2017 Doctoral Thesis

Dimensional reduction method for three-mode three-way data based on canonical covariance analysis

Graduate School of Culture and Information Science,
Doshisha University

48151003

Jun Tsuchida

Superviser Prof. Hiroshi Yadohisa

2017 Doctoral Thesis

Dimensional reduction method for three-mode three-way data based on canonical covariance analysis

Graduate School of Culture and Information Science,
Doshisha University

48151003

Superviser Prof. Hiroshi Yadohisa

Abstract

Three-mode three-way data exist in various research areas, such as psychology and marketing research. Three-mode three-way data are defined as using three finite sets of objects, variables, and conditions. It is not suitable to apply a multivariate data analysis method to three-mode three-way data, because such a method tends to regard the same variables under different conditions as different from each other. Hence, some researchers have extended multivariate data analysis for use with three-mode three-way data. Dimensional reduction methods for three-mode three-way data have been developed by several researchers within multivariate data analysis, because the numbers of objects and conditions of three-mode three-way data tend to be large. Well-known dimensional reduction methods for three-mode three-way data include Tucker 3 and parallel factor analysis (PARAFAC). Given two data sets, canonical correlation analysis is a famous dimensional reduction method within multivariate data analysis. Like multivariate data analysis, canonical correlation analysis is extended for use with three-mode three-way data. However, canonical correlation analysis seeks only common factors that maximize the correlation between factors, making us unable to interpret the factor separately from each data by these methods. To do so, canonical covariance analysis has been proposed. However, this method has not been extended for use with three-mode three-way data. Therefore, we propose new dimensional reduction methods for three-mode three-way data based on canonical covariance analysis in this study.

We describe a simple extension method for three-mode three-way data, which is called the basic method. The basic method is defined as maximizing covariance between common factors. However, the basic method does not distinguish which factor is a common factor. Moreover, in the basic method, it is assumed that the number of factors, including the common factor and those for each data, is the same between each data. We address this problem by introducing a connector matrix. We characterize this method by imposing some constraints on the connector matrix. A K-means type and spherical K-means type constrained connector matrix restrict elements of the connector matrix to 0 or 1. From this, it is easy to distinguish which factor is a common factor. The different points between the K-means type and spherical K-means type constrained connector matrix leads to the concept of the update rule for the connector matrix, which corresponds to K-means and spherical K-means, respectively. Regression analysis for three-mode three-way data is included in our proposed method. We propose a tandem analysis for three-mode three-way regression. Furthermore, we propose canonical covariance analysis for three-mode three-way data with a quantitative method based on non-metric principal component analysis.

This method overcomes the problem in previous works that it is not suitable for use with three-mode three-way data that has qualitative value. The parameter estimation method of our proposed method is the least squares method because the objective function of canonical covariance analysis for multivariate data corresponds to the least squares method. For parameter estimation, we provide objective functions, update formulas, partial derivative functions of objective functions, and algorithms. In addition, we include a simulation study and apply the proposed method to real data.

Contents

Chapter 1

Introduction

With sophisticated observation techniques, it is easy to obtain large-scale and complex data. Three-mode three-way data are one type of large-scale and complex data. Three-mode three-way data are defined by using three finite sets of objects, variables, and conditions. Thus, three-mode three-way data are often represented as cubic height, length, and depth corresponding to the number of objects, variables, and conditions, respectively. Figure 1.1 shows the image of three-mode three-way data. Three finite sets of objects, variables, and conditions are called also mode 1, 2, and 3, respectively. Each mode show the each finite set. The term "three-mode" means the data are described as a function whose domain is written by all three finite sets. The term "way" shows the number of domain dimensions of a function, which shows how data are collected.

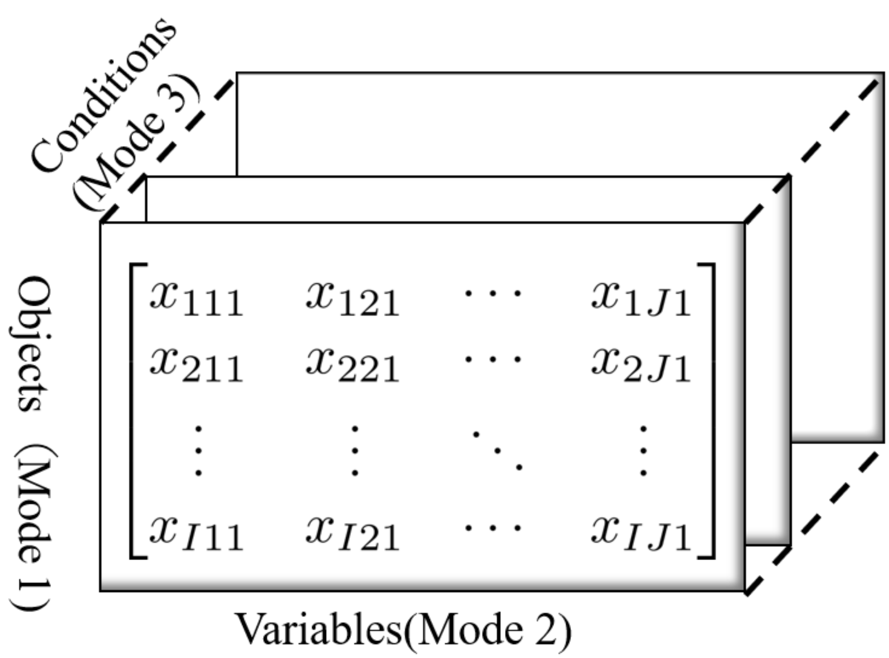


Figure 1.1: The image of three-mode three-way data

When we are interested in the effect of variables and of conditions, as well as the interaction between variables and conditions, we analyze three-mode three-way data. Why is it unsuitable to apply multivariate data analysis methods, which are proposed for two-mode two-way data, to three-mode three-way data? One of the reasons is that multivariate data analysis methods tend to regard the same variables under different conditions as different to each other. Therefore, it is difficult to interpret the main effect of variables, because we obtain several results for the same variables. The same can be said for conditions. Thus, when we are interested in the effect of variables and conditions, we need to use the information of the three finite sets. Moreover, although we can apply multivariate data analysis methods to each condition data decided from three-mode three-way data, it is difficult to interpret the overall result and to compare each result, because multivariate data analysis methods seek the best result for each data, not for whole. Therefore, when we are interested in the relationship between whole variables and conditions, we should use the three-mode three-way structure.

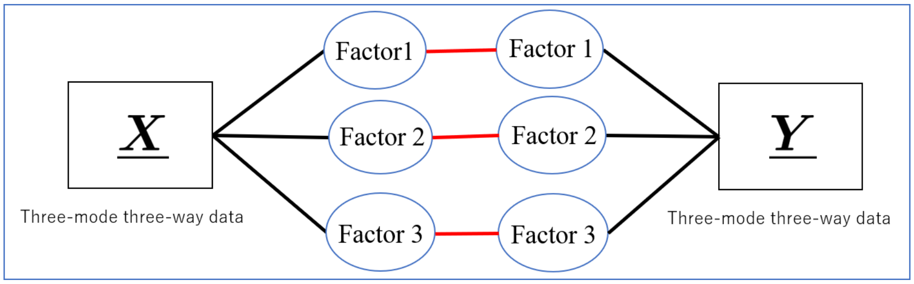
Three-mode three-way data exist in various research areas. One example of three-mode three-way data is semantic differential data, which are typical data in the field of psychology. Another type of three-mode three-way data is multivariate longitudinal data, because they are described by object $\times$ variable $\times$ time. Multivariate longitudinal data exist in many research areas, such as medicine and economics. Point-of-sale data, which are typical data for marketing research, is another type of three-mode three-way data. Point-of-sale data are described by item $\times$ variable $\times$ store. These data are analyzed by using several three-mode three-way methods. For example, Lundy et al. (1989) apply three-way factor analysis to the rating data of TV shows. Kroonenberg (1983) shows some examples of three-way principal component analysis (PCA) for semantic differential data.

Dimensional reduction methods of the multivariate analysis method for three-mode three-way data have been developed by several researchers, because the numbers of objects and conditions of three-mode three-way data tend to be large. Well-known dimensional reduction methods for three-mode three-way data include Tucker 3 (Tucker, 1966; Kroonenberg, 1983), and PARAFAC (Harshman, 1970). These methods are extended to high-order cases (De Lathauwer et al., 2000; Kolda & Bader, 2009). Given only one data set, these methods have been proven effective. Given two data sets, we can apply these methods to each data set individually, but we cannot interpret the relationship between these data sets.

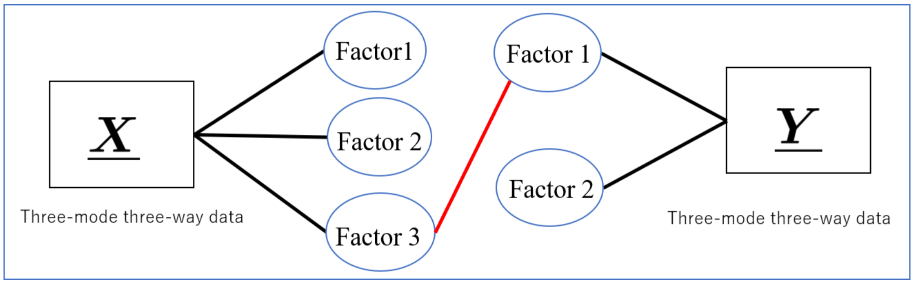
On the other hand, canonical covariance analysis proposes investigating the relationships between two variable sets (e.g., Tenenhaus & Tenenhaus (2011)). Canonical covariance analysis, which is one such dimensional reduction method for two-mode two-way data, attempts to find a subspace such that the covariance between variables is maximized. When data sets are converted via Mahalanobis transformation, canonical covariance analysis is equivalent to canonical correlation analysis (Hotelling, 1936). In the multivariate data case, Carroll (1968) and Kettenring (1971) extend canonical correlation analysis to a multi data set. Tenenhaus & Tenenhaus (2011) extend canonical covariance analysis to a multi data set. However, these extensions assume that extension methods are applied to multivariate data. In the three-mode three-way case, there are many types of extension of canonical correlation analysis. Luo et al. (2015) proposes a tensor canonical correlation analysis method based on the PARAFAC model. On the other hand, Zhao et al. (2013) proposes a tensor partial least squares (PLS) method based on the Tucker model. The PLS method is very similar to the canonical correlation analysis method. When some assumptions are restricted in the PLS method, it is equivalent to canonical correlation analysis (Sun et al., 2009). One purpose of these methods is to seek common factors between data sets. A common factor is represented as a linear combination of variables and conditions. The concept of this method is the same as that of canonical correlation analysis. Thus, these methods do not consider the variances of each data set; that is, these methods do not allow us to interpret the factor separately from each data.

In this study, we propose new dimensional reduction methods for three-mode three-way data based on canonical covariance analysis. Canonical covariance analysis is the simultaneous analysis of PCA and canonical correlation analysis because the objective function of canonical covariance analysis is described by the sum of the objective function of PCA for each data and of the canonical correlation analysis. Our proposed methods allow us to interpret common factors and factors for each data simultaneously. First, we describe a simple extension method for three-mode three-way data, which is called the basic method. This method is defined as maximizing covariance between common factors. However, the basic method does not distinguish which factor is a common factor. Moreover, in the basic method, it is assumed that the number of factors, including the common factor and those for each data, is the same between each data. We address this problem by introducing a connector matrix. The connector matrix indicates how to represent the common factor by using each factor. Figure 1.2 shows an image of the connector matrix. The red lines between factors indicate the connections. Figure 1.2(a) shows the case of a one-to-one relationship between factors. In this case, the connector matrix indicates all factor are common factors. Moreover, a one-to-one relationship between factors corresponds to the canonical covariance analysis for the three-mode three-way data when the value of the connector matrix is 1. Figure 1.2(b) shows that the connector matrix indicates which factor is a common factor. In this case, the connector matrix indicates that factor 3 of data $\underline{\boldsymbol{X}}$ and factor 1 of $\underline{\boldsymbol{Y}}$ are common factors. Other factors are factors for each data, that is, factors maximizing variance.

We impose some constraints on the connector matrix in order to interpret it. One of the constraints is a K-means (MacQueen, 1967) type. A K-means type constrained method restricts elements of the connector matrix to 0 or 1. By imposing this constraint, it is easy to distinguish which factor is common. Another constraint is a spherical K-means (Dhillo & Modha, 2001) type. Elements of the connector matrix are restricted to 0 or 1, like for the K-means type. The different points between spherical K-means and K-means represent the objective function. In the objective function of the K-means type constrained method, it is assumed that the variances of factors have similar values. On the other hand, the objective function of the spherical K-means constrained method evaluates the differences in the angle between factors. Therefore, when there are large differences between the variance of each factor, it is suitable to apply the spherical K-means method



(a) 1 to 1 case



(b) Selecting dimension case

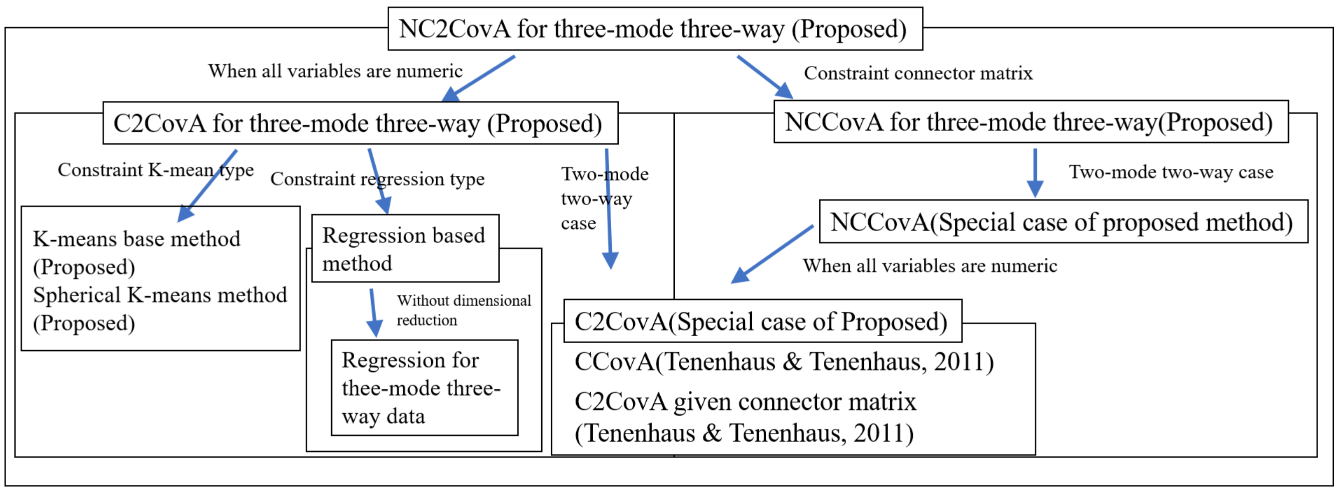
Figure 1.2: Image of connector matrix

to three-mode three-way data. The other constraint is regression type. When we set an appropriate number of dimensions of parameters and constrained parameters, then canonical covariance analysis for three-mode three-way data is equivalent to regression for three-mode three-way data. In this case, the connector matrix corresponds to the coefficient matrix. For easy interpretation, we describe dimensional reduction for regression analysis. Regression analysis with dimensional reduction has been proposed by many researchers. Zhao et al. (2013) proposes a simultaneous method of dimensional reduction for response and exploration. Rabusseau & Kadri (2016) proposes a method of dimensional reduction for exploration. These methods are simultaneous methods between regression and dimensional reduction. Differentiating our method from those of previous works is tandem analysis. Tandem analysis is two-step analysis. The first step is a dimensional reduction step. In the first step, we apply a dimensional reduction method to data sets. In the second step, we deliberately apply the method to data sets after dimensional reduction. In this case, the second step is the application of the regression method to data sets. We propose tandem analysis using the concept of PLS for the dimensional reduction for the regression based on Tsuchida & Yadohisa (2017).

In our proposed methods and those of previous works, it is assumed that all values are quantitative. However, there is a case in which qualitative values exist in three-mode three-way data. For example, longitudinal survey data sometimes have qualitative values, such as Japan's national census. In addition, item-rating data and semantic differential data are types of qualitative data. When the values do not satisfy the interval scale, the assumption of our proposed methods and previous works is not satisfied. In other words, it is not suitable to apply this method to data with qualitative values. In the multivariate data case, many researchers address this problem by using a quantitative method. Multiple corresponding analysis (Greenacre & Blasius, 2006; Gifi, 1990) is one of the famous dimensional reduction methods for categorical multivariate data. Non-metric principal component analysis (NPCA) (Young et al., 1978) is one of the dimensional reduction methods for categorical data. NPCA is one of the special cases of multiple correspondence analysis (Gifi, 1990). On the other hand, NPCA is simply an extension of PCA for categorical data. Therefore, we could simply extend the three-mode three-way canonical covariance analysis for categorical data by using the concept of NPCA. Extension using the concept of NPCA is based on Tsuchida & Yadohisa (2016b).

We use least squares as the estimation method for all of our proposed methods, because the objective function of canonical covariance analysis for multivariate data corresponds to the least squares method. However, explicit update formulas for the parameters of canonical covariance analysis for three-mode three-way data are not obtained simultaneously. Hence, we adopt an alternative least squares algorithm for the parameter estimation. In addition, some methods need a partial derivative function of the objective function for the gradient method proposed by Jennrich (2001). Therefore, we provide the derivative function of some objective function.

The reminder of this thesis is organized as follows. In Chapter 2, some notations used in this thesis are defined. In addition, some properties of three-mode three-way data are described. In Chapter 3, dimensional reduction methods based on the canonical covariance method are described. First, we define the model and objective function of the basic and connector matrix methods. These methods are the core of this study. Then, the algorithms for these methods are introduced. Next, we define the model and objective functions of the constrained connector matrix and we introduce algorithms for these methods. Finally, we describe the canonical covariance analysis for three-mode three-way data and provide a quantification method. Figure 1.3 shows the relationship between previous methods in the literature and this study's proposed methods. In Chapter 4, we provide estimation and prediction accuracy through numerical examples. In Chapter 5, we provide an example analysis using country investigation data and questionnaire survey data. In Chapter 6, we provide concluding remarks.



CCovA: Canonical Covariance Analysis (corresponding to Subsection 3.1.1)

C2CovA: Connector Canonical covariance analysis (corresponding to Subsection 3.1.2)

NCCovA: Nonmetric Canonical covariance analysis (corresponding to Subsection 3.3.1)

NC2CovA: Non-metric Connector Canonical covariance analysis (corresponding to to Subsection 3.3.1)

Figure 1.3: Relationship between proposed methods and previous work

Chapter 2

Notation

In this chapter, we introduce some notations and properties in preparation for our analysis. First, we define n-mode m-way data. Then, the three-mode three-way norm is introduced.

Definition 2.1. **n-mode m-way data**
$\mathcal{U}$ is a set of permutations, whose length is m with repetition of the elements in n finite index sets, that is, $\mathcal{U} = \prod_{i=1}^{m} N_{j_i}$. N_j $(j = 1,\ 2, \cdots,\ n;\ n \leq m)$ are finite index sets, and are called mode. j_i is the ith element of the set defined as the m combination with repetition from $\{1, 2, \cdots n\}$. Data $\boldsymbol{S}$ is represented as a function of $\mathcal{U}$ for a real number. In other words, data $\boldsymbol{S}$ are described as follows:

$$\boldsymbol{S} : \mathcal{U} \longrightarrow \mathbb{R}.$$

In this case, $\boldsymbol{S} = \boldsymbol{S}(\mathcal{U})$ is the n-mode m-way data and is described as $\boldsymbol{S} = (s_u) = (s_{(i_1,\, i_2, \cdots,\, i_m)}) \in \mathbb{R}^{\prod_{i=1}^{m} |N_{j_i}|}$, where u is an element of $\mathcal{U}$. $|A|$ refers to the cardinality of A and $i_1,\ i_2, \cdots,\ i_m$ are elements corresponding to the finite index set u.

Multivariate data are included in n-mode m-way data.

Example 2.1. ***Two-mode two-way data***
We set $N_1 = \{1,\ 2, \cdots, I\}$, $N_2 = \{1,\ 2, \cdots, J\}$ and $\mathcal{U} = N_1 \times N_2$. Then, two-mode two-way data $\boldsymbol{S}$ are described as

$$\boldsymbol{S} = (s_{(i,\, j)}) \in \mathbb{R}^{I \times J},$$

where i are elements of N_1, and j is an element of N_2. When N_1 and N_2 are a set of objects and a set of variables respectively, the two-mode two-way data are called multivariate data.

Using the concept of mode and way, we summarize the typical data in Table 2.1. We note that the number of modes and ways shape the data; however, the meaning of a value depends on the mode and the way. For example, two-mode two-way data include not only multivariate data but also time-series data. Multivariate data have a set of variables and a set of objects as the mode. On the other hand, time-series data have a set of time and a set of objects as the mode. The order of time plays an important role in analyzing time-series

Table 2.1: Relationship between typical data and the combination of mode and way

	One-way	Two-way	Three-way	$\cdots$	m-way
One-mode	summarized data	(dis)similarity data	(dis)similarity between three objects' data	$\vdots$	(dis)similarities in m objects' data
Two-mode	undefined	multivariate data	individual difference (dis)similarity data	$\vdots$	(dis)similarity between m objects of two data sets
Three-mode	undefined	undefined	time-series multivariate data	$\vdots$	(dis)similarity between m objects of three data sets
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n-mode	undefined	undefined	undefined	$\cdots$	sequential data under many conditions

data. The property of mode is one of the greatest differences between multivariate data and time-series data.

In this study, we focus on three-mode three-way data whose mode has no order. We introduce the notation of three-mode three-way data from Kiers (2000).

Example 2.2. ***Three-mode three-way data***
When $n = m = 3$, n-mode m-way data are called three-mode three-way data. three-mode three-way data are described in bold underlined capitals, such as $\underline{\boldsymbol{X}}$. When the cardinalities of modes of $\underline{\boldsymbol{X}}$ are I, J, and K, then we denote $\underline{\boldsymbol{X}} = (x_{ijk}) \in \mathbb{R}^{I\times J\times K}(i = 1,\ 2, \cdots,\ I; j = 1,\ 2, \cdots,\ J; k = 1,\ 2, \cdots,\ K)$, respectively.

The shape of three-mode three-way data is cubic; that is, one value of three-mode three-way data is defined by a three-dimension vector whose elements are i, j, and k. We define the sum and scalar multiple operator for three-mode three-way data, and the norm of three-mode three-way data.

Definition 2.2. ***Summation, scalar multiple of three-mode three-way data***
Given two three-mode three-way data $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ ($\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}} \in \mathbb{R}^{I\times J\times K}$), and constant value $a \in \mathbb{R}$, the sum of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ is defined as follows:

$$\underline{\boldsymbol{X}} + \underline{\boldsymbol{Y}} = (x_{ijk} + y_{ijk}).$$

Scalar multiple by a is defined as follows:

$$a\underline{\boldsymbol{X}} = (ax_{ijk}).$$

Definition 2.3. ***Norm of three-mode three-way data***
The norm of three-mode three-way data $\|\underline{\boldsymbol{X}}\|$ is defined as follows:

$$\|\underline{\boldsymbol{X}}\| = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}^2}$$

Definition 2.4. ***Inner product of three-mode three-way data***
The inner product of three-mode three-way data $\underline{\boldsymbol{X}}$ *and* $\underline{\boldsymbol{Y}}$ $(\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}} \in \mathbb{R}^{I\times J\times K})$ *is defined as follows:*

$$< \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}} >= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}y_{ijk}$$

For the definition of multiplying the three-mode three-way data with the matrix, we define the transformation operator of the three-mode three-way data to the matrix.

Definition 2.5. ***Unfolding mode***
Given three-mode three-way data $\underline{\boldsymbol{X}}$*, we define the unfolding mode 1* $\boldsymbol{X}_1$ *as follows:*

$$\boldsymbol{X}_1 = (x_{iI_{\boldsymbol{X}}(j,k)}) \in \mathbb{R}^{I\times JK},$$

where $x_{iI_{\boldsymbol{X}}(j,k)} = x_{ijk}$ $(i = 1,\ 2,\cdots,\ I;\ j = 1,\ 2,\ \cdots,\ J;\ k = 1,\ 2,\cdots,\ K)$*, and* $I_{\boldsymbol{X}}(j,k) = j + (k-1)J$*. In addition, we define the unfolding modes 2* $\boldsymbol{X}_2$ *and 3* $\boldsymbol{X}_3$ *as follows:*

$$\boldsymbol{X}_2 = (x_{jJ_{\boldsymbol{X}}(i,\ k)}) \in \mathbb{R}^{J\times IK},\ \ \boldsymbol{X}_3 = (x_{kK_{\boldsymbol{X}}(i,\ j)}) \in \mathbb{R}^{K\times IJ},$$

where $x_{jJ_{\boldsymbol{X}}(i,\ k)} = x_{ijk}$ $(i = 1,\ 2,\cdots,\ I;\ j = 1,\ 2,\ \cdots,\ J;\ k = 1,\ 2,\cdots,\ K)$, $x_{kK_{\boldsymbol{X}}(i,\ j)} = x_{ijk}$ $(i = 1,\ 2,\cdots,\ I;\ j = 1,\ 2,\ \cdots,\ J;\ k = 1,\ 2,\cdots,\ K)$, $J_{\boldsymbol{X}}(i,k) = i + (k-1)I$*, and* $K_{\boldsymbol{X}}(i,j) = i + (j-1)I$*.*

Figure 2.1 depicts the unfolding mode 1, 2 and 3. The most different point in the matrixes made by unfolding each mode is the arrangement of the row. The row vector of matrixes made by unfolding each mode corresponds to each mode. The number of dimensions of the row vector whose matrixes are made by unfolding modes 1, 2, and 3 are JK, IK, and IJ, respectively.

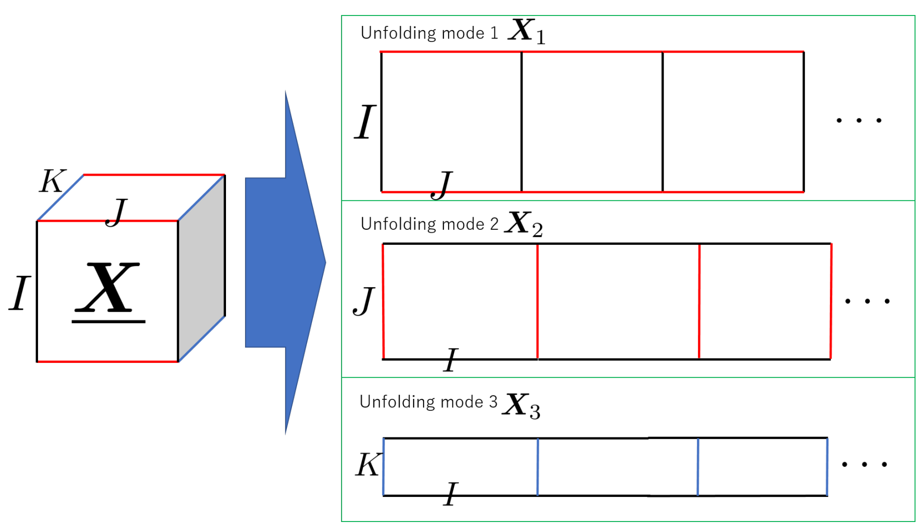


Figure 2.1: Representation of unfolding mode

Example 2.3. ***Example of unfolding mode***

$\boldsymbol{X}_1$, $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$ are obtained by unfolding the $\underline{\boldsymbol{X}} \in \mathbb{R}^{4\times 2\times 3}$. These matrixes are shown as follows:

$$
\boldsymbol{X}_1 = \begin{pmatrix}
x_{1I_{\boldsymbol{X}}(1,1)} & x_{1I_{\boldsymbol{X}}(2,1)} & x_{1I_{\boldsymbol{X}}(1,2)} & x_{1I_{\boldsymbol{X}}(2,2)} & x_{1I_{\boldsymbol{X}}(1,3)} & x_{1I_{\boldsymbol{X}}(2,3)} \\
x_{2I_{\boldsymbol{X}}(1,1)} & x_{2I_{\boldsymbol{X}}(2,1)} & x_{2I_{\boldsymbol{X}}(1,2)} & x_{2I_{\boldsymbol{X}}(2,2)} & x_{2I_{\boldsymbol{X}}(2,3)} & x_{2I_{\boldsymbol{X}}(2,3)} \\
x_{3I_{\boldsymbol{X}}(1,1)} & x_{3I_{\boldsymbol{X}}(2,1)} & x_{3I_{\boldsymbol{X}}(1,2)} & x_{3I_{\boldsymbol{X}}(2,2)} & x_{3I_{\boldsymbol{X}}(3,3)} & x_{3I_{\boldsymbol{X}}(2,3)} \\
x_{4I_{\boldsymbol{X}}(1,1)} & x_{4I_{\boldsymbol{X}}(2,1)} & x_{4I_{\boldsymbol{X}}(1,2)} & x_{4I_{\boldsymbol{X}}(2,2)} & x_{4I_{\boldsymbol{X}}(4,3)} & x_{4I_{\boldsymbol{X}}(2,3)}
\end{pmatrix}
= \begin{pmatrix}
x_{111} & x_{121} & x_{112} & x_{122} & x_{113} & x_{123} \\
x_{211} & x_{221} & x_{212} & x_{222} & x_{213} & x_{223} \\
x_{311} & x_{321} & x_{312} & x_{322} & x_{313} & x_{323} \\
x_{411} & x_{421} & x_{412} & x_{422} & x_{413} & x_{423}
\end{pmatrix},
$$

$$
\boldsymbol{X}_2' = \begin{pmatrix}
x_{1J_{\boldsymbol{X}}(1,1)} & x_{2J_{\boldsymbol{X}}(1,1)} \\
x_{1J_{\boldsymbol{X}}(2,1)} & x_{2J_{\boldsymbol{X}}(2,1)} \\
x_{1J_{\boldsymbol{X}}(3,1)} & x_{2J_{\boldsymbol{X}}(3,1)} \\
x_{1J_{\boldsymbol{X}}(4,1)} & x_{2J_{\boldsymbol{X}}(4,1)} \\
x_{1J_{\boldsymbol{X}}(1,2)} & x_{2J_{\boldsymbol{X}}(1,2)} \\
x_{1J_{\boldsymbol{X}}(2,2)} & x_{2J_{\boldsymbol{X}}(2,2)} \\
x_{1J_{\boldsymbol{X}}(3,2)} & x_{2J_{\boldsymbol{X}}(3,2)} \\
x_{1J_{\boldsymbol{X}}(4,2)} & x_{2J_{\boldsymbol{X}}(4,2)} \\
x_{1J_{\boldsymbol{X}}(1,3)} & x_{2J_{\boldsymbol{X}}(1,3)} \\
x_{1J_{\boldsymbol{X}}(2,3)} & x_{2J_{\boldsymbol{X}}(2,3)} \\
x_{1J_{\boldsymbol{X}}(3,3)} & x_{2J_{\boldsymbol{X}}(3,3)} \\
x_{1J_{\boldsymbol{X}}(4,3)} & x_{2J_{\boldsymbol{X}}(4,3)}
\end{pmatrix}
= \begin{pmatrix}
x_{111} & x_{121} \\
x_{211} & x_{221} \\
x_{311} & x_{321} \\
x_{411} & x_{421} \\
x_{112} & x_{122} \\
x_{212} & x_{222} \\
x_{312} & x_{322} \\
x_{412} & x_{422} \\
x_{113} & x_{422} \\
x_{113} & x_{123} \\
x_{213} & x_{223} \\
x_{313} & x_{323} \\
x_{413} & x_{423}
\end{pmatrix},
$$

$$
\boldsymbol{X}_3' = \begin{pmatrix}
x_{1K_{\boldsymbol{X}}(1,\,1)} & x_{2K_{\boldsymbol{X}}(1,\,1)} & x_{3K_{\boldsymbol{X}}(1,\,1)} \\
x_{1K_{\boldsymbol{X}}(2,\,1)} & x_{2K_{\boldsymbol{X}}(2,\,1)} & x_{3K_{\boldsymbol{X}}(2,\,1)} \\
x_{1K_{\boldsymbol{X}}(3,\,1)} & x_{2K_{\boldsymbol{X}}(3,\,1)} & x_{3K_{\boldsymbol{X}}(3,\,1)} \\
x_{1K_{\boldsymbol{X}}(4,\,1)} & x_{2K_{\boldsymbol{X}}(4,\,1)} & x_{3K_{\boldsymbol{X}}(4,\,1)} \\
x_{1K_{\boldsymbol{X}}(1,\,2)} & x_{2K_{\boldsymbol{X}}(1,\,2)} & x_{3K_{\boldsymbol{X}}(1,\,2)} \\
x_{1K_{\boldsymbol{X}}(2,\,2)} & x_{2K_{\boldsymbol{X}}(2,\,2)} & x_{3K_{\boldsymbol{X}}(2,\,2)} \\
x_{1K_{\boldsymbol{X}}(3,\,2)} & x_{2K_{\boldsymbol{X}}(3,\,2)} & x_{3K_{\boldsymbol{X}}(3,\,2)} \\
x_{1K_{\boldsymbol{X}}(4,\,2)} & x_{2K_{\boldsymbol{X}}(4,\,2)} & x_{3K_{\boldsymbol{X}}(4,\,2)}
\end{pmatrix}
= \begin{pmatrix}
x_{111} & x_{112} & x_{113} \\
x_{211} & x_{212} & x_{213} \\
x_{311} & x_{312} & x_{313} \\
x_{411} & x_{412} & x_{413} \\
x_{121} & x_{122} & x_{123} \\
x_{221} & x_{222} & x_{223} \\
x_{321} & x_{322} & x_{323} \\
x_{421} & x_{422} & x_{423}
\end{pmatrix}.
$$

Using the unfolding mode, we obtain the relationship between the norm of three-mode three-way data and the three-mode three-way data unfolding mode.

Proposition 2.1. *Given three-mode three-way data $\underline{\boldsymbol{X}} \in \mathbb{R}^{I\times J\times K}$, the following equation holds*

$$
\|\underline{\boldsymbol{X}}\| = \|\boldsymbol{X}_1\|_F = \|\boldsymbol{X}_2\|_F = \|\boldsymbol{X}_3\|_F,
$$

where $\|\cdot\|_F$ is the Frobenius norm; that is, $\|\boldsymbol{X}_1\|_F = \sqrt{\mathrm{tr}(\boldsymbol{X}_1'\boldsymbol{X}_1)}$.

Proof. From the definition of the norm of three-mode three-way data and the definition of unfolding mode,

$$\|\underline{\boldsymbol{X}}\| = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}^2} = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{iI_{\boldsymbol{X}}(j,k)}^2} = \sqrt{\mathrm{tr}(\boldsymbol{X}_1'\boldsymbol{X}_1)} = \|\boldsymbol{X}_1\|_F$$

holds. Other equations are proved in the same way. □

Hereafter, we denote the norm for matrix $\|\boldsymbol{X}\| = \|\boldsymbol{X}\|_F$.

Proposition 2.2. *Given two three-mode three-way data* $\underline{\boldsymbol{X}} \in \mathbb{R}^{I\times J\times K}$ *and* $\underline{\boldsymbol{Y}} \in \mathbb{R}^{I\times J\times K}$*, we obtain the following equation:*

$$< \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}} >= \mathrm{tr}(\boldsymbol{X}_1'\boldsymbol{Y}_1) = \mathrm{tr}(\boldsymbol{X}_2'\boldsymbol{Y}_2) = \mathrm{tr}(\boldsymbol{X}_3'\boldsymbol{Y}_3).$$

Proof. From the definition of the inner product and the definition of unfolding mode,

$$< \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}} >= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}y_{ijk} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{iI_{\boldsymbol{X}}(j,k)}y_{iI_{\boldsymbol{Y}}(j,k)} = \mathrm{tr}(\boldsymbol{X}_1'\boldsymbol{Y}_1)$$

holds. Other equations are proved in the same way. □

Now, we define the transform operator matrix to three-mode three-way data for introducing the mode product.

Definition 2.6. ***Folding mode of three-mode three-way data***
Given $\boldsymbol{X}_1 = (x_{i\ell}) \in \mathbb{R}^{I\times JK}$*, we define the folding operator* $[\cdot]_{J,K}^{(1)}$ *as* $[\boldsymbol{X}_1]_{J,K}^{(1)} = \underline{\boldsymbol{X}} = (x_{ijk}) \in \mathbb{R}^{I\times J\times K}$*, where* $j = \mathrm{mod}(\ell, J), k = \mathrm{int}(\ell/J)$*,* $\mathrm{mod}(\ell,\ o)$ *is* ℓ *modulo* o*, and* $\mathrm{int}(\ell)$ *is the integer part of* ℓ*. Given* $\boldsymbol{X}_2 = (x_{j\ell}) \in \mathbb{R}^{J\times IK}, \boldsymbol{X}_3 = (x_{k\ell}) \in \mathbb{R}^{K\times IJ}$*, we also define the folding operator* $[\cdot]_{I,K}^{(2)}$ *as* $[\boldsymbol{X}_2]_{I,K}^{(2)} = \underline{\boldsymbol{X}} = (x_{ijk}) \in \mathbb{R}^{I\times J\times K}$*, where* $i = \mathrm{mode}(\ell,\ I)$ *and* $k = \mathrm{int}(\ell/I)$*, and* $[\cdot]_{I,J}^{(3)}$ *as* $[\boldsymbol{X}_3]_{I,J}^{(3)} = \underline{\boldsymbol{X}} = (x_{ijk}) \in \mathbb{R}^{I\times J\times K}$ *where* $i = \mathrm{mode}(\ell,\ I)$*, and* $j = \mathrm{int}(\ell/I)$*.*

Figure 2.2 depicts the folding operator. The debate about the folding operator involves the matrix, which is made on the assumption of the unfolding operator. Thus, we should indicate the number of dimensions of each mode and the mode corresponding to the row.

