Complex analysis for three-way asymmetric relational data

Axel Grorud* ◯ Naohito Chino** Ryozo Yoshino***
*Université de Provence **Aichi Gakuin University ***ISM

1 Introduction

A number of methods for the analysis of three-way data have been proposed, given N by m data matrices,$\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_q$, or m by m cross-product matrices, $\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_q$. Kiers (1991) considered them as variants of PCA (principal component analysis) and examined their hierarchy from an algorithmic point of view.

However, there are few who proposed efficient algorithms for the analysis of three-way data of square asymmetric matrices, $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_q$, although there are several proponents who presented models for these asymmetric three-way data (Chino, 1980; Harshman, 1978; Harshman, Green, Wind, & Lundy, 1982; Kiers, 1995; Okada & Imaizumi, 1992; Young & Lewyckyj, 1979; Zielman, 1991).

Kiers (1991) distinguished between the *direct fitting model* and the *derived fitting model* in examining the hierarchy of extant various methods for the analysis of three-way data of the forms $\boldsymbol{X}'_k s$ and $\boldsymbol{S}'_k s$ discussed above. According to his terminology, all of the asymmetric models cited above may be classified as the direct fitting model. At present, there exists no derived fitting model as far as we know.

There are a few models which can be classified with neither of them. One is DYNASCAL, proposed by Chino & Nakagawa (1990), which is a *dynamical system model* for asymmetric longitudinal relational data matrices. Another is the idea of a *stochastic process model* for latent variables to be estimated from the three-way asymmetric data under consideration (Grorud, 1994).

In this paper we propose a model for the analysis of three-way asymmetric relational data. This is a natural generalization of STATIS (Structuration des tableaux à trois indices de la statistique), therefore, we call our model *GSTATIS* (Generalized STATIS). A prototype of STATIS was originally proposed by L'Hermier des Plantes (1976), and developed further by Glaçon

(1981), Lavit (1985, 1988), and Lechevallier (1987).

GSTATIS can be considered as a derived fitting model according to Kiers terminology. It might be considered also as a stochastic process model (Grorud, 1994). At present, however, we shall consider it as a general descriptive method for the analysis of three-way asymmetric data.

2 Three steps of STATIS

Before proceeding, we shall make a brief account of the three-step analysis in STATIS because it is yet not so well-known in English speaking society, although Kiers (1991) has introduced it partly. Step 1 is called the *inter-structure analysis* which uncovers a multidimensional inter-structure among q occasions. Step 2 is called the *intra-structure analysis* which discloses a holistic multidimensional intra-structure among m objects. Step 3 is called the *trajectory analysis* which estimates the trajectory of each object through occasions in the holistic intra-structure depicted in step 2.

In step 1, STATIS computes a kind of correlation or covariance matrix of q occasions, using cross-product matrices, $\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_q$. In general, the coefficient called Rv coefficient between occasion k and occasion l is defined as

$$Rv(\boldsymbol{S}_k,\, \boldsymbol{S}_l) = \frac{tr(\boldsymbol{S}_k \boldsymbol{S}_l)}{\left\{tr(\boldsymbol{S}_k)^2 tr(\boldsymbol{S}_l)^2\right\}^{1/2}}. \quad (1)$$

Then, the Rv correlation matrix between q occasions is analyzed by PCA, and a multidimensional inter-structure among occasions is uncovered.

In step 2, STATIS constructs a weighted sum of cross-product matrices, $\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_q$, whose weights are the elements of the eigenvector corresponding to the largest eigenvalue of the Rv correlation matrix computed in Step 1. Let the eigenvector be $\boldsymbol{\alpha}=(\alpha_1, \alpha_2, \cdots, \alpha_q)^t$. Then, the weighted sum matrix $\boldsymbol{W}$ is written as

$$\boldsymbol{W} = \sum_{k=1}^{q} \alpha_k \, \boldsymbol{S}_k. \quad (2)$$

PCA of the above matrix $\boldsymbol{W}$ enables us to uncover the holistic intra-structure among objects.

In step 3, STATIS *projects* the multidimensional structure of objects at each occasion, which is obtained by PCA of each of the cross-product matrices, *onto* the holistic intra-structure among objects uncovered in step 2, to yield the trajectory of each object through occasions in a holistic structure.

Let the coordinate vector of n objects on dimension t at occasion k be $\boldsymbol{x}_{tk}$. Moreover, let the matrix $\boldsymbol{Y}_p$ be composed of the p column eigenvectors associated with the first p eigenvalues of the matrix $\boldsymbol{W}$ in step 2. Then it is easy to verify that the *orthogonal projector* $\boldsymbol{P}$ can be directly written as

$$\boldsymbol{P} = \boldsymbol{Y}_p \, (\boldsymbol{Y}_p^t \boldsymbol{Y}_p)^{-1} \boldsymbol{Y}_p^t. \quad (3)$$

Using this operator, we can define $\boldsymbol{z}_{tk}=\boldsymbol{P}\,\boldsymbol{x}_{tk}$. The vectors, $\boldsymbol{z}_{tk}$, k=1, $\cdots$, q, constitute the very trajectory discussed above.

