A structured overview. *Statistical Methods in Medical Reseach*, **13** (5), 979-981.

Vicari, D. (2015). CLUSKEXT: CLUstering model for SKew-symmetric data including EXTernal information. *Advances in Data Analysis and Classification*, 1-22.

Vera, J. F., Macias, R., and Angulo, J. M. (2008). Non-stationary spatial covariance structure estimation in oversampled domains by cluster differences scaling with spatial constraints. *Stochastic Environmental Research and Risk Assessment*, **22**, 95-106.

Vera, J. F., Macias, R., and Heiser, W. J. (2013). Cluster differences unfolding for two-way two-mode preference rating data. *Journal of Classification*, **30**, 370-396.

Vicari, D. (2014). Classification of asymmetric proximity data. *Journal of Classification*, **31**, 386-420.

Vichi, M., and Kiers, H. A. (2001). Factorial k-means analysis for two-way data. *Computational Statistics & Data Analysis*, **37**, 49-64.

Wang, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika*, **97**, 893-904.

Weeks, D. G., and Bentler, P. M. (1982). Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika*, **47**, 201-208.

Yadohisa, H., and Niki, N. (1999). Vector field representation of asymmetric proximity data. *Communications in Statistics-Theory and Methods*, **28**, 35-48.

Yamamoto, M., and Hwang, H. (2014). A general formulation of cluster analysis with dimension reduction and subspace separation. *Behaviormetrika*, **41**, 115-129.

Young, F. W. (1975). An asymmetric Euclidean model for multi-process asymmetric data. *In US-Japan Seminar on MDS*, San Diego, U.S.A.

Young, F. W., De Leeuw, J. A. N., and Takane, Y. (1980). Quantifying qualitative data, In E.D. Lantermann and H.Feger (Eds), Similarity and choice. Bern: Huber. *Nature*, **231**, 39-41.

Zielman, B., and Heiser, W. J. (1993). Analysis of asymmetry by a slide-vector. *Psychometrika*, **58**, 101-114.

Zielman, B., and Heiser, W. J. (1996). Models for asymmetric proximities. *British Journal of Mathematical and Statistical Psychology*, **49**, 127-146.