Using unfolding mode and folding mode, we define the multiplier between the three-mode three-way data and matrix.

Definition 2.7. ***Mode product of three-mode three-way data***
Given three-mode three-way data $\underline{\boldsymbol{X}} \in \mathbb{R}^{I\times J\times K}$ *and matrixes* $\boldsymbol{U} \in \mathbb{R}^{I\times r}, \boldsymbol{V} \in \mathbb{R}^{J\times r},$ *and* $\boldsymbol{W} \in \mathbb{R}^{K\times r}$*, the mode product of 1, 2, and 3 is defined as follows:*

$$\begin{aligned}\underline{\boldsymbol{X}} \times_1 \boldsymbol{U} &= [\boldsymbol{U}'\boldsymbol{X}_1]_{J,K}^{(1)} \in \mathbb{R}^{r\times J\times K},\\ \underline{\boldsymbol{X}} \times_2 \boldsymbol{V} &= [\boldsymbol{V}'\boldsymbol{X}_2]_{I,K}^{(2)} \in \mathbb{R}^{I\times r\times K},\\ \underline{\boldsymbol{X}} \times_3 \boldsymbol{W} &= [\boldsymbol{W}'\boldsymbol{X}_3]_{I,J}^{(3)} \in \mathbb{R}^{I\times J\times r}.\end{aligned}$$

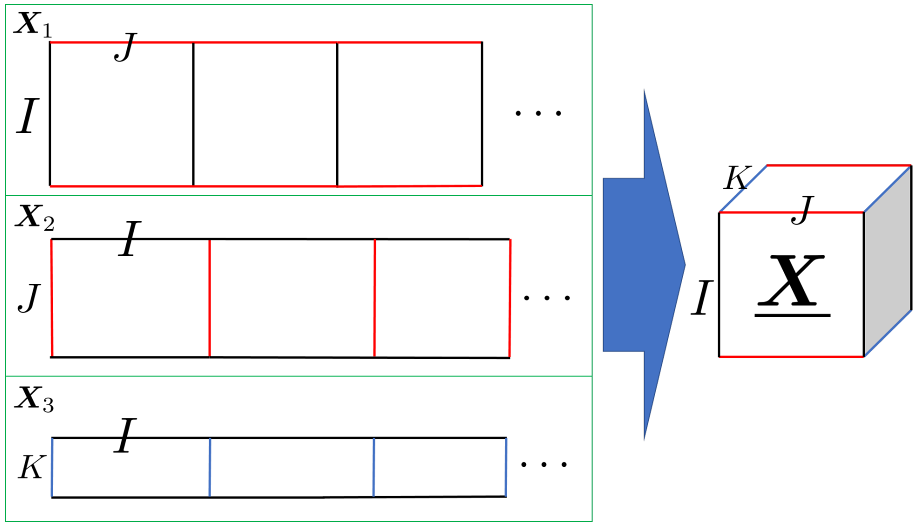


Figure 2.2: Representation of folding mode

The mode product is defined by using the matrix product. Thus, we obtain the relationships between the matrix calculation and the mode product.

Lemma 2.1. *Given the three-mode three-way data* $\underline{\boldsymbol{X}} \in \mathbb{R}^{I\times J\times K}$*, and three matrixes* $\boldsymbol{U} \in \mathbb{R}^{I\times r_1}$*,* $\boldsymbol{V} \in \mathbb{R}^{J\times r_2}$*, and* $\boldsymbol{W} \in \mathbb{R}^{K\times r_3}$*, we set* $\underline{\boldsymbol{Y}} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{U}$*. We obtain the following equations:*

$$\boldsymbol{Y}_1 = \boldsymbol{U}'\boldsymbol{X}_1,$$
$$\boldsymbol{Y}_2 = \boldsymbol{X}_2(\boldsymbol{I} \otimes \boldsymbol{U}),$$
$$\boldsymbol{Y}_3 = \boldsymbol{X}_3(\boldsymbol{I} \otimes \boldsymbol{U}).$$

When we set $\underline{\boldsymbol{Z}} = \underline{\boldsymbol{X}} \times_2 \boldsymbol{V}$*, we obtain the following equations:*

$$\boldsymbol{Z}_1 = \boldsymbol{X}_1(\boldsymbol{I} \otimes \boldsymbol{V}),$$
$$\boldsymbol{Z}_2 = \boldsymbol{V}'\boldsymbol{X}_2,$$
$$\boldsymbol{Z}_3 = \boldsymbol{X}_3(\boldsymbol{V} \otimes \boldsymbol{I}).$$

When we set $\underline{\boldsymbol{Q}} = \underline{\boldsymbol{X}} \times_3 \boldsymbol{W}$*, we obtain the following equations:*

$$\boldsymbol{Q}_1 = \boldsymbol{X}_1(\boldsymbol{W} \otimes \boldsymbol{I}),$$
$$\boldsymbol{Q}_2 = \boldsymbol{X}_2(\boldsymbol{W} \otimes \boldsymbol{I}),$$
$$\boldsymbol{Q}_3 = \boldsymbol{W}'\boldsymbol{X}_3,$$

where $\otimes$ *shows the Kronecker products and* $\boldsymbol{I}$ *is an identity matrix.*

Proof. $\boldsymbol{Y}_1 = \boldsymbol{U}'\boldsymbol{X}_1$ is the definition of mode product. From the definition of mode product, the element of $\underline{\boldsymbol{Y}}$ is obtained as follows:

$$y_{ijk} = \sum_{\ell=1}^{I} u_{\ell i} x_{\ell I_{\boldsymbol{X}}(j,k)}.$$

On the other hand, with regard to the element of $\boldsymbol{X}_2(\boldsymbol{I} \otimes \boldsymbol{U})$,

$$\sum_{\ell=1}^{I} x_{jJ_{\boldsymbol{X}}(\ell,k)} u_{\ell i} = y_{jJ_{\boldsymbol{Y}}(i,k)}$$

holds. Similarly, with regard to the element of $\boldsymbol{X}_3(\boldsymbol{I} \otimes \boldsymbol{U})$,

$$\sum_{\ell=1}^{I} x_{kK_{\boldsymbol{X}}(\ell,j)} u_{\ell i} = y_{kK_{\boldsymbol{X}}(i,j)}$$

holds.

Next, we consider proofs regarding $\underline{\boldsymbol{Z}}$. From the definition of mode product, the element of $\underline{\boldsymbol{Z}}$ is obtained as follows:

$$z_{ijk} = \sum_{\ell=1}^{J} v_{\ell j} x_{\ell J_{\boldsymbol{X}}(i,k)}.$$

On the other hand, with regard to the element of $\boldsymbol{X}_3(\boldsymbol{V} \otimes \boldsymbol{I})$,

$$\sum_{\ell=1}^{J} x_{kK_{\boldsymbol{X}}(i,\ell)} v_{\ell j} = z_{kK_{\boldsymbol{X}}(i,j)}$$

holds.

The proofs for the other equations are obtained in the same way as above. □

Proposition 2.3. *Given the three-mode three-way data* $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J \times K}$*, and three matrixes* $\boldsymbol{U} \in \mathbb{R}^{I \times r_1}$*,* $\boldsymbol{V} \in \mathbb{R}^{J \times r_2}$*, and* $\boldsymbol{W} \in \mathbb{R}^{K \times r_3}$*, we set* $\underline{\boldsymbol{Y}} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W}$*. We obtain the following equations:*

$$\begin{aligned} \boldsymbol{Y}_1 &= \boldsymbol{U}'\boldsymbol{X}_1(\boldsymbol{W} \otimes \boldsymbol{V}), \\ \boldsymbol{Y}_2 &= \boldsymbol{V}'\boldsymbol{X}_2(\boldsymbol{W} \otimes \boldsymbol{U}), \\ \boldsymbol{Y}_3 &= \boldsymbol{W}'\boldsymbol{X}_3(\boldsymbol{V} \otimes \boldsymbol{U}). \end{aligned}$$

Proof. From Lemma 2.1, $\underline{\boldsymbol{Z}} = (\boldsymbol{X}_1 \times_1 \boldsymbol{U} \times_2 \boldsymbol{V})$ is obtained as follows:

$$\underline{\boldsymbol{Z}} = [\boldsymbol{V}'\boldsymbol{X}_2(\boldsymbol{I} \otimes \boldsymbol{U})]^{(2)}_{I,K}.$$

In the same way, we set $\underline{\boldsymbol{Q}} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{U}$, and then $\underline{\boldsymbol{Y}} = (\boldsymbol{X}_1 \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W})$ is obtained as follows:

$$\begin{aligned} \underline{\boldsymbol{Y}} &= [\boldsymbol{W}'\boldsymbol{Q}_3(\boldsymbol{V} \otimes \boldsymbol{I})]^{(3)}_{I,J} \\ &= [\boldsymbol{W}'\boldsymbol{X}_3(\boldsymbol{I} \otimes \boldsymbol{U})(\boldsymbol{V} \otimes \boldsymbol{I})]^{(3)}_{I,J} \\ &= [\boldsymbol{W}'\boldsymbol{X}_3(\boldsymbol{V} \otimes \boldsymbol{U})]^{(3)}_{I,J}. \end{aligned}$$

Therefore, we obtain $\boldsymbol{Y}_3 = \boldsymbol{W}'\boldsymbol{X}_3(\boldsymbol{V} \otimes \boldsymbol{U})$. From Lemma 2.1, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are obtained as follows:

$$\begin{aligned} \boldsymbol{Y}_1 &= \boldsymbol{Z}_1(\boldsymbol{W} \otimes \boldsymbol{I}) = \boldsymbol{Q}_1(\boldsymbol{I} \otimes \boldsymbol{V})(\boldsymbol{W} \otimes \boldsymbol{I}) = \boldsymbol{U}'\boldsymbol{X}_1(\boldsymbol{W} \otimes \boldsymbol{V}), \\ \boldsymbol{Y}_2 &= \boldsymbol{Z}_2(\boldsymbol{W} \otimes \boldsymbol{I}) = \boldsymbol{V}'\boldsymbol{Q}_2(\boldsymbol{W} \otimes \boldsymbol{I}) = \boldsymbol{V}'\boldsymbol{X}_2(\boldsymbol{W} \otimes \boldsymbol{U}). \end{aligned}$$

□

Proposition 2.3 shows the relationship between the operator for matrix and mode product. Many multivariate data analyses for three-mode three-way data are described as mode product. Using this proposition, we rewrite the multivariate analysis for the multivariate data matrix.

Example 2.4. Three-mode three-way regression analysis
Given two three-mode three-way data $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J \times K}$ *and* $\underline{\boldsymbol{Y}} \in \mathbb{R}^{I \times L \times M}$*, the model of three-mode three-way regression analysis is as follows:*

$$\underline{\boldsymbol{Y}} \simeq \underline{\boldsymbol{X}} \times_2 \boldsymbol{U} \times_3 \boldsymbol{V},$$

where $\boldsymbol{U} \in \mathbb{R}^{J \times L}$ *and* $\boldsymbol{V} \in \mathbb{R}^{K \times M}$*. This model formula is also described as follows:*

$$\boldsymbol{Y}_1 \simeq \boldsymbol{X}_1(\boldsymbol{V} \otimes \boldsymbol{U}).$$

Example 2.5. Three-mode three-way principal component analysis
Given three-mode three-way data $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J \times K}$*, the model of three-mode three-way principal component analysis is described as follows:*

$$\underline{\boldsymbol{X}} \simeq \underline{\boldsymbol{G}} \times_1 \boldsymbol{A} \times_2 \boldsymbol{B} \times_3 \boldsymbol{C},$$

where $\boldsymbol{G} \in \mathbb{R}^{r_a \times r_b \times r_c}$*,* $\boldsymbol{A} \in \mathbb{R}^{r_a \times I}$*,* $\boldsymbol{B} \in \mathbb{R}^{r_b \times J}$*, and* $\boldsymbol{C} \in \mathbb{R}^{r_c \times K}$*. This model formula is also described as follows:*

$$\boldsymbol{X}_1 \simeq \boldsymbol{A}'\boldsymbol{G}_1(\boldsymbol{C} \otimes \boldsymbol{B}).$$

This formulation is proposed by Tucker (1966) and is called the Tucker 3 model.

Chapter 3

Three-mode three-way canonical covariance analysis

In this chapter, we describe the models and algorithm of the proposed methods for three-mode three-way data. In Section 3.1, we explain the basic method for three-mode three-way data. Because the basic method is simply an extension of two-mode two-way data to three-mode three-way data, this method has some problems. To address some of these problems, we introduce the connector matrix. In Section 3.2, we explain some constrained connector methods, which are broadly divided into clustering- and regression-based methods. The former is based on K-means and spherical K-means. The latter corresponds to simultaneous analysis of regression and dimensional reduction. The difference between the two methods is related to update rules for the connector matrix. The update rules of K-means and spherical K-means correspond to covariance and correlation, respectively. In section 3.3, we explain the quantification method for three-mode three-way canonical covariance analysis. The quantification method is based on NPCA. We could apply canonical covariance analysis to three-mode three-way categorical data by quantification method.

3.1 Basic method and connector matrix method

In this section, we explain the basic method of the canonical covariance method for three-mode three-way data. First, we discuss the two-mode two-way canonical covariance analysis method. Then, we define the model of the basic method. Finally, we describe the connector matrix method by introducing the connector matrix.

3.1.1 Basic method

In this subsection, we introduce the basic method of three-way three-mode canonical covariance analysis. For this purpose, we first introduce the two-mode two-way canonical covariance method based on Tenenhaus & Tenenhaus (2011); Tenenhaus et al. (2017). Then, we extend this model to three-mode three-way data analysis.

3.1.1.1 Model and objective function

Given two-mode two-way data $\boldsymbol{X} \in \mathbb{R}^{I\times J_x},\ \boldsymbol{Y} \in \mathbb{R}^{I\times J_y}$, the model of canonical covariance analysis is described as follows:

$$\boldsymbol{Y}\boldsymbol{B} = \boldsymbol{X}\boldsymbol{A} + \boldsymbol{E},$$

where $\boldsymbol{A} \in \mathbb{R}^{J_x\times r},\ \boldsymbol{B} \in \mathbb{R}^{J_y\times r}$ are weight matrixes for $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. $\boldsymbol{E} \in \mathbb{R}^{I\times r}$ is an error matrix and r is the number of factors. When $\boldsymbol{A}$ and $\boldsymbol{B}$ are column orthogonal, this model is called a canonical covariance model. On the other hand, when $\boldsymbol{A}$ and $\boldsymbol{B}$ satisfy $\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{B} = \boldsymbol{I}$, this model is called the canonical correlation model. The objective function f of the canonical covariance model is defined as follows:

$$\begin{aligned} f(\boldsymbol{A},\ \boldsymbol{B}|\ \boldsymbol{X},\ \boldsymbol{Y}) &= \mathrm{tr}(\boldsymbol{A}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B}) \\ \text{subject to } &\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B} = \boldsymbol{I} \end{aligned} \tag{3.1}$$

This objective function is also defined as follows:

$$\begin{aligned} g(\boldsymbol{A},\ \boldsymbol{B}|\ \boldsymbol{X},\ \boldsymbol{Y}) =&\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\boldsymbol{A}'\|^2 + \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{B}\boldsymbol{B}'\|^2 \\ &+ \|\boldsymbol{X}\boldsymbol{A} - \boldsymbol{Y}\boldsymbol{B}\|^2 \\ \text{subject to } &\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B} = \boldsymbol{I} \end{aligned} \tag{3.2}$$

Proposition 3.1.
Maximizing the objective functions (3.1) *is equivalent to minimizing the objective function* (3.2).

Proof. From the constraint of parameters, we obtain the following equation:

$$\begin{aligned} g(\boldsymbol{A},\boldsymbol{B}|\boldsymbol{X},\boldsymbol{Y}) =&\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\boldsymbol{A}'\|^2 + \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{B}\boldsymbol{B}'\|^2 + \|\boldsymbol{X}\boldsymbol{A} - \boldsymbol{Y}\boldsymbol{B}\|^2 \\ =&\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X}) - \mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{Y}'\boldsymbol{Y}) - \mathrm{tr}(\boldsymbol{B}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{B}) \\ &+ \mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{B}) - 2\mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B}) \\ =& -2\mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B}) + \text{const.} \longrightarrow \text{minimize} \end{aligned}$$

where const. is constant value independent from parameters. When maximizing $\mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B})$, we obtain the estimated parameter. Therefore, the proposition holds. □

Proposition 3.1 does not hold in the canonical correlation model.

Proposition 3.2.
Maximizing the objective function f defined as follows:

$$\begin{aligned} f(\boldsymbol{A},\ \boldsymbol{B}|\ \boldsymbol{X},\ \boldsymbol{Y}) &= tr(\boldsymbol{A}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B}) \\ \textit{subject to } &\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{B} = \boldsymbol{I} \end{aligned} \tag{3.3}$$

is not equivalent to the objective function defined as follows:

$$\begin{aligned} g(\boldsymbol{A},\ \boldsymbol{B}|\ \boldsymbol{X},\ \boldsymbol{Y}) =&\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\boldsymbol{A}'\|^2 + \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{B}\boldsymbol{B}'\|^2 \\ &+ \|\boldsymbol{X}\boldsymbol{A} - \boldsymbol{Y}\boldsymbol{B}\|^2 \\ \textit{subject to } &\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{B} = \boldsymbol{I} \end{aligned} \tag{3.4}$$

Proof. From the constraint of parameters, we obtain the following equation:

$$\begin{aligned} g(\boldsymbol{A},\ \boldsymbol{B}|\ \boldsymbol{X},\ \boldsymbol{Y}) =& \|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{A}\boldsymbol{A}'\|^2+\|\boldsymbol{Y}-\boldsymbol{Y}\boldsymbol{B}\boldsymbol{B}'\|^2 \\ &+\|\boldsymbol{X}\boldsymbol{A}-\boldsymbol{Y}\boldsymbol{B}\|^2 \\ =& -\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}\boldsymbol{A})-\mathrm{tr}(\boldsymbol{B}\boldsymbol{B}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{B}\boldsymbol{B}')-\mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B})+const. \\ =& -\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})-\mathrm{tr}(\boldsymbol{B}'\boldsymbol{B})-\mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B})+const. \end{aligned}$$

where const. is constant value independent from parameters. Therefore, minimizing the objective function 3.4 is the same problem of maximizing the objective function f defined as follows:

$$f(\boldsymbol{A},\ \boldsymbol{B}|\ \boldsymbol{X},\ \boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{B})+\|\boldsymbol{A}\|^2+\|\boldsymbol{B}\|^2$$
$$\text{subject to } \boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}=\boldsymbol{B}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{B}=\boldsymbol{I}.$$

□

Proposition 3.2 also shows that the objective function of canonical correlation analysis is equivalent to $\|\boldsymbol{X}\boldsymbol{A}-\boldsymbol{Y}\boldsymbol{B}\|^2$. Therefore, when $\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{Y}'\boldsymbol{Y}=\boldsymbol{I}$ hold, the objective function of canonical covariance analysis is equivalent to the objective function of canonical correlation analysis.

We extend the two-mode two-way canonical covariance model to the three-mode three-way canonical covariance model using the concept of the Tucker model. Given two three-mode three-way data $\underline{\boldsymbol{X}} \in \mathbb{R}^{I\times J_x\times K_x}$ and $\underline{\boldsymbol{Y}} \in \mathbb{R}^{I\times J_y\times K_y}$, the model formula of the three-mode three-way model is defined as follows:

$$\boldsymbol{Y}_1(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)=\boldsymbol{X}_1(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)+\boldsymbol{E}, \tag{3.5}$$

where $\boldsymbol{B}_x \in \mathbb{R}^{J_x\times r_b}$, $\boldsymbol{B}_y \in \mathbb{R}^{J_y\times r_b}$, $\boldsymbol{C}_x \in \mathbb{R}^{K_x\times r_c}$, $\boldsymbol{C}_y \in \mathbb{R}^{K_y\times r_c}$ are weight matrixes for mode 2 of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$, and mode 3 of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$. The weight matrixes are a column orthogonal matrix. Mode 2 and 3 show the variable and conditions, respectively, for popular cases of data analysis.

The objective function g_b of the three-mode three-way data is defined as follows:

$$\begin{aligned} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y|\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}) =& \|\boldsymbol{X}_1-\boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes\boldsymbol{B}_x\boldsymbol{B}_x')\|^2+\|\boldsymbol{Y}_1-\boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y'\otimes\boldsymbol{B}_y\boldsymbol{B}_y')\|^2 \\ &+\|\boldsymbol{X}_1(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)-\boldsymbol{Y}_1(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)\|^2 \end{aligned} \tag{3.6}$$
$$\text{subject to } \boldsymbol{B}_x'\boldsymbol{B}_x=\boldsymbol{B}_y'\boldsymbol{B}_y=\boldsymbol{I}, \boldsymbol{C}_x'\boldsymbol{C}_x=\boldsymbol{C}_y'\boldsymbol{C}_y=\boldsymbol{I}.$$

The objective function (3.6) is equivalent to maximizing the covariance problem. This proposition is proved in the same way as Proposition 3.1. However, this method has indeterminate parameters. When we set $\boldsymbol{B}_x^{(*)}=\boldsymbol{B}_x\boldsymbol{S}$, $\boldsymbol{B}_y^{(*)}=\boldsymbol{B}_y\boldsymbol{S}$, $\boldsymbol{C}_x^{(*)}=\boldsymbol{C}_x\boldsymbol{T}$, $\boldsymbol{C}_y^{(*)}=\boldsymbol{C}_y\boldsymbol{T}$ by using the orthonormal matrixes $\boldsymbol{S}$ and $\boldsymbol{T}$, the objective function (3.6) takes the same value. Therefore, we can use rotation methods for interpretation.

Proposition 3.3. *We assume that $\boldsymbol{S}$ and $\boldsymbol{T}$ are orthonormal. When we set $\boldsymbol{B}_x^{(*)}=\boldsymbol{B}_x\boldsymbol{S}$, $\boldsymbol{B}_y^{(*)}=\boldsymbol{B}_y\boldsymbol{S}$, $\boldsymbol{C}_x^{(*)}=\boldsymbol{C}_x\boldsymbol{T}$, $\boldsymbol{C}_y^{(*)}=\boldsymbol{C}_y\boldsymbol{T}$, we obtain the following equation:*

$$g_b(\boldsymbol{B}_x^{(*)},\ \boldsymbol{B}_y^{(*)},\ \boldsymbol{C}_x^{(*)},\ \boldsymbol{C}_y^{(*)}|\ \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})=g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y|\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}),$$

where g_b is defined as equation (3.6).

Proof. From the definition of the objective function,

$$\begin{aligned} g_b(\boldsymbol{B}_x^{(*)},\ \boldsymbol{B}_y^{(*)},\ \boldsymbol{C}_x^{(*)},\ \boldsymbol{C}_y^{(*)} |\ \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}}) =& \|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x^{(*)}\boldsymbol{C}_x^{(*)\prime} \otimes \boldsymbol{B}_x^{(*)}\boldsymbol{B}_x^{(*)\prime})\|^2 \\ &+ \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)}\boldsymbol{C}_y^{(*)\prime} \otimes \boldsymbol{B}_y^{(*)}\boldsymbol{B}_y^{(*)\prime})\|^2 \\ &+ \|\boldsymbol{X}_1(\boldsymbol{C}_x^{(*)} \otimes \boldsymbol{B}_x^{(*)}) - \boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)} \otimes \boldsymbol{B}_y^{(*)})\|^2 \end{aligned} \tag{3.7}$$

holds. The first term of the right-hand side of equation (3.7) is

$$\begin{aligned} \|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x^{(*)}\boldsymbol{C}_x^{(*)\prime} \otimes \boldsymbol{B}_x^{(*)}\boldsymbol{B}_x^{(*)\prime})\|^2 =& \|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{T}\boldsymbol{T}'\boldsymbol{C}_x' \otimes \boldsymbol{B}_x\boldsymbol{S}\boldsymbol{S}'\boldsymbol{B}_x')\|^2 \\ =& \|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{B}_x\boldsymbol{B}_x')\|^2. \end{aligned}$$

The second term of the right-hand side of equation (3.7) is

$$\begin{aligned} \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)}\boldsymbol{C}_y^{(*)\prime} \otimes \boldsymbol{B}_y^{(*)}\boldsymbol{B}_y^{(*)\prime})\|^2 =& \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{T}\boldsymbol{T}'\boldsymbol{C}_y' \otimes \boldsymbol{B}_y\boldsymbol{S}\boldsymbol{S}'\boldsymbol{B}_y')\|^2 \\ =& \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{B}_y\boldsymbol{B}_y')\|^2. \end{aligned}$$

The third term of the right-hand side of equation (3.7) is

$$\begin{aligned} &\|\boldsymbol{X}_1(\boldsymbol{C}_x^{(*)} \otimes \boldsymbol{B}_x^{(*)}) - \boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)} \otimes \boldsymbol{B}_y^{(*)})\|^2 \\ =& \mathrm{tr}((\boldsymbol{T} \otimes \boldsymbol{S})'(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)(\boldsymbol{T} \otimes \boldsymbol{S})) \\ &- 2\mathrm{tr}((\boldsymbol{T} \otimes \boldsymbol{S})'(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)(\boldsymbol{T} \otimes \boldsymbol{S})) \\ &+ \mathrm{tr}((\boldsymbol{T} \otimes \boldsymbol{S})'(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)'\boldsymbol{Y}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)(\boldsymbol{T} \otimes \boldsymbol{S})) \\ =& \mathrm{tr}((\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)(\boldsymbol{T} \otimes \boldsymbol{S})(\boldsymbol{T} \otimes \boldsymbol{S})') \\ &- 2\mathrm{tr}((\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)(\boldsymbol{T} \otimes \boldsymbol{S})(\boldsymbol{T} \otimes \boldsymbol{S})') \\ &+ \mathrm{tr}((\boldsymbol{C}_y \otimes \boldsymbol{B}_y)'\boldsymbol{Y}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)(\boldsymbol{T} \otimes \boldsymbol{S})(\boldsymbol{T} \otimes \boldsymbol{S})') \\ =& \mathrm{tr}((\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)) \\ &- 2\mathrm{tr}((\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)) \\ &+ \mathrm{tr}((\boldsymbol{C}_y \otimes \boldsymbol{B}_y)'\boldsymbol{Y}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)) \\ =& \|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x) - \boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\|^2. \end{aligned}$$

When we summarize all terms, the proposition holds. □

3.1.1.2 Algorithm

Here, we explain the algorithm of three-mode three-way canonical covariance. It is difficult to optimize the objective function (3.6) because it is not a convex function of all parameters. Therefore, we adopt the update algorithm as an alternative least squares algorithm.

When other parameters are given, we adopt the algorithm proposed by Jennrich (2001) as an update algorithm of $\boldsymbol{B}_x,\ \boldsymbol{B}_y, \boldsymbol{C}_x$, and $\boldsymbol{C}_y$. Jennrich's approach requires the derivative function of the objective function. Given this, the update formula is described by using singular value decomposition of the derivative function.

Proposition 3.4. *We set g_b as the objective function, as defined in equation* (3.6)*. The partial derivative functions of the objective function with respect to each parameter are obtained as follows:*

$$\frac{\partial}{\partial \boldsymbol{B}_x} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) = -2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y, \tag{3.8}$$

$$\frac{\partial}{\partial \boldsymbol{B}_y} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) = -2\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x, \tag{3.9}$$

$$\frac{\partial}{\partial \boldsymbol{C}_x} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) = -2\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y, \tag{3.10}$$

$$\frac{\partial}{\partial \boldsymbol{C}_y} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) = -2\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x, \tag{3.11}$$

Proof. From the definition of g_b and Proposition 2.1 and 2.3, we obtain the following equations:

$$\begin{aligned}
&g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) \\
=&\|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{B}_x\boldsymbol{B}_x')\|^2 + \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{B}_x\boldsymbol{B}_x')\|^2 \\
&+ \|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x) - \boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\|^2 \\
=&\|\boldsymbol{X}_2 - \boldsymbol{B}_x\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\|^2 + \|\boldsymbol{Y}_2 - \boldsymbol{B}_y\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{I})\|^2 \\
&+ \|\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{I}) - \boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y \otimes \boldsymbol{I})\|^2 \qquad (3.12) \\
=&\|\boldsymbol{X}_3 - \boldsymbol{C}_x\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{I})\|^2 + \|\boldsymbol{Y}_3 - \boldsymbol{C}_y\boldsymbol{C}_y'\boldsymbol{Y}_3((\boldsymbol{B}_y\boldsymbol{B}_y' \otimes \boldsymbol{I}))\|^2 \\
&+ \|\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{I}) - \boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y \otimes \boldsymbol{I})\|^2. \qquad (3.13)
\end{aligned}$$

Then, we calculate the partial derivative function of the objective function with respect to $\boldsymbol{B}_x$. First, we obtain the following equation:

$$\begin{aligned}
&\|\boldsymbol{X}_2 - \boldsymbol{B}_x\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\|^2 + \|\boldsymbol{Y}_2 - \boldsymbol{B}_y\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{I})\|^2 \\
&+ \|\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{I}) - \boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y \otimes \boldsymbol{I})\|^2 \\
=&\mathrm{tr}(\boldsymbol{X}_2'\boldsymbol{X}_2) - 2\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x) + \mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x) \\
&+ \mathrm{tr}(\boldsymbol{Y}_2'\boldsymbol{Y}_2) - 2\mathrm{tr}(\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y) + \mathrm{tr}(\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y) \\
&+ \mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x) - 2\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y) \\
&+ \mathrm{tr}(\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y) \\
=&\mathrm{tr}(\boldsymbol{X}_2'\boldsymbol{X}_2) + \mathrm{tr}(\boldsymbol{Y}_2'\boldsymbol{Y}_2) - 2\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2'(\boldsymbol{C}_x\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2\boldsymbol{B}_y).
\end{aligned}$$

Therefore, the partial derivative function of the objective function with respect to $\boldsymbol{B}_x$ is obtained as follows:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{B}_x} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) =&\frac{\partial}{\partial \boldsymbol{B}_x}\{\mathrm{tr}(\boldsymbol{X}_2'\boldsymbol{X}_2) + \mathrm{tr}(\boldsymbol{Y}_2'\boldsymbol{Y}_2) \\
&- 2\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y)\} \\
=&-2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y.
\end{aligned}$$

The partial derivative function of the objective function with respect to $\boldsymbol{B}_y$ is also obtained as follows:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{B}_y} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y|\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}) =& \frac{\partial}{\partial \boldsymbol{B}_y}\{\mathrm{tr}(\boldsymbol{X}_2'\boldsymbol{X}_2) + \mathrm{tr}(\boldsymbol{Y}_2'\boldsymbol{Y}_2) \\ & - 2\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_y \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y)\} \\ =& - 2\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x.\end{aligned}$$

Then, we calculate the partial derivative function of the objective function with respect to $\boldsymbol{C}_x$. For this purpose, we rewrite equation (3.13) as

$$\begin{aligned}& \|\boldsymbol{X}_3 - \boldsymbol{C}_x\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{I})\|^2 + \|\boldsymbol{Y}_3 - \boldsymbol{C}_y\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y' \otimes \boldsymbol{I})\|^2 \\ & + \|\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{I}) - \boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y \otimes \boldsymbol{I})\|^2 \\ =& \mathrm{tr}(\boldsymbol{X}_3'\boldsymbol{X}_3) - 2\mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x) + \mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x) \\ & + \mathrm{tr}(\boldsymbol{Y}_3'\boldsymbol{Y}_3) - 2\mathrm{tr}(\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y) + \mathrm{tr}(\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y) \\ & + \mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x) - 2\mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y) \\ & + \mathrm{tr}(\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y) \\ =& \mathrm{tr}(\boldsymbol{X}_3'\boldsymbol{X}_3) + \mathrm{tr}(\boldsymbol{Y}_3'\boldsymbol{Y}_3) - 2\mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3'(\boldsymbol{B}_x\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3\boldsymbol{C}_y).\end{aligned}$$

Thus, the partial derivative function of the objective function with respect to $\boldsymbol{C}_x$ is obtained as follows:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{C}_x} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y|\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}) =& \frac{\partial}{\partial \boldsymbol{C}_x}\{\mathrm{tr}(\boldsymbol{X}_3'\boldsymbol{X}_3) + \mathrm{tr}(\boldsymbol{Y}_3'\boldsymbol{Y}_3) \\ & - 2\mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y)\} \\ =& - 2\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y.\end{aligned}$$

The partial derivative function of the objective function with respect to $\boldsymbol{C}_y$ is also obtained as follows:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{C}_y} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y|\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}) =& \frac{\partial}{\partial \boldsymbol{C}_y}\{\mathrm{tr}(\boldsymbol{X}_3'\boldsymbol{X}_3) + \mathrm{tr}(\boldsymbol{Y}_3'\boldsymbol{Y}_3) \\ & - 2\mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y)\} \\ =& - 2\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x.\end{aligned}$$

□

Given a sufficient big constant α, using Jennrich's approach, the update formulas of the parameters are obtained as follows:

$$\begin{aligned}\boldsymbol{B}_x^{(t+1)} &= \boldsymbol{U}_{bx}\boldsymbol{V}_{bx}', \\ \boldsymbol{B}_y^{(t+1)} &= \boldsymbol{U}_{by}\boldsymbol{V}_{by}', \\ \boldsymbol{C}_x^{(t+1)} &= \boldsymbol{U}_{cx}\boldsymbol{V}_{cx}', \\ \boldsymbol{C}_y^{(t+1)} &= \boldsymbol{U}_{cy}\boldsymbol{V}_{cy}',\end{aligned}$$

where $\boldsymbol{U}_{bx}$ and $\boldsymbol{V}_{bx}$ are the left and right matrixes, respectively, of the singular value decomposition of the matrix

$$\alpha\boldsymbol{B}_x^{(t)} - \frac{\partial}{\partial \boldsymbol{B}_x} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y|\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}),$$

$\boldsymbol{U}_{by}$ and $\boldsymbol{V}_{by}$ are also the left and right matrixes, respectively, of the singular value decomposition of the matrix

$$\alpha \boldsymbol{B}_y^{(t)} - \frac{\partial}{\partial \boldsymbol{B}_y} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}),$$

$\boldsymbol{U}_{cx}$ and $\boldsymbol{V}_{cx}$ are also the left and right matrixes, respectively, of the singular value decomposition of the matrix

$$\alpha \boldsymbol{C}_x^{(t)} - \frac{\partial}{\partial \boldsymbol{C}_x} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}),$$

and $\boldsymbol{U}_{cy}$ and $\boldsymbol{V}_{cy}$ are also the left and right matrixes, respectively, of the singular value decomposition of the matrix as follows:

$$\alpha \boldsymbol{C}_y^{(t)} - \frac{\partial}{\partial \boldsymbol{C}_y} g_b(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}).$$

Algorithm 1 shows the algorithm of the three-mode three-way canonical covariance analysis. The setting of α is important. If α is not sufficiently big, This algorithm does not ensure a monotonous decrease. Even if α is sufficiently big, the algorithm 1 yields only a monotonous decrease, not yield global optima. To obtain global optima, we generate many initial values of parameters.

Algorithm 1 Algorithm of basic method

Set the number of dimension r_b, r_c, and stop condition ε
Set initial values $\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_y^{(0)}, \alpha$
$t \leftarrow 0$
$S^{(0)} \leftarrow g_b(\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_y^{(0)} | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}})$
repeat
  $t \leftarrow t + 1$
  $\boldsymbol{B}_x^{(t)} \leftarrow \boldsymbol{U}_{bx}\boldsymbol{V}_{bx}'$ using $\boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)}$
  $\boldsymbol{B}_y^{(t)} \leftarrow \boldsymbol{U}_{by}\boldsymbol{V}_{by}'$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)}$
  $\boldsymbol{C}_x^{(t)} \leftarrow \boldsymbol{U}_{cx}\boldsymbol{V}_{cx}'$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_y^{(t-1)}$
  $\boldsymbol{C}_y^{(t)} \leftarrow \boldsymbol{U}_{cy}\boldsymbol{V}_{cy}'$ using $\boldsymbol{B}_x^{(t)}, \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)}$
  $S^{(t)} \leftarrow g_b(\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)} | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}})$
until $|S^{(t-1)} - S^{(t)}| \leq \varepsilon$

3.1.2 Connector matrix method

In the three-mode three-way canonical covariance method, it is assumed that the number of dimensions of weight matrix for the same mode is the same. This assumption is not suitable when there is a big difference between the two data sets in the number of dimensions of the same mode. The canonical covariance method considers the variance of the data. Thus, when the number of dimensions is large, then the canonical covariance method tends to maximize the variance of the canonical variable. To overcome this problem, we introduce the connector matrixes.