3 GSTATIS

As is the case for STATIS, GSTATIS has three steps. A major difference between STATIS and GSTATIS is in the starting data matrices. That is, GSTATIS assumes that these are asymmetric (dis-)similarity matrices at q occasions *measured at a ratio level*, whereas STATIS assumes that we have either the N by m data matrices, $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_q$, or the m by m cross-product matrices computed from these data matrices.

A second major difference between them is that STATIS merely examines the *real* structure of objects as well as occasions, while GSTATIS deals not only with the real structure but also with the *complex* structure of objects and occasions.

The major reason for introducing complex numbers is that it not only clarifies but also simplifies metric structures of asymmetric matrices.

Complex treatment of the asymmetric relational data was originally introduced by Escoufier & Grorud (1980), and amplified by Chino and Shiraiwa (Chino, 1991, Chino & Shiraiwa, 1993). In those two-way methods, the observed asymmetric relational data matrix, $\boldsymbol{A}$, of order, say N is uniquely transformed to yield an empirical Hermitian matrix such that

$$\boldsymbol{H} = \boldsymbol{A}_s + i\,\boldsymbol{A}_{sk}, \tag{4}$$

where $\boldsymbol{A}_s$ and $\boldsymbol{A}_{sk}$ are, respectively, the symmetric part and the skew-symmetric part of $\boldsymbol{A}$. Escoufier & Grorud (1980) considered a complex uni-dimensional approximation of the matrix $\boldsymbol{H}$. This approximation is equivalent to the real two-dimensional approximation:

$$a_{ij}(s) \approx \lambda_1\,(u_{i1}u_{j1} + v_{i1}v_{j1}), \tag{5}$$

$$a_{ij}(sk) \approx \lambda_1\,(v_{i1}u_{j1} - u_{i1}v_{j1}). \tag{6}$$

This is what they call a *complex coding*. Chino (1991) considered a higher dimensional approximation and called it *HCM* (Hermitian Canonical Model).

Chino & Shiraiwa (1993) examined higher dimensional approximations of the ... and implicit *finite-dimensional complex Hilbert space structure*, if the matrix $\boldsymbol{A}$ is measured at a *ratio level* and if the matrix $\boldsymbol{H}$ constructed from $\boldsymbol{A}$ is *positive (negative) semi-definite*. Otherwise, $\boldsymbol{A}$ has an *indefinite metric structure*. This model is called *HFM* (Hermitian form model).

Thus, prior to the GSTATIS analysis we analyze each of the asymmetric relational data matrices, $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_q$, via HFM including HCM and the complex coding and examine the complex space structures contained in each of them. In HFM, *each of the complex dimensions* can be treated as if it were *a real two-dimensional space*, as can be seen in the application section. HFM is nothing but an eigenvalue-eigenvector decomposition (thus PCA) of the empirical Hermitian matrix, viewed from an algorithmic point.

Now we shall discuss the three steps of GSTATIS. In step 1, GSTATIS computes an Rv coefficient between the Hermitian matrices at occasion k and occasion l defined by

$$Rv(\boldsymbol{H}_k,\,\boldsymbol{H}_l) = \frac{tr(\boldsymbol{H}_k\boldsymbol{H}_l)}{\left\{tr(\boldsymbol{H}_k)^2 tr(\boldsymbol{H}_l)^2\right\}^{1/2}}. \tag{7}$$

It is easy to prove that the numerator of the right-hand side of eq. (7) is *real*, although each matrix is *complex*. Furthermore, it is also not difficult to prove that the coefficient ranges from -1 to 1. Apparently, the Rv coefficient defined above is a natural generalization of not only the Rv coefficient in STATIS but also the matrix correlation discussed in Ramsay, ten Berge, & Styan (1984) to a complex case.

In step 2, GSTATIS computes a weighted sum of Hermitian matrices, $\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_q$, whose weights are the elements of the eigenvector corresponding to the largest eigenvalue of the Rv coefficient matrix defined in step 1. It should be noticed that, in this case, the eigenvector is real. Let the eigenvector be $\boldsymbol{\beta}=(\beta_1, \beta_2, \cdots, \beta_q)^t$. Then, the weighted sum matrix $\boldsymbol{V}$ is written as

$$\boldsymbol{V} = \sum_{k=1}^{q} \beta_k\,\boldsymbol{H}_k. \tag{8}$$

It is well known that this type of matrix is also Hermitian. Therefore, HFM (thus, PCA of Hermitian matrix) of the above matrix $\boldsymbol{V}$ enables us to uncover a multidimensional complex intra-structure among objects, i.e., finite-dimensional complex (FDC) Hilbert space structure or an indefinite metric structure.