3.1.2.1 Model and objective function

Given two three-mode three-way data $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J_x \times K_x}$ and $\underline{\boldsymbol{Y}} \in \mathbb{R}^{I \times J_y \times K_y}$, the model of connector canonical covariance analysis for three-mode three-way data is defined as follows:

$$\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{D}_y = \boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{D}_x + \boldsymbol{E}, \tag{3.14}$$

where $\boldsymbol{B}_x \in \mathbb{R}^{J_x \times r_{bx}}$, $\boldsymbol{B}_y \in \mathbb{R}^{J_y \times r_{by}}$, $\boldsymbol{C}_x \in \mathbb{R}^{K_x \times r_{cx}}$, $\boldsymbol{C}_y \in \mathbb{R}^{K_y \times r_{cy}}$ are weight matrixes for mode 2 of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$, and mode 3 of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$. The weight matrixes are column orthogonal. $\boldsymbol{D}_x \in \mathbb{R}^{r_{cx} r_{bx} \times c_c c_b}$ and $\boldsymbol{D}_y \in \mathbb{R}^{r_{cx} r_{bx} \times c_c c_b}$ are connector matrixes for $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$, respectively. In this case, $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ have no structure of three-mode three-way data. Thus, we assume $\boldsymbol{D}_x = \boldsymbol{D}_{cx} \otimes \boldsymbol{D}_{bx}$, $\boldsymbol{D}_y = \boldsymbol{D}_{cy} \otimes \boldsymbol{D}_{by}$, where $\boldsymbol{D}_{bx} \in \mathbb{R}^{r_{bx} \times c_b}$, $\boldsymbol{D}_{by} \in \mathbb{R}^{r_{by} \times c_b}$ are connector matrixes of mode 2 $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$, respectively, and $\boldsymbol{D}_{cx} \in \mathbb{R}^{r_{cx} \times c_c}$, $\boldsymbol{D}_{cy} \in \mathbb{R}^{r_{cy} \times c_c}$ are connector matrixes of mode 3 $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$, respectively. r_{bx}, r_{by}, r_{cx} and r_{cy} are the number of dimensions of the weight matrix for variables of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ and for conditions of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$, respectively, and c_b and c_c are the number of connected factors of variable and conditions, respectively. We assume elements of connector matrixes are real numbers in general. However, for our interpretation, we often set the range of elements of connector matrixes as binary.

Introducing connector matrixes, the numbers of dimensions of the weight matrixes are different to each other. This characteristic is important for analysis of three-mode three-way data. In many cases, the numbers of dimensions of two three-mode three-way data are different. Therefore, the assumption that there is the same optimum number of dimensions for the PCA term is too strict. Using connector matrixes, we could loosen the assumption. On the other hand, the number of parameters increases.

The objective function g_c of connector canonical covariance analysis for three-mode three-way data is obtained as follows:

$$\begin{aligned} & g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y | \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) \\ =& \|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{B}_x\boldsymbol{B}_x')\|^2 + \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{B}_y\boldsymbol{B}_y')\|^2 \\ & + \|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{D}_x - \boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{D}_y\|^2 \end{aligned} \tag{3.15}$$

$$\text{subject to } \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I},\ \boldsymbol{B}_y'\boldsymbol{B}_y = \boldsymbol{I}, \boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I},\ \boldsymbol{C}_y'\boldsymbol{C}_y = \boldsymbol{I}.$$

This objective function is not equivalent to maximizing covariance in general. If connector matrixes are set as the indicator matrix, the objective function (3.15) equals the objective function (3.6). From this property, we regard the connector matrix method as one of the extensions of the basic method. Moreover, this method has rotational indeterminacy.

Proposition 3.5. *Given two orthonormal matrixes $\boldsymbol{S}$ and $\boldsymbol{T}$, when we set*

$$\begin{aligned} &\boldsymbol{B}_x^{(*)} = \boldsymbol{B}_x\boldsymbol{S},\ \boldsymbol{B}_y^{(*)} = \boldsymbol{B}_y\boldsymbol{S},\ \boldsymbol{C}_x^{(*)} = \boldsymbol{C}_x\boldsymbol{T},\ \boldsymbol{C}_y^{(*)} = \boldsymbol{C}_y\boldsymbol{T}, \\ &\boldsymbol{D}_x^{(*)} = (\boldsymbol{T} \otimes \boldsymbol{S})'\boldsymbol{D}_x,\ \boldsymbol{D}_y^{(*)} = (\boldsymbol{T} \otimes \boldsymbol{S})'\boldsymbol{D}_y, \end{aligned}$$

we obtain the following equation:

$$g_c(\boldsymbol{B}_x^{(*)},\ \boldsymbol{B}_y^{(*)},\ \boldsymbol{C}_x^{(*)},\ \boldsymbol{C}_y^{(*)}, \boldsymbol{D}_x^{(*)},\ \boldsymbol{D}_y^{(*)}|\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) = g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}),$$

where g_c *is defined as* (3.15).

Proof. From the definition of objective function g_c,

$$\begin{aligned}
&g_c(\boldsymbol{B}_x^{(*)},\ \boldsymbol{B}_y^{(*)},\ \boldsymbol{C}_x^{(*)},\ \boldsymbol{C}_y^{(*)} \boldsymbol{D}_x^{(*)},\ \boldsymbol{D}_x^{(*)}|\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) \\
=&\|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x^{(*)}\boldsymbol{C}_x^{(*)\prime} \otimes \boldsymbol{B}_x^{(*)}\boldsymbol{B}_x^{(*)\prime})\|^2 \\
&+ \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)}\boldsymbol{C}_y^{(*)\prime} \otimes \boldsymbol{B}_y^{(*)}\boldsymbol{B}_y^{(*)\prime})\|^2 \\
&+ \|\boldsymbol{X}_1(\boldsymbol{C}_x^{(*)} \otimes \boldsymbol{B}_x^{(*)})\boldsymbol{D}_x - \boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)} \otimes \boldsymbol{B}_y^{(*)})\boldsymbol{D}_y\|^2
\end{aligned} \tag{3.16}$$

holds. The first and second terms are proved in the same way as Proposition 3.3. The third term on the right-hand side of equation (3.16) is described as follows:

$$\begin{aligned}
&\|\boldsymbol{X}_1(\boldsymbol{C}_x^{(*)} \otimes \boldsymbol{B}_x^{(*)})\boldsymbol{D}_x - \boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)} \otimes \boldsymbol{B}_y^{(*)})\boldsymbol{D}_y\|^2 \\
=&\mathrm{tr}(\boldsymbol{D}_x^{(*)\prime}(\boldsymbol{C}_x^{(*)} \otimes \boldsymbol{B}_x^{(*)})'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{C}_x^{(*)} \otimes \boldsymbol{B}_x^{(*)})\boldsymbol{D}_x^{(*)}) \\
&- 2\mathrm{tr}(\boldsymbol{D}_x^{(*)\prime}(\boldsymbol{C}_x^{(*)} \otimes \boldsymbol{B}_x^{(*)})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)} \otimes \boldsymbol{B}_y^{(*)})\boldsymbol{D}_y^{(*)}) \\
&+ \mathrm{tr}(\boldsymbol{D}_y^{(*)\prime}(\boldsymbol{C}_y^{(*)} \otimes \boldsymbol{B}_y^{(*)})'\boldsymbol{Y}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y^{(*)} \otimes \boldsymbol{B}_y^{(*)})\boldsymbol{D}_y^{(*)}) \\
=&\mathrm{tr}(\boldsymbol{D}_x'(\boldsymbol{T}\boldsymbol{T}' \otimes \boldsymbol{S}\boldsymbol{S}')(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)(\boldsymbol{T}\boldsymbol{T}' \otimes \boldsymbol{S}\boldsymbol{S}')\boldsymbol{D}_x) \\
&- 2\mathrm{tr}(\boldsymbol{D}_x'(\boldsymbol{T}\boldsymbol{T}' \otimes \boldsymbol{S}\boldsymbol{S}')(\boldsymbol{C}_x \otimes \boldsymbol{B})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)(\boldsymbol{T}\boldsymbol{T}' \otimes \boldsymbol{S}\boldsymbol{S}')\boldsymbol{D}_y) \\
&+ \mathrm{tr}(\boldsymbol{D}_y'(\boldsymbol{T}\boldsymbol{T}' \otimes \boldsymbol{S}\boldsymbol{S}')(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)'\boldsymbol{Y}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)(\boldsymbol{T}\boldsymbol{T}' \otimes \boldsymbol{S}\boldsymbol{S}')\boldsymbol{D}_y) \\
=&\mathrm{tr}(\boldsymbol{D}_x'(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{D}_x) \\
&- 2\mathrm{tr}(\boldsymbol{D}_x'(\boldsymbol{C}_x \otimes \boldsymbol{B})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{D}_y) \\
&+ \mathrm{tr}(\boldsymbol{D}_y'(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)'\boldsymbol{Y}_1'\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{D}_y) = \|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{D}_x - \boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{D}_y\|^2.
\end{aligned}$$

When we summarize all terms, Proposition 3.5 holds. □

There are many points to consider about rotation methods of connector matrix canonical covariance analysis. One of the purposes of applying the rotation method is "interpretation." When connector matrixes are identical, we consider only the weight matrixes for the interpretation. In this case, it is difficult for rotation matrixes to be the same $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$. In other words, the rotation matrix is common between $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$. We need criteria for simplicity for both matrixes simultaneously. One of the solutions is the sum of the criteria, such as Varimax. It is easy to obtain the rotation matrixes by using Jennrich's approach.

It is more difficult to make criteria for simplicity for the connector matrixes with simultaneous rotation than for those without. Since the rotation matrixes for the connector matrixes are described as the Kronecker product, we consider simplicities of many matrixes. On the other hand, researchers are often interested in weight matrixes, because a purpose of this analysis is to seek subspace. Therefore, we suggest use of the same rotation method as that for the case in which connector matrixes are identical. Another solution is constraining the connector matrixes. The constrained method is discussed in the next subsection.

3.1.2.2 Algorithm

Here, we explain the algorithm of connector matrix canonical covariance analysis for three-mode three-way data. It is difficult to optimize the objective function (3.15) because it (3.15) is not convex for all parameters. Therefore, we adopt the update algorithm as an alternative least squares algorithm.

The update algorithm for $\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x$, and $\boldsymbol{C}_y$ is the same algorithm as that of the non-connector case. Thus, we need the partial derivative function of parameters.

Proposition 3.6. *We set g_c as the objective function as defined in equation* (3.15)*. The partial derivative function of g_c with respect to each parameter is obtained as follows:*

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{B}_x} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})\\ =&2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx}\otimes \boldsymbol{I})\Big((\boldsymbol{C}_x\boldsymbol{D}_{cx}\otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{bx}' - (\boldsymbol{C}_y\boldsymbol{D}_{cy}\otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{bx}'\Big)\\ &-2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x\end{aligned} \tag{3.17}$$

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{B}_y} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})\\ =&2\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy}\otimes \boldsymbol{I})\Big((\boldsymbol{C}_y\boldsymbol{D}_{cy}\otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{by}' - (\boldsymbol{C}_x\boldsymbol{D}_{cx}\otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{by}'\Big)\\ &-2\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y'\otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y\end{aligned} \tag{3.18}$$

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{C}_x} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})\\ =&2\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx}\otimes \boldsymbol{I})\Big((\boldsymbol{B}_x\boldsymbol{D}_{bx}\otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cx}' - (\boldsymbol{B}_y\boldsymbol{D}_{by}\otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cx}'\Big)\\ &-2\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x'\otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x\end{aligned} \tag{3.19}$$

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{C}_y} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})\\ =&2\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by}\otimes \boldsymbol{I})\Big((\boldsymbol{B}_y\boldsymbol{D}_{by}\otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cy}' - (\boldsymbol{B}_x\boldsymbol{D}_{bx}\otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cy}'\Big)\\ &-2\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y'\otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y\end{aligned} \tag{3.20}$$

Proof. From the definition of g_c and Propositions 2.1 and 2.3, we obtain the following equation:

$$\begin{aligned}&g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y \mid \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})\\ =&\|\boldsymbol{X}_1-\boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{B}_x\boldsymbol{B}_x')\|^2+\|\boldsymbol{Y}_1-\boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y'\otimes \boldsymbol{B}_y\boldsymbol{B}_y')\|^2\\ &+\|\boldsymbol{X}_1(\boldsymbol{C}_x\otimes \boldsymbol{B}_x)\boldsymbol{D}_x-\boldsymbol{Y}_1(\boldsymbol{C}_y\otimes \boldsymbol{B}_y)\boldsymbol{D}_y\|^2\\ =&\|\boldsymbol{X}_2-\boldsymbol{B}_x\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\|^2+\|\boldsymbol{Y}_2-\boldsymbol{B}_y\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y'\otimes \boldsymbol{I})\|^2\\ &+\|\boldsymbol{D}_{bx}'\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx}\otimes \boldsymbol{I})-\boldsymbol{D}_{by}'\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy}\otimes \boldsymbol{I})\|^2 \qquad (3.21)\\ =&\|\boldsymbol{X}_3-\boldsymbol{C}_x\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x'\otimes \boldsymbol{I})\|^2+\|\boldsymbol{Y}_3-\boldsymbol{C}_y\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y'\otimes \boldsymbol{I})\|^2\\ &+\|\boldsymbol{D}_{cx}'\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx}\otimes \boldsymbol{I})-\boldsymbol{D}_{cy}'\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by}\otimes \boldsymbol{I})\|^2. \qquad (3.22)\end{aligned}$$

Equations (3.21) and (3.22) are obtained by assuming that $\boldsymbol{D}_x=\boldsymbol{D}_{cx}\otimes\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_y=\boldsymbol{D}_{cy}\otimes\boldsymbol{D}_{by}$.

First, we derive the partial derivative function of $\boldsymbol{B}_x$. The first and second terms of equation (3.21) are the same as the first and second terms of equation (3.12). Thus, we can rewrite the first and second terms of equation (3.21) as follows:

$$
\begin{aligned}
&\|\boldsymbol{X}_2 - \boldsymbol{B}_x\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\|^2 + \|\boldsymbol{Y}_2 - \boldsymbol{B}_y\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{I})\|^2 \\
=&\mathrm{tr}(\boldsymbol{X}_2'\boldsymbol{X}_2) - \mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{I})(\boldsymbol{C}_x \otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x) \\
&+ \mathrm{tr}(\boldsymbol{Y}_2'\boldsymbol{Y}_2) - \mathrm{tr}(\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y \otimes \boldsymbol{I})(\boldsymbol{C}_y \otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y).
\end{aligned}
$$

The third term of equation (3.21) is rewritten as follows:

$$
\begin{aligned}
&\|\boldsymbol{D}_{bx}'\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I}) - \boldsymbol{D}_{by}'\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})\|^2 \\
=&\mathrm{tr}(\boldsymbol{D}_{bx}'\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}) \\
&- 2\mathrm{tr}(\boldsymbol{D}_{bx}'\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}) \\
&+ \mathrm{tr}(\boldsymbol{D}_{by}'\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}).
\end{aligned}
$$

Therefore, the partial derivative function of g_c with respect to $\boldsymbol{B}_x$ is obtained as follows:

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{B}_x} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}}) \\
=&\frac{\partial}{\partial \boldsymbol{B}_x}\{-\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{I})(\boldsymbol{C}_x \otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x) \\
&+ \mathrm{tr}(\boldsymbol{D}_{bx}'\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}) \\
&- 2\mathrm{tr}(\boldsymbol{D}_{bx}'\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by})\} \\
=&- 2\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{I})(\boldsymbol{C}_x \otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x \\
&+ 2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cx}'\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{bx}' \\
&- 2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cy}'\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{bx}'.
\end{aligned}
$$

In addition, we obtain the partial derivative function of g_c with respect to $\boldsymbol{B}_y$ as follows:

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{B}_y} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}}) \\
=&\frac{\partial}{\partial \boldsymbol{B}_y}\{-\mathrm{tr}(\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y \otimes \boldsymbol{I})(\boldsymbol{C}_y \otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y) \\
&+ \mathrm{tr}(\boldsymbol{D}_{by}'\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}) \\
&- 2\mathrm{tr}(\boldsymbol{D}_{by}'\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I})(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx})\} \\
=&- 2\boldsymbol{Y}_2(\boldsymbol{C}_y \otimes \boldsymbol{I})(\boldsymbol{C}_y \otimes \boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y \\
&+ 2\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cy}'\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{by}' \\
&- 2\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cx}'\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{by}'.
\end{aligned}
$$

Next, we derive the partial derivative function of $\boldsymbol{C}_x$. In the same way as the partial derivative function of $\boldsymbol{B}_x$ is calculated, equation (3.22) is rewritten as follows:

$$\begin{aligned}
&\|\boldsymbol{X}_3 - \boldsymbol{C}_x\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{I})\|^2 + \|\boldsymbol{Y}_3 - \boldsymbol{C}_y\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y' \otimes \boldsymbol{I})\|^2\\
&+ \|\boldsymbol{D}_{cx}'\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I}) - \boldsymbol{D}_{cy}'\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})\|^2\\
=&\mathrm{tr}(\boldsymbol{X}_3'\boldsymbol{X}_3) - \mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{I})(\boldsymbol{B}_x \otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x)\\
&+ \mathrm{tr}(\boldsymbol{Y}_3'\boldsymbol{Y}_3) - \mathrm{tr}(\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y \otimes \boldsymbol{I})(\boldsymbol{B}_y \otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y)\\
&+ \mathrm{tr}(\boldsymbol{D}_{cx}'\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx})\\
&- 2\mathrm{tr}(\boldsymbol{D}_{cx}'\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy})\\
&+ \mathrm{tr}(\boldsymbol{D}_{cy}'\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy}).
\end{aligned}$$

Therefore, we obtain the partial derivative function of g_c with respect to $\boldsymbol{C}_x$ as follows:

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{C}_x} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})\\
=&\frac{\partial}{\partial \boldsymbol{C}_x}\{-\mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{I})(\boldsymbol{B}_x \otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x)\\
&+ \mathrm{tr}(\boldsymbol{D}_{cx}'\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx})\\
&- 2\mathrm{tr}(\boldsymbol{D}_{cx}'\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy})\}\\
=&- 2\boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{I})(\boldsymbol{B}_x \otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x\\
&+ 2\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{bx}'\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cx}'\\
&- 2\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{by}'\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cx}'.
\end{aligned}$$

We also obtain the partial derivative function of g_c with respect to $\boldsymbol{C}_y$ as follows:

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{C}_y} g_c(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_x,\ \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})\\
=&\frac{\partial}{\partial \boldsymbol{C}_y}\{-\mathrm{tr}(\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y \otimes \boldsymbol{I})(\boldsymbol{B}_y \otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y)\\
&+ \mathrm{tr}(\boldsymbol{D}_{cy}'\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy})\\
&- 2\mathrm{tr}(\boldsymbol{D}_{cy}'\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by} \otimes \boldsymbol{I})(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx})\}\\
=&- 2\boldsymbol{Y}_3(\boldsymbol{B}_y \otimes \boldsymbol{I})(\boldsymbol{B}_y \otimes \boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\\
&+ 2\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{by}'\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cy}'\\
&- 2\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{bx}'\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cy}'.
\end{aligned}$$

□

The update formula for $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ is the same as that of the regression model, because the third terms of equations (3.21) and (3.22) are the same as those of the regression with given other parameters. Therefore, we obtain Proposition 3.7.

Proposition 3.7. *The update formulas of $\boldsymbol{D}_{bx},\ \boldsymbol{D}_{by},\ \boldsymbol{D}_{cx},\ \boldsymbol{D}_{cy}$ are obtained as follows:*

$$\boldsymbol{D}_{bx} = (\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cx}'\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x)^+\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cy}'\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}, \quad (3.23)$$

$$\boldsymbol{D}_{by} = (\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cy}'\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{B}_y)^+\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy}\boldsymbol{D}_{cx}'\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}', \quad (3.24)$$

$$\boldsymbol{D}_{cx} = (\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{bx}'\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x)^+\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{by}'\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy}, \quad (3.25)$$

$$\boldsymbol{D}_{cy} = (\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{by}'\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{C}_y)^+\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by}\boldsymbol{D}_{bx}'\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx}, \quad (3.26)$$

where $\boldsymbol{A}^+$ is Moore–Penrose pseudo inverse matrix $\boldsymbol{A}$.

Proof. For the third term of equation (3.21),

$$\begin{aligned}&\|\boldsymbol{D}_{bx}'\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx}\otimes\boldsymbol{I})-\boldsymbol{D}_{by}'\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy}\otimes\boldsymbol{I})\|^2\\=&\|(\boldsymbol{C}_x\boldsymbol{D}_{cx}\otimes\boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{B}_x\boldsymbol{D}_{bx}-(\boldsymbol{C}_y\boldsymbol{D}_{cy}\otimes\boldsymbol{I})'\boldsymbol{Y}_2'\boldsymbol{B}_y\boldsymbol{D}_{by}\|^2\end{aligned}$$

holds. The right-hand side of this equation is the same as that of the regression's objective function about $\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_{by}$. Therefore, we obtain the update formula for $\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_{by}$ with given other parameters.

The update formula for $\boldsymbol{D}_{cx}$ and $\boldsymbol{D}_{cy}$ is obtained in the same way as $\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_{by}$. For the third term of equation (3.21),

$$\begin{aligned}&\|\boldsymbol{D}_{cx}'\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx}\otimes\boldsymbol{I})-\boldsymbol{D}_{cy}'\boldsymbol{C}_y'\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{D}_{by}\otimes\boldsymbol{I})\|^2\\=&\|(\boldsymbol{B}_x\boldsymbol{D}_{bx}\otimes\boldsymbol{I})'\boldsymbol{X}_3'\boldsymbol{C}_x\boldsymbol{D}_{cx}-(\boldsymbol{B}_y\boldsymbol{D}_{by}\otimes\boldsymbol{I})'\boldsymbol{Y}_3'\boldsymbol{C}_y\boldsymbol{D}_{cy}\|^2\end{aligned}$$

holds. The right-hand side of this equation is the same as that of the regression's objective function about $\boldsymbol{D}_{cx}$ and $\boldsymbol{D}_{cy}$. Therefore, we obtain the update formula for $\boldsymbol{D}_{cx}$ and $\boldsymbol{D}_{cy}$ with given other parameters. □

Using Proposition 3.6 and 3.7, we obtain the update algorithm as algorithm 2. $\boldsymbol{U}_{bx}$, $\boldsymbol{U}_{by}$, $\boldsymbol{U}_{cx}$, $\boldsymbol{U}_{cy}$, $\boldsymbol{V}_{bx}$, $\boldsymbol{V}_{by}$, $\boldsymbol{V}_{cx}$, and $\boldsymbol{V}_{cy}$ in algorithm 2 are obtained by Jennrich's approach. Even if α is sufficiently big, the algorithm 2 yields only a monotonous decrease, not global optima. To obtain global optima, we generate many initial values of parameters.

3.2 Constrained connector matrix method

In this section, we introduce the constrained connector matrix method. The basic and connector matrix methods have rotational indeterminacy. One of the solutions for rotational indeterminacy is to constrain parameters. In this section, we explain the two types of constraints. The first one is K-means type, in which the elements of the connector matrix take 0 or 1. The second constraint type is regression, which corresponds to low rank regression for three-mode three-way data.

3.2.1 K-means based method

In this subsection, we explain the K-means type constrained case. One of the purposes of this constraint is to divide the factor into two types. The first is a common factor that shows the maximizing covariance. The second is an independent factor that shows the maximizing variance of each data.

3.2.1.1 Model and objective function

We set $\boldsymbol{D}_x=\boldsymbol{D}_{cx}\otimes\boldsymbol{D}_{bx}$, $\boldsymbol{D}_y=\boldsymbol{D}_{cy}\otimes\boldsymbol{D}_{by}$, and $\boldsymbol{D}_{bx}\in\{0,1\}^{r_{bx}\times c_b}$, $\boldsymbol{D}_{by}\in\{0,1\}^{r_{by}\times c_b}$, $\boldsymbol{D}_{cx}\in\{0,1\}^{r_{cx}\times c_c}$, $\boldsymbol{D}_{cy}\in\{0,1\}^{r_{cy}\times c_c}$. In this case, the estimation method of the connector matrixes is the same as that of K-means. The model of the K-means type constraint

Algorithm 2 Algorithm of connector matrix method

Set the number of dimensions $r_{bx},\ r_{by},\ r_{cx},\ r_{cy},\ c_b, c_c$, and stop condition ε

Set initial values $\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_y^{(0)},\ \boldsymbol{D}_{bx},\ \boldsymbol{D}_{by},\ \boldsymbol{D}_{cx},\ \boldsymbol{D}_{cy},\ \alpha$

$t \leftarrow 0$

$S^{(0)} \leftarrow g_c(\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$

repeat

  $t \leftarrow t+1$

  Update $\boldsymbol{D}_{bx}^{(t)}$ based on (3.23) using

  $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{by}^{(t-1)},\ \boldsymbol{D}_{cx}^{(t-1)},\ \boldsymbol{D}_{cy}^{(t-1)}$

  Update $\boldsymbol{D}_{by}^{(t)}$ based on (3.24) using

  $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{cx}^{(t-1)},\ \boldsymbol{D}_{cy}^{(t-1)}$

  Update $\boldsymbol{D}_{cx}^{(t)}$ based on (3.25) using

  $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{by}^{(t)},\ \boldsymbol{D}_{cy}^{(t-1)}$

  Update $\boldsymbol{D}_{cy}^{(t)}$ based on (3.26) using

  $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{by}^{(t)},\ \boldsymbol{D}_{cx}^{(t)}$

  $\boldsymbol{D}_x^{(t)} \leftarrow \boldsymbol{D}_{cx}^{(t)} \otimes \boldsymbol{D}_{bx}^{(t)}$

  $\boldsymbol{D}_y^{(t)} \leftarrow \boldsymbol{D}_{cy}^{(t)} \otimes \boldsymbol{D}_{by}^{(t)}$

  $\boldsymbol{B}_x^{(t)} \leftarrow \boldsymbol{U}_{bx}\boldsymbol{V}_{bx}'$ using $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

  $\boldsymbol{B}_y^{(t)} \leftarrow \boldsymbol{U}_{by}\boldsymbol{V}_{by}'$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

  $\boldsymbol{C}_x^{(t)} \leftarrow \boldsymbol{U}_{cx}\boldsymbol{V}_{cx}'$using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

  $\boldsymbol{C}_y^{(t)} \leftarrow \boldsymbol{U}_{cy}\boldsymbol{V}_{cy}'$using $\boldsymbol{B}_x^{(t)}, \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)}, \boldsymbol{C}_y^{(t-1)}\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

  $S^{(t)} \leftarrow g_c(\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$

until $|S^{(t-1)} - S^{(t)}| \leq \varepsilon$

method is the same as that of the connector matrix method. In other words, the model formula is as follows:

$$\boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{D}_y = \boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{D}_y + \boldsymbol{E}.$$

The objective function of this method is obtained as follows:

$$\begin{aligned} g_{ck}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) =& \|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{B}_x'\boldsymbol{B}_x')\|^2 \\ &+\|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{B}_y\boldsymbol{B}_y')\|^2 \\ &+\|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{D}_x - \boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{D}_y\|^2 \end{aligned} \tag{3.27}$$

$$\begin{aligned} \text{subject to } & \boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I},\ \boldsymbol{C}_y'\boldsymbol{C}_y = \boldsymbol{I}, \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I},\ \boldsymbol{B}_y'\boldsymbol{B}_y = \boldsymbol{I}, \\ & \boldsymbol{D}_{bx} \in \{0,\ 1\}^{r_{bx}\times c_b},\ \boldsymbol{D}_{by} \in \{0,\ 1\}^{r_{cx}\times c_b}, \\ & \boldsymbol{D}_{cx} \in \{0,\ 1\}^{r_{cx}\times c_c},\ \boldsymbol{D}_{cy} \in \{0,\ 1\}^{r_{cy}\times c_c}, \\ & \boldsymbol{D}_{bx}'\boldsymbol{1}_{r_{bx}} = \boldsymbol{D}_{by}'\boldsymbol{1}_{r_{by}} = \boldsymbol{1}_{c_b}, \boldsymbol{D}_{cx}'\boldsymbol{1}_{r_{cx}} = \boldsymbol{D}_{cy}'\boldsymbol{1}_{r_{cy}} = \boldsymbol{1}_{c_c}, \end{aligned}$$

where $\boldsymbol{1}_n$ is the n-dimension vector whose elements are all 1. The different points between the connector method and constrained method are the range of the connector matrix. The K-means type constraint regards the common factor as the object, and the factor as a cluster. Therefore, the cluster, which has at least one object, is the common factor. On the other hand, the null cluster, which has no object, indicates an independent factor. From objective function g_{ck} and the constraint of the connector, the independent factor considers only the first and second terms, because the weights for the third term are 0.

3.2.1.2 Algorithm

Here, we explain the algorithm of K-means type connected connector matrix canonical covariance analysis for three-mode three-way data. We also adopt an alternative least squares method. The constrained weight matrixes of objective function (3.27) are the same as those of (3.15). Thus, we obtain the update formula in the same way as for the non-constrained connector matrix method. On the other hand, update connector matrixes are different because the constraints of connector matrixes are different from objective function 3.27.

Connector matrixes depend on the third term of objective function (3.15), and therefore, we focus on this term. Given other parameters, the third term is the same as the objective function of K-means. Thus, the update formula of the connector matrix is the same of K-means. In other words,

$$d_{\ell q}^{(bx)} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I}_n)]_{\ell^*} - \boldsymbol{d}_q^{(by)\prime}\boldsymbol{B}_y'\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{D}_{cy} \otimes \boldsymbol{I}) \right\|\right) \\ 0 & \text{(otherwise)} \end{cases}$$

$$(q = 1,\ 2, \ldots, c_b), \tag{3.28}$$

$$d_{\ell q}^{(cx)} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{C}_x' \boldsymbol{X}_3 (\boldsymbol{B}_y \boldsymbol{D}_{bx} \otimes \boldsymbol{I}_n)]_{\ell^*} - \boldsymbol{d}_q^{(cy)\prime} \boldsymbol{C}_y' \boldsymbol{Y}_3 (\boldsymbol{B}_y \boldsymbol{D}_{by} \otimes \boldsymbol{I}) \right\| \right) \\ 0 & \text{(otherwise)} \end{cases}$$

$$(q = 1,\ 2, \ldots, c_b), \tag{3.29}$$

where $d_{\ell q}^{(cx)}$ and $d_{\ell q}^{(bx)}$ are the (ℓ, q) element of $\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_{cx}$, respectively, and $\boldsymbol{d}_q^{(by)}$ and $\boldsymbol{d}_q^{(by)}$ are the q-th column vector of $\boldsymbol{D}_{by}$ and $\boldsymbol{D}_{cy}$, respectively. $[\boldsymbol{A}]_{\ell^*}$ shows the ℓ^*-th column vector of $\boldsymbol{A}$. In the same way as for $\boldsymbol{D}_x$, the update $\boldsymbol{D}_y$ is obtained as follows:

$$d_{\ell q}^{(by)} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{B}_y \boldsymbol{Y}_2 (\boldsymbol{C}_y \boldsymbol{D}_{cy} \otimes \boldsymbol{I})]_{\ell^*} - \boldsymbol{d}_q^{(bx)\prime} \boldsymbol{B}_x' \boldsymbol{X}_2 (\boldsymbol{C}_x \boldsymbol{D}_{cx} \otimes \boldsymbol{I}) \right\| \right) \\ 0 & \text{(otherwise)} \end{cases}$$

$$(q = 1,\ 2, \ldots, c_c), \tag{3.30}$$

$$d_{\ell q}^{(cy)} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{C}_y' \boldsymbol{Y}_3 (\boldsymbol{B}_y \boldsymbol{D}_{by} \otimes \boldsymbol{I}_n)]_{\ell^*} - \boldsymbol{d}_q^{(cx)\prime} \boldsymbol{C}_x' \boldsymbol{X}_2 (\boldsymbol{B}_x \boldsymbol{D}_{bx} \otimes \boldsymbol{I}_n) \right\| \right) \\ 0 & \text{(otherwise)} \end{cases}$$

$$(q = 1,\ 2, \ldots, c_c), \tag{3.31}$$

where $d_{\ell q}^{(by)}$ and $d_{\ell q}^{(cy)}$ are the (ℓ, q) element of $\boldsymbol{D}_{by}$ and $\boldsymbol{D}_{cy}$, respectively, and $\boldsymbol{d}_q^{(bx)}$ and $\boldsymbol{d}_q^{(bx)}$ are the q-th column vector of $\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_{cx}$, respectively.

3.2.2 Spherical K-means based method

In this subsection, we describe spherical K-means constrained connector matrix method. The common factor of the K-means type constrained connector matrix method maximizes the covariance between two three-mode three-way data. On the other hand, spherical the K-means type constrained connector matrix method maximizes the correlation between two three-mode three-way data.

3.2.2.1 Model and objective function

The K-means type method has a problem when factor variances are different in each dimension. The K-means type constraint method tends to select the combination of factors with small correlation, because covariance is represented by the product of each variance and correlation. This property is a disadvantage when we interpret the common factor, which maximizes the correlation between two data. To overcome this problem, we introduce a normalized term, which makes the criteria of connecting correlation. In other

Algorithm 3 Algorithm of K-means constrained connector matrix method

Set the number of dimensions $r_{bx},\ r_{by},\ r_{cx},\ r_{cy},\ c_b, c_c$, and stop condition ε

Set initial values $\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_y^{(0)},\ \boldsymbol{D}_{bx},\ \boldsymbol{D}_{by},\ \boldsymbol{D}_{cx},\ \boldsymbol{D}_{cy},\ \alpha$

$t \leftarrow 0$

$S^{(0)} \leftarrow g_{ck}(\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$

repeat

$t \leftarrow t+1$

Update $\boldsymbol{D}_{bx}^{(t)}$ based on (3.28)

using $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{by}^{(t-1)},\ \boldsymbol{D}_{cx}^{(t-1)},\ \boldsymbol{D}_{cy}^{(t-1)}$

Update $\boldsymbol{D}_{by}^{(t)}$ based on (3.30)

using $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{cx}^{(t-1)},\ \boldsymbol{D}_{cy}^{(t-1)}$

Update $\boldsymbol{D}_{cx}^{(t)}$ based on (3.29)

using $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{by}^{(t)},\ \boldsymbol{D}_{cy}^{(t-1)}$

Update $\boldsymbol{D}_{cy}^{(t)}$ based on (3.31)

using $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{by}^{(t)},\ \boldsymbol{D}_{cx}^{(t)}$

$\boldsymbol{D}_x^{(t)} \leftarrow \boldsymbol{D}_{cx}^{(t)} \otimes \boldsymbol{D}_{bx}^{(t)}$

$\boldsymbol{D}_y^{(t)} \leftarrow \boldsymbol{D}_{cy}^{(t)} \otimes \boldsymbol{D}_{by}^{(t)}$

$\boldsymbol{B}_x^{(t)} \leftarrow \boldsymbol{U}_{bx}\boldsymbol{V}_{bx}'$ using $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$\boldsymbol{B}_y^{(t)} \leftarrow \boldsymbol{U}_{by}\boldsymbol{V}_{by}'$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$\boldsymbol{C}_x^{(t)} \leftarrow \boldsymbol{U}_{cx}\boldsymbol{V}_{cx}'$using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$\boldsymbol{C}_y^{(t)} \leftarrow \boldsymbol{U}_{cy}\boldsymbol{V}_{cy}'$using $\boldsymbol{B}_x^{(t)}, \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)}, \boldsymbol{C}_y^{(t-1)}\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$S^{(t)} \leftarrow g_{ck}(\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$

until $|S^{(t-1)} - S^{(t)}| \leq \varepsilon$

words, a constrained connector matrix is regarded as spherical K-means (Dhillo & Modha, 2001). The objective function g_{cs} is obtained as follows:

$$\begin{aligned}
&g_{cs}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) && (3.32)\\
=&\|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{B}_x'\boldsymbol{B}_x')\|^2\\
+&\|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{B}_y\boldsymbol{B}_y')\|^2\\
+&\|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{N}_x\boldsymbol{D}_x - \boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{N}_y\boldsymbol{D}_y\|^2 && (3.33)
\end{aligned}$$

$$\begin{aligned}
\text{subject to } &\boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I},\ \boldsymbol{C}_y'\boldsymbol{C}_y = \boldsymbol{I}, \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I},\ \boldsymbol{B}_y'\boldsymbol{B}_y = \boldsymbol{I},\\
&\boldsymbol{D}_{bx} \in \{0,\ 1\}^{r_{bx}\times c_b},\ \boldsymbol{D}_{by} \in \{0,\ 1\}^{r_{by}\times c_b},\\
&\boldsymbol{D}_{cx} \in \{0,\ 1\}^{r_{cx}\times c_c},\ \boldsymbol{D}_{cy} \in \{0,\ 1\}^{r_{cy}\times c_c},\\
&\boldsymbol{D}_{bx}'\boldsymbol{1}_{r_{bx}} = \boldsymbol{D}_{by}'\boldsymbol{1}_{r_{by}} = \boldsymbol{1}_{c_b}, \boldsymbol{D}_{cx}'\boldsymbol{1}_{r_{cx}} = \boldsymbol{D}_{cy}'\boldsymbol{1}_{r_{cy}} = \boldsymbol{1}_{c_c},
\end{aligned}$$

where $\boldsymbol{N}_x,\ \boldsymbol{N}_y$ are normalized matrixes. These matrixes are defined as follows:

$$\begin{aligned}
\boldsymbol{N}_x = \mathrm{diag}\Bigg(&\frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{x1} \otimes \boldsymbol{b}_{x1})\|}, \frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{x2} \otimes \boldsymbol{b}_{x1})\|}, \cdots \frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xr_{cx}} \otimes \boldsymbol{b}_{x1})\|}, \frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{x1} \otimes \boldsymbol{b}_{x2})\|}\\
&\cdots, \frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xi} \otimes \boldsymbol{b}_{xj})\|}, \cdots, \frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xr_{cx}} \otimes \boldsymbol{b}_{xr_{bx}})\|}\Bigg),\\
\boldsymbol{N}_y = \mathrm{diag}\Bigg(&\frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y1} \otimes \boldsymbol{b}_{y1})\|}, \frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y2} \otimes \boldsymbol{b}_{y1})\|}, \cdots \frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yr_{cy}} \otimes \boldsymbol{b}_{y1})\|}, \frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y1} \otimes \boldsymbol{b}_{y2})\|}\\
&\cdots, \frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yi} \otimes \boldsymbol{b}_{yj})\|}, \cdots, \frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yr_{cy}} \otimes \boldsymbol{b}_{yr_{bx}})\|}\Bigg),
\end{aligned}$$

where $\boldsymbol{b}_{xi},\ \boldsymbol{b}_{yi},\ \boldsymbol{c}_{xi}$, and $\boldsymbol{c}_{yi}$ are the i-th column vector of $\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x$, and $\boldsymbol{C}_y$, respectively.

3.2.2.2 Algorithm

When we use the spherical constrained method, the update formulas for weight matrixes $\boldsymbol{B}_x,\ \boldsymbol{B}_x,\ \boldsymbol{C}_x$, and $\boldsymbol{C}_y$ are a little different from those of the connector matrix method.

Proposition 3.8. *The partial derivative functions of $g_{cs}$with respect to $\boldsymbol{b}_{xi}$ and $\boldsymbol{b}_{yi}$ are* *obtained as follows:*

$$\frac{\partial}{\partial \boldsymbol{b}_{xi}} g_{cs}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}})$$

$$= -2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi}$$

$$-2\sum_{j=1}^{r_{by}}\sum_{k=1}^{r_{cx}}\sum_{\ell=1}^{r_{cy}}\sum_{o=1}^{c_b}\sum_{p=1}^{c_c}\Big\{d_{io}^{(bx)}d_{jo}^{(by)}d_{kp}^{(cx)}d_{\ell p}^{(cy)}\frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell}\otimes\boldsymbol{b}_{yj})\|\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes\boldsymbol{b}_{xi})\|^2}$$

$$\times\{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes\boldsymbol{b}_{xi})\|\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{y\ell}'\otimes\boldsymbol{I})'\boldsymbol{Y}_2\boldsymbol{b}_{yj} - (\boldsymbol{c}_{xk}\otimes\boldsymbol{b}_{xi})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell}\otimes\boldsymbol{b}_{yj})\frac{\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{xk}'\otimes\boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi}}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes\boldsymbol{b}_{xi})\|}\},$$

$$\frac{\partial}{\partial \boldsymbol{b}_{yi}} g_{cs}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\ \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}})$$

$$= -2\boldsymbol{Y}_2(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{b}_{yi}$$

$$-2\sum_{j=1}^{r_{bx}}\sum_{k=1}^{r_{cy}}\sum_{\ell=1}^{r_{cx}}\sum_{o=1}^{c_b}\sum_{p=1}^{c_c}\Big\{d_{io}^{(by)}d_{jo}^{(bx)}d_{kp}^{(cy)}d_{\ell p}^{(cx)}\frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{x\ell}\otimes\boldsymbol{b}_{xj})\|\|\boldsymbol{Y}_1(\boldsymbol{c}_{yk}\otimes\boldsymbol{b}_{yi})\|^2}$$

$$\times\{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yk}\otimes\boldsymbol{b}_{yi})\|\boldsymbol{Y}_2(\boldsymbol{c}_{yk}\boldsymbol{c}_{x\ell}'\otimes\boldsymbol{I})'\boldsymbol{X}_2\boldsymbol{b}_{xj} - (\boldsymbol{c}_{yk}\otimes\boldsymbol{b}_{yi})'\boldsymbol{Y}_1'\boldsymbol{X}_1(\boldsymbol{c}_{x\ell}\otimes\boldsymbol{b}_{xj})\frac{\boldsymbol{Y}_2(\boldsymbol{c}_{yk}\boldsymbol{c}_{yk}'\otimes\boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{b}_{yi}}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yk}\otimes\boldsymbol{b}_{yi})\|}\}.$$

The partial derivative functions of g_{cs} with respect to $\boldsymbol{c}_{xi}$ and $\boldsymbol{c}_{yi}$ are obtained as follows:

$$\frac{\partial}{\partial \boldsymbol{c}_{xi}} g_{cs}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y\ |\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}})$$

$$= -2\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{c}_{xi}$$

$$-2\sum_{j=1}^{r_{cy}}\sum_{k=1}^{r_{bx}}\sum_{\ell=1}^{r_{by}}\sum_{o=1}^{c_c}\sum_{p=1}^{c_b}\Big\{d_{io}^{(cx)}d_{jo}^{(cy)}d_{kp}^{(bx)}d_{\ell p}^{(by)}\frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yj}\otimes\boldsymbol{b}_{y\ell})\|\|\boldsymbol{X}_1(\boldsymbol{c}_{xi}\otimes\boldsymbol{b}_{xk})\|^2}$$

$$\times\{\|\boldsymbol{X}_1(\boldsymbol{c}_{xi}\otimes\boldsymbol{b}_{xk})\|\boldsymbol{X}_3(\boldsymbol{c}_{xk}\boldsymbol{c}_{yj}'\otimes\boldsymbol{I})'\boldsymbol{Y}_3\boldsymbol{c}_{yj} - (\boldsymbol{c}_{xi}\otimes\boldsymbol{b}_{xk})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{c}_{yj}\otimes\boldsymbol{b}_{y\ell})\frac{\boldsymbol{X}_3(\boldsymbol{b}_{xk}\boldsymbol{b}_{xk}'\otimes\boldsymbol{I})\boldsymbol{X}_3'\boldsymbol{c}_{xi}}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xi}\otimes\boldsymbol{b}_{xk})\|}\},$$

$$\frac{\partial}{\partial \boldsymbol{c}_{yi}} g_{cs}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y|\ \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}})$$

$$= -2\boldsymbol{Y}_3(\boldsymbol{B}_y\boldsymbol{B}_y' \otimes \boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{c}_{yi}$$

$$-2\sum_{j=1}^{r_{cx}}\sum_{k=1}^{r_{by}}\sum_{\ell=1}^{r_{bx}}\sum_{o=1}^{c_c}\sum_{p=1}^{c_b}\Big\{d_{io}^{(cy)}d_{jo}^{(cx)}d_{kp}^{(by)}d_{\ell p}^{(bx)}\frac{1}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xj}\otimes\boldsymbol{b}_{x\ell})\|\|\boldsymbol{Y}_1(\boldsymbol{c}_{yi}\otimes\boldsymbol{b}_{yk})\|^2}$$

$$\times\{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yi}\otimes\boldsymbol{b}_{yk})\|\boldsymbol{Y}_3(\boldsymbol{c}_{yk}\boldsymbol{c}_{xj}'\otimes\boldsymbol{I})'\boldsymbol{X}_3\boldsymbol{c}_{xj} - (\boldsymbol{c}_{yi}\otimes\boldsymbol{b}_{yk})'\boldsymbol{Y}_1'\boldsymbol{X}_1(\boldsymbol{c}_{xj}\otimes\boldsymbol{b}_{x\ell})\frac{\boldsymbol{Y}_3(\boldsymbol{b}_{yk}\boldsymbol{b}_{yk}'\otimes\boldsymbol{I})\boldsymbol{Y}_3'\boldsymbol{c}_{yi}}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yi}\otimes\boldsymbol{b}_{yk})\|}\}.$$

Proof. First, we calculate the derivative function of g_{cs} with respect to $\boldsymbol{b}_{xi}$. From the definition g_{cs}, we obtain the equation about $\boldsymbol{b}_{xi}$ as follows:

$$
\begin{aligned}
& g_{cs}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) \\
=& \|\boldsymbol{X}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{B}_x\boldsymbol{B}_x')\|^2 \\
& + \|\boldsymbol{Y}_1 - \boldsymbol{Y}_1(\boldsymbol{C}_y\boldsymbol{C}_y' \otimes \boldsymbol{B}_y\boldsymbol{B}_y')\|^2 \\
& + \|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{N_x}\boldsymbol{D}_x - \boldsymbol{Y}_1(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)\boldsymbol{N_y}\boldsymbol{D}_y\|^2 \\
=& -\operatorname{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x) \\
& + \operatorname{tr}(\boldsymbol{N}_x(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)\boldsymbol{N}_x\boldsymbol{D}_x\boldsymbol{D}_x') \\
& - 2\operatorname{tr}(\boldsymbol{N}_x(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{B}_y \otimes \boldsymbol{C}_y)\boldsymbol{N}_y\boldsymbol{D}_y\boldsymbol{D}_x') \\
& + \text{const.},
\end{aligned} \tag{3.34}
$$

where const. is constant independent from $\boldsymbol{b}_{xi}$. When the second term of equation (3.34) is described as element-wise, we obtain the following equation:

$$
\begin{aligned}
& \operatorname{tr}(\boldsymbol{N}_x(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)\boldsymbol{N}_x\boldsymbol{D}_x\boldsymbol{D}_x') \\
=& \sum_{i=1}^{r_{bx}}\sum_{j=1}^{r_{cx}}\sum_{k=1}^{c_b}\sum_{\ell=1}^{c_c} d_{ik}^{(bx)} d_{j\ell}^{(cx)} \frac{(\boldsymbol{b}_{xi} \otimes \boldsymbol{c}_{xj})'\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{b}_{xi} \otimes \boldsymbol{c}_{xj})}{\|\boldsymbol{X}_1(\boldsymbol{b}_{xi} \otimes \boldsymbol{c}_{xj})\|^2} \\
=& \sum_{i=1}^{r_{bx}}\sum_{j=1}^{r_{cx}}\sum_{k=1}^{c_b}\sum_{\ell=1}^{c_c} d_{ik}^{(bx)} d_{j\ell}^{(cx)} = c_c c_b.
\end{aligned}
$$

Therefore, the second term of equation (3.34) is constant. In the same way as we rewrite the second term, we can rewrite the third term of equation (3.34) as follows:

$$
\begin{aligned}
& -2\operatorname{tr}(\boldsymbol{N}_x(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{B}_y \otimes \boldsymbol{C}_y)\boldsymbol{N}_y\boldsymbol{D}_y\boldsymbol{D}_x') \\
=& -2\sum_{i=1}^{r_{bx}}\sum_{j=1}^{r_{by}}\sum_{k=1}^{r_{cx}}\sum_{\ell=1}^{r_{cy}}\sum_{o=1}^{b_c}\sum_{p=1}^{c_c} d_{io}^{(bx)} d_{jo}^{(by)} d_{kp}^{(cx)} d_{\ell p}^{(cy)} \frac{(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj})}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj})\|\|\boldsymbol{X}_1(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})\|}.
\end{aligned} \tag{3.35}
$$

Therefore, by the quotient rule, the derivative function g_{cs} with respect to $\boldsymbol{b}_{xi}$ for the third term of equation (3.34) is obtained as follows:

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{b}_{xi}} - 2\operatorname{tr}(\boldsymbol{N}_x(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{B}_y \otimes \boldsymbol{C}_y)\boldsymbol{N}_y\boldsymbol{D}_y\boldsymbol{D}_x') \\
=& -2\sum_{j=1}^{r_{by}}\sum_{k=1}^{r_{cx}}\sum_{\ell=1}^{r_{cy}}\sum_{o=1}^{c_b}\sum_{p=1}^{c_c} \Bigg\{ d_{io}^{(bx)} d_{jo}^{(by)} d_{kp}^{(cx)} d_{\ell p}^{(cy)} \frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj})\|} \\
& \times \frac{\partial}{\partial \boldsymbol{b}_{xi}} \frac{(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj})}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})\|} \Bigg\} \\
=& -2\sum_{j=1}^{r_{by}}\sum_{k=1}^{r_{cx}}\sum_{\ell=1}^{r_{cy}}\sum_{o=1}^{c_b}\sum_{p=1}^{c_c} \Bigg\{ d_{io}^{(bx)} d_{jo}^{(by)} d_{kp}^{(cx)} d_{\ell p}^{(cy)} \frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj})\|\|\boldsymbol{X}_1(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})\|^2} \\
& \{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})\|\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{y\ell}' \otimes \boldsymbol{I})'\boldsymbol{Y}_2\boldsymbol{b}_{yj} \\
& -(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj}) \frac{\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{xk}' \otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi}}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})\|}\} \Bigg\}
\end{aligned} \tag{3.36}
$$

where we use the following equations:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{b}_{xi}}\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes \boldsymbol{b}_{xi})\| =&\frac{\partial}{\partial \boldsymbol{b}_{xi}}(\|(\boldsymbol{c}_{xk}\otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{b}_{xi}\|^2)^{\frac{1}{2}}\\ =&\frac{\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{xk}'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi}}{\|(\boldsymbol{c}_{xk}\otimes \boldsymbol{I})'\boldsymbol{X}_2'\boldsymbol{b}_{xi}\|}\\ =&\frac{\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{xk}'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi}}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes \boldsymbol{b}_{xi})\|},\\ \frac{\partial}{\partial \boldsymbol{b}_{xi}}(\boldsymbol{c}_{xk}\otimes \boldsymbol{b}_{xi})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell}\otimes \boldsymbol{b}_{yj}) =&\frac{\partial}{\partial \boldsymbol{b}_{xi}}\mathrm{tr}(\boldsymbol{b}_{xi}'\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{y\ell}\otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{b}_{yj})\\ =&\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{y\ell}'\otimes \boldsymbol{I})\boldsymbol{Y}_2'\boldsymbol{b}_{yj}.\end{aligned}$$

On the other hand, the derivative function of the first term of equation (3.34) is obtained as follows:

$$-\frac{\partial}{\partial \boldsymbol{B}_x}\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x) = -2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x.$$

Therefore, the i-th column vector of $-2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x$ is the derivative function of the first term of equation (3.34) with respect to $\boldsymbol{b}_{xi}$. The i-th column of $-2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x$ is obtained as follows:

$$-2[\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{B}_x]_i = -2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi},$$

where $[\boldsymbol{A}]_i$ is the i-th column vector of $\boldsymbol{A}$. Thus, by summarizing the equations, we obtain the derivative function as follows:

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{b}_{xi}}g_{cs}(\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y, \boldsymbol{D}_x,\ \boldsymbol{D}_y|\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}})\\ =&-2\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x'\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi}\\ &-2\sum_{j=1}^{r_{by}}\sum_{k=1}^{r_{cx}}\sum_{\ell=1}^{r_{cy}}\sum_{o=1}^{c_b}\sum_{p=1}^{c_c}\Big\{d_{io}^{(bx)}d_{jo}^{(by)}d_{kp}^{(cx)}d_{\ell p}^{(cy)}\frac{1}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell}\otimes \boldsymbol{b}_{yj})\|\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes \boldsymbol{b}_{xi})\|^2}\\ &\{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes \boldsymbol{b}_{xi})\|\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{y\ell}\otimes \boldsymbol{I})'\boldsymbol{Y}_2\boldsymbol{b}_{yj}\\ &-(\boldsymbol{c}_{xk}\otimes \boldsymbol{b}_{xi})'\boldsymbol{X}_1'\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell}\otimes \boldsymbol{b}_{yj})\frac{\boldsymbol{X}_2(\boldsymbol{c}_{xk}\boldsymbol{c}_{xk}\otimes \boldsymbol{I})\boldsymbol{X}_2'\boldsymbol{b}_{xi}}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xk}\otimes \boldsymbol{b}_{xi})\|}\}.\end{aligned}$$

Other derivative functions are obtained in the same way as $\boldsymbol{b}_{xi}$. □

Updating the connector matrix of spherical K-means is the same as that of k-means. However, the update formulas of $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ do not minimize the squared loss between factors but maximize the correlation between factors.

Proposition 3.9. *The update formulas of $\boldsymbol{D}_{bx}$, $\boldsymbol{D}_{by}$, $\boldsymbol{D}_{cx}$, and $\boldsymbol{D}_{cx}$ are obtained as* *follows:*

$$d_{\ell q}^{(bx)} = \begin{cases} 1 & \left(\ell = \arg\max_{\ell^*} \sum_{j=1}^{r_{by}} \sum_{k=1}^{r_{cx}} \sum_{o=1}^{r_{cy}} \sum_{p=1}^{c_c} d_{jq}^{(by)} d_{kp}^{(cx)} d_{op}^{(cy)} \frac{(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{x\ell^*})' \boldsymbol{X}_1' \boldsymbol{Y}_1 (\boldsymbol{c}_{yo} \otimes \boldsymbol{b}_{yj})}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yo} \otimes \boldsymbol{b}_{yj})\| \|\boldsymbol{X}_1(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{x\ell^*})\|}\right) \\ 0 & (otherwise) \end{cases}$$

$$(q = 1,\ 2, \ldots, c_b), \tag{3.37}$$

$$d_{\ell q}^{(by)} = \begin{cases} 1 & \left(\ell = \arg\max_{\ell^*} \sum_{j=1}^{r_{bx}} \sum_{k=1}^{r_{cy}} \sum_{o=1}^{r_{cx}} \sum_{p=1}^{c_c} d_{jq}^{(by)} d_{kp}^{(cx)} d_{op}^{(cy)} \frac{(\boldsymbol{c}_{yk} \otimes \boldsymbol{b}_{y\ell^*})' \boldsymbol{Y}_1' \boldsymbol{X}_1 (\boldsymbol{c}_{xo} \otimes \boldsymbol{b}_{xj})}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xo} \otimes \boldsymbol{b}_{xj})\| \|\boldsymbol{Y}_1(\boldsymbol{c}_{yk} \otimes \boldsymbol{b}_{y\ell^*})\|}\right) \\ 0 & (otherwise) \end{cases}$$

$$(q = 1,\ 2, \ldots, c_b), \tag{3.38}$$

$$d_{\ell q}^{(cx)} = \begin{cases} 1 & \left(\ell = \arg\max_{\ell^*} \sum_{j=1}^{r_{cy}} \sum_{k=1}^{r_{bx}} \sum_{o=1}^{r_{by}} \sum_{p=1}^{c_b} d_{jq}^{(cy)} d_{kp}^{(bx)} d_{op}^{(by)} \frac{(\boldsymbol{c}_{x\ell^*} \otimes \boldsymbol{b}_{xk})' \boldsymbol{X}_1' \boldsymbol{Y}_1 (\boldsymbol{c}_{yj} \otimes \boldsymbol{b}_{yo})}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{yj} \otimes \boldsymbol{b}_{yo})\| \|\boldsymbol{X}_1(\boldsymbol{c}_{x\ell^*} \otimes \boldsymbol{b}_{xk})\|}\right) \\ 0 & (otherwise) \end{cases}$$

$$(q = 1,\ 2, \ldots, c_c), \tag{3.39}$$

$$d_{\ell q}^{(cy)} = \begin{cases} 1 & \left(\ell = \arg\max_{\ell^*} \sum_{j=1}^{r_{cx}} \sum_{k=1}^{r_{by}} \sum_{o=1}^{r_{bx}} \sum_{p=1}^{c_b} d_{jq}^{(cx)} d_{kp}^{(by)} d_{op}^{(bx)} \frac{(\boldsymbol{c}_{y\ell^*} \otimes \boldsymbol{b}_{yk})' \boldsymbol{Y}_1' \boldsymbol{X}_1 (\boldsymbol{c}_{xj} \otimes \boldsymbol{b}_{xo})}{\|\boldsymbol{X}_1(\boldsymbol{c}_{xj} \otimes \boldsymbol{b}_{xo})\| \|\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell^*} \otimes \boldsymbol{b}_{yk})\|}\right) \\ 0 & (otherwise) \end{cases}$$

$$(q = 1,\ 2, \ldots, c_c), \tag{3.40}$$

where $d_{\ell q}^{(bx)}$, $d_{\ell q}^{(by)}$, $d_{\ell q}^{(cx)}$ *and* $d_{\ell q}^{(cy)}$ *are the* (ℓ, q) *element of* $\boldsymbol{D}_{bx}$, $\boldsymbol{D}_{by}$, $\boldsymbol{D}_{cx}$, *and* $\boldsymbol{D}_{cy}$, *respectively.* $[\boldsymbol{A}]_\ell$ *is the* ℓ*-th column vector of* $\boldsymbol{A}$.

Proof. From equations (3.34) and (3.35), minimizing the objective function g_{cs} is equivalent to maximizing the objective function as follows:

$$\begin{aligned} &\mathrm{tr}(\boldsymbol{N}_x(\boldsymbol{B}_x \otimes \boldsymbol{C}_x)' \boldsymbol{X}_1' \boldsymbol{Y}_1 (\boldsymbol{B}_y \otimes \boldsymbol{C}_y) \boldsymbol{N}_y \boldsymbol{D}_y \boldsymbol{D}_x') \\ =& \sum_{i=1}^{r_{bx}} \sum_{j=1}^{r_{by}} \sum_{k=1}^{r_{cx}} \sum_{\ell=1}^{r_{cy}} \sum_{o=1}^{c_b} \sum_{p=1}^{c_c} d_{io}^{(bx)} d_{jo}^{(by)} d_{kp}^{(cx)} d_{\ell p}^{(cy)} \frac{(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})' \boldsymbol{X}_1' \boldsymbol{Y}_1 (\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj})}{\|\boldsymbol{Y}_1(\boldsymbol{c}_{y\ell} \otimes \boldsymbol{b}_{yj})\| \|\boldsymbol{X}_1(\boldsymbol{c}_{xk} \otimes \boldsymbol{b}_{xi})\|}. \end{aligned}$$

From the constraint of $\boldsymbol{D}_{bx}$, $\boldsymbol{D}_{by}$, $\boldsymbol{D}_{cx}$, and $\boldsymbol{D}_{cy}$, we choose the maximizing term corresponding to the dimension when other parameters are given. Thus, the element takes the value of 1 when the row dimension of the element is the maximum dimension of the correlation between the canonical vector in the same column. Therefore, we obtain the update rules as (3.37), (3.38), (3.39), and (3.40). □

We obtain Algorithm 4 by summarizing the update formula. Algorithm 4 ensure a monotonically decreasing. However, algorithm 4 does not yield the global optima of g_{cs}, because g_{cs} is not convex. Therefore, we need many initial values to obtain the global optima of g_{cs}.

3.2.3 Regression-based method

When we set $\boldsymbol{D}_y = \boldsymbol{I}, \boldsymbol{C}_y = \boldsymbol{I},\ \boldsymbol{B}_y = \boldsymbol{I}$, the model is the same as regression. However, $\boldsymbol{D}_x$ also have rotational indeterminacy. For this problem, we use tandem analysis. First,

Algorithm 4 Algorithm of spherical K-means constrained connector matrix method

Set the number of dimensions $r_{bx},\ r_{by},\ r_{cx},\ r_{cy},\ c_b, c_c$, and stop condition ε

Set initial values $\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_y^{(0)},\ \boldsymbol{D}_{bx},\ \boldsymbol{D}_{by},\ \boldsymbol{D}_{cx},\ \boldsymbol{D}_{cy},\ \alpha$

$t \leftarrow 0$

$S^{(0)} \leftarrow g_{cs}(\boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_y^{(0)},\ \boldsymbol{D}_x,\ \boldsymbol{D}_y \mid \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$

repeat

$t \leftarrow t + 1$

Update $\boldsymbol{D}_{bx}^{(t)}$ based on 3.37 using

$\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{by}^{(t-1)},\ \boldsymbol{D}_{cx}^{(t-1)},\ \boldsymbol{D}_{cy}^{(t-1)}$

Update $\boldsymbol{D}_{by}^{(t)}$ based on 3.38 using

$\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{cx}^{(t-1)},\ \boldsymbol{D}_{cy}^{(t-1)}$

Update $\boldsymbol{D}_{cx}^{(t)}$ based on 3.39 using

$\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{by}^{(t)},\ \boldsymbol{D}_{cy}^{(t-1)}$

Update $\boldsymbol{D}_{cy}^{(t)}$ based on 3.40 using

$\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_{by}^{(t)},\ \boldsymbol{D}_{cx}^{(t)}$

$\boldsymbol{D}_x^{(t)} \leftarrow \boldsymbol{D}_{cx}^{(t)} \otimes \boldsymbol{D}_{bx}^{(t)}$

$\boldsymbol{D}_y^{(t)} \leftarrow \boldsymbol{D}_{cy}^{(t)} \otimes \boldsymbol{D}_{by}^{(t)}$

$\boldsymbol{B}_x^{(t)} \leftarrow \boldsymbol{U}_{bx}\boldsymbol{V}_{bx}'$ using $\boldsymbol{B}_x^{(t-1)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$\boldsymbol{B}_y^{(t)} \leftarrow \boldsymbol{U}_{by}\boldsymbol{V}_{by}'$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t-1)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$\boldsymbol{C}_x^{(t)} \leftarrow \boldsymbol{U}_{cx}\boldsymbol{V}_{cx}'$using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t-1)},\ \boldsymbol{C}_y^{(t-1)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$\boldsymbol{C}_y^{(t)} \leftarrow \boldsymbol{U}_{cy}\boldsymbol{V}_{cy}'$using $\boldsymbol{B}_x^{(t)}, \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)}, \boldsymbol{C}_y^{(t-1)}\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)}, \alpha$

$S^{(t)} \leftarrow g_{cs}(\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)} | \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$

until $|S^{(t-1)} - S^{(t)}| \leq \varepsilon$

we apply canonical covariance analysis to three-mode three-way data. Then, we apply regression to three-mode three-way data.

3.2.3.1 Model and objective function

When we set $\boldsymbol{D}_y = \boldsymbol{I}, \boldsymbol{C}_y = \boldsymbol{I},\ \boldsymbol{B}_y = \boldsymbol{I}$, the model of the connector canonical covariance method is obtained as follows:

$$\boldsymbol{Y}_1 = \boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\boldsymbol{D}_x + \boldsymbol{E}. \tag{3.41}$$

This model formula is the same as the regression model, and thus, we call this method the regression constrained connector method. We set $\boldsymbol{W}_c = \boldsymbol{C}_x\boldsymbol{D}_{cx}$ and $\boldsymbol{W}_b = \boldsymbol{B}_x\boldsymbol{D}_{bx}$, then we rewrite the model formula (3.41) as

$$\begin{aligned} \boldsymbol{Y}_1 =& \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{B}_x\boldsymbol{D}_{bx}) + \boldsymbol{E} \\ =& \boldsymbol{X}_1(\boldsymbol{W}_c \otimes \boldsymbol{W}_b) + \boldsymbol{E} \\ =& \boldsymbol{X} \times_2 \boldsymbol{W}_b \times_3 \boldsymbol{W}_c + \boldsymbol{E}. \end{aligned} \tag{3.42}$$

Model formula (3.42) is the same as constrained regression, such as low-rank regression, such as the PLS method. When we set $\boldsymbol{F}_{x1} = \boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)$, we obtain another model formula as follows:

$$\begin{aligned} \boldsymbol{Y}_1 =& \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{B}_x\boldsymbol{D}_{bx}) + \boldsymbol{E} \\ =& \boldsymbol{F}_{x1}(\boldsymbol{D}_{cx} \otimes \boldsymbol{D}_{bx}) + \boldsymbol{E}. \end{aligned} \tag{3.43}$$

Model formula (3.43) is the same as principal component regression (PCR). The model formulas of partial least squares (PLS), low-rank regression, and PCR are the same. These methods restrict the weight matrixes from being column orthogonal. Using the least squares method as the estimation method, objective function g_r is obtained as follows:

$$g_r(\boldsymbol{B}_x,\ \boldsymbol{C}_x,\ \boldsymbol{D}_{bx},\ \boldsymbol{D}_{cx}|\underline{\boldsymbol{Y}},\ \underline{\boldsymbol{X}}) = \|\boldsymbol{Y}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{B}_x\boldsymbol{D}_{bx})\|^2 \tag{3.44}$$

$$\text{subject to } \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I},\ \boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I}.$$

This objective function has rotational indeterminacy.

Proposition 3.10. *Given orthogonal rotation $\boldsymbol{S}$ and $\boldsymbol{T}$, when we set $\boldsymbol{B}_x^* = \boldsymbol{B}_x\boldsymbol{S},\ \boldsymbol{C}_x^* = \boldsymbol{C}_x\boldsymbol{T},\ \boldsymbol{D}_{cx}^* = \boldsymbol{S}'\boldsymbol{D}_{bx},\ \boldsymbol{D}_{cx}^* = \boldsymbol{T}'\boldsymbol{D}_{cx}$,*

$$g_r(\boldsymbol{B}_x^*,\ \boldsymbol{C}_x^*,\ \boldsymbol{D}_{bx}^*,\ \boldsymbol{D}_{cx}^*|\underline{\boldsymbol{Y}},\ \underline{\boldsymbol{X}}) = g_r(\boldsymbol{B}_x,\ \boldsymbol{C}_x,\ \boldsymbol{D}_{bx},\ \boldsymbol{D}_{cx}|\underline{\boldsymbol{Y}},\ \underline{\boldsymbol{X}})$$

holds, where g_r is the objective function (3.44)*.*

Proof.

$$\begin{aligned} g_r(\boldsymbol{B}_x^*,\ \boldsymbol{C}_x^*,\ \boldsymbol{D}_{bx}^*,\ \boldsymbol{D}_{cx}^*|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}}) =& \|\boldsymbol{Y}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x^*\boldsymbol{D}_{cx}^* \otimes \boldsymbol{B}_x^*\boldsymbol{D}_{bx}^*)\|^2 \\ =& \|\boldsymbol{Y}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{T}\boldsymbol{T}'\boldsymbol{D}_{cx} \otimes \boldsymbol{B}_x\boldsymbol{S}\boldsymbol{S}'\boldsymbol{D}_{bx})\|^2 \\ =& \|\boldsymbol{Y}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{B}_x\boldsymbol{D}_{bx})\|^2 \\ =& g_r(\boldsymbol{B}_x,\ \boldsymbol{C}_x,\ \boldsymbol{D}_{bx},\ \boldsymbol{D}_{cx}|\underline{\boldsymbol{Y}},\ \underline{\boldsymbol{X}}) \end{aligned}$$

□

It is difficult to develop the criteria for the rotation because the simultaneous method has parameters with different purposes. One approach to the solution for rotation is tandem analysis, because the parameters for rotation are fixed by the two-step approach. Tandem analysis is a two-step approach. The first step is dimensional reduction of the data. The second step is multivariate analysis. In this case, the second step corresponds to regression analysis. One of the popular methods of tandem analysis for regression is principal component regression (PCR). The first step of PCR is to apply principal component analysis to three-mode three-way data. In other words, $\boldsymbol{C}_x$, and $\boldsymbol{B}_x$ are obtained by optimizing the objective function g_p as follows:

$$g_p(\boldsymbol{B}_x,\ \boldsymbol{C}_x|\underline{\boldsymbol{X}}) = \|\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\|^2 \longrightarrow \text{maximize}$$

$$\text{subject to } \boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I}, \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I}.$$

PCR does not consider the relationship between independent data $\underline{\boldsymbol{X}}$ and dependent data $\underline{\boldsymbol{Y}}$. To overcome this problem, we adopt the criteria of partial least squares regression (PLSR) for dimensional reduction method. The criteria of PLSR for dimensional reduction are to maximize covariance between data. The objective function of the regression constrained connector matrix of the first step is obtained as follows:

$$g_{pl}(\boldsymbol{B}_x,\ \boldsymbol{C}_x|\ \underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}}) = \|\boldsymbol{Y}_1'\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\|^2 \longrightarrow \text{maximize} \tag{3.45}$$

$$\text{subject to } \boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I}, \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I}.$$

Objective function (3.45) is regarded as maximizing the squared covariance between $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$. In this sense, PLSR is very similar to the canonical covariance method. The second step is regression using estimated values in first step.

3.2.3.2 Algorithm

Here, we explain the algorithm of the regression constrained connector matrix method. The regression constraint method is a type of tandem analysis. Therefore, the algorithm has two loops, that is, a dimensional reduction loop and a regression loop.

First, we explain the dimensional reduction loop. The update formula of $\boldsymbol{B}_x$ and $\boldsymbol{C}_x$ are obtained by eigenvalue decomposition

Proposition 3.11. *The update formula of $\boldsymbol{B}_x$ and $\boldsymbol{C}_x$ in the regression constrained connector matrix method are obtained as follows:*

$$\boldsymbol{B}_x = \boldsymbol{P}_{bx}, \tag{3.46}$$

$$\boldsymbol{C}_x = \boldsymbol{P}_{cx}, \tag{3.47}$$

where $\boldsymbol{P}_{bx}$ and $\boldsymbol{P}_{bx}$ are eigenvector matrixes of

$$\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{Y}_1)(\boldsymbol{C}_x \otimes \boldsymbol{Y}_1)'\boldsymbol{X}_2', \textit{ and } \boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{Y}_1)(\boldsymbol{B}_x \otimes \boldsymbol{Y}_1)'\boldsymbol{X}_3',$$

respectively.

Proof. From the definition of g_{pl}, we obtain the following equation:

$$\begin{aligned} g_{pl}(\boldsymbol{B}_x,\ \boldsymbol{C}_x|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}}) =&\|\boldsymbol{Y}_1'\boldsymbol{X}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)\|^2 \\ =&\|\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{Y}_1)\|^2 \qquad (3.48)\\ =&\|\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{Y}_1)\|^2. \qquad (3.49)\end{aligned}$$

From orthogonal constraint $\boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I}$ and equation (3.48), we obtain the update formula by using the Lagrange multiplier. The derivative function of g_{pl} with respect to $\boldsymbol{B}_x$ is obtained as follows:

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{B}_x}\{\|\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x \otimes \boldsymbol{Y}_1)\|^2 - \mathrm{tr}(\boldsymbol{\Lambda}(\boldsymbol{B}_x'\boldsymbol{B}_x - \boldsymbol{I}))\} \\ =&\frac{\partial}{\partial \boldsymbol{B}_x}\{\mathrm{tr}(\boldsymbol{B}_x'\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_2'\boldsymbol{B}_x) - \mathrm{tr}(\boldsymbol{\Lambda}(\boldsymbol{B}_x'\boldsymbol{B}_x - \boldsymbol{I}))\} \\ =&2(\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_2'\boldsymbol{B}_x - \boldsymbol{B}_x\boldsymbol{\Lambda}).\end{aligned}$$

When we set the value of the derivative function as 0, we obtain the following equation:

$$\begin{aligned} 2(\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_2'\boldsymbol{B}_x - \boldsymbol{B}_x\boldsymbol{\Lambda}) =&0 \\ \Longleftrightarrow \boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_2'\boldsymbol{B}_x =&\boldsymbol{B}_x\boldsymbol{\Lambda}.\end{aligned}$$

This equation indicates that $\boldsymbol{B}_x$ are an eigenvector of $\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{C}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_2$. We obtain the update formula of $\boldsymbol{C}_x$ in the same way as $\boldsymbol{B}_x$. From orthogonal constraint $\boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I}$ and equation (3.49), we obtain the update formula by using the Lagrange multiplier. The derivative function of g_{pl} with respect to $\boldsymbol{C}_x$ is obtained as follows:

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{C}_x}\{\|\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x \otimes \boldsymbol{Y}_1)\|^2 - \mathrm{tr}(\boldsymbol{\Lambda}(\boldsymbol{C}_x'\boldsymbol{C}_x - \boldsymbol{I}))\} \\ =&\frac{\partial}{\partial \boldsymbol{C}_x}\{\mathrm{tr}(\boldsymbol{C}_x'\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_3'\boldsymbol{C}_x) - \mathrm{tr}(\boldsymbol{\Lambda}(\boldsymbol{C}_x'\boldsymbol{C}_x - \boldsymbol{I}))\} \\ =&2(\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_3'\boldsymbol{C}_x - \boldsymbol{C}_x\boldsymbol{\Lambda}).\end{aligned}$$

When we set the value of the derivative function as 0, we obtain the following equation:

$$\begin{aligned} 2(\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_3'\boldsymbol{C}_x - \boldsymbol{C}_x\boldsymbol{\Lambda}) =&0 \\ \Longleftrightarrow \boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_3'\boldsymbol{C}_x =&\boldsymbol{C}_x\boldsymbol{\Lambda}. \qquad (3.50)\end{aligned}$$

Equation (3.50)indicates that $\boldsymbol{C}_x$ are an eigenvector of $\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{B}_x' \otimes \boldsymbol{Y}_1\boldsymbol{Y}_1')\boldsymbol{X}_3$. □

Given the estimated value of $\boldsymbol{C}_x$ and $\boldsymbol{B}_x$, we explain the regression step. This step is the same as the estimation step of the non-constrained connector matrix. Thus, we obtain the update formulas of $\boldsymbol{D}_{cx}$ and $\boldsymbol{D}_{bx}$ as an explicit function when other parameters are given.