In step 3, GSTATIS projects the complex multidimensional structure of objects at each occasion, which is obtained by HFM of each of the Hermitian matrices, onto the holistic intra-structure among objects disclosed in step 2, to yield the trajectory of each object through occasions in the holistic structure.

Suppose that the complex coordinate vector of objects on dimension t at occasion k is $\boldsymbol{y}_{tk}$. Moreover, let the matrix $\boldsymbol{U}_p$ be composed of the p eigenvalues of the matrix $\boldsymbol{V}$ in step 2. Then it is easy to prove that the *orthogonal projector* $\boldsymbol{Q}$ can be directly written as

$$\boldsymbol{Q} = \boldsymbol{U}_p\,(\boldsymbol{U}_p^*\boldsymbol{U}_p)^{-1}\boldsymbol{U}_p^*, \qquad (9)$$

where $\boldsymbol{U}_p^*$ is the *conjugate transpose* of $\boldsymbol{U}_p$.

Using $\boldsymbol{Q}$, we can define $\boldsymbol{t}_{tk}=\boldsymbol{Q}\,\boldsymbol{y}_{tk}$. The vectors, $\boldsymbol{t}_{tk}$, $k=1,\cdots,q$, constitute the very trajectory discussed above. In this case, however, it should be noticed that this vector is not real but *complex* in general. Considering the minimum length property of orthogonal projection, it might be necessary for the holistic space uncovered in step 2 to be *FDC* Hilbert space. Otherwise, we can not necessarily define norms in that space.

4 An application

In this section we shall show one possible result of application of GSTATIS to a set of Japanese occupational mobility data, which appeared in Naoi & Seiyama (1992). This set of data is composed of four square asymmetric matrices, each of which is the cross-classification of father's and son's 8 occupational categories in each year. The four matrices correspond, respectively, to the cross-classifications of the occupational mobility in years, 1955, 1965, 1975, and 1985. For ex-

Table 1: Cross-classification of occupational status in 1955.

father/son	1	2	3	4	5	6	7	8
1.SP	42	1	19	2	4	5	1	9
2.MA	4	12	12	8	3	5	0	4
3.CL	12	1	25	10	10	3	4	14
4.SE	20	12	32	69	29	20	4	13
5.SK	7	6	27	17	88	35	18	19
6.SS	5	3	12	7	18	35	7	15
7.NS	2	2	7	5	11	13	19	9
8.AG	37	25	67	64	98	71	43	689

Prior to the GSTATIS analyses, we shall examine the asymmetric structure contained in each of the cross-classification tables via HFM. Table 2 shows all the eigenvalues of the empirical Hermitian matrices constructed from the four tables. It indicates that the 1955 data has a complete *FDC* Hilbert space structure because the eigenvalues are all positive, whereas the other three have, to be precise, indefinite metric structures. However, we can say that all of the four data have *FDC* Hilbert space structure, as far as we approximate them using the first four dimensions. For space limitations, we shall drop the configuration of objects obtained via HFM in each of the four years.

Table 2: Eigenvalues of the empirical Hermitian matrix of the four years in their order.

order/year	1955	1965	1975	1985
1	710.1	451.9	494.0	351.0
2	112.1	84.4	91.7	88.0
3	61.8	59.0	66.9	58.6
4	36.6	35.5	43.4	44.9
5	24.4	17.9	−29.4	25.1
6	14.8	10.8	29.1	−15.6
7	12.5	4.4	8.0	14.8
8	3.8	−1.8	3.3	9.2

Fig. 1 depicts one possible result of the inter-structure analysis in step 1 of

Eigenvalues of the weighted sum matrix, $\boldsymbol{V}$, were all positive in step 2. This means that the holistic space in the intra-structure analysis in step 2 has a *FDC* Hilbert space structure. Fig. 2 shows the configuration of 8 occupational categories in the second complex dimension.

For space limitations, we shall not discuss the results of step 3 in this paper.

```
              3   | Dim.3
           (1975)|
                 |
                 |2
               (1965)
                 |
-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-- Dim.2
                 |
                 |
                 |
                 |              1
                 |            (1955)
    4(1985)      |
```

Fig. 1. Inter-structure analysis among four years.

```
  Imaginary
    axis |
         |                 4
         |        5      (SE)
         |      (SK)
         |
         |
         |
         |          2(MA)
     (NS)7
-+---+---+---+---+---+---+---+ Real
     (SS)6        1     3       axis
         |      (SP)  (CL)
         |
         |
         |
         |
   (AG)  |
     8   |
```

Fig. 2. Intra-structure analysis among 8 categories.