Proposition 3.12. *Given $\hat{\boldsymbol{B}}_x$ and $\hat{\boldsymbol{C}}_x$, which are the estimated value of $\boldsymbol{B}_x$ and $\boldsymbol{C}_x$, respectively, the update formula of $\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_{cx}$ are obtained as follows:*

$$\boldsymbol{D}_{bx} = (\hat{\boldsymbol{B}}_x'\boldsymbol{X}_2(\hat{\boldsymbol{C}}_x\boldsymbol{D}_{cx}\boldsymbol{D}_{cx}'\hat{\boldsymbol{C}}_x' \otimes \boldsymbol{I})\boldsymbol{X}_2\hat{\boldsymbol{B}}_x)^+(\hat{\boldsymbol{B}}_x'\boldsymbol{X}_2(\hat{\boldsymbol{C}}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})\boldsymbol{Y}_2') \qquad (3.51)$$

$$\boldsymbol{D}_{cx} = (\hat{\boldsymbol{C}}_x'\boldsymbol{X}_3(\hat{\boldsymbol{B}}_x\boldsymbol{D}_{bx}\boldsymbol{D}_{bx}'\hat{\boldsymbol{B}}_x' \otimes \boldsymbol{I})\boldsymbol{X}_3\hat{\boldsymbol{C}}_x)^+(\hat{\boldsymbol{C}}_x'\boldsymbol{X}_3(\hat{\boldsymbol{B}}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})\boldsymbol{Y}_3') \qquad (3.52)$$

Proof. From the definition of objective function g_r,

$$g_r(\boldsymbol{B}_x,\ \boldsymbol{C}_x,\ \boldsymbol{D}_{bx}\boldsymbol{D}_{cx}|\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}}) = \|\boldsymbol{Y}_1 - \boldsymbol{X}_1(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{B}_x\boldsymbol{D}_{bx})\|^2 \tag{3.53}$$

$$= \|\boldsymbol{Y}_2 - \boldsymbol{D}'_{bx}\boldsymbol{B}'_x\boldsymbol{X}_2(\boldsymbol{C}_x\boldsymbol{D}_{cx} \otimes \boldsymbol{I})\|^2 \tag{3.54}$$

$$= \|\boldsymbol{Y}_3 - \boldsymbol{D}'_{bx}\boldsymbol{C}'_x\boldsymbol{X}_3(\boldsymbol{B}_x\boldsymbol{D}_{bx} \otimes \boldsymbol{I})\|^2 \tag{3.55}$$

holds. Equations (3.53), (3.54), and (3.55) are the same as Proposition 3.7 when we set $\boldsymbol{D}_{cy} = \boldsymbol{I},\ \boldsymbol{D}_{by} = \boldsymbol{I},\ \boldsymbol{C}_y = \boldsymbol{I},\ \boldsymbol{B}_y = \boldsymbol{I}$. Therefore, we obtain the update formulas of $\boldsymbol{D}_{bx}$ and $\boldsymbol{D}_{cx}$ by the special case of Proposition 3.7. □

Algorithm 5 Algorithm of regression constraint connector matrix method

Set the number of dimensions $r_{bx},\ r_{cx}$, and stop condition ε
Set initial values $\boldsymbol{B}_x^{(0)},\ \boldsymbol{C}_x^{(0)}$
$t \leftarrow 0$
$S^{(0)} \leftarrow g_{pl}(\boldsymbol{B}_x^{(0)},\ \boldsymbol{C}_x^{(0)}, |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$
repeat
  $t \leftarrow t + 1$
  Update $\boldsymbol{B}_x$ using $\boldsymbol{C}_x^{(t-1)}$
  Update $\boldsymbol{C}_x$ using $\boldsymbol{B}_x^{(t)}$
  $S^{(t)} \leftarrow g_{pl}(\boldsymbol{B}_x^{(t)},\ \boldsymbol{C}_x^{(t)}, |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$
until $|S^{(t-1)} - S^{(t)}| \leq \varepsilon$
Set initial values $\boldsymbol{D}_{bx}^{(0)},\ \boldsymbol{D}_{cx}^{(0)}$
$\hat{\boldsymbol{C}}_x \leftarrow \boldsymbol{C}_x^{(t)}$
$\hat{\boldsymbol{B}}_x \leftarrow \boldsymbol{B}_x^{(t)}$
$t \leftarrow 0$
$S^{(0)} \leftarrow g_r(\hat{\boldsymbol{B}}_x,\ \hat{\boldsymbol{C}}_x, \boldsymbol{D}_{bx}^{(0)}, \boldsymbol{D}_{cx}^{(0)}\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$
repeat
  $t \leftarrow t + 1$
  Update $\boldsymbol{D}_{bx}$ using $\boldsymbol{D}_{cx}^{(t-1)}$
  Update $\boldsymbol{D}_{cx}$ using $\boldsymbol{D}_{bx}^{(t)}$
  $S^{(t)} \leftarrow g_r(\hat{\boldsymbol{B}}_x,\ \hat{\boldsymbol{C}}_x, \boldsymbol{D}_{bx}^{(t)}, \boldsymbol{D}_{cx}^{(t)}\ |\underline{\boldsymbol{X}},\ \underline{\boldsymbol{Y}})$
until $|S^{(t-1)} - S^{(t)}| \leq \varepsilon$

3.3 Category quantification method

In this section, we explain three-mode three-way canonical covariance analysis for categorical data. First, we introduce the method based on non-metric principal component analysis (NPCA) (e.g., Gifi (1990); Young et al. (1978)).

3.3.1 NPCA-based method

NPCA has been proposed for categorical data by Young et al. (1978). This method is a special case of Hayashi's quantification method. One of its advantages is that it is easy to interpret the qualification parameter.

3.3.1.1 Model and objective function

In the three-mode three-way canonical covariance method, it is assumed that the data are numerical. When the data are categorical data, we should not apply the three-mode three-way canonical covariance method to three-mode three-way data, because there is arbitrariness in the values of qualitative data. For example, for a gender variable, we can choose 1 or 0 to reflect male or female. However, the results of the three-mode three-way canonical covariance method are different from this case. Moreover, the variance of variable and covariance between qualitative data is changed when we change 0 to 1 and 1 to 0. To overcome these problems, we extend three-mode three-way canonical covariance to categorical data by using the concept of NPCA. The difference between NPCA and using the dummies variable is that NPCA adjusts the variance of the whole of the qualitative variable to 1. This property of NPCA makes interpretation easy because we can understand the weight matrix for the qualitative variable, rather than each item of the qualitative variable.

Given two three-mode three-way categorical data $\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}$, the objective function of the NPCA-based method is obtained as follows:

$$
\begin{aligned}
g_{npc}(\underline{\boldsymbol{F}}_x,\ \underline{\boldsymbol{F}}_y,\ \boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{Q},\ \boldsymbol{W},\ |\underline{\boldsymbol{X}}^\dagger,\ \underline{\boldsymbol{Y}}^\dagger) =& \|\boldsymbol{X}_1^\dagger \boldsymbol{Q} - \boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\|^2 \\
&+ \|\boldsymbol{Y}_1^\dagger \boldsymbol{W} - \boldsymbol{F}_1^{(y)}(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)'\|^2 \\
&+ \|\boldsymbol{F}_x - \boldsymbol{F}_y\|^2 \qquad (3.56)
\end{aligned}
$$

$$
\begin{aligned}
\text{subject to} \quad & \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{B}_y'\boldsymbol{B}_y = \boldsymbol{I},\ \boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{C}_y'\boldsymbol{C}_y = \boldsymbol{I}, \\
& \boldsymbol{X}_{k_xj_x}^\dagger \boldsymbol{q}_{k_xj_x} = \boldsymbol{J}\boldsymbol{X}_{k_xj_x}^\dagger \boldsymbol{q}_{k_xj_x},\ \frac{1}{I}\boldsymbol{q}_{k_xj_x}'\boldsymbol{X}_{k_xj_x}^{\dagger\,\prime}\boldsymbol{X}_{k_xj_x}^\dagger \boldsymbol{q}_{k_xj_x} = 1 \\
& \quad (j_x = 1,\ 2,\ \dots J_x; k_x = 1,\ 2,\ \dots,\ K_x), \\
& \boldsymbol{Y}_{k_yj_y}^\dagger \boldsymbol{w}_{k_yj_y} = \boldsymbol{J}\boldsymbol{Y}_{k_yj_y}^\dagger \boldsymbol{w}_{k_yj_y},\ \text{and}\ \frac{1}{I}\boldsymbol{w}_{k_yj_y}'\boldsymbol{Y}_{k_yj_y}^{\dagger\,\prime}\boldsymbol{Y}_{k_yj_y}^\dagger \boldsymbol{w}_{k_yj_y} = 1 \\
& \quad (j_y = 1,\ 2,\ \dots,\ J_y; k_y = 1,\ 2,\ \dots,\ K_y),
\end{aligned}
$$

where $\underline{\boldsymbol{X}}^\dagger,\ \underline{\boldsymbol{Y}}^\dagger$ are made by changing the categorical variable to a dummy variable. $\boldsymbol{X}_{j_x,\,k_x}^\dagger$ and $\boldsymbol{Y}_{j_y,\,k_y}^\dagger$ are the dummy matrix of j_x-th categorical variable under condition k_x of $\underline{\boldsymbol{X}}$ and j_y-th categorical variable under condition k_y of $\underline{\boldsymbol{Y}}$, respectively. $\boldsymbol{J}$ is a centering matrix, and $\boldsymbol{q}_{k_xj_x}$ and $\boldsymbol{w}_{k_yj_y}$ are a qualification vector of j_x-th categorical variable under condition k_x of $\underline{\boldsymbol{X}}$ and j_y-th categorical variable under condition k_y of $\underline{\boldsymbol{Y}}$, respectively. $\boldsymbol{Q} = \text{B-diag}(\boldsymbol{q}_{11},\ \boldsymbol{q}_{12},\cdots,\ \boldsymbol{q}_{1J_x},\ \boldsymbol{q}_{21},\ \boldsymbol{q}_{22},\cdots,\ \boldsymbol{q}_{k_xj_x},\cdots,\ \boldsymbol{q}_{K_x1},\ \boldsymbol{q}_{K_x2},\ \cdots,\ \boldsymbol{q}_{K_xJ_x})$, $\boldsymbol{W} = \text{B-diag}(\boldsymbol{w}_{11},\ \boldsymbol{w}_{12},\cdots,\ \boldsymbol{w}_{1J_y},\ \boldsymbol{w}_{21},\ \boldsymbol{w}_{22},\cdots,\ \boldsymbol{w}_{k_yj_y},\cdots,\ \boldsymbol{w}_{K_y1},\ \boldsymbol{w}_{K_y2},\ \cdots,\ \boldsymbol{w}_{K_yJ_y})$. B-diag($\boldsymbol{A}$) is a block diagonal matrix defined as follows:

$$
\text{B-diag}(\boldsymbol{A}) = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{a}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{a}_p \end{pmatrix},
$$

where $\boldsymbol{A} = (\boldsymbol{a}_1,\ \boldsymbol{a}_2, \cdots,\ \boldsymbol{a}_p)$. The constraint of the qualification vector corresponds to z-standardized variable. Therefore, NPCA is regarded as changing the categorical variable to z-standardized variable.

However, this method has two problems. First, we must set the number of dimensions of $\boldsymbol{B}_x$, $\boldsymbol{B}_y$, and $\boldsymbol{C}_x$, $\boldsymbol{C}_y$ to be the same. In other words, we must assume that the number of unique factors is the same, an assumption that is not suitable for real-world data analysis. The second problem is that the third term of objective function (3.56) considers all the factors; that is, it is difficult to determine which are the common factors. To overcome this problem, we introduce the K-means type connector matrix.

$$\begin{aligned}
&g_{ccca}(\underline{\boldsymbol{F}}_x,\ \underline{\boldsymbol{F}}_y,\ \boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,\ \boldsymbol{C}_y,\ \boldsymbol{Q},\ \boldsymbol{W},\ \boldsymbol{D}_x,\ \boldsymbol{D}_y \mid \underline{\boldsymbol{X}}^{\dagger},\ \underline{\boldsymbol{Y}}^{\dagger}) \\
=&\|\boldsymbol{X}^{\dagger}\boldsymbol{Q} - \boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\|^2 + \|\boldsymbol{Y}^{\dagger}\boldsymbol{W} - \boldsymbol{F}^{(y)}(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)'\|^2 \\
+&\|\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x - \boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\|^2,
\end{aligned} \tag{3.57}$$

$$\begin{aligned}
\text{subject to}\quad & \boldsymbol{B}_x'\boldsymbol{B}_x = \boldsymbol{I},\ \boldsymbol{B}_y'\boldsymbol{B}_y = \boldsymbol{I},\ \boldsymbol{C}_x'\boldsymbol{C}_x = \boldsymbol{I},\ \boldsymbol{C}_y'\boldsymbol{C}_y = \boldsymbol{I}, \\
& \boldsymbol{X}^{\dagger}_{k_xj_x}\boldsymbol{q}_{k_xj_x} = \boldsymbol{J}_I\boldsymbol{X}^{\dagger}_{k_xj_x}\boldsymbol{q}_{k_xj_x},\ \frac{1}{I}\boldsymbol{q}'_{k_xj_x}\boldsymbol{X}^{\dagger\prime}_{k_xj_x}\boldsymbol{X}^{\dagger}_{k_xj_x}\boldsymbol{q}_{k_xj_x} = 1 \\
& \quad (j_x = 1,\ 2,\ \ldots,\ J_x; k_x = 1,\ 2,\ \ldots,\ K_x), \\
& \boldsymbol{Y}^{\dagger}_{k_yj_y}\boldsymbol{w}_{k_yj_y} = \boldsymbol{J}_I\boldsymbol{Y}^{\dagger}_{k_yj_y}\boldsymbol{w}_{k_yj_y},\ \frac{1}{I}\boldsymbol{w}'_{k_yj_y}\boldsymbol{Y}^{\dagger\prime}_{k_yj_y}\boldsymbol{Y}^{\dagger}_{k_yj_y}\boldsymbol{w}_{k_yj_y} = 1 \\
& \quad (j_y = 1,\ 2,\ \ldots,\ J_y; k_y = 1,\ 2,\ \ldots,\ K_y), \\
& \boldsymbol{D}_x = \boldsymbol{D}_{cx} \otimes \boldsymbol{D}_{bx},\ \boldsymbol{D}_y = \boldsymbol{D}_{cy} \otimes \boldsymbol{D}_{by}, \\
& \boldsymbol{D}_{bx} \in \{0,\ 1\}^{r_{bx}\times c_b},\ \boldsymbol{D}_{by} \in \{0,\ 1\}^{r_{by}\times c_b}, \\
& \boldsymbol{D}_{cx} \in \{0,\ 1\}^{r_{cx}\times c_c},\ \boldsymbol{D}_{cy} \in \{0,\ 1\}^{r_{cy}\times c_c}, \\
& \boldsymbol{D}_{bx}'\boldsymbol{1}_{r_{bx}} = \boldsymbol{D}_{by}'\boldsymbol{1}_{r_{by}} = \boldsymbol{1}_{c_b}, \boldsymbol{D}_{cx}'\boldsymbol{1}_{r_{cx}} = \boldsymbol{D}_{cy}'\boldsymbol{1}_{r_{cy}} = \boldsymbol{1}_{c_c}.
\end{aligned}$$

When we set $r_{bx} = r_{by},\ r_{cx} = r_{cy},\ c_b = r_{bx}, c_c = r_{cx},\ \boldsymbol{D}_x = \boldsymbol{I}$, and $\boldsymbol{D}_y = \boldsymbol{I}$, the objective function g_{ccca} is equal to the objective function g_{npc}.

3.3.1.2 Algorithm

Here, we explain the algorithm of the NPCA-based method using the K-means type constraint connector matrix. When we set $r_{bx} = r_{by},\ r_{cx} = r_{cy},\ c_b = r_{bx}, c_c = r_{cx},\ \boldsymbol{D}_x = \boldsymbol{I}$, and $\boldsymbol{D}_y = \boldsymbol{I}$, objective function g_{ccca} equals objective function g_{npc}. Therefore, the NPCA-based method is a special case of using the K-means type constrained connector matrix method.

Proposition 3.13. *When we set* $\underline{\boldsymbol{Z}}^{(x)} = [\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}]^{(1)}_{J_x,K_x}$ *and* $\underline{\boldsymbol{Z}}^{(y)} = [\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}]^{(1)}_{J_y,K_y}$. *The update* *formula of* $\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,$ *and* $\boldsymbol{C}_y$ *is obtained as follows:*

$$\boldsymbol{B}_x = \boldsymbol{U}_{bx}\boldsymbol{V}'_{bx} \tag{3.58}$$

$$\boldsymbol{B}_y = \boldsymbol{U}_{by}\boldsymbol{V}'_{by} \tag{3.59}$$

$$\boldsymbol{C}_x = \boldsymbol{U}_{cx}\boldsymbol{V}'_{cx} \tag{3.60}$$

$$\boldsymbol{C}_y = \boldsymbol{U}_{cy}\boldsymbol{V}'_{cy} \tag{3.61}$$

where $\boldsymbol{U}_{bx}$ *and* $\boldsymbol{V}_{bx}$ *are left and right singular matrixes of*

$$\boldsymbol{Z}_2^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})'\boldsymbol{F}_2^{(x)\prime}.$$

$\boldsymbol{U}_{by}$ *and* $\boldsymbol{V}_{by}$ *are left and right singular matrixes of*

$$\boldsymbol{Z}_2^{(y)}(\boldsymbol{C}_y \otimes \boldsymbol{I})\boldsymbol{F}_2^{(y)\prime}.$$

$\boldsymbol{U}_{cx}$ *and* $\boldsymbol{V}_{cx}$ *are left and right singular matrixes of*

$$\boldsymbol{Z}_3^{(x)}(\boldsymbol{B}_x \otimes \boldsymbol{I})\boldsymbol{F}_3^{(x)\prime}.$$

$\boldsymbol{U}_{cy}$ *and* $\boldsymbol{V}_{cy}$ *are left and right singular matrixes of*

$$\boldsymbol{Z}_2^{(y)}(\boldsymbol{B}_y \otimes \boldsymbol{I})\boldsymbol{F}_3^{(y)\prime}.$$

Proof. First, we explain the update formula of $\boldsymbol{B}_x$. The term related to $\boldsymbol{B}_x$ is the first term. The first term of objective function g_{ccca} is rewritten as follows:

$$\begin{aligned}
\|\boldsymbol{X}_1\boldsymbol{Q} - \boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\|^2 =& \|\boldsymbol{Z}_1^{(x)} - \boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\|^2 \\
=& \|\boldsymbol{Z}_2^{(x)} - \boldsymbol{B}_x\boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})'\|^2 \\
=& \mathrm{tr}(\boldsymbol{Z}_2^{(x)\prime}\boldsymbol{Z}_2^{(x)}) - 2\mathrm{tr}(\boldsymbol{Z}_2^{(x)\prime}\boldsymbol{B}_x\boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})') \\
&+ \mathrm{tr}((\boldsymbol{C}_x \otimes \boldsymbol{I})\boldsymbol{F}_2^{(x)}\boldsymbol{B}'_x\boldsymbol{B}_x\boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})') \\
=& \mathrm{tr}(\boldsymbol{Z}_2^{(x)\prime}\boldsymbol{Z}_2^{(x)}) - 2\mathrm{tr}(\boldsymbol{B}_x\boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})'\boldsymbol{Z}_2^{(x)\prime}) \\
&+ \mathrm{tr}((\boldsymbol{C}_x \otimes \boldsymbol{I})\boldsymbol{F}_2^{(x)\prime}\boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})').
\end{aligned}$$

Given parameters except $\boldsymbol{B}_x$, we obtain the update formula of $\boldsymbol{B}_x$ by maximizing $\mathrm{tr}(\boldsymbol{B}_x\boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{I})'\boldsymbol{Z}_2^{(x)\prime})$. From the TenBerge theorem (ten Berge, 1993), $\mathrm{tr}(\boldsymbol{B}_x\boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x\otimes \boldsymbol{I})'\boldsymbol{Z}_2^{(x)\prime}) \leq \mathrm{tr}(\boldsymbol{D})$holds. $\boldsymbol{D}$ is the diagonal matrix whose elements are singular values of $\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}' = \boldsymbol{F}_2^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})'\boldsymbol{Z}'_2$. When we set $\boldsymbol{B}_x = \boldsymbol{U}\boldsymbol{V}'$, the equation holds. Therefore, we obtain the update formula. The update formula of $\boldsymbol{B}_y$ is obtained in the same way as $\boldsymbol{B}_x$.

Next, we explain the update formula of $\boldsymbol{C}_x$, which is obtained in a very similar way to $\boldsymbol{B}_x$. The term related to $\boldsymbol{C}_x$ is the first term. The first term of objective function g_{ccca} is rewritten as follows:

$$\begin{aligned}
\|\boldsymbol{X}_1\boldsymbol{Q} - \boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\|^2 =& \|\boldsymbol{Z}_1^{(x)} - \boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\|^2 \\
=& \|\boldsymbol{Z}_3^{(x)} - \boldsymbol{C}_x\boldsymbol{F}_3^{(x)}(\boldsymbol{B}_x \otimes \boldsymbol{I})'\|^2 \\
=& \mathrm{tr}(\boldsymbol{Z}_3^{(x)\prime}\boldsymbol{Z}_3^{(x)}) - 2\mathrm{tr}(\boldsymbol{Z}_3^{(x)\prime}\boldsymbol{C}_x\boldsymbol{F}_3^{(x)}(\boldsymbol{B}_x \otimes \boldsymbol{I})') \\
&+ \mathrm{tr}((\boldsymbol{B}_x \otimes \boldsymbol{I})\boldsymbol{F}_3^{(x)}\boldsymbol{C}'_x\boldsymbol{C}_x\boldsymbol{F}_3^{(x)}(\boldsymbol{B}_x \otimes \boldsymbol{I})') \\
=& \mathrm{tr}(\boldsymbol{Z}_3^{(x)\prime}\boldsymbol{Z}_3^{(x)}) - 2\mathrm{tr}(\boldsymbol{B}_x\boldsymbol{F}_3^{(x)}(\boldsymbol{C}_x \otimes \boldsymbol{I})'\boldsymbol{Z}_3^{(x)\prime}) \\
&+ \mathrm{tr}((\boldsymbol{B}_x \otimes \boldsymbol{I})\boldsymbol{F}_3^{(x)\prime}\boldsymbol{F}_3^{(x)}(\boldsymbol{B}_x \otimes \boldsymbol{I})').
\end{aligned}$$

Given another parameter except $\boldsymbol{C}_x$, we obtain the update formula of $\boldsymbol{C}_x$ by maximizing $\mathrm{tr}(\boldsymbol{C}_x\boldsymbol{F}_3^{(x)}(\boldsymbol{B}_x\otimes\boldsymbol{I})'\boldsymbol{Z}_3^{(x)\prime})$. From the TenBerge theorem, $\mathrm{tr}(\boldsymbol{C}_x\boldsymbol{F}_3^{(x)}(\boldsymbol{B}_x\otimes\boldsymbol{I})'\boldsymbol{Z}_3^{(x)\prime}) \leq \mathrm{tr}(\boldsymbol{D})$ holds. $\boldsymbol{D}$ is a diagonal matrix whose elements are singular values of $\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}' = \boldsymbol{F}_3^{(x)}(\boldsymbol{B}_x\otimes \boldsymbol{I})'\boldsymbol{Z}_3'$. When we set $\boldsymbol{C}_x = \boldsymbol{U}\boldsymbol{V}'$, the equation holds. Therefore, we obtain the update formula. The update formula of $\boldsymbol{C}_y$ is obtained in the same way as $\boldsymbol{C}_x$. □

Proposition 3.14. *The update formulas of* $\underline{\boldsymbol{F}}^{(x)}$ *and* $\underline{\boldsymbol{F}}^{(y)}$ *are obtained as follows:*

$$\boldsymbol{F}_1^{(x)} =(\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)+\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_x')(\boldsymbol{I}+\boldsymbol{D}_x\boldsymbol{D}_x')^{-1}, \tag{3.62}$$

$$\boldsymbol{F}_1^{(y)} =(\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)+\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_y')(\boldsymbol{I}+\boldsymbol{D}_y\boldsymbol{D}_y')^{-1}. \tag{3.63}$$

Proof. First, we explain about the update formula of $\boldsymbol{F}_1^{(x)}$.

$$\begin{aligned}
&\|\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}-\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)'\|^2+\|\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x-\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\|^2\\
=&-2\mathrm{tr}(\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)'\boldsymbol{Q}'\boldsymbol{X}_1^{\dagger\prime})-2\mathrm{tr}(\boldsymbol{F}_1^{(x)\prime}\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_x')\\
&+\mathrm{tr}(\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_x'\boldsymbol{F}_1^{(x)\prime})+\mathrm{tr}(\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x'\boldsymbol{C}_x\otimes\boldsymbol{B}_x'\boldsymbol{B}_x)\boldsymbol{F}_1^{(x)})+\mathrm{const.},
\end{aligned}$$

where const. is constant independence from $\boldsymbol{F}_1^{(x)}$. Thus, the partial derivative function of g_{ccca} with respect to $\boldsymbol{F}_1^{(x)}$ is obtained as follows:

$$\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{F}_1^{(x)}}\{\|\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}-\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)'\|^2+\|\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x-\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\|^2\}\\
=&-2\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)-2\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_x'+2\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_x'+2\boldsymbol{F}_1^{(x)}.
\end{aligned}$$

When we set the partial derivative function of g_{ccca} with respect to $\boldsymbol{F}_1^{(x)}$ as 0, we obtain the following equation:

$$\begin{aligned}
&-2\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)-2\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_x'+2\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_x'+2\boldsymbol{F}_1^{(x)}=0\\
\Longleftrightarrow&\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_x'+\boldsymbol{F}_1^{(x)}=\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)+\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_x'\\
\Longleftrightarrow&\boldsymbol{F}_1^{(x)}=(\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)+\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_x')(\boldsymbol{I}+\boldsymbol{D}_x\boldsymbol{D}_x')^{-1}.
\end{aligned}$$

The update formula of $\boldsymbol{F}_1^{(y)}$ is obtained in the same way as $\boldsymbol{F}_1^{(x)}$.

$$\begin{aligned}
&\|\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}-\boldsymbol{F}_1^{(y)}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)'\|^2+\|\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x-\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\|^2\\
=&-2\mathrm{tr}(\boldsymbol{F}_1^{(y)}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)'\boldsymbol{W}'\boldsymbol{Y}_1^{\dagger\prime})-2\mathrm{tr}(\boldsymbol{F}_1^{(y)\prime}\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_y')\\
&+\mathrm{tr}(\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_y'\boldsymbol{F}_1^{(y)\prime})+\mathrm{tr}(\boldsymbol{F}_1^{(y)}(\boldsymbol{C}_y'\boldsymbol{C}_y\otimes\boldsymbol{B}_y'\boldsymbol{B}_y)\boldsymbol{F}_1^{(y)})+\mathrm{const.},
\end{aligned}$$

where const. is constant independence from $\boldsymbol{F}_1^{(y)}$. Thus, the partial derivative function of g_{ccca} with respect to $\boldsymbol{F}_1^{(y)}$ is obtained as follows:

$$\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{F}_1^{(y)}}\{\|\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}-\boldsymbol{F}_1^{(y)}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)'\|^2+\|\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x-\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\|^2\}\\
=&-2\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)-2\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_y'+2\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_y'+2\boldsymbol{F}_1^{(y)}.
\end{aligned}$$

When we set the partial derivative function of $\boldsymbol{F}_1^{(y)}$ as 0, we obtain the Following equation:

$$\begin{aligned}
&-2\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)-2\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_y'+2\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_y'+2\boldsymbol{F}_1^{(y)}=0\\
\Longleftrightarrow &\boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\boldsymbol{D}_y'+\boldsymbol{F}_1^{(y)}=\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)+\boldsymbol{F}_1^{(y)}\boldsymbol{D}_x\boldsymbol{D}_y'\\
\Longleftrightarrow &\boldsymbol{F}_1^{(y)}=(\boldsymbol{Y}_1^{\dagger}\boldsymbol{W}(\boldsymbol{C}_y\otimes\boldsymbol{B}_y)+\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x\boldsymbol{D}_y')(\boldsymbol{I}+\boldsymbol{D}_y\boldsymbol{D}_y')^{-1}.
\end{aligned}$$

□

Proposition 3.15. *The update formula of $\boldsymbol{q}_{k_xj_x}$ is obtained as follows:*

$$\boldsymbol{q}_{k_xj_x}=\sqrt{I}(\boldsymbol{X}_{k_xj_x}^{\dagger}{}'\boldsymbol{X}_{k_xj_x}^{\dagger})^{-\frac{1}{2}}\boldsymbol{u}_1^{(qx)}, \tag{3.64}$$

where $\boldsymbol{u}_1^{(qx)}$ is the first dimension left singular vector of $(\boldsymbol{X}_{k_xj_x}^{\dagger}{}'\boldsymbol{X}_{k_xj_x}^{\dagger})^{-\frac{1}{2}}\boldsymbol{X}_{k_xj_x}^{\dagger}{}'\boldsymbol{J}_n(\sum_{\ell}^{r_{cx}}c_{k_x\ell}^{(x)}\boldsymbol{F}_{(\ell)}^{(x)}\boldsymbol{b}_{j_x}^{(x)})$. $c_{k_x\ell}^{(x)}$ is $(k_x,\ \ell)$ element of $\boldsymbol{C}_x$. $\boldsymbol{b}_{j_x}^{(x)}$ is the j_x-th row vector of $\boldsymbol{B}_x$. $\boldsymbol{F}_{(\ell)}^{(x)}$ is the matrix corresponding to dimension ℓ of $\boldsymbol{C}_x$.

The update formula of $\boldsymbol{w}_{k_yj_y}$ is obtained as follows:

$$\boldsymbol{w}_{k_yj_y}=\sqrt{I}(\boldsymbol{Y}_{k_yj_y}^{\dagger}{}'\boldsymbol{Y}_{k_yj_y}^{\dagger})^{-\frac{1}{2}}\boldsymbol{u}_1^{(wy)}, \tag{3.65}$$

where $\boldsymbol{u}_1^{(wx)}$ is the first dimension left singular vector of $(\boldsymbol{Y}_{k_yj_y}^{\dagger}{}'\boldsymbol{Y}_{k_yj_y}^{\dagger})^{-\frac{1}{2}}\boldsymbol{Y}_{k_yj_y}^{\dagger}{}'\boldsymbol{J}_n(\sum_{\ell}^{r_{cy}}c_{k_y\ell}^{(y)}\boldsymbol{F}_{(\ell)}^{(y)}\boldsymbol{b}_{j_y}^{(y)})$. $c_{k_y\ell}^{(y)}$ is $(k_y,\ \ell)$ element of $\boldsymbol{C}_y$. $\boldsymbol{b}_{j_y}^{(y)}$ is the j_x-th row vector of $\boldsymbol{B}_y$. $\boldsymbol{F}_{(\ell)}^{(y)}$ is the matrix corresponding to dimension ℓ of $\boldsymbol{C}_y$

Proof. First, we explain about the update formula of $\boldsymbol{q}_{k_xj_x}$. From definition $\boldsymbol{Q}$, $\boldsymbol{q}_{k_xj_x}$ are independent from each other. Thus, the update formula of $\boldsymbol{q}_{k_xj_x}$ can be calculated individually. The term that is related to $\boldsymbol{q}_{k_xj_x}$ is the first term of g_{ccca}. The first term of g_{ccca} is rewritten as follows:

$$\|\boldsymbol{X}_1^{\dagger}\boldsymbol{Q}-\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)'\|^2=-2\mathrm{tr}(\boldsymbol{Q}'\boldsymbol{X}^{\dagger}{}_1'\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)')+\mathrm{const.} \tag{3.66}$$

From equation (3.66), we consider the minimization problem as the $\boldsymbol{Q}$ that maximizes $\mathrm{tr}(\boldsymbol{Q}'\boldsymbol{X}^{\dagger}{}_1'\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)')$. From the definition of $\boldsymbol{Q}$, in order to maximize $\mathrm{tr}(\boldsymbol{Q}'\boldsymbol{X}^{\dagger}{}_1'\boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x\otimes\boldsymbol{B}_x)')$, we consider each value of $\boldsymbol{q}_{k_xj_x}$. Objective function g^* for $\boldsymbol{q}_{k_xj_x}$ is obtained as follows:

$$g^*(\boldsymbol{q}_{k_xj_x}\mid\boldsymbol{C}_x,\ \boldsymbol{B}_x,\ \boldsymbol{F}_x,\ \underline{\boldsymbol{X}}^{\dagger})=\mathrm{tr}(\boldsymbol{q}_{k_xj_x}'\boldsymbol{X}_{k_xj_x}^{\dagger}{}'\sum_{\ell}^{r_{cx}}c_{k_x\ell}^{(x)}\boldsymbol{F}_{(\ell)}^{(x)}\boldsymbol{b}_{j_x}^{(x)}),$$

From the constraint on $\boldsymbol{q}_{k_xj_x}$, this objective function g^* is very similar to the objective function of canonical correlation analysis. From the constraint $\boldsymbol{X}_{j_xk_x}^{\dagger}\boldsymbol{q}_{j_xk_x}=\boldsymbol{J}\boldsymbol{X}_{j_xk_x}^{\dagger}\boldsymbol{q}_{j_xk_x}$, $\boldsymbol{X}_{j_xk_x}^{\dagger}\boldsymbol{q}_{j_xk_x}$ is the element of complementary space of $\boldsymbol{1}$. Therefore, first, the $\boldsymbol{X}_{j_xk_x}^{\dagger}$ is projected $\boldsymbol{J}$ space. Then, we search the parameters maximizing g^*. Thus, we change the objective function g^* to g_1^* as follows:

$$g_1^*(\boldsymbol{q}_{k_xj_x}\mid\boldsymbol{C}_x,\ \boldsymbol{B}_x,\ \boldsymbol{F}_x,\ \underline{\boldsymbol{X}}^{\dagger})=\mathrm{tr}(\boldsymbol{q}_{k_xj_x}'\boldsymbol{X}_{k_xj_x}^{\dagger}{}'\boldsymbol{J}\sum_{\ell}^{r_{cx}}c_{k_x\ell}^{(x)}\boldsymbol{F}_{(\ell)}^{(x)}\boldsymbol{b}_{j_x}^{(x)}).$$

When we set $\boldsymbol{q}^*_{k_xj_x} = \frac{1}{\sqrt{I}}(\boldsymbol{X}^{\dagger}_{k_xj_x}{}'\boldsymbol{X}^{\dagger}_{k_xj_x})^{1/2}\boldsymbol{q}_{k_xj_x}$, equation $\boldsymbol{q}^*_{k_xj_x}{}'\boldsymbol{q}^*_{k_xj_x} = 1$ holds from the constraint case of g_{ccca}. Therefore, we can rewrite g^*_1 as follows:

$$g^*_1(\boldsymbol{q}_{k_xj_x} \mid \boldsymbol{C}_x,\ \boldsymbol{B}_x,\ \boldsymbol{F}_x,\ \underline{\boldsymbol{X}^{\dagger}}) = \sqrt{I}\mathrm{tr}(\boldsymbol{q}^*_{k_xj_x}{}'(\boldsymbol{X}^{\dagger}_{k_xj_x}{}'\boldsymbol{X}^{\dagger}_{k_xj_x})^{-\frac{1}{2}}\boldsymbol{X}^{\dagger}_{k_xj_x}{}'\boldsymbol{J}\sum_{\ell}^{r_{cx}} c^{(x)}_{k_x\ell}\boldsymbol{F}^{(x)}_{(\ell)}\boldsymbol{b}^{(x)}_{j_x}).$$

g^*_1 is the same as the objective function of the weight matrix in the categorical canonical covariance case. Therefore, we obtain the update formula of $\boldsymbol{q}^*_{k_xj_x}$ as

$$\boldsymbol{q}^*_{k_xj_x} = \boldsymbol{u}^{(qx)}_1,$$

where $\boldsymbol{u}^{(qx)}_1$ is the first dimension left singular vector of $(\boldsymbol{X}^{\dagger}_{k_xj_x}{}'\boldsymbol{X}^{\dagger}_{k_xj_x})^{-\frac{1}{2}}\boldsymbol{X}^{\dagger}_{k_xj_x}{}'\boldsymbol{J}\sum_{\ell}^{r_{cx}} c^{(x)}_{k_x\ell}\boldsymbol{F}^{(x)}_{(\ell)}\boldsymbol{b}^{(x)}_{j_x}$. From the definition of $\boldsymbol{q}^*_{k_xj_x}$, the update formula of $\boldsymbol{q}_{k_xj_x}$ is obtained as follows:

$$\boldsymbol{q}_{k_xj_x} = \sqrt{I}(\boldsymbol{X}^{\dagger}_{k_xj_x}{}'\boldsymbol{X}^{\dagger}_{k_xj_x})^{-\frac{1}{2}}\boldsymbol{u}^{(qx)}_1.$$

The update formula of $\boldsymbol{w}_{k_yj_y}$ is obtained in the same way as $\boldsymbol{q}_{k_xj_x}$. □

The update formulas of $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ are the same as the K-means algorithm.

Proposition 3.16. *The update formulas of $\boldsymbol{D}_{bx}$, $\boldsymbol{D}_{by}$, $\boldsymbol{D}_{cx}$, and $\boldsymbol{D}_{cy}$ are obtained as follows:*

$$d^{(bx)}_{\ell q} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{F}^{(x)}_2(\boldsymbol{D}_{cx} \otimes \boldsymbol{I})]_{\ell^*} - \boldsymbol{d}^{(by)}_q{}'\boldsymbol{F}^{(y)}_2(\boldsymbol{D}_{cy} \otimes \boldsymbol{I}) \right\| \right) \\ 0 & (otherwise) \end{cases} \quad (q = 1,\ 2, \ldots, c_b), \tag{3.67}$$

$$d^{(by)}_{\ell q} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{F}^{(y)}_2(\boldsymbol{D}_{cy} \otimes \boldsymbol{I})]_{\ell^*} - \boldsymbol{d}^{(bx)}_q{}'\boldsymbol{F}^{(x)}_2(\boldsymbol{D}_{cx} \otimes \boldsymbol{I}) \right\| \right) \\ 0 & (otherwise) \end{cases} \quad (q = 1,\ 2, \ldots, c_b), \tag{3.68}$$

$$d^{(cx)}_{\ell q} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{F}^{(x)}_3(\boldsymbol{D}_{bx} \otimes \boldsymbol{I})]_{\ell^*} - \boldsymbol{d}^{(cy)}_q{}'\boldsymbol{F}^{(y)}_3(\boldsymbol{D}_{by} \otimes \boldsymbol{I}_n) \right\| \right) \\ 0 & (otherwise) \end{cases} \quad (q = 1,\ 2, \ldots, c_c), \tag{3.69}$$

$$d^{(cy)}_{\ell q} = \begin{cases} 1 & \left(\ell = \arg\min_{\ell^*} \left\| [\boldsymbol{F}^{(y)}_3(\boldsymbol{D}_{by} \otimes \boldsymbol{I})]_{\ell^*} - \boldsymbol{d}^{(cx)}_q{}'\boldsymbol{F}^{(x)}_3(\boldsymbol{D}_{bx} \otimes \boldsymbol{I}_n) \right\| \right) \\ 0 & (otherwise) \end{cases} \quad (q = 1,\ 2, \ldots, c_c), \tag{3.70}$$

where $d^{(bx)}_{\ell q}$, $d^{(by)}_{\ell q}$, $d^{(cx)}_{\ell q}$ and $d^{(cy)}_{\ell q}$ are the (ℓ, q) element of $\boldsymbol{D}_{bx}$, $\boldsymbol{D}_{by}$, $\boldsymbol{D}_{cx}$, and $\boldsymbol{D}_{cy}$, respectively. $[\boldsymbol{A}]_\ell$ is the ℓ-th column vector of $\boldsymbol{A}$.

Proof. From the definition of g_{ccca}, the third term depends on $\boldsymbol{D}_{bx},\ \boldsymbol{D}_{by},\ \boldsymbol{D}_{cx}$, and $\boldsymbol{D}_{cy}$. We rewrite the third term of g_{ccca} as follows:

$$\begin{aligned}\|\boldsymbol{F}_1^{(x)}\boldsymbol{D}_x - \boldsymbol{F}_1^{(y)}\boldsymbol{D}_y\|^2 &= \|\boldsymbol{F}_1^{(x)}(\boldsymbol{D}_{cx}\otimes\boldsymbol{D}_{bx}) - \boldsymbol{F}_1^{(y)}(\boldsymbol{D}_{cy}\otimes\boldsymbol{D}_{by})\|^2 \\ &= \|\boldsymbol{D}_{bx}'\boldsymbol{F}_2^{(x)}(\boldsymbol{D}_{cx}\otimes\boldsymbol{I}) - \boldsymbol{D}_{by}'\boldsymbol{F}_2^{(y)}(\boldsymbol{D}_{cx}\otimes\boldsymbol{I})\|^2 \quad (3.71) \\ &= \|\boldsymbol{D}_{cx}'\boldsymbol{F}_3^{(x)}(\boldsymbol{D}_{bx}\otimes\boldsymbol{I}) - \boldsymbol{D}_{cy}'\boldsymbol{F}_3^{(y)}(\boldsymbol{D}_{by}\otimes\boldsymbol{I})\|^2 \quad (3.72)\end{aligned}$$

From the constraint of $\boldsymbol{D}_{bx},\ \boldsymbol{D}_{by}$, equation (3.71) is equivalent to K-means given other parameters. Thus, the update formulas of $\boldsymbol{D}_{bx},\ \boldsymbol{D}_{by}$ are obtained in the same way as K-means.

From the constraint of $\boldsymbol{D}_{cx},\ \boldsymbol{D}_{cy}$, equation (3.72) is equivalent to K-means given other parameters. Thus, the update formulas of $\boldsymbol{D}_{cx},\ \boldsymbol{D}_{cy}$ are obtained in the same way as K-means. □

Summarizing the update formulas, we obtain the algorithm of categorical canonical covariance analysis for three-mode three-way data as algorithm 6.

Algorithm 6 Algorithm of the NPCA-based method

Set the number of dimensions $r_{bx},\ r_{by},\ r_{cx},\ r_{cy}, c_b,\ c_c$, and stop condition ε
Set initial values $\boldsymbol{B}_x^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_y^{(0)},\ \underline{\boldsymbol{F}}_x^{(0)},\ \underline{\boldsymbol{F}}_y^{(0)},\ \boldsymbol{D}_{bx}^{(0)},\ \boldsymbol{D}_{by}^{(0)}, \boldsymbol{D}_{cx}^{(0)}$, and $\boldsymbol{D}_{cy}^{(0)}$,
$t \leftarrow 0$
$S^{(0)} \leftarrow g_{ccca}(\underline{\boldsymbol{F}}_x^{(0)},\ \underline{\boldsymbol{F}}_y^{(0)},\ \boldsymbol{B}_x^{(0)},\ \boldsymbol{B}_y^{(0)},\ \boldsymbol{C}_x^{(0)},\ \boldsymbol{C}_y^{(0)},\ \boldsymbol{Q}^{(0)},\ \boldsymbol{W}^{(0)},\ \boldsymbol{D}_x^{(0)},\ \boldsymbol{D}_y^{(0)} \mid \underline{\boldsymbol{X}}^\dagger,\ \underline{\boldsymbol{Y}}^\dagger)$
repeat
  $t \leftarrow t+1$
  Update $\boldsymbol{B}_x$ and $\boldsymbol{B}_y$ using $\boldsymbol{C}_x^{(t-1)},\ \underline{\boldsymbol{F}}_x^{(t-1)}, \boldsymbol{C}_y^{(t-1)},\ \underline{\boldsymbol{F}}_y^{(t-1)}$
  Update $\boldsymbol{C}_x$ and $\boldsymbol{C}_y$ using $\boldsymbol{B}_x^{(t)},\ \underline{\boldsymbol{F}}_x^{(t-1)}, \boldsymbol{B}_y^{(t)},\ \underline{\boldsymbol{F}}_y^{(t-1)}$
  Update $\underline{\boldsymbol{F}}_x$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)}, \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \underline{\boldsymbol{F}}_y^{(t-1)}, \boldsymbol{D}_x^{(t-1)},\ \boldsymbol{D}_y^{(t-1)}$
  Update $\underline{\boldsymbol{F}}_y$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)}, \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \underline{\boldsymbol{F}}_x^{(t)}, \boldsymbol{D}_x^{(t-1)},\ \boldsymbol{D}_y^{(t-1)}$
  Update $\boldsymbol{D}_{bx}$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)}, \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \underline{\boldsymbol{F}}_x^{(t)}, \underline{\boldsymbol{F}}_y^{(t)},\ \boldsymbol{D}_{cx}^{(t-1)},\ \boldsymbol{D}_y^{(t-1)}$
  Update $\boldsymbol{D}_{cx}$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)}, \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \underline{\boldsymbol{F}}_x^{(t)}, \underline{\boldsymbol{F}}_y^{(t)},\ \boldsymbol{D}_{bx}^{(t)},\ \boldsymbol{D}_y^{(t-1)}$
  Update $\boldsymbol{D}_{by}$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)}, \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \underline{\boldsymbol{F}}_x^{(t)}, \underline{\boldsymbol{F}}_y^{(t)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_{cy}^{(t-1)}$
  Update $\boldsymbol{D}_{by}$ using $\boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)}, \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \underline{\boldsymbol{F}}_x^{(t)}, \underline{\boldsymbol{F}}_y^{(t)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_{by}^{(t)}$
  $S^{(t)} \leftarrow g_{ccca}(\underline{\boldsymbol{F}}_x^{(t)},\ \underline{\boldsymbol{F}}_y^{(t)},\ \boldsymbol{B}_x^{(t)},\ \boldsymbol{B}_y^{(t)},\ \boldsymbol{C}_x^{(t)},\ \boldsymbol{C}_y^{(t)},\ \boldsymbol{Q}^{(t)},\ \boldsymbol{W}^{(t)},\ \boldsymbol{D}_x^{(t)},\ \boldsymbol{D}_y^{(t)} \mid \underline{\boldsymbol{X}}^\dagger,\ \underline{\boldsymbol{Y}}^\dagger)$
until $|S^{(t-1)} - S^{(t)}| \le \varepsilon$

Chapter 4

Simulation studies

In this chapter, we compare canonical covariance analysis for three-mode three-way data with that for two-mode two-way data using several evaluations. In the K-means based method case, we compare our proposed method with two-mode two-way canonical covariance analysis and no connector matrixes with the mean squared loss between the true and estimated weight matrixes. In the regression-based method case, we compare our proposed method with two-mode two-way methods and three-mode three-way regression using mean squared loss of prediction. In the quantification method case, we compare our proposed method with two-mode two-way canonical covariance analysis, no connector matrixes, and no quantification method using the mean squared loss between the true and estimated weight matrixes.

4.1 Constrained connector method

In this section, we describe a numerical example for the contained connector method. The purpose of introducing the constraint for parameters is different between the K-means type and the regression type. Therefore, we separate these two types. In the K-means type situation, we evaluate the squared error of parameters. In the regression type situation, we focus on the prediction error.

4.1.1 K-means type

In this subsection, we explain a numerical example for the K-means type. In this example, we consider that some canonical vectors are highly correlated in the low-dimensional space, and some canonical vectors are not correlated with other canonical vectors. In this case, we show that the connector matrix K-means type constrained case is better than the non-constrained case.

4.1.1.1 Data generation

We set the number of objects $I = 300$, the numbers of variables $J_x = 15,\ J_y = 16$, and the numbers of conditions $K_x = 16,\ K_y = 12$. We also set the numbers of dimensions for factor $r_{bx} = r_{by} = r_{cx} = r_{cy} = 3$. $\boldsymbol{X}_1,\ \boldsymbol{Y}_1$ is generated as follows:

$$\boldsymbol{X}_1 = \boldsymbol{U}(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)' + \boldsymbol{E}_x,\ \boldsymbol{Y}_1 = \boldsymbol{V}(\boldsymbol{C}_y \otimes \boldsymbol{B}_y)' + \boldsymbol{E}_y, \tag{4.1}$$

where the i-th row vector of $\boldsymbol{U}$ and $\boldsymbol{V}$ simultaneously follow multivariate normal distribution with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$. In other words,

$$(\boldsymbol{u}_i, \boldsymbol{v}_i) \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma}),\ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_{uv}' \\ \boldsymbol{\Sigma}_{uv} & \boldsymbol{I} \end{pmatrix},\ \boldsymbol{\Sigma}_{uv} = (\sigma_{ij}),$$

$$\sigma_{ij} = \begin{cases} 0.8 & (i,j) \in \{(1,9),\ (3,7),\ (4,6),\ (6,4)\} \\ 0 & \text{otherwise} \end{cases}.$$

The reason that part of σ_{ij} is set as 0.8 and the other part as 0 is that we set $\boldsymbol{D}_{bx}$, $\boldsymbol{D}_{by}$, $\boldsymbol{D}_{cx}$, and $\boldsymbol{D}_{cy}$ as follows:

$$\boldsymbol{D}_{bx} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{D}_{by} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix},$$
$$\boldsymbol{D}_{cx} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{D}_{cy} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Figure 4.1 shows the heat map of the covariance matrix, which shows the true covariance matrix after dimensional reduction. Therefore, the red block on the bottom left shows the covariance between $\boldsymbol{X}_1$ and $\boldsymbol{Y}_1$ after dimensional reduction. From Figure 4.1, the connector matrix is set to satisfy connecting between $\boldsymbol{b}_{x1}$ and $\boldsymbol{b}_{y3}$, $\boldsymbol{b}_{x3}$ and $\boldsymbol{b}_{y1}$, $\boldsymbol{c}_{x2}$ and $\boldsymbol{c}_{y2}$, and $\boldsymbol{c}_{x3}$ and $\boldsymbol{c}_{y1}$ in this condition. Therefore, there are four common factors in this setting while the other factors apply to each dataset. Because this covariance setting is a two-factor type of mixed setting, we expect the K-means based method to have the best result. The main contribution of this numerical study is that we confirm it is suitable for applying our proposed methods to datasets with similar assumed settings.

The elements of $\boldsymbol{E}_x$ and $\boldsymbol{E}_y$ follow normal distribution with mean 0 and variance $\sigma^2 = 0.1^2$. We set parameters $\boldsymbol{B}_x$, $\boldsymbol{B}_y$, $\boldsymbol{C}_x$, and $\boldsymbol{C}_y$ as follows:

$$\boldsymbol{B}_x = (\boldsymbol{b}_{x1},\ \boldsymbol{b}_{x2},\ \boldsymbol{b}_{x3}),\ \boldsymbol{B}_y = (\boldsymbol{b}_{y1},\ \boldsymbol{b}_{y2},\ \boldsymbol{b}_{y3}),$$
$$\boldsymbol{C}_x = (\boldsymbol{c}_{x1},\ \boldsymbol{c}_{x2},\ \boldsymbol{c}_{x3}),\ \boldsymbol{C}_y = (\boldsymbol{c}_{y1},\ \boldsymbol{c}_{y2},\ \boldsymbol{c}_{y3}),$$

where $\boldsymbol{b}_{xi},\ \boldsymbol{b}_{yi},\ \boldsymbol{c}_{xi}$ and $\boldsymbol{c}_{yi}$ are the i-th column vector of $\boldsymbol{B}_x, \boldsymbol{B}_y,\ \boldsymbol{C}_x$, and $\boldsymbol{C}_y$, respectively. the column vectors of $\boldsymbol{B}_x, \boldsymbol{B}_y,\ \boldsymbol{C}_x$, and $\boldsymbol{C}_y$ are set as follows:

$$\boldsymbol{b}_{x1} = (\boldsymbol{1}_5',\ \boldsymbol{0}_{10}')', \boldsymbol{b}_{x2} = (\boldsymbol{0}_5',\ \boldsymbol{1}_5',\ \boldsymbol{0}_5')',\ \boldsymbol{b}_{x3} = (\boldsymbol{0}_{10}',\ \boldsymbol{1}_5')',$$
$$\boldsymbol{b}_{y1} = (\boldsymbol{1}_5',\ \boldsymbol{0}_{11}')', \boldsymbol{b}_{y2} = (\boldsymbol{0}_5',\ \boldsymbol{1}_5',\ \boldsymbol{0}_6')',\ \boldsymbol{b}_{y3} = (\boldsymbol{0}_{10}',\ \boldsymbol{1}_6')',$$
$$\boldsymbol{c}_{x1} = (\boldsymbol{1}_5',\ \boldsymbol{0}_{11}')', \boldsymbol{c}_{x2} = (\boldsymbol{0}_5',\ \boldsymbol{1}_6',\ \boldsymbol{0}_5')',\ \boldsymbol{c}_{x3} = (\boldsymbol{0}_{11}',\ \boldsymbol{1}_5')',$$
$$\boldsymbol{c}_{y1} = (\boldsymbol{1}_4',\ \boldsymbol{0}_8')', \boldsymbol{c}_{y2} = (\boldsymbol{0}_4',\ \boldsymbol{1}_3',\ \boldsymbol{0}_5')',\ \boldsymbol{c}_{y3} = (\boldsymbol{0}_7',\ \boldsymbol{1}_5')',$$

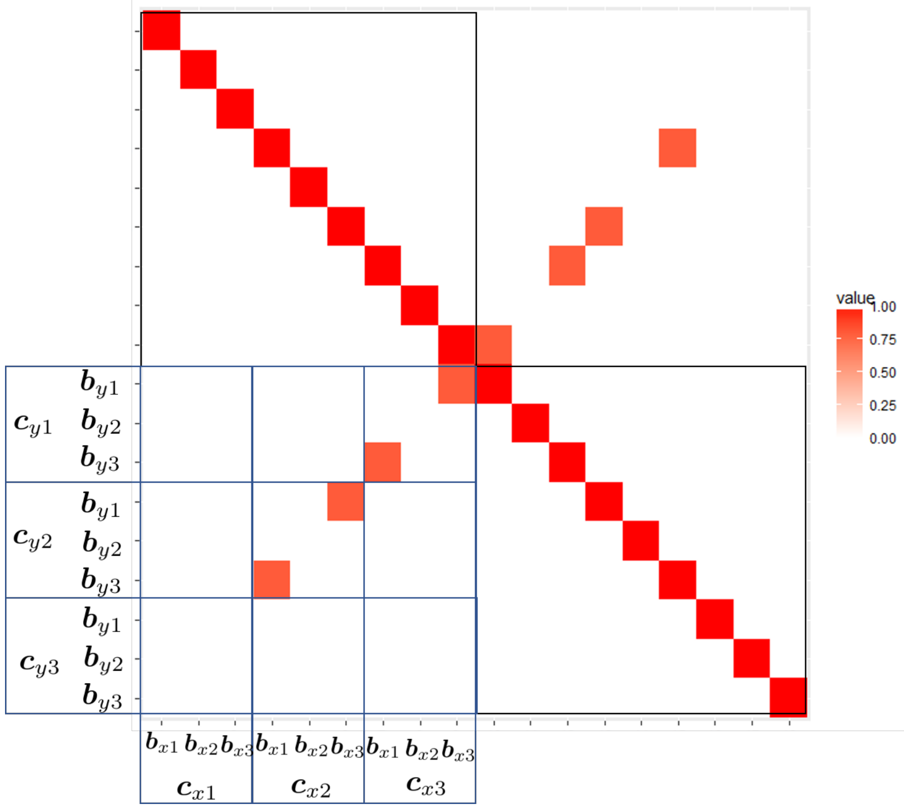


Figure 4.1: Heatmap of covariance matrix for $(\boldsymbol{u}, \boldsymbol{v})$

where $\mathbf{0}_d$ and $\mathbf{1}_d$ is the d dimensional vector whose elements are all 0 and 1, respectively.

We compare K-means type connector canonical covariance analysis for three-mode three-way data with non-connector canonical covariance analysis for three-mode three-way data and canonical covariance analysis for multivariate data. For each estimator, we calculate the mean of squared error, defined as follows:

$$\frac{1}{R}\sum\left(\|\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x - \boldsymbol{C}_x \otimes \boldsymbol{B}_x\|^2 + \|\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y - \boldsymbol{C}_y \otimes \boldsymbol{B}_y\|^2\right),$$

where R is iteration time. We set the iteration time R as 100. When we evaluate the mean squared error for the two-mode two-way method, we set $\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x = \hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y = \hat{\boldsymbol{B}}$. Moreover, we set the number of dimensions as the same number of parameters. In other words, $\hat{\boldsymbol{A}} \in \mathbb{R}^{15*12\times 9}, \hat{\boldsymbol{B}} \in \mathbb{R}^{16*12\times 9}$, $\hat{\boldsymbol{D}}_{bx} \in \{0,1\}^{3\times 2}, \hat{\boldsymbol{D}}_{by} \in \{0,1\}^{3\times 2}$, $\hat{\boldsymbol{D}}_{cx} \in \{0,1\}^{3\times 2}$, $\hat{\boldsymbol{D}}_{cy} \in \{0,1\}^{3\times 2}$.

We evaluate the following four cases. The first case has no noise variable. In other words, $\boldsymbol{X}_1$ and $\boldsymbol{Y}_1$ are generated by equation 4.1. The second and third cases use data with five additional variables and conditions, respectively. The five variables and conditions are not correlated other parameters. The fourth case uses data with the five additional variables and conditions. The noise that is not correlated with other parameters follows the same distribution as that of $\boldsymbol{E}_x$ and $\boldsymbol{E}_y$. Figure 4.2 depicts these cases.

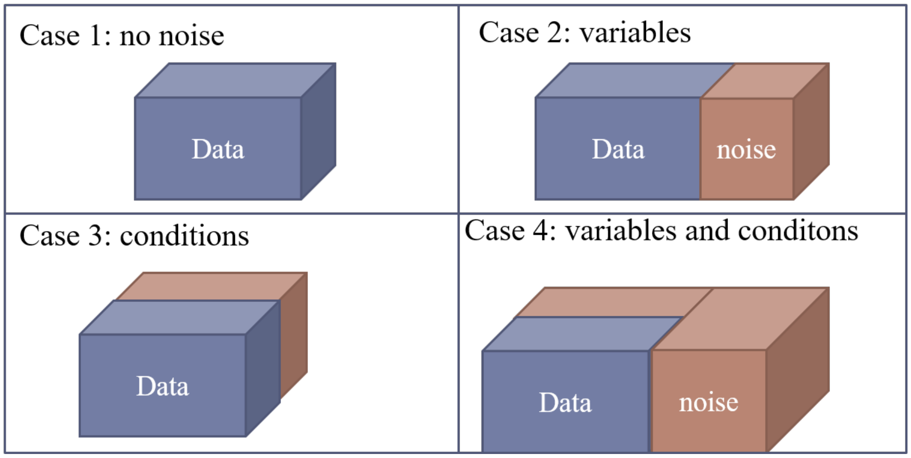


Figure 4.2: Representation of unfolding mode

4.1.1.2 Result

Figure 4.3 shows the boxplots of all cases. The boxplots show, from the left, the K-means type constrained connector matrix set as $\boldsymbol{I}$, and the two-way canonical covariance method. From Figure 4.3, applying the K-means type constrained case to these data is the better of the two methods, because the squared error of parameters is smaller than that of other methods. The reason for the difference between the three-way methods is the covariance structure of the simulation study. The covariance matrix between the canonical vectors is very sparse. However, the assumption of connector matrix set as $\boldsymbol{I}$ shows that the canonical vector is highly correlated with at least one other canonical vector.

Table 4.1 shows the simulation result of the mean of squared error and the standard deviation of squared error. The K-means based method has the best result among them because the setting is the same of the assumed K-means based method. The difference between the K-means based method and constrained connector matrix set as $\boldsymbol{I}$ is smaller than the difference between the K-means based method and the two-mode two-way method. We guess this explains why the numerical study is in a three-mode three-way setting. From this fact, when we assume that data have a three-mode three-way structure, it is suitable to apply multivariate data analysis to a dataset. The standard deviation of the two-way method is smallest among the methods because the solution of the two-way method is obtained by singular value decomposition. Therefore, it tends to yield a stabler estimator than the other method does. This is the drawback of the three-way three-mode methods.

4.1.2 Regression type

In this subsection, we explain the numerical example for the regression type connector matrix. One of the purposes of applying PLS to data is prediction. Thus, we compare the PLS method with three-mode three-way regression, the two-way PLS method, and two-way regression by prediction error. The main purpose of this numerical study is to

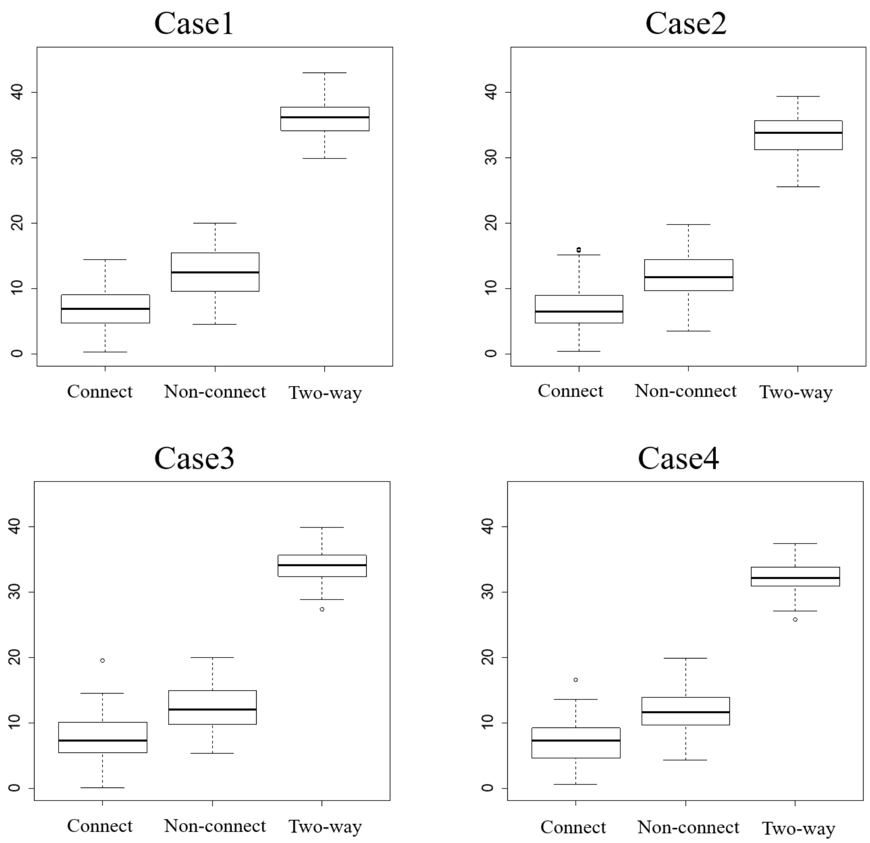


Figure 4.3: Boxplots of all case results

Table 4.1: Mean squared error of parameters under all cases

	Connect	Non-connect	Two-way
Case 1	7.048(3.207)	12.373(3.925)	35.937(2.613)
Case 2	7.077(3.743)	11.961(3.664)	33.574(2.985)
Case 3	7.774(3.437)	12.339(3.472)	34.056(2.280)
Case 4	7.115(3.286)	11.746(3.097)	32.255(2.334)

confirm that the regression-based method is better than two-mode two-way regression and two-mode two-way PLS in the sense of prediction. The result of the numerical example shows that the three-way PLS method is a little better than the other methods are.

4.1.2.1 Data generation

We generate the PLS scores $\boldsymbol{T}_1 \in \mathbb{R}^{I\times 3*3}$. The row vectors of $\boldsymbol{T}_1$ follow an identical and independent multivariate normal distribution with the mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}$. In the first case, the weight matrixes of $\boldsymbol{C}_x \in \mathbb{R}^{16\times 3}$ and $\boldsymbol{B}_x \in \mathbb{R}^{15\times 3}$ generate uniform distributions with min $= -1$ and max $= 1$. However, these matrixes are not orthogonal. By using singular value decomposition, we obtain the orthogonal matrixes. In the second case, the weight matrixes of $\boldsymbol{C}_x \in \mathbb{R}^{16\times 3}$ and$\boldsymbol{B}_x \in \mathbb{R}^{15\times 3}$ are generated as follows:

$$\boldsymbol{B}_x = (\boldsymbol{b}_{x1},\ \boldsymbol{b}_{x2},\ \boldsymbol{b}_{x3}),\ \boldsymbol{C}_x = (\boldsymbol{c}_{x1},\ \boldsymbol{c}_{x2},\ \boldsymbol{c}_{x3}),$$

$$\boldsymbol{b}_{x1} = (\mathbf{1}_5',\ \mathbf{0}_{10}')',\ \boldsymbol{b}_{x2} = (\mathbf{0}_5',\ \mathbf{1}_5',\ \mathbf{0}_5')',\ \boldsymbol{b}_{x3} = (\mathbf{0}_{10}',\ \mathbf{1}_5')',$$
$$\boldsymbol{c}_{x1} = (\mathbf{1}_5',\ \mathbf{0}_{11}')',\ \boldsymbol{c}_{x2} = (\mathbf{0}_5',\ \mathbf{1}_6',\ \mathbf{0}_5')',\ \boldsymbol{c}_{x3} = (\mathbf{0}_{11}',\ \mathbf{1}_5')',$$

The independent variables $\boldsymbol{X}_1 \in \mathbb{R}^{I\times 15*16}$ are defined as $\boldsymbol{X}_1 = \boldsymbol{T}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)' + \boldsymbol{E}$, where $\boldsymbol{E}$ is a noise matrix with elements following an identical and independent normal distribution with a mean of 0 and standard deviation is set by sd. The response variables $\boldsymbol{Y}_1 \in \mathbb{R}^{I\times 12*10}$ are generated as $\boldsymbol{T}_1(\boldsymbol{C}_x \otimes \boldsymbol{B}_x)'\boldsymbol{A} + \boldsymbol{E}_y$. $\boldsymbol{E}_y$ is generated in the same manner as $\boldsymbol{E}$. $\boldsymbol{A}$ is the regression coefficient matrix whose elements follow independently and identically uniform distributions with a minimum value of -1 and a maximum value of 1. We set the number of training samples $I = (300, 500)$ and the standard deviation of the noise $sd = (0.1, 0.5, 1)$. We use the following evaluation criterion for the mean squared prediction error:

$$\frac{1}{R}\sum\left(\frac{1}{(I_p * 12 * 10)}\|\boldsymbol{Y}_{1_p} - \hat{\boldsymbol{Y}}_{1_p}\|^2\right),$$

where I_p is the number of test data sets. We set I_p as the same number of $\boldsymbol{Y}_1$ in all cases. $\boldsymbol{Y}_{1_p}$ is generated in the same manner as $\boldsymbol{Y}_1$, and $\boldsymbol{Y}_{1_p}$ is the predictor when using the estimated parameters. The following methods are compared: the combination of three-mode three-way PLS and two-mode two-way linear regression, three-mode three-way regression, two-mode two-way PLS, and two-mode two-way linear regression. The combined method of three-mode three-way PLS and two-mode two-way linear regression uses the PLS score estimated for the three-mode three-way data set as independent variables. R is iteration time. We set R as 100.

4.1.2.2 Result

Figures 4.4 and 4.5 show the boxplots of the prediction error. The boxplot shows, from the left, the combination of three-mode three-way PLS and three-mode three-way linear regression, and the combination of three-mode three-way PLS and two-mode two-way linear regression, three-mode three-way regression, two-mode two-way PLS, and two-mode two-way linear regression. From Figures 4.4 and 4.5, prediction error depends not on dimensional reduction but on whether the data are three-way or not. The reason for this result is based on whether the predictor has a three-way structure or not. The three-mode three-way method has a three-way structure. In this case, the predictor generates the same structure as the data. In other words, the data are described as the mode product. Therefore, the difference between the three-way methods is small.

Tables 4.2 and 4.3 show the mean squared prediction error. The table shows, from the left, the combination of three-mode three-way PLS and three-mode three-way linear regression, and the combination of three-mode three-way PLS and two-mode two-way linear regression, three-mode three-way regression, two-mode two-way PLS, and two-mode two-way linear regression. From Tables 4.2 and 4.3, under all settings, the combination of three-mode three-way PLS and three-way regression has the best result among the methods. Except for the combination of three-mode three-way PLS and three-way regression, three-way regression has the best result under all settings. There are few differences between the combined methods of three-mode three-way PLS and three-way regression. The reason for this result is that it is possible to estimate the regression coefficient without dimensional reduction.

Table 4.2: Mean prediction error under setting 1: Simple weight case

	Three-Three	Three-Two	Three-way reg	Two-Two	Two-way reg
$n = 300,\ sd = 0.1$	0.120(0.005)	0.123(0.005)	0.121(0.005)	2.326(0.629)	0.616(0.031)
$n = 300,\ sd = 0.5$	0.598(0.022)	0.615(0.023)	0.606(0.023)	2.693(0.543)	3.071(0.151)
$n = 300,\ sd = 1.0$	1.171(0.041)	1.204(0.041)	1.185(0.042)	2.970(0.496)	6.000(0.303)
$n = 500,\ sd = 0.1$	0.121(0.005)	0.123(0.005)	0.122(0.005)	2.485(0.610)	0.234(0.009)
$n = 500,\ sd = 0.5$	0.596(0.025)	0.606(0.025)	0.601(0.025)	2.804(0.743)	1.151(0.050)
$n = 500,\ sd = 1.0$	1.177(0.038)	1.197(0.038)	1.185(0.039)	3.056(0.493)	2.270(0.075)

Table 4.3: Mean prediction error of setting 2: Random weight case

	Three-Three	Three-Two	Three-way reg	Two-Two	Two-way reg
$n = 300,\ sd = 0.1$	0.121(0.005)	0.124(0.005)	0.122(0.006)	2.321(0.631)	0.616(0.032)
$n = 300,\ sd = 0.5$	0.598(0.027)	0.615(0.027)	0.606(0.027)	2.608(0.603)	3.073(0.163)
$n = 300,\ sd = 1.0$	1.177(0.043)	1.211(0.044)	1.190(0.043)	2.916(0.534)	6.030(0.275)
$n = 500,\ sd = 0.1$	0.120(0.005)	0.122(0.005)	0.121(0.005)	2.153(0.525)	0.231(0.010)
$n = 500,\ sd = 0.5$	0.597(0.025)	0.607(0.025)	0.602(0.025)	2.606(0.570)	1.150(0.048)
$n = 500,\ sd = 1.0$	1.169(0.042)	1.189(0.043)	1.177(0.043)	2.903(0.527)	2.254(0.083)

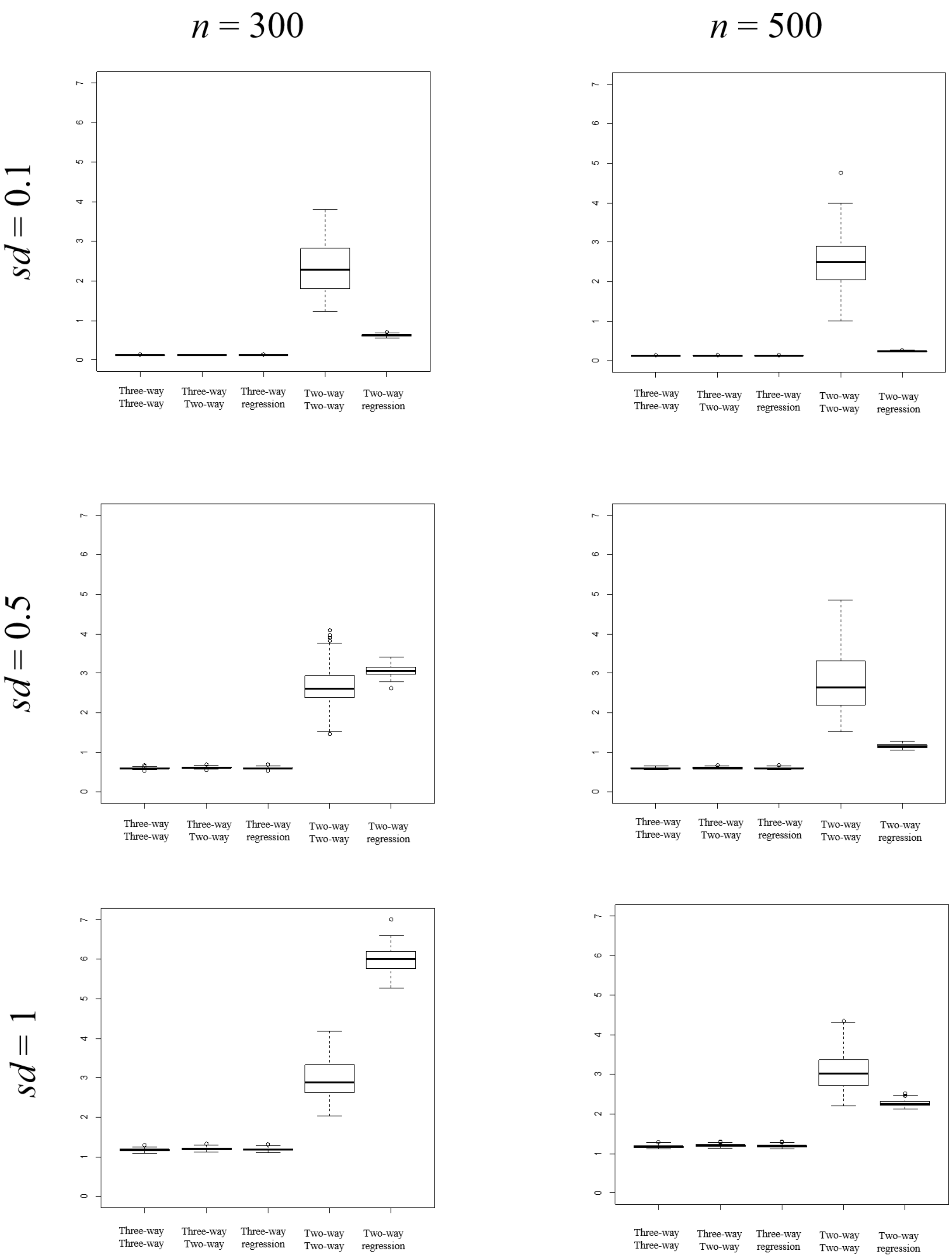


Figure 4.4: Boxplots of the prediction error in setting 1

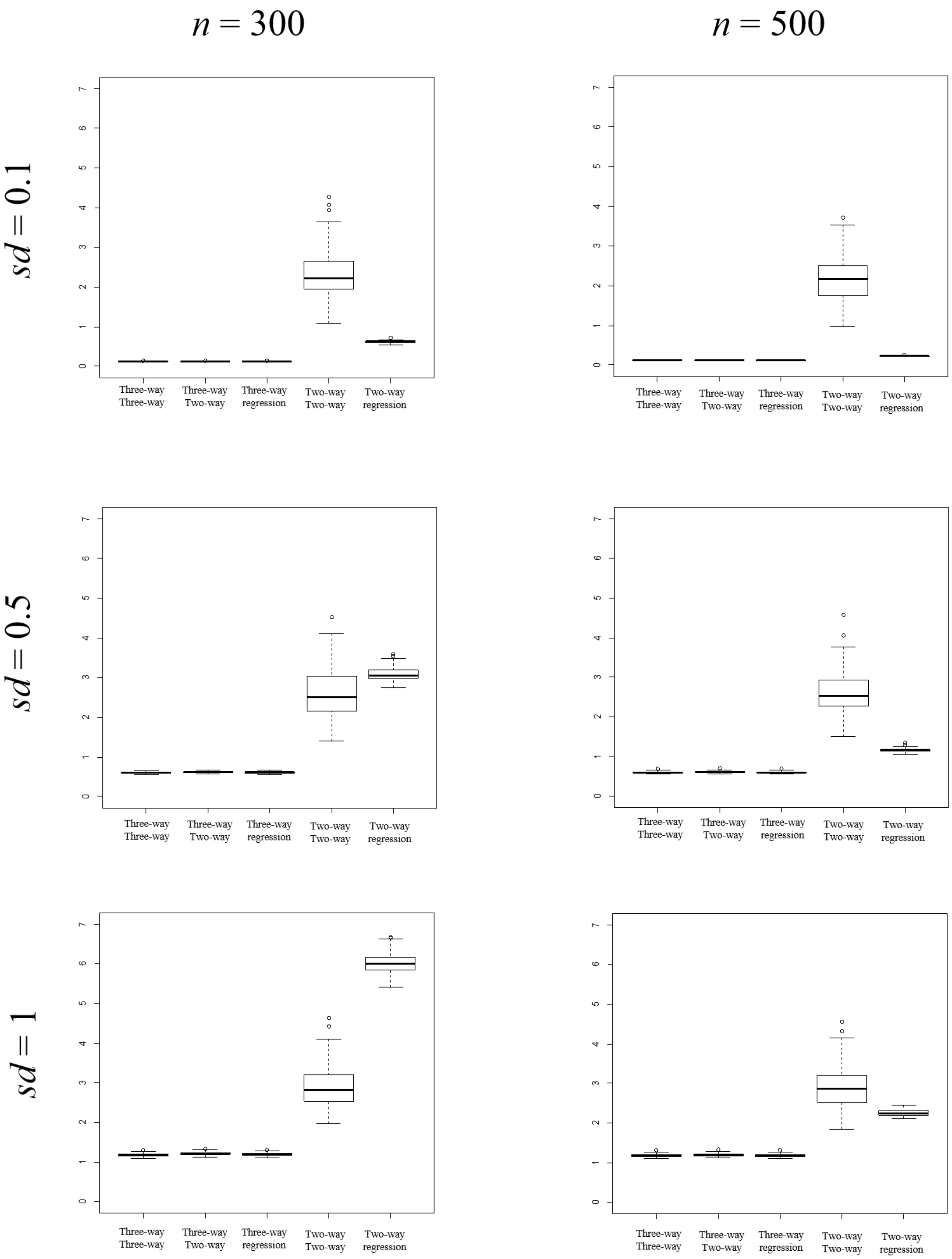


Figure 4.5: Boxplots of the prediction error in setting 2

4.2 Category quantification method

4.2.1 NPCA-based method

In this subsection, we consider the situation in which all values are categorical. In other words, all the variables are categorical variables under all conditions of data. In this numerical example, we consider two types of setting. In the first, the variable has order. In the second setting, the variable has no order. The main purpose of this numerical study is to confirm that the category quantification method has a better result than the non-quantification method does under the case of categorical variables. In both settings, the NPCA-based method has the best result among the compared methods.

4.2.1.1 Data generation

We compare categorical canonical covariance analysis for three-mode thee-way data with canonical covariance analysis for three-mode three-way and for two-mode two-way data. We set the true parameters $\boldsymbol{B}_x,\ \boldsymbol{B}_y,\ \boldsymbol{C}_x,$ and $\boldsymbol{C}_y$ as follows:

$$\begin{aligned}\boldsymbol{B}_x &= (\boldsymbol{b}_{x1},\ \boldsymbol{b}_{x2},\ \boldsymbol{b}_{x3}),\ \boldsymbol{B}_y = (\boldsymbol{b}_{y1},\ \boldsymbol{b}_{y2},\ \boldsymbol{b}_{y3}),\\ \boldsymbol{C}_x &= (\boldsymbol{c}_{x1},\ \boldsymbol{c}_{x2},\ \boldsymbol{c}_{x3}),\ \boldsymbol{C}_y = (\boldsymbol{c}_{y1},\ \boldsymbol{c}_{y2},\ \boldsymbol{c}_{y3}),\end{aligned}$$

$$\begin{aligned}\boldsymbol{b}_{x1} &= (\boldsymbol{1}_5',\ \boldsymbol{0}_{10}')',\ \boldsymbol{b}_{x2} = (\boldsymbol{0}_5',\ \boldsymbol{1}_5',\ \boldsymbol{0}_5')',\ \boldsymbol{b}_{x3} = (\boldsymbol{0}_{10}',\ \boldsymbol{1}_5')',\\ \boldsymbol{b}_{y1} &= (\boldsymbol{1}_5',\ \boldsymbol{0}_{11}')',\ \boldsymbol{b}_{y2} = (\boldsymbol{0}_5',\ \boldsymbol{1}_5',\ \boldsymbol{0}_6')',\ \boldsymbol{b}_{y3} = (\boldsymbol{0}_{11}',\ \boldsymbol{1}_6')',\\ \boldsymbol{c}_{x1} &= (\boldsymbol{1}_5',\ \boldsymbol{0}_{11}')',\ \boldsymbol{c}_{x2} = (\boldsymbol{0}_5',\ \boldsymbol{1}_6',\ \boldsymbol{0}_5')',\ \boldsymbol{c}_{x3} = (\boldsymbol{0}_{11}',\ \boldsymbol{1}_5')',\\ \boldsymbol{c}_{y1} &= (\boldsymbol{1}_4',\ \boldsymbol{0}_8')',\ \boldsymbol{c}_{y2} = (\boldsymbol{0}_4',\ \boldsymbol{1}_3',\ \boldsymbol{0}_5')',\ \boldsymbol{c}_{y3} = (\boldsymbol{0}_7',\ \boldsymbol{1}_5')'.\end{aligned}$$

Then, to satisfy the constraint, we normalize the loading matrixes. $\boldsymbol{F}_1^{(x)}$ and $\boldsymbol{F}_1^{(y)}$ are generated as follows:

$$(\boldsymbol{f}_i^{(x)},\ \boldsymbol{f}_i^{(y)}) \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(0,\ \boldsymbol{\Sigma}),\ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_{f_xf_y}' \\ \boldsymbol{\Sigma}_{f_xf_y} & \boldsymbol{I} \end{pmatrix},\ \boldsymbol{\Sigma}_{f_xf_y} = (\sigma_{ij}),$$

$$\sigma_{ij} = \begin{cases} 1 & (i = j) \\ 0.8 & ((i,\ j) \in \{(1,\ 9),\ (3,\ 7),\ (4,\ 6),\ (6,\ 4)\}) \\ 0 & (\text{otherwise}) \end{cases}.$$

This setting represents the case in which there are two common factor loadings for variables and conditions. Thus, there are four common factors in these data. This covariance matrix setting is the same as that of the K-means type. To generate data sets $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$, we first set the score data sets $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$ as follows:

$$\boldsymbol{X}_1^* = \boldsymbol{F}_1^{(x)}(\boldsymbol{C}_x' \otimes \boldsymbol{B}_x') + \boldsymbol{E}_x,\ \boldsymbol{Y}_1^* = \boldsymbol{F}_1^{(y)}(\boldsymbol{C}_y' \otimes \boldsymbol{B}_y') + \boldsymbol{E}_y$$
$$\boldsymbol{E}_x = (\varepsilon_{ij}^{(x)}),\ \boldsymbol{E}_y = (\varepsilon_{ij}^{(y)}),\ \varepsilon_{ij}^{(x)} \overset{\text{i.i.d.}}{\sim} N(0, sd^2),\ \varepsilon_{ij}^{(y)} \overset{\text{i.i.d.}}{\sim} N(0, sd^2).$$

We set $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{X}}$ under two conditions. Under the first condition, we generate $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ as follows:

$$x_{ij_xk_x} = \begin{cases} 1 & (x_{ij_xk_x} \leq \text{Quantile}(\boldsymbol{x}_{j_xk_x}, 0.25)) \\ 2 & (\text{Quantile}(\boldsymbol{x}_{j_xk_x}, 0.25) < x_{ij_xk_x} \leq \text{Quantile}(\boldsymbol{x}_{j_xk_x}, 0.45)) \\ 3 & (\text{Quantile}(\boldsymbol{x}_{j_xk_x}, 0.45) < x_{ij_xk_x} \leq \text{Quantile}(\boldsymbol{x}_{j_xk_x}, 0.85)) \\ 4 & (\text{Quantile}(\boldsymbol{x}_{j_xk_x}, 0.85) < x_{ij_xk_x}) \end{cases}$$

$$(i = 1,\ 2,\ \ldots,\ n; j_x = 1,\ 2,\ \ldots,\ 15; k_x = 1,\ 2,\ \ldots,\ 16),$$

$$y_{ij_yk_y} = \begin{cases} 1 & (y_{ij_yk_y} \leq \text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.2)) \\ 2 & (\text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.2) < y_{ij_yk_y} \leq \text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.4)) \\ 3 & (\text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.4) < y_{ij_yk_y} \leq \text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.6)) \\ 4 & (\text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.6) < y_{ij_yk_y} \leq \text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.8)) \\ 5 & (\text{Quantile}(\boldsymbol{y}_{j_yk_y}, 0.8) < y_{ij_yk_y}) \end{cases}$$

$$(i = 1,\ 2,\ \ldots,\ n; j_y = 1,\ 2,\ \ldots,\ 16; k_y = 1,\ 2,\ \ldots,\ 12),$$

where $\boldsymbol{x}_{j_xk_x}$ is the I-dimensional vector of variable j_x under condition k_x of $\underline{\boldsymbol{X}}$, $\boldsymbol{y}_{j_yk_y}$ is the I-dimensional vector of variable j_y under condition k_y of $\underline{\boldsymbol{Y}}$, and quantile $(\boldsymbol{x}, h)$ is the function returning the h-quantile of $\boldsymbol{x}$. We set the number of objects as 300 and 500, and the standard deviation sd of noise as 0.1 and 0.3.

Under the second case, the dividing rule by using the quantile is the same. However, we set the value as random. Therefore, the variable does not retain the order of score.

We set the number of dimensions of $\boldsymbol{B}_x$, $\boldsymbol{B}_y$, $\boldsymbol{C}_x$, and $\boldsymbol{C}_y$ as 3. For two-mode two-way analysis, we set the number of dimensions $\boldsymbol{A}_x$ and $\boldsymbol{A}_y$ as nine, because there are nine dimensions of $\boldsymbol{C}_x \otimes \boldsymbol{B}_x$ and $\boldsymbol{C}_y \otimes \boldsymbol{B}_y$.

For each estimator, we calculate the mean of squared error, defined as follows:

$$\frac{1}{R}\sum\left(\|\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x - \boldsymbol{C}_x \otimes \boldsymbol{B}_x\|^2 + \|\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y - \boldsymbol{C}_y \otimes \boldsymbol{B}_y\|^2\right),$$

where R is iteration times. We set the reputation time R as 50. When we evaluate the mean squared error for the two-mode two-way method, we set $\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x = \hat{\boldsymbol{A}}_x$ and $\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y = \hat{\boldsymbol{A}}_y$.

4.2.1.2 Result

Figures 4.6 and 4.7 show boxplots of settings 1 and 2, respectively. The boxplots show, from the left, three-mode three-way categorical canonical covariance using the K-means type constrained connector, three-mode three-way categorical canonical covariance setting with the connector matrix as $\boldsymbol{I}$, two-mode two-way categorical canonical covariance analysis, three-mode three-way canonical covariance using the K-means type constrained connector matrix, three-mode three-way canonical covariance setting with the connector matrix as $\boldsymbol{I}$, and two-mode two-way categorical canonical covariance analysis. From Figures 4.6 and 4.7, the categorical canonical covariance method constrained K-means connector matrix has the best result under all conditions and all settings. On the other hand, except for the categorical canonical covariance method constrained K-means connector matrix method, the results of the three-mode three-way method are not different. Moreover, the method, assuming all values are numerical, is better than that based on the NPCA method. The reason for this result is that the number of categories is the same and the cutting rule is the same. Therefore, the metric canonical covariance method could complete the variances.

Tables 4.4 and 4.5 show the mean squared error under settings 1 and 2, respectively. Tables 4.4 and 4.5 show, from the top, three-mode three-way categorical canonical covariance using K-means type constrained connector, three-mode three-way categorical canonical covariance setting with the connector matrix as $\boldsymbol{I}$, two-mode two-way categorical canonical covariance analysis, three-mode three-way canonical covariance using K-means type constrained connector matrix, three-mode three-way canonical covariance setting with the connector matrix as $\boldsymbol{I}$, and two-mode two-way categorical canonical covariance analysis. From Tables 4.4 and 4.5, standard deviation of the two-way method is smaller than that of the other method, because the solution of the two-mode two-way method is simple. On the other hand, the standard deviation of the qualification method for three-mode three-way data is bigger than that for the other method. One reason for this result is that the qualification method has many local minimums. The qualification method for three-mode three-way data considers many relationships between interactions of variable and conditions. Moreover, the qualification method has many parameters. This is the drawback of qualification method for three-mode three-way data.

From the above results, when data have three-mode three-way structure and categorical variables, the quantification method is more suitable for analysis than the two-mode two-way and non-quantification methods are.

Table 4.4: Mean squared error of setting 1: Transform the order variable

	$n = 300$		$n = 500$	
	$sd = 0.1$	$sd = 0.3$	$sd = 0.1$	$sd = 0.3$
Non-metric connect three-way	7.507(3.322)	7.548(3.162)	8.294(3.941)	7.629(3.532)
Non-metric three-way	11.459(3.446)	11.428(3.374)	12.167(3.509)	11.328(3.249)
Non-metric two-way	26.388(1.593)	26.351(1.672)	26.402(1.291)	26.281(1.435)
Metric connect three-way	11.388(2.987)	11.540(2.960)	12.036(2.470)	11.302(2.358)
Metric three-way	11.841(2.890)	11.247(3.092)	10.796(3.071)	11.443(2.844)
Two-way	26.725(1.044)	26.395(0.977)	26.302(0.974)	26.072(0.768)

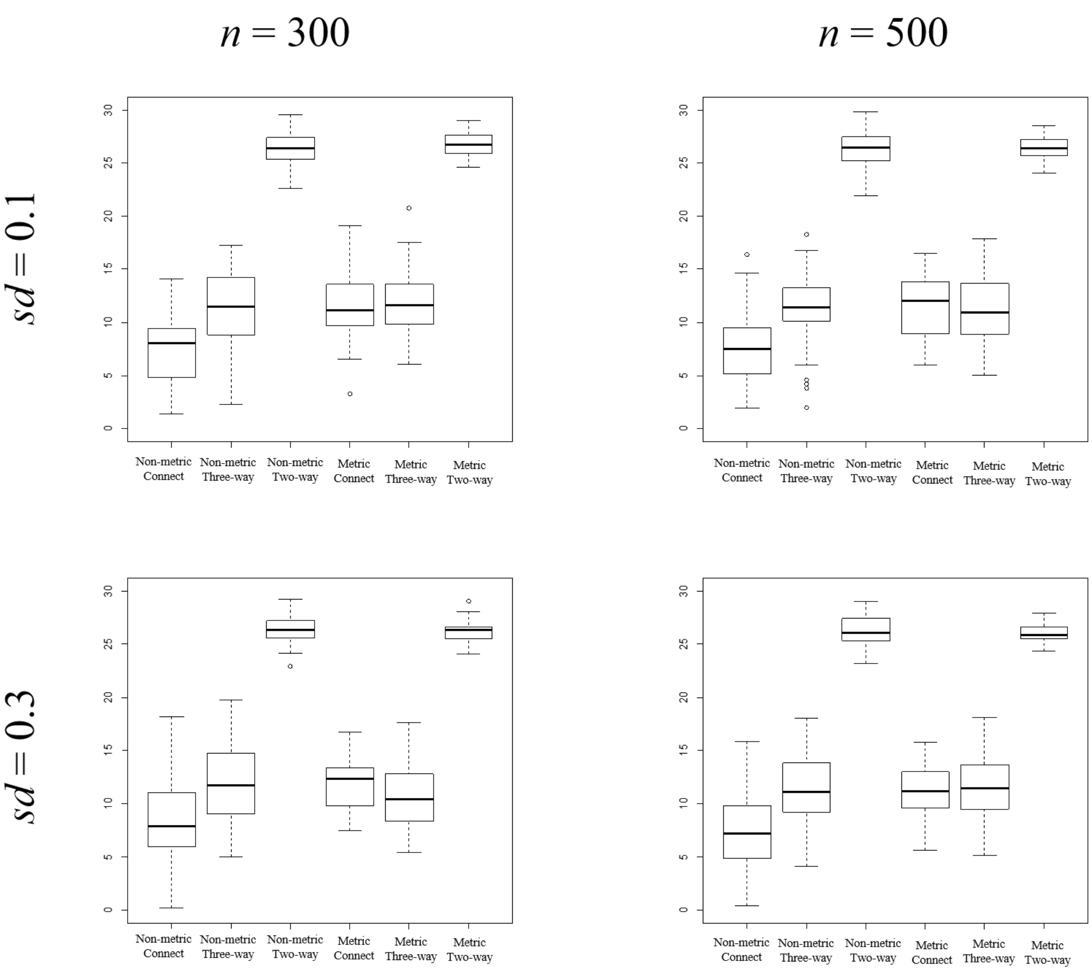


Figure 4.6: Squared error of parameters under setting 1

Table 4.5: Mean squared error of setting 2: Transform the categorical variable

	$n = 300$		$n = 500$	
	$sd = 0.1$	$sd = 0.3$	$sd = 0.1$	$sd = 0.3$
Non-metric connect three-way	7.482(3.263)	7.613(3.078)	7.031(3.548)	8.038(4.007)
Non-metric three-way	11.303(3.299)	10.919(3.048)	11.482(2.992)	11.823(3.020)
Non-metric two-way	26.238(1.775)	26.108(1.389)	26.307(1.568)	26.001(1.297)
Metric connect three-way	11.431(2.594)	12.781(3.137)	11.866(2.751)	12.758(2.911)
Metric three-way	11.282(2.901)	14.582(4.117)	12.355(2.950)	14.448(3.282)
Two-way	26.300(1.136)	26.102(0.724)	26.695(1.010)	26.224(0.628)

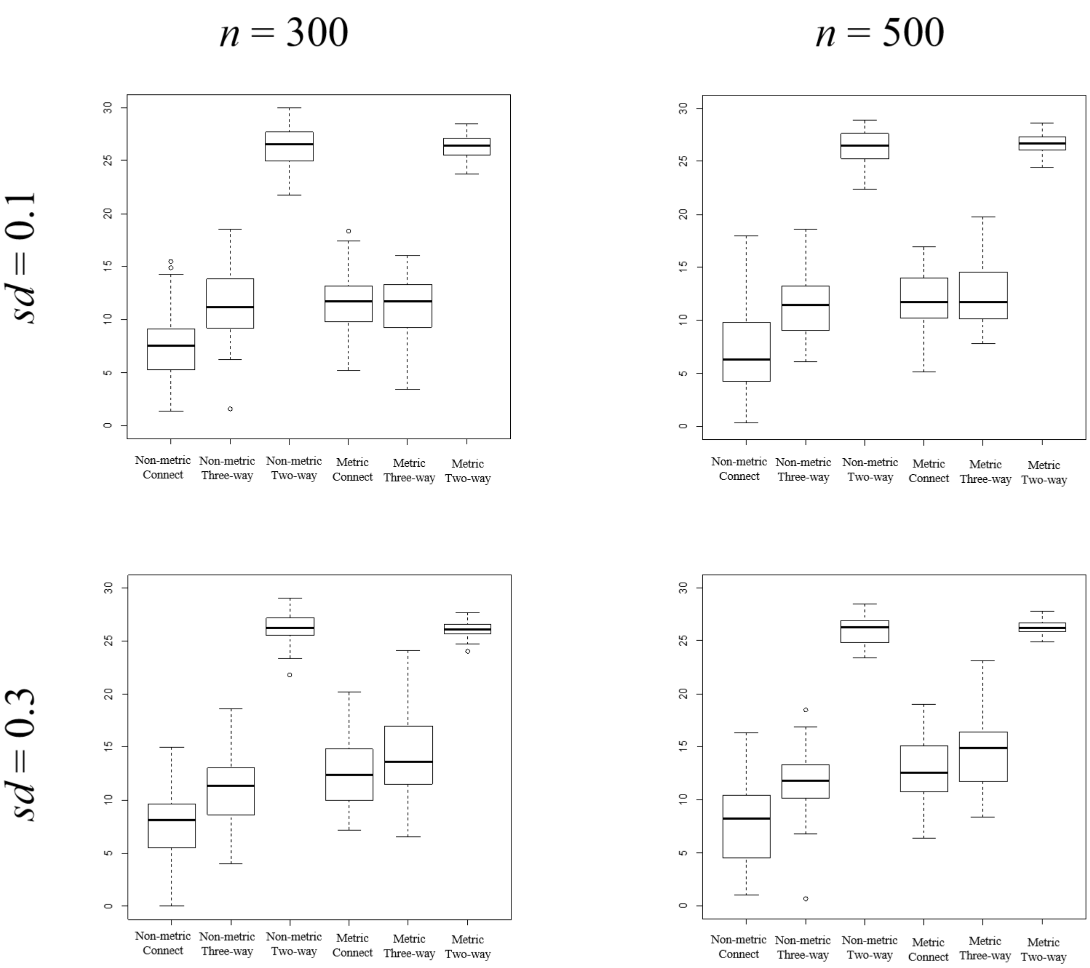


Figure 4.7: Squared error of parameters under setting 2

Chapter 5

Real data examples

5.1 UNESCO and World Bank data

In this section, we describe applications of the three-mode three-way canonical covariance constrained connector method. We use the same data for the real data example of the K-means type and the regression type constrained methods. First, we explain the data abstraction. Then, we describe the results of the K-means type and regression type constrained methods. We propose some hypotheses about these data. First, education and economic indexes are related to each other. Second, some factors exist for each data, such as expenditure for the education factor. Applying the proposed method and the Tucker 3 method to these data sets, we can compare the proposed method with Tucker 3 method.

5.1.1 Data description

In this subsection, we describe the abstraction of data used for the real data example. We use two data sets from different sources. The first source is the United Nations Educational, Scientific and Cultural Organization (UNESCO) (`http://data.uis.unesco.org/`). UNESCO data are survey data of an education index of developing countries. The other source is economic data from World Bank open data (`https://data.worldbank.org/`).

Objects correspond to countries. There are 42 countries, because we select the countries that have no missing values. The data from UNESCO and World Bank open data have 14 and 4 conditions, respectively. Conditions correspond to the survey periods. The survey period of UNESCO data is from 2000 to 2013. The survey period of World Bank open data is 1990, 2000, 2010, and 2012. Variables of these data sets are shown in Tables 5.1 and 5.2, respectively. UNESCO and World Bank open data have six and four variables, respectively.

Before applying the method to the data sets, we make model 1 column-wise standardized

Table 5.1: Variables of UNESCO data

Variable	Description	For shortn
GDP rate on education	Government expenditure on education, total GDP	GDPoE
Years of education	Mean of years schooling in population 25+ years	YoE
Government expenditure rate on education	Ratio of education expenditure and total government expenditure	GERoE
Literacy	Literacy rate of population 15–24 years	Lit
Rate of out-of-school adolescents of secondary school age	Rate of out-of-school people (aged 13—18 years) in secondary school age	RoOAoS
Rate of out-of-school children of primary school age	Rate of out-of-school children in primary school days	RoOCoP

Table 5.2: Variables of World Bank open data

Variable	Description	For short
Access electricity	Access to electricity % of population	AE
GDP per capita	GDP per capita (current US $)	GDPpC
Child mortality rate	Mortality rate under 5 years (per 1000 live births)	CMR
Average life span	Life expectancy at birth (years)	ALS

data. In other words,

$$\sum_{i=1}^{I} x_{ijk} = 0,\ \frac{1}{I}\sum_{i=1}^{I} x_{ijk}^2 = 1$$
$$(j = 1, 2,\ \cdots,\ J;\ k = 1,\ 2, \cdots,\ K)$$

is satisfied. This standardized method is the same as the two-mode two-way case. The objective functions of the K-means type and regression constrained type constrained methods are defined by using the matrix unfolding model. This is the reason we choose this standardized method.

5.1.2 Result of K-means type constrained

We set the numbers of dimensions for parameters as follows: for UNESCO variables $r_{bx} = 3$, for UNESCO conditions $r_{cx} = 4$, for World Bank variables $r_{by} = 2$, for World Bank conditions $r_{cy} = 2$, the number of connections for variables $c_b = 1$, and the number of connections for conditions $c_c = 1$. The numbers of dimensions for UNESCO variables $r_{bx} = 3$ because we hypothesize the existence of three factors, one of which is a common factor, and the others are expenditure and education factors. The World Bank variables $r_{by} = 2$, because we hypothesize the existence of two factors, one of which is a common factor, and the other a general factor. Hence, we choose the number of connections for variables $c_b = 1$. The numbers of dimensions for UNESCO conditions $r_{cx} = 4$, because we hypothesize that there is one common factor as well as three factors for UNESCO data. One of these three factors is a general factor. The others are before and after the 2008 financial crisis. The numbers of dimensions for World Bank conditions $r_{cy} = 2$, because we hypothesize that there are two factors, one of which is a common factor and the other a general factor.

Tables 5.3 and 5.4 show the weight matrixes for variables of each data set. We interpret low dimensional space by the weight matrix of variables. First, we describe UNESCO data. GDPoE and GERoE are higher values among Dim1. Thus, we refer to Dim1 as expenditure for the education factor. GERoE and RoOCoP take higher absolute values among Dim2. The value of RoOCoP is negative. However, higher RoOCoP means that many children of primary school age do not attend school. Thus, when the score of Dim2 is higher, the country has large expenditure on education and many primary school children attend school. On the other hand, RoOAoS takes a small absolute value. Thus, we refer to Dim2 as government expenditure on the primary school factor. Absolute values of all variables are higher among Dim3. Lit and YoE are positive values. On the other hand, other variables have negative values. Thus, when the score of Dim3 takes a higher value, the country shows a good value for its education index. On the other hand, the country spends less money on education than it does on other requirements. Thus, we refer to Dim 3 as the schooling education factor.

Then, we interpret the weight matrixes for World Bank open data. All variables take high absolute values in Dim1. AE, GDPpC, and ALS have negative values. On the other hand, CMR has a positive value. CMR is the ratio of child mortality. A high CMR indicates that the country has poor child mortality. Therefore, Dim1 is referred to as a general index. GDPpC takes the highest absolute value among Dim2. Therefore, Dim2 is referred to as the GDP factor.

Tables 5.5 and 5.6 show the connector matrixes of the UNESCO and World Bank open data, respectively. From Tables 5.5 and 5.6, we understand that government expenditure on the primary school factor and the GDP factor is connected. This result shows that GDP and government expenditure are correlated. This result is self-evident because GDP includes government expenditure.

The goodness of fit for the K-means based method is now defined as follows:

$$\frac{\|\boldsymbol{X}_1(\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x)\|^2 + \|\boldsymbol{Y}_1(\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y)\|^2 + 2\mathrm{tr}(\hat{\boldsymbol{D}}_x'(\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y)\hat{\boldsymbol{D}}_y)}{2(\|\boldsymbol{X}_1\|^2 + \|\boldsymbol{Y}_1\|^2)}$$

This goodness of fit is taken in the range from 0 to 1 from the facts that $\|\boldsymbol{X}_1(\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x)\|^2 \leq \|\boldsymbol{X}_1\|^2$, $\|\boldsymbol{Y}_1(\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y)\|^2 \leq \|\boldsymbol{Y}_1\|^2$, and $2\mathrm{tr}(\hat{\boldsymbol{D}}_x'(\hat{\boldsymbol{C}}_x \otimes \hat{\boldsymbol{B}}_x)'\boldsymbol{X}_1'\boldsymbol{Y}_1(\hat{\boldsymbol{C}}_y \otimes \hat{\boldsymbol{B}}_y)\hat{\boldsymbol{D}}_y) \leq \|\boldsymbol{X}_1\|^2 + \|\boldsymbol{Y}_1\|^2$ hold when $c_c = c_b = 1$. Therefore, this goodness of fit is not a general index but a special case index. In this result, the goodness of fit is very good at 0.995.

Tables 5.7 and 5.8 show the weight matrixes for conditions of the UNESCO data and World Bank open data, respectively. Table 5.7 shows that Dim1, Dim2, and Dim4 are the factors between 2005 and 2008, between 2009 and 2013, and between 2000 and 2002, respectively, because these years take higher absolute values in each dim. Table 5.7 shows that Dim2 displays general factors of the survey years because all years take high values.

From Tables 5.9 and 5.10, which show the connector matrixes for conditions of the UNESCO data and World Bank open data, Dim3 of the UNESCO data and Dim1 of the World bank open data are connected. 2000, 2003, 2004, and 2008 have higher absolute values in Dim3 of the UNESCO data. On the other hand, 2000 has higher value in Dim1 of the World Bank open data. It is easy to interpret this as the same survey year taking the higher value. On the other hand, 2003 and 2004 take negative values in the UNESCO data. We guess that the reason for this result is the Iraq War of 2003. This data set includes 42 developing countries in East and South Asia, Africa, and Central and South America. Therefore, we guess these countries had unstable political situations in these years.

Figures 5.1 and 5.2 show the network representation of weight matrixes for variables and conditions, respectively. Positive and negative are represented as blue and red, respectively. The width of lines and alpha value show the absolute value. A deep wide line shows that the absolute value is higher. When the variable and dimension are not connected by any lines, the absolute value is smaller than 0.2. From these figures, we distinguish which factors are common factors at a glance.

Table 5.3: Weight matrix for variables of UNESCO data

Variable	Dim1	Dim2	Dim3
GDPoE	0.618	−0.459	−0.391
YoE	0.254	−0.342	0.435
GERoE	0.513	0.534	−0.440
Lit	0.321	−0.068	0.406
RoOAoS	−0.371	0.051	−0.366
RoOCoP	−0.223	−0.616	−0.408

Table 5.4: Weight matrix for variables of World Bank open data

	Dim1	Dim2
AE	−0.498	−0.329
GDPpC	−0.475	−0.878
CMR	0.512	−0.238
ALS	−0.514	−0.254

Table 5.5: Connector matrix for variables of UNESCO data

	Connect
Dim1	0
Dim2	1
Dim3	0

Table 5.6: Connector matrix for variables of World Bank open data

	Connect
Dim1	0
Dim2	1

Table 5.7: Weight matrix for conditions of UNESCO data

Year	Dim1	Dim2	Dim3	Dim4
2000	0.054	−0.245	0.554	0.442
2001	0.018	−0.177	0.173	0.511
2002	−0.008	−0.167	−0.071	0.438
2003	−0.156	−0.112	−0.455	0.297
2004	−0.201	−0.121	−0.450	0.197
2005	−0.328	−0.108	−0.172	0.153
2006	−0.451	−0.113	−0.035	−0.012
2007	−0.461	−0.147	0.049	−0.117
2008	−0.420	−0.223	0.411	−0.211
2009	−0.132	−0.301	−0.034	−0.207
2010	−0.056	−0.341	−0.034	−0.246
2011	0.205	−0.429	0.026	−0.145
2012	0.359	−0.468	−0.203	−0.132
2013	0.200	−0.385	−0.050	−0.025

Table 5.8: Weight matrix for conditions of World Bank open data

Year	Dim1	Dim2
1990	−0.071	0.499
2000	0.786	0.561
2010	−0.419	0.469
2012	−0.450	0.465

Table 5.9: Connector matrix for variables of UNESCO data

	Connect
Dim1	0
Dim2	0
Dim3	1
Dim4	0

Table 5.10: Connector matrix for variables of World Bank open data

	Connect
Dim1	1
Dim2	0

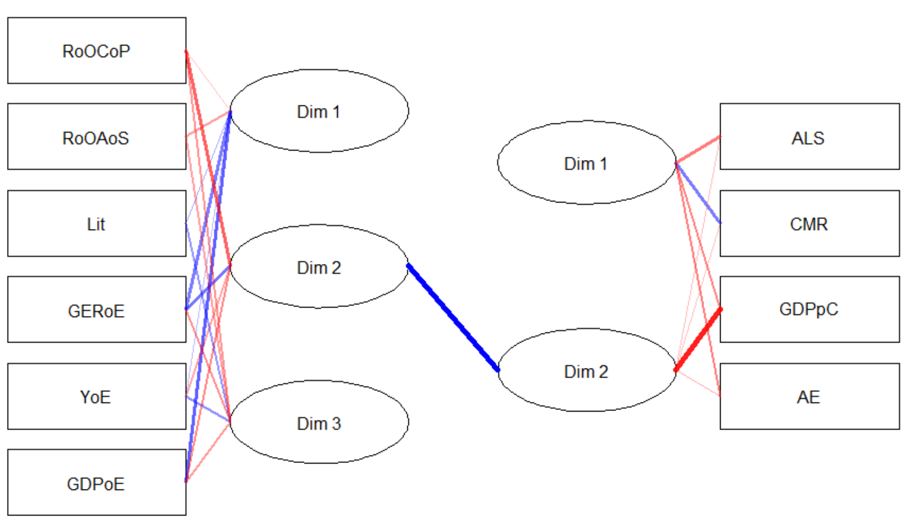


Figure 5.1: The network representation of weight matrixes for variables

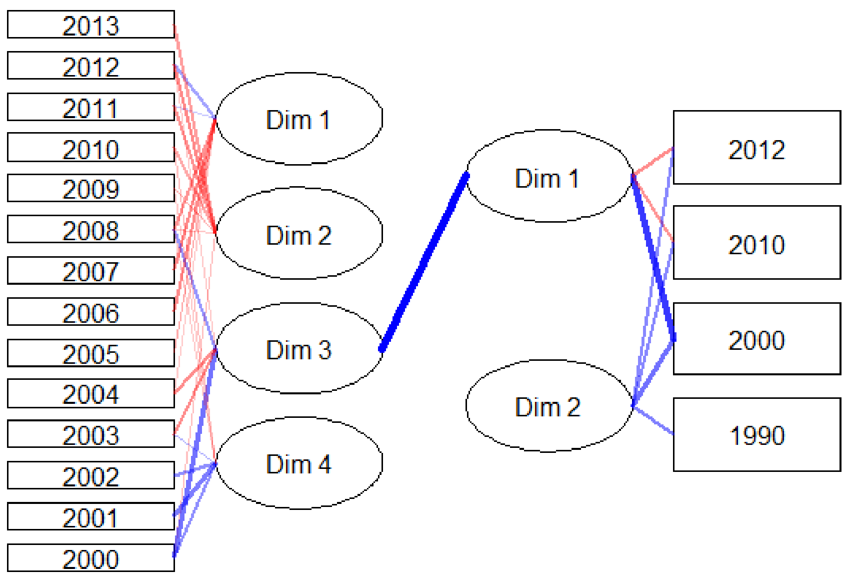


Figure 5.2: The network representation of weight matrixes for conditions

We also apply the Tucker 3 method to these data sets for comparison with our proposed method. The number of dimensions are the same as for the K-means type. Tables 5.11 and 5.12 show the weight matrixes for variables of UNESCO and the World Bank, respectively. From Table 5.11, YoE, Lit, RoOAoS, and RoOCoS have higher absolute values among Dim 1. On the other hand, GDPoE and GERoE have higher absolute values among Dim 2. Therefore, we obtain education and expenditure factors from the Tucker 3 method. Dim 3 is similar to Dim 2 of the K-means based method. From Table 5.12, we obtain general and expenditure factors. These results are very similar to the K-means based method. However, we cannot determine which factors are related to other factors. In this regard, the K-means based method is better.

Tables 5.14 and 5.13 show the condition matrixes for variables of the UNESCO and World Bank, respectively. We interpret the results from Table 5.14 to mean that Dim 1 is a general factor. However, the other factors are not clearly divided. Although Tucker 3 obtains a general factor, interpretation of the other factor is more difficult than in the K-means based method. From Table 5.13, we obtain similar factors to the K-means based method.

Figures 5.3 and 5.4 show the network representation of weight matrixes for variables and conditions, respectively. Positive and negative are represented as blue and red, respectively. The width of lines and alpha values show the absolute values. A deep wide line shows that absolute value is higher. When the variable and dimension are not connected by any lines, the absolute value is smaller than 0.2. From these figures, we obtain which factors are similar to those of the K-means based method at a glance.

The Tucker 3 method is one of the dimensional reduction methods for a data set. Therefore, it is difficult to interpret which is a common factor. On the other hand, by using the K-means based method, it is easy to interpret which is common factor. This is the advantage of the K-means based method.

Table 5.11: Weight matrix for variables of UNESCO data (Tucker 3)

Variable	Dim1	Dim2	Dim3
GDPoE	−0.036	−0.721	0.489
YoE	0.499	0.010	0.346
GERoE	−0.110	−0.676	−0.534
Lit	0.514	−0.072	0.067
RoOAoS	−0.500	0.134	−0.061
RoOCoS	−0.473	0.002	0.590

Table 5.12: Weight matrix for variables of World Bank open data (Tucker 3)

	Dim1	Dim2
AE	−0.517	0.286
GDPpC	−0.422	−0.905
CMR	0.525	−0.219
ALS	−0.528	0.226

Table 5.13: Weight matrix for conditions of World Bank open data (Tucker 3)

Year	Dim1	Dim2
1990	−0.503	0.053
2000	−0.493	−0.837
2010	−0.503	0.367
2012	−0.501	0.403

Table 5.14: Weight matrix for conditions of UNESCO data (Tucker 3)

Year	Dim1	Dim2	Dim3	Dim4
2000	−0.261	−0.525	−0.116	−0.453
2001	−0.264	−0.487	−0.243	−0.083
2002	−0.266	−0.371	0.013	0.357
2003	−0.269	−0.139	0.316	0.393
2004	−0.270	0.021	−0.158	0.019
2005	−0.271	−0.018	0.332	0.225
2006	−0.269	0.114	0.348	0.070
2007	−0.270	0.179	0.286	−0.296
2008	−0.270	0.164	0.283	−0.538
2009	−0.269	0.261	−0.171	−0.046
2010	−0.266	0.332	−0.229	−0.009
2011	−0.266	0.153	−0.037	0.243
2012	−0.265	0.224	−0.565	0.110
2013	−0.265	0.073	−0.084	0.001

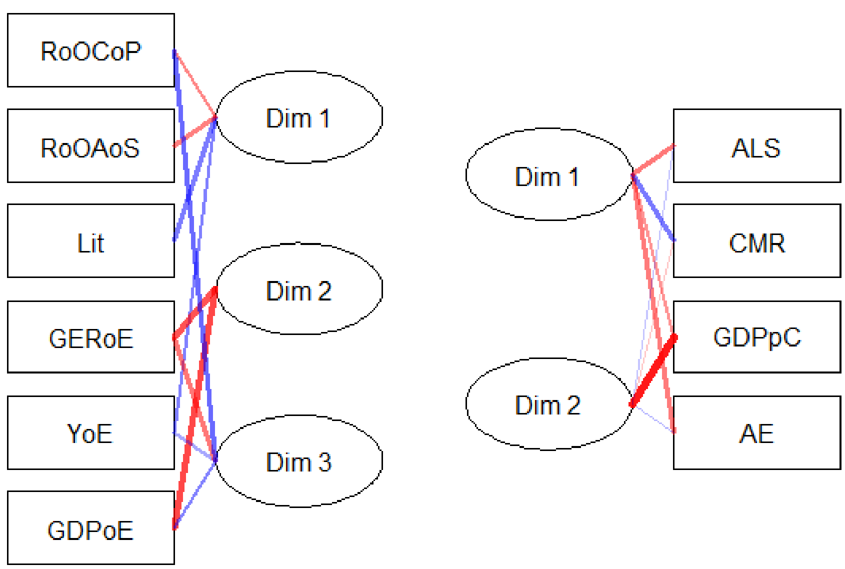


Figure 5.3: The network representation of weight matrixes for variables (Tucker3)

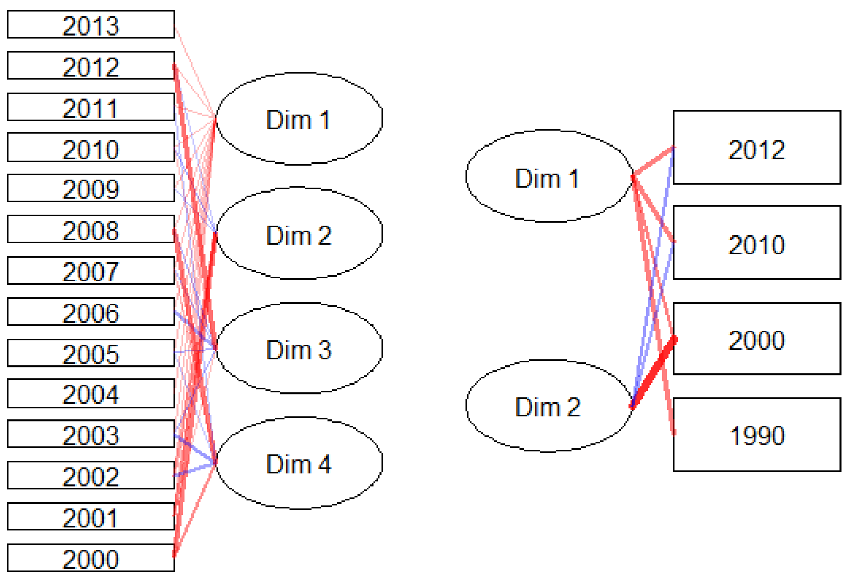


Figure 5.4: The network representation of weight matrixes for conditions (Tucker3)

5.1.3 Result of regression type constrained case

We set the response data sets as World Bank open data, and the explanatory data sets as UNESCO data. We set the numbers of dimensions for parameters as for UNESCO variables $r_{bx} = 2$, and as for UNESCO conditions $r_{cx} = 2$. The numbers of dimensions for UNESCO variables $r_{bx} = 2$ and for UNESCO conditions $r_{cx} = 2$, because we hypothesize that there is one general factor and one characteristic factor. GDPoE and GERoE are variables corresponding to expenditure. Therefore, we guess that the expenditure variables of UNESCO are strong in relation to GDPpC. the survey years of UNESCO are the same as those of the World Bank. Therefore, we guess that there are general factors as well as a common year factor.

Tables 5.15 and 5.16 show the weight matrixes for variables and conditions of the UNESCO data. From Table 5.15, we regard Dim1 as the education factor, because YoE and Lit take positive and higher absolute values, and RoOAoS and RoOAos take negative and higher absolute values. On the other hand, Dim2 is referred to as expenditure for the education factor, because GDPoE and GERoE take higher absolute values.

From Table 5.16, Dim1 is referred to as the general factor, because all years take similar values. On the other hand, it is difficult to interpret Dim2. The years from 2000 to 2002 take negative and higher absolute values. However, The years from 2006 to 2008 take positive and higher absolute values. We guess that Dim2 shows that policy of countries changed until the Iraq war began. Moreover, 2008 began with the Gaza War, and we guess this is the reason the years 2006 to 2008 take positive and higher absolute values.

Table 5.15: Weight matrix for variables of UNESCO data

Variable	Dim1	Dim2
GDPoE	−0.03	−0.72
YoE	0.50	0.01
GERoE	−0.11	−0.68
Lit	0.51	−0.07
RoOAoS	−0.50	0.13
RoOCoP	−0.47	0.00

Tables 5.17 and 5.18 show the coefficient matrixes for variables and conditions of the UNESCO data. The rows of Tables 5.17 and 5.18 correspond to the variables and conditions of the World Bank data, respectively. From Table 5.17, the education factor corresponds to all variables except CMR. However, AE and GDPpC take negative values. Expenditure is smaller when the education score is larger. This result is similar to the K-means type constrained case. The coefficient vector for the expenditure factor is in contrast to that of the education factor. When the score of the expenditure factor is larger, then AE and GDPpC tend to take higher values. This result shows that a country that spends substantial money on education has enough to developing its economy among the objects.

From Table 5.18, the coefficient matrixes of Dim1 and Dim2 have the same tendency.

Table 5.16: Weight matrix for conditions of UNESCO data

Year	Dim1	Dim2
2000	−0.26	−0.39
2001	−0.26	−0.48
2002	−0.27	−0.41
2003	−0.27	−0.13
2004	−0.27	−0.10
2005	−0.27	0.13
2006	−0.27	0.23
2007	−0.27	0.34
2008	−0.27	0.39
2009	−0.27	0.18
2010	−0.27	0.21
2011	−0.27	0.03
2012	−0.26	−0.03
2013	−0.26	0.00

However, the scale of the coefficient vector for Dim1 and Dim2 is different. It is difficult to interpret this difference as corresponding to important of exploration because the scales of the scores of Dim1 and Dim2 are different. Thus, now, we interpret this result as lack of interaction between the years of the UNESCO and World Bank data. In other words, there are no specific years in which the absolute value is large.

Figures 5.5 and 5.6 show the network representation. Red and blue show negative and positive value, respectively. The width of lines and arrows show the absolute values of the weight and coefficient matrixes, respectively. When the variable and dimension are not connected by any line, the absolute value is smaller than 0.2. On the other hand, the width of the arrows is relatively evaluated. From Figures 5.5 and 5.6, it is easy to interpret the same results as shown in from the tables.

Table 5.17: Coefficient matrix for variables

	Dim1	Dim2
AE	−0.03	0.04
GDPpC	−0.02	0.04
CMR	0.00	0.01
ALS	0.02	−0.04

Now, we compare this result with that of the Tucker 3 method. The PLS' Dim 1 and Dim 2 for variables is similar to the Tucker's Dim 1 and Dim 2. On the other hand, PLS' Dim 2 for conditions is unlike any Dim of Tucker's, although the PLS' Dim 1 for conditions is similar to Tucker's Dim 1. The reason for this result is that PLS maximizes

Table 5.18: Coefficient matrix for conditions

Year	Dim1	Dim2
1990	1.47	−8.03
2000	1.19	−7.18
2010	1.35	−7.29
2012	1.26	−7.26

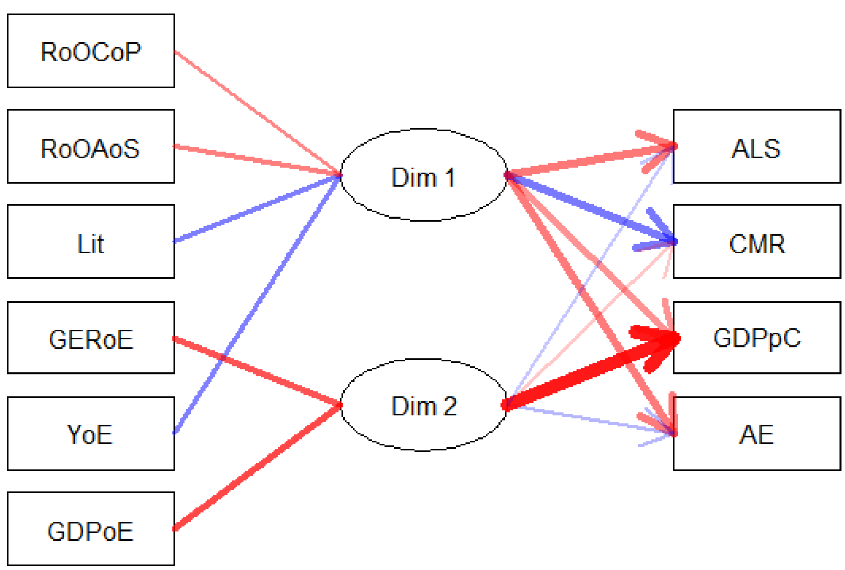


Figure 5.5: The network representation of weight and coefficients matrixes for variables

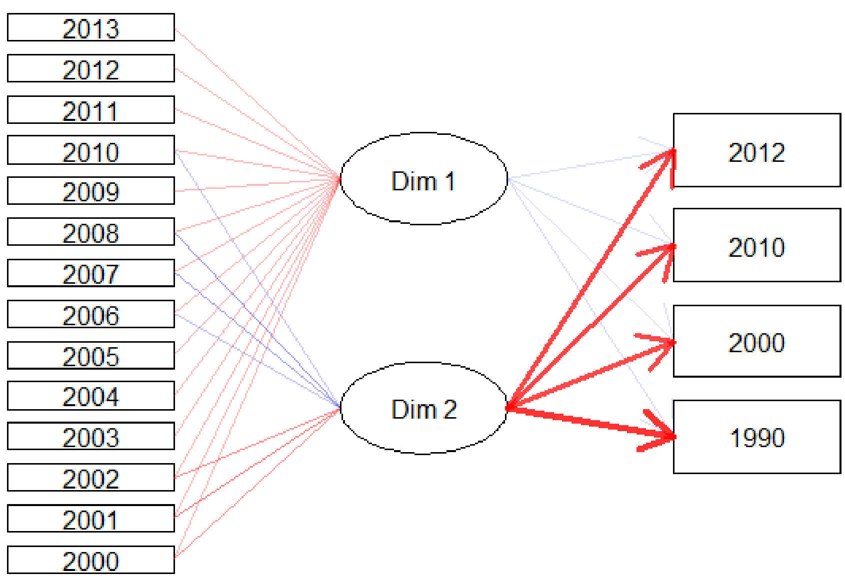


Figure 5.6: The network representation of weight and coefficients matrixes for conditions

the covariance between data sets.

5.2 Cognitive and purchase survey for beer and malt-free beer-like alcoholic beverage

In this section, we describe the result of applying the NPCA-based method to a cognitive and purchase survey for beer and beer-like drinks surveyed by Nomura Research Institute, Ltd. These data are provided by Nomura Research Institute, Ltd. for the Marketing Analysis Contest 2017 (`https://www.is.nri.co.jp/contest/`). First, we explain the data abstraction. Then, we describe the result by the NPCA-based method. We propose some hypotheses for these data. The first is that cognitive and purchase behavior are related to each other. The second is that some factors exist for each data, such as a beer factor and a malt-free beer-like alcoholic beverage factor. Applying the proposed method and the Tucker 3 method with the quantification method to these data sets, we compare the proposed method by the results.

5.2.1 Data description

In this subsection, we describe the data abstraction used as the real data example. These data are questionnaire survey data about cognitive attitude and purchase behavior. Each survey investigates the same participants twice using the same questionnaire. Pre- and post-surveys are conducted in February and March 2017. There are 3000 survey participants. Before applying the NPCA-based method, we delete participants who have missing values. Then, the number of objects declines to 2639. The question in the questionnaire related to cognitive attitude is "Do you want to purchase item x?" This question is measured on a 5-point scale. On the other hand, the questions related to purchase behavior are "Do you know item x? How often do you drink item x in one month?" This question is measured on a 5-point scale. Table 5.19 shows the list of 11 items of this survey.

5.2.2 Result of NPCA-based method

In this subsection, we describe the result applying the NPCA-based method. We set the numbers of dimensions for the parameters as follows: for cognitive attitude of item $r_{bx} = 3$, for cognitive attitude of item of conditions $r_{cx} = 2$, for purchase behavior of item $r_{by} = 3$, for purchase behavior of item of conditions $r_{cy} = 2$, the number of connections for variables $c_b = 1$, and the number of connections for conditions $c_c = 1$. The number of dimensions for variables is set to 3, because we hypothesize that there are three factors, one of which is a common factor. The others are beer and malt-free beer that tastes like an alcoholic beverage. Hence, we choose the number of connectors for the variable as 1. The reason the number of dimensions for conditions is set as 2 is that we are not interested in the difference between conditions. The reason that the number of connections for conditions is set as 1 is that the number of conditions as 2.

Table 5.19: Item list of beer and malt-free beer-like alcoholic beverages

Name	Category
Super Dry	Beer
The Premium Malts	Beer
KIRIN ICHIBAN	Beer
YEBISU Beer	Beer
The MALT'S	Beer
LAGER BEER	Beer
Sapporo Black Label	Beer
Mugi To Hop the Gold	Malt-free beer-like alcoholic beverage
Kin Mugi	Malt-free beer-like alcoholic beverage
Nodo Goshi Nama	Malt-free beer-like alcoholic beverage
Clear Asahi	Malt-free beer-like alcoholic beverage

The goodness of fit for the NPCA-based method is now not defined, because the score $\boldsymbol{F}_{(x)}$ and $\boldsymbol{F}$ is not constrained. Moreover, we do not have an explicit optimal estimator. Although we could use goodness of fit for the method without quantification like we do for the quantification method when we set $\boldsymbol{X}_1$ as $\boldsymbol{Z}_1^{(x)}$, we do not determine that the range of the value is from 0 to 1. This is left for future work.

Tables 5.20 and 5.21 show the weight matrixes for cognitive attitude and purchase behavior of an item, respectively. From Table 5.20, Dim1 of cognitive attitude is called the beer factor, because the items belonging to beer take large absolute values. On the other hand, Dim2 of cognitive attitude is called the malt-free beer-like alcoholic beverage factor, because the items belonging to malt-free beer-like alcoholic beverages take large absolute values. From Table 5.21, Dim2 of purchase behavior is called the beer factor, because the items belonging to beer take large absolute values. On the other hand, Dim3 of purchase behavior is called the malt-free beer-like alcoholic beverage factor, because the items belonging to malt-free beer-like alcoholic beverages take large absolute values.

Tables 5.22 and 5.23 show the connector matrixes of cognitive attitude and purchase behavior, respectively. From Tables 5.22 and 5.23, we could interpret there being a connection between Dim3 of cognitive attitude and Dim1 of purchase behavior. The MALT'S and The Premium Malts take large absolute values in Dim3 of cognitive attitude. On the other hand, The MALT'S, Super Dry, and the malt-free beer-like alcoholic beverages have large absolute values in the Dim1 of purchase behavior. Super Dry is a famous beer and The MALT'S is cheaper than other beer brands. Moreover, malt-free beer-like alcoholic beverages tend to have lower prices than do beer drinks. On the other hand, The Premium Malts is one of the most expensive beer brands in Japan. Therefore, we interpret this result as a good cognitive attitude toward cheaper and higher priced beer, although the purchase behavior tends to favor malt-free beer-like alcoholic beverages.

Tables 5.24 and 5.25 show the weight matrixes for conditions. In both data sets, Dim 1 and Dim2 correspond to the pre- and post-survey, respectively. Tables 5.26 and 5.27

Table 5.20: Weight matrix for cognitive attitude of item

Item	Dim1	Dim2	Dim3
Super Dry	−0.242	0.358	0.266
The Premium Malts	0.046	−0.262	0.559
KIRIN ICHIBAN	−0.217	0.372	0.113
YEBISU Beer	0.127	−0.312	0.256
The MALT'S	−0.021	0.244	−0.636
LAGER BEER	−0.253	0.390	0.282
Sapporo Black Label	−0.231	0.369	0.174
Mugi To Hop the Gold	0.435	0.232	0.055
Kin Mugi	−0.433	−0.237	−0.099
Nodo Goshi Nama	0.431	0.234	0.090
Clear Asahi	0.442	0.234	0.054

Table 5.21: Weight matrix for purchase behavior of item

Item	Dim1	Dim2	Dim3
Super Dry	−0.309	−0.441	0.080
The Premium Malts	0.060	−0.392	−0.018
KIRIN ICHIBAN	−0.019	−0.453	0.052
YEBISU Beer	−0.082	0.421	−0.001
The MALT'S	0.696	−0.122	−0.209
LAGER BEER	0.176	−0.331	−0.062
Sapporo Black Label	−0.123	0.333	0.103
Mugi To Hop the Gold	−0.335	−0.161	0.553
Kin Mugi	−0.254	−0.054	−0.441
Nodo Goshi Nama	0.379	0.072	0.440
Clear Asahi	−0.208	−0.010	−0.489

Table 5.22: Connector matrix for cognition attitude of item

	Connect
Dim1	0
Dim2	0
Dim3	1

Table 5.23: Connector matrix for purchase behavior of item

	Connect
Dim1	1
Dim2	0
Dim3	0

show the connector matrixes for conditions. From Tables 5.26 and 5.27, the post-survey is connected.

Therefore, the relationship between Dim3 of cognitive attitude and Dim1 of purchase behavior exists in the post-survey research. We cannot say whether the same relationship exists in the pre-survey research.

Figure 5.7 shows the weight matrixes and connection matrixes for variables. Positive and negative are represented as blue and red, respectively. The width of lines and alpha values show the absolute values. Therefore, a deep wide line shows a higher absolute value. When the variable and dimension are not connected by any line, the absolute value is smaller than 0.2. From this figure, we distinguish which factors are common factors at a glance.

Table 5.24: Weight matrix for conditions of cognition attitude of item

	Dim1	Dim2
Pre	−0.967	−0.254
Post	0.254	−0.967

Table 5.25: Weight matrix for conditions of purchase behavior of item

	Dim1	Dim2
Pre	−0.947	0.320
Post	0.320	0.947

Table 5.26: Connector matrix for conditions of cognitive attitude of item

	Connect
Dim1	0
Dim2	1

Table 5.27: Connector matrix for conditions of purchase behavior of item

	Connect
Dim1	0
Dim2	1

We also apply the Tucker 3 method with the quantification method proposed by Nakamura (2015). We set the number of dimensions as the same as those of the NPCA-based method. Tables 5.28 and 5.29 how the result of the weight matrix for cognitive attitude and purchase behavior of item, respectively. From Table 5.28, we interpret Dim 2 and Dim 3 as beer and malt-free beer-like alcoholic beverage factors, respectively. Super Dry has the highest absolute value among Dim 1. Hence, we interpret Dim 1 as the Super Dry factor. These factors are the difference with the NPCA-based method, except for the malt-free beer-like alcoholic beverage factor. The cause of this result is that NPCA has two factors for each data; that is, we set the number of dimensions as 2. From Table 5.29, we interpret that Dim1 and Dim2 are Beer and Malt-free beer-like alcoholic beverage factors, respectively. The Premium Malts and The MALT'S have bigger absolute values among Dim 3. Therefore, we consider Dim 3 a malts factor.

Figure 5.8 shows the weight matrixes and connection matrixes for variables. Positive and negative are represented as blue and red, respectively. The width of lines and alpha values show the absolute values. Therefore, a deep wide line shows that absolute values are higher. When the variable and dimension are not connected by any line, the absolute value is smaller than 0.2. From this figure, it is easy to interpret the same factors from

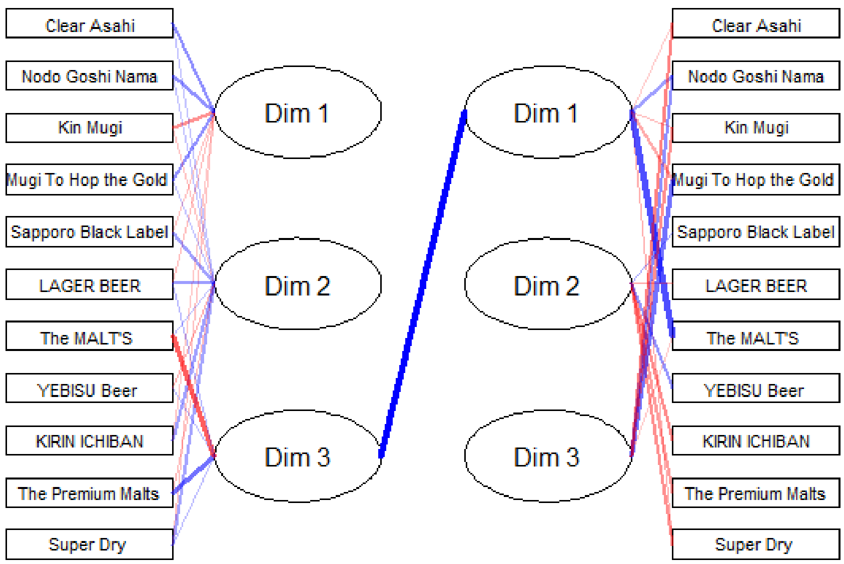


Figure 5.7: The network representation of weight matrixes for variables

the table.

The cause of the difference between the NPCA-based method and Nakamura's method is the malts factor. The NPCA-based method has no malts factor but instead has a cheaper bear and beverage factor. The cheaper bear and beverage factor is a common factor between cognitive attitude and purchase behavior that maximizes the covariance of the data. On the other hand, the malts factor maximizes the variance of the data.

Table 5.28: Weight matrix for cognitive attitude of item (Nakamrua)

Item	Dim1	Dim2	Dim3
Super Dry	0.654	−0.230	−0.323
The Premium Malts	−0.321	−0.208	0.140
KIRIN ICHIBAN	−0.227	−0.320	0.063
YEBISU Beer	−0.333	−0.235	0.198
The MALT'S	−0.222	−0.275	0.024
LAGER BEER	0.077	0.456	0.064
Sapporo Black Label	0.058	−0.609	−0.143
Mugi To Hop the Gold	−0.238	0.063	−0.389
Kin Mugi	−0.045	−0.113	−0.577
Nodo Goshi Nama	−0.216	0.089	−0.457
Clear Asahi	−0.379	0.261	−0.338

Table 5.29: Weight matrix for purchase behavior of item (Nakamrua)

Item	Dim1	Dim2	Dim3
Super Dry	-0.003	-0.372	0.337
The Premium Malts	0.017	0.319	-0.517
KIRIN ICHIBAN	-0.065	-0.447	-0.210
YEBISU Beer	0.076	0.402	-0.313
The MALT'S	-0.061	0.248	-0.528
LAGER BEER	0.019	0.402	0.236
Sapporo Black Label	-0.019	-0.406	-0.209
Mugi To Hop the Gold	0.505	0.021	0.224
Kin Mugi	-0.485	0.048	-0.008
Nodo Goshi Nama	-0.478	0.074	0.209
Clear Asahi	-0.516	0.031	0.074

Table 5.30: Weight matrix for conditions of cognition attitude of item (Nakamrua)

	Dim1	Dim2
Pre	−0.183	−0.983
Post	−0.983	0.183

Table 5.31: Weight matrix for conditions of purchase behavior of item (Nakamrua)

	Dim1	Dim2
Pre	−0.688	−0.726
Post	−0.726	0.688

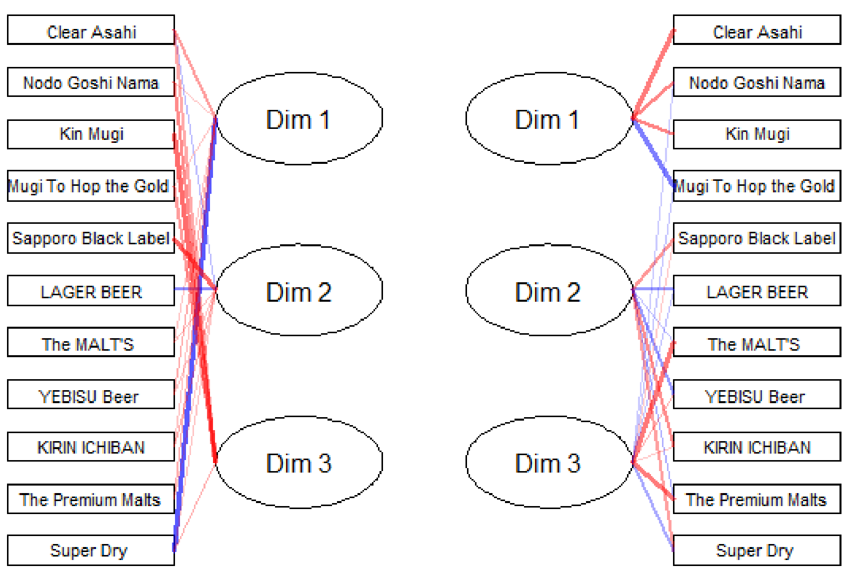


Figure 5.8: The network representation of weight matrixes for variables (Nakamura)

Chapter 6

Conclusion

In this study, we describe the dimensional reduction method for three-mode three-way data based on canonical covariance analysis. Canonical covariance analysis is regarded as simultaneous principal component analysis (PCA) and canonical correlation analysis. Therefore, canonical covariance analysis is a dimensional reduction technique. However, when we apply canonical covariance analysis to three-mode three-way data, we should interpret the same variable under different conditions as a different variable, because canonical covariance analysis has been proposed for two-mode two-way data. We address this problem by extending canonical covariance analysis to three-mode three-way data using the concept of the Tucker model (Tucker, 1966). Since the weight matrixes are represented as the Kronecker product, three-mode three-way canonical covariance considers the same variables under different conditions as the same variable. In general, the numbers of variables and conditions are different between data sets. Thus, the assumption that the number of factors corresponding to the principal component and common factor are the same between data sets is too strict for analyzing three-mode three-way data. Introducing the connector matrix, three-mode three-way canonical covariance could choose a different number of principal components. We describe several constrained connector methods in this paper. When we impose the connector matrix on the K-means type constraint, the update formula of the connector matrix is the same as that of the K-means. This method has no rotational invariant. On the other hand, the connector matrix depends on initial values. This point is one of the drawbacks of K-means. The regression type constrained method is one of the tandem analysis methods. The first step of the regression type constrained case is to seek the subspace maximizing the covariance. The second step is the regression step using the subspace that is obtained by the first step. Using this constraint, the prediction error is developed a little. These models assume that the data set has no categorical variable. To address this problem, we use the concept of NPCA. The NPCA-based method is one of the quantification methods for categorical data. When we choose squared loss as the loss function, the NPCA-based method maximizes the covariance between data and score. Given two three-mode three-way categorical data, the estimator of the NPCA-based method gives a better result than does the two-way method. Although these methods overcome some problems of two-mode two-way data, they have several drawbacks, as follows.

First, we do not have infallible criteria for how to decide the number of dimensions. One criteria is the information criteria, such as the Akaike information criterion (AIC)(Akaike, 1992), and Bayesian information criterion (BIC) (Schwarz, 1978). These criteria need a statistical model, which our models do not represent. While it is true that squared loss corresponds to maximum likelihood under independently normal distribution assumption, our models do not satisfy the AIC and BIC assumptions about nested models and identification. Therefore, we do not know whether the bias term is the same as the number of parameters. On the other hand, one of the criteria for a decision about the number of dimensions is scree plot. The scree plot is extended to three-mode three-way PCA by Timmerman & Kiers (2000). However, this method has been proposed for three-mode three-way PCA. Thus, we investigate whether this method could be applied to our method. Moreover, we need to extend the quantification method.

Second, the algorithms of our method are time consuming, because we use Jennrich's approach, which is iterative and needs singular value decomposition at each step. In addition, the quantification method needs an inverse matrix. Therefore, our algorithm takes much time when the numbers of objects, variables, and conditions are larger. One solution is the acceleration method for alternative least square algorithm (Kuroda et al., 2011). This acceleration method is a general method for an alternative least square algorithm. Therefore, we might apply this acceleration method to our algorithm. However, we do not investigate the fact that this algorithm does not satisfy the K-means type constrained case. Thus, we need to confirm that our method satisfies the assumption of the acceleration method.

Third, our methods are sensitive to initial values because the objective functions of the three-mode three-way method are not convex. In particular, the K-means type is very sensitive, because elements of connector matrixes take 0 or 1.One solution to this problem is that the connector matrix is fixed. For example, Tenenhaus & Tenenhaus (2011)estimated the canonical vector fixed connector matrix. Another solution to this problem is the use of a Bayesian technique. When there is prior knowledge of the connection and low-dimensional space, this knowledge is represented as prior distribution. Although our methods are based on canonical covariance analysis, we guess that the Bayesian PCA (Tipping & Bishop, 1999; Bishop, 1999) is helpful for introducing Bayesian estimation.

Fourth, it is difficult to interpret the parameters of three-mode three-way data analysis, because the number of kinds of parameters is bigger than the two-mode two-way case. One solution to this problem is sparse estimation. For example, sparse PCA (e.g., Zou et al. (2006)) estimates the sparse weight matrix. A sparse estimation dimensional reduction method for three-mode three-way data has been developed by many researchers. For example, Ikemoto & Adachi (2016) proposed a sparse estimation method for a core array of the Tucker 2 model. Tenenhaus & Tenenhaus (2011); Tenenhaus et al. (2017) proposed sparse canonical correlation analysis by using a regularization method. These methods are helpful for extending our model to the sparse estimation by regularization or constrained non-zero elements method. Moreover, the simultaneous network representation of weight matrixes for variables and conditions does not exist. Hence, we need to interpret the marginal values. For interpretation at a glance, we need to develop a representation method for simultaneous weight matrixes.

Fifth, three-mode three-way data are a special case of n-mode m-way data. In other words, we could extend our method to the n-mode m-way case. One direction for extension to the n-mode m-way is an extension to tensor decomposition (Kolda & Bader, 2009; De Lathauwer et al., 2000). However, in the context of tensor decomposition, it is often assumed that the number of modes and ways as $n = m$. Under this assumption, we cannot extend our method to an n-mode m-way method in the true sense. In the research area of multidimensional scaling, the n-mode m-way method was proposed by Carroll & Chang (1970). Tsuchida & Yadohisa (2016a) proposed n-mode m-way asymmetric multidimensional scaling. When we combine the concepts of tensor decomposition and multidimensional scaling, we can transform our method to an n-mode m-way method.

Finally, there is no guarantee that our method can estimate the true parameters. For example, we do not prove an asymptotic property, such as consistency. One future research direction for analyzing asymptotic properties is the use of the array variate random variable. Akdemir & Guota (2011) proposed a tensor normal distribution, which is a natural extension from multivariate normal distribution. Therefore, it is probably for extension theory of multivariate data to tensor data by using tensor normal distribution.

Acknowledments

I am deeply grateful to Prof. H. Yadohisa for his guidance, persistent help, and encouragement, without which this thesis would not have materialized. Prof. K. Kawasaki, Prof. K. Adachi, Prof. T. Yano, and Prof. M. Jin gave insightful comments and suggestions, which have helped to significantly improve the thesis. I thank Nomura Research Institute, Ltd. for providing the data in our application. I also thank all faculty members of the Graduate School of Culture and Information Science, and especially, members of the Fundamental Data Science Course, who give me many useful comments for study and life. Finally, I am grateful to all members of Yadohisa's Lab. for helpful discussion.

Finally, I would like to express appreciation to my family.

References

Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. In *Breakthroughs in statistics* (pp. 610–624). Springer.

Akdemir, D., & Guota, A. K. (2011). Array variate random variables with multiway kronecker delta covariance matrix structure. *Journal of Algebraic Statistics*, *2*, 98 - 113.

Bishop, C. M. (1999). Bayesian pca. In *Advances in neural information processing systems* (pp. 382–388). MIT Press.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the American Psychological Association*, 227–228.

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, *35*, 283–319.

De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, *21*, 1253–1278.

Dhillo, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning*, *42*, 143–175.

Gifi, A. (1990). *Nonlinear multivariate analysis*. John Wiley & Sons.

Greenacre, M., & Blasius, J. (2006). *Multiple correspondence analysis and related methods*. Chapman & Hall/CRC.

Harshman, R. (1970). Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*, 328–377.

Ikemoto, H., & Adachi, K. (2016). Sparse tucker2 analysis of three-way data subject to a constrained number of zero elements in a core array. *Computational Statistics & Data Analysis*, *98*, 1–18.

Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, *66*, 289–306.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, *58*, 433–451.

Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, *14*, 105–122.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*, 455–500.

Kroonenberg, P. M. (1983). *Three-mode principal component analysis.* DSWO Press.

Kuroda, M., Mori, Y., Iizuka, M., & Aakakihara, M. (2011). Acceleration of the alternating least squares algorithm for principal component analysis. *Computational Statistics & Data Analysis*, *55*, 143–153.

Lundy, M. E., Harshman, R. A., & Kruskal, J. B. (1989). A two-stage procedure incorporating good features of both trilinear and quadrilinear models. In *Multiway data analysis* (pp. 123–130). Elsevier.

Luo, Y., Tao, D., Wen, Y., Ramamohanarao, K., & Xu, C. (2015). Tensor canonical correlation analysis for multi-view dimension reduction. *IEEE Transactions on Knowledge adn Data Engineering*, *27*, 3111–3124.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297).

Nakamura, Y. (2015). Shitsuteki data ni taisuru hikeiryo sanso shuseibun bunseki ho [nonmeric three-mode principal component analysis for qualitative data]. *Kodokeiryogaku*, *42*, 105–115. (in Japanese.)

Rabusseau, G., & Kadri, H. (2016). Low-rank regression with tensor responses. In *30th conference on neural information processing systems* (pp. 1867–1875).

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, *6*, 461–464.

Sun, L., Ji, S., Yu, S., & Ye, J. (2009). On the equivalence between canonical correlation analysis and orthonormalized partial least squares. In *Ijcai* (Vol. 9, pp. 1230–1235).

ten Berge, J. M. (1993). *Least squares optimization in multivariate analysis.* DSWO Press.

Tenenhaus, A., & Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika*, *76*, 257–284.

Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. (2017). Regularized generalized canonical correlation analysis: a framework for sequential multiblock component methods. *Psychometrika*, *82*, 737–777.

Timmerman, M. E., & Kiers, H. A. L. (2000). Three-mode principal components analysis: choosing the numbers of comopnents and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology*, *53*, 1–16.

Tipping, M. E., & Bishop, C. M. (1999). Acceleration of the alternating least squares algorithm for principal component analysis. *Journal of the Royal Statistical Society. Series B*, *61*, 611–622.

Tsuchida, J., & Yadohisa, H. (2016a). Asymmetric multidimensional scaling of n-mode m-way categorical data using a log-linear model. *Behaviormetrika*, *43*, 103–138.

Tsuchida, J., & Yadohisa, H. (2016b). Connected categorical canonical covariance analysis for three-mode three-way data sets based on tucker model. *Procedia Computer Science*, *96*, 912–919.

Tsuchida, J., & Yadohisa, H. (2017). Partial least-squares method for three-mode three-way datasets based on tucker model. *Procedia Computer Science*, *114*, 234–241.

Tucker, L. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279–311.

Young, F., Takane, Y., & De Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, *43*, 279–281.

Zhao, Q., Caiafa, C. F., Mandic, D. P., Chao, Z. C., Nagasaka, Y., Fujii, N., ... Cichocki, A. (2013). Higher order partial least squares (hopls): a generalized multilinear regression method. *IEEE Transactions on Patttern Analysis and Machine Intelligence*, *35*, 1660–1673.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, *15*, 265–286.